Cybrarians Journal

Issue 62, June 2021 عدد 62، يونيه 2021

الوثائق الإدارية الجارية، والبيانات الضخمة في مكتبة الإسكندرية

د. نرمين إبراهيم على إبراهيم اللبان

أستاذ الوثائق، والأرشيف المساعد

قسم المكتبات والمعلومات– كلية الآداب– جامعة الإسكندرية – مصر

nermeenellaban@gmail.com

المستخلص :

سعت هذه الدراسة إلى استعراض مشروعات مكتبة الإسكندرية في مجال البيانات الضخمة، بالإضافة إلى تحليل مدي إمكانية استخدام تطبيقات البيانات الضخمة في بيئة المكتبة الوثائقية؛ للاستفادة من البيانات الحيوية التي تتضمنها وثائقها في المستقبل. وقد اعتمدت الباحثة على منهج دراسة الحالة في إجراء هذا البحث. وقد أظهرت الدراسة أن بيئة المكتبة الوثائقية صالحة بنسبة 72% لاستخدام تطبيقات البيانات الضخمة فيها؛ إذا اقتصر تقييم الباحثة على واقع وثائق المكتبة الورقية، والرقمية فقط، من دون وثائق الجيل الثاني. أما إذا كان التقييم يرتكز على تحليل بيئات وثائق المكتبة الثلاث: (الورقية والرقمية والجيل الثاني معًا) ففي هذه الحالة ينخفض معدل صلاحية بيئة المكتبة الوثائقية لاستخدام تقنيات البيانات الضخمة فيها إلى 59%. هذا وقد أوصت الدراسة بضرورة تحسين مكتبة الإسكندرية للبيئة الفنية، والتنظيمية المختصة بوثائقها؛ حتى يمكن توظيف تطبيقات البيانات الضخمة في تحليلها–عندما يتطلب الأمر في المستقبل. كما يمكنها –أي مكتبة

الإسكندرية – التعاون مع المؤسسات التي لديها تجارب في التعامل مع وثائق العمل؛ باعتبارها مصدرًا حيوياً للبيانات الضخمة؛ للاستفادة من خبراتها في هذا المجال.

الكلمات المفتاحية : البيانات الضخمة؛ إدارة الوثائق الجارية ؛ مكتبة الإسكندرية

Current Records and Big Data in the Bibliotheca Alexandrina

DR. Nermeen Ibrahim Ali Ibrahim Ellaban
Associate Professor of Records and Archives
Department of Library and Information Science,
Faculty of Arts, Alexandria University, Egypt

Abstract:

This study aims to shed light on the big data projects in BA as well as analyze the validity of the documentary environment in BA for using big data tools to benefit in the future from the vital data in BA's records. The researcher uses the case study and concludes that BA entered the era of big data through cultural-knowledge projects related to the web environment. Moreover, the research shows that the readiness rate of BA's documentary environment to utilize big data tools is (72%) —if the researcher depends only on assessing the digital and paper records environments without records 2.0. But in the case of accounting the assessments of records 2.0, the digital and paper records' environments together, the readiness rate of BA's documentary environment to use big data tools decreases to (59%). Finally, the study recommends that BA must improve the regulatory reality of its records so as not to lose an essential source of big data in the future. Furthermore, BA can cooperate with institutions that have a background in dealing with(using) records as a source of big data to benefit from their experiences in the field.

Keywords: Big Data; Current Records management; Bibliotheca Alexandrina

1.المقدمة:

يرجع ظهور مصطلح البيانات الضخمة big data إلى نهاية القرن العشرين؛ فقد ظهر باعتباره نتيجة للتطور التكنولوجي الهائل، الذي يشهده العصر الآن. وأصبحت هذه البيانات حقيقة ملموسة لا يمكن للمؤسسات المختلفة أن تتكرها؛ فهي نتيجة طبيعية للتوسع في استخدام تطبيقات الويب المختلفة في ممارسات العمل وإجراءاته. وأصبحت هذه البيانات تؤثر على كيفية إدارة الوثائق في المؤسسات، وفي عملية صنع القرار فيها.

1/1. مشكلة الدراسة

تمثلت مشكلة الدراسة في أن مكتبة الإسكندرية قد بدأت عصر البيانات الضخمة؛ عن طريق: أولًا. تبنيها التحول الرقمي في نطاق عملها الإداري، ثانيًا. اسهاماتها في مشروعات البيانات الضخمة، ثالثًا. استخدامها تطبيقات الويب 2,0 في التواصل مع الجمهور؛ مما يبشر – في المستقبل – بأن وثائقها قد تصبح مصدرًا مهمًا للبيانات الضخمة. ومن ثم تهدف هذه الدراسة إلى تسليط الضوء على دراسة واقع بيئة المكتبة الوثائقية؛ لتتعرف على مدي صلاحية هذه البيئة لاستخدام تطبيقات البيانات الضخمة فيها؛ تمهيدا للاستعانة بها في تحليل بيانات العمل، ووثائقه عند الضرورة في المستقبل.

1/2. أهمية الدراسة

تناولت هذه الدراسة ماهية البيانات الضخمة، وعلاقتها بمكتبة الإسكندرية، كما حرصت على تحليل بيئة المكتبة الوثائقية؛ لتقرير مدي صلاحيتها؛ لتطبيق تقنيات البيانات الضخمة فيها.

1/3. أهداف الدراسة

حاولت هذه الدراسة تحقيق الأهداف الآتية:

1–التعرف على ماهية البيانات الضخمة.

2-تحديد مصادر البيانات الضخمة في مكتبة الإسكندرية.

3-التعرف على البرمجيات التي تستخدمها مكتبة الإسكندرية في معالجة بياناتها الضخمة.

4-قياس مدى صلاحية بيئة المكتبة الوثائقية؛ لاستخدام تطبيقات البيانات الضخمة في تحليل وثائق العمل فيها.

1/4. تساؤلات الدراسة

حاولت الدراسة الإجابة عن التساؤلات الآتية:

1-ما خصائص البيانات الضخمة؟

2-ما المشاريع التي نفذتها مكتبة الإسكندرية في مجال البيانات الضخمة؟

3-ما طبيعة وثائق مكتبة الإسكندرية؟

4-ما التحديات التي تواجه مكتبة الإسكندرية في تحليل وثائقها بواسطة تطبيقات البيانات الضخمة؟

1/5. منهج الدراسة، وأدواته

قد اعتمدت الدراسة على منهج دراسة الحالة؛ للتعرف على واقع البيانات الضخمة في مكتبة الإسكندرية، ومصادرها وعلاقتها ببيئة المكتبة الوثائقية. وقد استعانت الباحثة بالملاحظة، والمقابلة الشخصية المقننة مع العاملين في مختلف قطاعات مكتبة الإسكندرية؛ لجمع معلومات حول مجالي إدارة وثائقها، ومصادر بياناتها الضخمة.

1/6 حدود الدراسة

تمثلت حدود الدراسة في: أولًا: حدود موضوعية: تبحث هذه الدراسة في مفهوم البيانات الضخمة، وعلاقتها بإدارة الوثائق. ثانيًا. الحدود المكانية: تناولت هذه الدراسة واقع البيانات الضخمة في مكتبة الإسكندرية، والتحديات التي قد تواجه معالجة وثائقها الجارية بواسطة تقنيات البيانات الضخمة. ثالثًا. الحدود النوعية: ركزت هذه الدراسة على العلاقة بين البيانات الضخمة، ووثائق المكتبة الإدارية الجارية؛ باعتبارها مصدرًا محتملًا للبيانات الضخمة في المستقبل؛ ولا سيما مع التوسع الرقمي الذي تتبناه المكتبة في مجال عملها الإداري. رابعًا. الحدود الزمنية: أُجريت هذه الدراسة في أثناء الفترة الممتدة من عام 2019عام حتى عام 2020م.

1/7. الدراسات السابقة

بدأ مصطلح البيانات الضخمة في الانتشار بين المهتمين بعلوم الويب منذ بداية القرن الحالي، وقد تبين للباحثة عن طريق تصفح الإنتاج الفكري في مجال البيانات الضخمة؛ أنه قد طرأ تطور في معالجة موضوع البيانات الضخمة عبر مرور السنوات؛ فنجد أن الدراسات التي ترجع إلى العقد الأول من القرن الواحد والعشرين ركزت على تحديد مفهوم البيانات الضخمة، وماهيتها، وقد برز هذا الاتجاه بوضوح في دراسة الباحث Douglas Laney، والتي نشرت عام 2001م بعنوان "3D Data Management: Controlling Data Volume, Velocity and Variety". وقد حدد فيها Douglas Laney سمات البيانات الضخمة، وعرفها بأنها البيانات التي تتصف بالكم الهائل، والتنوع، والسرعة العالية في التوالد والتدفق[1] وقد ظهر هذا الاتجاه أيضا في تقرير صدر من شركة جوجل في عام 2008 بعنوان "Big data: the next Google"، والذي أشار إلى أسباب نمو البيانات الضخمة؛ والتي تمثلت في التطور التكنولوجي في خدمات الإنترنت، وبيئات الويب؛ الأمر الذي ساعد على زيادة التفاعل بين الأفراد، والآلات من ناحية، وتعزيز تبادل البيانات المعقدة المختلفة بكميات هائلة، وبسرعة شديدة من ناحية أخرى؛ مما دفع إلى توالد البيانات الضخمة، وتدفقها[2]. ومع حلول العقد الثاني من هذا القرن؛ بدأت الدراسات، والأبحاث تهتم أكثر بمجال أرشفة البيانات الضخمة،

