استخدام البيانات الضخمة في مجال الإحصاءات الرسمية

The use of big data in the field of official statistics

إعداد

منذر محمد الأنصاري

Monther Mohammed Al-Ansari

مستشار أول ذكاء الأعمال السياحية - وزارة السياحة، المملكة العربية السعودية

Doi: 10.21608/ajahs.2023.278604

استلام البحث ١٣ / ١٢ / ٢٠٢٢

قبول البحث ٢٥ / ١٢ / ٢٠٢٢

الأنصاري ، منذر محمد (٢٠٢٣). استخدام البيانات الضخمة في مجال الإحصاءات الرسمية . ***المجلة العربية للآداب والدراسات الإنسانية***، المؤسسة العربية للتربية والعلوم والآداب، مصر، ٧(٢٥) يناير، ٤٠١ – ٤٢٢.

http://ajahs.journals.ekb.eg

استخدام البيانات الضخمة في مجال الإحصاءات الرسمية

المستخلص:

في اطار الثورة الصناعية الرابعة والتي نتج عنها انتاج كم هائل من البيانات عن طريق توظيف تقنية المعلومات في الحياة اليومية للأفراد و اعمال المنظمات الرسمية والخاصة في مختلف المجالات. وصبح من الممكن معالجة كميات كبيرة من البيانات المتنوعة بسرعة عالية واستخدامها في الوقت الفعلي. بينما تعتمد الطرق التقليدية لإنتاج الإحصاءات الرسمية من حيث مصادر البيانات بشكل رئيسي على بيانات المسوحات وبيانات السجلات الإدارية وتأخرها في تلبية احتياجات المستفيدين من الإحصاءات الرسمية برزت إمكانية الاستفادة من البيانات التي تنتجها الاجهزة الاليكترونية المحيطة بنا وكذلك الناتجة عن شبكة الانترنت وعبر شبكات التواصل الاجتماعي. وفى ضوء ما سبق يسعى البحث الحالي إلى دراسة دور البيانات الضخمة في تطوير الإحصاءات الرسمية. عبر استطلاع الدراسات السابقة وتجارب بعض الدول والمنظمات في استخدام البيانات الضخمة بغرض تطوير الإحصاءات الرسمية ، وذلك باستخدام منهج البحث الوصفي من خلال الدراسات السابقة والأدبيات التي تمكن الباحث من الاطلاع عليها. وقد خلصت الدارسة الى أهمية بناء قدرات الكوادر المتخصصة في مجالات تكنولوجيا المعلومات المختلفة؛ لتساعد في تحليل البيانات الضخمة و إنتاج الإحصاءات الرسمية منها. وضرورة رفع مستوى الوعي بأهمية البيانات الضخمة لدى العاملين في القطاع الاحصائي الرسمي من قيادات وموظفين في المؤسسات العامة والخاصة. أيضا، العمل على بناء استراتيجية وطنية لتطوير الإحصاء الرسمي باستخدام مصادر البيانات غير التقليدية ومنها البيانات الضخمة. يتم من خلالها استثمار البيانات المتوفرة لدى مختلف الجهات الحكومية والخاصة ووسائل التواصل لاجتماعي عبر مشاريع تجريبية لإنتاج إحصاءات رسمية. كما أكدت الدراسة على ضرورة وجود اطار وطني للحوكمة وضمان أمن وحماية خصوصية البيانات للمفردات الإحصائية.

Abstract:

Within the framework of the Fourth Industrial Revolution, ,and the resulting production of a huge amount of data by utilizing information technology in the people's daily lives, and the work of official and private organizations in various fields. It has become possible to process large amounts of diverse data at high speed and use it in real time. While the traditional methods of producing official statistics in terms of data sources depend mainly on survey data and administrative records data, and their delay in meeting the needs of the beneficiaries of official

statistics, the possibility of benefiting from the data produced by the electronic devices surrounding us as well as from the Internet and through social networks has emerged. In light of the above, the current research seeks to study the role of big data in developing official statistics. Through a review of previous studies and the experiences of some countries and international organizations in using big data for the purpose of official statistics. By using the descriptive exploratory research method through previous studies through the literature that the researcher was able to view. The study concluded the importance of building the capabilities of specialized cadres in the various fields of information technology; To help analyze big data and produce official statistics from it. Also, the need to raise awareness of the importance of big data among workers in the official statistical sector, including leaders and employees in public and private institutions. In addition to, work on building a national strategy for the development of official statistics using non-traditional data sources, including big data. Through it, the data available at various government and private agencies and social media are invested through pilot projects to produce official statistics. The study also emphasized the need for a national framework for governance and ensuring the security and protection of data privacy for statistical vocabulary.

مقدمة:

شهد العالم خلال السنوات القليلة الماضية تطورا كبيرا في تكنولوجيا المعلومات والاتصالات، وزيادة هائلة في كمية البيانات التي يتم إنتاجها وتخزينها وإتاحتها عبر الشبكات، نتج عنه تضخما في البيانات الرقمية المتوفرة عبر مختلف قنوات التواصل، فالبيانات التي كان ينظر إليها سابقًا وكأنها غير مهمة، يمكن الآن أن تضيف معلومات في غاية الأهمية عند دمجها مع بيانات من مصادر أخرى، فقد أصبح من الممكن معالجة كميات كبيرة من بيانات متنوعة بسرعة عالية في الوقت الفعلي، وقد ارتفع حجم البيانات المتاحة عبر العالم بسبب نشاط الثورة الرقمية عبر الأجهزة الذكية، وهو ما أشار إليه المتخصصون بطوفان البيانات الضخمة[١] .

حيث تعتمد الطرق التقليدية لإنتاج الإحصاءات الرسمية من حيث مصادر البيانات بشكل رئيسي على بيانات المسوحات وبيانات السجلات الإدارية. غير أنه في

عالمنا الحديث تزداد بشكل متسارع ومُطرد البيانات الناتجة عن شبكة الانترنت وعبر شبكات التواصل الاجتماعي وغيرها، وكذلك الكم الهائل من البيانات التي تنتجها الاجهزة الاليكترونية المحيطة بنا.

لقد أصبحت أدوات البيانات الضخمة حلاً واسع الانتشار. فلا تستخدم فقط في صناعات تكنولوجيا المعلومات المتخصصة، ولكن في جميع الحالات التي توجد فيها متطلبات التعامل مع البيانات التي لا يمكن التعامل معها من خلال حلول تكنولوجيا المعلومات التقليدية.

في الوقت الحالي تعمل الأمم المتحدة United Nations (UN)ولجنة الأمم المتحدة الاقتصادية لأوروبا (UNECE)بالإضافة إلى العديد من مكاتب الإحصاء الوطنية National Statistics Offices (NSO)على مشروعات تجريبية تهدف إلى استكشاف الاستخدامات الممكنة للبيانات الضخمة في إنتاج الإحصائيات الرسمية. يهدف هذا الحراك العالمي إلى الوصول منهجيات موحدة ومتفق عليها وتحديد المتطلبات الفنية وتأسيس إطار عمل تشريعي للوصول إلى مصادر البيانات الضخمة وتعيين المعارف والمهارات المطلوبة.

