مـــركـــز الإحـــصـــاء
STATISTICS CENTRE

مفاهيم عامة حول البيانات الكبيرة

أدلة المنهجية والجودة - دليل رقم (13)

13

www.scad.ae

قائمة المحتويات

المقدّمة 3

1 | مفهوم البيانات الكبيرة 4

1.1 تعريف البيانات الكبيرة 4

2.1 أبعاد البيانات الكبيرة 4

3.1 استخدام البيانات الكبيرة 5

4.1 مصادر البيانات الكبيرة 5

2 | البيانات الكبيرة والإحصاءات الرسمية 6

1.2 دور البيانات الكبيرة في الإحصاءات الرسمية 6

2.2 أهم التوصيات الدولية حول استخدام البيانات الكبيرة في الإحصاءات الرسمية 6

3.2 البيانات الكبيرة في مركز الإحصاء – أبوظبي 7

4.2 تبني مركز الإحصاء – أبوظبي للمفهوم الدولي للبيانات الكبيرة 7

5.2 فائدة البيانات الكبيرة لمركز الإحصاء – أبوظبي 7

6.2 تحديات البيانات الكبيرة في مركز الإحصاء – أبوظبي 8

7.2 نهج مركز الإحصاء – أبوظبي في البيانات الكبيرة 8

3 | المصادر 9

المنارة للاستشارات

www.manaraa.com

المقدّمة

أدى التقدم السريع الذي شهده العالم مؤخرا ويشهده حاليا في مجال تكنولوجيا المعلومات والاتصالات، إلى ثورة معلوماتية هائلة نجم عنها قواعد بيانات كبيرة، فتشير الدراسات التي قامت بها شركة إنتل (Intel) إلى أن حجم البيانات التي أنتجها البشر منذ بداية التاريخ وحتى عام 2003 يبلغ حوالي خمسة إكسابايت، إلا أن هذا الحجم تضاعف حوالي خمسمائة مرة خلال عام 2012 ليصبح 2.7 زيتابايت، هذا وتوقعت الدراسات آنذاك أن يتضاعف هذا الرقم إلى حوالي ثلاث مرات بحلول عام 2015. بذلك فقد أصبحت البيانات الكبيرة بمصادرها المختلفة واقعا يعيشه العالم بأسره، كما واعتمد قاموس أكسفورد مصطلح البيانات الكبيرة (Big Data) وأضافه للقاموس مع مصطلحات حديثة أخرى كالتغريدة (Tweet) وغيرها.

انطلاقا مما سبق، فإن النمو المتسارع في إنتاج البيانات ليس من حيث الحجم فقط لا بل من حيث المصدر والسرعة والتنوع واجهته ولا تزال تحديات كبيرة في كيفية التعامل مع هذه البيانات وتعظيم الاستفادة منها. بذلك أصبحت البيانات الكبيرة حدث الساعة من قبل معظم الجهات والمؤسسات الوطنية والإقليمية والدولية، ومجالا واسعا للبحث في إدارة قواعد البيانات، والمنهجيات والإجراءات الممكن تبنيها في سبيل استغلال البيانات الكبيرة بجميع مجالات الحياة.

تتناول هذه الوثيقة الإطار المفاهيمي، من حيث التعريف بمفهوم وأبعاد واستخدامات البيانات الكبيرة، ومن ثم تستعرض الوثيقة مصادر البيانات الكبيرة، وعلاقتها بالإحصاءات الرسمية والتوصيات الدولية حول ذلك. كما وتتناول الوثيقة نظرة مركز الإحصاء – أبوظبي للبيانات الكبيرة والفائدة المرجوة منها. وأهم التحديات. ومن ثم تصف الوثيقة نهج المركز في هذا المجال.

المنارة للاستشارات
www.manaraa.com

1 | مفهوم البيانات الكبيرة

1.1 تعريف البيانات الكبيرة:

منذ وقت ليس ببعيد كانت تنحصر البيانات ضمن فئة قواعد المعلومات المنظمة في مجلدات وملفات وجداول وغيرها، والتي أصبحت لا تشكل الآن أكثر من 10% من إجمالي المعلومات في العالم. ولكن دخول مصادر معلومات جديدة ومتسارعة وغير منظمة مثل رسائل البريد الإلكتروني، مقاطع الفيديو، منشورات فيسبوك، والتغريدات، ورسائل الدردشة على الواتساب، وغيرها من المصادر الأخرى، أدى إلى تكوين قواعد بيانات كبيرة.

