استخدام البيانات الضخمة في احتساب مؤشرات أهداف التنمية المستدامة في المنطقة العربية

التحديات والفرص

الأمم المتحدة
الاسكوا
ESCWA

ازدهارُ البلدان كرامةُ الإنسان

الأمم المتحدة
الاسكوا
ESCWA

ازدهارُ البلدان **كرامةُ الإنسان**

رؤيتنا

طاقاتٌ وابتكار، ومنطقتُنا استقرارٌ وعدلٌ وازدهار

رسالتنا

بشغَف وعزْم وعَمَل: نبتكِر، ننتج المعرفة، نقدِّمُ المشورة،
نبني التوافق، نواكب المنطقة العربية على مسار خطة عام 2030.

يداً بيد، نبني غداً مشرقاً لكلِّ إنسان.

E/ESCWA/CL4.SIT/2021/TP.2

اللجنة الاقتصادية والاجتماعية لغربي آسيا

استخدام البيانات الضخمة في احتساب مؤشرات أهداف التنمية المستدامة في المنطقة العربية

التحديات والفرص

الأمم المتحدة
بيروت

© 2021 الأمم المتحدة
حقوق الطبع محفوظة

تقتضي إعادة طبع أو تصوير مقتطفات من هذه المطبوعة الإشارة الكاملة إلى المصدر.

توجّه جميع الطلبات المتعلقة بالحقوق والأذون إلى اللجنة الاقتصادية والاجتماعية لغربي آسيا (الإسكوا)، البريد الإلكتروني: publications-escwa@un.org.

المؤلف: ماريا سيمونا أندريانو وجيوفاني سافيو.

النتائج والتفسيرات والاستنتاجات الواردة في هذه المطبوعة هي للمؤلفين، ولا تمثل بالضرورة الأمم المتحدة أو موظفيها أو الدول الأعضاء فيها، ولا ترتب أي مسؤولية عليها.

ليس في التسميات المستخدمة في هذه المطبوعة، ولا في طريقة عرض مادتها، ما يتضمن التعبير عن أي رأي كان من جانب الأمم المتحدة بشأن المركز القانوني لأي بلد أو إقليم أو مدينة أو منطقة أو لسلطات أي منها، أو بشأن تعيين حدودها أو تخومها.

الهدف من الروابط الإلكترونية الواردة في هذه المطبوعة تسهيل وصول القارئ إلى المعلومات وهي صحيحة في وقت استخدامها. لا تتحمل الأمم المتحدة أي مسؤولية عن دقة هذه المعلومات مع مرور الوقت أو عن مضمون أي من المواقع الإلكترونية الخارجية المشار إليها.

جرى تدقيق المراجع حيثما أمكن.

لا يعني ذكر أسماء شركات أو منتجات تجارية أن الأمم المتحدة تدعمها.

المقصود بالدولار دولار الولايات المتحدة الأمريكية ما لم يُذكر غير ذلك.

تتألف رموز وثائق الأمم المتحدة من حروف وأرقام باللغة الإنكليزية. والمقصود بذكر أي من هذه الرموز الإشارة إلى وثيقة من وثائق الأمم المتحدة.

مطبوعات للأمم المتحدة تصدر عن الإسكوا، بيت الأمم المتحدة، ساحة رياض الصلح،
صندوق بريد: 8575-11، بيروت، لبنان.

الموقع الإلكتروني: www.unescwa.org.

21-00217

شكر وتقدير

أعدت هذا التقرير ماريا سيمونا أندريانو (جامعة مركاتوروم في روما) وجيوفاني سافيو، بتنسيق من إسماعيل لبّد، خبير الإحصاء في شعبة الإحصاء في الإسكوا. يود المؤلفان أن يعربا عن امتنانهما لحيدر فريحات، كبير مستشاري التكنولوجيا في الإسكوا، وفؤاد مراد، مدير البرامج الأول في الإسكوا، والمشاركين في ورشة العمل الإقليمية حول تكامل البيانات الضخمة والمعلومات الجغرافية المكانية لتجميع مؤشرات أهداف التنمية المستدامة في البلدان العربية، من 13-15 تشرين الأول/أكتوبر 2020، على تعليقاتهم على العرض الذي قُدم أثناءها.

كما يشكر المؤلفان زينة سنو وندى مدلل ومحمد حصري من شعبة الإحصاء في الإسكوا على دعمهم.

موجز تنفيذي

بعد اعتماد الجمعية العامة للأمم المتحدة في أيلول/سبتمبر 2015 لأهداف التنمية المستدامة، يطلب من الأجهزة الإحصائية الوطنية القيام بثورة في البيانات، ما يشكل تحدياً كبيراً لمهامها: تأهيل مصفوفات أهداف التنمية المستدامة بجميع أبعادها الاقتصادية والاجتماعية والبيئية.

تؤدي الحاجة إلى زيادة التغطية ودقة التوقيت وتصنيف المؤشرات لتناسب متطلبات بيانات أهداف التنمية المستدامة، إلى بذل جهد كبير جداً – خاصة بالنسبة للبلدان الأقل نمواً إحصائياً في جميع أنحاء العالم – لإعادة النظر في طُرق عمل المنظمات الإحصائية والتخطيط للمستقبل.

في الآونة الأخيرة، أضافت أزمة كوفيد-19 تحدياً آخر. إذ يعطل الوباء العمليات الروتينية لجمع الإحصاءات الأساسية وتجميعها، ما يعرض التعدادات والمسوحات وبرامج البيانات الأخرى المخطط لها للإلغاء أو تأجيل التنفيذ.

يحلل هذا التقرير الفنّي كيف يمكن لمصادر المعلومات غير التقليدية، ولا سيما تلك التي قدمتها ثورة البيانات الضخمة، أن تفيد الإحصائيين الرسميين في سد الثغرات الموجودة وتقديم رؤى قيمة بشأن رصد التنمية المستدامة والإبلاغ عنها.

يقدم هذا التقرير تفاصيل عن تعريفات البيانات الضخمة، ويستعرض المشاريع الجارية والابتكارية باستخدام مصادر البيانات الضخمة للإحصاءات الرسمية، ويناقش إيجابيات وسلبيات استخدام هذه السلالة الجديدة من مصادر البيانات الإحصائية، ويحلل الأمثلة الملموسة لاستخدام معلومات الاستشعار عن بعد للحصول على رؤى حول المؤشرات الإحصائية البارزة مثل معدلات الفقر والنشاط الاقتصادي والأسعار، كما يقدم أمثلة على كيفية استخدام "المعلومات أعلاه" للحصول على معلومات عن آثار الوباء الأخير.

يخلُص التقرير إلى أن للبيانات الضخمة إمكانية جمع المعلومات والبيانات حول عدد من مجالات التركيز، بما في ذلك التنقل، والنقل، والسياحة، والأسعار، والفساد والجريمة، واستهلاك الطاقة، والكثافة السكانية، واستخدام الأراضي، والرفاه، وسوق العمل. رغم أن استخدامها يشكل تحديات أخلاقية وقانونية وفنية وتحديات تتعلّق بالسمعة بالنسبة للإحصائيين، إنما يمكن أن يحسن توقيتها ودقتها، وأن يصل إلى تفاصيل أكبر، ويزيد من قدرات التصنيف، ويسد الثغرات في البيانات الرسمية للعديد من مؤشرات أهداف التنمية المستدامة بحلول عام 2030.

رسائل رئيسية

- *بعد اعتماد الجمعية العامة للأمم المتحدة في أيلول/سبتمبر 2015 لأهداف التنمية المستدامة، يطلب من مكاتب الإحصاء الوطنية القيام بثورة في البيانات، ما يشكّل تحدياً هائلاً أمام تنفيذ مهامها: تأهيل مصفوفات أهداف التنمية المستدامة بجميع أبعادها الاقتصادية والاجتماعية والبيئية.*

- *في الآونة الأخيرة، أضافت أزمة كوفيد-19 تحديا آخر، إذ يعطل الوباء العمليات الروتينية لجمع الإحصاءات الأساسية وتجميعها، مما يؤخّر التعدادات والمسوحات وبرامج البيانات الأخرى المخطط لها.*

- *يمكن لمصادر المعلومات غير التقليدية، ولا سيما تلك التي قدمتها ثورة البيانات الضخمة، أن تفيد الإحصائيين الرسميين في سد الثغرات الموجودة وتقديم رؤى قيمة بشأن رصد التنمية المستدامة والإبلاغ عنها.*

- *يقدم هذا التقرير تفاصيل عن تعريفات البيانات الضخمة، ويستعرض المشاريع الجارية والابتكارية باستخدام مصادر البيانات الضخمة للإحصاءات الرسمية، ويناقش إيجابيات وسلبيات استخدامها.*

- *يخلُص التقرير إلى أن للبيانات الضخمة إمكانات كبيرة في المساعدة على جمع البيانات حول عدد من الظواهر. على الرغم من أن استخدامها يشكل تحديات أخلاقية وقانونية وفنية وتحديات تتعلّق بالسمعة بالنسبة للإحصائيين، إنما يمكنه أن يحسن توقيتها ودقتها، وأن يسد الثغرات في البيانات الرسمية للعديد من مؤشرات أهداف التنمية المستدامة الاقتصادية والاجتماعية والبيئية.*

المحتويات

الصفحة

شكر وتقدير iii
موجز تنفيذي v
رسائل رئيسية vii
مقدمة 1
1. تعريفات البيانات الضخمة 3
2. مشروع يتعلق بالبيانات الضخمة قيد التنفيذ 8
3. الاستشعار عن بُعد والهدفان 1 و8 15
4. تطبيقات أخرى: أجهزة الاستشعار 20
5. التحديات والفرص 22
6. الخلاصة 24
المراجع 25

قائمة الجداول

الجدول 1. خصائص البيانات الضخمة 5
الجدول 2. مخزون مشروع البيانات الضخمة 13
الجدول 3. المشاريع الابتكارية، التقرير السنوي لمبادرة النبض العالمي للأمم المتحدة 2019 14

قائمة الأشكال

الشكل 1. خصائص البيانات الضخمة الـ Vs الأربعة (IBM Big Data & Analytics Hub) 4
الشكل 2. رسم بياني راداري مع تقييم (نظري) لمختلف Vs لمصدر البيانات الضخمة، أي السرعة المنخفضة، والمصداقية والحجم شديدي الارتفاع، والتقلبات المتوسطة والتصور 5
الشكل 3. تغطية البيانات حسب أهداف التنمية المستدامة 9
الشكل 4. آنية البيانات حسب أهداف التنمية المستدامة 9
الشكل 5. تأثير كوفيد-19 على الأنشطة الإحصائية 10
الشكل 6. كيف يعمل الاستشعار عن بُعد بشكل عملي 15
الشكل 7. أضواء في الليل كما تم التقاطها بواسطة قمر صناعي تابع للإدارة الوطنية للملاحة الجوية والفضاء (وكالة ناسا) 16
الشكل 8. أمثلة لتقدير فجوات الفقر عند 5.50 دولار في اليوم لبلدان أمريكا اللاتينية ومنطقة البحر الكاريبي، باستخدام أضواء الليل 17
الشكل 9. أمثلة على تقدير معدل الفقر المدقع باستخدام الأضواء الليلية لبلديات سانتياغو، شيلي، في عام 2015 18
الشكل 10. أضواء الليل قبل وأثناء وبعد تأثير الوباء في بيجين، الصين 19

مقدمة

بعد اعتماد الجمعية العامة للأمم المتحدة في أيلول/سبتمبر 2015 أهداف التنمية المستدامة، طُلب من الأجهزة الإحصائية الوطنية (NSOs) والنُظُم الإحصائية الوطنية (NSSs) في جميع أنحاء العالم، وكذلك مكاتب الإحصاء في المنظمات الدولية (IOs)، القيام بثورة في البيانات. في الواقع، يجب على المتخصصين في الإحصاء توسيع نطاق وتفصيل البيانات المنتجة تقليدياً، وأن يقيسوا الظواهر الاقتصادية والاجتماعية والبيئية الجديدة بحلول عام 2030، دون إغفال أحد.