وفى ضوء ما سبق يسعى البحث الحالى إلى دراسة دور البيانات الضخمة في تطوير الإحصاءات الرسمية. عبر استطلاع الدراسات السابقة وتجارب بعض الدول والمنظمات في استخدام البيانات الضخمة بغرض الإحصاءات الرسمية.

مشكلة الدراسة:

تختلف البيانات الضخمة في طبيعتها عن المصادر التقليدية للبيانات من حيث:

١. **الحجم** الضخم والهائل لهذه البيانات (Volume)
٢. **سرعة** تواتر انتاج وانتشار هذه البيانات بشكل سريع ومستمر وآني(Velocity)
٣. **تنوع** هذه البيانات ما بين مهيكلة وغير مهيكلة ونصف مهيكلة Variety))

وتتمثل مشكلة الدراسة في التعرف على إمكانية ومجالات استخدام البيانات الضخمة في الإحصاءات الرسمية حيث أنها تمث مصدر مستحدث للبيانات في الإحصاء الرسمي.

أهداف الدراسة:

١. استكشاف دور البيانات الضخمة في إنتاج البيانات الإحصائية الرسمية.
٢. التعرف على التقنيات المستخدمة في إدارة البيانات الضخمة.
٣. التعرف على طرق استثمار البيانات الضخمة و الوثائق الإحصائية.
٤. التعرف على تطبيقات استخدام البيانات الضخمة في انتاج الإحصاءات الرسمية.
٥. التعرف على تحديات البيانات الضخمة في المركز الوطني للإحصاء و المعلومات.

أهمية الدراسة:

تبرز أهمية الدراسة في استطلاع التطبيقات المحتملة لمصادر البيانات البيانات الضخمة وهي ضمن مصادر البيانات الغير تقليدية وذلك بالاستفادة من الحجم الهائل للبيانات الناتجة عن الثورة الرقمية . ومنافستها للمصادر التقليدية للإحصاءات الرسمية، حيث تعتمد الطرق التقليدية لإنتاج الإحصاءات الرسمية بشكل رئيسي على بيانات المسوحات وبيانات السجلات الإدارية كمصادر وحيدة للبيانات في الإحصاءات الرسمية على المستوى العالمي وفي المملكة العربية السعودية. غير انه في عالمنا الحديث تزداد بشكل متسارع ومُطرد البيانات الناتجة عن شبكة الانترنت وعبر شبكات التواصل الاجتماعي وغيرها، وكذلك الكم الهائل من البيانات التي تنتجها الاجهزة الاليكترونية المحيطة بنا.

منهجية الدراسة:

انطلاقا من مشكلة الدراسة وتساؤلها وأهدافها فإن المنهج المستخدم في هذه الدراسة منهج البحث الوصفي من خلال الدراسات السابقة والأدبيات التي تمكن الباحث من الاطلاع عليها.

أولاً: ماهية البيانات الضخمة:

عرف واتسون (٢٠١٤) البيانات الضخمة انها مجموعة متنوعة من البيانات بكميات كبيرة يصعب التعامل معها بشكل جيد. كما عرفها *Bieraugel (2016)* بأنها البيانات التي لا يمكن تخزينها أو تحليلها من قبل الأجهزة والبرمجيات التقليدية.
اما التعريف الإجرائي للبيانات الضخمة هي تلك البيانات الكبيرة الحجم التي تخزن بطريقة غير منظمة لا تجعل الاستفادة منا اما ميسورا.٢

1)مفهوم البيانات الضخمة

عرفها كانوبي (Canopy,2015) بأنها "مصطلح يستخدم لوصف مجموعات من البيانات الكبيرة للغاية أو على درجة عالية من التعقيد، أو التي تتطلب قدرا كبيرا من المعالجة السريعة– التي تسمى أحيانا م شاكل الحجم / التنوع/ السرعة، والتي يصبح من الصعب أو من المستحيل التعامل معها باستخدام قواعد البيانات التقليدية والأدوات التحليلية التقليدية، علما بأن معالجة بيانات بهذا الحجم يتطلب برمجيات تعمل بشكل متوازي على العشرات أو المئات أو حتى الآلاف من الخوادم."[٣]

كما تعرف البيانات الضخمة على أنها: "مجموعة من البيانات الواسعة والبالغة التعقيد لها خصائصها المميزة (مثل الحجم، السرعة، التنوع، التباين، صحة البيانات)، من الصعب معالجتها باستخدام التكنولوجيا الحالية لتحقيق الاستفادة منها، وتكمن التحديات التي تصاحب هذا النوع من البيانات في معالجتها وتوفيرها وتحليلها وتخزينها والبحث فيها ومشاركتها ونقلها وتصويرها وتحديثها بالإضافة إلى الاحتفاظ بالخصوصية المصاحبة لها ". [٤]. من خلال التعريفات السابقة يمكن تعريف البيانات الضخمة بأنها "تنوع ضخم من البيانات متعددة الأشكال (المقروءة، والمسموعة،

والبصرية، إلخ) تنتجها مصادر متنوعة، وللاستفادة منها ينبغي معالجتها عبر وسائط تكنولوجية غير تقليدية ."أي أن: هذا المصطلح يشير إلى تضخُّم حجم البيانات من ناحية عددها وسرعتها والتنوع في إنتاجها، تأتي هذه البيانات من الصور الرقمية وأشرطة الفيديو، ومواقع التواصل الاجتماعي، والأجهزة الذكية، والهواتف الخلوية الذكية".

الفرق بين البيانات والمعلومات والمعرفة

البيانات: يمكن تعريفها بأنها مجموعة من الحروف، أو الكلمات، أو الأرقام، أو الرموز، أو الصور المتعلقة بموضوع ما. والبيانات في حد ذاتها ليس لها معنى أو قيمة، وهي الصورة الخام للمعلومة [٥]، مثال بيانات الموظفين وصورهم.

أما المعلومات، فهي البيانات التي تم تحليلها ومعالجتها، بحيث أصبحت تحمل معنى وقيمة، ويمكن الاستفادة منها في اتخاذ القرارات، مثل الحصول على إجمالي عدد الموظفين ومعدل أعمارهم ومعدل سنوات الخبرة، ... الخ.

والمعرفة، هي عملية تحليل المعلومات المختلفة وربطُها ببعضها البعض ووجود فَهم واضح لها ومزجها مع الخبرة،

رسم توضيحي(١) الفرق بين البيانات والمعلومات والمعرفة:

2) أنواع البيانات الضخمة

يمكن أن تُقسم البيانات الضخمة الى ثلاثة أنواع: [٦]

(1) **بيانات مهيكلة:** وهي البيانات المنظمة في صورة جداول او قواعد بيانات تمهيدا لمعالجتها

(2) **بيانات غير مهيكلة:** تشكل النسبة الأكبر من البيانات وهى البيانات التى يولدها الأشخاص يوميا من كتابات نصية وصور وفيديو ورسائل ونقرات على مواقع الانترنت الخ

(3) **بيانات شبه مهيكله:** تعتبر نوعا من البيانات المهيكلة الا ان البيانات لا تصمم في جداول او قواعد بيانات.