لا يوجد هناك تعريف واضح للبيانات الكبيرة، لأن مفهوم حجم البيانات غير محدد بمكان أو زمان. في ظل ما نشهده حاليا من تسارع في تطور تكنولوجيا المعلومات والاتصال ما هو ضخم من بيانات في الوقت الحالي قد لا يكون ضخما في الوقت القادم. كذلك ما هو ضخم من بيانات بالنسبة لشخص أو مؤسسة معينة قد لا يعد ضخما بالنسبة لشخص آخر أو مؤسسة أخرى.

أطلق معهد ماكنزي العالمي (Mackenzie) في عام 2011 تعريفا للبيانات الكبيرة، على أنها هي مجموعة البيانات بحجم يفوق قدرة قواعد البيانات التقليدية من التقاط، تخزين، إدارة، وتحليل تلك البيانات. وبذات السياق أطلق مصطلح البيانات الكبيرة في مجال تقنية المعلومات على مجموعة من حزم البيانات الكبيرة جدا والمعقدة والتي يصعب التعامل معها بواسطة نظم إدارة قواعد البيانات التقليدية.

من جانب آخر، تصف الأمم المتحدة البيانات الكبيرة بأنها «مصادر البيانات ذات الأحجام الضخمة والسرعات العالية والتنوع في البيانات، والتي تتطلب أدوات وأساليب جديدة لالتقاطها، وحفظها، وإدارتها، ومعالجتها بطريقة فعالة.

ببساطة هي البيانات التي لا يمكن تخزينها أو معالجتها باستخدام قواعد البيانات التقليدية نظرًا لكبر حجمها وتعدد مصادرها وتغيرها بشكل سريع وتنوعها أيضًا.

وعلى مستوى الدول اعترفت الحكومات في عدد متزايد من البلدان بأهمية البيانات الكبيرة وأنشأت مجموعات ممارسين وفرق عمل لدراسة استخدام تلك البيانات وتأثيراتها المحتملة.

2.1 أبعاد البيانات الكبيرة:

إضافة إلى المفهوم المتعارف عليه في البيانات الكبيرة وهو كبر حجم البيانات، عرفت شركة غارتنر (Gartner) للأبحاث أبعاد أخرى تميز البيانات الكبيرة، وهي على درجة من الأهمية:

- بعد الحجم الضخم للبيانات، والتحديات المتعلقة بعمليات الجمع والتخزين والتحليل لقواعد البيانات.
- بعد تعدد وتنوع البيانات من حيث مصادرها المختلفة التي ذكرت سابقا، إضافة إلى تنوع تركيبها فهناك:
  - البيانات المنظمة (Structured) وهي البيانات المخزنة في قواعد بيانات منظمة، يميزها إمكانية البحث فيها وتحليلها، كما يمكن إدارتها باستخدام الوسائل التقنية التقليدية. ويمثل هذا النوع من البيانات نحو 10% فقط من إجمالي «البيانات الكبيرة». التي تكون ضمن قواعد بيانات منظمة، تمكن من البحث فيها وتحليلها، كما يمكن إدارتها باستخدام الوسائل التقنية التقليدية.
  - البيانات غير المنظمة (unstructured): هي كل ما لا يمكن تصنيفه بسهولة كالصور والرسوم البيانية، ومقاطع الفيديو، وصفحات الويب، وملفات PDF، والعروض التقديمية، ورسائل البريد الالكتروني، والتغريدات، ومنشورات الفيسبوك، ورسائل الدردشة، وغيرها. ورغم أن هذه الأنواع من الملفات لها هيكل داخلي يخصها، لكنها تعتبر «غير منظمة» لأن بياناتها لا تتسق تماما كقواعد البيانات.
  - البيانات شبه المنظمة (Structured Simi): وهي بيانات شبه منظمة ويمكن تصنيفها بين النوعين السابقين، وهي خليط بين الإثنين، لكنها تفتقر إلى بنية منتظمة مثل برامج معالجة النصوص.
- إضافة لما سبق تعتبر سرعة تواتر حدوث البيانات، وسرعة الوصول إليها وتحليلها أحد الأبعاد أو الخصائص المميزة للبيانات الكبيرة، فهناك معلومات مثل التغريدات يتم الحصول عليها وحصرها خلال فترات قياسية قصيرة جدا.
- من جانب آخر قامت شركة IBM مؤخرا بإضافة بعدا رابعا هو مدى موثوقية ودقة قواعد البيانات الكبيرة بحيث تمكن من الاستخدام الرسمي لها من قبل أصحاب القرار.

