



مركز البحوث والدراسات

ثورة البيانات

البيانات الكبيرة، والبيانات المفتوحة، والبنى التحتية للبيانات،
والنتائج المترتبة عنها

تأليف

بروفيسور: روب كيتشن

ترجمة: د. محمد بن أحمد غروي

راجع الترجمة: ماجد بن عبدالعزيز البريشن

بسم الله الرحمن الرحيم



مركز البحوث والدراسات

ثورة البيانات

البيانات الكبيرة، والبيانات المفتوحة، والبنى التحتية للبيانات،
والنتائج المترتبة عنها

تأليف

بروفيسور: روب كيتشن

ترجمة: د. محمد بن أحمد غروي

راجع الترجمة: ماجد بن عبدالعزيز البريثن

١٤٣٩هـ - ٢٠١٨م

بطاقة الفهرسة

③ معهد الإدارة العامة، ١٤٣٩هـ.

فهرسة مكتبة الملك فهد الوطنية أثناء النشر.
ثورة البيانات المفتوحة والبنى التحتية للبيانات، والنتائج
المرتبة عنها / روب كيتشن، محمد بن أحمد غروي، ماجد
بن عبدالعزيز البريثن - الرياض، ١٤٣٩هـ

٣٦٤ص؛ ١٧سم × ٢٤سم.

ردمك: ٩٩٦٠-١٤-٢٦٥-٥

١- الإنترنت أ- غروي محمد بن أحمد (مترجم) ب- البريثن،
ماجد بن عبدالعزيز (مراجع) ج- العنوان.

ديوي ٠٠٤,٦٧ ١٤٣٩/١٢٧٣

رقم الإيداع: ١٤٣٩/١٢٧٣

ردمك: ٩٩٦٠-١٤-٢٦٥-٥

هذه ترجمة لكتاب:

THE
DATA
REVOLUTION
BIG DATA, OPEN DATA INFRASTRUCTURES
& THEIR CONSEQUENCES

Rob Kitchin

حول المؤلف البروفيسور روب كيتشن

هو باحث رئيسي في المجلس الأوروبي للبحوث المتقدمة (European Research Council) في جامعة إيرلندا الوطنية في ماينوث. وقد قام بتأليف وتحرير كتب عديدة أخرى ناهزت ثلاثة وعشرين كتاباً، وكان قد حاز على الميدالية الذهبية للعلوم الاجتماعية من الأكاديمية الملكية الأيرلندية للعام ٢٠١٣. وهو الباحث الرئيسي للحاوية الرقمية الأيرلندية والمرصد البحثي لعموم الجزيرة الأيرلندية.

قائمة المحتويات

الصفحة	العنوان
٤	حول المؤلف
٥	قائمة المحتويات
٦	قائمة الجداول البيانية
٧	قائمة الأشكال الرسومية
٩	شكر وتقدير
١٣	ملاحظة
١٥	مقدمة
٢١	الفصل الأول: تصور لمفهوم البيانات
٦٣	الفصل الثاني: البيانات الصغيرة، والبنى التحتية للبيانات، ووسطاء البيانات
٩٥	الفصل الثالث: البيانات المفتوحة والبيانات المرتبطة
١٢٥	الفصل الرابع: البيانات الكبيرة
١٤٧	الفصل الخامس: مُمكّنات البيانات الكبيرة ومصادرها
١٧٩	الفصل السادس: آليات تحليل البيانات
٢٠١	الفصل السابع: المبررات الحكومية والتجارية للبيانات الكبيرة
٢٢٥	الفصل الثامن: إعادة صياغة أبحاث العلوم، والعلوم الاجتماعية، والعلوم الإنسانية
٢٥٥	الفصل التاسع: القضايا الفنية والتنظيمية
٢٨١	الفصل العاشر: المخاوف الأخلاقية، والسياسية، والاجتماعية، والقانونية
٣١١	الفصل الحادي عشر: فهم ثورة البيانات
٣٢٣	المراجع

قائمة الجداول

الصفحة	الجدول
٢٧	١-١ مستويات قياس البيانات
٣١	١-٢ المستويات الستة لبيانات نظام رصد ومراقبة الأرض الخاص بوكالة الفضاء الأمريكية
٥٩	١-٣ الأدوات والعناصر لمجاميع البيانات
٦٤	٢-١ مقارنة البيانات الصغيرة والبيانات الكبيرة
٧٥	٢-٢ أنواع وأمثلة البنى التحتية للبيانات
	٢-٣ مجموعة منتقاة من مؤسسات تقديم المشورة، والضغط لكسب التأييد، والتنسيق للمحافظة على البيانات وإتاحتها للوصول والمشاركة وإعادة الاستخدام في العلوم الاجتماعية والإنسانية
٧٦	٢-٤ فوائد مستودعات البيانات والبنى التحتية للبيانات
٨١	٣-١ الخصائص المثالية للبيانات المفتوحة المتعلقة بتعريف الانفتاح
٩٩	٣-٢ مبادئ منظمة الحوكمة المفتوحة للبيانات المفتوحة
١٠١	٣-٣ المستويات الخمسة للبيانات المفتوحة والبيانات المرتبطة
١٠٥	٣-٤ نماذج التمويل لمبادرات البيانات المفتوحة
١١٤	٤-١ مقاييس البيانات الرقمية
١٢٨	٦-١ مهام وأساليب التنقيب عن البيانات
١٨٥	٧-١ أنماط ذكاء البيانات الكبيرة للشركات
٢١٢	٧-٢ فوائد البيانات الكبيرة لعشر صناعات مختارة
٢١٦	٨-١ نماذج العلم الأربعة
٢٢٧	٩-١ الخبرات اللازمة لبناء البنى التحتية للبيانات وإجراء بحوث البيانات الكبيرة
٢٧٦	١٠-١ تصنيف الخصوصية
٢٨٧	١٠-٢ مبادئ الممارسة العادلة للمعلومات
٢٩٠	١٠-٣ أنواع المعلومات المحمية
٢٩١	١٠-٤ المبادئ السبعة الأساسية للخصوصية من خلال التصميم

قائمة الأشكال الرسومية

الصفحة	الشكل	
٣٦	هرم المعرفة	١-١
٤٨	الأسئلة المتعلقة بالأفراد خلال التعداد السكاني الإيرلندي للسكان ما بين ١٨٤١-١٩٩١	١-٢
٦١	التقاطع بين أدوات مجاميع البيانات	١-٣
١٩٠	التوزيع الجغرافي لتغريدات التخوف من المثلثية في الولايات المتحدة الأمريكية	٦-١
١٩١	المواقع اللحظية لمسار الرحلات الجوية	٦-٢
١٩٢	لوحة معلومات مدينة لندن الخاصة بالمركز البريطاني للتحليل العمراني المتقدم	٦-٣
١٩٣	مجموعة أدوات التحليلات الجغرافية السورية التي وضعها المركز الوطني للتحليلات السورية في جامعة ينكوبينج (Linköping University)	٦-٤
١٩٤	استخدام مجموعة أدوات التحليلات الجغرافية السورية في بناء الرواية القصصية الجماعية	٦-٥
٢١٥	التسويق والبيانات الكبيرة	٧-١
٢٢١	مركز العمليات لوسط مدينة ريودي جانيرو	٧-٢

^

شكر وتقدير:

بدأت نشأة هذا الكتاب في أوائل الشهر السابع (يوليو/ آيار) من العام ٢٠١٢ خلال مناقشة بمقهى في أدنبرة مع روبرت روجك (Robert Rojek) من شركة سايج للنشر (SAGE). حيث كنت قد اقترحت عليه البحث عن شخص لكتابة مصنف عن البيانات الكبيرة، والبيانات المفتوحة، والبنى التحتية للبيانات، مع تقديم الأفكار وإسداء النصح عن ذلك الشخص الذي قد يكون مناسباً لصياغة النص في مثل هذا الموضوع. وكان يرى في حينه أنني ذلك الشخص المناسب لتلك المهمة. فقررت في وقت لاحق، بعد مضي عدة أشهر، الموافقة بين خطط كتاباتي والجمع بينها، بحيث بدأت في صياغة ما بدا في متناول اليد خلال فترة زمنية قصيرة جداً ويتناول التحليل النقدي لمشهد البيانات المتغير باطراد. تطور الكتاب مع مرور الوقت إلى مخطوط كامل يسعى إلى الدراسة المستفيضة للاتجاهات والمناقشات الناشئة بتأن وإنصاف. وبالتزامن مع ذلك، بقي روبرت متابعاً متحمساً ومصداً للعديد من المواد المثيرة للاهتمام، وكانت مساعدته موضع تقدير كبير جداً. وفي سايج أيضاً، ساعد زميله كيري ديكنز (Keri Dickens) في الدفع بالكتاب إلى عملية الإنتاج والنشر، التي تم توجيهها على نحو مثير للإعجاب من قبل كاثرين هاو (Katherine Haw).

قام مارتن دودج (Martin Dodge) وتريسي ب. لاوربولت (Tracey P. Lauriault) بقراءة مفصلة ونقد هادف للمخطوط بالكامل وعلى نحو متناه في السلاسة واللفظ. وقام مارك بويل (Mark Boyle) بقراءة كامل المسودة الثانية، كما قدم غافن مكاردل (Gavin McArdle) وإيفلين روبرت (Evelyn Ruppert) نقداً مفيداً لبعض فصول الكتاب، وإضافة لذلك شارك عدد من الزملاء والأصدقاء في مداخلات نافعة ومناقشات مفيدة، حيث دلني العديد منهم إلى مواد ذات صلة، منهم مارك غراهام (Mark Graham)، تايلور شيلتون (Taylor Shelton)، مات زوك (Matt Zook)، مات ويلسون (Matt Wilson)، ليف مانوفيتش (Lev Manovich)، سيان اوكالاهان (Cian O'Callaghan)، سونغ - يوه بيرنج (Sung-Yueh Perng)، أيلين أوه كارول (Aileen O'Carroll)، جين جراي (Jane Gray)، ساندرا كولينز (Sandra Collins)، جون كيتنغ (John Keating)، شارون ويب (Sharon Webb)، جوستين

جليسون (Justin Gleeson)، أوييف داوولنغ (Aoife Dowling)، يوغان مكارثي (Eoghan McCarthy)، مارتن تشارلتون (Martin Charlton)، تيم مكارثي (Tim McCarthy)، جان ريجبي (Jan Rigby)، روب برادشو (Rob Bradshaw)، آلان مور (Alan Moore)، دراش ماك دونشا (Darach Mac Donncha)، وجيم وايت (Jim White). كما تلقت ردوداً مفيدة أيضاً خلال العروض التقديمية في جامعة دورهام، وجامعة كلارك، وجامعة هارفارد. واهتمت رونا برادشو (Rhona Bradshaw) وأورلا دان (Orla Dunne) بإدارة أعمال المكتب عندما كنت أحاول الانكباب على إجراء البحوث وكتابة الفصول. كما تفضلت جوستين جليسون (Justin Gleeson) بإنتاج بعض الرسوم البيانية. فأنا مدين لكم جميعاً بالعرفان. كما أود أيضاً شكر الكثير من الناس المشاركين على موقع التواصل الاجتماعي - التويتر وتغريداتهم التي أشارت إلى مواد مثيرة للاهتمام، وانخراطهم في مناقشات علمية دقيقة ذات صلة. وأخيراً، الشكر موصول إلى كورا (Cora) التي قامت بتثبيتتي وقدمت دعماً رائعاً لي.

وتمّ دعم جزء من الأبحاث التي أجريت في كتابة هذا الكتاب من خلال جائزة المجلس الأوروبي للبحوث المتقدمة، "المدينة المبرمجة" (ERC-2012-AdG-323636; www.nuim.ie) والتمويل المقدم من هيئة التعليم العالي في برنامج البحوث في المستوى الثالث للدورة الخامسة، والذي يهدف إلى إنشاء حاوية رقمية لأيرلندا.

ويمكن الاطلاع على نسخة من مراجع الكتاب من خلال الرابط التالي: <http://thedatarevolutionbook.wordpress.com>. كما يجري بانتظام إضافة مصادر للمعلومات وتجارب حول ثورة البيانات على الرابط <http://www.scoop.it/t/the-programmable-city>. وبالطبع، فإن التعليقات والردود مرحب بها من خلال البريد الإلكتروني Rob.Kitchin@nuim.ie أو التغريد على حساب موقع التواصل الاجتماعي - تويتر @robkitchin.

لقد سبق نشر بعض المواد الموجودة في هذا الكتاب على شكل أوراق بحثية أو على النشرات اليومية، على الرغم من أنه قد تم تحديثها، وإعادة صياغتها، والتوسع فيها. ومنها:

- Dodge, M. and Kitchin, R. (2005) 'Codes of life: identification codes and the machine-readable world', *Environment and Planning D: Society and Space*, 23(6): 851–81.
- Kitchin, R. (2013) 'Big data and human geography: opportunities, challenges and risks', *Dialogues in Human Geography*, 3(3): 262–7.
- Kitchin, R. (2014) 'The real-time city? Big data and smart urbanism', *GeoJournal* 79(1): 1–14.
- Kitchin, R. (2014) 'Big data, new epistemologies and paradigm shifts', *Big Data and Society*, 1(1) April–June, 1–12.
- Kitchin, R. and Lauriault, T. (2014) *Small Data, Data Infrastructures and Big Data. The Programmable City Working Paper 1*. Available at SSRN: <http://ssrn.com/abstract=2376148>.
- Kitchin, R. and Lauriault, T. (in press) 'Small data in an era of big data', *Geo Journal*.
- Figure 1.1 is adapted from InformationisBeautiful.net with the permission of David McCandless.
- Figure 1.2 is reproduced with the permission of The Statistical and Social Inquiry Society of Ireland.
- Table 2.4 is included with the permission of Neil Beagrie, Brian Lavoie and Matthew Woollard and under a creative commons licence for Fry et al., <http://repository.jisc.ac.uk/279/>.
- Table 3.1 is reproduced from <http://opendefinition.org/od/> under a creative commons licence.

- Table 3.3 is included with the permission of Michael Hausenblas, <http://5stardata.info/>.
- Table 4.1 is reproduced with the permission of The Economist. The Economist Newspaper Limited, London, issued March 11, 2014.
- Figure 6.1 is reproduced with the permission of Monica Stephens.
- Table 6.1 is reproduced with the permission of Taylor and Francis.
- Figure 6.2 is reproduced with the permission of Flightradar24.com.
- Figure 6.3 is reproduced with the permission of Andrew Hudson-Smith.
- Figures 6.4 and 6.5 are reproduced with the permission of Professor Mikael Jern, National Center for Visual Analytics, Linköping University, <http://ncva.itn.liu.se>.
- Table 7.1 Forms of big data corporate intelligence is included with the permission of McKinsey & Company.
- Table 7.2 and Figure 7.1 are reproduced courtesy of International Business Machines Corporation, © International Business Machines Corporation.
- Figure 7.2 is reproduced from <http://ipprio.rio.rj.gov.br/centro-de-operacoes-rio-usa-mapas-feitospelo-ipp/> under a creative commons license.
- Tables 10.2 and 10.3 are included with the permission of John Wiley & Sons.
- Table 10.4 is included with the permission of Ann Cavoukian, Ph.D., Information and Privacy Commissioner, Ontario, Canada.

ملاحظة:

يُعبّر عن مصطلح البيانات (Data) في هذا الكتاب في صيغة الجمع، مع استخدام مصطلح وحدة بيان (Datum) للدلالة على المثل المفرد (Singular Instance). بحسب ما هو موضح في قاموس أوكسفورد للغة الإنجليزية (Oxford English Dictionary- OED): في اللغة اللاتينية، البيانات هي صيغة جمع لكلمة بيان، وتاريخياً وفي المجالات العلمية المتخصصة، على الخصوص، يتم التعامل أيضاً مع مصطلح «البيانات» للدلالة على الجمع في اللغة الإنجليزية، حيث يأتي في سياق فعل الجمع (were)، كما في «تم جمع البيانات وتصنيفها».

ومع ذلك، يستخدم هذا المصطلح على نحو متزايد في صيغة المفرد في وسائل الإعلام العامة والمحادثات اليومية، كما جرى تفصيل ذلك في قاموس أوكسفورد للغة الإنجليزية:

ومع ذلك، في العموم لا يجري التعامل مع مصطلح «البيانات» بصيغة الجمع في الاستخدام غير العلمي الحديث، ولكن بدلاً من ذلك، يجري التعامل معه كاسم إجمالي - للبند غير المحدود - على غرار كلمة «معلومات»، التي تأخذ الفعل بصيغة المفرد. وأصبح من المقبول في اللغة الإنجليزية القياسية على مدى عدد من السنوات وعلى نطاق واسع الجمل من قبيل: «تم جمع البيانات» في سياق فعل المفرد «was».

لذا، يتبع هذا الكتاب الاصطلاح العلمي المعمول به. ومع ذلك، تم إبقاء النص على حاله حيثما اتفق ورود استخدامه بصيغة المفرد عند الاقتباس. ودون الجدل في أي الصيغتين هو الأصح، فإن انحياز النحاة هو لصيغة فعل الجمع، ولكن الاستخدام العام الشائع هو أكثر انفتاحاً ومرونة.

مقدمة:

هناك تاريخ طويل للحكومات، والأعمال، والعلوم، والمواطنين فيما يتعلق بإنتاج واستخدام البيانات من أجل رصد العالم، وتنظيمه، والانتفاع منه، وجعله أكثر منطقاً وذا مغزى. وقد كانت البيانات في العادة مكلفة وتستغرق وقتاً طويلاً في الإنتاج والتحليل والتفسير وتوفر في العموم صورة ثابتة، غالباً مشوشة، وجزئية عن الظاهرة قيد الدراسة. ونظراً للندرة النسبية، كانت البيانات ذات الجودة العالية سلعة ذات قيمة، وعلى هذا الأساس كان التكتّم الشديد والحرص عليها وتداولها وتبادلها بأثمان باهظة. ولكن هذا الوضع بدأ يتغير بشكل جذري تماماً. فالبيانات لم تفقد شيئاً من قيمتها، ولكن جرى في نواح أخرى تحول في طبيعتها وطرق إنتاجها من خلال مجموعة ما وصفها Christensen (1997) بالابتكارات المربكة، والتي شكلت تحدياً للوضع الراهن في كيفية إنتاج البيانات وإدارتها وتحليلها وتخزينها والاستفادة منها. فبدلاً من أن تكون البيانات نادرة ومحدودة في قدرة الوصول إليها، غدت على نحو متزايد طرق إنتاج البيانات كالطوفان؛ سيلاً عميقاً وواسعاً من البيانات المتنوعة في طبيعتها، والمرتبطة منطقياً، والمتوافرة عند الحاجة إليها في الوقت المناسب، والمنخفضة نسبياً في التكلفة، وخارج نطاق الأعمال في طبيعتها، مع تنامي فتحها وازدياد قابلية وسهولة الوصول إليها. فتورة البيانات جارية في الوقت الراهن، وهي تعمل بالفعل على إعادة تشكيل كيفية إنتاج المعرفة، وإجراء الأعمال، وحوكمة التشريعات والقوانين.

لقد تأسست هذه الثورة في الموجة الأخيرة من تقنيات المعلومات والاتصالات (ICTs)، وذلك من خلال عدد كبير من الأجهزة الرقمية التي نصادفها اليوم في المنازل، وأماكن العمل، والأماكن العامة؛ كأجهزة الاتصال المحمولة، والحوسبة الموزعة، والحوسبة السحابية؛ ووسائل التواصل الاجتماعي؛ وإنترنت الأشياء (أجهزة ووسائل استشعار مترابطة ومتصلة من خلال الشبكة العالمية - الإنترنت). فهذه الوسائط والمنصات التقنية الجديدة أضحت تقود وتوجه أكثر من أي وقت مضى المزيد من جوانب الحياة اليومية - العمل، والاستهلاك، والسفر، والاتصال، والترفيه - وأضحت العوامل التي نعيش فيها تُفهم بهيئة بيانات وتُسير من خلال التقنيات المعتمدة على البيانات. وعلاوة على ذلك، فإن هذه الوسائط والمنصات التقنية الجديدة تعمل مادياً ومنطقياً على إعادة تشكيل طرق إنتاج، وتداول، وتفسير

البيانات، وإنتاج ما اصطلح على تسميته «البيانات الكبيرة Big data»، حيث كميات هائلة من البيانات الرقمية المتغيرة والمتنوعة، والتي من السهولة بمكان الجمع بينها، وتبادلها، وتوزيعها من خلال شبكات تقنية المعلومات والاتصالات، وتحليلها بواسطة جيل جديد من أساليب تحليل البيانات (Data Analytics) المعدّة لتتلاءم مع وفرة البيانات على العكس من ندرة البيانات المعتادة. ويتضح حجم طوفان البيانات الناشئة من خلال الادعاء بأنه "من فجر الحضارة وحتى العام ٢٠٠٣، أنتجت البشرية خمسة إكسا بايت (Exabyte) من المعلومات فقط؛ في حين نحن بصدد إنشاء القدر نفسه من البيانات كل يومين" (Hal Varian, chief economist with Google, مقتبس من Smolan and Erwit 2012).

ولا تُعدّ البيانات الكبيرة المكون الوحيد لثورة البيانات، فبالأحرى كان هناك العديد من المبادرات المتصلة بعضها ببعض من قبيل التحول الرقمي (Digitization)، ربط وتوسيع مجموعات البيانات المنتجة تقليدياً (البيانات الصغيرة) مع بعضها عبر بنى تحتية شبكية؛ وحركة البيانات المفتوحة التي تسعى لجعل أكبر قدر ممكن من البيانات متوفرة بشكل مفتوح للجميع للاستخدام؛ والهياكل المؤسسية الجديدة التي تسعى لتأمين المبادئ التوجيهية والسياسات المشتركة فيما يتعلق بأنماط تنسيق البيانات، ونماذج هيكلتها، والبيانات الوصفية عنها، وحقوق الملكية الفكرية، وتراخيص الاستخدام القانوني لها، ومعايير المشاركة بها. وتُشكل هذه المبادرات مجتمعة مجموعة من مجاميع البيانات الجديدة - مزيج من نظم الفكر، وأشكال المعرفة، والتمويل، والاقتصادات السياسية، والتشريعات الحاكمة والجوانب القانونية والمادية الناظمة، والبنى التحتية، والممارسات، والمنظمات والمؤسسات، والمجتمعات المحلية والشخصيات الاعتبارية، والأماكن، والأسواق - التي تؤطر لكيفية إنتاج البيانات والحدود القصوى لاستخداماتها الممكنة.

ويظهر أثر البيانات الكبيرة، والبيانات المفتوحة، والبنى التحتية للبيانات بشكل جلي في مجالات العلوم، وقطاع الأعمال، والقطاع الحكومي، والمجتمع المدني. فالتخصصات الجديدة التي تم استحداثها بدأت الآن بالتعامل مع تدفق البيانات الهائل بعد التعود على العمل في جزر البيانات المتفرقة، والسعي نحو استخراج المعلومات، واستخلاص النتائج من عدد قليل نسبياً من المشاهدات (H.J. Miller 2010). وقد اقترنت هذه التخصصات

بمجالات جديدة، مثل علوم البيانات (Data Science)، والحوسبة الاجتماعية (Social Computing)، والعلوم الإنسانية الرقمية (Digital Humanities)، والعلوم الاجتماعية الحاسوبية (Computational Social Sciences)، والتي تعنى بشكل واضح ببناء البنى التحتية للبيانات وإيجاد طرق مبتكرة لتحليل وفهم معنى البيانات عند توسعها وكبر حجمها. ففي قطاع الأعمال، توفر البيانات الكبيرة وسائل جديدة لإدارة جميع مناحي أنشطة الشركة بصورة مرنة وفعالة ويستفاد منها لتحقيق أرباح إضافية من خلال تعزيز الإنتاجية، والقدرة التنافسية، والمعرفة بالسوق. كما أضحت البيانات نفسها سلعة هامة، يتم شراؤها وبيعها مراراً في السوق العالمية بمليارات الدولارات. وبالنسبة للحكومات، وعلى نطاق واسع، توفر البيانات الديناميكية المتغيرة رؤى جديدة حول عملياتها، فضلاً عن إعادة تشكيل وسائل الحكم وتنظيم المجتمع. فمن خلال تفحص قواعد البيانات المفتوحة، يستخلص المواطنون والمنظمات غير الحكومية (NGOs) النتائج الخاصة بهم، متحدين المنظمات التجارية الربحية والأجندات الحكومية، ويقدمون رؤى بديلة للكيفية التي ينبغي بها تنظيم المجتمع وإدارته.

لقد أثارَت هذه الفرص الجديدة طفرة حقيقية ملموسة في ما يمكن تسميته بـ "تعزيز البيانات" (Data Boosterism)؛ من خلال حشد الدعوات المبشرة بفوائد وآفاق البيانات الكبيرة، والبيانات المفتوحة، والبيانات الصغيرة الموسعة. والبعض من هذه الدعوات له ما يبرره، وبعضها الآخر مجرد غثّ وتهويل دعائي لا حقيقة له البتة. في المقابل، أصبحت مصطلحات: البيانات الكبيرة والبيانات المفتوحة، أمماتاً ثقافية قوية في الانتشار والتأثير (ميمات - Memes)، وليست مجرد وسيلة لوصف البيانات ولكنها رمز لخطاب أوسع مدى يُستخدم في حشد الدعم لها وانتشار اعتمادها والبدء في تطبيقها. فمع هذا الترويج والتعزيز أضحى من السهل الانجراف نحو تبني هذه المصطلحات دون انتقاد أو تمحيص للتغيرات التي تحدث والتي أثار العديد منها مخاوف أخلاقية، ومحاذير سياسية وقانونية. ورغم ذلك، فقد كشف التاريخ فيما مضى عن سوابق من الابتكارات المربكة المتعلقة بالمعلومات - على سبيل المثال، التحول الجذري في إنتاج المعرفة في أعقاب اختراع الطباعة. ومن حيث الواقع، فقد صاحب بداية كل عصر جديد من العلم تقنيات جديدة أدت إلى زيادة مفرطة في البيانات وكانت بمثابة شرارة التحول نحو طرق جديدة لإنتاج، وتنظيم،

وتخزين، وتحليل، وتفسير البيانات (Darnton 2000). على سبيل المثال، من ملاحظات (Strasser 2012)، فإن الاستكشافات في عصر النهضة، والتي تعززت بفعل أدوات علمية أفضل في الملاحظة، ورسم الخرائط، قد أسفرت عن كميات هائلة من الاكتشافات الجديدة التي أدت إلى وسائل جديدة للتصنيف، وتقنيات جديدة في التحليل والتخزين، وأفكار علمية جديدة.

ونظراً للمرحلة المبكرة نسبياً في ثورة البيانات الحالية، فإنه ليس من المؤكد تماماً كيف ستتكشف وتستقر التحولات الحالية، وما الذي سيكون عليه نطاق عواقب التغيرات التي تحدث من منظور أوسع. فالواضح هو أن هناك حاجة ملحة لمحاولة الإدراك والشعور بما يحدث. وهكذا، فإن الهدف من هذا الكتاب هو تقديم إجمالي، وتحليل مبدئي ودقيق للبيانات وثورة البيانات الجارية حالياً، فهذا الكتاب يسعى، من جهة، لبيان شتى الطرق والوسائل المتعلقة بإعادة تشكيل إنتاج ومعالجة وتحليل وتبادل البيانات، وما يعنيه هذا لكيفية إنتاج واستخدام المعلومات والمعرفة؛ وهو من جهة أخرى يُهدد للنقاش والتفكير النقدي حول البيانات: طبيعتها، وكيفية تأطيرها من النواحي الفنية والفلسفية والأخلاقية والاقتصادية والتجميعات التقنية والمؤسسية المحيطة بها. وبدلاً من أن تتحدد فوائد البيانات الكبيرة، والبيانات المفتوحة، والبنى التحتية للبيانات من خلال حالة عاطفية، أو نقد لاذع يشجب نتائج العواقب الأكثر سلبية لها، يقدم الكتاب تقييماً نقدياً موضوعياً للتغيرات التي تحدث.

ويستند التحليل المقدم في هذا الكتاب على مشاركة واسعة النطاق مما سبقه من مؤلفات في مختلف العلوم، والعلوم الإنسانية، والعلوم الاجتماعية، ومن مطبوعات الثقافة الشعبية، والإصدارات الصحافية، والدوريات الصناعية، وعلى خبرة مباشرة في العمل على أنظمة حفظ أرشيفية وبنى تحتية ومشاريع تحليلات لبيانات على نطاق مؤسسي واسع. وينقسم الكتاب إلى أحد عشر فصلاً. يقدم الفصل الأول لمحة عامة ونقداً فكرياً حول مفهوم البيانات وكيفية فهم قواعد البيانات والبنى التحتية للبيانات. ويبحث الفصل الثاني في الدور المستمر للبيانات الصغيرة والكيفية التي تمّ بها توسيع نطاقها إلى أرشيفات/ محفوظات رقمية وبنى تحتية، إلى أن وصلت إلى درجة التداول وبيعها من خلال وسطاء

البيانات. كما يناقش الفصل الثالث الدافع للتوجه نحو استحداث البيانات المفتوحة والبيانات المرتبطة التي يجري تقاسمها ومشاركتها وإعادة استخدامها على نطاق أوسع بمرور الوقت. ويبيّن الفصلان الرابع والخامس طبيعة البيانات الكبيرة بالتفصيل وعوامل التمكين للبيانات الكبيرة ومصادرها المساعدة على انتشارها، في حين يقدّم الفصل السادس لمحة عامة عن مجموعة جديدة من تحليلات البيانات التي ترمي إلى فهم البيانات الصغيرة الموسعة والبيانات الكبيرة. كما يدرس الفصلان التاليان (السابع والثامن) الحجج المستخدمة لتعزيز البيانات الكبيرة والترويج لها وبيان تأثيرها في الحوكمة والأعمال، والسبل التي تقوم بها ثورة البيانات على إعادة تشكيل كيفية تصوّر البحوث وممارستها. ويناقش الفصلان التاسع والعاشر التحديات التقنية والتنظيمية، والأخلاقية، والسياسية، والقانونية لثورة البيانات. ويحدد الفصل الأخير بعض الاستنتاجات الكلية ويقدم خارطة طريق لإجراء المزيد من البحث والتأصيل.

الفصل الأول

تصور لمفهوم البيانات (Conceptualising Data)

يتمّ فهم البيانات عادة على أنها المادة الخام التي تنتجها طرق تجريد العوالم من حولنا والتعبير عنها من خلال فئات، ومقاييس، وأشكال تمثيلية أخرى - كالأرقام، والحروف، والرموز، والصور، والأصوات، والموجات الكهرومغناطيسية، والثنائية الرقمية - تشكل اللبنة الأساسية في بناء صروح المعلومات والمعرفة. ومع أن البيانات تُعد بطبيعتها مُعبرة ومُمثلة للأشياء (على سبيل المثال، القياسات المتعلقة بظاهرة ما كعمر الشخص، طوله، وزنه، لونه، ضغط دمه، رأيه، عاداته، مكان إقامته، وما إلى ذلك)، إلا أنه من الممكن أن تكون أيضاً ضمنية (من خلال الغياب بدلاً من الوجود)، أو مُشتقة على سبيل المثال، البيانات التي يتم إنتاجها من البيانات الأخرى كالنسبة المئوية للتغير مع مرور الوقت والتي يتم احتسابها من خلال مقارنة البيانات من فترتين زمنيّتين مختلفتين، ويمكن أن تكون البيانات مُسجلة ومُخزنة إما في شكل تناظري (Analog) أو من خلال ترميزها في شكل رقمي كأرقام ثنائية (Bits). وتتصف البيانات بأنها ذات جودة جيدة إذا كانت منفصلة وواضحة بمعنى أن تكون كل وحدة بيان (Datum) مفردة، ومستقلة وقابلة للانفصال، ومعرفة بشكل واضح، وتراكمية قابلة للحصر (يمكن أن تبنى في مجموعات)، ومرتبطة ببيانات وصفية لها (بيانات عن البيانات)، ويمكن ربطها بمجموعات بيانات أخرى لتقديم رؤى وأفكار غير متوفرة من مجموعة بيانات واحدة (Rosenberg 2013). وتتمتع البيانات باستخدام قوي وقيمة عالية كونها توفر المدخلات الأساسية لكافة مراحل التحليل التي يجري توظيفها من قبل الأفراد، والمؤسسات، وقطاعات العلوم والأعمال لفهم وتفسير العالم الذي نعيش فيه، والتي تُستخدم بدورها لاستحداث الابتكارات، المنتجات، السياسات، والمعارف التي تشكل الكيفية التي يعيش بها الناس حياتهم.

ومن هنا، تعد البيانات مورداً رئيسياً في العصر الحديث. ومع ذلك، ونظراً لفائدتها وقيمتها، ومقدار الجهد والموارد المخصصة لإنتاجها وتحليلها، فمن اللافت للنظر كيف تم إيلاء القليل من الاهتمام للتصور المفاهيمي والإدراك المبدئي للبيانات في حدّ ذاتها، وفي

المقابل، هناك الآلاف من المقالات والكتب المكرسة لفلسفة المعلومات والمعرفة. فكما أننا نميل في العادة للتركيز على المباني والأحياء عندما ننظر إلى المَدُن، بدلاً من الطوب والإسمنت المستخدم في بنائها، فكذلك هو الحال مع البيانات. وعلاوة على ذلك، وبالمثل، كما اعتدنا التفكير بالطوب والإسمنت بصفتهما مجرد وحدات بنوية بدلاً من عناصر تم تصنيعها داخل مصانع من قبل شركات ملزمة ضمن اعتبارات لوجستية ومالية وقانونية، وأنه يتم توزيعها، وتخزينها، وتداولها، كذلك حالنا إلى حد كبير في التعامل مع البيانات. ونتيجة لذلك، وعندما تكون البيانات هي محور التساؤل والاهتمام، عادة ما يتم النظر، بالمعنى التقني إلى حد كبير، إلى الكيفية التي ينبغي بها إنتاج وتحليل البيانات، أو الكيفية التي يمكن بها الاستفادة منها في إنتاج رؤى وقيم، وليس زيادة النظر في طبيعة البيانات من منظور مفاهيمي وفلسفي.

ومع الأخذ بهذه الملاحظة في الاعتبار، فإن الهدف الرئيسي من هذا الكتاب هو مركب ثلاثي الأبعاد، فيه: توفير شرح تفصيلي عن طبيعة البيانات ومجاميعها على نطاق أوسع؛ وإيضاح الكيفية التي تتحول فيها هذه المجاميع وتتحوّل مع تطوير بني تحتية جديدة للبيانات، والبيانات المفتوحة، والبيانات الكبيرة؛ والتفكير المعمق في الآثار المترتبة عن هذه المجاميع الجديدة للبيانات فيما يتعلق بكيفية استشعار العالم المحيط بنا والتعاطي معه. ولتوفير منهاج مفاهيمي وتصور مبدئي أولي، سيتم في هذا الفصل فحص أمطاط البيانات وأطرها، وطبيعتها، والقواعد الفلسفية لها بالتفصيل. وبعيداً عن كونها وحدات بنوية بسيطة، تكشف المناقشة عن أن البيانات أكثر تعقيداً مما كنا نظن. ففي حين يقبل العديد من المحللين البيانات على ظاهرها، ويعاملونها كما لو أنها محايدة وموضوعية وقابلة مسبقاً للتخمين والتحليل في طبيعتها، إلا أن البيانات مؤطرة في الواقع من الناحية الفنية، والأخلاقية، والاقتصادية، والزمانية، والمكانية، والفلسفية. فالبيانات لا يمكن أن توجد بشكل مستقل عن الأفكار، والأدوات، والممارسات، والسياقات، والمعارف المستخدمة لإنتاجها، ومعالجتها، وتحليلها (Bowker 2005; Gitelman and Jackson 2013). ولذا، فإن الحُجّة المُقدمة هي أن فهم البيانات وثورة البيانات الجارية حالياً يتطلب تحليلاً أكثر دقة من ذلك الذي تستعرضه الكثير من الأدبيات ذات الصلة بالبيانات المفتوحة والبيانات الكبيرة في وقتنا الحاضر.

ماهية البيانات:

بشكل اشتقائي، فإن كلمة البيانات «Data» مشتقة من الكلمة اللاتينية «Dare»، التي تعني «أن تُعطى». وبهذا المعنى، فإن البيانات هي عناصر خام يمكن استخراجها من أو تُعطى بواسطة ظاهرة معينة يجري قياسها وتسجيلها بطرق مختلفة. ولكن البيانات تُشير في الاستخدام العام إلى تلك العناصر التي تُؤخذ وتُستخرج من خلال الملاحظات، والحسابات، والتجارب، وحفظ السجلات (Borgman 2007). أما من الناحية الفنية، فما نفهمه من بيانات أنها في الواقع التقاطات «Capta» والمشتقة من الكلمة اللاتينية «Capere»، والتي تعني «الأخذ»؛ أي تلك الوحدات من البيانات التي تم اختيارها وانتقاؤها من مجموع كل البيانات المحتملة (Kitchin and Dodge 2011). كما قال (1950: ix، مقتبس من Jensen (1952: 278) Becker :

كان حادثاً مؤسفاً في التاريخ أن يستخدم المصطلح بيان «Datum» بدلاً من مصطلح لقطه «Captum» للدلالة على وصف وحدة الظاهرة في العلوم. وهي ليست المناولات العلمية التي منحتها الطبيعة للعالم، ولكن المناولات التي أخذت أو انتقيت من الطبيعة من قبل العالم نفسه وفقاً لهدفه.

من ثم، وبالمعنى الدقيق للكلمة، ينبغي على الكتاب أن يُعلنون بثورة اللقطة «The Capta Revolution». ومع ذلك، ولأن مصطلح البيانات «Data» قد أصبح متأسلاً في لغة الاستخدام الأكاديمي وقطاع الأعمال لتعني اللقطة «Capta»، وبدلاً من الخلط واللغظ أكثر في هذه المسألة فمن المنطقي المتابعة والاستمرار في استخدام مصطلح البيانات «Data» مع أن اعتماد مصطلح اللقطة «Capta» أكثر ملاءمة في الاستخدام. وبعد تجاوز تسليط الضوء على جذور الاشتقاق اللغوي لهذا المصطلح، فإن هذه المناقشة الوجيزة تبدأ في تسليط الضوء على أن البيانات التي يتم جمعها من خلال القياس هي دائماً لمجموعة منتقاة من مجموع كل البيانات المحتملة - والتي اخترنا التقاطها من كل المعطيات الممكنة. وعلى هذا النحو، تكون البيانات جزئية بطبيعتها، انتقائية، وممثلة، مع أهمية المعايير المميزة المستخدمة في التقاطها.

وقد لاحظ علماء آخرون أن ما تم فهمه على أنه «البيانات» قد تغير مع مرور الزمن وتطور العلم. فقد بينَ (Rosenberg 2013) بأن مصطلح «البيانات» استخدم لأول مرة باللغة الإنجليزية في القرن السابع عشر. وارتبط كمفهوم إلى حد كبير بالحدثة والنمو والتطور المصاحب للعلوم، والطرق الجديدة في إنتاج، وعرض، ومناقشة المعرفة في القرنين السابع عشر والثامن عشر والتي انتقلت بعيداً عن اللاهوت، والوعظ، والعواطف إلى الحقائق والأدلة واختبار النظرية من خلال التجربة (Poovey 1998; Garvey 2013; Rosenberg 2013). ومع مرور الوقت، جاء فهم البيانات على أنها سابقة للتحليل، وهي ما قبل اتخاذ صفة الواقعية، وتختلف في طبيعتها عن الحقائق، والأدلة، والمعلومات، والمعرفة، ولكنها عنصر مؤسس في تكوين جميع هذه العناصر على الرغم، في الغالب، من اختلاط التعاريف والمصطلحات المستخدمة في وصف البيانات، الحقائق، الأدلة، المعرفة، والمعلومات بحسب ما أشار إليه (Rosenberg 2013: 18) حين ذكر أن:

الحقائق وجودية (Ontological)، والأدلة معرفية (Epistemology)، والبيانات بلاغية (Rhetorical). وتكون وحدة البيان «Datum» حقيقة أيضاً، كما قد تكون الحقيقة في واقع الأمر دليلاً... على أن وجود وحدة البيان «Datum» يُعد مستقلاً دون أي اعتبار للحقيقة الوجودية المقابلة (Corresponding Ontological Truth). وعند نفي الحقيقة وإثبات عدم صحتها، فإنها تتوقف عن كونها حقيقة. فالبيانات الكاذبة هي بيانات مع ذلك.

وبحسب المصطلحات/التعابير البلاغية، فالبيانات هي تلك التي توجد قبل النقاش أو التفسير الذي يحولها إلى حقائق، وأدلة، ومعلومات (Rosenberg 2013). ومن هذا المنظور، تحمل البيانات مفاهيم مميزة: فهي تكون بذلك مجردة، ومنفصلة، وتراكمية (يمكن الجمع بينها) (Rosenberg 2013)، وذات معنى مستقل عن الشكل، واللغة، والمنتج لها، والسياق (بمعنى، أن تحمل البيانات معناها سواء تم تخزينها بصورة تناظرية (Analog) أو رقمية (Digital)، أو تم عرضها على الورق أو الشاشة أو تم التعبير عنها بأية لغة، أو تم تقييدها بأتماط محددة غير متغيرة، مثل إشارة عدد الحلقات في جذوع الأشجار لعمر الشجرة والذي يكون دوماً مساوياً له) (Floridi 2010). ويؤكد (Floridi 2008) أن دعم استقلال البيانات

قائم على ثلاثة أنواع من الحيادية هي: التصنيفية (Taxonomic) كون البيانات تُمثل كيانات علائقية معرفة تحديداً بالنسبة إلى بيانات خاصة أخرى؛ والطوبوغرافية (Typological) لأنه يمكن للبيانات اتخاذ أشكال مختلفة منفصلة بعضها عن بعض، على سبيل المثال، الشكل الأساسي، الثانوي، الوصفي، التشغيلي، والمشتق؛ والجينية (Genetic) كون البيانات تمتلك دلالات مستقلة عن مدى فهمها، فعلى سبيل المثال، تشكل الهيروغليفية على حجر الرشيد بيانات بغض النظر عن حقيقة أنه عند اكتشافه لم يمكن لأحد أن يفسرها.

وفي واقع الأمر ليس كل من يفكر أو يشتغل بالبيانات يحمل هذه النظرة البلاغية الضيقة. ففهم البيانات لم يتطور مع مرور الوقت فحسب، بل اختلف هذا الفهم بحسب منظور الفهم ووجهة النظر للبيانات. على سبيل المثال، أوضح (Floridi 2008) أن البيانات من المنظور المعرفي هي مجموعات من الحقائق، ومن المنظور المعلوماتي البيانات هي معلومات، ومن منظور الحوسبة التقنية البيانات هي مجموعات من العناصر الثنائية التي يمكن معالجتها ونقلها إلكترونياً، ومن المنظور العام (Diaphoric) البيانات هي عناصر مجردة بتميز واضح وملموس من البيانات الأخرى، ففي الحالة الأولى، توفر البيانات الأساس لمزيد من الاستدلال أو تشكل الأدلة التجريبية. وفي المجموعة الثانية، تشكل البيانات معلومات تمثيلية/ممثلة يمكن تخزينها ومعالجتها وتحليلها، ولكن لا تشكل بالضرورة حقائق. أما في المجموعة الثالثة، فتشكل البيانات المدخلات والمخرجات من الحوسبة ولكن ينبغي معالجتها لتحويلها إلى حقائق ومعلومات على سبيل المثال، يحتوي القرص الرقمي المدمج (DVD) على غيغابايت من البيانات، وهي ليست وقائع أو معلومات بحد ذاتها (فلوريدي، ٢٠٠٥). في المنظور الرابع، تكون البيانات ذات مغزى لأنها التقاط ودلالة على التنوع والتغيير على سبيل المثال، أمط من النقاط، والحروف الأبجدية والأرقام، والموجات الطولية التي تقدم إشارة يمكن تفسيرها. وكما هو مبين أدناه، قد يشمل فهم البيانات مناظير أخرى مثل الفهم القائم على اعتبار البيانات بناء اجتماعياً، أو امتلاك البيانات وجوداً نسبياً، أو تحميلها مضامين إيديولوجية، أو اعتبارها سلعة يتم تداولها الاتجار بها، أو أنها تشكل منفعة عامة، وغير ذلك من المفاهيم الموجودة. والنقطة الأساسية هنا، أن البيانات لم تكن أبداً ببساطة مجرد بيانات. حيث تختلف الكيفية التي يتم فيها تصور البيانات واستخدامها من قبل أولئك الذين يجمعون، ويحللون، ويستخلصون النتائج منها.

أنواع البيانات:

سواء كانت البيانات بطبيعتها قابلة للتخمين (تم التكهّن قبل التحقق منها) ومطابقة بلاغياً أو لا، فمن الواضح أن البيانات متنوعة في خصائصها، مما يحدّد بعبارات واضحة كيفية التعامل معها وما يمكن القيام به من خلالها. وبصفة عامة، تختلف البيانات بحسب النموذج (النوعي أم الكمي)، والهيكّل (المنظم، أو شبه المنظم، أو الغير منظم)، والمصدر (مسجّلة، أم مشتقة، أم مستنفذة، أم عابرة)، والمنتج (رئيسي، أو ثانوي، أو عام)، والنوع (بيانات تأشيرية، أو بيانات خاصة، أو بيانات وصفية).

البيانات الكمية (Quantitative Data) والبيانات النوعية (Qualitative Data):

من الممكن أن تتخذ البيانات العديد من الأشكال المادية متضمناً ذلك الأرقام، والنصوص، والرموز، والصور، والصوت، والموجات الكهرومغناطيسية، أو حتى الفراغ أو الصمت (الفراغ هو في حدّ ذاته بيانات). وتنقسم هذه البيانات في العادة إلى فئتين رئيسيتين: البيانات الكمية والبيانات النوعية، فالبيانات الكمية تتكون من سجلات رقمية، وبشكل عام تكون هذه البيانات واسعة النطاق وتتعلق بالخصائص الفيزيائية للظواهر (مثل الطول، الارتفاع، المسافة، الوزن، المساحة، والحجم)، أو تكون تمثيلية وتتعلق بالخصائص غير المادية للظواهر (مثل الطبقة الاجتماعية، المستوى التعليمي، الحرمان الاجتماعي، وتصنيفات مستوى المعيشة). وللبيانات الكمية أربعة مستويات مختلفة للقياس والتي تحدد كيفية معالجتها وتحليلها (Kitchin and Tate 1999)، وانظر أيضاً الجدول 1-1. ويمكن تحليل هذه البيانات باستخدام العروض المرئية، ومجموعة متنوعة من طرق الإحصاء الوصفي والاستدلالي، وتستخدم باعتبارها مدخلات للنماذج التنبؤية ونماذج المحاكاة.

الجدول (١-١)

مستويات قياس البيانات

المثال	التعريف	مستويات القياس
غير متزوج، متزوج، مطلق، أرمل	فئوية بطبيعتها، مع تسجيل المشاهدات في وحدات قياس منفصلة.	البيانات الاسمية (Nominal data)
منخفضة، متوسطة، عالية.	ترتب المشاهدات في فئات منظمة، حيث بعض المشاهدات هي أكبر من غيرها.	البيانات النظامية الترتيبية (Ordinal data)
درجة الحرارة على المقياس المئوي (Celsius scale).	تمتد القياسات على طول مقياس يتضمن الفترة أو المسافة بين نقطة أصل ثابتة وأخرى متغيرة. ولا تتغير طبيعة فترة المشاهدات في هذا النوع من القياسات عند الإضافة أو المضاعفة بمقدار ثابت. ويمكن أن تكون قيمة البيانات متصلة (على سبيل المثال، الوقت أو الطول) أو منفصلة/متقطعة (على سبيل المثال، عدد مرات المشاهدة) في الطبيعة.	بيانات الفترة (أو المسافة) (Interval data)
علامة الاختبار على مقياس من صفر-١٠٠	مماثلة لبيانات الفترة عدا أنها ذات نقطة أصل صفرية حقيقية، كما أن المضاعفة بمقدار ثابت لن تغير الطبيعة النسبية للمشاهدة.	البيانات النسبية (Ratio data)

وفي المقابل، فإن البيانات النوعية هي بيانات غير عددية، مثل النصوص، الصور، الفن، الفيديو، الأصوات، والموسيقا. وبرغم أنه يمكن تحويل البيانات النوعية إلى بيانات كمية، إلا أن ذلك يتضمن خفصاً وتجريداً كبيرين وخسارة لثراء البيانات الأصلية نتيجة لعملية التحويل. ولذا، فإن عملية تحليل البيانات النوعية تُمارس في العادة على المواد الأصلية، سعياً لاستخلاص وبناء المعنى والفهم بدلاً من إخضاع البيانات ومجرد إمرارها لتقنيات حاسوبية. ومع ذلك، فقد تم إحراز تقدّم كبير فيما يتعلق بمعالجة وتحليل البيانات النوعية حاسوبياً من خلال عدد من التقنيات مثل التعلّم الآلي والتنقيب في البيانات (انظر الفصل السادس).

البيانات المنظمة (Structured Data)، والبيانات غير المنظمة (Unstructured Data)، والبيانات الشبه المنظمة (Semi-structured Data):

البيانات المنظمة هي تلك البيانات التي يمكن تنظيمها، وتخزينها، ونقلها بسهولة من خلال نموذج بيانات محدد، مثل قائمة الأرقام والنصوص الواردة في جدول أو قاعدة بيانات علائقية ذات تنسيق ثابت موحد (على سبيل المثال، الاسم، تاريخ الميلاد، العنوان، الجنس،... إلخ). ويمكن معالجة هذه البيانات، والبحث فيها، والاستعلام عنها، والجمع بينها، وتحليلها بشكل مباشر نسبياً باستخدام حساب التفاضل والتكامل والخوارزميات، كما يمكن تمثيلها صورياً باستخدام أشكال مختلفة من الرسومات البيانية والخرائط، ومعالجتها بسهولة من قبل أجهزة الحاسب الآلي. أما البيانات الشبه المنظمة فهي بيانات ذات تنظيم/ هيكل فضفاض وليس لها نموذج أو مخطط مسبق، ومن ثم لا يمكن تقييدها ضمن قاعدة بيانات علائقية. وتكون بنية البيانات الشبه المنظمة غير متسقة، وضمنية، ومرنة، وغالباً ما تتداخل في شبكة ذات شكل هرمي، ولكنها في ذات الوقت تتمتع بمجموعة متسقة بشكل كبير من حقول البيانات، ويتم تمييز كل منها بوسم مميز، ومن ثم يمكن فصل المحتوى بصورة دلالية وتوفير بيانات وصفية ذاتية التعريف بصورة مرنة واستخدامها وسيلة لفرز وترتيب وهيكلية البيانات. ومن الأمثلة على البيانات الشبه المنظمة صفحات الويب الموسومة بلغة التوسيف الموسعة (Extensible Markup Language- XML) وهي الصفحات التي تستخدم لغة التوسيف الموسعة (XML) لتنسيق وترميز المستندات بصورة قابلة للقراءة آلياً وبشياً في آن واحد؛ Franks 2012؛ انظر البيانات المرتبطة في الفصل الثالث).

في المقابل، لا تمتلك البيانات غير المنظمة نموذج بيانات محدداً أو هيكلًا تعريفياً مشتركاً. ولكن قد يكون لكل عنصر على حدة شكل أو هيكل محدد خاص به، مثل عناصر النص الروائي أو الصورة، ولكن ليس كل البيانات ضمن مجموعة البيانات لها الهيكل نفسه. على هذا النحو، وبينما يمكن في الغالب البحث والاستعلام عن البيانات غير المنظمة، إلا أنه ليس من السهل الجمع بينها وتحليلها باستخدام الحاسب الآلي. وتكون البيانات غير المنظمة نوعية بطبيعتها، ولكن من الممكن في أغلب الأحيان تحويلها إلى بيانات منظمة من خلال الفرز والتصنيف. وحتى وقت قريب نسبياً، كانت مجموعات البيانات الكبيرة

جداً تحوي في العموم بيانات منظمة كونها أسهل بكثير في المعالجة والتحليل والتخزين على هذه الحال. أما في عصر البيانات الكبيرة، فقد تكونت العديد من مجموعات البيانات الضخمة التي تحوي بيانات شبه منظمة أو غير منظمة، مثل التي أنتجتها المشاركات الاجتماعية الجماعية على مواقع التواصل الاجتماعي مثل الفيسبوك، التغريدات على تويتر، الصور ومقاطع الفيديو المرفوعة على الشبكة العالمية، والمدونات (Blogs)، إذ تشير بعض التوقعات إلى معدل زيادة في نمو هذه البيانات غير المنظمة أو شبه المنظمة بمقدار خمسة عشر ضعفاً عن البيانات المنظمة (Zikopoulos et al. 2012)، وقد صاحب ذلك تطوراً في تصميم قواعد البيانات مثل قواعد البيانات غير العلاقية (NoSQL) والتي لا تستخدم النماذج الجدولية أو قواعد البيانات العلائقية، انظر الجدول رقم 5، وتقنيات التعلم الآلي التي تساعد في التخزين والتحليل (انظر الفصل السادس).

البيانات الملتقطة (Captured Data)، والبيانات المستنفدة/ المستهلكة (Exhaust Data)، والبيانات العابرة (Transient Data)، والبيانات المشتقة/ المستمدة (Derived Data):

هناك طريقتان رئيسيتان لإنتاج البيانات: الأولى هي أنه من الممكن الحصول على البيانات بصورة مباشرة من خلال شكل من أشكال القياس مثل الملاحظة، والمسح الميداني، والتجارب المخبرية والميدانية، وحفظ السجلات (على سبيل المثال، ملء الاستمارات أو كتابة المدونات)، والكاميرات، والمساحات الضوئية، وأجهزة الاستشعار. وفي هذه الحالات، عادة ما تكون البيانات هي المنتج المطلوب للقياس؛ أي أن النية هي إنتاج بيانات مفيدة. وفي المقابل، يتم إنتاج البيانات المستنفدة بالأصل بواسطة جهاز أو نظام، وتكون ناتج وظيفة رئيسية وليست المخرج الرئيسي (Manyika et al. 2011). فعلى سبيل المثال، فإن جهاز تسجيل الخروج الإلكتروني المصمم لاحتساب إجمالي البضائع التي تم شراؤها ومعالجة عملية الدفع لها من الممكن أن يُنتج كذلك بيانات تستخدم لمراقبة المخزون، وأداء العاملين، وعمليات شراء العملاء. وتنتج العديد من الأنظمة المحوسبة مثل هذه البيانات المستنفدة والتي أصبح الكثير منها مصدراً قيماً للبيانات. وفي حالات أخرى، فإن البيانات المستنفدة تبقى بيانات عابرة بطبيعتها؛ أي أنه لم يسبق أن جرى فحصها أو معالجتها ولذا فإنه يجري

التخلص منها، إما لأنها ضخمة جداً أو غير منظمة بطبيعتها، أو مكلفة في المعالجة والتخزين، أو أنه يوجد نقص في تقنيات استخلاص القيمة منها، أو لقلّة الاستخدام الإستراتيجي أو التكتيكي لها (Zikopoulos et al. 2012; Franks 2012). ومن الأمثلة على ذلك ما ذكره Manyika et al. (2011: 3) من أن "مقدمي خدمات الرعاية الصحية... يتجاهلون ٩٠٪ من البيانات التي ينتجونها (كالفيديوهات المسجلة أثناء العمليات الجراحية)".

وتُعدّ البيانات الملتقطة والبيانات المستنفدة بيانات «خام» بمعنى أنه لم يتمّ تحويلها أو ضمّها مع غيرها من البيانات. وفي المقابل، يتم إنتاج البيانات المشتقة من خلال عمليات إضافية من المعالجة والتحليل للبيانات التي تمّ التقاطها. فعلى سبيل المثال، قد تكون البيانات هي تعداد حركة الأفراد على أحد التقاطعات المرورية والبيانات المشتقة منها هي التعداد الإجمالي أو التعداد لكل ساعة، فالبيانات الأخيرة استمدت من البيانات السابقة. وتكون البيانات التي جرى التقاطها في كثير من الأحيان مدخلاً إلى نموذج العملية، في حين تكون البيانات المشتقة هي المخرج عن نموذج العملية. فعلى سبيل المثال، قد تكون بيانات الكثافة المرورية مدخلاً إلى نموذج النقل والمواصلات على أن يكون المخرج هو بيانات التوقع أو المحاكاة (مثل حساب حركة المرور المتوقعة في أوقات مختلفة أو في ظل ظروف مختلفة). ففي حال وجود نموذج، من المرجح أن تضم بيانات التعداد/الكثافة المرورية مع بيانات أخرى ملتقطة أو مشتقة (مثل نوع السيارة، وعدد الركاب... إلخ) لإنشاء بيانات مشتقة جديدة قد تشكل مدخلاً للنموذج. ويتمّ إنتاج البيانات المشتقة لعدة أسباب، بما فيها الحدّ من حجم البيانات والتقليل منها إلى قدر مقبول من الممكن إدارته لإنتاج مقاييس أكثر فائدة وذات معنى. وفي بعض الأحيان، قد تتمّ معالجة البيانات الأصلية الملتقطة إلى مستويات مختلفة من الاشتقاق اعتماداً على الاستخدام المقصود. ويُعدّ نظام وكالة الفضاء الأمريكية - ناسا الخاص برصد ومراقبة الأرض من الأمثلة على ذلك حيث يُقدم بياناته في ستة مستويات تبدأ من البيانات الملتقطة وغير المعالجة، وتمضى خلال درجات متزايدة من مستويات المعالجة والتحليل وصولاً إلى مخرجات النموذج القائمة على تحليل بيانات المستوى الأقل (Borgman 2007؛ انظر جدول ١-٢).

الجدول (١-٢)

المستويات الستة لبيانات نظام رصد ومراقبة الأرض الخاص بوكالة الفضاء الأمريكية

مستوى البيانات	الوصف
المستوى	بيانات غير منظمة، وغير معالجة بالدقة الكاملة وبالحالة الخام وتحتوي كافة إضافات بيانات الاتصالات (على سبيل المثال، بيانات التزامن، ترويسة الاتصال، البيانات المكررة)
المستوى الأول أ	بيانات غير منظمة، وغير معالجة بالدقة الكاملة، ولكن معنونة زمنيًا ومذيلة أو متبوعة بشروحات إضافية، تشمل معاملات معايرة ومقاييس راديوية وجغرافية ومؤشرات تحتسب وتضاف للبيانات في المستوى ٠ ولكن لا تطبق عليها.
المستوى الأول ب	هي بيانات المستوى الأول أ ولكن تمّ معالجتها إلى وحدات الاستشعار
المستوى الثاني	متغيرات جيوفيزيائية مشتقة بنفس دقة وموقع مصدر بيانات المستوى الأول
المستوى الثالث	متغيرات تمّ تعيينها في مقاييس شبكة منتظمة للفرغ - الزمن، في العادة مع بعض المكملات والتدقيق على سلامتها
المستوى الرابع	مخرج النموذج أو النتائج من التحليل لبيانات المستويات الأقل (المتغيرات المشتقة من قياسات متعددة).

المصدر: مقتبس من الموقع الإلكتروني على العنوان :

<https://earthdata.nasa.gov/data/standards-and-references/processing-levels>

البيانات الرئيسية (Primary Data)، البيانات الثانوية (Secondary Data)، والبيانات العامة (Tertiary Data):

يتم إنتاج البيانات الرئيسية (الأساسية) من قبل الباحثين وأدواتهم من خلال تصاميم البحوث التي يضعونها. أما البيانات الثانوية فهي البيانات التي يتم إنشاؤها من قبل شخص ثم تتاح للآخرين لإعادة استخدامها وتحليلها. وعلى ذلك، يمكن أن تكون البيانات الأساسية لشخص ما هي بيانات ثانوية عند شخص آخر. أما البيانات العامة فهي شكل من أشكال البيانات المشتقة، مثل التعداد، والفئات، والنتائج الإحصائية، وتصدر البيانات العامة غالباً من قبل الوكالات الإحصائية بدلاً من البيانات الثانوية لضمان السرية للأشخاص الذين تشير إليهم هذه البيانات. على سبيل المثال، يحظر نشر البيانات الأساسية للتعداد السكاني الأيرلندي على أنها بيانات ثانوية إلا بعد مائة عام من إنتاجها؛ ولكن يتم الإفراج عن البيانات على صورة ملخصات إجمالية وبيانات تصنيفية عامة. ويسعى العديد من الباحثين والمؤسسات لإنتاج البيانات الأساسية لأنها مصممة خصيصاً لاحتياجاتهم الخاصة ومركّز عليها، في حين لا تتوفر خيارات التصميم هذه لأولئك الذين يقومون على تحليل البيانات الثانوية أو العامة. وعلاوة على ذلك، على أولئك الذين يستخدمون البيانات الثانوية والعامة كمدخلات في دراساتهم الخاصة أن يتوثقوا من صلاحية البحث الأصلي وصحته.

ويجمع الباحثون في كثير من الأحيان بين البيانات الأساسية والبيانات الثانوية والعامة لإنتاج بيانات مشتقة ذات قيمة أكبر. على سبيل المثال، قد تسعى متاجر التجزئة لإنشاء مجموعة بيانات مشتقة تدمج بيانات المبيعات الأساسية مع البيانات السكانية الجغرافية العامة (Tertiary geodemographics data)، وهي بيانات عن أي نوع من الناس الذين يعيشون في مناطق مختلفة والمستمدة من التعداد السكاني والبيانات العامة والتجارية الأخرى من أجل تحديد أماكن لاستهدافها بالمواد التسويقية. إن البيانات الثانوية والعامة ذات قيمة لأنها تمكن من إجراء الدراسات التكرارية وبناء مجموعات بيانات أكبر حجماً وأكثر ثراءً وتعقيداً. وهي تنتج لاحقاً ما أطلق عليه (Crampton et al. 2012) مصطلح تضخيم البيانات (Data Amplification) والذي يعني أن الجمع بين البيانات يتيح رؤى

أكبر بكثير من خلال الكشف عن الروابط، العلاقات، والأنماط التي تبقى خافية في حال ظلت البيانات معزولة، ونتيجة لذلك، فإن بيانات السوق الثانوية والعامه هي صناعة بمليارات الدولارات (انظر الفصل ٢).

البيانات التأشيرية (Indexical Data)، البيانات الخاصة (Attribute Data)، والبيانات الوصفية (Metadata):

تختلف البيانات أيضاً في النوع، فالبيانات التأشيرية هي تلك البيانات التي تتيح التعريف بالبيانات وربطها بغيرها من البيانات الأخرى، وتشمل معرفات فريدة من نوعها، مثل أرقام جوازات السفر وبطاقات الضمان الاجتماعي، وأرقام بطاقات الائتمان، والأرقام التسلسلية المصنعة، ومعرفات الكائن الرقمي، وعناوين الربط الشبكي (IP and MAC)، وأرقام الشحن وطلبات الشراء، إضافة إلى الأسماء، والعناوين، والرموز البريدية. إن البيانات التأشيرية مهمة لأنها تساعد في حصر كميات كبيرة من البيانات غير المفهرسة بعضها مع بعض وتتبعها من خلال معرفات مشتركة، والمساعدة على التمييز بينها، وضمها، وتقسيمها وإعادة تجميعها، والبحث فيها وغيرها من أشكال المعالجة والتحليل. وكما هو مبين في الفصل الرابع، فقد أضحت البيانات التأشيرية شائعة على نحو متزايد وبصورة متخصصة، مما يزيد من درجة الارتباط والعلائقية بين مجموعات البيانات. أما البيانات الخاصة فهي بيانات تمثل جوانب من الظاهرة، ولكنها ليست تأشيرية بطبيعتها، وعلى سبيل المثال، بالنسبة إلى شخص ما، قد تكون البيانات التأشيرية هي بصمات الأصابع أو تسلسل الحمض النووي (DNA)، وقد تكون بيانات الخاصة له مثل العمر، الجنس، الطول، الوزن، لون العينين، فصيلة الدم، وغيرها. مع ملاحظة أن الغالبية العظمى من البيانات التي يتم إنشاؤها وتخزينها في الأنظمة هي بيانات الخاصة.

البيانات الوصفية هي بيانات عن البيانات، ويمكن أن تشير البيانات الوصفية إلى محتوى البيانات أو مجموعة البيانات كاملة، وقد تشمل البيانات الوصفية عن المحتوى أسماء ووصف حقول بيانات محددة على سبيل المثال، رؤوس الأعمدة في جدول البيانات وتعريفات البيانات. وتساعد هذه البيانات الوصفية مستخدم مجموعة البيانات على فهم مكونات البيانات وتركيبها وكيف ينبغي استخدامها وتفسيرها، وتساعد في الدمج بين

مجموعات البيانات، وقابلية التشغيل البيئي وتبادل البيانات بينها، والاكتشاف، والقدرة على الحكم على مصدر البيانات وسندها. أما البيانات الوصفية التي تشير إلى مجموعة البيانات كلها فلها ثلاثة أشكال (NISO 2004). فالشكل الأول يتعلق بالبيانات الوصفية التوضيحية والتي تشير إلى بيانات التحديد/التعريف والاكتشاف المتعلقة بمجموعة البيانات، وتشمل عناصر مثل الاسم، والمؤلف، والناشر، والموضوع، والوصف. ويتضمن الشكل الثاني البيانات الوصفية الهيكلية والتي تشير إلى بيانات التنظيم والتغطية لمجموعة البيانات. ويحتوي الشكل الثالث البيانات الوصفية الإدارية التي تعكس متى وكيف تم إنشاء مجموعة البيانات، وتفاصيل الجوانب الفنية للبيانات، مثل صيغة تنسيق الملف، ومن الذي يملك وله صلاحية استخدام البيانات. ومن المعايير الشائعة للبيانات الوصفية والذي يجمع الأنواع الثلاثة للبيانات الوصفية معيار دبلن كور ([http://](http://dublincore.org)) (Dublin Core) (<http://dublincore.org>). ويشترط هذا المعيار تمتع مجموعة البيانات بخمسة عشر حقلاً للبيانات الوصفية هي: العنوان، المنشئ، الموضوع، الوصف، الناشر، المساهم، التاريخ، النوع، الصيغة، المعرف، المصدر، واللغة، العلاقة، التغطية / النطاق، والحقوق / الصلاحيات. وبرغم أن البيانات الوصفية تُعد أحد المكونات الأساسية لكافة مجموعات البيانات، إلى أنه يجري إهمالها في ممارسات دعم الإتاحة وإعادة الاستخدام والمشاركة للبيانات (Data Curation)، وخصوصاً من الباحثين الذين يجمعون البيانات الأساسية لاستخدامهم الخاص دون مشاركة الآخرين بها.

البيانات (Data)، والمعلومات (Information)، والمعرفة (Knowledge)، والحكمة (Wisdom):

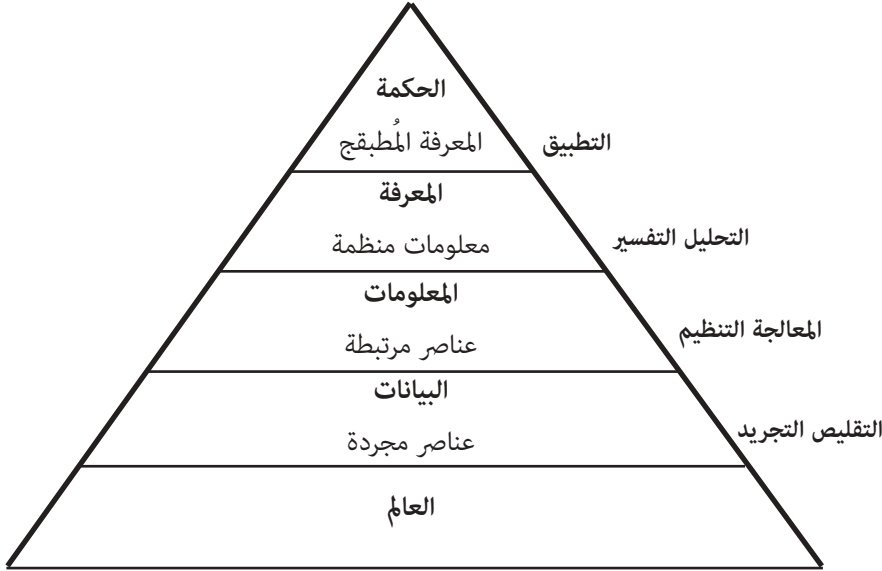
ما يوحد الأنواع المختلفة من البيانات التي تم التطرق إليها في الأجزاء السابقة هو أنها تشكل القاعدة أو الأساس الذي يقوم عليه هرم المعرفة: فالبيانات تسبق المعلومات، والمعلومات تسبق المعرفة، والمعرفة تسبق الفهم والحكمة (Adler 1986; Weinberger 2011). وتتميز كل طبقة من الهرم من خلال عملية التقطير (التقليل، التلخيص والتجريد، المعالجة والتجهيز، التنظيم، التحليل، التفسير، والتطبيق) التي تضيف التنظيم، المعنى، والقيمة من خلال الكشف عن العلاقات والحقائق حول العالم (انظر الشكل ١-١).

وفي حين أن ترتيب المفاهيم داخل الهرم هو لا جدال فيه على العموم، إلا أن طبيعة المفاهيم والاختلافات بينها كثيراً ما تختلف بين المدارس الفكرية. فالمعلومات، على سبيل المثال، هي المفهوم الذي يفهم على نحو مختلف بين العلماء، فبالنسبة للبعض، المعلومات هي تراكم البيانات المرتبطة بها، وبالنسبة للآخرين هي البيانات إضافة للمعنى، أو الإشارة ذات المعنى في ضوضاء البيانات، أو البناء متعدد الأوجه، أو البيانات العامة حيث تم إعادة صياغة البيانات الأساسية من خلال نموذج تحليلي. فبالنسبة لعالم الفيزياء، البيانات هي مجرد سلسلة من الأصفار والآحاد، أي (Bits) خام؛ وتعد ضوضاء (Noise). ولذلك فإن المعلومات بالنسبة لعالم الفيزياء تتكون عندما يتم تنظيم هذه الأعداد من الصفر والواحد في أنماط متميزة؛ فهذه هي الإشارة المفيدة (von Baeyer 2003). فموجات الأثير وكابلات الاتصالات هي ممتلئة بالمعلومات المتدفقة - كإشارات الراديو والتلفاز والمحادثات الهاتفية وحزم بيانات شبكة الانترنت - التي تعني أنماط ذات مغزى من البيانات ضمن حزمة أوسع من الضوضاء. وبالنسبة لآخرين، فالمعلومات ذات مفهوم أوسع. فقد حدّد (Floridi 2010: 74)، على سبيل المثال، ثلاثة أنواع من المعلومات :

- الواقعية (Factual): معلومات تعكس الحقيقة (على سبيل المثال، الأمطار، بصمات الأصابع، وحلقات جذوع الأشجار).
- التعليمية (Instrucional): معلومات لأجل الحقيقة (على سبيل المثال، الأوامر، الخوارزميات، والوصفات).
- الدلالية (Semantic): معلومات عن الحقيقة (على سبيل المثال، الجداول الزمنية للقطارات، الخرائط، والسير الذاتية).

شكل (١-١)

هرم المعرفة (منقول من Adler 1986 and McCandless 2010)



فالنوع الأول هو أساساً بيانات ذات معنى، وهو ما يطلق عليه عادة الحقائق. وهي البيانات التي يتم تنظيمها وتركيبها ضمن نظام قياس أو مرجع خارجي يوفر بطبيعته أساساً لإنشاء المعنى الأولي الذي يحمل بعض الحقيقة. وتمتد المعلومات أيضاً إلى ما بعد البيانات والحقائق من خلال إضافة قيمة تساعد على التفسير. وعليه، أوضح (Weinberger 2011: 2) أن: "المعلومات بالنسبة للبيانات، كالنيذ بالنسبة لحقل العنب: استخراج نواتج التقطير اللذيذة!". ولذا يمكن الحصول على هذه القيمة من خلال فرز، وتصنيف، وربط، وإضافة المحتوى الدلالي من خلال بعض الشكل النصي أو الصوري والذي يشير إلى شيء ما و/ أو يرشد إلى ما ينبغي القيام به (على سبيل المثال، الإشارة الضوئية التحذيرية على لوحة معلومات مؤشرات عمل السيارة التي تشير إلى فراغ البطارية وحاجتها لإعادة الشحن، Floridi, 2010). وقد بين (2002: ملخصاً في Case 40 Borgman 2007: أن الاختلافات في تعريف المعلومات تتوقف على خمس مسائل:

عدم اليقين (Uncertainty)، أو ما إذا كان شيء ما قد يحد من عدم اليقين في تأهيل البيانات لتكون معلومات؛ المادية (Physicality)، أو ما إذا كان شيء ينبغي أن يؤخذ على شكل مادي مثل كتاب، أو كائن، أو موجات صوتية لكلمة خطيب لتأهل البيانات لتكون معلومات؛ البنية / العملية (Structure/Process)، أو ما إذا كان يلزم مجموعة من الترتيبات أو العلاقات؛ القصد والعزيمة (Intentionality) أو ما إذا كان شخص ما قد أوعز بأن شيئاً ما تم التواصل معه ليتأهل ليكون معلومات؛ والحقيقة (Truth)، أو ما إذا كان ينبغي أن يكون الشيء صحيحاً لتأهل ليكون معلومات.

وبغض النظر عن كيفية تصور ذلك، فقد لاحظ (Floridi 2010) أنه بالنظر إلى إضافة المعلومات المعنى إلى البيانات، فهي تكتسب القيمة الربحية كسلعة، غير أنها، كنوع متميز من السلع، تمتلك ثلاثة خصائص رئيسية مميزة (والتي تتميز بها البيانات أيضاً):

غير تنافسية (Non-rivalrous): حيث من الممكن لأكثر من مستخدم أن يمتلك نفس المعلومات دون التأثير على مدى استفادة المستخدمين الآخرين منها (على عكس السلع المادية المستهلكة).

غير قابلة للاستثناء أو الإقصاء (Non-excludable): حيث يتم تقاسمها ومشاركتها بسهولة ويلزم بذل جهد كبير في السعي للحد من مثل هذه المشاركة (مثل فرض اتفاقيات حقوق الملكية الفكرية أو إضافة حواجز الدفع (Paywalls) قبل الاستخدام).

التكلفة الهامشية (Zero Marginal Cost): بمجرد إتاحة المعلومات من المرة الأولى، فإن تكلفة الاستنساخ أو إعادة الإنتاج في كثير من الأحيان لا تكاد تذكر.

ومع التشديد على خصائص كون المعلومات مورد يتميز بعدم المنازعة وعدم الاستثناء، ولأن المعلومات تعد سلعاً قيمة، تسعى العديد من الكيانات لضبطها والحد من تداولها والسيطرة عليها ومن ثم زيادة قيمتها. ومعظم هذه القيمة تضاف من خلال العمليات التي تفرض خلال دورة حياة المعلومات (Floridi 2010) والمشتتة على المراحل التالية:

الوجود (Occurrence): الاكتشاف، التصميم، التأليف.

الإرسال (Transmission): التشبيك، التوزيع، النفاذ، الاسترجاع، النقل.

المعالجة والإدارة (Processing and Management): الجمع، التحقق، التعديل، التنظيم، الفرز، التصنيف، التصفية، التحديث، التخزين.

الاستخدام (Usage): الرصد والمراقبة، النمذجة، التحليل، الشرح، التخطيط، التنبؤ، اتخاذ القرارات، التوجيه، التعليم، التعلم.

ومن خلال المعالجة والإدارة والاستخدام يتم تحويل المعلومات إلى معرفة أكثر قيمة.

وكما هو الحال مع جميع المفاهيم في هرم المعرفة، وبالمثل، يختلف فهم مفهوم المعرفة. فبالنسبة للبعض، المعرفة هي الدراية التي تحول المعلومات إلى تعليمات (Weinberger 2011: 3). فعلى سبيل المثال، يمكن ربط المعلومات الدلالية إلى وصفات وتعليمات (أولاً قم بهذا، ثم افعل ذلك..) أو نموذج شرطي لإجراءات استنتاجية (إذا كان الحال هو كذا وكذا فافعل ذلك، وإلا افعل هذا) (Floridi 2010). في هذا التأطير، تكون المعلومات هي البيانات الهيكلية، والمعرفة هي معلومات إجرائية قابلة للتنفيذ (Weinberger 2011). وبعبارة أخرى، "المعرفة هي الوصفة التي تحول المعلومات إلى الخبز القابل للأكل، بينما البيانات هي العناصر المكونة للدقيق والخميرة" (Zelany 1987، منقول عن Weinberger 2011). وبالنسبة للآخرين، فإن المعرفة هي أكثر من مجرد مجموعة من التعليمات؛ ويمكن أن تكون المهارات العملية، وأبعد بكثير عن مجرد وسيلة لمعرفة كيفية إجراء أو تحقيق المهمة، أو النظام الفكري الذي يربط بشكل متسق المعلومات مع بعضها البعض للكشف عن صورة أوسع لظاهرة ما. وينطوي إيجاد المعرفة على تطبيق العمليات المعرفية المعقدة على المعلومات مثل الإدراك، والجمع والتركيب، والاستخلاص، والدمج والتكوين، والاستنتاج والربط. وللمعرفة قيمة أكبر من قيمة المعلومات لأنها توفر أساساً لفهم، وتوضيح، ورسم رؤى حول العالم، والتي من الممكن استخدامها في صياغة السياسات والإجراءات. أما الحكمة، التي تحتل قمة هرم المعرفة، فهي القدرة على تطبيق المعرفة برجاحة وروية.

وبرغم أنه لم تتأصل بثبات علاقة كل أشكال المعرفة بالبيانات، على سبيل المثال، التخمين والآراء والمعتقدات، إلا أن البيانات تُشكل بوضوح مادة القاعدة الرئيسية لكيفية استشعار الواقع والعالم من حولنا. حيث تشكّل البيانات المدخلات الأساسية في العمليات مثل الجمع، الفرز، التصنيف، المطابقة، التشخيص / التنميط، والنمذجة التي تسعى إلى إنشاء المعلومات والمعرفة من أجل فهم الظواهر، والتنبؤ بها، وضبطها والسيطرة عليها. كما يمكننا إنتاج البيانات بمرور الوقت وفي أماكن مختلفة من تتبع، وتقييم، ومقارنة الظواهر عبر الأزمنة والأماكن والنطاقات. وهكذا، وعلى الرغم من أنه ينظر إلى المعلومات والمعرفة إلى كونها مفاهيم ذات درجة أعلى وقيمة أكبر، تبقى البيانات، مع ذلك، عنصراً أساسياً مع قيمة كامنة كبيرة يجرى إدراكها وتحصيلها عند تحويلها إلى معلومات ومعارف. فالدافع المنطقي الرئيسي لحركة البيانات المفتوحة، والتي بحثت في الفصل الثالث، هو القدرة على الوصول إلى القيمة الكامنة في مجموعات بيانات القطاعات الإدارية والعامّة.

تأطير البيانات (Framing Data):

حتى الآن في هذا الفصل، فقد بدأت بالفعل صياغة البيانات من الناحية النظرية من حيث مناقشة أنطولوجيا البيانات (ماهية وجودها)، وأشكالها المختلفة، وأين تتموضع داخل هرم المعرفة. كما أن هناك عدداً لا يحصى من الطرق الأخرى التي يمكن بها فهم البيانات والتفكير بها، على سبيل المثال، من المنظور التقني الخاص بجودة البيانات، وصحتها، وموثوقيتها، ونسبتها، وقابليتها للاستخدام، وكيف يمكن معالجتها وتنظيمها، وتقاسمها، وتحليلها؛ أو من المنظور الأخلاقي المتعلق بالأسباب التي دفعت إلى إنتاج البيانات والاستخدامات التي يتم توظيفها بها؛ أو من المنظور السياسي أو الاقتصادي والذي يأخذ بعين الاعتبار التصور المعياري للبيانات ومواقع الطعن والتنازع والتنافس فيها كمنفعة عامة، ورأس مال سياسي، وملكية فكرية، أو سلعة تجارية، وكيف يتم تنظيمها وتداولها؛ أو من المنظور المكاني و/ أو الزماني المتعلق بكيفية إنتاج البيانات وتطوير استخداماتها ومحاوَر تركيزها من قبل الأنظمة والمجموعات التقنية، والأخلاقية، والسياسية، والاقتصادية عبر المكان والزمان؛ أو من المنظور الفلسفي الذي يأخذ بعين الاعتبار المفاهيم المبدئية والمعرفية للبيانات. وسنرجع لاستذكار العديد من القضايا، التي تمّ بحثها في هذا القسم، لاحقاً في جميع أجزاء الكتاب المقبلة.

المنظور الفني (Technically):

في جميع التخصصات، يتم أخذ البيانات بعين الاعتبار من وجهة النظر التقنية المعيارية. والسؤال المطروح على المحك هو إلى أي مدى تنتج طرق التقاط البيانات وقياسها بيانات محددة، ونظيفة، ودقيقة، وكيف يمكن وينبغي معالجة، وتنظيم، ومشاركة، وتحليل هذه البيانات بطرق تحافظ على سلامتها، من أجل ضمان موثوقيتها وصلاحية الاستنتاجات المستخلصة منها. ولذا كان هنالك دائماً شكوك فيما يتعلق بصحة البيانات لأنها بطبيعتها مجردة، وعمومية، وتقريبية عند إنتاجها (Goodchild 2009). وحيث ذلك، فقد توجه الكثير من الاهتمام إلى قضايا تتعلق بمفاهيم تمثيل البيانات (Data representativeness)، وعدم اليقين بصحتها (Uncertainty)، ودرجة الموثوقية بها (Reliability)، والأخطاء الممكنة المصاحبة لها (Error)، والتحيز المنتقص لمصادقية دلالاتها (Bias)، والمعايرة (Calibration) في تصميم البحوث وتطبيقاتها، وتسجيل هذه المعلومات كبيانات وصفية. وبالنظر للبيانات على أنها علامة بديلة عن بعض جوانب ظاهرة ما - كالضوء يمثل نجماً، والخصائص الفيزيائية التي تمثل مبنى، والكلمات التي تمثل أفكار شخص - يتعلق التمثيل (Representativeness) بمدى التقاط البيانات لهذه الظاهرة التي تسعى إلى تمثيلها، وإلى أي مدى تمثل عينة البيانات التي تم إنشاؤها المجتمع الكامل للظاهرة. وفيما يتعلق بالسؤال السابق، فإن السؤال الرئيسي هو إلى أي مدى يمكن أن نكون على ثقة من أن التقنيات العلمية قد التقطت بدقة الظاهرة قيد الدراسة. لقد مثل هذا التساؤل مشكلة خاصة في مجال العلوم الاجتماعية والإنسانية وأثبت أنه كان من الصعب حلها. فعلى سبيل المثال، لوحظ بشكل كبير أن ما يقول الناس أنهم سيفعلون وما يفعلونه في الغالب ما يكون مختلفاً تماماً عن الواقع، وما يفعله الناس في كثير من الأحيان قد لا يقصدونه. ولذلك برز سؤال حول المدى الذي تمثله بيانات المقابلة الشخصية للسلوك البشري، أو مدى تمثيل الفكر الواعي لهم. وبالمثل، هناك مخاوف بشأن مدى التقاط وتمثيل المؤشرات الرئيسية للتعبير بدرجة كافية عن كيفية أداء مجال ما. على سبيل المثال، إلى أي مدى تُعطي المؤشرات المتعلقة بعدد مرات الاقتباس، ومؤشر إتش (H-index)، وبراءات الاختراع المسجلة دلالة على أداء عالي الجودة من قبل موظفي الجامعة (مع الاحترام لأعضاء هيئة

التدريس في مجال العلوم الإنسانية فإن هذه المؤشرات تُعد سيئة للغاية؟ ولذا كان الحل في محاولة تطوير تصاميم بحثية أكثر تعقيداً لمواجهة أوجه القصور في هذه الأساليب المختلفة، أو تجاهل أوجه القصور المتعلقة بها إلى حد كبير.

وبالنسبة لما يتعلق بمدى تمثيل عينة لمجتمع دراسة ما، فقد نقرر، اعتماداً على خمسين نجمة، إنتاج بيانات فلكية هائلة، ومفصلة، وعلى مدى طويل من أجل فهم أفضل لطبيعتها. ولكن إلى أي مدى يمكن أن نكون على ثقة بأن هذه الخمسين نجمة تمثل مجمل خصائص المليارات الأخرى من النجوم الموجودة؟ وحتى في عصر البيانات الكبيرة، التي تسعى إلى أن تكون شاملة وليست انتقائية في إنتاج البيانات (انظر الفصل الرابع)، فإن البيانات هي بطبيعتها عينة (فليس كل الناس يستخدمون وسائل التواصل الاجتماعي، أو يتسوقون باستخدام البطاقات الائتمانية، وبالتأكيد لا يمتلك كثير من الناس في جميع أنحاء العالم حق الوصول إلى الهواتف وأجهزة الحاسب)، وهذا يعني أن البيانات هي ممثلة لمجموعة من الناس، حتى لو كانت هذه المجموعة كبيرة جداً. ومرة أخرى، كان الحل هو ابتكار مجموعة من تقنيات أخذ العينات التي تسعى إلى ضمان التمثيل الدقيق في ظل مختلف الظروف (والتي تعتمد غالباً على عينة عشوائية)، واستحداث أساليب إحصائية تقوم بحساب مدى الثقة التي يمكننا أن نكون عليها حيال تمثيل عينة ما لمجتمع الدراسة (Kitchin and Tate 1999).

وتتعلق الموثوقية بالتكرارية أو الاتساق في الحصول على النتيجة نفسها من توظيف آلية البحث. وقد وصف (Golledge and Stimson 1997) ثلاثة أنواع من الموثوقية: (١) الموثوقية التَّصَوُّريَّة (quixotic reliability)، حيث تؤدي طريقة واحدة للملاحظة والاستكشاف إلى قياس غير متغير باستمرار. (٢) الموثوقية الزمنية المتعاقبة (Diachronic Reliability) وفيها يكون استقرار الملاحظات عبر الزمن. (٣) الموثوقية التزامنية (Synchronic Reliability)، وفيها يكون التشابه في الملاحظات خلال الفترة الزمنية نفسها. وتُعد الموثوقية مهمة لأنه من المسلم به أنه كلما كان القياس أكثر اتساقاً في إنتاج البيانات، أمكن بشكل أكبر وصف هذه البيانات بأنها بيانات موثوقة.

والخطأ هو الفرق بين القيمة المقاسة والقيمة الحقيقية، ويمكن أن يشمل الغيابات البيانات المفقودة، والأخطاء مثل الخطأ في الإدخال / الترميز أو الخطأ في التصنيف أو سوء تطبيق الأسلوب التقني، وسوء الفهم. أما التحيز فهو نوع معين من الخطأ، حيث تنحرف البيانات بسبب نمط ثابت من الخطأ. وعادة ما يحدث التحيز بسبب الطريقة، أو الأداة، أو أسلوب أخذ العينات المستخدم لإنشاء البيانات والتي لها تأثير مفرط على البيانات التي يتم إنتاجها، أو يمكن تقديمها بسبب موقف إيديولوجي أو تطلعات الباحث في كثير من الأحيان بطريقة غير متعمدة (Kitchin 1996). ومن الممكن أن تتسبب آليات المعالجة مثل التجميع أو الإجمال في إحداث تحيز من خلال الحد من التباين في مجموعة البيانات مما يؤدي إلى أخطاء المغالطة المنطقية للاستدلال البيئي (Ecological Fallacy Errors)، بمعنى أنه على افتراض تمثيل القيمة الإجمالية لمجموع الأفراد بدقة على سبيل المثال إذا كان لدينا شخصين وزنهما ٥٠ كيلوجراماً وآخرين وزنهما ١٥٠ كيلوجراماً فسيكون متوسط الوزن الإجمالي هو ١٠٠ كيلوجرام، ولكن لا أحد في المجموعة يزن هذا الوزن في الواقع (Kitchin and Fotheringham 1997). ويتعلق مفهوم عدم اليقين بمدى ثقة الباحث بخصوص دقة البيانات وأي تحليل قائم على أساسها. أما فيما يتعلق بالبيانات الكمية، فيتصل بدرجة اليقين بالاختبار الإحصائي وذلك بالنظر إلى البيانات المدخلة، وعادة تُحتسب على أساس الاحتمالات ويُعبّر عنها بمستويات الثقة (Goodchild 2009). أما مفهوم عدم اليقين بالنسبة للبيانات النوعية، فيجري في الغالب تقيمه والحكم عليه من قبل خبير بناء على الخبرة السابقة.

وتتمثل دعائم معالجة هذه المخاوف بالإيمان بأن مثل هذه القضايا تنشأ بسبب الضعف البشري في تصميم البحث أو أوجه القصور وعدم كفاءة الأدوات والمعدات أو الطرق المستخدمة والتي من الممكن التغلب عليها من خلال الحلول التقنية. وهذا يعني أنه من الممكن معالجتها عن طريق تحسين نوعية الإجراءات والمعدات المستخدمة، وتطبيق أنظمة التوحيد القياسي (التقييس - Standardization) والتي توفر معايير جودة البيانات المعروفة مثل تلك التي أقرتها المنظمة الدولية للمعايير آيزو (ISO)، والتي تعمل على إيجاد سبل للتعويض عن عدم اليقين والخطأ والتحيز في وسائل التحليل المستخدمة.

المنظور الأخلاقي (Ethically):

الأخلاق معنيّة بالفكر والممارسة المتعلقة بمفاهيم القيم مثل العدالة، والمساواة، والنزاهة، والصدق، والاحترام، والحقوق، والواجبات، والرعاية. ويعمل كل مجتمع وفق مزيج من القيم الأخلاقية الفطرية البديهية والعامة غير الرسمية، المسلم بها، والمواقف الأخلاقية المقننة جيداً والمنصوص عليها في قواعد، ومبادئ، وسياسات، وتراخيص، وقوانين تخضع للإنفاذ من قبل أجهزة الدولة والوكالات الأخرى. وكثيراً ما تكون هذه المواقف الأخلاقية محل نزاع بين مجموعات مختلفة تأخذ وجهات نظر متناقضة عن القيم نفسها وإلى أي مدى ينبغي إخضاع المواقف الأخلاقية للتشريع القانوني، ومناقشتها كممارسة في الفلسفة الأخلاقية. كما يوجد مثل هذا الطعن فيما يتعلق بالبيانات، وخاصة ما يتعلق بما يتم إنشاؤه من بيانات ووسائل الإنتاج لها، وكيفية مشاركة البيانات وتقاسمها، وتداولها، وحمايتها، وإلى الحدود التي ينتهي عندها العمل بها.

ففي حين، تعد بعض البيانات معتدلة نسبياً، على سبيل المثال القياسات المتعلقة بالطقس، فإن بعض البيانات الأخرى تعد حساسة للغاية، كالبيانات المتعلقة بالأفراد والتي يمكن استخدامها لإنتاج صورة مفصلة عن الحياة المعيشية لهم، وضبط وتنظيم تلك الحياة على إيقاع معين. وقد يلحق إنتاج البيانات ضرراً في بعض الحالات، على سبيل المثال الأضرار التي قد تنتج عن إجراء مقابلات مع ناجين من جرائم الحرب وما قد تتسبب فيه تلك المقابلات من ضغوطات نفسية وضيق. وفي هذا المقام، هناك تساؤلات حول المدى الذي من الممكن أن يشكل فيه إنتاج البيانات، ومختلف أشكال المراقبة والترصد للبيانات (Dataveillance) أي المراقبة والرصد من خلال معالجة وتحليل سجلات البيانات، وتحليل البيانات تعدياً على الخصوصية وحقوق الإنسان الأخرى، وتساؤلات أيضاً حول المدى الذي يمكن أن تستخدم فيه هذه البيانات بفعالية لفرز الأفراد اجتماعياً بمعنى تقديم خدمات تفضيلية على أساس خصائص الأفراد المميزة لهم (Graham 2005). وتتعاظم هذه المخاوف بالنظر إلى سهولة جمع ومشاركة وتداول البيانات الرقمية، ونحن نعيش في عصر توليد البيانات ومراقبتها بصورة طاغية وعلى نطاق واسع. ولعله ليس من المستغرب بعد ذلك أن وكالات تمويل البحث العلمي ومؤسسات التعليم العالي الآن تقيّم بشكل دوري

الأبعاد الأخلاقية للمشاريع البحثية نظراً لآثارها المحتملة على نطاق أوسع، وإضافة لذلك فقد سنت الدول تشريعات، مثل قوانين حماية البيانات وقوانين الخصوصية، في محاولة لمنع سوء الاستخدام والتجاوزات المسيئة للبيانات، ولقد تمّ مناقشة هذه المسائل وما يتصل بها بشكل مفصل في الفصل العاشر.

المنظور السياسي والاقتصادي (Politically and Economically):

مثل الاهتمام بالمنظور الأخلاقي للبيانات البداية للكشف عن الطرق التي تعمل على تأطير البيانات من خلال الاهتمامات السياسية والاقتصادية على نطاق أوسع. وتتعلق هذه الاهتمامات بماهية البيانات التي يتم إنشاؤها، وكيفية معالجتها، وتحليلها، وتوظيفها بالنسبة إلى: كيفية تصوّر البيانات وتوصيفها معيارياً ضمن المجموعات السكانية وتوظيفها من قبل الدول، ومفاهيم الكيفية التي ينبغي بها تنظيم البيانات وتقنينها؛ والمناقشات التي تطرح ضمن مجموعات الاختصاص التي تؤيد أو تعارض إنتاج البيانات وتطبيقاتها؛ واتخاذ القرار حول التمويل والاستثمار في البيانات؛ وانتشار الرأسمالية والطرق التي تستخدم البيانات لإدارة المساعي وزيادة القيمة والأرباح؛ والتداول للبيانات كسلعة مع ظهور سوق للبيانات تُقدر قيمته بمليارات الدولارات والتي يشارك فيها مجموعة متنوعة من اللاعبين كالمنتجين، المجمعين، البائعين، المحللين، والمستهلكين، انظر الفصل الثاني. فمنتجو البيانات ينبغي عليهم الأخذ بعين الاعتبار الرأي العام والسياسي، والاعتبارات الأخلاقية والبيئية التنظيمية والتمويل المتاحة وسلامة استثماراتهم إزاء الموارد. كما ينبغي على أولئك القائمين على الساحتين التشريعية والتمويل المالي أن يكون لديهم سعة الأفق واتخاذ القرارات حول كيفية تشكيل المشهد الذي يعمل به منتجو ومستخدمو البيانات، وكذلك النظر في أنظمة العمل المتعلقة بالبيانات الخاصة بهم وما يكشفونه عن جداول أعمالهم وأولوياتهم وأساليب الإدارة والحوكمة الخاصة بهم (Lauriault 2012).

وفي كلتا الحالتين، هناك مجموعة متنوعة من العلاقات السياسية والاقتصادية العقلانية في المشهد، مع أصوات متنافسة تسعى للتأثير على الرأي ومشهد البيانات على النطاق الأوسع. فعلى سبيل المثال، فإن حركة البيانات المفتوحة تمثل البيانات كسلعة عامة تتشكّل من العموم وتكون في متناول الجميع بحرية مطلقة باستثناء البيانات الحساسة أمنياً،

والشخصية الخاصة أو يمكن الوصول إليها من خلال اتفاقيات الاستخدام العادل. وفي المقابل، يعرض قطاع الأعمال البيانات كسلعة تجارية قيمة تحتاج من جهة إلى الحماية من خلال أنظمة الملكية الفكرية (حقوق الطبع والنشر، براءات الاختراع، وحقوق الملكية) ولكن، من جهة أخرى لا ينبغي ربطها بشدة بالمحاذير الأخلاقية التي تمنع من استغلالها لتحقيق المكاسب المالية. أما بالنسبة للمجتمعات والدول، فالبيانات هي الوسائل التي يمكن من خلالها تقنين، وممارسة، ومنازعة الأجندات السياسية والأعمال عبر تمكين بناء الإثباتات - أي الروايات المسندة والدعوات المضادة التي لها قيمة بلاغية أكبر من مجرد التندر أو دغدغة المشاعر (Wilson 2011; Garvey 2013). وبعبارة أخرى، تعد البيانات بالنسبة لعالم الفلسفة الفرنسي (Foucault's (1981 شكلاً من أشكال السلطة / المعرفة؛ أي وسيلة يمكن من خلالها ضبط وتنظيم الشعوب، والظواهر، والأقاليم (Lauriault 2012). وتتقاطع هذه الاهتمامات والمصالح البديلة في كثير من الأحيان بطرق متناقضة، من حيث أنه قد يكون لها أجندات مختلفة تماماً، على سبيل المثال دعم الأعمال التجارية الكبيرة لحركة البيانات المفتوحة فيما يتعلق بالبيانات العامة (انظر الفصل الثالث). وبعبارة أخرى، تظهر البيانات جلياً وتتموضع ضمن الاقتصادات السياسية المعقدة، وفي الوقت نفسه فهي تستخدم لتشكيل مثل هذه القضايا والنظم والمجموعات.

وعلاوة على ذلك، تشكل البيانات مورداً اقتصادياً، وهي عنصر أساسي في المرحلة المقبلة من الاقتصاد القائم على المعرفة، وتعمل على إعادة تشكيل نمط الإنتاج القائم على البيانات (انظر الفصل السابع). ومنذ أواخر الثمانينيات (1980s)، يرى بعض العلماء مثل Castells (1988, 1996) أن آخر دورة للرأسمالية يدعمها إنتاج المعرفة التي تنشئ منتجات وأشكال عمالة جديدة، وتسهل إعادة الهيكلة الاقتصادية، وتعزز الإنتاجية والقدرة التنافسية، والكفاءة، والاستدامة، وتراكم رأس المال. وتعد البيانات الكبيرة، على وجه الخصوص، هي أحدث التطورات المتعلقة بتعميق وتعزيز هذه الدورة، وتوفير ثورة من الأدلة التي يتم استخدامها من قبل الشركات من جهة، لرصد وتقييم أداء الشركة اللحظي، والحد من الهدر والاحتيايل، وتحسين إستراتيجية الشركات، والتخطيط وصنع القرار، ومن جهة أخرى، لتصميم سلع جديدة، وتحديد واستهداف أسواق جديدة، وتطبيق التسعير المتغير، وإدراك الإمكانيات غير المستغلة، واكتساب ميزة تنافسية (Manyika et al. 2011; Zikopoulos et al. 2012). وفي سبيل

القيام بذلك، يمكن إنتاج وتحليل البيانات من تشغيل الشركات بشكل أكثر ذكاء فيما يتعلق بكيفية تنظيمها وتشغيلها، وتعزيز المرونة والابتكار، والحد من المخاطر والتكاليف والخسائر التشغيلية، وتحسين تجربة العملاء، وتعظيم العائد من الاستثمار والأرباح. وكقائد لعملية تراكم رأس المال، ستسهم البيانات الكبيرة في إيجاد أعمال (وظائف) جديدة وفي الجولة المقبلة من التنمية غير المتوازنة، ومن هذا المنظور يمكن فهم البيانات على أنها وكيل لمصالح رأس المال واهتماماته.

المنظور الزماني والمكاني (Temporally and Spatially):

تمتلك البيانات أبعاداً زمانية ومكانية على حدٍ سواء. حيث تختلف ماهية البيانات التي يتم إنتاجها والطرق التي يتم بها معالجتها، وتحليلها، وتخزينها، أو التخلص منها باختلاف الزمان والمكان؛ أي أن للبيانات ومجاميعها المحيطة بها دلالات تاريخية وجغرافية. كما تتغير كيفية معالجة البيانات وتحليلها مع مرور الوقت، متأثرة بالتغيرات والتحسينات التنظيمية في التعداد والإدارة، والقوانين الجديدة المتعلقة بتداول البيانات وحمايتها، والتقنيات الجديدة، والطرق الجديدة في فرز البيانات وتحليلها، والاختلافات الإحصائية الجغرافية مثل حدود المقاطعة المحلية أو الوطنية الجديدة، والأساليب الإحصائية الجديدة. وعلاوة على ذلك، يمكن لمجاميع البيانات التي تستخدم في ولاية معينة أن تكون مختلفة تماماً عن تلك التي تستخدم في أخرى، وحتى داخل حدود الولاية الواحدة، فقد تختلف الطريقة التي يقوم فيها كيان ما بإنتاج البيانات وإدارتها بسبب التقنيات المؤسسية أو الشخصية.

وبالنظر في التعدادات السكانية، يتكوّن التعداد من مسح شامل لمنطقة وساكينها، ويتم إجراء المسح في العادة كل عشر سنوات، والهدف من ذلك هو جمع معلومات أساسية حول المقيمين في منطقة ما وخصائصهم على سبيل المثال، العمر، الجنس، الحالة الاجتماعية، التركيبة الأسرية، الدين، العرق، الطبقة الاجتماعية، وما إلى ذلك وجوانب عيشهم فيها (عملهم، ومكان الإقامة، وما إلى ذلك). ويتطلب تمكين قياس التغيير في التعدادات السكانية الاستمرارية فيما يتعلق بالأسئلة المطروحة فيها والكيفية التي تدار بها هذه التعدادات. وفي ذات الوقت، يتطلب التقاط البيانات الجديدة موضع الاهتمام

والتي تعكس تغيرات أوسع في المجتمع، إجراء تغييرات على آلية هذه التعدادات، مثل إضافة أسئلة جديدة أو التعديل على الأسئلة الأساسية (انظر الشكل ٢-١: ولاحظ كيف أنه حتى عندما تمّ الحفاظ على الأسئلة عبر التعدادات السكانية المتعاقبة، فإن كيفية صياغتها كانت غالباً ما تكون مختلفة تماماً في كل مرة). وعلاوة على ذلك، فإن الكيفية التي يدار بها التعداد السكاني كانت تتشكل في التعدادات المتعاقبة من قبل قوى مؤسسية، وسياسية، واقتصادية، وتطورات تقنية جديدة : انظر (Linehan 1991) للتعرف على تاريخ التعداد السكاني الايرلندي من ١٨٢١ - ١٩٩١، و(2012 Lauriault) لتحليل التعداد الكندي من ١٨٧١ - ٢٠١١. علاوة على ذلك، يتعرض التعداد للطعن والتفاوض حيث تتنافس مصالح خاصة لضمّ، أو تغيير، أو إزالة الأسئلة منها. وفي بعض الحالات، يمكن أن تكون التغييرات المطلوبة جذرية بشكل متطرف، مثل القرار الألماني بوقف إجراء التعداد السكاني في الثمانينات (انظر Hannah 2011). ونتيجة لذلك، يركز التعداد الوطني دائماً تحت وطأة التوتر الناتج عن الاستمرارية والتغيير، ومع ذلك، يتطور مع مرور الوقت وفي مناطق جغرافية مختلفة. وحتى الآن، ومع ذلك، فقد كان هناك عدد قليل من مجاميع البيانات المميزة تاريخياً وجغرافياً (Alder 2002; Desrosières 1988; Hannah 2011; Hewitt 2010; Lauriault 2012; Poovey 1998; Porter 1995).

المنظور الفلسفي (Philosophically):

بالنسبة للبعض، تعد البيانات على المستوى الوجودي (Ontological level) حميدة (غير خطيرة). فالبيانات هي مجرد بيانات، وعناصر أساسية يتم استخراجها من العالم بطرق محايدة وموضوعية تخضع لقيود تقنية. فهي "لا تمتلك أي معانٍ ضمنية كامنّة فيها، ولا تقدّم بالضرورة أي تفسيرات أو آراء، ولا تحمل خصائص دلالية قد تكشف عن أهميتها وصلاتها" (Pérez- Montoro and Díaz Nafría 2010)، فهي تخمينية قابلة للتحليل. ومن هذا المنظور، فإن جهاز الاستشعار (sensor) على سبيل المثال ليس لديه أي سياسة أو جدول أعمال. فهو ببساطة يقيس الضوء أو الحرارة أو الرطوبة،... الخ - حيث يقوم بإنتاج القراءات التي تعكس حقيقة الواقع عن العالم المحيط ما لم يشوبه خلل فني. وبعبارة أخرى، تُنتج أجهزة الاستشعار نظرة موضوعية واقعية عن العالم وتكشف عن الأشياء كما هي في الواقع، حيث إن قياس واقعية الشيء هو مستقل عن عملية القياس له (Desrosières 1998). أما في إطار عمليات القياس التي يلعب فيها العنصر البشري دوراً مركزياً - في التجارب المخبرية، أو إجراء مسح ميداني أو مقابلة شخصية - فإن عمليات القياس تطبّق شكلاً من أشكال الموضوعية الميكانيكية التي تلتزم بقواعد محدّدة، وطرق منهجية دقيقة لإنتاج بيانات منفصلة، غير متصلة، ونزيهة على نحو شفاف، وخالية من تحيز الباحث وهواه وتفضيلاته، ومن ثم فهي مستقلة عن العادات، والتقاليد، والثقافة، والمعرفة، والسياق (Porter 1995). وعلى هذا النحو، فإن العلم عند ممارسته بشكل صحيح لا ينبغي أن يتلبس بأي دوافع سياسية أو أجندات خفية بحيث يمكن بعد ذلك أن تؤخذ البيانات على ظاهرها وعلاقتها. وفي الواقع، تقترح المصطلحات المستخدمة عادة لتفصيل كيفية التعامل مع البيانات عمليات تقنية حميدة مثل: «الجمع»، «الإدخال»، «الترجمة»، «التخزين»، «المعالجة»، و«التنقيب» (Gitelman and Jackson 2013). ولذا فإن استخدامات البيانات هي التي تخضع للإملاءات السياسية. وبعبارة أخرى، إن الناس هم من يفسدون البيانات ويعملون على تجريدها لمصالحهم الخاصة، وليس للعلم بحدّ ذاته.

وبالنسبة لآخرين، فإن هذا الرأي لا يمكن تبريره والدفاع عنه، إذ تحدّد طبيعة البيانات بفعالية من خلال الكيفية التي يمكننا بها تصوّر البيانات، وقياسها، والتعامل معها. فبالنسبة لهم، البيانات لم تكن موجودة قبل إنتاجها؛ إلا أنها لا تنشأ من العدم. ويتمّ إنتاج البيانات من خلال القياس، والتجريد والاقتراب، وتعميم التقنيات التي تمّ وضعها ضمن تصوّر لتنفيذ مهمة ما وتسجّل في الأشكال والمقاييس التي تتفق مع المعايير التي اخترعها الناس على سبيل المثال، النظام المترى. فالبيانات هي وحدات معرفية، جُعلت على نموذج تمثيلي لتمكين وإتاحة العمل المعرفي، وقياس وتسجيل البيانات عن نفس الظاهرة من خلال العديد من الطرق والأساليب، التي يقدّم كل منها مجموعة مختلفة من البيانات التي يمكن تحليلها وتفسيرها باستخدام وسائل مختلفة (Poovey 1998). فالكيفيات التي يتمّ إنشاء البيانات بها لا يمكن حصرها نظراً لتعدد التصاميم والنقاشات والتعديلات التي تطرأ على البروتوكولات، والعمليات التنظيمية، ومعايير القياس، والفئات، والمعايير الخاصة بالبيانات في ظل الفوضى المصاحبة لعملية توليد هذه البيانات وإنتاجها. فلو أخذنا حالة قياس التعداد السكاني لبلد ما: فإن هنالك العديد من القرارات التي يجب أن تُتخذ مثل تلك التي تحدّد من يضاف للتعداد ومن يستثنى منه على سبيل المثال، أن يشمل الزوار، والأجانب المقيمين بصورة مشروعة أو غير مشروعة، وأولئك الذين يتجنبون المشاركة عن عمد أو لا، وما إلى ذلك وأين يجب أن يتمّ تعدادهم على سبيل المثال، تعدادهم تبعاً لإقامتهم ليلة إجراء التعداد أو مكان إقامتهم الدائمة في العادة؛ فجميع أنواع القواعد والإجراءات يتمّ وضعها قبل التعداد، ومع ذلك يظل هناك أمور لم يتخذ قرارات بخصوصها واختلاف عبر العدّادين في طريقة التطبيق (Porter 1995).

ونتيجة لذلك، فإن الكيفية المتبعة لتعريف البيانات وتحديد وجودها ليست عملية فنية محايدة، ولكن معايير سياسية وأخلاقية يجري التنازع عليها، ولها نتائج قابلة للتحليل، والتفسير، والعمل في وقت لاحق (Bowker and Star 1999). ومع ذلك، لمرة أخرى، تعمل تبويبات البيانات على استبعاد التباينات والتقلبات وتحديد ما يتاح ويحجب داخل مجموعة البيانات، وعلى الرغم من ذلك، تكون هذه العملية من التقارب والاستقرار معرضة على الدوام للممانعة والرفض وإعادة الصياغة بسبب تعدد الجهات الفاعلة، والموضوعات والعمليات موضع المناقشة والرد، والتناقض الحاصل في البيانات

والذي لا يمكن تطبيقه بسهولة وبما يتناسب مع نظام ما (Bowker and Star 1999). وعلاوة على ذلك، فإنه بمجرد إنتاج البيانات يكون من الممكن فرز، وتقسيم، وتكعيب هذه البيانات بطرق متعددة في عدد من الفئات. وبعبارة أخرى، فإن البيانات ليست مستقلة عن نظام الفكر والأدوات التي تدعم إنتاجها (Bowker and Star 1999)، فهذه النظم الفكرية مُطعّمة بافتراضات فلسفية ومعتقدات تُمارس بشكل مختلف. فالواقع، وكما لاحظ (Borgman 2007: 38) أنه في الوقت الذي يسعى فيه العلم إلى تصوير نفسه بالعالمية والعموم، تبدو ممارسات العلوم محلية ومتباينة على نطاق واسع، ومن ثم فإن البيانات لا تمثّل ببساطة واقع وحقيقة العالم؛ فالبيانات هي وحدات بنيوية حول العالم (Desrosières 1998; Poovey 1998). من هذا المنظور «يتمّ إنتاج المعرفة العلمية - بدلاً من براءة اكتشافها» (Gitelman and Jackson 2013: 4). كما في التالي:

إذا كان إنتاج البيانات عن طريق أداة مثل شبكة استشعار، فإن تفسير هذه البيانات يتطلب فهم الأداة - على سبيل المثال، ما الذي تستكشفه أجهزة الاستشعار؟ وتحت أي ظروف؟ وكم عدد تكرار مرات الملاحظة؟ ومع أي نوع من المعايير؟ (Borgman 2007: 183).

ومع ذلك نجد أن العلم يحاول في كثير من الأحيان جعل البيانات حرة ومستقلة عن مثل هذا الفهم السياقي، وتنظيم ومشاركة البيانات من خلال قواعد البيانات التي يجري التحسين فيها والتخفيف من حدة الفوضى عند إنشائها وحماية المستخدمين من معرفة كيفية إنتاج البيانات وتنظيمها (Gitelman and Jackson 2013). ولذلك، يُبيّن (Ribes and Jackson 2013: 165) أن الاستخدامات العلمية للبيانات والقائمة على أنها محايدة وموضوعية هي مجرد تخيلات لا حقيقة لها؛ أنها «تفترض وتصور العالم بناء على البيانات المتداولة دون النظر إلى مصدرها، وحالتها، وجوهرها، وتاريخها، ومن ثم تبرز هذه البيانات على أنها حرة للتنقل عبر العالم كعملة غير متميزة وعالمية». في المقابل، هما يؤكدان أن «البيانات هي أكثر ارتباطاً من ذلك».

ونتيجة لذلك، فالبيانات غير مسبقة التحليل، وليست عناصر موضوعية مستقلة. كما بيّنها (Gitelman and Jackson 2013: 2)، متبعين في ذلك بوكر (Bowker 2005) عندما

ذكروا أن "البيانات الخام هي مصطلح متناقض؛ «فالبيانات هي على الدوام "مطهوه" سلفاً، ولم تكن يوماً "خاماً" أبداً. فهناك حاجة لتصوّر البيانات كما هي على الحقيقة في الوجود والعمل بموجب ذلك" (Gitelman and Jackson 2013: 3). فالبيانات - على حدّ سواء - هي اجتماعية من باب أنها تقع ضمن السياق، ومادية من باب أن لديها شكلاً (كأرقام ثنائية، أو رموز، أو أعداد... وما إلى ذلك)، ويجري تخزينها على الورق، والأشرطة الممغنطة، والأقراص الصلبة، وما إلى ذلك (Wilson 2011; Gitelman and Jackson 2013). كلا المنظورين يعمل على تشكيل القواعد التأسيسية للبيانات، فعلى سبيل المثال، تعتمد البيانات الكبيرة على النقاشات الأخلاقية، والاقتصادية، والسياسية في المجالات المختلفة التي تدعم البدء في اعتمادها وتنفيذها (انظر الفصل السابع)، وتعتمد على قواعد البيانات، والحواسيب، والخوادم، وشبكات الاتصال التي تمكّن بدورها عمليات إنتاج البيانات، ومعالجتها، ومشاركتها، وتحليلها، وتخزينها (انظر الفصل الخامس). وتسهّل هذه الوسائط فساد البيانات (تعفّنها)، وسوء تموضعها، أو نسيانها، أو حذفها، أو تأكلها من خلال تلف الوحدة التخزينية للأرقام الثنائية - بت (أي تعطلّ وسائط التخزين نتيجة تحلل الأشرطة الممغنطة أو تضرر الأقراص الصلبة) (Boellstorff 2013). وفي الواقع لم تكن البيانات قطّ مطهّوة فقط، ولكنها أيضاً منفتحة لتقبّل أساليب «غير مخطط لها، وغير متوقعة، وعرضية»، و«متحولة بأشكال شبة إنسانية معقّدة، وطرق زمانية ناشئة لا تتبع على الدوام وصفة حسابية محددة مسبقاً» (Boellstorff 2013).

ونظراً للطبيعة الاجتماعية والمادية للبيانات، فإننا ننشط في إعادة صياغة وتشكيل «عواملنا المادية، والتقنية، والجغرافية، والتنظيمية، والاجتماعية إلى نوع من البيئات التي يمكن للبيانات أن تزدهر فيها... والوصول إلى العلاقة التكافلية مع البيانات» (Ribes and Jackson 2013: 152). ومن ثم، وبينما يُنظر إلى البيانات بوصفها شيئاً موجوداً - شيئاً حول الواقع (Shah 2013، التأكيد الأصلي)، فإن الفهم الأكثر إنتاجية للبيانات هو اعتبارها أحد مكوّنات الواقع وأحد مُنتجتي الواقع في ذات الوقت. إن البيانات ليست محض فكرة تمثيلية مجردة، بل هي تأسيسية، وإنتاجية، وتحليلية، وتفسيريها له نتائج وعواقب تتبعها. وكما بيّن (Gitelman and Jackson 2013: 2): «إذا كانت البيانات تخضع لنا بطريقة أو بأخرى، فإننا نحن أيضاً نخضع للبيانات».