3) مصادر البيانات الضخمة

يمكن تحديد مصادر البيانات الضخمة وفق المصادر التالية: [٧]

١- المصدر الأول: **البيانات التي تنتجها المؤسسات:** البيانات التي يتم إنتاجها من قبل الشركات والمؤسسات خلال عَملها بشكل يومي، ويتم تخزينها في قواعد بيانات أو ملفات تكون خاصة بالشركة أو المؤسسة.

-**المصدر الثاني: البيانات التي ينتجها الإنسان:** البيانات الناتجة عن الإنسان سواءً منها المستندات الموجودة على أجهزة الحاسب الآلي أو تلك المخزنة في الأجهزة المحمول المختلفة كالرسائل النصية وتطبيقات الأجهزة الذكية، إضافة إلى تلك البيانات التي يدرجها الإنسان في الشبكة العالمية للمعلومات مثل البريد الإلكتروني ومواقع التواصل الاجتماعي ، من مشاركة للصور والمحادثات والتعليقات وتكون أغلبها بيانات غير منظمة

-**المصدر الثالث: البيانات التي تنتجها الآلات**: هي جميع البيانات التي يتم إنتاجها تلقائيا من دون تدخل الإنسان، مثل بيانات كاميرات المراقبة وأجهزة الاستشعار والأقمار الاصطناعية .

المصدر الرابع: **أجهزة متابعة البيانات:** مثل متابعة البيانات المستمدة من الهواتف النقال، ونظام تحديد المواقع.

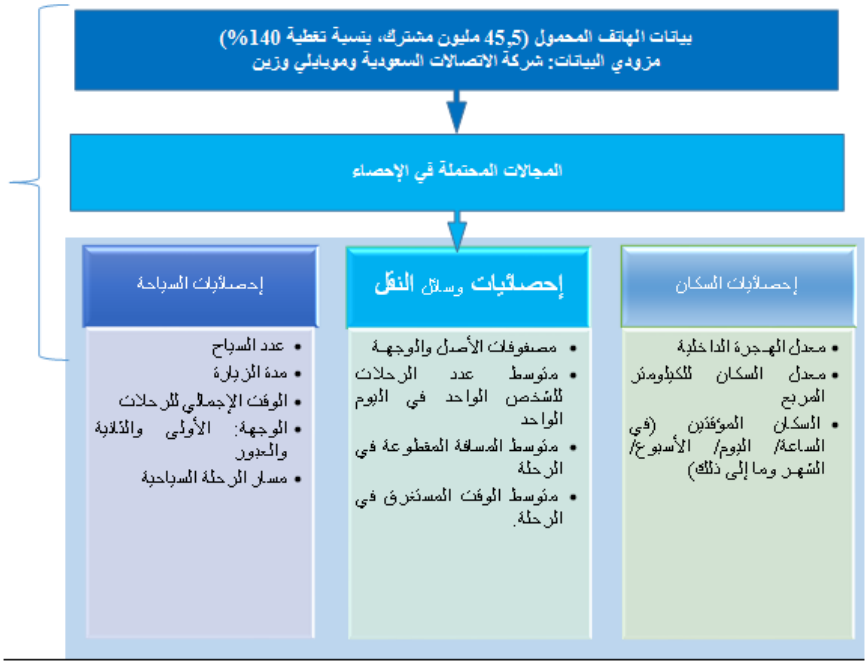
مصادر وسائل التواصل الاجتماعي

- تظهر الإحصاءات التي نشرتها وزارة الاتصالات وتكنولوجيا المعلومات في شهر مارس ٢٠١٦ وجود ١٨.٣ مليون مستخدم سعودي مشترك في منصات التواصل الاجتماعي، بمعدل استخدام يبلغ ٥٨% من مجمل السكان. كما أن متوسط الزمن الذي يقضيه الفرد في وسائل التواصل الاجتماعي يبلغ ٢٦٠ دقيقة، وذلك عند الذين يدخلون مواقع التواصل الاجتماعي من خلال الهواتف الذكية (بما يعني أكثر من ٤ ساعات في اليوم). ويمكن اعتبار بيانات مواقع التواصل الاجتماعي بمثابة مجموعة مناقشة كبيرة تولد بيانات غنية ومفصلة، والتي يمكن استخدامها لقياس العديد من الإحصاءات الاقتصادية والاجتماعية. ومثالًا على المنتجات الإحصائية الجديدة التي يمكن انتاجها بالاعتماد على مواقع التواصل الاجتماعي هي [٨]:
  - **مؤشر ثقة المستهلك.**
  - **مؤشر الرفاهية الذاتي (Subjective Well-being Index).**
  - **مؤشر التلاحم الوطني.**
  - **مؤشر السعادة.**
  - **والمنتجات أعلاه تندرج تحت صنف المؤشرات الجديدة التي يمكن حسابها**

بناءً على البيانات الضخمة. فيعرف مؤشر الرفاهية الذاتية على أنه مؤشر هام يتفوق على المقاييس النقدية (مثل الناتج المحلي) في قياس رفاهية الشعوب. ومن جانبها، عملت منظمة التعاون الاقتصادي والتنمية على تطوير إرشادات حول كيفية قياس الرفاهية موضوعيًا مستخدمة في ذلك الدراسات المسحية.[٩]

بيانات الهاتف المحمول .

رسم توضيحي(2) مجالات استخدام البيانات الضخمة في الإحصاء:



الأدوار والمهارات

يُلخص الجدول التالي الأدوار المختلفة المطلوبة لاستخدام وإدارة بنية البيانات الضخمة، كل دور يتطلب مجموعة مختلفة من المهارات والتي سترد في الجدول التالي :

جدو (١) أدوار ومهارات المسئولين عن إدارة البيانات الضخمة

الدور	الوظيفة	المهارات
المـــدراء	إنشاء وإدارة البنية من حيث الأجهزة والبرمجيات.	مطلوب مهارات Linux المتطورة. تكون مهارات الشبكات مطلوبة فقط في حالة وجود عناقيد لعقد المادية. تحتاج إلى تطوير مهارات محددة خاصة بتكوين وإدارة Hadoop. لا يشترط وجود هذا الدور في حالة النشر من خلال السحابة (Cloud).
المبرمجون	تطوير برامج Spark	برمجة بايثون وتطوير مهارات محددة خاصة ببرمجة Spark.
المحللون الإحصائيون	تحليل مجموعات البيانات وإعطاء المواصفات للمبرمجين حول الأساليب الإحصائية.	الإحصاءات العامة. يوصي بمعرفة لغة قواعد البيانات لتحليل الخدمة الذاتية.
علماء البيانات	تحليل البيانات باستخدام تقنيات تحليلية متقدمة.	برمجة بايثون، وخوارزميات التعلم الآلي.

أشكال البيانات الضخمة:

تتمثل البيانات الضخمة فيما يلي: (١٠)

١ -**بيانات الويب:** يمكن التقاط بيانات عمل الويب على مستوى الفرد كالمشاهدات للصفحات، وعمليات البحث، وقراءة التعليقات، ويمكن أن تعزز الأداء في مجالات مثل أفضل العروض، وبناء النماذج، وتقسيم الافراد، والإعلانات المستهدفة.