المنارة للاستشارات
www.manaraa.com

3.1 استخدام البيانات الكبيرة

مما لا شك فيه أن إتاحة البيانات الكبيرة بأنواعها المختلفة للاستخدام سيؤدي إلى فتح آفاق واسعة للبحث والتطوير على مختلف الأصعدة. تقول شركة IBM أن البيانات الكبيرة تعطيك فرصة لاكتشاف رؤى مهمة في البيانات، وتضيف شركة اوراكل Oracle بأن البيانات الكبيرة تتيح للشركات أن تفهم زبائنها بعمق أكثر. إن الفائدة المتوقعة من استخدام البيانات الكبيرة ستكون في مختلف الجوانب والمجالات، وعلى سبيل المثال لا الحصر:

- دعم البحث العلمي، وذلك في مختلف الحقول والمجالات، من خلال توفيرها لقواعد بيانات على مستويات تفصيلية للمجتمعات المبحوثة، وضمن النطاق الزمني المحدد لمرجعية البيانات. مما ينعكس إيجابا على دقة وكفاءة نتائج الأبحاث والدراسات.
- تشجيع الإبداع والابتكار، من خلال استغلال قواعد البيانات الكبيرة الحالية وإجراء دراسات تفصيلية عليها لتطوير وابتكار خدمات جديدة تساهم في تحسين مستوى أداء المؤسسة أو الشركة.
- الاستخدام في المجالات الاقتصادية المختلفة كاستغلال قواعد البيانات في تحفيز الاستثمار وخلق فرص العمل، إضافة إلى تشجيع التنافسية في مؤسسات الأعمال مما يزيد من كفاءة وجودة الخدمات والسلع المنتجة.
- إتاحة البيانات الكبيرة لتحسين مستوى الخدمات المختلفة، ففي مجال الرعاية الصحية مثلا استغلالها في مجال البحث حول أساليب توفير الرعاية الصحية المثلى، وغيرها من القضايا.

وعلى صعيد آخر، فإن البيانات الكبيرة تعتبر أيضا محرك للابتكار، وهي تفتح آفاق جديدة للحكومة وأصحاب الأعمال لتحسين العمليات والخدمات، وتساعد في استشراف مجالات جديدة لفرص النمو في كافة القطاعات، وتتيح الفرصة أمام شركات الأعمال لاكتشاف وتصميم وتقديم خدمات ومنتجات جديدة لتلبية احتياجات غير ملبّاة من قبل. وإن الأفكار التي سيتم ابتكارها بفضل نشر وتبادل البيانات، ومجموعات المهارات التي سيتم تطويرها لتصنيف وتحليل البيانات، ستسهم في التحول من اقتصاد معرفي إلى اقتصاد إبداعي، وسينتقل الأفراد والهيئات من إنتاج البيانات إلى الاستفادة من البيانات بما يصبّ في مصلحة الجميع.

4.1 مصادر البيانات الكبيرة

تنتج البيانات حالياً بشكل تلقائي ومستمر في صيغة رقمية بطرق مختلفة كثيرة. ويمكن أن تكون هذه المصادر المختلفة للبيانات محل اهتمام لغرض استخدامها في الإحصاءات الرسمية من أجل قياس بعض الاتجاهات الاجتماعية أو البيئية أو المالية أو الاقتصادية بمزيد من الدقة وحسن التوقيت.