وتحليلها، وأمنها، واستخداماتها في المؤسسات المختلفة. ففي عام 2012م ظهرت دراسة بعنوان "Perspectives on Big Data and Big Data Analytics." أشارت إلى أهمية البيانات الضخمة في صنع القرارات في الشركات؛ ولا سيما في مجال التسويق، كما استعرضت البرمجيات المستخدمة من قبل الشركات في تحليل هذه البيانات؛ ولا سيما برنامج أباتشي هادووب (**Apache** Hadoop) مفتوح المصدر[3].[4] وفي عام 2013م ظهرت دراسة أخري بعنوان "Big Data Challenges" تطرقت هذه الدراسة إلى مجال أرشفة البيانات الضخمة، والتحديات التي قد تواجهها المؤسسات عند أرشفتها –منها على سبيل المثال : محدودة سعة التخزين المتاحة للبيانات، وضعف البنية التكنولوجية التحتية للمؤسسة –والتي يجب أن تتعامل معها المؤسسة؛ حتي لا تعيق عملية أرشفة بياناتها الضخمة[5]. وفي عام 2015م نُشر بحث بعنوان " Data Archival using Hadoop Distributed File system."تناول المخاطر التي قد تتعرض لها المؤسسة؛ إذا ما استخدمت طرق أرشفة البيانات الورقية التقليدية في أرشفة بياناتها الضخمة؛ مما يؤثر سلبًا على دقة نتائج تحليل هذه البيانات، ودقة عملية صناعة القرار في المؤسسة. كما استعرض هذا البحث ميزات استخدام برنامج Apache Hadoop في معالجة البيانات الضخمة وأرشفتها[6]. وفي عام 2017م ظهرت دراسة مهمة بعنوان " "Big Data Issues for Federal Records Managers." تناولت التحديات الرئيسة التي قد يواجها مديرو الوثائق في التعامل مع أرشفة البيانات الضخمة، وتقييمها، وشروط الاحتفاظ بها في مؤسساتها، وتكلفة حفظها؛ ولا سيما أن هذه البيانات تلعب في العصر الحالي دورًا خطيرًا في عملية صنع القرار في المؤسسات[7].وفي عام 2018م ظهرت دراسة عربية بعنوان "واقع استخدام البيانات الضخمة في هيئة الوثائق، والمحفوظات الوطنية: دراسة استطلاعية " تناولت البيانات الضخمة من منظور وثائقي أرشيفي ؛فقد استعرضت العلاقة بين الأرشيفات الوطنية، والبيانات الضخمة؛ فعن طريق تسليط على واقع البيانات الضخمة في هيئة الوثائق، والمحفوظات الوطنية العمانية، والصعوبات التي قد تواجه الهيئة في تطبيق تقنيات تحليل البيانات الضخمة فيها، وقد أكدت الدراسة على أهمية تطوير الإمكانات المادية، والتكنولوجيا، والبشرية لدي الهيئة؛ لكي تستطيع إدارة بياناتها الضخمة المستقبلية[8]. وفي عام 2019 م نشرت دراسة عربية بعنوان "دور البيانات الضخمة في دعم التنمية المستدامة بالدول العربية" وتمتاز هذه الدراسة بأنها تناولت العلاقة بين البيانات

الضخمة، وأهداف التنمية المستدامة، وكيف يمكن للدول الاستفادة من البيانات الضخمة في تحقيق أهدافها التنموية في نواحي الحياة المختلفة ؛ من : الاجتماعية، والاقتصادية، والثقافية...إلخ؛ للارتقاء بالبيئة الفكرية، والاجتماعية، والصحية المختصة بالمواطن؛ وانتهت الدراسة إلى أهمية الشراكة، وتوفير مصادر البيانات الضخمة، ووثائقها لمؤسسات الدول ؛لكي تستطيع توظيف في البيانات الضخمة في مجال التنمية المستدامة بطريقة فعالة ؛ وتحسين صناعة القرارات بها[9]. وتختلف دراسة الباحثة عما سبق؛ بأنها تستعرض واقع البيانات الضخمة في مكتبة الإسكندرية، وعلاقتها بوثائقها الإدارية الجارية، ومدى صلاحية بيئة المكتبة الوثائقية لاستخدام تقنيات البيانات الضخمة في تحليل وثائق عملها الجارية.

2.البيانات الضخمة

هي مجموعة معقدة، وضخمة، ومتنوعة من البيانات؛ التي قد تكون مهيكلة structured، أو شبه مهيكلة semi-structured، أو غير مهيكلة unstructured؛ ولا يمكن معالجتها، وتحليلها ببرمجيات تحليل البيانات التقليدية، وتحتاج إلى برمجيات متخصصة[10]. وتتمتع البيانات الضخمة بمجموعة من السمات؛ ولكن حدد المتخصصون سبع خصائص أساسية؛ ينبغي للبيانات الضخمة أن تتمتع بها، وقسموها إلى مجموعتين؛ الأولى: مجموعة خصائص الطبيعة **Natural**، وتضم ثلاث خصائص؛ هي: خاصية الحجم Volume: فالبيانات الضخمة تمتاز بالكم الهائل، والذي قد يصل إلى زيتا بايت- والذي يعادل 1021 بايت- وهذا يتطلب سعات تخزينية كبيرة؛ قد تعجز بعض المؤسسات أن توفرها؛ مما يجعلها تفقد بياناتها الضخمة. أما بالنسبة إلى خاصية السرعةvelocity : فالبيانات الضخمة تنتج، وتتدفق، وتستخرج بسرعة شديدة؛ تعجز برامج معالجة البيانات التقليدية التعامل معها؛ ومن ثم فإن تقادم هذه البيانات، أو تعديلها، أو مسحها يمثل تحدياً كبيرًا يواجه صانعي القرار[11]. في حين أن خاصية التنوع Variety؛ تشير إلى ثلاث جزئيات، هي: أولًا. جزئية التنوع الشكلي: ويقصد بها تنوع النمط الذي تظهر عليه البيانات الضخمة؛ فقد تكون عبارة عن نص، أو صورة، أو تسجيل صوتي ...أو قد تتاح في أكثر من شكل في الوقت نفسه على سبيل المثال: مقاطع الفيديو التي تجمع بين الصورة، والصوت. ثانيًا جزئية تنوع المحتوي: ويقصد بها أن البيانات الضخمة قد تتفاوت بشدة في التعبير

عن تفاصيل موضوعية مختلفة؛ فعلى سبيل المثال: قد تمتلك المؤسسة كمًا ضخمًا من البيانات النصية، ولكن مدلولها الموضوعي متفاوت بشدة؛ ومن ثم تصنف هذه البيانات تحت فئة البيانات الضخمة. ثالثًا: جزئية التنوع في البنية : فقد تتمتع البيانات الضخمة بإحدى هذه البنيات، أو أكثر؛ فقد تكون مهيكلة *structured* ؛ أي منظمة، ومرتبة وفقًا لنماذج تنظيميةorganizational models محددة –مثل البيانات المتاحة عبر قواعد البيانات، أو تكون شبه مهيكلة *Semi-structured* وهي نوع من البيانات المنظمة، لكنها لا تلتزم بالنماذج أو قوالب ثابتة، ورسمية – مثل الجداول – في عرضها ؛ ولكن تستخدم عناصرelements، والفئات classes، والتيجان Tags؛ لعرضها بوضوح وسلاسة؛ ومثل تسجيلات XML .أما بالنسبة للبيانات غير المهيكلة *Unstructured* ؛ فهي البيانات التي تعرض بطريقة حرة، ولا تستخدم قوالب تنظيمية لعرضها، مثل: الصور، والنصوص النصية المتاحة عبر الشبكات الاجتماعية[12].

على جانب أخر تركز مجموعة خصائص المعالجة **Processing**–والتي تضم أربع خصائص– على المتطلبات التي يجب أن تضعها المؤسسات في حسبانها عند معالجة هذه البيانات؛ فنجد خاصية الصدق **veracity**، والتي يقصد بها أنه ينبغي للمؤسسة أن تتأكد من دقة من بياناتها الضخمة، وحداثتها، وموثوقية مصدرها، وصحة سياقها قبل أن تستخدمها في صنع قراراتها. أما خاصية التباينvariability **:** فيقصد بها أنه يجب على المؤسسة إدراك أن سياق الذي تتداول عبره البيانات الضخمة قد يؤثر على مدلولها – فعلى سبيل المثال استخدام كلمة عظيم في سياق المدح يعكس شعورًا إيجابي، وقد يكون لها تأثير سلبي عند استخدامها في سياق ذم شخص ما، وتوبيخه. في حين أن خاصية القيمة **value** تعني أن البيانات الضخمة ليس لها قيمة في حد ذاتها. ولكن قيمتها تعود إلى أهميتها في مجال، أو نشاط معين؛ ومن ثم تصنفها المؤسسات وفقًا لمجال استخدامها إلى البيانات الاقتصادية والسياسية، والاجتماعية، والتعليمية...أما خاصية التمثيل**Visualization** : فيقصد بها أن المؤسسة يجب أن تستعين بتطبيقات التمثيل البصري الرسومي–مثل الأشكال التوضيحية، والرسوم البيانية –للتعبير عن نتائج تحليل البيانات الضخمة وشرحها بكفاءة؛ لأن التمثيل الرسومي –مقارنة بالنصوص – أكثر نجاحًا في توصيل المعلومات إلى صناع القرار.[13]

3.البيانات الضخمة، ومكتبة الإسكندرية

تعد مكتبة الإسكندرية من أقدم المكتبات العامة، التي ظهرت في التاريخ؛ فترجع جذورها إلى العصر البطلمي، وقد ضمت مكتبة الإسكندرية القديمة بين جدرانها آلاف من المؤلفات؛ لا سيما مؤلفات الحضارتين الإغريقية والفرعونية، ولكنها –للأسف تعرضت لعدد من الحرائق قضت عليها تمامًا في عصر الإمبراطورية الرومانية، وفي القرن العشرين بدأت تظهر دعوة لإحياء هذه المكتبة مرة أخري، وإعادة إنشائها. وقد نالت هذه الدعوة القبول من قبل منظمة اليونسكو؛ التي تعاونت مع الحكومة المصرية في إعادة بناء هذه المكتبة في مكانها السابق في مدينة الإسكندرية، وقد ساهم عدد كبير من دول العالم في إثراء مقتنيات هذه المكتبة بإهدائها عدد ضخم من الكتب، والمؤلفات، وفي عام 2002م فتحت مكتبة الإسكندرية الحديثة أبوابها للجمهور [14].وتتشكل المكتبة من عدد من القطاعات، منها: قطاع تكنولوجيا المعلومات والاتصالات، وقطاع المكتبات، والقطاع الإداري والمالي، وقطاع التواصل الثقافي، وقطاع البحث الأكاديمي...إلخ. وكل قطاع يتكون من عدد من الإدارات، والمراكز – فعلي سبيل المثال: قطاع التواصل الثقافي: يضم مركز المخطوطات، وإدارة سفارات المعرفة، وإدارة المعارض والمقتنيات الفنية، ومركز الدكتور زاهي حواس للمصريات، ووحدة الوثائق التاريخية ...إلخ. [15]وتضم مكتبة الإسكندرية كمًا متنوعًا، وضخمًا من مصادر التراث الثقافي المختلفة من: الكتب، والمخطوطات، والوثائق التاريخية، والصور، والقطع المتحفية، والتسجيلات الصوتية ...إلخ. ويزور مكتبة الإسكندرية سنويًا الملايين من المصريين، والأجانب الذين ينتمون إلى جنسيات مختلفة.