٢ -**البيانات النصية:** البريد الإلكتروني والأخبار، ومنشورات فيسبوك والوثائق... إلخ. وهي واحدة من أكبر أنواع تطبيق معظم البيانات على نطاق واسع.

٣ -**بيانات التوقيت والمواقع:** تجعل تقنيات GPS والهاتف النقال بالإضافة إلى معلومات الوقت والموقع مصدرا مزودا للبيانات الضخمة، وينبغي أن تُعامل بحذر شديد لخصوصيتها.

٤ -**بيانات الشبكة الاجتماعية:** من خلال تطبيقات الويب ٢ عبر مواقع شبكات التواصل الاجتماعي يمكن القيام بتحليل الترابط للكشف عن شبكة من مستخدم معين، حيث يمكن أن يعطي تحليل الشبكة الاجتماعية أفكارا حول الإعلانات التي قد تجذب المستخدمين المحددين.

4)قياس حجم البيانات

يمكن قياس حجم البيانات باستخدام مجموعة من وحدات القياس المختلفة تبدأ بأقل شيء وهي البت(Bit) والبايت (Byte) والكيلو بايت(Kilobyte) وصولاً لأعلى شيء متعارف عليه وهو دوميجمقروبايت (Domegemegrottebyte)(١١)

البت (Bit) : هي أصغر وحدة تخزين ممكنة، تمثل مرور أو عدم مرور تيار كهربائي ، وقيمة البت تكون ١ أو ٠ حيث يمثل ١ مرور تيار كهربائ ي، و ٠ عدم مروره[١٢].

وحدة قياس سعة التخزين وهي البايت (Byte) ومضاعفاتها مبينة، كما هو موضح في الجدول التالي.

جدول (٢) معايير قياس حجم البيانات:

وحدة القياس	الاختصار	الحجم
البت Bit	b	0 أو 1
البايت Byte	B	8 بت
كيلوبايت Kilobyte	kB	1024 بايت
ميجابايت Megabyte	MB	1024 كيلو بايت
جيجابايت Gigabyte	GB	1024 ميجابايت
تيرابايت Terabyte	TB	1024 جيجابايت
بيتابايت Petabyte	PB	1024 تيرابايت
إكسابايت Exabyte	EB	1024 بيتابايت
زيتابايت Zettabyte	ZB	1024 اكسابايت
يوتابايت Yottabyte	YB	1024 زيتابايت
زينوتابايت Xenottabyte	XB	1024 يوتابايت
شايلينوبايت Shilentnobyte	SB	1024 زينوتابايت
دوميجمقروبايت Domegemegrottebyte	DB	1024 شايلينوبايت

أساليب تخزين البيانات :

تختلف وتتنوع أساليب تخزين البيانات ما بين:

١ .**بيانات هيكلية**، يتم تخزينها بطريقة منظمة، مثل البيانات الموجودة في قواعد البيانات وتتميز بسهولة المعالجة والتحليل.

٢ .**بيانات غير هيكلية:** تمثل النسبة الأكبر من البيانات، وتكون موجودة ومخزنة بطريقة غير منظمة مثل البيانات الموجودة على الشبكة العالمية للمعلومات (الإنترنت) خاصة منها مواقع التواصل الاجتماعي وتطبيقات الهواتف الذكية ولا يمكن معالجتها بسهولة.

٣ .**بيانات شبة هيكلية:** تخزن بهيئة وترتيب معين مُختلف عن قواعد البيانات [١٣].

رسم توضيحي(٣) أساليب تخزين البيانات:

بيانات هيكلية STRUCTURED DATA	البيانات غير هيكلية UNSTRUCTURED DATA	بيانات شبه هيكلية SEMI-STRUCTURED DATA
بيانات تم تنظيمها في جدول أو قاعدة بيانات. مثل البيانات الخاصة بالموظفين أو عملاء البنك أو معاملات المؤسسة المختلفة.	تمثل النسبة الأكبر من البيانات، مثل البيانات التي تنتجها المواقع الإلكترونية يوميا من نصوص وصور وفيديوهات وخاصة مواقع التواصل الاجتماعي ، وكذلك تطبيقات الأجهزة الذكية.	تعد على أنها نوع من البيانات المهيكلة ولكنها ليست مبينة في جداول أو قواعد بيانات، ولكنها تخزن في ملفات بشكل منظم مثل : JSON,HTML and XML

ثامناً: أهمية استخدام البيانات الضخمة:

تحليل البيانات الضخمة يساعد في التوصل إلى فهم أكثر شمولاً وتفصيلاً لخصائص واحتياجات الأفراد والمجموعات المختلفة، وذلك عن طريق إدراج أجزاء متعددة ومتباينة من البيانات في عملية التحليل، وبالتالي يمكن أن تكون الخدمات والعروض المقدمة لهذه الجماعات أو الأفراد أكثر تخصصا وملائمة، ويكون التواصل والتفاعل معها أكثر كفاءة، ونتيجة لذلك من الممكن أن يرتفع مستوى الرضا بين الأفراد، كما أنه يمكن للبيانات الضخمة أن تعزز كفاءة ودقة التنبؤات، ومن الممكن أيضا استخدام البيانات الضخمة لاكتشاف فرص خفض التكاليف وما يمكن اكتسابه من تعزيز الكفاءة؛ من خلال التحليل الشامل لمختلف العمليات التنظيمية والتشغيلية، هذا إضافة إلى المكاسب المتعلقة بالإنتاجية، وزيادة فرص التوصل إلى مختلف أشكال الابتكارات(١٤)

ثانيا: تطور أنظمة المعالجة للبيانات

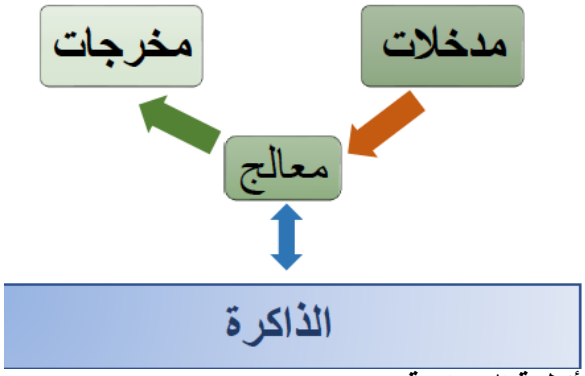
لقد تطورت طرق وآليا ت معالجة البيانات، حيث مرت بعدة مراحل ارتباط ا بالتطور الذي شهدته البنية الهيكلية سواءً على صعيد تطور الأجهزة ومكوناتها ومواصفاتها الفنية وآلية ربطها ببعضها البع ض أو وفقا للتطور في أنظمة التشغيل نفسها. يمكن تلخيص هذا التطور إلى أهم ثلاث مراحل على النحو التالي:

المرحلة الأولى: الأنظمة التسلسلية أو المتتالية (Serial Systems)