لتأهيل مصفوفات أهداف التنمية المستدامة بجميع أبعادها، تواجه الأجهزة الإحصائية الوطنية والنُظُم الإحصائية الوطنية تحدياً كبيراً. لا تطرح تغطية المؤشرات مشاكل خطيرة فحسب، بل ينطوي التصنيف بجميع أشكاله (الزمني والجغرافي وحسب الجنس والدخل والعمر وما إلى ذلك) على متطلبات حازمة للبيانات، مما يجبر المنظمات الإحصائية على إعادة النظر في كيفية عملها وخططها المستقبلية.

في الآونة الأخيرة، أضافت أزمة كوفيد-19 تحدياً آخر، إذ يعطل الوباء العمليات الروتينية التي تجريها الأجهزة الإحصائية الوطنية والنظام الإحصائي الوطني لجمع الإحصاءات الأساسية، مما يؤخر التعدادات والمسوحات وبرامج البيانات الأخرى المخطط لها.

بسبب هذه التحديات الهائلة، هناك توافق متزايد في الآراء بين أصحاب المصلحة على أن البيانات الضخمة، أيا كان شكلها، قد تعزز قيمة مصادر البيانات والإحصاءات التقليدية في رصد الرفاه المستدام، وتيسر خطة التحول الذي ينبغي للمنظمات أن تنفذها في السنوات المقبلة.

مع ذلك، احتلّت البيانات الضخمة أولى اهتمام الإحصاءات الرسمية خلال السنوات الخمس إلى العشر الماضية، وأثار الكثيرون اعتبارات حول قيمتها المضافة الحقيقية والفعالة بالنسبة للإحصاءات الرسمية، مع تردد العديد من المنظمات الوطنية والدولية في استخدامها. استناداً إلى تلك الآراء، يجب أن تكون البيانات الضخمة قد اجتازت الآن مرحلة التوقعات المبالغ بها ووصلت إلى منحنى تنازلي نحو الاندثار.

يهدف هذا التقرير الفني إلى تحديد نطاق مفهوم البيانات الضخمة في الإحصاءات وتحديد إيجابيات وسلبيات ومزايا وعيوب استخدام البيانات الضخمة للإحصاءات الرسمية. كما يستعرض النقاط الرئيسية للمناقشة، في محاولة لمعرفة الدور الذي يمكن أن تلعبه البيانات الضخمة للإحصاءات في المستقبل القريب، مع التركيز بشكل خاص على المنطقة العربية. يستمد التقرير مواد من عرض تقديمي أُلقي خلال اجتماع افتراضي نظمته لجنة الأمم المتحدة الاقتصادية والاجتماعية لغربي آسيا بعنوان "ورشة عمل إقليمية حول تكامل البيانات الضخمة والمعلومات الجغرافية المكانية لتجميع مؤشرات أهداف التنمية المستدامة في البلدان العربية"، في الفترة من 13 إلى 15 تشرين الأول/أكتوبر 2020.

بعد الملخص التنفيذي وهذه المقدمة، يناقش التقرير التحديات الفعلية التي تواجهها منظمات الإحصاءات في تجميع مؤشرات أهداف التنمية المستدامة ويحلل

تعريفات البيانات الضخمة في القسم 2. ويتضمن القسم 3 مراجعة للمشاريع الفعلية الجارية في البيانات الضخمة للإحصاءات الرسمية، في حين يستعرض القسمان 4 و5 أمثلة على استخدام مصادر البيانات الضخمة للاستشعار عن بعد لرسم خرائط مؤشرات الفقر، ونظرة عامة موجزة عن مصادر الاستشعار ذات التطبيقات الملموسة لهذه الأرقام. ويتضمن القسم 6 مناقشة حول التحديات والفرص المتعلقة بالبيانات الضخمة في تجميع أهداف التنمية المستدامة. ويرد في القسم الأخير من التقرير موجز للنقاط الرئيسية.

1. تعريفات البيانات الضخمة

أصبحت البيانات الضخمة نتاجاً ثانوياً للرقمنة المتزايدة لحياتنا المعاصرة، وهي ظاهرة من المرجح أن تستمر لسنوات عديدة قادمة. لا يوجد تعريف مقبول بصورة موحدة للبيانات الضخمة. مع ذلك، من المهم تقديم تعريف من خلال مراعاة كامل النظام البيئي الذي ينتجها ويستخدمها. في الواقع، لا تعني البيانات الضخمة "الكثير من البيانات"، وعلى الرغم من الاسم، فلا يشكّل الحجم السمة المحددة الوحيدة، إذ يجب أن تكون مصحوبة بالإشارة إلى ميزاتها الأخرى، ما يسمى بـ "Vs".

نحن نفضّل تعريف البيانات الضخمة الذي تقدمه مؤسسة TechAmerica، على الرغم من أنه تعريف عام إلى حد ما. وينص على ما يلي: "البيانات الضخمة هي مصطلح يصف كميات كبيرة من البيانات عالية السرعة والمعقدة والمتغيرة التي تتطلب أساليب وتقنيات متقدمة لتمكين استيعاب المعلومات وتخزينها وتوزيعها وإدارتها وتحليلها".

تتميز البيانات الضخمة بشكل عام بما يسمى بالـ Vs الأربعة، وهي السرعة والحجم والمصداقية والتنوع، كما هو موضح في الرسوم البيانية لـ IBM Big Data & Analytics Hub (الشكل 1).

يشير بعض المؤلفين أيضاً إلى Vs أخرى ذات الصلة من أجل وصف البيانات الضخمة بشكل كامل. ومن بين هذه الخصائص: قيمتها (المعلومات والرؤى التي توفرها البيانات الضخمة)، وقابليتها للاستمرار (التقييم السريع والفعال من حيث التكلفة المتعلقة بمتغير معين)، والتباين (بسبب التعريفات المتغيرة، والمخالفات في البيانات، ووجود العديد من أبعاد البيانات الناتجة عن أنواع ومصادر بيانات متباينة متعددة)، وأسلوب العرض (طريقة عرض البيانات بطريقة قابلة للقراءة ويمكن النفاذ إليها). يحاول الشكل 2 أدناه تصوير حالة افتراضية لمصدر بيانات كبيرة عام من حيث تقييم خصائص V's المختلفة التي تميز المصدر.

تم تلخيص معاني الـ Vs التقليدية الأربعة في الجدول 1، المستمد من مؤسسة TechAmerica Foundation (2012). كما أشير بشكل صحيح في Manske, Sangokoya, Pestre and Letouzé (2016)، لا تشير البيانات الضخمة فقط إلى البيانات، ولكن أيضاً إلى النظام البيئي بأكمله الذي ينتجها ويستخدمها. وهذا يؤدي إلى تعريف C's الثلاثة للبيانات الضخمة، والتي تتميز باتحاد فتات البيانات الضخمة (نوع جديد من البيانات الناتجة بشكل سلبي)، والقدرة (باعتبارها قدرة تقنية وبشرية على إنتاج رؤى من هذه البيانات)، والمجتمع (جهات فاعلة جديدة من القطاع الخاص ومجتمع البحث، على سبيل المثال).

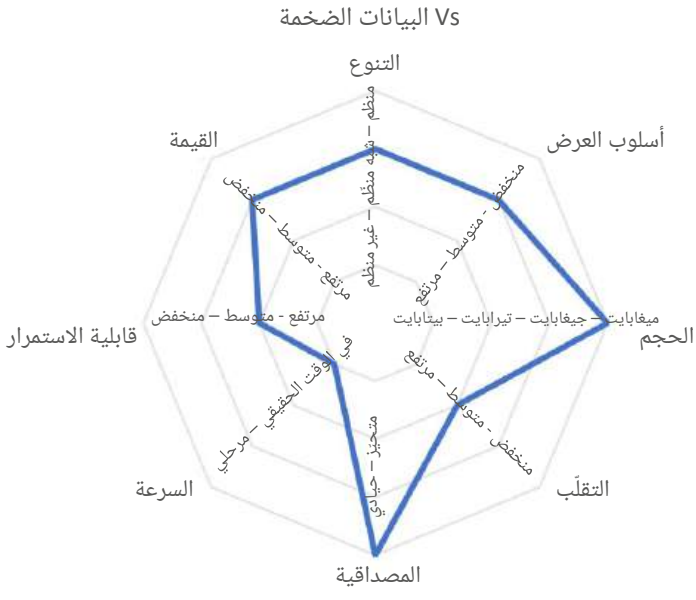
يشير التقلب إلى "التقنية المتغيرة أو بيئات الأعمال التي يتم فيها إنتاج البيانات الضخمة، والتي يمكن أن تؤدي إلى تحليلات ونتائج غير صحيحة، وكذلك إلى هشاشة البيانات الضخمة كمصدر للبيانات" (Hammer and others, 2017, p. 8). للوهلة الأولى، قد تبدو Vs الإضافية غريبة

لأنها ليست في حد ذاتها خصائص مميزة للبيانات أو جوهرية بالنسبة لها. مع ذلك، تعتبر التقلبات والمصداقية من الإضافات الهامّة جداً لفهم المساهمة التي يمكن أن تقدمها البيانات الضخمة في تجميع الإحصاءات ومؤشرات أهداف التنمية المستدامة. يقدم A 6-V وهو التعريف الذي يتضمن "القيمة"، بمعنى الإفادة المستمدة من البيانات، تعريفاً رئيسياً، لأنه يقدم مفهوم التكلفة والفائدة، بمعنى وجوب موازنة تكاليف الاستثمار في البيانات الضخمة بعناية مقابل ما يمكن أن تقدمه هذه البيانات من الناحية العملية (الشكل 1). وعلى غرار التقلب والمصداقية، لا تعتبر القيمة سمة جوهرية، ولكن، كما ذكر أعلاه، تعتبر إضافة هذا البعد، مفيدة مع ذلك.

الشكل 1. خصائص البيانات الضخمة الـ Vs الأربعة (IBM Big Data & Analytics Hub)

المصدر: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS.