وتهتم مكتبة الإسكندرية بدعم المجال الثقافي على مستوي المحلي، والعالمي، ولذلك تحرص دائمًا على مواكبة التطورات العلمية، والثقافية، والتكنولوجية أولًا بأول؛ لكي توظفها في تحسين أداء عملها، ورفع كفاءة خدماتها لمستفيديها من مختلف دول العالم؛ ومن ثم قد أولت مجال البيانات الضخمة اهتمامًا كبيرًا، وقد انضمت إلى المتعاملين مع البيانات الضخمة عن طريق مشروعين؛ هما: مشروع أرشيف الويب **The BA Internet Archive**، وموسوعة الحياة **Encyclopedia of Life**. أما بالنسبة إلى وسائل الإيجاد المختصة بمكتبة الإسكندرية –مثل فهارسها – فلا يمكن اعتبارها في الوقت الحالي مصدرًا من مصادر البيانات

الضخمة؛ ولاسيما أنها لا تنطبق عليها سمات البيانات الضخمة؛ التي تركز على تنوع البيانات؛ فهي تتمتع بقلة كمية البيانات التي تتيحها، وببنيتها المهيكلة الموحدة، وواصفاتها المحددة؛ ومن ثم يمكن مكتبة الإسكندرية استخدام تقنيات معالجة البيانات التقليدية في تحليلها؛ ومن دون استخدام برمجيات البيانات الضخمة المعقدة.[16]

ويدور مشروع المكتبة المختص بأرشيف الإنترنت حول أرشفة مواقع الإنترنت عن طريق التقاط لقطات snapshots لصفحات مواقعه، وحفظها. وترجع أصول هذا المشروع إلى عام 2002م عندما عقدت المكتبة اتفاق مع مبادرة أرشيف الإنترنت الأمريكية، والتي أُنشئت في سان فرانسسكو بواسطة مهندس الحاسبات الآلية، ورائد الإنترنت Brewster Kahle؛ بهدف أرشفة مواقع الويب archiving web. وقد نص الاتفاق على: 1-أن تمنح مبادرة أرشيف الإنترنت الأمريكية مكتبة الإسكندرية نسخة احتياطية backup من أرشيف مواقع الإنترنت الذي يغطي الفترة الممتدة من عام 1996م إلى عام 2001م. 2-إنشاء موقعين للمكتبة على غرار موقع مبادرة أرشيف الإنترنت الأمريكي الأصلي؛ كما أهدت مبادرة أرشيف الإنترنت الأمريكية مكتبة الإسكندرية -وفقًا لهذا الاتفاق-الجيل الأول من البرمجيات، والأجهزة التي تستخدم في أرشفة مواقع الإنترنت[17]. وفي عام 2006م طورت مكتبة الإسكندرية بالتعاون مع مبادرة أرشيف الإنترنت الأمريكية البنية التحتية المختصة بمشروعها "أرشيف الإنترنت"؛ لكي ترفع السعة التخزينية له. وفي عام2008م عززت المكتبة البنية التحتية المختصة بهذا المشروع بأجهزة مصرية الصنع؛ لكي ترفع السعة التخزينية مرة أخرى، وفعلًا أصبحت سعة المشروع التخزينية تبلغ الآن 4,9 بيتابايت ((4.9 petabits أي ما يعادل تقريبًا 5018تيرابايت. مما دفع مكتبة الإسكندرية للاستفادة من هذه السعة التخزينية في أرشفة المحتوي الرقمي الذي تمتلكه، والمتمثل-غالبًا- في الكتب التي قامت المكتبة برقمنتها من قبل. كما يمتاز مشروع أرشيف الإنترنت المختص بمكتبة الإسكندرية بأنه يغطي محتوى متنوع موضوعيًا؛ ومتاح بلغات مختلفة مثل اللغة الإنجليزية، واللغة الفرنسية ...، وخاصة اللغة العربية، ومن ثم لا يمكن اعتبار مشروع مكتبة الإسكندرية لأرشفة الإنترنت بمثابة نسخة احتياطية من

مبادرة أرشيف الإنترنت الأمريكية، ولكنه كيانًا ثقافيًا وإعلاميًا يخدم كل دول العالم؛ ولا سيما قارة أفريقيا، ومنطقة الشرق الأوسط[18].

أما بالنسبة إلى مشروع موسوعة الحياة: فهو عبارة عن موسوعة بيولوجية رقمية تعاونية collaborative على الخط المباشر متاحة بعدد من اللغات –من الإنجليزية، والعربية، والتركية، والفرنسية... تختص بتوثيق الحياة الطبيعية، وكائناتها على سطح الأرض من بكتريا، وحيوانات، وطيور، ونباتات، وفطريات، وغيرها من الكائنات الحية الأخرى؛ التي تحيا في أنحاء العالم المختلفة. وتعد هذه الموسوعة من أهم المنصات التعليمية العلمية البحثية في المجال البيولوجي. ويرجع تأسيس هذه الموسوعة إلى عام 2008م برعاية المتحف الوطني للتاريخ الطبيعي التابع لمعهد سميثسونيان Smithsonian Institution's National Museum of Natural History؛ وعند إنشائها كانت تعرض معلومات عن ثلاثين ألف 30,000كائن فقط، أم الآن فهي تغطي حوالي مليونين 2000,000 كائن يعيش على سطح الأرض[19]. وتقدم هذه الموسوعة محتوى نصيًا، وصوتيًا، وبصريًا؛ يوثق الحياة على سطح الأرض، وقد جُمع هذا المحتوي من مصادر المعلومات المختلفة من كتب، وقواعد البيانات، ومواقع الويب...وغيرها، بالإضافة إلى معرفة الأفراد غير الموثقة في مصادر المعلومات. وتعد مكتبة الإسكندرية عضوًا في اللجنة التنفيذية لهذا المشروع إلى جانب عددًا من المؤسسات العلمية، والأكاديمية؛ منها على سبيل المثال : المتحف الوطني للتاريخ الطبيعي التابع لمعهد سميثسونيان، أطلس حياة أستراليا Atlas of Living Australia، جامعة هارفرد Harvard University، معمل الأحياء البحرية Marine Biological Laboratory . وفي عام 2018م تم إطلاق منصة platform الإصدارة الثالثة Third versionمن الموسوعة؛ والتي تميزت بتطوير سمات البحث المتقدم Advanced search القائم على الربط بين المصطلحات البيولوجية المختلفة؛ لكي تعزز نتائج البحث، ورفع كفاءة الاسترجاع. [20]

وينتج عن هذين المشروعين –سابقي الذكر–سنويًا كمًا هائلًا، ومتنوعًا، وسريعًا من البيانات؛ نتيجة للإثراء المستمر في محتواهما؛ ولذلك تلجأ المكتبة إلى الاستعانة بتطبيقات تحليل البيانات الضخمة؛ لإصدار تقارير متابعة العمل المختصة بهذين المشروعين؛ للوقوف على التطور في محتواهما كمًا، ونوعًا.

وتطبق مكتبة الإسكندرية عددًا من المراحل الآلية المتتالية؛ لمعالجة بياناتها الضخمة، وتحليلها –مستخدمة تطبيق Apache Hadoop مفتوح المصدر open source. وتبدأ عملية معالجة بيانات المكتبة الضخمة بمرحلة تزويد البيانات Data Acquisition، فيقوم برنامج Hadoop بجمع بيانات المشروع المراد إصدار تقرير عنه من الخوادم Servers المختصة به، ويقسمها إلى كتل blocks، موزعة على ملفات Files؛ كل ملف يضم كتلة، أو أكثر من البيانات، ثم يقوم هادوب بواسطة نظام الملفات الموزعة **Hadoop Distributed File System (HDFS**[21] بتوزيع هذه الملفات على شبكة من Nodes.[22] فيقوم كل Node بواسطة نموذج MapReduce[23] بمعالجة ملف البيانات المختص به. فيقوم node بتطهير ملفه (مرحلة تطهير البيانات Data Cleansing)، وتنقيته من البيانات الخاطئة، أو الغامضة، أو القديمة المستبعدة، والتي تم تزويد برنامج هادوب بسماتها. ثم يقوم Node بتصنيف بيانات ملفه (مرحلة اكتشاف البيانات Data Exploring)، ويتم فيها تحديد طبيعة البيانات – مثل: اللون، والحجم – وخصائصها – مثل: أسود كبير – تمهيدًا لتجميع البيانات المتشابهة معًا. ثم تبدأ مرحلة تحليل البيانات Data Analysis باستخدام Aprache Pig[24]، والذي يحلل الروابط بين البيانات المختلفة وفقًا لخريطة البيانات Data Map المعدة مسبقًا؛ والتي تعبر عن العلاقات، والروابط بين البيانات المختلفة، بعد انتهاء هذه المرحلة تقوم المكتبة بحفظ مخرجات هذه العملية في قواعد البيانات غير العلائقية NoSQL[25]؛ والتي تتميز بقدرتها الفائقة على حفظ كمًا هائلًا، ومتنوعًا من البيانات، وعلاقاتها معًا؛ وهذه القواعد هي: (HBase)[26]، و (MongoDB)[27] Neo4j.[28] ثم تبدأ مكتبة الإسكندرية مرحلة تمثيل البيانات Data Visualization؛ فتستخدم برنامج التمثيل البياني (Gephi)؛ لتمثيل نتائج تحليل البيانات الضخمة بطريقة رسومية – مثل الرسوم البيانية، والأشكال التوضيحية؛ تمهيدًا لإعداد التقارير النهائية؛ ولا سيما أن الصورة– مقارنة بالنص– تعبر عن المضمون بسهولة، وبوضوح. وفي النهاية تُصدر التقارير، وتقدم إلى مدير المشروع الذي يعتمد عليها في متابعة سير العمل في المشروع، واتخاذ القرارات بشأنه. وعندما ينتهي العمل بهذه التقارير يتم الاحتفاظ بها إلى ما لا نهاية في نظام إدارة محتوى إلكتروني (Content Management System (CMS خاص ومستقل يختص بالقطاع التكنولوجيات؛ لاستردادها عند الحاجة لها.[29]

وقد تبين للباحثة في أثناء الدراسة أن مكتبة الإسكندرية لا تصدر تقارير دورية عن مشروعاتها المتعلقة بالبيانات الضخمة، ولكن تعدها –فقط– عند الحاجة إليها. كما تبين –أيضاً– أن استخدام برمجيات تحليل البيانات الضخمة، يقتصر استخدامها على نطاق العمل الرسمي داخل المكتبة، ولا يسمح باستخدامها من قبل الباحثين المترددين على المكتبة في إجراء أبحاثهم العلمية.[30]

4- البيئة الوثائقية، ومكتبة الإسكندرية

وجدت الباحثة أن وثائق المكتبة تصنف وفقًا لطبيعتها إلى ثلاث فئات؛ هي: الوثائق الرقمية Digital records، والوثائق الورقية paper records، ووثائق الجيل الثاني2,0records؛ ومن ثم تتنوع البيئة الوثائقية المختصة بالمكتبة، وفقًا لطبيعة هذه الوثائق، وفيما يلي توضيح لهذه الوثائق، وبيئاتها.