مع ظهور أول حاسب آلي عام ١٩٤٥ ، كانت الأجهزة تعتمد بشكل أساسي في تنفيذ المهام على استخدام "معالج مركزي (Central Processing Unit (CPU))" حيث يتم تنفيذ تعليمات البرامج ومعالجة البيانات في الذاكرة الإلكترونية بشكل تسلسلي، من هنا تأتي تسميتها بهذا الإسم، علما وأن ذلك كان يتطلب وقتا كبيراً جداً لتنفيذ التعليمات بالتتابع واحدة تلو الأخرى .(١٥)

الرسم البياني التالي يوضح آلية عمل الأنظمة التسلسلية

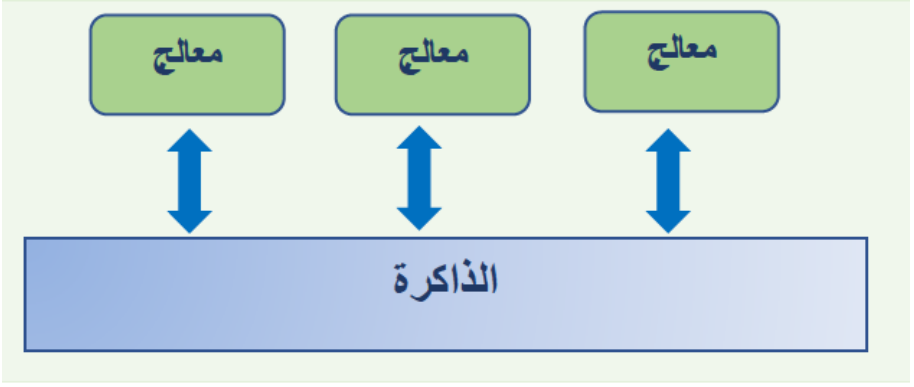
رسم توضيحي (٤) آلية عمل الأنظمة التسلسلية:



المرحلة الثانية: الأنظمة المتوازية (Parallel Systems)

بسبب الوقت الكبير المطلوب لتنفيذ البرامج ومعالجة البيانات، فضلا عن عدم القدرة على تنفيذ أكثر من مهمة في نفس الوق ت، فقد ظلت فكرة زيادة كفاءة أجهزة الحاسب الآلي وتعزيز مواصفاتها لمنحها قدرة أعلى في الأداء وتحسين زمن المعالجة والاستجابة، هي الشغل الشاغل للمختصين، وبالفعل نجحوا في ذلك، عن طريق زيادة القدرة على الأداء المتزامن، أي قدرة الجهاز على أداء أكثر من مهمة في نفس الوقت وذلك عن طريق زيادة عدد المعالجات وحجم الذاكرة، وهو ما يعرف بالأنظمة المتوازية، التي تسببت في طفرة في الكثير من المجالات خصوصا منها الفضاء والطيران. والتي يمكن تجسيدها بالرسم التالي :

رسم توضيحي (٥) آلية عمل الأنظمة المتوازية:



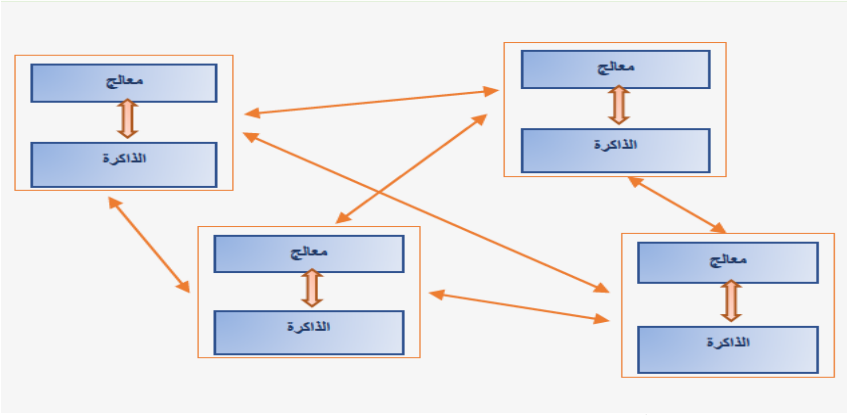
المرحلة الثالثة: الأنظمة الموزعةDistributed System :

مع بداية الثمانينات التي عرفت تطور الشبكات والقدرة على ربط أكثر من جهاز في شبكة واحدة، فقد ظهر بُعد جديد في التفكير في اتجاه آخر وهو تنفيذ تلك المهام

الكبيرة التي مازالت تحتاج إلى وقت معالجة كبير وتتطلب مساحة تخزينية وذاكرة أكبر بشكل مشترك، متزامن وموزع بين أكثر من جهاز واحد عرف بتقنية الأنظمة الموزعة.

كما يبين الشكل التالي، فالأنظمة الموزعة عبارة عن أنظمة تعمل على العديد من الأجهزة مرتبطة مع بعض بشبكة، أي أنها أنظمة تشغيل تظهر للمستخدم وكأنه يعمل على نظام مركزي، إلا أنه في الحقيقة يعمل على عدة وحدات معالجة مركزية تعمل على مستوى التوزيع.

رسم توضيحي (٥) آلية عمل الأنظمة الموزعة:



الإحصائات الوطنية :

أولاً:خلفية :

يعتمد نموذج العمل لمكاتب الاحصاءات الوطنية في الوقت الحالي على مصدرين من مصادر البيانات وهما: المسوحات والسجلات الإدارية. وقد أدى تطور تكنولوجيا جمع وتخزين البيانات وانتشار منصات وسائل التواصل الاجتماعي إلى إحداث ثورة في إنتاج البيانات مما أدَّى الى ظهور البيانات الضخمة، تتميز تلك البيانات بالضخامة في الحجم، والآنية في الانسياب ((Velocity، والتنوع (.(Variety

وفي الوقت الحالي، أصبحت أدوات البيانات الضخمة حلاً واسع الانتشار. فلا تستخدم فقط في صناعات تكنولوجيا المعلومات المتخصصة، ولكن في جميع الحالات التي توجد فيها متطلبات التعامل مع البيانات التي لا يمكن التعامل معها من خلال حلول تكنولوجيا المعلومات التقليدية.

في الوقت الحالي تعمل الأمم المتحدة United Nations (UN)ولجنة الأمم المتحدة الاقتصادية لأوروبا (UNECE)بالإضافة إلى العديد من مكاتب الإحصاء الوطنية National Statistics Offices (NSO)على مشروعات تجريبية تهدف إلى استكشاف الاستخدامات الممكنة للبيانات الضخمة في إنتاج الإحصائيات الرسمية.

يهدف هذا الحراك العالمي إلى الوصول منهجيات موحدة ومتفق عليها وتحديد المتطلبات الفنية وتأسيس إطار عمل تشريعي للوصول إلى مصادر البيانات الضخمة وتعيين المعارف والمهارات المطلوبة.

تعتمد الطرق التقليدية لإنتاج الإحصاءات الرسمية من حيث مصادر البيانات بشكل رئيسي على بيانات المسوحات وبيانات السجلات الإدارية. غير انه في عالمنا الحديث تزداد بشكل متسارع ومُطرد البيانات الناتجة عن شبكة الانترنت وعبر شبكات التواصل الاجتماعي وغيرها، وكذلك الكم الهائل من البيانات التي تنتجها الاجهزة الاليكترونية المحيطة بنا.