الشكل 2. رسم بياني راداري مع تقييم (نظري) لمختلف Vs لمصدر البيانات الضخمة، أي السرعة المنخفضة، والمصداقية والحجم شديدي الارتفاع، والتقلبات المتوسطة والتصور



الجدول 1. خصائص البيانات الضخمة

الخاصية	التوصيف	الصفة	الدافع
الحجم	الكمية الهائلة من البيانات التي تم إنشاؤها أو كثافة البيانات التي يجب استيعابها وتحليلها وإدارتها لاتخاذ القرارات بناءً على التحليل الكامل للبيانات.	ووفقاً لـ Digital Universe Study (دراسة الكون الرقمي) التي أجرتها مؤسسة البيانات الدولية (IDC)، "الكون الرقمي" في العالم هو بصدد توليد 1.8 زيتابايت من المعلومات – مع استمرار النمو المطّرد – من المتوقع أن تبلغ 35 زيتابايت في عام 2020.	زيادة في مصادر البيانات، وأجهزة الاستشعار ذات دقة عالية.
السرعة	مدى سرعة إنتاج البيانات وتغييرها والسرعة التي يجب تلقي البيانات بها وفهمها ومعالجتها.	• النفاذية: المعلومات متى يريدها المستخدم وأين، وبأي طريقة، عند نقطة التأثير؛ • قابلية التطبيق: تتحول المعلومات القيمة ذات الصلة لمؤسسة ما، بوتيرة غزيرة، إلى ظاهرة في الوقت الحقيقي؛	• زيادة مصادر البيانات؛ • تحسين توصيل الناتج؛ • قوة حوسبة محسنة لأجهزة توليد البيانات.

الخاصية	التوصيف	الصفة	الدافع
		• قيمة الوقت: ينتج عن التحليل في الوقت الحقيقي قرارات محسّنة تستند إلى البيانات.	
التنوع	يؤدي ظهور المعلومات الواردة من مصادر جديدة، داخل وخارج جدران المؤسسة أو المنظمة، إلى خلق ضغوط تكاملية وإدارية وحوكمية ومعمارية على تكنولوجيا المعلومات.	• المُنظَم: 15 في المائة من البيانات اليوم هي مُنظمة، صفوف وأعمدة؛ • غير المنظم: 85 في المائة هي معلومات غير مُنظمة أو معلومات قام أشخاص بإنشائها؛ • شبه مُنظم: أصبح الجمع بين البيانات المنظمة وغير المنظمة أمراً بالغ الأهمية؛ • التعقيد: حيث تتحرك مصادر البيانات وتستقر.	• الخليوي؛ • وسائل التواصل الاجتماعي؛ • أشرطة الفيديو؛ • الدردشة؛ • علم الجينوميات؛ • اجهزه الاستشعار.
المصداقية	جودة البيانات المستلمة ومصدرها.	قد تكون جودة البيانات الضخمة جيدة أو سيئة أو غير محددة بسبب عدم اتساق البيانات وعدم اكتمالها، والغموض، والكمون، والتضليل، وتقديرات النموذج.	تتطلب القرارات المستندة إلى البيانات إمكانية التتبع والتبرير.

المصدر: TechAmerica Foundation, 2012.

تصنف أنواع البيانات الضخمة وفقاً لتعريف يستند في معظمه إلى مصادر البيانات (المجلس الاقتصادي والاجتماعي للأمم المتحدة، 2013)، على النحو التالي:

- مصادر البيانات والمعلومات الناشئة عن إدارة برنامج ما، سواء كان حكومياً أم لا، مثل السجلات الطبية الإلكترونية، وزيارات المستشفيات، وسجلات التأمين، والسجلات المصرفية، وبنوك الطعام؛
- مصادر تجارية أو معاملات ناشئة عن المعاملة بين كيانين، على سبيل المثال، معاملات بطاقات الائتمان والمعاملات عبر الإنترنت (بما في ذلك عبر الأجهزة المحمولة)؛
- مصادر شبكة الاستشعار، مثل بيانات الكاميرا، والتصوير عن طريق الأقمار الصناعية، وأجهزة استشعار الطرق، وأجهزة الاستشعار المناخية، مثل تلك المتعلقة بمصادر بيانات الاستشعار عن بعد؛
- تتبع مصادر الأجهزة، مثل تتبع البيانات من الهواتف المحمولة والنظام العالمي لتحديد المواقع (GPS)؛
- مصادر البيانات السلوكية، مثلاً، عمليات البحث عبر الإنترنت (حول منتج أو خدمة

أو أي نوع آخر من المعلومات)، والمشاهدات عبر الإنترنت؛

- مصادر بيانات الرأي، مثال التعليقات على وسائل التواصل الاجتماعي؛
- بيانات نظام المعلومات الجغرافية (GIS) ومعلومات من مصادر وأنواع مختلفة.

يجب عدم تصنيف البيانات الإدارية، التي تنظمها الإدارات العامة عادة بطريقة منظمة (يشير إليها أحياناً الإحصائيون على أنها سجلات إدارية)، ضمن البيانات الضخمة، لكن يمكن أن تصبح كذلك في حال زادت خصائص السرعة والحجم، بالشكل المناسب في المستقبل.

2. مشروع يتعلق بالبيانات الضخمة قيد التنفيذ

رغم إحراز تقدم جيد في زيادة توافر البيانات القابلة للمقارنة على المستوى الدولي لأهداف التنمية المستدامة، لا تزال هناك فجوات هامة في البيانات من حيث التغطية الجغرافية والتوقيت المناسب ومستوى التصنيف.

يكشف تحليل للمؤشرات المعترف بها في قاعدة البيانات العالمية لمؤشرات أهداف التنمية المستدامة، المتاح على الرابط https://unstats.un.org/sdgs/indicators/database أن لدى أقل من نصف البلدان أو المناطق الـ 194 بيانات قابلة للمقارنة على المستوى الدولي بالنسبة لأربعة من أهداف التنمية المستدامة السبعة عشر (الشكل 3). وهذا أمر مقلق بشكل خاص بالنسبة للهدف 5 (المتعلق بالمساواة بين الجنسين)، وكذلك الأهداف 12 و13 و14 (الاستهلاك والإنتاج المسؤولان، والعمل المناخي، والحياة تحت الماء، على التوالي). الأهم من ذلك أن البلدان التي تتوافر لديها بيانات لا تتوفر لديها سوى قليل من عمليات الرصد مع مرور الوقت، ما يعسّر على واضعي السياسات رصد التقدم وتحديد الديناميات.

يتوفر عدد من أهداف التنمية المستدامة بفارق زمني طويل (الشكل 3). بالنسبة لما لا يقل عن 50 في المائة من البلدان أو المناطق في قاعدة البيانات، تعود أحدث البيانات المتاحة لمؤشرات الفقر (الهدف 1) لعام 2016 أو قبل ذلك. وتنطبق الحالة ذاتها بالنسبة للمؤشرات المتعلقة بالمساواة بين الجنسين (الهدف 5)، والمدن المستدامة (الهدف 11)، والسلام والعدل والمؤسسات القوية (الهدف 16). ويعني الاستمرار على هذا النحو أن المعلومات المتعلقة بأهداف التنمية المستدامة لعام 2030، لن تكون متاحة إلا بعد أربع سنوات من تحليل المقصد، أي في عام 2034.

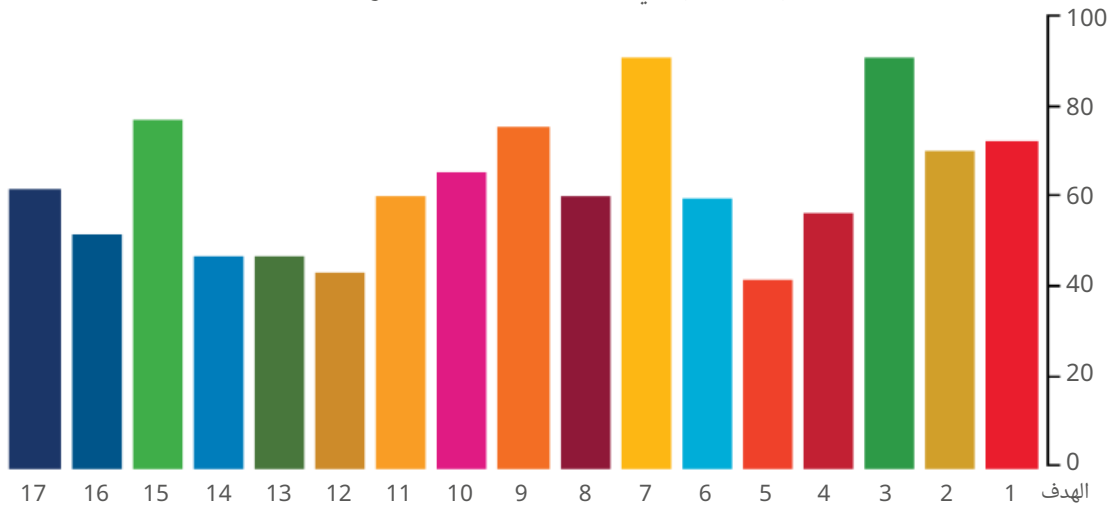
بينما تحاول الحكومات احتواء انتشار كوفيد-19، يتم تعطيل عمليات جمع البيانات المنتظمة. ما يهدد قدرة العديد من الأجهزة الإحصائية الوطنية على تقديم إحصاءات شهرية وفصلية رسمية، فضلاً عن البيانات اللازمة لرصد التقدم المحرز في أهداف التنمية المستدامة (الشكل 4).

يبين مسح أجرته الأمم المتحدة والبنك الدولي مؤخراً (حصد ردوداً من 122 بلداً) تأثير الوباء على عمليات عدد من الأجهزة الإحصائية الوطنية: أغلقت 65 في المائة من مقراتها الإدارية جزئياً أو كلياً، وأعطى 90 في المائة منها تعليمات للموظفين بالعمل من منازلهم، وتوقفت 96 في المائة منها جزئياً أو كلياً عن جمع البيانات وجهاً لوجه. في شمال أفريقيا وغرب آسيا، أشار 62 في المائة من البلدان التي شملها المسح إلى أن إنتاج الإحصاءات المنتظمة، بما فيها البيانات الشهرية والفصلية، فضلاً عن سجلات الأعمال التجارية، قد تأثر، وأشارت النسبة نفسها إلى أنها تواجه صعوبة في تلبية الالتزامات الدولية للإبلاغ عن البيانات.

ووفقاً لنتائج المسح، شهدت تسع من أصل عشرة أجهزة إحصائية وطنية في البلدان المنخفضة الدخل والبلدان ذات الدخل المتوسط الأدنى انخفاضاً في التمويل وهي تكافح من أجل الحفاظ على عملياتها العادية خلال الجائحة.