حيث يتمّ التقاط البيانات من العالم، ولكن في المقابل تعمل في العالم. إن البيانات لم تكن يوماً، ولا يمكن أن تكون، حميدة (Shah 2013). بدلاً من ذلك، من الضروري فهم «البيانات على أنها مؤطرة متأثرة بغيرها ومؤطرة مؤثرة على غيرها» (Gitelman and Jackson 2013: 5). وبعبارة أخرى، هناك حاجة إلى ما هو أكثر من مجرد الاعتراف العام من قطاعي العلوم والأعمال لأهمية التصور المفاهيمي للبيانات.

التفكير النقدي حول قواعد البيانات والبنى التحتية للبيانات:

في العادة، يتمّ تجميع البيانات في مجاميع (Datasets) في محاولة لفهم البيانات، والتي غالباً ما يتم تنظيمها وتخزينها في قواعد بيانات (وهي وسيلة لتنظيم وتخزين البيانات التي يمكن الاستعلام عنها بطرق متعددة) وبنى تحتية للبيانات مثل نظم المحفوظات الأرشيفية (Archives) وحاويات البيانات (Repositories) (انظر الفصول الثاني والثالث والخامس). وكما كان الحال مع التصور المفاهيمي للبيانات بشكل عام، فمن المهم كذلك التفكير بشكل نقدي حول طبيعة قواعد البيانات والبنى التحتية للبيانات، وإنتاجها من الناحيتين الاجتماعية والتقنية (Socio-Technical)، وكيف أنها تعكس العقلانية حول العالم في الوقت نفسه لأنها تتكاثر وتعزز هذه العقلانية. وكان هذا التفكير النقدي غالباً إلى حد كبير فيما يتعلق بالبيانات الكبيرة، والبيانات المفتوحة، والبيانات الصغيرة الموسّعة، مع التركيز حتى الآن عليه ليكون ذا طبيعة تقنية وآلية بشكل أكبر.

وتمتدّ الفرضية المطروحة التي تمّ تبنيها وتطويرها في جميع أنحاء هذا الكتاب من الحجة المبنية في القسم الأخير، وهي الافتراض بأن قواعد البيانات والبنى التحتية للبيانات ليست ببساطة محايدة، أو وسائل تقنية لتجميع وتبادل البيانات؛ وليست بالتأكيد منتجات لتخزين البيانات الملتقطة عن العالم، ولكنها حزم من العمليات الطارئة والمترابطة التي تقوم بتأدية الأعمال في العالم (Star and Ruhleder 1996; Kitchin and Dodge 2011). إنها أنظمة اجتماعية تقنية معقدة كجزء لا يتجزأ ضمن المشهد المؤسسي الأكبر من الباحثين، والمؤسسات، والشركات، التي تشكل أدوات أساسية في إنتاج المعرفة والحكم ورأس المال.

ويتمّ تصميم وبناء قواعد البيانات على تقبل أنواع معينة من البيانات وتمكين أنواع معينة من التحليل، ولذا فهنالك آثار عميقة مترتبة على الكيفية التي يتم بها هيكلتها على

الاستفسارات والتحليلات الممكن أدائها؛ كما أن شروط صياغتها من شأنها أن تحدد العمل الذي يمكن القيام به عليها ومن خلالها (Ruppert 2012). فعلى سبيل المثال، يختلف تخزين البيانات في قواعد البيانات اختلافاً تاماً عن الشكل السردى من حيث ما تمّ تخزينه، وكيف يتم تخزينه، وكيف من الممكن استدعاؤها وإعادة استخدامها (Bowker 2005). وينتج عن قواعد البيانات إهمال وتجاهل لبعض البيانات، إضافة إلى الثغرات الكامنة في البيانات، وذلك كلما تم الكشف عن علاقات بين البيانات وتوفير الإجابة عن بعض الأسئلة التي يتعين الإجابة عنها؛ فهي تقيّد وتتيح من خلال وجودها (الأنطولوجيا)، وتنتج العديد من العلاقات في ذات الوقت الذي يتم فيه إخفاء علاقات أخرى (Ruppert 2012; Vis 2013). واستناداً إلى Derrida، فقد جادل (Bowker 2005: 12) بأن قواعد البيانات والمحفوظات الأرشيفية هي المتحكم (jussive): فهي "تقول لنا ما نستطيع وما لا نستطيع أن نقوله" من خلال تحديد ما يمكن تذكره وما يجري تجاهله ونسيانه. فهذا التذكر والتناسي يتم تحديده في الأصل من خلال أنظمة وممارسات مبنية على قوانين سياسية وفلسفية.

ومن هنا فإن التبويبات داخل قواعد البيانات غير ثابتة وغير طبيعية، ولكن يتم إنشاؤها من قبل جهات فاعلة ذات أهداف محددة تعمل ضمن مجتمعات الممارسة، أساليب الحوكمة، والقيود التقنية الفنية. ومن ثم، فإن قواعد البيانات هي تعبير عن السلطة / المعرفة وهي تقوم بإقرار تشريعها وإعادة إنتاج هذه العلاقات (Ruppert 2012)، على سبيل المثال، تحديد معدل التأمين لشخص ما أو ما إذا كان بإمكانه السفر بين البلدان. وعلاوة على ذلك، فإن قواعد البيانات هي كيانات ديناميكية متغيرة تقوم بإجراء «عملية مستمرة من التمييز والتفريق» (Ruppert 2012: 129) من خلال التفاعل مع مجاميعها المرتبطة بها (المنشئون، المستخدمون، البرمجيات، الأجهزة، الشبكات، وما إلى ذلك).

وفي الوقت نفسه، عملت قواعد البيانات على فصل عملية تحليل البيانات عن البيانات نفسها من خلال تمكين إجراء استعلامات وحسابات معقدة دون الحاجة إلى إجراء هذه التحليلات التي تتطلب الاطلاع والعمل مع البيانات نفسها أو حتى فهم الكيفية التي تمّ من خلالها تجميع البيانات وتنظيمها (Gitelman and Jackson 2013). وقد تم تعزيز هذا الفصل من خلال آليات مثل التوحيد العياري (Standardization) للصيغ والبيانات الوصفية والتعامل مع البيانات دون النظر إلى السياق الواردة فيه وبعيداً عن

التأثير السياسي (Wilson 2011). والأهم من ذلك، فإن هذا الفصل والتحرير للبيانات قد مكّن السلطة / المعرفة المتعلقة بقواعد البيانات من الانتقال والانتشار وليتم توظيفها من قبل الآخرين بعيداً عن الأعمال الداخلية المعقدة فيها وتاريخها وسياسة الإنتاج لها (بالطريقة نفسها التي يستفيد فيها السائق من السيارة من دون معرفته كيفية صنع كل أنظمتها المعقدة أو ماذا تعمل أو كيفية تفاعلها لتشكيل تجربة القيادة). ومن ثم يمكن للباحثين الاستفادة من قواعد البيانات الحكومية مثل التعداد السكاني أو المسح التجاري أو المؤشرات الاقتصادية دون معرفة الدوافع السياسية لِمَ وكيف تمّ بناء قواعد البيانات، أو الجوانب الفنية لإنتاجها، أو وجود الألفة الشخصية مع الظواهر الملتقطة. على سبيل المثال، استخدام الوزارة الإيرلندية للبيئة لقواعد بيانات عقارية غير مكتملة في أيرلندا في مرحلة ما بعد الانهيار الاقتصادي 2008 (متوفرة على الرابط:

<http://www.housing.ie/OurServices/Unfinished-Housing-Developments.aspx>

حيث كان للمرء أن يجري البحث، والتصور، والاستخلاص للنتائج حول العقارات في أيرلندا دون أي معرفة مسبقة عن تاريخ وسياسات المسح، وكيف تمّ القيام به، ودون زيارة أي من تلك العقارات (انظر Kitchin et al., 2012a, b). فمثل هذا التحرير والفصل مكّن من العمل مع قواعد البيانات على أنها غير متغيرة ومستقرة وقابلة للتحويل والنقل عبر المكان والزمان (Latour 1989).

وتستضيف البنى التحتية للبيانات قواعد البيانات وتصل بينها لتشكّل بنية تقنية اجتماعية أكثر تعقيداً. وكما هو الحال مع قواعد البيانات، لا يوجد شيء كامن أو مُعطى حول كيفية تكوين هذه الهياكل الأرشيفية والتشاركية. وفي الواقع، وكما نوقش في جميع أنحاء الكتاب، تعرّض تصميم وإدارة البنى التحتية للبيانات إلى تحديات تقنية وسياسية تم التصدي لها من خلال نقاشات فوضوية ومفاوضات عبثية تمت في سياق أتصف بالعديد من الأجنداث (جداول الأعمال) والقبولة الحكومية، فالحلول التي تمّ إيجادها والمتعلقة بالمعايير، والبروتوكولات، والسياسات، والقوانين كانت بطبيعتها ذات آثار تقريبية كونها هدفت إلى تثبيت أرضية مشتركة وتعميم الممارسات ونشرها بين المطورين والمستخدمين (Lauriault 2012)، والتمويه على وتخفيف حدة التوتر والتباين بين قابلية الاستخدام وإتاحته والحدّ من التخصيص وتقييد الابتكار، ومنع الطرق البديلة لهيكل وترتيب البيانات

(Star and Ruhleder 1996). وبالنظر إلى حدة التوترات، ينبغي التأكيد على عمليات التقريب باستمرار وبشكل متكرر من خلال التطبيق، والإدارة، وأنظمة الحوكمة (Star and Lampland, 2009). لذلك، أكد (Star and Ruhleder 1996: 112) بأنه «لا وجود لمركز مطلق يصدر منه التحكم والمعايير؛ وكذلك، لا أطراف مطلقة ... مع كون البنية التحتية شيء ظاهر للناس في الممارسة العملية، ومتصلاً بالأنشطة والهيكل».

فهذا الظهور، وعلى الرغم من عدم تركزه بشكل كامل، ليس شكلاً حراً وهو يتشكل من خلال علاقات بنيوية واسعة، فقد جادل (Graham and Marvin 2001) بأن البنى التحتية هي تأسيسية نتيجة «تراكمات بعيدة المدى من التمويل، والتكنولوجيا، والدراية، والسلطة التنظيمية والجيوسياسية» (ص. 12) والمحافظة على «الهندسة الفنية الاجتماعية للسلطة» (ص. 11) من المصالح الاجتماعية المتحجرة. وتشمل هذه التراكمات أنظمة من التشريعات تسعى من الناحية القانونية ومن خلال أشكال الحوكمة لتحديد كيفية إدارة وتحليل ومشاركة البيانات، على سبيل المثال قوانين حماية البيانات (انظر الفصل العاشر). ولذلك، اقترح (Starr 1987: 8) أن البنية التحتية للبيانات لديها:

نوعان من الهياكل التنظيمية - الاجتماعية والمعرفية: فالتنظيم الاجتماعي يتكون من العلاقات الاجتماعية والاقتصادية لأفراد العينة الممثلة، وأجهزة الدولة، والشركات الخاصة، والمنظمات المهنية والدولية، وغيرهم من المشاركين في إنتاج تدفقات البيانات من مصادرها الأصلية وحتى نقاط التحليل، والتوزيع، والاستخدام، في حين يشير التنظيم المعرفي لهيكل البيانات نفسها، بما في ذلك حدود الاستعلام، والافتراض عن الواقع الاجتماعي، وأنظمة التصنيف، وطرق القياس، والقواعد الرسمية لتفسير وعرض البيانات.

كما أكد (Dourish and Bell 2007)، أنه لا يمكن النظر بصورة صرفة لقواعد البيانات والبنى التحتية كوسائل وأدوات ثانوية لأنها تشمل بطبيعتها نواحي ثقافية، واقتصادية، ومعرفية، وهي غارقة في الأهمية على المستوى الاجتماعي. ولذلك اقترحا منظورين يمكن من خلالها فهم البنى التحتية للبيانات: المنظور الأول يتضمن قراءة اجتماعية سياسية يُدرس من خلالها «بلورة العلاقات المؤسسية» (ص. 416). أما المنظور الثاني فهو القراءة التجريبية التي تدرس " كيفية تشكيل البنى التحتية للبيانات لإجراءات الأفراد وخبراتهم"

(ص. ٤١٧). وفي كلتا الحالتين، تُفهم البنى التحتية للبيانات ككيانات مترابطة. ويُعيد هذا الترابط تشكيل العالم بشكل عرضي بناء عليها، وهي بدورها تتشكل من قبل العالم كذلك. ولكي نصل إلى استخدام قواعد البيانات والبنى التحتية للبيانات والاعتماد عليها لفهم العالم والقيام بأعمال فيه، فإن ممارساتنا المنطقية والمادية تتكيف وتتمحور كاستجابة لهما (Star and Ruhleder 1996). ولا ينعكس العالم في البيانات فقط، بل يتغير بها؛ «فالعمل على إنتاج وحفظ وتبادل البيانات ومشاركتها يقوم بإعادة تشكيل العوامل التنظيمية، والتكنولوجية، والثقافية من حولهم» (Ribes and Jackson 2013: 147).

وبعبارة أخرى، لا تدعم قواعد البيانات والبنى التحتية للبيانات البحوث ببساطة، ولكنها تُغير وبشكل جذري الممارسات والنظم البحثية - الأسئلة المطروحة، وكيفية طرحها، وطريقة الإجابة عليها، وكيفية تطبيق الإجابات، ومن يجري البحث وكيف يعملون بصفة باحثين (انظر الفصل الثامن). فعلى سبيل المثال، وفي دراستها لتطور التعداد السكاني الكندي وأطلس كندا، بيّنت (Lauriault 2012) بالتفصيل كيف تطور كل تعداد بشكل متواتر ومتكرر على أساس نماذج للعالم شكلت طرُقاً لتصور وعرض كندا. وقد جادلت بأن محفوظات البيانات والبيانات نفسها تشكل «نظام ذاكرة خارج الجسم المؤسسي يسمح بسرد قصص عن طبيعة كندا... من خلال الخرائط والرسوم البيانية والنماذج والإحصاءات التي تعتمد على أجهزة الاستشعار، والبيانات، وقابلية التشغيل البيئي، ومعايير رسم الخرائط على شبكة الإنترنت، والبوابات الإلكترونية، والبيانات الوصفية والنماذج، والعلوم، والهياكل المفتوحة» (ص. ٢٧). في المقابل، تعدّل هذه القصص النماذج الأساسية ومن ثم البنية التحتية للبيانات، مما يعمل على تحويل الوسائل التي يتم من خلالها إنشاء القصص.

ويتطلب تفهم قواعد البيانات والبنى التحتية للبيانات التفكيك والتفريغ بعناية لطبيعة نشوئها، وتداخلها، وعلائقيتها، وسياقيتها (Star and Ruhleder 1996). وهذا يعني البحث فيما وصفه (Bowker and Star 1999: 34) بأنه انعكاس الأساس (Infrastructural Inversion) أي أن ندرك أعماق الترابط بين الشبكات التقنية والمعايير الفنية، من جهة، والعمل الحقيقي للسياسة وإنتاج المعرفة من جهة أخرى. وكما بيّنت (Lauriault 2012)، فإن هذا يتطلب أيضاً تحليلاً تاريخياً يوثق كيف تتطور قواعد البيانات والبنى التحتية للبيانات مع مرور الوقت وتغير المكان.

مجاميع البيانات وثورة البيانات (Data Assemblages and the Data Revolution) :

كان الجدل والحجة الرئيسية المعروضة في هذا الفصل بأن التفكير بالبيانات لا يكون بالطريقة المباشرة، ولا يمكن أن توجد البيانات بشكل مستقل عن الأفكار، والآليات، والتقنيات، والنظم، والناس، والسياقات، بغض النظر عن كونها في كثير من الأحيان يتم تصورها بهذه الطريقة (Lauriault 2012; Ribes and Jackson 2013). فالبيانات يتم إنشاؤها بوصفها نتاجاً لكثير من العقول العاملة في مختلف الحالات والأوضاع، وهي مؤطرة وتتشكل في ظروف وهيكل مختلفة.

ولذا فإن من طرق فهم البيانات هو التفكير بها كقضية مركزية لتجمّع فني اجتماعي معقد، ويتكوّن هذا التجمّع للبيانات من العديد من الأدوات والعناصر التي تزداد تشابكاً، وتطوراً، وتتحوّر مع مرور الزمن وتبدّل المكان (انظر الجدول 3-1). وتصوغ كل أداة مصحوبة بعناصرها ما هو ممكن، وما هو مطلوب، وما هو متوقع من البيانات. وعلاوة على ذلك، فهي تتفاعل بعضها مع بعض ويشكل بعضها بعضاً من خلال شبكة معقدة ومتغيرة من العلاقات المتعددة الأوجه (انظر الشكل 3-1). وكما ادّعى كل من Ribes and Jackson (2013)، فإن هذه الأدوات لا تقوم بتأطير ماهية وكيفية إنتاج البيانات وإلى أي حدّ من الممكن توظيفها فحسب، ولكنها بحدّ ذاتها منظمة ومُدارة لتقوم بإنتاج مثل هذه البيانات. وعلى ذلك، فإن البيانات ومجاميعها تتشارك التحديد (Co-determinous) وتتبادل عملية التشكيل كل منها للآخر (Mutually Constituted)، وترتبط ببعضها من خلال مجموعة من الممارسات والعلاقات المنطقية والمادية الطارئة، والعلائقية، والسياقية. وعلى ذلك، يختلف كل تجميع للبيانات في نسق ترتيب العناصر وسياقها، ولكنها تتشارك في قواسم مشتركة وتأثير كل منها في الآخر بسبب اعتماد الأدوات على بعضها وتداخلها والاتفاقيات التي تغطي مجاميع البيانات، ومع نشوء أفكار ومعارف جديدة، واختراع التقنيات، وتطوير المهارات، وانفتاح الأسواق، فإن مجاميع البيانات تتطور، وتتحوّر، وتندمج، وتتفكك. ونتيجة لذلك، هناك تنوع كبير من تجمّعات البيانات عبر النطاقات والاختصاصات المختلفة.

جدول رقم (٣-١)

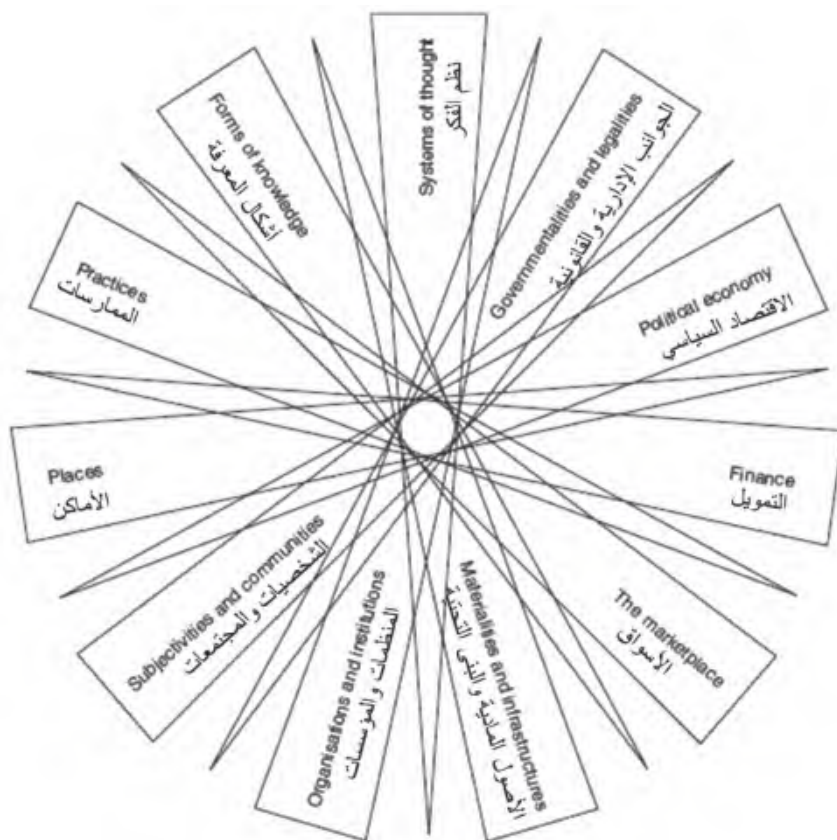
الأدوات والعناصر لمجاميع البيانات

العناصر	الجهاز / النظام
طرائق التفكير، الفلسفيات، النظريات، النماذج، الإيديولوجيات، المنطقيات،... إلخ.	نظم الفكر
النصوص البحثية، الأدلة، المجلات، المواقع الإلكترونية، الخبرة، الخُطب، منتديات الدردشة،... إلخ.	أشكال المعرفة
نماذج الأعمال، الاستثمار، رأس المال الاستثماري، المنح، التبرعات والصدقات، الربح،... إلخ.	التمويل
السياسة، الأنظمة الضريبية، الرأي العام والرأي السياسي، الاعتبارات الأخلاقية،... إلخ.	الاقتصاد السياسي
معايير البيانات، صيغ الملفات، متطلبات النظم، اللوائح البروتوكولية، القوانين، الترخيص، أنظمة الملكية الفكرية،... إلخ.	الجوانب الإدارية والقانونية
الأوراق / الأقلام، الحاسبات الآلية، الخدمات الرقمية، أجهزة الاستشعار، المساحات الضوئية، قواعد البيانات، شبكات الاتصال، الخوادم،... إلخ.	الأصول المادية والبنى التحتية
التقنيات، طرق التنفيذ، السلوكيات المستفاد، الاتفاقيات العلمية،... إلخ.	الممارسات
سجلات الأرشفة، الشركات، المستشارون، المصنّعون، بائعو التجزئة، الجهات الحكومية، الجامعات، المؤتمرات، النوادي والمجتمعات، واللجان والمجالس، المجتمعات الممارسة،... إلخ.	المنظمات والمؤسسات
منتجو البيانات، القِيَمون على البيانات، المديرين، المحللون، العلماء، السياسيون، المستخدمون، المواطنون،... إلخ.	الشخصيات والمجتمعات
المختبرات، المكاتب، مواقع العمل، مراكز البيانات، مجموعات الخوادم، مجتمعات الأعمال،... والقوى البشرية المتصلة بها.	الأماكن
للبيانات، ولمشتقاتها (على سبيل المثال، النصوص، الجداول، المخططات، الخرائط)، والمحللون، وبرمجيات التحليل، والتفسيرات،... إلخ.	الأسواق

فهذا الكتاب يبحث في تجمعات البيانات الناشئة، والمتطورة، المنتجة للبيانات المفتوحة، والبنى التحتية للبيانات، والبيانات الكبيرة، وللقيام بذلك يطرح الكتاب ثلاث مسائل رئيسية للمناقشة: أولاً، هناك حاجة لتطوير طرق مبدئية وفلسفية لفهم البيانات. فهناك نقص ملحوظ في التفكير النقدي والبحثي عن البيانات وفي البيانات بحد ذاتها من حيث التكوين والتشغيل للمجاميع المحيطة بها، وعلى الخصوص بالمقارنة مع التركيز والاهتمام اللذين تحظى به مفاهيم المعلومات والمعرفة. وذلك على الرغم من أن البيانات هي عنصر أساسي في إنتاج المعرفة. ثانياً، هناك ثورة بيانات جارية تعمل على تشكيل لحظة حاسمة في تطور وتحوُّر تجمعات البيانات. ونظراً لالتقاء العديد من التطورات الناشئة في مجال الحوسبة، والتقنيات المنهجية، والمجالين السياسي والاقتصادي، فإن حجم البيانات، وتنوعها، وسرعتها، ودقتها، وتوافرها، والكيفية التي يتم بها معالجة البيانات وتحليلها وتخزينها وتوظيفها لزيادة التبصر وتحصيل القيمة، قد شهد تحولاً جذرياً. ثالثاً، نظراً لمختلف التحديات التقنية والأخلاقية والعلمية التي أثارها ثورة البيانات، فإن هناك حاجة ملحة لتطوير فهم مفصل لتجمعات البيانات الجديدة والناشئة التي يتم إنشاؤها، ولذا تهدف الفصول العشرة التالية إلى توفير نظرة إجمالية على نطاق واسع، وشامل، ودقيق لهذه التجمعات وتبسيط الضوء على القضايا التي تتطلب مزيداً من الاهتمام والبحث.

شكل (١-٣)

التقاطع بين أدوات مجاميع البيانات



الفصل الثاني

البيانات الصغيرة (Small Data)، والبنى التحتية للبيانات (Data Infrastructures)، ووسطاء البيانات (Data Brokers)

قبل بضع سنوات مضت كان مصطلح «البيانات الصغيرة» نادر الاستخدام، إن لم يكن معدوماً. وقد ذاع انتشاره مصطلحاً بحكم الواقع على النقيض مما سُمي «البيانات الكبيرة». فكلّ البيانات كانت في الواقع بيانات صغيرة ومن ثم لم تتطلب تمييزها بهذا الإصطلاح. وسيجري لاحقاً مناقشة ما قام على تأسيس البيانات الكبيرة بالتفصيل في الفصل الرابع، ولكن من الضروري تسليط الضوء على الفرق بين البيانات الصغيرة والبيانات الكبيرة، وذلك قبل التوجه إلى مناقشة كيفية تحوّل البيانات الصغيرة من خلال البنى التحتية الجديدة للبيانات والأسواق المقدّرة بـبلايين الدولارات، كما سيتم لاحقاً في الفصل التالي مناقشة انفتاحها لتقبل المشاركة وإعادة الاستخدام.

فمن الناحية التقليدية، ونظراً للتكاليف والصعوبات المصاحبة لعمليات إنتاج ومعالجة وتحليل وتخزين البيانات، فقد تم إنتاج البيانات الصغيرة بطرق مضبوطة بأحكام باستخدام تقنيات أخذ العينات التي تحدّد من نطاقها، وصلاحياتها الزمنية، وحجمها (H.J. Miller 2010). وبينما يمكن أن تكون بعض قواعد البيانات كبيرة جداً في الحجم، مثل قاعدة بيانات التعداد الوطني، ولجعل ممارسة تجميع هذه البيانات قابلة للإدارة، فإن عملية إنتاجها عادة ما تتم خلال بضع سنوات وتكون محدودة النطاق (في حالة التعداد، يجري في العادة مرة واحدة كل عشر سنوات ويكون مقتصرًا على ثلاثين إلى أربعين سؤالاً فقط). وعلاوة على ذلك، فقد كانت دقة البيانات الصغيرة العامة (Small Tertiary Data) في كثير من الأحيان ضعيفة جداً (كانت، على سبيل المثال، على مستوى المقاطعات أو الدول وليس الأفراد والأسر)، كما أن الأساليب المستخدمة لإنتاجها غير مرنة على الإطلاق (على سبيل المثال، من المستحيل تعديل الأسئلة أو الإضافة لها أو الإزالة منها بمجرد التهيئة للتعداد وإعداده والبدء بإدارته لأن ذلك سينعكس سلباً وبشكل جديّ على القدرة على المعالجة والتحليل). وفي المقابل، تتميز البيانات الكبيرة

بكونها تُنتج بشكل مستمر، وتسعى لأن تكون شاملة وغارقة في التفاصيل ودقيقة على مستوى النطاق الذي تغطيه، ومرنة وقابلة للتطوير في عملية إنتاجها (انظر الجدول 1-2 والفصل الرابع).

جدول رقم (٢-١)

مقارنة البيانات الصغيرة والبيانات الكبيرة

البيانات الكبيرة	البيانات الصغيرة	
ضخمة جداً	محدودة	الحجم
كامل مجتمع الدراسة	عينات	درجة التفصيل / الشمولية
محكمة وقوية	من تقريبية وضعيفة إلى مترابطة وقوية	الدقة والتحديد
قوية	من ضعيفة إلى قوية	المنطقية / العلائقية
سريعة، مستمرة	بطيئة، مجمدة التأخير / محزّمة	السرعة
واسعة	من محدودة إلى واسعة	التنوع
عالية	منخفضة إلى متوسطة	المرونة وقابلية التوسع

وفي الوقت الذي تعد فيه البيانات الصغيرة واحات صغيرة ضمن صحارى البيانات، تنتج البيانات الكبيرة طوفان بيانات حقيقي. وقد أدى ذلك بالبعض إلى التساؤل عما إذا كانت البيانات الكبيرة قد تؤدي إلى زوال البيانات الصغيرة أو تضائل مكانة الدراسات التي تعتمد على البيانات الصغيرة نظراً لمحدوديتها في الحجم، والنطاق، والصلاحية الزمانية. فقد لاحظ Sawyer (2008) أن وكالات تمويل البحوث الأكاديمية، على سبيل المثال، توجه موارد التمويل المحدودة الخاصة بها إلى مجالات البيانات الغنية على نحو متزايد، وذلك ربما بسبب التداخل بين حجم البيانات وسرعتها مع مقدار التبصر والتوظيف والقيمة، وأن هذه الوكالات تحول الأسئلة البحثية التي من الصعب أن تنتج بيانات كبيرة لجهات التمويل الأخرى المتعددة وتمنحها وضع هامشي داخل وخارج هذه المؤسسات الأكاديمية. وبعيداً عن البحوث الأساسية، فقد استكمل هذا التحرك من خلال توجه العديد من وكالات التمويل نحو البحوث التجريبية التطبيقية ذات الشراكة الصناعية. وقد أدى إعادة الترتيب

لألويات البحث إلى سوء الفهم لطبيعة البيانات الكبيرة وقيمة البيانات الصغيرة، فالبيانات الكبيرة ربما تهدف إلى أن تكون شاملة، ولكن كما هو الحال مع جميع البيانات فإن كليهما عبارة عن تمثيل (Representation) وعينة (Sample)، فماهية البيانات التي يتم التقاطها تتشكل في الواقع من خلال التالي:

- مجال إطار عرض / أخذ العينات (حيث يتم توظيف أجهزة التقاط البيانات وماهية إعدادات التهيئة لها/ ومعاملاتها؛ مثل من يستخدم المكان أو وسائل الإعلام، وعلى سبيل المثال من الذي ينتمي إلى وسيلة التواصل الاجتماعي الفيسبوك أو يتسوق في محلات وول مارت... إلخ).
- التقنية والمنصة المستخدمة (المسوحات المختلفة، وأجهزة الاستشعار، والعدسات، والمطالبات النصية، والمخططات والتصاميم... إلخ، وكلها تنتج فروقاً وتباينات في ماهية البيانات التي يتم إنشاؤها).
- السياق الذي يتم إنشاء البيانات فيه (الأحداث الجارية تعنى أن البيانات تقع دائماً نسبة إلى الطرف الذي تقع فيه).
- تبويبات (أنطولوجية) البيانات المستخدمة (كيفية معايرة البيانات وتصنيفها).
- البيئة التنظيمية فيما يتعلق بالخصوصية، وحماية وأمن البيانات.

وتلتقط البيانات الكبيرة في العموم ما هو سهل التقاطه - أي البيانات التي يتم التعبير عنها علناً (ما كُتب، ومُرر، ومُسح ضوئياً، وما تمّ استشعاره... إلخ؛ وتصرفات الناس وسلوكهم؛ وحركة الأشياء) - وكذلك البيانات المستنفذة والتي هي ناتج وظيفة رئيسية وليست المخرج الرئيسي، والبيانات الوصفية التي تمّ إنشاؤها تلقائياً. وتأخذ البيانات الكبيرة هذه البيانات بالقيمة الاسمية، على الرغم من أنها قد لا تكون قد تم تصميمها للردّ على أسئلة محددة والبيانات الناتجة قد تكون عشوائية وغير دقيقة (انظر الفصل التاسع).

أما البيانات الصغيرة فقد تكون محدودة في الحجم والسرعة، ولكن لديها تاريخاً طويلاً من التطوير، مع منهجيات مقررة وأطوار للتحليل، وسجل حافل من الإجابات الناتجة للأسئلة العلمية. ويمكن لدراسات البيانات الصغيرة أن تكون مصممة بشكل أفضل بكثير للرد على

أسئلة بحثية محددة والاستكشاف المفصل والمعتمق للطرق المتنوعة، والسياقية، والمنطقية وغير المنطقية التي يتعامل بها الناس، ويتم فهم العالم بها، وكيفية عمل العمليات العلمية (انظر الفصلين الثامن والتاسع). ويمكن أن تركز البيانات الصغيرة على حالات محددة وتُخبر عن قصص فردية، وسياقية، ودقيقة. وبسبب أنها تستهدف ظاهرة معينة، فهي تركز بشكل مكثف على قضية وعادة ما تكون أقل حاجة للموارد بشكل مكثف: وتسعى دراسات البيانات الصغيرة على التنقيب عن الذهب من خلال العمل في شق ضيق للغاية، في حين تسعى دراسات البيانات الكبيرة إلى استخراج شذرات من خلال التعدين في حفرة مفتوحة وشفط وغرلة مساحات شاسعة من الأراضي. ومن ثم فقد برهن (2008) Sawyer على أن التمويل يحتاج إلى استهداف الإجابة عن الأسئلة الحرجة، سواء كانت معالجتها باستخدام بيانات صغيرة أو كبيرة، وليس مجرد أن توجّه نحو المشاريع التي يمكنها الوصول إلى كميات هائلة من البيانات على أمل أنها تحوي بطبيعتها رؤى وأفكاراً مفيدة.

وهكذا، يمكن القول إنه من الممكن تحصيل المزيد من القيمة والرؤى من البيانات الصغيرة من خلال توسيع نطاقها إلى مجموعات بيانات أكبر، وجعلها متاحة لإعادة الاستخدام من خلال البنى التحتية للبيانات الرقمية والتي تجعلها إضافة إلى ذلك قابلة لمختلف أشكال التحليل الحسابي، والحفاظ عليها للأجيال القادمة. ونتيجة لذلك، وكما هو الحال مع البيانات الكبيرة، فقد استثمرت موارد كبيرة في السنوات الأخيرة لإنشاء مثل هذه البنى التحتية للبيانات، والتي أحدثت تغييرات كبيرة في مشهد البيانات.

حاويات البيانات (Data Holdings)، محفوظات البيانات (Data Archives)، والبنى التحتية للبيانات (Data Infrastructures):

تمّ جمع البيانات معاً وتخزينها لفترة طويلة من التاريخ المسجّل، وقد كانت هذه الممارسات غير رسمية ورسمية على حدّ سواء في طبيعتها، فالممارسات غير الرسمية تتكوّن ببساطة من جمع البيانات وتخزينها، في حين تتكوّن الممارسات الرسمية من مجموعة الممارسات الفاعلة ذات العلاقة بإدارة دورة حياة البيانات (Data Curation) والهياكل المؤسسية المصمّمة لضمان الحفاظ على البيانات من أجل الأجيال القادمة. وقد يكون من الأفضل وصف الممارسات غير الرسمية بأنها حاويات البيانات، في حين أن الممارسات الرسمية

هي محفوظات (أرشفات) البيانات. ويجري، في كثير من الأحيان، الخلط بينهما أو تستخدم بالتبادل، جنباً إلى جنب مع غيرها من المصطلحات مثل النسخ الاحتياطي (Backup)، وبنك المعلومات (Information Bank)، ومستودع البيانات (Data Repository)، ولكن المحفوظات الأرشيفية تمتلك خصائص مميزة جداً عند أمناء المحفوظات (Archivists) المهنيين.

فالمحفوظات هي مجموعات رسمية منظمة من البيانات التي يتم هيكلتها وتوثيقها ونشرها بشكل فعال، مصحوبة بالبيانات الوصفية المناسبة لها، مع التخطيط الفعال لقضايا المحافظة عليها وإتاحة الوصول إليها وإمكانات اكتشافها (Lauriault et al. 2013). فالمحفوظات تهدف بشكل واضح وصريح وطويل الأمد، للحفاظ على البيانات لإعادة استخدامها في المستقبل. إن مفهوم الحفاظ، في هذا السياق، لديه استخدام خاص محدد على أنه مجموعة من الأنشطة المدارة والمدعومة «بالمبادئ، السياسات، القواعد، والإستراتيجيات الرامية إلى إطالة أمد الكائن (البيانات) من خلال المحافظة عليه في حالة مناسبة للاستخدام، سواء على شكله الأصلي أم في شكل مُستدام أكثر ثباتاً، مع الإبقاء على الشكل الفكري للكائن» (InterPARES 2, 2013). ويسعى الأرشيف للحفاظ على مجموعة سجلات كاملة، وليس مجرد البيانات؛ أي أنها تشمل كل الوثائق الداعمة والبيانات الوصفية وغيرها من المواد ذات الصلة التي تحتوي على تفاصيل المنشأ والسياق فيما يتعلق بكيفية إنتاج البيانات والطريقة التي ينبغي أن تعامل وتحلل وتفسر بها. فالنهج المتبع في الحفاظ على البيانات مبني على الإدراك التام بأن التقنيات والبروتوكولات والمبادئ التوجيهية لأفضل الممارسات هي عرضة للتغيير والتقدم، بمعنى أن البيانات سوف تحتاج للتنقل والترحيل عبر منصات وتقنيات بسبب الابتكارات الجديدة التي تأتي في السياق، كما يُدرك النهج المتبع بأنه دون ممارسات فعالة لإتاحة وإعادة الاستخدام والمشاركة للبيانات فإن البيانات معرضة للتلف، أو الضياع، أو الحرمان من بياناتها الوصفية السياقية والوثائق الداعمة (Borgman 2007; Dasish 2012). ويسعى المحفوظ الأرشيفي إلى توفير موارد جديدة بالثقة، من خلال الحفاظ على سلامة البيانات (Data Integrity) مع مرور الوقت.

وفي بعض الحالات، قد يلزم مؤتمنو البيانات (Data Custodians) بمسؤولية قانونية لحفظ وأرشفة البيانات بشكل مناسب لإعادة الاستخدام الحالي والمستقبلي، كما قد يكون

المؤمن على الأرشفة والقائم عليها نفسه سلطة قانونية تشريعية. فعلى سبيل المثال، فإن المحفوظات الأرشيفية الوطنية والوكالات الإحصائية الوطنية هي الكيانات القانونية التي عهد إليها جمع وإنتاج وأرشفة قواعد بيانات محددة مثل السجلات الشخصية للأفراد والحسابات القومية، والتعدادات والمسوح السكانية. وينبغي أن تمثل المحفوظات أيضاً للتشريعات المتعلقة بحقوق الوصول إلى البيانات، والخصوصية، والمبادئ الأخلاقية، وحقوق الطبع والنشر، والملكية الفكرية، وتحديد المسموح لهم للوصول إلى البيانات وما يمكنهم القيام به من خلال تحديد الصلاحيات على البيانات. على سبيل المثال، في العديد من البلدان، من الممكن نشر سجلات التعداد في شكل إجمالي، مع تقييد نشر السجلات الشخصية إلى ما بعد مئة عام، فالالتزام بإتاحة الوصول وإعادة الاستخدام للبيانات لا يستلزم بالضرورة العمل على نموذج مفتوح (انظر الفصل الثالث)، ولكن ينبغي أن يتم ضمان توافر البيانات في المستقبل للمجتمع المستهدف بطريقة قانونية (CCSDS, 2012). إن المجتمع المستهدف هو عادة مجموعة معينة من العلماء أو المتخصصين في مجال السياسات الذين لديهم في العادة احتياجات محددة تماماً من البيانات، ومتطلبات وظيفية ودعم مما يعني أن حلول الأرشفة العامة ستكون ذات فائدة محدودة (Lauriault et al. 2013).

ومن ثم، فإن محفوظات البيانات ليست مجرد مخازن بيانات أو أنظمة نسخ احتياطي (والتي تحفظ البيانات في حال حدوث أعطال على المدى القصير مثل عطل الوسط التخزيني أو الفشل التقني)، ولكنها تُخطّط، وتنسّق، وتُدار، وتُوظّف بفعالية من قبل موظفين متفرغين ومتخصصين يعملون على إضافة القيمة وضمان الاستمرارية. فهؤلاء الأفراد، إضافة إلى الكيان الإداري الوظيفي، مسؤولون عن الأرشفة إضافة إلى ثلاث مجموعات أخرى من الأطراف الفاعلة في عملية الأرشفة وهم: منتجو البيانات، ومستهلكو البيانات، والإدارة (CCSDS 2012). فالمنتجون هم منشئو البيانات أو الكيان القانوني المسؤول عن إنتاجها؛ أما المستهلكون فهم مستخدمو المحفوظات، وأصحاب المصلحة المعنيون بها أو مجتمع معين؛ وأما الإدارة فهي الهيئة التي تشرف على العملية، وتوفّر للأرشيف ميثاقاً ونطاقاً، ولكنها لا تأخذ دوراً نشطاً في عمليات الأرشفة اليومية (CCSDS 2012; Lauriault et al. 2013). وتشكّل هذه الدوائر مجتمعة كيفية بناء وتشغيل وتطوير الأرشيف مع مرور الوقت.

في المقابل، فإن حاويات البيانات هو مجموعة غير رسمية وغير نظامية من البيانات. وقد يعتقد بعض منتجي البيانات أن النسخ الاحتياطي أو أنظمة التخزين تشكل محفوظات أرشيفية بحكم الواقع، ولكنها تفتقر إلى الجهاز المؤسسي، والتنظيم، والسياسات الرسمية النظامية (بالنسبة إلى المعايير، وإتاحة الوصول إليها، وإمكانية إعادة استخدامها، وحقوق الملكية الفكرية، والمعايير الأخلاقية، وإلى آخره)، وما يطمح إلى تحقيقه الأرشيف (Lauriault et al. 2013). فقد تكون حاويات البيانات منتظمة وموجهة، ولكن بدون قواعد صارمة أو أي خطط طويلة المدى تتعلق بالمحافظة على البيانات وحفظها، ودون أي التزام لمشاركة البيانات وتبادلها أو جعلها متاحة لإعادة الاستخدام. وعلى الرغم من أن حاويات البيانات قد تلتزم بشكل جيد بمعايير البيانات المتوقعة، إلا أن هذا الالتزام يفتقر إلى الاعتماد والمصادقة عليه ويفتقر إلى الوثائق الأخرى والبيانات الوصفية المناسبة.

وقد فقدت معظم البيانات التي تم إنتاجها على مر التاريخ أو دُمّرت بسبب تخزينها في حاويات بيانات، وليس أرشيفاً، أو لأنه تقرر الحفاظ على المعلومات المستقاة من هذه البيانات (مثل المقالات والكتب) التي كانت تعد أكثر قيمة، وتخزينها في المكتبات. وبشكل عام، تم الإبقاء فقط على مجموعات البيانات الأكثر قيمة، مثل تلك المرتبطة بالمساعي العلمية والثقافية الأساسية والسجلات الحكومية، والمعاملات الاقتصادية، والعقود القانونية. وفي حالات النماذج، الفنون، الكائنات... إلخ والتي يمكن اعتبارها صوراً معينة من البيانات فقد تم تنسيقها وتخزينها بالأساس في المتاحف وصلات العرض جنباً إلى جنب مع البيانات الوصفية المرتبطة بها. وقد تم تخزين البيانات الخاصة بمعظم العلماء بشكل غير رسمي على ملفات وصناديق أو على مختلف محركات الأقراص الصلبة في مكاتبهم أو في منازلهم. وعندما يتقاعدون خارج الخدمة أو يتوفون يتم تدمير محركات الأقراص الصلبة في مكاتبهم أو في منازلهم. قاموا بإنتاجها. كما أن الغالبية العظمى من البيانات التي تم إنتاجها لأطروحات الدكتوراه تفقد بعد الانتهاء منها، وغالباً لا يتم نشر حتى المعلومات المستقاة من البحوث. وعلاوة على ذلك، فإن مُمَوِّلي البحوث تاريخياً لم يطالبوا المشاريع التي قاموا بتمويلها بحفظ البيانات وتخزينها، أو إن طلبوا ذلك فقد كان لفترة قصيرة من الزمن. وكما هو الحال في العديد من المشاريع التي يمولها الاتحاد الأوروبي، يُشترط الاحتفاظ بجميع مواد المشروع بما في ذلك البيانات لمدة سبع سنوات، ولكن يمكن بعد ذلك التخلص منها، وهو ما يتم في

كثير من الحالات، بسبب تكاليف التخزين والظن الغالب بأن المقالات الورقية الناتجة تمثل قيمة أكبر.

ومع تطور حلول التخزين الرقمي، التي قللت التكلفة والمساحة التخزينية للاحتفاظ بالبيانات، فقد بدأ التخزين الطويل الأمد على نطاق واسع للبيانات الاعتيادية المنخفضة التكلفة (انظر الفصل الخامس). فقد ظهرت أولى قواعد البيانات الرقمية التي ضمت مجموعات البيانات في الخمسينيات والستينيات من القرن الماضي وكانت تنسق البيانات في هياكل مسطحة أو هرمية منظمة (Driscoll 2012). ومع ذلك، فقد كانت هذه الترتيبات غير فعالة للبحث والاستعلام عن البيانات، كما تم تطوير قواعد البيانات العلائقية (Relational databases) ابتداءً في عام ١٩٧٠ من قبل إدغار كود (Edgar Codd) وقد نظم خصائص البيانات المميزة (Attribute) في مجموعة من الجداول المترابطة، حيث يتشارك كل جدول بحقل بيانات واحد على الأقل مع جدول آخر، فعلى سبيل المثال، قد يكون رقم الضمان الاجتماعي هو الحقل المشترك بين جدولين (المفتاح التأشير - Indexical Key) أحدهما يحتوي على التفاصيل الشخصية (مثل العمر، الجنس، الحالة الاجتماعية، والعنوان) ويحتوي الجدول الآخر على المتعلقات المالية، ويسمح رقم الضمان الاجتماعي بربط البيانات للاستعلام في كلا الجدولين معاً وإنتاج بيانات جديدة مشتقة منهما. وإذا كان المفتاح المشترك عنواناً أو تاريخاً، فيمكن عندها وضع البيانات جغرافياً أو زمانياً. لقد مكنت قواعد البيانات العلائقية تنظيمًا واستعلاماً أكثر فعالية وتعقيداً عن البيانات الهيكلية باستخدام لغات الاستعلام الهيكلية (SQL). وجنباً إلى جنب مع قواعد البيانات العلائقية، سمح تطوير البيانات المجدولة (Spreadsheets) بتنظيم وتخزين كميات كبيرة من البيانات الحسابية إضافة لتطبيق الصيغ والمعادلات عليها مما يؤدي إلى إنتاج بيانات مستقاة جديدة. وقد استكملت هذه الأشكال من قواعد البيانات، في السنوات الأخيرة، بقواعد البيانات غير الهيكلية (بلا استعلام -noSQL) التي يمكنها التعامل مع كميات كبيرة من البيانات غير الهيكلية (انظر الفصل الخامس).

ويكون إنشاء ومحو مجموعات البيانات سهلاً في الحالات التي تنتج فيها البيانات ابتداءً بصورة رقمية (نشأت بالكود الثنائي -binary code)، وأحياناً بطرق تلقائية ومؤتمتة، كما

أنها عرضة للخسارة بشكل كبير بسبب التقادم في البرمجيات والمعدات. وقد بين O'Carroll (2012) and Webb أنه: «في حين يمكن لأي أحد التقاط كتاب كُتب قبل مئة سنة، والنظر فيه، وقراءة صفحة منه، فإنه لن يكون ذلك بالمثل صحيحاً لقرص مرن يحتوي على ملفات نصية لبرمجيات وورد بيرفكت أنشئت قبل ٢٠ سنة فقط». وفي المقابل، ينبغي أن تحوّل الأشكال التناظرية/ التماثلية للبيانات إلى الصورة الرقمية الثنائية على سبيل المثال، مسحها ضوئياً أو إعادة تسجيلها وإدخالها، والتي يمكن أن تكون مكلفة وتستغرق وقتاً طويلاً، على الرغم من إمكانية أتمتة هذه العملية آلياً إلى حد ما. وقد حدّت التكاليف والنفقات المرتبطة بالرقمنة (Digitisation) من حيث التكلفة المادية، وقت الموظفين، والمعدات المتخصصة، من تطبيقها في العديد من المحفوظات الأرشيفية التناظرية التي تحتفظ بها المتاحف والمكتبات والمجموعات الشخصية الخاصة. وبينما كافتحت هذه المؤسسات لتمويل أنشطة الرقمنة، فإن جهات خيرية غير ربحية مثل أرشيف الإنترنت (<http://archive.org>) وتجارية مثل قوقل (Google) تساعد على القيام بمثل هذه الأنشطة، وذلك باستخدام مواردها الذاتية أو تبرعات المحسنين من عامة الأفراد مما يجعلها متاحة مجاناً للجميع (انظر الفصل الخامس).

وفي جميع الحالات، فإن البيانات ضمن حاويات البيانات الرقمية والمحفوظات الأرشيفية يمكن مشاركتها وإعادة استخدامها بسهولة وبتكلفة هامشية منخفضة، وذلك على الرغم من أنه من الممكن الحدّ فيما يتعلق بالوصول وإعادة الاستخدام بواسطة سياسات حقوق الملكية الفكرية. وعلاوة على ذلك، فإنه من الممكن معالجتها وتحليلها من خلال الأدوات الحسابية القوية، إذ أصبحت الإجراءات والحسابات التي كان من الصعب القيام بها يدوياً أو باستخدام التقنيات التناظرية على سبيل المثال، المسطرة الحاسبة التناظرية (Slide rules) ممكنة في غضون بضعة أجزاء من الثانية، وأكثر في بناء واختبار النماذج وأنظمة المحاكاة المعقدة. كما يمكن أيضاً ربط حاويات البيانات الرقمية والمحفوظات الأرشيفية معاً بسهولة نسبياً ومدّها إلى أشكال أخرى من البنى التحتية للبيانات.

فالبنية التحتية للبيانات هي الوسيلة المؤسسية والمادية والرقمية لتخزين وتبادل واستهلاك البيانات عبر التقنيات المرتبطة ببعضها البعض، وعلى مدى العقدين الماضيين على

وجه الخصوص، فقد تمّ بذل جهد كبير لتطوير وتعزيز هذه البنى التحتية، والتي تأخذ عدة أشكال: الكتالوجات، والأدلة، والبوابات، ومراكز تبادل المعلومات، والمستودعات، والمستودعات الرقمية الموثوقة، والبنى التحتية السيبرانية (Cyber-Infrastructures) (Lauriault et al. 2007). وغالباً ما تستخدم هذه المصطلحات بالتبادل ويجري الخلط بعضها ببعض، وذلك على الرغم من أنها تمثل أنواع مختلفة من الكيانات (انظر الجدول 2-2 على سبيل المثال).

فالكتالوجات، والأدلة، والبوابات هي موارد مركزية تفصل وتربط حاويات البيانات والمحفوظات الأرشيفية التي تحتفظ بها المؤسسات على المستوى الفردي. وقد توفرّ قوائم مفصلة إلى حدّ ما عن مجموعات البيانات التي تضمّها، كما يمكن أن تكون بمثابة مجاميع للبيانات الوصفية، ولكن لا يستضيف الموقع نفسه البيانات (O'Carroll et al. 2013). أما مستودع البيانات أو مركز تبادل المعلومات فهو مكان مشترك لتخزين حاويات البيانات والمحفوظات الأرشيفية والوصول إليها. وقد توفر هذه المستودعات والمراكز بعض خدمات البيانات من قبيل البحث والاسترجاع، وإدارة البيانات ومعالجتها، ولكن عملية إنتاج كل حاوية أو أرشيف تتم بشكل مستقل، ولذا فهي ربما تفتقر إلى صيغ ومعايير وبيانات وصفية وسياسات مشتركة للبيانات. ومع ذلك، يسعى مستودع البيانات لضمان التزام كل أرشيف أو حاوية بمجموعة من المواصفات المطلوبة ويقوم بالمراجعة، والتدقيق، والتصديق على منح الشهادات لضمان سلامة البيانات وتعزيز الثقة بين المستخدمين (Dasish 2012). وعرّف O'Carroll et al. (2013) نوعين من المستودعات. مستودعات الموقع الواحد التي تستضيف كافة مجموعات البيانات في موقع واحد، ويمكن الوصول إليها من خلال واجهة على شبكة الإنترنت، على الرغم من إمكانية احتفاظها بنسخ احتياطية أو مواقع مثيلة في أماكن متعددة. والمستودعات المتعددة المواقع التي تستضيف البيانات ضمن هيكل موحد يسمح بالوصول إلى مجموعات البيانات القائمة في عدد من المواقع المؤسسية.

فالمستودع الرقمي الموثوق (TDR - Trusted Digital Repository) هو مستودع مرخص بموقع واحد أو متعدد المواقع تتمثل مهمته في توفير وصول آمن وموثوق به للبيانات التي يديرها على المدى البعيد (RLG and OCLC 2002). ومن أجل التأهل

كمستودع موثوق، فصلت (5: 2002) RLG and OCLC الخصائص المطلوب توفرها في البنية التحتية لهذه المستودعات وهي كالتالي:

- تقبل المسؤولية عن الصيانة طويلة الأمد للموارد الرقمية بالنيابة عن المودعين، ولما فيه مصلحة المستخدمين الحاليين والمستقبليين، والمستهلكين أو المجتمعات المعنية.
- امتلاك نظام مؤسسي لا يدعم فقط استمرارية نظام المحافظة على البيانات على المدى البعيد، ولكن يحافظ أيضاً على المعلومات الرقمية التي تقع ضمن نطاق مسؤوليته.
- توضيح المسؤولية المالية وضمان الاستدامة.
- تصميم أنظمتها وفقاً للاتفاقيات والمعايير المقبولة والمعتمدة لضمان الإدارة المستمرة، والوصول، وأمن المواد المودعة داخلها.
- وضع منهجيات لتقييم المستودع، وذلك للتأكد من تلبيةه للتوقعات المجتمع من حيث الموثوقية.
- إمكانية الاعتماد على البنية التحتية الخاصة بها وضمان تحمّل مسؤولياتها على المدى الطويل تجاه المودعين والمستخدمين بشكل علني وصریح.
- امتلاك السياسات، والممارسات، والأداء الذي يمكن تدقيقه وقياسه.

ولكي يمكن لمستودع ما الحصول على هذه المتطلبات، أكدت RLG and OCLC (13: 2002) أنه ينبغي أن يكون متوافقاً مع النموذج المرجعي لنظام المعلومات الأرشيفي المفتوح (CCSDS 2012)-(OAIS Open Archival Information System) وأن يمتلك المسؤولية الإدارية، والسلامة التنظيمية، والاستدامة المالية، والملائمة التقنية والإجرائية، وأمن النظام، والمساءلة الإجرائية، وتساعد كل هذه الخصائص معاً على ضمان وبناء الثقة في البيانات ومستودعها.

وتُعد البنية التحتية السيبرانية أكثر من مجرد مجموعة من المحفوظات الأرشيفية الموثوقة فهي تُمكن من تبادل البيانات والتشارك في موارد البنية التحتية باهظة الثمن. وتتألف البنية التحتية السيبرانية من مجموعة متكاملة من المعدات المتخصصة والتقنيات

الشبكية (تشمل المراصد والمرافق البحثية التجريبية، وشبكات الاستشعار، وما إلى ذلك)، والخدمات والأدوات لبرمجيات التشغيل المتبادل والبرمجيات الوسيطة، والخدمات المشتركة (المتعلقة بإدارة البيانات ومعالجتها)، وأدوات التحليل مثل التمثيل البصري للبيانات (على سبيل المثال، تطبيقات الرسوم البيانية ورسم الخرائط)، والسياسات المشتركة (فيما يتعلق بحق الوصول للبيانات واستخدامها وحقوق الملكية الفكرية، وغيرها) التي تمكن من نشر البيانات وتوزيعها، وربطها بعضها مع بعض وتحليلها (هيئة البنية التحتية السيبرانية 2007 Cyberinfrastructure Council). كما تسعى إلى تعزيز تنمية المجتمع الثقافي الذي يدعم التعاون المباشر بين الأفراد متجاوزاً الحدود الجغرافية والمؤسسية (2007 Cyberinfrastructure Council). وبينما تستخدم أحياناً لوصف البنية التحتية التي تمكن عمل مستودع بيانات موحد، فإنها تستخدم هنا للدلالة على البنية التحتية للبيانات التي تتشارك في مواصفات تقنية مشتركة متعلقة بالنماذج، والمعايير، والبروتوكولات. وبعبارة أخرى، هناك قواعد صارمة متعلقة بتطبيق المعايير على البيانات والالتزام بها داخل البنية التحتية. وتشمل مثل هذه البنى التحتية السيبرانية تلك التي تنفذها الوكالات الإحصائية الوطنية والبنى التحتية المكانية الوطنية للبيانات، إضافة إلى البنى التحتية العلمية واسعة النطاق التي تتعلق بأبحاث الجينات، وعلم الفلك، والأرصاد الجوية وغيرها، والتي تتطلب تخزين كافة البيانات ومشاركتها مع الامتثال لمعايير محددة من أجل تحقيق أقصى قدر من قابلية التشغيل البيئي للبيانات وضمان جودة البيانات، والإخلاص والنزاهة التي تعزز الثقة. وهذه الأخيرة لها أهمية خاصة في مجال العلوم، والسياسات التي تقوم على أساسها مثل تتبّع تغيرات المناخ وتنبؤاته، حيث تعد سلامة البيانات أمراً بالغ الأهمية في المناقشات الجارية حالياً.

وإضافة إلى بوابات البيانات، والمستودعات، والبنى التحتية، تم استحداث عدد من المنظمات والهيكل المؤسسية التي تقدم المشورة بشأن أفضل الممارسات في مجال الحفظ الرقمي، وإتاحة ومشاركة البيانات، والتحقق من صحة العمل والمعايير المتعلقة بالأنواع المختلفة من بنوك البيانات، والدفع باتجاه تنميتها وتطويرها، وتنسيق الجهود المختلفة بشأن ذلك (انظر الجدول 2-3). وتؤدي هذه المنظمات والهيكل المؤسسية عملها من خلال أنشطة عدة مثل تقديم البرامج التدريبية والتعليمية، واستضافة الندوات والمؤتمرات، وإنتاج

الكتيبات والأدلة، وتوفير التمويل اللازم لمبادرات محدّدة. وتكتسب هذه المنظمات أهمية نظراً لإسهاماتها في خلق مجتمعات الممارسة بين مختصي البيانات، وتسهيل المفاوضات بين الهيئات وتوسيع نطاق الجهود المؤسسية، ووضع مبادئ توجيهية ومعايير مشتركة. ويركز بقية هذا القسم على مستودعات البيانات البحثية والبنى التحتية، ويفصّل في الأسس المنطقية لها والتحديات التي تواجهها، كما ستتم مناقشة البنى التحتية الخاصة والمتعلقة بشركات البيانات التجارية في وقت لاحق في هذا الفصل.

جدول رقم (٢-٢)

أنواع وأمثلة البنى التحتية للبيانات

النوع	الأمثلة
حاوية البيانات Data Holding	يُطلق على مجموعات غير رسمية من ملفات البيانات التي تكون على جهاز حاسب شخصي، ويسمى نظام النسخ الاحتياطي في بعض الأوقات.
محفوف البيانات (أرشيف البيانات) Data Archive	أرشيف لوني لبيانات الصور (LONI Image Data Archive-IDA) هو أرشيف آمن لصور الجهاز العصبي والبيانات السريرية الطبية ذات الصلة بالمشاريع البحثية في جميع أنحاء القارة الأمريكية الشمالية والقارة الأوروبية. أرشيف الموسيقى العالمية الموجود في مكتبة موسيقا لوب (Loeb) في جامعة هارفارد الذي يقدّم تخزيناً آمناً ووصولاً متاحاً لمحتوى الأرشيف والتسجيلات التجارية الصوتية والمرئية للموسيقى من جميع أنحاء العالم.
الكتالوجات، الأدلة، البوابات، البيانات الوصفية، والمجمعون Catalogues , Directories , Portals , Metadata , Aggregators	تمكّن خدمة البيانات الوطنية الأسترالية المستخدمين من تصفح أكثر من أربعمئة ألف من مجموعات البيانات وأكثر من ثلاثمئة ألف مشروع تقوم بإنشاء مجموعات البيانات. تتيح المكتبة الرقمية العامة الأمريكية الوصول إلى مليونين وخمسمائة ألف سجل من سجلات البيانات الوصفية من مكاتب، ومحفوظات، ومتاحف مقرها الولايات المتحدة. يوروبيانا هي بوابة إلكترونية ومجمّع للبيانات الوصفية لأكثر من مئة وثلاثين من المكتبات والمتاحف والمكتبات الرقمية للوسائط المتعددة في جميع أنحاء أوروبا. محفوظات بيانات المجلس الأوروبي للعلوم الاجتماعية (CESSDA) هي بوابة إلكترونية ومجمّع للبيانات الوصفية منذ ما يزيد عن ثلاث وعشرين سنة لمستودعات البيانات الأوروبية الإحصائية للعلوم الاجتماعية التي تضم أكثر من ثمانية وعشرين ألف من مجموعات البيانات.

<p>خدمات أرشفة وشبكات البيانات (DANS) هو الأرشيف الوطني الهولندي لبيانات البحوث الرقمية. إضافة للأرشفة، وتقدّم المبادئ التوجيهية للتدريب والاعتماد.</p> <p>أرشيف بيانات المملكة المتحدة (UKDA) هو مستودع بيانات الموقع الواحد لبيانات بحوث العلوم الاجتماعية والإنسانية الممولة من قبل مجلس البحوث الاقتصادية والاجتماعية (ESRC)، ولجنة نظم المعلومات المشتركة (JISC)، وجامعة إسيكس (Essex).</p>	<p>مستودعات بيانات الموقع الواحد</p> <p>Single-site Repositories</p>
<p>المستودع الرقمي الأيرلندي (DRI) هو مستودع رقمي موثوق للبيانات المعاصرة والتاريخية، الاجتماعية والثقافية، وتقوم عليه المؤسسات الأيرلندية، ويوفّر إتاحة الوصول، والحفاظ على البيانات للجهات المعنية ذات المصلحة واكتشافها.</p> <p>أوبونير (Openaire) هو مستودع رقمي متعدد المواقع للحصول على البيانات العلمية المفتوحة في منطقة البحث الأوروبية المكوّنة من واحد وأربعين منظمة وثلاثة وثلاثين بلداً أوروبياً.</p> <p>المكتبة الرقمية لتكساس (TDL) هي مجموعة من مؤسسات التعليم العالي التي توفر منصة وخدمات مشتركة لتوفير الوصول المفتوح للبيانات والحفاظ عليها.</p>	<p>مستودعات بيانات متعددة المواقع</p> <p>Multi-site Repositories</p>
<p>توفّر البنية التحتية للبيانات الجغرافية المكانية الكندية (CGDI) الوصول الموحد، والمتسق، والدقيق، والقابل للتشغيل المتبادل للبيانات المكانية عبر القطاعات والنطاقات الكندية.</p>	<p>البنية التحتية السيبرانية</p> <p>Cyber- Infrastructures</p>

المصدر : مقتبس بتصريف من O'Carroll et al.(2013).

جدول (٣-٢)

مجموعة منتقاة من مؤسسات تقديم المشورة، والضغط لكسب التأييد، والتنسيق للمحافظة على البيانات وإتاحتها للوصول والمشاركة وإعادة الاستخدام في العلوم الاجتماعية والإنسانية

المؤسسة بالإنجليزية	الموقع الإلكتروني للمؤسسة	المؤسسة بالعربية
The Federation of All European Academies (ALLEA)	www.allea.org	اتحاد جميع الأكاديميات الأوروبية
Ariadue	www.ariadne-eu.org	آرياديو
British and Irish Sound Archives ((BISA	www.bisa-web.org	المحفوظات الصوتية البريطانية والإيرلندية
Consortium of European Social Science Data Archives	www.cessda.org	الاتحاد الأوروبي لمحفوظات البيانات للعلوم الاجتماعية
(Coalition for Networked Information (CNI	www.cni.org	التحالف من أجل معلومات مترابطة
Databib	www.databib.org	داتابيب
DataONE	dataone.org	داتاون
Common Language Resources and (Technology Infrastructure (CLARIN	www.clarin.eu	موارد اللغة المشتركة والبنية التحتية التقنية
Committee on Data for Science and Technology of the International Council for Science	www.codata.org	لجنة البيانات للعلوم والتقنية للمجلس الدولي للعلوم
Digital Services Infrastructure for Social (Sciences and Humanities (DASISH	www.dasish.eu	الخدمات الرقمية للبنية التحتية للعلوم الاجتماعية والإنسانية
(Digital Curation Centre (DCC	www.dcc.ac.uk	مركز كرايشن الرقمي
(Digital Library Federation (DLE	www.diglib.org	اتحاد المكتبة الرقمية
(Digital Preservation Coalition (DPC	www.dpconline.org	تحالف الحفظ الرقمي
(Digital Production Partnership (DPP	www.digitalproductionpartnership.co.uk	الشراكة للإنتاج الرقمي

International Federation of Television (Archives (FIAT/IFTA	www.fiatifta.org	الاتحاد الدولي لمحفوظات التلفزيون
International Association for Social Science Information Services and (Technology (IASSIST	www.iassistdata.org	الرابطة الدولية لخدمات المعلومات والتكنولوجيا للعلوم الاجتماعية
IcarusNet	www.icarusweb.arhiv.hr	شبكة أكاروس
International Association of Sound and Audiovisual Archives	www.iasa-web.org	الرابطة الدولية للمحفوظات الصوتية والسمعية - البصرية
Inter-university Consortium for (Political and Social Research (ICPSR	www.icpsr.umich.edu/ icpsrweb/ICPSR	اتحاد الجامعات للبحوث السياسية والاجتماعية
Network of Expertise in Long-term (Storage of Digital Resources (NESTOR	www.langzeitarchivierung.de/ /Subsites/nestor/DE/Home	شبكة الخبرة في التخزين على المدى الطويل للموارد الرقمية
Open Planets Foundation	www.openplanetsfoundation. org	مؤسسة الكواكب المفتوحة
PrestoCentre	www.PrestoCentre.eu	مركز بريستو
(Research Data Alliance (RDA	www.rd-alliance.org	ائتلاف بيانات البحوث
(Text Encoding Initiative (TEI	www.tei-c.org	مبادرة ترميز النصوص

المصدر : مقتبس بتصرف من (O'Carroll et al. (2013).

الأسس المنطقية للبنى التحتية للبيانات البحثية:

بالنظر إلى أن بناء البنى التحتية للبيانات هي ممارسة ليست باليسيرة، ومكلفة من حيث الموارد والوقت الذي يستلزمه إنشاؤها والمحافظة عليها بشكل مستمر، فينبغي أن تكون هناك أسباب معتبرة لإجراء مثل هذا المسعى والمضي فيه. ويستعرض الجدول (٤-٢) قائمة لأكثر من ثلاثين من هذه الأسباب مقسمة على ثلاثة أبعاد - مباشرة / وغير مباشرة، وفي المدى القريب / وعلى المدى البعيد، والعامّة / والخاصة بحسب تعريف (Beagrie et al. (2010.

ويمكن أن تترجم هذه الأسباب تقريبياً إلى المكاسب العلمية والمالية، والتأثير التراكمي للفوائد، والمستفيدين من هذه البنى التحتية.

وتتمحور المناقشات والحجج العلمية لتخزين، ومشاركة، وتوسيع البيانات داخل البنى التحتية للبيانات حول وعود من الاكتشافات والابتكارات الجديدة من خلال الجمع بين مجموعات البيانات والتوظيف الجماعي للعقول المشاركة. فمجموعات البيانات الفردية ذات قيمة في حد ذاتها، ولكن عند ضمها جنباً إلى جنب مع مجموعات بيانات أخرى أو تفحصها بطرق وأساليب جديدة، فإن رؤى وأفكار جديدة من المحتمل أن تظهر وأسئلة جديدة من الممكن أن تتم الإجابة عليها (Borgman 2007)، فبسبب الطبيعة التراكمية، من خلال الجمع بين مجموعات البيانات، يمكن الزعم بتسارع وتيرة بناء المعرفة (Lauriault et al. 2007). وعلاوة على ذلك، يصبح من الممكن من خلال الحفاظ على البيانات تتبع الاتجاهات والأمط مع مرور الوقت، وكلما ازدادت فترة التسجيل للبيانات، زادت معها القدرة على بناء النماذج وأنظمة المحاكاة ومن ثم زيادة الثقة في الاستنتاجات التي يتم التوصل إليها (Lauriault et al. 2007). لذا، فإنه ومع مرور الوقت، تزداد القيمة التراكمية لمستودعات البيانات والبنى التحتية بحيث تصبح البيانات متاحة على نطاق واسع وبشكل أكثر يسراً، وذلك من حيث النطاق والصلاحية الزمانية.

ومن المرجح أن تعمل هذه المشاركة الاستراتيجية إلى بث روح التعاون بين الباحثين والفرق في التخصصات الجديدة، وتعزيز المهارات بصورة أكبر من خلال إمكانية الوصول إلى أنواع جديدة من البيانات (Borgman 2007). وعلاوة على ذلك، فإن تبادل المعلومات واعتماد معايير وبروتوكولات وسياسات البنية التحتية من شأنه أن يرفع من مستوى جودة البيانات وأن يفسح المجال لأطراف أخرى لإنتاج بيانات مُشتقة والتحقق من البيانات المتاحة، ومن ثم زيادة سلامة البيانات (Lauriault et al. 2007). كما تعمل البنى التحتية أيضاً على إتاحة بيانات عالية الجودة في التعليم وتحسين المصادر التربوية والتعليمية لتدريب الجيل القادم من الباحثين. كما أن البنى التحتية للبيانات هي الوسيلة الوحيدة لإدارة البيانات والمشاريع بكفاءة وفعالية في الحالات التي تعتمد فيها المشاريع العلمية بشكل كبير على أدوات تقنية المعلومات والاتصالات الباهظة الثمن، والتي ينتج عنها كميات هائلة من

البيانات، وكذلك في المشاريع التي تضم فرقاً بأعداد كبيرة من الباحثين المنتشرين في جميع أنحاء العالم لتحليل هذه البيانات، لذا فإن البنى التحتية للبيانات تمثل الوسيلة الوحيدة الفاعلة والفعالة لإدارة مثل هذه النوعية من البيانات والمشاريع (Borgman 2007).

وتتمحور الفوائد المالية للبنى التحتية للبيانات حول الموازين الاقتصادية التي تنشأ من خلال تقاسم الموارد ومشاركتها وتجنب تكرارها، والآثار المترتبة على إعادة الاستخدام للبيانات المكلفة، والثروة الناتجة من خلال الاكتشافات الجديدة، وإنتاج مجتمعات أكثر كفاءة. فالبحث عملية مكلفة في العادة، حيث تنفق وكالات التمويل المختلفة بشكل جماعي مليارات الدولارات سنوياً لتمويل النشاط البحثي، وينطوي جزء كبير من هذا النشاط على ازدواجية غير ضرورية في جهود إنتاج البيانات، ومعالجتها، وإدارتها، وتحليلها. وبدلاً من إنشاء عدد كبير من المحفوظات الأرشيفية الخاصة، فإنه من المنطقي إنشاء عدد أقل من البنى التحتية المخصصة التي تقوم بتقديم وتحضير البيانات الأساسية على صورة موحدة، وتنتج فعالية كبيرة في الجهد، وتُمكن كذلك وصولاً أوسع للبيانات للأفراد الباحثين والمؤسسات كون تكاليف الدخول إلى حقل أو مجال بحثي في العادة باهظة (Fry et al. 2008). وإضافة إلى ذلك، ستصبح عمليات ونتائج المشاريع البحثية أكثر شفافية ويُمكن الحكم على مدى تمثيلها القيمة مقابل المال. فضلاً عن الحد من الهدر والحفاظ على الثمار الناتجة من هذه الجهود وتقاسمها، فإنها ستعمل على تعظيم العائد من الاستثمار من خلال تمكين أكبر قدر من المكاسب التي يمكن الحصول عليها من البيانات (Lauriault et al. 2007). فعلى سبيل المثال، قد ينتج عن إتاحة البيانات على نطاق واسع صناعات جديدة قائمة على المعرفة وتكون شركات منبثقة ومنفصلة من خلال تسويق البحوث الممولة من القطاع العام (ائتلاف العلوم 2013 Science Coalition)، كما يمكن للبيانات ضمن البنى التحتية للبيانات أن تؤدي إلى طرق أكثر فعالية في الحوكمة والإدارة المجتمعية. على سبيل المثال، توفر العديد من البنى التحتية للبيانات المكانية الوطنية، إضافة إلى دعمها للبحوث المكانية، أدوات فعالة للإدارة والحكومة الإقليمية، وتمكّن من الاستعداد للطوارئ والاستجابة للكوارث بصورة أكثر فعالية، وتسهيل التخطيط بين الدول.

وبحسب ما تم إيضاحه من خلال النقاش حتى الآن، فإن الفوائد المحتملة المكتسبة من البنى التحتية للبيانات هي فوائد ممتدة وغير محصورة بالقطاع العام أو الخاص (Beagrie et al. 2010). وقد ظهر جلياً في الواقع أن استراتيجية مشاركة البيانات وتقاسمها تعود بالفائدة على جميع الأطراف - الباحثين، المؤسسات، الممولين، والمؤسسات العامة والخاصة. فعلى مدى العقدين الماضيين، وبالنظر إلى المكاسب المتوقعة من مشاركة البيانات وتبادلها، عملت الوكالات البحثية للحكومات الوطنية والهيئات الإقليمية مثل الاتحاد الأوروبي، جنباً إلى جنب مع المنظمات الخيرية، على الاستثمار بنطاق واسع في تمويل مجموعة كبيرة من المبادرات المختلفة. ومع ذلك، وكما هو مبين أدناه، فهناك مقاومة لمثل هذه المشاريع، وعلى الأخص من قبل الباحثين أنفسهم، ومن الواضح أنه على الرغم من أن هناك فوائد لا شك فيها، فهناك أيضاً تكاليف مرتبطة بإنشاء البنى التحتية للبيانات.

تحديات بناء البنى التحتية للبيانات:

على الرغم من الفوائد المتوقعة والمُدركة من البنى التحتية للبيانات، إلا أنها لم تلق ترحيباً عالمياً، كما أنها أثبتت صعوبة بنائها وتنفيذها، فهناك مجموعتين رئيسيتين من القضايا الفاعلة - التقنية والاجتماعية/ السياسية - التي تعوق تطوير البنى التحتية للبيانات وتنميتها، وستواصل القيام بذلك لبعض الوقت.

جدول رقم (٤-٢)

فوائد مستودعات البيانات والبنى التحتية للبيانات

الفوائد المباشرة	الفوائد غير المباشرة (تجنب التكاليف)
فرص جديدة للبحث.	عدم تكرار إنشاء البيانات / الازدواجية في البيانات
التواصل العلمي / إتاحة الوصول إلى البيانات.	عدم فقدان فرص البحث المستقبلية.
إعادة تحديد الأهداف وإعادة استخدام البيانات.	انخفاض تكاليف المحافظة على البيانات في المستقبل.
زيادة الإنتاجية البحثية	إعادة توظيف البيانات للمهتمين الجدد.
تحفيز أنواع جديدة من التواصل / التعاون.	إعادة توظيف المنهجيات المستخدمة.
إتاحة البيانات للتعليم ومشاريع الطلاب.	الاستخدام من قبل مهتمين جدد.

نقل المعرفة للصناعة.	حماية العائد على الاستثمار المبكر.
تحسين قاعدة المهارات.	الأدوات والمعايير يحتمل أن تزيد من مستوى جودة البيانات.
زيادة الإنتاجية / النمو الاقتصادي.	تقليل الاستفسارات العشوائية المتعلقة بالبيانات
التثبت من البحث / سلامة البحث	
الوفاء بالالتزامات	
فوائد على المدى القريب	الفوائد على المدى الطويل
قيمة للباحث والطلاب الحاليين.	تأمين القيمة للباحثين والطلاب المستقبليين.
عدم فقدان البيانات بتغير الباحثين.	إضافة القيمة مع مرور الوقت وتمو المجموعة وتطورها للحد الأقصى.
توسيع إتاحة الوصول حيث التكاليف الباهظة للباحثين / المؤسسات.	زيادة سرعة البحث والوقت اللازم لإدراك آثاره.
إعادة الاستخدام على المدى القصير للبيانات المنسقة جيداً.	يحفز الأسئلة البحثية الجديدة، وبخاصة المتعلقة بالبيانات المرتبطة والمشتقة.
التخزين الآمن لبيانات البحوث المكثفة.	
تعزيز توافر البيانات للنشر والمشاركة.	
الفوائد الخاصة	الفوائد العامة
فوائد للرعاة / الممولين من البحث / المحفوظات.	مدخل للبحوث المستقبلية.
فوائد للباحثين والمؤسسات.	تحفيز البحوث الجديدة.
الوفاء بالالتزامات الخاصة بالمنح.	تحفيز الشركات الجديدة وتوظيف المهارات العالية.
زيادة الوضوح / الاقتباس.	الشفافية في تمويل البحوث.
الاتجار بالبحوث	

المصدر: Beagrie et al. 2010; Fry et al 2008.

إن إنشاء بنية تحتية عاملة وقابلة للتشغيل المتبادل للبيانات ليس تحدياً هيناً، خاصة إذا كان النظام موثقاً ويقدم مستويات مختلفة من إتاحة الوصول اعتماداً على تعريف المستخدم و ضمانات فيما يتعلق بمصدر البيانات وسلامتها، ونزاهتها، وصلاحيتها، فذلك يتطلب تعاوناً مكثفاً وواسع النطاق والموائمة بين منتجي البيانات ومالكي المحفوظات الأرشيفية فيما يتعلق بالتالي:

- إجراءات إنتاج البيانات.
- نماذج البيانات ومعايير البيانات.
- البيانات الوصفية والتوثيق.
- إجراءات تحضير وتنقية وإدخال البيانات.
- جودة البيانات و ضمانها.
- سياسات الحفاظ على البيانات، والنسخ الاحتياطي، والتدقيق عليها.
- البرمجيات والمعدات.
- أمن وحماية المعلومات.
- سياسات الوصول، ومنح التراخيص، والاستخدام، وإعادة الاستخدام، والخصوصية، والجوانب الأخلاقية.
- سياسة الملكية، وحقوق الطبع والنشر، وحقوق الملكية الفكرية.
- الترتيبات الإدارية وآليات تنظيم الإدارة والحوكمة.
- تمويل البنية التحتية وخدماتها وإدارتها.

(Borgmann 2007; Data Seal of Approval 2010; O'Carroll and Webb 2012)

فالبنية التحتية للبيانات لن تنجح إلا إذا كان لديها رؤية طويلة الأجل، وهناك التزام لإدارة المحتوى نيابة عن المودعين والمستخدمين وتحمل مسؤوليتها، مع التطبيق التام لجميع العوامل المذكورة أعلاه بشكل فعال (O'Carroll and Webb 2012).

فالقضايا التي تمت الإشارة إليها ليست مجرد قضايا تقنية، ولكنها أيضاً اجتماعية وسياسية، ذلك لأنها تتطلب التفاوض والاتفاق على طريقة تصميمها واختيارها ومتطلبات إدارتها القائمة بين العديد من الأطراف - الممولين، والمصممين، والمبدعين، والمديرين والمودعين، ومجموعات الضغط والتوجيه، والمساهمين وأصحاب المصالح الخاصة الأخرى - في سياق مختلف التعليمات، واللوائح، والقوانين المتعلقة بإنتاج، وتخزين، ومشاركة البيانات. ومن ثم، فهي تتشكل بواسطة الأجنات الشخصية والشخصيات، والقيم المؤسسية، والالتزامات والوصايا، والثقافات، والهياكل، والعمليات، والتمويل، والقضايا الحكومية، والتبعيات الإجرائية حيث أنه مجرد أن يبدأ المشروع في مسار معين، فسيكون من الصعب غالباً توجيهه في مسار مختلف كلياً عنه (Lauriault et al. 2007; Feenberg in Lauriault 2012). كما يمكن أن يكون تجميع ومواءمة المحفوظات الأرشيفية المعتمدة على مسارات تقنية مختلفة عملية في غاية الصعوبة، وطويلة الأمد، ومكلفة لأنها تتطلب من تلك المشاريع إجراء تغييرات لتحديث بياناتها المتقدمة لتتوافق مع المواصفات المتفق عليها في المواثيق (Lauriault 2012). وهذا هو السبب في اعتماد العديد من المشاريع الأرشيفية على بنية مستودع البيانات بدلاً من السعي لتصبح بنية تحتية إلكترونية/ سيرانية.

ويأتي بعد الصعوبات المتعلقة بإنشاء هياكل تبادل البيانات، إقناع أصحاب البيانات أنفسهم بتبادل سلحتهم الثمينة (بياناتهم)، فهناك مبدأ أساسي في البحوث الأكاديمية وهو إتاحة الوصول المفتوح؛ بمعنى، أن جميع جوانب إنتاج المعرفة ينبغي أن تكون متاحة بحرية للآخرين لتفحصها ونقدها. وهذا المبدأ لم يُعمل به على النحو الأمثل أبداً، إذ تختفي الأبحاث الأكاديمية في الغالب خلف حواجز الأجور ويتردد الباحثون في مشاركة بياناتهم وتبادلها مع الآخرين. وكما أشار Borgman (2007)، فإن مشاركة البيانات وتبادلها شائع في عدد قليل من التخصصات مثل علم الفلك (Astronomy)، علم الجينات (Genomics)، وعلم البيانات العمرانية بهيئتها الرقمية (Geomatics) والتي تعتمد على فرق علمية كبيرة وموزعة جغرافياً؛ أما في التخصصات الأخرى فيجري في أحيان نادرة مشاركة البيانات وقد لا تجري على الإطلاق. وقد خلصت Borgman إلى أن «السرّ الصغير القدر» وراء تشجيع مشاركة البيانات، وتبادلها هو عدم التبادل لكثير من البيانات في الواقع (Borgman 2012: 1059). وذلك لأن هناك عدداً من المثبطات التي تحول دون تبادل البيانات ومنها:

- عدم وجود مكافآت للقيام بذلك.
- الجهد اللازم لإعداد وأرشفة البيانات.
- نقص الخبرة والموارد والأدوات اللازمة لأرشفة البيانات.
- مخاوف من القدرة على استخراج قيمة من البيانات قبل الغير، وذلك من حيث الأوراق العلمية وبراءات الاختراع، ونظراً للجهد المستثمر في إنتاج البيانات.
- المخاوف بشأن الكيفية التي سيتم بها استخدام البيانات، خاصة إذا كانت تتعلق بأشخاص، أو كيف يمكن أن يساء التعامل معها أو يساء تفسيرها.
- المخاوف بشأن الاستفسارات بخصوص إنتاج البيانات والطلبات التي من شأنها أن تستلزم أعمالاً إضافية.
- مخاوف من القضايا المتعلقة بالكشف عن البيانات ونتائج البحوث التي يمكن تقويضها من خلال تفسيرات بديلة لنفس البيانات.
- قضايا الملكية الفكرية.
- الخوف أنه لن يتم استخدام البيانات، وبالتالي تشكّل الأرشفة جهداً ضائعاً. (Borgman 2007, 2012; Strasser 2013).

فمن الواضح أن ضمان البيانات من خلال أرشفتها لإعادة استخدامها مستقبلاً سوف يتطلب أكثر من مجرد إنشاء مستودعات البيانات والبنى التحتية لها؛ كون ذلك يتطلب تغييراً ثقافياً في الممارسات البحثية. ويتكشف هذا التغيير للعيان باستخدام إستراتيجية العصا والجزرة، فمن ناحية بدأ استخدام الحوافز لتشجيع الباحثين على أيداع وتبادل البيانات، مثل تشجيع الاقتباس والإسناد عند إعادة استخدام البيانات (Borgman 2012)، وتوفير الأموال الكافية لأرشفة البيانات ضمن جوائز المنح. ومن ناحية أخرى، بدأت وكالات الأبحاث في إجبار الباحثين على إيداع وتبادل البيانات، مع مراعاة الجوانب الأخلاقية وحقوق الملكية الفكرية، كشرط لتمويل البحوث.

وسطاء البيانات والأسواق (Data Brokers and Markets):

يمتلك القطاع الخاص دافعاً مختلفاً جداً لإنتاج وإعادة استخدام البيانات عن ذلك الذي عند الأكاديميين، فالبنى التحتية للبيانات البحثية تسعى لتحقيق الصالح العام من خلال إنشاء مشاعات للبيانات لتسهيل وتشجيع تبادل البيانات وإعادة استخدامها. وفي المقابل، يلتقط وسطاء البيانات (يسمون أحياناً مجمعي أو مختزلي أو بائعي البيانات) البيانات ويجمعونها معاً ويعيدون تعبئتها (تحزيمها) في البنى التحتية للبيانات التي يملكها القطاع الخاص لعرضها للإيجار (للاستخدام لمرة واحدة أو لاستخدامها تحت شروط الترخيص) أو إعادة البيع على أسس ربحية. وتشمل المنتجات قوائم العملاء/الزبائن المحتملين الذين يستوفون معايير معينة (هما في ذلك تفاصيل مثل الأسماء، العناوين، أرقام الهاتف، وعناوين البريد الإلكتروني، وكذلك معلومات مثل «الجنس، العمر، العرق، وجود الأطفال، الدخل، قيمة المنزل، ملكية بطاقة الائتمان، الوضع الائتماني، أنماط الشراء، والهوايات»؛ (CIPPIC 2006: ii)، والتدقيق على الخلفية الأمنية، ومجموعة من منتجات البيانات المشتقة التي أضاف لها الوسطاء قيمة من خلال التكامل، ومنتجات تحليل البيانات التي تستخدم في معظم الأحوال لحملات الدعاية والتسويق المصغرة، وتقييم الجدارة الائتمانية والتصنيف الاجتماعي للأفراد، وتوفير خدمات البحث عن المفقودين، وتقديم تحليلات الأعمال المفصلة (CIPPIC 2006). وفي حالة القوائم، فهي في معظمها مستأجرة وفي كثير من الحالات لا يستلم المستأجر القائمة، حيث يعمل مكتب خدمة كطرف ثالث على إعداد وإرسال البريد بالنيابة عنهم (CIPPIC 2006). وبالمثل، مع تحليلات البيانات، حيث يتلقى العملاء نتائج تحليل البيانات، وليس ذات البيانات الأساسية.

إن عمليات اختزال البيانات وإعادة بيعها، وما يرتبط بها من تحليل للبيانات وخدمات ذات قيمة مضافة، هي صناعة بمليارات الدولارات، مع كميات هائلة من البيانات والمعلومات المستمدة التي تُوجر، وتُباع، وتُشتري بشكل يومي عبر مجموعة متنوعة من الأسواق - تجارة التجزئة، والمالية، والصحة، والسياحة، والخدمات اللوجستية، والأعمال الذكية، والعقارات، والأمن الخاص، والافتراع السياسي، وإلى آخره. وتتعلق هذه البيانات بجميع جوانب الحياة اليومية وتشمل الإدارة العامة، والاتصالات، واستهلاك السلع ووسائل الإعلام، والسفر،

والترفيه، والجريمة، وتفاعلات وسائل التواصل الاجتماعية، وإلى آخره، فوسطاء البيانات المتخصصين هم متواجدون منذ فترة طويلة، يجمعون البيانات من اشتراكات وسائل الإعلام (على سبيل المثال، الصحف، والمجلات)، وتجار التجزئة عبر طلبات البريد، واستطلاعات الرأي، والدراسات الاستقصائية، ووكالات السفر، والمؤتمرات، والمسابقات، وتسجيلات وضمانات المنتجات، وشركات معالجة الدفع والسداد، والسجلات الحكومية، وإلى آخره (CIPPIC 2006). وفي السنوات الأخيرة، وخاصة منذ ظهور الإنترنت، توسعت هذه الصناعة بشكل سريع، مع التنوع في البيانات التي تم التقاطها واختزالها، والمنتجات والخدمات التي تقدمها، وغالباً ما تجمع معا على حدّ سواء البيانات الصغيرة والكبيرة.

فبيع البيانات للوسطاء أصبح يمثل تدفقاً مهماً للدخل للعديد من الشركات، فعلى سبيل المثال، يبيع تجار التجزئة، في كثير من الأحيان، البيانات المتعلقة بالعمليات مثل تفاصيل بطاقة الائتمان، ومشترتي العملاء وبرامج الولاء الخاصة بالمتجر، وإدارة علاقات العملاء، ومعلومات الاشتراك. كما تبيع المواقع الإلكترونية بيانات النقر (الاختيارات) المتعلقة بكيفية تصفح شخص ما لموقع على شبكة الإنترنت والوقت الذي يقضيه على مختلف الصفحات. وبالمثل تستخلص شركات وسائط الإعلام، مثل محطات الصحف والإذاعة والتلفزيون، البيانات الواردة في محتواها (على سبيل المثال، الأخبار، والإعلانات). وتستخلص شركات وسائط التواصل الاجتماعي البيانات الوصفية وبيانات المحتوى لمستخدميها، وتستخدم هذه البيانات لتحديد المواصفات الأساسية للمستخدمين ومنتجات البيانات الخاصة بهم، أو بيع البيانات لوسطاء البيانات. على سبيل المثال، تستخدم شركة فيسبوك الملفات التعريفية للمليارات من مستخدميها، وشبكاتهم الاجتماعية، والمحتوى الذي يقومون بتحميله (المفضلات، والتعليقات، والصور، والفيديوهات، وما إلى ذلك) لتمكين مجموعة من المنتجات الإعلانية مثل الإعلانات الموجهة للجماهير المتماثلة (Lookalike Audiences)، والجماهير المخصصة المدارة (Managed Custom Audiences)، وفئات الشريك (Partner Categories) وذلك بالشراكة مع وسطاء البيانات الكبيرة والمسوقين مثل داتا لوجيكس (Datalogix)، وأبسيلون (Epsilon)، وأكسيوم (Acxiom)، وبلو كاي (BlueKai) من أجل دمج بيانات الشراء الخاصة بهم التي لا تخص فيسبوك مع البيانات السلوكية التي يوفرها فيسبوك (Edwards 2013).

فبطرق مختلفة، يتخلى الأفراد عن البيانات الشخصية الخاصة بهم، عن قصد أو عن غير قصد، وبدرجات متفاوتة: بصفة مشترين، أو مشتركين، أو مسجلين، أو أعضاء، أو حاملي بطاقات، أو مانحين، أو متسابقين، أو مشاركين في الاستطلاعات، أو حتى مجرد مستفسرين (CIPPIC 2006: ii). وعلاوة على ذلك، وحيث إن إنشاء وإدارة، وتحليل البيانات هي مهمة متخصصة، تستعين العديد من الشركات بجهات خارجية من شركات معالجة البيانات وتحليلاتها للوفاء بمتطلبات البيانات الخاصة بها. وتستطيع هذه الشركات، من خلال تقديم هذا النوع من خدمات البيانات لمختلف العملاء، بناء مجموعات بيانات واسعة النطاق، والتي يمكن الجمع بينها وحزمها لاستخدامها في إنتاج بيانات مشتقة جديدة، والتي بدورها توفر المزيد من الأفكار والرؤى مقارنة بتلك التي يمكن تحصيلها من مصدر واحد فقط للبيانات، وإضافة إلى هذه البيانات التي يحصل عليها وسطاء البيانات من القطاع الخاص، يعمل هؤلاء الوسطاء كذلك على الحصول على مجاميع بيانات القطاع العام سواء التي تتعلق بالأفراد أو تلك التي تخص مجموعة منهم كهيئة معينة وأماكنهم مثل بيانات سجلات الملكية والتعداد السكاني، كما يقومون باستئجار/ أو شراء القوائم المتعلقة بالأفراد من الجمعيات الخيرية والمنظمات غير الحكومية.

ومن خلال تجميع البيانات من مصادر متنوعة، يُنشئ وسطاء البيانات بنية تحتية من البيانات الضخمة والمتراصة، فعلى سبيل المثال، تشتهر شركة إبسيلون (Epsilon) بامتلاك بيانات عن ثلاثمائة مليون بطاقة ولاء من أعضائها في جميع أنحاء العالم، إضافة إلى بنك معلومات يضم بيانات ذات صلة بـمئتين وخمسين مليون مستهلك في الولايات المتحدة وحدها (Edwards 2013). كما تشتهر شركة أكسيوم (Acxiom) بإنشائها بنكا للمعلومات يتعلق بخمسمائة مليون مستهلك فعّال في جميع أنحاء العالم (تقريباً مئة وتسعون مليون شخص، ومئة وستة وعشرون مليون أسرة في الولايات المتحدة)، مع نحو ألف وخمسمائة معلومة عن كل شخص، وتنفذ خوادمها أكثر من خمسين تريليون عملية في السنة، وتجاوز إجمالي تدوير رأس المال للشركة المليار دولار (Singer 2012a). كما تدير أيضاً قواعد بيانات منفصلة، أو تعمل مع سبعة وأربعين شركة من كبرى الشركات المئة الأكبر في العالم (Singer 2012a). وتزعم شركة داتا لوجيكس (Datalogix) بتخزينها بيانات متعلقة بمشتريات لم تتم عن طريق الإنترنت تبلغ قيمتها أكثر من تريليون دولار (Edwards 2013).

وتشمل شركات الوساطة والتحليل للبيانات شركات أخرى مثل شركة أليانس داتا سيستمز (Alliance Data Systems)، وإي بيرو (eBureau)، وتشويس بوينت (ChoicePoint)، وكورلوجيك (Corelogic)، وإيكيفاكس (Equifax)، وإكسبيريان (Experian)، آي دي أنالتيكس (ID Analytics)، وإنفوجروب (Infogroup)، وإنوفيس (Innovis)، وإنتليوس (Intelius)، وريكورديد فيوتشر (Recorded Future)، وسيست (Seisint)، وترانز يونيون (TransUnion). وتتجه كل شركة من هذه الشركات إلى التخصص في أنواع مختلفة من البيانات ومنتجات البيانات والخدمات، فعلى سبيل المثال، تقيم شركة إي بيرو العملاء المحتملين بالنيابة عن شركات بطاقات الائتمان، والمقرضين، وشركات التأمين، والمؤسسات التعليمية، كما تقدم شركة إنتليوس خدمات البحث عن الأشخاص والتحري عن خلفياتهم (Singer 2012a).

وبشكل عام، فإن مبتغى وسطاء البيانات وشركات التحليل هو مجموعة واسعة من البيانات (الصغيرة والكبيرة)، والمتعلقة بأكبر شريحة ممكنة من السكان، وأن تكون هذه البيانات مترابطة بشكل كبير ويمكن تمييزها (تحتوي تعريف فريد) بحسب طبيعتها، وكلما استطاع وسيط البيانات تصدير ودمج المزيد من البيانات، كان عمل منتجاته أكثر نجاحاً وعلى النحو الأمثل، واكتسب ميزة تنافسية عن منافسيه الآخرين. ويُمكن من خلال جمع البيانات معا وهيكلتها بشكل مناسب لوسطاء البيانات اشتقاق بيانات جديدة، وملفات تعريفية عن الأفراد والمناطق، والقيام بإجراء النمذجة التنبؤية لتوقع ما قد يفعله الأفراد في ظل ظروف مختلفة وفي أماكن مختلفة. وهذا من شأنه أن يُمكن من تحديد المستهدفين والمخاطر بشكل أكثر فاعلية بين مختلف المستهلكين وأن يُعطي مؤشراً على مدى التصرف المحتمل لهؤلاء المستهدفين والدفع بهم لاختيار استجابة معينة (على سبيل المثال، اختيار وشراء عنصر معين)، فشركة أكسيوم تسعى، على سبيل المثال، لربط البيانات غير المتوافرة على الإنترنت (Offline Data)، والبيانات المتوافرة على الإنترنت (Online Data)، وحتى البيانات المتنقلة (Mobile Data) من أجل تكوين نظرة شاملة متعددة الزوايا عن المستهلكين، حيث تستخدم هذه البيانات لإنشاء ملفات تعريفية مفصلة ونماذج تنبؤية متماسكة (Singer 2012a). فهذه المعلومات والنماذج مفيدة جداً لشركات الشراء لأنها تمكنهم من تركيز جهودهم المتعلقة بالتسويق والمبيعات، وتزيد من فرص نجاح عملياتهم،

في ذات الوقت الذي تمكنهم من الحد من النفقات من حيث الهدر والخسارة التي قد تنتج عن استثمارات محفوفة بالمخاطر، ومن ثم تسعى هذه الشركات لتصبح أكثر فعالية وكفاءة في عملياتها باستخدامها لمثل هذه المنتجات.

ومن المثير للاهتمام، أنه بالرغم من امتلاك وسطاء البيانات وشركات التحليل لكميات كبيرة ومتنوعة من البيانات الشخصية التي تُستخدم منتجاتها لتصنيف واستهداف الأفراد والأسر اجتماعياً، فقد كان هناك نقص ملحوظ في الدفع بالاهتمام النقدي لعمليات هذه الشركات. ففي الواقع، هنالك ندرة في التحليل الأكاديمي والإعلامي عن هذه الشركات والآثار المترتبة على عملها ومنتجاتها، ويعود السبب في ذلك جزئياً إلى أن هذه الصناعة بعيدة عن الأضواء وسرية نسبياً، ولا ترغب في توجيه انتباه الجمهور إليها، وزعزعة ثقة الجمهور في أصولها وأنشطتها، والتي قد تؤدي إلى حملات عامة للشفافية، والمساءلة، والتنظيم. فوسطاء البيانات بشكل عام غير منظمين إلى حد كبير في الوقت الحاضر، كما أنهم غير ملزمين قانونياً بتزويد الأفراد بإمكانية الوصول إلى البيانات المحفوظة عندهم، ولا هم ملزمون بتصحيح الأخطاء المتعلقة بهؤلاء الأفراد (Singer 2012b). وإضافة إلى أن بيانات ومنتجات هذه الشركات يُمكن أن يكون لها تأثير عميق في الخدمات والفرص المقدمة للأفراد، مثل ما إذا كان سيتم تقديم عرض لفرصة عمل، أو قبول تقديم طلب قرض، أو إصدار بوليصة تأمين، أو الموافقة على الاستئجار، وتكلفة السلع والخدمات المقدّرة على أساس المخاطر المتوقعة والقيمة بالنسبة لشركة ما (Lyon 2002).

إن قلق البعض، بمن فيهم (Edith Ramirez 2013) رئيسة لجنة التجارة الاتحادية (Federal Trade Commission) في الولايات المتحدة، يتمثل في ممارسة هذه الشركات شكلاً من أشكال حتمية البيانات (Data Determinism) التي لا يجري فيها فقط تصنيف الأفراد والحكم عليهم على أساس ما فعلوه، ولكن على التنبؤ بما قد يفعلونه في المستقبل باستخدام خوارزميات هي أبعد ما تكون عن الصحة والكمال، بحيث قد تحمل في ثناياها تحيزات تتعلق بالعرق، والإثنية، والجنس، والحياة الجنسية. وحتى الآن، فهم صناديق سوداء تفتقر إلى الرقابة المجدية والإجراءات المُعالجة، وتستخدم بيانات منخفضة الجودة على الأغلب، ومن ثم تكون عرضة للخطأ (انظر أيضاً للفصل العاشر)، وعلاوة على ذلك

فإن هذه الشركات تستخدم البيانات لغير الأغراض التي تمّ إنشاؤها لأجلها، وبالتأكيد ليست لأغراض أولئك الذين تشير البيانات إليهم وتمثل ما قد يرغبون فيه، فالذين يتقدمون ببياناتهم للحصول على بطاقة ولاء المتجر لا يتوقعون بالضرورة أن يتم بيع بياناتهم واستخدامها لأغراض تتجاوز علاقتهم مع المتجر، ولا يتوقعون استخدامها بطرق قد تعرضهم للعقاب والمحاسبة بعد ذلك (Ramirez 2013). وبينما يرحب بعض المستهلكين بالعروض الشخصية من المتاجر أو الإعلانات الموجهة على المواقع الإلكترونية، يعدها البعض الآخر تطفلاً واستغلالية (Singer 2012a). وعلاوة على ذلك، فإن كنز البيانات التي قد يكون لها قيمة في المستقبل كإجراء للمضاربة السوقية، هو انتهاك لقوانين الحد من البيانات التي تنص على وجوب الإبقاء على البيانات محددة القيمة فقط (Raley 2013). ويُعد ووسطاء البيانات هدفاً رئيسياً للمجرمين العازمين على الاحتيال وسرقة الهويات، وذلك بالنظر إلى حجم السجلات الشخصية الحساسة مثل الأسماء، والعناوين، وأرقام الهوية (مثل الضمان الاجتماعي، وجواز السفر، ورخصة القيادة)، وأرقام بطاقات الائتمان المتوافرة لديهم. وفي الواقع، فقد تابعت لجنة التجارة الاتحادية (Federal Trade Commission- FTC) أكثر من أربعين قضية لخروقات أمنية خطيرة للبيانات، أغلبها ضد ووسطاء البيانات، لفشلهم في توفير ضمانات أمنية كافية على البيانات (Ramirez 2013).

وفي الآونة الأخيرة، وفي أعقاب الخروقات الأمنية والكشف عن صفقات بيع مشبوهة واستخدام بيانات حساسة (Dwoskin 2013)، جنباً إلى جنب مع الكشف عن برنامج بريزم التجسسي لوكالة الأمن القومي (NSA PRISM) ووصول الحكومة إلى سجلات تجارية (Greenwald and MacAskill 2013)، إضافة للأخبار عن فضيحة قرصنة الهاتف العالمي واستجواب ليفينسون (Leveson) في المملكة المتحدة (<http://www.levesoninquiry.org.uk>)، فقد انصبّ المزيد من الاهتمام على ووسطاء البيانات وممارساتهم التجارية. ففي أواخر عام 2012، استدعت لجنة التجارة الاتحادية تسعة من ووسطاء البيانات لاكتشاف المزيد حول ماهية البيانات والمعلومات المستمدة منها التي يقومون بإنتاجها وجمعها عن الناس وكيفية توظيفهم هذه البيانات وطرق بيعها، وأصدرت تقريراً يدعو للخصوصية عند التصميم (انظر الفصل العاشر)، وخيارات مبسطة، وقدر أكبر من الشفافية (لجنة التجارة الاتحادية 2012). وعلاوة على ذلك، فقد تمّ

تطبيق الضغط السياسي لاعتماد أنظمة أشد رقابة وأكثر شفافية. أما في أوروبا، فيسعى قانون الاتحاد الأوروبي الجديد لحماية البيانات لتنظيم كيفية استخدام البيانات، ومنح المواطنين الحق في الوصول إلى البيانات والتنازع عليها والحق في تناسيها وحذفها (انظر الفصل العاشر). وسيعتمد مدى نجاح مثل هذه الأعمال على إعادة تشكيل أعمال وسطاء البيانات والذي سيحدثنا عنه الزمن لاحقاً، وفي غضون ذلك، هناك العديد من الأسئلة المتعلقة بوسطاء البيانات والأسواق وآثارها على المواطنين والمستهلكين التي تحتاج إلى استكشاف للإجابة عنها (انظر الفصل السادس).

الخلاصة:

على الرغم من الدعوات المنمقة لمؤيدي البيانات الكبيرة وأولويات التمويل للوكالات البحثية، فإن البيانات الصغيرة ستستمر بالبقاء باعتبارها عنصراً حيوياً في فضاء البحث العلمي، فدراسات البيانات الصغيرة قد تكون محدودة في الحجم، ومحصورة من حيث عيناتها وتنوعها، وثابتة، وغير مرنة بمجرد تنفيذها، ولكنها خدمت العلوم بشكل جيد حتى وقتنا الحاضر لأنها مصممة خصيصاً للرد على أسئلة مركزة بإحكام. ويُتيح إنتاج البيانات الصغيرة، جنباً إلى جنب، مع أدوات التحليل الجيدة، للباحثين التنقيب الدقيق وبشكل فعال في مستويات محكمة من البيانات العالية الجودة من أجل فهم العالم. وفي المقابل، وكما نوقش في الفصل الثامن، تسعى العديد من دراسات البيانات الكبيرة إلى الإجابة عن الأسئلة التي يمكن الإجابة عنها من البيانات المنتجة؛ فكما قيل: الذيل الآن هو الذي يهز الكلب.

ومع ذلك، فمن الواضح أن رؤى وقيمة أكبر بكثير يمكن الحصول عليها من البيانات الصغيرة من خلال توسيع نطاقها في البنى التحتية للبيانات والاحتفاظ بها للأجيال القادمة. وفي حين أن البيانات ولفترات طويلة قد جُمعت معاً، وجرى تخزينها وأرشفتها، إلا أن تطور التقنيات الرقمية قد غير بشكل جذري قدرتنا على إدارة، وتنظيم، ومعالجة، وتحليل، ومشاركة، وإعادة استخدام البيانات، وخاصة تلك البيانات التي ولدت رقمية في الأصل، فقد كانت البيانات الصغيرة فيما سبق نادرة ومقيّدة، ولكنها غدت متوافرة بشكل أكبر وأكثر انفتاحاً (انظر الفصل الثالث). وبتوسيع نطاق هذه البيانات يمكن تحقيق جدوى اقتصادية وذلك فيما يتعلق بخدمات البيانات، وأدوات التحليل والموارد التربوية، وتوظيف

البيانات في استخدامات جديدة، وإنتاج اكتشافات جديدة. ونتيجة لذلك، يُبذل الآن جهد كبير لإنتاج البنى التحتية للبيانات في جميع المجالات الأكاديمية. وهو الجهد الذي من شأنه أن يُتيح بشكل أكبر بكثير البيانات للمجتمعات المعنية وأن يُغيّر طبيعة عمل العلوم (انظر الفصل الثامن).

ويتضح المدى الذي يمكن من خلاله لهذا التوسع إنتاج قيمة من خلال النمو السريع لقطاعي وساطة البيانات والتحليلات المصاحبة لها. حيث يمكن لوسطاء البيانات، من خلال الجمع بين مخازن كبيرة من البيانات الصغيرة التي تحتفظ بها المؤسسات العامة والشركات الخاصة ودمجها جنباً إلى جنب مع مجالات تدفق البيانات الكبيرة، إنتاج مختلف أنواع الملفات التعريفية التفصيلية للأفراد والمجموعات والتي يمكن استخدامها للاستهداف المركز للأسواق، وتقييمها، وفرزها، مما يوفر معلومات ذات قيمة عالية للعملاء. كما تكشف هذه الممارسات عن بعض القضايا والمخاطر الناجمة عن التوسع في نطاق البيانات، وبالتحديد استخدامها لغير الأغراض التي جُمعت لأجلها، وأمنها، والفرص الجنائية في حال تم سرقة البيانات، حيث لم يُفهم أي من هذه القضايا والمخاطر بشكل تام حتى الآن، مع ضعف الجانب التنظيمي والإجراءات المضادة.

فهناك العشرات من الأسئلة التي تتطلب اهتماماً بشأن الدور المستقبلي وطبيعة دراسات البيانات الصغيرة، وتوسيع نطاق البيانات الصغيرة إلى أشكال مختلفة من البنى التحتية للبيانات والآثار المترتبة عليها، والعواقب الناتجة عن حفظ البيانات والوساطة فيها على أساس الربح. ومع ذلك، فمن الواضح أن التحليل النقدي للتغيرات التي تحدث ليست مواكبة للتطورات التقنية والتجارية، فالواقع أنه حتى الآن كان هناك عدد قليل من الانطباعات النقدية المبنية على أساس التفكير من خلال الآليات والدلالات في كيفية إنتاج البيانات الصغيرة، وتوسيع نطاقها واستخدامها، والتنظير لأدواتها - قواعد البيانات والبنى التحتية للبيانات - التي يتم استخدامها لتنظيم وتبادل وتحليل البيانات. ونتيجة لذلك، هناك ندرة في المعرفة المتعلقة بالأنظمة الاجتماعية التقنية الناشئة والمعقدة التي يتم إنشاؤها، مع عدد قليل من دراسات الحالة المعقدة. ومن ثم، يتضح أن هناك حاجة للكثير من البحث والتأمل لفهم البنى التحتية للبيانات في جميع أشكالها.

الفصل الثالث

البيانات المفتوحة (Open Data) والبيانات المرتبطة (Linked Data)

بالنظر إلى التكاليف والنفقات والموارد المطلوبة لإنتاج مجموعات البيانات وقيمتها في الكشف عن معلومات عن العالم من حولنا، فقد تمّ بشكل عام تقييد الوصول إليها بطريقة أو بأخرى، فعلى سبيل المثال يُتاح الوصول للمستخدمين المسجلين المعتمدين فقط، أو من خلال المطالبة برسوم اشتراك، أو عن طريق حصر الكيفية التي يمكن فيها استخدام البيانات بواسطة السياسات والتراخيص. وحتى عندما كانت مجموعات البيانات متاحة ومفتوحة نسبياً، فقد كانت تتطلب معدات وأدوات متخصصة لفهمها والوصول إليها، مثل أجهزة الحاسب، والبرمجيات، والمهارات المتعلقة بطرق الإحصاء ورسم الخرائط والمعرفة السياقية للحقل أو الموضوع الذي يجري التطرق إليه، مما يتعدى بكثير قدرات عموم الناس. ونتيجة لذلك، كانت البيانات والمعلومات المستمدة منها تقليدياً مغلقة إلى حدّ كبير بطبيعتها - أي مقفل عليها داخل مؤسسة أو محفوظ أرشيفي. وفي الواقع، فقد مثل عدم إتاحة الوصول إلى مجموعات البيانات التي يمكن أن تساعد في الإجابة عن أسئلة معينة مصدر إرباط على مدى قرون للباحثين، والمحللين، والصحفيين، ومنظمات المجتمع المدني.

• ومن هذا المنطلق سعت حركة البيانات المفتوحة لتحويل هذا الوضع جذرياً، ليس فقط من خلال فتح البيانات وإتاحتها لإعادة استخدامها على نطاق أوسع فقط، ولكن أيضاً من خلال توفير أدوات بحثية سهلة الاستخدام والتي لا تتطلب الحاجة إلى مهارات تحليلية متخصصة. وبُنيت هذه الحركة على ثلاثة مبادئ هي: الانفتاح، والمشاركة، والتعاون (البيت الأبيض 2009 White House)؛ فمن خلال الشفافية، والمشاركة، والعمل معاً يمكن إدراك قيمة البيانات للمجتمع. وتهدف الحركة إلى تحرير القدرة على إنتاج المعلومات والمعرفة، بدلاً من حصر إمكانية الاستفادة من قوة البيانات على منتجها أو أولئك الذين يمكنهم دفع ثمن إتاحة الوصول إلى البيانات. وقد تركز الاهتمام، على وجه الخصوص، على فتح البيانات التي تمّ إنتاجها من قبل أجهزة الدولة وغالباً ما يطلق عليها معلومات القطاع

العام (Public Sector Information - PSI) أو البحوث الممولة من القطاع العام (انظر الفصل الثاني)، مع الأخذ بالحسبان بأن هذه البحوث قد تم تمويلها من قبل الخزينة العامة لفائدة عامة الناس، مع تركيز أقل على فتح البيانات التي تمّ إنشاؤها من قبل القطاع الخاص، والتي قد يكون لديها قيمة خاصة أكبر عند منتجها على وجه الخصوص.

وقد تطورت حركة البيانات المفتوحة خلال العقدين الماضيين، بالتزامن مع - ولكن مع فصل واضح إلى حدّ كبير عن - حركة الحق في المعلومات (Right to Information-RTI) المعنية باتخاذ قرارات الشفافية وصنع السياسات من قبل الوكالات والمعروفة أيضاً باسم حرية المعلومات (Freedom of Information)، وحركتي المصدر المفتوح والعلوم المفتوحة المعنيتين بإعادة تشكيل البرمجيات والنشر إلى أشكال مفتوحة وشفافة فيما يتعلق بالترخيص، وحقوق الطبع والنشر، وحقوق الملكية الفكرية. فمنذ أواخر سنة 2000م، اكتسبت حركة البيانات المفتوحة مكانة بارزة وزخم ثابت بشكل ملحوظ، ابتداء من حملة صحيفة الغارديان في المملكة المتحدة تحت مُسمى "تخليص بياناتنا- www. Free Our Data" ودعوة منظمة التعاون الاقتصادي والتنمية (Organization for Economic Cooperation and Development - OECD) حكومات الدول الأعضاء لفتح بياناتها في عام 2008 م، وإطلاق الحكومة الأمريكية في العام 2009 للموقع الإلكتروني (www.data.gov)، وهو موقع على شبكة الإنترنت صُمم لتوفير الوصول إلى مجموعات البيانات غير الحساسة والتاريخية التي تحتفظ بها أجهزة الدولة على مستوى الولايات ومستوى الحكومة الفيدرالية بالولايات المتحدة الأمريكية، وتطوير البيانات المرتبطة والترويج ل الويب الدلالي (Semantic Web) كعنصر موحد لتكنولوجيا الانترنت في المستقبل، والذي يُمكن البيانات المفتوحة والبيانات المرتبطة من أن تتصل في الغالب بشكل طردي (Berners-Lee 2009). ومنذ عام 2010 م، حذت حذوها عشرات الدول والمنظمات الدولية على سبيل المثال، دول الاتحاد الأوروبي (-European Union EU)، وبرنامج الأمم المتحدة الإنمائي (-United Nations Development Programme UNDP) مما جعل الآلاف من مجموعات البيانات المقيدة سابقاً متاحة ومفتوحة بطبيعتها للاستخدام غير التجاري والتجاري (انظر DataRemixed 2013). وقد أسهم في مثل هذا التحول في الموقف مجموعات الضغط الدولية والوطنية المؤثرة مثل مؤسسة المعرفة المفتوحة

(Open Knowledge Foundation) ومؤسسة ضوء الشمس (Sunlight Foundation)، ورافقه ضغط من المجموعات والشركات العاملة في مجال الاقتصاد المعرفي، بالإضافة إلى كبار موظفي الخدمة المدنية المقتنعين بالحجج المستخدمة، وعشرات الجماعات المحلية التي تسعى للاستفادة من البيانات المدنية.