4/1- الوثائق الرقمية[31]

قد أظهرت الدراسة أن الوثائق الرقمية أصبحت جزءًا واقعيًا من منظومة المكتبة التوثيقية؛ وهذا يرجع إلى تبني المكتبة سياسة تقليل الممارسات الورقية في إنجاز مهام العمل. وتنتج وثائق المكتبة الرقمية من مصدرين: أولًا. الوثائق الناتجة عن" نظام المسار Track System"، وهو نظام لميكنة المعاملات المالية والإدارية، التي يمارسها موظفو مكتبة الإسكندرية بشكل روتيني، مثل: وثائق طلب الإجازة، وثائق مفردات مرتبات الموظفين. فيقوم الموظف بإرسال الطلب الإداري، أو المالي المراد إجراءه إلى مديره عبر Track System؛ الذي يبدي رأيه فيه؛ عندئذ تصدر وثيقة توثق قرار المدير، تحمل كود فريد يحدد هويتهاId–مثل 3c73499rt98736– ثم يتم حفظها في قاعدة البيانات المختصة بأرشيف النظام نفسه بعد وصف محتواها، وسياقها الإداري الذي صدرت خلاله آليًا؛ ومن دون ربطها ببقية وثائق المكتبة الإدارية. ثانيًا. رسائل البريد الإلكتروني: تعتبر مكتبة الإسكندرية رسائل البريد الإلكتروني نوعًا من المكاتبات الرسمية بين العاملين، وتحافظ على رسائل البريد الإلكتروني، من دون وصف أو تصنيف، في قاعدة بيانات مخصصة للبريد الإلكتروني تتصل ببوابة العمل الداخلية للمكتبة لمدة 180 يومًا، ثم يتم تحذفها. وتتبع المكتبة النظام المركزي في حفظ وثائقها الإلكترونية،

ويتولى الإشراف على أرشفتها المسؤولون عن تكنولوجيا المعلومات في المكتبة، ويرجع السبب إلى ذلك إلى اعتبار الوثائق الرقمية جزء من منظومة التحول الرقمي الإداري الذي يشرف عليه القطاع التكنولوجي في المكتبة[32]. كما تبين من الدراسة أن وثائق المكتبة الرقمية تنتمي إلى فئة الوثائق النصية، وأغلبها باللغة الإنجليزية.

4/2- الوثائق الورقية

أما إذا تطرقنا إلى فئة الوثائق الورقية ؛فمكتبة الإسكندرية لا تزال تستخدم الوثائق الورقية في إنجاز المعاملات المالية والإدارية والتنظيمية، والهندسية ... التي يترتب عليها آثار قانونية مهمة – مثل: بروتوكولات التعاون، وعقود الشراء، وعقود التعيين، واللوحات الهندسية؛ ويرجع السبب وراء عدم اكتفاء المكتبة بالوثائق الرقمية من دون الورقية في ممارسات العمل؛ إلى أنها لم تطبق منظومة التوقيع الرقمي في وثائقها الرقمية حتي الآن؛ ومن ثم لا تتمتع وثائقها الرقمية بنفس حجية الوثائق الورقية القانونية؛ لأن القانون المصري يعترف بقانونية، ورسمية الوثائق الرقمية التي تحمل توقيع رقمي فقط. وتتنوع وثائق المكتبة الورقية بين الوثائق النصية: مثل القرارات الإدارية، والبصرية مثل اللوحات الهندسية، وتتنوع وثائق المكتبة الورقية لغويًا؛ فعلى الرغم من أن أغلبها باللغة العربية؛ ولكنها تحتفظ بوثائق مكتوبة باللغة الإنجليزية، والفرنسية ... والتي تتناول –عادة – الاتفاقيات، وبروتوكولات التعاون بين المكتبة والمؤسسات الدولية الأجنبية. [33]

وتتبع المكتبة النظام المركزي، واللامركزي في إدارة وثائقها الورقية؛ حيث اتضح أن الوثائق المتعلقة بالجوانب الإدارية، والمالية، والقانونية المختصة بمكتبة الإسكندرية وقطاعاتها؛ تُحفظ حفظًا مركزيًا. فنجد أن الوثائق المالية –مثل: المناقصات، والميزانيات، والتقارير المالية –تحفظ حفظًا مركزيًا في وحدة الوثائق التابعة للقطاع الإداري والمالي، في حين تحفظ وثائق الأفراد وملفاتهم–مثل: أوراق التعيين، والترقية، والندب، والإجازات– حفظًا مركزيًا في وحدة السجلات التابعة لقطاع الموارد البشرية والخدمات العامة، أما الوثائق القانونية –مثل: القضايا، والعقوبات – تحفظ حفظًا مركزيًا في أرشيف إدارة الشئون القانونية. أما بالنسبة للوثائق المتعلقة فقط بأنشطة قطاع، أو إدارة office، أو مركز محدد في المكتبة؛ في تحفظ لا مركزيًا؛ فكل قطاع،

أو إدارة office، أو مركز مسؤول عن حفظ وثائق أنشطته المتخصصة، وضبطها – فعلى سبيل المثال مركز المخطوطات يحتفظ بوثائقه المتعلقة بورش العمل التي ينظمها، وخطط عمله– كما أن المكتبة لا تطبق نظامًا موحدًا في تصنيف وثائقها الورقية؛ بل يتفاوت نظام التصنيف من نوع لأخر من الوثائق، ومن قطاع لأخر، ومن إدارة لأخرى؛ ويرجع السبب في ذلك إلى رغبة المسؤولين في تطبيق نظام تصنيف يتوافق مع احتياجات العمل وأنشطته داخل كل قطاع، وإدارة[34]؛ فتصنف ملفات الأفراد بكود الفرد الذي يمنح له عند تعينه، وتصنف وثائق القطاع الإداري موضوعيًا؛ لتعبر عن الأنشطة، والمعاملات الإدارية المختصة بالمكتبة، أما الوثائق المالية فترتب زمنيًا بالسنة المالية، ومن زاوية أخري هناك قطاعات تستخدم الترميز الرقمي للتعبير عن الفئات التصنيفية لوثائقها المصنفة، في حين أن هناك قطاعات تكتفي بالمسمى اللفظي للنشاط؛ للتعبير عن الفئات التصنيفية لوثائقها. كما تبين من الدراسة أن المكتبة لا تطبق نظام معياري موحد في وصف وثائقها الورقية؛ بل تكتفي –غالبًا–بالبيانات الرئيسة عن الوثيقة فقط، وهي: عنوان الوثيقة، وتاريخ إنشائها، ومضمونها من دون وصف سياقها. كما تبين –أيضًا– أن عملية أرشفة الوثائق الورقية رقميًا تتفاوت من قطاع لأخر، ويرجع السبب في ذلك إلى أن القطاعات التي ترغب في تعزيز الحجية القانونية لوثائقها؛ لا تطبق الرقمنة على وثائقها؛ لأن النسخة الرقمية من الوثائق الورقية ليس لها حجية قانونية وفقًا للقانون المصري؛ لذلك تكتفي بحفظ أصول الوثائق الورقية في ملفات، ومن دون حفظها إلكترونيًا – مثل وثائق المناقصات، وعقود الشراء. في حين أن القطاعات التي تحتاج إلى تيسير استرجاع وثائقها؛ لمتابعة تطورات العمل مثل –قطاع الموارد البشرية – تميل إلى حفظ وثائقها رقميًا في نظام إدارة المحتوي الإلكتروني المختص بها إلى جانب حفظ أصولها الورقية في ملفات[35].

وتجدر الإشارة إلى أن نظم إدارة المحتوي الإلكتروني المختصة بقطاعات المكتبة يمكن الربط بينها؛ لإنشاء شبكة متماسكة، ومتكاملة لإدارة مصادرها، ووثائقها المختلفة معًا.

4/3 وثائق الجيل الثاني 2.0 Records

أما بالنسبة إلى وثائق الجيل الثاني 2.0 Records؛ فيقصد بها البيانات، والوثائق المنتجة عبر تطبيقات الويب 2.0 2,0 applications of web – مثل شبكات التواصل الاجتماعي–بطريقة رسمية في سياق عمل المؤسسة. وقد أظهرت الدراسة أن مكتبة الإسكندرية شأنها شأن الكثير من المؤسسات التي تستخدم مواقع: توتير، وفيس بوك، ولينكد إن Twitter, LinkedIn, and Facebook باعتبارها قنوات إعلامية للتواصل مع الجمهور فقط، ومن دون النظر لقيمتها الرسمية لإنتاج الوثائق التي تعكس الرأي العام، وتعزز المشاركة التعاونية مع الجمهور. وقد أظهرت الدراسة أن استخدام مكتبة الإسكندرية لشبكات التواصل الاجتماعي قد نتج عنه كمًا من وثائق الجيل الثاني؛ فقد اتضح في أثناء تصفح صفحة مكتبة الإسكندرية على الفيس بوك وجود مقاطع فيديو توثق أنشطة المكتبة التي نفذتها، ومن ثم تعد هذه المقاطع بمثابة وثائق رسمية يجب أن تحرص المكتبة على أرشفتها. كما اتضح–أيضا–من الدراسة أن مكتبة الإسكندرية لا تطبق أي قواعد أو معايير لأرشفة محتوى صفحاتها، ووثائقها المتاحة عبر هذه التطبيقات. ولا ننكر أن المكتبة حاولت عبر مشروع أرشيف الإنترنت أن تؤرشف موقع الفيس بوك باعتباره مصدرًا معرفيًا–وليس باعتباره مصدرًا وثائقيًا رسميًا – باستخدام برنامج Heritrix؛ وهو برنامج زاحف ويب web crawling، لكن المحاولة باءت بالفشل؛ بسبب البنية التكنولوجية المعقدة المختصة بموقع الفيس بوك التي أعاقت عملية الزحف إلى محتوي الفيس بوك، والتقاط لقطات منها[36]. وتتنوع وثائق المكتبة للجيل الثاني بين النصية، والسمعية، والبصرية؛ مثل: مقاطع الفيديو، والصور، والتسجيلات الصوتية، كما تتفاوت وثائق المكتبة للجيل 2.0 لغويًا؛ فقد تكون باللغة العربية، أو الإنجليزية، أو الفرنسية[37].