الشكل 3. تغطية البيانات حسب أهداف التنمية المستدامة

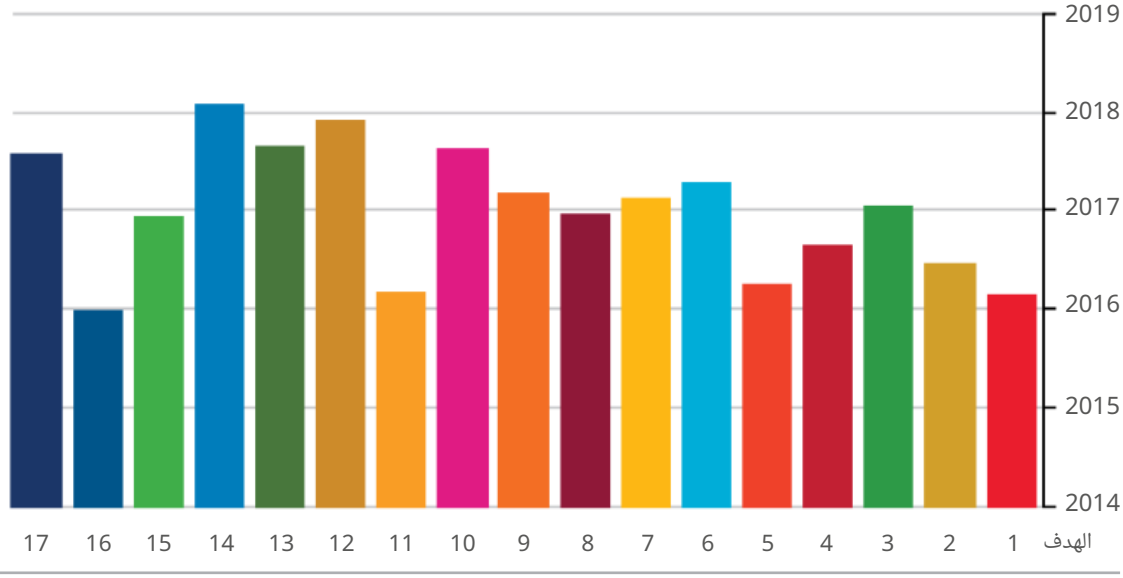
تغطية البيانات: نسبة البلدان أو المناطق التي تملك بيانات (المتوسط المرجح للمؤشرات) حسب الهدف (بالنسبة المئوية)



المصدر: تقرير أهداف التنمية المستدامة لعام 2020. https://unstats.un.org/sdgs/report/2020/.

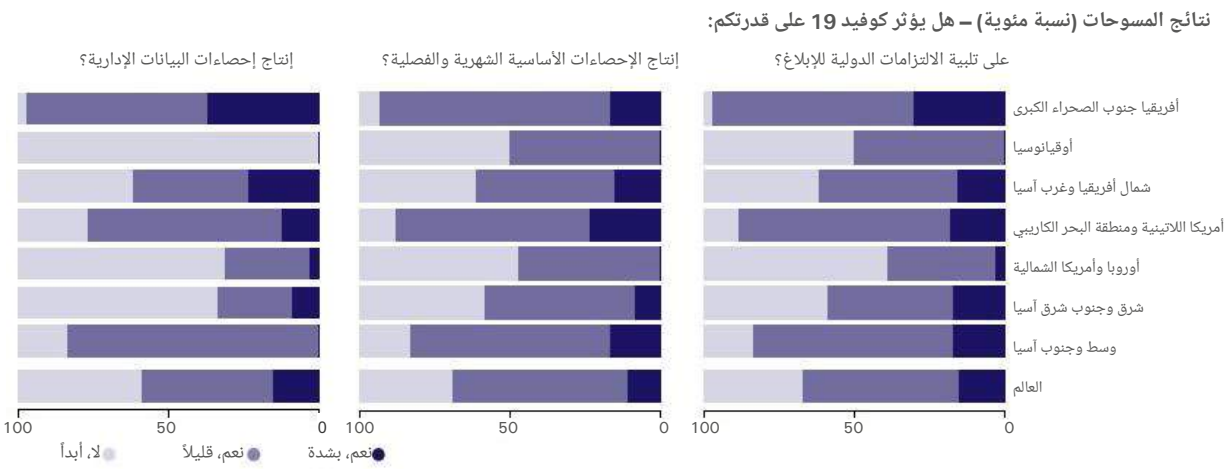
الشكل 4. آنية البيانات حسب أهداف التنمية المستدامة

آنية البيانات: آخر سنة تتوفر عنها البيانات (المتوسط المرجح للبلد الوسيط حسب المؤشر) حسب الهدف



المصدر: تقرير أهداف التنمية المستدامة لعام 2020. https://unstats.un.org/sdgs/report/2020/.

الشكل 5. تأثير كوفيد-19 على الأنشطة الإحصائية



المصدر :تقرير أهداف التنمية المستدامة 2020. https://unstats.un.org/sdgs/report/2016/goal-10.

دور البيانات الضخمة"

في مواجهة هذا الوضع، سؤال واضح جداً يطرح نفسه: هل هناك أي دور يمكن أن تؤديه البيانات الضخمة لمساعدة جهود الأجهزة الإحصائية الوطنية والنظام الإحصائي الوطني في جميع أنحاء العالم، وتيسير أعمالها الرامية إلى تزويد المستخدمين بصورة أكثر اكتمالاً وفي الوقت المناسب لتطورات أهداف التنمية المستدامة، ولا سيما في منطقة الإسكوا؟

دعونا ننظر الآن في لمحة عامة واسعة جداً للمصادر النموذجية التي يمكن استخدامها في تجميع مؤشرات أهداف التنمية المستدامة التي تم الحصول على جزء كبير منها من الرسوم البيانية لمبادرة النبض العالمي (UN Global Pulse) للأمم المتحدة.

- الهدف 1: يمكن أن تقدم أنماط الإنفاق على خدمات الهاتف النقال مؤشرات بديلة عن مستويات الدخل، الاستشعار عن بعد لأضواء الليل والبيانات السكانية؛
- الهدف 2: الاستعانة بمجموعة كبيرة من المصادر الخارجية أو تتبع أسعار المواد الغذائية المدرجة على الإنترنت؛
- الهدف 3: رسم خرائط لحركة مستخدمي الهواتف النقالة؛
- الهدف 4: المواطنون الذين يبلغون عن أسباب تسرب الطلاب؛
- الهدف 5: المعاملات المالية التي تشير إلى أنماط الإنفاق والاستجابات للصدمات؛
- الهدف 6: أجهزة استشعار متصلة بمضخات مياه لتتبع الحصول على المياه النظيفة؛
- الهدف 7: عدادات ذكية تتيح للشركات تعديل مستويات الكهرباء والمياه والغاز الموزعة من أجل الحد من الهدر وضمان الإمداد المناسب في فترات الذروة؛
- الهدف 8: يمكن لأنماط حركة البريد العالمية والأضواء الليلية الناجمة عن الاستشعار عن بعد أن توفر مؤشرات على النمو الاقتصادي والتحويلات والتجارة؛
- الهدف 9: يمكن استخدام البيانات المستمدة من النظام العالمي لتحديد المواقع في

مراقبة حركة السير وتحسين وسائل النقل العامة؛

- الهدف 10: يمكن لتحليلات تحويل المنطوق إلى مكتوب في الإذاعة المحلية أن تكشف عن ممارسات التحيّز؛
- الهدف 11: يمكن للاستشعار عن بعد بواسطة الأقمار الصناعية أن يكشف عن التعدي على الأراضي والغابات والمتنزهات؛
- الهدف 12: يمكن لأنماط البحث على الإنترنت أو معاملات التجارة الإلكترونية أن توفر معلومات عن الانتقال إلى المنتجات الموفرة للطاقة؛
- الهدف 13: قد يساعد المزج بين الصور الملتقطة بواسطة الأقمار الصناعية والمعلومات الواردة من إفادات الشهود والبيانات المفتوحة على رصد إزالة الغابات؛
- الهدف 14: يمكن لبيانات تتبع السفن البحرية التي يتم الحصول عليها عن طريق الاستشعار عن بعد، أن تكشف عن عمليات الاتجار وأنشطة صيد الأسماك غير المشروعة وغير المنظّمة؛
- الهدف 15: يمكن للبيانات حول وسائل التواصل الاجتماعي أن تساعد على إدارة الكوارث بتقديم معلومات آنية عن أماكن وجود الضحايا وعن آثار حرائق الغابات وقوتها؛
- الهدف 16: قد يكشف تحليل المشاعر على وسائل التواصل الاجتماعي عن الرأي العام حول الحوكمة أو الخدمات العامة أو حقوق الإنسان؛
- الهدف 17: يمكن لإقامة شراكات من أجل تمكين الجمع بين الإحصاءات وبيانات الهواتف النقالة والإنترنت، أن تتيح عالماً عصرياً حقيقياً متواصلاً بشكل أفضل وآني.

يمكن الاطلاع على مزيد من الرؤى بشأن المشاريع الدولية الجارية باستخدام مصادر البيانات الضخمة لرصد أهداف التنمية المستدامة في بوابة جرد مشاريع البيانات الضخمة التابعة للأمم المتحدة. (تم فحصه في 9 تشرين الأول/أكتوبر 2020). يبين الجدول 2 المشاريع المبلّغ عنها حسب مصدر البيانات الضخمة والمجال المواضيعي.

على الرغم من كون المشاريع تخمينية وطموحة أحياناً، حيث يتوقع بعضها استخدام المزيد من المصادر، يمكن الحصول بسهولة على فكرة عن أهم مصادر البيانات المستخدمة، ومجالات تركيزها.

تكاد مصادر البيانات الضخمة، مثل استخراج البيانات عبر الويب وبيانات الماسح الضوئي والهاتف النقّال، تحتكر مصادر المعلومات المستخدمة في المخزون. المصدران الأول والثالث هما المصدران الرئيسيان المستخدمان في منطقة الإسكوا، استناداً إلى الردود على الاستبيان الذي سلمته المنظمة إلى البلدان الأعضاء فيها قبل ورشة العمل الإقليمية حول تكامل البيانات الضخمة والمعلومات الجغرافية المكانية لتجميع مؤشرات أهداف التنمية المستدامة في البلدان العربية. في الوقت الحالي، توفر التكنولوجيا وتكنولوجيا المعلومات والاتصالات والرقمنة مصادر جديدة للبيانات الضخمة ذات سمات مختلفة من Vs. تنتج حيادية ومقاربة واندماج التكنولوجيا مصادر أكثر تقلباً للبيانات التي يمكن أن تستخدمها الأجهزة الإحصائية الوطنية والنظام الإحصائي الوطني، والمجتمع الدولي، بما في ذلك أهداف التنمية المستدامة.

تشمل المصادر الرئيسية الأخرى المستخدمة في المنطقة العربية وسائل التواصل الاجتماعي، التي تستخدم، في مخزوننا، في ستة مشاريع جارية.

فيما يتعلق بمجال المشروع، يتم التدقيق أكثر في مجال الأسعار وأسواق العمل والتنقل والسياحة، في حين يتم تحليل مجالات الأسعار والبيئة/الطاقة أكثر في البلدان الأعضاء في الإسكوا.

إذا نظرنا بعمق في نتائج الاستبيان للبلدان العربية، ولا سيما في مكون البيانات الضخمة، الذي خُصص له قسم كامل، تشير نتائج المسح إلى أن خمسة بلدان فقط (36 في المائة)، تتبع إجراءات التحقق من إمكانات مصادر البيانات الضخمة لأغراض إحصائية، بما في ذلك تطور مؤشرات أهداف التنمية المستدامة.

حددت هذه البلدان المصادر الرئيسية للبيانات الضخمة المستخدمة في ما يلي: الإنترنت، والهاتف النقّال، وصور الأقمار الصناعية أو الجوية، ووسائل التواصل الاجتماعي، وأجهزة استشعار الطرق، واستخدام وسائل النقل العام، وغيرها مثل بطاقات الائتمان، وعدادات استهلاك الكهرباء/عداد الكهرباء.