حيث تختلف هذه البيانات في طبيعتها عن المصادر التقليدية:

٤. **الحجم** الضخم والهائل لهذه البيانات (Volume)

٥. **سرعة** تواتر انتاج وانتشار هذه البيانات بشكل سريع ومستمر وآني(Velocity)

٦. **تنوع** هذه البيانات ما بين مهيكلة وغير مهيكلة ونصف مهيكلة Variety))

ومن هذا المنطلق نشأ مفهوم ومصطلح "البيانات الضخمة"، وهذا أدى الى توليد فرص تجارية جديدة في القطاع الخاص وكذلك فإن البيانات الضخمة أيضا تمثل مجال هام جدا كمدخل في إنتاج الإحصاءات الرسمية، حيث تقدم البيانات الضخمة مصادر جديدة وثرية للبيانات التي يمكن أن تمثل أهمية في إنتاج الإحصاءات ومنها على سبيل المثال:

١. بيانات متوافرة على الانترنت عن الاسعار والمنتجات والوظائف وغيرها.

٢. بيانات مواقع التواصل الاجتماعي وما تمثله من آراء واهتمامات المواطنين.

٣. بيانات عن سلوك المواطنين على صفحات الانترنت من خلال البحث عن منتج معين أو علاج لمرض محدد.

٤. بيانات ناشئة عن الأجهزة الاليكترونية مثل البيانات الصادرة عن استخدام أجهزة المحمول ونظام تحديد المواقع.

٥. بيانات صادرة عن أجهزة الاستشعار مثل صور الاقمار الصناعية وإشارات المرور وحساسات المناخ.

يمكن تصنيف نطاق استخدام البيانات الضخمة إلى ثلاثة أنواع مختلفة:

١. **إنتاج منتجات إحصائية جديدة:** وفي مثل هذه الحالة، يتم تطوير المنتجات الجديدة باستخدام مصدر البيانات الضخمة. هذا يعني أن المنتج لا يمكن حسابه

من المسوحات والسجلات الادارية. مثال على ذلك إحصاءات المرور على مستوى الطرق باستخدام بيانات الكاميرات.

٢. **استكمال المنتجات الإحصائية الموجودة:** في مثل هذه الحالة من يتم مج مصادر البيانات الضخمة مع المصادر التقليدية، وذلك لخفض التكلفة أو زيادة الدقة. مثال على ذلك استخدام بيانات نقاط البيع في المحلات التجارية لحساب جزء من أسعار العناصر المستخدمة في حساب مؤشر سعر المستهلك (مثلا المواد الغذائية) على أن يتم تكملة العناصر غير المتوفرة في المحلات التجارية الكبيرة عن طريق المسح.

٣. **استبدال المنتجات الإحصائية الحالية:** باعتبارها امتداد لحالة الاستخدام السابقة، يمكن أن تحل مصادر البيانات مصادر البيانات الضخمة محل المصادر التقليدية؛ فعلى سبيل المثال، تستطيع المكاتب الإحصائية إنتاج مؤشر ثقة المستهلك كاملًا من خلال استخدام بيانات وسائل التواصل الاجتماعي.

استخدام تقنيات البيانات الضخمة للإحصاء

في هذا القسم سوف نناقش الاستخدام المتخصص لأدوات البيانات الضخمة (Big Data Technology)داخل الهيئة العامة للإحصاء. بدءً من عرض عام للاستخدامات المحتملة، سوف نستعرض التجارب الدولية ونقدم الدروس المستفادة منها.

- يمكننا تصور أربعة استخدامات مُحتملة:

1) **تجميع البيانات الضخمة:** يتم تجميع البيانات كما هي مباشرة في قاعدة البيانات الضخمة، حيث يتم تجهيزها مسبقا لتنقيحها و/ أو أخذ العينات. ثم يتم تخزين واستخراج البيانات المنقحة في RDBMSحيث يمكن توزيعها على المستخدمين و/ أو معالجتها بشكل أوسع. وتتمثل ميزة استخدام أدوات البيانات الضخمة في مرحلة مبكرة من المعالجة في أنها قادرة على التعامل بشكل جيد مع الحجم الكبير للبيانات من مصادر البيانات الضخمة، مما يسمح بالإبقاء على كل من البيانات بحالتها الأصلية والإصدارات التي تتم معالجتها بشكل مختلف دون الحاجة إلى حذف البيانات من أجل توفير المساحة. يمكن للمحللين العودة إلى مصدر البيانات الأصلي في أي وقت لاختبار حلول مختلفة أو استخدام نفس مجموعة البيانات لأغراض مختلفة.

2) **تفريغ قاعدة البيانات:** تقوم نظم إدارة قواعد البيانات العلائقية بتخزين ومعالجة البيانات حسب الحاجة، وعندما يصبح الحجم زائداً بشكل حرج، فإن أدوات

البيانات الضخمة تتولي تحميل البيانات على قواعد البيانات، على سبيل المثال تخزين جميع البيانات التاريخية، هذا الحل يناسب الحالات التي يكون فيها مصادر البيانات الضخمة لعملية إحصائية تتكرر مع مرور الوقت، مع تزايد البيانات باستمرار. حيث تسمح بالمحافظة على المساحة المستخدمة لتخزين البيانات وأداء قاعدة البيانات تحت السيطرة، بعد تخزين البيانات في قواعد البيانات تبقى متاحة دائما للتحليل في اي وقت دون الحاجة للاتصال بالشبكة.

3) **المعالجة عالية الكثافة:** يمكن لأدوات البيانات الضخمة استضافة نسخ متعددة من مجموعات البيانات التي لا يمكن اعتبارها "بيانات ضخمة" من حيث طبيعتها، ولكن حجمها كبير بقدر كاف لحدوث مشاكل في الأدوات التقليدية عندما تخضع للمعالجة المعقدة. وكمثال على ذلك، يمكننا أن ننظر الى مشكلة ربط السجلات، حيث يجب أن نربط مجموعة بيانات تمثل عدد السكان (من الأفراد أو المؤسسات) إلى مجموعات بيانات أخرى وعدد كبير من المقارنات. وفي مثل هذه الحالات، يمكن لأدوات البيانات الضخمة أن تستغل خاصية التماثل لتحقيق الأداء السريع ولكي تستطيع أن تنجح أيضاً في مواجهة المشاكل الصعبة مثل تطبيق أساليب التعلم الآلي.