5–العلاقة بين الوثائق الإدارية، والبيانات الضخمة في مكتبة الإسكندرية

أصبحت البيانات الضخمة واقعًا ملموسًا في مكتبة الإسكندرية؛ ومن ثم وجب إلقاء الضوء على العلاقة بين وثائق المكتبة الإدارية الجارية، والبيانات الضخمة. وقد تبين من الدراسة أن وثائق المكتبة الإدارية الجارية – الآن–تتمتع بتنوع المحتوي، والبنية، ولكنها تتصف ببطء النمو، والاسترجاع، وتحتوي كميات قليلة من

البيانات لا تصل إلى تيرابايت، ومن ثم لا تعد مصدرًا حاليًا للبيانات الضخمة في المكتبة. ولكن من المتوقع مع التحول الرقمي الذي تتبناه مصر، ومع تطبيق منظومة العمل الإلكتروني في المكتبة؛ أن تصبح وثائق المكتبة الإدارية مصدرًا للبيانات الضخمة في المستقبل. ومن ناحية أخري تعد البيانات الضخمة في مكتبة الإسكندرية مصدرًا حيويًا لإنتاج الوثائق المختصة بمشروعي (أرشيف مكتبة الإنترنت (Web) وموسوعة الحياة)؛ حيث يتولد من تحليل بيانات هذين المشروعين عددًا من الوثائق المتمثلة في التقارير. وقد أظهرت الدراسة أن هذه التقارير تحفظ حفظًا مركزيًا إلى ما لانهاية في نظام إدارة محتوي إلكتروني مستقل، ومنفصل عن بقية نظم إدارة المحتوي الإلكتروني المستخدمة من قبل بعض القطاعات في حفظ نسخ إلكترونية من وثائقها الإدارية –مثل القطاع الإداري والمالي، وقطاع الموارد البشرية؛ ومن ثم يمكن القول على الرغم من وجود الوثائق الإدارية، والبيانات الضخمة جنبًا إلى جنب في المكتبة في الوقت الحالي؛ إلا أن كلًا منهما يخضع إلى إشراف مستقل، وضبط منفصل؛ يحدد قواعده القطاع الإداري، أو الفني المسئول عنها.[38]

6- التحديات التي قد تواجه وثائق المكتبة؛ باعتبارها مصدرًا للبيانات الضخمة.

تتوقع الباحثة أن مع التحول الرقمي الذي تطبقه المكتبة في مجال العمل، والشراكات المؤسسية المعلوماتية الذي تتبناها، والتوسع في توظيف تكنولوجية الاتصالات في تلبية خدمات المواطنين من قبل الدولة المصرية؛ كلها عوامل تبشر باحتمالية أن تصبح وثائق المكتبة الجارية مصدرًا حيويًا للبيانات الضخمة؛ ومن ثم كان الحاجة لتحليل واقع بيئة المكتبة الوثائقية؛ لمعرفة مدي استعدادها لاستخدام برمجيات البيانات الضخمة فيها.

وقد حددت الباحثة في بحث سابق لها بعنوان " البيانات الضخمة، وإدارة الوثائق "[39]– والذي قدمته في ملتقي الأرشيفيين الثاني؛ الذي نظم من قبل منظمة التنمية الإدارية العربية –إحدى مؤسسات جامعة الدول العربية– في عام 2018م – إحدى عشر (11) عنصرًا؛ تمثل التحديات التي يجب أن تعالجها أي مؤسسة قبل أن تستخدم تطبيقات البيانات الضخمة في تحليل وثائقها الإدارية الجارية[40]. وفيما يلي جدول يحدد التحديات التي قد تواجه كل فئة من فئات وثائق المكتبة عند معالجتها بتطبيقات البيانات الضخمة في المستقبل.

جدول 1يعرض العناصر المستخدمة في تقييم بيئات وثائق مكتبة الإسكندرية الثلاث

عناصر التقييم(التحديات)	الوثائق الرقمية	الوثائق الورقية	وثائق الجيل الثاني [41]
التشريعات	√	√	√
أخصائيي الوثائق	√	√	√
حوكمة الوثائق	√	√	√
وصف الوثائق	√	√	√
التجميعات الموضوعية	√	√	√
التقييم، ومدد الحفظ	√	√	√
عدم التكرار	√	√	√
الصيغ الرقمية التفاعلية	√	√	√
البنية التحتية للحفظ طويل الأجل	√	√	√
ملكية المعلومات والخصوصية			√
الذاتية في مقابل المهنية			√

وفيما يلي تحليل لواقع بيئات وثائق المكتبة الثلاث وفقًا لهذه التحديات؛ لمعرفة النواحي الإيجابية، والسلبية في بيئة المكتبة الوثائقية.

7. تحليل واقع بيئة المكتبة الوثائقية

لا ننكر أن التعامل مع وثائق المكتبة الجارية باعتبارها مصدرًا للبيانات الضخمة؛ يتطلب توافر عددًا من الأمور في بيئة المكتبة الوثائقية؛ لضمان معالجة وثائقها بكفاءة بواسطة تطبيقات البيانات الضخمة؛ لتحقيق الاستفادة القصوى من بياناتها في العمل.

فإذا تطرقنا إلى **عنصر التشريعات Legislation. (التحدي الأول)**، والذي يعد من أهم التحديات التي تواجه التعامل مع الوثائق باعتبارها منبعًا للبيانات الضخمة. لا ننكر أن تعارض تشريعات البيانات، والوثائق الرسمية المحلية مع استراتيجيات أي مؤسسة وخططها المعلوماتية؛ قد يعيق معالجة بياناتها، ووثائقها بكفاءة؛ إذ قد يفقد معلوماتها المصداقية؛ لا سيما أن تشريعات البيانات والوثائق تركز دائمًا على الحفاظ على أصالة البيانات، وموثوقيتها. ولذلك حرصت مكتبة الإسكندرية على تطبيق التشريعات المصرية المختصة بمجالي البيانات، والوثائق[42]، كما تحرص على عدم اتخاذ أي قرارات تتعلق بشئونها الوثائقية تتعارض مع هذه القوانين. وتعتبر التشريعات المصرية الوثائق مصدرًا مطلقًا للبيانات بكل أنواعها، ويمكن استخدام أي وسائل لمعالجتها، ومن ثم لا يخالف معالجة المكتبة لوثائقها بواسطة تقنيات البيانات الضخمة التشريعات المصرية الوثائقية. ومع ذلك لم تتناول التشريعات الوثائقية المصرية مجال أرشفة وثائق الجيل الثاني الرسمية، مما ترتب عليه عدم التزام المؤسسات المصرية – ومنها مكتبة الإسكندرية –بأرشفة وثائقها للجيل الثاني، بل يُترك الأمر لرغبة كل مؤسسة؛ وبالتالي تفقد الدولة، ومؤسساتها مصدرًا مهمًا ومتجددًا للبيانات؛ قد يساعدها في رسم استراتيجيات العمل، وخططه المستقبلية.

أما إذا تناولنا **عنصر "أخصائيي الوثائق Documentation Specialists"(التحدي الثاني)** فلا ننكر أن وعي أخصائيي الوثائق بمفهوم البيانات الضخمة، ومصادرها، وبرمجيات معالجتها؛ يلعب دورًا مهمًا في

نجاح تحليل الوثائق بواسطة تطبيقات البيانات الضخمة في أي مؤسسة[43]. وقد أظهرت الدراسة أن أخصائيي الوثائق في مكتبة الإسكندرية لديهم دراية حول مفهوم البيانات الضخمة، وبعض مصادرها –لا سيما مواقع الويب، وشبكات التواصل الاجتماعي–ولكن ليس لديهم خلفية كافية عن كيفية التعامل مع الوثائق الإدارية الجارية باعتبارها مصدرًا للبيانات الضخمة[44]؛ مما يعيق وضعهم استراتيجيات، وخطط تلبي المتطلبات المستقبلية المختصة بتحليل الوثائق بواسطة برمجيات البيانات الضخمة.

أما بالنسبة لعنصر **حوكمة الوثائق Records Governance (التحدي الثالث)** فيجب أن تحرص كل مؤسسة على وضع إطار العمل، والقواعد المختصة بإدارة وثائقها، وضبطها؛ لكي تضمن تنظيم وثائقها بدقة؛ ومن ثم يمكن تحليلها بسهولة وبكفاءة. وعلى الرغم من أن قطاعات مكتبة الإسكندرية قد حددت قواعد إدارة وثائقها الورقية؛ لكنها لم تضع سياسة موحدة مكتوبة يلتزم بها كل قطاعاتها في ضبط وثائقها Records management، كما أن التواصل ضعيف بين وحدات الحفظ المركزي، ووحدات الحفظ اللامركزي في المكتبة؛ مما لا يحقق اتساق في ضبط وثائقها الورقية على مستوي المكتبة وقطاعاتها كلها[45]، في حين أن وثائقها الرقمية أكثر حظًا؛ حيث وضعت المكتبة سياسة موحدة لضبطها، وأرشفتها على مستوي المكتبة كلها؛ بينما لا تطبق المكتبة أي نوعًا من حوكمة على وثائقها للجيل الثاني. فقد تبين من الدراسة أن المكتبة – للأسف – لا تطبق أي سياسة، ولا تتخذ أي إجراء يتعلق بإدارة الوثائق الجيل الثاني Records management 2,0[46]؛ مما يعيق ضبطها للمحتوي صفحاتها على شبكات التواصل الاجتماعي؛ مما قد يفقدها محتوي متفرد لا يتوافر في وثائقها الورقية، أو الرقمية، والذي يتنوع بين آراء مستفيدين، وإعلانات عن أنشطة، ومقاطع فيديو، وصور، أو مقاطع صوتية توثق عقد فاعليات أو أنشطة. لقد أظهرت الدراسة أن المكتبة لم تهتم بوصف وثائقها للجيل الثاني، وتصنيفها، وتقييمها، ومن ثم ستواجه المكتبة صعوبة في الفصل بين وثائق الجيل الثاني المهمة، والمستبعدة، الأمر الذي سيجعل عملية معالجة هذه الوثائق بواسطة تطبيقات البيانات الضخمة تستغرق وقتًا أطول، وتكلفة أكثر، بل ستتأثر عملية تحليل هذه الوثائق سلبًا؛ مما ينعكس على دقة نتائج التحليل، وعملية صنع القرار.