تشارك الوكالات الإحصائية في ستة بلدان (55 في المائة) في مشاريع معالجة البيانات الضخمة باستخدام تكنولوجيات وأدوات مختلفة مثل أدوات وبرمجيات تصور البيانات ونُظُم المعلومات الجغرافية وعناقيد هادوب (Hadoop Clusters).

تستخدم ثلاثة من هذه البلدان فقط (23 في المائة) مصادر البيانات الضخمة لقياس مؤشرات أهداف التنمية المستدامة، وهي نسبة الأفراد الذين يمتلكون هاتفاً نقالاً حسب الجنس، والنسبة المئوية للسكان المشمولين بشبكة الهاتف النقال، حسب التكنولوجيا، ونسبة الأفراد الذين يستخدمون الإنترنت.

تقام مشاريع البيانات الضخمة في ثمانية بلدان، بالشراكة مع المنظمات الدولية والشركاء في مجال التكنولوجيا.

وتبين نتائج المسح أن ثلاثة فقط (23 في المائة) من الأجهزة الإحصائية الوطنية قد طورت أساليب جديدة لحساب التقديرات أو إطاراً منهجياً يتصل تحديداً باستخدام مصادر البيانات الضخمة.

تشمل التحديات والعقبات التي تواجهها هذه الوكالات، والتي تحول دون استخدام مصادر البيانات الضخمة في إنتاج الإحصاءات الرسمية، بما في ذلك مؤشرات أهداف التنمية المستدامة، وجود إطار قانوني غير ملائم، ومحدودية النفاذ إلى مجموعات البيانات، وموارد بشرية غير ماهرة كما يجب لتتمكن من النفاذ إلى مصادر البيانات الضخمة وإدارتها (57 في المائة)، وارتفاع تكاليف النفاذ إلى تلك البيانات (50 في المائة)، والافتقار إلى أدوات التكنولوجيا، وإلى رأي إحصائيين خبراء في البيانات الضخمة (43 في المائة)، وصعوبة تطبيق الأساليب والمنهجيات (36 في المائة).

وفقاً لآخر تقرير سنوي لمبادرة النبض العالمي للأمم المتحدة، هناك بعض المشاريع الابتكارية القائمة تستحق الذكر، تهدف إلى تحسين فهم أهداف التنمية المستدامة باستخدام البيانات الضخمة. غير أنه تجدر الإشارة إلى أن هذه المشاريع تهدف أساساً إلى تحليل الحقائق في البلدان الآسيوية والأفريقية (الجدول 3). وفيما يلي وصف موجز لهذه المشاريع.

ضمن أحد هذه المشاريع، عملت مبادرة النبض العالمي للأمم المتحدة مع مكتب الأمم المتحدة لخدمات الرقابة الداخلية على تحليل بيانات تويتر لفهم مدى فعّالية حملات وسائل التواصل الاجتماعي، وما إذا كان بالإمكان تحسينها.

وقد أطلقت وكالات الأمم المتحدة العاملة في ليسوتو مبادرة أخرى لتحسين فهم تصورات الناس للمكاسب الاجتماعية والاقتصادية والبيئية التي يحققها البلد للمساعدة في زيادة العمل على أهداف التنمية

المستدامة. استخدمت المبادرة الذكاء الجماعي لنمذجة أشكال جديدة لجمع البيانات (من خلال استطلاعات للتصورات وتحليل وسائل التواصل الاجتماعي) من أجل بناء آليات لتقديم الآراء التي يمكن أن تحسن عملية صنع القرار وتقارير المواطنين.

طورت مبادرة أخرى لوحة معلومات تفاعلية للتصور والتحليل تستخدم بيانات من آخر تعداد حول الزراعة لتحديد صغار المزارعين، وبالتالي توليد رؤى يمكن أن ترفد السياسات المتعلقة بالمؤسسات الصغيرة والمتوسطة في قطاع الزراعة.

هناك مشروع آخر يجمع بيانات من مصادر داخلية وخارجية عن الأسباب المحتملة للنزوح، بما في ذلك معلومات عن أحداث النزاع وواقعات الوفيات، والأجور وأسعار السلع الأساسية، والحالات الشاذة المرتبطة بالمناخ، وتدفقات النزوح التاريخية. يتم اختبار نماذج التعلم الآلي المختلفة لإجراء التنبؤات، ويتم تقديم النتائج في لوحة متابعة توضح الوافدين التاريخيين، إلى جانب التنبؤات للنماذج الثلاثة الأعلى أداءً. في التكرار الثاني، يتم توسيع لوحة المتابعة من خلال توحيد مجموعة بيانات الإدخال، وتمديد شكل أفق التنبؤ، وبناء لوحة متابعة داخلية ثانية تسمح للمستخدمين بمقارنة التنبؤات عبر جميع النماذج، واستكشاف بيانات الإدخال، وتقييم أداء الخوارزميات.

في مشروع يركز على نوع الجنس، تُستخدم أداة لتحليل المحتوى الإذاعي، وأنشئت لوحة متابعة لمفاهيم نوع الجنس في الوقت الحقيقي لكشف المناقشات والمواضيع المتعلقة بالعنف الجنسي والعنف القائم على الجنس في بلد معين. تستخدم أداة التعرف الآلي على الكلام تقنية الذكاء الاصطناعي لسحب تسجيلات المحادثات الإذاعية حول مواضيع معينة ونسخها. وبمجرد تحديد مقتطفات المعلومات ذات الصلة – في هذه الحالة المتعلقة بالعنف ضد المرأة والفتاة – يتم تحليلها وتصويرها باستخدام لوحة المتابعة للسماح للسلطات بالوصول إلى معلومات أدق وفي الوقت المناسب.

الجدول 2. مخزون مشروع البيانات الضخمة

مصدر البيانات الضخمة	عدد المشاريع	موضوع المشروع	عدد المشاريع
استخراج البيانات عبر الويب	24	الأسعار	33
الماسح الضوئي	21	سوق العمل	11
الهاتف النقال/سجلّات تفاصيل المكالمات	17	تنقل الموظفين	10
التصوير عبر الأقمار الصناعية	8	السياحة	10
وسائل التواصل الاجتماعي	6	البيئة/الطاقة	9
جهاز استشعار الطريق	5	وسائل النقل	8
العداد الذكي	5	الجغرافية المكانية	5
السجلات الصحية	3	محطات الرصد دون السمعي/تكنولوجيا المعلومات والاتصالات	4
البطاقات الائتمانية	3	الزراعة	3
تعريف السفن	2	الإحصائيات الحيوية/التسجيل المدني	2

الجدول 3. المشاريع الابتكارية، التقرير السنوي لمبادرة النبض العالمي للأمم المتحدة 2019

أهداف التنمية المستدامة	الموضوع	مصدر البيانات	موضوع المشروع
جميع الأهداف	التصور عبر الإنترنت والتواصل الفعال	• بيانات تويتر، وسائل التواصل الاجتماعي؛ • مسوحات التصور، وسائل التواصل الاجتماعي.	• قياس فعالية حملات المناصرة على وسائل التواصل الاجتماعي؛ • دعم الفرق القُطرية لفهم التصورات عبر الإنترنت لأهداف التنمية المستدامة.
2	الأمن الغذائي	البيانات الجزئية للتعداد الزراعي.	رسم خرائط لصغار المزارعين في إندونيسيا لرفد السياسات.
12,16,17	الأثر المجتمعي الذكاء الاصطناعي	وسائل التواصل الاجتماعي (خطابات الأمم المتحدة).	فهم المخاطر في النصوص التي تم إنشاؤها بواسطة الذكاء الاصطناعي.
1,2,3,8,11,16	الحدّ من مخاطر الكوارث	• الهاتف النقال؛ • التنقل والسكان وعلم الاقتصاد والنزاعات وغيرها.	• فهم الحركة السكانية بعد كوارث سولاويزي الوسطى عام 2018؛ • استخدام الذكاء الاصطناعي لنمذجة النزوح في الصومال.
5,10	سلامة المرأة وحقوقها	• وسائل التواصل الاجتماعي (بيانات الإذاعة)؛ • بيانات التنقل والمقابلات.	• إنهاء العنف ضد المرأة والفتاة في أوغندا؛ • بعد حلول الظلام: تشجيع العبور الآمن للمرأة المسافرة ليلاً.
3	الصحة والرفاه	وسائل التواصل الاجتماعي (بيانات الإذاعة).	استخدام البث الإذاعي لتعزيز الكشف المبكر عن المخاطر الصحية.

من أجل استكشاف فائدة استخراج البث الإذاعي للإشارات المبكرة للمخاطر الصحية، قام مشروع آخر بتحليل البيانات المستمدة من البث الإذاعي باستخدام تكنولوجيا تحويل المنطوق إلى مكتوب التي تم تطويرها لتدوين الحديث الإذاعي تلقائياً كنص. بحثت الدراسة في الإشارات الرئيسية المتعلقة بالصحة استناداً إلى كلمات مفتاح محددة. وأظهرت النتائج الأولية وجود ارتباطات محتملة بين المواضيع الصحية التي ناقشها الناس عبر الإذاعة والمقاييس الصحية المسجلة على أرض الواقع.

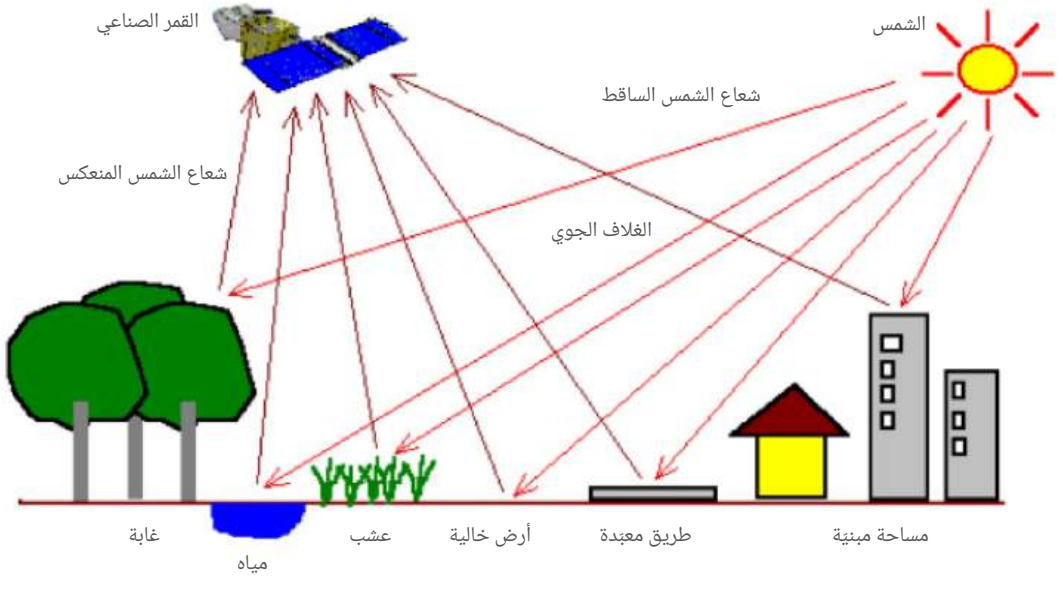
ليست هذه سوى أمثلة قليلة على الكيفية التي يمكن بها للبيانات الضخمة أن تسهم في تجميع أفضل لمؤشرات التنمية المستدامة في مجال محدد وفي بلد معين أو منطقة محلية، مما يوفر لواضعي السياسات رؤى قيّمة تساعدهم على توجيه تدخلاتهم.