والسبب الأهم لزيادة حجم البيانات هو أنها تستمر بالتولد بشكل أكبر بكثير من السابق من خلال عدة أجهزة ومصادر، والأهم أن معظم تلك البيانات ليست منظمة، كتغريدات تويتر والفيديوهات على يوتيوب وتحديثات الحالة على فيس بوك وغيرها، ما يعني أنه لا يمكن استخدام أدوات إدارة قواعد البيانات وتحليلها التقليدية مع هذه البيانات لأنها ببساطة ليست وفق الهيكل الذي تتعامل معه كجداول.

يمكن تصنيف مصادر البيانات الكبيرة على النحو التالي:

- المصادر الناشئة عن إدارة أحد البرامج، سواء كان برنامجا حكوميا أو غير حكومي، على سبيل المثال السجلات الطبية الإلكترونية، وزيارات المستشفيات، وسجلات التأمين، والسجلات المصرفية، وبنوك الطعام
- المصادر التجارية أو ذات الصلة بالمعاملات، الناشئة عن معاملات بين كيانين، على سبيل المثال معاملات البطاقات الائتمانية والمعاملات التي تجرى عن طريق الإنترنت (بوسائل منها الأجهزة المحمولة)
- مصادر شبكات أجهزة الاستشعار، على سبيل المثال، التصوير الساتلت (Satellite image)، وأجهزة استشعار الطرق، وأجهزة استشعار المناخ
- مصادر أجهزة التتبع، على سبيل المثال تتبع البيانات المستمدة من الهواتف المحمولة والنظام العالمي لتحديد المواقع.
- مصادر البيانات السلوكية، على سبيل المثال، مرات البحث على الإنترنت (عن منتج أو خدمة ما أو أي نوع آخر من المعلومات)، ومرات مشاهدة إحدى الصفحات على الإنترنت.
- مصادر البيانات المتعلقة بالآراء، على سبيل المثال، التعليقات على وسائط التواصل الاجتماعي.

وهذه المعلومات تتسم بخصائص فريدة للغاية تميّزها عن البيانات المستمدة من المصادر التقليدية. وتحقق البيانات المستسقاة من هذه المصادر المبتكرة مستويات عالية في التوزيع، وهي ذات هياكل فضفاضة وكميات كبيرة، وكثيرا ما تتاح في وقتها. والبيانات الكبيرة هي مصادر بيانية يمكن وصفها بأنها" بيانات تتسم بضخامة كميتها وسرعتها الفائقة وشدة تنوّعها بحيث تتطلب أشكالاً فعالة من حيث التكلفة ومبتكرة لفهمها على نحو أعمق واستخدامها على نحو أفضل في عملية اتخاذ القرارات".

المنارة للاستشارات
www.manaraa.com

2 | البيانات الكبيرة والإحصاءات الرسمية

1.2 دور البيانات الكبيرة في الإحصاءات الرسمية

هناك حقيقة واضحة، إن البيانات الكبيرة موجودة وستبقى، وهي تحظى بدعم دولي رفيع المستوى، لا سيما من الأمم المتحدة، التي دعت إلى «ثورة بيانات» لدعم أهداف إنمائية لما بعد عام 2015.

بالرغم من أن هناك أمثلة واعدة تشير إلى أن البيانات الكبيرة قد تستخدم في نهاية المطاف في الإحصاءات الرسمية. إلا إنه لازال هناك نقاشات وجدل حول قدرة البيانات الكبيرة على مواجهة تحديات ومشاكل الإنتاج الإحصائي الرسمي، فمثلا هل لدى البيانات الكبيرة القدرة على توفير البيانات الإحصائية الرسمية ضمن التوقيت الزمني المحدد، وهل لديها القدرة على مواجهة ارتفاع تكاليف إنتاج العمل الإحصائي الرسمي، وتوفير البيانات والمؤشرات ضمن الدقة والكفاءة المطلوبة، وفقا لأطر تأكيد جودة البيانات الإحصائية المتوفرة لدى مكاتب الإحصاءات الوطنية، ووفقا لتوصيات المبادئ الأساسية للإحصاءات الرسمية.