أما إذا استعرضنا مجال **وصف الوثائق describing records (التحدي الرابع)**؛ فيجب أن نؤكد على أن عملية وصف الوثائق، وسياقها context تعد من أهم العمليات الفنية المتعلقة بضبط الوثائق في أي مؤسسة؛ لأنها تعكس طبيعة الوثائق ومحتواها؛ والظروف التي ظهرت فيها. وقد تبين في أثناء الدراسة أن عملية الوصف تختلف من فئة لأخري من فئات وثائق المكتبة؛ حيث نجد أن وصف وثائق المكتبة الرقمية قد غطى محتوي الوثائق، وسياقها بدقة وبالتفصيل، في حين أن المكتبة ركزت في وصف وثائقها الورقية على الحد الأدنى من البيانات الرئيسة –كما أشرنا من قبل–ومن دون التطرق لسياقها على الإطلاق[47]، بينما لم تقم المكتبة بوصف محتوي وثائق الجيل الثاني وسياقها نهائيًا. كما لاحظت الباحثة أن مكتبة الإسكندرية لا تطبق واصفات بيانات metadataموحدة في وصف وثائقها الورقية، والرقمية بل تختلف فيما بينها. كما نجد أن واصفات بيانات الوثائق الرقمية شاملة، وموحدة مقارنة بواصفات بيانات الوثائق الورقية التي قد تختلف أيضًا من قطاع إداري لأخر في المكتبة؛ مما يجعل عملية تحليل الوثائق التي تغطي الموضوع، أو السياق نفسه عملية مرهقة؛ إذا كانت هذه الوثائق تتنوع بين الورقية، والرقمية.

كما يعد عنصر " **التجميعات الموضوعيةObjective aggregations**" **(التحدي الخامس)** من أهم التحديات التي يجب أن تضعها المؤسسات في حسبانها عند معالجة وثائقها بواسطة تقنيات البيانات الضخمة؛ فلا شك في أن تصنيف الوثائق إلى تجميعات وفق محتواها يسرع عملية معالجتها بواسطة تطبيقات الضخمة، ويزيد من دقة نتائج تحليلها، لا سيما إذا استخدمت الكلمات المفتاحية Keywords، والتيجان Tags للتعبير عن محتوى الوثائق، وموضوعاتها. هذا وقد أظهرت الدراسة أن المكتبة حريصة على تصنيف وثائقها، وتقسيمها إلى تجميعات وفق موضوعاتها، والبيانات التي تتضمنها –مثال: وثائق الميزانيات، واللوحات الهندسية المختصة بمبني المكتبة، واتفاقيات التعاون الدولي مع المؤسسات الثقافية–[48]ولكن للأسف لا تميل المكتبة إلى استخدام الكلمات المفتاحية والتيجان للتعبير عن محتوي وثائقها– ولا سيما في مجال الوثائق الورقية– ومن ثم قد يقيد هذا الأمر قدرة تطبيقات البيانات الضخمة في تحليل وثائق المكتبة عند الحاجة. وقد حاولت المكتبة التغلب

على موضوع محدودية التوسيم Tagging في مجال الوثائق الرقمية بوصف محتوى هذه الوثائق، وسياقها وصفًا تفصيليًا دقيقًا مستخدمة الألفاظ المتداولة في بيئة عملها الإداري؛ لتحقق الغرض نفسه من عملية التوسيم.

أما بالنسبة إلى عنصر "**التقييم، ومدد الحفظ Appraisal and Retention Periods**"**(التحدي السادس) فهو** يتكامل مع سمتي المصداقية، والقيمة في البيانات الضخمة؛ فعندما تحرص المؤسسات على تقييم وثائقها بصفة دورية؛ لاستبعاد ما ليس له قيمة للعمل؛ يوفر لها بيانات صحيحة وحديثة؛ مما يحقق المصداقية في بياناتها؛ ومن ثم يساعدها على اتخاذ قرارات إدارية سليمة. هذا وقد أظهرت الدراسة أن المكتبة –للأسف– لا تمارس عملية التقييم والاستبعاد على وثائقها بصفة دورية، ولا تلتزم– أحيانًا– بمدد الحفظ؛ فقد لاحظت الباحثة أن هناك وثائق قد انتهت مدة حفظها؛ ولا زالت المكتبة تحتفظ بها؛ مثل: رسائل البريد الإلكتروني التي تعود إلى منتصف عام 2019م –كما أنها لا تحرص على عقد اللجنة الدائمة للمحفوظات؛ والتي تمثل همزة الوصل بين المكتبة ودار الوثائق القومية في مجال التقييم والاستبعاد؛ مما يؤثر سلبًا على ضبط وثائقها. وبلا شك أن تراخي المكتبة في ممارسة التقييم والاستبعاد قد يضرر بمعالجة وثائقها بواسطة تطبيقات البيانات الضخمة؛ بسبب اختلاط البيانات القديمة أو المرفوضة بنسختها الحديثة أو المعتمدة؛ مما يؤثر على دقة نتائج تحليل محتوى هذه الوثائق.[49]

ويجب على المؤسسات أن تحرص – أيضًا–على تحقيق عنصر **عدم التكرار Non-repetition (التحدي السابع)** والذي يقصد به الاحتفاظ بنسخة واحدة فقط من الوثيقة، والتخلص من نسخها المتكررة أولًا بأول. ويعتقد البعض أن عملية تنقية البيانات data cleansing قد تعالج مشكلة تكرار نسخ الوثائق المتعددة، ولكنها ستستغرق وقت أطول؛ ولا سيما إذا كانت النسخ المكررة كثيرة. وبلا شك ان التحول الرقمي في مكتبة الإسكندرية قد ساعد على تخلص وثائقها الرقمية من مشكلة تكرار النسخ؛ ولكن مازالت وثائق المكتبة الورقية تعاني من هذه المشكلة؛ ولا سيما مع تطبيق الحفظ اللامركزي في قطاعات المكتبة –مثل عندما يعقد القطاع الثقافي اتفاق مع مؤسسة ثقافية؛ يتم إعداد ثلاث نسخ من الاتفاق تحفظ النسخة الأولي في القطاع نفسه، والثانية في وحدة الوثائق الإدارية، والمالية، والثالثة في إدارة الشئون القانونية؛ مما يؤثر سلبًا على نتائج تحليل

هذه الوثائق؛ إذا لم يتم فرز هذه النسخ واستبعادها قبل معالجتها[50]. كما تبين من الدراسة أن وثائق المكتبة للجيل الثاني تعاني –أحيانًا–من التكرار ؛ فقد لاحظت الباحثة أن المنشور الواحد post قد يتكرر نشره أكثر من مرة –ولا سيما على الفيس بوك–مما يضر عملية معالجتها بواسطة تطبيقات البيانات الضخمة.

أما إذا تناولنا عنصر "**الصيغ الرقمية التفاعليةInteractive Digital Formats**" **(التحدي الثامن)** فينبغي لنا الإشارة إلى أن هذا التحدي يواجه الوثائق الورقية أكثر من الوثائق الرقمية أو الجيل الثاني؛ فمن أكبر التحديات التي قد تواجه أي مؤسسة عندما تؤرشف وثائقها الورقية إلكترونيًا ؛ هو تحديد الصيغة التي تتاح عليها الإصدارة الإلكترونية لهذه الوثائق؛ ولا ننكر أن هناك علاقة طردية بين دقة تحليل الوثائق الورقية آليًا، ومدى تفاعل البرامج مع الإصدارة الإلكترونية المختصة بها ؛ فكلما زاد تفاعل البرمجيات المختلفة مع النسخة الإلكترونية من الوثيقة الورقية، وقدرتها على معالجة محتواها، ونصها بحرية، كلما زادت كفاءة نتائج تحليلها. وقد تبين من الدراسة أن وثائق المكتبة الورقية عندما يتم أرشفتها إلكترونيًا –غالبًا–يتم حفظها في صيغة غير تفاعلية؛ لا تتدعم معالجة نصها آليًا [51]– مثل صيغة JPEG – مما يؤثر على معالجة هذه الوثائق بواسطة تطبيقات البيانات الضخمة في المستقبل؛ وبلا شك أن عمليتي تصدير Export والاستيراد Import بين الصيغ المختلفة يمكن أن تعتمد عليه المكتبة في التغلب على هذا التحدي، ومع ذلك يفضل أن تستخدم المكتبة منذ البداية صيغ تفاعلية– مثل PDF– عند أرشفة وثائقها الورقية رقميًا ؛ لتقليص التكلفة، وإهدار الوقت الناتج عن عمليتي تصدير Export والاستيراد Import بين الصيغ المختلفة؛ ولا سيما إذا كانت تطبق على كم كبير من الوثائق في الوقت نفسه.