3. الاستشعار عن بُعد والهدفان 1 و8

في هذا القسم والأقسام التالية، نركز على مصدرين رئيسيين للبيانات الضخمة، ويبدو أن هذين المصدرين هما الأهم في مساعدة خبراء الإحصاء على رصد مؤشرات أهداف التنمية المستدامة والإبلاغ عنها، وهما بيانات الاستشعار عن بُعد وأجهزة الاستشعار. الميزة الرئيسية لاستخدام معلومات الاستشعار عن بُعد هي كون بيانات الاستشعار عن بُعد متاحة مجاناً بشكل عام، وتضم طائفة واسعة من التطبيقات ذات الصلة. يتم تثبيت أجهزة الاستشعار بشكل عام على الأقمار الصناعية، والتي تكتشف الطاقة المنعكسة من الغابات والمياه والأعشاب والتربة والطرق المعبدة والمناطق المبنية (الشكل 6).

هناك أدبيات متزايدة، بشكل خاص، تستخدم المعلومات المتعلقة بالأضواء الليلية (على سبيل المثال، تلك الموضحة في الشكل 7) للدلالة على العديد من الظواهر المهمة، بدءاً من العثور على الارتباط الشديد بين شدة أضواء الليل وعدد من التغييرات التي يسببها الإنسان: أي في المجال الاقتصادي، الناتج المحلي الإجمالي والأسعار؛ وفي المجال الاجتماعي، معدلات الفقر؛ وفي المجال الديمغرافي، السكان وحركة الهجرة، وفي المجال البيئي، الانبعاثات، والتلوث وتدهور الأراضي؛ وفي مجالات أخرى، الحروب والتهريب والأنشطة غير الرسمية والسياحة والتوسع الحضري.

الشكل 6. كيف يعمل الاستشعار عن بُعد بشكل عملي



الشكل 7. أضواء في الليل كما تم التقاطها بواسطة قمر صناعي تابع للإدارة الوطنية للملاحة الجوية والفضاء (وكالة ناسا)

فيما يتعلق بالهدف 1، بشأن الفقر، استخدم Elvidge and others (2009) ضمن ورقة أساسية البيانات السنوية للسكان LandScan وبيانات نظام مسح الخط التشغيلي لبرنامج الأرصاد الجوية الدفاعي DMSP-OLS من وكالة ناسا (الأضواء أثناء الليل)، وكلاهما بدقة كيلومتر مربع واحد، بهدف اشتقاق مؤشر الفقر المعطى $PI = \frac{Pop}{NL}$ ، من أجل الحصول على معايرة بين PI ومعدلات الفقر الرسمية المستمدة من مؤشرات التنمية العالمية للبنك الدولي. ثم تم تطبيق المعاملات المقدرة على بيانات خلايا السكان وأضواء الليل للحصول على خرائط للفقر على مستوى جغرافي أدق.

تعتبر البيانات المتاحة عن الفقر على الصعيد الوطني نادرة جداً في جميع أنحاء العالم، والمعلومات المفصلة حسب الجنس والمنطقة الجغرافية والقطاع ومستوى الدخل غير متاحة عملياً. ولذلك، يبدو شعار "عدم إغفال أحد" مجرد أضغاث أحلام إذا لم يكن مصحوباً بإجراءات محددة لتحسين جودة البيانات وتوافرها.

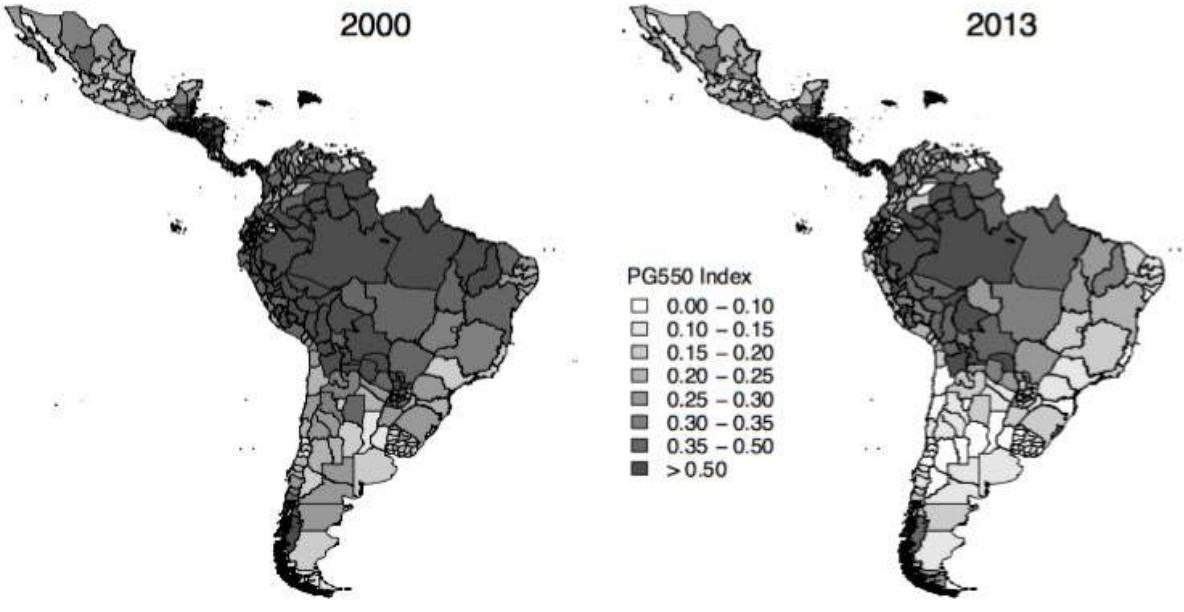
بصرف النظر عن توافر البيانات، الذي يكشف عدم إجراء جميع البلدان في العالم مسوحات للأسر – والتي تعتبر المصدر الإحصائي الرئيسي المستخدم لتقديرات الفقر – تُظهر التقديرات العالمية للفقر

مشاكل معروفة جيداً، وتشمل: ارتفاع تكاليف جمع البيانات وتكاليف معالجتها. عدم دقة توقيت البيانات، وتوافر البيانات، واختلاف توقيت وتواتر جمع البيانات، وعدم اليقين في دورة المسح على المستوى القُطري، والافتقار إلى إمكانية مقارنة المسوحات فيما بين البلدان، والتأثيرات المختلفة لأخطاء القياس على المستوى الوطني الذي قد يعتمد بدوره على عدد من القضايا الحرجة.

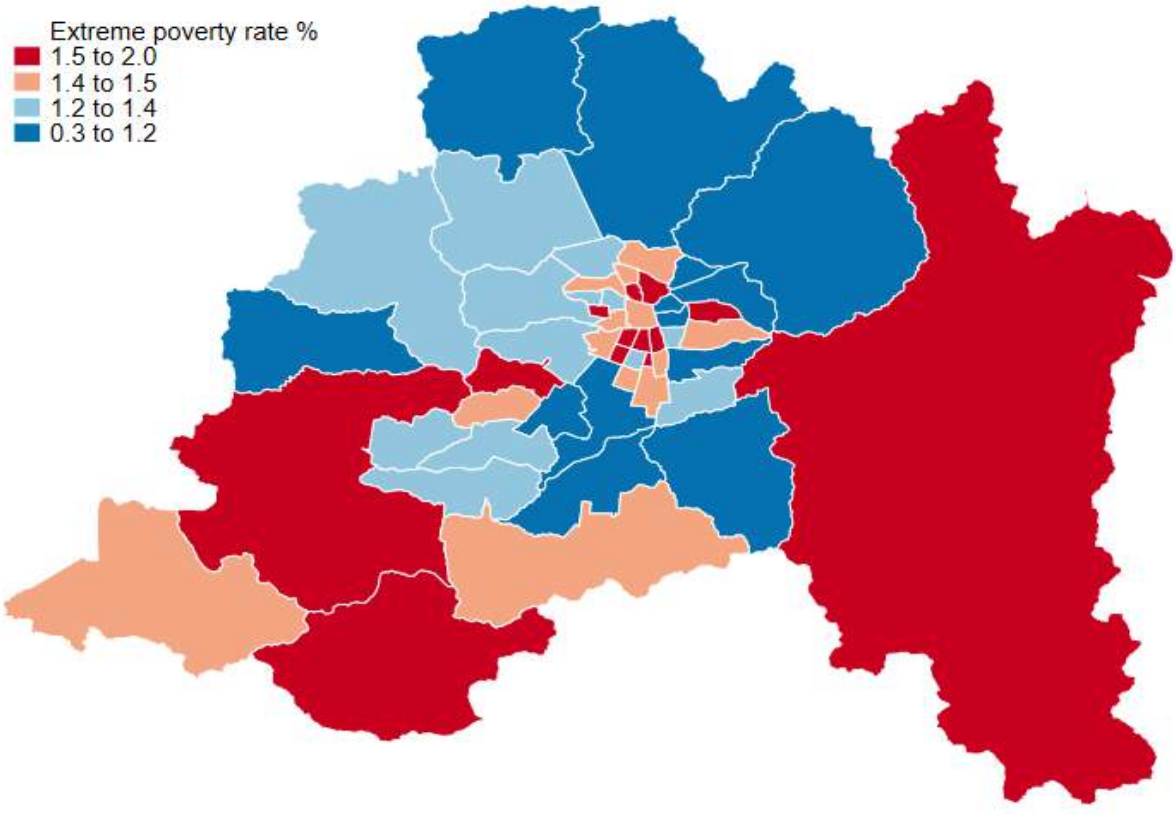
عندما تتوافر مسوحات للأسر المعيشية، فقد يمثل استخدام البيانات الجزئية لوضع تقديرات موثوقة على المستوى المكاني الدقيق من خلال تقنيات تقدير المناطق الصغيرة مهمة استثنائية، نظراً إلى احتمال انخفاض موثوقية المعلومات المتاحة. إحدى أكثر المسائل الإحصائية صعوبة هي الافتقار إلى سلسلة زمنية متجانسة من الإحصاءات دون الوطنية عن الفقر، التي تعتبر ذات أهمية قصوى في توجيه المعونة والتدخلات في مجال السياسات وموارد التنمية المستدامة، لا سيما عندما يعتبر "الإقليم هاماً".

بعد Elvidge and others (2009)، استخدم Andreano and others (2020) و Cecchini and others (2020) إطارين مختلفين للاقتصاد القياسي – لوحة كسرية ونماذج لوجيستية متعددة الحدود – للحصول على خرائط السلسلة الزمنية لتقديرات معدلات الفقر المتعلقة بمؤشر أهداف التنمية المستدامة 1.1.1 لبلدان أمريكا اللاتينية ومنطقة البحر الكاريبي، باستخدام المعلومات الخاصة بالأضواء الليلية والسكان فقط، مع عدم توافر التفاصيل الجغرافية والزمنية في الإحصاءات الصادرة رسمياً من قِبل الأجهزة الإحصائية الوطنية في المنطقة. وترد أمثلة لنتائج هذه التقديرات في الشكل 8 والشكل 9.