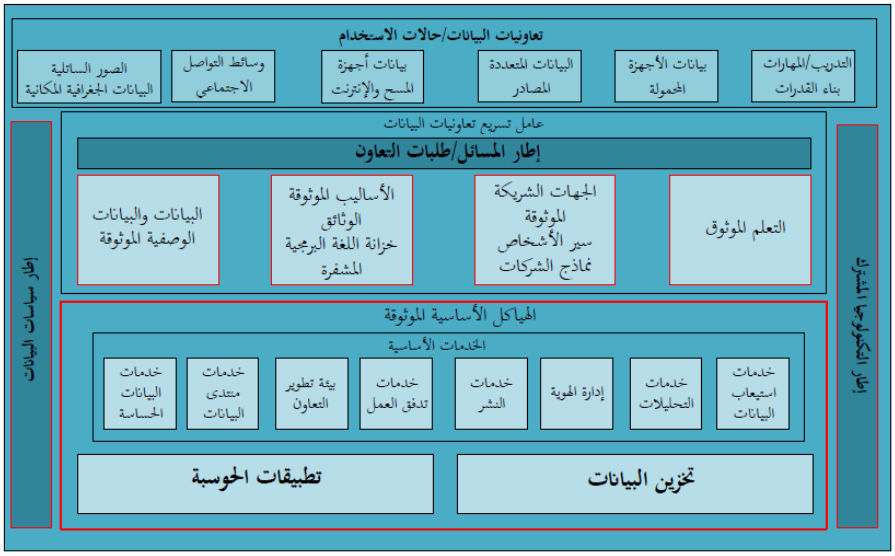
4) **معالجة البيانات الغير مُهيكلة:** لا تستطيع أدوات البيانات الضخمة فقط معالجة البيانات في شكل جداول تقليدية ولكن أيضا البيانات الغير مُهيكلة. يمكن لبعض أدوات البيانات الضخمة (على وجه التحديد - NoSQLقواعد بيانات المستندات) فهرسة محتوى الملفات النصية وتسهيل عملية تصنيفها وفقا لمحتواها. وهذا مفيد خصوصا عند حساب إحصاءات من مصادر مثل صفحات الويب التي يتم تنزيلها من الإنترنت أو رسائل وسائل التواصل الاجتماعي.

5) وهناك مسألة هامة يجب أخذها في الاعتبار، إن إدخال حلول تقنيات جديدة الى الهيئة يتطلب النظر بعناية الى كيفية ادماج هذه التقنيات في بيئة الهيئة التقنية، حيث اعتاد تقنييها وإحصائيها على استخدام أدوات البرمجيات الإحصائية مثل SASأو R. من الناحية التقنية، فإن جميع البرامج الإحصائية الحديثة تعطي إمكانية الوصول إلى أدوات البيانات الضخمة، على الأقل Hadoop، وتوفر واجهات سهلة الاستخدام لذلك. ومن هذا المنطلق، يمكن استعمال أدوات البيانات الضخمة بطريقة مماثلة لنظام RDBMS: حيث يتم التعامل مع البيانات الضخمة من قبل أدوات البيانات الضخمة ويتم تغذية البرنامج الإحصائي بالنتائج التي تم تنقيحها والتي يمكن أن يتم اعتبارها مدخلات لتحليل إضافي.

مقارنة مصادر البيانات

يوضح الجدول أدناه تصنيف أدوات البيانات الضخمة وفقاً لمصادر البيانات التي تم تناولها في دراسة المنتجات والمنهجيات. بالنسبة لكل مصدر بيانات، يوضح الشكل (٦) الأدوات التي يوصى بها.

الشكل (٦) إطار عام لمصادر وهياكل البيانات الإحصائية:



ويشير الجدول أيضا إلى تخصص كل تقنية فيما يتعلق بنوع معين من البيانات، منظمة (عموديا بعدد ثابت من الصفات وانتشار القيم العددية)، منظمة بشكل حر (عموديا بعدد متغير من الصفات و/ أو انتشار المحتوى النصي) أو غير منظمة تماما (النص الخام أو الملفات الثنائية). لا يتم سرد قواعد البيانات ذات الرسم البياني وقواعد بيانات القيمة الرئيسية في سياق الإنتاج الإحصائي نظرًا لأهميتهم الهامشية.

مجموعة العمل العالمية لاستخدام البيانات الضخمة في الإحصاءات الرسمية (GWG on BD))

إتفقت اللجنة الإحصائية للأمم المتحدة بجلستها الخامسة والأربعون المنعقدة عام ٢٠١٤ على إنشاء مجموعة عمل عالمية لاستخدام البيانات الضخمة في الإحصاءات الرسمية للتحقق من منافع وتحديات استخدام البيانات الضخمة في إنتاج الإحصاءات الرسمية . وفي هذا السياق، أقرت المجموعة والمجتمع الاحصائي الحاجة الى العمل على القضايا المهمة والضرورية للاستغلال الصحيح والامثل

للبيانات الضخمة وعلى رأس هذه القضايا قضية الوصول الى منهجيات متفق عليها وقضة التشريعات المتعلقة بالوصول للبيانات والبنية التحتية التقنية (ملحق ١ يسرد تفاصيل مهام مجموعة).

تتكون المجموعة من ٩ منظمات عالمية هـي: •البنك العالمي
•النبض العالمي التابعة للأمم المتحدة
•قسم الإحصاء التابع الأمم المتحدة
•لجنة الأمم المتحدة الاقتصادية لأوروبا
•اللجنة الاقتصادية والاجتماعية لآسيا والمحيط الهادئ
•منظمة التعاون الاقتصادي والتنمية
•الاتحاد الدولي للاتصالات
المركز الإحصائي لدول مجلس التعاون لدول الخليج العربي
•الإحصاء الاوروبي
•بالاضافة الى العديد من الدو حيث تم تقسيم المجموعة الى ٧ فرق عمل كالتالي:

١. الوصول للبيانات والشراكات.
٢. البيانات الضخمة وأهداف التنمية المستدامة.
٣. بيانات الهاتف المحمول.
٤. بيانات الأقمار الصناعية والبيانات الجغرافية المكانية.
٥. بيانات مواقع التواصل الاجتماعي.
٦. التدريب والمهارات وبناء القدرة.
٧. لجنة المنصة العالمية لخدمات البيانات.

اعمال لجنة الأمم المتحدة الاقتصادية لأوروبا:

- عملت لجنة الأمم المتحدة الاقتصادية لأوروبا على مشروع للبيانات الضخمة كمبادرة ضمن عمل الفريق رفيع المستوى المعني بتحديث أنظمة الإحصاءات الرسمية. ويهدف مشروع لجنة الأمم المتحدة الاقتصادية لأوروبا المعني بالبيانات الضخمة على إمكانية استخدام البيانات الضخمة في الاحصاءات الرسمية والتقنيات الضرورية للاستفادة من تلك المصادر، بالاضافة الى قابلية تطبيق المعايير والأساليب الإحصائية الحالية على البيانات الضخمة.
- تتبنى هذه المجموعة المنهج العملي الذي يُمكن للمشاركين من اختبار مصادر وأدوات البيانات الضخمة على المنصة التعاونية المشتركة والمسماة ساندبوكس Sandbox. ويمكن تعريف ساندبوكس على أنها بيئة تقنية تمكن مكاتب الاحصاء في أوروربا على وجه الخصوص والعالم أجمع من إجراء دراسات تجريبية على البيانات الضخمة، وتستضيفه مكتب الاحصاء المركزي التابع لايرلندا والمركز الإيرلندي للحوسبة المتطورة (ICHEC).