ومن التحديات الصعبة التي يجب أن توليها المؤسسات العناية عندما تتعامل مع وثائقها بمثابة مصدرًا للبيانات الضخمة؛ هو تحدي البنية التحتية للحفظ طويل الأجل "Long–term Preservation **(التحدي التاسع)**؛ فعدم توافر بنية تحتية تكنولوجيات تدعم الحفظ طويل الأجل للوثائق والبيانات؛ وتحافظ على أصالتها ومصداقتها؛ قد يتسبب في عجز المؤسسة في الاستفادة من بيانات وثائقها وتحليلها بالجودة المطلوبة. ولذلك حرصت مكتبة الإسكندرية على امتلاك بنية تكنولوجيا قوية؛ تعزز الحفظ طويل الأجل لمحتواها الرقمي؛ بل

تحرص على تطويرها وإثراء قدرتها التخزينية؛ لكي تستوعب كميات هائلة من البيانات، والمصادر المختلفة، ومن ثم فالمكتبة مؤهلة تكنولوجيًا للتعامل المستقبلي مع وثائقها باعتبارها مصدرًا للبيانات الضخمة.[52]

وبعد استعراض التحديات العامة المشتركة بين بيئات وثائق المكتبة الثلاث؛ سنعرض التحديات الخاصة ببيئة وثائق المكتبة للجيل الثاني فقط؛ والتي تتمثل في:

أولًا. **ملكية المعلومات والخصوصية Ownership and Privacy:** من أكبر التحديات التي تواجه استخدام تطبيقات الويب 2,0 هو تحديد **ملكية المعلومات، والخصوصيةOwnership and Privacy (التحدي العاشر)** فكثير من الأفراد قد يعترض على استخدام المؤسسة لمعلوماتهم في أداء أنشطتها على اعتبارها – من وجهة نظرهم – ملكًا لهم، وعلى جانب أخر ترى المؤسسة أن من حقها استخدام هذه المعلومات؛ لأنها دونت في سياق التفاعل الرسمي معها. كما يجب أن تراعي المؤسسة خصوصية المستفيد، وأن لا تستخدم أي محتوي منشور من قبله عبر تطبيقات 2,0 بطريقة قد تتسبب له بأضرار شخصية أو اجتماعية، ومن ثمَّ ينبغي للمؤسسة أن تحرص على إعداد سياسة رسمية مختصة بالخصوصية تنشر على الخط المباشر، وتحدد فيها طبيعة محتوى تطبيقات 2,0 التي يمكن للمؤسسة الاستفادة منه في العمل، والمحتوي المستبعد مع الإشارة إلى التشريعات التي تنظم مجال الخصوصية، وحقوق الملكية المعلوماتية في الدولة، مع مراعاة أن لا يسمح للفرد بالتفاعل مع المؤسسة عبر صفحاتها المتاحة على تطبيقات الويب2,0 إلا بعد موافقة الفرد على بنود هذه السياسة[53]. وقد أظهرت الدراسة أن مكتبة الإسكندرية –للأسف–لم تنشر أي سياسة خصوصية مكتوبة عبر صفحاتها عبر تطبيقات التواصل الاجتماعي، أو موقعها الرسمي؛ مما يعيق معالجة المكتبة لمحتوي هذه التطبيقات بواسطة تطبيقات تحليل البيانات الضخمة، ويقيد استفادتها من مقترحات المستفيدين وآراءهم المنشورة عبر هذه التطبيقات؛ بل قد تتعرض إلى مساءلة قانونية في حالة استخدامها آراء المستفيد أو مقترحاته بغير إذنه.

ثانيًا. تحدي "**الذاتية في مقابل المهنية**self vs. Professional " **(التحدي الحادي عشر)** والذي يعد من الموضوعات التي يجب أن تعالجها المؤسسة قبل استخدامها لتطبيقات الويب2,0 في عملها الرسمي–خاصة– وأن هذه التطبيقات تدعم التواصل الرسمي، وغير الرسميّ بين الأفراد وفق السياق التي تستخدم فيه، ومن ثمَّ فهي قد تُستخدم في أغراض شخصية ومهنية في الوقت نفسه، مما يُحدث نوعًا من التداخل واللبس في طبيعة محتوى هذه التطبيقات[54]؛ ولذلك حرصت مكتبة الإسكندرية على توعية موظفيها بالالتزام بالمهنية، والموضوعية، والشفافية عند تواصلهم مع المستفيدين خلال صفحات المكتبة الرسمية المتاحة عبر تطبيقات الويب2.0، وعدم مناقشة أي موضوعات شخصية، أو عرض توجهاتهم الذاتية. وقد تبين من تصفح صفحات المكتبة عبر Twitter, LinkedIn, and Facebookالالتزام الشديد من قبل موظفي المكتبة بالمهنية، والمحايدة في الإجابة على استفسارات المستفيدين، أو نشر المنشورات، أو إدارة المناقشات العلمية عبرها.

وبعد الانتهاء من تحليل واقع إدارة الوثائق في مكتبة الإسكندرية، والتحديات التي قد تواجهها المكتبة عند استخدامها برمجيات البيانات الضخمة في بيئتها الوثائقية؛ عندئذ ستقوم الباحثة بقياس مدى صلاحية بيئات المكتبة الوثائقية لاستخدام تطبيقات البيانات الضخمة فيها..

8.تقييم صلاحية بيئات المكتبة الوثائقية؛ لاستخدام تطبيقات البيانات الضخمة

وفقًا للتحليل السابق؛ حاولت الباحثة تقييم مدى صلاحية بيئات المكتبة الوثائقية؛ لاستخدام تطبيقات البيانات الضخمة في المستقبل، وقد اعتمدت في هذا التقييم على قياس معدل توافر عناصر التقييم (التحديات) سابقة الذكر في بيئات إدارة وثائق المكتبة المختلفة: من الرقمية، والورقية، والجيل الثاني، واستخدمت الدراسة المقياس الآتي في قياس درجة توافر هذه العناصر:

يتوافر تمامًا= 2(الحد الأقصى)

يتوافر إلى حد ما =1

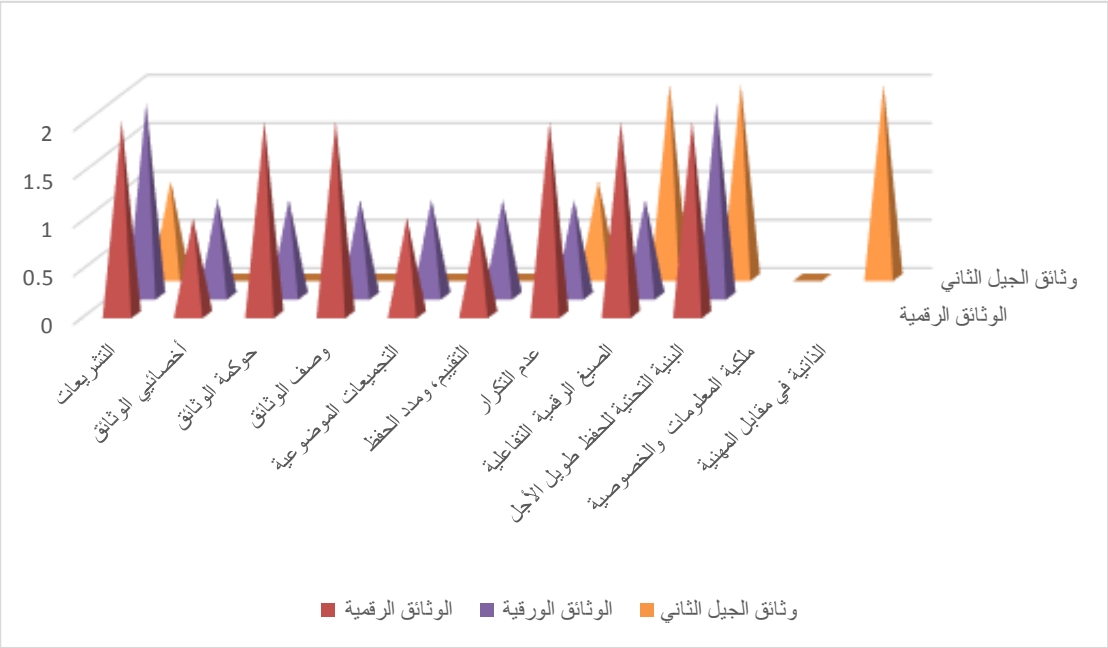
لا يتوافر =صفر (الحد الأدنى)

وقد طبقت الباحثة عناصر التقييم (التحديات)التسعة الأولى على فئات وثائق المكتبة الثلاث كلها، في حين اقتصر تطبيق التحديين (العاشر، والحادي عشر) على وثائق الجيل الثاني –فقط–؛ لأنها تتناسب مع طبيعتها –من دون الوثائق الورقية، والرقمية. وفيما يأتي جدول يوضح مدى توافر عناصر التقييم (التحديات) في بيئات إدارة وثائق المكتبة المختلفة.

جدول 2 يوضح مدى توافر عناصر التقييم (التحديات) في بيئات إدارة وثائق المكتبة المختلفة.

عناصر التقييم(التحديات)	الوثائق الرقمية	الوثائق الورقية	وثائق الجيل الثاني
التشريعات	2	2	0
أخصائيي الوثائق	1	1	1
حوكمة الوثائق	2	1	0
وصف الوثائق	2	1	0
التجميعات الموضوعية	1	1	0
التقييم، ومدد الحفظ	1	1	0
عدم التكرار	2	1	1
الصيغ الرقمية التفاعلية	2	1	2

البنية التحتية للحفظ طويل الأجل	2	2	2
ملكية المعلومات والخصوصية	_____	_____	0
الذاتية في مقابل المهنية	_____	_____	2
مجموع النقاط	15	11	8



شكل 1 يوضح مدى توافر عناصر التقييم (التحديات) في بيئات إدارة وثائق المكتبة المختلفة.

وقد اعتمدت الدراسة على معادلة النسبة المئوية في قياس صلاحية بيئة المكتبة الوثائقية؛ لاستخدام تطبيقات البيانات الضخمة فيها. وتعد هذه المعادلة من أشهر المعادلات؛ التي يُعتمد عليها في قياس كفاءة الأداء في كثير من المجالات الإدارية، والتنظيمية؛ لسهولتها، ووضوحها في التعبير عن الواقع الفعلي؛ فقد اعتمد عليها الأرشيف الوطني الأمريكي في إعداد تقارير التقييم الذاتي self-assessment reports المختصة ببرامج تقييم إدارة الوثائق في الحكومة الفيدرالية An Assessment of Records Management Programs in the Federal Government[55]. وتم تطبيق معادلة النسبة المئوية في قياس صلاحية بيئات المكتبة الوثائقية لاستخدام تطبيقات البيانات الضخمة فيها عن طريق: أولًا. يُحسب عدد النقاط الفعلية، التي تحصل عليها بيئة/ بيئات الوثائق عند تقييمها. ثانيًا. يضرب هذا العدد في رقم 100. وأخيرًا. يُقسَّم الناتج على مجموع نقاط التقييم الإجمالية المختصة ببيئة /بيئات الوثائق المراد تقييمها؛ والتي تمثل الحد الأقصى الذي يجب أن تبلغه هذه البيئات في عملية التقييم؛ وينتج مجموع نقاط التقييم الإجمالية عن حاصل ضرب عدد عناصر التقييم المختصة بكل بيئة من بيئات الوثائق في الحد الأقصى من النقاط التي ينبغي لكل عنصر من عناصر التقييم هذه أن يحققها، وهو نقطتان (2)، وفيما يلي مجموع نقاط التقييم الإجمالية المختصة بكل بيئة من بيئات الوثائق المختصة بمكتبة الإسكندرية، هي:

- ***مجموع نقاط التقييم الإجمالية المختصة ببيئة الوثائق الرقمية= 9عناصر×2= 18نقطة***
- ***مجموع نقاط التقييم الإجمالية المختصة ببيئة الوثائق الورقية= 9عناصر×2= 18نقطة***
- ***مجموع نقاط التقييم الإجمالية المختصة ببيئة وثائق الجيل الثاني= 11عناصر×2= 22نقطة***