وفي حين يتم عرض حجج حركة البيانات المفتوحة بطريقة بديهية، باستخدام مفردات الاستعارة البلاغية مثل الشفافية، والمساءلة، والمشاركة، والابتكار، والنمو الاقتصادي، فإن الانفتاح السريع للبيانات الحكومية والعلمية لم يلق ترحيباً عالمياً. فالواقع، أن العديد من الآثار الاجتماعية والسياسية والاقتصادية لفتح البيانات يجري حالياً الكشف عنها ومناقشتها وانتقادها. وعلاوة على ذلك، وللمفارقة إلى حد ما، تعزز نطاق ومدة حقوق الملكية الفكرية على مدى نصف القرن الماضي، بمعنى أن نمو حركة البيانات المفتوحة تم بالتوازي مع زيادة في حقوق الملكية الفكرية (Pollock 2006)، فالمقاومة التي واجهتها فكرة توفير حرية الوصول إلى البيانات العلمية نوقشت في الفصل السابق، وسيتم في القسم الأخير من هذا الفصل مناقشة الانتقادات الموجهة لدعوى البيانات المفتوحة. فبدية، ومع ذلك، يحدّد هذا الفصل خصائص البيانات المفتوحة والبيانات المرتبطة، ومختلف الطرق التي هيأت لقضية البيانات المفتوحة، والاقتصاديات المرتبطة بتقديم بيانات مفتوحة، وعلى الرغم من أن فتح البيانات الحكومية ما يزال جزئياً، فليس هنالك شك في أن تحولاً كبيراً يحدث بشأن الكيفية التي تتم من خلالها عرض ومشاركة البيانات بشكل عام.

البيانات المفتوحة (Open Data):

عند أحد المستويات، يُعد تعريف مصطلح البيانات المفتوحة بسيطاً نسبياً، فقد حدّده بولوك (2006) على سبيل المثال كالتالي: «تكون البيانات مفتوحة إذا كان لأي شخص الحرية في استخدامها، وإعادة استخدامها، وإعادة نشرها - فقط شرط موافقتها لمطلب إسنادها و/ أو مشاركتها على حدّ سواء». ولكن هذا التعريف يُلّمح إلى الطبيعة المتنوعة لما يمكن أن يكون عليه المقصود من لفظة «الفتح» وكيف أنها يمكن أن تعني أشياء مختلفة بالنسبة لمختلف الوكالات في سياق حقوق الملكية الفكرية، فقد يشير الانفتاح إلى الاستخدام أو إعادة الاستخدام، أو إعادة الصياغة، أو إعادة التوزيع، أو إعادة النشر والتوزيع، أو

إعادة البيع، وقد يكون لكلٍ منها شروط وأحكام تختص بها، فعلى سبيل المثال قد يكون للمستخدم القدرة على استخدام مجموعة البيانات بحرية بموجب ترخيص، ولكن ذلك لا يعني إمكانية إعادة التعامل مع هذه البيانات بهدف الربح التجاري أو إعادة البيع، أو أي استخدام قد يتطلب الإسناد. وبعبارة أخرى، قد يكون الوصول إلى مجموعة البيانات مفتوحاً، ولكن ليس بالضرورة أن يكون كل ما يمكن فعله مع البيانات الممكن الوصول إليها مفتوح أيضاً. وفي حالات أخرى، قد تحتفظ المنظمات ببياناتها وتتحكم بإتاحة الوصول إليها بنفسها، ولكنها تتيح بحرية البيانات الوصفية المتعلقة بها فقط، أو قد تسمح المنظمة لبعض المستخدمين بالوصول إلى البيانات، ولكنها لا تمكنهم من إعادة نشرها إلا إذا كان ذلك سيحقق قيمة مضافة لها وبدون الوصول إلى البيانات الأولية الأساسية.

ومع ذلك، سعت عدد من المنظمات إلى تحديد الخصائص المثالية للبيانات المفتوحة، على سبيل المثال يؤكد تعريف الانفتاح بأن العمل هو مفتوح إذا استوفت حالته في النشر، والشروط المنصوص عليها في الجدول رقم (٣-١)، والتي تفرض بعض القيود على الوصول، والاستخدام، وإعادة الصياغة، وإعادة التوزيع والنشر، وتشجع كذلك بشكل نشط على الاستخدام الربحي للبيانات المفتوحة دون أي تعويض مالي لمنشئ البيانات الأصلي. وبالمثل، حدّدت منظمة الحوكمة المفتوحة (OpenGovData) تسعة مبادئ للبيانات المفتوحة تتعدى مجرد نمط التوزيع إلى النظر أيضاً في طبيعة البيانات نفسها (انظر الجدول رقم 2-3). وبالإضافة إلى هاتين المجموعتين من الشروط / المبادئ، اقترحت وكالات أخرى مثل منظمة التعاون الاقتصادي والتنمية (OECD) (2008) والحكومة الأسترالية (Fitzgerald 2010) حاجة البيانات المفتوحة إلى أن تكون مصحوبة بقوائم الأصول وآليات اكتشاف البيانات، من أجل إصدار المؤشرات المتعلقة بجودة البيانات وموثوقيتها، واستخدام النماذج والمعايير التي تتيح قابلية التشغيل البيئي للبيانات المفتوحة، وتوفير بنى تحتية سهلة الاستخدام للبيانات لتسهيل النشر المنتظم وإيجاد مجموعات بيانات جديدة، بالإضافة إلى عدد من الأدوات الأساسية والتخصصية التي من شأنها أن تحسن الاستخدام والتحليل للبيانات. ويُعد ضمان التشغيل البيئي مهماً بشكل خاص لأنه يسمح بربط مجموعات البيانات المختلفة، مما يُساهم في بناء مجموعات بيانات جديدة أكثر تعقيداً واكتشاف رؤى جديدة (مؤسسة المعرفة المفتوحة 2012 Open Knowledge Foundation). كما لا توجد قيود فيما يتعلق بمجالات التركيز للبيانات المفتوحة - حيث يمكن أن تتعلق بأي مجال

من المجالات الاجتماعية، والاقتصادية، والتجارية، والثقافية (وسائل الإعلام والمكتبات والتراث)، والبيئية، أو الظواهر العلمية - ولكن، بشكل عام، ينصب التركيز حتى الآن على فتح البيانات التي تتصل بالسياسات العامة عالية المستوى ولها قيمة تجارية عند إعادة استخدام مثل البيانات المتعلقة بالاقتصاد، والنقل، والبيانات الجغرافية.

جدول رقم (٣-١)

الخصائص المثالية للبيانات المفتوحة المتعلقة بتعريف الانفتاح

إتاحة الوصول Access	ينبغي أن يكون العمل متاحاً لمجموعه ككل وبلا تكلفة إعادة إنتاج تتجاوز المعقول كما ينبغي أن تتوفر بشكل مناسب قابل للتعديل
إعادة التوزيع Redistribution	ينبغي أن لا تُقيّد الرخصة أي طرف عن بيع العمل أو توزيعه سواء فيما يتعلق بالعمل نفسه أو كان ذلك العمل كجزء من حزمة مركبة من أعمال من عدّة مصادر مختلفة. وينبغي أن لا يتطلب الترخيص إتاوة أو رسوماً أخرى لهذا البيع أو التوزيع.
إعادة الاستخدام Reuse	ينبغي أن يسمح الترخيص بالتعديل على العمل وإجراء عمليات الاشتقاق منه ويسمح كذلك بتوزيع الأعمال المشتقة وفقاً لأحكام العمل الأصلي.
غياب القيود التقنية Absence of technological restrictions	ينبغي تقديم العمل على الصورة التي لا يوجد معها عقبات تقنية تحول دون استخدامه أو إعادة توزيعه.
الإسناد Attribution	قد يُلزم الترخيص الإسناد إلى المساهمين والمبدعين للعمل الأصلي كشرط لإعادة التوزيع وإعادة الاستخدام، وإذا تمّ فرض هذا الشرط فينبغي ألا تتطلب عملية الإسناد إجراءات مرهقة ومُعقدة.

من الممكن أن يشترط الترخيص حمل الإصدارات المعدلة للبيانات أسماء تختلف عن العمل الأصلي كشرط لإعادة توزيعها.	Integrity والصحة والنزاهة
ينبغي عدم تمييز الترخيص ضد أي شخص أو مجموعة من الأشخاص وضمان إتاحة الوصول للجميع.	عدم التمييز ضد الأشخاص أو المجموعات No discrimination against persons or groups
ينبغي عدم تقييد أي شخص من الاستفادة من العمل في أي مجال معين. على سبيل المثال، لا ينبغي تقييد العمل من استخدامه في الأعمال التجارية.	عدم التمييز ضد مجالات معينة في العمل No discrimination against fields of endeavor
ينبغي تطبيق الحقوق المتعلقة بالعمل على جميع الذين يتم إعادة توزيعه عليهم دون الحاجة لتنفيذ ترخيص إضافي من قبل تلك الأطراف.	توزيع الترخيص of licence Distribution
ينبغي أن لا تعتمد الحقوق المتعلقة بالعمل على كونه جزءاً من حزمة معينة.	عدم تعيين الترخيص على حزمة أعمال Licence must not be specific to a package
ينبغي أن لا يضع الترخيص قيوداً على الأعمال الأخرى التي يتم توزيعها جنباً إلى جنب مع العمل المرخص. على سبيل المثال، ينبغي أن لا يصر الترخيص على أن تكون كل الأعمال المستمدة منه مفتوحة أيضاً.	عدم تقييد الترخيص توزيع الأعمال الأخرى not restrict the distribution of other works Licence must not restrict the distribution

المصدر: مقتبس من <http://opendefinition.org/od>

جدول رقم (٣-٢)

مبادئ منظمة الحوكمة المفتوحة للبيانات المفتوحة

تكون كافة البيانات متاحة، مع مراعاة أحكام وشروط الخصوصية، والأمن أو قيود الصلاحيات.	يجب أن تكون البيانات كاملة Data must be complete
تُنشر البيانات كما تمّ جمعها من المصدر، مع أفضل مستوى ممكن من التفاصيل، وليست بأشكال مجملة أو معدّلة.	يجب أن تكون البيانات أساسية (أولية) Data must be primary
إتاحة البيانات في أسرع وقت ممكن عند الحاجة لها للحفاظ على قيمة البيانات.	يجب أن تكون البيانات في الوقت المناسب Data must be timely
إتاحة البيانات لأكثر عدد ممكن من المستخدمين ولأوسع مجموعة من الأغراض.	يجب أن تكون البيانات في المتناول Data must be accessible
تنظم البيانات بشكل مناسب للسماح بالمعالجة الآلية لها	يجب أن تكون البيانات قابلة للمعالجة الآلية Data must be machine-processable
تتوفر البيانات لأي أحد، مع عدم اشتراط التسجيل.	يجب عدم تمييز عملية الوصول Access must be non-discriminatory
تتوفر البيانات في صيغة لا يكون فيها سيطرة حصرية عليها لأي كيان.	يجب أن تكون صيغ البيانات غير مخصصة الملكية Data formats must be non-proprietary

<p>لا تخضع البيانات لأية أحكام في حقوق التأليف والنشر وبراءات الاختراع والعلامات التجارية. وقد يُسمح ببعض قيود الخصوصية والأمن والصلاحيات المناسبة بحسب ما تفرضه التشريعات الأخرى.</p>	<p>يجب أن يكون ترخيص البيانات مجانياً Data must be licence-free</p>
<p>يجب أن يتم تعيين شخص كضابط اتصال للرد على الناس الذين يحاولون استخدام البيانات أو تقديم شكاوى حول انتهاكات المبادئ الرئيسية. ويجب أن يكون هناك هيئة أخرى ذات اختصاص لتحديد ما إذا تم تطبيق المبادئ بشكل مناسب.</p>	<p>يجب أن يكون الامتثال قابلاً للمراجعة وإعادة النظر Compliance must be reviewable</p>

المصدر: مقتبس من <http://www.opengovdata.org>

وإلى حدٍ كبير، تعد هذه الخصائص المثالية البيانات المفتوحة مُنتجاً، وليست خدمة. وفي المقابل، فقد جادل Gurstein (2013) حول الحاجة إلى إعادة التفكير بالبيانات المفتوحة كعملية خدمية (Service process) - بوصفها تفاعلاً وتنطوي على علاقة بين مزودي البيانات والمستخدم النهائي. بالنسبة له، لا يكفي أن تكون البيانات بسيطة متاحة من قبل المنظمات لإعادة استخدامها، بل ينبغي أن يكون فتح البيانات موجّهاً أكثر باعتباره خدمة، مع الأخذ بعين الاعتبار حاجات وتوقعات المستخدمين النهائيين منها، فهذه إعادة لصياغة المفاهيم تقترح على أولئك الذين ينتجون البيانات المفتوحة اتباع طرقاً مختلفة لإدارة البيانات والمستخدمين النهائيين، كالأثار المترتبة على قضايا إعادة تحديد المصدر الخاص بالبيانات. بالنسبة له، يتطلب هذا النوع من النهج مشاركة المستخدمين النهائيين في التخطيط، والتطوير، وإدارة المشاريع الجارية للبيانات المفتوحة، وكذلك استخدام مقاييس لتقييم مدى نجاح مثل هذه المشاريع في تعزيز الصالح العام، وهذا يتطلب برامج بناء القدرات الرامية إلى رفع مهارات المستخدمين ليكونوا قادرين على إدارة ومعالجة وتحليل البيانات بشكل مناسب وفعال. ورغم أنه من السهل أن نتفق على أنه ينبغي تسليم البيانات المفتوحة كخدمة، إلا أن البيانات المفتوحة بطبيعتها لا تنتج دخلاً أو تنتج القليل مقابل تمويل هذه الترتيبات الخدمية فضلاً عن التكاليف المتعلقة بفتح البيانات، فالواقع أن ذلك قد يكون من الطموح إلا إذا تم تطوير نماذج تمويل فعالة (كما نوقش أكثر بالكامل أدناه).

البيانات المرتبطة (Linked Data):

تتمثل فكرة البيانات المرتبطة بتحويل شبكة الانترنت من شبكة الوثائق إلى شبكة البيانات من خلال إنشاء الشبكة الدلالية (Semantic web) (Berners-Lee 2009; P. Miller, 2010)، أو ما اصطلح عليه (Goddard and Byrne 2010) بمصطلح «المقروءة آلياً على شبكة الإنترنت» (Machine-Readable Web). وتدرك هذه النظرة بأن كافة المعلومات التي يتم تبادلها على الإنترنت تحتوي على تنوع غني للبيانات - أسماء، وعناوين، وتفاصيل المنتج، والحقائق، والأرقام، إلى آخره. ومع ذلك، لا يتم بالضرورة تحديد هذه البيانات رسمياً على هذا النحو، كما أنها ليست منتظمة رسمياً في طريقة يمكن جمعها واستخدامها بسهولة، فالواقع، أن معظم وثائق شبكة الإنترنت غير منظمة إلى حد كبير في طبيعتها. ومن خلال ترميز الوثائق وهيكلتها باستخدام معرفات فريدة واستخدام لغة التوصيف (Markup Language)، يمكن جعل البيانات التي تحتويها هذه الوثائق مرئية، وتمكين الآخرين تلقائياً من دمجها، ومعالجتها، وفهمها، وربطها مع البيانات الأخرى ذات الصلة (P. Miller, 2010).

ومن أجل تمكين شبكة الإنترنت الدلالية من أن تعمل كبيانات مرتبطة، بين Berners-Lee (2009) حاجة الوثائق المنشورة على شبكة الإنترنت للمواءمة مع أربعة توقعات سلوكية. أولاً، يجب تعريف الكائنات والموارد ضمن كل وثيقة باستخدام لغة التوصيف الموسعة (XML- Extensible Markup Language) وتسميتها بشكل واضح لا لبس فيه باستخدام معرفات الموارد الموحدة (Uniform Resource Identifiers - URIs). ثانياً، ينبغي استخدام الهيكل الضمني لشبكة الإنترنت لربط الوثائق والمعلومات (على سبيل المثال، استخدام معيار نقل النصوص التشعبية - أتش تي تي بي (HTTP) ومعرفات الموارد الموحدة (URIs) بحيث يمكن التنقيب في الأسماء والبحث فيها). ثالثاً، تسهيل اكتشاف المعلومات حول كائن مسمى أو مورد (على سبيل المثال، عندما يقوم شخص بالبحث عن معرف لكائن ما، تُقدّم له معلومات مفيدة باستخدام معايير إطار عمل وصف الموارد (Resource Description Framework - RDF). فإطار عمل وصف الموارد هو وثيقة تفصل وتصف طبيعة معرفات الموارد ضمن نطاق ما وينبغي أن يُحال كل معرف إلى هذه الوثيقة لتحصيل معلوماته (Goddard and Byrne 2010). ويضمن استخدام هذه المعايير إطاراً مشتركاً لجمع البيانات عبر الإنترنت، بدلاً من اعتماد مجموعة كبيرة من

الأساليب المتعارضة (Dietrich 2012). رابعاً، توفير وصلات للربط بين الكائنات والموارد ذات الصلة، إذا كانت معروفة، بحيث يمكن للمستخدمين اكتشاف المزيد من الأشياء.

وتمكّن هذه التوقعات السلوكية من جعل الوثائق مقروءة آلياً ويمكن معالجتها بحيث يكون التعرف على طبيعة الكائنات فيها ممكناً (على سبيل المثال، الولايات المتحدة هي بلد، وباراك أوباما هو شخص)، وإزالة الغموض عن المفاهيم (على سبيل المثال، مارس هو قطعة شوكولاتة، والمريخ (مارس) هو كوكب؛ أو أن باراك أوباما هو مؤلف وثيقة مقابل أن باراك أوباما هو موضوع وثيقة)، وإنشاء روابط وعلاقات داخل البيانات (على سبيل المثال، باراك أوباما هو رئيس الولايات المتحدة) (Goddard and Byrne 2010). وبهذه الطريقة، يصبح من الممكن لوثائق شبكة الإنترنت أن تكون قابلة للقراءة آلياً بحيث لا يقتصر عمل برنامج القراءة على فهم المحتوى ولكن يمكنه أن يستمد أيضاً بيانات مشتقة جديدة من خلال الاستدلال من المحتوى (Goddard and Byrne 2010).

وعندما تُنشر الوثائق بهذه الطريقة، يمكن استخلاص المعلومات على شبكة الإنترنت وإعادة تقديمها كبيانات وربطها بطرق لا حصر لها اعتماداً على الغرض. ولكن، وبحسب ما دوّن P. Miller (2010) فإن «البيانات المرتبطة يمكن أن تكون بيانات مفتوحة، ويمكن ربط البيانات المفتوحة، ولكن من الممكن أيضاً للبيانات المرتبطة أن تحمل تراخيصاً أو غيرها من القيود التي تحول دون اعتبارها مفتوحة»، أو إتاحة البيانات المفتوحة بطرق قد لا تجعل من السهل ربط بعضها ببعض. وبشكل عام، فإن أي وثائق مرتبطة ليست متوافرة على شبكة الإنترنت أو تقع ضمن نطاق نظام اشتراك مدفوع هي أيضاً بيانات مفتوحة بطبيعتها. فبحسب ما يرى (Berners-Lee 2009)، ينبغي أن تكون البيانات المفتوحة والبيانات المرتبطة متكافئة بشكل مثالي، وإضافة لذلك فقد حدد خمسة مستويات لمثل هذه البيانات، كل منها أكثر فائدة وقيمة تدريجياً (انظر الجدول رقم 3-3). وكان تطلعه إلى ما أسماه بيانات الخمسة نجوم (المستوى الخامس) - شبكة دلالية عاملة بالكامل. ولكن تجدر الإشارة إلى أنه بالنسبة إلى العديد من المنظمات، فإن مجرد الحصول على مستويات أعلى من المستوى الأول يُعدّ إنجازاً، وكون المستوى الخامس بعيد المنال عنها دون عمالة ماهرة، وموارد إضافية، وأدوات جديدة.

الجدول (٣-٣)

المستويات الخمسة للبيانات المفتوحة والبيانات المرتبطة

المستوى	الصورة / الشكل	الفوائد	التكاليف
1	غير مقروءة آلياً	البيانات متوافرة.	يتم قفل البيانات في الوثيقة ويصعب تحريرها لاحقاً.
2	مقروءة آلياً ولكن بصيغة حصرية (على سبيل المثال، ملف إكسل).	يمكن تحليل البيانات باستخدام البرمجيات الحصرية / الاحتكارية؛ كما يمكن تصديرها إلى صيغ أخرى.	تعتمد على برامج حصرية للوصول إليها واستخدامها.
3	مقروءة آلياً باستخدام صيغة عامة غير حصرية (على سبيل المثال، صيغة سي إس في - CSV)	يمكن تحليل البيانات باستخدام أية حزمة برمجية.	هل البيانات على الشبكة، ليست البيانات على الشبكة، وليست مرتبطة بطبيعتها، ومن ثم هي معزولة.
4	مقروءة آلياً باستخدام صيغة عامة غير حصرية واستخدام المعارف ووصف الموارد (URIs and RDF)	يمكن الوصول إلى البيانات من أي مكان على شبكة الإنترنت، ومن السهل ربطها جنباً إلى جنب مع غيرها من البيانات، وإمكانية وصلها بالأدوات والمكتبات القائمة.	يمكن أن تزيد زمن تحضير البيانات وإدارة البيانات وتبادلها.
5	مقروءة آلياً باستخدام صيغة عامة غير حصرية واستخدام المعارف ووصف الموارد ومرتبطة بالبيانات والبيانات الوصفية الأخرى	كما في المستوى الرابع، ولكن تصبح البيانات أكثر اكتشافاً وتمنح المستخدمين وصولاً كاملاً إلى مخطط (أنطولوجيا) البيانات	تحتاج إدارة فعالة للبيانات للحفاظ على الروابط الداخلة والخارجة.

المصدر: معدلة عن <http://sstardata.inio> بواسطة Michael Hausenblas.

الدَّعْوَى نحو البيانات المفتوحة:

في حين كان لدى بعض البلدان، مثل الولايات المتحدة، تقليد عريق في جعل بعض بيانات القطاع العام متاحة بحرية، وخاصة البيانات ذات الفائدة العالية مثل بيانات التعداد، ورسم الخرائط، والطقس، فإن إتاحة الوصول، في بعض البلدان الأخرى، يتم بشكل عام من خلال نموذج تغطية التكاليف الذي يتيح البيانات بمقابل مالي وبموجب اتفاقية ترخيص معتمدة. ففي المملكة المتحدة، تخضع الوثائق الحكومية لحقوق التأليف والنشر الملكية (Crown Copyright) وتُحكم البيانات ذات الفائدة العالية بواسطة الصناديق التجارية (على سبيل المثال، بيانات الخرائط ضمن هيئة المسح الجغرافية، وبيانات أحوال الطقس في مكتب الأرصاد الجوية) التي تعمل بصفة كيانات احتكارية، وتتحكم في الوصول إلى مجموعات بيانات رئيسية (Arthur and Cross 2006; Pollock 2006). وعلى مدى العقدين الماضيين، تم تجميع مجموعة من الحجج لبناء دَعْوَى إعادة صياغة دور هذه الصناديق وفتح بيانات القطاع العام للجمهور. وبصفة عامة، يمكن تقسيم هذه الحجج والاستحقاقات في خمسة أشكال، والتي ينظر إليها على أنها ذات فائدة لكل من حامل البيانات والمستخدم النهائي.

أولاً: حيث أن الكثير من البيانات في إطار الهيئات العامة والمنظمات غير الحكومية تتعلق بالعمليات التشغيلية لتلك الهيئات، فهي توفر وسيلة يمكن من خلالها قياس مدى نجاح مختلف البرامج والأنشطة الخاصة بها. ولذا فإن فتح هذه البيانات للتدقيق والرقابة العامة سيجعل أعمال المنظمة وصنع القرار فيها تتمتع بالشفافية ويمكن استخدامه لتعزيز المساءلة وتقييم القيمة الناتجة مقابل المال المستثمر (Janssen 2013; Gordon 2012). وقد تزايدت أهمية هذه الشفافية والمساءلة في الخطاب العام في عصر سياسات التقشف ومحدودية الموارد، وفي سياق المناقشات الجارية حول المشتريات، والهدر، والعجز الظاهر في الخدمات العامة.

ثانياً: يسمح إتاحة الوصول للمستخدمين النهائيين لبيانات المنظمة، على زعم البعض، برفع مستوى الوعي حيال القضايا المثارة، مما يسهل الاختيار وصنع القرار فيما يتعلق بالخدمات العامة، ويشجع على المشاركة الفاعلة والواعية في المجال العام

(Janssen 2012; Yiu 2012). وعلاوة على ذلك، فإن إتاحة الوصول للمستخدمين النهائيين لبيانات المنظمة يعزّزُ المواطنة الفاعلة والمشاركة السياسية في تشكيل كيفية أداء وتنظيم الحكم المحلي (Huijboom and Van der Broek 2011). وهذه، بدورها، تشجّع الابتكار الاجتماعي، وتعمل على تعزيز العلاقات بين أفراد المجتمع، ورفع مستوى النقاش العام (Yiu 2012)، ونتيجة لذلك، تمكّن البيانات المفتوحة من بناء الديمقراطية القائمة على المشاركة.

ثالثاً: يشجّع فتح البيانات حول المنظمة وأدائها على استخدامها هذه البيانات للاستفادة في تحسين الكفاءة التشغيلية والإنتاجية من خلال المراقبة واتخاذ القرارات بالأدلة والبيانات المثبتة. وعلاوة على ذلك، تحصل وحدات الأعمال داخل المؤسسة على حق الوصول إلى البيانات وإمكانية استخدامها عبر المنظمة بكاملها، مما يؤدي إلى رؤى ومعارف جديدة ومزيد من التفكير المشترك والكفاءة (Northcutt 2012; Verwayen et al. 2011). وإضافة لذلك، فإنه يمكنهم اكتساب ملاحظات وردود قيّمة وتلقّي النصح والمشورة من الوكالات الخارجية التي تقوم باستخدامها، وتحليلها، وتفسيرها، مما يرفع من مستوى الجودة، والنزاهة، والفائدة من البيانات. وتؤدي هذه التحليلات الداخلية والخارجية إلى تعزيز الحوكمة التنظيمية وإدارة المجتمع على نطاق أوسع، على سبيل المثال التصدي للغش والتزوير والجرائم الأخرى من خلال إنشاء خدمات أكثر فعالية (Huijboom and Van der Broek 2011). كما أنها تمكن من الرصد والتقييم المستمرين لأثر السياسات والبرامج الجديدة (مؤسسة المعرفة المفتوحة 2012 Open Knowledge Foundation).

رابعاً: يُمكن توفير البيانات المفتوحة من إثراء العلامة المميزة، وإتاحة البيانات بحرية يرفع من شأن المنظمة، ويدلّل على ريادتها، وروح المبادرة فيها، ويخدم مهمتها العامة، ويزيد مستوى الاتصال والتفاعل مع العملاء، والمستخدمين النهائيين ويقود التوجه إلى الخدمات المؤسسية (Verwayen et al. 2011)، وفي المقابل، يبني هذا الثقة والسمعة، ويعمل على تكوين انطباع عن المعايير الفاضلة للمؤسسة.

خامساً: بينما تمتلك بعض مجموعات البيانات المنشئة والقائمة في العلن قيمة اقتصادية هامشية، على سبيل المثال تلك المتعلقة بالتراث الثقافي، فإن الكثير من مجموعات البيانات الأخرى لديها قيمة تجارية كبيرة. ويمكن استخدام البيانات التي تم إنشاؤها علناً لإضافة قيمة إلى البيانات التجارية القائمة، وإنشاء تطبيقات وخدمات جديدة ومن ثم أسواقاً جديدة، وتحسين المعرفة واتخاذ القرارات للأعمال (Janssen 2012; Yiu 2012). ولذا فليس من المستغرب عندئذ حرص المصالح الصناعية على الدعوة لفتح البيانات المنتجة علناً لإعادة استخدامها في القطاع التجاري، وخاصة البيانات التي طالما عُرفت باحتوائها وإمكانية إنشائها لقيمة كتلك التي تُدار من قبل صناديق تداول القطاع العام (على سبيل المثال، بيانات الخرائط). فهناك من يزعم بأن تقييد الوصول إلى البيانات العامة، والتي تم إنشاؤها من أجل المصلحة العامة، يخنق الابتكار ويعمل على فقدان وزن مميّز للاقتصاد؛ أي أن حصر استخدامها فقط لمن لهم القدرة على تحمّل تكلفتها، هو تسعير لإخراج الآخرين من السوق (Pollock 2006; Yiu 2012). ولذا فإن الاستثمار العام في إنتاج البيانات لم يستغل بالشكل المناسب كما تم فقدان العديد من الفرص التي من شأنها أن تضيف قيمة على نشاطات القطاع الخاص (Yiu 2012). وإضافة إلى ذلك، فقد تأثر الارتفاع على الاقتصاد الشامل نتيجة فقدان القيمة الناتجة عن فتح البيانات العامة (Northcutt 2012).

وتتفاوت كيفية احتساب أوزان هذه الحجج الخمس عندما يتم تقديمها عبر السلطات القضائية والسياقات المختلفة؛ ومع ذلك، فهي تشكّل مجتمعة نظاماً ذا منطق قوي حقّق نجاحاً ملحوظاً في إقناع المنظمات لفتح البيانات الخاصة بها، لأسباب ليس أقلها أنها تتوافق مع العمليات الجديدة للخصخصة (Newliberalism) التي تعزز تسويق الخدمات العامة (Bates 2012). وفي الواقع، فإن المحرك للحجج المتعلقة بفتح البيانات في الغالب هي المصالح السياسية والتجارية التي تهدف إلى دفع المؤسسات العامة والمنظمات غير الحكومية للحصول على وصول إلى سلعة ثمينة، بدلاً من السعي لتحدي مفهوم الملكية الفكرية في حدّ ذاتها (Verwayen et al. 2011).

في المقابل، ومما لا يثير الدهشة في هذا السياق، أن الحجج والدعاوى لقطاع الأعمال لإتاحة بياناتها بحرية لإعادة استخدامها على نطاق أوسع لم يجر تطويرها والترويج لها

على نحو كاف. ومع ذلك، فإن العديد من الحجج نفسها لفتح بيانات القطاع العام تصلح لبيانات الأعمال أيضاً، أو على الأقل عناصر مختارة من هذه البيانات. ففتح البيانات من شأنه أن يمكن الشركات من الاستفادة من التحليل والرؤى التي تنتج بواسطة الغير من المستخدمين النهائيين، والجمع بين البيانات ومجاميع البيانات الأخرى التي تعمل على إثرائهم بطرق مثمرة، وتعزيز التعاون مع الشركاء والموردين مما يصنع فعالية في مجال الخدمات اللوجستية والمنتجات الجديدة، وبناء علاقة مع العملاء تساعد على توسيع قاعدة العملاء وتعزيز الملف التعريفي للشركة وسمعتها (Deloitte 2012). وعلاوة على ذلك، فإنها قد تساعد على تطوير الشراكات بين القطاعين العام والخاص، وفي هذا السياق تمّ اعتماد نهج الابتكار المفتوح، والذي حقق نجاحاً فيما يتعلق بالبرمجيات المفتوحة، والعمل على مبدأ أن تقاسم الموارد سينتج المزيد من القيمة التجارية دون الحرص على اكتنازها بغيرة والتحرز عليها. ففي الحالات التي تلقت الشركات الخاصة أموالاً عامة للمساعدة في البحث والتطوير، يمكن إقامة دَعْوَى على أن أي بيانات تنتج عن مثل هذه المحاولات ينبغي أن تكون مفتوحة إما بشكل كامل وإما بشكل انتقائي كوسيلة لإنتاج سلعة عامة لمثل هذا الاستثمار.

اقتصاديات البيانات المفتوحة:

هناك عدد من العوامل الكابحة / المشبّطة لفتح البيانات، فبعض هذه العوامل تتصل بالمخاوف المتعلقة بقضايا الضمان والمسؤولية، والمخاوف بشأن الخصوصية والأمن، والحرص المحتمل على جودة البيانات أو مداها، والكفاءة الفنية لبناء واجهات مناسبة لبرمجة التطبيقات (Applications Programming Interfaces- APIs) وإدارة البنية التحتية. وحتى الآن، فالأمر الأكثر إثارة للقلق، مع ذلك، هو الشأن المالي. فالبيانات قد تكون غير متنازعة بطبيعتها، مما يعني أنه من الممكن توزيعها بتكلفة هامشية، نظرياً على الأقل، ولكن ينبغي الدفع مقابل النسخة الأولية جنباً إلى جنب مع إدارة البيانات القائمة وخدمة العملاء (Pollock 2006). وعلى هذا النحو، قد تمثل البيانات المفتوحة مورداً مجانياً للمستخدمين النهائيين، ولكن إنتاجها ومشاركتها والمحافظة عليها هي بالتأكيد ليست دون تكلفة معتبرة (خصوصاً فيما يتعلق بالتقنيات المناسبة والموظفين المهرة). ففي كثير من

الحالات، كانت هذه البيانات أيضاً مصدراً رئيسياً للدخل للمنظمات، وفي حالة الشركات ميزة تنافسية. زمن ثم يركز السؤال الأساسي حول كيفية تمويل مشاريع فتح البيانات على نحو مستدام في ظل عدم وجود مصادر الدخل المباشر لها. ولذا هناك وجهان لهذه المعضلة في التمويل: فمن ناحية هناك التمويل اللازم من قبل الدولة لإنتاج البيانات وجعلها مفتوحة. ومن ناحية أخرى هناك التمويل للحفاظ على المبادرات التي يقودها المواطنون، والذي يعتمد على العمل التطوعي والمنح، ونماذج الأعمال التي من شأنها تمكين الشركات التي تستخدم البيانات المفتوحة من الازدهار والتطور.

ويختلف تمويل خدمات البيانات الحكومية بين البلدان والوكالات، ففي كثير من الحالات، تُفرض الضرائب لإنتاج ومعالجة البيانات. ولكن في بعض النظم القانونية والحالات، تكون خدمات البيانات معقدة بسبب أربعة عوامل: الأول، أن إجراءاتها يتم بالتعاقد عليها مع أطراف ثالثة لإدارتها وتشغيلها بالنيابة عن الدولة، حيث يضيف الطرف الثالث قيمة مخصصة احتكارية أو يجعل البيانات متاحة مقابل رسم مالي. وقد حدث هذا مؤخراً مع الرموز البريدية الأيرلندية المرتقبة التي سيتم إدارتها من قبل إحدى الشركات نيابة عن الدولة وستمول هذه العملية من خلال بيع وترخيص البيانات. الثاني، أن البائعين الآخرين يتكثرون بنشاط في سبيل وقف تقديم البيانات المفتوحة لأنه يدمر نموذج أعمالهم. ثالثاً، أن بعض أجهزة الدولة تعمل باعتبارها صناديق تداول تجارية، فهي لا تحصل على كل تمويلها من عائدات الضرائب، ولكن جزءاً كبيراً من دخلها قائم على بيع البيانات. فهئية المساحة الأيرلندية، على سبيل المثال، تعمل بهذه الطريقة كون أقل من نصف دخلها يأتي مباشرة من الدولة في شكل إعانة، ومن المسلم به أن بعض هذه الدفعات التي تتلقاها تأتي من وكالات حكومية أخرى، ولكنها تأتي أيضاً من المؤسسات الخاصة وعمليات الشراء الفردية. فإتاحة جميع البيانات الخاصة بها مجاناً يقوّس قدرتها على العمل وتمويل الخدمات الجارية لها. رابعاً، أن جعل البيانات مفتوحة لا يتمثل بنشرها في شكل محدد من قبل الدولة فقط. فالكثير من البيانات تحتاج إلى إعادة توظيفها وتنسيقها لتمكينها من أن تكون مفتوحة (على سبيل المثال، إبهام المصدر لها، وتجميعها) ووضع أنظمة جديدة في المكان المناسب لتحقيق هذا الهدف. وهذه ليست عملية هيئية، وفي أوقات التقشف وخفض النفقات يعني هذا إعادة تخصيص التمويل لتغطية ثمن هذا العمل، الذي هو ضروري

أيضاً للخدمات الأساسية. وتعني هذه التعقيدات الأربعة أن التصريحات، في كثير من الأحيان، بأن جميع بيانات الدولة قد تم بالفعل تغطية كلفها من خلال فرض الضرائب، وأنه ينبغي أن تكون متاحة بحرية، هي ليست بسيطة كما يُراد لها أن تكون. أما فيما يتعلق بالمبادرات التي يقودها المواطنون، فهذه أيضاً لديها تكاليف فيما يتعلق بالتوظيف والمعدات والخدمات إذا أُريد لها أن تكون مستدامة على المدى الطويل. فكيفية تأمين هذه الموارد خارج نطاق العمل التطوعي والهدايا تُشكل تحدياً صعباً. حيث تعتمد هذه المنظمات، في الوقت الحاضر، إلى حد كبير على التبرعات الخيرية وتمويل الدولة، والتي لا يحمل أيّاً منها ضمانات للتجديد والاستمرار. ومن ثم، فمن النتائج المحتملة المترتبة على التخفيضات أو التقلبات في القاعدة المالية لخدمات البيانات المفتوحة الانخفاض في جودة البيانات، وبطء الاستجابة، والابتكار، والأداء العام (Pollock 2006).

وعموماً يُجادل الدعاة للبيانات المفتوحة بأن الخيار الأفضل لتأمين قاعدة مالية مستقرة للبيانات المفتوحة من داخل الدولة وخارجها هو عن طريق الإعانات الحكومية المباشرة لتكلفتها. ويرى أنصار هذا النهج أن زيادة الإنفاق العام يعوّض في أربع طرق: الأولى، أن إتاحة الوصول المباشر إلى البيانات يحدّ من بعض التكاليف التي يتكبّدها المنتجون، مثل التوظيف المطلوب للتسويق، والمبيعات، والتواصل مع العملاء، ومراقبة شروط وأحكام الترخيص (Pollock 2006). الثانية، أنه يمكن للنموذج المفتوح الاستفادة من العمل الإضافي الحر والابتكار من جموع المستخدمين التي تضيف قيمة معتبرة إلى مجموعات البيانات وللمنظمات من ناحية جودة البيانات، وتحليلها، والمعرفة المستمدة، والمنتجات الجديدة والابتكارات، والعلاقات والشراكات التجارية الجديدة (de Vries et al. 2011). الثالثة، أن البيانات المفتوحة تنتج قيم فائضة متنوعة للمستهلكين، كما تولّد سلعاً عامة مهمة تستحق استثمار النفقات العامة فيها (Pollock 2009). الرابعة، أن البيانات المفتوحة تؤدي إلى منتجات مبتكرة جديدة من شأنها إنشاء أسواق جديدة، وهذا بدوره سوف ينتج مداخيل للشركات وإيرادات ضريبية إضافية. وستكون هذه الإيرادات الضريبية فائضة عن التكاليف الحكومية الإضافية في فتح البيانات، وبعبارة أخرى فإن نهج التكلفة الصفرية أو الهامشية هي أكثر فائدة على المدى الطويل من إستراتيجيات تغطية التكاليف واستردادها (المفوضية الأوروبية 2012) (European Commission 2012).

ولذا، يُعتقد أن فتح البيانات ينطوي على إمكانية إيجاد عدة عشرات المليارات من اليورو سنوياً في منطقة الاتحاد الأوروبي وحدها (مؤسسة المعرفة المفتوحة Open Knowledge Foundation 2012). ومن الأمثلة المستخدمة لدعم هذا الاعتقاد هو قرار الولايات المتحدة بإتاحة بيانات نظام تحديد المواقع العالمية والطقس للجُمهور بحرية (Pollock 2011; de Vries et al. 2006). حيث يدعم كل منها صناعات بمليارات الدولارات والتي تولّد إيرادات ضريبية كبيرة وتوظّف أعداداً كبيرة من العمال. وقد قدّرت الإيكونوميست (The Economist) في العام 2013م بأن ثلاثة ملايين وظيفة في الولايات المتحدة وحدها تعتمد على نظام تحديد المواقع. فلو أن أحداً قارن استخدام بيانات الطقس العامة في الولايات المتحدة وأوروبا، فهناك فرق شاسع في استخدامها وإنتاج قيمة مضافة عنها، حيث يشير Pollock (2006) إلى أنه في حين "تساوي الاقتصاديين تقريباً في الحجم، فإن صناعة أحوال الطقس التجارية في الولايات المتحدة هي أكبر بعشر مرات من نظيرتها في أوروبا كما أن صناعة إدارة مخاطر الطقس الوليدة هي أكبر مائة مرة منها في أوروبا". وبعبارة أخرى، تمّ استخدام البيانات العامة المتاحة علناً بالمجان أو بقيمتها الهامشية بطرق أنتجت عنها قيمة إضافية كبيرة. فأسواق البيانات المفتوحة تتعلق على حد سواء بأغراض مرتفعة (وهي قاعدة صغيرة من العملاء الذين يدفعون رسوماً كبيرة لقيمة مضافة عالية؛ على سبيل المثال، المستهدفين بشكل كبير بالتنبؤات الجوية)، وأغراض منخفضة (وهي تستهدف أعداداً كبيرة من المستخدمين طلباً لقيمة مضافة منخفضة حيث الخدمة مجانية، مدعومة من إيرادات الإعلانات؛ على سبيل المثال، تطبيقات حركة المرور) (de Vries et al. 2011). ويتضح التأثير المحتمل لفتح بيانات القطاع العام في أوروبا من خلال دراسة أجريت على واحد وعشرين مشروعاً للبيانات المفتوحة في عشرة بلدان والتي وجدت أن الانتقال إلى التكلفة الصفرية أو الهامشية لتكلفة العمليات التشغيلية زادت عدد المستخدمين ما بين ألف إلى عشرة آلاف في المائة، كما جذبت العديد من الأنواع الجديدة من المستخدمين، ولا سيما الشركات الصغيرة والمتوسطة (de Vries et al. 2011). وهذا يفصح عن أنه على الرغم من أن هناك بعض المكاسب الكبيرة، فقد بينّ Eaves (2013) أن غالبية الإيرادات الناشئة ستنتج عن سلسلة طويلة من الادخار والكفاءة والزيادة المتواضعة من التدوير.

وعلى الرغم من هذه الحجج، لم يصدر عن هيئة المحلفين حكم قاطع حول ما إذا كان فتح جميع بيانات القطاع العام ممكناً اقتصادياً وبشكل مستدام، وخاصة في المدى القصير. فهناك بالتأكيد حالة تتعلق بأن البيانات المفتوحة من الممكن تسويقها، مما يؤدي إلى منتجات عالية ومنخفضة المستوى والتي يمكن بالتأكيد أن تنتج مصادر للدخل وفرص العمل. ومع ذلك، وكما بين (de Vries et al. (2011)، فإن متوسط ما يكسبه مطورو التطبيقات من مبيعاتها سنوياً ليس سوى ثلاثة آلاف دولار فقط، مع الأخذ بالحسبان أن ثمانين بالمائة من تطبيقات الأندرويد المدفوعة يتم تنزيلها أقل من مئة مرة. وبالإضافة إلى ذلك، فقد لاحظوا أنه حتى التطبيقات الناجحة مثل ماي سيتي واي (MyCityWay) التي تم تحميلها أربعين مليون مرة، لم تنتج أرباحاً بعد. وبدلاً من ذلك، يستثمر أصحاب رؤوس الأموال في مشاريع محتملة الربح مع الحاجة إلى نموذج مستدام للأعمال. وقد يكون من الجيد أن يستغرق الأمر وقتاً لتطوير ابتكارات وأسواق جديدة؛ فعلى سبيل المثال، استغرقت الصناعات القائمة على نظم تحديد المواقع سنوات عديدة لتزدهر وتنضج بعد قرار جعل البيانات متاحة علناً والذي اتخذ في العام 1984م. كما قد يكون الحال أن بعض البيانات لها قيمة كبيرة جداً ولكن دون جدوى اقتصادية، مما يستوجب عادة عدم استمرار إعانتها مالياً إذا أريد لها أن تبقى مفتوحة بطبيعتها.

وفي ظل غياب طرح الاكتتابات الحكومية العامة لجميع التكاليف المتعلقة بفتح البيانات الخاصة بها، يجري النظر في حلول تمويلية محتملة أخرى ولا سيما نماذج الأعمال المختلفة. فقد بين (Ferro and Osella (2013) ثمانية نماذج مختلفة لتستخدم، أو قد تستخدم لتمويل مبادرات البيانات المفتوحة (انظر الجدول ٤-٣). ولا تستبعد بعض هذه النماذج المختلفة بعضاً. في حين تعمل نماذج شبه - مفتوحة أخرى على توفير البيانات مجاناً لإعادة الاستخدام غير التجاري، وإلزام معيدي الاستخدام للأغراض الربحية بدفع مقابل مادي، أو الدخول في شراكات بين القطاعين العام والخاص حيث يزود القطاع العام البيانات وتقدم الشركات الخاصة التمويل والخدمات المضافة لإتاحة الوصول وحقوق إعادة الاستخدام (OECD 2008). وفي نموذج آخر، يمكن بناء تحالف يمتلك مجموعة البيانات، وقرق العمل، والموارد، والأدوات، ويسهل بناء القدرات، ولكن يتقاضى رسوم العضوية من أعضاء التحالف لتغطية الخدمات ذات القيمة المضافة المشتركة. ويتوقف اعتماد أي نموذج، أو

مجموعة من النماذج، على المجموعة المستهدفة لتحمل تكلفة إنتاج وصيانة البيانات - المستخدمين، المحدثين / البائعين او الحكومة - وإذا كانت المنظمة تسعى لاسترداد التكاليف كاملة أو جزء هامشياً منها، أو حتى قيمة أكبر من التكاليف الكاملة التي يمكن استثمارها مرة أخرى في الخدمة (Pollock 2009; Ferro and Osella 2013).

الجدول (٣-٤)

نماذج التمويل لمبادرات البيانات المفتوحة

الوصف	النموذج
تقدّم للمستخدمين النهائيين منتجاً عالياً المستوى أو الخدمات التي تضيف قيمة إلى البيانات (على سبيل المثال، البيانات المشتقة، الأدوات والتحليل) بمقابل مادي، كدفعة مقطوعة ثابتة، أو رسوماً متكررة، أو الدفع - مقابل - الاستخدام، ودون استخدام حقوق احتكارية. وهذا يتيح لمنتج البيانات الحصول على مزايا المبادرة والسبق في مجال تسويق وبيع السلع التكميلية.	المنتج المميز / خدمة Premium product / service
تقدّم للمستخدمين النهائيين مجموعة متدرجة من الخيارات، متضمناً ذلك خياراً مجانياً يشمل عناصر أساسية (على سبيل المثال، ميزات محدودة أو عينات من مجموعات البيانات)، مع خيارات أكثر تقدماً تحقق قيمة مضافة بمقابل رسم مالي. تفتح المنتج / الخدمة على سوق أكبر اتساعاً وأكثر شعبية والمزيد من الاستخدام العرضي الممتد، مع الاحتفاظ بالخدمات / المنتجات المدفوعة، عالية المستوى، المتاحة للمستخدمين الأكثر تخصصاً.	المنتج المجاني / خدمة Freemium product / service
تقدّم منتجات / خدمات للمستخدمين النهائيين مجاناً دون مقابل، حيث تُدعم من خلال سبل التمويل الأساسية الأخرى أو غيرها من المنتجات / الخدمات.	المنتجات مفتوحة المصدر Open source

<p>يعرض المنتج او الخدمة ابتداء بالمجان أو بقيمة مخفضة (ماكينة الحلاقة) مما يشجع على المتابعة بالاستخدام المدفوع بمقابل (الشفرات). وقد يكون إتاحة الوصول إلى واجهات برمجة التطبيقات بالمجان، ولكن يكلف هذا الاستخدام المحوسب المستخدمين على نموذج دفع المستحقات أولاً بأول (بحسب الاستخدام)، مع دعم هذا النموذج الأخير (الشفرات) للنموذج الأول (ماكينة الحلاقة).</p>	<p>نموذج الهدية الترويجية (البنى التحتية على نسق ماكينة الحلاقة والشفرات) Infrastructural razor and blades</p>
<p>تقدّم خدمات ذات قيمة مضافة عن طريق تنظيف، وتكرير، وتوحيد، وربط البيانات مما يوفر مجموعة قياسية من الأدوات التحليلية، وإتاحتها من خلال متجر شامل (متجر الشباك الواحد) وتحميل المستخدمين التكلفة من خلال التسعير المجاني او المميز (تتصاعد تدريجياً بخيارات 1 و 2 لاكتساب الكفاءة).</p>	<p>المنصة الموجهة بالطلب Demand-oriented platform</p>
<p>يُعطى مطورو البرمجيات حرية الوصول إلى البيانات لإنشاء الخدمات والمنتجات التي يمكنهم بيعها إلى هيئات القطاع العام، والتي يمكنها بدورها أن تجعلها متاحة للآخرين من خلال نموذج المنتجات المجانية.</p>	<p>المنصة الموجهة بالتزويد Supply-oriented platform</p>
<p>تُقدّم المنتجات / الخدمات مجاناً، ولكن يتلقى المستخدمون إعلاناً عند استخدام المنتج / الخدمة (المدرة للدخل) أو تقدّم المنتجات / الخدمات من قبل شركات مختلفة وتوصف بالعلامات التجارية لها لتشجيع استخدام منتجاتها وخدماتها الأخرى (الدعم المتبادل).</p>	<p>التوزيع المجاني مع الإعلان Free with advertising</p>
<p>يتم إنشاء منتج / خدمة مخصصة للعميل وتخصّص لاستخدامه الخاص، مع دفع العميل رسوماً لمرة واحدة أو اشتراك يتضمن تكاليف الصيانة والتحديث.</p>	<p>تطوير العلامة البيضاء White-label development</p>

(المصدر: جمعت من (Ferro and Osella (2013

المخاوف المتعلقة بالبيانات المفتوحة:

باستثناء الكيفية اللازمة لتمويل مستدام لمبادرات البيانات المفتوحة، والخسائر المحتملة للشركات التي قامت ببناء نماذج أعمال قائمة على إعادة بيع بيانات القطاع العام (de Vries et al. 2011)، سيظهر للوهلة الأولى أن هناك عدداً قليلاً من السلبيات الناتجة عن فتح البيانات العامة للجمهور لإعادة استخدامها، وأن هناك العديد من الإيجابيات. ومع ذلك، فقد بدأ في الظهور مؤخراً عدد من الانتقادات المضادة التي تزعم أن حركة البيانات المفتوحة ليست محمودة من الناحية السياسية أو الاقتصادية، حيث إن بعض العناصر زائفة في أهدافها، وعلاوة على ذلك، أن هناك عدداً من التبعات الضارة التي يمكن أن تنجم عن فتح البيانات ويتضرر منها بعض المواطنين. ولا توحى هذه الانتقادات بفكرة التخلي عن التوجه نحو فتح البيانات، ولكن تؤكد على وجوب أن تكون مبادرات البيانات المفتوحة أكثر وعياً وتعقلاً في إتاحة البيانات، وكيفية استخدامها، وطرق تمويلها. ويمكن تقسيم هذه الانتقادات إلى ثلاث فئات رئيسية: تسهيل البيانات المفتوحة للتوجه للخصخصة (تأييد الرأسمالية المطلقة وعدم تدخل الدولة في الاقتصاد) والهيكلية التسويقية للخدمات العامة؛ وأنها تروج لسياسة توسعية وتزيد سلطة المتسلط؛ وأنها تفتقر إلى الاستدامة، والأدوات، وسهولة الاستخدام.

الخصخصة والهيكلية التسويقية للخدمات العامة:

برهن Jo Bates (2012) على أن «المبادرات المفتوحة مثل البيانات الحكومية المفتوحة (Open Government Data - OGD) تظهر كعملية تاريخية، وليست وقائع محايدة». فكما هو الحال مع كل المبادرات السياسية، فإن سياسة البيانات المفتوحة ليست ببساطة بديهية أو محايدة، بل هي تركز على عقيدة سياسية واقتصادية. فحركة البيانات المفتوحة متنوعة وتتكون من مجموعة من الدوائر الانتخابية ذات أجندات وأهداف مختلفة، وليس الدافع من ورائها مجرد أي طرف واحد منها. ومع ذلك، بينت Bates بأن عامل الشد السياسي في حالة حركة البيانات المفتوحة في المملكة المتحدة كان ضئيلاً حتى بدأت الشركات الكبرى تنشط بتكتلها في حملات دعم البيانات المفتوحة، وبدأت المبادرات الحكومية المفتوحة تنسجم مع برامج التقشف القسري والهيكلية التسويقية للخدمات العامة. فبالنسبة لها،

قد استولت الأحزاب السياسية وقطاع الأعمال على حركة البيانات المفتوحة لصالح المصالح الرأسمالية المهيمنة تحت ستار أجندة / جدول أعمال الشفافية (Bates 2012).

وبعبارة أخرى، فإن جدول الأعمال الحقيقي لقطاع الأعمال هو الحصول على البيانات المنتجة بتكلفة باهظة دون أي مقابل، ومن ثم الوصول إلى البنية التحتية المدعومة بشكل كبير مما يمكنهم من الاستفادة من الأرباح، وفي الوقت ذاته، إزالة القطاع العام من السوق وإضعاف موقعه بصفته منتجاً لمثل هذه البيانات. ولذا، فبسبب اختفاء الدخل من خدمات البيانات (كما في الحالات التي جرى تمويلها من الرسوم وليس من دافعي الضرائب)، ومن ثم اختفاء التمويل اللازم لدعم الإنتاج والإدارة المحلية الداخلية، تم دفع هيئات القطاع العام إلى الاستعانة بمصادر خارجية لمثل هذه الخدمات من القطاع الخاص على أسس تنافسية أو التخلي عن إنتاج البيانات للقطاع الخاص التي عليهم بعد ذلك شراؤها منهم (Gurstein 2013). وهنا، يتعين شراء خدمات البيانات والبيانات المشتقة من البيانات المفتوحة المتاحة بحرية من منشئ البيانات، في الوقت نفسه الذي يتم إفراغ المعارف الأولية لبيانات المنظمة. وعلاوة على ذلك، وحيث إن البيانات المفتوحة تتعلق في كثير من الأحيان بأنشطة الهيئة نفسها، وخصوصاً عند اشتغالها على مؤشرات قياس الأداء الرئيسية، فهي تعمل على تسهيل إصلاح القطاع العام وإعادة تنظيمه مما يعزز الخصخصة، وروح الإدارة العامة الجديدة ومصالح القطاع الخاص (McClean 2011; Longo 2011). فمثل هذه العمليات، كما بينت (Bates 2013)، هي جزء من إستراتيجية سياسية مرسومة لفتح وإتاحة كافة الخدمات العامة تقريباً للمنافسة بين مزودي الخدمات من القطاع الخاص والقطاعات الأخرى، حيث يعمل فتح البيانات الخاصة بالخدمات العامة على تمكين مستخدمي الخدمة من استحداث خيارات معلومة ضمن سوق الخدمات العامة على أساس التطبيقات التي تعتمد على البيانات التي ينتجها مجموعة من مطورو القطاعات التجارية وغير التجارية (2013، التعبير الأصلي). وفي مثل هذه الحالات، فإن جدول أعمال الشفافية التي يروج لها السياسيون وقطاع الأعمال هو مجرد تعبير مجازي، فإذا كان أي من الطرفين مهتماً حقاً في الشفافية والحكومة المفتوحة فإنه سيكون داعماً بالتساوي للحق في حركة المعلومات والعمل المتعلق بإفشاء الأسرار (Janssen 2012) وكذلك تخفيف القيود عن حقوق الملكية الفكرية على نطاق أوسع (Shah 2013). وبدلاً من ذلك، فإن الحكومات والشركات بشكل عام مقاومة لهما على حد سواء.

السياسات التوسعية وتمكين المُمكن:

هناك جدل ذو صلة يتمثل في أن الغالب في حركة البيانات المفتوحة مدفوع من وجهة نظر فنية واقتصادية، ويركز بشكل كبير على إتاحة الوصول إلى البيانات، وليس على سياسات البيانات نفسها، وما قد تكشف عنه، أو كيفية استخدامها، ولمصالح من تصبّ (Shah 2013). وبعبارة أخرى، تسعى الحركة بشكل كبير إلى تقديم صورة تظهرها على أنها محببة وبديهية، وتعزيز الاعتقاد بأن فتح البيانات بطبيعته شيء جيد في حد ذاته. وبالنسبة لآخرين، فإن إتاحة الوصول إلى البيانات هو مجرد وجه واحد فيما يتعلق بمفهوم الانفتاح، وعلى القدر نفسه من الأهمية هو ما تتكوّن البيانات منه، والكيفية التي يمكن بها إنشاء مجتمع أكثر اعتدالاً وإنصافاً. فإذا كانت حركة البيانات المفتوحة تخدم فقط المصالح الرأسمالية من خلال فتح البيانات العامة، وتحفظ بالبيانات الاحتكارية الخاصة خلف أنظمة جدران الدفع (بمقابل رسوم الاشتراك) وحماتها بأنظمة الملكية الفكرية، وتمكن بصورة أكبر أولئك المُمكنين بالفعل وحرمان الآخرين منها، فإنها تكون قد فشلت في جعل المجتمع أكثر ديمقراطية وانفتاحاً (Gurstein 2011; Shah 2013).

ففكرة أن البيانات محايدة وموضوعية في طبيعتها تُطرح ضمناً في معظم المناقشات المتعلقة بالبيانات المفتوحة وأن لكل إمكانية الوصول لهذه البيانات واستخدامها (Gurstein 2011; Johnson 2013). بيد أن القضية ليست كذلك، فمع الاحترام لفتح البيانات نفسها يؤكد Johnson (2013) وجود درجة عالية من الامتياز الاجتماعي والقيم الاجتماعية الكامنة ضمناً في بيانات القطاع العام فيما يتعلق بماهية البيانات المنتجة، وتعلّقها بمن ولماذا (لا سيما في المجالات التي تعمل كأنظمة تخصصية، مثل أنظمة الرعاية الاجتماعية، وأنظمة إنفاذ القانون)، التي تتمثل مصالحها ضمن مجموعة البيانات وأيضاً تلك التي يتم استبعاد مصالحها. وهكذا فإن هياكل القيمة متأصلة في مجموعات البيانات والتي تشكّل بدورها التصميم والتفسير لها وربما تعمل على نشر المظالم وتعزيز المصالح المهيمنة.

فالمواطنون متفاوتون من حيث إمكانية الوصول إلى الأجهزة والبرامج المطلوبة لتحميل ومعالجة مجموعات البيانات المفتوحة، كما أن لهم مستويات مختلفة من المهارات اللازمة

لتحليل، وتأطير، وتفسير البيانات (Gurstein 2011). وحتى لو امتلكت بعض المجموعات القدرة على تقديم فهم مقنع للبيانات، إلا أنهم لا يملكون بالضرورة الاتصالات اللازمة لاكتساب التأييد الجماهيري والتأثير على النقاش، أو الحنكة السياسية بمقابل خصم جيد التجهيز والدهاء. ومن ثم، فإن الديمقراطية المحتملة للبيانات المفتوحة مفرطة في التفاؤل، إذا أخذ بالحسبان معظم المستخدمين الذين لديهم درجة عالية من المعرفة التقنية وملف تعريف سياسي قائم (McClean 2011)، فالواقع أنه يمكن للبيانات المفتوحة أن تعمل على زيادة تمكين السلطة وإعادة إنتاجها وتعميق اختلال توازن القوى (Gurstein 2011). ومن الأمثلة التي يستشهد كثيراً بها على هذا الأخير هو رقمنة سجلات الأراضي في ولاية كارناتاكا (Karnataka)، الهند، حيث عمل مشروع البيانات المفتوحة، والذي تمّ الترويج له باعتباره مبادرة لصالح الفقراء، على حرمان الفقراء من خلال تمكين ذوي الموارد المالية والمهارات من الوصول إلى البيانات التي سبق حجبها وإعادة اقتطاع أراضيهم بوضع اليد بغية الاستيلاء عليها (Gurstein 2011; Slee 2012; Donovan 2012). لقد سهّلت البيانات المفتوحة في هذه الحالة، وبعيداً عن العمل على مساعدة كل المواطنين، على التغيير في حقوق ملكية الأرض ونقل الثروة من الفقراء إلى الأغنياء. وبعبارة أخرى، فإن فتح البيانات لا يعني عملية متصلة في الديمقراطية وتعزيزها، بل إن الواقع أنه يُمكن للبيانات المفتوحة أن تعمل أداة للسلطة التأديبية (Johnson 2013).

ومن المهم هنا أن التنبّه إلى أن إنتاج البيانات الحكومية يكون لأغراض الحوكمة، وهي تتألف من نوعين رئيسيين: الأنشطة المتعلقة بأنشطة الدولة والأنشطة المتعلقة بالمواطنين، والأماكن، والأعمال. فالنوع الأول يهتم بكيفية عمل الدولة، وعند إتاحة هذه البيانات بشفافية، يمكن استخدامها لتقييم الأداء والمساءلة. وقد يقف قلة ضدّ جعل مثل هذه البيانات متاحة للجمهور، على الرغم من الحاجة لتقدير مثل سياسات المكاشفة هذه، خاصة إذا كانت التدابير المستخدمة ذات تأثير مشوّه لتقديم الخدمة وتعمل على التلاعب ببيانات التقييم (بالتوافق مع قانون كامبل -Campbell's Law- انظر الفصل السابع)، وهذا يؤثر بدوره سلباً في ماهية الخدمات المقدمة وجعل الحياة سيئة لأولئك الأشخاص الذين يتلقون الخدمة. ويتكوّن النوع الثاني من السجلات الشخصية والمؤسسية الحساسة للغاية، ولم يتم إنشاؤها بهدف مشاركتها وتبادلها مع الآخرين. وفي الواقع،

يتوقع المواطنون لها أن تكون محمية بموجب قوانين الخصوصية وحماية البيانات. وحتى عندما يتم تجريد هذه البيانات من البيانات التي تشير إلى هوية أشخاص معينين أي جعلها مجهولة (Anonymised) وعرضها مجملة، فإنه يمكن للبيانات أن تكون على درجة من الحساسية والسياسية بصورة كبيرة. فخذ، على سبيل المثال، بيانات الرعاية الاجتماعية والصحية المنقحة نسبياً والمجمعة في وحدات مكانية (على سبيل المثال، مستوى الحي)، فهذه البيانات تُمثل أداة لتوجيه التدخلات الموجهة الرامية إلى معالجة المعوقات الاجتماعية، كما أن لها إسهامات مفيدة في تحليل البيانات التي تسعى اجتماعياً إلى فرز المواطنين والتعريف عنهم فيما يتعلق بالقروض الائتمانية والتأمين ضد المخاطر، ويمكن من ثم استخدامها لإنشاء ملفات تعريف مناطقية والتي توصم موضعاً ما وتحد من الاستثمار فيه (انظر الفصل العاشر). وبعبارة أخرى، يمكن إعادة توظيف البيانات لأغراض مختلفة بطرق شتى والتي لها مخرجات متباينة، وهناك أسباب مشروعة لتوخي الحذر فيما يتعلق بما يتم الإفراج عنه من البيانات الحكومية ومقاومة شعار التبسيط المستخدم من قبل بعض المدافعين عن البيانات المفتوحة "إنها بياناتنا، لقد دفعنا مقابلها، وينبغي أن يكون لنا حق الوصول إليها".

الاستدامة والفائدة وسهولة الاستخدام:

تركز الاهتمام بشكل كبير، حتى الآن، على جانب توفير الوصول إلى البيانات وإنشاء مبادرات البيانات المفتوحة، وكان هناك عدم إيلاء اهتمام كاف لاستدامة المبادرات والكيفية التي يتم بها استخدام البيانات وتوظيفها. ففي دراسة أجريت على عدد من مشاريع البيانات المفتوحة المختلفة أوضحت (Helbig et al. (2012 بأن العديد منها تركز بشدة على الناحية التقنية الفنية، ويمكن فهم هذه المشاريع على أنها «تزيد قليلاً عن المواقع المرتبطة بملفات بيانات متنوعة، ودون أي مراعاة لسهولة الاستخدام، وجودة المحتوى، أو عواقب استخدامها». والنتيجة هي مجموعة من مواقع البيانات المفتوحة التي تعمل أكثر كحاويات للبيانات أو مكبات البيانات، بحيث تفتقر إلى الصفات المتوقعة في البنية التحتية العاملة للبيانات والمنظمة تنظيماً جيداً، مثل البيانات النظيفة، ذات الجودة العالية، والتي جرى التحقق من صحتها، وقابليتها للتشغيل المتبادل بما يتوافق مع معايير البيانات ولها التعريف الوصفي

المناسب ومجموعات تسجيل كاملة (الوثائق المرتبطة بها)؛ أي سياسات المحافظة والنسخ الاحتياطي والتدقيق؛ وسياسات إعادة الاستخدام والخصوصية والأخلاقيات، والترتيبات الإدارية وآليات تنظيم الإدارة والحوكمة؛ والاستقرار المالي و خطة طويلة الأجل للتنمية والاستدامة (انظر الفصل الثاني). كما تفتقد العديد من المواقع للأدوات المناسبة والمواد السياقية لدعم تحليل البيانات، وعلاوة على ذلك فإن غالب مجموعات البيانات المنشورة تكون صعبة الأخذ والاستفادة منها، حيث تتألف من تلك التي يسهل نشرها وتحتوي على بيانات غير حساسة كما أن لديها فائدة منخفضة نسبياً. وفي المقابل، غالباً ما تبقى البيانات الأكثر صعوبة وتطلباً لجعلها مفتوحة في حالة مغلقة، بسبب قضايا حساسيتها أو لأنها تتطلب المزيد من العمل الإداري للامتثال لقوانين حماية البيانات (Chignard 2013).

ويتمثل جزء من المشكلة في أن العديد من مواقع البيانات المفتوحة هي استجابة عنيفة وجاهرة لظاهرة ناشئة، وقد تمّ بناؤها من قبل متحمسين ومنظمات ممن لديهم خبرة قليلة في أرشفة البيانات أو الاستخدام السياقي للبيانات التي تمّ فتحها. وقد تمّ دعمها والترويج لها من خلال الملتقيات المكثفة لمطوري البرمجيات (الهاكاثونات، Hackathons) ومنتديات العطل الأسبوعية لمجموعات الاختصاص (Data Dives)، التي تعيد طرح وإنتاج العديد من هذه القضايا. وكما أكدّ (McKeon 2013) and Porway (2013) فإن هذه الأحداث والتجمعات التي تدعو المبرمجين والأطراف المعنية الأخرى لبناء تطبيقات باستخدام البيانات المفتوحة لها فائدة محدودة. ففي حين أنها تركز الاهتمام على البيانات وتُعد وسيلة جيدة للتواصل، إلا أن أولئك الذين يقومون بالبرمجة غالباً ما يفتقرون إلى المعرفة السياقية المعمّقة فيما يتعلق بما تشير البيانات إليه، أو ينتمون إلى مجموعة سكانية معينة ليست بالضرورة معبّرة عن شريحة مجتمعية أوسع نطاقاً (على سبيل المثال، الشباب والمتعلمون والتقنيون)، ويظنون أن المشاكل الهيكلية العميقة يمكن حلّها من خلال الحلول التقنية. إنهم «كيان مؤلف من مجتمع صغير من المتطوعين المتحمسين، وليسوا من الأشخاص الذين لديهم مصلحة كبيرة في رؤية المشروع ينجح» (McKeon 2013). وعلاوة على ذلك، لا تزال الحلول التي تمّ إنشاؤها في هذه اللقاءات على إصدارها الأول، مع قليل من المتابعة بعد انتهاء الحدث ودون تحديث أو صيانة أو تطوير لاحق. وكان جواب (Porway's 2013) لبناء حلول تشرع لتغيير مناسب هو الجمع بين أولئك الذين يفهمون

البيانات والمشكلات القائمة على أرض الواقع مع أولئك الذين يستطيعون برمجة التطبيقات وبنائها والمحافظة عليها، وقد اقترح (McKeon 2013) عقد اجتماعات تخطيطية مسبقة وتوصيات توجيهية بعد الحدث تهدف إلى استدامة المخرجات وإنشاء مجتمع على المدى الطويل. وبعبارة أخرى، هناك حاجة للتوفيق بين الخبرات الموجودة والخبرات في تطوير التطبيقات. والحاجة نفسها مطلوبة مع مواقع البيانات المفتوحة، حيث تتطلب تعلم الدروس المستفادة من أولئك الذين عملوا لفترة طويلة سابقة في الأرشيف وقطاع أعمال تبادل البيانات، متضمناً ذلك أمناء المحفوظات الأرشيفية والعلماء والجغرافيون.

ونتيجة لهذه القضايا العويصة المختلفة، وبدلاً من إنشاء دورة إيجابية (Virtuous Cycle) بحسب ما تفترض حركة البيانات المفتوحة، حيث الإفراج عن المزيد والمزيد من مجموعات البيانات، في المزيد من الأشكال والصيغ، ينتج عنه مزيد من الاستخدام، ومن ثم الإفراج عن المزيد من البيانات مستقبلاً، إلا أن العديد من المواقع، كما لاحظت (Helbig et al. 2012) عانت من انخفاض معدلات التصفح لها وتراجعها لأنها لا تشجع على الاستخدام أو لا تقدم التسهيلات للمستخدمين في هذا الخصوص، وتقتصر خدماتها على عوامل أخرى مثل ممارسات إدارة البيانات وعرض الجهد المؤسسي المبذول والسياسات الداخلية المعمول بها. فبعد الشرارة الأولى من الاهتمام، تراجع استخدام البيانات بشكل ملحوظ للغاية بسبب الكشف عن القيود المفروضة على البيانات ومحدوديتها ونضال المستخدمين من أجل إيجاد طرق لكيفية استخدام وتحليل البيانات بشكل مريح. وقد لاحظ (McClean 2011)، على سبيل المثال، أن التحليل الناشئ عن البيانات المفتوحة كان له تأثير محدود على النقاشات السياسية، وخلص فيما يتعلق بنظام المعلومات المجمعة على الإنترنت (كوينز COINS) الخاص بالبيانات المالية الحكومية في المملكة المتحدة، إلى أنه بعد

هبة قصيرة من اهتمام وسائل الإعلام في منتصف عام 2010، وفي أعقاب إطلاقه.... فإن التقارير التي تشير صراحة إلى نظام كوينز غدت الآن نادرة للغاية، كما أن العاملين في الصحافة الذين كانوا الأكثر اهتماماً بالحصول على حق الوصول إليه قد صرحوا بأنه لم تثبت فائدته كمحرك فاعل للصحافة على وجه الخصوص.

وحيث يتم نشر البيانات دورياً على سبيل المثال، بصورة فصلية أو سنوية، فإن الاستخدام يميل إلى أن يكون دورياً وغالباً مرتبطاً بمشاريع محددة مثل التقارير الاستشارية بدلاً من وجود نمط أكثر ثباتاً من الاستخدام الذي قد يتوقعه المرء لبيانات تتغير بطبيعتها آتياً بصورة مستمرة أو لديها قيمة اقتصادية عالية. ففي مثل هذه الحالات، لاحظت Helbig et al. (2012) أن مجموعة من حلقات ردود الفعل السلبية قد أبطأت توفير البيانات واستخدامها، ومن ثم زيادة خفض الاستخدام. ولذا، فبعد بعض «المكاسب السريعة» الأولية، يكمن الخطر في تحوّل أي دورة إيجابية من كونها إيجابية إلى سلبية، وتقويض الأساس المنطقي لتمويل الحكومة المركزية لمثل هذه المبادرات وكونها عرضة للتوقف المفاجئ. إن المفتاح لتجنب نشوء مثل هذه الدورة السلبية هو التأكد من تركيز أي مبادرة بأكبر قدر ممكن على جانب الطلب كما في جانب العرض، وأنها توفّر للمستخدمين بيانات قابلة للتشغيل المتبادل والأدوات التحليلية وغيرها من الخدمات التي تسهل استخدام وإضافة قيمة للبيانات، بدلاً من مجرد ربط الملفات.

الخلاصة:

عند أحد المستويات، تُعد قضية البيانات المفتوحة والبيانات المرتبطة بديهية - حيث تُسهّم البيانات المفتوحة في صنع الشفافية والمساءلة والمشاركة والاختيار والابتكار الاجتماعي والكفاءة والإنتاجية وتعزيز الحوكمة والابتكار الاقتصادي وتكوين الثروات. في حين تعمل البيانات المرتبطة على تحويل المعلومات الموزعة عبر شبكة الإنترنت إلى شبكة دلالية يمكن من خلالها القراءة الآلية للبيانات وربطها معاً. وعليه تحمل البيانات المفتوحة والبيانات المرتبطة في طياتها الكثير من الوعود والآمال والقيمة مع اعتبارها مجازفة محفوفة بالمخاطر، فواقع الحال بالنسبة للبيانات المفتوحة والبيانات المرتبطة أكثر تعقيداً، كما أن الأسس الاقتصادية لها ليست واضحة على الإطلاق. ويبدو أن للبيانات المفتوحة والبيانات المرتبطة تكاليف هامشية قليلة، ولكن الحاجة قائمة في عملية إنتاجها والطاقت الفني والجهاز المؤسسي لتسهيل استخدامها والحفاظ عليها مما يلزم تكاليف حقيقية من حيث العمالة، والمعدات، والموارد. وفي الواقع، فإن البيانات المفتوحة بعيدة كل البعد عن كونها مجانية، ولكن حتى الآن، لم يتم وضع نموذج متماسك لتمويل مستدام

لمبادرات البيانات المفتوحة دون إعانة حكومية. كما أن فوائد البيانات المفتوحة من حيث إنتاج قيمة مضافة على فائض الاستهلاك والمشاريع التجارية الجديدة، والمنتجات المبتكرة، والتكاليف قد قابلها إيرادات ضريبية إضافية افتراضية إلى حد كبير أو بدأت بالتحقق فقط، وعلاوة على ذلك، فإن العواقب السلبية المحتملة لفتح البيانات لم تستكشف بعد بالكامل.

فهناك حاجة إلى الدفع بمزيد من الاهتمام النقدي لكيفية تطور مشاريع البيانات المفتوحة كظاهرة تقنية اجتماعية معقدة ومرتبطة بطائفة متنوعة من البرامج، والمعنيين، وأصحاب المصلحة. وحتى الآن، تمحورت الجهود حول العمل السياسي والتقني المتعلق بإقامة مشاريع البيانات المفتوحة، ولم يتم العمل بشكل كاف على دراسة هذه التحركات المادية والتي تلامس العديد من الجوانب وعواقبها. ونتيجة لذلك، فإننا نفتقر إلى دراسات الحالة التفصيلية لتطبيق مشاريع البيانات المفتوحة، والتجمعات المحيطة بها، والتي تعمل على صياغتها، والطرق الفوضوية، والطارئة والعلائقية التي تتكشف بمرور الوقت. ولن تظهر صورة أكثر اكتمالاً للبيانات المفتوحة والبيانات المرتبطة إلا من خلال مثل هذه الدراسات التي ستكشف عن كل من الإيجابيات والسلبيات لمثل هذه المشاريع وتوفر إجابات عن أسئلة معيارية بشأن الكيفية التي ينبغي تنفيذها بها ومعالجة قضايا الاستدامة وسهولة الاستخدام والسياسات المتعلقة بها.

الفصل الرابع

البيانات الكبيرة (Big Data)

تتبع (2012) Francis Diebold منشأ استخدام مصطلح «البيانات الكبيرة» إلى منتصف التسعينيات (1990s)، حيث استخدم لأول مرة من قبل جون ماسي (John Mashey)، كبير العلماء السابق المتقاعد من شركة سيليكون غرافيكس، للإشارة إلى معالجة وتحليل مجموعات البيانات الضخمة. وقد كان المصطلح قليل الجذب في البداية. ففي العام ٢٠٠٨ استخدم عدد قليل جداً من الناس مصطلح «البيانات الكبيرة»، سواء في المجالات الأكاديمية أو الصناعية. وبعد خمس سنوات أصبح مصطلح «البيانات الكبيرة» التعبير الطنان الأكثر شيوعاً في الاستخدام في دوائر الأعمال ووسائل الإعلام الشعبية، مع تعليقات دورية في الصحف الأكثر انتشاراً، مثل صحيفة النيويورك تايمز (New York Times) والفاينانشال تايمز (Financial Times)، ومقاطع وأعداد خاصة في المجلات العامة والمجلات العلمية مثل عالم الاقتصاد (Economist) والوقت (Time) والطبيعة (Nature)، والعلوم (Science). لقد كان هذا الانتشار وزيادة الاستخدام المرتبطة به هو ما أدى بالفعل إلى إعلان غارتر (Gartner) في يناير (2013) عن انتقال تقنية البيانات الكبيرة على طول دورة الدعاية (Hype cycle) (كما يطلق عليها نموذج النضج التقني) من مرحلة «ذروة تضخم التوقعات» إلى «مستوى خيبة الأمل» (Sicular 2013)، مع إعلان بعض الدعاة سابقاً عن موت «البيانات الكبيرة» بصفته مصطلحاً ذا معنى، حيث أصبح مصطلحاً متشعباً وواسع النطاق للغاية في التعريف (على سبيل المثال، de Goes 2013)، فبعض أوائل الشركات التي تبنت المفهوم تصارع من أجل تحويل الاستثمارات في البيانات الكبيرة إلى عوائد ربحية والبعض يشككون في فوائدها المحتملة. ومع ذلك، ظل قطاع الأعمال والقطاع الحكومي العام وممولو البحوث على قناعتهم بارتفاع منحني التنوير (Slope of Enlightenment) للبيانات الكبيرة مجدداً إلى مستوى الإنتاجية (Plateau of Productivity) عبر دورة الإعلان أو نموذج النضج التقني، بل وأكثر من ذلك، أنها وجدت لإحداث تغيير جذري في كيفية عمل العلوم وإجراء الأعمال (Sicular 2013). انظر أيضاً للفصلين السابع والثامن).

وكما نوقش في الفصل الثاني، فقد وُجدت ولفترة طويلة من الزمن مجموعات ضخمة من بيانات عبر الحكومات والصناعات والأوساط الأكاديمية التي تم استخراج المعلومات منها من أجل تقديم مزيد من الرؤى والمعرفة، فالإدارات والوكالات الحكومية أنتجت بشكل روتيني كميات ضخمة من البيانات. على سبيل المثال، قامت إدارة المحفوظات والسجلات الوطنية في الولايات المتحدة في عام ٢٠١٣ بتخزين ما يقارب ٤,٥ مليون قدم مكعب من الوثائق المادية من وكالات السلطة التنفيذية الأمريكية، والمحاكم، ومجلس النواب، ومقر الرئاسة (وهي خمسة في المائة من سجلات الحكومة الاتحادية) مما يضيف ثلاثين ألف قدم بصورة خطية سنوياً (Ellis 2013)، بالإضافة إلى خمسمائة تيرابايت من البيانات الرقمية. كما تمّ بالمثل جمع بيانات للشركات عن عملياتها، والأسواق، والعملاء وقواعد بيانات هائلة من البيانات العلمية التي تمّ جمعها ومشاركتها من بداية العلم الحديث. لذلك، ما المقصود من مصطلح «البيانات الكبيرة»، بالنظر إلى كميات البيانات الناتجة في المجالات السابقة؟

فكما هو شأن العديد من المصطلحات المستخدمة للإشارة إلى الاستخدام المتسارع في التقنيات والممارسات، لا يوجد تعريف أكاديمي محدد أو صناعي متفق عليه لمصطلح البيانات الكبيرة، إلا أن الكثير من التعاريف الشائعة تشير إلى الثلاثية المكررة للأحرف الابتدائية (3Vs) للخصائص المميزة للبيانات الكبيرة: الحجم (Volume)، والسرعة (Velocity)، والتنوع (Variety) (Laney 2001; Zikopoulos et al. 2012). فالبيانات الكبيرة تتميز بأنها:

- ضخمة في الحجم، أي تتكوّن من تيرابايت أو بيتابايت من البيانات.
- عالية السرعة، أي أنها تنشأ بشكل آني أو فوري، أو قريباً من الزمن الحقيقي.
- مختلفة من حيث تعدد تنوعها، حيث تتنوع البيانات بين المنظمة وغير المنظمة في طبيعتها، كما تتنوع في الغالب من حيث مرجعيتها الزمانية والمكانية.

فخلال زمن ما قبل البيانات الكبيرة، تقيّد قواعد البيانات بين هذه الخصائص الثلاث: حيث كان من الممكن لقاعدة البيانات الجمع بين خاصيتين اثنتين في آن واحد (ضخمة

وسريعة، متنوعة وسريعة، وكبيرة ومتنوعة (Croll 2012). إلا أنه مع تعزيز قدرة المعالجة الحاسوبية، والتصاميم الجديدة لقواعد البيانات، والتخزين الموزع (انظر الفصل الخامس)، أصبح تحقيق الثلاثة مجتمعة في وقت واحد ممكناً مما يتيح أشكالاً جديدة للتحليل وتقديم عروض مفصلة للغاية للنظم الضخمة والمتغيرة باستمرار. وعدا عن الخصائص الثلاث السابقة، أشارت المؤلفات الصادرة مؤخراً إلى عدد من الخصائص الرئيسية الأخرى، حيث تتميز البيانات الكبيرة أيضاً بالميزات التالية:

- شاملة في نطاقها، حيث تسعى لتغطية كامل مجتمع أو نظم الدراسة (ن = الجميع)، أو على الأقل توفر حجم عينات أكبر من تلك التي يجري استخدامها في دراسات البيانات التقليدية أو الصغيرة.
- متناهية في دقتها، حيث تهدف إلى أن تكون مفصلة قدر الإمكان، ومفهرسة بشكل فريد في تحديد الهوية.
- علائقية بطبيعتها، كونها تحتوي على حقول بيانات مشتركة تمكن من الربط بين مجموعات بيانات مختلفة.
- مرنة، حيث تحمل سمات التفرع (يمكن إضافة حقول بيانات جديدة بسهولة) مع قابلية التوسع (يمكن التوسع في حجمها بسرعة).

boyd and Crawford 2012; Dodge and Kitchin 2005; Marz and Warren)
(2012; Mayer- Schonberger and Cukier 2013).

وبالنظر إلى الحملة الدافعة إلى التحويل الرقمي (الرقمنة- Digitisation) للبيانات التقليدية الصغيرة والتوسع فيها ضمن البنى التحتية الرقمية التي تتميز بالضخامة والتنوع مثل المحفوظات الأرشيفية الوطنية، التعدادات السكانية، والمجموعات الثقافية والاجتماعية التراثية؛ انظر الفصل الثاني)، فقد كانت السرعة والخصائص الإضافية هي ما ميزت البيانات الكبيرة وجعلت منها ابتكاراً استثنائياً (Christensen's 1997) وغير بصورة جذرية طبيعة البيانات وما يمكن عمله من خلالها (انظر الجدول ١-٢). فعلى سبيل المثال، يتميز المسح الأسري الوطني بكونه أكبر الحجم، والدقة المتناهية، وقوة العلائقية، إلا أنه يفتقر إلى السرعة

(تتغير البيانات مرة واحدة في السنة)، والتنوع (يتكوّن من ثلاثين سؤالاً منظماً)، والشمول (حيث تقتصر العينة على فرد واحد من عشرين أسرة)، والمرونة (كون حقول البيانات ثابتة ولا تتغير خلال المسوحات المتعاقبة للتمكن من تحليل السلاسل الزمنية). ففي هذا الفصل تمّ إيراد الخصائص السبعة للبيانات الكبيرة وبناقش الفصل التالي العوامل المساعدة للبيانات الكبيرة ومصادرها.

الحجم (Volume):

شهد العقد الماضي انفجاراً في كمية البيانات التي تمّ إنتاجها ومعالجتها بشكل يومي. وكما عنونت مجلة وايرد (Wired) بشكل رئيس في عددها الخاص للعام 2008: نحن ندخل «عصر التيرابايت» (وفي الواقع، لقد ولجنا بالفعل في عصر الزيتابايت؛ 70^2 بايت). وقد سعت العديد من الدراسات لتقدير وتتبع الأحجام المعنية (على سبيل المثال، Hilbert and López 2009; Gantz and Reinsel 2011; Short et al. 2011). واستخدمت هذه الدراسات منهجيات وتعريف مختلفة، ولكنها أجمعت بالكلية على أن معدل نمو البيانات كان رهيباً في نطاقه، وعلاوة على ذلك، فهي قائمة على النمو باطراد أكبر في المستقبل المنظور. إن أبسط طريقة لتوضيح هذا النمو هو إعطاء بعض الأمثلة على تقديرات عملية لأحجام البيانات وبعض التقديرات المتعلقة بكيانات محددة، ولتوفير إطار مرجعي، يفصل الجدول (٤-١) بصورة مختصرة كيفية قياس حجم البيانات.

الجدول (٤-١)

مقاييس البيانات الرقمية

الوحدة	القياس	ما الذي تعنيه
بت (b)	صفر أو 1	هي اختصار لتعبير «الخانة الثنائية»، حيث تستخدمه الحاسبات لتخزين ومعالجة البيانات على صورة الرمز الثنائي (0 أو 1)
بايت (B)	ثمانية خانات ثنائية (8 بت)	مساحة تخزينية كافية لتمثيل رموز الأبجدية الإنجليزية والأرقام باستخدام الرموز الثنائية

من كلمة «الألف» (Thousand) باليونانية. فالصفحة الواحد من النص المطبوع تقدّر بحوالي 2 كيلوبايت	1000 أو 210 بايت	كيلوبايت (KB)
من كلمة «ضخم» (Large) باليونانية. فكامل أعمال شكسبير الأدبية يقدر حجمها ب 5 ميجابايتات في حين يقدر حجم الملف الصوتي الموسيقي بحوالي 4 ميجابايت	1000 كيلوبايت؛ 220 بايت	ميغا بايت (MB)
من كلمة «هائل» (Giant) باليونانية. ففيلم من ساعتين يمكن ضغطه في ملف حجمه 1-2 غيغابايت	1000 ميجابايت؛ 230 بايت	غيغابايت (GB)
من كلمة 'عملاق' (Monster) باليونانية. فالكتب المفهرسة في مكتبة مجلس النواب الأمريكي (الكونغرس) يقدر حجمها بحوالي 15 تيرابايت	1000 غيغابايت؛ 240 بايت	تيرابايت (TB)
تقدّر بيانات كل الرسائل التي تمّ توصيلها من قبل خدمة البريد الأمريكي في العام 2010 بنحو 5 بيتابايت.	1000 تيرابايت؛ 250 بايت	بيتابايت (PB)
مكافئة لبيانات عشرة بلايين نسخة من مجلة عالم الاقتصاد (The Economist).	1000 بيتابايت؛ 260 بايت	إكسابايت (EB)
يقدر المجموع الكلي للمعلومات في الوجود في عام 2010 بحوالي 1.2 زيتابايت	1000 إكسابايت؛ 270 بايت	زيتابايت (ZB)
حالياً هي قيمة أكبر من التخيل	1000 زيتابايت؛ 280 بايت	يوتابايت (YB)
تمّ وضع الاختصارات من قبل المنظمة الحكومية الدولية intergovernmental organization، والمكتب الدولي للأوزان والمقاييس International Bureau of Weights and Measures. تمّ إضافة وحدتي يوتابايت وزيتابايت في العام 1991؛ المصطلحات للكلمات الأكبر لم يتمّ وضعها بعد.		

المصدر: مجلة عالم الاقتصاد (2010) (The Economist)، تمّ إعادة إنتاجها بإذن عالم الاقتصاد. صحيفة عالم الاقتصاد المحدودة، لندن، الإصدار 11 مارس 2014.

وقد بينَ Zikopoulos et al. (2012) أنه في العام 2000م تمّ تخزين 800000 بيتابايت من البيانات في العالم. وبحسب التقرير السنوي ل Short et al. (2011: 7) - كمّ كمّ المعلومات - عالجت خدمات العالم للعام 2008م، 9.57 زيتابايت من المعلومات... وكان هذا بمعدل 12 غيغابايت من المعلومات يومياً للعامل العادي، أو نحو 3 تيرابايتات من

المعلومات لكل عامل سنوياً. وقامت الشركات في العالم بمعالجة ٦٣ تيرابايتاً من المعلومات سنوياً باستثناء المصادر غير المحوسبة، وبحلول عام ٢٠١٠ م، قدّرت مؤسسة إم جي آي (MGI) (بحسب ما ذكر 3: 2011 Manyika et al.) بأن الشركات عالمياً قد قامت بتخزين أكثر من ٧ إكسابايت من البيانات الجديدة على الأقراص الصلبة ... في حين قام المستهلكون بتخزين أكثر من ٦ إكسابايتات من البيانات الجديدة على أجهزة من قبيل أجهزة الحاسب الشخصية وأجهزة الحاسب المحمولة. كما قدّروا كذلك أنه في «عام ٢٠٠٩ م، كان ما يقرب من كل القطاعات في الاقتصاد الأمريكي لا يقلُّ بياناتها المخزّنة بالمتوسط عن ٢٠٠ تيرابايت من البيانات لكل شركة فيها أكثر من ألف موظف. كما امتلكت العديد من القطاعات أكثر من 1 بيتابايت من البيانات المخزّنة لكل شركة». وفي عام ٢٠١٣ م، أشار Neelie Kroes، مفوض الاتحاد الأوروبي لجدول الأعمال الرقمي، بأن ١,٧ مليون بليون بايت من البيانات يجري توليدها في الدقيقة الواحدة على مستوى العالم (Rial 2013).

وبناء على مراجعة (2011) Manyika et al. لنمو حجم البيانات، فقد توقعوا ارتفاعاً بنسبة ٤٠ في المائة في نمو البيانات التي يتمّ إنشاؤها على مستوى العالم سنوياً. وقدّر (2011) Gantz and Reinsel بأن «كمية المعلومات التي تمّ إنشاؤها وإعادة نسخها (تكرارها) على الإنترنت ستتجاوز ١,٨ زيتابايت (١,٨ تريليون غيغابايت) في العام ٢٠١١ مخزّنة في ٥٠٠ كوادريليون (مليون مليار) ملف بيانات». ويبيّن هذا الذي ذكره نمواً بمقدار تسعة أضعاف خلال خمس سنوات فقط، مع نمو متوقع في ذات الوقت إلى أكثر من الضعف كل سنتين. ونتيجة لذلك، فقد توقعوا في تقريرهم بأنه في العقد التالي:

سيزداد عدد الخوادم (الافتراضية والفيزيائية) في جميع أنحاء العالم بعشرة أضعاف، في حين ستزداد كمية المعلومات المُدارة في مراكز بيانات الشركات بمقدار خمسين ضعفاً، وسيزداد أعداد الملفات التي سوف تضطر مراكز البيانات إلى التعامل معها بمقدار 75 ضعفاً، على الأقل.

وبمثل هذا النمو الهائل في إنتاج البيانات، ادعت شركة آي بي إم (IBM) (2012) بأن «٩٠٪ من البيانات في العالم اليوم قد تمّ إنشاؤها في العامين الماضيين فقط»

كما توقع (Zikopoulos et al. 2012) أن تقارب أحجام البيانات حدَّ ٣٥ زيتابايت بحلول عام ٢٠٢٠م.

ويمكننا إدراك وقوع هذا الانفجار في إنتاج البيانات الرقمية من خلال النظر في بعض الأمثلة المحددة. فـ شركة تك أمريكا (TechAmerica) قدّرت بأنه يجري يومياً إرسال ما يقرب من 114 مليار رسالة بريد إلكترونية و٢٤ مليار رسالة نصية، كما يتم إجراء ١٢ مليار مكالمات هاتفية على مستوى العالم (Strohm and Homan 2013). ووفقاً لشركة سيسكو (CISCO)، ففي العام ٢٠١٣ م كان هنالك تقريباً ما يناهز ١٠ مليارات عنصر (أجهزة اتصال ومستشعرات) مشكّلة شبكة إنترنت الأشياء (Internet of Things -IoT)، والتي يقوم كلُّ منها بإنتاج كميات متغيرة من البيانات، مع الزيادة المتوقعة لهذا الرقم ليقف الخمسين ملياراً بحلول عام ٢٠٢٠ م (Farber 2013). أما ما يتعلّق بحركة الاتصال على الإنترنت، فقد كانت شركة جوجل تعالج ٣ مليارات طلب استعمال للبحث يومياً، وتقوم بتخزينها جميعاً (Mayer-Schonberger and Cukier 2013)، وقرابة ٢٤ بيتابايت من البيانات كل يوم (Davenport et al. 2012). وفي العام ٢٠١١ م، قضى المستخدمون النشطون لموقع التواصل الاجتماعي فيسبوك أكثر من ٩,٣ مليارات ساعة خلال شهر (Manyika et al. 2011)، وبحلول عام ٢٠١٢ م، ذكرت الشركة بأنها كانت تعالج ٢,٥ مليار من مختلف أنواع بيانات المحتوى (وصلات، مخازن، صور، أخبار، وما إلى ذلك)، و٢,٧ مليار من ردود الإعجاب وثلاثمائة مليون عملية تحميل للصور يومياً (Constine 2012). كما تمّ إنتاج أكثر من أربعمائة مليون تغريدة يومياً في العام ٢٠١٢م، بمعدل زيادة مقداره مئتان في المائة سنوياً، مع احتواء كل تغريدة على ثلاثة وثلاثين بنداً من البيانات الوصفية لها (Mayer-Schonberger and Cukier 2013). وتُعد الكثير من هذه البيانات غير منظمة بطبيعتها. وبالمثل هنالك انفجار في حجم إنتاج البيانات المنظمة قد وقع أيضاً، على سبيل المثال، وفيما يتعلق ببيانات تجارة التجزئة الخاصة بالأوراق المالية والمبيعات، والتي تمّ جمعها من خلال السلاسل السوقية (اللوجستية) ومنافذ البيع، أنتجت سلسلة متاجر وول مارت (Walmart) أكثر من ٢,٥ بيتابايت من البيانات المتعلقة بأكثر من مليون عملية خاصة بالعملاء في كل ساعة في العام ٢٠١٢م، «أي يفوق ما يعادل ١٦٧ مرة من المعلومات الواردة في جميع الكتب المصنفة في مكتبة الكونغرس»؛ (Open Data Center Alliance 2012: 6).

كما أنتجت متاجر تسكو (Tesco) البريطانية أكثر من ١,٥ مليار بنداً جديداً من بنود البيانات كل شهر في العام ٢٠١١ م (Manyika et al. 2011).

وبالمثل، تنتج الحكومات والهيئات العامة كميات هائلة من البيانات حول مواطنيها والدول الأخرى. على سبيل المثال، بدأت هيئات التنقل والعبور في مراقبة التدفق المستمر للناس من خلال أنظمة النقل، على سبيل المثال، جمع زمن ومكان استخدام بطاقات السفر مسبقاً الدفع مثل بطاقة أويستر في لندن. كما يجري الآن تقديم العديد من نماذج وطرق دفع الضرائب أو طلبات الحصول على الخدمات الحكومية من خلال الانترنت، ففي العام 2009م، أنتجت حكومة الولايات المتحدة ٨٤٨ بيتابايت من البيانات (مؤسسة التقنية الأمريكية TechAmerica Foundation 2012). كما تعمل وكالات الاستخبارات الست عشرة التي تشكل أمن الولايات المتحدة، جنباً إلى جنب مع فروع الجيش الأمريكي، على مراقبة وتخزين وتحليل كميات هائلة من البيانات كل ساعة، من خلال آلاف المحللين العاملين في تدقيق وتفسير النتائج. وللحصول على فكرة عن حجم بعض مشاريع الاستخبارات العسكرية، فإن مشروع النظام المستقل للمراقبة الأرضية الصورية الحقيقية (ARGUS-IS)، والذي كشفت عنه وكالة مشاريع البحوث المتطورة الدفاعية (داربا DARPA) وجيش الولايات المتحدة في العام ٢٠١٣ م، هو منصة للمراقبة بالفيديو بدقة ١,٨ غيغابايت والتي يمكنها بيان التفاصيل الصغيرة حتى ست بوصات من ارتفاع ٢٠٠٠٠ قدم (ستة كيلومترات) (Anthony 2013). وتجمع المنصة ١,٨ مليار بكسل، في ١٢ إطاراً في الثانية { لقطة في الثانية }، مما ينتج ما يقارب ٦٠٠ غيغابايت في الثانية. وهذا يعادل نحو ٦ بيتابايتات من البيانات يومياً. وباستخدام الحاسوب العملاق، يمكن للنظام إجراء التحليل آنياً من خلال تتبع ما يصل إلى ٦٥ جسماً متحركاً ضمن مجال الرؤية. وهذا هو مشروع واحد فقط في ترسانة المشاريع الاستخباراتية المماثلة وذات الصلة.

وبالمثل، فيما يتعلق بالمشاريع العلمية، يتكوّن تسلسل الجينوم البشري من نحو ١٠٠ غيغابايت من البيانات (Vanacek 2012): وهذا يتضاعف عبر الآلاف من الأفراد مما يوسع قاعدة البيانات قريباً إلى تيرابايت وبيتابايت من البيانات. وعندما بدأ التشغيل في مسح سلوان الجوي الرقمي (Sloan Digital Sky Survey) في العام ٢٠٠٠، أنتج

المراقب (التلسكوب) في نيومكسيكو من بيانات الرصد والمراقبة في أول شهرين أكثر مما سبق جمعه في تاريخ علم الفلك حتى ذلك الوقت (Cukier 2010). وفي عام ٢٠١٠ م، بلغ حجم أرشيفها ما يقارب ١٤٠ تيرابايت من البيانات، وهي كمية سيتم قريباً جمعها كل خمسة أيام من قبل المراقب الكبير للمسح الشامل (Large Synoptic Survey Telescope) والذي سيبدأ العمل به في شيلي في العام 2016 م (Cukier 2010). وأكثر ضخامة من هذا، مصادم هادرون الكبير (Large Hadron Collider) والذي يقوم بإنتاج ٤٠ تيرابايت من البيانات في الثانية الواحدة (The Economist, 2010). وفي هذا، وغيره من الحالات، يتم إنشاء بيانات هائلة للغاية لا يمكن تحليلها أو تخزينها، بحيث تعد بيانات عابرة. وفي الواقع، لا وجود لهذه القدرة التخزينية لكل هذه البيانات، فعلى الرغم من التوسع السريع في القدرات التخزينية، إلا أن هذا التوسع لا يمكنه مواكبة إنتاج البيانات (Gantz et al. 2007; Manyika et al. 2011).

الشمول (Exhaustivity):

تستخدم عملية انتقاء العينات في دراسات البيانات الصغيرة من أجل إنتاج مجموعة بيانات تمثيلية لمجتمع الدراسة من جميع البيانات المحتملة في وقت معين ومكان معين. وتستخدم عملية انتقاء العينات من مجتمع الدراسة بسبب كبر هذا المجتمع وتعذر جمع كافة البيانات المتعلقة به لمحدودية الوقت والموارد المخصصة. في المقابل، تسعى مشاريع البيانات الكبيرة إلى التقاط بيانات كامل مجتمع الدراسة (ن = الكل)، أو على الأقل، حجم عينة أكبر من المعتادة في دراسات البيانات الصغيرة (Mayer-Schonberger and Cukier 2013). فمن ناحية، تُعد هذه البيانات نتاجاً ثانوياً للتقنيات التي استخدمت في إنتاج البيانات، جنباً إلى جنب مع النمو الهائل في القدرة على تخزين البيانات (انظر الفصل الخامس)، ومن ناحية أخرى، هناك قناعة بأن "الأكثر هو أفضل" وسيقدم تمثيلاً أكبر ومزيداً من الدقة والصحة في التحليل.

وفي النظم المفتوحة مثل المشاريع العلمية الكبيرة، كتلك التي تقيس البيانات المناخية للإبلاغ عن الطقس ونمذجة الأرصاد الجوية، أو جمع البيانات الفلكية باستخدام مراقب قوي، يكون التوجه نحو مجموعات بيانات أكبر بكثير، مع زيادة أحجام العينات عبر العديد

من المتغيرات ما أمكن ذلك. ولا يعني ذلك في علم الفلك، على سبيل المثال، مجرد جمع بيانات الضوء، ولكن أيضاً البيانات من مختلف ألوان الطيف الكهرومغناطيسي، وبدقة عالية لأكثر قدر ممكن. أما في حالة النظم المغلقة، مثل الفيسبوك أو شراء السلع من متجر على شبكة الإنترنت مثل أمازون أو إرسال رسائل البريد الإلكتروني، فمن الممكن تسجيل كافة التفاعلات والعمليات التي تحدث، بالإضافة إلى مستوى عدم التفاعل. وفي هذه الحالات، هذا هو الواقع بطبيعة الحال، فموقع التواصل الاجتماعي الفيسبوك يقوم بتسجيل كل نشر، أو تسجيل لإعجاب، أو رفع لصورة يتم تحميلها، أو وصلة إلى موقع آخر، أو رسالة مباشرة، أو لعبة لعبت، أو فترات الغياب، وما إلى ذلك لكل مستخدميه الذين ناهزوا المليار أو نحو ذلك. وبالمثل، لا تسجل شركة أمازون فقط تفاصيل كل عملية شراء والمشتريين، ولكن أيضاً جميع الروابط التي تم زيارتها وكل السلع التي تم استعراضها من خلال موقعها، بالإضافة إلى كافة البنود التي تم وضعها في سلة التسوق ولم يتم شراؤها. كما يتم تسجيل كافة رسائل البريد الإلكتروني من قبل الخادمت التي تستضيف صندوق البريد الإلكتروني للعميل، وتخزين كامل البريد الإلكتروني وجميع البيانات الوصفية المرتبطة به على سبيل المثال، الذين تم إرسال البريد الإلكتروني لهم أو استلامه منهم، والوقت / التاريخ، والموضوع، والملفات المرفقة. وحتى لو تم تحميل البريد الإلكتروني محلياً على جهاز المستخدم وتم حذفه فما زالت نسخة محفوظة منه على الخادم، هذا مع احتفاظ معظم المؤسسات والشركات بهذه البيانات لعدد من السنوات.

وبالمثل بالنسبة للأشكال الأخرى من البيانات، فقد نمت البيانات المكانية بشكل كبير في السنوات الأخيرة، وذلك من خلال الاستشعار الآتي عن بُعد والصور الرادارية، إلى المشاريع الكبيرة مكتظة المصادر مثل مشروع خرائط الشوارع المفتوحة (OpenStreetMap)، والمسارات المكانية الرقمية التي تم إنشاؤها بواسطة مستقبلات أنظمة تحديد المواقع الجغرافية المضمنة داخل الأجهزة. ويهدف الصنفان الأولان إلى أن يكونا شاملين مكانياً، بحيث تجمع التضاريس للكوكب الأرضي بأسره، ورسم خرائط البنية التحتية لدول بأكملها وتوفير مجموعات بيانات مشتركة ومرخصة للخرائط. في حين يوفر المشروع الثالث القدرة على تتبع وتعقب الحركة عبر الفضاء بمرور الوقت؛ وبناء مسارات الزمن - الفضاء التي يمكن تجميعها لتوفير نماذج الزمن - الفضاء للسلوك عبر مدن ومناطق بأكملها. وتمكن مجتمعة

من النمذجة المفصلة للأماكن وطرق التنقل، والمقارنة المسحية عبر الفضاء، واستهداف التسويق لمجتمعات معينة، وتقديم خدمات جديدة معتمدة على الموقع، والبيانات التي يمكن الجمع بينها لاشتراكها في المرجعيات المكانية لإنشاء قواعد بيانات وتطبيقات جديدة يمكن البحث مكانياً فيها (على سبيل المثال، جمع البيانات حول منطقة لإنشاء ملفات تعريف للحي).

وبالنظر إلى التقدم الحاصل في ساعات التخزين (انظر الفصل الخامس)، يبدو أننا قد وصلنا إلى مرحلة بحيث يكون من الأسهل في كثير من الحالات تسجيل كل شيء، دون الحاجة إلى فرز البيانات، وتدقيقها، وأخذ عينات منها، وتسجيل ما يحتمل أنه مفيد منها فقط (ومن الذي يعرف ما يثبت احتمال أنه مفيد في المستقبل؟). وكما أشار Zikopoulos et al. (2012): «لا عجب أننا نغرق في البيانات، فإذا كنا نستطيع تتبع وتسجيل شيء ما، فإننا بالعادة نفعل». وفي الواقع، أشار Dumbill (2012: 7) إلى أن المبدأ الأساسي خلف البيانات الكبيرة هو «عندما تستطيع، احتفظ بكل شيء». ويتجلى هذا بالاعتقاد بأنه كلما كان المزيد من البيانات متاحاً، سنحت فرص أفضل لصنع رؤية أكثر صحة ونفاذاً... وفرص العثور على «المنتجين» لنظرية جديدة (John Seely Brown مقتبس من 8: 2010 Bollier). وعلى الرغم من ذلك، تسهم إستراتيجية السعي نحو الشمولية بشكل كبير في طوفان البيانات، ويتمثل التحدي الأكبر في رؤية النمط الشامل والصورة الكبيرة من جملة التفاصيل وتمييز الغث من السمين، وإثارة مجموعة من الأسئلة الأخلاقية فيما يتعلق بنطاق البيانات التي يتم إنتاجها والاحتفاظ بها، والاستخدامات التي يجري توظيفها فيها أو من الممكن توظيفها فيها (انظر الفصل العاشر). كما أنها تثير تساؤلات معرفية أساسية (Floridi 2012). على سبيل المثال، وبالنظر إلى شموليتها، تساءل Callebaut (2012) عما إذا كانت تحليلات البيانات الكبيرة هي علم ما بعد الاختزال (Post-reductivist)، وسيتم مناقشة هذه الأسئلة بمزيد من التفصيل في الفصل الثامن.

الدقة والتبويب/ الفهرسة (Resolution and Indexicality):

بالإضافة إلى شمولية البيانات، أصبحت البيانات الكبيرة أكثر تناهياً في الدقة بكثير، وهاتان الخاصيتان يدفعان جميعاً في الاتجاه نحو التبويب/ الفهرسة القوية (عنونة بوسم مميز وتحديد للهوية) (Dodge and Kitchin 2005)، وتعد صور الاستشعار

عن بُعد مثلاً على تعزيز الدقة، ففي أواخر الثمانينيات، كانت أعلى دقة للصور المتاحة لمعظم الباحثين غير الحكوميين عن سطح الأرض هي لتلك الصور التي التقطتها أقمار لاندسات (Landsat Satellites) الصناعية، حيث تمثل كل نقطة ضوئية (بكسل) قطعة من الأرض تقدر 30 x 30 متراً، في حين أن الكثير من الصور المتاحة على جوجل إيرث ذات دقة تصل إلى 2,5 x 2,5 متر، مما يتيح عرض المزيد من التفاصيل وتحليلها. وبالمثل، فيما يتعلق بمخرجات بيانات التعداد السكاني، فقد زاد علو درجة البيانات في العديد من الدول. ففي الحالة الأيرلندية، كان نشر بيانات التعداد حتى وقت قريب للدوائر الانتخابية (3,409 منطقة بمتوسط 1,350 نسمة في كل منها، حيث كانت الكثافة السكانية لكل دائرة انتخابية أعلى بكثير في المدن والبلدات وأقل في المناطق الريفية). وفي العام 2011، تم وضع بيانات التعداد بتوزيع جغرافي جديد سُمي بالتجمعات الصغيرة، التي بلغ عددها 18,488. تمثل الوحدات الجديدة بالعادة بيانات من 80 - 150 من الأسر (Gleeson et al. 2009). وقد أتاحت التجمعات الصغيرة إجراء تحليل التعداد على مستوى الحي أو الشارع، بدلاً من مساحات شاسعة جداً، ومناطق ذات أعداد متساوية تقريباً من السكان التي يمكن مقارنتها، مما يوفر فهماً أكبر للتجمعات السكانية والاقتصاد الأيرلندي. وإمعاناً في تحري المزيد من الدقة، يجمع العديد من وكلاء / سمسرة البيانات كميات كبيرة من البيانات المتعلقة بالأفراد والأسر التي تمكن الشركات من استهدافهم بشكل فردي بالسلع والخدمات (انظر الفصل الثاني).

وقد صاحبت الزيادة في دقة البيانات، الزيادة في تحديد الهويات للأشخاص، والمنتجات، والمعاملات، والأقاليم بحيث أصبحت بياناتها أكثر تبويماً وفهرسة بطبيعتها (انظر الفصل الخامس). فعلى سبيل المثال، تحمل معظم السلع المعروضة للبيع في متاجر التجزئة الكبرى في الوقت الحاضر الرمز الشريطي (الباركود)، ويميّز هذا الرمز المنتج، ولكن ليس كل وحدة منه على حدة - حيث تشترك كل زجاجات صابون شعر الرأس (الشامبو) من نفس العلامة التجارية (الماركة) في نفس الرمز - مما يعني أنه لا يمكن التمييز بينها بشكل فردي. وفي المقابل، تتميز زجاجة صابون شعر الرأس الموسومة بشريحة تحديد الهوية باستخدام موجات الراديو (Radio Frequency Identification Device- RFID) بشكل فريد، لأن لكل شريحة رمز فريد من الممكن قراءته عن بُعد بواسطة جهاز استشعار الذبذبات

الراديوي. ونتيجة لذلك، يمكن تتبع كل زجاجة من مكان صنعها من خلال سلسلة التوريد وحتى المتجر وسلة العميل، مما يوجد مسار للتدقيق المفصل. وبعبارة أخرى، فقد أصبح من الممكن تتبع الأمور الفردية بدقة عبر الزمان والمكان، بما في ذلك أولئك الذين يعملون على نقل أشياء من مكان لآخر. وبالمثل، بالنسبة للمعلومات، وخاصة تلك التي في شكل رقمي، حيث يتم تعريفها بشكل فريد من خلال رموز إدارة الحقوق الرقمية (Digital Rights Management). فعلى سبيل المثال، يمكن تعيين معرفات الكائن الرقمي (Digital Object Identifiers- DOIs) إلى الأعمال الإبداعية المتوفرة على الإنترنت (كالتقارير، المجلات، الصور، الصوت، وملفات الفيديو)، فمعرفات الكائن الرقمي هي محدد هوية دائم مرتبط بالبيانات الوصفية، مثل عنوان الإنترنت (Universal Resource Locator- URL) الذي يشير إلى موقع الملف. ويعزز استخدام المعرفات الفريدة العلائقية والقدرة على ربط وضم البيانات معاً، كما يوفر وسائل عملية لفرز ورصد ومطابقة وتمثيل الكائنات (Lyon 2005; Dodge and Kitchin 2005; Graham 2003a؛ انظر أيضاً الفصل العاشر).

العلائقية (Relationality):

تتمثل العلائقية بمدى القدرة على ضمّ مجموعات بيانات مختلفة بروابط بعضها مع بعض وكيف يمكن استخدام هذه الروابط للإجابة عن أسئلة جديدة. فالعلائقية هي في صميم قواعد البيانات العلائقية (انظر الفصلين الثاني والخامس)، وهي القدرة على إنشاء بيانات غاية في العلائقية التي تدفع بسوق البيانات الواسع وأرباح وكلاء البيانات وشركات الملفات التعريفية (انظر الفصل الثاني). إن الدرجة العالية من العلائقية هي ما يجعل التعداد السكاني مفيداً جداً لفهم التركيبة السكانية لأي أمة، وكيف تتغير عبر الزمان والمكان. وتختلف دراسات البيانات الصغيرة في مدى علائقيتها، أي ميل تلك التي تنطوي على البيانات المنظمة لدرجات أعلى من الترابط من تلك البيانات غير المنظمة، وهذا يستدعي القول إنه ينبغي وجود شكل من أشكال العلائقية بين البيانات لاستخلاص التفسيرات والاستنتاجات الشاملة منها.

وعلى الرغم من أن البيانات الكبيرة لا تستخدم غالباً بنية قاعدة بيانات علائقية (انظر الفصل الخامس)، فإن ميزة أساسية في طبيعتها هي العلائقية القوية. وكما فصل

(2: 2011) Boyd and Crawford فإن «البيانات الكبيرة مترابطة بشكل أساسي، وتتبع قيمتها من الأماط التي يمكن استخلاصها من خلال صنع روابط بين أجزاء من البيانات، عن فرد، أو أفراد على علاقة بآخرين، أو مجموعات من الناس، أو ببساطة عن هيكل تنظيم المعلومات نفسها». إلا أنها وخلافاً لقواعد البيانات العلائقية، فهي تتقن بالتساوي التعامل مع البيانات غير الرقمية (الرمزية)، والبيانات غير المنظمة كما البيانات المنظمة، والربط بين النوعين معاً وتعظيم القيمة المستفادة من التقاطعات بينها. ومن ثم، أصبح من الممكن الربط بين مجموعات متنوعة من البيانات - الشخصية، والعملية، والتفاعلية، والاجتماعية، والمالية، والمكانية، والزمنية،... إلخ - وتحليلها على أساس فردي وجماعي للعلاقات والأماط.

وقد ثبت جلياً، كمثال على قوة هذه العلائقية، الاستخدام المكثف للبيانات الكبيرة في الحملات الانتخابية للرئيس الأمريكي أوباما في العام ٢٠٠٨م و٢٠١٢م. فعلى النحو الذي فضّله (Issenberg 2012)، سعى فريق أوباما لتحديد وتتبع جميع جوانب حملاتهم الانتخابية في ٢٠٠٨م و٢٠١٢م، ووضع سلسلة كاملة من المقاييس التي تم تسجيلها والتنقيب فيها بشكل مستمر للحصول على معلومات، وأماط واتجاهات مفيدة. واشتمل ذلك على مراقبة صارمة للإجراءات الخاصة بهم، مثل وضع الإعلانات عبر وسائل الإعلام المختلفة، وإجراء حملات بريد الإلكترونيّة، والاتصال الهاتفي بالناخبين المحتملين، وطرق الأبواب وفرز الاصوات الانتخابية للمناطق، وتنظيم الاجتماعات والمسيرات، وتتبع الذين كانوا قد تحدثوا إليهم وماذا قالوا أو ما التزموا به. وقد عزّزوا هذه المعلومات بمئات التجارب العشوائية واسعة النطاق المصمّمة لاختبار فعالية مختلف طرق إقناع الناس على التصويت لصالح أوباما أو التبرع بالأموال لحملته. وجمع فريق أوباما بين جميع المعلومات التي نتجت فيما تعلق بالناخبين وبيانات التسجيل، والتعداد السكاني، والبيانات الحكومية الأخرى، ومسوحات الاقتراع، وبيانات جرى شراؤها من طيف واسع من الموردين، بما في ذلك وسطاء البيانات ووكالات التصنيف الائتماني، وشركات التلفاز الخطي (Cable TV). وكانت النتيجة مجموعة من قواعد البيانات الضخمة عن كل الناخبين في البلاد تتكون مما لا يقل عن ثمانين من المتغيرات (Crovitz 2012)، وأكثر من ذلك، أنها ترتبط بالخصائص الديموغرافية للناخبين المحتملين، وتاريخ انتخابهم، وكل حالة من حالات تواصلهم مع

حملة أوباما وردود أفعالهم، والتاريخ الاجتماعي والاقتصادي، وأنماط سلوكهم واستهلاكهم، ووجهات النظر والآراء التي تمّ الإعراب عنها، مع تحديث قواعد البيانات يومياً خلال الحملة بكل بيانات جديدة يجري إنتاجها أو شراؤها. إما في الحالات التي لا يعرف فيها محللو أوباما الانتماء السياسي للناخب، ولا يمكنهم فيها الوصول إلى ذلك من خلال الاتصال المباشر معه، فإنهم يوظفون خوارزمية متطورة تستخدم المتغيرات التي بحوزتهم عن شخص ما للتنبؤ عن التصويت المحتمل المفضل له (Issenberg 2012). وكانت النتيجة مليارات من أجزاء البيانات المترابطة والتي استخدمت للتعريف بالناخبين بشكل فردي، وتقييم مدى احتمالية قيامهم بالتصويت والكيفية التي يصوتون بها، والكيفية التي قد يستجيبون بها لمختلف السياسات والقصص. لقد أنشأ الترابط بين البيانات في حملة أوباما ما اصطلح عليه (Crampton et al. 2012) بمصطلح «تأثير تضخيم المعلومات» (Information Amplifier Effect)، حيث مجموع البيانات أكثر من الأجزاء.