فيشير مثلا استخدام شركة Price Stats تقنيات جمع البيانات عبر الإنترنت لجمع الآلاف من أسعار السلع الاستهلاكية من متاجر حول العالم، وتصدر تقديرات يومية للتضخم في العديد من الدول. وهذه التقديرات ومؤشرات الأسعار تكون مشابهة للأرقام التي تأتي بها الإحصاءات الرسمية. وربما في وقت ما ستحل محل الإحصاءات الرسمية مثل المؤشرات الشهرية عن سعر المستهلك.

إن مثل هذه النقلة لن تعني انتهاء دور مكاتب الإحصاء الوطنية، بل على العكس. ستظل الإحصاءات الرسمية مهمة لغرض القياس من أجل إعطاء قيم يمكن مقارنتها دوريًا بالتقديرات من مصادر مختلفة. وأيضًا لدعم تطوير نماذج البيانات الكبيرة المناسبة. وفي ظل محدودية الموارد العامة المتاحة للإحصاءات الرسمية، تواجه مكاتب الإحصاء الوطنية في الدول النامية تحديات كبيرة في تلبية الاحتياجات الإحصائية. لكن الفوائد العظيمة المرجوة من المقاربات الجديدة تتيح حافزًا للابتكار.

وعلى صعيد آخر، ينظر الكثيرون للبيانات الكبيرة على أنها منافسة للإحصاءات الرسمية، غير مدركين أن البيانات الكبيرة تعتبر مكونا مهما للوصول إلى نظام إحصائي أمثل. إن تحقيق ذلك يضع الجميع أمام سؤال هو كيف تتطور الإحصاءات الرسمية لتظل ملائمة في عصر البيانات الكبيرة. بذلك يستوجب من مكاتب الإحصاء الوطنية أن تبدأ بالعمل على استكشاف البيانات الكبيرة وتطويعها للاستخدامات الرسمية، من خلال وضع هدف مسبق يركز على سد فجوة البيانات في الإحصاءات الرسمية أو جعل الإحصاءات الرسمية المتاحة أكثر موائمة للوقت. فالبيانات الكبيرة هي فرصة يجب اغتنامها، لا مصدر للتهديد. والفوائد المرجوة من ورائها هائلة.

بهدف مواجهة مختلف تحديات الإحصاءات الرسمية بتطويع واستخدام البيانات الكبيرة، قدمت العديد من المبادرات الأكاديمية في هذا المجال مثل «ميديا لاب» التابعة لمعهد ماساتشوستس للتقنية. كما وأفصحت العديد من الجهات الفاعلة في مجال التنمية مثل المنتدى الاقتصادي العالمي عن دعمها لمساعدة مكاتب الإحصاء الوطنية على بناء القدرات التحليلية للبيانات الكبيرة في الدول.

2.2 أهم التوصيات الدولية حول استخدام البيانات الكبيرة في الإحصاءات الرسمية:

أظهرت العديد من المؤسسات والهيئات الدولية والإقليمية اهتماما كبيرا في تطويع واستخدام البيانات الكبيرة لأغراض الإحصاءات الرسمية، وما ينعكس على ذلك من جودة للمنتج الإحصائي، وأهم هذه التوصيات:

- بناءً على ملخص اجتماعات اللجنة الاقتصادية لأوروبا والمكتب الإحصائي للمفوضية الأوروبية (المكتب الإحصائي للجماعات الأوروبية)، ومنظمة التعاون والتنمية في الميدان الاقتصادي، واللجنة الاقتصادية والاجتماعية لآسيا والمحيط الهادئ على المنظمات الإحصائية توحيد جهودها لمواجهة المسائل المشتركة المتعلقة باستخدام البيانات الكبيرة والعمل من أجل إيجاد حلول مشتركة، من قبيل إنتاج إحصاءات يكون عمرها المتوقع قصيراً، مع الأخذ بمنهجيات متعددة التخصصات في معالجة البيانات الكبيرة، والاتفاق على وضع نظام موحد لتصنيف مختلف أنواع البيانات الكبيرة.
- تشجيع أعضاء المنظومة الإحصائية الأوروبية لتسخيرها في استخدام الإحصاءات الرسمية من وضع استراتيجية للبيانات الكبيرة، وتبادل الخبرات، والتعاون على مستوى المنظومة الإحصائية الأوروبية وخارجه.
- بيان جدوى الإنتاج الكفؤ للمنتجات الإحصائية الجديدة و"المعتادة" على حد سواء، باستخدام مصادر البيانات الكبيرة، والنظر في إمكانية استنساخ تلك النهج في سياقات وطنية مختلفة.