الشكل 8. أمثلة لتقدير فجوات الفقر عند 5.50 دولار في اليوم لبلدان أمريكا اللاتينية ومنطقة البحر الكاريبي، باستخدام أضواء الليل



الشكل 9. أمثلة على تقدير معدل الفقر المدقع باستخدام الأضواء الليلية لبلديات سانتياغو، شيلي، في عام 2015



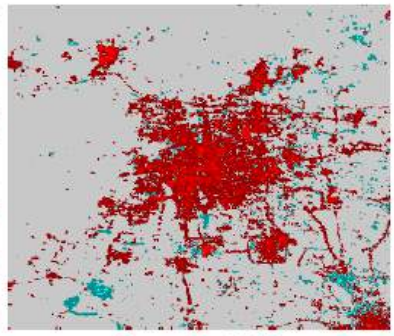
في تطبيق مواز آخر، وباستخدام نماذج بيانات الألواح المطبقة على الأضواء الليلية والبيانات السكانية، يتم الحصول على تقديرات الناتج المحلي الإجمالي وتعادل القوة الشرائية في سلاسل زمنية على مستوى جغرافي دقيق لبلدان أوروبا الشرقية والدول الأعضاء في رابطة الدول المستقلة (Andreano and others, 2019).

قد توفر الأضواء الليلية رؤى وفهما قَيّمين للظروف الاجتماعية والاقتصادية وتغيراتها في وقت الوباء، كما هو الحال بالنسبة لوباء كوفيد-19 الأخير.

يبين الشكل 10 تعتيم الأضواء الليلية واستعادتها في بيجين بالصين، قبل حدوث الوباء وأثناءه وبعده. من الواضح أنه يمكن لأضواء الليل أن توفر لمحة عامة واضحة عن تأثير الأحداث على الأرواح والظروف في المنطقة المتضررة (Elvidge and others, 2020).

الشكل 10. أضواء الليل قبل وأثناء وبعد تأثير الوباء في بيجين، الصين

بيجين، الصين
تعتيم الأضواء الليلية فى شباط/فبراير 2020

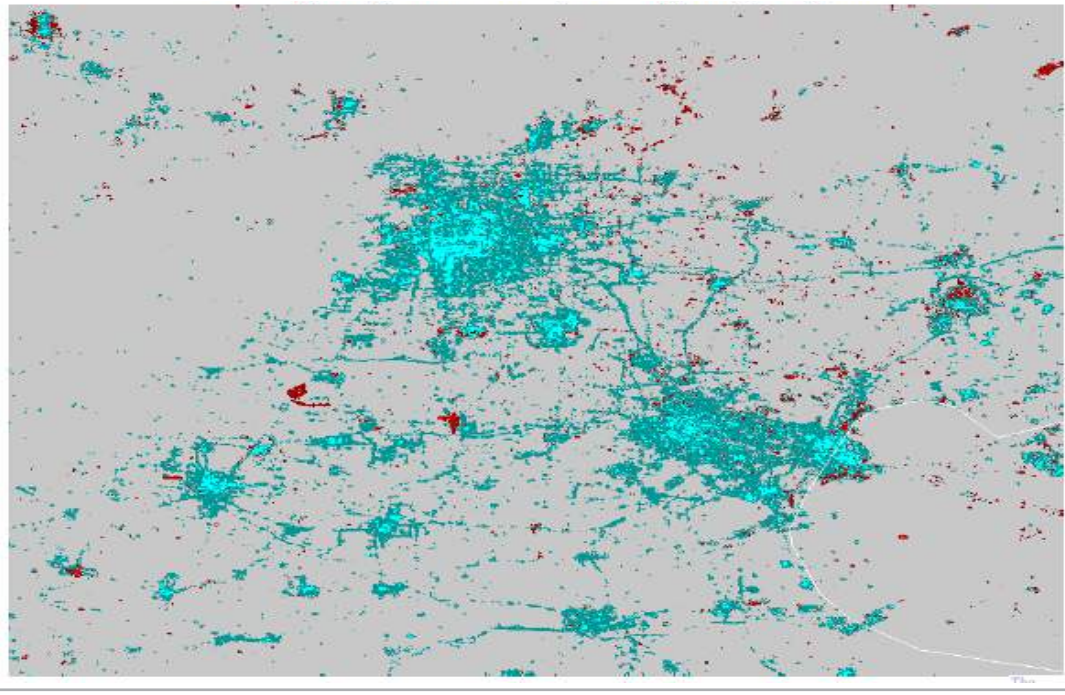


كانون الأول/ديسمبر 2019
عدا شباط/فبراير 2020
أحمر= إضاءة خافتة (معتم)
سماوي= إضاءة مشعّة

كانون الأول/ديسمبر 2018
عدا آذار/مارس 2019
أحمر= إضاءة خافتة (معتم)
سماوي= إضاءة مشعّة

شباط/فبراير 2020
مركب خال من الغيوم

بيجين، الصين
شباط/فبراير 2020 عدا آذار/مارس 2020
استعادة الأضواء بشكل كبير
أحمر= إضاءة خافتة (معتم) - سماوي= إضاءة مشعّة



4. تطبيقات أخرى: أجهزة الاستشعار

أصبحت أجهزة الاستشعار مصدراً هاماً آخر للبيانات الضخمة لرصد أهداف التنمية المستدامة. وتتوفر أمثلة عديدة للتطبيقات الملموسة في الوقت الحاضر. تعتبر الزراعة والمياه والبيئة والطبيعة والمدن والنقل والمنازل مجالات الاهتمام الرئيسي. سنذكر هنا بإيجاز حالتين مثيرتين للاهتمام تهدفان على وجه التحديد إلى معالجة جمع البيانات من خلال أجهزة الاستشعار وتقدير مؤشرات محددة لأهداف التنمية المستدامة.

كثيراً ما توصف أجهزة الاستشعار في المدن ووسائل النقل والمنازل من خلال مصطلح "المدينة الذكية"، إشارة إلى الإجراءات الرامية إلى تحسين مدن اليوم، وظيفياً وهيكلياً، باستخدام تكنولوجيا المعلومات والاتصالات كونها تشكّل بنية تحتية. تشكّل أجهزة الاستشعار العنصر الأساسي نحو هذه التحسينات، والتي يتم تركيبها في جميع أنحاء المدينة لرصد التدفقات الحضرية بأوسع معانيها والكشف عن أي قضايا تحتاج إلى إصلاح أو تحسين في حياة المدينة. وتتألف هذه المجموعة من أجهزة استشعار في المنازل، تهدف إلى تحسين الاستخدام الفعّال للموارد وكذلك سلامة المباني المدنية.

أحد الأمثلة الهامة على ذلك مشروع "تحسين تخطيط النقل من خلال تحليلات البيانات في الوقت الحقيقي"، الذي ينفذه مختبر مبادرة النبض العالمي التابع للأمم المتحدة في جاكرتا، إندونيسيا. تشتهر جاكرتا بازدحام المرور فيها، ما يتسبب بإطالة فترة التنقل وتعقيد جدولة وسائل النقل العام، وبغيرها من الآثار. يتعاون فريق المدينة الذكية داخل حكومة جاكرتا ومختبر النبض في جاكرتا لاستكشاف بيانات موقع الحافلات في الوقت الحقيقي وبيانات الطلب على الخدمة ومعلومات حركة المرور في الوقت الحقيقي. يهدف المشروع إلى تعزيز تخطيط النقل واتخاذ القرارات التشغيلية داخل حكومة جاكرتا من خلال تحليلات البيانات في الوقت الحقيقي، وبالتالي توفير أدلة إحصائية على مؤشر أهداف التنمية المستدامة 11.2.1 – نسبة السكان الذين يتمتعون بالوصول المريح إلى وسائل النقل العام، حسب الجنس والعمر والأشخاص ذوي الإعاقة.

كما يتزايد استخدام أجهزة الاستشعار في المناطق الريفية والطبيعة والمياه، بما في ذلك في البلدان النامية. يمكن النظر إلى أجهزة الاستشعار في المناطق النائية على أنها خطوة نحو هدف "عدم إغفال أحد" لأنها توفر بيانات من المناطق التي تفتقر إلى المعلومات. يشمل ذلك أيضاً التنبؤ بالكوارث باعتبارها مجالاً ذا أولوية للتمكين من الإنذار المبكر، على سبيل المثال، بالفيضانات والانهيارات الأرضية والانهيارات الجليدية.

يتناول أحد الأمثلة، على وجه التحديد، مؤشر أهداف التنمية المستدامة 15.5.1 (مساحة الغابات كنسبة من إجمالي مساحة اليابسة)، مشروع وادي الدرون (وادي الطائرات بدون طيار) الذي ينفذه طلاب جامعة نيويورك في أبو ظبي في الإمارات العربية المتحدة. ويستفيد النظام من تكنولوجيا الطائرات بدون طيار التجارية والبرمجيات المسجلة الملكية للحفاظ على الحياة البرية وحماية البيئة. يجمع مشروع وادي الدرون بيانات في المناطق التي قد يؤدي فيها نشر البنية التحتية للاتصالات إلى إفساد التراث الطبيعي أو يمثل خطراً بشرياً على استرداد البيانات مادياً، وذلك عن طريق استرجاع المعلومات من أجهزة القياس العلمي الأرضية. في حديقة وادي الوريعة

الوطنية، تحلق الطائرة بدون طيار فوق الجبال وعبر الوديان لتنزيل الصور الجوّية التي تلتقطها مصائد الكاميرات الأرضية التي تلتقط تلقائياً صوراً للحياة البرية أثناء مرورها أمام أجهزة استشعار الحركة الخاصة بها. تخدم طائرة وادي الدرون جهود الحفاظ على الحياة البرية لجمعية الإمارات للطبيعة من خلال زيادة معدل تحليل الخبراء لبيانات الصور الجوية للحياة البرية والحد من المخاطر البشرية المرتبطة بالطريقة الحالية للمشي لمسافات طويلة لاستعادة الصور من مصائد الكاميرا البعيدة. كما يلغي وادي الدرون الحاجة إلى استخدام طائرة عمودية مكلفة للوصول إلى مصائد الكاميرات خلال أشهر الصيف، حيث تشكل درجات الحرارة المرتفعة ظروفاً خطيرة للمشي لمسافات طويلة.

5. التحديات والفرص

ومن خلال فهرس مشاريع البيانات الضخمة، والمناقشة في القسمين 4 و5، من الواضح أن البيانات الضخمة يمكن أن تساعد بطريقة عميقة في جمع المعلومات عن عدد من مجالات التركيز، بما في ذلك التنقل والنقل والسياحة والأسعار والفساد والجريمة واستهلاك الطاقة والكثافة السكانية واستخدام الأراضي والرفاه والمدن وسوق العمل.

حالياً، أصبح تحليل البيانات الضخمة مألوفاً جداً بالنسبة للقطاع الخاص، حيث يجري تطوير تحليل سلوك المستهلك والخدمات الشخصية والتمارين والأدوات التنبؤية واستخدامها للتسويق والإعلان والتنبؤ والإدارة. يمكن اعتماد تقنيات مماثلة لاكتساب رؤى في الوقت الحقيقي حول رفاه الناس وتوجيه تدخلات المعونة إلى الفئات الضعيفة.