السرعة (Velocity):

هناك فرق جوهري بين البيانات الصغيرة والبيانات الكبيرة يتمثل في الطبيعة الديناميكية لإنتاج البيانات، فالبيانات الصغيرة تتكوّن عادة من الدراسات التي تكون مؤطرة بقالب جامد من حيث الزمان والمكان. وحتى في الدراسات المطوّلة، يتم التقاط البيانات في أوقات متفرقة ذات فاصل زمني (على سبيل المثال، كل بضعة أشهر أو سنوات). فعلى سبيل المثال، يتم إجراء التعدادات السكانية بشكل عام كل خمس أو عشر سنوات، وفي المقابل يتم إنشاء البيانات الكبيرة على أسس أكثر استمرارية بكثير، وفي كثير من الحالات بشكل آني أو لحظي وبشكل مباشر بالتقارب من الوقت الحقيقي للوقوع (Real-Time). وعوضاً عن قدر ضئيل ومتفرق من البيانات، التي تُجمع أو تعالج بمشقة، فإن البيانات الكبيرة تتدفق بسرعة، ولذلك كان هناك انتقال في التعامل من المعالجة بالدفعات (Batch processing) إلى البيانات المتدفقة (Streaming Data) (Zikopoulos et al. 2012). فمن ناحية، يسهم هذا في مسألة كبر حجم البيانات من خلال إنتاج بيانات بسرعة أكبر، ومن ناحية أخرى، فهو يجعل كامل دورة البيانات أكثر ديناميكية بكثير، ويبرز قضايا كيفية إدارة نظام البيانات المتغير باستمرار.

وتتبع السرعة من التكرار المستمر لتسجيل المشاهدات مع مرور الوقت و/أو تغيير المكان (Jacobs 2009) مع كون العديد من الأنظمة في وضع التشغيل الدائم باستمرار (Dodge and Kitchin 2005). فعلى سبيل المثال، تسجل المواقع الإلكترونية بيانات كافة الزيارات التي تمت للموقع والأنشطة التي قام بها الزائر خلال تصفحه على الموقع؛ كما تقوم المعدات الطبية بمراقبة العلامات الحيوية باستمرار، وتسجل الكيفية التي يستجيب بها الجسم للعلاج وتُطلق إنذاراً في حال تجاوزت هذه العلامات حدّ معين؛ وتتعبق شركات الهاتف المحمول موقع، وهوية، وطبيعة استخدام الأجهزة الموصولة على شبكاتها كل بضع ثوان؛ وبالمثل ترصد شبكات استشعار الطقس مؤشرات الغلاف الجوي كل بضع دقائق وترسل نتائجها إلى قاعدة بيانات مركزية لإدراجها في الاستطلاعات الجوية؛ ويسجل جهاز استشعار الذبذبات الراديوي (Transponder) على طول طرق المدينة البرية والسكك الحديدية هوية الحافلات والقطارات التي تمرّ، مما يتيح لسلطة النقل العام معرفة أماكن مركباتها في أي وقت، واحتساب الوقت المقدّر للوصول في محطات التوقف المختلفة؛ كما تراقب متاجر التجزئة مبيعات آلاف المنتجات المختلفة للآلاف من العملاء، حيث تستخدم البيانات لمعرفة أوقات إعادة تعبئة الرفوف وطلب دفعات جديدة من الموردين؛ ويتواصل الناس بعضهم مع بعض من خلال مواقع وسائل التواصل الاجتماعي في سبيل لا ينتهي من التبادلات والعلاقات البينية المتشابكة؛ ويرصد المرقاب (Telescope) باستمرار السماء لقياس التقلبات في موجات الراديو من أجل فهم أفضل لطبيعة الكون، ففي كل هذه الحالات، هناك تيار مستمر من البيانات التي تتطلب إدارة وتحليلاً مستمراً.

ويشكّل نقل وإدارة كميات كبيرة من البيانات المنتجة باستمرار تحدياً تقنياً بسبب التحديات الناشئة عن محدودية قدرة سعة التخزين مما يوجد اختناقات بشكل سريع في حركة مرور وتبادل البيانات. على سبيل المثال، وكما أن فيديوهات موقع اليوتيوب (YouTube) قد تتوقف أو يتأخر عرضها بسبب كون سعة قناة الاتصال بالإنترنت ليست كافية لتواكب سرعة تدفق البيانات المطلوب، فإن الأثر نفسه يمكن أن يقع في حالة جمع ومعالجة البيانات، حيث تكون الأنظمة عاجزة عن مواكبة تدفق البيانات. وتشمل حلول هذه المشكلة زيادة سعة قناة الاتصال، وتوظيف تقنيات فرز وضغط البيانات مما يقلل من حجمها ومن ثم وقت معالجتها، وتحسين كفاءة خوارزميات المعالجة وتقنيات إدارة

البيانات. كما يمثل تحليل سيل البيانات المتدفقة تحدياً أيضاً، لأن النظام لا يسكن في أي لحظة وعند أي نقطة، كما في حالات الأسواق المالية، حيث يمكن أن يكون تحليل الصفقات لمايكرو ثانية من الزمن قيماً للغاية. وهنا، تستخدم خوارزميات معقدة، إلى جانب العرض المرئي للبيانات الديناميكية وهي في حالة تغير مستمر، لتتبع وتقييم النظام.

التنوع (Variety):

قد تختلف البيانات الصغيرة والبيانات الكبيرة في طبيعتها، كونها منظمة، أو غير منظمة، أو شبه منظمة، وتتألف من الأرقام، النصوص، الصور، الفيديو، وغيرها من أنواع البيانات الأخرى. إلا أن هذه الأنواع المختلفة من البيانات يرجح أن تكون في البيانات الكبيرة أكثر قابلية للجمع والربط بينها، وإمكانية وصل البيانات المنظمة بالبيانات غير المنظمة. فعلى سبيل المثال، تتألف نشرات الفيسبوك من النصوص التي ترتبط في الغالب بصور أو ملفات فيديو أو غيرها من المواقع، وهي تجتذب تعليقات مستخدمي الفيسبوك الآخرين؛ كما قد تجمع شركة ما بين البيانات المالية المتعلقة بالمبيعات واستطلاعات العملاء التي تعبر عن مدى ثقتهم بالمنتج، وفي المقابل تكون البيانات الصغيرة أكثر انفصلاً وترتبط، على كل حال، من خلال مفاتيح تعريفية وحقول بيانات مشتركة. ويتمثل السبق الأساسي بالنسبة للبيانات الكبيرة في كيفية اختلافها عن الأشكال التي سبقتها لإدارة البيانات الرقمية، والتي كانت احترازية للغاية في معالجة وتخزين البيانات الرقمية باستخدام قواعد البيانات العلائقية، مما مكن من إجراء مختلف أنواع التحليل الإحصائي. إلا أنها كانت أضعف بكثير في التعامل مع نماذج البيانات الغير رقمية، عدا عن مجرد تخزينها في ملفات مجردة أو مضغوطة. وكما لاحظ اتحاد مراكز البيانات المفتوحة (7: 2012) Data Center Alliance بأنه في السابق كان يجري تجاهل البيانات غير المنظمة، أو في أحسن الأحوال، استخدامها بشكل يفتقر إلى الكفاءة. ومع ذلك، فإن التقدم الحاصل في الحوسبة الموزعة، وتصميم قواعد البيانات باستخدام هياكل عوضاً عن لغات الاستعلام الهيكلية (NoSQL) (انظر الفصل الخامس)، والتنقيب في البيانات وتقنيات اكتشاف المعرفة (انظر الفصل السادس)، قد زادت بشكل كبير القدرة على إدارة ومعالجة واستخراج المعلومات من البيانات غير المنظمة. وفي الواقع، يجري على نطاق واسع تداول أن ما يقارب ثمانين في المائة من جميع

البيانات الكبيرة هي غير منظمة بطبيعتها، وكما أبان (Grimes 2011)، فقد أصبح هذا الرقم مسلماً به مع قليل من الإثبات.

المرونة (Flexibility):

يكون تصميم البحث وإدارة البيانات، مع مشاريع البيانات الصغيرة، غير مرن نسبياً بمجرد بدء العمل الميداني والتحليل، وذلك بالنظر إلى الخدمات اللوجستية، والتكاليف، والحاجة إلى التمثيل في عينة صغيرة الحجم. فعلى سبيل المثال، فإنه من الضروري على كل شخص تمّ جمع بيانات التعداد له تعبئة النموذج ذاته للتأكد من إمكانية قابلية مقارنة البيانات عبر كافة السكان. وبمجرد طباعة النماذج، لا يمكن إضافة حقول بيانات جديدة، مما يعني ثبات البيانات الممكن استخراجها من هذه النماذج، وبالمثل تميل قواعد البيانات العلائقية التي تُحفظ بها البيانات إلى أن تكون قالباً ثابتاً ومحدودة النطاق. وعلى نحو مماثل، يكون تصميم البحث عادة، في التجارب العلمية والدراسات البيئية، غير مرن بمجرد بدئها، وذلك بهدف تمكين إجراء عمليات المقارنة والتكرار. أما في الدراسات التي تعتمد المقابلات أو الأنشطة الاجتماعية، فمن الممكن للباحث، على الرغم من ذلك، أن يكون أكثر مرونة في النهج المتبع، وأن يكون لديه نموذج حرّ من الأسئلة وإمكانية التكيف مع الأوضاع المستجدة. ويمكن أن يكون ترميز وإدارة وتحليل هذه البيانات مرناً نسبياً، ولكن هذا يعود جزئياً إلى محدودية حجم ونطاق مجموعة البيانات.

في المقابل، تمّ تصميم نظم البيانات الكبيرة لتكون مرنة بطبيعتها، بحيث تحمل سمات قابلية الاستطالة والتفرع (يمكن إضافة حقول بيانات جديدة بسهولة) وقابلية التوسع (يمكن التوسع بسرعة) بغض النظر عن الحجم (Marz and Warren 2012). ويعني استخدام قواعد البيانات غير الهيكلية للاستعلام (NoSQL) إمكانية إدارة البيانات المتغيرة بسرعة عالية، والتكيف مع حقول البيانات الجديدة (انظر الفصل الخامس)، وهذا يعني أنه من الممكن التكيف مع إنتاج البيانات على أسس تدويرية، وأداء الاختبار على التكيف. فعلى سبيل المثال، تقوم شركات جوجل، والفيسبوك، ومنصات أخرى على الإنترنت بتعديل تصاميمها، وجمع بيانات حول كيفية استجابة المستخدمين لهذه التغييرات على سبيل المثال، مراقبة الزيارات العابرة للموقع، وتحليل النتائج واستخدامها

لتقديم مزيد من التعديلات المحسنة للتصميم بغية تشجيع إجراءات وردود معينة بشكل مستمر. ولأن أعداد الأشخاص الذين يستخدمون هذه المواقع هائلة، وأحجام العينات الخاصة بهم كبيرة جداً، فهذا يعني أن بإمكانهم إجراء التغييرات دون الخوف من فقدان التمثيل. على سبيل المثال، وبالعودة إلى حملة باراك أوباما الانتخابية، فقد أجرى فريق حملته تجارب متتالية عن مدى فعالية التعديلات المحسنة لموقع الحملة (BarackObama.com) لزيادة المشاركة، والعمل التطوعي، والتبرعات. فأحد الاختبارات قيّم آثار التغيير على زر «اشترك» إلى «مزيد من المعلومات»، و «انضم إلينا الآن»، و«اشترك الآن»: وأصبح من الواضح على مدى ثلاثمائة ألف زيارة أن استخدام زر «انضم إلينا الآن» قد أدى إلى عشرين في المائة زيادة في الزائرين المقبلين على التسجيل في الموقع (Issenberg 2012).

فمثل هذه المواقع الكبيرة الحجم ينبغي أن تكون أيضاً قابلة للتوسع والتطوير، وقادرة على التعامل مع الطفرات الناشئة في طلب وإنتاج البيانات، حيث يتسبب مقدار حركة مرور البيانات في انهيار قاعدة البيانات العلائقية التقليدية العاملة على خادم واحد. فعلى سبيل المثال، يمكن أن يزداد عدد التغريدات التي يتعامل معها تطبيق التواصل الاجتماعي تويتر بشكل ملحوظ، مع عشرات الآلاف من التغريدات التي يجري نشرها كل بضع ثوان أثناء الأحداث الكبيرة، مثل حفل افتتاح الألعاب الأولمبية أو خلال المباراة النهائية. وكان الحل هو تركيب أنظمة عتاد تتألف من أجزاء موزعة بحيث تخزن البيانات في قواعد بيانات مقسّمة عبر العديد من الخوادم، مما يمكن من توسيع نطاق التخزين بحسب الحاجة. وعلاوة على ذلك، ففي بعض الأنظمة، مثل تويتر، يمكن تعيين المرونة من قبل المستخدمين لتحديد ما إذا كانت تشمل البيانات أم لا. على سبيل المثال، يقرر المستخدمون في العديد من تطبيقات وسائل التواصل الاجتماعي والمتنقلة إدراج أو عدم إدراج مواقعهم، وكذلك البيانات التعريفية الرئيسية الأخرى المتعلقة بالهوية (Gorman 2013).

الخلاصة:

البيانات الكبيرة هي ظاهرة حديثة، وبالنظر إلى التطبيق والانتشار المتسارعين لها، فقد برزت نقاشات مستمرة حول ما يشكل البيانات الكبيرة والخصائص المرتبطة بها. وتُعد بعض التعاريف، مثل تلك التي تنص على أن البيانات الكبيرة هي أي مجموعة بيانات كبيرة جداً بالقدر الذي لا يمكن معه تخزينها في جدول بيانات إكسل أو على جهاز واحد (Strom 2012)، مبتذلة للغاية وغير مفيدة، حيث تخفض مستوى مفهوم البيانات الكبيرة إلى خاصية الحجم فقط. لقد أصبح من الواضح أن لدى البيانات الكبيرة عدد من الخصائص الملزمة لها والتي تجعلها مختلفة نوعياً عن الأشكال السابقة من البيانات، وتمّ في هذا الفصل مناقشة الخصائص الأساسية السبعة للبيانات الكبيرة: الحجم، والسرعة، والتنوع، والشمول، والدقة / الفهرسة، والعلائقية، والمرونة / قابلية التوسع التي تميزها عن البيانات الصغيرة (انظر الجدول ١-٢).

فهذه هي دورة المستوى الأول فيما يتعلق بتقديم تقييم توصيفي لطبيعة البيانات الكبيرة. وهناك حاجة إلى مزيد من العمل لتقييم البيانات الكبيرة الناتجة عن مصادر متعددة لتحديد ما إذا كان هناك أصناف متنوعة في طبيعة البيانات الكبيرة. فعلى سبيل المثال، قد تحمل بعض البيانات خمساً أو ستاً من هذه الخصائص، ولكنها لا تستوفي أو ضعيفة في واحدة أو اثنتين، وقد تفتقر مجموعة بيانات إلى التنوع (منظمة بشكل كبير)، أو الحجم (صغيرة، ولكن شاملة مع حجم العينة $n =$ الجميع)، أو قليلة السرعة (يتم إنشاء البيانات بانتظام ولكن في كل شهر عوضاً أن تكون بشكل مستمر)، أو تفتقر إلى الفهرسة (كونها مجهولة المصدر أو مجمعة)، مع حملها للخصائص الأخرى على سبيل المثال. فمن الواضح أن مثل هذه البيانات ليست بيانات صغيرة كما هو مبين في الفصل الثاني، ولكنها في ذات الوقت ليست بيانات كبيرة كما تُفهم بمعناها الضيق الذي يصفها بكل الخصائص السبع، ولكن يمكن اعتبارها على الرغم من ذلك شكلاً من أشكال البيانات الكبيرة. وبعبارة أخرى، هناك حاجة لإنتاج تصنيف للبيانات الكبيرة قائم على أدلة تجريبية قوية مع أمثلة الحالات التي من شأنها أن تساعدنا على التفكير بصورة أكبر في طبيعة هذه البيانات. ويحتاج هذا إلى أن يكون مصحوباً بدراسة للخصائص الأخرى، مثل نوعية البيانات وصحتها، ودقتها، ومنشئها (انظر الفصل التاسع).

كما تثير الخصائص السبع للبيانات الكبيرة تساؤلات بشأن الآثار المترتبة على طوفان مثل هذه البيانات، وما الذي يعنيه القدرة على الوصول إلى بيانات كبيرة جداً، وشاملة، ومتغيرة، ودقيقة، ومفهرسة، ومتنوعة، وعلائقية، ومرنة، وقابلة للتوسع بالنسبة للمجتمع والحكومة ورجال الأعمال؟ وإلى أي مدى يمكن أن توفر هذه البيانات رؤى ثابتة عن حالة الإنسان أو المساعدة على تناول بعض القضايا الاجتماعية، والسياسية، والاقتصادية، والبيئية الأكثر إلحاحاً والتي تواجه كوكب الأرض؟ أو، بدلاً من خدمة الصالح العام، فهل ستستخدم هذه البيانات لمزيد من خدمة المصالح الخاصة في الغالب؟ أو خدمة مصالح الدولة؟ وكيف سوف تغير هذه البيانات في نظرية المعرفة العلمية في جميع المجالات (الآداب والعلوم الإنسانية، والعلوم الاجتماعية، والعلوم الفيزيائية والحياتية، والهندسة)؟

فالفصل السابع والثامن والعاشر ستناقش هذه القضايا بالتفصيل، وتقدم التفكير النقدي لآثار وعواقب البيانات الكبيرة.

الفصل الخامس

مُمكّنات البيانات الكبيرة ومصادرها

برز النمو السريع للبيانات الكبيرة نتيجة التطور المتزامن معه لعدد من التقنيات المساعدة، والبنى التحتية، والأساليب والممارسات، وتضمينها بشكل سريع في الأعمال اليومية، والممارسات، والمساحات الاجتماعية. وتضمنت هذه البنية التحتية المعرفية الجديدة إطلاق طيف واسع من مجموعة متنوعة من تقنيات المعلومات والاتصالات، وخاصة خدمة الإنترنت الثابتة والمتنقلة عبر الهاتف المحمول؛ وإدخال البرمجيات في جميع الكائنات، والآلات والأنظمة، ناقلة إياها من «الغباء» إلى «الذكاء»، فضلاً عن إنشاء أجهزة وأنظمة رقمية بحتة بالكامل؛ وتطوير الحوسبة الشمولية الواسعة الانتشار والقدرة على الوصول إلى الشبكات والحوسبة في العديد من البيئات، وعلى ذات خطى هذا التحرك؛ التقدم في تصميم قواعد البيانات ونظم إدارة المعلومات؛ وعروض التخزين الموزّع والتخزين الأبدي للبيانات بأسعار معقولة؛ والأشكال الجديدة من تحليلات البيانات المصممة للتعامل مع وفرة البيانات بدلاً من ندرتها.

وإضافة إلى تمكين الوصول والمشاركة والتخزين للبيانات، تعد البنية التحتية والمعرفية الجديدة في الغالب الوسيلة التي يتم إنتاج الكثير من البيانات الكبيرة منها. ففي الواقع، وبلا شك أن الحالة في وقتنا الحاضر تتمثل في أننا نعيش في مجتمعات تستخدم على نحو متزايد تقنيات البيانات المكثفة والمنتجة لها وتتنظم على أساسها. كما أصبحت الممارسات الشخصية في الاتصال، والاستهلاك، والإنتاج، والسفر، والحياة المنزلية، أكثر فأكثر تتوسطها الأجهزة الرقمية التي تنتج الكثير من البيانات المستنفذة القيمة. وعلاوة على ذلك، يتم الآن مراقبة وتنظيم الأماكن التي نعيش فيها على نحو متزايد بمجاميع كثيفة من التقنيات والبنى التحتية المدعّمة بالبيانات، مثل أنظمة المرور وإدارة المباني، ونظم المراقبة والشرطة، وقواعد البيانات الحكومية، وإدارة العملاء وسلاسل الدعم اللوجستية، والأنظمة المالية ونظم المدفوعات. وستناقش العوامل المساعدة ومصادر البيانات الكبيرة بالتفصيل في هذا الفصل.

العوامل المساعدة للبيانات الكبيرة:

تُعد البيانات الكبيرة مُحصلة للتطور والتقارب بين مجموعة متقدمة من التقنيات في مجال الحوسبة منذ نهاية الحرب العالمية الثانية. ويشمل ذلك إنتاج أجهزة الحاسب المركزية في الخمسينيات والستينيات؛ وشبكات الإنترنت الوليدة في السبعينيات والثمانينيات التي ربطت الحاسبات بعضها ببعض؛ والانتشار الواسع النطاق للحاسبات الشخصية في الثمانينيات والتسعينيات؛ والنمو الهائل للإنترنت في التسعينيات، وتطوير الصناعات القائمة على الويب، جنباً إلى جنب مع النمو الهائل في الهواتف النقالة والأجهزة الرقمية مثل أجهزة الألعاب والكاميرات الرقمية؛ وتطوير الحوسبة السحابية، والموزعة، والنقالة، وتقنيات الإصدار الثاني لشبكة الويب (Web 2.0) في بدايات الألفية الثانية؛ وتفشي الحوسبة الشمولية واسعة الانتشار (Ubiquitous and Pervasive Computing) خلال هذا العقد. وخلال هذه الفترة وقع عدد من المؤثرات التحويلية، فزادت القدرة الحاسوبية أضعافاً مضاعفة، وتم ربط الأجهزة معاً، وتداخلت الأنظمة الرقمية بازدياد متعاضد في شؤون الحياة اليومية وعملياتها، وأصبحت البيانات أكثر من أي وقت مضى مفهرسة وقابلة للقراءة الآلية، كما توسعت أنظمة تخزين البيانات وأصبحت موزعة.

الحوسبة (Computation):

بينما قدّمت الحواسيب الرقمية المركزية الأولية في الخمسينيات والستينيات قدرة حاسوبية أكثر كفاءة من تلك التي عند البشر والأجهزة التناظرية التي استخدموها (مثل المعداد، والآلات الحاسبة الميكانيكية، والآلات الحاسبة بالبطاقات المثقوبة، وأجهزة الحاسب التناظرية، إلخ)، إلا أن قوة المعالجة لها كانت محدودة، ومن ثم كانت أنواع العمليات التي يمكن أداؤها مقيدة، وعلاوة على ذلك كانت تلك الحواسيب كبيرة الحجم ومكلفة. على سبيل المثال، كان بإمكان حاسب الإينياك (Electronic Numerical Integrator Analyser and Computer- ENIAC)، الذي طُوّر في منتصف الأربعينيات، أداء خمسة آلاف عملية جمع وطرح في الثانية، ولكن اقتصر ذاكرته الرئيسية عالية السرعة على عشرين كلمة (أي ثمانين بايت)، لآلة تزن ثلاثين طناً، وتحتوي ١٨,٠٠٠ أنبون فراغي، و١٥,٠٠٠ مرحّل، ومئات الآلاف من المقاومات والمكثفات والمحاثات،

الربط (Networking):

بذات القدر من الأهمية للنمو الكبير في القوة الحاسوبية لمعالجة البيانات الرقمية كان الربط بين الحاسبات بعضها مع بعض والذي سهل نقل وتبادل البيانات فيما بينها. وقد وقع أول اتصال شبكي بين جهازي حاسب عبر شبكة الإنترنت في نوفمبر عام ١٩٦٩ م، وقد تم الربط بين نقطة في جامعة كاليفورنيا في لوس أنجيلوس (UCLA) ونقطة أخرى في جامعة ستانفورد (Salus 1995). ثم نمت الشبكة ببطء، من ١٣ نقطة في يناير ١٩٧١ م، إلى ٢٣ في أبريل ١٩٧٢ م، إلى ٦٢ في يونيو ١٩٧٤ م، و ١١١ بحلول مارس ١٩٧٧ م، واقتصرت حينها على متعاقد ووكالة مشاريع البحوث المتطورة الدفاعية (Hart et al. 1992; O'Neill) (ARPA) (1995). وتعمل شبكة الإنترنت من خلال تقطيع البيانات والتعليمات في حزم متساوية الحجم ونشرها عبر النظام، مع عنوان كل حزمة بمعرف وعنوان المرسل إليه المطلوب، ثم تمرر كل حزمة من نقطة اتصال إلى أخرى إلى أن تصل إلى الوجهة المقصودة حيث يتم إعادة تجميعها هناك. ولتمكين اتصال الأجهزة والشبكات التي تعمل على سرعات مختلفة على مدار الساعة وتستخدم حزمًا مختلفة الحجم، اعتمدت مجموعة معيار (بروتوكول) الشبكة تي سي بي / آي بي (TCP/IP) (معيار التحكم بالإرسال / بروتوكول الإنترنت) في عام ١٩٧٣ (O'Neill 1995)، ولا يزال هذا المعيار يستخدم حتى يومنا هذا. كما يؤكد هذا المعيار على حيادية الشبكة، حيث يتم التعامل مع كافة البيانات على قدم المساواة ودون تمييز بينها أو تفاضل في التكلفة (Wu 2011).

وبحلول أواخر السبعينيات، تم إنشاء الشبكات المحلية الداخلية، كما تم إنشاء الشبكات الأخرى دولياً وتم ربطها بشبكة الإنترنت الوليدة، إضافة إلى شبكات الهواة المستقلين والشركات، كما أصبح من الممكن إرسال حزم البيانات عبر الأقمار الصناعية. وبالإضافة إلى نقل البيانات والملفات، تم إنشاء البريد الإلكتروني واللوحات الإعلانية (Kitchin 1998). وخلال الثمانينيات، نمت البنية التحتية، مع لاعبين جدد من المؤسسات والشركات، جنباً إلى جنب مع تطور شبكات الإنترنت الداخلية (الشبكات الخاصة). وفي عام ١٩٩٢ م اخترع تيم بيرنرز لي (Tim Berners-Lee) من سيرن (CERN) في جنيف الشبكة العالمية (الويب)، وقدّم وسيلة سهلة الاستعمال أكثر بكثير للوصول إلى الإنترنت واستخدامها. وخلال

التسعينيات وبدايات الألفية الجديدة، تمّ تطوير تقنيات شبكية جديدة مثل شبكات اتصال الميدان القريب (NFC) والتواصل المباشر (Proximate Communication) باستخدام معايير البلوتوث، والتغطية المحلية اللاسلكية الواي فاي (WiFi)، والشبكات الوطنية للاتصالات المتنقلة (GSM/3G). ووفقاً لقوانين جورج غيلدر (George Gilde) (2000) في الاتصالات، فإن إمدادات العالم من عرض النطاق الترددي (قدرتها على نقل البيانات) تتضاعف كل ستة أشهر، مع مزيد من القدرة الإضافية التي تقدّمها الشبكات اللاسلكية. ونتيجة لهذه التطورات، فقد أصبح الربط متاحاً بين الأجهزة الحاسوبية من خلال شبكة الإنترنت بشكل أسهل على نحو متزايد، وأسرع وعلى نطاق أوسع.

الحوسبة الشمولية واسعة الانتشار:

صاحب التوسع في النفاذية وقدرة الوصول إلى شبكات تقنية المعلومات والاتصالات وزيادة ساعات قنوات الاتصال (عرض النطاق الترددي)، تنوّع وتوسّع في الأجهزة الرقمية والتي تتصل إما بشكل مباشر أو غير مباشر (حيث يتمّ تحميلها على جهاز آخر أولاً) بشبكة الإنترنت لتبادل البيانات، أو التعليمات، أو الحصول على التحديثات البرمجية. ففي بعض الحالات، حلّت التقنيات الرقمية الجديدة محلّ الأشكال السابقة من البيانات التماثلية / التناظرية، فعلى سبيل المثال، تمّ استبدال أنظمة الهاتف التماثلي بالهواتف المحمولة والهواتف الذكية، واستبدلت كاميرات المراقبة التناظرية التي تسجّل على شريط فيديو بكاميرات رقمية تلتقط الصور الرقمية، ويجري الاستعاضة عن الموسيقى المخزنة على الشريط المغناطيسي أو الفينيل بملفات صوتية رقمية، والنماذج الورقية بأخرى إلكترونية على الإنترنت. من هنا، تم تحويل الأجهزة والأنظمة التي كانت تُعدّ نسبيًا «غبية» لتصبح «ذكية» ببعض الطرق، وعادة من خلال استخدام البرامج المرتبطة بها (Dodge and Kitchin 2005). وتعني كلمة «الذكية» هنا بشكل عام بأن الجهاز قد تمّت برمجته ليكون واعياً للطريقة الخاصة التي يستخدم بها وقادراً على اتخاذ قرارات مستقلة، وآنية، ومؤتمتة من مجموعة الخيارات المحددة من خلال تطبيق خوارزميات على البيانات المنتجة (انظر Dodge and Kitchin 2007a). وخير مثال على تجميع تقني «غبي» أصبح على نحو متزايد «ذكي» هو السيارة، فالسيارة الحديثة يتمّ تحميلها

بأجهزة رقمية، ومستشعرات ومحركات، والتي تراقب وتنظم أنظمتها المختلفة وتتداخل مع تجربة القيادة (Thrift 2004). ويمكن أن تنتقل البيانات المخزنة من هذه الأجهزة عبر شبكات الاتصالات عن بُعد للشركات أو تحميلها من قبل مهندس لإجراء تحليل تشخيصي (Dodge and Kitchin 2007a). وفي حالات أخرى، هناك أجهزة لا مكافئ تناظرياً لها، حيث تنتج بيانات جديدة كلياً في شكلها، مثل تلك الناتجة عن أنواع جديدة من التقنيات الرقمية، والمسح الضوئي، وأجهزة الاستشعار على سبيل المثال، مساحات التصوير بالرنين المغناطيسي MRI في مجال الصحة، وإضافة إلى جعلها «ذكية»، فإن الأنظمة تعمل في الغالب على جمع أكبر قدر من البيانات قدر الإمكان، وتسعى لتكون هذه البيانات شاملة وليست انتقائية. وهنا، قد يتم جمع بيانات عرضية أو غير ضرورية لتوفير الخدمة أو العملية، مثل التفاصيل الشخصية أو الموقع الجغرافي. فغالباً ما تُحدد الإعدادات الافتراضية للأجهزة بشكل يتيح للناس إمكانية اختيار الانسحاب من أنظمة إنتاج البيانات من خلال عقود الاستخدام للبرمجيات والمواقع، مع النص على العقوبات إذا اختاروا القيام بذلك، مثل تردي الخدمات، والتكاليف الإضافية، أو حجب الخدمة بالكلية.

وبالمثل، حصل هذا النمو في الأجهزة الرقمية في المنزل، وفي الشارع، وفي أماكن العمل، وعبر شبكات النقل، وفي غيرها، مما دفع المعلقين للافتراض بأننا قد ولجنا عصر (البرمجيات في كل مكان - Greenfield 2006) (EveryWare). إذن، هذا هو عصر الحوسبة واسعة الانتشار في كل مكان، حيث يتم توزيع القدرة الحاسوبية وإتاحتها في كل مكان، مع العديد من الأجهزة الحاسوبية التي تعمل لكل شخص (Lucas et al. 2012)، فالحوسبة المنتشرة (Pervasive Computing) هي إضافة للقدرة الحاسوبية وإتاحة الوصول إلى شبكات تكنولوجيا المعلومات والاتصالات للكائنات والبيئات اليومية الثابتة لجعلها تفاعلية و«ذكية» (Dourish 2001). وفي المقابل، فإن الحوسبة في كل مكان (Ubiquitous Computing) هي القدرة الحاسوبية التي تتحرك مع الشخص بغض النظر عن البيئة؛ كالأجهزة التي يحملها الناس مثل الهواتف الذكية وأجهزة الملاحة عبر الأقمار الصناعية التي تعتمد على الوصول إلى شبكات تقنية المعلومات والاتصالات لتقديم الخدمات. وإذا كان شعار الحوسبة المنتشرة هو الحوسبة «في كل شيء»، فإن شعار الحوسبة في كل مكان هو الحوسبة «في كل مكان»، مع إبراز الحوسبة المنتشرة لعمليات التباعد والتفرع (تضمين البرمجيات في مزيد من

الأجهزة) وإظهار الحوسبة في كل مكان لأوجه التقارب (جهاز رقمي واحد للقيام بمزيد من المهام) (Kitchin and Dodge 2011). وبينما يلزم الحوسبة المنتشرة أن تكون على إدراك موقعي/موضعي (Situationally Aware) لتنفيذها بنجاح، تتطلب الحوسبة في كل مكان الإدراك المستمر للسياق والموقع.

قابلية الفهرسة والقراءة الآلية (Indexical and Machine-Readable Identification):

على مدى نصف القرن الماضي، أصبحت الظواهر في العالم أكثر وأكثر تحديداً بشكل فريد ومن السهولة التقاطها كبيانات، بسبب التقنيات الرقمية وتقنيات العنونة ووضع العلامات الجديدة، مما زاد في درجة دقتها وعلائقيتها ومن ثم مستوى التفصيل في تحليلها، بالإضافة إلى تمكين ربط مجموعات بيانات أكثر (انظر الفصل الرابع). فعلى سبيل المثال، تتميز كل الأجهزة الرقمية التي تساهم في الحوسبة المنتشرة والحوسبة في كل مكان بمعرف فريد مُفهرس، بحيث يُعَيَّن رمز تعريف حصري لكل منها مما يتيح رصدها وتتبعها على الدوام. وعلاوة على ذلك، فإن جميع عناصر إنترنت الأشياء (Internet of Things- IoT)، كوكبة واسعة من الأجهزة المتصلة بالشبكة، المحركات، وأجهزة الاستشعار، تنتج بيانات مفهرسة يمكن تمييزها بشكل فريد (Gershenfeld et al. 2004). وإضافة لذلك، فإن المعارف الفريدة التي تم ربطها بالناس منذ زمن طويل، مثل جوازات السفر، والبطاقات الصحية، وسجلات التأمينات الاجتماعية، قد جرى تعزيزها من خلال أنظمة تحديد الهوية الرقمية والحيوية الجديدة التي تربطها بقواعد بيانات رقمية تتحقق من هويتهم (Dodge and Kitchin 2005). وهذه تشمل أسماء المستخدمين وكلمات السر، وبطاقات السحب المغناطيسية، والقارئ الرقمية لبصمة الأصابع، وقزحية العين، والحمض النووي. وفي كل حالة، تعد هذه المعارف بمثابة المصافحة الرقمية، على الرغم من أنها تختلف في مستواها من حيث الدقة والثقة. فعلى سبيل المثال، يمكن لأسماء المستخدمين وكلمات المرور أن تُسرق أو تُزور بسهولة، في حين أن الحمض النووي هو أصيل وفريد لكل شخص ويحقق المعايير الأربعة الأولى من قائمة كلارك (Clarke 1994b) المتعلقة بالخصائص المرغوب فيها لتحديد هوية الكائن البشري بشكل فعّال وهي: شمولية التغطية / النطاق (يجب أن يمتلك الجميع الخاصية نفسها)؛ التفرد (يجب أن يمتلك كل شخص قيمة واحدة فقط

لهذه الخاصية، ولا يتساوى فيها أي شخصين؛ الدوام (يجب أن لا تتغير قيمة الخاصية بمرور الزمن، ولا تكون قابلة للتغيير أيضاً)؛ اللزوم (بمعنى أن تُعد الخاصية صفة طبيعية لا يمكن إزالتها ولا غنى عنها). وهذه الصفات تتوافق مع الصفات السبع الأخرى وهي: القدرة على التحصيل / الجمع؛ والقدرة على التخزين؛ والتفرد؛ والدقة؛ والبساطة؛ والتكلفة والتناسب؛ والقبول.

وبالمثل، فإن رموز التحديد المكانية هي أنظمة تحدّد مواقع الأشخاص، والأماكن، والكائنات في جميع أنحاء العالم، وتشمل خطوط الطول والعرض، وإحداثيات الخرائط الشبكية، والوحدات الإدارية مثل الرموز البريدية (Zip Codes). وتختلف هذه الأنظمة في نطاقها ودقتها (درجة التفصيل)، وقد أصبحت أكثر دقة وتحديداً من أي وقت مضى. فعلى سبيل المثال، تُوسم البيانات الإدارية بالنسبة إلى الملكيات الفردية أو قطع الأراضي بدلاً من الأحياء؛ كما يسجّل أي جهاز يحوي نظام تحديد المواقع العالمي (Global Positioning System- GPS) الموقع في نطاق بضعة أمتار؛ وبدأت أنظمة الخرائط الوطنية بإسناد عدد كبير من الكائنات الناشئة بفعل بشري أو طبيعي في المخططات الخرائطية بمعرفات إقليمية فريدة. ويوجد معرفات فريدة أخرى شائعة على نحو متزايد للمنتجات (انظر الفصل الرابع)، والمعاملات (على سبيل المثال، طلبات الشراء)، والتفاعلات (على سبيل المثال، البريد الإلكتروني والمكالمات الهاتفية).

وفضلاً عن كونها أكثر قابلية للفهرسة والتبويب، أصبحت رموز تحديد الهوية قابلة للقراءة الآلية أيضاً على نحو متزايد (Dodge and Kitchin 2005). على سبيل المثال، يحتوي الرمز الشريطي (Barcode) على رمز تعريف يرتبط بنظام معلومات يحتفظ ببيانات وصفية إضافية تصف الكائن عند مسحه بقارئ ليزري (مثل نوع المنتج / النموذج، تاريخ ومكان الصنع، الأسعار، وما إلى ذلك). كما يمكن استشعار رقائق تحديد الهوية باستخدام موجات الراديو (RFID) عن بُعد من قبل مستجيب راديوي والتعرف عليها. ويمكن كذلك التعرف على المركبات بواسطة برنامج التعرف التلقائي على أرقام لوحاتها، وقد تمّ تطوير تقنيات خوارزمية جديدة للتعرف وتحديد هوية الوجوه وطريقة المشي (كيف يمشي الناس). وهكذا أصبحت عناصر / كائنات العالم مفتوحة لجمع والتقاط بياناتها وتحديد

هويتها تلقائياً، مما يوسّع بشكل كبير نطاق البيانات المنتجة ومدى دقتها بالنسبة إلى نظام ما، مما قدّم أشكالاً جديدة من التنظيم والحوكمة (Dodge and Kitchin 2007a) وعزّز عمل الأنظمة بطرق أكثر آلية (Kitchin and Dodge 2011). لذا فإن الاتجاه السائد هو المزيد والمزيد من التحديد للظواهر والممارسات لتصبح معرّفة بشكل فريد وقابلة للقراءة آلياً، مما يزيد الدقة، والعلائقية، والأتمتة.

تخزين البيانات (Data Storage):

كان هناك تحوّل كبير، على مدى العقدين الماضيين، في القدرة على التخزين الطويل الأمد لكميات هائلة من البيانات، ففي البداية تمّ تخزين البيانات الرقمية في صورة تناظرية باستخدام البطاقات المثقوبة وذلك قبل أن يتم الانتقال إلى الشريط المغناطيسي والأقراص الصلبة. وقد استكملت هذه الوسائل باستخدام وسائط التخزين الرقمية المتنقلة والتي تُعرف باسم الأقراص المرنة (Floppy Disks)، وتلا ذلك ظهور الأقراص المدمجة (CDs)، وأقراص الفيديو الرقمية (DVDs)، ومحركات الأقراص الخارجية والذاكرة المتنقلة (Flash/ External Drives). وقد دفعت تقنيات الأقراص الصلبة (Hard-disk Technologies)، على وجه الخصوص، بالقدرة على تخزين البيانات الرقمية بشكل كبير، وذلك مع النمو الكبير للسعة التخزينية المقدرّة بالخانة الثنائية لكل بوصة مربعة (بت/إنش²)، في حين انخفضت التكلفة لكل غيغابايت بشكل كبير وبعامل يقدر بعشرة ملايين إلى واحد من عام 1956 م إلى 2000 م (Gilheany 2000) واستمرت الأسعار في الهبوط، مما جعل التخزين الرقمي أرخص تكلفة بكثير وأقل مساحة من الأشكال التناظرية (على سبيل المثال، الورق أو الأشرطة الفلمية) (Grochowski and Halem 2003). فالآن، أصبحت التكاليف رخيصة جداً وبالشكل الذي يُمكن من تخزين كميات هائلة من البيانات بمقابل بخس، ودون الحاجة تقريباً لحذف الملفات القديمة. ونتيجة لذلك، فإن المعاملات التي تجري اليوم (على سبيل المثال، رسالة أرسلت إلى قائمة البريد الإلكتروني أو صفحة موقع التواصل الاجتماعي الفيسبوك، أو المدفوعات إلى متجر باستخدام بطاقة الائتمان، أو المحادثة مع أحد الأصدقاء على الهاتف) من الممكن تسجيلها وحفظها بشكل جيد إلى ما بعد وفاة الشخص المعني بها، مع إمكانية الإشارة إليها وتحليلها في أي وقت في المستقبل.

ونتيجة لذلك، وكما نوقش في الفصل السابق، كان النمو في حجم البيانات المخزنة بالأرقام المطلقة والنسبية استثنائياً، وخاصة منذ العام 2000م. فعلى سبيل المثال، قدّر Hilbert and López (2009) أن خمسة وعشرين في المائة فقط من البيانات تم تخزينها في شكل رقمي في عام 2000، مع الاحتفاظ بما تبقى بأشكال تناظرية مثل الكتب، والمجلات، والصور، والأشرطة الممغنطة. وبحلول عام 2007 م، شكّلت البيانات الرقمية ما نسبته أربع وتسعون في المائة من مجموع البيانات المخزنة، ومنذ ذلك الحين تواصل النمو للحصة النسبية من البيانات الرقمية، خاصة مع تطور التخزين والخدمات الموزعة من خلال الحوسبة السحابية ومراكز البيانات. وتأخذ الحوسبة السحابية أحد شكلين يعملان في الغالب بصورة تعاونية: سُحِبَ المرافق الخدمية (Utility Cloud)، وسُحِبَ البيانات التخزينية (Farber et al. 2011) (Data Clouds). فالسُحِب الخدمية تُوفّر القدرات والإمكانات المتعلقة بتقنية المعلومات كخدمات حسب الطلب حيث يُمكن الوصول إليها عن طريق شبكة الإنترنت وذلك بشكل مستقل عن موقع طالب الخدمة. ويشمل ذلك «البنية التحتية كخدمة» (Infrastructure as a Service- IaaS) مثل التخزين والخوادم والشبكات، «المنصة كخدمة» (Platform as a Service- PaaS) وتضم بيئة التنفيذ لتطوير تطبيقات مخصصة وقواعد البيانات، و«البرمجيات كخدمة» (Software as a Service- SaaS) والتي تمكن المستخدمين من الوصول إلى التطبيقات ومعالجة البيانات عن بُعد (Farber et al. 2011; Hancke et al. 2012). في حين تُمكن سحب البيانات من ربط، تخزين، ومعالجة كميات هائلة من البيانات التي قد تنتج عبر المؤسسة عن بُعد، بالاعتماد على القدرة الحاسوبية لمئات من الأجهزة، وتحليلها عن طريق سُحِب المرافق الخدمية (Farber et al. 2011). ومن ثم، يمكن للأفراد والشركات الاستفادة من التخزين والقدرة الحاسوبية دون الحاجة إلى القيام باستثمارات رأسمالية كبيرة، فضلاً عن كونها قادرة على الاستفادة من هذه الموارد من أي مكان يتوافر فيه اتصال بالإنترنت (Bryant et al. 2008).

وبالمثل، كان هناك تحوّل في كيفية تنسيق البيانات وهيكلتها عند تخزينها، وبسبب الزيادة الناشئة في حجم، وسرعة، وتنوع البيانات، فقد عانت قواعد البيانات العلائقية للبقاء ومواجهة الموقف. فقد صُممت لبيانات على درجة عالية من التنظيم، والسلامة،

والمنظمة، والثابتة. وبدلاً عنها، ظهر شكل جديد من قواعد البيانات هو قواعد البيانات غير الهيكلية (NoSQL)، مما يتيح تخزين البيانات الكبيرة. وتميل قواعد البيانات غير الهيكلية إلى أن تكون أقل تعقيداً في الشكل من قواعد البيانات العلائقية، ولكن يمكنها التعامل مع كل من البيانات المنظمة (Structured Data) وغير المنظمة (Unstructured Data)، وهي قابلة للتوسع بسرعة، ويمكنها التعامل فعلاً مع العديد من مجموعات بيانات سريعة التغير وهائلة في الحجم، وتمتاز بالمرونة في كيفية تنظيمها، مما يتيح هياكل بيانات مفصلة بحسب الطلب (Driscoll 2012; Marz and Warren 2012). كما تميل قواعد البيانات غير الهيكلية إلى تخزين البيانات الابتدائية الخام بدلاً من المشتقة التي يتم هيكلتها وفقاً لعلاقة محددة سلفاً (Driscoll 2012). فعلى سبيل المثال، تُخزّن قواعد البيانات غير الهيكلية معلومات خام عن زيارة صفحة الويب (المشاهدة) بدلاً من تخزين تعداد المشاهدات للصفحات المشتقة منها، وهي إستراتيجية أكثر قوة من حيث أنها تقلّل من الأخطاء المحتملة في قواعد البيانات، ولكن هذه الإستراتيجية لا يمكن تطبيقها مع قواعد البيانات العلائقية بسبب النمو السريع والحجم الناتج من ذلك النمو والذي لا يمكن التعامل معه من خلالها (Marz and Warren 2012). ويتم في العادة توزيع البيانات وتكرارها عبر العديد من الأجهزة في قواعد البيانات غير الهيكلية، بدلاً من مركزيتها في مكان واحد (ومن ثم حل مشكلة مجموعات البيانات الكبيرة جداً كونها كبيرة جداً عن أن يجري احتواؤها على جهاز واحد)، وتُقسّم طلبات الاستعلام ويجري تنفيذها في الوقت نفسه على عدد من الأجهزة جنباً إلى جنب ثم يُعاد تجميع النتائج (ومن ثم تسريع العمليات الحسابية) (Dumbill 2012). وتتلخص هذه الإستراتيجية من خلال تقنية التعيين والاختزال (ReduceMap) التي تمّ تطويرها من قبل شركة جوجل حيث يتم تعيين البيانات والاستعلامات على عدة خادمات ومن ثمّ جمع النتائج الجزئية واختزالها معاً (Dumbill 2012). ويُعد هادوب (Hadoop)، الذي تمّ وضعه بداية من قبل شركة ياهو (Yahoo)، مثالاً على قواعد البيانات غير الهيكلية ومفتوحة المصدر وشائعة الاستخدام في تطبيق نموذج التعيين والاختزال. فالحل البديل لمثل هذا النهج يتمثل باستخدام نموذج أكثر محدودية للبيانات، ودرجة أقل من العلائقية في مجموعة البيانات، وقدرة أقل على الاستكشاف باستخدام لغة الاستعلام العلائقية (SQL) (Driscoll 2012). وهذا يعني أن

قواعد البيانات غير الهيكلية مثالية للتخزين الفعّال والاسترجاع السريع لكميات هائلة من البيانات، ولكنها أقل فائدة لفحص العلاقات بين عناصر البيانات (مركز تحالف البيانات المفتوحة Open Data Center Alliance 2012)

مصادر البيانات الكبيرة (Sources of Big Data):

وَفَر التوسع الجذري والتكامل بين الحوسبة، والشبكات، والأجهزة الرقمية، وتخزين البيانات منصة قوية لانطلاق البيانات الكبيرة، وإضافة لذلك الوسيلة التي يتم بها إنتاج، ومعالجة، ومشاركة، وتحليل البيانات الكبيرة. ويمكن تصنيف هذا الإنتاج للبيانات في ثلاث فئات رئيسية. البيانات الموجهة (Directed Data) والتي يتم إنتاجها بوسائل مراقبة تقليدية، حيث تتركز أنظار التقنية على شخص أو مكان من قبل مشغل بشري. أما في حالة البيانات الآلية (Automated Data)، فتنتج البيانات عن وظيفة كامنة في الجهاز أو النظام بطريقة تلقائية، في حين يتم تداول البيانات الطوعية (Volunteered Data) وإهداؤها من قبل الناس للنظام.

البيانات الموجهة (Directed Data):

توجد المراقبة المنظمة والهيكلية حيثما وجدت مجموعة من الناس الذين يشرفون على أشخاص آخرين كالمسؤولين عن إنفاذ القانون، والمعلمين، والأطباء، ومسؤولي الرعاية الاجتماعية، والبيروقراطيين، ورؤساء العمل والذين يعملون على مراقبة الآخرين شخصياً كالمواطنين، والتلاميذ، والمرضى، والعمال أو من خلال العدسات التقنية كالمسوح مثل التعداد السكاني، ونماذج البيانات الحكومية، وإيصالات الضرائب، وعمليات التفتيش، وكاميرات مراقبة الدوائر التلفزيونية المغلقة (CCTV)، حيث كانت المراقبة المنظمة والهيكلية منذ فترة طويلة سمة من سمات المجتمعات، ومكوّناً أساسياً من مكونات الدولة وحوكمة الشركات (Lyon 2007). فهذا النموذج للتحكم (التداخل المنطقي، الأدوات، المؤسسات، وأدوار وإجراءات الحوكمة) يُتيح مركزية التنظيم والتحكم عبر مجموعة واسعة من المجالات، ويساعد في الحفاظ على النظام، وإنتاج حكومة جيدة، وإدارة فعّالة، وعمل مربح، ومجتمعات مستدامة ومستقرة، من خلال المحاسبة الفعّالة للأشخاص والتهديب الذاتي

أيضاً (بمعنى، تعديل الناس سلوكهم ليتوافق مع التوقعات والقواعد). وتُستكمل هذه النظم بطرق تحديد هوية الأفراد، مثل بصمات الأصابع، والصور، وأرقام التأمين الوطني أو الضمان الاجتماعي، وجوازات السفر ورخص القيادة، التي تجعل من السهل تتبع واقتفاء أثر الناس، كما تمّ توسيع نطاق المراقبة ليشمل الحيوانات والبيئات، واستخدام الكائنات وحركتها.

وحتى وقت قريب، كانت كل عمليات المراقبة جزئية وتناظرية بطبيعتها وتنتج مستويات مختلفة من البيانات المسجّلة، بدءاً من الملاحظات غير المسجّلة إلى السجلات التفصيلية أو التسجيل المستمر، ويجري تطبيقها أحياناً على عينات وفي حالات قليلة على المجتمع بأكمله. وكانت البيانات المسجّلة انتقائية، تُجمع من خلال التقارير أو الإبلاغ الذاتي على الورق أو على شريط مغناطيسي أو فيلم، في مكان وزمان معيّنين. وتميل البيانات إلى أن تكون منظمة للغاية في طبيعتها، ومقتصرة على مجموعة من حقول البيانات المحددة مسبقاً. وكانت السجلات ضخمة ومكلفة في التخزين، مما يعني عدم الاحتفاظ الطويل الأمد بها إلا لكمية محدودة من البيانات، إضافة إلى كون الكثير منها إما عبارة عن عينة أو مشتقة. وحتى في الحالات التي قد تكون المراقبة فيها مستمرة، مثل الدوائر التلفزيونية المغلقة، فقد تمتدّ اللقطات نفسها على فترات زمنية طويلة تتعدى قدرة الشريط التسجيلي على تغطيتها ويجري في العموم إعادة التسجيل عليها بعد فترة معينة (غالباً بعد سبعة أيام أو شهر). كما كان تحليل البيانات الناتجة من مختلف أشكال المراقبة المستخدمة يتمّ بواسطة مفسّر بشري إلا إذا جرى تحويلها رقمياً في وقت لاحق. وفي حين أن كميات البيانات التي تنتجها هذه المراقبة يمكن أن تكون ضخمة، وربما تستوفي واحدة أو اثنتين من الخصائص المرتبطة بالبيانات الكبيرة مثل قابلية الفهرسة والشمولية والعلائقية أو السرعة، إلا أنه لا يمكنها بأي حال أن تستوفي مجموعة الصفات المرتبطة بالبيانات الكبيرة كاملة.

وفي حين لا تزال العديد من أنظمة المراقبة والحوكمة مستمرة في كونها جزئية وتناظرية، كان هناك تحرك في الآونة الأخيرة لاستبدالها والتوسع فيها بنظائر رقمية مكافئة لها والتي تنتج الآن بيانات كبيرة. فعلى سبيل المثال، يجري استبدال أنظمة الدوائر التلفزيونية المغلقة التماثلية الثابتة بأخرى رقمية متحركة حيث يمكن تغيير مجال الرؤية، والتقليل من

المحددات على التخزين، وتعزيز التفسير البشري بتحليلات خوارزمية حسابية، مثل التعرف على الوجه أو طريقة المشي، كما يمكن ربط مثل هذه البيانات مع قواعد بيانات أخرى. وهنا، يوجّه تركيز الكاميرا، وعمليات التحليل والتفسير، من قبل مشغّل بشري، وبمساعدة من البرمجيات. على سبيل المثال، من الممكن مراقبة وضبط العديد من الكاميرات في غرفة عمليات الشرطة وربطها بسجلّ الحوادث الحية من أجل الاستجابة بكفاءة وتوجيه الموارد المناسبة لمواقع معينة بشكل استباقي. وبالمثل، في مراقبة الهجرة والجوازات، حيث يجري أنياً جمع معلومات الركاب التفصيلية وتدقيقها مع مختلف قواعد البيانات الأمنية والحدودية بشكل آني أو لحظي (Real-Time)، وإنتاج بيانات جديدة مثل المسح الضوئي للصور الشخصية وبصمات الأصابع أو قزحية العين. وتستخدم العديد من الحكومات المحلية أنظمة إدارية لتسجيل مشاركات المواطنين مع خدماتها أنياً ورصد إن كان تعامل الموظفين مع أي قضايا قد تمّ ضمن أطر زمنية محددة، مع مراقبة مديري الإشراف للنظام لإعادة توجيه الموارد بحسب الحاجة.

وتتكوّن بعض الأنواع الأخرى الأقل انتظاماً في إنتاج البيانات الكبيرة الموجهة من: التصوير الجوي الرقمي عن طريق الطائرات أو الطائرات بدون طيار، أو الفيديو المكاني، أو الليدار (الرادار بتقنية الليزر) لكشف الضوء وتحديد مدى البعد، أو الحرارية أو غيرها من أنواع الأشعة الكهرومغناطيسية للبيئات التي تُمكن من الرسم المتحرك للخرائط الثنائية والثلاثية الأبعاد للمناظر الطبيعية المفهرسة مكانياً بشكل آني أو لحظي (Real-Time) من خلال نظام تحديد المواقع العالمي (GPS). وتنتج مثل هذه الفيديوهات والمسوح بالأشعة كميات هائلة من البيانات المتنوعة عالية السرعة. فعلى سبيل المثال، يمكن لمسح الليدار توليد مليون نقطة بيانية، مع دقة بيانية تصل إلى بضعة سنتيمترات، مع إمكانية إجراء أكثر من مسح خلال الثانية الواحدة. إن وضع ماسح ليدري في مركبة وقيادتها في الأنحاء، أو الطيران به فوق منطقة معينة ينتج تمثيلاً ثلاثي الأبعاد مفصلاً للغاية للتضاريس الطبيعية لتلك المنطقة.

البيانات الآلية (Automated Data):

في حين وسّعت البيانات الموجهة أنظمة المراقبة والحوكمة التقليدية، فقد حوّل إنتاج البيانات الآلية بشكل جذري كيفية إنتاج البيانات وتوظيف استخداماتها، ومكّن من إنشاء أنظمة جديدة للتنظيم والإدارة التلقائية (Dodge and Kitchin 2007a, Kitchin and Dodge 2011؛ انظر الفصل العاشر). فمن خلال مجموعة من التقنيات الرقمية يتم إنشاء البيانات تلقائياً بواسطة مجموعة من النظم الآلية وبرقابة بشرية قليلة. وعلاوة على ذلك، يتم في الغالب معالجة وتحليل مثل هذه البيانات تلقائياً وبشكل مستقل، إذ يجري العمل عليها بواسطة خوارزميات برمجية لتحقيق نتيجة معينة تهدف إلى تنظيم ظاهرة ما. وهناك طرق عديدة لإنتاج البيانات الآلية، بعضها قد يكون إنتاجه ثانوياً عن النظام وليس من صميم عمله.

المراقبة الآلية (Automated Surveillance):

كما أصبحت تقنيات المراقبة رقمية بطبيعتها ومرتبطة بعضها ببعض، فقد أصبح من الممكن أيضاً أتمتة جوانب متعددة لنظام المراقبة، وإضافة تقنيات جديدة، لتتبع أكثر فعالية وكفاءة في مراقبة وتتبع الاستخدام للأنظمة والأماكن المختلفة. وتعدّ قراءة العدادات الذكية مثلاً على النموذج اليدوي من المراقبة التي غدت آلية بشكل متزايد. هنا، يتم استخدام تقنية قراءة العداد آلياً (Automatic Meter Reading- AMR) لمراقبة استخدام الآلة والتواصل معها دون الحاجة إلى القراءة اليدوية التقليدية (Hancke et al. 2013)، وعلاوة على ذلك، يمكنها القيام بهذه المهام بشكل متواصل مما يمكن المورد من تتبع الاستخدام بشكل آني أو لحظي، وهذا له فائدة في مطابقة الطلب مع العرض وفي العثور على أخطاء / تسريبات في النظام. كما أنها توفر وسيلة للقيام بالفوترة الآلية، مما يحدّ من نفقات التوظيف. وفي كثير من الحالات، من الممكن للمستهلك أيضاً أن يراقب استهلاكه الخاص لخدمة معينة، مثل الكهرباء أو الغاز أو المياه، فعلى سبيل المثال، يمكن استخدام تطبيق للهاتف الذكي للاتصال بعداد عن بعد باستخدام تقنية اتصال المجال القريب على سبيل المثال، البلوتوث، ومن ثم تغيير الاستهلاك تبعاً لذلك.

وبالمثل، تعزز أنظمة إدارة المرور من القدرة على تنظيم وضبط السائقين من خلال الاستفادة من تقنية التعرف الآلي على أرقام لوحات المركبات (Automatic Number Plate Recognition- ANPR) (Dodge and Kitchin 2007a). فهنا، تسمح الخوارزميات ببيانات الصور من الكاميرات الرقمية للكشف عن لوحات الترخيص. ومن جهة، يمكن استخدام هذه المعلومات لتتبع المركبات خلال عبورها على طرقات المدينة وتقديم مدخلات إلى نظم النقل الذكية (Intelligent Transportation Systems- ITS)، ومن جهة أخرى، يمكن إحالة هذه البيانات إلى قاعدة بيانات مالكي المركبات لإدارة الغرامات والعقوبات المتعلقة بالمخالفات المرورية. فعلى سبيل المثال، بالنسبة للمثال السابق، يتم فحص لوحة الترخيص لجميع المركبات الداخلة إلى منطقة الزحام المروري في لندن ومطابقتها مع قاعدة بيانات تلك المركبات التي سددت رسوم الازدحام، أما تلك التي لم تسدد خلال فترة أربع وعشرين ساعة فيتم تغريم مالكيها تلقائياً من خلال عملية الإدارة الآلية (يتمتع النظام باستقلالية في إصدار الغرامات دون الحاجة لرقابة بشرية). ويمكن للنظام بالمثل استخدام كاميرات السرعة لإصدار قسائم تجاوز السرعة للسائقين المخالفين.

وفي حالات أخرى، تمّ تسهيل المراقبة الآلية من خلال استخدام رموز الهوية المقروءة آلياً لتسجيل ماهية الأنشطة المرورية/ المجهولة على شبكة المراقبة. فعلى سبيل المثال، تمّ استبدال التذاكر الورقية التي يتزايد تبادلها في السفر على نظام النقل من غير حاملها بالبطاقات الذكية، التي ينبغي التدقيق عليها لدخول وخروج المحطات ووسائل النقل. ويتم إنشاء سجل عند نقاط المسح ومن ثم يمكن تتبع كل حركات حاملي البطاقات الذكية. وفي حالة نظام النقل في لندن، فإن استخدام بطاقات أويستر (Oyster) يعني أنه من الممكن تتبع الرحلات الفردية عبر نظام الحافلات والسكك الحديدية المستخدمة من قبل مليوني راكب يومياً (Batty et al. 2012). وبالمثل، من الممكن الآن المراقبة التلقائية لأشكال متنوعة من الخدمات العامة المقدمة التي كان يجري رصدها عرضاً، مثل جمع النفايات، من خلال استخدام رقائق تحديد الهوية باستخدام موجات الراديو (RFID) التي تثبت على حاويات النفايات ويتم مسحها عن بُعد عند نقاط الجمع. وأصبح من الممكن قياس كميات التخلص من النفايات باستخدام مقاييس مرفقة بمجموعة مركبات وتحميل الأسر بصورة فردية تكلفة التخلص من النفايات على أساس الحجم بدلاً من الرسم الموحد

دون تمييز. كما أصبح من الممكن أيضاً الحصول على فهم مفصّل لأنماط التخلص من النفايات داخل منطقة ما وإعادة تخطيط طرق وأوقات الجمع.

الأجهزة الرقمية (Digital Devices):

كما سبق الإشارة إليه، كان هناك نمو هائل في عدد الأجهزة الرقمية المستخدمة حول العالم والتي تساعد على تسهيل، وتعزيز، وإدارة العديد من جوانب الحياة اليومية. فهذه الأجهزة لا تقوم بمجرد معالجة البيانات الرقمية والتعليمات فحسب، ولكنها تنتج بيانات رقمية بشكل نشط وكوظيفة أساسية لها أو ثانوية، أو الاثنين على حدّ سواء. ومن الأمثلة على الأجهزة التي تنتج بيانات رقمية كوظيفة أساسية: الكاميرات، وأشرطة الفيديو، والمراقب ووحدة تحديد المواقع (GPS)، وأشكال مختلفة من المعدات الطبية مثل كاميرات المنظار، والمساحات بالموجات فوق الصوتية، وموازن الحرارة الرقمية، إلخ. أما أمثلة الأجهزة التي تنتج بيانات رقمية كوظيفة ثانوية، فتشمل الهواتف النقالة ومستقبلات البث التلفزيوني عبر الكابل أو الأقمار الصناعية، والتي تنتج بيانات متعلقة بكيفية استخدامها (على سبيل المثال، الوقت، المكان، الشخص المتصل / القناة المشاهدة)، مع البيانات المتصلة بها والتي جرى جمعها في سجلات يتم إرسالها إلى أطراف ثالثة (مثل مصنعي الأجهزة ومقدمي الخدمات). أما الأمثلة على الأجهزة التي تنتج البيانات الأساسية والثانوية معاً فتشمل الهواتف الذكية التي يمكن استخدامها لتسجيل الملاحظات، والتقاط الصور، وكتابة البريد الإلكتروني والتعليقات على وسائل التواصل الاجتماعي،... إلخ، بالإضافة إلى تسجيل استخدام مختلف التطبيقات. وتشكّل هذه الأجهزة سلالة جديدة من الكائنات، تُسمى لوجكتس (Logjects) (Kitchin and Dodge 2011) - كائنات ترصد وتتبع تاريخ استخدامها الخاص. وفي كثير من الحالات، يكون من الممكن تسجيل موقع هذه الأجهزة. فالهاتف الجوال والهواتف الذكية يمكن تتبعها من الفضاء عن طريق التثليث (حساب المثلثات) عبر هوائيات الهاتف، وأيضاً عن طريق مستقبلات تحديد المواقع المدمجة فيها، بالإضافة إلى مراقبة السرعة والاتجاه من خلال المدوار (Gyroscopes)، ومقاييس التسارع (Accelerometers)، والبوصلات (Compasses) على شكل أقراص أسطوانية، أو أجهزة ملاحية بالأقمار الصناعية أو أي أجهزة أخرى. وعلاوة على ذلك، فإن العديد من هذه

الأجهزة يمكن أن تتفاعل بشكل مستقل بعضها مع بعض وتقوم بتبادل البيانات. ومع تكاثر الأجهزة المعقدة لكافة المهام وفي كل الأماكن، يتم إنتاج المزيد والمزيد من البيانات الغنية بالبيانات الوصفية القابلة للفهرسة، مما يتيح تحليلاً أكثر تطوراً واستحداث إستراتيجيات ومنتجات جديدة قائمة على البيانات.

البيانات المستشعرة (Sensed Data):

تعد المستشعرات والمحركات من أنواع الأجهزة الرقمية. وهي ما يمكن أن تكون ضمناً أو جزءاً من هياكل مختلفة لقياس مخرجات معينة مثل مستويات الضوء، والرطوبة، ودرجة الحرارة، والغاز، والمواد الكيميائية، والمقاومة الكهربائية، والصوتيات، وضغط الهواء، والحركة، والسرعة،... إلخ. وتُعد أجهزة الاستشعار والمحركات صغيرة، وغير مكلفة في العموم، ويمكنها إنتاج تيار مستمر من البيانات، ويمكن أن تكون سلبية وتقرأ بواسطة الماسحات الضوئية، أو يمكن أن تكون نشطة، بحيث تبتّ البيانات على فترات منتظمة عبر شبكات الحاسب المحلية أو الواسعة، أو قد يكون لديها قدرة الاتصال من خلال حقل الاتصالات القريب (NFC) والذي يتيح الاتصالات في اتجاهين (Hancke et al. 2013). وبوضعها على جسر، يمكن للمستشعرات أن تقيس وتتواصل لحساب معدلات التآكل (عن طريق قياس المقاومة الكهربائية) ومقدار الضغط على مواد البناء باستخدام أجهزة استشعار الانبعاثات الصوتية (للكشف عن انتشار الموجات الصوتية) وأجهزة استشعار التخضر المغناطيسي (للكشف عن تغيرات في الحث المغناطيسي)، ومن ثم يقلل ذلك من الحاجة إلى التفتيش اليدوي المكلف (Hancke et al. 2013). وبالمثل، يمكن للمستشعرات داخل نظام المياه قياس نوعية المياه، وضغطها، ومقدار تدفقها، وتمكين الإدارة الآنية وصيانة الأنابيب لحظة بلحظة. وعلى المركبات، يمكن للمستشعرات رصد حجم العمل، والإجهاد، والتضاريس. فشركة هيتاشي تستخدم هذه المجسات لرصد مركباتها الثقيلة المعقدة والمكلفة، وكذلك تفعل الشيء نفسه شركة النقل السريع (UPS) مع أسطولها من عربات النقل والشاحنات بحيث توظفها لتشغيل خدماتها بكفاءة وبصورة استباقية (Plumridge 2012; Mayer-Schonberger and Cukier 2013). ومن خلال ربط رقائق تحديد الهوية باستخدام موجات الراديو (RFID) بالمنتجات أصبح من الممكن رصد وتتبع حركة الوحدات الفردية من المصنع

أو المزرعة إلى المستهلك، والحدّ من السرقة والهدر. وبالمثل، تحتوي بطاقة الضريبة على شريحة تحديد الهوية باستخدام موجات الراديو التي تتواصل مع قارئ ترددات راديوي على حواجز تحصيل الرسوم، مما يؤدي إلى سرعة أعلى في حركة المركبات والدفع الآلي، بالإضافة إلى حساب معدل حركة المركبات أو عدد مواقف السيارات المتاحة لإدارة حركة المرور. كما تثبت رقائق تحديد الهوية باستخدام موجات الراديو على الحافلات والقطارات والتي تتواصل مع قارئ الموجات الراديوية على طول مساراتها مما يجعل من الممكن تتبع موقع المركبات في أي لحظة آلياً.

كما يمكن إنشاء شبكة استشعار كثيفة من خلال توزيع العديد من أجهزة الاستشعار وربط البيانات التي تنتجها، مما يتيح رصد أوضاع مختلفة عبر نظام أو مكان ما. وقد توالى الأمثلة على هذه الشبكات وأجهزة الاستشعار في مدينة سانتاندر (Santander) في إسبانيا كجزء من مشروع المدينة الذكية (SmartSantander)، وفيه 12000 من أجهزة الاستشعار، وتتكون من خمسة أنواع، تمّ توزيعها بشكل إستراتيجي في جميع أنحاء المدينة لقياس الضوضاء، والحرارة، ومستويات الإضاءة المحيطة، وتركيز غاز أول أكسيد الكربون، والأماكن المتوافرة لوقوف السيارات وركنها (Bielsa 2013). وترتبط أجهزة الاستشعار معاً من خلال الاتصالات اللاسلكية وتوظيف البرمجة عبر الأثير (OTAP)، مما يمكن من برمجتها وتحديث برمجياتها عن بُعد (Bielsa 2013). كما يجري تطوير وادي بلان أي تي (PlanIT) في البرتغال، وهو في الوقت الحاضر تحت الإنشاء ومصمّم لإيواء 225 ألف نسمة، ويهدف إلى إنشاء بيئة قائمة على أكثر من 100 مليون جهاز استشعار مدمج فيها والتي سوف تنتج بيانات لرصد مجموعة متنوعة من البنى التحتية والبيئات (Marchetti 2012). وهكذا، يُنظر إلى المستشعرات بوصفها عنصراً أساسياً في التطورات التي تشهدها المدينة الذكية.

بيانات المسح الضوئي (Scan Data):

بالإضافة إلى الأجهزة الرقمية وأجهزة الاستشعار التي تنتج البيانات بشكل نشط والتي هي جزء من شبكة إنترنت الأشياء، من الممكن رصد وتتبع الكائنات الأخرى رقمياً من خلال رموز الهوية المقروءة آلياً. وقد تجاوز استخدام الرمز الشريطي (Barcode) قطاع البيع بالتجزئة ليشمل السلع المصنّعة والاستثمارات الرسمية، كما وضعت العديد من المنظمات

الكبرى والقطاعات الصناعية نموذجها الخاص من الرمز الشريطي ومعاييرها لتخصيص الأرقام (Dodge and Kitchin 2005). وتمثل البيانات في الرموز الشريطية ثنائية الأبعاد أفقياً ورأسياً مجموعة من النقاط المرقطة ويمكن استخدامها لتحديد كائن بشكل فريد. وتستخدم بعض خدمات البريد هذه الرموز الشريطية الثنائية الأبعاد لتمكين المتابعة والتعقيب للرسائل والطرود البريدية بشكل فردي. وبالمثل، تحتوي الشرائط المغناطيسية المثبتة على بطاقات الائتمان والولاء وبطاقات السماح بالعبور من خلال الأبواب على معرف فريد يختص بصاحب البطاقة التي يتم قراءتها عند تمريرها على القارئ. وقد جرى استبدالها لاحقاً برقائق إلكترونية يمكن التحقق من البيانات التفصيلية فيها باستخدام الأرقام السريّة الشخصية لمستخدميها. كما يتم إنتاج البيانات الخاصة بالاستعمال، في كل مرة يجري فيها مسح الرمز الشريطي أو الشريط المغناطيسي أو الرقاقة الإلكترونية، متضمناً ذلك أين ومتى تمّ الاستخدام ولأي غرض. وبالنظر إلى أن الغرض من هذا المسح في كثير من الحالات هو لشراء السلع، فهي توفر وسائل مهمة لإنتاج معلومات عالية الدقة والتفصيل للصفقات المبرمة.

البيانات التفاعلية (Interaction Data):

على الرغم من أن بعض مستخدمي الإنترنت ينظرون إليها باعتبارها وسطاً مجهول المعالم إلى حدّ ما، إلا أن كل تفاعلات المستخدمين الجارية عبر شبكات المعلومات والاتصالات تُنتج بيانات تفاعلية. وحقيقة، فإن إنتاج البيانات هو قائم بطبيعته على العديد من أشكال الاتصال والاستخدام في شبكات تقنية المعلومات والاتصالات. على سبيل المثال، يتعقب مزودو خدمات الإنترنت المواقع التي تتصل بها حسابات المستخدمين الفردية والأجهزة، وتستخدم العديد من المواقع، وخاصة مواقع الشركات، سجلّ التتبع وتعريف الارتباط للمتصفح (كوكيز Cookies) لإنتاج بيانات النقر (اختيارات المستخدم)، لتتبع كيفية تنقل المستخدم خلال الموقع على شبكة الإنترنت وما الميزات التي جرى اختيارها. وبشكل عام، تستخدم مثل هذه البيانات لمراقبة كيفية أداء الموقع الإلكتروني، وإجراء أبحاث عن السوق وتتبع مسار إنتاجية الموظف، مع إمكانية بيع هذه البيانات إلى أطراف أخرى. ويتم إنشاء بيانات النقر هذه بصورة تلقائية ما لم يُعدّ المستخدم تطبيق المتصفح لعدم قبول ملفات تعريف الارتباط (وهذا قد يضعف استخدام الموقع). وبالمثل، يتم تخصيص رموز

تعريفية فريدة في جميع الصفقات المالية القائمة على تكنولوجيا المعلومات والاتصالات، سواء الخدمات المصرفية الشخصية الإلكترونية، أو عمليات سحب أموال من أجهزة الصراف الآلي، أو عمليات المتاجرة في الأسهم، ويجري تخزين بيانات هذه المعاملات تباعاً. وتحتوي جميع رسائل البريد الإلكتروني على رؤوس في مقدمة حزم بياناتها والتي تقوم بتحديد البريد الإلكتروني بشكل فريد، وتحتوي على معلومات تشمل المرسل والمتلقي، والخوادم والأجهزة المستخدمة. وعلى المنوال نفسه، يتم تمييز جميع المكالمات الهاتفية بمعلومات المتلقي للمكاملة، ووقت إجرائها ومدة بقائها، وما إذا كان الهاتف محمولاً والموقع. ونتيجة لذلك يتم إنشاء كميات هائلة من البيانات بشكل روتيني بشأن التفاعلات الحاصلة عبر شبكات تكنولوجيا المعلومات والاتصالات.

البيانات التطوعية (Volunteered Data):

على النقيض من المراقبة الموجهة إلى الأشخاص أو الأشياء من قبل الأفراد والهيئات، أو تلك التي يتم التقاطها تلقائياً باعتبارها ميزة متأصلة لجهاز أو نظام ما، فإن كثيراً من البيانات الكبيرة يتم إنتاجها من قبل الناس بشكل تطوعي فعّال. ففي مثل هذه الحالات، يُنتج الأفراد البيانات ويقومون بإدخالها للعمل على الاستفادة من خدمة (مثل وسائل التواصل الاجتماعي)، أو المشاركة في مشروع جماعي مثل مشروع خريطة الشارع المفتوح (<http://www.openstreetmap.org>) أو موسوعة الإنترنت ويكيبيديا (<https://www.wikipedia.org>). ويُسمى مثل هذا العمل بالإنتاج الاجتماعي (Prosumption) حيث تختلط أمشاط الإنتاج والاستهلاك جزئياً بعضها في بعض، مع اكتساب الأفراد دوراً في إنتاج الخدمة أو المنتج الذي يستهلكونه (Ritzer and Jurgenson 2010). فعلى سبيل المثال، يجري إنتاج واستهلاك محتوى مواقع التواصل الاجتماعي في ذات الوقت من قبل الأفراد المستخدمين عن طريق إدخال التعليقات، تحميل الصور ومقاطع الفيديو، والمشاركة في مناقشة وتبادل المشاعر («الإعجاب» أو «عدم الإعجاب» للشيء). ومن خلال جهودهم الجماعية، يُنشئ الأفراد المنتجات والخدمات التي يشاركون فيها ويضيفون قيمة لها ويستخرجون القيمة منها، بحسب ما أشار إليه Benkler (2006) بمصطلح «ثروة الشبكات». وفي معظم الأحوال، يقوم أفراد

الإنتاج الاجتماعي بهذا العمل الإضافي بمقابل مادي قليل أو دون أي تعويض، وذلك إما للحصول على متعة المهمة، أو الشعور بالسلطة، أو توفيرهم المال والوقت مع انخفاض تكلفة الخدمة والتي غالباً ما تصبح أكثر مرونة بطبيعتها مثل الخدمات المصرفية عبر الإنترنت أو الحجز في رحلة طيران من المنزل. في المقابل، يتلقى مزود الخدمة أو متاجر التجزئة عملاً دون تكلفة، بالإضافة إلى الخبرات والآراء والمعرفة واكتساب الكفاءة ومعلومات ذات قيمة من خلال القرب أكثر من العملاء / الزبائن والقدرة على تسييل بياناتهم (Ritzer and Jurgenson 2010; Dodge and Kitchin 2013). فبعض أشكال الإنتاج الاجتماعي هي أقرب إلى الخدمة الذاتية ومقيّدة بطبيعتها، مع قيام أفرادها بالأعمال التي كانت مسندة إلى موظفين، في حين أن البعض الآخر أكثر انفتاحاً وتفاعلية. وهنا، سيتم مناقشة خمسة أشكال للبيانات التطوعية باختصار: المعاملات (Transactions)، ووسائل التواصل الاجتماعي (Social Media)، والمراقبة الحية (Sousveillance)، والتعهد الجماعي (Crowdsourcing)، وعلم المواطن (Citizen Science).

المعاملات (Transactions):

يتطلب الشراء عن طريق الإنترنت من المشتري إدخال البيانات الشخصية له فيما يخص عملية الدفع والعنوان وغيرها من البيانات الوصفية الضرورية الأخرى باعتبارها جزئياً عملية للتحقق وتبادل المعلومات ولكن أيضاً لمعرفة المزيد عن المستهلك. وعادة ما تكون التفاعلات في مثل هذه التبادلات محدودة إلى حد ما، وتقتصر على اختيار ماهية بنود الشراء، والعديد من حقول البيانات الإلزامية لوضع اللمسات الأخيرة وإتمام عملية الشراء. وبعد تبادل المعلومات، من الممكن طلب المزيد من المعلومات من المستخدمين، إما لتقييم أداء الموقع وخدمة العملاء، أو لتصنيف ومراجعة بيانات البند الذي تم شراؤه. وتضيف بيانات المراجعة هذه قيمة للموقع من خلال اقتراح كيفية تحسين التصميم وكذلك تزود المستهلكين الآخرين بردود الأفعال (Feedbacks) عن مدى جودة وكفاءة المنتجات. فعلى مواقع إلكترونية مثل (<http://www.tripadvisor.com>)، يمكن لأفراد الإنتاج المجتمعي تقييم واستعراض الفنادق وخدمات السفر الأخرى، مثل البيانات التي يجري تقديمها طوعاً على الموقع عن حركة المرور، وبث الإعلانات والإشارة إلى مواطن الربح المحتملة، مما يكون

له تأثير ملحوظ في اختيارات المسافرين الآخرين. كما يوفر الموقع بيانات مفيدة حول الشخص الذي تطوَّع لإجراء الاستعراض والتقييم، مثل خيارات أسلوب الحياة والإنفاق خلال السفر التي يمكن تسهيلها مالياً عن طريق بيعها إلى أطراف ثالثة. وبالمثل، تتطلب تعبئة النماذج الإلكترونية الحكومية من المواطن تقديم معلومات للموقع بشكل تطوعي. وهنا، يكون هنالك تبادل للبيانات يتعدى حدود عملية الدفع للمبالغ النقدية، فبعض البيانات الإضافية الأخرى يمكن أن تكون مرتبطة بالمعاملات، مثل بيانات مسار النقر على الصفحات ومعلومات بطاقة الولاء، مما يتيح بناء ملف مبدئي للشخص يمكن البناء عليه مع مرور الزمن أو يتم ربطه مع مصادر البيانات الأخرى مثل بيانات التوزيع الجغرافي للسكان.

التواصل الاجتماعي (Social Media):

قبل العام ٢٠٠٣م، كانت شبكة الإنترنت إلى حدٍ كبير وسطاً للبتّ الإعلامي يمكن من خلالها للناس تقديم المعلومات أو شراء البضائع. وكانت هناك لوحات الإعلانات والقوائم البريدية التي يستطيع الناس من خلالها التواصل، ولكنها كانت قائمة على استخدام النصوص فقط، ومحددة بطريقة عرض بدلاً من كونها وسيلة مفتوحة وعامة. وبينما يمكن للأفراد إنتاج أشكال أخرى من المحتوى مثل صفحات الويب، إلا أن ذلك يتطلب المهارة لإنشاء مثل هذه الصفحات وشراء اسم النطاق، وكانت هذه الصفحات ثابتة إلى حدٍ كبير وينبغي تحريرها بشكل رئيسي على الطريقة اليدوية ليجري تحديثها. واعتباراً من العام ٢٠٠٤م، وعلى الرغم من ذلك، بدأت طبيعة شبكة الإنترنت بالتغيّر لتصبح أكثر تفاعلية وتشاركية ودينامية، وقد تمّ الإشارة إلى هذا التحول على أنه الانتقال من إصدار شبكة الويب الأول (Web 1.0) إلى الإصدار الثاني (Web 2.0) (O'Reilly 2005).

وعوضاً عن محتوى الويب الذي يجري تقديمه إلى حدٍ كبير من قبل المختصين ومصممي المواقع الإلكترونية، فقد أصبح لأي شخص إمكانية إنشاء محتوى يمكن تحديثه بشكل مستمر. وعلاوة على ذلك، يمكن الوصول إلى هذا المحتوى والتفاعل معه من خلال مجموعة متنوعة من المنصات مثل الهواتف الذكية والأجهزة اللوحية. وبدلاً من مجرد تصفح المعلومات أو تقييدها في مناقشات ضيقة نسبياً، أصبح من الممكن للناس المشاركة بسهولة وبشكل فعّال في إنتاج عالم الإنترنت وتصميم وتحرير المحتوى، والإسهام بحرية في

الخدمات التي تتبادل المواد مع الآخرين. وفي حين تم اعتبار الإصدار الأول لشبكة الويب (Web 1.0) وسطاً للقراءة فقط، فإن الإصدار الثاني لشبكة الويب (Web 2.0) هو وسط للقراءة والكتابة، يمكن للناس من خلاله إضافة قيمة للمواقع عندما يقومون باستخدامها. وتُميل مواقع الإصدار الثاني من الويب وخدماتها إلى التركيز على النشر من العديد - إلى - العديد، والتواصل الاجتماعي، وتوفير معلومات أكثر ثراءً واتساقاً. وتشتمل الأمثلة على خدمات الشبكات الاجتماعية على سبيل المثال الفيسبوك وتويتر؛ والصور ومواقع مشاركة الفيديو على سبيل المثال، فليكر ويوتيوب؛ والمدونات؛ وعمليات المزج باستخدام واجهات برمجة التطبيقات المفتوحة (APIs) لدمج البيانات من مصادر مختلفة لإنشاء تطبيقات جديدة على سبيل المثال، فورسكوير، وييلب (Yelp). وتعتمد هذه المواقع كلها على المشاركة الفعالة من جمهور متحمس لتبادل المعلومات حول حياتهم والقيام بأعمال من قبيل الكتابة، التحرير، التطبيق، التعديل، النشر، المشاركة، الوسم، التواصل، ... إلخ (Beer and Burrows 2007). وعند القيام بذلك، فهم لا يتطوعون بالعمل فقط، ولكن بالكثير من البيانات عن أنفسهم، متضمناً ذلك الملفات التعريفية، الصور، المواقع، الآراء، التفضيلات، القيم، وشبكة اتصالاتهم الاجتماعية. ولأنه قد جرى تملك هذه المواقع من قبل مجموعة شركات تأسست حديثاً، فإنه يجري تداول البيانات في أيدي مملوكة للقطاع الخاص والتي تسعى بدورها لإنتاج نماذج جديدة من تراكم رأس المال عن طريق استخراج القيمة منها (انظر الفصلين الثاني والسابع).

المراقبة الذاتية (Sousveillance):

المراقبة الحية هي الرصد الذاتي وإدارة صحة الفرد الشخصية وحياته من خلال التقنيات الرقمية الحميمة على سبيل المثال، معدات اللياقة البدنية، والحوسبة القابلة للارتداء لتسجيل البيانات المتعلقة بالفرد (Mann et al. 2003). وعلى النقيض من المراقبة الإشرافية (Surveillance)، والتي يتم فيها مراقبة الفرد من موقع خارجي بواسطة شخص آخر، يجري توظيف وضبط المراقبة الذاتية عن وعي وإدراك الفرد لتحقيق الذات، وتوفير منظور داخلي أولي عن حياته. وعلى مدى العقد الماضي، تطورت حركة المراقبة الذاتية من أناس يراقبون ويسجلون بنشاط بياناتهم الشخصية والتي تُعرف أيضاً باسم حركة القياس

الكمّي للذات أو التتبع الذاتي كمياً (Quantified Self). وبشكل عام، يراقب الأفراد جوانب الصحة واللياقة البدنية، ويجمعون بيانات الاستهلاك (على سبيل المثال، المواد الغذائية / السعرات الحرارية)، والحالات الجسدية (مثل ضغط الدم والنبض)، والحالات العاطفية (على سبيل المثال، المزاج ومستويات الإثارة) والأداء (على سبيل المثال، أميال المشي / الركض / الدوران المقطوعة، وعدد ساعات النوم، وأنواع النوم)، بالتعاون مع عدد من الشركات التي تقدّم التقنيات والخدمات المرتبطة بالمراقبة الذاتية. وتشمل شركات مراقبة الأنشطة الرياضية والنوم، على سبيل المثال، شركات من أمثال فت بت، ونايك بلس، فيول باند، وجوبون (Fitbit, Nike+, Fuelband, and Jawbone) والتي تعمل على مزامنة البيانات الناتجة مع تطبيقات متابعة وتعقب تسمح للمستخدم برصد وتحليل بياناته الشخصية. ولا يجري بالضرورة مشاركة البيانات التي تمّ إنتاجها على نطاق أوسع، على الرغم من إمكانية حدوث ذلك، ولكن يجري مشاركتها مع الشركات المزوّدة للخدمة، مما يوفر لهم اطلاعاً نافذاً على البيانات الشخصية.

وتعدّ هذه التقنيات في المرحلة الأولى من التطوير وهناك رؤى أكثر شمولاً لتسجيل بيانات الحياة (Life Logging) التي تعمل على إنشاء سجل رقمي موحد لخبرات الأفراد، من خلال جمع متعدد الطرق عبر أجهزة استشعار رقمية، وتخزينها بشكل دائم كأرشيف شخصي متعدد الوسائط (Mann et al. 2003)، وذلك مع عدد من نماذج الأبحاث التي يجري تطويرها حالياً. وتهدف سجلات بيانات الحياة إلى إنشاء سجل متصل، وقابل للبحث والتحليل لماضي الشخص الذي يشمل كل عمل، أو كل حدث، أو كل محادثة، أو كل موقع تمّت زيارته، أو أي تعبير مادي متصل بحياة الفرد، فضلاً عن الظروف الفيزيولوجية داخل الجسم والأوضاع الخارجية المحيطة به (على سبيل المثال، الوُجهة، ودرجة الحرارة، ومستويات التلوّث) (Dodge and Kitchin 2007b) - "أي مجمل المعلومات التي تتدفق خلال حياة الإنسان" (Johnson 2003: 85). ومن الواضح أن إنتاج هذه السجلات عن الحياة يثير عدداً من الأسئلة بشأن الخصوصية، وملكية البيانات المنتجة، وطريقة استخدام هذه البيانات (Dodge and Kitchin 2007b).

التعهد الجماعي (Crowdsourcing):

التعهد الجماعي هو الإنتاج المشترك لوسائط الإعلام، والأفكار، والبيانات من خلال العمل التطوعي من قبل العديد من الناس لحل مهمة معينة. وبينما يمكن القول بأن محتوى وسائل التواصل الاجتماعي هي نتاج تعهد جماعي بمعنى أنه يتم الحصول عليه من قبل عدد كبير من الناس، إلا أن الغرض من هذه المحتوى موزع وواسع الانتشار ويفتقر إلى التركيز. فبدلاً من ذلك، يركز التعهد الجماعي على الإنتاج التعاوني للمعلومات وإنشاء حلول لقضايا معينة من خلال الاعتماد على الطاقة، والمعارف، والمهارات، والعمل التوافقي والجماعي لحشد من الناس (Howe 2008). وقد بين Howe (2008) أن هناك أربعة تطورات تدعم نمو التعهد الجماعي: نهضة الهواة (بحسب المعايير المهنية في الغالب)، وظهور حركة البرمجيات مفتوحة المصدر، وزيادة وفرة أدوات الإنتاج خارج إطار الشركات المصنعة، وصعود مجتمعات محلية حيوية ومنظمة وفقاً لاهتمامات الناس وميولهم على شبكة الإنترنت، فهذه التطورات مجتمعة مكّنت من تجميع موزع وديمقراطي للأفكار والأدوات والمواد. وهناك ثلاثة أنواع رئيسية من التعهد الجماعي التي تعمل في الوقت الحاضر: تلك التي تنتج حلاً بشكل جماعي على سبيل المثال، خريطة الشارع المفتوح (<http://www.openstreetmap.org>)، ويكيبيديا (<https://www.wikipedia.org>)؛ وتلك التي تستخدم الجمهور لتقييم حلول أو منتجات متعددة على سبيل المثال، موقع شركة أمازون (<https://www.amazon.com>)، ومستشار الرحلات (<https://www.tripadvisor.com>)؛ وتلك التي تسعى إلى حل واحد أو أكثر من الجمهور على سبيل المثال، إنوسنتف (<https://www.innocentive.com>)، والممتلكات المكثفة لمطوري البرمجيات (Hackathons).

وبخصوص النوع الأول، أصبح إنتاج المعلومات (خريطة مفتوحة المصدر، موسوعة قابلة للتعديل والوصول إليها بحرية) تعاونياً، وعملاً من إنتاج الأفراد، مع إعطاء المشاركين الحقوق والتسهيلات لتعديل وتنقيح وإسهاب الشرح لمساهمات الآخرين (Benkler 2006). فمن خلال العمل عبر شبكة الإنترنت يمكن لأي أحد له قدرة النفاذ إليها من أن يسهم ويضيف إلى مجمّع الخبرات المحتملة ونطاق المشاريع مقابل تكلفة أقل، وضمان ممارسة مختلف وجهات النظر والتقنيات على المشروع (Weinberger 2011). وفي الواقع،

فإن إنتاج خريطة مفصلة للعالم أو موسوعة شاملة مع عدد قليل من المهنيين هي مهمة صعبة ومكلفة، وببساطة شاقة للغاية بالنسبة لمجموعات صغيرة من المتطوعين. ولكن بالنظر إلى آلاف المشاركين، يمكن تجزئة المهمة عليهم وحلّها. وسيكون المقابل الذي سيجنه المشاركون هو إنتاج بيانات مفتوحة المصدر خالية من حقوق الطبع والنشر والرخص المكلفة (كما هو الحال مع بيانات الوكالة الوطنية لرسم الخرائط) (Dodge and Kitchin 2013). وبالمثل، يمكن لشركة أمازون وغيرها من شركات التجزئة على الإنترنت، من خلال تمكين المستخدمين من استعراض المنتجات التي يشترونها، استخدام قوة الجماهير لمراجعة ومشاركة آرائهم حول هذه المنتجات. كما أن الفائدة التي تعود على المشاركين هي تبادل الرأي فيما يتعلق بالمزايا النسبية للسلع والخدمات.

ويُعد استخدام المسابقات لتحفيز الجماهير على اقتراح حلول مقابل جائزة من أشكال التعهيد الجماعي الخاصة والذي تستخدمه بعض الشركات لتقديم الخبرات والأفكار (Weinberger 2011). فعوضاً عن اللجوء للمناقصات والاستعانة بمجموعة محدودة من الخبراء المحتملين، تفتح شركات مثل إنوسنتيف (Innocentive) وتوب كودر (TopCoder) المشكلة من خلال الإعلان عنها لأي شخص مهتم، وتقدّم في العادة من عشرة آلاف إلى مئة ألف دولار، وأحياناً أكثر من ذلك بكثير، لإيجاد حلول لمشاكل العملاء. وهنا، يكون الجمهور هو مصدر الحل، ولكن لا يعمل الأفراد بعضهم مع بعض بشكل جماعي، وتكون الفائدة بالنسبة للشركات هي خيارات الحل المقدمة وللمشاركين هي العائد المالي إذا تمّ اختيار الحل الخاص بأحدهم. وعلى نحو مماثل، تسعى مسابقات بيانات التطبيقات المفتوحة التي تعقدّها المجالس البلدية إلى تقديم مكافآت أكثر تواضعاً مقابل تطبيقات أكثر فائدة وإثارة للاهتمام. ويمكن لهذه المسابقات أن تأخذ شكل الملتقيات المكثفة لمطوري البرمجيات (Hackathons) التي يعمل الناس فيها معاً في فرق لتطوير حلول تطبيقات، وعادة ما تمتد على مدى يوم أو خلال عطلة نهاية الأسبوع، وتتمثل الفائدة العائدة على البلدية في استخدام بياناتها من أجل الصالح العام، وشهرة المواطنين، والمشاركة المدنية.

علم المواطنة (Citizen Science):

علم المواطنة هو شكل من أشكال التعهيد الجماعي حيث تعمل مجتمعات أو شبكات من المواطنين بصفة مراقبين في بعض مجالات العلوم (Goodchild 2007: 218). هنا، يقوم الناس بإنتاج وإعداد ومعالجة الملاحظات العملية التجريبية، والقياسات التفصيلية للظواهر بالمجان، والتي تُعد، حاسمة ومفصلية، وذات قيمة حقيقية كنقاط بيانات للعلوم الصحيحة. ويكون علماء المواطنة ذوي فائدة بشكل خاص من حيث الرصد لأنهم قد يكونون موزعين جغرافياً ومُدرجين ضمناً في المكان عبر الزمن حيث من الصعب على مستويي التنسيق والتكلفة استخدام مساعدي البحوث المهنيين لجمع القياسات الميدانية. ومن الأمثلة على ذلك، قياسات هواة الرصد الجوي الناتجة من محطات الحدائق الخلفية للأرصاد الجوية، وهواة علم الفلك الذين يراقبون السماء ليلاً بدقة. وقد أصبح علم المواطنة راسخاً خلال القرن العشرين، وذلك دعماً لرواج العلم وتعميمه، ولكن لا يزال النشاط ضئيلاً ويعتمد على درجة معقولة من المهارات والمعارف والمعدات المتخصصة مثل مراقب ذي نوعية جيدة، وفوق ذلك قدر كبير من التحفيز والالتزام المستمرين.

وقد صنّف (Bonney et al 2009; cited in Miller-Rushing et al. 2012). أشكال علم المواطنة في ثلاث فئات:

- التساهمية (Contributory): وتصمم بشكل عام من قبل العلماء ولأجلها يسهم الأفراد عامة بالبيانات في المقام الأول.
- التعاونية (Collaborative): وتصمم بشكل عام من قبل العلماء ولأجلها يسهم الأفراد عامة بالبيانات، ولكن قد يساعدون أيضاً على تحسين تصميم المشروع، وتحليل البيانات، ونشر النتائج.
- الإنشاء المشترك (Co-created): وتصمم من قبل العلماء وأعضاء من عامة الجمهور يعملون معاً ولأجلها يتم إشراك بعض عامة المشاركين على الأقل في معظم أو جميع خطوات العملية العلمية؛ ويشمل ذلك أيضاً البحوث التي وضع تصورهما وتم تنفيذها من قبل علماء هواة (غير مهنيين) بشكل كامل.

وفي العادة، يختلف مستوى الخبرة ونوع المشاركة بحسب الفئة، مع مشاركة المواطنين العلماء فقط من ذوي المهارات الأكثر تقدماً في مشاريع الإنشاء المشترك. ويمكن أن تتخذ المشاركة عدداً من الأشكال، بما في ذلك:

- الحوسبة الموزعة: حيث يضيف أحد المشاركين قوة معالجة حاسوبية من أجهزة الحاسب الشخصية الخاصة به إلى المشروع، مما يسمح بمعالجة البيانات باستخدام الموارد الحاسوبية منها (التساهمية).
- النسخ / التحويل / الرقمنة: تحويل شكل من أشكال البيانات إلى آخر، على سبيل المثال، تحويل وثائق مكتوبة بخط اليد إلى نص رقمي (التساهمية).
- قياسات الرصد: ينتج العلماء البيانات، ولكن يُستخدم علم المواطنة لإجراء القياسات باستخدام أدوات محددة يجري توريدها لهم (التساهمية).
- جمع البيانات: يجمع أفراد المواطنين البيانات ومن ثمّ يقومون بمشاركتها مع العلماء (التساهمية).
- تحليل الرصد: يقوم أفراد المواطنين بإجراء بعض التحليل، بما في ذلك توفير بعض التفسيرات لما قد تعنيه النتائج، بدلاً من مجرد الاكتفاء بإنتاج أو قياس البيانات (التعاونية).
- تصميم البحث وتشغيله: يقوم أفراد المواطنين بدور نشط في تصميم وتشغيل المشروع، فضلاً عن نشر النتائج والاستنتاجات (الإنشاء المشترك).

(Open Scientist 2013)

الخلاصة:

تطرق هذا الفصل إلى العوامل المساعدة للبيانات الكبيرة ومصادرها. وتمّ بيان أن إنتاج البيانات الكبيرة قد تيسّر بفضل التقاء خمسة ابتكارات تقنية من بداية الألفية الجديدة فصاعداً وهي تنامي قوة المعالجة الحاسوبية، وكثافة التشبيك الحاسوبي، والحوسبة المنتشرة والحوسبة في كل مكان، والتعرّف وتحديد الهوية من خلال زيادة قابلية الفرز والقراءة آلياً، وأخيراً التخزين الموزع الواسع النطاق. وقد أدّت هذه التطورات مجتمعة إلى مجموعة

متنوعة من النظم الاجتماعية التقنية التي تنتج البيانات الكبيرة. وقد قدّم هذا الفصل مسار تتبّع أولي لهذه الأنظمة، حيث قسّمها إلى ثلاث فئات واسعة النقاط من أشكال إنتاج البيانات: الموجهة، والآلية، والتطوعية. وينبغي النظر إلى مثل هذا المسار باعتباره خطوة أولية في محاولة وضع خريطة لمشهد يتكشف باستمرار. وما هو مطلوب حالياً يتمثل في مجموعة مفصلة من الدراسات التجريبية التي تدرس بشكل أكثر عمقاً الطرق المختلفة التي يتم بها حالياً إنشاء ومعالجة واستخدام البيانات الكبيرة. كما تحتاج مثل هذه الدراسات إلى رسم أصول التقاء هذه التقنيات الممكنة وفك مجاميع البيانات التي تمّ أنشائها وكيف تطبق عملياً لإنتاج أشكال معينة من البيانات الكبيرة (انظر الفصلين 1 و11). وهذا يشمل تفكيك عناصر النظام المرتبط بها (انظر الفصل 7). وعند القيام بذلك، يمكن الحصول على أفكار ورؤى أكثر عمقاً لكل نظام اجتماعي تقني ومختلف العمليات والضوابط والسياسات المتصلة به، والعمل المتنوع الذي يقوم به في هذا العالم.

كما ينبغي لمجموعة مصاحبة من الدراسات أن تختبر ما قد يحدث عند ربط النظم الاجتماعية التقنية معاً لإنشاء أنظمة جديدة أكبر والتي من خلالها تنتج الآثار الهامة لتضخم البيانات. كما أن التقنيات التي تقوم عليها البيانات الكبيرة، والبيانات نفسها من خلال علاقتها، هي أيضاً جاهزة لمثل هذه التوسعات. في الواقع، تكون مجاميع اجتماعية تقنية (Socio-Technical) معقدة، مثل سلسلة متاجر التجزئة الكبيرة كـ Walmart)، تقوم وعلى نحو متزايد، بربط العديد من الأنظمة مثل إدارة سلسلة التوريد (Supply Chain Management)، تخطيط موارد المؤسسات (Enterprise Resource Planning)، إدارة علاقات العملاء (Customer Relationship Management)، إدارة المخزون (Store Management)، وأنظمة الأمن (Security Systems) والتي تُنتج أشكالاً من البيانات الموجهة، والآلية، والتطوعية لإنشاء منظمات معقدة ومتطورة موجهة بالبيانات والتي تترتب عليها آثار متنوعة على الموردين والموظفين والمستهلكين. وبالمثل، تحاول مراكز السيطرة والتشغيل للمدن الذكية دمج أنواع مختلفة من البيانات التي تمّ جمعها من مصادر مختلفة في مركز واحد حيث يتم الجمع بينها وتحليلها بشكل آني أو لحظي (Real-Time)، مما يغيّر بشكل جذري الطريقة التي تُدار وتُحكم بها التجمعات المدنية (انظر الفصل السابع).

وقد أثار مثل هذا التوسع وإعادة الضبط العديد من الأسئلة الاجتماعية والسياسية والأخلاقية والمعياريّة المتعلقة بأنواع الأنظمة والأماكن التي نريد العيش داخلها، وقد بدأنا بالكاد أن نسأل ونجيب عن هذه الأسئلة (انظر الفصل العاشر).

الفصل السادس

آليات تحليل البيانات (Data Analytics)

ليست البيانات بحدّ ذاتها مفيدة، كما أنها ليست ذات فائدة إلا إذا أمكن استخلاص معنى وقيمة منها. وبعبارة أخرى، فإن ما يمكن فعله حيال البيانات هو أمر أهم من مجرد عملية إنتاجها؛ إذ تركز جميع العلوم على تحقيق معنى وقيمة من البيانات. إن إيجاد واستشعار المعنى من البيانات الصغيرة الموسّعة والبيانات الكبيرة ينطوي على تحديات جديدة. ففي حالة البيانات الصغيرة الموسّعة، يكمن التحدي في ربط قواعد البيانات المتنوعة للحصول على رؤى جديدة، وانفتاح البيانات لتقبّل المناهج التحليلية الجديدة التي يجري استخدامها مع البيانات الكبيرة. أما بالنسبة للبيانات الكبيرة، فيتمثل التحدي في التكيّف مع الوفرة والشمولية (متضمناً ذلك الكميات الكبيرة من البيانات ذات الفائدة والقيمة المنخفضة)، والمناسبة الزمانية والديناميكية، والفضوى وعدم اليقين، والعلائقية العالية، والطبيعة شبه المنظمة وغير المنظمة، وحقيقة أن يتم إنتاج الكثير من البيانات الكبيرة دون أي سؤال محدد في العقل لتجيب عنه أو أنها في صفة مُنتج هو نتيجة ثانوية لنشاط آخر. فالواقع، وحتى وقت قريب، تمّ تصميم تقنيات تحليل البيانات بشكل أساسي لاستخلاص رؤى من مجموعات بيانات شحيحة، وثابتة، ونظيفة، وضعيفة العلائقية، وجرى جمعها بطرق علمية وملتزمة بافتراضات صارمة مثل الاستقلالية، والثبات، والعيارية، كما جرى إنتاجها وتحليلها مع وضع سؤال محدد في الاعتبار (H.J. Miller 2010).

حتى وقت قريب كان تطوير أدوات للربط بين مجموعات البيانات المختلفة وتحليل البيانات الكبيرة ضعيفاً؛ بسبب التعقيد المرتبط بتطوير هذه الأدوات من الناحية الحاسوبية. ولم يجرِ تحليل مجموعات البيانات الضخمة بشكل منتظم سوى في السنوات الأربعين الماضية أو نحو ذلك. وحتى ذلك الحين كانت هذه التحليلات قائمة على مبادرات خاصة قادرة على تحمّل الموارد اللازمة. وبدون طاقة حاسوبية عالية واسعة الانتشار ويمكن النفاذ إليها، سيكون إيجاد واستشعار المعنى من طوفان البيانات الهائل مكلفاً

للاغاية أو صعب التطبيق ويستغرق وقتاً طويلاً. وبالطبع، هناك نوعٌ ما من الجدل الدائر في الواقع هنا؛ إذ دون الحوسبة الواسعة الانتشار في كل مكان لم يكن إنتاج البيانات الكبيرة ليكون في المقام الأول. ومع ذلك، كما ذكر (Hastie et al. 2009: xi)، فإنه بالنظر إلى «ظهور الحواسيب وعصر المعلومات، فقد تفاقمت المشاكل الإحصائية من حيث الحجم والتعقيد».

وتمثل الحل لتحديات معالجة وتحليل البيانات الصغيرة الموسّعة والبيانات الكبيرة مجموعة جديدة من آليات إدارة المعلومات والتخزين (انظر الفصل الخامس) وتطوير أدوات تحليل البيانات الكبيرة. فمنذ بداية الحوسبة استمرت هذه الآليات والأدوات التحليلية الجديدة في التطور، وهي تقوم على اختبارات إحصائية، ونماذج، وأساليب تمثيل صوري رصينة متّبعة، كما أنها تعمل على إنشاء أساليب وطرق جديدة تضرب جذورها في البحوث المتعلقة بالذكاء الاصطناعي والأنظمة الخبيرة التي هدفت إلى إنتاج تعليم الآلة بحيث يمكنها حاسوبياً وبصورة تلقائية التنقيب، والكشف عن الأنماط، وبناء النماذج التنبؤية. فمثل هذه الأدوات التحليلية أضحت ملائمة وبشكل مثالي في معالجة واستخلاص المعلومات من مجموعات بيانات ضخمة ومتصلة ببعضها، فقد أصبحت تشغل حيزاً كبيراً من الاستثمار في البحوث من أجل توسيع وإنتاج طرق معالجة جديدة للبيانات والخوارزميات الإحصائية والنمذجة وتقنيات التمثيل الصورية (المؤسسة الوطنية للعلوم (National Science Foundation 2012). وتسعى هذه الأدوات التحليلية، كما هي مطبّقة في مجال الأعمال التجارية والعلوم، للإجابة عن أربع مجموعات أساسية من الأسئلة (Minelli et al. 2013):

- الوصف (Description): ماذا ومتى حدث ذلك الشيء؟ وكم مرة يحدث؟
- التفسير (Explanation): لماذا حدث ذلك؟ وما هو تأثيره؟
- التنبؤ (Prediction): ماذا يمكن أن يحدث بعد ذلك؟ وماذا لو فعلنا هذا أو ذاك؟
- العلاج (Prescription): ما الحل أو النتيجة المُثلَى؟ وكيف يمكن تحقيق ذلك؟

وتُستمد الإجابة عن هذه الأسئلة من أربع فئات رئيسة من آليات التحليل: التنقيب عن البيانات والتعرّف على الأنماط، التمثيل الصوري للبيانات وآليات التحليل البصرية، التحليل الإحصائي، والتنبؤ والمحاكاة والتحسين. وسيتم مناقشة كل من هذه الفئات باختصار، ولكن ينبغي تقديم مرحلة ما قبل التحليل وتعلّم الآلة أولاً، لأنها مركزية لجميع الفئات الأربعة.

ما يسبق التحليلات (Pre-analytics):

تحتاج كل تحليلات البيانات إلى أن تكون البيانات المطلوب تحليلها معدّة سلفاً للتحليل؛ بمعنى أنه قد جرى تجهيزها والتحقق منها، وقد بيّن H.J. Miller (2010) and Han et al. (2011) أربع عمليات فيما يتعلّق بالبيانات الموسّعة والكبيرة والتي تطبّق عادة بالتسلسل، على الرغم من أنه لا يلزم تنفيذها بأي ترتيب معين، إضافة إلى إمكانية تكرار تنفيذها:

- اختيار البيانات (Data selection): تحديد مجموعة جزئية من المتغيرات الأكثر قيمة، وإطار العينة المحتمل لهذه المتغيرات. حيث لن تكون كل المتغيرات الناتجة ذات أهمية، ولا بالضرورة كل نقاط البيانات، فإدراج مثل هذه البيانات غير الضرورية في نماذج قد يولّد التكرار وعدم الكفاءة.
- المعالجة المسبقة للبيانات (Data pre-processing): تنظيف البيانات المختارة لإزالة الضوضاء أو الأخطاء أو التحيّز، أو التعامل مع الحقول المفقودة أو التناقضات وعدم السلامة، وهيكلية البيانات لإدخالها في عملية التحليل.
- اختزال البيانات وإسقاطها (Data reduction and projection): التقليل من أبعاد البيانات من خلال التحويلات (على سبيل المثال، التسوية والصقل (Smoothing)، وبناء السمات والخصائص، والتجميع، والتطبيع، والتسلسل الهرمي للمفهوم، والأساليب الإحصائية مثل تحليل مستوى الانحدار وتحليل المكونات الرئيسية) لمكافحة تمثيل البيانات ولكن بصورة أكثر كفاءة.
- إثراء البيانات (Data enrichment): دمج البيانات المختارة مع بيانات أخرى (مثل بيانات التعداد السكاني وبيانات السوق) من أجل الوصول إلى رؤى يمكن الاستفادة منها بشكل أكبر.

فمن جهة، صُممت كل خطوة من هذه الخطوات من أجل زيادة نوعية البيانات المستخدمة في التحليل، ومن جهة أخرى، باستثناء إثراء البيانات، من أجل تقليل حجم البيانات التي على التحليل التعامل معها. فالأول تمّ تصميمه لتعزيز صحة وصلاحيّة الاستنتاجات المستخلصة بالنظر إلى الفوضى الكامنة وعدم اليقين في البيانات الكبيرة (انظر الفصل التاسع)، في حين يهدف الثاني إلى زيادة الكفاءة التشغيلية مع الأخذ بعين الاعتبار أحجام البيانات وسرعاتها والأحمال الحاسوبية. ومن الممكن أن يكون من الصعب القيام بهذه المهام عند إجراء التحليلات بشكل آني أو لحظي، وخاصة فيما يتعلق بتنظيف البيانات. ونتيجة لذلك، وفيما عدا التمثيل الصوري للمؤشرات، لا يجري تحليل البيانات الكبيرة آنيًا أو بشكل لحظي، ولكن يجري التحليل على سلاسل تمثل عينات كبيرة جدًا لبيانات نظيفة، ومختزلة، ومخصّبة (ثرية) وخلال أزمان مختلفة. أما إذا كان المطلوب تحليلات بشكل آني أو لحظي، فتطبّق في العادة التحليلات المسبقة للبيانات مقدّمًا على عينة من أجل تحديد طبيعة البيانات وكيفية التعامل معها للانتقاء، والاختزال، والتنظيف على عجل.

ويعد إثراء البيانات مهمة ذات قيمة عالية لأنها تنتج أثرًا مضخمًا (Crampton et al. 2012) يُمكن من اكتساب رؤى لا يمكن الحصول عليها من خلال قاعدة بيانات واحدة. وبحسب من تمّت مناقشته في الفصل الثاني، فإن المبرر الرئيسي لإنشاء البنى التحتية للبيانات تمثل في إعادة استخدام البيانات. وإن التحدي يتمثل في إنشاء طرق لربط البيانات التي أنتجت لأغراض مختلفة بعضها ببعض، والتي قد تختلف أيضًا في البيانات الوصفية المصاحبة لها، ومعايير البيانات، ووحدات القياس، والإجراءات، والفئات، والنطاق، والتزامن، وتنسيقات الملفات، أو إنشاء طرق للربط لتقليل حدوث المغالطات المحتملة الناتجة عن الاستدلال (Creation of Potential Ecological Fallacies). وهذه ليست مهمة يسيرة، ولكن تعمل الأساليب الحاسوبية على جعلها أسهل من خلال استخدام الخوارزميات التي يمكنها البحث، المطابقة، الدمج، إعادة الجمع من خلال أنواع مختلفة من التحويلات، وإعادة تهيئة البيانات. وبالنتيجة، يمكن التنقيب والتحليل في مجموعة بيانات جديدة باستخدام الفئات الأربع الرئيسية للتحليلات التي سيجري نقاشها أدناه.

ويمكن أن تكون أعمال التحليلات المسبقة مملّة باستغراقها وقتاً طويلاً للغاية، ولكنها مع ذلك مهمة جداً ولا يمكن تجاهلها. وبالنظر إلى الزيادة المنقطعة النظير في مختلف أنواع البنى التحتية الجديدة للبيانات والبيانات الكبيرة، فقد أصبحت منطقة ثرية للبحث حيث يسعى علماء البيانات إلى إيجاد طرق أكثر إنتاجية وكفاءة وفعالية لإتمام، وخصوصاً أتمته، هذا العمل.

تعليم الآلة أو التعلّم الآلي (Machine learning):

يمكن إجراء تحليل الأعداد الكبيرة جداً من سجلات البيانات في إطار زمني مناسب عن طريق الخوارزميات الحاسوبية. وفي حين أن الكثير من عمليات تحليل البيانات الكبيرة يمكن إجراؤها بالطرق نفسها المتبعة في تحليل البيانات الصغيرة، حيث يتخذ المحللون القرارات المتعلقة بالكيفية التي يتم فيها الاستكشاف من خلال هذه الخوارزميات، فإن الهدف من أغلب البحوث هو تطوير عمليات مؤتمتة يمكنها تقييم البيانات والتعلم منها ومن تحليلها. ويطلق على هذه العمليات المؤتمتة مصطلح تعلّم الآلة أو التعلّم الآلي وتشكّل أحد فروع الذكاء الاصطناعي. ويهدف تعلّم الآلة إلى التطوير المتكرر لفهم مجموعة البيانات؛ والتعلم التلقائي لإدراك أنماط معقّدة وبناء نماذج توضح وتتوقّع مثل هذه الأنماط؛ وتحسين النتائج (Han et al. 2011).