المنارة للاستشارات
www.manaraa.com

3.2 البيانات الكبيرة في مركز الإحصاء – أبوظبي

ضمن إطار عمله يسعى مركز الإحصاء – أبوظبي إلى توفير البيانات الإحصائية من مختلف مصادرها، وذلك انسجاما مع التوصيات الدولية الخاصة بالإحصاءات الرسمية وأهمها المبادئ الأساسية للإحصاءات الرسمية، والتي تركز على جودة وكفاءة البيانات المنتجة، إضافة إلى التقليل ما أمكن من العبء على المستجيبين وتخفيض التكاليف المادية الخاصة بجمع البيانات.

انطلاقا مما سبق، وكما هو الحال في العديد من الأجهزة الإحصائية الرسمية في الدول المتقدمة، أبدى مركز الإحصاء – أبوظبي اهتماما بالغا في موضوع البيانات الكبيرة كمصدر من مصادر البيانات الإحصائية الرسمية مستقبلا، وهناك عدد من الخطط التي بدأ المركز في العمل عليها والتي تتناول تطويع البيانات الكبيرة للاستخدام في الإحصاءات الرسمية بأمارة أبوظبي.

4.2 تبني مركز الإحصاء – أبوظبي للمفهوم الدولي للبيانات الكبيرة

إذا ما استعرضنا على مستوى دول العالم الإطار المرجعي للمؤسسات والجهات التي تناولت بحث مفهوم وتعريف البيانات الكبيرة سواء كانت هذه المؤسسات أو الجهات ضمن القطاع الخاص أو العام أو على المستوى الدولي نلاحظ كل من تعريف ومفهوم البيانات الكبيرة يتشابه فيما بينها إلى حد كبير. وينسجم بشكل كلي مع ما ورد من تعريف وأبعاد لدى شركة غارتنر (Gartner).

في ضوء ما سبق من تعدد تعريف ومفهوم البيانات الكبيرة، إن مركز الإحصاء – أبوظبي وبصفته أحد المؤسسات الإحصائية الرسمية يحرص على تبني مفهوم البيانات الكبيرة المعترف به من قبل كل من لجنة الأمم المتحدة الاقتصادية لأوروبا UNECE، ومكتب الإحصاء في الأمم المتحدة UNSD. حيث ينص التعريف على ما يلي:

« البيانات الكبيرة هي مصادر البيانات ذات الحجم الكبير، وذات التنوع والسرعة العالية، والتي تتطلب أدوات وطرق جديدة مبتكرة وفعالة وبأقل تكلفة، لالتقاطها، الإشراف عليها، إدارتها، ومعالجتها»

من جانب آخر إن مؤسسات الأمم المتحدة أعلاه لم تقف عند حد وضع تعريف أو مفهوم للبيانات الكبيرة فحسب ولكنها شكلت فريق دولي (Global Working Group GWG)، حيث عقد هذا الفريق اجتماعه الأول في الصين 2014. حيث حدد مهام عمله وأهمها:

- تبادل أفضل الممارسات في استخدام البيانات الكبيرة للإحصاءات الرسمية.
- تحسين وتطوير كفاءة أدوار المؤسسات الدولية والإقليمية والدول المانحة والمتلقية، في مجال البيانات الكبيرة.
- العمل على مصادر البيانات الكبيرة ضمن أمثلة عملية وملموسة.
- معالجة مسألة تكلفة البيانات الكبيرة
- وضع استراتيجيات اتصال مناسبة بشأن البيانات الكبيرة.
- الوصول إلى مبادرات البيانات الكبيرة، والمجتمع المدني، والقطاع الخاص والقطاع، والأوساط البحثية والأكاديمية

5.2 فائدة البيانات الكبيرة لمركز الإحصاء – أبوظبي

من المعلوم أن استخدامات البيانات الكبيرة في الإحصاءات الرسمية حول العالم لا زالت في بداياتها، إلا أنه وفي الآونة الأخيرة، بدأنا نلحظ في بعض الدول تجارب ناجحة في عملية تطويع البيانات الكبيرة للاستخدامات الإحصائية، لما في تطبيق ذلك من فوائد على الأنظمة الإحصائية والأجهزة الإحصائية تحديدا.