يمكن استخدام البيانات الضخمة بالاقتران مع مصادر البيانات التقليدية، أو كبديل لها، لتحسين الإحصاءات القائمة وتعزيزها. وهناك اهتمام متزايد بالجمع بين مختلف مصادر البيانات الضخمة (أي البيانات المتنقلة ورصد الأرض) من أجل الاستفادة قدر الإمكان من الإمكانات وأوجه التآزر والتكامل التي توفرها نُظُم المعلومات المختلفة وغير المتجانسة.

قد توفر البيانات الضخمة أيضاً حلولاً للثغرات في البيانات في العالم النامي، حيث فشلت في ذلك حتى الآن النُهُج التقليدية لجمع البيانات. هناك توافق متزايد في الآراء على أن البيانات الضخمة، التي تمثل معاملات حقيقية، قد تكون في بعض الحالات أفضل من بيانات المسوحات. وفي هذا الصدد، تم التأكيد أيضاً على أن البيانات الضخمة قد توفر بيانات أضبط، مع مصداقية أكبر من بيانات المسح التقليدية.

تشمل مزايا البيانات الضخمة التي أبرزتها الأدبيات وفرة في التكاليف، وتحسن التوقيت والآنية، والحد من الأعباء، وإمكانية التعمق في التفاصيل، وأحياناً دقة أكبر وقابلية للمقارنة على الصعيد الدولي، وتنوع أكبر، وسلسلة زمنية جديدة من المؤشرات المتاحة لمنتجي البيانات.

وقد تتيح البيانات الضخمة أيضاً فرصاً لإعادة التفكير في دور الإحصاءات الرسمية وإعادة تحديد موقعها في ضوء نظام بيانات بيئي، أوسع وأكثر اكتمالاً.

العديد من مصادر البيانات الضخمة هي فوق وطنية أو عالمية في نطاقها. ويتيح هذا الجانب المعولم من البيانات الضخمة فرصاً مثيرةً، وإن كانت حساسة من الناحية الاستراتيجية، لإعادة التفكير في نماذج الإنتاج الوطنية. فعلى سبيل المثال، في حين يعتبر التحول من نموذج إنتاج وطني إلى نموذج إنتاج دولي تعاوني أمر رائع، فإنه لا يزال يطرح العديد من التحديات. بما أن التشريعات الإحصائية وحماية البيانات غالباً ما تكون ضعيفة في العديد من البلدان النامية، فقد يكون التركيز على البيانات الضخمة قبل معالجة هذه القضايا الأساسية أمراً خطيراً.

علاوة على ذلك، فإن العديد من البيانات الضخمة هي ملكيات خاصة، مما يعني وجود عوائق قانونية، وتكاليف للنفاذ، وقضايا تتعلّق بالسرية.

ثمة جانب آخر يستحق الاهتمام وهو تركيز ملكية بعض مصادر البيانات الضخمة في فئة قليلة من الناس، ولا سيما تلك التي يمكن الحصول عليها من الأنشطة الرقمية. إذ يهدد وجود الملكية لدى فئات غير الأجهزة الإحصائية الوطنية والنظام الإحصائي الوطني والمنظمات الدولية، سمعة المنظمات الرسمية، وقد يكون من السهل الطعن في أولوية هذه المنظمات بوصفها جهات فريدة من نوعها لتقديم البيانات، وبالتالي تتعرض الكيانات الرسمية لمزيد من الضعف.

عموماً، تعتبر التكاليف والاستثمار على الموظفين (تكوين رأس المال البشري) عناصر مهمة، كما يجب على المنظمات مواجهة التغييرات التكنولوجية السريعة، تلك التي تنمو بها البيانات الضخمة.

من المهم أيضاً ذكر ندرة التمثيل وانخفاض الدقة وضعف البيانات الوصفية باعتبارها سلبيات عند التعامل مع البيانات الضخمة. لا تفي العديد من مصادر البيانات الضخمة بجميع متطلبات جودة البيانات التي يرغب جهاز الإحصاء الوطني في الحصول عليها.

على سبيل المثال، غالباً ما يشار إلى بيانات Google Trend كمصدر بيانات هام يستخدم لأغراض التنبؤ/البث الآني. مع ذلك، لا ينشئ هذه المصادر خبراء إحصائيون، ولا تُنشأ لأغراض إحصائية. فهي تمثل ببساطة عينة مختارة ذاتياً (غير احتمالية)، مع آليات توليد غير معروفة في أغلب الأحيان. لذلك، ليس هناك ما يضمن تمثيلية البيانات إلا إذا كانت تغطي جميع السكان المعنيين، كما هو الحال بالنسبة لبيانات الاستشعار عن بُعد عبر الأقمار الصناعية (Andreano and others, 2019).

كما قد تطرح التقلبات وعدم الاستقرار المتأصلين في مصادر البيانات الضخمة أيضاً قضايا هامة، في حين تفضل الأجهزة الإحصائية الوطنية استقرار البيانات المنشورة، ويشعر المستخدمون بالانزعاج إزاء البيانات التي تخضع لتنقيحات كبيرة. ذكرت بلدان الإسكوا جزءاً من هذه التحديات في إجاباتها على الاستبيان الذي تمت مناقشته أعلاه. حددت البلدان المسائل التالية حسب أهميتها: (أ) عدم كفاية الإطار القانوني، ومحدودية النفاذ إلى البيانات. (ب) عدم مهارة الموارد البشرية كما يجب، وارتفاع تكاليف النفاذ إلى البيانات. (ج) نقص الأدوات التكنولوجية وتصوّر الخبراء الإحصائيين للبيانات الضخمة.

يشكّل الأمن المعلوماتي مسألة هامة أخرى تتعلق بالبيانات الضخمة. تأتي البيانات الضخمة في كثير من الحالات كمصدر مفتوح. وعلى الأغلب، لم يعتبر عنصر الأمان كخاصية أساسية فيها عند تصميمها، وقد يسبب ذلك مشاكل عندما تكون مخازن المعلومات حساسة أو سرية، مثل معلومات العملاء أو أرقام بطاقات الائتمان أو معلومات الاتصال. كل هذه القضايا تعني، من بين أمور أخرى، أنه يجب على الوكالات التي تستخدم البيانات الضخمة أن تولي اهتماماً كبيراً لاستخدام تطبيقات الطرف الثالث وتأمين وصول المستخدم المميز واتفاقيات مستوى الخدمة. قد تشمل أدوات الأمن التشفير، وإدارة المفاتيح المركزية، ومراقبة نفاذ المستخدم، وتعقّب التسرب والوقاية منه، فضلاً عن إنشاء أنظمة الأمن المادي.

6. الخلاصة

قد نكون فقط في بداية عصر البيانات الضخمة: "ثورة البيانات" وهي في جزء منها "تطور البيانات" التي تمثلها البيانات الضخمة في الواقع. وقد تتيح البيانات الضخمة، إذا ما "عُولجت" على النحو الصحيح، بعض الفرص الهامة للأجهزة الإحصائية الوطنية والنظام الإحصائي الوطني والمنظمات الدولية، ولا سيما تحسين التوقيت والدقة، وزيادة التفاصيل وقدرات التصنيف، والقدرة على سد الثغرات في البيانات الرسمية للعديد من مؤشرات أهداف التنمية المستدامة بحلول عام 2030.

قد يلائم ذلك البلدان النامية إحصائياً، بشكل خاص، وخاصة في أوقات وجود قيود خارجية معاكسة (مثل كوفيد-19 والكوارث الطبيعية والكوارث الناجمة عن النشاط البشري).

تقدم البيانات الضخمة مجموعة واسعة من البدائل، تغطي عدداً من المجالات الرئيسية للإحصاءات الرسمية في المجالات الاقتصادية والاجتماعية والبيئية. مع ذلك، يطرح استخدامها تحديات ذات طبيعة مختلفة يجب النظر فيها بعناية، وهي تحديات أخلاقية وقانونية وفنية وتحديات تتعلق بالسمعة.

يجب أن توضع اعتبارات الجودة دائماً على رأس جدول أعمال المؤسسات المكلفة بتوفير البيانات الرسمية. تشمل "الجودة"، ربما في المقام الأول من منظور المستخدم، بُعد "التوافر".

وفي مواجهة متطلبات التوافر، ينبغي أن تُجري الأجهزة الإحصائية الوطنية والنظام الإحصائي الوطني اختباراً نزيهاً حول مزايا وعيوب مصادر البيانات التقليدية المصاحبة (مثل المسوحات الرسمية والتعدادات) مع المعلومات التي يمكن جمعها من مصادر البيانات الضخمة، من أجل استكمال أوجه القصور والفشل في الإحصاءات الرسمية بالمعلومات التي يمكن الحصول عليها من مصادر الأدلة المبتكرة.

المراجع

C. D. Elvidge, P. C. Sutton, T. Ghosh, B. T. Tuttle, K. E. Baugh, B. Bhaduri and E. Bright, “A global poverty map derived from satellite data” (Computers & Geosciences, Elsevier, 2009).

J. E. Steele, P. R. Sundsøy, C. Pezzulo, V. A. Alegana, T. J. Bird, J. Blumenstock, J. Bjelland, K. Engø-Monsen, Y.-A. de Montjoye, A. M. Iqbal, K. N. Hadiuzzaman, X. Lu, E. Wetter, A. J. Tatem and L. Bengtsson, “Mapping poverty using mobile phone and satellite data” (Journal of the Royal Society Interface, 2017).

M. S. Andreano, R. Benedetti, F. Piersimoni, P. Postiglione and G. Savio, “Sampling and modelling issues using big data in now-casting”, in New Statistical Developments in Data Science (eds. Verde R., Ferrari F., Petrucci A. and Racioppi F.), Springer Verlag, 2019.

M. S. Andreano, R. Benedetti, F. Piersimoni, P. Postiglione and G. Savio, “Mapping GDP and PPPs at sub-national level through earth observation in Eastern Europe and CIS Countries”, Voprosy Statistiki, Rosstat, 2019.

M. S. Andreano, R. Benedetti, F. Piersimoni and G. Savio, “Mapping poverty indices for Latin American and the Caribbean countries through satellite remote sensing” (Social Indicators Research, Springer, 2020).

S. Cecchini, G. Savio and V. Tromben, “Shedding light on territorial poverty in Chile”, submitted to Regional Science Policy and Practice, Wiley, 2020.

C. D. Elvidge, F.-C. Hsu, T. Ghosh, and M. Zhizhin, “World tour of COVID-19 impacts on nighttime lights”, Earth Observation Group, Payne Institute for Public Policy Colorado School of Mines, 21 April 2020. Available at https://payneinstitute.mines.edu/wp-content/uploads/sites/149/2020/04/World-Tour-of-COVID-19-Impacts-on-Nighttime-Lights-3.pdf.

TechAmerica Foundation, “Demystifying Big Data: A practical guide to transforming the business of government” (Technical Report, 2012).

United Nations Economic and Social Council, “Big Data and modernization of statistical systems – Report of the Secretary-General”, Forty-fifth session of the UN Statistical Commission, NY, 4-7 March 2014, doc. E/CN.3/2014/11. December. New York, pp. 1-16, 2013.

J. Manske, D. Sangokoya, G. Pestre, and E. Letouzé, “Opportunities and requirements for leveraging Big Data for Official Statistics and the Sustainable Development Goals in Latin America”, White Paper Series Data-Pop Alliance, pp. 1-71, 2016.

21-00217