يتشكل مشروع لجنة الأمم المتحدة الاقتصادية لأوروبا للبيانات من خمس فرق عمل على النحو التالي:

١- فريق عمل الشراكات.
٢- فريق عمل الخصوصية.
٣- فريق عمل الجودة.
٤- فريق عمل ساندبوكس.Sandbox
٥- فريق عمل المهارات والتدريب.

التوصيات:

١. بناء على مخرجات الدراسة و ما توصل إليه الباحث من نتائج يوصي بالآتي:
٢. بناء القدرات والكوادر المتخصصة في مجالات تكنولوجيا المعلومات المختلفة؛ لتساعد في تحليل البيانات الضخمة وإنتاج الإحصاءات الرسمية منها.
٣. رفع مستوى الوعي بأهمية البيانات الضخمة لدى العاملين في القطاع الاحصائي الرسمي من قيادات وموظفين في المؤسسات العامة والخاصة.
٤. العمل على بناء استراتيجية وطنية لتطوير الإحصاء الرسمي باستخدام مصادر البيانات الغير تقليدية ومنها البيانات الضخمة.
٥. استثمار البيانات المتوفرة لدى مختلف الجهات الحكومية والخاصة ووسائل التواصل لاجتماعي عبر مشاريع تجريبية لإنتاج إحصاءات رسمية.
٦. ضرورة وجود إطار وطني للحوكمة وضمان أمن وحماية خصوصية البيانات للمفردات الإحصائية.
٧. التركيز على الاستفادة من برمجيات تحليل البيانات والتعرف إلى حاجة بيئة العمل الإحصائي، و المواءمة بين نوع البيانات و مجالات العمل الإحصائي التي يحتاجها.
٨. تشجيع القطاع الخاص للاستثمار في تطوير تقنيات البيانات الضخمة.

مراجع البحث

أولا : المراجع العربية

١- ابتسام، حملاوي . (2018) واقع وآفاق البيانات الضخمة في الوطن العربي : دراسة حالة المملكة العربية السعودية، المجلة الدولية ل لدراسات الاقتصادية، المركز الديمقراطي العربي، ألمانيا، برلين، العدد (3)، ص122-137

٢- الأكلبي، علي (٢٠١٧) ، تحويل البانات الضخمة الى قيمة مضافة ، مجلة مكتبة الملك فهد الوطنية، مج٢٣،ع٢،رجب-ذو الحجة ١٤٣٨هـ

٣- البار، عدنان مصطفى. (2017) البيانات الضخمة ومجالات تطبيقها. متاح https://www.kau.edu.sa/GetFile.aspx?id٢٨٥٢٦٠&fn=Article-of-the-

٤- الطيب، زينب والرباعي، سليم ان إبراهيم. (2018)، الأدوار الجديدة لأخصائي المعلومات للتعامل مع البيانات الضخمة، ، Journal of Information Studies & Technology مج ١٦ ، ١- ١٥ .

٥- مقناني، صبرينة وشبيلية، مقدم . (2019)، دور البيانات الضخمة في دعم التنمية المستدامة بالدول العربية Journal of Information Studies & Technology - ، مج 4، 1-14 .

٦- وزارة تكنولوجيا الإعلام والاتصال . (2014)، البيانات الضخمة : تحقيق التوازن بين المزايا والمخاطر، قطر.

ثانيا : المراجع الأجنبية :

1- Canopy, L. (2015). Big data for marketing: When is big data the right choice?Retrived from -(file///c/:/Users/user/Downloads/canopyfile-whitepaper_big-ata-formarketing.pdf(-

2- Meguenani, S. & Mokadem, C. (2019). Role of big data in supporting sustainable development in Arab countries. Journal of Information Studies and Technology, 1 (4), 1-14. https://doi.org/10.5339/jist.2019.4

3- Mercier, M. (2019). "Contribution to High Performance Computing and Big Data Infrastructure Convergence", Université Grenoble Alpes, Oct

4- Sanders, John. (2016). "Defining Terms: Data, Information and Knowledge", Pages (1-3)

5- Ward, B. (2004). "How Linux Works: What Every Superuser Should Know?"

6- Weiss, A. (2018). "Big Data Shocks: An Introduction to Big Data for Librarians and Information Professionals" , Chicago: American Library Association , page 19

الهوامش:

[1]مقناني، صبرينة وشبيلية، مقدم . (2019)، دور البيانات الضخمة في دعم التنمية المستدامة بالدول العربية.
Journal of Information Studies & Technology - ، مج 4، 1-14 .

[2]الأكلبي، علي (٢٠١٧) ، تحويل البانات الضخمة الى قيمة مضافة ، مجلة مكتبة الملك فهد الوطنية، مج٢٣،ع٢،رجب-ذو الحجة ١٤٣٨هـ

([3])Canopy, L. (2015). Big data for marketing: When is big data the right choice?Retrived from -(file///c/:/Users/user/Downloads/canopyfile-whitepaper_big-ata-formarketing.pdf(-

[4]البار، عدنان مصطفى. (2017) البيانات الضخمة ومجالات تطبيقها. متاح في
: &fn=Article-of-the-٢٨٥٢٦٠https://www.kau.edu.sa/GetFile.aspx?id=
- .pdf٢٠١٧-November٠١-Week-Adnan-Albar-

([5])Sanders, John. (2016). "Defining Terms: Data, Information and Knowledge", Pages (1-3)

([6])Meguenani, S. & Mokadem, C. (2019). Role of big data in supporting sustainable development in Arab countries. Journal of Information Studies and Technology, 1 (4), 1-14. https://doi.org/10.5339/jist.2019.4

([٧])ابتسام، حملاوي . (2018) واقع وآفاق البيانات الضخمة في الوطن العربي : دراسة حالة المملكة العربية السعودية، المجلة الدولية ل لدراسات الاقتصادية، المركز الديمقراطي العربي، ألمانيا، برلين، العدد (3)، ص137-122

[٨]وزارة الاتصالات وتكنولوجيا المعلومات، المملكة العربية السعودية في شهر مارس ٢٠١٦ **http://www.mcit.gov.sa/En/MediaCenter/Pages/News/News-22032016_982.aspx**

[9] موقع منظمة التعاون والتنمية**OECD**
http://www.oecd.org/statistics/oecd-guidelines-on-measuring-subjective-well-being-9789264191655-en.htm

([١٠])الطيب، زينب والرباعي، سليم ان إبراهيم. (2018)، الأدوار الجديدة لأخصائي المعلومات للتعامل مع البيانات الضخمة، ، Journal of Information Studies & Technology مج ١٦ ، ١- ١٥ .

(١١) Weiss, A. (2018). "Big Data Shocks: An Introduction to Big Data for Librarians and Information Professionals" , Chicago: American Library Association , page 19

(12)Ward, B. (2004). "How Linux Works: What Every Superuser Should Know?"

(13)Ward, B. (2004). "How Linux Works: What Every Superuser Should Know?"

(١٤)وزارة تكنولوجيا الإعلام والاتصال . (2014)، البيانات الضخمة : تحقيق التوازن بين المزايا والمخاطر، قطر.

(15)Mercier, M. (2019). "Contribution to High Performance Computing and Big Data Infrastructure Convergence", Université Grenoble Alpes, Oct