ويتكوّن تعلّم الآلة بشكل عام من نوعين رئيسيين: مُراقب (باستخدام بيانات التدريب) وغير مُراقب (باستخدام التنظيم الذاتي). في التعلّم المُراقب، يتمّ تدريب نموذج لمطابقة المدخلات مع مخرجات معلومة. على سبيل المثال، يمكن تدريب النموذج لمطابقة الرموز البريدية المكتوبة بخط اليد مع مثيلاتها المطبوعة آلياً، أو التنبؤ بنتائج معينة، فهي «مراقبة» بمعنى أن بيانات التدريب موجودة لتوجيه عملية التعلّم (Hastie et al. 2009). في المقابل، تسعى نماذج التعلّم غير الخاضعة للرقابة إلى تعليم نفسها على اكتشاف الأنماط وإيجاد هياكل البيانات دون استخدام بيانات التدريب. وبشكل عام، يمكن تحقيق ذلك من خلال تحديد مجاميع وعلاقات بين البيانات التي لم يكن معروفاً بينها مسبقاً خصائص للتشابه أو الارتباط. فعلى سبيل المثال، يُمكن أن يتعلم النموذج على كيفية تقسيم العملاء إلى مجموعات ذاتية متماثلة والتنبؤ بالمشترى لتلك المجموعات (Han et al. 2011). في كلتا الحالتين، يتم

إنشاء نموذج من خلال عملية التعلم التي شكلتها القواعد والأوزان التي توجه كيفية بناء نموذج فيما يتعلق بالبيانات (Hastie et al. 2009). وتبدأ عملية بناء نموذج من بناء بسيط ثم يتم التعديل مراراً وتكراراً على ذلك البناء باستخدام قواعد التعلم، بالمثل كما يُطبق مع الطفرات الجينية، إلى أن يتطور البناء إلى نموذج قوي (Siegel 2013: 122). وهناك شكلان للتعلم المُراقب وغير المُراقب وهما التعلم شبه المُراقب الذي يشتمل على استخدام كلٍ من التدريب والبيانات التي لا تحمل عناوين (وهي البيانات التي يُمكن الحصول عليها بسهولة ولكنها غير مسمّاة كالصور والتسجيلات الصوتية، والفيديو، المقالات الإخبارية، وغيرها من البيانات الخالية من أي ملصقات تتيح التعرف عليها بشكل أكبر)، والتعلم النشط الذي يتيح للمستخدمين القيام بدور نشط في توجيه نموذج التعلم (Han et al. 2011).

وقد جرى استخدام التعلّم الآلي في جميع عمليات تحليلات البيانات الكبيرة الأربع، على الرغم من أن هذه التحليلات ليست مقتصرة في عملها على هذه الطريقة. ويتم إجراؤها، في كثير من الحالات، من قبل الباحث بشكل مباشر، وحتى ضمن التعلم الآلي، يبقى المحلل ودوره مهماً في الحكم على العملية وتوجيهها وتقييم النتائج الوسيطة. وكما لاحظ (H.J. Miller 2010) فإن التعلّم الآلي ليس مجرد علم تلقائي وسهل المنال، بل يتطلب خبرة بالمجال وفكراً عميقاً؛ وهي المهارات التي لا يزال العقل البشري أفضل بها من أجهزة الحاسب الآلي. وبينما تم إحراز تقدّم كبير في تطوير تقنيات التعلّم الآلي، فإنه لا يزال من العلوم الناشئة التي مازالت بحاجة إلى الكثير من البحث لتحسين فعالية ومتانة النماذج المنتجة.

وفيما يلي، سيتم مناقشة كل فئة من فئات التحليلات الرئيسية على حدة، على الرغم من أنه تجدر الإشارة إلى أنها غالباً ما تستخدم جنباً إلى جنب. على سبيل المثال، ينبغي أن يوفر التنقيب عن البيانات والتعرف على الأنماط أسس التنبؤ/التوقع أو التحسين، ويُمكن استخدام الإحصاءات في التنقيب عن البيانات للكشف عن الأنماط أو في حساب التوقع/التنبؤ، وقد يستخدم التمثيل الصوري في التنقيب عن البيانات، أو لتحصيل المخرجات من أنظمة المحاكاة... إلخ.

التنقيب عن البيانات والتعرّف على الأنماط (Data Mining and Pattern Recognition):
 التنقيب عن البيانات هي عملية استخلاص البيانات وأنماط من مجموعات البيانات الكبيرة (Manyika et al. 2011). ويقوم التنقيب عن البيانات على مفهوم امتلاك كل مجموعات البيانات الضخمة معلومات مفيدة ذات معنى على صورة غير عشوائية، وذات صلاحية سارية المفعول، وقابلة للرواية، ومفيدة، ومفهومة في نهاية المطاف (Han et al. 2011). وبالمثل، يستخدم التعلّم الآلي المراقب وغير المراقب للكشف والتصنيف والتقسيم عن العلاقات والارتباطات والاتجاهات ذات المغزى بين المتغيرات. وهي تقوم بذلك باستخدام سلسلة من التقنيات بما في ذلك معالجة اللغات الطبيعية، والشبكات العصبية، وأشجار اتخاذ القرار، والأساليب الإحصائية (دون معاملات وذات المعاملات). وتختلف طريقة الاختيار بين نوع البيانات (المنظمة، وغير المنظمة، وشبه المنظمة) والهدف من هذا التحليل (انظر الجدول ١ - ٦).

الجدول رقم (١ - ٦)

مهام وأساليب التنقيب عن البيانات

الآليات	الوصف	مهمة التنقيب عن البيانات
التحليل العنقودي (Cluster Analysis)	تحديد قائمة من المجموعات الضمنية التي تصف البيانات	التجزئة أو التجميع (Segmentation or Clustering)
تصنيف النظرية الافتراضية (Bayesian Classification) استقراء شجرة القرارات (Decision Tree Induction) الشبكات العصبية الذكية (Artificial Neural Networks) آلة دعم التمييز (Support Vector Machine)	التنبؤ باسم الفئة التي تنتمي إليها مجموعة من البيانات استناداً إلى بعض مجموعات بيانات التدريب.	التصنيف (Classification)

قوانين الارتباط (Association Rules) تصنيف النظرية الافتراضية (Bayesian Classification)	إيجاد روابط بين كائنات البيانات؛ التنبؤ بقيمة بعض الحقول بالاعتماد على قيمة حقول أخرى	الربط (Association)
التحليل العنقودي (Cluster Analysis) كشف الانحراف (Outlier Detection) تحليل التطور (Evolution Analysis)	إيجاد عناصر البيانات المعرضة للانحراف عن المتوقع	الانحراف (Deviation)
الانحدار (Regression) استخلاص تسلسل النمط (Sequence Pattern Extraction)	الاتجاهات والمنحنيات التي تلخص قواعد البيانات، عادة عبر الزمن	الاتجاهات (Trends)
القواعد والقوانين الموجزة (Summary Rules) الاستقراء الموجه بالسمات (Attribute-Oriented Induction)	دمج مواصفات البيانات	التعميم (Generalization)

(المصدر: 7: 2009 Miller and Han)

إن معظم الآليات المدرجة في الجدول (1-6) تتصل بالبيانات المنظمة كتلك الموجودة في قواعد البيانات العلائقية. على سبيل المثال، من الممكن تطبيق نماذج التجزئة على قواعد بيانات عملاء التجزئة ومشترياتهم لتقسيمهم بحسب معرفات مختلفة بالاعتماد على خصائصهم وأنماط سلوكهم من أجل تقديم عروض / خدمات مختلفة لكل مجموعة مختلفة. وفي تحليل الشبكات الاجتماعية، يُمكن تحليل الروابط بين الأفراد لفهم المتغيرات الاجتماعية وكيفية تدفق المعلومات فيما بينهم. وفي كشف الارتباطات، يمكن استخدام مجموعة متنوعة من نماذج الانحدار في الكشف عن الروابط لحساب الاقترابات بين المتغيرات، ومن ثم الكشف عن الأنماط

الخفية التي يمكن الاستفادة منها لتحقيق مكاسب تجارية (على سبيل المثال، تحديد ما يتم شراؤه كوحدة من البضائع وإعادة تنظيم المخزون لتعزيز الشرائية) (انظر الفصل السابع).

وتؤثر البيانات غير المنظمة، والتي تكون على شكل نصوص أو صور أو أصوات، تحديات خاصة فيما يتعلق بالتنقيب عن البيانات. فآليات معالجة اللغات الطبيعية تهدف إلى تحليل اللغة الإنسانية التي يجرى التعبير عنها من خلال الكلمة المكتوبة والمنطوقة. وهي تستخدم دلالات وتصنيفات من أجل التعرف على الأنماط واستخراج المعلومات من الوثائق. وتشتمل الأمثلة على استخراج العنصر الذي يستخرج البيانات الوصفية من النص بصورة تلقائية من خلال البحث عن أنواع معينة من النصوص والتعابير، مثل أسماء الأشخاص والمواقع والتواريخ والمصطلحات التخصصية والتراكيب المنطقية؛ واستخراج العلاقات التي تحدد تلقائياً العلاقات بين العناصر الدلالية، وربطها معاً (على سبيل المثال، اسم الشخص وتاريخ ميلاده أو مكانه، أو رأيه ببند ما) (McCreary 2009). ومن التطبيقات النموذجية لمثل هذه الآليات التحليل الوجداني الذي يسعى إلى تحديد الطبيعة العامة للآراء وقوتها حول قضية ما، على سبيل المثال، ما يقوله الناس عن المنتج على وسائل التواصل الاجتماعي. كما يمكن تعقب أماكن نشوء هذه الآراء باستخدام البيانات الوصفية الموضعية (Graham et al. 2013) والتنقيب عن نشر المعلومات داخل وسائل التواصل الاجتماعي، على سبيل المثال، تبادل عناوين المواقع الإلكترونية الأكثر تفضيلاً ومشاركتها بين مختلف المستخدمين (Ohlhorst 2013). فهذه المعلومات مفيدة للشركات، مثل الوكالات الإعلانية، والمسوقين، والخدمات المالية، والتي تسعى إلى الاستفادة من الاتجاهات الناشئة والوصول المنظم في الوقت المناسب (على سبيل المثال، لوضع إعلانات تتناسب مع الأحداث الجارية؛ لبيع وشراء الأسهم قبل ردة الفعل العنيفة للسوق).

أما الصور فيتم إنشاؤها للتخزين والعرض، وليس للمحتوى والبحث (Ohlhorst 2013). فكشف وتصنيف واستخراج أنماط داخلها، مثل التعرف على الوجه أو مكان، ليس بالأمر السهل، ولكن يتم التعامل مع ذلك من خلال تقنيات المسح التصويري، والاستشعار عن بعد، ومعالجة الصور وتقنيات الاستبصار الآلية، متضمناً ذلك التعرف على الأشياء ومطابقة القالب باستخدام مجموعات التدريب، وطرق التجميع، والشبكات العصبية. وتتفاقم

مشكلة التنقيب في الصور عند محاولة استخراج، ومقارنة، وفهرسة الأنماط لعدد هائل من الصور (Zhang et al. 2001). وبرغم أن التنقيب في الصور لا يزال من المواضيع الناشئة، إلا أنه أصبح أكثر تقدماً في السنوات الأخيرة. فعلى سبيل المثال، فإن الموقع الإلكتروني على شبكة الإنترنت: (<http://imagevision.com>) يدعي بأنه سيكون قادراً على تصنيف 50400 دقيقة من الفيديو في الساعة الواحدة، لكل خادم، وذلك باستخدام خوارزميات التعلم الآلي للكشف عن بعض الخصائص مثل العري وشعارات الشركات.

التمثيل الصوري للبيانات وآليات التحليل الصورية (Data Visualisation and Visual Analytics):

شاعت مقولة تنص على أن صورة واحدة تغني عن ألف كلمة. وبالمثل، جرى استخدام السجلات البصرية لاختزال ووصف مجموعات البيانات من خلال المخططات والرسومات الإحصائية، والرسوم البيانية، والخرائط، والرسوم المتحركة. وتكشف هذه الأساليب البصرية وتوصل على نحو فعال هيكل المتغيرات، وأنماطها، واتجاهاتها، والروابط بينها. وبالنظر إلى الحجم الهائل للبيانات الكبيرة وسرعتها، فليس من المستغرب إذاً أن التمثيل الصوري قد صار وسيلة شعبية لاستشعار البيانات وإيصال هذا الشعور.

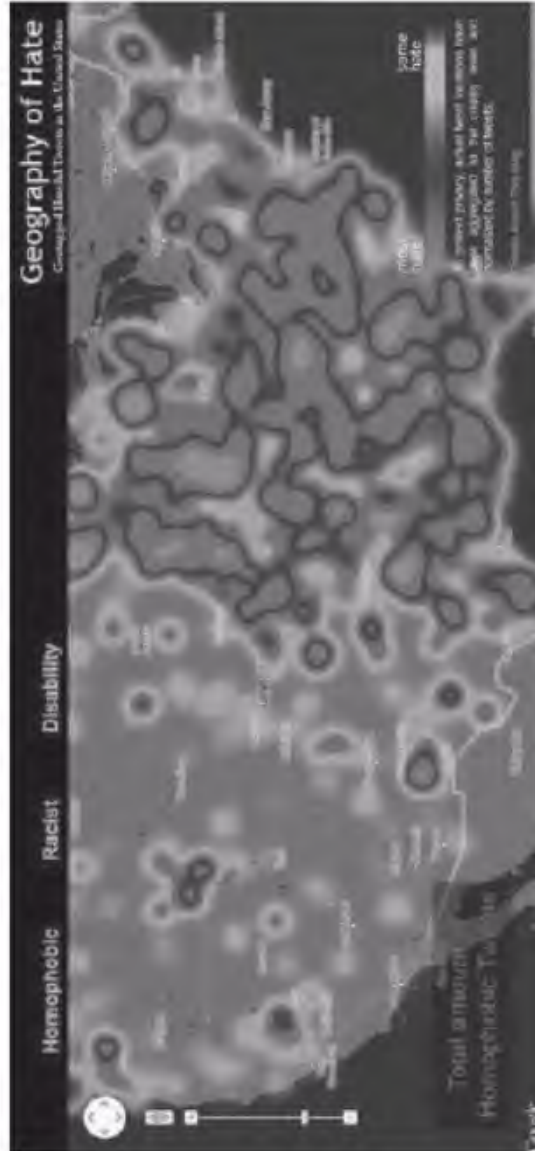
ويمكن استخدام التماثيل الصورية الناشئة ضمن عالم التقنية الرقمية للتنقل والاستعلام عن البيانات، مما يتيح للمستخدمين الحصول على لمحة عامة عن كامل مجموعة البيانات، بالتركيز على البنود المهمة، واستبعاد البيانات غير المهمة، واختيار عنصر بياني أو مجموعة من البيانات والحصول على مزيد من التفاصيل، وعرض العلاقات بين العناصر، واستخراج مجموعات فرعية من التفاصيل عند الحاجة (Shneiderman 1996). وعند عمل ذلك، يمكن الكشف عن خصائص وبنية مجموعات البيانات، وأكثر من ذلك يمكن استخدامها لتوضيح وفهم النماذج ومجموعات البيانات المعقدة التي يتعذر تمثيلها بطريقة مجردة وواضحة ومفهومة (مثل الظواهر الذرية، والظواهر الفلكية الثلاثية الأبعاد)، ورسم عشرات الآلاف من نقاط البيانات للكشف عن البنية، والتجميعات، والثغرات، والقيم الشاذة التي قد تظل بدلاً من ذلك مجهولة في الخفاء (Shneiderman 1996). على سبيل المثال، استيعاب وفهم ملايين التغريدات على شبكة التواصل الاجتماعي تويتر ليست مهمة سهلة، فيمكن للمرء

الحصول على فكرة تقريبية عن القصد أو المواضيع التي تبدو واعدة، ولكن الحصول على نظرة عامة وشاملة هي مهمة غاية في الصعوبة، وكيفية تغيير التوجه على مستوى الجماهير والأماكن. فأحد الحلول يتمثل في رسم خريطة تغريدات موزعة جغرافياً ومصنفة بحسب الميول والأهواء، وهو نهج جرى الأخذ به من خلال فريق عمل الموقع الإلكتروني (<http://floatingsheep.org>). لقد أنشؤوا عشرات الخرائط حول توجهات المواضيع والتغريدات، واشتمل ذلك على الشكل (١ - ٦) والذي يعرض التوزيع الجغرافي لتغريدات التخوف من المثليين في الولايات المتحدة من يونيو ٢٠١٢ م وحتى إبريل ٢٠١٣ م. كما أنهم عرضوا محتوى موسوعة الويكيبيديا وجوجل بالنسبة لمواضع التوزيع الجغرافي حول العالم.

ويشجع استخدام التمثيل الصوري أيضاً ليكون وسيلة للمراقبة الآنية المستمرة لظاهرة ما بشكل آني أو لحظي، مما يتيح تعقب ورصد العديد من المتغيرات عبر الزمان والمكان، والمقارنة بينها، وتحديد التغيير الحاصل. فلوحدات التحكم والعرض (Dashboards) للبيانات المتغيرة تكون غالباً معروضة على شاشات الحاسب الآلي في غرف التحكم الحديثة، وهي تُلخص على نحو صوري الأنظمة المتغيرة باستمرار للمشغلين الآدميين، مع سلاسل زمنية ورسوم بيانية وخرائط للكشف عن الأحداث (انظر ليك Lake 2013). للمقارنة بين ٢٤ لوحة عرض). وعلى سبيل المثال، يمكن أن توفر البيانات من مختلف أنحاء نظام النقل خريطة لتدفق حركة المرور وتقارير الحوادث بشكل آني أو لحظي؛ أو مواقع الطائرات خلال مرورها عبر منطقة ما (انظر الشكل ٢ - ٦)، كما قد توفر بيانات رادار الطقس خريطة لهطول الأمطار بشكل آني أو لحظي ومحاكاة للساعات القليلة الماضية. فمثل هذه البيانات الصورية لا تساعد ضباط إدارة حركة المرور والراصدات الجويين فقط، ولكن أيضاً المواطنين الذين يمكنهم الوصول إليها ومتابعة الأوضاع الجارية باستخدام الحاسب الآلي أو الهواتف الذكية وتبديل سلوكهم لتجنب طرق معينة أو التعامل بشكل مناسب مع المستجدات. ومثال على نموذج أولي للوحة تحكم خاصة بمدينة عامة تضم مجموعة من البيانات الآنية أو اللحظية عن الطقس، وتلوث الهواء، ومعوّقات حركة مرور النقل العام، وتوافر الدراجات العامة، ومستوى ارتفاع المنسوب المائي النهري، وارتفاع الطلب على الطاقة الكهربائية، وسوق الأوراق المالية، واتجاهات التغريدات على شبكة التواصل الاجتماعي تويتر، وتغذية كاميرات مراقبة حركة المرور تلك المعروضة في الشكل (٣-٦).

الشكل رقم (١ - ٦)

التوزيع الجغرافي لتغريدات التخوف من المثلية في الولايات المتحدة الأمريكية



المصدر: http://users.humboldt.edu/mstephens/hate/hate_map.html#

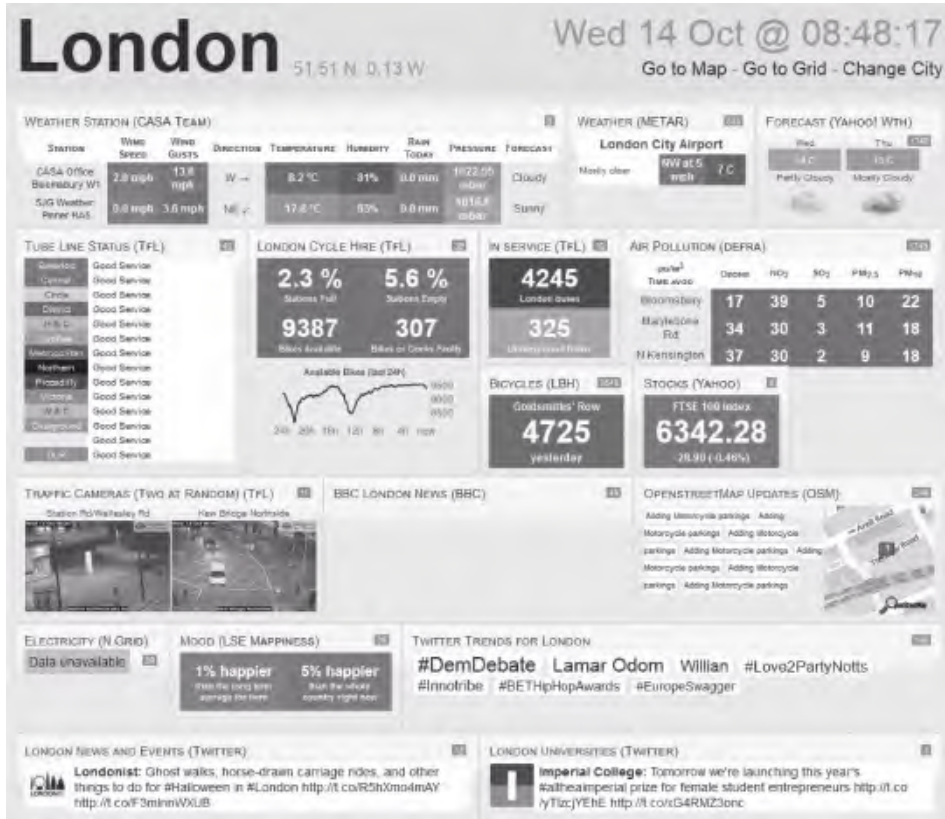
الشكل رقم (٢ - ٦)
المواقع اللحظية لمسار الرحلات الجوية



المصدر: <http://www.flightradar24.com/>

الشكل رقم (٣ - ٦)

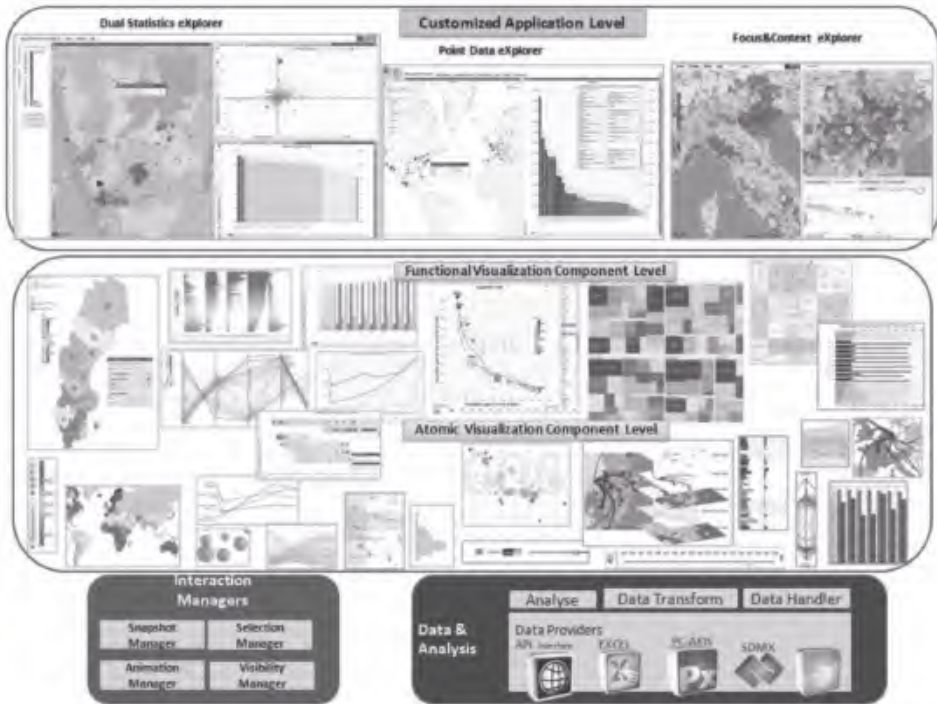
لوحة معلومات مدينة لندن الخاصة بالمركز البريطاني للتحليل العمراني المتقدم



المصدر: <http://citydashboard.org/london/>

الشكل رقم (٤ - ٦)

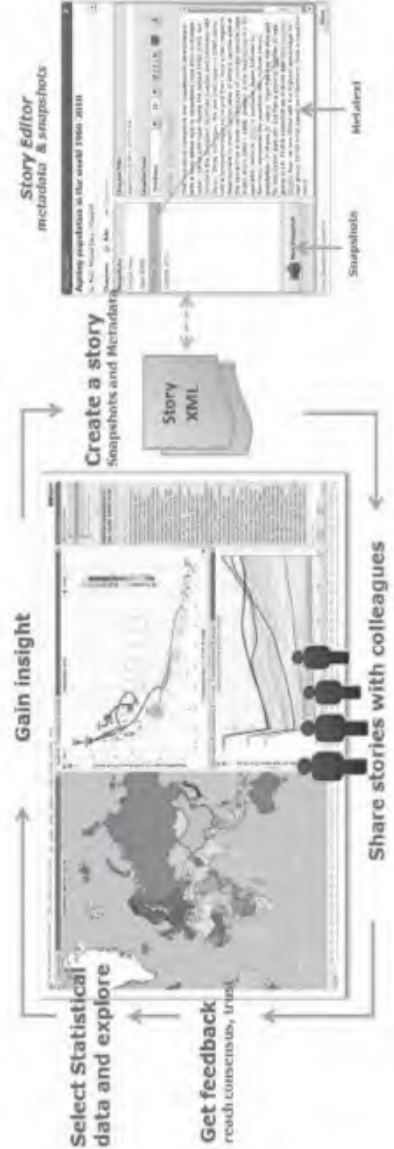
مجموعة أدوات التحليلات الجغرافية الصورية التي وضعها المركز الوطني للتحليلات الصورية في
جامعة لينكوبينج (Linköping University)



المصدر <http://ncva.itn.liu.se/tools?l=en>

الشكل رقم (٥ - ٦)

استخدام مجموعة أدوات التحليلات الجغرافية الصورية في بناء الرواية القصصية الجماعية



المصدر <http://ncva.itn.liu.se/tools?l=en>

ومن الممكن أيضاً استخدام التمثيل الصوري كشكل من أشكال الاستنتاج التحليلي. وهنا، لا يصف أو يعرض التمثيل الصوري مجرد البيانات فقط، ولكن يجري استخدامها أداة تحليلية. ويتم توجيه هذه الطريقة، والتي يطلق عليها مصطلح التحليلات الصورية، في الغالب بواسطة مجموعة من الأشخاص والخوارزميات التي تعمل باتساق لاستخراج البيانات، وبناء النماذج البصرية وتفسيرها، والمضي قدماً في المزيد من التحليل الإحصائي (Keim et al. 2010). لقد أصبحت آليات التحليل البصرية مجالاً مهماً وجديداً من مجالات البحث، وهو الأمر الذي أسهم في تطوير حقل التمثيل الصوري للمعلومات من خلال دمج عناصر التنقيب عن البيانات، وعلوم الإحصاء والإدراك (Thomas and Cook 2006). فالعديد من أنواع الرسومات البصرية المختلفة شائع الاستخدام، وهي في العموم تفاعلية وتتيح للمستخدم التعامل معها والتلاعب في الصور من أجل البحث والكشف عن الأنماط والروابط. وبالإضافة إلى ذلك، فإن الشاشات المعروضة للمستخدم متصلة ببعضها في الغالب بحيث إن التفاعل مع إحداها ينعكس على الأخرى، مما يمكن من تفحص جوانب من البيانات من أكثر من منظور واحد في ذات الوقت. فعلى سبيل المثال، يعرض الشكل (٤-٦) أنواع آليات التحليل الصورية المتاحة ضمن أدوات آليات التحليل الصورية (GAV)، الموضوع من قبل المركز الوطني للتحليلات الجغرافية الصورية في جامعة لينكوبينج (Linköping University) في السويد. بالإضافة إلى توفيرها مجموعة من الأدوات التحليلية، تمكن أدوات المجموعة من تحليل البيانات واستكشافها بشكل تعاوني، وبناء التطبيقات، والتداخل مع تطبيقات الخرائط كخرائط جوجل، وبناء وتبادل القصص السردية. وينشئ الأخير بعداً اجتماعياً للأدوات، مما يمكن الزملاء وغيرهم من تبادل تفسيرات حسية متعلقة بالتمثيل الصوري (انظر الشكل ٥ - ٦).

التحليل الإحصائي (Statistical analysis):

هناك تاريخ طويل من الأساليب الإحصائية التي يجري تطبيقها على البيانات الكمية في محاولة لفهمها. فالإحصاء الوصفي يفصل خصائص وتوزيع نقاط البيانات ومستوياتها من الخطأ وعدم اليقين. ويشتمل الإحصاء الوصفي على تحليل السلاسل الزمنية التي تنص على كيفية تغير البيانات بمرور الزمن، ونظرية الرسم البياني التي تدرس بشكل رياضي تنظيم

هياكل الشبكة، والإحصاء المكاني الذي يصف هندسة وأنماط التجميع، والتشتت، والانتشار، في حين يسعى الإحصاء الاستدلالي إلى شرح، وليس مجرد وصف، الأنماط والعلاقات التي قد توجد داخل مجموعة البيانات، واختبار قوة وأهمية الارتباطات بين المتغيرات. كما يشمل الإحصاء البارامتري (المعلمي) الذي يجري استخدامه لتقييم الفرضيات باستخدام بيانات الفاصل الزمني وبيانات مستوى النسبة، مثل الارتباط والانحدار؛ والإحصاءات غير المعلمية المستخدمة لاختبار الفرضيات باستخدام بيانات المستوى الاسمية أو الترتيبية؛ والإحصاءات الاحتمالية التي تحدّد احتمال حدوث حالة ما، مثل تحليل النظرية الافتراضية (Bayesian). ويجري تطبيق تقنيات الإحصاء الوصفي والاستدلالي التي تستخدم تقليدياً في تحليل البيانات الصغيرة في البيانات الكبيرة أيضاً، وكما جرى نقاشها في الفصل التاسع فإن هذا ليس سهلاً على الدوام لأن الكثير من هذه التقنيات جرى وضعها لرسم رؤى من البيانات الشحيحة وليست المتوافرة بشكل فائض. ومع ذلك، توفّر هذه التقنيات وسائل للاستفادة من كميات البيانات الهائلة، كما يضمن حجم البيانات الهائل إمكانية تجزئتها إلى مجموعات بطرق عديدة دون أن تصبح صغيرة جداً مما لا يمكن معه حساب الاختلافات بين مجموعتي الضبط والمعالجة بصورة ذات معنى. وتستخدم الإحصاءات بشكل روتيني في المساعدة على التنقيب عن البيانات والتنبؤ بها وتحسينها (انظر الجدول ١ - ٦).

التوقع والمحاكاة والتحسين (Prediction, Simulation and Optimisation):

هناك طريقة أساسية في الحصول على القيمة من البيانات وهي استخدامها لمحاولة معرفة أو توقع ما سيحدث في ظل ظروف مختلفة. على سبيل المثال، قد ترغب شركة ما في توقع ردود أفعال العملاء تجاه منتج معين أو حملة ما، أو قد تحتاج الحكومة المحلية إلى محاولة التنبؤ بكيفية عمل البنى التحتية للنقل إذا جرى إغلاق عنصر مهم ذو حساسية، أو قد يسعى العلماء إلى التنبؤ بموعد احتمال حدوث انهيار أرضي وتحت أية ظروف. إن مثل هذه المعلومات مفيدة جداً للمنظمات من حيث تمكينها من التخطيط لحالات الطوارئ المختلفة، وللشركات من حيث إنتاج أثر لزيادة نمو الأرباح (انظر الفصل السابع). وفي جميع الحالات، تُبنى النماذج باستخدام المعرفة القائمة سلفاً عن كيفية عمل النظام، والتي تعالج البيانات لتقدير النتائج المحتملة في ظل سيناريوهات مختلفة. وكما في التنقيب عن

البيانات، هناك العديد من الطرق المختلفة التي يمكن استخدامها لإنتاج النماذج التنبؤية. ولكل منها نقاط قوتها وضعفها، وهي تنتج تنبؤات أقل خطأً أو أكثر دقة بالاعتماد على نوع المشكلة والبيانات (Seni and Elder 2010). ومع ذلك، من الصعب الحكم مسبقاً على أن أي نوع من النماذج، وإصداراته المختلفة، سيعمل بشكل أفضل على أي مجموعة بيانات مُعطاة. وقد تمثل الحل لهذه المسألة في استخدام نهج مركب يستفيد من الكمية الهائلة من الطاقة الحاسوبية المتاحة للمحللين الآن (Siegel 2013). فعوضاً عن اختيار نهج واحد وبناء عدد قليل من النماذج، فالنهج المركب يبنى نماذج متعددة باستخدام مجموعة متنوعة من التقنيات للتنبؤ بالظاهرة نفسها. وبعد ذلك، بدلاً من تحديد النتائج (التقديرات) من النموذج الأفضل أداء، يجري العمل على الجمع بين التقديرات من جميع النماذج لإنتاج إجابة واحدة شاملة. ويؤدي تجميع النتائج إلى نتائج أكثر قوة كما تعوض العملية نقاط الضعف في كل نموذج. على سبيل المثال، قد يبنى نهجاً مركباً لتوقع سلوك العملاء سلسلة من نماذج الانحدار، والشبكات العصبية، والجار الأقرب، وشجرة القرارات. فكل نموذج قد يكون أفضل من غيره في التنبؤ بأنواع معينة من المستهلكين، ولكن يمكن تقليل الفروقات من خلال الجمع بين مخرجات النموذج، مما يزيد التنبؤ قوة (Franks 2012; Siegel 2013). فباستخدام طريقة تضم العديد من النماذج المنسجمة يمكن تطبيق المئات من الخوارزميات المختلفة على مجموعة بيانات، ومن ثم ضمان إنتاج أفضل نموذج تنبؤي ممكن.

فالمحاكاة هي عبارة عن نماذج تسعى إلى نمذجة العمليات والأنظمة في العالم الحقيقي. والهدف من ذلك هو تحديد كيفية عمل نظام ما وكيفية سلوكه وفقاً لسيناريوهات مختلفة، وتقييم أدائها إحصائياً بهدف تحسين كفاءتها وفعاليتها (Robinson 2003). ومن الأمثلة الشعبية لعبة الحاسب الآلي سيم سيتي (SimCity) التي تحاكي كيف يمكن لمدينة أن تنمو وتتطور في ظل ظروف يختارها اللاعبون، قائمة على أساس النموذج الكامن وراء العمليات المدنية المعروفة. وبالمثل، تستند توقعات الطقس على محاكاة لكيفية تطور الطقس بالنظر إلى الظروف السائدة والمعرفة العلمية. وهناك العديد من الأنواع المختلفة لنماذج المحاكاة، والتي استفاد كثير منها من التعلم الآلي من أجل تحسين عمل النموذج بشكل تلقائي والتعامل مع الخصائص الناشئة مثل الأحداث غير المتوقعة. إن نموذج سيم سيتي هو

نموذج قائم على التوكيل (Agent-based) (Batty 2007). ويتكوّن هذا النموذج من بيئة تُمكن من تعيين خصائص معينة للميزات الفردية فيها مثل المباني والطرق. ويتم تعيين خصائص هذه البيئة بواسطة وكلاء، وعند تشغيل النموذج، يسعى الوكلاء إلى حل مهمة ما، والتفاعل مع البيئة وغيرهم من الوكلاء الآخرين بالاعتماد على خصائصهم المسندة إليهم. وبالمحصلة، عند قيام الوكلاء بمهامهم، يقومون بتغيير البيئة، وهي في هذه الحالة المدينة التي يسكنونها، وإنتاج نظام معقد ومتغير. ومن ثم فإن النظام يعمل من أسفل إلى أعلى، مع أمّاط مكانية وزمانية تتغيّر من تفاعلات الوكلاء الفردية مع البيئة. وتستخدم مثل هذه النماذج لمحاكاة القضايا المدنية خارج الألعاب لنمذجة استخدام الأراضي وتخطيط النقل ووضع خطط الطوارئ للكوارث (Batty 2007).

ويهتمّ التحسين بتحديد الإجراء الأمثل لتحسين الأداء (عادة من خلال خفض التكلفة أو زيادة العائد أو المخرجات). ويمكن حساب التحسين باستخدام وتقييم نماذج التنبؤ والمحاكاة، أو تطويرها من خلال أنواع أخرى من الخوارزميات أو الاختبار الإحصائي. فعلى سبيل المثال، تستخدم الخوارزميات الجينية، وهو نوع معين من التعلّم الآلي، الأفكار المنتقاة من الطبيعة، مثل الوراثة، والطفرات، والاختيار والتهجين، لوضع وتطوير حلول ممكنة لمشكلة ما (Mitchell 1996). كما يسعى علم الشبكات العصبية، وهو نهج آخر مستوحى بيولوجيا، لمحاكاة كيفية عمل الدماغ البشري، باستخدام عناصر المعالجة العالية الترابط لحساب وتقييم وحلّ مشكلة ما (Picton 2000). كما يمكن استخدام اختبار أ/ب (A/B Testing) بشكل دوري لتقييم وتطوير نظام ما، من خلال مقارنة مجموعة ضوابط مع مجموعات اختبار مختلفة من أجل تحديد العلاجات (على سبيل المثال، النصوص، والتصاميم، والصور، والألوان المستخدمة على موقع إلكتروني على الإنترنت) لتحسين هدف معين (على سبيل المثال، معدلات التبرع).

الخلاصة:

كان الهدف من هذا الفصل تقديم لمحة عامة عن مجموعة من الآليات الناشئة وتحليلات البيانات التي يتم استخدامها لمعالجة وتحليل وتطبيق البيانات الصغيرة الموسّعة والبيانات الكبيرة. لقد تمّ نبذ مثل هذه الأساليب لعدد من السنوات، ولكنها أصبحت أكثر انتشاراً

ويمكن الوصول إليها مع تطورها ونضجها، وتوفّر المعالجة الحاسوبية المطلوبة لتصبح متاحة على نطاق أوسع. ومع ذلك، يطرح حجم وسرعة البيانات الكبيرة تحديات جديدة لآليات التحليل وما قبلها بسبب عدم تصميم العديد من هذه الآليات التقليدية والخوارزميات لتناسب مع المليارات من الملاحظات، أو لتتعامل مع مجموعات بيانات سريعة التغير (Batty et al. 2012). ومن ثم فمن الواضح أن هناك حاجة إلى مزيد من البحوث لتكييف، وصقل وتوسيع الآليات الحالية وابتكار أساليب جديدة يمكن لها إضفاء معنى من البيانات الكبيرة والبنى التحتية للبيانات واستخلاص القيمة منها. وعلاوة على ذلك، أضحت الأجيال الجديدة والقديمة من العلماء ومحلي الأعمال، على حد سواء، بحاجة إلى التعرف على مصادر البيانات الجديدة وخصائصها، وأن تتعلّم مجموعة جديدة من المهارات إذا كانت تريد أن تكون قادرة على الدراسة والاستفادة من رؤى وأفكار طوفان البيانات القائم، متضمناً ذلك معالجة البيانات المنظمة والبيانات غير المنظمة، والتمييز الأساسي، والتنقيب في البيانات، والتمثيل الصوري، والنمذجة، والمحاكاة (انظر الفصل التاسع).

إن هذه التحديات لا تقوم على مجرد أسس تقنية أو موارد بشرية. فإنّ آليات التحليل تُعدُّ انعكاساً لطريقة معينة في الرؤية والاستشعار للعالم المحيط؛ فهي مظهر من مظاهر المعرفة الخاصة. وكما نوقش في الفصل الثامن، تثير البيانات الصغيرة الموسعة والبيانات الكبيرة وآليات التحليل المرتبطة بها تساؤلات جوهرية بشأن تنظيم وممارسات العلوم (Floridi 2012)، مع العديد من المقترحات التي توحى بالدخول في نموذج علمي جديد. ونتيجة لذلك، لا يتطلب التصدي لطوفان البيانات مجرد فكر جديد حول طرق التحليل، ولكن تفكيراً متأنياً فيما يتعلق بفلسفة العلوم (Leonelli 2012). ويستدعي جزء كبير من هذه المهمة التفكير النقدي في التأطير الواسع للتحليلات وآليات الطرق والمناهج المحددة، والتفكير في كيفية إنتاج آليات التحليل الجديدة أنواعاً معينة من المعارف والرؤى للعالم المحيط بنا، ومن ثم العمل على معرفة عمل هذه المعارف الجديدة في العالم، متضمناً ذلك نمذجة كل الأشياء التي تشير إليها (Beer and Burrows 2013; Ruppert 2013)؛ انظر الفصل الحادي عشر). وبعبارة أخرى، هنالك حاجة ملحة وبصورة عاجلة للبحوث الفنية والفلسفية فيما يتعلق بآليات التحليل الناشئة من أجل استشعار البيانات الموسّعة والبيانات الكبيرة.

الفصل السابع

المبررات الحكومية والتجارية للبيانات الكبيرة

(The Governmental and Business Rationale for Big Data)

لم تتكشف ثورة البيانات دون فكر وبطريقة مجهولة، فكما هو حال الثورات الأخرى، جرى الدفع بها بسبب مجموعة قوية من الحجج، والتمهيد لها من قبل مجموعة من المتحمسين المؤمنين بفوائد الطرق الجديدة للمعرفة والتعامل في هذا العالم وتحالف أصحاب المصالح الخاصة الذين يجنون الأرباح من ظهورها. وفي حين أن الحجج التي استخدمت يمكن أن تكون متنوعة جداً، وغالباً ما تنشأ عن وجهات نظر مختلفة، إلا أنها تشكل في مجموعها ما أطلق عليه Foucault (1977) مصطلح نظام الخطابة والمجادلة. أي مجموعة من النقاشات المتداخلة التي تبرر وتغذي التطورات الجديدة وتعمل على تطبيعها واستنساخها وإعادة استخدامها. فهذا النظام الخطابي يقدم الأساس المنطقي لتبني الأفكار والتقنيات الجديدة، وإضفاء الشرعية على مساعي تطويرها وسبل تنفيذها. وتدعو النقاشات المستخدمة في النظام إلى الترويج لرسالتهم وجعلها قناعة عامة، لترغيب الناس والمؤسسات باعتناق منطقهم، وإلى الاعتقاد والعمل فيما يتعلق بهذا المنطق (Dodge and Kitchin 2005). كما عملت على إنتاج أجواء معينة يتم من خلالها "تغذية أنواع خاصة من الفكر والممارسة بحيث تبدو طبيعية ومرغوباً فيها" (Amin and Thrift 2013: 5)، وهي تغرس نوعاً خاصاً من الاستجابة العاطفية (Anderson 2009). فمثل هذه الخطابات والأجواء سعت إلى إعادة صياغة العالم وفق رؤية خاصة، وإعادة تشكيل "القضايا، الفاعلين، المخاطر، المصالح، والرغبات التي تنتشر في الحقل السياسي والمشهد الاقتصادي" (Barnett 2013: 4).

وكما تمّ النقاش في الفصلين الثاني والثالث، فقد سعت نظم الجدل والمحاور المتداخلة إلى الدفع بعمليات تطوير البنى التحتية للبيانات والبيانات المفتوحة مستخدمة النقاشات والمداولات حول تبادل ومشاركة البيانات، واستنساخها وإعادة استخدامها، وحرية النفاذ

والوصول إليها، والحكومة المفتوحة، والشفافية، والمساءلة، وتنظيم المشاريع الاجتماعية، واقتصاديات السعة ووفورات حجم الإنتاج. وكما تمّ بيانه في هذه الحالات، فإن منحى النقاش وطريقة التداول قد تكون مدفوعة من قبل مجموعة من الأطراف المعنية التي قد تستخدم الحجج نفسها ولكن لتحقيق مرام مختلفة (على سبيل المثال، تستخدم منظمات المجتمع المدني وقطاعات الأعمال الكبيرة خطابات الشفافية والمساءلة لإتاحة بيانات الحكومة وفتحها، ولكن أحدهما لأسباب مدنية والأخرى للانتفاع المالي من البيانات). فالأنظمة الخطابية إذاً لم تكن حصراً نتاجاً فوقيّاً من الدول أو الأعمال، ولكنها انتشرت وسُنّت من خلال العديد من القنوات الأصغر حجماً (Foucault 1977, 1978). وقد جرى قبولها دون إشكالية من قبل الجماهير، من المواطنين والمؤسسات الذين كان من السهل استمالتهم من خلال منطقتهم. وبالأحرى، فإن الخطابات مفتوحة للتحدي، والرد عليها، ومقاومتها، وتفنيدها، ومخالفتها، والتصدي لها، وإظهار عوارها، وتقديم رؤى بديلة جديدة. وفي كالتين، يستند المرؤجون أو المعارضون لرؤى جديدة على السجلّ الزمني الممتد من الماضي، إلى الحاضر، وحتى المستقبل لدعم حججهم، وذلك باستخدام الأدلة والشواهد من جولات سابقة من التغير الاجتماعي والاقتصادي، إلى جانب التحديات الحاضرة، والتكهنات حول الكيفية التي يمكن أن تشكل الكيفية التي ستكون عليها الأمور خلال سنوات قادمة، في ظل سيناريوهات مختلفة، (Kinsley 2011, 2012). ومن ثم، جرى تداول المشهد الخطابي عبر جولات من التفاوض باسم الصراع على الأفكار والسبل لهيكل وتمثيل وعمل وتنظيم المجتمعات، في الحاضر والمستقبل.

وفي هذا الفصل، سيجري دراسة المبررات المنطقية للبيانات الكبيرة للجهات الحكومية وقطاع الأعمال. وقد تمّ تنظيم النقاش حول أنماط خاصة من النظم والأنشطة الاجتماعية بدلاً من تنظيم الفصل بحسب مختلف الخطابات التي يشيع توظيفها للترويج لإشاعة اعتماد البيانات الكبيرة على سبيل المثال، التبصّر والحكمة، الإنتاجية، القدرة التنافسية، الكفاءة، الفعالية، الفائدة، الاستدامة، والتوريق المالي (Securitization). فالخطب تتملك تأثيرها الأقوى عندما يجري تعبئتها وتوجيهها في مهام وقضايا معينة، والمثير في حالة البيانات الكبيرة هو أن جدالها يمتد إلى جميع القطاعات والشرائح - الاجتماعية، والسياسية، والاقتصادية، والبيئية؛ فهي واعدة من خلال تقديمها وسائل جديدة كلياً في فهم

وإدارة جميع جوانب حياة الإنسان. وهنا، سيتم مناقشة هذا الوعد فيما يتعلق بأربع مهام رئيسية هي: حكم الناس وتنظيمهم، وإدارة المنظمات، وتعزيز القيمة وإنتاج رأس المال، وأخيراً إنشاء أماكن أفضل. وهذه مهام توضيحية فقط، على سبيل الحصر وليست شاملة، حيث إن هناك العديد من التطبيقات الأخرى الممكنة، بما في ذلك القضايا البيئية (انظر Mayer-Schonberger and Cukier 2013; Siegel 2013; Smolan and Erwitte 2012). وستتم مناقشة كيفية إعادة تشكيل البيانات الكبيرة للبحوث في الفصل التالي، كما سيجري مناقشة الدعوات المضادة لتطبيق البيانات الكبيرة في الفصل العاشر.

حكم الشعب وتنظيمه (Governing People):

الدولة هي مولدٌ ومستخدم رئيس للبيانات، فمنذ عصر التنوير سعت الدولة لإنشاء طرق أكثر منهجية، من جهة، لإدارة وحكم المجاميع السكانية، ومن ناحية أخرى، لتقديم الخدمات للمواطنين. فإحدى هذه الطرق الرئيسية التي هدفت من خلالها إلى أداء هذه الأدوار بكفاءة وفعالية كانت من خلال التدقيق والتقدير الكمي والنوعي للمجتمع - أي بناء معرفة مفصلة لإنفاذ ولايتها القضائية والاختصاصية من خلال تطوير النظم الإدارية على نطاق واسع، بدعم من أجهزة فنية وتقنية متنوعة، تعمل على قياس وتتبع فهرسة وتمييز البيانات عن الأفراد، والشركات، والمؤسسات، والمكونات الأخرى (انظر Alonso and Starr 1987; Barnes and Hannah 2001; Desrosières 1998; Kitchin and Dodge 1995; Porter 2011). إن إنتاج، وإدارة، وتخزين، وتحليل البيانات من المهام الرئيسية في جميع مؤسسات الدولة، فهي تُستخدم في تقييم الالتزامات والاستحقاقات للموضوعات السيادية وغير السيادية، والكشف والمراقبة عن الخروقات وعدم الامتثال، والتهرب، والغش، فيما يتعلق بالمعايير والالتزامات القانونية المعمول بها (Rose 1996). فعلى سبيل المثال، يتم في الدوائر الحكومية، وخاصة تلك المتعلقة بالتمويل والتجارة والرعاية الاجتماعية، تحديد جميع المواطنين والكيانات وتتبعهم، ورصد القضايا والاتجاهات من خلال التحديثات على سجلاتهم، وتعيين ملفاتهم التعريفية وربط بياناتها بهم، وتعقب المدفوعات، وإدارة انضباطهم والخدمات المقدمة لهم. أما في المدارس، فيؤخذ التسجيل لبياناتهم، واحتساب وتقييم الدورات الدراسية وعلامات الاختبارات، ويتم مكافأة الطلاب

ومعاقبتهم وفقاً لأدائهم وسلوكهم. وفي قطاع الصحة، يتم قياس حالة المريض، وتسجيلها، والعمل على أساسها، واحتساب وتقييم أداء المرافق الصحية. كما تُعنى وكالات حكومية متخصصة أخرى بإنتاج وإدارة وتحليل ونشر البيانات التي لها بالغ الأثر في إفادة العمل الحكومي. على سبيل المثال، الوكالات الإحصائية، وهيئات المسح ورسم الخرائط الجغرافية، وأجهزة الاستخبارات التي تتعهد ببرامج مستمرة لإنتاج البيانات من خلال أنظمة قياس مثل المسح والمراقبة.

ولتعزيز ممارساتها الحسابية، سعت الدول، من جهة، لتبرير وترشيد نظمها وعملياتها على الموضوعات، ومن جهة أخرى، لتطبيق وسائل إدارية أكثر تعقيداً وتفصيلاً من أي وقت مضى (McNay 1994; Miller 2001). فالمسعى الأول يشمل الخطابات حول الحكومة الفعالة والقيمة مقابل المال، والمواطنة والتمكين، والعدالة ومكافحة الجريمة، والأمن والسلامة، والتي تتماشى مع فكرة تناول المشاكل والقضايا العامة من خلال المنطق الجماعي والمبادئ المشتركة. وتُجابه هذه الخطابات من قبل أولئك الذين لديهم الرغبة في تقليل مستوى رقابة الدولة وسلطتها مقابل الحريات الشخصية، والرقابة والضبط للنظم المجتمعية والاقتصادية، أو أن تكون أكثر انفتاحاً واستجابة وإنتاجية في الممارسات الحكومية، متضمناً ذلك إتاحة البيانات التي يحملونها بشكل أكبر وأكثر حرية (انظر الفصل الثالث).

أما المسعى الثاني فيتضمن تنفيذ أنظمة وتقنيات وبنى تحتية مصممة لجمع ومعالجة البيانات على نحو فعال وأكثر تطوراً من أي وقت مضى، متضمناً ذلك الأنظمة الأكثر ثباتاً وقدرة في التعريف والتحديد (التواقيع والصور وبصمات الأصابع، وأرقام الضمان الاجتماعي وجواز السفر، وقوانين الضرائب، وأرقام الترخيص والحسابات، والمعلومات البيومترية، والمستخدمين وكلمات السر والرموز البريدية... إلخ)، والتي يمكن تشغيلها على نحو متزامن بشكل متزايد وعن بُعد (Dodge and Kitchin 2005). وعليه، يؤكد Hannah (1997: 178) على أن أحد أشكال الحاكمية الأكثر فعالية يكون في الغالب من خلال تكوين "ليف معقد من مجموعة بسط النفوذ والهيمنة...من خلال التعدادات والبيروقراطية والسياسات، وما إلى ذلك، وكم هائل من المؤسسات التخصصية مثل المدارس وأماكن العمل، والأراضي الممتدة المجهولة المثبتة بملايين من نقاط الانطلاق كأجهزة الصراف الآلي، وكاميرات المراقبة، ونماذج تعبئة الطلب".

ولذا، فإن البيانات الكبيرة بأحد معانيها ليست شيئاً جديداً على الدول - وهي التي تعاملت طويلاً مع كميات هائلة من البيانات في إدارة الاختصاصات القضائية والقانونية، كجزء جرى معالجته واتخاذ الإجراءات المتعلقة به في الوقت المناسب في حينه أو قريباً منه كلما تفاعلت الوكالات مع الموضوعات. ومع ذلك، يُنظر إلى البيانات الكبيرة باعتبارها أحدث مجموعة من التقنيات التي يمكنها توسيع وتحسين عمل الدول من خلال زيادة الانضباطية والشمولية للممارسات الحسابية لأعمالها. كما يتبنى البائعون، في بعض الحالات، خطاب الدفع بحلول البيانات الكبيرة بهدف بيع خدمات الدولة، وفي حالات أخرى، تأخذ الدولة بزمام المبادرة بشكل نشط. وفي الوقت الحاضر، هناك نوعان من الإستراتيجيات الخطابية الرئيسية المستخدمة لتشجيع استخدام البيانات الكبيرة في السلطات الاختصاصية، يتعلق النوع الأول بالتحسينات على الإدارة العامة وتقليل التكاليف من خلال تعزيز الكفاءة التشغيلية، وتقليل تكلفة الأخطاء والاحتيايل في إدارة المنافع، وزيادة في الإيرادات الضريبية عن طريق تضيق الفجوة الضريبية... وتحسين تخصيص التمويل في البرامج والخدمات ذات الجودة الأعلى، وزيادة المساءلة في القطاع العام، وإطلاع أفضل للمواطنين"، مع تقديرات مقترحة بعائدات لحلول البيانات الكبيرة من الكفاءة التشغيلية المتعلقة بإدارة القطاع العام في القارة الأوروبية تعادل مئة وخمسين مليار دولار أمريكي (Manyika et al. 2011: 61). وسيجري لاحقاً مناقشة هذه الحجج بشكل أكبر فيما يتعلق بإدارة المنظمات وصنع أماكن أفضل.

ويهتم النوع الثاني بأمن الدولة ومكافحة الجريمة، إذ تقوم جميع الدول بالمراقبة والرصد لأغراض الأمن والسلامة ومنع الجريمة والاعتقال من خلال أجهزة الشرطة وجمع المعلومات الاستخباراتية على نطاق أوسع. وقد تمّ تطوير الكيفية التي تنفذ بها هذه الأنشطة بما يتماشى مع التقنيات الجديدة - كالتنصت على المكالمات الهاتفية واعتراض الموجات الراديوية، والدوائر التلفزيونية المغلقة، والتصوير الجوي، وصور الأقمار الصناعية، ومراقبة الإنترنت - والتي كثيراً ما جرى التبرير لها عند وقوع أحداث رئيسية مثل الحرب الباردة، والحوادث الإرهابية (Innes 2001). وفي الآونة الأخيرة، قدّمت أحداث الحادي عشر من سبتمبر وما تبعها من الحرب على الإرهاب المبرر المنطقي لتغيير جذري في عمليات المراقبة والتأمين التي تقودها الدولة على نطاق واسع (Amoore 2006). وكان جلياً لبعض

الوقت أن تحليلات البيانات الكبيرة كانت عنصراً أساسياً في تحسين هذه الأنشطة، وخاصة فيما يتعلق بمراقبة الحدود وفحص الهجرة (Kitchin and Dodge 2006; Clements). وقد كشفت تسريبات يونيو 2013 بشأن برنامج وكالة الأمن القومي الأمريكي بريزم (PRISM) المدى الذي وصلت إليه ممارسات مراقبة البيانات، وممارسات البيانات الكبيرة، والآليات التحليلية المستخدمة لجمع ومعالجة البيانات المتعلقة بالاتصالات اليومية لجميع المواطنين مثل المكالمات اليومية، الرسائل النصية، رسائل البريد الإلكتروني، واستخدام الإنترنت، بدلاً من اقتصرها على المواطنين الأجانب والقوى الخارجية أو لأهداف محددة (Greenwald and MacAskill 2013; Risen and Lichtblau 2013). فهذه التسريبات كانت هي الأحدث في سلسلة الكشف عن تجسس الولايات المتحدة على مواطنيها بشكل خفي (Solove 2007; Bamford 2009). والواقع، أن من الواضح أن الوكالات الاستخباراتية، والمتعاقدين العسكريين-الصناعيين، هم في طليعة مطوري الجيل القادم من أدوات البيانات الكبيرة لإدراك الكميات الهائلة من البيانات عالية السرعة، وقد استثمرت بكثافة في بناء الكفاءات والتقنيات والمرافق على سبيل المثال، جرى استثمار 1.7 مليار دولار أمريكي في مركز البيانات القائم في يوتا لمعالجة وتخزين البيانات الكبيرة التي تم جمعها من الولايات (Carroll 2013). وبينما تسعى الدول إلى ترشيد تلك المراقبة ذات النطاق الواسع على مواطنيها من خلال الخطابات العامة الموجهة للأمن والسلامة، فإن الحقيقة الماثلة هي أنها قد طوّعت الكثير من أجهزتها سراً لتنفيذها ودون اتباع للإجراءات القانونية الواجب العمل بها مما يقوّض المثل العليا للديمقراطية القائمة على الشفافية والثقة (Clements 2013). ومع ذلك، فمن غير المرجح على المدى القصير أن يقوم الغضب الشعبي العارم على هذه الممارسات والتجاهل الصارخ للحقوق الأساسية مثل الخصوصية، على إبطاء أو إعادة تشكيل طريقة التعامل مع البيانات الكبيرة في هذا الجانب من الحكم، ولا لثني الخطابات التي استخدمت لتبريرها.

إدارة المنظمات (Managing Organisations):

تماماً كما تغيرت الدول وتفرعاتها الكثيرة في الحكم كالإدارات، السلطات المحلية، الوكالات، الجيش، مقدمي الرعاية الصحية... إلخ مع مرور الوقت، واعتمادها ممارسات وتقنيات إدارية جديدة لإعادة تشكيل الكيفية التي يجري بها تنظيمها وتشغيلها، بالمثل، قامت شركات الأعمال والكيانات المؤسسية الأخرى، مثل المنظمات غير الحكومية والهيئات التطوعية والمجتمعية، والجمعيات الخيرية، والأحزاب السياسية... إلخ. فالبيانات المتعلقة بمهامها الأساسية تشكل أحد الجوانب الأساسية لكيفية فهم هذه المنظمات لنفسها ودوائرها الانتخابية، وكيفية اتخاذ القرارات بشأن عملياتها الحالية والمستقبلية. كما أن البيانات تقدم الإثبات والدليل للإنتاج المعرفي الذي يستخدم لإدارة المنظمة بشكل أكثر كفاءة، وفعالية، وتنافسية، وإنتاجية، وتحديد المنتجات والأسواق الجديدة (انظر القسم التالي).

وهكذا، أصبحت نظم المعلومات على مدى السنوات الستين الماضية من البنى التحتية الأساسية لجميع المنظمات على اختلاف أحجامها، مما يمكنها من تتبع وإدارة المجاميع البشرية المعقدة كالموظفين، الداعمين، المانحين، والعملاء، والمكونات، والسلع الأساسية، والبنى التحتية عبر الزمان والمكان، وتمكينها من توسيع نطاق عملياتها من المستوى المحلي إلى المستوى العالمي (Castells 1996; Dicken 2003). فنظم المعلومات تتيح تشغيل المنظمات بشكل أكثر ذكاءً، وإنتاج وفورات كبيرة من تكاليف قاعدتها التشغيلية. وهناك ثلاثة من هذه النظم المعلوماتية المبنية على البيانات والتي تم اعتمادها على نطاق واسع: نظم تخطيط موارد المنظمات (ERP)، ونظم إدارة سلسلة التوريد (SCM)، ونظم إدارة علاقات العملاء (CRM). وتسهّل هذه الأنظمة قدرًا أكبر من التنسيق والضبط داخل المؤسسة، ومع المنظمات الأخرى، والعملاء. على سبيل المثال، ينظم ويجمع نظام تخطيط موارد المنظمات العديد من قواعد البيانات والأنظمة المتعلقة بالشراء، التخزين، الجرد، النقل، التسويق، المحاسبة، إدارة شؤون الموظفين وإعداد قوائم المرشحين، إدارة المشاريع، والعلاقات مع العملاء في نظام واحد شامل يضمن بصورة سهلة وتلقائية توافر البيانات والعمليات من جزء واحد من الأعمال لجزء آخر (Dery et al. 2006; Grant et al. 2006). كما تُستخدم نظم إدارة سلسلة التوريد (SCM) بأكثر قدر من الكفاءة لتنظيم المشتريات، والحركة، وإدارة وتخزين المواد من الموردين من الشركة وحتى

الزبائن (Chopra and Meindl 2012). في حين تسعى نظم إدارة علاقات العملاء (CRM) إلى بناء علاقات شخصية مع العملاء من خلال جمعها أنماط الحجوزات والمشتریات السابقة لهم، وعرض التفضيلات والخدمات الممكنة الاستفادة منها خلال زيارتهم، والاستفادة منها في تقديم مكافآت الولاء، والعروض الخاصة، والترقيات، وتقديم التوجيهات للموظفين والأنظمة ليظهروا بمظهر العارف بالعميل (Sigala 2005). وقد أكد (Brynjolfsson et al. 2011) من خلال دراسة شملت 179 شركة من شركات التداول العام الكبيرة على اعتماد هذه الشركات في اتخاذ القرارات المبينة على البيانات (Data Driven Decision- DDD) حيث يكون الناتج والإنتاجية أعلى بما يُراوح بين 5-7% عما هو متوقع عند الاعتماد على استثماراتهم الأخرى وطرق استخدام تقنية المعلومات.

ويُنظر إلى البيانات الكبيرة من قبل العديد على أنها المرحلة التالية في العملية الجارية التي تسعى إلى الكفاءة التنظيمية والبنى التي تعزز الميزة التنافسية، مع الحد من المخاطر، والتكاليف، والخسائر التشغيلية، وتحسين تجربة العملاء. فالذي توفره البيانات الكبيرة هو الإمكانية الشاملة والمفصلة لغاية إدراك جميع جوانب عمل المؤسسة بشكل آني ولحظي. وكما افترضت جماعة الضغط، مؤسسة تك أمريكا (TechAmerica Foundation 2012: 12)، فإن البيانات الكبيرة ستضع الأساس: لاستبدال أو دعم عمليات اتخاذ القرار البشرية بخوارزميات آلية، مما يعمل على الحد من أوجه القصور داخل المنظمات؛ ويوجد الشفافية؛ ويحسن الأداء من خلال تمكين التجريب لاكتشاف الحاجات وإظهار التباينات؛ وتحسين العائد عن الاستثمار (ROI) لاستثمارات تقنية المعلومات، وتحسين عملية اتخاذ القرار والتشغيل الذي؛ وتوفير قدرات تنبؤية لتحسين مخرجات المهام؛ والحد من التهديدات الأمنية والجريمة؛ والقضاء على الهدر، والاحتيايل، وإساءة الاستخدام والمعاملة؛ وابتكار نماذج أعمال وخدمات مستفيدين جديدة.

وبعبارة أخرى، ستوفر البيانات الكبيرة معرفة لحظية غاية في الدقة المفصلة لكبار المديرين في جميع مناحي منظماتهم والتي يمكنهم من خلالها صياغة دورات عمل مُحكمة. وستمكن البيانات الكبيرة المنظمات من أن تكون أكثر فطنة، وملائمة، ومرنة، ومبتكرة، وذكية في كيفية تنظيمها وتشغيلها (Manyika et al. 2011).

ومثال على كيفية التوظيف الناجح للبيانات الكبيرة من قبل منظمة غير تجارية لإعادة هيكلة وإدارة عملياتها الحملة الانتخابية لباراك أوباما عن الحزب الديمقراطي. فكما أشرنا في الفصل الرابع، فقد تبنت الحملة الانتخابية لأوباما البيانات الكبيرة، من خلال جمع مجموعة هائلة من البيانات من مصادر متعددة وإنتاج بيانات اقتراع معتبرة لحظياً في الوقت المناسب، من أجل إدارة رد الفعل، وتشكيل الإستراتيجية، والإجراءات، ومهام العاملين والمتطوعين. واستناداً إلى تحليل الأداء عام ٢٠٠٨، فقد نمت مجموعة تحليلات البيانات للحملة الانتخابية لأوباما للعام ٢٠١٢ في الحجم خمسة أضعاف عنها في حملة ٢٠٠٨ متضمنة رواداً تقنيين جرى التعاقد معهم من سوق الصناعة (Scherer 2012). وقد حسنت هذه المجموعة العلائقية بين البيانات التي تم جمعها من مصادر مختلفة بحيث أمكن ربطها معاً بشكل أكثر فعالية. كما طوّروا تطبيقات للحملة الانتخابية، وجرى استخدام شبكات التواصل الاجتماعي مثل فيسبوك للدفع باتجاه التسجيل للانتخاب والحصول على التصويت لصالحها، وتثبيت ملفات تعريف الارتباط (Cookies) الخاصة بهم وتلك المتعلقة بأطراف تم التعاقد معها على أجهزة الزائرين لموقعهم الإلكتروني لجمع البيانات عنهم وتتبع عاداتهم على الشبكة (Crovitz 2012; Kaye 2012). كما قاموا بتحسين تصنيفاتهم التعريفية وتوقعات النماذج التنبؤية وكيفية استخدام المعلومات الناتجة عن تحليلاتهم لتوجيه الحملة، بالإضافة إلى اختبار وصقل طرق زيادة التمويل لدعم الحملة مالياً (Scherer 2012). كما أضافوا وعالجوا البيانات الجديدة بشكل مستمر مع تشغيل لأنظمة المحاكاة للتنبؤ بالنتائج وأفضل الردود. وكما صرح أحد مسؤولي الحملة: "لقد شغلنا نموذج المحاكاة ٦٦,٠٠٠ ألف مرة في كل ليلة" لتحديد احتمالات الفوز في كل ولاية متأرجحة". وفي كل صباح يصل إلينا الناتج، هذه هي فرص نجاحكم في الفوز في هذه الولايات، وبهذه الطريقة تم تخصيص الموارد (مقتبس عن Scherer 2012). لقد حصل فريق أوباما على مزايا مهمة عن منافسيهم من خلال التشغيل المستمر لمجموعات قواعد البيانات المتغيرة من خلال خوارزميات متطورة، وعليه أشار (2012: 246) Issenberg إلى أن حملات أوباما الانتخابية كانت "المجموعة السياسية المثالية: المؤسسة الممولة جيداً، القائمة على البيانات، والممارسة التجريبية الصارمة"، وهذا هو وعد الدعاة للبيانات الكبيرة بأن الفوائد العائدة عنها كالتى اكتسبها

الحزب الديموقراطي من البيانات الكبيرة، لها أن تتحقق لجميع المنظمات وغيرها من شركات قطاع الأعمال.

تحصيل القيمة وإنتاج رأس المال (Leveraging Value and Producing Capital):

فضلاً عن جعل الشركات أكثر كفاءة وفعالية في إدارتها، يتم الترويج بكثافة للبيانات الكبيرة لتكون وسيلة لرفع مستوى القيمة الناتجة من أنشطة الشركة وعلاقتها مع الموردين والمستهلكين. وفي الواقع، وكما أشار (Gantz and Reinsel (2011)، فإن المحرك الرئيسي للبيانات الكبيرة ليس تقنياً؛ ولكنه مالي مع الوعد بالمزيد من الكفاءة والأرباح. ومن ثم فإن اعتماد حلول البيانات الكبيرة هو إستراتيجية لإدراك تراكم رأس المال غير المستغل بطريقة تقدّم عائداً كبيراً من الاستثمار وتوفير ميزة تنافسية (Manyika et al. 2011; Zikopoulos et al. 2012). وهناك العديد من الطرق التي يمكن من خلالها لحلول البيانات الكبيرة أن تُضفي الذكاء للشركات وبطريقة يمكن معها أن تنمو مبيعاتها وأرباحها. فمع الأخذ لقطاع البيع بالتجزئة كمثال، حدّد (Manyika et al. (2011 ستة عشر حلاً مقسّمة إلى خمسة أنواع - التسويق، والترويج، والعمليات، وسلسلة التوريد، و نماذج الأعمال الجديدة - يمكن لكل منها تحصيل مكاسب محتملة في مجال المبيعات والكفاءة (انظر الجدول ٧-١). وهي تشمل إنتاج بيانات عبر الشركة وتطبيق آليات التحليل من أجل تقسيم السوق، والتصدي لعزوف العملاء والموظفين، وتحسين المدخلات المختلفة (على سبيل المثال، المكونات، والعمالة، والمرافق) وتؤدي من ثم إلى بناء مختلف التشكيلات والنماذج التنبئية للإجابة عن مجموعة متنوعة من الأسئلة من قبيل:

- هل يتم الاتصال بالعميل أم لا (التسويق المستهدف)؟
- هل يتم تقديم عروض للاحتفاظ بالعميل أم لا (المحافظة على العملاء)؟
- أي نوع من الإعلان أو اختيار الكلمات / الصور أو المنتج تقدم للعميل (انتقاء المحتوى)؟
- أي قناة ينبغي الاتصال من خلالها بالعميل (اختيار القناة)؟
- هل يتم تقديم عرض أعلى أو أقل سعراً للعميل (التسعير/الخصم المتغير)؟
- هل يتم شطب جزء أكبر من الدين عن المقترض (التحصيل)؟

- هل يتم تقديم حد ائتماني أعلى أو أدنى أو معدل فائدة للعميل (مخاطر الائتمان)؟ (Siegel 2013).

وللتركيز على نوع واحد من أنواع الحلول الخمسة التي قدّمها (Manyika et al. 2011)، وهو التسويق، يمكن استخدام عدد من آليات التحليل وأدوات إدارة البيانات الكبيرة الموزعة في أربعة محاور هي: التوحيد والاختزال، التقسيم والتجزئة، المطابقة، والتحسين، وذلك لإدارة ومعالجة وتحليل وتفسير البيانات من أجل تلبية الأهداف (انظر شكل ١ - ٧). إن هدف التسويق السلوكي هو بيع البضائع والخدمات بصورة أكثر فعالية للأفراد والعملاء وبناء علاقة طويلة المدى معهم من أجل الحفاظ على عاداتهم وذلك من خلال الانسجام مع خصائصهم، وسلوكهم، وميولهم، ورغباتهم (Schwartz and Solove 2011). فمن خلال بناء الملفات التعريفية الفردية للأشخاص ودمجها مع التحليلات التنبؤية، تسعى الشركات إلى تقديم تجربة شخصية من خلال التسويق الجزئي للمنتجات والعروض الخاصة. وفي سبيل ذلك، تسعى الشركات إلى توقع نية العميل في الشراء وتسهيل تحقيقهم لهذه النية بلطف، مع السعي لبناء ولاء العملاء في الوقت نفسه (Minelli et al. 2013). وفي ذات المعنى، زعم Bill Stensrud أن "الموزعين يستخدمون البيانات الكبيرة في محاولة للحصول على المستهلكين" (مقتبس في 24: Bollier 2010).

وقد تمّ بالفعل تنفيذ مثل هذا النهج للبيانات الكبيرة من قبل عدد من تجار التجزئة مثل وول مارت وتيسكو. حيث تنتج مثل هذه السلاسل من المتاجر كميات هائلة من البيانات يومياً من خلال نظم تخطيط موارد المنظمات (ERP)، وإدارة سلسلة التوريد (SCM)، وإدارة علاقات العملاء (CRM) الخاصة بها، بما في ذلك ربط منافذ المبيعات بالزبائن من خلال بطاقات الولاء والائتمان (في حالة وول مارت هناك ما يقارب ٢٦٧ مليون معاملة يومياً)، واستخدامها لوضع إستراتيجيات أفضل للأسعار، ومخططات التخزين، وإعداد قوائم الموظفين، ومراقبة المخزون، والحملات الإعلانية، وتكاليف المورد، وتحديد مواقع جديدة للتخزين والمنتجات... إلخ، واقتطاع حصة من سوق منافسيها عند القيام بذلك (Bollier 2010). وعضواً عن إجراء مسح عينية ضيقة ودراسات استقصائية باهظة التكلفة، لاستطلاع آراء المشاركين عن ماذا سيفعلون، وما هي القيمة المعتمدة بنظرهم، تقوم

البيانات الناتجة عن العمليات والتصفح على الشبكة بالكشف عن هذه المعلومات عبر جميع المستهلكين. فالفوائد المتوقعة مرتفعة، حيث يشير (Manyika et al. (2011 إلى أن البيانات الكبيرة يمكن أن تزيد هوامش التشغيل لتجار التجزئة بنسبة ٦٠ في المائة، وفي حين أن هذه الحلول المتعلقة بمبيعات التجزئة والتسويق من الممكن أن تتوفر في مختلف القطاعات، فإن درجة بروزها يختلف من قطاع لآخر بين مختلف القطاعات الأكثر استفادة من حلول معينة (انظر الجدول ٧-٢).

الجدول رقم (١ - ٧)

أهماط ذكاء البيانات الكبيرة للشركات

<p>البيع العابر للحدود (باستخدام المواقع الجغرافية للزبائن، وتاريخ الشراء، والتفضيلات، والمعطيات الأخرى لزيادة متوسط حجم الشراء)</p> <p>التسويق القائم على الموقع (باستخدام الموقع الحالي بشكل آني أو لحظي، على سبيل المثال نظام تحديد المواقع الجغرافية في الهواتف الذكية، واستهداف الزبائن الموجودين في المتاجر أو بالقرب منها)</p> <p>تحليل السلوك داخل المتاجر (تتبع الهواتف الذكية، وسلال التسوق، والمراقبة المرئية للمساعدة في تحسين مخطط توزيع المتجر، وتنوع المنتجات، ومواضع الأرفف لزيادة المبيعات)</p> <p>التقسيم الجزئي للعملاء (استخدام تاريخ الشراء والتصفح لاستهداف العملاء بشكل فردي بالإعلانات والعروض الخاصة)</p> <p>التحليل العاطفي (تتبع ورصد آراء العملاء على مواقع التواصل الاجتماعي لتحديد ميلهم تجاه منتج ما وتعديل الحملات التسويقية بشكل مناسب)</p> <p>تعزيز تجربة العملاء مع تعدد القنوات (إنشاء تجارب وفهارس على شبكة الإنترنت مصممة بشكل فردي بحسب العملاء)</p>	<p>التسويق (Marketing)</p>
---	--------------------------------

<p>تحسين التشكيلة (تخصيص البضائع لحفظها في أيّ من المتاجر استناداً إلى المواقع الجغرافية المحلية، وانطباع المشترين وسلوكهم، والحقائق المحلية الأخرى)</p> <p>تحسين التسعير (باستخدام السلوكيات الشرائية والجغرافية لاتخاذ قرارات التسعير في مختلف المواقع لزيادة المبيعات / الأرباح)</p> <p>تحسين التموضع والتصميم (باستخدام السلوك داخل المتجر والتفاعلات على الموقع الإلكتروني لتوجيه عملية تصميم الموقع / المتجر وأماكن وضع البضائع لزيادة المبيعات)</p>	<p>التبضع (Merchandising)</p>
<p>شفافية الأداء (التحليل المفصل، بشكل آني أو لحظي على نحو متزايد، لتحديد مدى أداء المنتجات، والمتاجر، والموظفين)</p> <p>تحسين مدخلات الموظفين (احتساب المستويات المثلى للتوظيف وجدولة المناوبات لتفادي فرط التوظيف أو شح القوى العاملة)</p>	<p>العمليات (Operations)</p>
<p>إدارة المخزون (تحسين توقع المخزون من خلال رصد تواريخ المبيعات، ودورات المبيعات الفصلية، والأحداث والظروف المحلية لضمان مستويات مناسبة من المخزون المطلوب)</p> <p>تحسين التوزيع والدعم اللوجستي (تحديد مواقع البضائع بشكل آني أو لحظي، واحتساب مسارات الطرق الأكثر ملاءمة وكفاءة من حيث الوقت والوقود، ومراقبة سلوك السائقين، وجدولة أعمال الصيانة)</p> <p>مفاوضات الموردين (باستخدام البيانات المتعلقة بتفضيلات المستهلكين ومشترياتهم لإبلاغ مزودي المنتجات)</p>	<p>سلسلة التوريد (Supply Chain)</p>
<p>خدمات مقارنة الأسعار (تقديم خدمات مقارنات شفافة من أطراف ثالثة للزبائن لأسعار المنتجات بشكل آني أو لحظي عبر مختلف تجار التجزئة)</p> <p>الأسواق القائمة على الويب (تجار التجزئة الذين يقدمون خدمات بقيمة مضافة مثل الوصول إلى الباعة الآخرين، ومعلومات المنتجات، والمراجعات الناتجة عن المستهلكين، ومقارنات الأسعار)</p>	<p>نماذج الأعمال الجديدة (New Business Models)</p>

المصدر : جُمعت من 67-71. Manyika et al.2011.

مقتبس من "البيانات الكبيرة: آفاق جديدة للابتكار والمنافسة والإنتاجية"، (معهد ماكينزي العالمي Mckinsey Global Institute (May 2011)
(www.mckinsey.com/mgi.Mckinsey&Company).

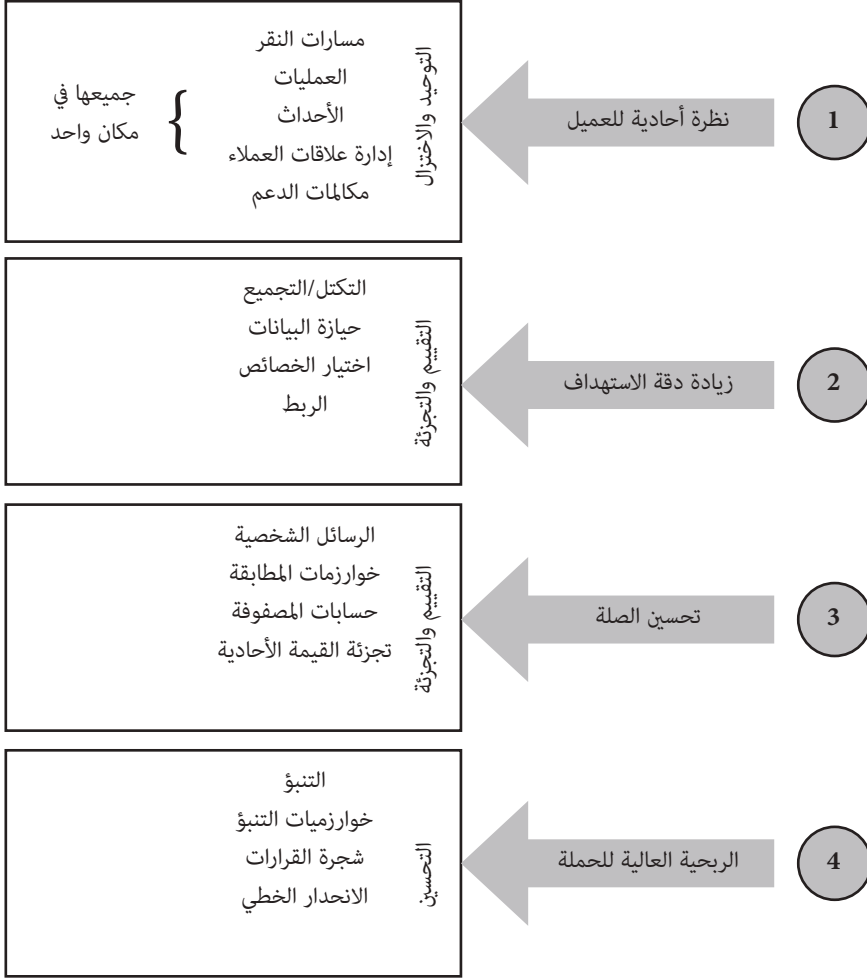
وإضافة إلى الدفع بالفوائد المتوقعة، سعى دعاة البيانات الكبيرة أيضاً إلى دعم اعتمادها من خلال تسليط الضوء على عواقب فشل الاستثمار فيها. على سبيل المثال، أشار Manyika et al. (2011: 6) إلى أنهم: "يتوقعون أن يرون، عبر القطاعات، تحقق القيمة التي تعود على المستخدمين الرئيسيين للبيانات الكبيرة على حساب المتقاعسين عنها"، وبعبارة أخرى، سوف يتخلف أولئك الذين لا يستطيعون التأقلم واعتماد حلول البيانات الكبيرة عن الركب ويتعرضون للتعثّر والأفول؛ لأنهم سيفقدون حصتهم في السوق. وبغض النظر عما إذا كان استخدام الخطابات للدفع أو الجذب، فقد أضحت البيانات الكبيرة للأعمال منتشرة داخل التيار الرئيسي ووسائل الإعلام التجارية منذ العام 2012م، وبالنظر لفوائدها الموصوفة، يبدو أن البدء في تطبيقها على نطاق واسع سيمثل المرحلة التالية للابتكار التقني وذلك بصورة مؤكدة.

الشكل رقم (١ - ٧)

التسويق والبيانات الكبيرة

قدرة المنصة

أهداف المسوقين



المصدر: مقتبس من Claverie-Berge 2012.

تم إعادة الطباعة بإذن من المؤسسة الدولية لآلات الأعمال، © المؤسسة الدولية لآلات الأعمال (International Business Machines Corporation).

الجدول رقم (٢ - ٧)

فوائد البيانات الكبيرة لعشر صناعات مختارة

التأمين	صناعة السيارات
تحديد التصرف القدام الأمثل نمذجة الكوارث الغش في المطالبات	المراقبة المتقدمة للحالة تجميع وتسعير السيارات الجديدة مطالبات الضمان
النفط والغاز	الخدمات المصرفية
المراقبة المتقدمة للحالة مراقبة وتحسين الحفر مراقبة وتحسين الإنتاج	تحديد التصرف القدام الأمثل تحليل سجل النظام إدارة مخاطر السيولة
التجزئة	الطاقة والخدمات
التحسين الآلي للأسعار التشكيل المحلي تحسين التسويق	تحليلات العداد الذكي التنبؤ بتوزيع الأحمال وجدولتها الصيانة بحسب الحالة
الاتصالات	الحكومة
ضمان الخدمة الأفضل ذكاء الشبكة استحداث منتجات وخدمات جديدة	منع الجرائم والتهديدات والتنبؤ بها تحليلات الاحتيال والنفائات وإساءة المعاملة
السفر والنقل	الرعاية الصحية
تحليلات العملاء وولائهم التسويق الصيانة التنبؤية	التدبير والتصرف بناء على الصحة العامة للمجتمع إشراك المستهلكين في رعايتهم الصحية

المصدر: أي بي إم (IBM (n.d))

<http://www-01.ibm.com/software/data/bigdata/industry.html>

إيجاد أماكن أفضل:

اخترع الناس واعتمدوا على مر التاريخ على تقنيات جديدة صممت لإنشاء الثروة وتحسين الحياة اليومية وتعزيزها. وقد أثرت العديد من هذه التقنيات بشكل كبير على طبيعة وهيكل وتنظيم وبنّ الحياة المدنية والريفية، في الخدمات الخاصة المُقدمة، والنقل، والاتصالات، وبناء البنية التحتية. فهذه التقنيات تحمل في كثير من الأحيان تأثيرات متناقضة وسلبية على سبيل المثال، المصانع التي صنعت الثروات، وفي الوقت نفسه التلوث البيئي أيضاً، أو أن تكون مبالغاً فيها بحيث تتجاوز الطاقة الاستيعابية على سبيل المثال، الاكتظاظ أو الازدحام المروري، أو تنتج التبعيات التي لا يُمكن أن تُطاق على سبيل المثال، المسافات الطويلة بين مواقع السكن ومواقع العمل. إن تقنيات المعلومات والاتصالات (تكنولوجيا المعلومات والاتصالات) هي أحدث موجة من الأجهزة والبنى التحتية الجديدة التي أضحت جزءاً لا يتجزأ في نسيج البيئة التي نعيش بها، ومعها مكوّن مهمّ من مكونات منطقتها الاستطراذي وهو أنها تعمل على توفير الحلول للمشاكل الناجمة عن الدورات السابقة من التنمية التي قادتها التقنية - بحيث تجعل منها أكثر كفاءة واستدامة، ونظافة، وأمنًا، وإنتاجية... إلخ - وكذلك إنشاء طرق جديدة تمامًا للعمل في العالم المحيط. ونتيجة لذلك، تبنت المدن تقنية المعلومات والاتصالات، على مدى العقود القليلة الماضية، باعتبارها عنصراً أساسياً في إستراتيجياتها الإنمائية (Castells 1996; Graham and Marvin 1996; Kitchin 1998).

إن مفهوم المدن الذكية هو تجسيد لهذه الرؤية التي تقودها تقنيات المعلومات والاتصالات للتنمية المدنية، والذي يصف الأماكن التي تتكون بشكل متزايد وتُراقب بواسطة الحوسبة المنتشرة والحوسبة في كل مكان، ويقود ضبط اقتصادها وإدارتها الإبداع، والابتكار، وزيادة الأعمال، من قبل أناس أذكاء. وهذه المدن الذكية مجهزة بشكل مكثف ويمكن فهمها وضبطها بشكل آني أو لحظي؛ فهي تنتج، وتتبادل، وتدمج، وتستهلك، وتعمل على البيانات الكبيرة (Kitchin 2014). إن مثل هذه البيانات الكبيرة توفر فهماً شاملاً، ومتغيراً، ومفصلاً بدقة لهذه المدن والمدخلات للأنظمة التي من شأنها بناء أماكن أكثر ملاءمة للعيش، وأمنة، ووظيفية، وتنافسية، ومستدامة (Hancke et al. 2013; Townsend 2013). فمثل

هذه الرؤية للمدن الذكية يجري الترويج لها بشكل كبير من قبل عدد من أكبر مزودي خدمات البرمجيات في العالم وشركات الأجهزة الرقمية (على سبيل المثال، سيسكو، وآي بي أم، ومايكروسوفت، وإنتل، وسيمنز، وأوراكل، وساب) والتي اعتمدت بحماس من قبل المؤسسات البلدية والوطنية والعبارة للحدود التي تتوقع لتقنيات المدن الذكية أن تُنتج تقدم اجتماعي-اقتصادي وتجدد المراكز المدنية لتصبح مراكز للابتكار والعمل (Kourtiti et al. 2012). وفي حين تمّ بناء بعض مشاريع المدن الذكية من الألف إلى الياء (على سبيل المثال، سونغدو (Songdo)، أو مدينة مصدر (Masdar City))، فإن معظم هذه المشاريع تدريجية، وتتألف من إعادة تجهيز للبنية التحتية القائمة مع تقنيات رقمية وحلول للبيانات؛ لأن الوظيفة الرئيسية للبيانات الكبيرة في كلتا الحالتين هي توفير تحليلات بشكل آلي أو لحظي لإدارة كيفية جوانب وظيفة المدينة وتنظيمها.

لقد استخدمت هذه المراقبة وآليات تحليل البيانات الآنية واللحظية لعدد من السنوات في بعض القطاعات. على سبيل المثال، قيام العديد من المدن ببناء أنظمة النقل الذكية، والتي تستخدم البيانات الناتجة عن شبكة من الكاميرات والنواقل والمتعلقة بحركة المرور حول نظام ما لتقديم تغذية راجعة لمركز تحكم مركزي، وتستخدم لمراقبة وتنظيم التدفق، وتعديل تسلسل حركة المرور، وحدود السرعة المسموح بها، وإدارة العقوبات المتعلقة بالمخالفات المرورية بشكل تلقائي (Dodge and Kitchin 2007a). وبالمثل، فقد ترصد وتراقب الشرطة مجموعة من الكاميرات وسجلات الحوادث الحية من أجل توجيه الموارد المناسبة لمواقع معينة للاستجابة السريعة وبكفاءة. كما يمكن جمع البيانات المتعلقة بالظروف البيئية من خلال شبكة استشعار موزعة في جميع أنحاء المدينة، على سبيل المثال، قياس وتتبع تلوث الهواء، ومستويات المياه، أو النشاط الزلزالي. وتستخدم العديد من الحكومات المحلية نظم إدارية لتسجيل مشاركات الجمهور وتفاعله مع خدماتها. وفي جميع الحالات تقريباً، كانت هذه الأنظمة معزولة عن بعضها البعض وتتعامل مع قضية واحدة ويتم التحكم بها من قبل وكالة واحدة.

وفي الآونة الأخيرة، كانت هناك محاولة لجمع كل هذه الأنواع من المراقبة والتحليل في مركز واحد، وأن تستكمل بآليات تحليل عامة ومفتوحة على نطاق أوسع. على سبيل المثال،

قام مركز العمليات في ريودي جانيرو، البرازيل، والذي تم إنشائه بالشراكة بين حكومة المدينة وشركة آي بي أم، بتجهيز نظام على مستوى كافة أنحاء المدينة حيث يجمع البيانات من ثلاثين وكالة معاً في مركز موحد لتحليلات البيانات، بما في ذلك حركة المرور والنقل العام، والبلدية وخدمات المرافق، وخدمات الطوارئ، ونقاط جمع بيانات الطقس، والمعلومات المرسلة من قبل الموظفين والجمهور عبر الهاتف، والإنترنت، والراديو (انظر الشكل 7.2، Singer 2012c). هنا، تعمل الخوارزميات وفريق من المحللين على معالجة، وتصوّر، وتحليل، ورصد كمية هائلة من بيانات الخدمة الحية، جنباً إلى جنب مع البيانات المجمعة مع مرور الزمن وكميات ضخمة من البيانات الإدارية التي تم إصدارها على أسس دورية بصورة أكبر، وغالباً ما يتم المزج بين مجموعات البيانات معا للتحقيق في جوانب معينة من حياة المدينة والتغيرات الحاصلة فيها مع مرور الوقت، وبناء النماذج التنبئية فيما يتعلق بالتنمية والإدارة اليومية للمدينة وإدارة حالات الكوارث كالفيضانات. ويكمل هذا عن طريق منصة العمليات الافتراضية التي تمكن المسؤولين في المدينة من تسجيل الدخول من الميدان والوصول إلى المعلومات بشكل آني أو لحظي. فعلى سبيل المثال، يمكن للشرطة في مشهد الحادث المروري استخدام المنصة لمعرفة عدد سيارات الإسعاف المرسلة للموقع وموعد وصولها، وتحميل معلومات إضافية على الموقع (Singer 2012c). وكان الهدف المعلن من عمدة المدينة، إدوارد بايس (Eduardo Paes)، هو إسقاط الحواجز بين الإدارات وضمّ بيانات كل منها إلى الآخر لمساعدة المنظومة بأكملها (Singer 2012c). كما يجري تطوير مراكز مماثلة في أماكن أخرى، يصابها مجموعة من التطبيقات للمواطنين للوصول والاستفادة من بعض مناحي البيانات.

أما بالنسبة لأولئك العاملين على تطوير واستخدام آليات تحليل متكاملة وآنية لبيانات المدينة، فإن مثل هذه المراكز والتطبيقات توفر وسائل فعالة لاستشعار وفهم العيش في المدينة وإدارتها في الوقت الراهن، وتصوّر وتوقع السيناريوهات المستقبلية، وإجراء توزيع الموارد بالاعتماد على الدليل وتسهيل التفكير المترابط عبر الحكومة (Flowers 2013; Kitchin 2014). وبدلاً من اتخاذ القرارات استناداً إلى الكلام المتداول، أو الحدس والظن، أو المحسوبة، أو أدلة دورية / جزئية، فمن الممكن تقييم ما يحدث في أي وقت، والرد والتخطيط المناسب له.

الشكل رقم (٢ - ٧)

مركز العمليات لوسط مدينة ريودي جانيرو



المصدر : <http://ipprio.rio.rj.gov.br/centro-de-operacoes-rio-usa-mapas-feitos-pelo-ipp>

وعلاوة على ذلك، يوفر استخدام عينات كبيرة وربط أشكال متنوعة من البيانات تحليلاً أكثر عمقا، وشمولية، وقوة. وقد أصبح من الممكن للدعاة لمثل هذه الأنظمة تطوير، وتشغيل، وتنظيم، وخوض غمار الحياة في المدينة على أساس قوي، ودليل منطقي غير ضعيف، لا دليل انتقائي مبني على أسس إيديولوجية سياسية (Flowers 2013). وهكذا، فثمة من يقول، إن استخدام هذه البيانات الكبيرة سيوفر الأساس لمدينة أكثر فعالية، واستدامة، وتنافسية، وإنتاجية، وانفتاحا، وشفافية. إلا أن آخرين يحذرون من أن البيانات الكبيرة تشير إلى احتمالات الوصاية "Big Brother"، وأشكال الحكم القائم على التقنية (Technocratic)، وخصخصة (Corporatisation) إدارة المدينة (انظر الفصل العاشر).

الخلاصة:

رسم هذا الفصل الدوافع الدعائية للبيانات الكبيرة فيما يتعلق بأربع مهام رئيسية هي: حكم الشعب، وإدارة المنظمات، وتعزيز القيمة وإنتاج رأس المال، وإيجاد أماكن أفضل للحياة (حيث سيجري مناقشة إمكاناتها بالنسبة للعلوم في الفصل التالي). وقد بينت المناقشة الكيفية التي يتم بها تأطير البيانات الكبيرة سياسياً واقتصادياً (انظر الفصل الأول) وكيف يتم وضع مجموعة قوية من البيئات والحجج العقلانية لدعم البدء في تنفيذ واعتماد تقنيات البيانات الكبيرة وحلولها. وتتضح قوة الأنظمة الخطائية التي يجري بناؤها من خلال النظر في الحجج المضادة - حيث من الصعب أن نقول إننا نريد أن نبقى في حالة أقل تبصراً وحكمة، وإنتاجية، وتنافسية، وكفاءة، وفعالية، واستدامة، وأمنًا... إلخ. فإذا كانت البيانات الكبيرة تقدم كل هذه الفوائد، فإن دعاة البيانات الكبيرة يجدون من غير المنطقي عدم المضي قدماً في تطوير نظم البيانات الكبيرة ودعمها. وبطبيعة الحال، فإن الحجج التي قدمت هي ضيقة وانتقائية وتبتعد عمداً عن تسليط الضوء على الآثار السلبية المحتملة فيما يتعلق بالحريات المدنية، ومراقبة البيانات، والفرز الاجتماعي، وأمن البيانات، وانفلات زمام التحكم وفرض السيطرة، والحكم الاستباقي، والحكم التكنوقراطي (القائم على التقنية) والخصخصة، وغلق التقنيات وعدم إتاحتها

بصورة أكبر (انظر الفصل العاشر). فهذه هي وجهة نظر أصحاب المصالح الخاصة، ولا سيما أولئك الذين يسعون لبيع تقنيات البيانات الكبيرة، والحكومات التي تسعى نحو رؤية الليبرالية الجديدة للحكم والتنظيم (الليبرالية الجديدة: هي فكر أيديولوجي يمثل تأييد الرأسمالية المطلقة وعدم تدخل الدولة في الاقتصاد.)، وليست وجهة نظر المواطنين أو المجتمعات التي من الممكن أنها لا تزال من دعاة البيانات الكبيرة والحوسبة في كل مكان، ولكنهم يتصورون أن استخدامهما سيكون بطرق تحريرية، وتمكينيه، وتشاركية مع إخضاع التأثيرات الأكثر سلبية للرقابة والتنظيم.

لقد تميّز التحليل المعروض في هذا الفصل بالإيجاز، مع توفير لمحة شاملة لبعض الحجج التي تطرح. وما هو مطلوب في الوقت الحاضر، من خلال دراسات الحالة المحددة، هو تفصيل وتمحيص للأنظمة الخطابية التي تتكشف ويجري بناؤها. على سبيل المثال، فيما يتعلق بالترويج للبيانات الكبيرة في المدن الكبيرة، فهناك حاجة لرسم وتجزئة الكيفية التي تروّج بها مختلف الخطابات الداعمة فيما يتعلق بأماكن محددة من خلال المجالات التجارية والمواقع الإلكترونية، والصحف اليومية، والمواد الترويجية للشركات، وسجلات الدولة، والبيانات السياسية... إلخ، وتقاطعات الأجناس المختلفة ومصالح المعنيين. وهذا يحتاج إضافة إلى تحليل الكيفية التي يتم بها التصدي لمثل هذه الخطابات برؤى بديلة، وتحليل المعارضة والتداخل بين الحجج الخطابية وتتبع تسلسل الكيفية التي تحولت بها مع مرور الوقت، والتحليل المقارن للمشهد الخطابي عبر مختلف البقاع، أن يصاحبه تحليل لآثار الخطابات على تشكيل الكيفية التي يعمل بها المجال في الممارسة والنتائج المترتبة على تنفيذ حل البيانات الكبيرة.

وبالنسبة فيما يتعلق بالنقطة الأخيرة، بين (Salmon 2014) بالتفصيل أنه مجرد أن يتم تطبيق حل قائم على البيانات فإنه من النادر أن يحقق الفوائد التي توقعها المتحمسون والداعمون لذلك الحل، وغالبًا ما ينتج عن الحل مجموعة من القضايا الجديدة. وقال إنه توجّه لقانون كامبل (Campbell's Law)، الذي صاغه عالم الاجتماع دونالد كامبل (Donald T. Campbell 1976)، ليشرح هذه الظاهرة. فقانون كامبل ينص على: "إنه كلما كثر استخدام أي مؤشر اجتماعي كمي لصنع القرار الاجتماعي، كلما كان الموضوع

تحت ضغوط الفساد وأكثر عرضة للتشويه وإفساد العمليات الاجتماعية التي يهدف إلى رصدها." أو بحسب صياغة سلمون نفسه: "كلما جرى تشغيل حقل بنظام ما بشكل أكبر، أوجد هذا النظام المزيد من الدوافع للجميع (الموظفين، والعملاء، والمنافسين) لتغيير سلوكهم بطرق منحرفة ضارة - مما يوفر أكثر من الشيء الذي تم تصميم النظام لقياسه وإنتاجه، سواء كان لذلك أي قيمة في الواقع أم لا".

وبعبارة أخرى، يبدأ الناس في اللعب على النظام بعقلانية واعتبار المصلحة الذاتية، ولكن في كثير من الأحيان بطرق غير متوقعة لإفساد المقاييس، والخوارزميات، وعمليات صنع القرار الآلي. على سبيل المثال، أشار Porter (1995: 44)، إلى أنه "عندما يتم الحكم على مديري الأعمال من خلال حساباتهم، فإنهم يبدوون التعلّم على كيفية تحسين حساباتهم، ربما من خلال بعض الحيل من قبيل تأجيل الصيانة اللازمة وغيرها من التكاليف ذات المدى الطويل". ونتيجة لذلك، فإنه في حين يُمكن أن تتحول كيفية إدارة وتشغيل المنظمات والحكومة، لتصبح قائمة على البيانات ومقاييس الأداء، فإنه من المحتمل أن يحدث ذلك آثاراً متنوعة وغير متوقعة ويُفعل أشكالاً معينة من الحاكمية، ليس بالضرورة أن تكون إيجابية في مجملها.

ولذا، تحتاج دراسات البيانات الكبيرة وآثارها إلى الانفتاح لعرض السياسات والأجندات الكامنة للبيانات الكبيرة في مختلف السياقات وتوفير أساس لمزيد من المحادثات المعيارية حول هذا النوع من عوالم البيانات الكبيرة التي قد نرغب في العيش بها. فهذه الأنواع من المحادثات هي قيد التطوير في الوقت الحاضر، ويرجع ذلك بشكل جزئي إلى حداثة التطورات وسرعة وقوعها، وأيضاً بسبب هيمنة الأطروحات المبالغ بها عن البيانات الكبيرة ومكائد أصحاب المصلحة المعنيين بها. ومع ذلك، تظل الحاجة قائمة لمثل هذه الأحاديث بالنظر إلى أن البيانات الكبيرة تعيد تشكيل الطريقة التي يجري بها حكم المواطنين، وتنظيم المواقع، وإدارة المنظمات، وعمل الاقتصاديات، وممارسة العلوم.

۲۲۴

الفصل الثامن

إعادة صياغة بحوث العلوم، والعلوم الاجتماعية، والعلوم الإنسانية

The Reframing of Science, Social Science and (Humanities Research)

في حديث لمجلة الإيكونوميست في العام 2010 م، ذكر سنان آرال (Sinan Aral)، أستاذ إدارة الأعمال في جامعة نيويورك، أن "الثورات في العلوم غالباً ما يسبقها ثورات في القياس" (مقتبس عن Cukier 2010). إن هذا الفصل يستكشف مدى انطباق ملاحظة آرال فيما يتعلق بالبيانات الكبيرة، والبنية التحتية للبيانات، والبيانات المفتوحة، وتأثيراتها فيما يتعلق بكيفية مقارنة البحوث والاكتشافات الأكاديمية من منظور فلسفي ومنهجي. وعلى وجه الخصوص، يقوم هذا الفصل بدراسة كيفية تحدى توافر البيانات الكبيرة والبنية التحتية للبيانات إلى جانب الأدوات التحليلية الجديدة للنظريات والعلوم المعرفية القائمة في مختلف التخصصات - الكيفية التي يتم بها طرح الأسئلة وكيفية الإجابة عنها - ويؤدي إلى نشوء مجالات وتخصصات جديدة. وهكذا، يستكشف هذا الفصل ما إذا كانت التغيرات التي تطرأ على الطريقة التي يتم بها إنتاج وإدارة وتحليل المادة الخام للمعلومات والمعرفة ستؤدي إلى تغيير جذري في فهم وممارسة المبادئ الأساسية التي تقوم عليها العلوم، والعلوم الاجتماعية والإنسانية. وبعبارة أخرى، هل تقود ثورة البيانات إلى نقلة نوعية في المجالات الرئيسية للعلوم والبحوث؟

يمثل النموذج (Paradigm)، كما حدده كوهن (Kuhn 1962)، وسيلة مقبولة لاستجواب العالم وجمع المعرفة المشتركة بين نسبة كبيرة من الباحثين في اختصاص معين في لحظة زمنية معينة. ويتشارك الباحثون العاملون ضمن نطاق هذا النموذج في فلسفة واحدة، واستخدام مجموعة من الأساليب الشائعة، وطرح الأسئلة والإجابة عنها للعمل على بناء المعرفة بصورة تدريجية. ومن ثم، فإنهم يميلون إلى تفضيل التصنيفات، والمعارف، والنظريات، والأساليب،

والأطر الأخلاقية والعقائدية نفسها أو القريبة من بعضها البعض. ومع ذلك، فقد أشار كوهن، وبصورة متكررة، إلى بروز طرق جديدة للتفكير تتحدى النظريات والمناهج المقبولة والمعمول بها. على سبيل المثال، فإن النظرية النسبية لأينشتاين جابهت، وإلى حد كبير، حلت محل نظرية نيوتن في الفيزياء. وبالمثل، أحدثت نظرية التطور لداروين تغييراً جذرياً في الفكر النظري للعلوم البيولوجية، كما تحدت المذاهب الدينية في معتقدات النشوء والخلق. وفي كلتا الحالتين، فقد تمسك بعض الناس بالطرق المتبعة في التفكير وتشبثوا بها، في حين انتقل آخرون إلى الفكر الجديد، والذي سرعان ما أصبح الفكر المهيمن. وفي بعض المجالات الأكاديمية، هناك القليل من الأدلة على نماذج عاملة، لا سيما في العديد من العلوم الاجتماعية، مثل علم الاجتماع والجغرافيا البشرية، حيث تم تطبيق مجموعة متنوعة من المناهج الفلسفية على سبيل المثال، الوضعية (Positivism)، والظواهرية (Phenomenology)، والبنوية (Structuralism)، وما بعد البنوية (Poststructuralism) ... وغيرها، وقامت العديد من هذه المعسكرات النظرية بطرح آراء متعارضة بشأن أفضل السبل لفهم العالم. وفي مجالات أخرى، مثل العلوم، كان الاجتماع أكثر على وحدة المنهج العلمي المتبع، معززاً باختبار الفرضيات للتحقق منها أو نفيها. وهذا لا يعني، مع ذلك، عدم تحول أو تغيير الطريقة التي تعمل على تصوير ونشر المنهج العلمي بشكل دوري، ووجود بعض النظريات المتنافسة فيما يتعلق بشرح ظواهر معينة، فالنظريات يمكن أن تختلف حول الظواهر ولكنها تشترك في نفس النهج الأوسع للمساعي العلمية.

وقد بين جيم جراي (Jim Gray)، على سبيل المثال (كما هو مفصّل في Hey et al. 2009)، تطوّر العلوم من خلال أربعة نماذج، حيث أشار إلى أن النموذج الرابع ما زال في بداياته ولكن هو نتيجة لثورة المعلومات المنتشرة (انظر الجدول 8.1). وبخلاف طرح كوهن، تحدث التحولات في النموذج لأن النموذج السائد في العلم لا يمكنه تفسير بعض الظواهر المعينة أو الإجابة عن الأسئلة الأساسية، ومن ثم تبرز المطالبة بإعادة صياغة أفكار جديدة، فتحولات جراي تأسست على التطورات الجارية في أشكال البيانات وتطورات الأساليب التحليلية الجديدة. وسيستكشف هذا الفصل إلى أي مدى يأتي وجود النموذج الرابع في العلوم، وإلى أي مدى تقود ثورة البيانات إلى نظريات معرفية بديلة في العلوم الإنسانية والعلوم الاجتماعية. وبالنظر إلى المراحل الأولية نسبياً في ثورة البيانات المنتشرة،

فرمها لن تتفاجأ بتباين الآراء حول تأثير البيانات الكبيرة، والبيانات المفتوحة، والبنى التحتية للبيانات، وتحليلات البيانات الجديدة على إنتاج نظريات معرفية جديدة، وسيستمر وجود مثل هذه التباين في الآراء. ولكن، لا مجال للشك في أن البيانات الكبيرة، والبيانات المفتوحة، والبنى التحتية للبيانات، وتحليلات البيانات الجديدة تطرح قضايا معرفية أساسية وذلك لأن استخلاص المعلومات الصحيحة والمفيدة من طوفان البيانات لا يُعد ببساطة قضية فنية يمكن التعامل معها من خلال حلول تقنية فحسب (Floridi 2012).

الجدول رقم (١ - ٨)

نماذج العلم الأربعة

النموذج	الطبيعة	الشكل	الوقت
الأول	العلوم التجريبية	التجريبية، وصف الظواهر الطبيعية	قبل عصر النهضة
الثاني	العلوم النظرية	النمذجة والتعميم	قبل ظهور الحواسيب
الثالث	العلوم الإحصائية	محاكاة الظواهر المعقدة	قبل البيانات الكبيرة
الرابع	العلوم الاستكشافية	يتسم بكثافة البيانات، الاستكشاف الإحصائي والتنقيب عن البيانات	في الوقت الحالي

المصدر: تم التجميع من Hey et al. 2009.

النموذج الرابع للعلم (The Fourth Paradigm in Science):

في حين أن جيم جراي (Jim Gray) تصور أن النموذج الرابع في العلوم يتسم بكثافة البيانات واتساع المنهج العلمي المؤسس، إلا أن بعض العلماء الآخرين يرون أن البيانات الكبيرة أفسحت الطريق لبزوغ عصر جديد للفلسفة التجريبية (Empiricism)، حيث يمكن لحجم البيانات الذي تصحبه أساليب معينة أن يكشف عن الحقيقة الكامنة فيها، وأن يُمكن البيانات من أن تتحدث عن نفسها وهي متحررة من النظرية. وقد اكتسبت وجهة النظر الأخيرة الاعتمادية خارج المجمعات العلمية، وخاصة داخل دوائر الأعمال، ولكن أفكارها تأصلت وتجزرت في المجال الجديد لعلوم البيانات والعلوم الأخرى. فكلما المنهجين

يوفران تحدياً للمنهجية العلمية التقليدية إضافةً إلى طرح نموذج بديل محتمل، وفي هذا الجزء، سيتم دراسة الادعاءات العلمية لكلا المنهجين بشكل نقدي.

عودة ظهور الفلسفة التجريبية (The Re-Emergence of Empiricism):

بالنسبة للبعض، فإن فلسفة النموذج الرابع هي ما أطلق عليه Brooks (2013 a) «منهج البيانات» (Data-ism)، وهو محاولة التقاط أي شيء على شكل بيانات واستخلاص مستوى معين من التبصر منها: "فأي شيء يمكن قياسه يجب أن يتم قياسه... البيانات هي منظور شفاف وموثوق يسمح لنا بتنقية العاطفية والأيدلوجية؛ إن البيانات ستساعدنا على فعل أشياء رائعة - مثل التنبؤ بالمستقبل". إن من شأن مثل هذه الادعاءات المتعلقة بالبيانات أن تعزز كثيراً من الدعايات بشأن البيانات الكبيرة داخل مجتمع الأعمال، ويتم التعبير عنها بشكل عام من خلال الصياغة التجريبية - التي يمكن لها من خلال الحجم الكافي للبيانات أن تعبر البيانات عن نفسها. ومثل هذه الفلسفة التجريبية متضمنة بشكل أمثل في ادعاءات (2008) Chris Anderson، وهو رئيس تحرير سابق في مجلة وآيرد (Wired)، والذي ضربت دعوته التأليبية بأن البيانات الكبيرة ما هي إلا مؤشر على "انتهاء النظرية" على وتر حساس للكثير من المعلقين. لقد ذكر Anderson في مقطوعة استفزازية أن "طوفان البيانات يجعل المنهج العلمي مهملاً"، وأن الأنماط والعلاقات المتضمنة داخل البيانات الكبيرة تقدم بشكل متواصل معرفة مهمة وثاقبة عن العمليات الاجتماعية، والسياسية، والاقتصادية، والظواهر المعقدة. وقد أكد ذلك قائلاً:

هناك الآن طريقة أفضل. فالبيتابايت تسمح لنا بأن نقول: "إن الارتباط كافٍ" (Correlation is enough). ويمكننا أن نتوقف عن البحث عن النماذج. ويمكننا أيضاً أن نحلل البيانات بدون أي افتراضات بشأن ما يمكن أن يظهره التحليل لنا. ويمكننا أن نرمي الأرقام في أكبر تجميعات الإحصاء التي شهدتها العالم يوماً، وأن ندع الخوارزميات الإحصائية تجد الأنماط إذا كان العلم لا يستطيع إيجاد هذه الأنماط... فالارتباط يحل محل السببية، ويمكن للعلم أن يتقدم ويتطور حتى بدون النماذج المترابطة أو النظريات الموحدة، أو أي تفسير آلي على الإطلاق. لذا، لا يوجد أي سبب للتعليق بالطرق القديمة.

وعلى نحو مشابه، أكد (Prensky 2009) على أنه "لم يعد يتعين على العلماء الآن القيام بتخمينات تنم عن الثقافة، أو أن يقوموا بتأسيس افتراضات ونماذج، ويختبروا هذه الافتراضات وتلك النماذج من خلال التجارب والأمثلة المعتمدة على البيانات. فبدلاً من ذلك، يمكنهم التنقيب في المجموعة الكاملة من البيانات بحثاً عن الأنماط التي تُظهر التأثيرات وتُقدم النتائج العلمية بدون إجراء أي تجريب إضافي". ومن ثم فلقد أكدت (Dyche 2012) أن: "التنقيب في البيانات الكبيرة يكشف العلاقات والأنماط التي لم نكن نعرفها كي نبحث عنها". واستمرت في الادعاء قائلة: "هذه الأنماط محددة للغاية وتبدو وكأنه تم التحكم في تحديدها من خلال الافتراضات". وبالمثل، أكد (Steadman 2013)، على ما يلي:

يُتيح منهج البيانات الكبيرة تجاه جمع المعلومات المبني على المعرفة للمحلين الحصول على القرار الكامل فيما يتعلق بالشئون عالمية النطاق. فلا يتم فقد أي شيء عند النظر عن قرب بالغ إلى جزء واحد محدد من البيانات؛ ولا يتم فقد أي شيء عند محاولة الحصول على منظور واسع للغاية بشأن موقف حيث تم فقدان التفاصيل الدقيقة. فالخوارزميات تجد الأنماط في حين أن الافتراضات تتبع من البيانات، ولا يجب على المحلل أن يزعج نفسه حتى بطرح افتراض بعد الآن. لقد تحول دورها من دور استباقي إلى دور تفاعلي، حيث إن الخوارزميات هي من تقوم بالعمل السياقي... فالخوارزميات ستقوم بتحديد الأنماط ثم تقوم بوضع النظريات، ولذا فهناك حاجة متناقضة للقلق بشأن وضع افتراض في البداية ثم اختبار هذا الافتراض من خلال عينة من البيانات.

وبعبارة أخرى، فبدلاً من اختبار إذا ما كانت بعض الأنماط المفترضة المعينة أو العلاقات توجد داخل مجموعة بيانات، فإن الخوارزميات مصممة للعمل على البيانات الكبيرة من أجل اكتشاف ارتباطات ذات دلالة بين البيانات دون أن يتم توجيهها من قبل الافتراضات. وكما قال (Croll 2012:56): "في النموذج القديم المتسم بندرة البيانات، كان يتعين علينا أن نقرر ما الذي ينبغي جمعه أولاً ثم نقوم بجمعه. أما مع النموذج الجديد المتسم بوفرة البيانات، فنحن نقوم بالجمع أولاً ثم نسأل الأسئلة لاحقاً".

إن الأمثلة المستخدمة في توضيح مثل هذا الموقف عادةً ما تكون نابعة من التسويق والتجزئة. فعلى سبيل المثال، وصف Dyche (2012) حالة سلسلة تجزئة حيث قام بتحليل صفقات الشراء لمدة اثني عشر عاماً من أجل ملاحظة العلاقات المحتملة غير الملحوظة بين المنتجات التي انتهى بها الأمر في سلة المشتريات للمتسوقين. وقد أدى اكتشاف الارتباطات بين البنود المعينة إلى تعيين الأماكن الملائمة للمنتج الجديد وإجراء بعض التعديلات على إدارة مساحة الأرفف مع زيادة تبلغ ١٦٪ في الإيرادات لكل حقيبة تسوق في الشهر التجريبي الأول. لم تكن هناك أي افتراضات بأن المنتج (س) على سبيل المثال غالباً ما يتم شراؤه مع المنتج (ص) والتي تم اختبارها فيما بعد. لقد تم بكل بساطة طرح بعض التساؤلات بشأن البيانات من أجل اكتشاف ماهية العلاقات الموجودة التي لم يكن من الممكن ملاحظتها في السابق. وبالمثل، فإن نظام التوصيات في موقع أمازون يقدم اقتراحات بشأن منتجات أخرى من الممكن أن يكون المتسوق مهتماً بها دون معرفة أي شيء عن ثقافة أو تقاليد الكتب والقراءة لذلك المتسوق، فهذا النظام يحدد، ببساطة، أنماط الشراء عبر العملاء، وذلك من أجل تحديد ما إذا كان الشخص (س) يحب الكتاب (ص)، فمن المحتمل أنه سوف يحب الكتاب (ز) وذلك بالنظر إلى الأنماط الاستهلاكية له وللآخرين.