أهتم مركز الإحصاء – أبوظبي بموضوع البيانات الكبيرة من حيث التفكير باستخدامها كمصدر للإحصاءات الرسمية وذلك شأنه شأن المؤسسات الإحصائية في جميع الدول، وأن توجهه نحو تطويع البيانات الكبيرة في العمل الإحصائي الرسمي بالتأكيد سيعود عليه بالنفع من حيث:

- يمكن أن توفر البيانات الكبيرة كما أكبر بكثير من الإحصاءات المتاحة في الوقت المناسب، وخاصة عند توفر البيانات المطلوبة بصور آنية.
- توفير بيانات إحصائية تحقق بُعد «الوقتية» وهو أحد أهم أبعاد جودة البيانات الإحصائية. إذ أن توفر وإتاحة البيانات الكبيرة من مصادرها المختلفة هو بمتناول الجميع عند توفر الأدوات والتكنولوجيا المناسبة، مما يتيح معالجة هذه البيانات وتهيئتها للحصول على بيانات إحصائية بأسرع وقت ممكن.
- توفير مركز الإحصاء – أبوظبي الإحصاءات الرسمية من مصادرها المختلفة حسب التوصيات والمواثيق الدولية بأقل عبء على المستجيبين وأقل تكلفة مادية. إن اعتماد البيانات الكبيرة كمصدر من مصادر الإحصاءات الرسمية المستقبلية في مركز الإحصاء – أبوظبي تلبي هذان المتطلبان.

المنارة للاستشارات
www.manaraa.com

- إن اعتماد مركز الإحصاء – أبوظبي على البيانات الكبيرة مصدرا للبيانات عوضا عن المسوح الإحصائية يؤدي إلى تحسين كل من دقة وكفاءة النتائج التي تبنى على هذه البيانات، وذلك باعتبار أن الاعتماد على البيانات الكبيرة سيغني عن استخدام لمعاينة الإحصائية وما تفرزه من أخطاء معاينة تؤثر في النتائج النهائية.
- تمكن البيانات الكبيرة من توسيع مجال الظواهر الممكن اخضاعها للبحث والدراسة، وذلك من خلال توفير مركز الإحصاء – أبوظبي البيانات اللازمة، بأقل مستوى من الوقت والجهد.

6.2 تحديات البيانات الكبيرة في مركز الإحصاء – أبوظبي

وكما هو الحال في معظم المؤسسات الإحصائية حول العالم، يواجه مركز الإحصاء – أبوظبي تحديات مختلفة في تطويع البيانات الكبيرة لغايات الإحصاءات الرسمية، ويحاول دائما الاستفادة من تجارب وممارسات مختلف المؤسسات الإحصائية والتجارب الدولية، وذلك بهدف التغلب على هذه التحديات. هذا ويمكن تصنيف تحديات المركز بهذا المجال كالآتي:

تحديات فنية

إن كبر حجم البيانات الكبيرة يعتبر تحديا كبيرا للجميع سواء في مركز الإحصاء – أبوظبي أو في معظم الدول، ويتمثل هذا التحدي في القدرة على تخزين الكم الكبير من البيانات، إضافة إلى إجراءات معالجة البيانات والتقليل من أثر الأخطاء الإجرائية المتوقعة على النتائج، وأخيرا تحليل المؤسسة الإحصائية للبيانات.

إن التغلب على هذا النوع من التحدي يتطلب توفر الأدوات والأساليب الإلكترونية والفنية القادرة على استيعاب التعامل مع البيانات الكبيرة. ولمركز الإحصاء – أبوظبي شأن باقي المؤسسات الإحصائية على المدى القادم التوجه في الاستثمار في التكنولوجيا الحديثة بهدف بناء القدرات الفنية والإحصائية واستغلال البيانات الكبيرة في بناء قواعد بيانات إحصائية رسمية.