ويزعم Dyche أن هذا المنهج المفتوح للاكتشاف، بدلاً من أن يكون موجهاً، يُعد أكثر احتمالية لكشف الأنماط الضمنية غير المعروفة، وذلك فيما يتعلق بسلوكيات المستهلك، والعلاقة بين المنتجات، والمخاطر المالية، والتي يمكن الاستفادة منها لاحقاً. وفي حين أن من الممكن أن يكون من المفضل توضيح أي ارتباطات موجودة داخل البيانات وما السبب في كون هذه الارتباطات ذات دلالة، فإن مثل هذا التفسير يُعد غير ضروري بشكل كبير، فالمهم هو معرفة أنه يفعل ذلك. ومن ثم فلقد ادعى Siegel (2013: 90)، (هذا هو تأكيد) فيما يتعلق بالتحليلات التنبؤية: "نحن في العادة لا نعرف عن السببية، ونحن في الغالب لا نهتم بذلك بالضرورة... إن الهدف يركز بشكل كبير على التنبؤ أكثر من تركيزه على فهم العالم... إنه بحاجة إلى أن ينجح فحسب؛ إن التنبؤ يفوق التفسير". وكما ادعى Weinberger (2011: 33) فيما يتعلق بهنش (Hunch)، وهو موقع شبكي يقوم بتقديم توصيات بناء على الإجابات عن الأسئلة التي يبدو أنها لا ترتبط ارتباطاً وثيقاً بالأسئلة الأصلية التي تم توجيهها:

إن التحليل لا يدعم النظرية وهو لا يقدم أي نظرية. لا توجد لدى موقع هُنش (Hunch) أي فكرة لم أن الناس الذين، على سبيل المثال، يفضلون ارتداء صندل على الشاطئ والذين لم يقوموا بنفح الهندباء (نبتة بريه) في العام الماضي، ربما يحبون هذه الأفلام الأربعة (التي تم اقتراحها لهم). فالموقع لم يكن لديه أي افتراض، ولم يكن لديه أي تخمين. كل ما كان لديه ارتباط إحصائي فحسب.

إن بعض برامج تحليلات البيانات يتم بيعها من أجل هذا التصور بالضبط، فعلى سبيل المثال، فإن البعض يدعي أن برنامج التنقيب عن البيانات والعرض البصري آياسداي (Ayasdi)، لديه القدرة على اكتشاف الرؤى بشكل آلي بغض النظر عن التعقيد ودون توجيه أي أسئلة. ويمكن لعملاء آياسداي أن يتعلموا أخيراً الإجابات عن الأسئلة التي لم يكونوا يعرفون أن يسألوها في المقام الأول. ولعرض الأمر بصورة مبسطة، فإن آياسداي هو «سرنديبية رقمية» بمعنى أن آياسداي لديه القدرة على اكتشاف الأشياء المفيدة مصادفة (من أسطورة أمراء سرنديب الثلاثة) (Clark 2013). وعلاوةً على ذلك، فهناك من يدعي أن آياسداي أزال بشكل كلي دور العنصر البشري في عملية التنقيب عن البيانات - ومن ثم، فلقد أزال جميع التحيزات البشرية المتضمنة معه. وبدلاً من انتظار أن يتم توجيه سؤال إليه، أو أن يتم توجيهه إلى روابط بيانات موجودة معينة، فإن النظام سوف يقوم وبدون توجيهه باكتشاف الأنماط التي ربما لم يفكر المتحكم البشري أن يبحث عنها (Clark 2013).

إن هناك مجموعة قوية وجذابة من الأفكار التي تعمل في مثل هذه المجادلات والتي تجري على نحو متضاد مع المنهج الاستدلالي المسيطر والمهيمن داخل العلوم الحديثة. أولاً: أنه يمكن للبيانات الكبيرة أن تلتقط المجال بالكامل وأن تقدم قراراً كاملاً. ثانياً: أنه لا توجد هناك حاجة للنظرية، أو النماذج، أو الافتراضات الاستنباطية. ثالثاً: أنه من خلال تطبيق تحليلات البيانات بدون دراية يمكن للبيانات أن تتحدث عن نفسها وهي متحررة من الصياغة والانحياز البشريين، وأن أي أنماط أو علاقات داخل البيانات الكبيرة تكون صادقة وهادفة بشكل أصلي. رابعاً: أن المعنى يتجاوز السياق أو المعرفة المرتبطة بالمجال.

فهذه الأمور تعمل سويةً لكي تشير إلى أن هناك أسلوباً جديداً في العلوم يتم تأسيسه، وهو الأسلوب الذي تكون فيه طريقة العمل استقرائية بطبيعتها، وممثل هذا ادعى

(Gannon and Reid 2009) أن "التوافر الجاهز في البيانات المتعددة يحول المناهج العلمية من الطريقة العلمية التقليدية التي تقودها الافتراضات إلى العلوم المعتمدة على الاستكشاف". وبالمثل، يقترح Weinberger (2011: 127) أن الهدف التقليدي للعلوم المتمثل في تكوين النظريات التي تدعمها الحقائق وتفسر الحقائق في الوقت ذاته يتحول في الوقت الراهن إلى "تقديم نماذج معقدة تفسر العلاقات المعقدة" حيث "في بعض الحالات يكون من المستحيل كشف السبب وراء عمل تأثيرات معينة، وبدلاً من ذلك، هناك ارتباط قوي ملاحظ بين المحاكاة والعالم الحقيقي والذي يمنح الثقة بأن المحاكاة الأخرى سوف تقدم نتائج موثوقة". وقد استخدم Weinberger المثال المتعلق بوضع عمود على بعد متر واحد أمام أحد الأبواب في برنامج محاكاة لتدفق خروج الناس في حال الإخلاء والذي أدى إلى تدفق أفضل، ولكن دون وجود أي تفسير من النموذج للسبب الذي يجعل العمود يفعل ذلك. واقترح، بطرق أخرى متعددة، أننا لسنا بحاجة لأن نعرف السبب، وإنما نحتاج لأن نعرف أن العمود يفعل ذلك وأن تطبيق الفكرة سوف يساعد على إنقاذ حياة الناس، وقد أنتج هذا ل Andrejevic (2013: 26) حالة تسمى "المعرفة بدون الفهم".

مغالطات الفلسفة التجريبية (The Fallacies of Empiricism):

لا يوجد هناك أي شك بأن البيانات الكبيرة لديها، وسوف يكون لديها بشكل متزايد، تأثير على كيفية ممارسة وفهم العلوم. وعلى الرغم من ذلك، فإن رؤية الفلسفة التجريبية بشأن كيفية تكون هذا العلم الجديد تعتمد على التفكير المغالط - جميع الأفكار الأربع التي تعزز تشكيلها مشكوك فيها. فلربما تكون هذه الأفكار جاذبة بشكل كبير، ولكن التدقيق النقدي والفلسفي يكشف مشاكل خطيرة مع كل فرضية. إن من المهم أن يتم التحقق من كل فكرة، كل بدورها، قبل صياغة رؤية بديلة للعلم الذي تقوده البيانات والذي يمزج بين سمات الإبعاد، والاستقراء، والاستنتاج.

أولاً: يمكن للبيانات الكبيرة أن تلتقط المجال بالكامل وأن تقدم قراراً كاملاً. حيث إن البيانات الكبيرة تكافح من أجل الشمولية، فمناصروها يدعون أنه من الممكن رؤية كل شيء داخل المجال (على سبيل المثال جميع الأنشطة على موقع تويتر، أو جميع عمليات الشراء داخل سلاسل المتاجر الكبيرة)، وأنه من الممكن أن نفهم بشكل كامل ما نراه (Vis 2013). وعلى

الرغم من ذلك، وكما ناقشنا في الفصول ١ و٢ و٩، فإن البيانات الكبيرة ربما تسعى إلى أن تكون شمولية، ولكنها تشكل كلاً من التمثيل والعينة. فالواقع، أن جميع البيانات تقدم رؤى بصرية قليلة للعالم، وليس رؤى شاملة: أي رؤى من وجهات نظر معينة. باستخدام أدوات معينة، وذلك بدلاً من رؤى بصرية شاملة لا تخطئ (Amin and Thrift 2002; Haraway 1991). وعلاوةً على ذلك، فإن المجالات تتطور وتتغير، ومن ثم فإن البيانات الكبيرة ليس بإمكانها الوصول إلى كل مكان، وذلك يرجع من ناحية إلى أن البيانات الكبيرة لا يمكنها الإلمام بجميع الأماكن والأزمنة، ومن ناحية أخرى؛ فإن الكثير من الأماكن والأزمنة الجديدة تبقى لكي يتم اختراعها (Amin and Thrift 2002: 128). وعلى هذا النحو، تشكل البيانات الكبيرة "سلسلة من الأوامر الجزئية والكيانات المتمركزة، مع مقدرتها على التحديق في بعض الاتجاهات دوناً عن الأخرى" (Latour والمقتبس في Amin and Thrift 2002: 92)، ومن ثم فلقد استنتج (Vis 2013): "أنه من المهم أن نتذكر أن ما نراه تتم صياغته من قبل ما نحن قادرون على رؤيته أو في الواقع ما نريد رؤيته من خلال إطار عمل فكري محدد".

ثانياً: أنه لا توجد هناك حاجة للنظرية، أو النماذج، أو الافتراضات الاستنباطية. الافتراض هنا هو أن تكوين البيانات وتحليلات البيانات واللذان يعرزان استخدام البيانات الكبيرة يبدو أنهما يأتيان من لا شيء، وأنهما بشكل ما متحرران من "القوة التنظيمية للفلسفة" (Berry 2011: 8). فالأنظمة المتناقضة تُصمم من أجل التقاط أنواع معينة من البيانات، وغالباً ما يحصل ذلك من أجل أغراض محددة للغاية، وأن التحليلات والخوارزميات اللتين تدعمان البيانات الكبيرة تعتمدان على الاستدلال العلمي وأنه تم صقلها من خلال الاختبار العلمي. فالعديد من الأفكار المستخدمة في تصميم نظم البيانات الكبيرة مستمدة من الأدبيات العلمية واسعة النطاق والتجارب والمعارف لهؤلاء الذين يعملون على النظام. إن هناك الكثير من الوقت المستنفد في اختيار واختبار طرق متعددة لتحديد واستخلاص القيمة من البيانات المنتجة، وتقييم النتائج وتطوير التحليلات. وهناك جدول أعمال يقود مثل هذا العمل وذلك فيما يتعلق بمهية النتائج المرجوة (زيادة دوران رأس المال أو الأرباح، أو فهم نظام معين، إلخ)، والذي يستفيد من النظريات المبرهنة، وذلك فيما يتعلق بمهية الطرق الملحوظة لتحقيق مثل هذه التأثيرات، وما إذا كانت مثل هذه التأثيرات تحدث في الواقع ثم يتم اختبارها بعد ذلك.

وكما أشار Leonelli (2012)، فإن الإستراتيجية الاستقرائية لتحديد الأنماط داخل البيانات عادةً ما يقودها الاستدلال العلمي، وهي لا تحدث فراغاً علمياً. أو، كما ذكر Quintero (2012)، "لا يمكن تقديم التفسيرات ببساطة من التحليلات الإحصائية، وذلك دون الحاجة إلى نظرية سابقة من نوع ما". وقد أتبع ذلك قائلاً: "كونك منقاداً بالبيانات يعد أمراً مفيداً فحسب إذا كان لديك نظرية قوية يمكنك من خلالها أن تستمر في طريقك، إن أي شيء آخر سوف يتركك متجهاً إلى منحدر دون أن تراه". فالواقع، أن كلاً من الاستدلال الاستنتاجي والاستقرائي يتم تشكيلهما دوماً بشكل استطرادي ولا ينشأ أحدهما من لا شيء. ومن ثم فقد اقترح Popper (1979)، مقتبس من Callebaut (2012: 74) أن جميع العلوم تطبق منهج الأضواء الكاشفة على الاكتشاف العلمي، حيث يكون تركيز الضوء منقاداً من خلال النتائج السابقة والنظريات والتدريب، وكذلك من خلال التخمين المؤسس على التجارب والمعارف. إن الأمر ذاته ينطبق على أمازون (Amazon) وهُنش (Hunch)، وآياسداي (Ayasdi)، وجوجل (Google). فالكيفية التي اعتمد عليها أمازون في بناء نظام التوصيات على الاستدلال المنطقي، كانت مُعززة من خلال نموذج إرشادي يرافقه اختبار تجريبي تم تصميمه من أجل تحسين أداء الخوارزميات التي يستخدمها. وبالمثل، فإن جوجل يقوم ببحوث وتطوير واسع النطاق، وهو يعمل بالشراكة مع العلماء ويشترى المعرفة العلمية، سواء تم هذا من خلال تمويل البحوث داخل الجامعات أو من خلال شراء المعارف الإلكترونية للشركات الأخرى، وذلك من أجل تنقية وتوسيع المنفعة بشأن كيفية تنظيمه، وتقديمه، واستخلاصه للقيمة من البيانات. ونتيجة لذلك، فإذا وجدت الخوارزميات الإحصائية أنماطاً في البيانات، فإن ذلك يعود إلى أنه تم استخدام علم التعرف على الأنماط، جنباً إلى جنب مع المعرفة الخاصة بالمجال.

ثالثاً: يمكن للبيانات أن تتحدث عن نفسها وهي متحررة من الصياغة والانحياز البشريين. إن المفهوم القائل بأن إنتاج المعرفة من البيانات الكبيرة يحدث بشكل غير مرتبط بالعلم، هو على علاقة بالفكرة القائلة بأن تحليلات البيانات الكبيرة تُمكن البيانات من التحدث عن نفسها، وهي غير مثقلة بالسياقية أو أهواء التوضيح البشري. فليس فقط إنتاج البيانات هو ما يتم بشكل افتراضي ومتحرر من النظرية، فحتى تفسير ومعاني تلك البيانات يمكن بطريقة مشابهة أن يحدث في فراغ علمي، فمثل هذا المفهوم يُعد سمة من سمات التفكير

التجريبي، وهو أنه عندما يتم تقديم وتحليل البيانات بطريقة ملائمة، فإن البيانات، بنفسها ومن نفسها، تشكل دليلاً ملحاً يرتبط بظاهرة أو عملية معينة، بدلاً من تفسيرها. وفي حالة البيانات الكبيرة، فإن حجم البيانات يضيف إلى أهمية الدليل. فهذا "التقديس" للبيانات يمكن أن يؤدي، كما ادعى (Jenkins 2013)، إلى حتمية البيانات في المناقشات التي تتم من أجل البيانات الكبيرة. وهذا يعني أن البيانات تتفوق على النظرية، وأن البيانات أمر حقيقي، في حين أن النظرية ليست حقيقية (Hales 2013)، وأن البيانات تقول الحقيقة في حين أن النظرية مجرد تليفيق. وهنا، يوجد ثلاث قضايا يتعين التفكير فيها: القضية الأولى هي افتراض أن البيانات موضوعية، ومحايدة، ومتحررة من التحيز البشري. القضية الثانية أن الأنماط والعلاقات داخل البيانات الكبيرة هادفة وصادقة بشكل متأصل. القضية الثالثة أن تفسير مثل تلك الأنماط والعلاقات يفترق إلى أي نوع من الوضعية والموقعية.

وكما ناقشنا في الفصل الأول، فإن البيانات ليست ببساطة عناصر محايدة أو طبيعية بحيث يتم فصلها عن العالم بطرق محايدة وموضوعية، ويمكن قبولها بالقيمة الاسمية. فالبيانات لا توجد قبل تكوينها ولا تنشأ من أي شيء. ولكن الأخرى، أن البيانات يتم إنشاؤها داخل مجاميع البيانات المعقدة التي تشكل تركيبها بشكل فعال، ولا يمكن للبيانات في هذه الحالة التحدث عن نفسها، ولكنها على الدوام تتحدث، بشكل متأصل، من وضعية معينة (Crawford 2013). وعلاوة على ذلك، فإن ادعاء (Anderson 2008) بأن الارتباط يفوق السببية يشير إلى أن الأنماط الموجودة داخل مجموعة البيانات ذات معنى بشكل متأصل. وهذا افتراض يعرف جميع الإحصائيون المدربون أنه خطر وغير صحيح. فالارتباطات بين المتغيرات داخل مجموعة البيانات يمكن أن تكون عشوائية بطبيعتها ويوجد بينها ارتباط عرضي قليل، أو لا يوجد هذا الارتباط على الإطلاق (انظر الفصل التاسع). إن تفسير أي ارتباط على أنه ذو مغزى ربما يؤدي في هذه الحالة إلى مغالطات بيئية خطيرة. ويمكن أن يتفاقم هذا الأمر في حالة البيانات الكبيرة، وذلك يرجع إلى كون الموقف التجريبي يظهر وكأنه يعزز ممارسة تعريف البيانات - البحث عن أي ارتباط - ومن ثم يزيد من احتمالية اكتشاف ارتباطات عشوائية. وعلى الرغم من ذلك، ففي حين أن التعرف على الأنماط ربما يحدد علاقات شيقة محتملة، إلا أن صحة هذا الأمر تستلزم اختبارها على مجموعات بيانات أخرى وذلك من أجل التأكد من مصداقيتها وصحتها. وبعبارة أخرى، يتعين أن

تشكل العلاقات أساس الافتراضات التي يتم اختبارها على نطاق أوسع، والتي تستخدم بعد ذلك من أجل بناء وصقل النظرية التي تفسرها، ومن ثم فإن الارتباطات لا تفوق السببية، ولكنها بالأحرى يجب أن تشكل أساس بحوث إضافية من أجل معرفة ما إذا كانت مثل هذه الارتباطات دالة على السببية، وفي هذه الحالة فقط يمكننا أن ندرك مدى مغزى أسباب الارتباط.

وفي حين أن الفكرة القائلة بأن البيانات تتحدث عن نفسها وهي متحررة من التحيزات أو أن الصياغة ربما تبدو وكأنها فكرة جاذبة، إلا أن الواقع مختلف إلى حد ما. فبحسب ما أشار Gould (1981: 166)، فإن "البيانات غير الحية لا يمكنها أبداً أن تتحدث عن نفسها، ونحن دوماً نحاول أن نوجه بعض إطار العمل المفاهيمي، سواءً كان بديهياً ومشكلاً بصورة سيئة، أم مركباً بشكل محكم ومنهجي، نحو مهمة الاستقصاء، والتحليل، والتفسير". إن فهم البيانات دوماً مؤطر؛ ويتم فحصها من خلال منظور معين يصوغ كيف تتم ترجمتها، وحتى لو كانت العملية مؤتمتة بطريقة ما، فإن الخوارزميات المستخدمة لمعالجة البيانات عادةً ما تكون متشعبة بقيم معينة وتكون مقترنة بالسياق داخل منهج علمي معين. وكما أوضح Hales (2013): فإن "أي اختبار إحصائي أو خوارزمية تعلم آلي تعبر عن منظور ماهية النمط أو النظام... فأحد الخوارزميات سوف تجد نمطاً ما، في حين تجد خوارزمية أخرى شيئاً آخر". فإذا ترجم ذلك شخص ما، فسيكون من المحتمل أن يستفيد من معرفته، ومهاراته، وحدسه المتراكم من أجل تنقية رؤاه، حتى يقوم محلل مختلف بطرح نتائج مختلفة من نفس التحليلات. الفهم أو التفسير يتشكل في العادة من خلال أعين المشاهدين بغض النظر عن ادعائهم أنهم محايدون أو متحررون من القيمة الناتجة، وعلاوةً على ذلك، فإن التفسيرات ليست ثابتة، فهي تتغير كلما تعرضت إلى معارف جديدة من خلال بحوث، أو مناقشات، أو مجادلات إضافية. وعلى هذا، فإن البيانات لا تتحدث عن نفسها أبداً، وكما ادعى Silver (2012: 9): "نحن نتحدث بالنيابة عن البيانات".

رابعاً: المعنى يتجاوز السياق أو المعرفة المرتبطة بالمجال. إن من الجوانب المرتبطة بالفكرة القائلة بأن البيانات تتحدث عن نفسها المفهوم القائل بأن تفسير البيانات الكبيرة لا يتطلب معرفة سياقية أو مرتبطة بمجال معين. وبدلاً من ذلك، فالهم هو المقدرة على

توضيح المعنى داخل البيانات، فبسبب أن البيانات من المفترض أن تتحدث عن نفسها، يمكن لأي شخص ذو فهم معقول بالإحصاء أن يكون قادراً على تفسيرها. إن هناك شكوكاً قليلة فيما يتعلق بأن علماء البيانات والحاسب والعلماء الآخرين يتحركون داخل نطاق تخصصات بعضهم البعض مع انتشار البنى التحتية للبيانات والحوسبة. وفي الكثير من الحالات، فإنهم لا يقدمون أدوات جديدة فحسب، ولكنهم يشتركون بفعالية في تفسير البيانات. ومع ذلك، فإن هؤلاء العلماء لديهم القليل من الفهم العميق في المجال المناط بهم وكذلك بالنظريات التجريبية المتعمقة والمعارف التي تم تأسيسها عبر فترات زمنية طويلة. فعلى سبيل المثال، فيما يتعلق بالعلوم الاجتماعية والإنسانية، فإن الأساليب الجديدة من التحليلات والبنى التحتية للبيانات توجه من قبل علماء الحاسب والفيزياء. وفي الغالب تجعل النتائج أولئك المتشربين لوجهة النظر التخصصية يستسلمون وينقادون لهذه النتائج. ولتوضيح هذا الجانب، فإن المجال الناشئ للفيزياء الاجتماعية، والذي من خلاله يقوم الفيزيائيون وآخرون باتخاذ قرارات بشأن العمليات الاجتماعية والمكانية بناءً على تحليلات البيانات الكبيرة، وخاصة تلك المتعلقة بالمدن والقوانين المفترضة التي تعزز صياغتها ووظائفها (Bettencourt et al. 2007; Lehrer 2010; Lohr 2013)، غالباً ما يتجاهل عن عمد ثقافة علم الاجتماع الممتد لقرنين من الزمان، ويشمل ذلك تجاهل التحليلات الكمية والنماذج المبنية لما يقارب قرناً من الزمان. والنتيجة هي تحليل المدن الذي يكون اختزالياً إلى حد كبير، وبنوعه يتجاهل تأثيرات الثقافة، والسياسة، والحوكمة، ورأس المال، وتقاليد ثرية من العمل الذي سعى نحو فهم كيفية عمل المدن من النواحي الاجتماعية، والثقافية، والسياسية، والاقتصادية والنتيجة إعادة إنتاج الأنواع ذاتها من أوجه القصور التي أوجدتها العلوم الاجتماعية الكمية / الوضعية في منتصف القرن العشرين (Kitchin 2013). وهذا أيضاً يتعدى حدود التقاليد المتنوعة والثرية للإرث الفلسفي ويعمل على تعزيز المناهج الفلسفية التجريبية والوضعية بصفتها أفضل وسيلة لفهم المدن مع أنه من المحتمل ألا يتفق مع هذه الرؤية النظرية إلا القليل جداً من العلوم الاجتماعية التي تلت عصر الفلسفة الوضعية (Kitchin 2013).

إن هناك مجموعة متشابهة من المخاوف لدى أولئك العاملين في مجال العلوم. فقد أشار Strasser (2012) على سبيل المثال، إلى أنه في العلوم البيولوجية، ادعى متخصصو المعلوماتية الحيوية الذين لديهم أسلوب معين وضيق للغاية لفهم علم الأحياء، أن الأرض كان يسكنها في السابق الأطباء السريريون وعلماء الأحياء التجريبيون والجزيئيون. وأشار إلى أن: "تحويل التحليل إلى البيانات الرقمية... يُفسح المجال أمام الأسئلة المعرفية كتلك المتعلقة بمن يملك الشرعية ليكون منتجاً للمعرفة- المهتم بالآثار (أو الطبيب السريري أو عالم البيولوجيا الجزيئية) أو الإحصائي الذي يحلل البيانات ويقدم المحاكاة أو النموذج (2012: 7). ومن ثم فإن بعض علماء البيانات بدون شك يتجاهلون ملاحظات (2013) Porway، وهي:

بدون توافر خبراء في موضوع البحث من أجل توضيح المشاكل مسبقاً، فسوف نحصل على نتائج ضعيفة... وبدون شك فإن هناك حاجة إلى خبراء في موضوع البحث من أجل تقييم نتائج العمل وخاصةً عندما يتعامل الفرد مع بيانات حساسة عن السلوك البشري. وبوصفنا علماء بيانات، فنحن مؤهلون بشكل جيد للتعامل مع "ماذا" فيما يتعلق بالبيانات، ولكن لا يجب علينا إلا فيما ندر التعامل مع سؤال "لماذا" فيما يتعلق بالموضوعات التي لا نكون خبراء فيها.

وكما أشار Porway، فإن المطلوب بالفعل هو أن يعمل علماء البيانات وخبراء النطاق مع بعضهم البعض من أجل التأكد من أن تحليلات البيانات المستخدمة مفهومة وأن النتائج من مثل هذه التحليلات يتم تفسيرها بعقلانية وسياقية. وبالمثل، يدعو علماء الحاسب الضالعين حاسوبياً، (2009: 10 - 11) Lazer et al. إلى التعاون بين علماء الاجتماع الضالعين إحصائياً وبين علماء الحاسب الضالعين حاسوبياً، (2009: 10 - 11)، وفيما يتعلق بالأعمال، ادعى (2013) Minelli et al. أن فرق علماء البيانات ينبغي أن يعملوا مع خبراء عمليات الأعمال من أجل الاستفادة من الرؤى المناسبة (انظر أيضاً الجدول ١-٩).

العلم المُقاد بالبيانات (Data-driven Science):

يسعى العلم المقاد بالبيانات، بدلاً من كونه متجذراً في الفلسفة التجريبية، إلى التمسك بمعتقدات الطريقة العلمية، ولكنه أكثر انفتاحاً لاستخدام مزيج مختلط من المناهج

الاستدلالية، والاستقرائية، والاستنتاجية، من أجل تطوير فهم الظاهرة. وهذا يختلف عن التصميم الاستدلالي التجريبي التقليدي في أنه يسعى إلى طرح افتراضات ورؤى يتم توليدها من البيانات بدلاً من توليدها من النظريات (Kelling et al. 2009: 613). وبعبارة أخرى، فهو يسعى إلى دمج أسلوب من الاستقراء في تصميم البحوث، وذلك على الرغم من أن الشرح من خلال الاستقراء ليس هو الغاية المستهدفة (كما هي الحال مع المناهج التجريبية). وبدلاً من ذلك، فهو يشكل أسلوباً جديداً من أساليب توليد الفرضيات قبل استخدام المنهج الاستنتاجي. ولا تنشأ عملية الاستقراء من لا شيء، ولكنها تقع وتقترب بسياق نطاق نظري متطور للغاية. ومن ثم، فإن الإستراتيجية المعرفية المتبناة داخل العلم المقاد بالبيانات هي استخدام أساليب موجهة لاكتشاف المعرفة من أجل تحديد الأسئلة (الافتراضات) المحتملة الجديرة باختبارات وبحوث إضافية.

فالعملية يتم توجيهها من ناحية أن النظرية الحالية تستخدم من أجل توجيه عملية اكتشاف المعرفة، وذلك بدلاً من أن نأمل ببساطة أن نحدد جميع العلاقات داخل مجموعة البيانات وأن نفترض أنها ذات معنى بطريقة أو بأخرى. ومن ثم فإن كيفية تكوين البيانات أو إعادة تحديد الغرض منها تقوده افتراضات معينة تعززها خبرات ومعارف عملية ونظرية، مثل ما إذا كانت التقنيات وإعدادها سوف تلتقط أو تنتج مواد بحثية مفيدة وملاممة. فبدلاً من إنتاج البيانات بأي وسيلة ممكنة، وذلك باستخدام أي نوع من التقنيات المتاحة أو أي نوع من أطر التمثيل، يتم التفكير بعناية في إستراتيجيات تكوين البيانات وإعادة توظيفها، مع اتخاذ قرارات إستراتيجية من شأنها أن تؤدي إلى أنواع معينة من البيانات. وعلى نحو مشابه، يتم توجيه كيفية معالجة وإدارة وتحليل هذه البيانات من خلال الافتراضات المتعلقة بماهية الأساليب التي ربما توفر رؤى هادفة. فالبيانات لا يمكن إخضاعها لأي صياغة وجودية ممكنة، ولا يمكن إخضاعها كذلك لأي نوع من أساليب التنقيب عن البيانات، على أمل أنها ربما تكشف بعض الحقائق المخفية. وبدلاً من ذلك، يتم اتخاذ القرارات المبنية على النظرية بشأن الطريقة الأمثل للتعامل مع مجموعة البيانات حتى يمكنها أن تكشف المعلومات التي ربما تحمل فائدة محتملة وتكون جديرة بحوث إضافية. وبدلاً من اختبار صحة أي علاقة يتم كشفها، فإن الاهتمام يتركز على تلك العلاقات -بناءً على بعض المعايير - التي يبدو أنها تؤدي إلى المسار الصحيح أو من المحتمل بشكل

أكبر أن تؤدي إلى المسار الصحيح. وفي الواقع، فإن الكثير من العلاقات المفترضة داخل مجموعات البيانات يمكن تنحيها على أنها تافهة أو سخيفة من قبل متخصصي المجال، في حين يتم وسم البعض الآخر على أنها تستحق اهتماماً إضافياً (H.J. Miller 2010).

إن مثل عملية اتخاذ القرار هذه فيما يتعلق بطرق توليد البيانات والتحليل تعتمد على التفكير الاستدلالي. والاستدلال هو وسيلة للاستنتاج والتفكير المنطقي تم وضعه بواسطة (H.J. Miller 2010) (C.S. Peirce (1839 – 1914)). ويهدف الاستدلال إلى خاتمة تكون ذات معنى منطقي ومعقول، ولكنها ليست حاسمة في ادعاءاتها. فعلى سبيل المثال، لا توجد أي محاولة لاستنتاج ما هي أفضل طريقة لتوليد البيانات، ولكن يمكن بدلاً من ذلك تحديد طريقة ذات معنى منطقي بالنظر إلى ما هو معروف بالفعل عن إنتاج مثل هذه البيانات. ويستخدم الاستدلال بشكل شائع في العلوم وخاصة في تشكيل الافتراضات، على الرغم من أن مثل هذا الاستخدام غير معترف به على نطاق واسع. ولذا، فإن أي علاقات يتم كشفها داخل البيانات لا تنشأ من لا شيء، ولا تتحدث عن نفسها ببساطة كذلك. إن عملية الاستقراء - للرؤى الناشئة من البيانات - تتم صياغتها بشكل سياقي. وهذه الرؤى ليست نهاية مطاف الاستقصاء، إذ يتم ترتيبها وتفسيرها من خلال نظرية ما. وبدلاً من ذلك، فإن الرؤى توفر الأساس لتشكيل الافتراضات والاختبار الاستنتاجي لصحتها. وبعبارة أخرى، يُعد العلم المقاد بالبيانات نسخة معاد تشكيلها من الطريقة العلمية التقليدية، مما يوفر وسيلة جديدة لبناء نظرية، وإضافة إلى ذلك، فإن التغيير المعرفي هو هادف بحد ذاته.

ويدعي البعض أن العلم المقاد بالبيانات سوف يصبح النموذج الجديد للطريقة العلمية في عصر البيانات الكبيرة والمتدرجة لأن التفضيل المعرفي يتناسب مع استخلاص رؤى إضافية وقيمة سيفشل العلم التقليدي المقاد بالمعرفة في توليدها (Kelling et al. 2009; H.J. Miller 2012; Loikides 2010). إن العلم المقاد بالمعرفة، عند استخدامه لمنهج استنتاجي مباشر، له فائدة معينة في فهم وتفسير العالم تحت ظروف البيانات النادرة والحوسبة الضعيفة، ولكن الاستمرار في استخدام مثل هذا النهج لا يبدو أنه مفهوم بشكل كلي عندما تشير التطورات التقنية والمنهجية إلى أنه من الممكن الاضطلاع بتحليلات أكثر ثراءً للبيانات - تطبيق تحليلات جديدة للبيانات وتوفير القدرة على ربط بيانات كبيرة ومتفاوتة

سويًا بطرق كانت غير ممكنة لوقت قريب، والتي تقدم بيانات جديدة قيمة - والتحديد والتعامل مع الأسئلة بطرق جديدة ومؤثرة. فالعلم المقاد بالبيانات، بحسب ما يتم نقاشه، يتناسب أكثر مع استكشاف واستخلاص القيمة وفهم مجموعات البيانات الهائلة والمتراصة، وتعزيز البحوث متعددة الاختصاصات التي تدمج خبرات المجال من خلال الإطار النظري الاستهلاكي كونها أقل محدودية، وسوف تؤدي إلى نماذج ونظريات أكثر شمولية وعمومية للأنظمة المعقدة بالكامل بدلاً من عناصر منها (Kelling et al. 2009).

فعلى سبيل المثال، يدعي البعض أن العلم المقاد بالبيانات سوف يحول فهمنا عن الأنظمة البيئية (Bryant et al. 2008; Lehning et al. 2009). وسوف يتيح إمكانية التكامل بين بيانات آنية ولحظية وعالية الدقة وذلك في سبيل تقديم نماذج تفصيلية للبيئة المتغيرة على النحو الذي يتعارض مع نقاط ثابتة غير متغيرة عبر المكان والزمان. ويُقصد هنا التكامل بين البيانات التي يتم توليدها من العديد من المصادر كمحطات الطقس التقليدية والمتحركة، والأقمار الصناعية والتصوير الجوي، وادارات الطقس، ومحطات رصد التيار، ومحطات القياس، وملاحظات المواطنين، والاستشعار الجوي واختبار جودة المياه ومقاييس الغاز، وفحص التربة، والحساسات الموزعة التي تقيس الظواهر المختارة مثل درجة الحرارة والرطوبة. ومن شأن ذلك التكامل أن يساهم في تحديد العلاقات الخاصة بين الظواهر والعمليات التي تولد افتراضات ونظريات جديدة يمكن اختبارها بعد ذلك من أجل التحقق من صحتها. وهذا سوف يساعد أيضاً على تحديد وفهم نقاط الربط بين مجالات بيئية مختلفة مثل الغلاف الجوي (الجو)، والغلاف الحيوي (الأنظمة البيئية)، والغلاف المائي (أنظمة المياه)، والقشرة الأرضية (القشرة الصخرية للأرض)، والغلاف الترابي (التربة)، والمساعدة على دمج النظريات في مجاميع نظرية أكثر شمولية. ومن شأن هذا أن يوفر فهماً أفضل للعمليات المترابطة والمتنوعة في أرض الواقع والعلاقات الترابطية بينها وبين الأنظمة البشرية، ويمكن استخدامه كذلك من أجل توجيه النماذج والمحاكاة من أجل التنبؤ بالاتجاهات طويلة المدى والإستراتيجيات التكيفية الممكنة.

العلوم الاجتماعية الحاسوبية والعلوم الإنسانية الرقمية (Computational Social Sciences and Digital Humanities):

في حين أن معرفية العلم المقاد بالبيانات تبدو كما لو كانت وضعت لتحويل منهج البحث المطبق في العلوم المادية، والطبيعية، والحياتية، والهندسية، إلا أن مسارها في العلوم الاجتماعية والإنسانية يبدو أقل تأكيداً. فهذه المجالات من العلوم متنوعة بشكل كبير في أسسها الفلسفية، وخاصة مع قيام عدد قليل فقط من العلماء في هذه المجالات بتوظيف المعرفة المشتركة بين العلوم. فهؤلاء الذين يستخدمون الطريقة العلمية من أجل تفسير ومُذجة الظواهر الاجتماعية، بشكل عام، يستفيدون من أفكار الفلسفة الوضعية (على الرغم من أنهم ربما لا يستخدمون مثل هذا العنوان، Kitchen 2006). ومثل هذا العمل يميل إلى التركيز على المعلومات الواقعية والكمية - الظواهر المشاهدة تجريبياً والتي يمكن قياسها بشكل فعال كالأعداد والمسافات والتكلفة والوقت على النحو الذي يتعارض مع المجالات الأكثر مادية من الحياة الإنسانية مثل المعتقدات والأيدولوجيات - باستخدام الاختبار الإحصائي من أجل تأسيس علاقات سببية وبناء النظريات والنماذج التنبؤية والمحاكات. فالمناهج الوضعية مؤسسة بشكل جيد في الاقتصاديات، والعلوم السياسية، والجغرافيا البشرية، وعلم الاجتماع، ولكنها نادرة بشكل كبير في العلوم الإنسانية. وعلى الرغم من ذلك، فداخل هذه التخصصات المذكورة، كانت هناك حركة قوية خلال النصف الأخير من القرن الماضي تجاه المناهج ما بعد الوضعية، وخاصة في الجغرافيا البشرية وعلم الاجتماع.

وفيما يتعلق بعلماء الوضعية في العلوم الاجتماعية، فإن البنى التحتية للبيانات، والبيانات المفتوحة والبيانات الكبيرة توفر فرصاً هائلة من أجل تطوير نماذج أكثر تطوراً وأكثر دقة وعلى نطاق واسع للحياة الإنسانية. فثورة البيانات توفر إمكانية الانتقال من دراسات المجتمعات التي تتسم بالندرة في البيانات إلى الدراسات الثرية في البيانات، ومن اللقطات الثابتة إلى الانتشار الديناميكي، ومن التجميعات الرديئة إلى الدقة العالية، ومن النماذج البسيطة نسبياً إلى محاكاة أكثر تعقيداً وتطوراً (Kitchen 2013). إن هناك احتمالية لعصر جديد من العلوم الاجتماعية الحاسوبية التي تجري دراسات ذات اتساع، وعمق، ومدى، وإطار زمني

أكبر بكثير، وهي تغطي فترات زمنية متعددة بشكل متآصل، وذلك على النقيض من بحوث العلوم الاجتماعية الحالية (Lazer et al. 2009; Batty et al. 2012). وعلاوةً على ذلك، فإن حجم، وصحة، وتنوع، وشمولية، ودقة، وارتباطية، ومرونة، وتوسعية البيانات، بالإضافة إلى القوة المتزايدة للحوسبة وتحليلات البيانات الجديدة، تتعامل مع بعض انتقادات علماء الوضعية حتى اليوم، وخاصةً تلك المتعلقة بالاختزالية والشمولية، وذلك من خلال تقديم تحليلات أكثر دقةً وحساسيةً والتي يمكنها أن تهتم بالسياق والاحتمالية، ويمكن استخدامها من أجل صقل وتوسيع الفهم النظري للعالم الاجتماعي المكاني. وعلاوةً على ذلك، وبالنظر إلى شمولية البيانات، فمن الممكن اختبار صحة مثل هذه النظرية عبر العديد من المواقف والبيئات. وفي مثل هذه الظروف، فإن المعرفة عن الأفراد، والمجتمعات، والجماعات، والبيئات سوف تصبح أكثر تبصراً وفعلاً وذلك فيما يتعلق بصياغة السياسة والتعامل مع القضايا المتعددة التي تواجه البشرية. إن التحدي الكامن في مثل هذا السيناريو لم يعد هو البيانات، ولكن تأسيس طرق إحصائية وغمذجة جديدة لها المقدره على التعامل مع ملايين أو بلايين المشاهدات (Batty et al. 2012).

وبالنسبة لعلماء ما بعد الوضعية، فإن ثورة البيانات المنتشرة توفر كلاً من الفرص والتحديات. فالفرص تتمثل في رقمنة وربط البيانات التناظرية وغير المنظمة، والتي يُعد الكثير منها جديداً كوسائل التواصل الاجتماعي، والكثير منها كان من الصعب الوصول إليها في السابق كملايين الكتب، والمستندات، والصحف، والصور، والأعمال الفنية، والأشياء المادية... إلخ من عبر التاريخ والتي تم تحويلها إلى الصيغة الرقمية خلال العقدين السابقين من خلال مجموعة من المنظمات (Cohen 2008)؛ وإتاحة أدوات جديدة لمعالجة وإدارة وتحليل البيانات؛ وتوفير وسائل جديدة لتعزيز التعاون بين مختلف التخصصات وتجاوز نموذج العالم الواحد؛ وتوسعة الروابط البحثية مع المؤسسات غير الأكاديمية وتضمين جمهور جديد على نطاق أوسع (Sword 2008). وتخضع هذه الفرص للدراسة على نطاق واسع من خلال المجال الناشئ للإنسانيات الرقمية، والذي يجمع العلماء سويًا من جميع العلوم الإنسانية كالمؤرخين، واللغويين، والفنانين المبدعين، وعلماء الأدب والإعلام... إلخ مع أخصائي المكتبات، وعلماء الحاسب، والبيانات، والمعلومات.

وبشكل مبدئي، فإن العلوم الإنسانية الرقمية تتشكل من معالجة وتحليل البيانات التي تكونت في الأصل بشكل رقمي، ورقمنة وأرشفة المشروعات التي تسعى إلى تحويل النصوص والمواد التناظرية إلى صيغ رقمية بحيث يمكن تنظيمها والبحث فيها وإخضاعها لأشكال أساسية من التحليلات الآلية، والموجهة، والشاملة، مثل العروض المرئية الموحدة للمحتوى أو الروابط بين المستندات، والناس أو الأماكن (Schnapp and Presner 2009). ونتيجة لذلك، فلقد ادعى المؤيدون أن المجال قد تطور لكي يوفر أدوات أكثر تطوراً لمعالجة، والبحث، وربط، ومشاركة، وتحليل البيانات التي تسعى إلى إتمام ودعم طرق العلوم الإنسانية الحالية، وتسهيل الأشكال التقليدية من التفسير وبناء النظرية، وذلك بدلاً من استبدال الطرق التقليدية أو تقديم منهج وضعي أو تجريبي لعلماء العلوم الإنسانية (Berry 2011; Manovich 2011).

إن ما توفره العلوم الإنسانية الرقمية هو وسيلة لمعالجة ومشاركة البيانات، وزيادة التحليل التقليدي، وتعزيز البحوث عبر مدى وحجم أكبر بكثير من المصادر، وتوفير الوقت والجهد. وبدلاً من التركيز على حفنة من روايات العصر الفيكتوري أو صور من أوائل القرن العشرين أو اثنين من فناني عصر النهضة وأعمالهم، يصبح من الممكن البحث والربط عبر عدد كبير للغاية من الأعمال ذات الصلة، وبدلاً من التركيز على حفنة من المواقع الشبكية أو غرف الدردشة أو الصور أو الفيديوهات أو الصحف الإلكترونية، يصبح من الممكن دراسة الآلاف من مثل هذه الوسائط. ومن ثم فلقد ادعى (Manovich 2011) أن: "الدراسة المعززة بالحاسب لمجموعات بيانات ثقافية هائلة عادةً ما تكشف أمثاطاً جديدة للبيانات التي لا يمكن حتى من خلال القراءة المتأنية لأفضل دليل أن تكشفها". وبالطبع، فإن جيشاً من علماء العلوم الإنسانية لن يكون قادراً بشكل حريص على القراءة المتأنية لمجموعات بيانات هائلة في المقام الأول. وادعى (Moretti 2005) أن العلوم الإنسانية الرقمية ستطبق الصرامة المنهجية على التخصصات التي كانت في السابق غير نظامية وعشوائية في تركيزها ومنهجها المتبع. وادعى أن الأدبيات ذات الصلة بالأعداد، والأشكال، والوسائط المرئية من المحتمل أن توفر رؤى جديدة وتضفي رونقاً جديداً على المجالات المنهكة.

وفي الوقت ذاته، وكما أن هناك فرصاً، هناك أيضاً مجموعة من التحديات. فبالنسبة للمعارضين، فإن استخدام طرق العلوم الإنسانية الرقمية الجديدة سوف تعزز ما أطلق عليه (2005) Moretti مصطلح "القراءة عن بعد" حيث يمكن لأجهزة الحاسب أن تعمل مثل آلات القراءة (Ramsay 2010). فمثل هذا النهج يتعارض مع الممارسة الراسخة منذ أمد بعيد والمتعلقة بتوفير قراءة متأنية للمصدر. وقد ادعى (2010: 20) Culler على سبيل المثال بأن: "ما يتعارض مع القراءة المتأنية ليس القراءة عن بُعد ولكن شيئاً ما مثل القراءة غير المتقنة أو القراءة العرضية". واستمر قائلاً بأن القراءة المتأنية تتضمن الاهتمام بكيفية إنتاج أو نقل المعنى، وبنوع الإستراتيجيات والأساليب الأدبية والبلاغية المستخدمة من أجل تحقيق ما يُمكن للقارئ أن يأخذ به على أنه المؤثر على العمل أو الفقرة (ص 22)، وهذا شيء لا يمكن للقراءة عن بعد أن تقوم به، فقلق Culler ناتج عن أن العلوم الإنسانية الرقمية لا تسهل أو تعزز "القراءة عن بعد" فحسب، بل إنها تجعل "من الممكن إجراء البحوث الأدبية بدون القراءة على الإطلاق: مثل إيجاد كافة الحالات التي وردت فيها كلمتي "يتسول" و"متسول" في روايات كتبها اثنان من المؤلفين ثم كتابة النتائج" (ص 24). وعلى نحو مشابه، ذكرت (2009: 164) Trumpener بأن: "النموذج المقاد إحصائياً لتاريخ الأدب... يبدو أنه يتطلب يداً مجهولة غير مرئية. وعلاوة على ذلك، فإنه يحدد الأنماط ولكنه يتجنب تحديد السببية، والتي لا يمكن تحديدها من خلال تتبع محفوظات الناشرين، وقراءة مسودات المخطوطات الفردية في مكتبات الكتب النادرة ومحاولة فهم، من كتاب لكتاب، من قام بتحديد عنوان كل رواية: المؤلف، الناشر، أو وكيل الدعاية" (ص 164). وبالنسبة لها، فإن مثل هذا البحث يتضمن حركة حقيقية وليس استخدام مجموعة من الخوارزميات. كما أكدت أن تحليل آلاف النصوص يعني "أن أي نص معين سوف يصبح غير ذي صلة من الناحية الإحصائية" (ص 164). واستمرت قائلةً "أي محاولة لرؤية الصورة بالكامل تحتاج إلى دعم من قبل المعرفة الواسعة وفهم تاريخي فطن بكيفية عمل الأساليب والمؤسسات الأدبية والأدوات التفسيرية القاطعة. وفي معرض تقديرها للجزافية"، اختتمت كلامها (ص 170 - 171) قائلة إنه من "المهم أن معظمنا تخلى عن الاستمرار في البقاء في المكتبة... فاليأس من عدم توافر المعرفة الشاملة يجب ألا يدفعنا إلى التفكير المتطرف المعاكس بأنه يجب علينا أن نبدأ في معالجة الأدبيات بكميات هائلة لكي نحرز أي تقدم".

وبالمثل، فبالنسبة لـ (Marche 2012)، فإن الآثار الثقافية مثل الأدبيات لا يمكن معاملتها كونها مجرد بيانات. فالجزئية الأدبية المكتوبة لا يمكن أن تكون ببساطة مجرد ترتيب للحروف والكلمات، بل هي مقترنة بالسياق وتنقل المعنى ولها سمات تفوق الوصف. فالخوارزميات ضعيفة للغاية فيما يتعلق بالتقاط وفك تشفير المعنى أو السياق، وكما ادعى Marche فإن الخوارزميات تعامل جميع الأدبيات كما لو كانت هي نفسها. واستمر قائلاً: "إن التحليل الخوارزمي للروايات والمقالات في الجرائد يكون بالضرورة ضمن حدود الاختزالية. إن عملية تحويل الأدب إلى بيانات تزيل التفوق ذاته، وهي تزيل التذوق وتزيل جميع الدقة من النقد، وتزيل التاريخ المرتبط بإطلاق هذا العمل، فرواية "إلى المنارة" ما هي إلى رواية أخرى من ضمن ركام الروايات".

وعلاوةً على ذلك، فبالنسبة للآخرين، فبدلاً من العمل على المصادر الأصلية، فإن العلوم الإنسانية الرقمية تتعامل مع الأشياء الرقمية، التي تُنقل دفعة واحدة وتختلف من الناحية الكيفية عن أصولها (بمعنى أنها وصف للوصف) وذلك من شأنه أن يؤثر بشكل حتمي على تفسيرها.

وبالنسبة للكثيرين، فإن العلوم الإنسانية الرقمية، في هذه الحالة، تعزز التحليلات السطحية الضعيفة بدلاً من الرؤى العميقة المتخلخلة. ولذا، يتم نقدها على اختزالياتها المفرطة وعدم إتقان أساليبها، وتضحيتها بالتعقيد، والتدقيق، والسياق، والعمق، ونقدها للأشكال الوصفية، والآلية، والمدرجة، والمتسعة، والانطباق بأن التفسير لا يتطلب معرفة سياقية عميقة. ومن ثم، فلقد ادعى (Jenkins 2013) أن: "قيمة الفنون وجودة المسرحية أو الرسم لا يمكن قياسها على الإطلاق. فأنت بمقدورك وضع جميع أنواع البيانات في آلة: التواريخ، والألوان، والصور، وإيصالات شبك التذاكر، ولكن لا يمكن لأي منهم أن يفسر ما هو العمل الفني، وما الذي يعنيه، وما السبب في قوته، فالإجابة عن هذه الأسئلة تتطلب إنساناً وليس آلة. وفي حين أنني متأكد أن البيانات الكبيرة مفيدة، إلا أنها لا تنطبق على جميع مجالات الحياة البشرية، فهي لن تحسن الظروف البشرية. فالحب، والثقافة، والفن، والسياسة أمور ضرورية في حياتنا ولكن البيانات الكبيرة تخبرنا القليل عن هذه الأمور. وباختصار، لا تصدق جميع الدعايات عن البيانات، فهي ليست الإجابة عن الأسئلة الكبيرة".

وعلاوةً على ذلك، فإن التمويل المحدود المخصص للعلوم الإنسانية تتم إعادة تركيزه على مشروعات العلوم الإنسانية الرقمية وذلك على حساب الدراسات الأكثر تقليديةً مما يحرم بعض تخصصات العلم من التمويل. وتسري تأثيرات عمليات إعادة توزيع التمويل أيضاً داخل العلوم الإنسانية الرقمية ذاتها، حيث غالباً ما يتم توجيه التمويل تجاه بناء البنى التحتية للبيانات وإنشاء أدوات جديدة، وذلك بدلاً من تمويل مشروعات تستفيد من المجموعات الرقمية من أجل الإجابة على الأسئلة الأكثر جوهرية. والنتيجة توجيه التمويل تجاه هندسة البرمجيات، وتطوير أدوات تحليلية، ومعدات، ومختبرات، ومشروعات الموارد الأخرى، وذلك بدلاً من تمويل العلوم الإنسانية بذاتها.

ويمكن أن تنطبق ذات أنواع المناقشات هذه فيما يتعلق بالعلوم الاجتماعية الحاسوبية. فعلى سبيل المثال، في مناقشة تتعلق بتأليف التصوير في ويكيبيديا، ادعت فيرناندا فيجاس (Frenanda B. Viegas) والتي تعمل في مختبر الاتصالات المرئية في (IBM) أن التخطيط المرئي للبيانات يمكن أن "يساعد الباحثين على تحديد الأسئلة التي ربما يمكن استكشافها من الناحية الإحصائية - على سبيل المثال، ما مدى تكرار حدوث التخريب، وما مدى سرعة ارتداد النص؟" (مقتبس في 11: Bollier 2010). وعلى الرغم من ذلك، تعد مثل هذه الأسئلة بسيطة نسبياً. ففي هذه الحالة، ربما يكشف التخطيط المرئي أن موضوعاً ما تتم مناقشته بحماسة بالغة والمدى الذي يمكن أن يجعل محرري ويكيبيديا يتدخلون في عملية تحرير مدخلات البيانات، ولكن التخطيط المرئي يفعل القليل للإجابة على الأسئلة الأكثر نفعاً وتشويقاً وكشفاً، مثل أسباب وطبيعة النزاعات والقصص البديلة التي يتم إنتاجها، ولا تتم دراسة أي منها بشكل أمثل من خلال الإحصائيات. وعلى نحو مشابه، فإن خريطة لغة التغريدات في مدينة ما ربما تكشف أنماط التركيز الجغرافي لجماعات عرقية متعددة، ولكن الأسئلة المهمة هي من يشكل مثل هذه التركيزات، وما السبب في وجودها، وما هي العمليات التي أسهمت في التشكيل والتكاثر، وما هي نتائجها الاجتماعية والاقتصادية. إن تحديد الأنماط شيء؛ ولكن تفسير هذه الأنماط شيء مختلف تماماً، حيث يتطلب نظرية اجتماعية ومعرفة سياقية عميقة. وفي مثل هذه الحالات، فإن الخطر يكمن في أن البحوث تمارس ببساطة بسبب توافر البيانات، مما يحد بشكل كبير من الأسئلة التي يمكن توجيهها وذلك لأن البيانات لم يتم إنتاجها مع وجود مثل هذه الأسئلة على البال (Vis 2013). وكما

ادعى (Gonzalez Bailom 2013): "إن تخطيط التوزيع المكاني للمشاعر الإيجابية، أو التكرار الذي يتم من خلاله ذكر بعض الكلمات في التواصل على الإنترنت لا يخبرنا الكثير عن توافق هذه الأنماط مع الديناميكيات الاجتماعية التي تشكلها وتخلقها... إن قياس الأشياء بسبب أنه يمكن قياسها فحسب، لا يجعلها مشوقة أو ذات صلة".

وفي حالات مثل خريطة التغريدات، فإن الأنماط المكتشفة نادراً ما يتعين أن تكون نقطة النهاية وهو ما يبدو عليه الحال في بعض المشروعات (Rogers 2013). والأحرى أن تمثل الأنماط المكتشفة نقطة البداية مما يتطلب بالتأكيد أشكالاً أخرى من التحليلات وربما بعض مجموعات البيانات الأخرى.

وكما هو الحال مع الانتقادات الأولى للعلوم الاجتماعية الكمية، فإن العلوم الاجتماعية الحاسوبية يُزج بها في مهام من قبل ما بعد الوضعيين كونها آلية، ومفتتة ومحددة، وتُقلص الأفراد المتنوعين، والبنى الاجتماعية المعقدة متعددة الأبعاد، إلى مجرد نقاط بيانات (Wyly منشور في الصحافة). وعلاوةً على ذلك، فإن التحليل يفسد من قبل افتراضات الحتمية الاجتماعية كما أوضح Pentland (2012) ذلك بالأمثلة قائلاً: "إن نوعية الشخص الذي تكون عليه تتحدد بشكل كبير من قبل بيئتك الاجتماعية، ولذا إذا كان بإمكانك رؤية بعض سلوكياتك، فيمكنني أن أستنتج الباقي، وذلك فحسب من خلال مقارنة مع الأشخاص في جماعتك". وعلى النقيض من ذلك، يدعي البعض أن المجتمعات البشرية معقدة للغاية وفوضوية وغير متوقعة لكي يتم تحويلها إلى صيغ وقوانين، حيث توفر النماذج الكمية رؤى قليلة لتفسير الظواهر مثل الحروب، والإبادة الجماعية، والعنف الأسري، والعنصرية. وعلى الرغم من أن البيانات الكمية يمكنها أن توضح بعض هذه المجالات، فإنها توفر رؤى شاملة فحسب للأنظمة البشرية الأخرى مثل الاقتصاد، وهي تفسر بشكل غير كافٍ دور السياسة، والأيدلوجيا، والبنى الاجتماعية، والثقافة (Harvey 1972). إن الناس لا يتصرفون بطرق منطقية ومحددة سلفاً، ولكن بدلاً من ذلك فهم يعيشون حياة مليئة بالتناقضات، والمفارقات، والحوادث غير المتوقعة، فالكيفية التي يتم من خلالها تنظيم وعمل المجتمعات تختلف عبر الزمان والمكان، ولا يوجد شكل مثالي أو أفضل أو سمات عالمية. وفي الواقع، يوجد تنوع مدهش في الأفراد، والثقافات، وأساليب

الحياة على الكوكب. فتقليص مثل هذا التعقيد إلى الموضوعات المجردة التي تشغل النماذج العالمية يؤدي إلى الكثير من العنف الرمزي فيما يتعلق بخلقنا للمعرفة. وعلاوةً على ذلك، فإن المناهج الوضعية تتجاهل عن عمد المجالات الميتافيزيقية (ما وراء الطبيعة) للحياة البشرية (المتعلقة بالمعاني، والمعتقدات، والخبرات) والأسئلة المعيارية (المعضلات الأخلاقية والمعنوية بشأن كيف تكون الأشياء على النحو الذي يجب أن تكون عليه مقارنةً بالوضع الحالي لهذه الأشياء) (Kitchen 2006). وبعبارة أخرى، فإن المناهج الوضعية تركز فقط على أنواع معينة من الأسئلة، والتي تسعى إلى الإجابة عنها بطريقة اختزالية تتجاهل على ما يبدو ما الذي يعنيه أن يكون المرء إنساناً وأن يعيش في مجتمعات وأماكن خصبة التنوع، وهذا لا يعني أن المناهج الكمية غير مفيدة بل هي مفيدة بشكل واضح للغاية، وخاصةً فيما يتعلق بتفسير ونمذجة الأنظمة الآلية ولكن ينبغي فهم جوانب قصورها فيما يتعلق بفهم الحياة البشرية ودمجها مع المناهج الأخرى.

ومن ثم فلقد ادعى Brooks (2013b) أن تحليلات البيانات الضخمة تتصارع مع الطبيعة الاجتماعية كون الناس غير منطقيين ولا يتصرفون بشكل متوقع وكون الأنظمة البشرية معقدة بشكل غير معقول ولديها علاقات متناقضة ومتباينة. وهي تتصارع كذلك مع السياق كون البيانات منفصلة بشكل كبير عن السياق الاجتماعي، والاقتصادي، والسياسي، والتاريخي، وهي أيضاً تخلق أكواماً هائلة من القش كونها تتكون من ارتباطات غير طبيعية هائلة للغاية مما يجعل من الصعوبة بمكان تحديد مكان الإبرة أو ما نبحت عنه وكونها تواجه مشاكل عند مجابهة المشاكل الكبيرة وخاصةً المشاكل الاقتصادية والاجتماعية. وهي كذلك تفضل المستنسخ على الأصل كونها تحدد الاتجاهات ولكنها لا تحدد بالضرورة السمات المهمة التي ربما تتحول إلى اتجاه وكونها تحجب القيم والمصالح المتعلقة بمنتجي البيانات وهؤلاء الذي يحللون البيانات وأهدافهم.

وهم يفعلون ذلك، لأنه وكما قال Peter Could (1981: 174)، وهو مؤيد للمناهج الإحصائية في تحليل البيانات: "يبدو أن رموز لغاتنا الرياضية مقيدة ومحددة بشكل كبير. وفي الغالب هم يحدثون أضراراً ليس لأنهم يُقيدوننا فحسب، ولكنهم في الواقع يحون من الوجود ما يمكن أن يكون مستحقاً لانتباهنا واهتمامنا. وإذا ما تصورنا، وهذا ما نفعله في

الغالب، الطبيعة متعددة الأبعاد التي تصف تعقيد حياتنا المعاصرة... والفراغات الناتجة بواسطة الفلتر الخطي (Strainer Called) المسمى تحليل العامل (Factor Analysis)، فرمها نقوم بسحق المعلومات خارج الوجود وذلك تحت مسمى التبسيط. إنني لست متأكدًا تمامًا من السبب وراء فعل ذلك، عدا أن لدي شعورًا مؤكدًا تقريباً بأن توافر أجهزة حاسوب كبيرة وسريعة، وكذلك البرامج الجاهزة له يد في ذلك".

وبالنسبة لـ Gould، فإن فهم السلوك البشري والمجتمعات البشرية لا يمكن ولا يجب تحويله إلى تحليلات متكررة ومنهجية وآلية، ولقد دعا إلى التفكير المستدام بشأن أنواع الأساليب التي يتعين استخدامها مع أي نوع من البيانات، وتحت أي ظروف، وذلك من أجل الإجابة عن بعض الأسئلة المحددة، بدلاً من تشغيل البيانات من خلال مصنع "نقانق" إحصائي والذي ينتج لحوماً منخفضة الجودة من الألف إلى الياء بدلاً من تقطيع اللحوم المنتقاة.

وفيما يتجاوز الأسلوب المنهجي والمعرفي، فإن جزءاً من القضية هو أنه يبدو أن بعض تحليلات البيانات الكبيرة تحدث دون وجود أي أسئلة على البال، أو أن يكون التركيز مدفوعاً من قبل تطبيق طريقة أو محتوى لمجموعة البيانات بدلاً من سؤال محدد، أو أن تكون مجموعة البيانات مستخدمة من أجل البحث عن إجابات عن أسئلة لم يتم تصميمها لكي يتم الإجابة عنها في المقام الأول. وفيما يتعلق بالآخر، لم يتم إنتاج بيانات تويتر ذات المكان المحدد جغرافياً من أجل تقديم إجابات تتعلق بالتركيز الجغرافي لمجموعات اللغة في مدينة ما والعمليات التي تقود مثل هذا الارتباط الآلي المكاني. فنحن يجب علينا ألا نشعر بالدهشة ربما في حالة ما إذا كانت النتيجة لقطات سطحية فحسب، حتى وإن كانت لقطه رائعة، بدلاً من ذلك فإن ما يجب أن يدهشنا هو الرؤى المتعمقة والمتغلغلة في جغرافيات الجنس، واللغة، والتكتلات، والانعزال في أماكن معينة، ولهذه الأسباب، ستظل دراسات البيانات الصغيرة مفيدة.

وفي معرض الدفاع عن المنهج المستخدم من قبل العلوم الإنسانية الرقمية، ادعى (Manovich 2011) أن الهدف ليس استبدال الخبراء البشريين بأجهزة الحاسب، ولكن الهدف هو استخدام التحليلات من أجل استكشاف مجموعات البيانات الهائلة مثل بلايين

التغريدات، أو ملايين الفيديوهات على موقع يوتيوب، أو عشرات الآلاف من الروايات ومن ثم اختيار أهداف تمثيلية معينة من أجل التحليل اليدوي المُحكم. وبعبارة أخرى، فإن منهج العلوم الإنسانية الرقمية يفتح أمامنا إمكانية إجراء الدراسات السطحية على نطاق واسع عبر الأشخاص بدلاً من الدراسات المتعمقة التي تركز على أعداد قليلة من الأفراد (Manovixh 2011). وبالمثل، يدعي البعض أن هذا المنهج يوفر العمل الدؤوب للتعلم في آلاف صفحات الأخبار ومشاهدة مئات الساعات من البرمجة، لكي "يتمكن الباحثين من تركيز طاقاتهم بشكل فعلي على التفسير" (Harris 2012)، ولذا، اقترح Flaounas et al. (2013: 111) أن: "أتمتة الكثير من المهام في تحليل المحتوى الإخباري لن يحل محل الحكم البشري المطلوب من أجل أشكال التحليل النوعية والدقيقة، ولكنه سوف يسمح للباحثين بتركيز اهتمامهم على مدى أبعد بكثير من أحجام عينة الأشكال التقليدية من تحليل المحتوى. وبدلاً من قضاء الكثير من العمل الثمين على مرحلة ترميز البيانات الخام، يمكن للمحللين التركيز على تصميم التجارب والمقارنات من أجل اختبار فرضياتهم، وأن يتركوا لأجهزة الحاسب مهمة اكتشاف جميع المقالات في موضوع معين، وقياس العديد من سمات المحتوى مثل قابليتهم للقراءة، واستخدام أشكال معينة من اللغة والمصادر، إلخ. مجرد عدد قليل من المهام التي يمكن أتمتها في الوقت الراهن".

وبالمثل، فلقد ادعى (Selisker 2012) أن العلوم الإنسانية الرقمية لا تسعى إلى "أن تستبدل الأدب برموز الحاسب، أو أن تقايس التعقيد الأدبي بصيغ لا هدف لها. وبدلاً من ذلك، فإن هذه المشروعات تفكر فحسب بطرق إبداعية عن المشاكل التاريخية الأدبية التي لا يمكن التعامل معها بسهولة بدون الحواسيب". وبالنسبة لـ (Ramsay 2003; 2010)، يمكن أن تساعد الحوسبة في عملية التفسير، ولقد ادعى أن القراءة المتأنية تتضمن ممارسات اختيار وعزل وملاحظة الأنماط في النصوص، وأن الترميز ذو كفاءة عالية في مثل هذه الممارسات، وعلاوةً على ذلك، فإن الترميز يقوم بالأمر نفسه بطرق أكثر موضوعيةً واتساقاً من النظرة الذاتية للعالم، وبدلاً من استبدال التعرف على المعنى، فبالنسبة له فإن توظيف الخوارزميات يسهل هذه العملية.

إن دفاع العلوم الاجتماعية الحاسوبية أمر مختلف تماماً في الغالب، ففي حين أن جميع العلوم الإنسانية الرقمية تقر بقيمة القراءة المتأنية، وتؤكد أن القراءة عن بعد تكملها من خلال توفير التعمق والسياقية، فإن الأساليب الموضوعية للعلوم الاجتماعية تتعارض مع المناهج بعد الوضعية. إن الفرق بين العلوم الاجتماعية الحاسوبية والعلوم الإنسانية الرقمية بهذا الصدد يرجع إلى أن الإحصائيات المستخدمة في العلوم الإنسانية الرقمية وصفية بشكل كبير تحدد الأنماط وتصوغهم كأعداد وصور وخرائط. وعلى النقيض من ذلك، فإن العلوم الاجتماعية الحاسوبية تستخدم المنهج العلمي حيث تكمل الإحصائيات الوصفية بالإحصائيات الاستدلالية التي تسعى إلى تحديد السببية. وبعبارة أخرى، فإنها تحصل على الدعم من قبل المعرفة حيث يكون الهدف هو تقديم نماذج إحصائية متطورة تفسر وتحاكي وتتنبأ بالحياة البشرية. ومن الصعوبة بمكان التوفيق بين هذا وبين مناهج ما بعد الوضعية، فالدفاع في هذه الحالة يركز على فائدة وقيمة الطريقة والمناهج، وليس على توفير تحليل تكميلي لمجموعة من البيانات أكثر تَمْدَاداً.

وهناك بدائل لمثل هذا الموقف، مثل البديل الذي تم تبنيه في نظم المعلومات الجغرافية (GIS) والإحصائيات الجذرية (Radical Statistics)، وهؤلاء الذين يستخدمون مناهج ذات طرق مختلطة، وهذا يتضمن إما توظيف النماذج والإحصائيات الاستدلالية شريطة أن يكون هنالك إدراك لمواطن الضعف فيهما، أو، كما يحدث على نحو شائع، استخدام الإحصائيات الوصفية التي تدعمها دراسات البيانات الصغيرة. وهنا، يتم استخدام كلٍّ من البيانات الكبيرة والبيانات الصغيرة من أجل توفير رؤى مختلفة ولكن تكميلية. وبغض النظر عن مثل هذه التكميلية، فإن الكثير من علماء ما بعد الوضعية سوف يستمرون في مقاومة البيانات الكبيرة والمناهج الإحصائية بأشكالها المتعددة، وبلا شك فإن المجادلات سوف تتسارع وتيرتها في المستقبل المنظور.

الخاتمة:

هناك شكوك قليلة بأن ثورة البيانات أتاحت احتمالية إعادة التشكيل المعرفي للعلوم، والعلوم الاجتماعية والإنسانية، وأن مثل إعادة التشكيل هذه تحدث بالفعل بشكل فعال ونشط. فالبيانات الكبيرة وتوسع البيانات الصغيرة من خلال البنى التحتية للبيانات يساعدان على تنفيذ مناهج جديدة تجاه إنتاج وتحليل البيانات، مما يجعل من الممكن توجيه الأسئلة والإجابة عنها بطرق جديدة. وقد شجع هذا البعض على الادعاء بشكل جريء بأن ثورة البيانات هي ابتكار تدميري يعلن عن نقلة نوعية في كيفية إجراء البحوث، وهذه بالقطع هي الحالة حيث إن التحول من صحاري البيانات إلى طوفان البيانات وتحليلات البيانات الجديدة سيغير الممارسات البحثية. وعلى الرغم من ذلك، يتعين أخذ الحيطة عند القيام بتأكيدات تتعلق بإعادة تشكيل البحث وبأفضل طريق يمكن اتباعه في إعادة التشكيل هذه.

إن العلم يُعامل بشكل سيئ من قبل الادعاءات القائلة بأن ثورة البيانات تعني "أن البيانات يمكنها أن تتحدث عن نفسها" أو أنه يمكنها بدون أي مشاكل استخدام أساليب مثل تجريف البيانات حيث يسبق الارتباط السببية، (انظر الفصل التاسع). وتُعد المناهج التي يتم تطويرها داخل العلوم المقادة بالبيانات والتي تمزج بين الاستدلال، والاستنتاج والاستقراء أكثر إنتاجية بكثير، حيث تتم إعادة تشكيل الطريقة العلمية، وذلك بدلاً من تدميرها، إن الأسس الفلسفية للعلم المقاد بالبيانات، على الرغم من ذلك، لا تزال في أطوارها التمهيدية وهناك حاجة للتفكير المستفيض في معتقداتها المعرفية ومبادئها ومنهجيتها.

أما الموقف فيما يتعلق بالعلوم الاجتماعية والعلوم الإنسانية فهو أكثر تعقيداً بشكل ما وذلك إذا وضعنا في الاعتبار تنوع الأسس الفلسفية لهذه العلوم، حيث إنه من غير المحتمل أن تقود ثورة البيانات إلى تأسيس نموذج جديد. وبدلاً من ذلك، فإن البنى التحتية للبيانات وثورة البيانات سوف يعززان مجموعة البيانات المتوافرة للتحليل ويمكنان مناهج وأساليب جديدة، ولكنها لن تحل محل دراسات البيانات الصغيرة. وفيما يتعلق بالعلوم، هناك حاجة لتفكير نقدي أوسع في الآثار المعرفية لثورة البيانات على العلوم الاجتماعية والعلوم الإنسانية، وهذه المهمة بدأ بالكاد التعامل معها وذلك على الرغم من سرعة التغيير الحاصل في مشهد البيانات.

۲۰۶

الفصل التاسع

القضايا الفنية والتنظيمية

(Technical and Organizational Issues)

كما ناقشنا في الفصلين السابقين، توفرُّ البنى التحتية للبيانات ومبادرات البيانات المفتوحة وإنتاج البيانات الكبيرة فرصاً جديدة لقياس وفهم العالم. وهي أيضاً تثير قضايا فنية وتنظيمية، وتطرح أسئلة معيارية وأخلاقية. وفي هذا الفصل، سيتم التركيز على الأمر الأول، بما يعني التركيز على بعض القضايا مثل مجال مجموعات البيانات، والوصول إلى البيانات، وجودة البيانات، وتكامل البيانات، والتبادلية (القابلية للتبادل)، وتطبيق التحليلات والمغالطات البيئية، والمهارات والقدرات والإمكانات التنظيمية. فبعض هذه القضايا يمكن التعامل معها من خلال الحلول الإدارية والفنية، في حين أن البعض الآخر من هذه القضايا تمثل مشاكل عسيرة الحل ومن الصعب التعامل معها. وعلى الرغم من ذلك، فإذا كانت القضايا المتعلقة بجودة البيانات والمغالطات البيئية لا يمكن التعامل معها بشكل كاف، ففي هذه الحالة لا يهم ماهية النظريات المعرفية التي يتم تطبيقها، (انظر الفصل الثامن)، لأن التحليلات والتفسيرات ستكون محل ارتياب. وعلاوة على ذلك، فإذا كان الوصول إلى البيانات مقيداً بشكل كبير، أو إذا كانت الحكومات والشركات والمجتمع المدني يفتقرون للمهارات الكافية لإدراك المغزى من هذه البيانات، ففي هذه الحالة ستكون قيمة ومنافع البيانات المفتوحة والكبيرة المتدرجة مقيدة بشكل كبير للغاية. ونتيجة لذلك، فإن القضايا التي تمت مناقشتها في هذا الفصل لا يمكن تجاهلها أو إبعادها بسهولة، ومن ثم فهي تتطلب الاهتمام البالغ. وفي الفصل التالي، سيتحول التركيز إلى قضايا سياسية واجتماعية أعم وأشمل والتي تتعلق بإنشاء ومشاركة وتحليل واستخدام البيانات مع الوضع في الاعتبار الأبعاد الأخلاقية، والقانونية، والأمنية لثورة البيانات.

الصحاري والفيضانات (Deserts and Deluges):

في كافة النقاشات الدائرة بشأن ثورة البيانات يكون في الغالب من الصعب تحديد مصدر مجموعة مناسبة من البيانات فيما يتعلق بالكثير من الظواهر، وبخاصة تلك البيانات التي تغطي فترات زمنية طويلة ومتأصلة مكانياً. وبشكل جزئي، فإن هذه تعد قضية من قضايا التغطية من ناحية، ومن قضايا الوصول من ناحية أخرى، فمن جانب التغطية، فإن جميع مصادر البيانات محددة من ناحية العينة، والمدى الجغرافي، والوقت، ومتغيرات الصفات (الحقول) والمؤشرات. وفي الواقع، فإن العالم معقد بشكل كبير للغاية أنه يستحيل التقاط نطاق كامل وكافة فروقاته الدقيقة، وتناقضاته، وتبايناته.

إن البيانات التي يتم جمعها من خلال دراسات البيانات الصغيرة تعد عينات بشكل صريح للغاية، وهي تجمع بشكل غير مستمر، وهي ذات رقم محدد من المتغيرات وذلك بالنظر إلى تركيزها المحكم والتكلفة المرتبطة بالجمع والتحليل، فمعظم هذه الدراسات تمثل دراسات أو تجارب منفصلة تركز على قضايا وحالات معينة لمجموعة، مؤسسة، أنواع، أماكن، تقنيات... إلخ، وهي تفتقر إلى المكون الطولي حيث إنها لا تغطي فترات زمنية متعددة وطويلة. فالدراسات التي تقوم بدراسة التغيرات بمرور الوقت عادة ما يتم تكرارها على فترات محددة. وفي هذه الحالة، فإن فهمنا لظاهرة معينة عادة ما يعتمد على مجموعة مقيدة من البيانات، وحتى عندما يتم تجميع وترقية مجموعات البيانات داخل البنى التحتية للبيانات، فإن التغطية تكون جزئية وانتقائية. فعلى سبيل المثال، فإن مجموعات بيانات الإدارة العامة الأيرلندية ذات تغطية محدودة وهي مملوءة بالفجوات والسكنات عن قطاعات معينة مثل قطاعات الإسكان والنقل، وغالباً ما يتم الاحتفاظ بها داخل المؤسسات (Kichin et al. 2007).

وعلى نحو مشابه، فإن البيانات الكبيرة ربما تميل إلى أن تكون بيانات شاملة، ولكن وكما هو الحال مع مجموعات البيانات الأخرى، فإن كليهما تكون تمثيلاً وعينة، وهي تحل، بشكل جزئي فحسب، قضية صحراء البيانات التي يقصد بها أن تكون البيانات موزعة على مجاميع بيانات متفرقة يصعب استقاء البيانات منها أو إضافة بيانات إليها. وكما لاحظنا في الفصل الثاني، فإن البيانات التي يتم التقاطها تتشكل من قبل التقنية المستخدمة، والسياق

الذي تم فيه إنشاء هذه البيانات، وتوصيف البيانات الذي تم توظيفه. فالبيانات المنشئة تمثل عادةً ما يمكن التقاطه إلى حد ما بطريقة مباشرة، وهي تكون في غالب الأحيان منتجاً فرعياً للنظام، ويكون مجالها الفوري مقيداً من قبل النطاق كسلاسل الأسواق المركزية، وشبكات المستشعرات، ومواقع التواصل الاجتماعي. وفيما يتعلق بالعلوم، فإن البيانات الكبيرة يتم إنشاؤها حالياً لما يتعلق بمجموعة منتقاة من الموضوعات فقط، وهي بعيدة كل البعد عن أن يتم إنتاجها عبر جميع النطاقات والبؤر العلمية. وعلاوة على ذلك، فإن القرار فيما يتعلق بإنشاء البيانات الكبيرة لم يتشكل بعد، ولكنه يتطور ويتركز من أجل توفير مجموعة من الرؤى، بحيث تشكل كل واحدة منها بعضاً من العناصر التي يتم التقاطها كالاتساع، والعمق، والتفاصيل... إلخ. وحتى عندما يتم دمج بيانات كبيرة مع مجموعات البيانات الأخرى، فإن مجالهم يكون بعيداً كل البعد عن كونه مجالاً عالمياً. ولذا، فإن البيانات الكبيرة، في هذه الحالة، لا تعالج مسألة صحاري البيانات وذلك فيما يتعلق بالكثير من مجالات الدراسة.

وفي الوقت نفسه، وفي حين أن صحاري البيانات مازال تواجهها مستمراً، فإن البنى التحتية للبيانات والبيانات الكبيرة أنشأ طوفاناً من البيانات في بعض المجالات. وكما لاحظنا في الفصل الرابع، كان هناك نمو ملحوظ في حجم وسرعة إنتاج البيانات والتي تم تحديدها من أجل أن تنمو بسرعة عالية، وذلك مع إنشاء وبدء المصادر الجديدة، وتعدد مجموعات البيانات التي تم إنتاجها، والمقدرة على ربط البيانات بعضها البعض. ومن شأن هذا الطوفان أن يفرض عدداً من التحديات فيما يتعلق بالمناولة، والمعالجة، والتأمين، والتصنيف، والفهم لمثل هذه البيانات (Gantz and Reinsel 2011). وبدلاً من مجموعات البيانات الصغيرة وغير المستمرة، يحاول المحللون الاستفادة من الكميات الهائلة من البيانات التي يتم تدفقها من أنبوب سميكة. فمعظم هذه البيانات أصبحت بيانات عابرة وتم فقدها، ولم يتم أبداً تحليلها أو الاستفادة منها. وبعض هذه البيانات قد تم تخزينها في مستودعات البيانات التي تمثل إلى حد كبير مكبات غير مراقبة للنفايات (Asay 2013)، وهناك يتم الإبقاء عليها حتى تأتي أوقات يمكن لهذه البيانات أن تقدم فيها معلومات مفيدة. ونتيجة لذلك، وبينما تزداد نسبة البيانات بشكل تصاعدي، فإن نسبة البيانات التي يتم معالجتها وتحليلها تقل بشكل كبير (Zikopoulos et al 2012). وقد أشار Wu (2012) إلى أنه في الوقت التي

تزداد فيه البيانات، يتم الحصول على معلومات أقل كنسبة من هذه البيانات. وبعبارة أخرى، تزداد ضوضاء البيانات بشكل كبير للغاية ولكن إشارات المعلومات تنمو بشكل أقل بكثير، وعلاوةً على ذلك، يصبح الأمر أكثر صعوبة لعزل هذه البيانات في خضم هذه الضوضاء (انظر Silver 2012). وهذا بلا جدال وظيفة البيانات الكبيرة وعلم البيانات حيث لا يزالان في أطوارهما الأولية، ومع الاستغلال الكامل للأساليب الحديثة والتي بإمكانها مراقبة ومعالجة وتحليل كميات هائلة من البيانات بشكل آني أو لحظي إلا أن هذا من المتعذر تحقيقه بشكل كبير. وحتى مع نضوج وتطور تحليلات البيانات الكبيرة، فإنه من المرجح أن تعاني من أجل تطوير طرق ومناهج يمكن من خلالها استخلاص التحليلات المفصلة وليست التحليلات السطحية؛ والتعامل مع أسئلة "لماذا" وكذلك أسئلة "ماذا؟" وتقديم إيضاحات وليس شروحات (انظر الفصل الثامن). وعلاوةً على ذلك، يتعين تنويع هذه الطرق وذلك من أجل مجاراة الأشكال الجديدة من البيانات التي يتم إنشاؤها من قبل المصادر المبتكرة. وحتى مع تحقيق ذلك، فإن مجاراة طوفان البيانات واستخلاص إشارات ذات مغزى من الضوضاء سيكون كفاحاً مستمراً.

الوصول (Access):

إن العائق الأساسي الآخر أمام الاستعانة بمصادر مجموعات مناسبة للبيانات هو مدى قابلية الوصول إلى هذه المجموعات وقيود التراخيص التي تحدد معايير استخدام تلك المجموعات. وبشكل واضح، وكما ناقشنا في الفصلين الثاني والثالث، فإن الوصول يعتبر قضية أساسية فيما يتعلق بالبحوث والبيانات الممولة من قبل المؤسسات العامة والتي تنشئها الهيئات العامة، فمحفوظات البيانات والبنى التحتية للبيانات ومبادرات البيانات المفتوحة تُعني وتهتم بشكل كبير بمشاركة وإنتاج البيانات المتاحة للتحليل. وكما أوضحت المناقشات في الفصول المشار إليها سابقاً، فعلى الرغم من حقيقة أن البيانات غير تنافسية، وغير قابلة للاستثناء أو الإقصاء، ولها تكلفة هامشية عند استنساخها (Floridi 2010)، إلا أن البيانات مقيدة بشكل شائع وذلك عند توزيعها. وفي بعض الأحيان، فإن الوصول المقيد إلى البيانات يكون أمراً مرغوباً فيه، وذلك إذا وضعنا في اعتبارنا طبيعتها الحساسة والشخصية وكذلك البعد الأخلاقي المرتبط بنشر واستخدام هذه البيانات (انظر الفصل العاشر). وفي

أحياناً أخرى، يكون الوصول إلى البيانات مقيداً لأجل التأكد من أن هؤلاء الذين يحتفظون بهذه البيانات سيرفعون من قيمة هذه البيانات ويضيفون ميزات تنافسية لها أو يحققون أموالاً من خلال البيع أو الترخيص للوصول لهذه البيانات (انظر الفصل السابع). وفي أحيان أخرى، ربما ترغب إحدى الهيئات في تقييد توزيع البيانات نتيجة أنها تخشى ما يُمكن أن تكشفه هذه البيانات، والذي ربما يكون له تداعيات سياسية أو اقتصادية، أو يكون مرد ذلك الخوف من كشف تدني جودة البيانات التي كانت محفوظة.

وفي حين أن هذه المعوقات بدأت تتلاشى ببطء فيما يتعلق بالبيانات الممولة والمستحدثة من قبل المؤسسات العامة في الدول الديمقراطية الغربية، إلا أن مثل هذه التحركات تعد أكثر تقلباً مع تلك البيانات المستحدثة من قبل المؤسسات الخاصة. وفي الواقع، فإن هذا الأمر يعد متناقضاً إلى حد ما، إذ إنه على الرغم من طوفان البيانات الكبيرة الناشئ، إلا أن الوصول إلى مثل هذه البيانات يظل مقيداً ومحدوداً في الوقت الحالي، وذلك بوجود عدد محدود للغاية من الكيانات الغارقة في مثل هذا الطوفان والتي يمكنها الوصول إليه (Crawford 2012, King 2011). فالشركات من أمثال مشغلي الهواتف الخلوية، ومطوري التطبيقات، ومزودي وسائل التواصل الاجتماعي، والمؤسسات المالية، وسلاسل التجزئة، وشركات المراقبة والشركات الأمنية، لا يقع على كواهلها أي التزامات لمشاركة البيانات التي يستحدثونها من خلال عملياتهم التشغيلية بحرية. وعادةً ما يتم التفاوض للوصول إلى البيانات بشكل فردي وهذا يتضمن توقيع العديد من اتفاقيات حماية الملكية الفكرية واتفاقيات عدم الإفصاح، وكذلك الموافقة على الشروط والبنود التي تحدد كيف وفي أي غرض سيتم استخدام أو لا يتم استخدام مثل هذه البيانات. وحتى بعد كل هذا، فمن المحتمل أن تمثل البيانات التي يتم توفيرها عينة محدودة، كما من الممكن أن تفتقر هذه البيانات إلى المعلومات السياقية مثل الكيفية التي تم تعيين هذه البيانات من خلالها، أو تكون البيانات غير معالجة بشكل كبير أي لم يتم تنقيتها أو هيكلتها، أو تكون مُجمعة، أو مُجردة من المتغيرات المفيدة، أو معالجة بشكل أو بآخر، وذلك من شأنه أن يُقيد أساليب تحليل هذه البيانات. وفي بعض الحالات، يمكن أن يتم إتاحة كمية محدودة من البيانات للباحثين والجمهور وذلك من خلال واجهات برمجة التطبيقات (Application Programming Interfaces - APIs). فواجهات برمجة التطبيقات تتكون من مجموعة

من الأوامر التي يمكن استخدامها من أجل استعادة البيانات المخزنة في قواعد بيانات على خادم متاح الوصول إليه للجمهور (Manovich 2011). فعلى سبيل المثال، يسمح تويتر لعدد محدود من الشركات بالوصول إلى كافة بياناته المتدفقة عبر موقعه، وذلك مقابل رسوم أو لأغراض تجارية. وهو يتمتع بالصلاحيات التي تسمح له بإملاء شروطه فيما يتعلق بالأغراض التي يمكن فيها استخدام مثل هذه البيانات. فالباحثون يتم تقسيمهم إلى قسمين: القسم الأول يمكنه الوصول إلى ١٠٪ من التغريدات العامة، والقسم الثاني لا يمكنه الوصول إلا إلى ١٪ فحسب من التغريدات العامة، أو يمكنهم الوصول إلى مجموعات فرعية مختلفة من المحتوى (الحسابات المدرجة في القائمة البيضاء)، وذلك مع استبعاد التغريدات الخاصة والمحمية في جميع الأحوال (boyd and Crawford 2012). وفي حال المواقع الإلكترونية التي تحوي على سبيل المثال قوائم منتجات وتقييماتها (Product Listings and Reviews) وأدلة أعمال (Business Directories)، وكذلك في حال وسائل التواصل الاجتماعي والتي تتضمن الملفات الشخصية العامة (Public Profiles) والتغذية (Feeds)، فإن بعض الشركات مثل (www.80legs.com) (80legs) وشركة (Mozenda) (www.mozenda.com) توفر خدمات زاحف الشبكة (أي تصفح الشبكة العالمية بطريقة منهجية وآلية ومنظمة) وتجريف الشبكة (أي استخراج البيانات من مواقع الإنترنت عن طريق برامج مخصصة) لحصد بيانات مفصلة من هذه المواقع أو للوصول لبيانات تغذية مبنية مسبقاً.

وما يقلق في هذا الأمر هو أن المعلومات التي يمكن أن توفرها البيانات الكبيرة المملوكة للشركات الخاصة والتي تم بيعها بشكل تجاري سوف تكون محصورة على قطاع الأعمال فحسب، أو ربما يتم إتاحتها لمجموعة مميزة من الباحثين الأكاديميين الذين لا يمكن تكرار النتائج التي توصلوا إليها أو التحقق منها (Lazer et al. 2009; King 2011). ولا أحد من هذه المواقع يخدم المصلحة العامة، ولكن إذا وضعنا في اعتبارنا القيمة التجارية للبيانات محل المناقشة فإن زيادة الوصول إلى هذه البيانات لن يكون مهمة سهلة أبداً. فحقيقة أن البيانات الكبيرة الثرية من الناحيتين الاجتماعية والثقافية مملوكة للشركات الخاصة تُشير إلى أنه في الوقت الحالي لن يكون من السهل في الواقع إجراء بحوث العلوم الاجتماعية الحاسوبية والعلوم الإنسانية الرقمية (Manovich 2011). فكما ذكر كلاً من

boyd and Crawford (2012: 673) فإن "الكثير من الحماس الذي يحيط بالبيانات الكبيرة ينبع من الإدراك القائل بأن هذه البيانات توفر وصولاً سهلاً لكميات هائلة من البيانات. ولكن من يتمتع بمثل هذه القدرة للوصول إلى هذه البيانات؟ ولأي غرض يتم الوصول؟ وفي أي سياق يجري هذا الأمر؟ وما معوقات الوصول؟" فهذه أسئلة تتطلب الكثير من التفكير مع اتخاذ إجراءات مناسبة.

جودة وصحة وأصل البيانات (Data Quality, Veracity and Lineage):

جودة وصحة البيانات هما أمران مهمان للغاية كونهما يعززان المدى الذي يمكن أن يثق فيه المرء بالنتائج المستخلصة من تحليل البيانات. وترتبط جودة البيانات بمدى تنظيمها أي خلوها من الأخطاء والفراغات، ونقائها أي خلوها من التحيز، وتماسكها بمعنى أن تحوي القليل من التناقضات. بينما يُشير مصطلح صحة البيانات إلى موثوقية البيانات وإلى مدى تمثيلها للمعنى الذي تُشير إليه بدقة (إتقان) وأمانة (إخلاص). ولهذا فإن تأسيس وتوثيق جودة وصحة البيانات يعتبر شقاً رئيسياً في سُلالة أصل البيانات. فسُلالة أصل البيانات هي: المعلومات التي تصف مصدر الملاحظات، والمنهجيات التي خضعت لها البيانات عند جمعها وتصنيفها، والتعديلات، والتحويلات، والتحليل، والاشتقاقات التي تعرضت لها، وهي توفر أيضاً الافتراضات والمعايير التي تم تطبيقها عند أي مرحلة من مراحل حركة هذه البيانات، وكذلك أي تحيزات... وتوفر سُلالة أصل البيانات مجموعة بيانات مع أصلها وتسمح للمستخدم أن يحدد مدى ملاءمة البيانات للاستخدام (Lauriault 2012).

وهي أيضاً تصف بالتفصيل منشأ البيانات، ومن شأن الجودة والصحة وسُلالة الأصل الضعيفة للبيانات أن يقوضنَ درجة الثقة والمصدقية التي يمكن أن يضعها الشخص في مجموعة البيانات والتحليلات التي اعتمدت على هذه البيانات.

ومن الناحية التقليدية، فإن الدراسات القائمة على بيانات نادرة والتي يكون حجم العينات فيها صغيراً في الغالب، فإن جودة وصحة وأصل البيانات تكون على قدر كبير من الأهمية. فإذا كان الشخص سيعمل على استخلاص نتائج عن السكان بالكامل وذلك اعتماداً على عينة مثالية صغيرة، فإنه يتعين في هذه الحالة أن تكون العينة واضحة للغاية

وخالية من العيوب بأكبر قدر ممكن، ويتعين أن تكون على درجة عالية من الدقة والصحة. وعلى الرغم من ذلك، فإن عدد مجموعات البيانات يتم تشكيلها بشكل تام محدود للغاية، ويمكن استخدامها بدون أي شكل من أشكال تنظيف البيانات، بمعنى التحقق من الأخطاء وتصحيحها، وإضافة البيانات المفقودة، ومعايرة وإعادة تهيئة البيانات حتى يمكن ربطها مع بيانات أخرى أو لتكون هذه البيانات سهلة الاستجابة لمناهج تحليل معينة (انظر الفصل السادس). ولذا، فإنه يتم استهلاك الكثير من العمل خلال تصميم البحث من أجل تقييد التحيزات المنهجية والتحيزات المرتبطة بتحديد العينات كي يتم التأكد من أن البيانات دقيقة وذات مصداقية بقدر الإمكان قبل تحليل أو مشاركة هذه البيانات. وتعمل المحفوظات الأرشيفية والبنى التحتية بشكل نشط من أجل التأكد من أن تظل مثل هذه البيانات ذات مصداقية ودقيقة وصحيحة بمرور الوقت، مع المحافظة على سُلالة أصل البيانات من خلال الاستقرار المؤسسي، وعمليات إدارة البيانات وواصفات البيانات، وتمكين نقل البيانات عبر مختلف الصيغات والمنصات بما يحقق المؤامة بين الأجيال المختلفة من التقنيات (Lauriault 2012). وبدون مثل هذا العمل، فإن صحة البيانات المتضمنة داخل تلك البيانات سوف تتلاشى في القريب العاجل وذلك مع انعدام الثقة.

وعلى النقيض من ذلك، ادعى البعض أن دراسات البيانات الكبيرة ليست بحاجة إلى تطبيق معايير جودة وصحة البيانات ذاتها، وذلك يرجع إلى أن الطبيعة الكلية لمجموعات البيانات الكبيرة "ن=الجميع" (n=all) تجعلها تتخلص من التحيزات المتعلقة بتحديد العينات، وتقوم بالتعويض، أكثر من المطلوب، عن أي أخطاء أو ثغرات أو تناقضات في البيانات أو أي ضعف في صحة البيانات (Mayer-Schonberger and Cukier 2013). فالحجة التي استندت إليها وجهة النظر هذه أنه "مع وجود عدد قليل من الأخطاء المرتبطة بتحديد العينة، يمكننا تقبل أخطاء أكبر مرتبطة بالقياس" (ص 13)، "ويمكننا التسامح مع عدم الدقة" (ص 16). إن الثمن المقابل لـ "ن=الجميع" (n=all) هو قبول الفوضى التي نشأت بشكل كبير من جراء صعوبة تنظيف البيانات ذات السرعة العالية وغير المنظمة، وتقبل البيانات كونها "محتملة أكثر من كونها دقيقة" (ص 35). وعند النظر إلى الأمر بهذه الطريقة، ادعى (Mayer-Schonberger and Cukier 2013) أنه "كلما كان هنالك بيانات أكثر، كان ذلك أفضل" أو للتعبير عن هذا الأمر بشكل آخر، "إذا كان لديك بيانات

كثيرة جداً فإن ما هو جيد بدرجة كافية يُعد جيداً بدرجة كافية" (Helland 2011). ومن ثم فقد ادعى (Franks 2012: 211) أن المطلوب هو "بيانات نظيفة بدرجة كافية"، وهي البيانات التي تمكن المحللين من "الحصول على النتائج التي يمكنهم الوثوق بها". وبالطبع فإن هذا يفترض أن جميع استخدامات البيانات الكبيرة سوف تتسامح مع عدم الدقة، في حين أن الواقع يشير إلى أن الكثير من استخدامات البيانات الكبيرة تتطلب الدقة بشكل كبير، أو على الأقل بيانات ذات معايير الأخطاء التي يمكن حسابها أي بيانات موثوقة. فعلى سبيل المثال، فإن موقع خريطة الطريق المفتوح (OpenStreetMap) يهدف إلى أن يكون في مستوى منتجات الخرائط الاحترافية من حيث التفصيل والدقة، ويهدف كذلك إلى كسب ثقة مستخدميه كونه مصدرًا من مصادر المعلومات. وعلى النحو ذاته، تهدف موسوعة ويكيبيديا (Wikipedia) إلى أن تكون موسوعة دقيقة وموثوقة مثل الموسوعات الأخرى. وبطريقة مماثلة يتعين أن تكون البيانات المالية والبيانات الطبية دقيقة وصحيحة من أجل أن تغرس الثقة في نفوس التجار والأطباء والمرضى... والقائمة تطول. فمثل وجهة النظر هذه تفترض أيضاً أن سلاّة أصل البيانات وقضايا صحة وموثوقية البيانات ليست مهمة بالنسبة لمستخدمي البيانات.

وبغض النظر عن تسامح بعض مجموعات بيانات البيانات الكبيرة تجاه عدم الدقة، إلا أن تحذير المعطيات الرديئة تؤدي إلى مخرجات رديئة لا يزال سارياً. فمجموعات البيانات الكبيرة التي تنشئ بيانات معيبة ورديئة، وتم التلاعب بها وتمحيضة، أو تنشئ بيانات ذات مصداقية ضعيفة، سوف تؤدي إلى تحليلات ونتائج ذات مصداقية ضعيفة وسوف تكون الفوائد المرجوة منها قليلة للغاية بالنسبة لهؤلاء الذين يحللون ويستفيدون منها. ويمكن للبيانات الكبيرة، عن طريق طرق إنتاجها، أن تعاني من جميع العلل، فالبيانات يمكن أن تكون معيبة ورديئة من خلال أخطاء الأداة، حتى عبر التقنيات التي يُفترض أن تكون ذات دقة عالية للغاية مثل نظام التموضع العالمي (GPS) والذي تختلف دقته من جهاز لآخر، وعدد الأقمار الصناعية المتوفرة في وقت معين من الأوقات (Dodge and Kitchin 2013). أو من الممكن أن يكون قد تم التلاعب أو التزوير في البيانات من خلال حسابات مزورة أو القرصنة من أجل التأثير على الاتجاه العام وتوجيهه العامة نحو مسارات محددة (Vis 2013). فعلى سبيل المثال، هناك مئات الآلاف من الحسابات المزيفة على تويتر، وهي تشكل 4.0%.

من إجمالي حسابات المستخدمين على تويتر (De Micheli and Stroppa 2013)، وأيضاً فإن حتى ٤٥٪ من حسابات بعض العلامات التجارية لمتابعي تويتر هي حسابات آلية (Calzolari 2012)، وأكثر من درزنين من الخدمات تباع حسابات تويتر مزيفة (Baym 2013).

فالبيانات يُمكن أن تتسم بالتحيز جراء بعض الأسباب. إن الخصائص السكانية التي يتم اختيارها في العينات نادراً ما تكون شاملة وتتشوه إحصائياً من قبل الجنس، والنوع، والدخل، والموقع، والعديد من العوامل الاجتماعية والاقتصادية الأخرى. فليس جميع الأشخاص يستخدمون تويتر أو الفيسبوك أو يتسوقون من محلات معينة، أو يكونوا مشتركين في شبكة الاتصالات ذاتها... إلخ (Bollier 2010; Graham 2012; Crampton et al 2012). فالتقنية المستخدمة والمعايير المطبقة تؤثر على طبيعة البيانات. فعلى سبيل المثال، يمكن أن يتأثر تحديد المشاركات الأكثر قراءة أو مشاركة على وسائل التواصل الاجتماعي بشكل كبير من خلال خوارزميات الترتيب وليس الاهتمامات البسيطة (Baym 2013). وعلى النحو ذاته، فإن واجهات برمجة التطبيقات تنظم ماهية البيانات التي يتم استخلاصها، ففي تويتر، على سبيل المثال، يتم التقاط الوسم المرتبط بالحدث وليس جميع التغريدات ذات الصلة (Burns 2013)، وأيضاً فقد بين (Gonzales – Bailon et al 2012) أن اختلاف طرق الوصول إلى بيانات تويتر – البحث في واجهات برمجة التطبيقات مقابل تدفق واجهات برمجة التطبيقات – يؤدي إلى الحصول على مجموعات مختلفة من النتائج. ونتيجة لذلك، لا توجد أي ضمانات بأن فريقين من الباحثين، على سبيل المثال، الذين يحاولون جمع البيانات ذاتها في الوقت ذاته سوف يكونون قادرين في نهاية المطاف على الحصول على مجموعات البيانات ذاتها (Burns 2013). وعلاوة على ذلك، فإن اختيار البيانات الوصفية والمتغيرات التي يتم توليدها والبيانات الوصفية والمتغيرات التي يتم تجاهلها يؤدي إلى رسم صورة معينة. ومثل هذه التحيزات تؤدي إلى انحراف أو حتى إلى نشوء أفكار واتجاهات وأشخاص ونماذج مستترة لم يتم تصويرها أو تمثيلها في مجموعة البيانات (Graham 2012).

وفيما يتعلق بمصداقية البيانات، فإن الحال لا يتمثل دوماً في أن البيانات تعبر بصدق عما تم استخدام البيانات من أجل قياسه. فعلى سبيل المثال، هناك علامات استفهام

حول مدى حقيقة ودقة تعبير المشاركات في وسائل التواصل الاجتماعي عن وجهات نظر الأشخاص، وعلى مدى الوثوق في مثل هذه المشاركات. وقد حذر (Manovich 2011:6) قائلاً إن "المشاركات، والتغريدات، والصور التي تم تحميلها، والتعليقات، وجميع الأشكال المختلفة من المشاركات على الإنترنت لا تعتبر نوافذ شفافة بذاتها، وبدلاً من ذلك، فهي غالباً مرتبة بعناية وتتم إدارتها بشكل منتظم وذلك من خلال الانتقادات نفسها التي تنطبق على بيانات المقابلات ومجموعات التركيز" (Focus Group). وأبدى آخرون شكوكهم حيال صحة بيانات علم المواطن (Citizen Science) والتعهد الجماعي، وحذروا من جهل وعدم مقدرة الحشود على التنبؤ ومخاطر التفاوت (فيما يتعلق بالتغطية)، والمتغيرات (فيما يتعلق بالجودة والتجانس)، وتوليد البيانات المتحيزة المتأصلة في مشروعات المساواة الكبرى، التي تنفجر إلى القادة وتتشكل من مجموعة ضيقة من السكان ذوي التباين في المهارات والدوافع (Carr 2007). وفي الواقع، فإن وسائل التواصل الاجتماعي يهيمن عليها نخب رقمية غير ممثلة (Crutcher and Zook 2009) في حين أن هناك مجموعات كبيرة من المجتمع مثل الفقراء وكبار السن لا ينخرطون في وسائل التواصل الاجتماعي أو يفتقرون إلى الهواتف الذكية، ومن ثم فإن أي بيانات يتم جمعها لا تعكس بالضرورة وجهات نظر هؤلاء الأشخاص أو أحيائهم (Crawford 2013). ومن ثم فمن الضروري للغاية أن يتم تحديد ووضع مثل هذه التحيزات في الاعتبار وذلك عند تحليل البيانات من أجل كسب فهم أكثر شمولية للظاهرة، وخصوصاً إذا ما كان سيتم استخدام مثل هذه البيانات في أعمال السياسة العامة والقرارات التي تؤثر على حياة الناس كالنتائج الصحية.

وعلاوة على ذلك، يوجد بعض المخاوف بشأن جودة وتجانس المحتوى والبيانات الوصفية التي تم تكوينها عبر أفراد متنوعي المهارات والدوافع، ومخاوف بشأن كيفية توفير درجات موثقة من الاعتمادية وتكوين إحساس بالمصداقية (Dodge and Kitchin 2013). وقد أدى هذا بالبعض إلى افتراض أن عمل التمهيد الجماعي "الهواة" سيكون إجراؤه بشكل أفضل إذا بُذل على تصحيح وتنويع البيانات وليس على تكوين البيانات (Carr 2007). فالمثال الذي ألقى Carr الضوء عليه هو موسوعة ويكيبيديا، إذ على الرغم من كونها شهيرة وشاملة، إلا أنها تتوسع بشكل غير مدروس يماثل الاهتمامات المختارة للمشاركين، ويوجد بها مقالات غير كاملة وأحياناً مكتوبة بلغة ركيكة وتافهة ومطعون في صحتها بشكل

كبير، مما يقوض حجيتها وإمكانية استخدامها أو نفعها. وادعى Carr أنه "إذا كانت موسوعة ويكيبيديا غير مجانية، فمن غير المحتمل أن قراءها سيكونون أكثر تسامحاً تجاه تلك العيوب." (2007:4). ويمكن أن تُعاني خريطة الطريق المفتوح (OpenStreetMap) سوء التغطية في بعض الأماكن حيث يوجد عدد قليل من المتطوعين. ويوجد أيضاً بعض المخاوف بشأن استدامة أعمال التمهيد الجماعي التطوعية، حيث ادعى Carr 2007 أن الارتباطات التي تربط مجموعة افتراضية من الناس سويماً عادةً ما تكون ارتباطات سطحية، وهي تفتقر إلى العمق والالتزام الإيجابي، وعادةً ما تكون عرضةً للتشتت، وهي تعتمد بشكل كبير على مجموعة أساسية صغيرة لكي تحافظ على سير المشروع وتوفر الجزء الأكبر من العمل. وعلى النقيض من ذلك، لاحظ البعض أنه فيما يتعلق بـ OpenStreetMap، فإن جودة البيانات المقدمة تطابق جودة البيانات التي تقدمها الشركات الاحترافية وأن التغطية متنوعة (Haklay 2010, Mooney et al 2011).

فهذه المناقشة تلقي الضوء على أن كون مجموعة البيانات كبيرة للغاية من ناحية الحجم، فإن هذا لا يعني بالضرورة أنها عشوائية أو ممثلة للظاهرة أو صحيحة أو ذات مصداقية أو موثوقة. وكما لاحظ (boyd and Crawford (2012): "كون البيانات الكبيرة تزودنا بكميات هائلة من البيانات، فإن هذا لا يعني أن القضايا المنهجية لم تعد ذات صلة. ففهم العينة على سبيل المثال، يُعد أكثر أهمية الآن عما كان عليه الأمر في الماضي" (ص 668). وبالنسبة لـ (Kelling et al. (2009: 615 فإن هذه القضية تعتبر قضية أصل ومَنْشأ أي "المعلومات المتعلقة بأصل وتعريف وملكية وبنية" مجموعة البيانات التي يمكن محلي هذه البيانات من معرفة كيفية تكوين هذه البيانات ومعالجتها، وتمكنهم كذلك من تقييم جودة وصحة البيانات. وعلى الرغم من ذلك، فإن مصدر معظم مجموعات بيانات البيانات الكبيرة مقيد بهؤلاء الذين قاموا بإنشاء البيانات، مما يجعل من الصعوبة بمكان تقييم صحة الدراسات المنشورة التي اعتمدت على هذه البيانات. فأحد التحديات الرئيسية التي تواجه البيانات الكبيرة يتمثل في المقدرة على معرفة تفاصيل سُلالة الأصل والذي يشمل تقييمات مستويات الأخطاء والشكوك في مجموعة البيانات، وإنشاء وسائل فعالة للتحقق من وتنظيف ومعايرة البيانات عالية السرعة وشديدة التنوع.

دمج وتوافقية البيانات (Data Integration and Interoperability):

إن أحد الوعود التي انبثقت عن ثورة البيانات هو إنشاء مجموعات بيانات ذات ترابط قوي مما يمكن هذه المجموعات من أن تتحد سوياً من أجل إنشاء قيم ومعارف إضافية.

وكما اتضح من خلال تطور البنى التحتية للبيانات ومدمجي البيانات، فإن العديد من البيانات عرضةً لكي تندمج مع بعضها، واتضح أيضاً أن الكثير من البيانات تظل منعزلة في مستودعات أو يمكن دمجها فقط من خلال معالجة إضافية كبيرة. ومن أجل أن يتم دمج البيانات في مجموعات بيانات جديدة، أو استخدام هذه البيانات بالاشتراك مع أي بيانات أخرى، فإن هذه البيانات تتطلب حقولاً تأشيريه مشتركة وتوفر معايير البيانات، وبيانات وصفية متسقة، وتوافق على مستوى التنظيم والصيغة. وفي حين أن تقدماً كبيراً قد حدث في إدارة البيانات وتطبيق معايير البيانات التي قادتها منظمات مثل المنظمة الدولية للمعايير (ISO) ورابطة الشبكة العالمية (W3C) واتحاد المعلومات المكانية المفتوحة (OGC) في سبيل المحاذاة بين البيانات والبيانات الوصفية وتحسين التوافق بين المنصات التقنية المختلفة التي تعالج البيانات وتحفظ بها، وتطوير البيانات المترابطة، إلا أن الكثير من مجموعات البيانات لا يمكن ربط بعضها ببعض. فهي تشكل ما أطلق عليه Singh (2012) مصطلح "البيانات الجانحة" (Stranded Data).

إن معظم دراسات البيانات الصغيرة لا تزال تُنشئ بيانات جانحة كونها تستخدم لغات توصيف خاصة بها وأنظمة تنظيم البيانات التي تناسب الأغراض التي سيستخدمونها فيها، ولكن القليل من هذه الدراسات تضع في اعتبارها قضايا التوسع والربط. وهذه هي أيضاً الحال غالباً داخل المنظمات التي يوجد بها إدارات منفصلة تستخدم أنظمتها الإدارية والتصنيفية الخاصة. فعلى سبيل المثال، فإن كل واحدة من هيئات التخطيط التي يبلغ عددها ثمان وثمانين هيئة في إيرلندا تستخدم نظامها الخاص لاستخدام الأراضي وتصنيف المناطق، وعلاوة على ذلك، فهذه الهيئات توظف العديد من المنصات التقنية المختلفة لغرض إدارة وتخزين وتخطيط البيانات، فربط جميع هذه البيانات وتوحيدها في نظام وطني واحد ليس مهمة سهلة على الإطلاق. وبالمثل، فإن توسيع مجموعات البيانات الوطنية سوف يواجه مشاكل مماثلة. فعلى سبيل المثال، فإن ربط الإحصائيات السكانية

في إيرلندا وإيرلندا الشمالية سويًا من أجل إنشاء مجموعة بيانات واحدة للجزيرة بالكامل يفرض عددًا من التحديات الفنية وذلك يرجع إلى أن ٣٠٪ من الأسئلة متطابقة، و٣٥٪ أخرى من الأسئلة يمكن أن يتم مضاهاتها وذلك من خلال تقسيم خطط التصنيف وإعادة تجميعها في الخطط المشتركة، وحتى إذا تم هذا، فإن البيانات ستكون ذات جغرافيات إحصائية ضمنية مختلفة يتم تجميعها في وحدات مختلفة الأحجام (Kitchin et al 2007).

وتتضاعف هذه القضايا عبر قارة أوروبا، حيث تتبع كل دولة وكل مؤسسة إجراءات مختلفة (وحدات وتصنيفات)، وتعتمد فترات زمنية متفاوتة (عدد المرات التي يتم فيها إنشاء البيانات)، وتستند إلى أماكن متباعدة (الجغرافيا الإحصائية)، وتستعمل تقنيات مختلفة (الصيغ). ومن ثم فلقد استنتج (Rubert 2012: 118) أنه حتى مع وجود بيانات حكومية شاملة، إلا أنها في غالب الأحيان تتصارع تحت وطأة الأعداد التي لا تحصى من مجموعات البيانات المتضاربة وغير المتوافقة وغير الصالحة للمقارنة. وتسعى بعض المبادرات مثل توجيهاات إنسبير (INSPIRE) (البنية التحتية للمعلومات المكانية في المجتمع الأوروبي) إلى تكوين بيانات مشتركة عبر الجغرافيا الأوروبية الإحصائية المشتركة، ولكن الوصول إلى توافق في هذا الشأن يُعتبر عملية بطيئة وذلك يرجع إلى أن جميع الولايات لا ترغب في الانتقال بشكل أحادي (أي من طرف واحد) إلى نظام جديد للبيانات وخلال العملية يوقفون سجلاتهم الزمنية ويفقدون مقدرتهم على إجراء تحليلات زمنية متتالية، إضافة إلى أن كل ولاية لها أولويات مختلفة فيما يتعلق بالبيانات التي ترغب في إنشائها وتتبعها.

وتوجد هذه التحديات نفسها فيما يتعلق بالبيانات الكبيرة، والتي ربما تكون غير متجانسة وغير منظمة بشكل كبير للغاية، ومتفاوتة فيما يتعلق بالجودة، وهي أيضًا مرنة وقابلة للتوسع. فربط وتجميع مثل هذه البيانات ليس مهمة سهلة على الإطلاق وهي تتطلب تطوير مجموعة جديدة من الأدوات والأساليب، وكذلك إنشاء وتبني معايير جديدة للبيانات. فأحد الحلول المطروحة، والذي أشار إليه (Short et al 2011) هو دمج البيانات المخصص والذي يمكن المستخدمين من التحكم بسرعة وبشكل مباشر في أي من البيانات التي سيتم دمجها، وذلك على الرغم من أن هذا يتطلب أن تكون مجموعة البيانات تتمتع بسمات معينة مثل البنية التحتية الدلالية المؤثرة. إن ربط وتجميع مثل هذه البيانات من الممكن أن يؤدي إلى التلوث المتقاطع (Cross-Contamination) بسبب البيانات

ذات الجودة الضعيفة والحالة المتهالكة مما يؤدي إلى فساد وتلويث مجموعات البيانات الفعالة ويحولها إلى مجموعات بيانات ضعيفة. ومع ازدياد طوفان البيانات، فإن إيجاد الحلول التي تساعد على دمج البيانات وتوحيدها وتحافظ على سلامتها وفعاليتها سيمثل قضية ملحة، وذلك على الرغم من أن عدم المقدرة على ربط البيانات له جوانب إيجابية أيضاً وذلك فيما يتعلق بتقييد مراقبة البيانات وتحسين أمن البيانات (انظر الفصل العاشر).

التحليل الركيك والمغالطات البيئية (Poor Analysis and Ecological Fallacies):

جميع النتائج التي يتم التوصل إليها من الدراسات تكون دقيقة وسليمة وقوية بحسب البيانات التي تم استخدامها والتحليلات التي أُجريت، وإذا ما وضعنا جودة وصحة البيانات جانباً، فخلال تصميم أي بحث يكون من المهم للغاية استخدام منهجية ملائمة وآليات مناسبة. فالإخفاق في فعل ذلك سيؤدي إلى مغالطات بيئية خطيرة عند تفسير النتائج، بمعنى استخلاص النتائج المضللة والتي في الواقع لا تدعمها البيانات المستخدمة. وهناك تاريخ طويل من المساجلات التي تتعلق بمكامن القوة والضعف النسبية في العديد من الأنماط والأساليب المنهجية المختلفة، والظروف التي يمكن فيها تطبيق مثل هذه الأنماط والأساليب بشكل صحيح. وعلى وجه العموم، هناك الآن إرشادات وقواعد محددة بشكل واضح تتعلق بكيفية تكوين وتحليل أنواع مختلفة من البيانات عبر أنظمة متعددة. وعلى الرغم من ذلك، وفيما يتعلق بالبيانات الكبيرة، فإن علوم البيانات لا تزال في أطوارها التمهيدية، ومثل هذه المعايير لا تزال في طور التشكيل والصيغة، فآليات جديدة يتم الآن اختراعها، وتوسيعها، وتقييمها.

وهناك أربع قضايا منهجية تتعلق بالبيانات الكبيرة جذبت الانتباه، فأولى هذه القضايا تتمثل في الإجابات المتفاوتة بشكل عميق والتي تؤدي إليها الافتراضات والمقاربات داخل النماذج (Silver 2012). وثانياً، مدى قدرة الآليات والأساليب المستخدمة على إحداث الاستسقاط، أي رؤية واستخلاص أنماط من البيانات العشوائية التي لا معنى لها، أو إحداث فرط التعلم أو الملاءمة (Overfitting) والتي تحدث عندما يصف نموذج إحصائي خطأ عشوائياً أو ضجيجاً بدلاً من العلاقة الكامنة (boyd and Crawford 2012)، أو الخطأ في التعامل مع التشويش على أنه معلومات (Silver 2012; Siegel 2013). وثالث هذه

القضايا يتمثل في الاتهامات بأن بعض الآليات المعينة تؤدي إلى ما هو أكثر بقليل من تجريف البيانات (Webster 2011) (Data Dredging). وأخيراً، إذا كان بإمكان الآليات أن تتوسع خارج نطاق بيانات ضبط الاختبار إلى مليارات السجلات المتدفقة في العالم الواقعي (Bryant et al, 2008).

إن النماذج تهدف إلى توضيح والتعبير عن العلاقة بين المتغيرات وكيفية عمل النظام. فكل نموذج يتم تأسيسه على نظرية معينة، وبنائه باستخدام مجموعة من الافتراضات والمقاربات التي تبين كيفية تفاعل المتغيرات، في حين أنه يتم اختبار صحة النموذج من خلال البحوث التجريبية، فإذا كانت النظرية والافتراضات غير صحيحة، فإن جميع التنبؤات ستكون غير صحيحة في هذه الحالة كذلك، وحتى التغييرات الصغيرة للغاية على الافتراضات في نموذج يمكن ما أن تؤدي إلى نتائج مختلفة تماماً. ففي حالة الأزمة المالية العالمية التي حدثت في العام 2008، ادعى (Silver 2012) أن أحد العوامل المهمة للغاية في إحداث مثل هذه الأزمة كان استخدام نماذج تمويل دمجت المخاطر بحيث يمكن حساب احتمالية الدخل المحتمل مع الارتياح لأن المخاطر المحتملة من الصعب تقييمها. ونتيجة لذلك، اعتمدت النماذج على افتراضات غير يقينية ولكن تم معاملتها كما لو كانت مخاطر معروفة على وجه اليقين. والنتيجة كانت افتراض أن السندات المالية الجديدة التي يحيط بها ارتياح نظامي عالٍ للغاية هي أصول ذات مخاطر منخفضة، وهذا افتراض ثبت أنه خاطئ. وفي الأنظمة المعقدة، فإن مثل هذه الأخطاء لا يتم قياسها على أساس درجات ولكن يتم قياسها على أساس ترتيبها حسب الأهمية. وقد قدرت مؤسسات ستاندرد أند بورز وموديز (S&P and Moody's) مخاطر التخلف عن الدفع والمرتبطة بالتزامات الديون المضمونة (CDOs) بأقل من قيمتها بعامل من مائتين (Silver 2012: 45). وبغض النظر عن جودة البيانات الضمنية، فإذا كان النموذج خاطئاً، فأى نتائج يتم استخلاصها من هذا النموذج ستكون خاطئة كذلك.

لقد افترض (Taleb (2013) and Granville (2013) أن لعنة أو مأساة البيانات الكبيرة مُضاعفة، نتيجة العلاقات الإيجابية الخاطئة بين البيانات؛ وأنماط مجموعات البيانات التي تكون عرضية بشكل تام، وليس لها أي قوى تنبؤية، ولا يمكن تكرارها، وربما يمكن أن

تكون قناعاً لنماذج أضعف ذات أهمية، فمجموعات البيانات تشتمل غالباً على البيانات التي يبدو أنها مرتبطة بعضها البعض، حتى ولو كانت العلاقة بينها عشوائية في الواقع. ومع زيادة حجم مجموعة البيانات، تتضاعف الأخطاء الإيجابية، إلى الحد الذي تكون فيه أي علاقة تقريباً بين البيانات مهمة من الناحية الإحصائية وفق معايير اختبارات الارتباط التقليدية (Pentland 2012; Rajaraman et al. 2012; Taleb 2013). وعلى النحو ذاته، ففي أساليب التنقيب عن البيانات، مثل شجرة القرارات (تسلسل اتخاذ القرار)، يمكن أن يتسم النموذج بفرط الملاءمة كي يتم استنباط نتائج استثنائية تحول إلى نتائج قابلة للتعميم مما يؤدي هذا إلى حدوث مغالطات بيئية. فالبيانات الكبيرة في هذه الحالة لا تعني معرفة أفضل، ولكنها تعني بالأحرى ارتباكاً أكثر وذلك من جراء صعوبة تحديد دالة حقيقية وسط ركام من الدلالات. ولذا فإنه: "أحياناً يكون الأقل هو الأقوى تأثيراً" (Verhulst) تم الاقتباس من (Bollier 2012:14).

إن أحد الجوانب المهمة في مشكلة الأخطاء الإيجابية ليس حجم مجموعة البيانات فحسب، بل كيفية معالجة وتحليل مجموعة البيانات هذه. فأحد أجزاء إستراتيجية الكثير من تحليلات البيانات الكبيرة يتمثل في فحص مجموعة البيانات قبل صياغة الافتراضات (انظر الفصل السادس)، ومن الطرق المتبعة للقيام بهذا الأمر حساب الارتباطات بين جميع المتغيرات وذلك من أجل تحديد العلاقات المحتملة بين البيانات الجديرة بالدراسة الإضافية. وتشكل هذه العملية ما يُعرف بتجريف البيانات (Data Dredging) والتي تسمى أحياناً تصيد البيانات، وهي ممارسة لا يحبذها بشدة الكثير من الخبراء الإحصائيين (Webster 2011; Piatetsky-Shapiro 2012). فالمشكلة التي ترتبط بهذا المنحى مشكلة مزدوجة. فمن جهة، هي تحدد آلاف الأخطاء الإيجابية المحتملة وذلك دون تحديد أي من هذه الأخطاء عشوائي وأي منها ذو مغزى. ومن جهة أخرى، يمكن أن تؤدي إلى التنقل من متجر لآخر طلباً للأرخص وعرض النتائج التي تؤكد اعتقاد المحلل، مع وجود دليل من التحليل المستخدم من أجل أن يجعل الادعاء يبدو معقولاً ومدروساً (Webster 2013; Taleb 2013). فالسؤال الشائك الذي يطرح نفسه هنا، كما أشار (Granville 2013)، هو كيف يمكن التمييز بين دالة حقيقية ودالة عرضية وسط الكميات الهائلة من البيانات؟ وكيف يمكن ألا ينقاد المرء للاستسقاط وفرط التعلم؟

إن هناك بعض الحلول الممكنة. حيث اقترح (Rajaraman et al. 2015) استخدام مبدأ بونفيروني (Bonferroni)، والذي يسعى إلى تحديد الأحداث النادرة جداً لدرجة أنها لا يمكن أن تحدث في البيانات العشوائية، وذلك من أجل فرز الارتباطات الفعلية عن الارتباطات العشوائية. كما كان (Lusk 2013) من أنصار استخدام مستويات معيارية معززة، وذكر نقطة انطلاق معيار خمسة سيجما (Five Sigma) المستخدم حالياً في الفيزياء من أجل مواجهة الارتباطات الوهمية وإعلان الاكتشافات (احتمال كون النتائج خاطئة وناتجة عن التغيرات الإحصائية فقط هو واحد في كل ثلاثة ملايين وخمسمئة ألف أو 1:3,000,000). ومن الحلول المطروحة الأخرى هو التحقق الإضافي من خلال مجموعات بيانات أخرى، أو التقسيم العشوائي لمجموعة البيانات الحالية إلى عينات ومعرفة إذا ما كانت العلاقة أو الارتباط سيظل موجوداً عبر العينات (Jensen 2000). وفيما يتعلق بفرط التعلم، ناقش (Miller 2010) إستراتيجية استخدام الخلفية المعرفية، والتي تتكون من استخدام الحقائق المعروفة عن مجموعة بيانات مُستقاة من الافتراضات عن النظام، والحقائق الملاحظة، ومعرفة الخبر، أو النظرية، أو المقاييس الممتعة والمرغوب فيها التي تقيم البساطة واليقن والفائدة والحدثة للأنماط المستحدثة، وذلك من أجل فصل الأنماط الهادفة عن التشويش. وعلى الرغم من ذلك، فمن الواضح أن هناك حاجة لبذل أعمال إضافية من أجل تطوير أساليب لتقييم المغزى والدلالة في عصر البيانات الكبيرة.

وإذا وضعنا في الاعتبار أن تحليلات البيانات لا تزال في مراحلها الأولى من التطور، وتم تطبيقها بالدرجة الأولى على بيانات الضبط والاختبار، يظل أيضاً هناك بعض المخاوف بشأن فعالية مثل هذه الأساليب في مجارة البيانات الفوضوية وغير المنضبطة. وكما أشار (Byrant et al. 2008: 4)، فإن "الكثير من الخوارزميات لا تتوسع خارج نطاق مجموعات البيانات التي تتكون من بضعة ملايين من العناصر، أو لا يمكنها أن تتحمل التشويش والثغرات الإحصائية الموجودة في بيانات العالم الواقعي". فإذا كان هناك أي اختلالات أو تحيزات موجودة في الخوارزميات، فمن شأنها أن تؤدي إلى انحراف النتائج وتفسيرات هذه النتائج (byoed and Crawford 2012). ويرتبط بهذا التخوف أيضاً القلق من أن هناك اعتماداً مفرطاً والكثير جداً من الموثوقية الموضوعية في الخوارزميات وذلك من أجل إجراء التحليلات، واستخراج المنافع والمزايا التي تجلبها الخبرات والمعارف والفترة البشرية من

أجل فهم ظاهرة ما (Lohr 2012). ومن ثم، هناك حاجة لإجراء المزيد من العمل من أجل تأسيس وتقييم صحة التحليلات الجديدة والثقة في النتائج التي يمكن استخلاصها من هذه التحليلات.

ومن الانتقادات الأخرى الموجهة إلى تحليلات البيانات الكبيرة هو تركيز هذه البيانات على الآنية (هنا والآن)، وذلك بالنظر إلى حداثةها، وعلى وجه العموم، فإن البيانات الكبيرة تفتقر إلى العمق التاريخي. فشبكات الاستشعار الجديدة أو بيانات التمهيد الجماعي يمكن أن توفر بيانات مستمرة وثرية، ولكنها تتسم بقصر الفترات الزمنية التي تغطيها. ونتيجة لذلك، فإن الكثير من التحليلات تكون مناسبة في وقتها ولكنها تفتقر إلى العمق الزمني. وهذا أمر لا غبار عليه ومستحسن إذا كان الشخص مهتماً بالوقت الحاضر بتتبع حدث ممتد باستخدام وسائل التواصل الاجتماعي (على سبيل المثال حدث سياسي كبير مثل الربيع العربي)، أو أن يكون الشخص قائماً بتحليل عملية مادية لا تتغير بمرور الوقت (مثل تجربة فيزيائية)، ولكن سيكون الأمر مسبباً للمشاكل إذا تم استخدامه من أجل فهم العمليات التي تمتد لآلاف السنين أو أكثر (على سبيل المثال الأنظمة البيئية والجيولوجية). وينطبق هذا الانتقاد أيضاً، بالطبع، على دراسات البيانات الصغيرة، والتي تعتبر أيضاً لقطات تم اتخاذها في أماكن معينة وأوقات محددة.

المهارات والموارد البشرية (Skills and Human Resourcing):

في حين أن بعض تحليلات البيانات يتم تركها للخوارزميات، وخاصةً العمل المتعمق للمعالجة والحسابات، إلا أن أعمال التوجيه والتفسير لا تزال إلى حد كبير حكراً على المحللين البشر. فبالاعتماد على مهاراتهم وخبراتهم ومعارفهم، يتخذ الباحثون والمحللون القرارات المتعلقة بالأماكن المثلى التي ينبغي التركيز عليها، وكيفية صياغة وإجراء التحاليل، وإدراك المراد من النتائج والعمل على هذه النتائج. فالبشر يظلون هم الفاعلين الأساسيين في تأسيس وتشغيل والمحافظة على المشروعات التي تقودها البيانات، فعلى سبيل المثال، يتطلب تأسيس بنية تحتية للبيانات مهندسي برامج وأجهزة ومسؤولين رقميين، وأخصائيين في السياسة وخبراء مجالات علمية ومحللين. فهؤلاء العاملون يضمنون تصميماً جيداً للنظام كي ينتج بيانات مفيدة وصحيحة وموثوقة، ويوفرون الوسائل الضرورية لتخزين ومشاركة البيانات، ويؤمنون

مجموعة من الأدوات التحليلية والإدارية المناسبة، التي يمكنها استخلاص القيمة والمعرفة. وهم يعرفون أيضاً ماهية الأسئلة التي ينبغي توجيهها وما الذي يتعين فعله بالإجابات على تلك الأسئلة. ولذا، فإن المواهب البشرية بضاعة رائجة. ولكن وكما أشارت بعض التقارير، هناك نقص متزايد في الموظفين الذين يتمتعون بمهارات في البيانات وخاصةً محلي ومديري البيانات الذين بإمكانهم تحويل البيانات والمعلومات إلى معارف وعلوم (Gantz and Reinsel 2011; Manyika et al. 2011; Shah et al. 2012; Rooney 2012).

وقد أكد (Manyika et al. 2012:3) " أن الولايات المتحدة وحدها تعاني نقصاً يبلغ ١٤٠,٠٠٠ إلى ١٩٠,٠٠٠ شخص يتمتعون بمهارات تحليلية متعمقة وكذلك ١,٥ مليون مدير ومحلل من أجل تحليل البيانات الكبيرة واتخاذ القرارات بناءً على اكتشافاتهم". واستطردوا بالإشارة إلى أن مثل هذه المواهب "من الصعب إيجادها إذ تستغرق سنوات عديدة في التدريب أو إعادة التدريب" (ص ١٠). وقيم (Shah et al. 2012) عدد ٥٠٠٠ موظف في اثنتي عشرة شركة عالمية فيما يتعلق بمقدرتهم على معالجة تحليلات البيانات بشكل مثمر حيث قاموا بوضع الموظفين في ثلاث مجموعات. الأولى، التجريبيون غير المعترضين الذين يثقون في التحليلات أكثر من الأحكام الشخصية (٤٣٪). والثانية، متخذو القرار المتعمقون الذين يتجاهلون التحليلات ويتصرفون بما يمليه عليه حدسهم الفطري (١٩٪). والثالثة، المتشككون المطلعون (٣٨٪) الذين يوازنون بين الأحكام الشخصية والتحليلات، ويسعون إلى تفسير البيانات ووضعها في سياقها المناسب عند اتخاذ القرارات. وبشكل مقلق، فإن أولئك الذين يدعون البيانات تعبر عن نفسها وأولئك الذين يتجاهلون البيانات والتحليلات بشكل روتيني يفوقون في عددهم أولئك القادرين بشكل أفضل على فهم البيانات الكبيرة. وعلاوةً على ذلك، لاحظوا أن الشركات تعاني من أربعة مشاكل تحول دون الاستفادة من إمكانيات البيانات الكبيرة وهي:

- أولاً: أن المهارات التحليلية تتركز في عدد محدود للغاية من الموظفين ولا تنتشر في باقي الموظفين.
- ثانياً: أن حلول تقنية المعلومات تتركز بشكل تقليدي على التقنية وبشكل أقل بكثير على المعلومات في حين أن المعلومات هي المجال الذي يتم من خلاله جني القيمة.

- ثالثاً: أن تخزين البيانات يحدث كيفما اتفق ومن الصعوبة بمكان على أي موظف أن يحدد مكان المادة إذا لم يكن هذا الموظف هو المسئول عن جمعها.
- رابعاً: أن المديرين يفهمون قيمة الموهبة ورأس المال والعلامة التجارية، ولكن لا يفهمون قيمة المعلومات، ولا يبذلون جهوداً كافية من أجل فهم البيانات أو ترتيب هذه البيانات بشكل فعال وكفء.

ويوجد مثل هذا النقص في المهارات أيضاً في المؤسسات الأكاديمية، حيث إن خبراء المجالات المختلفة تدربوا في وقت ندرة البيانات وهم غير مستعدين بشكل كبير لعصر البنى التحتية للبيانات والبيانات الكبيرة وذلك باستثناء عدد لا يتجاوز أصابع اليد الواحدة من العلماء والمراكز.

إن القلق الذي ينتاب الكثير من المعلقين هو أن المنافع المحتملة من الشركات والعلوم التي تقودها البيانات لن يتم تحقيقها بشكل كامل بسبب نقص المواهب البشرية وخاصة علماء البيانات الذين تجتمع لديهم مهارات مبرمجي البرامج والإحصائيين والرواة المحترفين وذلك من أجل استخلاص شذرات الذهب المختبئة تحت جبال من البيانات (Cukier 2010)، والمديرين الذين يفهمون كيفية تحويل هذه الشذرات إلى قرارات حكيمة. وفيما يتعلق بالمديرين، وكما أشار (Shah et al. 2012: 23)، فإن "الاستثمار في التحليلات يمكن أن يكون بلا جدوى، ويمكن حتى أن يكون ضاراً، إلا إذا كان بإمكان الموظفين دمج هذه البيانات في العملية المعقدة لاتخاذ القرارات". فالجامعات بدأت الآن في إنشاء برامج جديدة ومراكز بحثية لعلوم البيانات، ويسعون إلى تعديل الدورات التدريبية الحالية لكي تشتمل على التدريب على مجموعة المهارات الجديدة هذه، وذلك في مسعى إلى تنقيح بعض الثغرات المتعلقة بالمهارات. ويتم تنظيم بعض هذه البرامج بالمشاركة مع بعض الشركات. فعلى سبيل المثال، بدأت شركة آي بي إم (IBM) بالفعل في العمل مع العديد من الجامعات من أجل تطوير دورات ومقررات تعليمية، وهناك بعض الشركات الأخرى مثل شركة تيراداتا (Teradata) التي تقدم برامج تدريبية معتمدة من الجامعات (Bertolucci 2013)، وذلك على الرغم من أن هذه البرامج من الممكن أن تؤدي إلى مخاطرة توجيه الطلاب تجاه رؤى وبرامج خاصة ببعض الشركات.

الجدول رقم (١ - ٩)

الخبرات اللازمة لبناء البنى التحتية للبيانات وإجراء بحوث البيانات الكبيرة

<ul style="list-style-type: none"> • معرفة نظرية وعملية عميقة بالمجال (المجالات) ذات الصلة. • المعرفة التامة بأنواع البيانات التي ستتم دراستها وأصل هذه البيانات وأهميتها للمجال (المجالات) ذات الصلة. • المقدرة على تحديد الفجوات المعرفية. • المعرفة التامة بالأدبيات والاتفاقيات النظامية. • المقدرة على تعليم الآخرين الذين يملكون خلفيات متعددة من أجل استيعاب كل ما ذكر أعلاه. 	<p>خبرات المجال</p> <p>Domain Expertise</p>
<ul style="list-style-type: none"> • فهم كيف يتم تجميع البيانات والمحافظة عليها، وكذلك فهم العلاقات بين الأشياء المادية والتمثيل الرقمي لهذه الأشياء (إذا كان ذلك ممكناً). • المعرفة التامة بنماذج البيانات و/ أو اتفاقيات وصف البيانات. • فهم كيف يتم الوصول إلى البيانات ذات الصلة وتخزين هذه البيانات. • المقدرة على تسهيل مشاركة البيانات وتصحيح الأخطاء بشكل يدوي، سواءً كان هذا أثناء أم بعد المشروع. • المقدرة على التنبؤ بالاستخدامات المستقبلية أو البديلة للبيانات. • فهم الأشكال الجديدة من المطبوعات التي يمكنها أن تحتوي على البيانات. 	<p>خبرات البيانات</p> <p>Data Expertise</p>
<ul style="list-style-type: none"> • فهم نقاط القوة والضعف في أدوات البحث الفردية. • اختيار وتعديل الأدوات المناسبة لدعم الأهداف البحثية. • التنبؤ بالمشاكل التي يمكنها أن تنشأ مع استخدام الأدوات المختارة لتنفيذ مهام المشروع. • توقع واكتشاف معدلات الأخطاء في خوارزميات البيانات وتحليلات البيانات، واختيار الطرق الإحصائية التي تفسر هذه الأخطاء كلما كان ذلك مناسباً. • تعليم الآخرين كيفية تفسير نتائج البحث. 	<p>الخبرات التحليلية</p> <p>Analytical Expertise</p>

<ul style="list-style-type: none"> • المقدررة على صياغة معالم المشروع. • المقدررة على تحديد الأهداف المناسبة والمواعيد النهائية وتنسيق مسارات العمل المتوازية إذا كان ذلك ضرورياً. • المقدررة على اختيار أفضل الإستراتيجيات المناسبة للاتصال والتوثيق للمشروع. • إجادة الأدوات البحثية التعاونية. • الإرادة القوية للعمل من أجل تحقيق النتائج التي تفيد جميع أعضاء الفريق. 	<p>خبرات إدارة المشاريع Project Management expertise</p>
---	--

المصدر: مقتبس من (Williford and Henry (2012: 16 – 20).

إن علم البيانات ليس هو الخبرة الوحيدة التي يتعين تأسيسها. وبحسب ما ادعى (Williford and Henry (2012)، فإن تأسيس بنى تحتية للبيانات وإجراء البحوث على البيانات الكبيرة يتطلبان دمج أربعة أنواع من الخبرات من أجل تحقيق النجاح (انظر الجدول ١-٩). وهذه الخبرات الأربعة كما ادعى (Rubert (2013 يمكن أن يتم تحقيقها فقط من خلال منهج متعدد التخصصات والقطاعات، مما يسمح بتوحيد المهارات والرؤى التي يمكنها أن تتقاطع مع المجالات وتحقق التعاون بين المؤسسات الأكاديمية والصناعية والحكومية. وحتى يأتي مثل هذا الوقت، أي عندما يتطور وينضج جيل جديد من المواهب المناسبة، وتتطور مناهج متعددة المجالات والأنظمة، فإن المشروعات ستظل عرضة للإخفاق أو سوء الأداء.

الخاتمة:

كشف هذا الفصل عن أوجه النقص الفنية في البيانات الصغيرة المتوسعة والبيانات الكبيرة. وناقش أن البيانات الكبيرة تحيط بها العديد من القضايا التي تحول دون الاستفادة منها وتطويرها في الوقت الراهن، ومن هذه القضايا على سبيل الذكر لا الحصر النطاق الضيق إلى حد ما لمجموعات البيانات المتوافرة، والوصول المحدود إلى البيانات، والجودة الضعيفة للبيانات، وصحة وموثوقية البيانات، وسُلالة أصل البيانات المحددة بشكل سيئ،

والتبادلية الضعيفة، والتحليلات المشكوك فيها، والمغالطات البيئية، وقاعدة المهارات المطورة بشكل ضعيف، والاستجابة المؤسسية. وقد أدى ذلك ب (Dembosky et al. (2012) للدعاء بأن "التحركات الأولية ربما تكون واعدة بشكل مفرط ولكن أداءها ربما يكون أقل من المتوقع"، وخصوصاً عندما يصبح هؤلاء الذين يتبنون حلول البيانات الكبيرة "مُتخمين بسبب التقنية التي لا تزال حديثة للغاية وغير مُنقحة، فالتقنية أضحت مرهقة ومثبطة أكثر من كونها مفيدة". وبمرور الوقت، فإن الكثير من هذه القضايا سيتم التعامل معها إلى حد ما وذلك من خلال الحلول الإدارية والفنية، ولكن ستظل هذه القضايا تشكل مخاوف وتتطلب اهتماماً مستمراً وعملاً إصلاحياً متجدداً. وينطبق الأمر ذاته أيضاً على مشروعات البنى التحتية للبيانات ومشروعات البيانات المفتوحة. وفي الواقع، وكما أشرنا إلى ذلك في الفصل الثالث، فإن الكثير من هذه القضايا التي تؤثر على البيانات الكبيرة تؤثر بشدة في مشروعات البيانات المفتوحة. فعلى سبيل المثال، فإن بعض مواقع البيانات المفتوحة لا تعد أكثر من كونها مقال نفايات للبيانات ذات المتغيرات الانتقائية، إذ إن هذه المواقع منظمة بشكل سيئ للغاية وهي ذات بنية ضعيفة تماماً وبياناتها غير نظيفة، وهي تفتقر إلى مجموعات التسجيل الكاملة، والتوثيق، والسياسات المتعلقة بالحفظ، والنسخ الاحتياطي، والتدقيق، وإعادة الاستخدام، والخصوصية والسياسات الأخلاقية.

فنحن يتعين علينا أن نكون حريصين للغاية من أجل أن ندرك أن ثورة البيانات لا تزال في أطوارها التمهيدية وهي تتمتع بخصائص الجنين فيما يتعلق بالنضج والتطور، وفي الوقت الراهن، يبدو أن التسابق يتجه نحو طرح نظم تهدف إلى استعراض الإمكانيات وتأمين حصة سوقية مبكرة دون بذل الجهد الكافي لتفحص المصدقية والفعالية والمنفعة، أو لتفحص القضايا الإدارية والموارد البشرية التي تتعلق بالجاهزية المؤسسية ومجموعات المهارات المناسبة. وفي الواقع، فإن شعار بعض المشروعات هو أن تتجه نحو ما هو مستهجن بشكل مبكر ثم تتعامل مع القضايا التي تنشأ على أساس انتشارها.

ونتيجةً لذلك، فإن البحوث المركزة بشكل كبير والمصممة من أجل تحديد طبيعة ومدى أوجه القصور الفنية والمؤسسية للبيانات الكبيرة والبيانات المفتوحة والبنى التحتية للبيانات لا تزال متأخرة بشكل واضح مقارنة بأعمال التطوير الأخرى. ولكن مثل هذه البحوث

تظل حيوية للغاية لإنشاء أنظمة تقدم اكتشافات ونتائج صحيحة، ومن شأنها أيضاً أن تضيف قيمة كافية للمؤسسات التي تتبنى وتستخدم هذه البحوث. ويتعين أن يتم إجراء هذه البحوث عبر قطاعات متعددة وعبر الأنواع المختلفة من البيانات وذلك من أجل تأسيس القضايا والحلول المتعلقة بالبيانات والأنظمة المختلفة. وعلاوة على ذلك، يتعين أن يتم تخصيص استثمارات من أجل زيادة وصقل مهارات الموظفين الحاليين وتدريب جيل جديد من علماء البيانات، ودون إجراء مثل هذه البحوث، فإن البيانات الكبيرة، والبيانات المفتوحة والبنى التحتية للبيانات ربما ستعاني أضرار السمعة التي ستؤدي في نهاية المطاف إلى بطء في نشرها وتبنيها. وهناك مجموعة أخرى من القضايا التي ستؤدي الدور ذاته وهي القضايا المتعلقة بالآثار الأخلاقية والاجتماعية والسياسية لأنظمة البيانات الجديدة، وهو ما يركز عليه الفصل التالي.

الفصل العاشر

المخاوف الأخلاقية، والسياسية، والاجتماعية، والقانونية

(Ethical, Political, Social, and Legal Concerns)

كما نوقش في الفصل السابع، يتم تكوين وتوظيف البيانات للعديد من الغايات التي تشمل حكم المجتمعات، وإدارة المنظمات، والتوسع في الأرباح، وتنظيم الأماكن. وفي جميع هذه الحالات، تُعدُّ البيانات مدخلات أساسية في الأنظمة التي يتم تنفيذها، للمفارقة، لغرض أن تكون المجتمعات أكثر أماناً، وأمناءً، وتنافسية، وإنتاجية، وكفاءة، وشفافية ومساءلة، على الرغم من أنه يتم القيام بذلك من خلال العمليات التي تراقب وتفرض النظام، وتكبح، وتقنع، وتجبر، وتستغل الناس. فهناك توازن دقيق، في هذه الحالة، بين استخدام البيانات بطرق تحريرية وتمكينية، وبين استخدام البيانات من أجل تحقيق أغراض الفرد الشخصية وإلحاق الضرر بالآخرين، أو استخدام البيانات بطرق تتعارض مع رغبات هؤلاء الذين تمثلهم هذه البيانات. ولا يتعلق الأمر، مع ذلك، بأن تستخدم البيانات ببساطة في أشياء جيدة أو أشياء سيئة، فالمشكلة أعقد من ذلك بكثير. وفي غالب الأحيان فإن النتائج التي تظهر على أنها متعارضة يتم تجميعها سوياً لدرجة أن الأشخاص يمكن تحريرهم أو إجبارهم في الوقت نفسه - وهم يحصلون على منافع شخصية في الوقت ذاته الذي يقعون فيه في مصيدة النظام الذي يسعى إلى تحقيق مكاسب من وراء مشاركتهم. وفي مصطلحات (Althusser (1971، تعمل مثل هذه الترتيبات من خلال الاستجواب، وتوريط الناس عن طريق منطقتها وذلك من خلال الإقناع والحوافز. فعلى سبيل المثال، تمنح بطاقات الولاء في الأسواق للعملاء خصومات في الوقت ذاته الذي تعمل هذه البطاقات على تأسيس الولاء تجاه المتجر وتكوين قدر من البيانات التي يتم استخدامها من أجل محاولة بيع المزيد من البضائع لهؤلاء العملاء، مما يؤدي إلى زيادة الأرباح. وبالمثل، فإن ثمن حصولك على الأمان من الهجمات الإرهابية هو المراقبة التي تنتهك خصوصيات جميع الأفراد في المجتمع، حيث يحصل المواطنون على الأمان مقابل خصوصيتهم.

إن تكوين البيانات والعمل الذي تقوم به مثل هذه البيانات محاط بشكل جوهري بمخاوف أخلاقية، واجتماعية، وسياسية. ومثل هذه المخاوف كانت لوقت طويل مثار إدراك ومجادلات داخل المنتديات العامة والمنتديات العلمية، مما أدى إلى صياغة مجموعة كبيرة من الإرشادات الأخلاقية والمهنية والتشريعات التي تحدد كيف يتم إنتاج، إدارة، مشاركة، واستخدام البيانات. فليس هناك أي شكل من أشكال إنتاج وتحليل البيانات خالٍ من مثل هذه المخاوف، وحتى المشاريع العلمية التي يبدو من الناحية الظاهرية أنها غير خطيرة ومحايدة، يمكن أن تثير أسئلة أخلاقية ويكون لها تداعيات اجتماعية وسياسية. فعلى سبيل المثال، ربما يبدو ظاهرياً أن قياس ومذجة الأنظمة البيئية يكون مقتصرًا على فهم الظواهر الطبيعية، ولكن من الممكن أن يتم استخدام ذلك أيضاً في تشكيل السياسات التي تؤثر في حياة الناس داخل مثل هذه الأنظمة. إن مثل هذه الحوارات الأخلاقية مستمرة وتتطور واحدة تلو الأخرى مع المجال الاستطراذي الأشمل الذي يتعلق بالاتجاهات والآراء الاجتماعية، وتطور التقنيات الجديدة. وفيما يتعلق بالأخيرة، فإن توسيع البيانات الصغيرة داخل البنى التحتية للبيانات، وإنشاء سوق البيانات الهائل، وفتح البيانات المؤسسية، وتدفعات البيانات الكبيرة يطرحان أسئلة أساسية أخلاقية، واجتماعية، وسياسية، إذ إنها تعدل بشكل جذري مشهد البيانات. إن حجم وتنوع البيانات التي يتم تكوينها عن جميع مجالات الحياة اليومية والعالم الذي نعيش فيه اليوم ينمو بشكل مطرد، وهذه البيانات هي أكثر من أي وقت مضى من حيث مناسبة توقيتها، وتصميمها، وشموليتها، وترابطها. فما كان في الماضي خاصاً أو مجهولاً أصبح الآن وبشكل متزايد مكشوفاً أمام مجموعة متنوعة من المصالح، وعمليات اتخاذ القرار داخل الحكومات، وأصبحت الأعمال توجه بالبيانات بشكل أكبر، ومدعومة بالدليل وأكثر اعتماداً على التقنية (تكنوقراطية).

يناقش هذا الفصل مجموعة مختارة من المخاوف الأخلاقية، والاجتماعية، والسياسية، والقانونية التي تثيرها ثورة البيانات، وهذه المخاوف تشمل مراقبة البيانات وتعقب البيانات، والخصوصية، وأمن البيانات، والتشخيص، والتصنيف الاجتماعي والخطوط الحمراء (Redlining) والتي تعني ممارسة الحرمان أو زيادة التكلفة في الخدمات لمناطق لها غالبية عرقية محددة بناء على البيانات، والتحكم في الانسلال، والحوكمة التوقعية، وحوكمة المؤسسات، والاعتماد التام على التقنية وحوكمة الشركات والإغلاق التقني، والملكية

وحقوق الملكية الفكرية. فالفصل يناقش كيفية التفكير في كل قضية من هذه القضايا، مع الآراء المتباينة داخل كل علم على حدة وبين العلوم المختلفة بعضها البعض، وكذلك داخل وبين الشركات والحكومات والمجتمع المدني، ومن لديه أجندات مختلفة، ومصالح مكتسبة، وحساسيات سياسية. ولذا، فإنه لا توجد إجابات سهلة لحل هذه القضايا التي يناقشها هذا الفصل، في حين أن الحلول تتكون دوماً من تسويات وحلول وسط. ومن المؤكد، على ما ذكرنا، أنه كلما تنتشر ثورة البيانات، فمن المرجح أن تزداد حدة المناقشات والجدال بشأن هذه المخاوف، وخصوصاً عند القيام بمحاولات من أجل صياغة تشريعات جديدة للتعامل مع التطورات التقنية التي تساعد على انبثاق طرق جديدة لإنشاء وتعزيز وتحليل البيانات، ومن ثم تؤدي إلى ظهور قضايا جديدة مما يجعل من التشريعات والقوانين القديمة تشريعات وقوانين عفا عليها الزمن.

تعقب البيانات ومراقبة البيانات (Data Shadows and Dataveillance):

كما ناقشنا في الفصول السابقة، يتم تكوين كمية هائلة من البيانات في الوقت الراهن وذلك فيما يتعلق بالمواطنين في جميع ولايات الدولة. إن إنتاج وتحويل مثل هذه الكمية من البيانات ينمو ويتزايد باطراد، حيث تسعى الدولة والشركات ومنظمات المجتمع المدني بشكل فعال للحصول على البيانات المتعلقة بمواطنيهم وعملائهم والأعضاء المشتركين في أنشطتهم. وفي الواقع، أصبح من الصعبه يمكن أن يؤدي الفرد جزءاً من حياته اليومية دون أن يترك أثراً لذلك بسبب الدور غير المباشر للتقنيات الرقمية والبرامج واستخدام المعارف الدالية (Kitchin and Dodge 2011). فحتى ولو لم يقيم المشتري باستخدام بطاقته الائتمانية لشراء البضائع من المتاجر، إلا أن حضور هذا المشتري قد تم تسجيله من خلال كاميرات المراقبة، وحتى لو استخدم الشخص اسم مستخدم مجهول على وسائل التواصل الاجتماعي فإنه يتم تسجيل عنوان بروتوكول الإنترنت (IP Address) وعنوان التحكم بالوصول للوسائط (MAC Address). ومن ثم فنحن نترك آثار البيانات بشكل روتيني في أعقابنا، على الرغم من أننا في الغالب لا نملك السيطرة الكاملة على شكل ومدى وكيفية استخدام هذه الآثار.

وكما أشار (Koops 2011)، على سبيل المثال، فإن هيئة حماية البيانات الهولندية قدرت أن المواطن الهولندي العادي مندرج في نحو ٢٥٠ - ٥٠٠ قاعدة بيانات، في حين أن المواطنين الأكثر نشاطاً من الناحية الاجتماعية مندرجون في نحو ١٠٠٠ قاعدة بيانات. ولا تشمل قواعد البيانات هذه على الآثار الرقمية للمواطنين فحسب (البيانات التي يتركها المواطنون أنفسهم في أعقابهم)، ولكنها تتضمن أيضاً بيانات الصور التخيلية (Data Shadows) الخاصة بالأفراد (البيانات عن المواطنين التي ينشئها الآخرون)، وهي توفر بشكل متزايد عملية تعقب البيانات الخاصة بالموقع والتفاعلات والمعاملات عبر المكان والزمان (Clarke 1994a; Lyon 2007; Dodge and Kitchin 2005; Koops 2011). وعلاوة على ذلك، ففي حين أن الآثار الرقمية والصور التخيلية للبيانات سريعة الزوال ولا تدوم إلا فترة محدودة للغاية، فإن النسخة الرقمية منهما تدوم بشكل كبير للغاية ويمكن تخزينها في قواعد البيانات لفترة غير محدودة (Koops 2011)، مما يؤدي إلى إنشاء عالم من الممكن له أن يتذكر جميع نقاط البيانات ولا ينسى على الإطلاق (Dodge and Kitchin 2007b). فهذه الآثار الرقمية والصور التخيلية للبيانات يتم تفتيتها وتقسيمها ونشرها عبر الكثير من المنظمات والخوادم، وهي تكون عرضة للدمج والتقسيم (Raley 2013). وفي أفضل الأحوال، تشكل الآثار الرقمية والصور التخيلية للبيانات القليل من الرؤى - رؤى محدودة من زوايا مرتفعة جزئية من مواقف محددة مع إسقاط الرؤى المحددة (Amin and Thrift 2002) - وذلك بدلاً من تمحيص هذه الآثار والتخيلات سويماً من أجل تكوين رؤية شاملة وعامة تحيط بكل شيء. وعلى الرغم من ذلك، وإذا وضعنا في اعتبارنا ارتباطية البيانات وقيمة تضخيم البيانات، فإن الاتجاه السائد سيكون تجاه تجميع وجهات النظر المحدودة من أجل تكون تصور أكثر قوة.

وإجمالاً، توفر الآثار الرقمية والصور التخيلية للبيانات سجلاً مفصلاً بحياة المرء اليومية، وأنماط استهلاكه، وعمله، وسفره، واتصالاته، ولعبه، وتفاعلاته مع المنظمات، وأفكاره واهتماماته. وليس الأفراد فحسب هم من يقعون تحت طائل هذه المراقبة الشاملة، ولكن هذه المراقبة تشمل أيضاً الأشياء والمؤسسات والتعاملات والمناطق (Dodge and Kitchin 2005). ولم يحدث أبداً في السابق أن يتم تكوين مثل هذه الكمية من البيانات عن حياة الناس الاقتصادية والاجتماعية بسهولة بالغة، وأن يتم إتاحة هذه البيانات على نطاق واسع

للغاية من خلال أسواق البيانات والبنى التحتية للبيانات ومبادرات البيانات المفتوحة. وفي الواقع، فإن اكتساب الوصول إلى البيانات الحساسة ليس مقتصرًا على الحكومات والشركات والمنظمات، وذلك بسبب أن مراقبة النظراء أصبحت أمرًا أكثر انتشارًا بشكل نسبي دون معرفة أو دون الحصول على إذن الشخص الخاضع للمراقبة (Andrejevic 2007). وبالطبع، كان هذا الأمر متاحًا على الدوام من خلال التطفل والقييل والقال، ولكنه أصبح أكثر سهولة في الوقت الحالي من خلال الإنترنت سواءً كان هذا من خلال البحث عن البيانات أو شراء تلك البيانات، أو من خلال استعراض سجلات الأجهزة الرقمية كالهواتف الخلوية، وبرامج تصفح الإنترنت، أو برامج الملاحاة بالأقمار الصناعية في السيارات... وغير ذلك، أو من خلال وضع أجهزة مراقبة رخيصة نسبيًا مثل الكاميرات الدقيقة المصغرة. وكما ناقشنا في الفصل الثاني، فإن سُماسرة البيانات يسمحون بدراسة تواريخ الأشخاص وذلك فيما يتعلق ببعض المسائل مثل الحالة الاجتماعية، والإفلاس، وملكية العقارات، والتاريخ الوظيفي، والسجلات الجنائية، والحالة الضريبية، ودعاوى العنف الجنسي أو العنف الأسري، والخلفية التعليمية... وغير ذلك.

ومن ثم فإن طوفان البيانات مكشوف أمام مراقبة البيانات والاستغلال، فمراقبة البيانات، تُعد أحد أساليب المراقبة التي يتم تفعيلها من خلال فرز وتدقيق مجموعات البيانات وذلك بغرض التحديد والمراقبة، والتتبع، والتنظيم، والتنبؤ، والتوجيه (Clarke 1988; Raley 2013)، وتعمل مراقبة البيانات على تحويل الكميات الهائلة من الأفعال والأفكار التي تمثلها هذه البيانات إلى صورة واضحة ومقروءة ونماذج واضحة للحكم على الأشياء (Curry et al. 2004: 359)، وهي مكون أساسي من الأشكال الحديثة للحكومة والحاكمية. وفيما يتعلق باستخدام البيانات، وكما ناقشنا في الفصل السابع، يتم استخدام البيانات من أجل توصيف واستهداف الأشخاص وذلك بهدف زيادة الأرباح. ومن ثم ففي غالب الأحيان يتم إعادة تحديد الغرض من البيانات واستخدامها في أغراض لم تكن مقصودة على الإطلاق عند تكوين هذه البيانات، ومن ثم فإن مراقبة الآثار الرقمية والصور التخيلية للبيانات يثيران العديد من المخاوف الإضافية وخاصة فيما يتعلق بقضايا مثل الخصوصية، والتنميط والفرز الاجتماعي، وإقرار أساليب متعددة للحكومة.

الخصوصية (Privacy):

الخصوصية هي حالة يتوقعها ويقدرها الكثير من الناس، وهي تُعدُّ حقاً من حقوق الإنسان الأساسية، وهي من الحقوق المقدسة في الكثير من القوانين الوطنية والدولية. وفي الولايات المتحدة، فإن الخصوصية مغطاة بشكل كبير في القوانين المسماة قوانين الخصوصية، أما في الاتحاد الأوروبي، فهي تقع ضمن نطاقات قوانين حماية البيانات (Minelli et al. 2013). والخصوصية مصطلح متعدد الأبعاد فيما يتعلق بمعناه، وهو يستخدم في غالب الأحيان في طرق معتمدة على السياق، ولكن على وجه العموم هو يشير إلى الممارسات المقبولة فيما يتعلق بالوصول إلى والإفصاح عن المعلومات الشخصية والحساسة (Elwood and Leszczynski 2011). ومن الوسائل المستخدمة في إيضاح الأبعاد المتعددة للخصوصية دراسة الوسائل العديدة التي يمكن من خلالها انتهاك الخصوصية والأضرار المرتبطة بهذه الانتهاكات، (Solove 2006، انظر الجدول رقم 10.1). فالذي يتضح لنا من خلال هذا التصنيف هو أن انتهاك الخصوصية يمكن أن يكون له العديد من التأثيرات على الرفاهية الجسدية والعاطفية للأفراد، ويجعل هؤلاء الأفراد عرضةً للأنشطة المضرة التي يقوم بها الآخرون، ويجعلهم كذلك عرضة لاختلال توازن السلطات غير المتماثل.

ويوجد بعض الشكوك بأن مفهوم الخصوصية يتغير. وكما تمت الإشارة أعلاه، فإن الناس عرضة الآن لمستويات من الفحص والتدقيق وكذلك أساليب المراقبة أكثر من أي وقت مضى. فالمهام التي كانت تعد في الماضي خاصة إلى حد كبير، يتم الآن مراقبتها وتسجيلها. على سبيل المثال، فإن البرامج التي كان يشاهدها الشخص على التلفاز لم تكن معروفة في الماضي إلا لمن هم حاضرون مع الشخص ذاته في الغرفة ذاتها، في حين أنه يمكن لشركات التلفاز الآن مراقبة ما الذي يتم مشاهدته على كل تلفاز على حدة وذلك بسبب استبدال البث التلفزيوني وحيد الاتجاه بالتواصل ثنائي الاتجاه. وما يشتره المرء كان في السابق معروفاً للعميل والشخص الذي يخدم هذا العميل فحسب والآن أضحت هذه المعلومات التي يتم جمعها بشكل روتيني من خلال إدراج النقود الرقمية في عملية الدفع متاحة لشرائح أكبر. والأماكن التي كان يسافر إليها الفرد لم تكن سابقاً معروفة إلى حد كبير لأي شخص ما عدا

المسافرين ومرافقيهم، ولكن في الوقت الراهن أضحى العديد من الأنظمة تكشف موقع الشخص والطرق التي اجتازها كنظم الملاحة باستخدام الأقمار الصناعية (GPS) المثبتة في السيارات أو أجهزة الهواتف الخلوية، ونظام تثلث الإشارات في أبراج شبكات الهاتف، وأنظمة النقل العام التي تعتمد على استخدام شريحة تحديد الهوية باستخدام موجات الراديو (RFID) في أنظمة أو أكشاك دفع رسوم المرور. ومثل هذه البيانات التي تنتجها هذه الأنظمة يتم إعادة تجميعها ونشرها وبيعها.

الجدول رقم (١ - ١٠)

تصنيف الخصوصية

الوصف	انتهاك الخصوصية	النطاق
مشاهدة أو الاستماع إلى، أو تسجيل الأنشطة التي يقوم بها الشخص	المراقبة	جمع المعلومات Information Collection
الأشكال المتعددة من الاستجواب أو التقصي من أجل الحصول على المعلومات	الاستجواب	
تجميع أجزاء مختلفة من البيانات عن شخص ما	التجميع	معالجة المعلومات Information Processing
ربط المعلومات بأشخاص معينين	التعريف	
الإهمال في حماية المعلومات المخزنة من التسيريات والوصول إليها بطريقة غير ملائمة	عدم الأمان	
يتم استخدام المعلومات التي تم تجميعها لغرض معين في أغراض أخرى دون الحصول على موافقة الشخص موضوع البيانات	الاستخدام الثانوي	
الإخفاق في السماح للشخص موضوع البيانات بمعرفة البيانات التي يمتلكها الآخرون عنه والمشاركة في معالجة واستخدام هذه البيانات، وهذا يشمل الإقصاء والمنع من التمتع بالمقدرة على الوصول إلى هذه البيانات وتصحيح الأخطاء الواردة فيها.	الاستبعاد	

انتهاك السرية	إخلاف الوعد في المحافظة على سرية بيانات الشخص	نشر المعلومات Information Dissemination
الإفصاح	الكشف عن المعلومات عن شخص ما والتي تؤثر في الطرق التي يحكم بها الآخرون على شخصيته	
الكشف	الكشف عن عري أو حزن أو الوظائف الجسدية لشخص ما	
زيادة إمكانية الوصول	زيادة إمكانية الوصول إلى المعلومات	
الابتزاز	التهديد بالكشف عن معلومات شخص ما	
الاستيلاء	استخدام هوية الشخص موضوع البيانات من أجل خدمة أهداف أو مصالح شخص آخر	
التشويه	نشر معلومات خاطئة أو مضللة عن شخص ما	
التطفل	الأعمال الانتهاكية التي تزعج هدوء أو انزعاج شخص ما	الانتهاك Invasion
التدخل المتعلق بالقرارات	التدخل في قرارات الشخص موضوع البيانات التي تتعلق بشئونه الخاصة.	

المصدر: تم تجميعه من (Solove 2006).

وعلاوةً على ذلك، فنحن نعيش في عالم أكثر انفتاحاً وشفافية بكثير عما اعتدنا عليه سابقاً. فالمعلومات التي كانت تُعدُّ خاصةً في السابق يتم مشاركتها الآن بشكل أكثر حرية كالسير الذاتية من خلال موقع لينكد إن (LinkedIn)، والصور والفيديوهات العائلية من خلال مواقع فليكر (Flickr) وإنستجرام (Instagram) ويوتيوب (YouTube)، والقصص الشخصية والعائلية من خلال موقع فيسبوك (Facebook) والمدونات، والأفكار والمعتقدات الشخصية من خلال تويتر (Twitter) وغرف الدردشة والتعليقات على الإنترنت. فما كان يتم مشاركته في السابق مع عدد لا يتجاوز أصابع اليد الواحدة من الأشخاص كأفراد العائلة، والأصدقاء المقربين، وأرباب العمل في المنزل أو المقهى المحلي أو مكتب للموارد البشرية يتم نشره الآن على المستوى العالمي مما يسمح لأي شخص بمشاهدته وتعديله (Minelli et al. 2013).

وعلى الرغم مما ذكر سابقاً، لا يتصرف كل شخص بسهولة وبحرية فيما يتعلق ببياناته الشخصية، حتى ولو كان الشخص يتصرف على هذا النحو، فإنه لا يتوقع بالضرورة أن هذه البيانات سوف يتم انتقاؤها، ومعالجتها، وتجميعها، وبيعها. وهم لا يتوقعون أيضاً أن تطبيقات البرامج التي لا يوجد بها أي عنصر من عناصر وسائل التواصل الاجتماعي، والتي يعتبرونها تطبيقات خاصة ربما تقوم بتكوين ونقل البيانات إلى مطوري تلك التطبيقات، أو أن بعض الشركات ربما تحاول أن تقوم بالهندسة العكسية لإستراتيجيات حماية الخصوصية مثل إخفاء الهوية من أجل إنشاء ملفات شخصية موحدة، ومع ذلك، فإن هذا الأمر يحدث بانتظام. على سبيل المثال، وجدت صحيفة وول ستريت جورنال (Wall Street Journal) في اختبار أجري على 101 تطبيق للهواتف الذكية، أن 56 تطبيقاً نقلت معرف الجهاز المتفرد للهاتف إلى شركات أخرى وذلك دون معرفة المستخدم أو الحصول على موافقته، وأن 47 تطبيقاً أرسلت موقع الهاتف، وأن 5 تطبيقات أرسلت البيانات الشخصية للمستخدمين، وأن 45 تطبيقاً لم يكن بها أي ارتباط مع سياسات حماية الخصوصية التي يمكن للمستخدمين مشاهدتها (Efrati et al. 2011). كما وجدت شركة الثقة الإلكترونية (TRUSTe) أن 19% فقط من أعلى 3340 تطبيقاً مرتبطون بسياسة الخصوصية وأن متاجر شركتي أبل وجوجل يشترطان ضرورة اشتغال التطبيقات على هذه السياسة (Coterill 2011). ومع ذلك، وكما أشار (Gralla et al 2011)، يمكن للتطبيقات أن تتتبع وتنقل "عادتك على الشبكة، وأن تبحث في قائمة جهات الاتصال لديك، وأن تجري مكالمات هاتفية دون أن تعرف، وأن تتتبع موقعك، وأن تفحص ملفاتك، وغير ذلك الكثير".

إن مشهد الخصوصية الآن في حالة تغير مستمر، وهو يخيب كلاً من التوقعات القانونية والاجتماعية، وبالنسبة للبعض، فإن مفهوم الخصوصية قد تعطل بشكل كبير (Rambam 2013; Rubenking 2008). ويرى البعض أنه من الصعوبة بمكان المحافظة على الخصوصية في التطبيق العملي، كونها تحجب وتخفي تجربة المستخدم، وهي معوق اقتصادي، ولا يبدو أن معظم الناس يعارضون أن يتم التنقيب في بياناتهم، وإذا لم يكن لديك شيء لتخفيه، فما المشكلة في أن يتم معرفة بياناتك؟ (Solove 2007; Raley 2013: 126). وبالنسبة للبعض الآخر، فإن الخصوصية حق يتعين حمايته لأنه من الحقوق التأسيسية لـ "المواطنة المطلعة والمتأملمة" وهو مكون أساسي أيضاً لحرية التعبير (Cavoukian 2009). وعلى هذا، يُرى أن

الخصوصية "سمة ضرورية وتأسيسية للأنظمة السياسية الديمقراطية المتحررة" (Cohen 2: 2012). فإذا تم إنكار الأشكال المتعددة من الخصوصية التي تم إيجازها في الجدول رقم (١٠-١)، فإن الناس سيكونون عرضة لأنواع مختلفة من الأضرار الشخصية وأشكال الحوكمة الأكثر ضرراً. ويبدو أن جميع المعلقين يؤيدون وجهة النظر القائلة بأن تشريعات الخصوصية لم تعد ملائمة للغرض التي شرعت من أجله، وأنه يتعين تحديث تلك التشريعات لكي تلائم الأوقات التي نعيشها الآن.

الجدول رقم (٢ - ١٠)

مبادئ الممارسة العادلة للمعلومات

المبدأ	الوصف
الإخطار (Notice)	يتم إخطار الأفراد بأن بياناتهم يتم إنشاؤها ويتم إعلامهم بالغرض الذي سيتم استخدام هذه البيانات فيه
الاختيار (Choice)	يتمتع الأفراد بحرية اختيار المشاركة أو يؤثرون عدم المشاركة وذلك فيما يتعلق بما إذا كانوا يرغبون في استخدام بياناتهم أو الإفصاح عنها وكيفية القيام بذلك
الموافقة (Consent)	يتم إنشاء البيانات والإفصاح عنها فحسب بعد موافقة الأفراد المعنيين
الأمن (Security)	يتم حماية البيانات من الفقد، وإساءة الاستخدام، والوصول غير المصرح به، والإفصاح عنها، وتعديلها أو تدميرها.
التكامل (Integrity)	تكون البيانات موثوقة ودقيقة وكاملة وحالية
الوصول (Access)	يمكن للأفراد الوصول إلى بياناتهم الشخصية والتحقق منها والتثبت منها
المساءلة (Accountability)	يكون مقتني البيانات مسئولاً عن ضمان تحقق المبادئ السابق ذكرها وأن يكون لديه الوسائل اللازمة من أجل ضمان الالتزام بهذه المبادئ

المصدر: (Minelli et al. (2013: 156).

الجدول رقم (٣ - ١٠)

أنواع المعلومات المحمية

المعلومات الشخصية المعرفة (PII): هي أي معلومات تُعرف شخصاً ما بصورة مباشرة أو غير مباشرة	المعلومات الحساسة: هي أي معلومات يُمكن للكشف غير المصرح لها أن يسبب الإحراج أو الضرر لشخص ما	المعلومات الأخرى التي يمكن استخدامها لتخمين هوية شخص ما
الاسم	العرق / الانتماء العرقي	الأشياء المفضلة
العنوان البريدي / الرمز البريدي	الآراء السياسية	هوية ملف تعريف الارتباط (cookie)
البريد الإلكتروني	المعتقدات الدينية / الفلسفية	عنوان بروتوكول الإنترنت الثابت
رقم الجوال / الهاتف	عضوية النقابات التجارية	
رقم الضمان الاجتماعي	المعلومات الصحية / الطبية	
رقم رخصة القيادة	الحالة الاجتماعية / الحياة الجنسية	
رقم الحساب المالي (البنكي)	العمر	
رقم بطاقة الصراف / بطاقة الائتمان	الجنس	
	السجل الجنائي	

المصدر: مقتبس من (Minelli et al. (2013: 159).

وفي الوقت الراهن، فإن تشريعات الخصوصية مبنية حول الحقوق الشخصية والموافقة المتعلقة بإنشاء، واستخدام، والإفصاح عن البيانات الشخصية (Solove 2013). وهي تتكون من سبعة مبادئ عالمية للخصوصية (Minelli et al. (2013: 156)، (انظر الجدول رقم ٢ - ١٠) وفي إطار هذه المبادئ، فإن الأفراد، من الناحية النظرية، يتم منحهم الحق للتحكم في بياناتهم الشخصية ومنح موافقتهم للآخرين فيما يتعلق بهذه البيانات. وعلاوة على ذلك، فإنه يتعين على مقتنبي البيانات الالتزام بقوانين حماية البيانات التي تقيد الإفصاح عن المعلومات الشخصية

المعرفة (Personally Identifiable Information- PII)، والمعلومات الحساسة، والبيانات الأخرى التي يمكن استخدامها للاستدلال على الشخصية (انظر الجدول رقم ٣-١٠)، وتؤيد تقليص البيانات بحيث يقتصر إنشاء البيانات على البيانات الضرورية لتحقيق غرض معين (Tene and Polonetsky 2012). وعلى وجه العموم، فإن الحل لاشتراطات المعلومات الشخصية المعرفة (PII) هو استخدام أساليب إخفاء الهوية مثل عدم تحديد الهوية (إخفاء المعلومات الشخصية المعرفة)، والأسماء المستعارة، والتجميع، بالإضافة إلى التشفير، والتخزين الآمن، والقيود الخاصة بالوصول إلى المعلومات (Coterill 2011).

وأخيراً، تجدر الإشارة إلى وجود العديد من المشاكل الإدراكية والبنوية المرتبطة بتشريعات الخصوصية الحالية.

وتحاول الدول بصورة روتينية التحايل على قوانين الخصوصية وذلك من أجل أغراض الأمن وجمع المعلومات الاستخباراتية، بحسب ما تم الكشف عنه مؤخراً بواسطة برامج تحليل وجمع البيانات السرية في الولايات المتحدة الأمريكية، والمملكة المتحدة، والعديد من الحكومات الأخرى (انظر الفصل السابع). كما أن الكثير من الشركات لا تشعر بأنها مجبرة على تقديم سياسة الخصوصية للأفراد وذلك لأن هذه الشركات تدعي أنها تجمع معلومات مجهولة المصدر، ومن ثم فهي لا تقع تحت طائلة ممارسات المعلومات العادلة. أو ربما تقدم هذه الشركات سياسة الخصوصية، وذلك عند الاستهلال الأولي للالتزامات (على سبيل المثال، عند تثبيت برنامج ما أو الاشتراك في خدمة ما) وتكون هذه السياسة مصاغة بلغة معقدة وغامضة، وعادةً ما تشتمل على الاحتفاظ بالحق في تعديل هذه السياسة في وقت لاحق دون تشاور إضافي مع الشخص (Rubinstein 2013).

وفيما يتعلق بالبيانات التي تم تكوينها بدون معرفة الشخص، فمن الصعوبة بمكان اكتشاف مثل هذه الممارسات ومواجهتها من الناحية القانونية. أما في حالة الخصوصية ذات الإدارة الذاتية، وكما شرح (Solove 2013)، فهي ببساطة مرهقة للغاية وبالغة الصعوبة على الأفراد كونها تتطلب قيامهم بتنظيم خصوصيتهم عبر الكثير من الكيانات، والموازنة بين التكاليف والمنافع المرتبطة بالموافقة على الشروط والبنود دون معرفة كيف سيتم استخدام هذه البيانات في الوقت الحالي وفي المستقبل، وأيضاً تقييم التأثيرات الشاملة

والتراكمية لبياناتهم التي سيتم دمجها مع مجموعات البيانات الأخرى. وبناء على ذلك، أشار (Solove 2013) إلى أن:

(1) الناس لا يقرؤون سياسات الخصوصية؛ (2) إذا قرأ الناس سياسة الخصوصية، فإنهم لا يفهمونها؛ (3) إذا قرأ الناس سياسة الخصوصية وفهموها، فإنهم في الغالب يفتقدون الخلفية الكافية لاتخاذ اختيار مبني على المعرفة؛ (4) وإذا قرأ الناس سياسة الخصوصية وفهموها، وكان بإمكانهم اتخاذ اختيار واع ومبني على المعرفة، فإن اختيارهم ممكن أن ينحرف وذلك من جراء العديد من الصعوبات المرتبطة باتخاذ القرار (Solove 2013: 1888).

ومن ثم فإن الموافقة تكون في غالب الأحيان ممارسة عديمة الجدوى (Rubenstein 2013: 2)، وذلك عندما يقوم الأفراد عن غير قصد بالتخلي عن الحقوق دون إدراك لمدى أو عواقب الفعل الذي يقومون به. ومن ثم، فإن سياسات الخصوصية في غالب الأحيان تخدم إخلاء المسؤولية للشركات بصورة أكبر مما تخدم عملية ضمان الخصوصية للمستهلكين، (Tens and Polonestky 2012). وعلاوة على ذلك، فمن الواضح أنه في عصر البيانات الكبيرة، فإن إستراتيجية إزالة التعرف على الأشخاص توفر شكلاً ضعيفاً فحسب من أشكال الخصوصية لأن من الممكن القيام بالهندسة العكسية للعملية وذلك من خلال التنقيب في مجموعات البيانات وتجميعها (Narayanan and Shmatikov 2010). فعلى سبيل المثال، يشرح نيف سينجر (Niv Singer) من شركة تراكسك (Traxc) (والمقتبس في Minelli et al. 2013: 31) عملية إزالة التعرف على الأشخاص كما يلي:

يمكن في بعض الأحيان أن يكون من الصعوبة بمكان توحيد الملفات الشخصية الاجتماعية لمستخدم واحد والذي ربما يستخدم العديد من الأسماء أو تعاملات مختلفة في كل واحدة من شبكات التواصل الاجتماعي، ولذا فلقد قمنا ببناء خوارزميات تقوم بالتنقيب في العوامل الأساسية التي تشتمل على محتوى المشاركات والموقع على سبيل المثال لا الحصر، وذلك من أجل توفير توحيد للهوية بشكل فعال إلى درجة كبيرة.

ونتيجة لهذه المخاوف، يدرس الاتحاد الأوروبي منذ فترة سياسات حماية البيانات (انظر المفوضية الأوروبية؛ Koops 2011; Hon et al. 2011; European Commission 2012، Rubinstein 2013). وتشتمل اقتراحات الاتحاد الأوروبي على أن تكون الموافقة صريحة وليست ضمنية أو مفترضة، وأن يكون وصول الأفراد إلى البيانات الخاصة بهم سهلاً وشاملاً، وأن يكون للأفراد الحق في قابلية النقل للبيانات بمعنى المقدرة على نقل البيانات الشخصية من مزود خدمة إلى مزود آخر، والحق في أن يتم نسيان الشخص، حيث يمكن للأفراد أن يطلبوا حذف بياناتهم إذا لم تكن هناك أي أسس مشروعة للاحتفاظ بهذه البيانات، وأن تنطبق هذه القواعد أيضاً على الشركات خارج الاتحاد الأوروبي إذا كانت هذه الشركات عاملة في سوق الاتحاد الأوروبي وتقدم خدماتها لمواطني دول الاتحاد الأوروبي (European Commission 2012). وبالمثل، تقدمت مفوضية التجارة الفيدرالية في الولايات المتحدة الأمريكية باقتراح يتضمن ثلاثة تعديلات من أجل توفير حماية أكثر فعالية للخصوصية: الخصوصية من خلال التصميم حيث يتم تأسيس الخصوصية بشكل متأصل في كل مرحلة من مراحل تطوير المنتج؛ منح الشركات والمستهلكين خيار مبسط يمنحهم المقدرة على اتخاذ القرارات بشأن بياناتهم الخاصة، وهذا يشمل تطبيق آلية لعدم التتبع والحصول على الموافقة الصريحة فيما يتعلق بالبيانات الحساسة، أو الحصول على موافقة قبل استخدام البيانات في غرض مختلف بشكل مادي عن الغرض الذي تم من أجله تكوين هذه البيانات؛ وشفافية أكثر فيما يتعلق بالبيانات وجمعها واستخدامها، وهذا يشمل الوصول المعقول إلى البيانات من قبل هؤلاء الذين تمثلهم تلك البيانات، مع منحهم المقدرة على تصحيح أو حذف البيانات (مفوضية التجارة الفيدرالية 2012 Federal Trade Commission). وتنشد إستراتيجية الخصوصية من خلال التصميم تطبيق إستراتيجية مختلفة عن الالتزام التشريعي والتنظيمي، وهي تقترح أن تكون الخصوصية هي الوضع الافتراضي للعملية؛ أي نظام يسعى بفعالية وبوضوح إلى ضمان الخصوصية (Cavoukian 2009). ومن شأن المبادئ السبعة المذكورة في الجدول رقم (٤ - ١٠) أن تعزز هذا المنهج.

الجدول رقم (٤ - ١٠)

المبادئ السبعة الأساسية للخصوصية من خلال التصميم

المبدأ	الوصف
استباقية وليست تفاعلية؛ الوقاية وليس العلاج	ينبغي أن تسعى نظم تكنولوجيا المعلومات لتوقع الخصوصية تحديد المخاوف المتعلقة بالخصوصية وتجنبها بدلاً من السعي إلى حل المشاكل الناتجة عن مخالفات الخصوصية بعد حدوثها
الخصوصية هي الوضع الافتراضي	الخصوصية محمية تلقائياً ولا تتطلب أي عمل من جانب الأفراد
الخصوصية مدمجة في التصميم	حماية الخصوصية هي سمة أساسية للتصميم والهندسة الهيكلية لأنظمة تقنية المعلومات وليست ميزة إضافية متدنية الأهمية
الفعالية الكاملة - عائد إيجابي وليس محصلة صفرية	جميع المصالح والأهداف المشروعة يتم استيعابها، بدلاً من أن يكون هناك مقايضات بين الخصوصية واعتبارات أخرى مثل الأمن
الأمن من النهاية للنهائية-حماية كاملة طوال دورة حياة النظام	تكون الخصوصية جزءاً لا يتجزأ من النظام من بداية إطلاقه إلى حين التخلص منه
الوضوح والشفافية - فلتكن مفتوحة على الدوام	تكون جميع الأجزاء المكونة للنظام وعملياته واضحة وشفافة للمستخدمين والشركات الموفرة على السواء وتخضع لتدقيق مستقل
احترام خصوصية المستخدم - أي جعلها تتمحور حول المستخدم	يجب أن يبنى النظام حول الأفراد، وأن يحمي مصالحهم، وأن يكون مُمكنًا لهم

المصدر: (Cavoukian 2009).

لقد اقترح العديد من العلماء المتخصصين في الخصوصية وأيضاً العديد من مجموعات الضغط الصناعية مناهج متعددة لكيفية التعامل مع قضية الخصوصية، فالعلماء المختصون بموضوع الخصوصية سبق لهم التطرق للمواضع التالية: الإصلاح التشريعي المصمم من أجل حماية حقوق المواطنين، والذي يشمل بعض القضايا كالحق في الصفحة البيضاء، وتواريخ الانتهاء للبيانات (Koops 2011). ومن الذي يكون مسئولاً عن ضمان الخصوصية، المطورون أم الوكالات أم المستخدمون، وعن الوسائل التقنية والإدارية لتنفيذ هذه الأمور (Coterill 2011). وما الذي يشكل المعلومات الخاصة وكيفية تأطير الخصوصية حول المخاطر ومدى الضرر بدلاً من تأطيرها حول تعريف المحتوى (Hen

et al. 2011). واقترح هؤلاء العلماء أيضاً دخول الأفراد في شركات مع المطورين حيث يكون بإمكانهم أن يختاروا بشكل أكثر استباقية ما البيانات التي يرغبون في نشرها، ولمن تُنشر هذه البيانات، وتحت أي ظروف تُنشر هذه البيانات. كما اقترحوا أن تتيح الشركات للمستخدمين الوصول إلى بياناتهم الشخصية بصيغة قابلة للاستخدام وذلك من أجل مصلحتهم الشخصية. وأخيراً، اقترحوا أن تقوم الشركات بمشاركة الثروات الناتجة عن تحويل البيانات الشخصية إلى أموال (Tene and Polonetsky 2012; Rubinstein 2013). وكمثال على هذه المشاركة المفيدة لكلا الطرفين في الثروات المتحصلة من بيع بيانات الشبكات الذكية حيث يتم استخدام البيانات التي تم تكوينها من خلال العدادات الذكية والمتعلقة باستهلاك أفراد الأسرة للكهرباء من قبل شركة الكهرباء من أجل ضمان كفاءة التزويد بالكهرباء، ويتم مع ذلك توفير تطبيقات لأفراد الأسرة لتمكينهم من مراقبة استخدامهم الشخصي وأن يعدلوا سلوكياتهم من أجل توفير المال. إن سوق الصناعة، إلى حد كبير، يُريد إما الاستمرار في الإمدادات الحالية على ما هي عليه أو تقليل هذه الإمدادات، مع إدارة الخصوصية من خلال تنظيم يقود السوق ولا يكبح الاستفادة الاقتصادية من البيانات. وفي هذه الحالة، فمن المتصور أن الشركات سوف تقوم بعملية التنظيم الذاتي خشية حدوث أضرار بسمعة تلك الشركات أو أن يعبر عملاء الشركات عن عدم رضاهم من خلال عدم التعامل مع هذه الشركات إذا انتهكت هذه الشركات خصوصية هؤلاء العملاء بشكل مبالغ فيه (Minelli et al. 2013; Mayer-Schonberger and Cukier 2013).

أمن البيانات (Data Security):

مع وضع قيمة البيانات في الاعتبار، وخاصة البيانات الشخصية التي يمكنها أن تسهل سرقة الهوية، أو البيانات التجارية التي يمكن قرصنتها أو استخدامها من أجل تحقيق ميزة تنافسية، أصبح أمن البيانات مجالاً هاماً في حماية البيانات. فالأفراد يمكنهم فتح العديد من الحسابات سواءً الحسابات الشخصية أو حسابات العمل، مع العديد من الهيئات والبايعين وذلك باستخدام العديد من رموز الهوية الشخصية وأسماء المستخدمين وكلمات السر. فالبيانات المتعلقة بهؤلاء الأفراد، والبيانات الشخصية الخاصة بهم، يتم تخزينها في الكثير من الأماكن (على سبيل المثال أجهزة الحاسب الآلي، والأجهزة الرقمية، والهواتف

الذكية، ومحركات الأقراص الصلبة الخارجية، وأجهزة تخزين البيانات المتنقلة مثل (USB Drive) والخوادم. وتكون هذه الحسابات والأجهزة الرقمية عرضة للقرصنة (الدخول على الحسابات أو الوصول إلى الأجهزة)، والبرمجيات الخبيثة (الرموز التي يتم تثبيتها بشكل سري على جهاز الحاسب والتي يمكنها الوصول إلى المعلومات الحساسة أو تسجيل ضغطات لوحة المفاتيح وترسل هذه الأشياء إلى طرف ثالث)، والانتحال (Phishing) (الرسائل الإلكترونية الاحتيالية أو الهندسة الاجتماعية من خلال المكالمات الهاتفية التي تحاول أن تجعل الشخص يتطوع بتقديم معلومات حسابه)، مما يؤدي إلى سرقة واختلاس البيانات.

وقد ناقش (Gantz and Reinsel 2011) بالتفصيل خمسة مستويات لأمن البيانات، وكل مستوى من هذه المستويات له محركات مختلفة إلى حد ما، وذلك على الرغم من أن جميع هذه المستويات تتطلب إجراءات أمن استباقية. وهذه المستويات تكون على النحو التالي:

- (١) الخصوصية: من أجل المحافظة على المعلومات وتقييد تداولها.
- (٢) الامتثال والالتزام: من أجل حماية البيانات التي من المحتمل أن تكون قابلة للكشف عند التقاضي، أو الخاضعة لشروط الاستبقاء.
- (٣) الاحتراز: من أجل حماية البيانات التي يمكنها أن تؤدي إلى أو تساعد على سرقة الهوية.
- (٤) السرية: من أجل التأكد من حماية المعلومات السرية مثل الأسرار التجارية.
- (٥) التأمين، من أجل حماية المعلومات عالية السرية مثل الصفقات التجارية، والملفات الشخصية، والسجلات الطبية، والمعلومات العسكرية، التي يمكن أن يكون لها تأثير يتجاوز السرقة الشخصية.

وتُعد إدارة المستويات الخمسة السابقة لأمن البيانات مهمة بالغة الأهمية بالنسبة للأفراد والشركات والمؤسسات. وعلى وجه العموم، يتم إنجاز هذه المهمة من خلال التحكم في الوصول عبر برامج الأمن وتصميم النظام / الشبكات (على سبيل المثال تثبيت جدار ناري، وبرامج لاكتشاف ومكافحة الفيروسات، والرسائل غير المرغوب فيها، وحصان طروادة، والبرمجيات الخبيثة)، والتشفير الذي يتطلب كلمة سر لفتح الملفات المغلقة. وإذا وضعنا في الاعتبار أنه

بمجرد أن يتم الوصول إلى البيانات فسوف يصبح من السهل نسخها ونشرها، فإن إدارة الحقوق الرقمية تسعى لتقييد مثل هذه الممارسات وأن تجعل من السهل أيضاً تتبع تلك الممارسات. ومع ذلك، وعلى الرغم من هذه التهديدات، فإن الأجهزة الرقمية والخدمات والبيانات، ونقاط الضعف المحتملة في تهيئتهم، تزيد بسرعة أكبر من المقدرة على تأمينهم (Gantz and Reinsel 2011). وفي حين أن الكثير من أجهزة الحاسب الشخصية وأجهزة الحاسب المحمولة آمنة بصورة نسبية مادامت برامج الحماية الخاصة بها تُحدَّث بشكل دوري، إلا أن أجهزة الحاسب اللوحية، والهواتف المحمولة، وأجهزة الفاكس، ومحركات الأقراص الصلبة الخارجية، والأجهزة الطرفية (مثل الطابعات، وأجهزة المسح الضوئي، وآلات تصوير المستندات الشبكية)، والعدد الهائل من الأجهزة التي تشكل إنترنت الأشياء عادة ما تتمتع بحماية قليلة (Rezendes and Stephenson 2013; Rose 2013). وقد أدى ذلك بـ (Mims 2013) إلى الادعاء بأنه بمجرد وجود أي شيء في بيئة تشتمل على رمز رقمي ويتم توصيله بالشبكة، فيمكن قرصنة هذا الشيء من خلال كائنات يُحتمل أن يتم التحكم بها عن بعد، أو يتم إجبار ذلك الشيء على تجاوز معايير التصميم، أو العمل بطرق غير مريحة أو خطيرة، أو يضلل المستخدمين مما يؤدي إلى حدوث الأخطاء والأضرار، أو يتجسس على الساكنين أو المستخدمين. وعلاوةً على ذلك، فإن الطرق المستخدمة من قبل القرصنة أو المحتالين أصبحت أكثر تعقيداً وعدوانية عن السابق، وأن صناعة الأمن دوماً في سباق من أجل مواكبة هؤلاء القرصنة والمحتالين (Goldberg 2012).

فمع تطور ثورة البيانات، ووفرة المزيد من الأجهزة التي تنتج، وتبادل، وتستخدم البيانات، يبدو أن القضايا والمخاوف الأمنية مُعرضة للزيادة وليس التقلص (Gantz et al. 2007). وهذا بدوره سيؤدي إلى تفاقم الجرائم مثل سرقة الهوية، وتقويض الثقة في نظم البيانات، ورفع سلسلة من المسائل القانونية المتعلقة بالمسؤولية والالتزامات تجاه حماية النظم عندما يتم العبث بالبيانات، واختلاسها وسرقتها (Weber 2010).

التشخيص، والتصنيف الاجتماعي والخطوط الحمراء (Profiling, Social Sorting and Redlining):

بعيداً عن توسيع محيط المراقبة، وانتهاك الخصوصية، ونقاط الضعف المتعلقة بأمن البيانات، تأتي المخاوف بشأن ماهية المجالات التي يتم فيها توظيف فيضان البيانات لتحقيق أشياء معينة. وكما أشرنا إلى ذلك في الفصل السابع، لم تتغير طبيعة البيانات فحسب، ولكن يوجد أيضاً تحول نوعي في الكيفية التي يتم بها تحليل البيانات، وفي أي الغايات يمكن استخدام البيانات وتحليلات البيانات، ومن لديه السلطة لاستخدام واستخلاص القيمة من البيانات. فالبيانات كانت تستخدم لوقت طويل من أجل تنميط وتقسيم وإدارة السكان، ولكن هذه العمليات أصبحت في الوقت الراهن أكثر تعقيداً، ودقة، وانتشاراً، وانتظاماً عن ذي قبل بكثير. وفي حين أن الهيئات الحكومية تكون ملفات شخصية للأفراد وذلك من أجل أغراض الأمن ورصد الاحتيال، إلا أن معظم النمو في عملية التنميط يتم من قبل الشركات التجارية والتي تسعى إلى فهم واستهداف قاعدة المستهلكين الحالية والمحتملة (انظر الفصل الثاني).

ففي الماضي، كانت الشركات تقوم بالتسويق الجماعي، ونشر رسالة عامة إلى حد ما وسط مجموعة كبيرة جداً من الجمهور وذلك من خلال الإعلان الجماعي باستخدام التلفاز، والراديو، والطباعة، والبريد الإلكتروني، وذلك من أجل الوصول إلى المجموعة المستهدفة لتلك الشركات (Schwartz and Solove 2011). وبمرور الوقت، تم تعزيز هذا الأمر من خلال التسويق المستهدف وذلك باستخدام التنميط الخام بشكل نسبي. وهنا، قامت مجموعة صغيرة من الشركات المتخصصة بتكوين تصنيفات عامة للسكان حيث تم تصنيف الأسر إلى أصناف ذات ملفات شخصية عامة، والتي كانت في العادة يتم تسميتها من خلال اختيار متغيرات ومواقع ديموغرافية، على سبيل المثال باستخدام بيانات الإحصاء السكاني أو البيانات التي تعكس أسلوب الحياة كاستخدام بيانات الاشتراكات في دوريات معينة على سبيل المثال (Goss 1995). وبدلاً من محاولة تنميط عملائهم الشخصيين، ترغب الشركات في شراء الملفات الشخصية وبيانات الاتصال وذلك من أجل تقسيم واستهداف أصناف معينة من السكان من خلال رسالة مؤطرة بدقة، وفي الوقت نفسه تحقيق الفعالية

في ميزانية التسويق من خلال تقليل مصروفات الإعلانات المبددة. وفي الآونة الأخيرة، بدأت شركات التنميط في إنشاء ملفات شخصية فردية بدلاً من الملفات الشخصية العامة وذلك من خلال تجميع البيانات من مصادر عدة مثل معاملات البطاقات الائتمانية ومعاملات بطاقات المتاجر، ومسارات النقر، والمشاركات في وسائل التواصل الاجتماعي، والعديد من أنواع البيانات الشخصية الأخرى (Siegel 2013). وعلاوةً على ذلك، تستفيد الشركات ذاتها من الكميات الهائلة من البيانات التي تنشئها عن العملاء على سبيل المثال، من خلال المبيعات ومسارات النقر وإدارة علاقات العملاء، وتجميع هذه البيانات مع المصادر الأخرى للبيانات وذلك من أجل تكوين الملفات الشخصية الخاصة بهم والاشتراك في التسويق السلوكي.

هنا من الناحية الظاهرية، يبدو أن التنميط التنبئي موقف مربح للجميع، للعملاء والبائعين، فالعملاء يحصلون على معاملة شخصية ومخصصة، في حين يحصل البائعون على مبيعات أكثر ويقللون من النفقات. ولكن وفي المقابل، فإنه يمكن استخدام التنميط التنبئي من أجل تصنيف الناس من الناحية الاجتماعية واستبعاد فئات منهم من خلال اختيار أصناف معينة منهم للحصول على وضع تفضيلي وتهميش واستبعاد الآخرين (Graham 1999; Leysnon and Thrift 2005). فعلى سبيل المثال، يمكن للشركة تنميط الناس فيما يتعلق بمخاطرهم الائتمانية وما هي احتمالية مقدرتهم على الوفاء بسداد مدفوعاتهم، أو فيما يتعلق بقيمة حياتهم المتوقعة إذا ما ظلوا في برنامج الولاء، وما هي احتمالية تغييرهم لعاداتهم (Minelli et al. 2013). ومن خلال استخدام مثل هذه الملفات الشخصية والتنميط، يمكن للشركات تحديد الأولويات لاهتماماتهم ومواردهم، وذلك يتم عادة من خلال التركيز على العملاء ذوي القيمة العالية. ومن ثم يتم قياس وتصنيف العملاء بشكل منتظم ويحصلون على خدمات تفضيلية، وبعضها يكون تمييزياً مثل استبعاد السكان الذين يُعدُّون بلا فائدة (لن يحققوا مكاسب للشركة)، أو لن يحققوا مكاسب كافية، أو أن المخاطر المرتبطة بهم عالية، مما يمكن أن يؤثر بدوره على الفرص الحياتية (Curry 1997; Danna and Gandy 2002; Wyly in press). فهذه الملفات الشخصية تُعتبر منتجات ويمكن بيعها لشركات أخرى تكون مهتمة في تقييم قيمة ومخاطر العملاء المحتملين، ومن ثم فإن النتائج السلبية يمكنها أن تطل الشخص عبر الأماكن وعبر الأزمنة.

ومثل هذه الممارسات التمييزية يمكن أن تشتمل على تسعير فعال وشخصي، فمن الأمور الشائعة بالفعل في سلاسل المتاجر أن تكون السلعة ذاتها مسعرة بطرق مختلفة عبر المتاجر المختلفة وذلك استناداً إلى خصائص وسمات السكان الذين يتسوقون من هذه المتاجر، أو أن يكون سعر المنتجات مختلفاً وفق الكمية المشتراة (مثلاً المنتج الواحد بدولار في حين أن الثلاثة منتجات بدولارين)، أو أن تختلف الأسعار باختلاف الجماعات (مثلاً حصول الطلاب أو المواطنين الكبار في السن على تخفيضات) (Varian 1996). وتتمثل رغبة الكثير من الشركات في تداول مثل هذه الممارسات على أسس فردية، ومعدلة لكي تناسب الملفات الشخصية، وذلك حتى يدفع أناس مختلفون أسعار مختلفة للمنتج ذاته (كما هي الحال مع أسعار تذاكر الطيران ولكن بناءً على نموذج مخصص). وتراوح الأسعار أيضاً بشكل فعال ووفق السياق، بناءً على الظروف. وقد بدأ بالفعل تطبيق مثل هذه الممارسات الشخصية والمرنة، بما في ذلك قطاع المتاجر. فعلى سبيل المثال، تطرق Clifford (2012) إلى أمثلة تبين أن المتسوقين يدفعون في المتجر ذاته أسعار مختلفة للمنتج ذاته وذلك استناداً إلى ملف العميل الشخصي. وفي بعض الحالات، تقوم المتاجر بتعديل الأسعار وقت تسوق العملاء وذلك بناءً على موقعهم داخل المتجر وتاريخ التسوق لهؤلاء العملاء حيث يتم منح القسائم الإلكترونية في الموقع عندما يقوم المتسوقون بمسح المنتج باستخدام التطبيقات في هواتفهم الخلوية. إن الهدف من وراء التسعير الشخصي والفعال هو الحصول على كامل فائض القيمة المتاح في المعاملة وذلك من خلال تسعير البضائع والخدمات لأقرب سعر ممكن من السعر الاحتياطي (سعر التحفظ) لدى الفرد (Tene and Polonetsky 2012: 17)، وبعبارة أخرى، فإن هذا مصمم للاستفادة من الإنفاق المثالي لمصلحة المتجر.

وبالنظر إلى قدرتهم على تشكيل وتصنيف خبرات المستهلكين والفرص الحياتية، فإن الملفات الشخصية للأفراد والبيانات وممارسات البيانات التي تدعمها بحاجة إلى اهتمام أكبر. وفي الوقت الحالي، فإن كلاً من البيانات المستخدمة والتحليلات المطبقة مبهمة وغامضة بالنسبة للمستهلكين والمواطنين، فإذا انتهى الأمر بالشخص مدرجاً على قائمة الممنوعين، أو تم معاملته بشكل سلبي، أو تم استبعاده من منتج ما من قبل شركة ما، على سبيل المثال تم وضعه في القائمة السوداء للحصول على الائتمان، فمن الصعوبة بمكان تحديد الأسس لهذا التمييز ومجابهة هذه الأسباب. فالنقاد يدعون بأننا قد ولجنا عصرًا حيث:

رهباً يتراءى للبعض أن لبنية البيانات أو الصورة التخيلية للبيانات ميزة وجودية، ولكنها بالفعل ذات ميزة وجودية. إن ما تقوله بنية بياناتك عنك أكثر حقيقة وصدقاً مما تقوله أنت عن نفسك، فبنية البيانات هي البنية التي تم الحكم عليك من خلالها في المجتمع، وهي البنية التي تملي وتفرض وضعك في العالم، فما نشاهده عند هذه الفترة من الزمن هو غلبة التمثيل على الكينونة (Critical Art Ensemble 1995).

وعلى الرغم من ذلك فنحن نعرف أن البيانات الكبيرة وبنية البيانات فوضويان - فهما ممتلئان بالثغرات، والمتناقضات والأخطاء (انظر الفصل التاسع) - ونعرف أيضاً أن البرامج والخوارزميات تعمل بطرق انتقائية ومتحيزة (Kitchin and Dodge 2011). ومن ثم، فإن الأنظمة التي تعتمد على البيانات الكبيرة، والتي لا تُعد شفافة وتفتقد إلى التدقيق اللازم وحق الرجوع الملائم، تطرح جميع الأنواع المحتملة من المشاكل وذلك عندما يمنحون الأفضلية لبنية البيانات. وفي الواقع، فإن الوقت أصبح ملائماً الآن أكثر من أي وقت مضى للرقابة القانونية والتنظيمية على التنميط التنبئي، وعلى الإجراءات التي تشوش على التمييز الوجودي للبيانات وتحمي الناس من الممارسات التمييزية والضارة.

الاستخدامات الثانوية والتحكم في الانسلاخ والحوكمة الاستباقية (Secondary Uses, Control Creep and Anticipatory Governance).

إن أحد الأسس التي تستند إليها سياسة الخصوصية وحماية البيانات في الاتحاد الأوروبي وأمريكا الشمالية هو مفهوم تقليص البيانات. وهذا المفهوم يشترط أن تقوم الهيئات والباحثون بإيجاد البيانات الضرورية فحسب لأداء مهمة ما، بمعنى أن يتم الاحتفاظ بالبيانات طالما كان ذلك ضرورياً لأداء هذه المهمة أو كما تمليه الاعتبارات القانونية، وأن البيانات التي يتم تكوينها يتعين استخدامها فحسب من أجل هذه المهمة (Tene and Polonetsky 2012). وبمعنى آخر، لا يتعين على جامعي البيانات أن يستولوا على أي شيء يمكنهم الاستيلاء عليه، ولا يتعين عليهم كذلك تخزين هذه البيانات لمدة غير محدودة أو أن يتيحوا هذه البيانات لاستخدامات ثانوية. ومن الواضح أن هذه التطلعات والطموحات تتعارض مع المبادئ والأسس المنطقية للبيانات الكبيرة وعمل أسواق البيانات التي تسعى إلى تكديس البيانات في حال كان من الممكن أن تكتسب هذه البيانات أي قيمة في المستقبل،

(Tene and Polonetsky 2012; Andrejevic 2013). فالحل أمام البائعين تمثل في إعادة تقديم البيانات من خلال إزالة الأجزاء التي تشير لهوية الأفراد من هذه البيانات أو إنشاء بيانات مشتقة، مع إخضاع مجموعة البيانات الأساسية فقط لعملية تقليص البيانات. ويمكن في هذه الحالة بيع البيانات المعاد تقديمها أو استخدامها بعدة طرق، ومثل هذا الأسلوب لإعادة تحديد الغرض من البيانات أمر شائع، حيث يجادل البائعون أنه لا توجد حاجة لإبلاغ الشخص الذي تشير إليه البيانات أو الحصول على موافقته، حيث إن هذه البيانات مجهولة المصدر، ومشتقة ومجمعة (Solove 2007). وهذا الحصاد المحتمل للكميات الهائلة من البيانات، والذي تم الحصول على معظمه دون معرفة أو فهم الأشخاص، والذي يتم استخدامه بعد ذلك في استخدامات ثانوية، يثير الكثير من الأسئلة الأخلاقية التي لا تتعلق بالخصوصية وحماية البيانات فحسب، بل تتعلق كذلك بالحوكمة، ومن الأمثلة الواضحة والمتعلقة بالحوكمة هو التحكم في الانسلا.

إن التحكم في الانسلا يحدث عندما تكون البيانات التي تم إنشاؤها لشكل من أشكال الحوكمة مشابه لشكل آخر (Innes 2001). وحدث هذا غالباً بشكل شديد الوضوح فيما يتعلق بالأمن وخصوصاً بعد أحداث الحادي عشر من سبتمبر، حيث تم إعادة توظيف البيانات الخاصة بخطوط الطيران والبيانات الإدارية الحكومية من أجل تنميط وتقييم مخاطر الأمن للركاب (Lyon 2003b). وعلى نحو مشابه، فلقد تم إعادة توظيف كاميرات الإشارات المرورية ورسوم الازدحام في لندن من أجل القيام بمهام تتعلق بالأمن، وذلك بدلاً من مراقبة الانتهاكات المرورية فحسب (Dodge and Kitchin 2007a). ومن الأمثلة التجارية على التحكم في الانسلا هو نظام الملاحاة في السيارات المستأجرة حيث يتم إعادة توظيف هذا النظام من مساعدة السائقين على إيجاد طريقهم فحسب إلى مراقبة وفرض غرامة على هؤلاء الذين يقودون خارج المسار أو يخرجون من الولاية (Elliott 2004). فالتحكم في الانسلا يقوض بشكل منهجي الأسباب المنطقية وراء تقليص البيانات، في حين أن نشره يشكل تهديدات واضحة للحريات المدنية، حيث يكون جميع المواطنين، سواء الأبرياء أو المذنبين، خاضعين لمراقبته وإجراءاته التأديبية.

وهناك مثال آخر على الاستخدام الثانوي للبيانات والذي ينتهك الحريات المدنية ألا وهو الحوكمة الاستباقية (التوقعية). وهنا، يتم استخدام التحليلات التنبؤية من أجل تقييم السلوكيات أو الأحداث المستقبلية المحتملة والتوجيه باتخاذ الإجراء المناسب. فمثل هذه الحوكمة الاستباقية تُعد سمة من سمات السفر بالجو خلال عدد من السنوات، حيث يتم تنميط المسافرين من أجل المخاطر والتحقق الأمني وذلك قبل البدء في الرحلة (Dodge and Kitchin 2004). وبالمثل، تستخدم العديد من قوات الشرطة في الولايات المتحدة الأمريكية التحليلات التنبؤية من أجل توقع موقع الجرائم المستقبلية أو من أجل توجيه ضباط الشرطة لزيادة دورياتهم في هذه المناطق (Siegel 2013). وعلى النحو ذاته، تستخدم بعض الشركات مثل هيوليت باكارد (Hewlett Packard) التحليلات التنبؤية من أجل تقييم ومعرفة مَنْ مِنَ الموظفين من المحتمل أن يغادر الشركة، ومن ثم التدخل بشكل استباقي من أجل تقليل نفقات استبدال الموظفين (Siegel 2013). وفي مثل هذه الحالات، فإن آثار البيانات الخاصة بالشخص تقوم بما هو أكثر من تتبع الأشخاص، إذ تقوم باستباق هؤلاء الأشخاص (Stalder 2002).

إن مثل هذه الأساليب من الحوكمة تفرض قضايا أخلاقية بسبب أنها توجه الاهتمام إلى مجموعات وأماكن معينة، وهي كذلك تسعى لمراقبة السلوكيات التي ربما لن تحدث على الإطلاق، وخلال هذه العملية، كما تقوم بإعادة تشكيل كيف يمكن للناس العمل من خلال الانضباط الذاتي (Harcourt 2006). وعلاوةً على ذلك، فإن مثل هذه التحليلات التنبؤية تعمل في غالب الأحيان على تفاقم التحامل والتمييز، وتعمل كذلك على تأسيس النبوءات المحققة لذاتها (التوجه نحو إشباع الذات) (Harcourt 2006; Siegel 2013). ومن ثم فلقد ادعى (Raley 2013: 128): "إن البيانات في هذا المجال تكون أدائية: فتجميع أجزاء وبتات (bits) من البيانات في ملفات شخصية لمشتبه به في الإرهاب، وإعادة تأسيس البيانات المجردة عند استهداف حياة واقعية سوف يكون لهما تأثير على إنتاج هذه الحياة، هذا الشخص، كمشتبه به بالإرهاب بالفعل". ومن ثم يتم وضع الفرد تحت بؤرة المراقبة وممارسات مراقبة الدولة، وجعله خاضعاً لنتائج وتداعيات كونه مشتبهاً به بالإرهاب، على الرغم من عدم التيقن من هذا الاشتباه. وبعبارة أخرى، تتم معاملة الشخص بطريقة مختلفة من جراء التوقع بأنه ربما يكون شيئاً ما، في حين أنه ربما يُكن هذا الشيء أو

لا يمكنه أو ربما يرتكب هذا الفعل أو لا يرتكبه. وفي حين أن النوايا وراء هذا العمل ربما تكون نوايا عظيمة وحسنة، إلا أن التوقع في هذه الحالة له تداعيات تتجاوز منع الأحداث المتوقعة (Harcourt 2006). وكما هو الحال مع الأشكال الأخرى من التنميط، فإن الحوكمة التنبؤية يتم توجيهها بشكل كبير من خلال خوارزميات الصندوق الأسود التي تعمل على بيانات ذات أصل غير معروف، وهي على وجه العموم منغلقة على حق الرجوع (النقد). فالحوكمة التنبؤية كانت مثار القليل للغاية من الاهتمام النقدي أو المناقشات وذلك بما يتعلق بتأثيراتها، وعلى الرغم من ذلك تم تطويرها بفعالية من أجل استخدامها في الحكومات والشركات.

أساليب الحوكمة والضوابط التقنية (Modes of Governance and Technological Lock-Ins):

بعيداً عن التحكم في الانسلاخ والتنميط التنبؤي، فإن ثورة البيانات لها تأثيرات محتملة فيما يتعلق بتنظيم وتشغيل الحوكمة. وإذا وضعنا في الاعتبار الشبكة المتسعة لجمع البيانات، وقراراتها الدقيقة، فإن قلق الكثير من المعلقين ينبع من أن عصر ظهور البيانات الكبيرة يسبق عصر ظهور ما أصطلح على تسميته وصاية الأخ الأكبر (Big Brother). وهذا يعني أننا في عصر أضحت فيه الدولة والمؤسسات تعرف وتتوقع الكثير جداً عن الأفراد من خلال مراقبة البيانات والتنميط التنبؤي لدرجة أنهم يمتلكون السلطة لفرض أشكال قاسية وضارة من التحكم الانضباطي.

وكما أشرنا إلى ذلك خلال هذا الفصل، فنحن نعيش وبدون شك في عصر المراقبة الزائدة بشكل مبالغ فيه، حيث أصبحت المراقبة أكثر اتصالاً من أي وقت مضى، وهي مدعومة من خلال تزايد ثقافة السيطرة التي ترغب في الأمن، والاستقرار وتقليل المخاطر (Lyon 2007). وهنا، يتم تشريع نوع جديد من الحوكمة، وهو نوع يؤول فيه تنظيم جوانب معينة من الحياة اليومية إلى الأنظمة التقنية المملوءة بالتوكيل الثانوي (Kitchin and Dodge 2011). فعلى سبيل المثال، فإن معرفة، ومعالجة وإدارة الانتهاكات المرورية تتم بشكل متزايد من خلال الأنظمة الآلية، والمسموح بها من خلال برنامج يقوم بمعالجة البيانات التي تتعلق بأرقام اللوحات، والسرعة، وحق الوصول، وإحالة هذه البيانات إلى قواعد بيانات

الملكية من أجل فرض غرامات أو خصم نقاط بشكل آلي (Dodge and Kitchin 2007a). وعلى النقيض من الأشكال التقليدية من المراقبة التي تسعى إلى تشجيع الانضباط الذاتي، فإن الأنظمة الآلية تقوم بإعادة تشكيل السلوك بشكل فاعل (Agre 1994; Kitchin and Dodge 2011). وفي مثل هذه الأنظمة، فإن آليات إنشاء البيانات وتنظيمها تعتبر مجالاً مهماً للغاية من النظام، وذلك بدلاً من كونها جزءاً خارجاً عن النظام. فعلى سبيل المثال، بدلاً من أن يقوم عمال صناديق المحاسبة بإجراء الانضباط الذاتي لمعدل عملهم لأنهم ربما أو ربما لا يعرفون أنه تتم مراقبتهم من خلال كاميرات المراقبة التلفزيونية ذات الدائرة المغلقة، فإن عمل مسح البضائع ومعالجة المدفوعات هي الوسائل التي يتم من خلالها مراقبة معدل العمل بشكل مستمر. وفي مثل هذه الأنظمة/ فإن المراقبة وآثار البيانات المرتبطة بهذه المراقبة أصبحت مستمرة، ومتغلغلة، ومواظبة، ومتفاعلة، مع سلوك الخاضع لها، ولكنها تقع خارج سيطرته (Cohen 2012).

إن البيانات التي تكونت بشكل آلي لا يستخدم جميعها من أجل تشريع الإدارة الآلية، ولكن هذا هو الاتجاه السائد بشكل متزايد، وخاصةً للأنظمة عالية التنظيم. وعلاوة على ذلك، فإن الأساليب الجديدة من التحكم، إلى الآن، لا يتم السماح بها بالطريقة الصارمة التي توقعتها رؤية جورج أورويل (George Orwell) في روايته "الأخ الأكبر"، فهذه الأساليب عادية وروتينية ومغرية وتشجع المواطنين على أن يشاركوا بفعالية وبتلقائية في ممارساتها (Kitchin and Dodge 2011; Cohen 2012). ونتيجةً لذلك، فإنه وكما ذكر (Solovo 2007)، فإن البيانات الكبيرة لا تتعلق بشكل كبير بمباشرة رؤية جورج أورويل ولكنها ترتبط بشكل أكبر برؤية كافكا (Kafka) كونها تطرح قضايا عن الإجراءات الواجب اتخاذها فيما يتعلق باللامبالاة، الأخطاء، الانتهاكات، الإحباط، نقص الشفافية، والاستعانة أكثر مما تطرحه عن التسلط والاستبداد. وعلى النقيض من ذلك، فبالنسبة لما قاله (Boellstroff 2013)، فإن المجاز الأساسي في البيانات الكبيرة والمراقبة ليس الرؤية الكلية، ولكن بالأحرى هو التسليم، فبدلاً من أن يكون الفرد خاضعاً على نحو صرف لمراقبة انتهاكية، يُسلم الناس بحماس ببياناتهم من خلال وسائل التواصل الاجتماعي، أو من خلال الانضمام في برامج بطاقات الولاء... إلخ. وبغض النظر عن الصياغة التصورية للعملية، فمن الواضح أن هذه الأنظمة الجديدة لها تأثيرات فيما يتعلق بالسلطة والسيطرة، وتقوم

بتشكيل التجارب والفرص الحياتية بطريقة مختلفة، وأن دور البيانات ومراقبة البيانات في مثل هذه الأنظمة يتطلب اهتماماً نقدياً إضافياً.

إن من الطرق التي يتم من خلالها تحويل الحوكمة من قبل التقنيات التي تدعمها البيانات أن نجعل الحوكمة أكثر تكنوقراطية بطبيعتها (معتمدة على التقنية بشكل كلي). فعلى سبيل المثال، فإن النزعة تجاه إدارة وتنظيم المدن من خلال أنظمة المعلومات والأنظمة التحليلية تعزز الأسلوب التكنوقراطي للحكومة المدنية التي تفترض أنه من الممكن قياس ومراقبة جميع المجالات المدنية ومعاملة هذه المجالات على أنها مشاكل فنية يمكن التعامل معها ومجابهتها من خلال الحلول الفنية (Kitchin 2014)، وهي توضح ما أُطلق عليه (Mattern 2013) مصطلح "العقلانية الأدائية"، وما أُطلق عليه (Morozov 2013) "مقاربة الحلول التقنية" حيث يمكن تجزئة المواقف الاجتماعية المعقدة إلى مشاكل محددة بشكل دقيق يمكن حلها أو البحث عن أفضل الحلول لها من خلال الحوسبة. ومرةً أخرى، فإن هذا يعد تجسيداً للبيانات الكبيرة، حيث يمكنها أن توفر إجابات لجميع المشاكل (Mattern 2013). ومن خلال التقاط وتمثيل الظواهر كبيانات بشكل آلي أو لحظي، يصبح من الممكن من الناحية الظاهرية تخطيط، وفهم، وإدارة، وإصلاح الموقف أو المشكلة حال حدوثها. وقد أوضح (Hill 2013)، أن "تفكير المدينة الذكية" ينم عن وجهة نظر تكنوقراطية بأن المدينة هي شيء يتعين علينا فهمه بالتفصيل كالمحرك أو محطة الطاقة النووية وذلك في حال كان لدينا بيانات كافية فقط. وأنه تبعاً لذلك يمكننا السيطرة على هذه المدينة من خلال القوة الغاشمة للعلم والهندسة. وفي الواقع، فقد اقترح (Matten 2013) أن تحضر البيانات الكبيرة يعاني من تراكم البيانات، وهو افتراض أن جميع التدفقات والأنشطة الهادفة يمكن الإحساس بها وقياسها. وداخل مثل هذا التفكير هناك افتراض واضح في الغالب وهو أن الكون تشكل من قبل المعايير القابلة للمعرفة والقابلة للتحديد والتي تؤكد لنا أنه لو كان بإمكاننا قياس جميع هذه المعايير، فسوف نكون قادرين على التنبؤ والاستجابة بتمكن تام وفقاً لذلك (Haque 2012). وعلى الرغم من ذلك، لن يكون بمقدور الحلول التقنية بذاتها حل المشاكل الهيكلية المتأصلة في المدن لأن هذه الحلول لا تتعامل مع أسباب تأصل تلك المشاكل. ولكن بدلاً من ذلك، فإن هذه الحلول تمكن فقط من إدارة مظاهر تلك المشاكل بشكل أكثر فعالية. وفي حين أن تقنيات المدن الذكية مثل

أساليب التحليل اللحظية والآنية يتم الدفع بها كوصفة للتعامل مع جميع قضايا الحوكمة المدنية، إلا أن هذه التقنيات في الغالب الأعم تداري الشقوق بدلاً من إصلاحها، إلا إذا واكب هذه التقنيات مجموعة أخرى من السياسات (Kitchin 2014). ويمكن أن ينطبق الأمر نفسه على الطرق التكنوقراطية المطبقة في بعض المجالات الأخرى، فأساليب الحوكمة التي تفوقها البيانات تضمن بالفعل تنفيذ إستراتيجيات مدعومة بالأسانيد ولكن يجب أن تكون مقترنة بالسياق، ومرنة، ومستهدفة بشكل ملائم، ولا يتم إدارتها ببساطة من خلال الأرقام والخوارزميات.

ويتزامن مع ذلك الانتقاد القائل أن الحوكمة أصبحت مبنية على معرفة تامة وتكنوقراطية مفرطة ومع الخوف من أن يتم الاستيلاء عليها وتشكيلها بشكل صريح من قبل مصالح الشركات من أجل تحقيق مكاسب شخصية. وبالعودة إلى المثال المتعلق ببرنامج المدينة الذكية ومثال الحوكمة البلدية، وكما أشرنا إلى ذلك في الفصل السابع، فإن عدداً من أكبر شركات خدمات البرامج والأجهزة يسعى بشكل فعال إلى التعاقد من الباطن على مجالات من الإدارة المدنية وأن تجعل من بضاعتهم جزءاً لا يتجزأ وذات دور جوهري في كيفية مراقبة وتنظيم العديد من مجالات الحياة في المدينة (Kitchin 2014; Townsend 2013). إن الهاجس والقلق النابع من خصخصة الإدارة المدنية ذو ثلاثة أوجه (Kitchin 2014):

- أولاً: أنه يعزز بشكل فعال الاقتصاد السياسي لليبرالية الجديدة والتحول نحو اقتصاد السوق في القطاع العام حيث يتم إدارة وظائف المدينة من أجل تحقيق الربح الخاص (Hollands 2008).
- ثانياً: أنه يؤسس للقيد التقني الذي يجعل المدن مدينة وممنونة بالفضل لمنصات ومزودي خدمات تقنية معينة خلال مدة زمنية طويلة، مما يخلق مواقف احتكارية (Hill 2013). والخطر في هذه الحالة يكمن في تأسيس تبعية للمسار المؤسسي مما لا يمكن إلغاؤه أو التحول عنه بسهولة (Bates 2012).
- ثالثاً: أنه يؤدي إلى حلول مغلقة أساسها "نظام واحد مناسب لجميع المدن الذكية" والتي لا تضع في اعتبارها إلا بشكل ضئيل للغاية خصوصية وتفرد الأماكن، والناس، والثقافات، وتقيد إدارة المدينة على أسلوب تكنوقراطي ذي رؤية ضيقة للحوكمة (Townsend 2013).

وإذا ما أستحضرنا الواقع، فإن شركة IBM في الوقت الحالي منتجاً يسمى "مركز عمليات IBM الذي" الذي يجمع العديد من الأنظمة التي تم تصميمها لمدينة ريو دي جانيرو في منتج واحد يمكن تطبيقه على أي مدينة (Singer 2012c). انظر الفصل السابع). ولا يعني هذا أن نقول إن القيود التي تفرضها الشركات أمر حتمي، ولكن من الواضح أن ذلك يُمثل رغبة عدد من أكبر الشركات العاملة في هذا المجال. وبالمثل، يتم استهداف أشكال أخرى من الحوكمة بشكل متزايد من خلال الشركات التي تسعى إلى إيجاد أسواق للحلول التقنية. ومرة أخرى، فإن تداعيات مثل هذه الأشكال من الحوكمة تتطلب الكثير من التدقيق.

الخاتمة:

ناقش هذا الفصل التداعيات الأخلاقية، والاجتماعية، والسياسية، والقانونية للتغيرات التي تحدث من جراء تكوين البيانات ودمج البيانات واستخدام البيانات. ومن الواضح أنه في حين أن مثل هذه الممارسات لها فوائد بالنسبة للحكومات والشركات والمواطنين، إلا أنها في الوقت نفسه لها تداعيات سلبية وتمييزية. وإذا وضعنا في الاعتبار السرعة التي يتغير بها مشهد البيانات، فإن مواكبة التطورات وتحديد الآثار المحتملة لهذه التطورات، ودراسة ردود الفعل الاجتماعية والقانونية المناسبة يعتبر تحدياً هائلاً. وفي الواقع، هناك الكثير من الأسئلة الأساسية المعيارية التي تتطلب دراسة تأملية والتي تتعلق بمن يمكنه تكوين مجموعات البيانات والوصول إليها ومشاركتها وتحليلها، ولأي غرض وفي أي سياق وتحت أي معوقات أو قيود (boyd and Crawford 2012: 673). إن الحاجة إلى إجابات لمثل هذه الأسئلة أمر ملح، وذلك إذا وضعنا في الاعتبار المدى والتأثيرات المترتبة على مراقبة البيانات وانتهاك الخصوصية والأمن غير المحكم للبيانات، والتأثيرات الضارة للفرز الاجتماعي والاستبعاد، والطبيعة والأساليب المتغيرة للحوكمة. فحتى لو كانت الاستجابة سريعة ومبالغ فيها، فإن الإجابات المترتبة ستكون متنوعة، حيث إن العديد من أصحاب المصلحة يسعون وراء إستراتيجيات بديلة من أجل التعامل مع المخاوف العديدة، بحسب ما يتضح من خلال المناهج المقترحة لتفتت وانتهاكات الخصوصية. ونتيجة لذلك، يمكن للمرء أن يتوقع الكثير من المناقشات في السنوات القادمة فيما يتعلق بإدارة العناصر العديدة في ثورة البيانات. وعلى وجه الخصوص، وكما ناقشنا في الفصل الثاني، فإنه من المحتمل أن

يوضع عمل سماسة البيانات تحت التدقيق المتزايد سواء من جانب الحكومة أو المواطنين. ومن منظور أكاديمي، من المهم للغاية أن يضطلع العلماء بدور فاعل في استقصاء ودراسة القضايا القانونية والسياسية والاجتماعية والأخلاقية التي تنشأ نتيجة لثورة البيانات، مع التتبع والإسهام في المناقشات العديدة المستمرة. يحتاج مثل هذا العمل إلى أن يشتمل من جانب على دراسات حالة تجريبية مفصلة توثق التأثيرات المادية والاستطردادية الناتجة عن كيفية توظيف البيانات، ومن جانب آخر على البحوث والرسائل المعيارية والشاملة التي تحدد تأثيرات تجميع العديد من البيانات والطرق البديلة الممكنة. وحتى هذه اللحظة، فإن معظم الإسهامات كهذا الفصل والأعمال التي يناقشها تكونت من الجانب الأخير مع وجود القليل جداً من الدراسات التجريبية على بعض الأنظمة الاجتماعية الفنية المحددة ونتائج هذه الدراسات. إن مثل هذه الدراسات مهمة للغاية لأنها تكشف الطرق الملموسة التي تم من خلالها تقديم مبررات لتجميع البيانات، وكذلك المهام التي يتم ممارستها وكيف يتم قبولها، وإعادة تنقيحها، ومقاومتها، وإفسادها، وانتهاكها من قبل المجتمعات، وهذه الدراسات توفر أيضاً قاعدة أدلة يتم من خلالها تصور كيف يمكن إعادة تأطير مثل هذه الأنظمة أو إعادة تحديد الغرض منها.

الفصل الحادي عشر: فهم ثورة البيانات

(Making Sense of the Data Revolution)

ناقش هذا الكتاب قضية أنه يتعين أن تكون هناك حاجة إلى تعاطٍ أكثر نقدياً وفلسفياً مع البيانات وأن ثورة البيانات ما زالت قائمة في الوقت الراهن. وفيما يتعلق بالأمر الأول، ادعى البعض أن البيانات تُمثل عناصر أساسية لإنتاج المعرفة، ومع ذلك، فحتى الوقت الراهن لم تحظ الصياغة الوجودية للبيانات والمجاميع المحيطة بإنتاج البيانات واستخدامها بالاهتمام الكافي. وبدلاً من ذلك، فإن التركيز كان منهجياً بصورة أكبر فيما يتعلق بتكوين البيانات، أو أن التركيز كان على الأشكال الاشتقاقية للمعلومات والمعارف. وفيما يتعلق بالأمر الثاني، ادعى البعض أن تدرج البيانات الصغيرة في البنى التحتية للبيانات وإتاحة الوصول للبيانات التي كان الوصول إليها محدوداً في السابق، وظاهرة البيانات الكبيرة تدمر الابتكارات ولا تدعمها أو تحافظ عليها، فهي تغير طبيعة البيانات من حيث حجم هذه البيانات، وسرعتها، وشموليتها، وعلائقتها ودلائيتها، ومرونتها وتوسعها، وتنوعها، وانفتاحها وتبادليتها، وهي السبب في ظهور مجاميع بيانات جديدة وطرق جديدة لفهم العالم الذي نعيش فيه. إن التأثيرات التدميرية لثورة البيانات لها تداعيات هائلة على الحكومات والشركات والمجتمع المدني وذلك من خلال إنشائها لمعارف وممارسات جديدة يمكن استخدامها من أجل إعادة صياغة كيف يمكن أن يتم حكم الناس، وكيف يمكن أن تتم إدارة المنظمات، وكيف يمكن أن تتم زيادة القيمة والفائدة، وكيف يمكن أن يتم تكوين رأس المال، وكيف يمكن أن يتم تحسين الأماكن، وكيف يمكن أن تتم إعادة تخطيط وممارسة العلم. وهي أيضاً أثارت العديد من التحديات وطرحت الكثير من الأسئلة، ومكنت لوسائل أكثر شيوعاً فيما يتعلق بمراقبة البيانات، وتقويض الخصوصية، وتنميط الأشخاص والأماكن والفرز الاجتماعي، ومهدت لسن أشكال جديدة من الحوكمة. فهذا الفصل الأخير يحدد خطة طريق دلالية من أجل إدراك وفهم البيانات وثورة البيانات وذلك بالنظر إلى الفجوات الحالية في الفكر المفاهيمي والمعرفة. ويشير هذا الفصل إلى أن مثل هذا الفهم بحاجة إلى أن يحدث بطريقتين: الأولى، من خلال التفكير الفلسفي الشامل والتحليل النقدي

والمفاهيمي. والثانية، من خلال البحوث التجريبية المفصلة التي تتعلق بأصل، وتشكيل، ووظيفية وتطور مجاميع البيانات.

فهم البيانات وثورة البيانات (Understanding Data and the Data Revolution):

إذا وضعنا في الاعتبار فائدة وقيمة البيانات، فهناك حاجة ملحة للنظر إليهم من وجهة النظر الفلسفية والمفاهيمية. فالفلسفة، كما ادعى Wittgenstein (1921)، هي التوضيح المنطقي للفكر، فكل مدرسة من مدارس الفكر تُعدُّ نظاماً للأفكار ومنظوراً لفهم العالم. وبمصطلح عام، تتشكل كل مدرسة وتختلف عن المدارس الأخرى من خلال أربعة مكونات - الوجودية (Ontology)، والمعرفية (Epistemology)، والأيدولوجية (Ideology)، والمنهجية (Methodology) - وهذه المكونات هي التي تحدد معاييرها. فالوجودية هي مجموعة من المعتقدات التي تتعلق بطبيعة الوجود، وهي تهتم بما هو موجود وما يمكن ملاحظته ومعرفته. وبالنسبة للبيانات، فالوجودية تتعلق بماهية البيانات. ما الطبيعة الأساسية للبيانات؟ أما المعرفية فهي تتعلق بكيفية استخلاص المعرفة أو الوصول إليها، وهي الافتراضات عن كيف يمكننا معرفة العالم بمعنى كيف أصبح بإمكاننا معرفة البيانات؟ وهي أيضاً تتعلق بكيفية حشد البيانات في ممارسة التعلم. أما الأيدولوجية فهي تتعلق بالسياسات الضمنية والغرض من إنتاج المعرفة، والمدى الذي يسعى من خلاله المنهج لكي يكون محايداً بمعنى أن يكون موضوعياً وغير تقويي، أو المدى الذي يسعى من خلاله المنهج بفعالية للتدخل وتغيير العالم. وحيث ذلك، ينطوي تحت مفهوم الأيدولوجية العديد من الأسئلة مثل هل البيانات سابقة للحقيقة؟ وهل هي سابقة للتحليل وبلاغية؟ أو هل البيانات اجتماعية أم سياسية؟ وهل تمت صياغة هذه البيانات بفعالية؟ أما المنهجية فهي مجموعة الإجراءات المستخدمة من أجل تفعيل سؤال أو نظرية ضمن الافتراضات الوجودية أو المعرفية للفلسفة الإجمالية للبحث. وهي تشكل اختيار الأساليب التي يتم من خلالها تكوين وتحليل البيانات، فالمنهجية والأساليب مرتبطان بشكل وثيق، ولكن ما طبيعة هذه العلاقة بينهما؟

لقد شرع الفصل الأول في تقديم إجابات أولية عن هذه الأسئلة وقدم أيضاً صياغة لفهم البيانات من منظور فلسفي مكن من فصل هذه البيانات عن كونها مجرد "مواد خام"

للمعلومات والمعارف. فالبعض ادعى أن البيانات اجتماعية ومادية، وأن البيانات لا تمثل العالم فحسب، ولكنها تنتج هذا العالم بشكل فعال. وادعى البعض أيضاً أن البيانات لا توجد بشكل مستقل عن الأفكار، والأساليب، والتقنيات، والناس، والسياقات التي تنتج، وتعالج، وتدير، وتحلل، وتخزن هذه البيانات. وفي الواقع، يتم تنظيم البيانات وتخزينها في قواعد البيانات والبنى التحتية للبيانات التي تشكل لب المجاميع الاجتماعية الفنية المعقدة. وكما تم الإشارة إليه بالتفصيل في الفصل الأول، تشتمل هذه الأنظمة على أنظمة الفكر، وأشكال المعرفة، والتمويل، والاقتصاد السياسي، وأنماط الحاكمية، والتشريعات، والماديات، والبنى التحتية، والممارسات، والمنظمات والمؤسسات، والذاتيات والمجتمعات، والأماكن والأسواق، وكل نظام من هذه الأنظمة يشتمل على عناصر عديدة (انظر الجدول ٣ - ١). فهذه العناصر تعمل سويًا بطرق متعددة، واتفاقية وارتباطية وذلك من أجل الإنتاج الاستطراذي والمادي لمجاميع البيانات (انظر الشكل ٣ - ١). لقد تم عبر هذا الكتاب مناقشة العديد من هذه المجاميع مثل البنى التحتية البحثية، والمحفوظات الوطنية، والبنى التحتية السيرانية، ومشروعات البيانات المفتوحة، وسماسرة البيانات، وأنظمة البيانات الحكومية، وأنظمة الأعمال. وكل واحدة من هذه المجاميع، بالرغم من أنها تتشارك في قواسم مشتركة، إلا أنها تتطور وتعمل بطرق مختلفة، وهي تتكون من مجموعة من الأنظمة والعناصر التي تتوسع بشكل مختلف (على سبيل المثال من المنظمات، والماديات المحلية، إلى الفرق المنتشرة، والقوانين الوطنية، وفوق الوطنية إلى الأسواق العالمية) ولكنها على الرغم من ذلك مرتبطة داخل كوكبة فريدة.