تحديات تغطية البيانات

في كثير من الأحيان ربما تكون هناك فئات من المجتمع المستهدف لا يتوفر لها بيانات أو ربما تتوفر بيانات ولكن بصورة غير مكتملة، إذ قد لا تغطي البيانات متغيرات معينة قيد البحث والدراسة. الأمر الذي يتطلب توفر الخبرات والإمكانيات الفنية والإحصائية إضافة إلى الأدوات والبرمجيات الخاصة بمعالجة البيانات والتعويض عن القيم المفقودة واختبار جودة البيانات الإحصائية كمخرج نهائي.

التحديات القانونية والأخلاقية

وكما هو معلوم في مركز الإحصاء – أبوظبي وكغيره من المؤسسات الإحصائية إن جمع واستخدام البيانات لغايات الإحصائية الرسمية يكون وفقا لقانون خاص بالإحصاء يركز على جوانب سرية وأمن البيانات والمتعاملين، كذلك وفقا لمعايير ومبادئ أخلاقية وسلوكية تضبط كافة الأطراف المعنية في العملية الإحصائية في المؤسسة.

إن التحدي الواضح في هذا المجال هو النظر في المعالجة اللازمة من حيث الوصول إلى البيانات وملكيتها وفق أطر المحافظة على حقوق الملكية.

7.2 نهج مركز الإحصاء – أبوظبي في البيانات الكبيرة

يتبع مركز الإحصاء نهجاً متدرجاً في إدارة مشروعه الخاص بالبيانات الكبيرة. وتتمثل الخطوة الأولى في تحديد البيانات المتوافرة وبحث كيفية استخدامها لإنتاج إحصاءات جديدة ملائمة للاحتياجات أو لتوفير بدائل أقل كلفة وأنسب من حيث التوقيت مقارنة بالمصادر التقليدية. وفي كثير من الحالات، يتطلب ذلك تطوير شراكات مع الشركات الخاصة التي تحوز على البيانات. كما سيتطلب إعداد مخرجات إحصائية جديدة واختبار وضمان جودتها. شرع مركز الإحصاء – أبوظبي في إعداد نهج إدارة البيانات الكبيرة، وبدأ بعدد من الخطوات أو الإجراءات من أهمها:

- تحديد جوانب البيانات الممكن الوصول لها واستخدامها، إضافة إلى تحديد الكيفية التي يمكن استخدامها لإنتاج إحصاءات رسمية جديدة. إضافة إلى تحديد المتطلبات الفنية والتقنية اللازمة.
- تحديد المتطلبات اللازمة المتعلقة بالتنسيق والتعاون مع مختلف الجهات وبناء شراكات مع القطاع الخاص. إضافة إلى التعاون مع المؤسسات الإحصائية الأخرى في هذا المجال لبناء القدرات واكتساب الخبرات في هذا المجال. واستقطاب الخبراء الإحصائيين المتخصصين في مجال البيانات الكبيرة. ولا بد من دور للأكاديميين في زيادة الفهم في هذا المجال وتقديم الإضاءات حول المكاسب الممكنة من البيانات الكبيرة وما ستقدمه للمركز.

المنارة للاستشارات
www.manaraa.com

المراجع:

1	https://unstats.un.org/unsd/statcom/doc1411--2014/BigData-A.pdf
2	https://www.scad.ae/ar/pages/bigdata.aspx
3	http://journals.sagepub.com/doi/full/10.11772053951714538417/
4	http://discoversociety.org/201530/07//what-does-big-data-mean-for-official-statistics/
5	http://www.tech-wd.com/wd/201324/07//what-is-big-data/
6	https://www.imf.org/external/arabic/pubs/ft/fandd/201609//pdf/das.pdf
7	https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.442015//mtg1/S43-_UNSD__Snyder__P.pdf
8	https://www.google.ae/url?sa=t&rct=j&q=&esrc=s&source=web&cd=8&cad=rja&uact=8&ved=0ahUKEwiJ29ea9bfUAhXHbRQKHYnRDZQQFghfMAc&url=http%3A%2F%2Fwww1.unece.org%2Fstat%2Fplatform%2Fdownload%2Fattachments%2F584921002%FBig%2BData%2BHLG%2BFinal.docx&usg=AFQjCNH5x4nReMvvw8JVk9eS9Urfc9ugEQ

المنارة للاستشارات

www.manaraa.com