إن مثل هذه الصياغة تمكننا من جهة من التفكير في إنتاج البيانات وكيف يمكن صياغة هذه البيانات بأشكال متعددة من النواحي الفنية، والسياسية، والاقتصادية، والأخلاقية، والمكانية، والزمانية، وصياغة استخدامات هذه البيانات والعمل الذي تقوم به في العالم، وتمكننا من الجهة الأخرى، من توجيه البحوث التجريبية التي تتعلق بمجاميع البيانات (انظر الجزء التالي). وهي، على الرغم من ذلك، مجرد مجموعة واحدة من الطرق المحتملة لفهم البيانات، ومما لا شك فيه أنه مرور الوقت سيتمكن العلماء من إنتاج مجموعة من المناظير التي يمكننا من خلالها فهم البيانات. إن تنوع وجهات النظر سوف يؤسس لعناصر تكاملية مثمرة للأفكار الجديدة ووجهات النظر المفاهيمية من أجل توجيه البحوث

التجريبية. وربما تشتمل هذه على تنظير البيانات من خلال منظار تأسيسي (بنوي) أكبر يركز على دور هذه البيانات في تشغيل الرأسمالية المتأخرة، أو يستفيد من مفاهيم ما بعد البنيوية لدولوز (أحد أهم الفلاسفة الفرنسيين في النصف الثاني من القرن العشرين) التي تتعلق بأساليب الحوكمة الجذمورية (الجدمور: مفاهيم تتوزع على شبكة، بحيث تلتقي مع بعضها دون اشتراط اللزوم)، أو الانتقادات النسائية ما بعد الاستعمار للنوع الاجتماعي والإنتاج المُسيّس وتوظيف البيانات. وبغض النظر عن المنظور، فالمطلوب هو تفكير عميق، وحرص ونقدي مع تنظير العمل من خلال دراسات الحالة التجريبية.

إننا لسنا بحاجة إلى استكشاف الصياغات المفاهيمية للبيانات فحسب، ولكننا بحاجة، كما ناقشنا في الفصل الثامن، إلى دراسة كيف تطرح ثورة البيانات بعض التحديات أمام الفلسفات الحالية للعلوم. فالبيانات الكبيرة وتحليلات البيانات توفر، على وجه الخصوص، إمكانية ظهور نموذج علمي جديد، أي نهج معرفي جديد يتميز بكثافة البيانات ويتسم بالاستكشافية بطبيعته وذلك من أجل تطوير النظرية العلمية. فالمحاولات الأولية للتفكير فيما يمكن أن يعنيه مثل هذا التحول النموذجي كانت في غالب الأحيان مرتبكة ومضللة، وتُعزز شكلاً من أشكال التجريبية التي يجدها الكثير من العلماء إبعادية وخطوة إلى الخلف وليست خطوة إلى الأمام. ويتعين استبدال هذا بمنظور أكثر فعالية لما يمكن أن يتكون منه العلم المساق بالبيانات من حيث معتقداته الفلسفية والمنهجية. لقد بدأ مثل هذا التفكير بالكاد، ومع ذلك فهناك حاجة ملحة إليه من أجل مواكبة إيقاع التغيير التقني ونشر المنهجيات المخصصة والبراجماتية (الواقعية)، وكذلك من أجل استبدال الأشكال التجريبية الضعيفة المنتشرة. فالتحول النموذجي أقل احتمالية في العلوم الإنسانية والاجتماعية، ولكن على الرغم من البيانات الكبيرة، فإن البيانات المفتوحة والبنى التحتية للبيانات توفر منهجيات وطرق جديدة للقضايا الاجتماعية، والثقافية، والسياسية، والاقتصادية، والتاريخية. فالفصل الثامن استعرض خطة تفصيلية بالمزايا والانتقادات النسبية المتعلقة بالعلوم الإنسانية الرقمية والعلوم الاجتماعية الحاسوبية التي يتم تطويرها، ولكن هناك حاجة ملحة لمناقشات كاملة وصریحة من شأنها استخراج النتائج من التحول المعرفي في الفكر والممارسة، ويجب أن تقوم مثل هذه المناقشات أيضاً بتقييم دور ومزايا دراسات البيانات الصغيرة، وكذلك المناهج الهجينة التي تمزج أساليب ووسائل البيانات الصغيرة والبيانات الكبيرة.

ومثل هذا التفكير يجب أن يواكبه تفكير آخر، موجه بشكل أكبر تجاه المعيارية، في تكوين واستخدامات البيانات، وإنشاء وتشغيل مجاميع البيانات. وكما ناقشنا في الفصل العاشر، تطرح ثورة البيانات عدداً من القضايا الشائكة وذلك فيما يتعلق بالمدى الذي يتم على أساسه التقاط الحياة اليومية في مجاميع البيانات، وكيف أن بنية البيانات للناس تفوق وتضع قيوداً على حياتهم بجميع أنواع الطرق غير المتوقعة. فعلى سبيل المثال، هل يجب أن يتحلى الناس بالمقدرة على الوصول إلى جميع البيانات المتعلقة بهم، وأن يكون لديهم الحق في تصحيح والاعتراض على هذه البيانات، أو تحليلات البيانات أو التنميط؟ وإلى أي مدى يجب أن يكون الناس قادرين على ممارسة الحق في أن يتم نسيانهم؟ وما هي الطرق التي يمكن من خلالها إخضاع ممارسة البيانات لرقابة وتنظيم مستقلين، وخاصة عندما تكون بياناتهم وخوارزمياتهم تؤثر في الفرص الحياتية للناس؟ وما هو مدى تأمين الأنواع المختلفة من البيانات؟ وما العقوبات المحتملة على انتهاكات هذا التأمين؟ وما الحدود المشروعة للحكومات في تجميع ودمج البيانات عن مواطني هذه الحكومات؟ وإلى أي مدى يمكن إشراك المؤسسات في أعمال الحكومات وأنظمة الحوكمة؟ فهذه القضايا تتعلق بماهية المجتمع المرغوب فيه في عصر وفرة البيانات، والإجابات عن هذه الأسئلة لها تداعيات وخاصة فيما يتعلق بإمكانية إعادة تصور وإعادة تهيئة مجاميع البيانات، وهذا يشمل الأشكال الجديدة من الحوكمة والتنظيم. فالعلماء القانونيون بدؤوا بالفعل في التفكير في بعض القضايا مثل نتائج خصوصية البيانات الكبيرة، وبدؤوا كذلك في تقديم طرق بديلة للتعامل مع أكثر الآثار ضرراً، ولكن من الواضح أن مثل هذه القضايا بعيدة كل البعد عن الحل، وأنها ستكون محل تركيز المناقشات المتطورة والصراعات التشريعية، وأن هناك حاجة ملحة للإسهام في هذا الحوار الفعال والصريح والمستمر.

ويتزامن مع تطوير فهم مفاهيمي ومعيارية أعمق وأفضل للبيانات، الحاجة إلى توفير وجهات نظر شاملة نقدية وكلية عن البنى التحتية للبيانات والبيانات المفتوحة والبيانات الكبيرة وتداعياتهم. لقد قدم هذا الكتاب تحليلاً شاملاً بدلاً من طرح فرضية تستند إلى بحث تجريبي شامل، واستفاد من الكتابات الأكاديمية والإعلامية والصناعية. فهذا النهج ذو فائدة لأنه يجمع الأفكار والملاحظات سوياً من مصادر متعددة وذلك من أجل تقديم نظرة أعم وأشمل عن المشهد المنتشر، مما يساعد على رسم وتصور الديناميكيات الأكبر، ودراسة

التأثيرات المحتملة لها. ومع ذلك، وعلى الرغم مما سبق، فإن المنظور العام الصعب، والتأملي، والشامل لثورة البيانات لا يزال ضعيفاً في الواقع العملي. وبدلاً من ذلك، فهناك مجموعة من الكتب وعلى وجه الخصوص في مجالات الإدارة، والأعمال، والمجالات التقنية تعمل بشكل كبير على دعم البيانات ولكنها تفتقر للمنظور النقدي. وعلى النقيض من ذلك، فالمطلوب تقارير متعددة الاختصاصات ومعقدة في تفكيرها. وفي حين أنه من المفيد دراسة ثورة البيانات من خلال منظور متعدد الاختصاصات والأوجه، إلا أن هذا المنظور مقيد بشكل جذري وذلك بالنظر إلى الأجندة الشاملة والآثار المترتبة. وعلى نحو مشابه، فإن المفهوم القائل بأن ثورة البيانات تنتج البيانات التي يمكن تفسيرها خارج سياقها أو المعارف خارج مجالها المخصص بحاجة إلى أن يتم إلغاؤه. وبدلاً من ذلك، فمن الضروري الاستفادة من الأفكار والمدارك عبر الاختصاصات، حيث يعمل العلماء الذين يتمتعون بخبرات مختلفة سويًا من أجل تقديم صورة أكثر شمولاً وعمومية (Ruppert 2012).

وعلاوةً على ما سبق، فإن التحليلات أوجدت حاجة لكي نكون حريصين ألا ننزلق إلى الجدليات التافهة التي إما تحدد بشكل انفعالي حالة البيانات الكبيرة، والبيانات المفتوحة، والبنى التحتية للبيانات، أو توفر نقداً مستحكماً يشجب أكثر نتائجها سلبية. فالموقف لا يكون إما أبيضاً أو أسوداً، أو جيداً أو سيئاً، إلا فيما ندر، وعلى النحو الذي تطرحه الجدليات الانفعالية. وبدلاً من ذلك، يتعين علينا أن نقر بأن التطورات الحادثة اتفاقية، وارتباطية، وسياقية، وهي تنتشر بطرق فوضوية وأحياناً تناقضية وتباينية. وكما أشرنا إلى ذلك في الفصل التاسع، تعمل مجاميع البيانات في الغالب على كل من الإيجاب والتحرير بشكل فوري ومتزامن، وأن مستهلكي البيانات أحياناً يكونون هم منتجي البيانات، وأن الأنظمة تسعى إلى استنطاق المستخدمين بطرق عديدة. وعلى هذا النحو، فإن المجاميع غالباً ما تكون متشابكة ويتعين أن يتم فتحها وتفريغها بشكل حذر وذلك من أجل إظهار تدرجاتها الرمادية المتعددة. وعندما توجد تأثيرات تعمل بطرق مضرّة وتمييزية، فنحن بحاجة إلى تحديد الاستجابات الاجتماعية والقانونية المناسبة التي تحمي الحقوق الفردية والجماعية، ولكنها لا ترفض الكل لعدم رضاها عن الجزء.

دراسة مجاميع البيانات (Researching Data Assemblages):

هناك ضرورة ملحة لإجراء بحوث تجريبية مفصلة عن تكون، وعمل، وتعزيز مجاميع البيانات وذلك من أجل مواكبة وتعزيز التحليلات النقدية والمفاهيمية الأعم والأشمل. ففي الوقت الحالي، لدينا فهم متعمق قليل بكل من البناء الإجمالي لمجاميع البيانات وأجهزتها وعناصرها الفردية. ونتيجة لذلك، هناك حاجة ملحة لإجراء دراسات حالة تتبع الترتيبات الفنية الاجتماعية لكافة المجاميع، مثل تجميع بيانات لسماح ما أو البنية التحتية لبيانات بحث ما، أو حركة البيانات المفتوحة في مدينة ما، أو التوثيق المفصل لمجالات معينة من أحد المجاميع، مثل علم الاجتماع، والاقتصاد السياسي لمطوري البرمجيات (Hackathons)، ومنتديات العطل الأسبوعية لمجموعات الاختصاص (Data Dives)، أو مجتمعات الممارسة داخل قطاع ما في البيانات الكبيرة. وبشكل نموذجي، فإن مثل هذه الدراسات ستكون دراسات مقارنة بطبيعتها، تقارن التكرارات في تجميع ما، على سبيل المثال، عبر الأماكن أو البيئات، أو عبر أنواع مختلفة من التجميع. ومن شأن مثل بحوث المقارنة هذه أن تمكن من تحديد العموميات والخصوصيات، وكذلك الطرق الاتفاقية والارتباطية المتنوعة التي تتسع من خلالها التجميع ليتم تشكيلها.

وهناك طرق عديدة يمكن من خلالها تفعيل مثل هذه البحوث، وهي تشمل تحليلات البيانات، كما تمت مناقشته في الفصل السادس، وذلك من أجل تحديد تركيب تجميع ما والارتباطات والعلاقات بين العناصر. وسيكون لمثل هذا النهج فوائد عديدة وذلك من خلال استخدام قوة التنقيب عن البيانات والتعلم الآلي من أجل تحديد الأنماط داخل وعبر مجاميع كبيرة للبيانات، وكذلك من أجل تأسيس قوة نسبية للروابط. وهذا من شأنه أن يوفر تفسيراً واسعاً ومقارناً للمجاميع وتشغيل هذه المجاميع. وعلى الرغم من ذلك، وفيما يتعلق بفهم السياق، والاتفاقية والارتباطية المتعلقة بتجميع ما -التداخل المعقد للكائن ومحيطه - فإنه من المحتمل، ربما بشكل ساخر، أن مجموعة من المنهجيات الخاصة بالبيانات الصغيرة قد توفر تبصراً إدراكياً تنويرياً دقيقاً. وتشتمل هذه المنهجيات، على سبيل المثال لا الحصر، على منهج علم الأصول (Genealogies)، ومنهج التفكيك (Deconstruction)، ومنهج الأعراق والثقافات (Ethnographies) ومنهج مشاركة الملاحظ (Observant Participation)، وهي تمثل مجموعة من المنهجيات التي أيدت استخدامها في السابق

من أجل فهم البرامج (Kitchin and Dodge 2011)، والتخطيط (Dodge et al. 2009; Kitchin et al. 2012 a). وتسعى كل منهجية إلى أن تحدد وتفكك بشكل دقيق الظواهر الاجتماعية، وكما بينت العناصر الأحد عشر، فإن مجاميع البيانات تعتبر اجتماعية بشكل كلي. ومن خلال استخدامها كمجموعة، فإن الإنتاج المادي والاستطراذي المنتشر لمجاميع البيانات يمكن تجزئته لأغراض علمية، مما يوفر ما أطلق عليه (Wang 2013) مصطلح "البيانات الكثيفة"، أي الأفكار والقصص السياقية عن البيانات وإنتاج البيانات.

إن علم الأصول يستخدم كثيراً من أجل تتبع الانتشار العرضي لنظام الأفكار أو لمجموعة من الأفعال عبر الأوقات والأزمان، وذلك بدلاً من إنتاج تأريخ غائي سليم (Crowley 2009). وهذا يوضح كيف أن المستقبل يُبنى على الماضي، ولكنه لا يتحدد بالضرورة من خلاله بطرق سببية مسببه بسيطة. وعلى هذا النحو، يتم توظيف علم الأصول من أجل الحل والفهم بشكل تاريخي للتكرارات المتناقضة والمتغيرة أحياناً، والمعقدة، والمتعددة لأحد مجاميع البيانات، أو أحد عناصرها وكذلك الفهم للانتشار المتطور والقائم للأفكار، والقرارات، والمعوقات، والأفعال، واللاعبين الذين أسهموا في تطور هذه المجاميع، جنباً إلى جنب مع النهايات المسدودة والإخفاقات الواضحة (Foucault 1977). وعند فعل ذلك، فإن هذا يحدد نقاط التجمع حيث يحتشد الناس أو الأفكار سويماً ويتسببون في ظهور مجاميع جديدة، وكذلك الطرق المعقدة والفوضوية التي تتطور من خلالها هذه المجاميع بعد ذلك. وفي حال ثورة البيانات، يمكن توظيف منهج علم الأصول من أجل تتبع تكوين وتطور البيانات المفتوحة والبنى التحتية للبيانات والبيانات الكبيرة بشكل أكثر عمومية، والتجسيد المحدد لها.

فعلى سبيل المثال، يمكن لشخص ما أن ينتج تأصيلاً مفصلاً لتطور حركة البيانات المفتوحة بتعبيرات عامة، من خلال تتبع ظهور اللاعبين، والمنظمات، ونماذج التمويل، والأفكار، والأحداث بمرور الوقت وعبر الزمن، والترابطات والتحالفات والخلافات بينهم، وتأثيراتهم على إعادة تشكيل مشهد البيانات من خلال إعادة تحديد السياسات التي تحيط بالحكومة التي أنتجت، واحتفظت، وكشفت مثل هذه البيانات من أجل استخدامات أعم وأشمل. أو يمكن للشخص أن يشرع في تحديد أصل كيان فردي مثل مؤسسة المعرفة المفتوحة، وهي

مؤسسة غير ربحية تأسست في العام 2004 في المملكة المتحدة، وامت بسرعة لكي تصبح حركة دولية مؤثرة تطالب بالبيانات المفتوحة والحكومة المفتوحة. وهنا، يمكن أن يسعى الفرد تجاه تتبع نشأة المنظمة، وتطور مهمتها وقيمتها، وتطور البنى التحتية، والحوكمة، والتمويل، والأنشطة، والبرامج، والإصدارات، وانتشار تأثيرها وتفاعلاتها مع الهيئات الأخرى وهكذا، وذلك من أجل فهم الطريقة الاتفاقية والارتباطية لتوسعها وعملها على سبيل المثال، وكذلك موقعها الحالي وسلطتها ضمن حركة البيانات المفتوحة الأعم.

إن إجراء تحليل علم الأصول عادةً ما يصاحبه الاستفادة من التفكير، وهذه الطريقة من التحليل تجرد وتكشف المعاني داخل النصوص المفهومة بشكل واسع مثل النصوص المكتوبة، والصور، والخرائط، والأحاديث، ومزيج من هذه الأشياء تشكل في العادة المحتوى للكتيبات والمواقع الإلكترونية. فمن المفهوم أن النصوص هي الوسيط لكل من الرسائل الصريحة والضمنية، ومن خلال التفسير التشريحي للنص، يمكن فهم والكشف عن معانٍ أعمق. ومن ناحية الجوهر، فإن التفكير يبحث فيما يتجاوز الإشارات السطحية والقيم الاسمية لكي يواجه القراءات المسلم بها للنص وذلك من أجل الكشف عن المعاني المخفية بداخله، أو ربط المعاني المبسطة بشكل محكم (Burman and MacLure 2005; Harley 1989). فالتفكير هو قراءة نقدية تستخدم أساليب مثل استبدال المعاني أو القوة المحتملة، أو تحديد نقاط التناقض أو التباين لأجل حل النصية البينية والتضمين داخل الماديات والسياقات الأعم، والتفكير في وضعية المتحدث والجمهور المستهدف له (Burman and MacLure 2005). وكما أشرنا إلى ذلك بالتفصيل في الفصول الثاني والثالث والسابع، فإن البيانات المفتوحة والبنى التحتية للبيانات والبيانات الكبيرة عادةً ما يواكبها جميعاً أنظمة خطابية تسعى إلى إقناع الناس، والشركات، والمؤسسات بمنطقها من أجل دعم وتبني والاستثمار في مبادئها ومنتجاتها. فالأنظمة الخطابية هذه ليست عاملية ولكنها معقدة وفوضوية، فهي مزيج من المصالح المتعاونة والمتنافسة. ومن خلال استخدام التفكير، فإن الأنظمة الخطابية لتجميع ما أو عناصر معينة في هذه التجميعات، مثل الفئات العديدة في حركة البيانات المفتوحة أو شركات معينة تعمل على تعزيز حلول البيانات الكبيرة على القضايا المدنية، يمكن تجريدتها وتفكيك رسالتها وتكوينها، مما يكشف عن المعاني، والأيدولوجيات (المعتقدات)، والسلطات الكامنة داخل تصميمها وعرضها.

وتوفر دراسات الأعراق والثقافات تحليلات شاملة وهائلة لمجتمعات الممارسة من خلال التوثيق التفصيلي لتنظيمها، وعملها، وثقافتها، وقواها المحركة الداخلية وارتباطاتها الخارجية (Herbert 2000). وعلى نحو نموذجي، يوظف المنهج الإثنوغرافي (وصف الأعراق والثقافات) مزيجاً من مراقبة المشاركين التي يتم القيام بها عبر فترة زمنية ممتدة، ومقابلات شخصية متعمقة مع مجموعة كبيرة من المستفيدين الداخليين والخارجيين، مع تعزيز كل ما سبق بأساليب أخرى مثل القراءات التفسيرية للوثائق والأدوات ذات الصلة مثل المواقع الإلكترونية، وتبادلات البريد الإلكتروني، وأماكن العمل... إلخ. ومن الناحية الجوهرية، يسعى علم دراسة الأعراق والثقافات إلى الفهم الدقيق للعالم الحياتي لمجتمع ما - علاقاته الاجتماعية، وإيقاعاته المتواترة، ومعانيه الثقافية، وأمط السلطة، واتخاذ القرار فيه، وطرق كينونته... إلخ - من أجل فهم كيف تشكل هذا المجتمع وكيف ينمو بشكل مستمر (Crang and Cook 2007). ويذهب الباحث بعيداً متجاوزاً تقديم وصف سطحي لتوثيق الطرق المعقدة التي تعمل من خلالها المجتمعات من خلال كونها متضمنة داخل العلاقات والممارسات التي يتم دراستها. ومن شأن دراسة الأعراق والثقافات لمجاميع البيانات أن تقدم بعد ذلك أوصافاً تفصيلية شاملة عن كيف تم تكوينها وكيف تعمل على أرض الواقع. فعلى سبيل المثال، فإن دراسة الأعراق والثقافات لمؤسسة المعرفة المفتوحة سوف تتكون من باحث تم تضمينه داخل تلك المنظمة، وهو يشارك في التخطيط للاجتماعات، والأنشطة، والفعاليات، ويراقب تفاعلات العمال ويجري مقابلات شخصية مع المستفيدين الأساسيين للحصول على مدارك مباشرة حول عمليات اللاعبين الأساسيين في حركة البيانات المفتوحة، وكيف يتم تأسيس وإدارة المكونات العديدة لمجاميع البيانات (بمعنى أنظمة الأفكار؛ والاقتصاد السياسي؛ وأمط الحاكمية؛ والماديات؛ والبنى التحتية؛ إلخ) داخلها وعبرها. وعلى نحو مشابه، يمكن للفرد أن يجري دراسة الأعراق والثقافات داخل بنية تحتية معينة للبيانات، أو على سمسار للبيانات، أو داخل شركة للبيانات الكبيرة، أو داخل أنواع معينة من الفعاليات مثل سباقات البرمجة أو المعارض والمؤتمرات الصناعية، أو فرق البيانات الكبيرة داخل الحملات الانتخابية... إلخ.

ويمكن أن يتم دعم مثل هذه الدراسات للأعراق والثقافات من خلال مشاركة الملاحظ، وهذا نوع من الدراسة الذاتية للأعراق والثقافات حيث يُجري الباحث

دراسة مستديمة لارتباطاتهم وارتباطات الناس الآخرين بظاهرة أو ممارسة ما (Crang and Cook 2007; Morton 2005). وفيما يتعلق بثورة البيانات، من الممكن أن يتم مشاركة الملاحظين بحيث يكونون جزءاً من فريق يبني البنى التحتية للبيانات، أو يكونون أعضاءً في حركة البيانات المفتوحة، أو يعملون كعلماء للبيانات يوظفون البيانات الكبيرة من أجل معالجة قضية معينة... إلخ. فتوظيف مثل هذا النهج يعتبر في الأصل تمريناً للانعكاس الذاتي، وهو تدريب شخصي وغير موضوعي، حيث يسعى الباحث جاهداً وبدقة بالغة لدراسة ممارساته وممارسات الآخرين في مجال ما حيث يكون لاعباً أساسياً في هذا المجال، فالفائدة الأساسية لمثل هذا المنهج تتمثل في أن الباحث يدرك بشكل كامل المشهد المتنوع والمعقد من النواحي الاجتماعية، والسياسية، والاقتصادية والذي يعمل من خلاله هذا الباحث ويعتبر أحد المشاركين الفاعلين فيه ويتعرض للعمليات المتعددة الجارية. ويظل الانتقاد الأساسي الموجه لمثل هذا المنهج هو أنه من الممكن بدرجة كبيرة أن ينثني الباحث ويوضع في مكان ما بشكل شخصي، مما يفقده التمثيل العريض، وربما تتم الاستفادة من هذا المنهج بشكل أمثل إذا ما تم دمج مع طرق أخرى توفر سياقاً أعم وأشمل.

إن هذه الطرق تمثل مجموعة جزئية من مجموعة من الطرق المحتملة لفهم مجاميع البيانات وعناصر هذه المجاميع، ومما لا شك فيه أن الكثير من هذه المناهج سوف يتم توظيفها من أجل إجراء بحوث تجريبية على طبيعة ثورة البيانات المنتشرة، إذ يوفر كل منهج رؤى مختلفة. وفي الواقع، فإن المنهج الذي يستخدم طرقاً مختلطة تمزج بين الرؤى السياقية المتعمقة للدراسات النوعية التي تستهدف البيانات الصغيرة وبين الدراسات الأعم التي توظف تحليلات البيانات ربما يكون سبيلاً مفيداً لاتباعه. وفي مثل هذه الدراسة، من الممكن استخدام تحليلات البيانات من أجل تقديم منظور شامل لمجاميع البيانات، أو الروابط أو التداخلات بين المجاميع، والتي يتم اتباعها بعد ذلك من خلال البحوث الأكثر تركيزاً باستخدام الطرق التي تم تفصيلها من أجل توفير رؤى متعمقة للعلاقات والعمليات السارية داخل وبين هذه العناصر. أو من الممكن أن يكون هذا المزيج مزيجاً كميّاً صرفاً فيما يتعلق بالصيغة، وذلك على النحو الذي تمت الإشارة إليه بالتفصيل في المناقشات بشأن علم البيانات في الفصل الثامن، حيث تم تبني التحليل التفسيري لتجميع البيانات بشكل

أكبر من اختبار الفرضيات التقليدي من أجل تحديد أكثر العناصر الفاعلة ظهوراً. ويمكن أن يعمل هذا المنهج أيضاً بشكل عكسي، حيث يتم استخدام البحوث المتعمقة المركزة من أجل توجيه انتشار تحليلات البيانات عبر عينة أكبر بكثير. وفي كلتا الحالتين، فإن المرحلتين الأولى والثانية تعملان بتناغم بعضهما مع بعض من أجل تعزيز المدارك والمعارف. ففي الحالة الأولى، يتم استخدام الصورة الأعم من أجل المساعدة في صياغة الفهم وتعميقه وقرنه بالسياق، أما في الحالة الثانية، فإن الفهم المتعمق يوفر الأساس لمحاولة تأسيس أو وضع تفسير أعم وأشمل.

أفكار ختامية:

لا تزال ثورة البيانات في طورها التمهيدي، ولكنها تنتشر بسرعة بالغة، فخلال عدد من السنوات لا يتجاوز أصابع اليد الواحدة، كان للبيانات المفتوحة والبيانات الكبيرة والبنى التحتية للبيانات تأثيرات كبيرة على مشهد البيانات، وتزامن هذا مع مناقشات صوتية حماسية تعلن عن التأثيرات الإيجابية الخطيرة لها. إن سرعة التطورات التي تحدث تعني أن التحليل النقدي الموسع، والمتعمق، والمدرّس يجد صعوبة بالغة في مواكبة تلك التطورات. ومما لا شك فيه أنه خلال السنوات القليلة القادمة سيكون هناك فيضان حقيقي من مثل هذه الدراسات التي تسعى إلى توثيق طبيعة وآثار مجاميع البيانات الناشئة. وكما ادعى البعض خلال هذا الفصل، فهناك حاجة ملحة لهذه الدراسات. إن هذا العمل سوف يحدد من جانب الطرق المناسبة لفهم ثورة البيانات، من خلال كل من التفكير المفاهيمي والدراسات النقدية. وعلى الجانب الآخر، سيوفر بيئة لتفكير أكثر معيارية فيما يتعلق بكيفية تطور وتنظيم وإدارة مشهد البيانات. وفي الوقت الراهن، فإن الأخير يكون في الغالب تفاعلياً، ومفهوماً بشكل ضعيف، وتسيطر عليه المصالح المكتسبة. لقد سعى هذا الكتاب إلى تقديم تحليل أولي ومفاهيمي بشكل متوازن، ونقدي وشامل والذي من الممكن أن يساعد على توجيه الدراسات التي تتبعه. ولوقت طويل للغاية، فإن البيانات وتكوين وعمل المجاميع التي تحيط بالبيانات قد تم الأخذ بها على أنها أمورٌ مسلم بها، في حين أن الاهتمام كان منصباً على المعلومات والمعارف المستقاة منها، فالوقت قد حان لتدارك هذا الإهمال.

المراجع:

- Adler, M.J. (1986) *A Guidebook to Learning: For a Lifelong Pursuit of Wisdom*. Macmillan, London.
- Agre, P. (1994) 'Surveillance and capture: two models of privacy', *Information Society*,10(2): 101-27.
- Alder, K. (2002) *The Measure of All Things - The Seven-Year-Odyssey that Transformed the World*. Abacus, London.
- Alonso, W. and Starr, P. (1987) *The Politics of Numbers*. Russell Sage Foundation, New York.
- Althusser, L. (1971) *Lenin and Philosophy and Other Essays*, translated by B. Brewster. New Left Books, London.
- Amin, A. and Thrift, N. (2002) *Cities: Reimagining the Urban*. Polity, London.
- Amin, A. and Thrift, N. (2013) *Arts of the Political: New Openings for the Left*. Duke University Press, Durham, NC.
- Amore, L. (2006) 'Biometric borders: governing mobilities in the war on terror', *Political Geography*, 25: 336-51.
- Anderson, B. (2009) 'Affective atmospheres', *Emotion, Space and Society*, 2: 77-81.
- Anderson, C. (2008) 'The end of theory: the data deluge makes the scientific method obsolete', *Wired*, 23 June, http://www.wired.com/science/discoveries/magazine/1607-/pb_theory (last accessed 12 October 2012).
- Andrejevic, M. (2007) *iSpy: Surveillance and Power in the Interactive Era*. University of Kansas, Lawrence, KS.
- Andrejevic, M. (2013) *Infoglut: How Too Much Information is Changing the Way We Think and Know*. Routledge, New York.
- Anthony, S. (2013) 'DARPA shows off 1.8-gigapixel surveillance drone, can spot a terrorist from 20,000 feet', *ExtremeTech*, 28 January 2013, <http://www.extremetech.com/extreme/146909-darpa-shows-off-18-gigapixel-surveillance-drone-can-spot->

- a-terrorist-from-20000-feet (last accessed 1 February 2013).
- Arthur, C. and Cross, M. (2006) 'Give us back our crown jewels', *The Guardian*, 9 March. <http://www.theguardian.com/technology/2006/mar/09/education.epublic> (last accessed 21 October 2013).
 - Asay, M. (2013) 'Big data and the landfills of the digital enterprise', *Readwrite*, 11 February. <http://readwrite.com/201311/02//big-data-and-the-landfills-of-our-digital-lives> (last accessed 19 February 2013).
 - Bamford, J. (2009) *The Shadow Factory: The Ultra-secret NSA from 911/ to the Eavesdropping on America*. Anchor Books, New York.
 - Barnes, T.J. and Hannah, M. (2001) 'The place of numbers: histories, geographies, and theories of quantification', *Environment and Planning D: Society and Space*, 19: 379–83.
 - Barnett, C. (2013) 'Theory as political technology', *Antipode* (online first), http://radicalantipode.files.wordpress.com/201307//book-review_barnett-on-amin-and-thrift.pdf (last accessed 2 August 2013).
 - Bates, J. (2012) "'This is what modern deregulation looks like": co-optation and contestation in the shaping of the UK's Open Government Data Initiative', *The Journal of Community Informatics*, 8(2) <http://www.cijournal.net/index.php/ciej/article/view/845916/> (last accessed 6 February 2013).
 - Bates, J. (2013) 'Opening up public data', *SPERI Comment*, 21 May. <http://speri.dept.shef.ac.uk/201321/05//opening-public-data/> (last accessed 18 September 2013).
 - Batty, M. (2007) *Cities and Complexity: Understanding Cities with Cellular Automata, Agent Based Models, and Fractals*. MIT Press, Cambridge, MA.
 - Batty, M., Axhausen, K.W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G. and Portugali, Y. (2012) 'Smart cities of the future', *European Physical Journal Special Topics*, 214: 481–518.
 - Baym, N.K. (2013) 'Data not seen: the uses and shortcomings of social media metrics', *FirstMonday*, 18(10), <http://firstmonday.org/ojs/index.php/fm/article/view/48733752/> (last accessed 3 January 2014).

- Beagrie, N., Lavoie, B. and Woollard, M. (2010) Keeping Research Data Safe 2, JISC, London and Bristol. <http://www.beagrie.com/jisc.php>.
- Becker, H. (1952) 'Science, culture, and society', *Philosophy of Science*, 19(4): 273–87.
- Beer, D. and Burrows, R. (2007) 'Sociology and, of and in Web 2.0: some initial considerations', *Sociological Research Online*, 12(5), <http://www.socresonline.org.uk/1217/5/.html>.
- Beer, D. and Burrows, R. (2013) 'Popular culture, digital archives and the new social life of data', *Theory, Culture and Society*, 30(4): 47–71.
- Benkler, Y. (2006) *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, New Haven, CT.
- Berners-Lee, T. (2009) *Linked Data*, <http://www.w3.org/DesignIssues/LinkedData.html> (last accessed 30 January 2013).
- Berry, D. (2011) 'The computational turn: thinking about the digital humanities', *Culture Machine*, 12, <http://www.culturemachine.net/index.php/cm/article/view/440470/> (last accessed 3 December 2012).
- Bertolucci, J. (2013) 'IBM, universities team up to build data scientists', *Information Week*, 15 January. <http://www.informationweek.com/big-data/big-data-analytics/ibm-universities-team-up-to-build-data-scientists/> (last accessed 16 January 2014).
- Bettencourt, L.M.A., Lobo, J., Helbing, D., Kuhnert, C. and West, G.B. (2007) 'Growth, innovation, scaling, and the pace of life in cities', *Proceedings of the National Academy of Sciences*, 104(17): 7301–06.
- Bielsa, A. (2013) 'The Smart City Project in Santander', *Sensors*, 1 March, <http://www.sensormag.com/wireless-applications/smart-city-project-santander-11152> (last accessed 27 May 2013).
- Boellstorff, T. (2013) 'Making big data, in theory', *First Monday*, 18(10), <http://firstmonday.org/ojs/index.php/fm/article/view/48693750/> (last accessed 3 January 2014).
- Bollier, D. (2010) *The Promise and Peril of Big Data*. The Aspen Institute. http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf (last accessed 1 October 2012).

- Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J. and Wilderman, C.C. (2009) Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. http://www.informalscience.org/documents/PPSR_report_20FINAL.pdf. A CAISE Inquiry Group Report, Center for Advancement of Informal Science Education (CAISE). Washington, DC (last accessed 17 January 2014).
- Borgman, C.L. (2007) *Scholarship in the Digital Age*. MIT Press, Cambridge, MA.
- Borgman, C.L. (2012) 'The conundrum of sharing research data', *Journal of the American Society for Information Science and Technology*, 63(6): 1059–78.
- Bowker, G. (2005) *Memory Practices in the Sciences*. MIT Press, Cambridge, MA.
- Bowker, G. and Star, L. (1999) *Sorting Things Out: Classification and Its Consequences*. MIT Press, Cambridge, MA.
- boyd, D. and Crawford, K. (2011) 'Six provocations for big data', SSRN, <http://ssrn.com/abstract=1926431> (last accessed 22 February 2013)
- boyd, D. and Crawford, K. (2012) 'Critical questions for big data', *Information, Communication and Society*, 15(5): 662–79.
- Brooks, D. (2013a) 'The philosophy of data', *New York Times*, 4 February, <http://www.nytimes.com/2013/05/02/opinion/brooks-the-philosophy-of-data.html> (last accessed 5 February 2013).
- Brooks, D. (2013b) 'What data can't do', *New York Times*, 18 February, <http://www.nytimes.com/2013/19/02/opinion/brooks-what-data-cant-do.html> (last accessed 18 February 2013).
- Bruns, A. (2013) 'Faster than the speed of print: reconciling "big data" social media analysis and academic scholarship', *First Monday*, 18(10), <http://firstmonday.org/ojs/index.php/fm/article/view/48793756/> (last accessed 3 January 2014).
- Bryant, R., Katz, R.H. and Lazowska, E.D. (2008) *Big-Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science and Society*. December, Computing Community Consortium, http://www.cra.org/ccc/docs/init/Big_Data.pdf (last accessed 12 October 2012).

- Brynjolfsson, E., Hitt, L.M. and Kim, H.H. (2011) Strength in Numbers: How Does Data- Driven Decision-making Affect Firm Performance? SSRN, <http://ssrn.com/abstract=1819486> (last accessed 22 February 2013).
- Burman, C. and MacLure, M. (2005) 'Deconstruction as a method of research', in B. Somekh and C. Lewin (eds), *Research Methods in the Social Sciences*. Sage, London, pp. 282-92.
- Callebaut, W. (2012) 'Scientific perspectivism: a philosopher of science's response to the challenge of big data biology', *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43: 69-80.
- Calzolari, M.C. (2012) 'Analysis of Twitter followers of leading international companies', 8 June, <http://www.camisanicalzolari.com/MCC-Twitter-ENG.pdf> (last accessed 3 January 2014).
- Campbell, D.T. (1976) *Assessing the Impact of Planned Social Change*. The Public Affairs Center, Dartmouth College, New Hampshire.
- Carr, N.G. (2007) 'The ignorance of crowds', *Strategy + Business Magazine*, 47: 1-5.
- Carroll, R. (2013) 'Welcome to Utah, the NSA's desert home for eavesdropping on America', *The Guardian*, 14 June. <http://www.guardian.co.uk/world/2013/jun/14/nsa-utah-data-facility> (last accessed 8 July 2013).
- Case, D.O. (2002) *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. Academic Press, Amsterdam.
- Castells, M. (1988) *The Informational City: Information Technology, Economic Restructuring and the Urban-regional Process*. Blackwell, Oxford.
- Castells, M. (1996) *Rise of the Network Society*. Blackwell, Oxford.
- Cavoukian, A. (2009) *Privacy by Design: A Primer*. <http://www.privacybydesign.ca/content/uploads/201310//pbd-primer.pdf> (last accessed 15 October 2013).
- CCSDS (2012) *Reference Model for an Open Archival Information System (OAIS)*. Consultative Committee for Space Data Systems, Washington, DC. <http://public.ccsds.org/publications/archive/650x0m2.pdf> (last accessed 21 October 2013).

- Chignard, S. (2013) 'A brief history of open data', Paris Tech Review, 29 March, <http://www.paristechreview.com/201329/03//brief-history-open-data/> (last accessed 18 September 2013).
- Chopra, S. and Meindl, P. (2012) Supply Chain Management: Strategy, Planning and Operation, 5th edition. Pearson, Harlow.
- Christensen, C.M. (1997) The Innovator's Dilemma. Harvard Business Review Press, Cambridge, MA.
- CIPPIC (2006) On the Data Trail: How Detailed Information about You Gets into the Hands of Organizations with whom You Have no Relationship. A Report on the Canadian Data Brokerage Industry. <http://www.cippic.ca/uploads/May106-/DatabrokerReport.pdf>, The Canadian Internet Policy and Public Interest Clinic, Ottawa (last accessed 17 January 2014).
- Clark, L. (2013) 'No questions asked: big data firm maps solutions without human input', Wired, 16 January, <http://www.wired.co.uk/news/archive/201316/01-/ayasdi-big-data-launch> (last accessed 28 January 2013).
- Clarke, R. (1988) 'Information technology and dataveillance', Communications of the ACM, 31(5): 498-512.
- Clarke, R. (1994a) 'The digital persona and its application to data surveillance', The Information Society, 10(2): 77-92.
- Clarke, R. (1994b) 'Human identification in information systems: management challenges and public policy issues', Information Technology and People, 7(4): 6-37.
- Claverie-Berge, I. (2012) 'Solutions big data IBM', IBM corporation. http://www-05.ibm.com/fr/events/netezzaDM_2012/Solutions_Big_Data.pdf (last accessed 9 July 2013).
- Clements, A. (2013) Democratic State Surveillance, Transparency and Trust. Cyberdialogue conference, University of Toronto, 17-18 March. <http://www.cyberdialogue.ca/201303//democratic-state-surveillance-transparency-and-trust-by-andrew-clement/> (last accessed 8 July 2013).
- Clifford, S. (2012) 'Shopper alert: price may drop for you alone', New York

- Times, 9 August, <http://www.nytimes.com/2012/10/08/business/supermarkets-try-customizing-prices-for-shoppers.html> (last accessed 8 August 2013).
- Codd, E.F. (1970) 'A relational model of data for large shared data banks', *Communications of the ACM*, 13(6): 377-87.
 - Cohen, D. (2008) 'Contribution to: The Promise of Digital History' (round table discussion), *Journal of American History*, 95(2): 452-91.
 - Cohen, J. (2012) 'What is privacy for?' Social Sciences Research Network, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2175406 (last accessed 16 July 2013)
 - Constine, J. (2012) 'How big is Facebook's data? 2.5 billion pieces of content and 500+ terabytes ingested every day', *TechCrunch*, 22 August, <http://techcrunch.com/2012/22/08/how-big-is-facebooks-data-25-billion-pieces-of-content-and-500-terabytes-ingested-every-day/> (last accessed 28 January 2013).
 - Cotterill, C. (2011) 'Location privacy: who protects?', *URISA*, 23(2): 49-59.
 - CRA (2003) *Grand Research Challenges in Information Systems*, The Computing Research Association. <http://archive.cra.org/reports/gc.systems.pdf> (last accessed 16 January 2014).
 - Crampton, J., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M.W. and Zook, M. (2012) *Beyond the Geotag? Deconstructing 'Big Data' and Leveraging the Potential of the Geoweb*, http://www.uky.edu/~tmute2/geography_methods/readingPDFs/2012-Beyond-the-Geotag-2012.10.01.pdf (last accessed 21 February 2013).
 - Crang, M. and Cook, I. (2007) *Doing Ethnographies*. Sage, London.
 - Crawford, K. (2013) 'The hidden biases of big data', *Harvard Business Review Blog*, 1 April, <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/> (last accessed 18 September 2013).
 - Critical Art Ensemble (1995) 'The mythology of terrorism on the Net', <http://www.critical-art.net/books/digital/tact2.pdf> (last accessed 31 May 2013).
 - Croll, A. (2012) 'Big data is our generation's civil rights issue, and we don't know

- it', in *Big Data Now: Current Perspectives from O'Reilly Media*. O'Reilly Media, Sebastopol, CA, pp. 55–9.
- Crovitz, L.G. (2012) 'Obama's "Big Data" Victory', *Wall Street Journal*, 18 November, <http://online.wsj.com/article/SB10001424127887323353204578126671124151266.html> (last accessed 19 November 2012).
 - Crowley, U. (2009) 'Genealogy method', in R. Kitchin and N. Thrift (eds), *International Encyclopedia of Human Geography*, vol. 4. Elsevier, Oxford, pp. 341–4.
 - Crutcher, M. and Zook, M. (2009) 'Placemarks and waterlines: racialized cyberscapes in post-Katrina Google Earth', *Geoforum*, 40(4): 523–34.
 - Cukier, K. (2010) 'Data, data everywhere', *The Economist*, 25 February, <http://www.economist.com/node/15557443> (last accessed 12 November 2012).
 - Culler, J. (2010) 'The closeness of close reading', *ADE Bulletin*, 149: 20–25.
 - Curry, M.R. (1997) 'The digital individual and the private realm', *Annals of the Association of American Geographers*, 87: 681–99.
 - Curry, M.R., Philips, D.J. and Regan, P.M. (2004) 'Emergency response systems and the creeping legibility of people and places', *The Information Society*, 20: 357–69.
 - Cyberinfrastructure Council (2007) 'Cyberinfrastructure vision for 21st century discovery', <http://www.nsf.gov/pubs/2007/nsf0728/index.jsp?org=EEC>, National Science Foundation, Washington, DC (last accessed 17 January 2014).
 - Danna, A. and Gandy, O.H. (2002) 'All that glitters is not gold: digging beneath the surface of data mining', *Journal of Business Ethics*, 40: 373–86.
 - Darnton, R. (2000) 'An early information society: news and the media in eighteenth-century Paris', *American Historical Review*, 105: 1–30.
 - Dasish (2012) Roadmap for Preservation and Curation in the Social Sciences and Humanities, http://dasish.eu/publications/projectreports/D4.1_-_Roadmap_for_Preservation_and_Curation_in_the_SSH.pdf (last accessed 15 October 2013).
 - Data Seal of Approval (2010) Quality Guidelines for Digital Research Data, http://www.datasealofapproval.org/media/filer_public/201327/09//dsa-booklet_1_

- june2010.pdf (last accessed 8 October 2013).
- DataRemixed (2013) Worldwide Open Data Sites, 8 August, <http://dataremixed.com/201308//worldwide-open-data-sites> (last accessed 4 November 2013).
 - Davenport, T.H., Barth, P. and Bean, R. (2012) 'How "Big Data" Is Different', MIT Sloan Management Review, 30 July, <http://sloanreview.mit.edu/the-magazine/2012-fall/54104/how-big-data-is-different/> (last accessed 28 January 2013).
 - de Goes, J. (2013) "'Big data" is dead. What's next?', VB/Big Data, 23 February, <http://venturebeat.com/201322/02//big-data-is-dead-whats-next/> (last accessed 26 February 2013).
 - De Micheli, C. and Stroppa, A. (2013) 'Twitter and the underground market', 11th Nexa Lunch Seminar, 22 May, http://nexa.polito.it/nexacenterfiles/lunch-11-de_micheli-stroppa.pdf (last accessed 3 January 2013).
 - de Vries, M., Kapff, L., Negreiro Achiaga, M., Wauters, P., Osimo, D., Foley, P., Szkuta, K., O'Connor, J. and Whitehouse, D. (2011) Pricing of Public Sector Information Study (POPSIS), <http://epsiplatform.eu/sites/default/files/models.pdf> (last accessed 11 August 2013).
 - Deloitte (2012) Open Data: Driving Ingenuity, Innovation and Growth. Deloitte, London. <http://www.deloitte.com/assets/DcomUnitedKingdom/Local%20Assets/Documents/Market%20insights/Deloitte%20Analytics/uk-insights-deloitte-analytics-open-data-june-2012.pdf> (last accessed 14 August 2013).
 - Dembosky, A., Waters, R. and Steel, E. (2012) 'Questions raised over power of big data', Financial Times, Tech Hub, 11 December, <http://www.ft.com/intl/cms/s/092/f7677a-40dd-11e2-aafa-00144feabdc0.html> (last accessed 11 December 2012).
 - Dery, K., Hall, R. and Wailes, N. (2006) 'ERPs as technologies-in-practice: social construction, materiality and the role of organisational factors', New Technology, Work and Employment, 21(3): 229–41.
 - Desrosières, A. (1998) The Politics of Large Numbers: A History of Statistical Reasoning, translated by C. Naish. Harvard University Press, Cambridge, MA.

- Dicken, P. (2003) *The Global Shift*. Sage, London.
- Diebold, F. (2012) 'A personal perspective on the origin(s) and development of "big data": the phenomenon, the term, and the discipline', http://www.ssc.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf (last accessed 5 February 2013).
- Dietrich, D. (2012) 'Linked data', ePSIplatform Topic Report, no: 2012 11, <http://epsiplatform.eu/content/topic-report-linked-data> (last accessed 14 August 2013).
- Dodge, M. and Kitchin, R. (2004) 'Flying through code/space: the real virtuality of air travel', *Environment and Planning A*, 36: 195–211.
- Dodge, M. and Kitchin, R. (2005) 'Codes of life: identification codes and the machine-readable world', *Environment and Planning D: Society and Space*, 23(6): 851–81.
- Dodge, M. and Kitchin, R. (2007a) 'The automatic management of drivers and driving spaces', *Geoforum*, 38(2): 264–75.
- Dodge, M. and Kitchin, R. (2007b) "'Outlines of a world coming in existence": pervasive computing and the ethics of forgetting', *Environment and Planning B*, 34(3): 431–45.
- Dodge, M. and Kitchin, R. (2013) 'Mapping experience and knowledge: crowdsourced cartography', *Environment and Planning A*, 45(1): 19–36.
- Dodge, M., Perkins, C. and Kitchin, R. (2009) 'Mapping modes, methods and moments: a manifesto for map studies', in M. Dodge, R. Kitchin and C. Perkins (eds), *Rethinking Maps*. Routledge, London, pp. 220–43.
- Donovan, K. (2012) 'Seeing like a slum: towards open, deliberative development', *Georgetown Journal of International Affairs*, 13(1): 97–104.
- Dourish, P. (2001) *Where the Action Is*. MIT Press, Cambridge, MA.
- Dourish, P. and Bell, G. (2007) 'The infrastructure of experience and the experience of infrastructure: meaning and structure in everyday encounters with space', *Environment and Planning B*, 34: 414–30.
- Driscoll, K. (2012) 'From punched cards to "big data": a social history of database populism', *communication +1*, 1(4), <http://scholarworks.umass.edu/cpo/vol1/>

- iss14/ (last accessed 22 February 2013).
- Dumbill, E. (2012) 'What is big data?' in Big Data Now: Current Perspectives from O'Reilly Media. O'Reilly Media, Sebastopol, CA. pp. 3–10.
 - Dwoskin, E. (2013) 'Data broker discloses what it knows', Wall Street Journal, 4 September, <http://blogs.wsj.com/digits/201304/09//data-broker-discloses-what-it-knows/>(last accessed 11 October 2013).
 - Dyche, J. (2012) 'Big data "eureka!" don't just happen', Harvard Business Review Blog, 20 November, http://blogs.hbr.org/cs/201211//eureka_doesnt_just_happen.html(last accessed 23 November 2012).
 - Eaves, D. (2013) 'The value of open data – don't measure growth, measure destruction', 25 April, <http://eaves.ca/201325/04//the-value-of-open-data-dont-measure-growth-measure-destruction/> (last accessed 18 September 2013).
 - Economist, The (2010) 'All too much: Monstrous amounts of data', 25 February, <http://www.economist.com/node/15557421> (last accessed 12 November 2012).
 - Economist, The (2013) 'Open data: a new goldmine', 18 May, <http://www.economist.com/news/business/21578084-making-official-data-public-could-spur-lots-innovation-new-goldmine> (last accessed 16 September 2013).
 - Edwards, J. (2013) 'Facebook is about to launch a huge play in "big data" analytics', Business Insider, 10 May, <http://www.businessinsider.com/facebook-is-about-to-launch-a-huge-play-in-big-data-analytics-20135-> (last accessed 18 September 2013).
 - Efrati, A., Thurm, S. and Searchy, D. (2011) 'Mobile-app makers face U.S. privacy investigation', Wall Street Journal, 5 April, <http://online.wsj.com/article/SB10001424052748703806304576242923804770968.html>(last accessed 17 July 2013).
 - Elliott, C. (2004) 'Some rental cars are keeping tabs on the drivers', New York Times, 13 January, <http://www.nytimes.com/200413/01//business/13gps.html> (last accessed 8 August 2013).
 - Ellis, J. (2013) 'Hiding in public: how the National Archives wants to open up its data to Americans', Nieman Journalism Lab, 25 February, <http://www.niemanlab.com>.

- org/201302//hiding-in-public-how-the-national-archives-wants-to-open-up-its-data-to-americans/ (last accessed 26 February 2013).
- Elwood, S. and Leszczynski, A. (2011) 'Privacy reconsidered: new representations, data practices, and the geoweb', *Geoforum*, 42: 6–15.
 - European Commission (2012) Commission Proposes a Comprehensive Reform of the Data Protection Rules, 25 January, http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm (last accessed 6 August 2013).
 - Farber, D. (2013) 'Counting the Internet of things in real time', *CNet*, 30 July, http://news.cnet.com/830176-57596162-3_11386-/counting-the-internet-of-things-in-real-time/ (last accessed 18 September 2013).
 - Farber, M., Cameron, M., Ellis, C. and Sullivan, J. (2011) *Massive Data Analytics and the Cloud: A Revolution in Intelligence Analysis*. Booz Allen Hamilton. <http://www.boozallen.com/media/file/MassiveData.pdf> (last accessed 16 July 2013).
 - Federal Trade Commission (2012) *Protecting Consumer Privacy in an Era of Rapid Change*, <http://www.ftc.gov/os/2012120326/03/privacyreport.pdf> (last accessed 14 October 2013).
 - Ferro, E. and Osella, M. (2013) *Eight Business Model Archetypes for PSI Re-Use*. Open Data on the Web, Workshop, 23–24 April 2013, Google Campus, Shoreditch, London. http://www.w3.org/201304//odw/odw13_submission_27.pdf (last accessed 13 August 2013).
 - Fitzgerald, A.M. (2010) *State of Play: PSI Reuse in Australia*. European Public Sector Information Platform, Topic Report No. 13, <http://epsiplatform.eu/sites/default/files/ePSIplatform%20Topic%20Report%20No.%201320%20Australia%20final.pdf> (last accessed 14 August 2013).
 - Flaounas, I., Ali, O., Lansdall-Welfare, T., De Bie, T., Mosdell, N., Lewis, J. and Cristianini, N. (2013) 'Research methods in the age of digital journalism', *Digital Journalism*, 1(1): 102–16.
 - Floridi, L. (2005) 'Is information meaningful data?', *Philosophy and Phenomenological Research*, 70(2): 351–70.

- Floridi, L. (2008) 'Data', in W.A. Darity (ed.), International Encyclopedia of the Social Sciences, 2nd edition. Detroit: Macmillan. Preprint online: <http://www.philosophyofinformation.net/publications/pdf/data.pdf>.
- Floridi, L. (2010) Information: A Very Short Guide. Oxford University Press, Oxford.
- Floridi, L. (2012) 'Big data and their epistemological challenge', Philosophy and Technology, 25(4): 435–7.
- Flowers, M. (2013) 'Beyond open data: the data-driven city', in B. Goldstein and L. Dyson (eds), Beyond Transparency: Open Data and the Future of Civic Innovation. Code for America Press, San Francisco, CA. pp. 196–209.
- Foucault, M. (1977) Discipline and Punish. Allen Lane, London.
- Foucault, M. (1978) The History of Sexuality, Volume One. Random House, New York.
- Foucault, M. (1981) Power/Knowledge: Selected Interviews and Other Writings 1972–1977. Harvester, London.
- Franks, B. (2012) Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics. Wiley, Hoboken, NJ.
- Fry, J., Lockyer, S., Oppenheim, C., Houghton, J.W. and Rasmussen, B. (2008) Identifying Benefits Arising from the Curation and Open Sharing of Research Data Produced by UK Higher Education and Research Institutes, JISC, London and Bristol. <http://repository.jisc.ac.uk/279/> (last accessed 21 October 2013).
- Gannon, D. and Reid, D. (2009) 'Parallelism and the cloud', in T. Hey, S. Tansley and K. Tolle (eds), The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research, Redmond, WA, pp. 131–5.
- Gantz, J. and Reinsel, D. (2011) Extracting Value from Chaos. IDC, Framingham, MA <http://idcdocserv.com/1142> (last accessed 1 October 2012).
- Gantz, J., Reinsel, D., Chute, C., Schlichting, W., McArthur, J., Minton, S., Xheneti, I., Toncheva, A. and Manfrediz, A. (2007) The Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2010. IDC, Framingham,

- MA, <http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf> (last accessed 26 October 2012).
- Garvey, E.G. (2013) “facts and FACTS”: abolitionists’ database innovations’, in L. Gitelman (ed), ‘Raw Data’ is an Oxymoron. MIT Press, Cambridge, MA, pp. 89–102.
 - Gershenfeld, N., Krikorian, R. and Cohen, D. (2004) ‘The internet of things’, *Scientific American*, October, 291(4): 76–81.
 - Gilder, G. (2000) *Telecosm: How Infinite Bandwidth will Revolutionize Our World*. Free Press, New York.
 - Gilheany, S. (2000) ‘Projecting the cost of magnetic disk storage over the next 10 years’, <http://www.teqniun.in/images/MagneticStorageOvertheNext10years.pdf> (last accessed 16 January 2014).
 - Gitelman, L. and Jackson, V. (2013) ‘Introduction’, in L. Gitelman (ed.), ‘Raw Data’ is an Oxymoron. MIT Press, Cambridge, MA, pp. 1–14.
 - Gleeson, J., Kitchin, R., Bartley, B. and Treacy, C. (2009) ‘New ways of mapping social inclusion in Dublin City’, <http://www.nuim.ie/nirsa/mappinginclusion.pdf>, Dublin City Partnership/Dublin City Development Board (last accessed 13 August 2013).
 - Goddard, L. and Byrne, G. (2010) ‘Linked data tools: semantic Web for the masses’, *First Monday*, 15(11), <http://firstmonday.org/ojs/index.php/fm/article/view/31202633/> (last accessed 14 August 2013).
 - Goldberg, J. (2012) ‘Big data ushers in era of security intelligence’, *Wired: Innovation Insights*, 30 November, <http://www.wired.com/insights/201211/big-data-ushers-in-era-of-security-intelligence/> (last accessed 3 December 2012).
 - Golledge, R. and Stimson, R. (1997) *Spatial Behaviour*. Guilford, New York.
 - González-Bailón, S. (2013) “Big data” and the capillaries of human geography’, *Dialogues in Human Geography*, 3(3): 292–6.
 - González-Bailón, S., Wang, N., Rivero, A., Borge-Holtoefer, J. and Moreno, Y. (2012) *Assessing the Bias in Communication Networks Sampled from Twitter*. Working Paper, <http://arxiv.org/abs/1212.1684> (last accessed 17 January 2014).

- Goodchild, M.F. (2007) 'Citizens as sensors: the world of volunteered geography', *GeoJournal*, 69: 211–21.
- Goodchild, M.F. (2009) 'Uncertainty', in R. Kitchin and N. Thrift (eds), *International Encyclopedia of Human Geography*, vol 12. Elsevier, Oxford, pp. 1–5.
- Gordon, G. (2013) 'Towards an overarching narrative on open government', *Open Government Partnership UK*, 12 February, <http://www.opengovernment.org.uk/towards-an-overarching-narrative-on-open-government/> (last accessed 19 February 2013).
- Gorman, S. (2013) 'The danger of a big data episteme and the need to evolve geographic information systems', *Dialogues in Human Geography*, 3(3): 285–91.
- Goss, J. (1995) "'We know who you are and we know where you live": the instrumental rationality of geodemographics systems', *Economic Geography*, 71: 171–98.
- Gould, P. (1981) 'Letting the data speak for themselves', *Annals of the Association of American Geographers*, 71(2): 166–76.
- Graham, M. (2012) 'Big data and the end of theory?' *The Guardian*, 9 March, <http://www.guardian.co.uk/news/datablog/2012/mar/09/big-data-theory> (last accessed 12 November 2012).
- Graham, M., Stephens, M. and Hale, S. (2013) 'Mapping the geoweb: a geography of Twitter', *Environment and Planning A*, 45: 100–02.
- Graham, S. (2005) 'Software-sorted geographies', *Progress in Human Geography*, 29(5): 562–80.
- Graham, S. and Marvin, S. (1996) *Telecommunications and the City*. Routledge, London.
- Graham, S. and Marvin, S. (2001) *Splintering Urbanism*. Routledge, London.
- Gralla, P., Sacco, A. and Faas, R. (2011) 'Smartphone apps: is your privacy protected?', *ComputerWorld*, http://www.computerworld.com/s/article/9218163/Smartphone_apps_Is_your_privacy_protected_ (last accessed 17 July 2013).
- Grant, D., Harley, B. and Wright, C. (2006) 'Editorial introduction: the work and organ-izational implications of Enterprise Resource Planning systems', *New Technology, Work and Employment*, 21(3): 196–8.

- Granville, V. (2013) 'The curse of big data', analyticbridge, 5 January, <http://www.analyticbridge.com/profiles/blogs/the-curse-of-big-data> (last accessed 28 January 2013).
- Greenfield, A. (2006) *Everyware: The Dawning Age of Ubiquitous Computing*. New Riders, Boston.
- Greenwald, G. and MacAskill, E. (2013) 'NSA Prism program taps in to user data of Apple, Google and others', *The Guardian*, 7 June, <http://www.guardian.co.uk/world/2013/jun/06/us-tech-giants-nsa-data> (last accessed 8 July 2013).
- Grimes, S. (2011) 'Unstructured data and the 80 percent rule', Clarabridge, <http://clarabridge.com/default.aspx?tabid=137&ModuleID=635&ArticleID=551> (last accessed 4 March 2013).
- Grochowski, E. and Halem, R.D. (2003) 'Technological impact of magnetic hard disk drives on storage systems', *IBM Systems Journal*, 42: 338–46.
- Gurstein, M. (2011) 'Open data: empowering the empowered or effective data use for everyone', *First Monday*, 16(2), <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/33162764/> (last accessed 6 February 2013).
- Gurstein, M. (2013) 'Should "Open Government Data" be a product or a service (and why does it matter?)', *Gurstein's Community Informatics*, 3 February, <http://gurstein.wordpress.com/201303/02//is-open-government-data-a-product-or-a-service-and-why-does-it-matter/> (last accessed 6 February 2013).
- Haklay, M. (2010) 'How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets', *Environment and Planning B*, 37: 682–703.
- Hales, D. (2013) 'Lies, damned lies and big data', *Aid on the Edge of Chaos*, 1 February 2013, <http://aidontheedge.info/201301/02//lies-damned-lies-and-big-data/> (last accessed 5 February 2013).
- Han, J., Kamber, M. and Pei, J. (2011) *Data Mining: Concepts and Techniques*, 3rd edition. Morgan Kaufmann, Waltham, MA.

- Hancke, G.P., de Carvalho e Silva, B. and Hancke Jr., G.P. (2012) 'The role of advanced sensing in smart cities', *Sensors*, 13: 393–425.
- Hannah, M. (1997) 'Space and the structuring of disciplinary power: an interpretative review', *Geografiska Annaler Series B, Human Geography*, 79: 171–80.
- Hannah, M. (2011) *Dark Territory in the Information Age*. Ashgate, Farnham.
- Haque, U. (2012) 'What is a city that it would be "smart"?', Volume #34: City in a Box <http://volumeproject.org/blog/201221/12//volume-34-city-in-a-box/>.
- Haraway, D. (1991) *Simians, Cyborgs and Women: The Reinvention of Nature*. Routledge, New York.
- Harcourt, B.E. (2006) *Against Prediction: Profiling, Policing and Punishing in an Actuarial Age*. Chicago University Press, Chicago.
- Harley, J.B. (1989) 'Deconstructing the map', *Cartographica*, 26: 1–20.
- Harris, D. (2012) 'Researchers mine 2.5m news articles to prove what we already know', *Gigaom*, 26 November 2012, <http://gigaom.com/201226/11//researchers-mine-25-m-news-articles-to-prove-what-we-already-know/> (last accessed 28 January 2013).
- Hart, J.A., Reed, R.R. and Bar, F. (1992) 'The building of the Internet: implications for the future of broadband networks', *Telecommunications Policy*, 16: 666–89.
- Harvey, D. (1972) *Social Justice and the City*. Blackwell, Oxford.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition. Springer, New York.
- Hausenblas, M. (2012) '5 * open data', <http://5stardata.info/> (last accessed 30 January 2013).
- Helbig, N., Cresswell, A.M., Burke, G.B. and Luna-Reyes, L. (2012) *The Dynamics of Opening Government Data: A White Paper*. Centre for Technology in Government, State University of New York, Albany. <http://www.ctg.albany.edu/publications/reports/opendata/opendata.pdf> (last accessed 6 February 2013).
- Helland, P. (2011) 'If you have too much data, then "good enough" is good

- enough', ACM Queue, <https://queue.acm.org/detail.cfm?id=1988603> (last accessed 4 July 2013).
- Herbert, S. (2000) 'For ethnography', *Progress in Human Geography*, 24(4): 550–68.
 - Hewitt, R. (2010) *Map of a Nation: A Biography of the Ordnance Survey*. Granta Books, London.
 - Hey, T., Tansley, S. and Tolle, K. (2009) 'Jim Gray on eScience: A transformed scientific method', in T. Hey, S. Tansley and K. Tolle (eds), *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, Redmond, WA, pp. xvii–xxxi.
 - Hilbert, M. and López, P. (2011) 'The world's technological capacity to store, communicate, and compute information', *Science*, 10(331): 703–05.
 - Hill, D. (2013) 'On the smart city: or, a "manifesto" for smart citizens instead', *City of Sound*, 1 February, <http://www.cityofsound.com/blog/201302//on-the-smart-city-a-call-for-smart-citizens-instead.html> (last accessed 5 February 2013).
 - Hollands, R.G. (2008) 'Will the real smart city please stand up?', *City*, 12(3): 303–20.
 - Hon, W.K., Millard, C. and Walden, I. (2011) *The Problem of 'Personal Data' in Cloud Computing – What Information is Regulated?*, Queen Mary University of London, School of Law Legal Studies Research Paper No. 752011/, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1783577 (last accessed 16 July 2013).
 - Houghton, J. (2011) *Costs and Benefits of Data Provision*. Report to the Australian National Data Service. Centre for Strategic Economic Studies, Victoria University. <http://ands.org.au/resource/houghton-cost-benefit-study.pdf> (last accessed 14 August 2013).
 - Howe, J. (2008) *Crowdsourcing*. Random House, New York.
 - Huijboom, N. and Van der Broek, T. (2011) 'Open data: an international comparison of strategies', *European Journal of ePractice*, 12 (March/April), http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf(last accessed 15 August 2013).
 - IBM (2012) *What is Big Data?*, <http://www-01.ibm.com/software/data/bigdata/> (last accessed 12 November 2012).

- IBM (n.d.) Big Data at the Speed of Business <http://www01.ibm.com/software/data/bigdata/industry.html> (last accessed 9 July 2013).
- Innes, M. (2001) 'Control creep', *Sociological Research Online*, 6(3), <http://www.socresonline.org.uk/63//innes.html> (last accessed 14th January 2014).
- Issenberg, S. (2012) *The Victory Lab: The Secret Science of Winning Campaigns*. New York, Crown.
- InterPARES 2 (2013) Dictionary Definitions. InterPARES 2 Terminology Database: http://www.interpares.org/ip2/ip2_terminology_db.cfm (last accessed 21 October 2013).
- Jacobs, A. (2009) 'The pathologies of big data', *ACM Queue*, 1 July, <http://queue.acm.org/detail.cfm?id=1563874> (last accessed 12 November 2012).
- Janssen, K. (2012) 'Open Government Data: right to information 2.0 or its rollback version?', *ICRI Working Paper*, 82012/, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2152566 (last accessed 14 August 2013).
- Jenkins, T. (2013) 'Don't count on big data for answers', *Scotsman*, 12 February, <http://www.scotsman.com/the-scotsman/opinion/comment/tiffany-jenkins-don-t-count-on-big-data-for-answers-12785890-> (last accessed 11 March 2013).
- Jensen, H.E. (1950) 'Editorial note', in H. Becker (1952), *Through Values to Social Interpretation*. Duke University Press, Durham, NC, pp. vii-xi.
- Jensen, D. (2000) *Data Snooping, Dredging and Fishing: The Dark Side of Data Mining*, <http://www.martinsewell.com/datamining/Jens00.pdf> (last accessed 20 February 2013).
- Johnson, S. (2003) 'Offloading your memories', *The New York Times*, 14 December, <http://www.nytimes.com/200314/12//magazine/14OFFLOADING.html> (last accessed 16 January 2014).
- Johnson, J.A. (2013) *From Open Data to Information Justice*. Paper Presented at the Annual Conference of the Midwest Political Science Association, 13 April, Chicago, Illinois. <http://papers.ssrn.com/abstract=2241092> (last accessed 16 August 2013).
- Kaye, K. (2012) 'Obama's approach to big data: do as I say, not as I do', *AdAge*, 16 November, <http://adage.com/article/digital/obama-s-approach-big-data-i-i/238346/>

(last accessed 19 November 2012).

- Keim, D., Kohlhammer, J., Ellis, G. and Mansmann, F. (2010) *Mastering the Information Age – Solving Problems with Visual Analytics*. Eurographics Association. <http://www.vismaster.eu/book/> (last accessed 16 August 2013).
- Kelling, S., Hochachka, W., Fink, D., Riedewald, M., Caruana, R., Ballard, G. and Hooker, G. (2009) 'Data-intensive science: a new paradigm for biodiversity studies', *BioScience*, 59(7): 613–20.
- King, G. (2011) 'Ensuring the data-rich future of the social sciences', *Science*, 331 (February): 719–21.
- Kinsley, S. (2011) 'Anticipating ubiquitous computing: logics to forecast technological futures', *Geoforum*, 42(2): 31–240.
- Kinsley, S. (2012) 'Futures in the making: practices to anticipate "ubiquitous computing"', *Environment and Planning A*, 44(7): 1554–69.
- Kitchin, R. (1996) 'Methodological convergence in cognitive mapping research: investigating configurational knowledge', *Journal of Environmental Psychology*, 16(3): 163–85.
- Kitchin, R. (1998) *Cyberspace: The World in the Wires*. John Wiley and Sons, Chichester.
- Kitchin, R. (2006) 'Positivistic geography and spatial science', in S. Aitken and G. Valentine (eds), *Approaches in Human Geography*. Sage, London. pp. 20–9.
- Kitchin, R. (2013) 'Big data and human geography: opportunities, challenges and risks', *Dialogues in Human Geography*, 3(3): 262–7.
- Kitchin, R. (2014) 'The real-time city? Big data and smart urbanism', *GeoJournal*, 79 (1): 1–14.
- Kitchin, R. and Dodge, M. (2006) 'Software and the mundane management of air travel', *First Monday*, 11(9), http://firstmonday.org/issues/special11_9/kitchin/ (last accessed 6 March 2014).
- Kitchin, R. and Dodge, M. (2011) *Code/Space: Software and Everyday Life*. MIT Press, Cambridge, MA.

- Kitchin, R. and Fotheringham, A.S. (1997) 'Aggregation issues in cognitive mapping research', *Professional Geographer*, 49(3): 269–80.
- Kitchin, R. and Tate, N. (1999) *Conducting Research in Human Geography: Theory, Methodology and Practice*. Prentice Hall, Harlow.
- Kitchin, R., Bartley, B., Gleeson, J., Cowman, M., Fotheringham, S. and Lloyd, C. (2007) 'Joined-up thinking across the Irish border: making the data more compatible', *Journal of Cross Border Studies*, 2: 22–33.
- Kitchin, R., Gleeson, J. and Dodge, M. (2012a) 'Unfolding mapping practices: a new epistemology for cartography', *Transactions of the Institute of British Geographers*, 38(3): 480–96.
- Kitchin, R., O'Callaghan, C., Boyle, M., Gleeson J. and Keaveney, K. (2012b) 'Placing neoliberalism: the rise and fall of Ireland's Celtic Tiger', *Environment and Planning A*, 44: 1302–26.
- Koops, B.J. (2011) 'Forgetting footprints, shunning shadows: a critical analysis of the "right to be forgotten" in big data practice', *SCRIPTed*, 8(3): 229–56.
- Kourtit, K., Nijkamp, P. and Arribas-Bel, D. (2012) 'Smart cities perspective – a comparative European study by means of self-organizing maps', *Innovation*, 25(2): 229–46.
- Kuhn, T. (1962) *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago.
- Lake, C. (2013) '24 beautifully-designed web dashboards that data geeks will love', 3 June, Econsultancy, <http://econsultancy.com/ie/blog/6284424--beautifully-designed-web-dashboards-that-data-geeks-will-love> (last accessed 12 November 2013).
- Laney, D. (2001) '3D data management: controlling data volume, velocity and variety', Meta Group, <http://blogs.gartner.com/doug-laney/files/201201//ad9493-D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (last accessed 16 January 2013).
- Latour, B. (1989) *Science in Action*. Harvard University Press, Cambridge, MA.
- Lauriault, T.P. (2012) *Data, Infrastructures and Geographical Imaginations: Mapping Data Access Discourses in Canada*. Ph.D. Thesis, Carleton University, Ottawa.

- Lauriault, T.P., Craig, B.L., Taylor, D.R.F. and Pulsifier, P.L. (2007) 'Today's data are part of tomorrow's research: archival issues in the sciences', *Archivaria*, 64: 123–79.
- Lauriault, T.P., Hackett, Y. and Kennedy, E. (2013) *Geospatial Data Preservation Primer*. Ottawa, Hickling, Arthurs and Low.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. and Van Alstyne, M. (2009) 'Computational social science', *Science*, 323: 721–33.
- Lehning, M., Dawes, N., Bavay, M., Parlange, M., Nath, S. and Zhao, F. (2009) 'Instrumenting the Earth: next-generation sensor networks and environmental science', in T. Hey, S. Tansley and K. Tolle (eds), *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research; Redmond, WA. pp. 45–51.
- Lehrer, J. (2010) 'A physicist solves the city', *New York Times*, 17 December, http://www.nytimes.com/2010/12/17/magazine/19Urban_West-t.html (last accessed 23 December 2013).
- Leonelli, S. (2012) 'Introduction: making sense of data-driven research in the biological and biomedical sciences', *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43: 1–3.
- Leyshon, A. and Thrift, N. (1999) 'Lists come alive: electronic systems of knowledge and the rise of credit-scoring in retail banking', *Economy and Society*, 28: 434–66.
- Linehan, T.P. (1991) 'History and development of Irish population censuses', *Journal of the Statistical and Social Inquiry Society of Ireland*, XXVI(IV): 91–125.
- Lohr, S. (2012) 'Sure, big data is great. But so is intuition', *New York Times*, 29 December. <http://www.nytimes.com/2012/30/12/technology/big-data-is-great-but-dont-for-get-intuition.html> (last accessed 3 January 2013).
- Lohr, S. (2013) 'SimCity, for real: measuring an untidy metropolis', *New York Times*, 23 February, <http://www.nytimes.com/2013/24/02/technology/nyucenter-develops-a-science-of-cities.html> (accessed 1 April 2013).
- Longo, J. (2011) '#OpenData: digital-era governance: thoroughbred or new public management Trojan horse?', *PP+G Review*, 2(2), <http://ppgr.files.wordpress.com>.

- com/201105//longo-ostry.pdf (last accessed 16 September 2013).
- Loukides, M. (2010) 'What is data science?', O'Reilly Radar, 2 June, <http://radar.oreilly.com/201006//what-is-data-science.html> (last accessed 28 January 2013).
 - Lucas, P., Ballay, J. and McManus, M. (2012) *Trillions: Thriving in the Emerging Information Ecology*. John Wiley, Hoboken, NJ.
 - Lusk, G. (2013) 'Big. Bad. Big data', The Bubble Chamber, 13 February, <http://thebubblechamber.org/201302//big-bad-big-data/> (last accessed 19 February 2013).
 - Lyon, D. (2002) 'Everyday surveillance: personal data and social classifications', *Information, Communication and Society*, 5: 242-57.
 - Lyon, D. (2003a) 'Surveillance as social sorting: computer codes and mobile bodies', in Lyon, D. (ed.), *Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination*. Routledge, London, pp. 13-30.
 - Lyon, D. (2003b) 'Airports as data filters: converging surveillance systems after September 11th', *Information, Communication and Ethics in Society*, 1: 13-20.
 - Lyon, D. (2007) *Surveillance Studies: An Overview*. Polity, Cambridge.
 - Mann, S., Nolan, J. and Wellman, B. (2003) 'Sousveillance: inventing and using wearable computing devices for data collection in surveillance environments', *Surveillance and Society*, 1(3): 331-55.
 - Manovich, L. (2011) 'Trending: the promises and the challenges of big social data', http://www.manovich.net/DOCS/Manovich_trending_paper.pdf (last accessed 9 November 2012).
 - Manyika, J., Chiu, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Hung Byers, A. (2011) *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute.
 - Marche, S. (2012) 'Literature is not data: against digital humanities', *Los Angeles Review of Books*, 28 October, <http://lareviewofbooks.org/article.php?id=1040&fulltext=1> (last accessed 4 April 2013).
 - Marchetti, N. (2012) 'In Portugal, a smart city from the ground up', *Sustainable*

- Cities Collective, 7 June, <http://sustainablecitiescollective.com/namarchetti/42438/portugal-smart-city-ground> (last accessed 27 May 2013).
- Marz, N. and Warren, J. (2012) *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*. MEAP edition. Manning, Shelter Island, New York.
 - Mattern, S. (2013) 'Methodolatry and the art of measure: the new wave of urban data science', *Design Observer: Places*, 5 November, <http://designobserver.com/places/feature/038174/> (last accessed 15 November 2013).
 - Mayer-Schonberger, V. and Cukier, K. (2013) *Big Data: A Revolution that will Change How We Live, Work and Think*. John Murray, London.
 - McCandless, D. (2010) 'Data, information, knowledge, wisdom', *Information is Beautiful*, 29 November, <http://www.informationisbeautiful.net/2010/data-information-knowledge-wisdom/> (last accessed 24 July 2013).
 - McClean, T. (2011) *Not with a Bang but a Whimper: the Politics of Accountability and Open Data in the UK*. Paper prepared for the American Political Science Association Annual Meeting. Seattle, Washington, 1–4 September 2011. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1899790 (last accessed 19 August 2013).
 - McCreary, D. (2009) 'Entity extraction and the semantic web', *Semantic Web*, 12 January, http://semanticweb.com/entity-extraction-and-the-semantic-web_b10675 (last accessed 19 July 2013).
 - McKeon, S.G. (2013) 'Hacking the hackathon', *Shaunagm.net*, 10 October, <http://www.shaunagm.net/blog/201310/hacking-the-hackathon/> (last accessed 21 October 2013).
 - McNay, L. (1994) *Foucault: A Critical Introduction*. Polity Press, Oxford.
 - Miller, H.J. (2010) 'The data avalanche is here. Shouldn't we be digging?', *Journal of Regional Science*, 50(1): 181–201.
 - Miller, H.J. and Han, J. (2009) 'Geographic data mining and knowledge discovery: an overview', in H.J. Miller and J. Han (eds), *Geographic Data Mining and Knowledge Discovery*. Taylor and Francis, London, pp. 3–32.
 - Miller, P. (2001) 'Governing by numbers: why calculative practices matter', *Social*

- Research, 68(2): 379–96.
- Miller, P. (2010) Linked Data and Government. ePSIplatform Topic Report no. 7, <http://epsi-platform.eu/content/topic-report-no-7-linked-data-and-government-0> (last accessed 14 August 2013).
 - Miller-Rushing, A., Primack, R. and Bonney, R. (2012) 'The history of public participation in ecological research', *Frontier in Ecology and the Environment*, 10(6): 285–90.
 - Mims, C. (2013) 'Coming soon: the cybercrime of things', *The Atlantic*, 6 August, <http://www.theatlantic.com/technology/archive/201308//coming-soon-the-cybercrime-of-things/278409/> (last accessed 7 August 2013).
 - Minelli, M., Chambers, M. and Dhiraj, A. (2013) *Big Data, Big Analytics*. Wiley, Hoboken, NJ.
 - Mitchell, M. (1996) *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA.
 - Mooney, P., Sun, H., Corcoran, P. and Yan, L. (2011) 'Citizen generated spatial data and information: risks and opportunities', *Proceedings of the 2nd International Conference on Network Engineering and Computer Science (ICNECS 2011)*, Xi'an, Shaanxi, China, <http://www.cs.nuim.ie/~pmooney/websitePapers/Mooney-et-all-Submitted-May2011.pdf> (last accessed 28 January 2013).
 - Moore, G.E. (1965) 'Cramming more components onto integrated circuits', *Electronics*, 38(8), http://download.intel.com/museum/Moores_Law/Articles-Press_Releases/Gordon_Moore_1965_Article.pdf (last accessed 20 May 2013).
 - Moretti, F. (2005) *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, London.
 - Morozov, E. (2013) *To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems That Don't Exist*. Allen Lane, New York.
 - Morton, F. (2005) 'Performing ethnography: Irish traditional music sessions and new methodological spaces', *Social and Cultural Geography*, 6: 661–76.
 - Najmi (2004) 'Generations of computer', *Techi/Warehouse*, 13 August, <http://www>.

- techiwarehouse.com/engine/a046ee08/Generations-of-Computer (last accessed 20 May 2013).
- Narayanan, A. and Shmatikov, V. (2010) 'Privacy and security: myths and fallacies of "personally identifiable information"', *Communications of the ACM*, 53(6): 24–6.
 - National Science Foundation (2012) Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA). Programme solicitation NSF 12499-<http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.pdf> (last accessed 25 February 2013).
 - NISO (National Information Standards Organization) (2004) Understanding Metadata. Bethesda, MD, <http://www.niso.org/publications/press/UnderstandingMetadata.pdf> (last accessed 28 June 2013).
 - Nordhaus, W.D. (2002) The Progress of Computing, version 5.2.2, Yale University. http://www.econ.yale.edu/~nordhaus/homepage/prog_083001a.pdf (last accessed 17 January 2014).
 - Northcutt, C. (2012) 'Open data creates value', NYI, 28 August, <http://www.nyi.net/blog/201208//open-data-creates-value/> (last accessed 28 January 2013).
 - OECD (2008) OECD Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information. OECD, <http://www.oecd.org/internet/ieconomy/40826024.pdf> (last accessed 14 August 2013).
 - O'Carroll, A. and Webb, S. (2012) Digital Archiving in Ireland: National Survey of the Humanities and Social Sciences. NUI Maynooth and Digital Repository of Ireland, Dublin.
 - O'Carroll, A., Collins, S., Gallagher, D., Tang, J. and Webb, S. (2013) Caring for Digital Content, Mapping International Approaches. NUI Maynooth, Trinity College Dublin, Royal Irish Academy and Digital Repository of Ireland, Dublin.
 - O'Neill, J. E. (1995) 'The role of ARPA in the development of the ARPANET, 1961–1972', *IEEE Annals of the History of Computing*, 17(4): 76–81.
 - O'Reilly, T. (2005) What is Web 2.0: Design Patterns and Business Models for the

- Next Generation of Software, 30 September, <http://www.oreillynet.com/pub/a/oreilly/tim/news/200530/09//what-is-web-20.html> (last accessed 17 January 2014).
- Ohlhorst, F.J. (2013) *Big Data Analytics: Turning Big Data into Big Money*. Wiley, New York.
 - Open Data Center Alliance (2012) *Big Data Consumer Guide*. Open Data Center Alliance, http://www.opendatacenteralliance.org/docs/Big_Data_Consumer_Guide_Rev1.0.pdf (last accessed 11 February 2013).
 - Open Knowledge Foundation (2012) *Open Data Handbook Documentation, 1.0.0*, <http://opendatahandbook.org/> (last accessed 16 August 2013).
 - Open Scientist (2013) 'The levels of citizen science involvement – part 1', 23 January, <http://www.openscientist.org/201301//the-levels-of-citizen-science.html> (last accessed 28 January 2013).
 - Pentland, A. (2012) 'Reinventing society in the wake of big data', *Edge*, 30 August, <http://www.edge.org/conversation/reinventing-society-in-the-wake-of-big-data> (last accessed 28 January 2013).
 - Pérez-Montoro, M.G. and Díaz Nafría, J.M. (2010) 'Data', in J.M. Díaz Nafría, M.G. Pérez-Montoro and F.S. Alemany (eds), *Glossary of Concepts, Metaphors, Theories and Problems Concerning Information*. León: Universidad de León. <http://glossarium.bitrum.unileon.es/Home/dato/data> (last accessed 13 January 2014).
 - Piatetsky-Shapiro, G. (2012) 'From data mining to big data and beyond', *Inside Analysis*, 18 April, <http://www.insideanalysis.com/201204//data-mining-and-beyond/> (last accessed 20 February 2013).
 - Picton, P.D. (2000) *Neural Networks*, 2nd edition. Palgrave Macmillan, Basingstoke.
 - Plumridge, B. (2012) 'Big data in a big brave world', *Wired*, 21 November, <http://www.wired.co.uk/news/archive/201221/11-/future-of-big-data?> (last accessed 23 November 2012).
 - Pollock, R. (2006) 'The value of the public domain', *IPPR*, <http://www.ippr.org/pub-lication/551526//the-value-of-the-public-domain> (last accessed 13 August 2013).

- Pollock. R. (2009) The Economics of Public Information. Cambridge Working Papers in Economics 0920. <http://www.econ.cam.ac.uk/research/repec/cam/pdf/cwpe0920.pdf>(last accessed 13 August 2013).
- Poovey, M. (1998) A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society. University of Chicago Press, Chicago, IL.
- Porter, T.M. (1995) Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton University Press, Princeton, NJ.
- Porway, J. (2013) 'You can't just hack your way to social change', Harvard Business Review Blog, 7 March, http://blogs.hbr.org/cs/201303//you_cant_just_hack_your_way_to.html (last accessed 9 March 2013).
- Prensky, M. (2009) 'H. sapiens digital: from digital immigrants and digital natives to digital wisdom', Innovate, 5(3), <http://www.innovateonline.info/index.php?view=article&id=705> (last accessed 12 October 2012).
- Quintero, E. (2012) 'The data-driven education movement', Shanker Blog, 22 October, <http://shankerblog.org/?p=7015> (last accessed 12 November 2012).
- Rajaraman, A., Leskovec, J. and Ullman, J.D. (2012) Mining of Massive Datasets. Available at <http://infolab.stanford.edu/~ullman/mmds.html>, also available from Cambridge University Press.
- Raley, R. (2013) 'Dataveillance and countervailance', in L. Gitelman (ed.), 'Raw Data' is an Oxymoron. MIT Press, Cambridge, MA, pp. 121–46.
- Rambam, S. (2008) Privacy is Dead, Get Over It. Presentation at the Last Hope Conference, New York. <http://www.youtube.com/watch?v=Vsxxsrn2Tfs> (last accessed 15 October 2013).
- Ramirez, E. (2013) 'The privacy challenges of big data: a view from the lifeguard's chair', Technology Policy Institute Aspen Forum, 19 August, <http://ftc.gov/speeches/ramirez/130819bigdataaspen.pdf> (last accessed October 11 2013).
- Ramsay, S. (2003) 'Toward an algorithmic criticism', Literary and Linguistic Computing, 18(2): 167–74.

- Ramsay, S. (2010) *Reading Machines: Towards an Algorithmic Criticism*. University of Illinois Press, Champaign, IL.
- Rezendes, C.J. and Stephenson, W.D (2013) 'Cyber security in the internet of things', Harvard Business Review Blog, 21 June, http://blogs.hbr.org/cs/201306//cyber_security_in_the_internet.html (last accessed 7 August 2013).
- Rial, N. (2013) 'The power of big data in Europe', New Europe, 24 May, <http://www.neurope.eu/article/power-big-data-europe> (last accessed 27 May 2013).
- Ribes, D. and Jackson, S.J. (2013) 'Data bite man: the work of sustaining long-term study', in L. Gitelman (ed.), 'Raw Data' is an Oxymoron. MIT Press, Cambridge, MA, pp. 147-66.
- Risen, J. and Lichtblau, E. (2013) 'How the US uses technology to mine more data more quickly', New York Times, 8 June, http://www.nytimes.com/201309/06//us/revelations-give-look-at-spy-agencys-wider-reach.html?hp&_r=3& (last accessed 8 July 2013).
- Ritzer, G. and Jurgenson, N. (2010) 'Production, consumption, prosumption: the nature of capitalism in the age of the digital "prosumer"', *Journal of Consumer Culture*, 10(1): 13-36.
- RLG and OCLC (2002) *Trusted Digital Repositories: Attributes and Responsibilities*. <http://www.oclc.org/research/activities/trustedrep.html> (last accessed 18 February 2013).
- Robinson, S. (2003) *Simulation: The Practice of Model Development and Use*. John Wiley & Sons, Chichester.
- Rogers, S. (2013) 'Twitter's languages of New York mapped', Guardian, 21 February, <http://www.guardian.co.uk/news/datablog/interactive/2013/feb/21/twitter-lan-guages-new-york-mapped> (last accessed 3 April 2013).
- Rooney, B. (2012) 'Big data's big problem: little talent', *Wall Street Journal: Tech Europe*, 26 April, <http://blogs.wsj.com/tech-europe/201226/04//big-datas-big-problem-little-talent/> (last accessed 12 November 2012).
- Rose, A. (2013) 'The internet of things has arrived - and so have massive security issues', *Wired*, 11 January, <http://www.wired.com/opinion/201301//securing-the-internet-of-things/> (last accessed 7 August 2013).

- Rose, N. (1996) *Inventing Our Selves: Psychology, Power and Personhood*. Cambridge University Press, Cambridge.
- Rosenberg, D. (2013) 'Data before the fact', in L. Gitelman (ed.), 'Raw Data' is an Oxymoron. MIT Press, Cambridge, MA, pp. 15–40.
- Rubenking, N.J. (2013) 'Privacy is dead. The NSA killed it. Now what?', PC Mag, <http://www.pcmag.com/article20,2817,2424193,00/.asp> (last accessed 15 October 2013).
- Rubinstein, I.S. (2013) 'Big data: the end of privacy or a new beginning?', International Data Privacy Law, online first, <http://idpl.oxfordjournals.org/content/early/201324/01//idpl.ips036.short> (last accessed 15 July 2013).
- Ruppert, E. (2012) 'The governmental topologies of database devices', *Theory, Culture Society*, 29: 116–36.
- Ruppert, E. (2013) 'Rethinking empirical social sciences', *Dialogues in Human Geography*, 3(3): 268–73.
- Salmon, F. (2014) 'Why the Nate Silvers of the world don't know everything', *Wired*, 7 January, <http://www.wired.com/business/201401//quants-dont-know-everything/> (last accessed 8 January 2014).
- Salus, P. (1995) *Casting the Net: From Arpanet to Internet and Beyond*. Addison Wesley, Reading, MA.
- Sawyer, S. (2008) 'Data wealth, data poverty, science and cyberinfrastructure', *Prometheus: Critical Studies in Innovation*, 26(4): 355–71.
- Schnapp, J. and Presner, P. (2009) *Digital Humanities Manifesto 2.0*. http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf (last accessed 13 March 2013).
- Scherer, M. (2012) 'Inside the secret world of the data crunchers who helped Obama win', *Time: Swampland*, 7 November, <http://swampland.time.com/201207/11//inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win/> (last accessed 19 November 2012).
- Science Coalition (2013) *Sparking Economic Growth: Companies Created from Federally Funded University Research, Fueling American Innovation and Economic*

- Growth, October 2013, <http://www.sciencecoalition.org/reports/Sparking%20Economic%20Growth%20FINAL%201013-21-.pdf> (last accessed 12 January 2014).
- Schwartz, P.M. and Solove, D. (2011) 'The PII problem: privacy and a new concept of personally identifiable information', *New York University Law Review*, 86: 1814–94.
 - Selisker, S. (2012) 'The digital inhumanities?', *Los Angeles Review of Books*, 5 November, <http://lareviewofbooks.org/article.php?type=&id=1146> (last accessed 4 April 2013).
 - Seni, G. and Elder, J. (2010) *Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions*. Morgan & Claypool, San Rafael, CA.
 - Shah, N. (2013) 'Big data, people's lives, and the importance of openness', *DMLcentral*, 24 June, <http://dmlcentral.net/blog/nishant-shah/big-data-peoples-lives-and-importance-openness> (last accessed 25 July 2013).
 - Shah, S., Horne, A. and Capellá, J. (2012) 'Good data won't guarantee good decisions', *Harvard Business Review*, 90(4): 23–5.
 - Shneiderman, B. (1996) 'The eyes have it: a task by data type taxonomy for information visualizations', *Proceedings IEEE Visual Languages*, 96: 336–43.
 - Short, J.E., Bohn, R.E. and Chaitanya, C. (2011) *How Much Information? 2010: Report on Enterprise Server Information*. Global Information Industry Center, UC San Diego. http://hmi.ucsd.edu/pdf/HMI_2010_EnterpriseReport_Jan_2011.pdf (last accessed 1 October 2012).
 - Sicular, S. (2013) 'Big data is falling into the trough of disillusionment', *Gartner*, 22 January, <http://blogs.gartner.com/svetlana-sicular/big-data-is-falling-into-the-trough-of-disillusionment/> (last accessed 26 February 2013).
 - Siegel, E. (2013) *Predictive Analytics*. Wiley, Hoboken, NJ.
 - Sigala, M. (2005) 'Integrating customer relationship management in hotel operations: managerial and operations implications', *International Journal of Hospitality Management*, 24(3): 391–413.
 - Silver, N. (2012) *The Signal and the Noise: The Art and Science of Prediction*. Penguin, London.

- Singer, N. (2012a) 'You for sale: mapping, and sharing, the consumer genome', New York Times, 17 June, <http://www.nytimes.com/2012/17/06/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html> (last accessed 11 October 2013).
- Singer, N. (2012b) 'F.T.C. opens an inquiry into data brokers', New York Times, 18 December, <http://www.nytimes.com/2012/19/12/technology/ftc-opens-an-inquiry-into-data-brokers.html> (last accessed 11 October 2013).
- Singer, N. (2012c) 'Mission control, built for cities: I.B.M. takes 'Smarter Cities' concept to Rio de Janeiro', New York Times, 3 March, <http://www.nytimes.com/2012/04/03/business/ibm-takes-smarter-cities-concept-to-rio-de-janeiro.html> (last accessed 9 May 2013).
- Singh, R. (2012) 'Crowdsourced geospatial data', GIM International, 26(9): 26–31.
- Slee, T. (2012) 'Seeing like a geek', Crooked Timber, 25 June, <http://crookedtimber.org/2012/25/06/seeing-like-a-geek/> (last accessed 18 September 2013).
- Smolan, R. and Erwitte, J. (2012) *The Human Face of Big Data*. Sterling, New York.
- Solove, D.J. (2006) 'A taxonomy of privacy', *University of Pennsylvania Law Review*, 154(3): 477–560.
- Solove, D.J. (2007) "'I've got nothing to hide" and other misunderstandings of privacy', *Social Sciences Research Network*, <http://ssrn.com/abstract=998565> (last accessed 16 July 2013).
- Solove, D. (2013) 'Privacy management and the consent dilemma', *Harvard Law Review*, 126: 1880–903.
- Stalder, F. (2002) 'Privacy is not the antidote to surveillance', *Surveillance and Society*, 1(1): 120–4.
- Star, S.L. and Ruhleder, K. (1996) 'Steps Toward an Ecology of Infrastructure: design and access for large information spaces', *Information Systems Research*, 7(1): 111–34.
- Star, S.L. and Lampland, M. (2009) 'Reckoning with standards', in M. Lampland and S.L. Star (eds) *Standards and Their Stories: How Quantifying, Classifying and Formalizing Practices Shape Everyday Life*. Cornell University Press, Ithaca, NY, pp. 3–24.

- Starr, P. (1987) 'The sociology of official statistics', in W. Alonso and P. Starr (eds), *The Politics of Numbers*. Russell Sage, New York, pp. 7–58.
- Steadman, I. (2013) 'Big data and the death of the theorist', *Wired*, 25 January, <http://www.wired.co.uk/news/archive/201325/01-/big-data-end-of-theory> (last accessed 30 January 2013).
- Strasser, B.J. (2012) 'Data-driven sciences: from wonder cabinets to electronic databases', *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43: 85–7.
- Strasser, C. (2013) 'Closed data ... excuses, excuses', *Data Pub: California Digital Library*, 24 April, <http://datapub.cdlib.org/201324/04//closed-data-excuses-excuses> (last accessed 18 September 2013).
- Strohm, C. and Homan, T. (2013) 'NSA spying row in Congress ushers in debate over big data', *Bloomberg*, 25 July, <http://www.bloomberg.com/news/201325-07-/nsa-spying-row-in-congress-ushers-in-debate-over-big-data.html> (last accessed 25 July 2013).
- Strom, D. (2012) 'Big data makes things better', *Slashdot*, 3 August, <http://slashdot.org/topic/bi/big-data-makes-things-better/> (last accessed 24 October 2013).
- Sword, K. (2008) 'Contribution to: The Promise of Digital History' (roundtable discussion), *Journal of American History*, 95(2): 452–91.
- Taleb, N. (2012) *Antifragile: Things That Gain From Disorder*. Random House, New York.
- Taleb, N. (2013) 'Beware the big errors of "big data"', *Wired*, 2 February, <http://www.wired.com/opinion/201302//big-data-means-big-errors-people/> (last accessed 9 February 2013).
- TechAmerica Foundation (2012) *Demystifying Big Data: A Practical Guide to Transforming the Business of Government*. <http://www.techamericafoundation.org/official-report-of-the-techamerica-foundations-big-data-commission> (last accessed 25 February 2013).
- Tene, O. and Polonetsky, J. (2012) 'Big data for all: privacy and user control in the age

- of analytics', Social Sciences Research Network, <http://ssrn.com/abstract=2149364> (last accessed 15 July 2013).
- Thomas, J.J. and Cook, K.A. (2006) 'A visual analytics agenda', *IEEE Computer Graphics & Applications*, 26: 10–13.
 - Thrift, N. (2004) 'Driving in the city', *Theory, Culture & Society*, 21(459–41) : (5/).
 - Townsend, A. (2013) *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. W.W. Norton & Co, New York.
 - Trumpener, K. (2009) 'Critical response I. Paratext and genre system: a response to Franco Moretti', *Critical Inquiry*, 36(1): 159–71.
 - Vanacek, J. (2012) 'How cloud and big data are impacting the human genome – touch-ing 7 billion lives', *Forbes*, 16 April, <http://www.forbes.com/sites/sap/2012/16/04//how-cloud-and-big-data-are-impacting-the-human-genome-touching-7-billion-lives> (last accessed 27 February 2013).
 - Varian, H. (1996) 'Differential pricing and efficiency', *First Monday*, 1(2–5), <http://firstmonday.org/ojs/index.php/fm/article/view/473394/> (last accessed 8 August 2013).
 - Verwayen, H., Arnoldus, M. and Kaufman, P.B. (2011) *The Problem of the Yellow Milkmaid: A Business Mode Perspective on Open Metadata*. Europeana, <http://pro.europeana.eu/documents/8585662/cbf1f78-e0364088--af2594684-ff90dc5> (last accessed 12 September 2013).
 - Vis, F. (2013) 'A critical reflection on big data: considering APIs, researchers and tools as data makers', *First Monday*, 18(10), <http://firstmonday.org/ojs/index.php/fm/article/view/48783755/> (last accessed 3 January 2013).
 - von Baeyer, H.C. (2003) *Information: The New Language of Science*. Harvard University Press, Cambridge, MA.
 - Wang, T. (2013) 'Big data needs thick data', *Ethnography Matters*, 13 May, <http://ethnographymatters.net/2013/13/05//big-data-needs-thick-data/> (last accessed 3 January 2014).

- Weber, R.H. (2010) 'Internet of things – new security and privacy challenges', *Computer Law & Security Review*, 26: 23–30.
- Webster, T. (2011) 'Social media data dredging', *BrandSavant*, 22 March, <http://brandsavant.com/social-media-data-dredging/> (last accessed 20 February 2013).
- Weinberger, D. (2011) *Too Big to Know*. Basic Books, New York.
- White House (2009) *Open Government Directive*. Executive Office of the President. <http://www.whitehouse.gov/sites/default/files/microsites/ogi-directive.pdf> (last accessed 19 September 2013).
- Williford, C. and Henry, C. (2012) *One Culture: Computationally Intensive Research in the Humanities and Social Sciences*. Council on Library and Information Resources, Washington DC. <http://www.clir.org/pubs/reports/pub151> (last accessed 6 January 2014).
- Wilson, M. (2011) 'Data matter(s): legitimacy, coding, and qualifications-of-life', *Environment and Planning D: Society and Space*, 29: 857–72.
- Wired (2008) 'The petabyte age: because more isn't just more – more is different', 23 June, http://www.wired.com/science/discoveries/magazine/1607-/pb_intro (last accessed 9 November 2012).
- Wittgenstein, L. (1921[1974]) *Tractatus Logico-Philosophicus*, translated by D.F. Pears and B.F. McGuinness. Routledge, London.
- Wu, M. (2012) 'The big data fallacy and why we need to collect even bigger data', *Tech Crunch*, 25 November, <http://techcrunch.com/2012/25/11/the-big-data-fallacy-data-%E289%A0-information-%E289%A0-insights/> (last accessed 28 January 2013).
- Wu, T. (2011) *The Master Switch: The Rise and Fall of Information Empires*. Vintage, New York.
- Wyly, E. (in press) 'Automated (post)positivism', *Urban Geography*.
- Yiu, C. (2012) *A Right to Data: Fulfilling the Promise of Open Public Data in the UK*. Policy Exchange Research Note, <http://www.policyexchange.org.uk/>

publications/category/item/a-right-to-data-fulfilling-the-promise-of-open-public-data-in-the-uk (last accessed 14 August 2013).

- Zelany, M. (1987) 'Management support systems: towards integrated knowledge management', *Human Systems Management*, 7: 59–70.
- Zhang, J., Hsu, M. and Li, L. (2001) 'Image mining: issues, frameworks and techniques', *CiteSeerx*, <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.4.8726> (last accessed 17 January 2014).
- Zikopoulos, P.C., Eaton, C., deRoos, D., Deutsch, T. and Lapis, G. (2012) *Understanding Big Data*. McGraw Hill, New York

المترجم في سطور

د. محمد بن أحمد علي غروي

المؤهل العلمي:

- دكتوراه علوم المعلومات؛ جامعة نيويورك الحكومية، الولايات المتحدة الأمريكية، ٢٠١٢م.

الوظيفة الحالية:

أستاذ تقنية المعلومات المساعد في قطاع تقنية المعلومات بمعهد الإدارة العامة بالرياض.

إنجاز العديد من الدراسات العلمية منها:

1. Gharawi, M., & Khoja, M. (2015). Assessing Basic Computer Applications' Skills of College-Level Students in Saudi Arabia. International Journal of Social, Behavioral, Educational, Economic and Management Engineering Vol: 9, No: 4.
2. Gharawi, M., Estevez, E., Janowski, T (2014). Identifying government chief information officer education and training needs: the case of Saudi Arabia. DG.O 2014: 280- 289
3. Deosthali, K., Gharawi, M. & Potnis, D. (2014). Assessing Employee Satisfaction of the Electronic Training Program of the Gulf Cooperation Council's Countries. In Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (pp. 488- 497).
4. Gharawi, M., & Dawes, S. (2013). Exploring the Influence of Contextual Distances on Transnational Public Sector Knowledge Networks: A Comparative Study of AIRNow-I Shanghai and the Hajj-MDSS Initiatives. Paper presented at the 46th Hawaii International Conference on System Sciences (HICSS-46), Maui, Hawaii.
5. Gharawi, M., & Pardo, T. (2012). Factors Influencing Exchange Activities in Transnational Knowledge Networks: The Case of the Hajj-MDSS. ACM International Conference Proceeding Series, 322, 113- 122.

-
6. Djoko, S., Jong, J., Nam, T. & Gharawi, M. (2012). Building the Academic Community of E-Government Research on Cross-Boundary Information Integration and Sharing. Paper presented at the 45th Hawaii International Conference on System Sciences (HICSS-45), Maui, Hawaii.
 7. Dawes, S., Gharawi, M., & Burke, B. (2012). Transnational Public Sector Knowledge Networks: Knowledge and Information Sharing in a Multidimensional Context. *Government Information Quarterly*, 29(1), S112-S120.
 8. Dawes, S., Gharawi, M., & Burke, B. (2011). Knowledge and Information Sharing in Transnational Knowledge Networks: A Contextual Perspective. Paper presented at the 44th Hawaii International Conference on System Sciences (HICSS-44), Kauai, Hawaii.
 9. Gharawi, M., & Dawes, S. (2010). Conceptualizing Knowledge and Information Sharing in Transnational Knowledge Networks. Paper presented at the 4th International Conference on Theory and Practice of Electronic Governance, Beijing, China.
 10. Rorissa, Demissie D., A., Gharawi, M. (2010). A Descriptive Analysis of Contents of Asian e-Government. In *E-Government Website Development: Future Trends and Strategic Models*. Ed Downey, SUNY College at Brockport: USA.
 11. Rorissa, A., Gharawi, M., Demissie D. (2010). A Tale of Two Continents: Contents of African and Asian E-government Websites. *Proceedings of the 43rd Hawaii International Conference on System Sciences (HICSS-43)*, Koloa, Kauai, Hawaii.
 12. Gharawi, M., Pardo, T., & Guerrero, S. (2009). Issues and Strategies for Conducting Cross-National E- government Comparative Research. *ACM International Conference Proceeding Series*, 322, 163- 170.
 13. Karshmer, A., and Gharawi, M. (2001). Computer Speech for Teaching Mathematics to the Blind, *The International Conference for Education and Training*, Prague, Czech Republic.
 14. Montilva, J., Hamzan, K., and Gharawi, M. (2000). The Watch Model for Developing Business Software in Small and Midsize Organizations, *Proceedings of the SCI*, Volume 1, Orlando, Florida.

أبرز الأنشطة العملية:

١. عضو اللجنة الرئيسية لمبادرة البرنامج الوطني للتدريب عن بُعد، ٢٠١٦م - الآن.
٢. عضو الفريق الوطني لبناء إطار حوكمة التعليم والتدريب الإلكتروني بالمملكة العربية السعودية، ٢٠١٦م.
٣. عضو الفريق الدولي لدراسة احتياج المهارات الوظيفية، صندوق الموارد البشرية بالتعاون مع جامعة هارفرد، ٢٠١٥-٢٠١٧م.
٤. عضو الفريق الوطني لقياس التحول للتعاملات الحكومية الإلكترونية (قياس) ٢٠١٥-الآن.
٥. عضو الفريق الدولي لجائزة الإنجاز في التعاملات الحكومية الإلكترونية (إنجاز) ٢٠١٤م.
٦. عضو الجمعية الدولية للحكومة الرقمية (DG.O) ٢٠١١ - الآن.
٧. مدير مشروع إثراء وتنويع مصادر المعرفة العربية ٢٠١٦-الآن؛ معهد الإدارة العامة، الرياض.
٨. مشرف على برنامج التدريب الإلكتروني ٢٠١٢م - ٢٠١٤م؛ معهد الإدارة العامة، الرياض.
٩. باحث ومدير برامج ٢٠٠٩م - ٢٠١١م؛ مركز أبحاث التقنية الحكومية، نيويورك، الولايات المتحدة الأمريكية.
١٠. تنفيذ عدد من الاستشارات للجهات الحكومية بالمملكة العربية السعودية ٢٠٠١م - ٢٠١٧م.
١١. تصميم وتطوير العديد من البرامج التدريبية في مجال تقنية المعلومات ٢٠٠١م - ٢٠١٧م.
١٢. المشاركة في العديد من المحافل والمؤتمرات المحلية والإقليمية والدولية.

مراجع الترجمة في سطور

د. ماجد بن عبدالعزيز البريثن

المؤهل العلمي:

- الدكتوراة في مجال أنظمة المعلومات والتخطيط الإستراتيجي للمنشآت، جامعة سالفورد بريطانيا (٢٠١١).

الوظيفة الحالية:

- أستاذ مساعد بقسم نظم المعلومات - كلية علوم الحاسب والمعلومات - جامعة الإمام محمد بن سعود الإسلامية.

الخبرات العلمية والعملية:

- عمل عضو هيئة تدريس بقسم تقنية المعلومات والاتصالات بكلية الملك فهد الأمنية.
- عمل محاضراً غير متفرغ بمركز خدمة المجتمع بجامعة الإمام محمد بن سعود.
- الإشراف ومناقشة عدد من رسائل الماجستير.
- حضور ونشر عدد من المشاركات العلمية بعدد من المؤتمرات والمجلات العالمية.
- عمل مستشاراً بهيئة الهلال الأحمر السعودي، ٢٠١٧ - حتى الآن.
- عضو فرق عمل ضمان الجودة لشركة QSR-International فيما يخص ضمان جودة واختبار برمجياتها (NVivo) تطبيق إدارة وتحليل البيانات الوصفية.
- العمل مستشار لأمانة مدينة الرياض في تطوير مشروع مركز الاتصال ٩٤٠.
- العمل مع إدارة تقنية المعلومات بالمؤسسة العامة للتقاعد على تحليل بعض المشاكل القائمة.
- العمل على تطوير أنظمة حاسوبية لشركة القصبي.
- العمل والإشراف على تطوير عدة أنظمة حاسوبية تطبيقية بكلية الملك فهد الأمنية.

-
- القيام بدراسة تقييم تطبيقات أنظمة المعلومات من حيث النجاح والفشل ومن بينها نظام الإسعاف بمدينة لندن وكارثة (LASDS) عام ١٩٩٢م.
 - دراسة تحليلية لتطبيق نظام ERP لإدارة عمل جمعية الهلال الأحمر التركية (TRCS).
 - القيام بإعداد دورة تدريبية خاصة لطلبة الدكتوراة في قسم IRIS بجامعة سالفورد فيما يخص العمل على نظام التحليل الوصفي للبيانات NVivo.
 - القيام بإعداد دورة تدريبية خاصة لطلبة الدكتوراة في قسم IRIS بجامعة سالفورد فيما يخص العمل على نظام الفهرسة الإلكترونية EndNote.

حقوق الطبع والنشر محفوظة لمعهد الإدارة العامة ولا يجوز اقتباس جزء من هذا الكتاب أو إعادة طبعه بأية صورة دون موافقة كتابية من المعهد إلا في حالات الاقتباس القصير بغرض النقد والتحليل، مع وجوب ذكر المصدر.

تم التصميم والإخراج الفني والطباعة في
الإدارة العامة للطباعة والنشر بمعهد الإدارة العامة - ١٤٣٩هـ

هذا الكتاب

يهدف كتاب ثورة البيانات إلى تقديم تحليل مبدئي ودقيق للبيانات وثورة البيانات الجارية حالياً. فالكتاب يسعى، من جهة، لبيان شتى الطرق والوسائل المتعلقة بإعادة تشكيل إنتاج ومعالجة وتحليل وتبادل البيانات، وما يعني هذا لكيفية إنتاج واستخدام المعلومات والمعرفة؛ وهو، من جهة أخرى، يُمهّد للنقاش والتفكير النقدي حول البيانات من حيث طبيعتها، وكيفية تأطيرها من النواحي الفنية والفلسفية والأخلاقية والاقتصادية والتجميعات التقنية والمؤسسية المحيطة بها، ويستند التحليل المقدم في هذا الكتاب إلى مشاركة واسعة النطاق وإلى ما سبقه من مؤلفات في مختلف العلوم، والعلوم الإنسانية، والعلوم الاجتماعية، ومن مطبوعات الثقافة الشعبية، والإصدارات الصحافية، والدوريات الصناعية؛ وإلى خبرة مباشرة في العمل على أنظمة حفظ أرشيفية وبنى ختية ومشاريع تحليلات لبيانات على نطاق مؤسسي واسع. وينقسم الكتاب إلى أحد عشر فصلاً، يقدم الفصل الأول لمحة عامة ونقداً فكرياً حول مفهوم البيانات وكيفية فهم قواعد البيانات والبنى التحتية للبيانات. في حين يبحث الفصل الثاني في الدور المستمر للبيانات الصغيرة والكيفية التي تم بها توسيع نطاقها إلى أرشيفات/محفوظات رقمية وبنى ختية، إلى أن وصلت إلى درجة التداول وبيعها من خلال وسطاء البيانات. كما يناقش الفصل الثالث الدافع للتوجه نحو استحداث البيانات المفتوحة والبيانات المرتبطة التي يجري تقاسمها ومشاركتها وإعادة استخدامها على نطاق أوسع بمرور الوقت. ويبين الفصلان الرابع والخامس طبيعة البيانات الكبيرة بالتفصيل وعوامل التمكين للبيانات الكبيرة ومصادرها المساعدة على انتشارها. ويقدم الفصل السادس لمحة عامة عن مجموعة جديدة من تحليلات البيانات التي ترمي إلى فهم البيانات الصغيرة الموسعة والبيانات الكبيرة. كما يدرس الفصلان السابع والثامن الحجج المستخدمة لتعزيز البيانات الكبيرة والترويج لها وبيان تأثيرها في الحوكمة والأعمال، والسبل التي تقوم بها ثورة البيانات لإعادة تشكيل كيفية تصوّر البحوث وممارستها. ويناقش الفصلان التاسع والعاشر التحديات التقنية والتنظيمية والأخلاقية، والسياسية، والقانونية لثورة البيانات. ويحدد الفصل الأخير بعض الاستنتاجات الكلية ويقدم خارطة طريق لإجراء المزيد من البحث والتأصيل.



9 9 6 0 1 4 2 6 5 5