وقد استعانت الباحثة بالبيانات التقييمية في جدول رقم (2) في تطبيق معادلة النسبة المئوية في قياس صلاحية بيئات المكتبة الوثائقية الثلاث؛ لاستخدام تطبيقات البيانات الضخمة:

-صلاحية بيئة المكتبة الوثائقية الرقمية =

(عدد النقاط الفعلية التي سجلتها بيئة الوثائق الرقمية ×100) ÷ مجموع نقاط التقييم الإجمالية المختصة ببيئة الوثائق الرقمية= (15×100) ÷18=83%

-صلاحية بيئة المكتبة الوثائقية الورقية=

(عدد النقاط الفعلية التي سجلتها بيئة الوثائق الورقية ×100) ÷ مجموع نقاط التقييم الإجمالية المختصة ببيئة الوثائق الورقية= (11×100) ÷18=61%

-صلاحية بيئة المكتبة المختصة بوثائق الجيل الثاني =

(عدد النقاط الفعلية التي سجلتها بيئة وثائق الجيل الثاني المختصة ×100) ÷ مجموع نقاط التقييم الإجمالية المختصة ببيئة وثائق الجيل الثاني= (8×100) ÷22=36%

ويتبين مما سبق أن بيئة الوثائق الرقمية قد حققت أعلى معدل؛ لإمكانية استخدام تطبيقات تحليل البيانات الضخمة معها؛ فقد بلغ معدل صلاحيتها نسبة (83%) في حين بلغت صلاحية بيئة الوثائق الورقية لاستخدام تطبيقات البيانات الضخمة (61%)، بينما حققت بيئة وثائق الجيل الثاني أقل معدل من الصلاحية؛ حيث بلغت (36%).

كما حاولت الباحثة قياس صلاحية بيئة المكتبة الوثائقية العامة؛ لاستخدام تطبيقات البيانات الضخمة فيها؛ للوقوف على الواقع بشكل كامل بواسطة حساب متوسط صلاحية بيئات وثائق المكتبة المختلفة عن طريق جمع معدلات صلاحية بيئات المكتبة الوثائقية المعنية معًا، ثم تقسيمها على عدد هذه البيئات. هذا وقد أظهرت الدراسة أن بيئة المكتبة الوثائقية صالحة لاستخدام تطبيقات تحليل البيانات الضخمة بلغت 72% في حالة اقتصار التقييم على بيئتي الوثائق الرقمية، والوثائق الورقية معًا، ومن دون بيئة وثائق الجيل الثاني. وهذا ما أثبتته المعادلة الآتية:

-متوسط صلاحية بيئة المكتبة الوثائقية[الجزئية]=

(صلاحية بيئة المكتبة الوثائقية الرقمية + صلاحية بيئة المكتبة الوثائقية الورقية) ÷ عدد هذه البيئات الوثائقية = (61%+83%) ÷2=72%

شكل 2يعرض متوسط صلاحية بيئة المكتبة الوثائقية (الجزئية)

في حين انخفض متوسط صلاحية بيئة المكتبة الوثائقية لاستخدام تطبيقات تحليل البيانات الضخمة إلى (59%) عندما شملت التقييم بيئات وثائق المكتبة الثلاث كلها. وهذا ما أثبتته المعادلة الآتية:

-متوسط صلاحية بيئة المكتبة الوثائقية[الكاملة]=

(صلاحية بيئة المكتبة الوثائقية الرقمية + صلاحية بيئة المكتبة الوثائقية الورقية + صلاحية بيئة المكتبة المختصة بوثائق الجيل الثاني) ÷ عدد هذه البيئات الوثائقية = (83%+61%+36%) ÷3=59%

شكل 3يعرض متوسط صلاحية بيئة المكتبة الوثائقية (الكاملة)

ومن ثم يتبين أن ضعف بيئة إدارة وثائق 2.0 في مكتبة الإسكندرية قد أثر بشكل مباشر في تراجع معدل صلاحية بيئة المكتبة الوثائقية العامة للتعامل مع تطبيقات البيانات الضخمة، وتراجعت صلاحيتها من جيد إلى مقبول.

9-النتائج، والاستنتاجات

ونستنتج مما سبق أن: أولًا. دخلت مكتبة الإسكندرية عصر البيانات الضخمة عن طريق المشروعات الثقافية المرتبطة ببيئة الويب. ثانيًا. أظهرت الدراسة أن تقارير البيانات الضخمة المتعلقة بهذه المشروعات تحفظ إلى ما لانهاية في نظام إدارة محتوي منفصل، ومستقل عن نظم إدارة المحتوي المختصة بوثائق المكتبة الإدارية الجارية الأخرى. ثالثًا. على الرغم من وثائق عمل المكتبة الجارية لا تُصنف باعتبارها مصدرًا للبيانات الضخمة -الآن-، ولكنها من المحتمل بشدة أن تصبح مصدرًا لهذه البيانات في المستقبل في ظل سياسات التحول

الرقمي الذي تتبناها المكتبة في أداء مهام عملها، وأنشطتها، وتقديم خدماتها. رابعًا. لا تمارس المكتبة عملية التقييم، والاستبعاد على وثائقها بطريقة مثالية، مما يؤثر سلبًا على نتائج تحليل هذه الوثائق عند الحاجة. خامسًا. أظهرت الدراسة أن المكتبة لا تطبق أي عملية من عمليات إدارة الوثائق الجيل الثاني records management 2,0، كما أنها تفتقد وجود سياسة خصوصية لها مكتوبة، ومعلنة تتعلق باستخدام تطبيقات الجيل الثاني من الويب؛ مما يؤثر سلبًا على قدرتها – أي المكتبة – على الاستفادة القصوى الآمنة من المحتوي، والآراء المتاحة عبر صفحاتها على هذه التطبيقات. سادسًا. اتضح أن المكتبة عندما ترقمن –أحياناً – وثائقها الورقية؛ فقد تحفظها في صيغ رقمية غير تفاعلية، لا يمكن معالجتها بتطبيقات تحليل البيانات. سابعًا. يقتصر وصف وثائق المكتبة الورقية على البيانات الوصفية الأساسية المختصة بالوثائق ومحتواها، ومن دون التطرق إلى سياقها context؛ مما يعيق معالجتها بواسطة تقنيات تحليل البيانات الضخمة في المستقبل. ثامنًا. اهتمت المكتبة بضبط وثائقها الرقمية بعناية –مقارنة بوثائقها الورقية، وثائقها للجيل الثاني – مما يجعلها أكثر استعدادًا لاستخدام تطبيقات تحليل البيانات الضخمة معها في المستقبل. تاسعًا. تمتلك المكتبة بنية تكنولوجية قوية تؤهلها لاستخدام تطبيقات تحليل البيانات الضخمة في معالجة وثائقها مستقبليًا بطريقة فعالة، عند الحاجة.

10-التوصيات

انتهت الدراسة إلى عدد من التوصيات تمثلت في: أنه ينبغي للمكتبة أن تصيغ السياسات، والخطط المنوطة بمجال إدارة وثائقها بكل فئاتها؛ لتحقق الاتساق في ضبطها، وتوحيد إجراءات أرشفتها في كل قطاعات المكتبة؛ لتيسير معالجتها، وتحليلها. كما يجب أن تحرص على تحسين الجوانب الفنية، والتنظيمية المختصة بوثائق الجيل الثاني؛ حتى لا تفقد المكتبة مصدرًا حيويًا للبيانات الضخمة في المستقبل. كما يجب أن تعد المكتبة لجنة تتشكل من مديري الوثائق Records Managers ومهندسي تكنولوجيا المعلومات؛ لتتولي هذه اللجنة وضع الاستراتيجيات والخطط المتعلقة بمعالجة الوثائق باعتبارها مصدرًا للبيانات الضخمة؛ كما تكون مسئولة – أيضًا–عن حل المشكلات المتعلقة بأرشفة البيانات الضخمة. بالإضافة إلى ذلك يجب أن تستخدم المكتبة

صيغ تفاعلية في أرشفة وثائقها الورقية إلكترونيًا؛ لتيسير معالجتها بواسطة تطبيقات البيانات الضخمة عند الحاجة. كما يجب أن تقنن واصفات البيانات metadata المختصة بالوصف، والسياق الأرشيفي، وتوحيدها على فئات وثائقها الثلاث؛ لتسهيل معالجتها آليًا. ومن ناحية أخري يجب أن تثري المكتبة وعي أخصائييها للوثائق حول متطلبات إدارة الوثائق باعتبارها مصدرًا للبيانات الضخمة، وكيفية أرشفة البيانات الضخمة نفسها؛ عن طريق عقد الدورات التدريبية، وورش العمل في مجال البيانات الضخمة وأرشفتها. كما يجب أن تعد المكتبة سياسة خصوصية مكتوبة، وأن تعلنها عبر موقعها الرسمي – على الأقل–وصفحاتها على شبكات التواصل الاجتماعي؛ لكي يمكنها الاستفادة القصوى من محتوي هذه الشبكات من دون التعرض لأي مساءلة قانونية. كما توصي الدراسة بتشجيع التعاون بين مكتبة الإسكندرية، والمؤسسات التي لديها تجارب في معالجة الوثائق الإدارية الجارية بواسطة تطبيقات البيانات الضخمة؛ للاستفادة من خبراتهم في هذا المجال.

11- الاستشهادات

[1] Laney, Douglas " 3D Data Management: Controlling Data Volume, Velocity and Variety". META Group Research Note, 6(2001). available at: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. (Accessed July 26,2019)

[2] Buxton, Bill et al. Big data: the next Google. Interview by Duncan Graham-Rowe nature. 455.10.1038/455008a.researchgate.2008. available at: https://www.researchgate.net/publication/273353769_Big_data_the_next_Google_Interview_by_Duncan_Graham-Rowe (accessed July 19,2019)

[3] برنامج أباتشي هادوب (**Apache** Hadoop): هو برنامج مفتوح المصدر، ويعد من أشهر البرامج المختصة بمعالجة البيانات الضخمة وتحليلها، وهو مكتوب بلغة الجافا. للمزيد من المعلومات أنظر الرابط الآتي: https://hadoop.apache.org/

[4] Geanina, Elena et al. "Perspectives on Big Data and Big Data Analytics." Database Systems Journal, April 2012: 314.

[5] TOLE, Alexandru Adrian. "Big Data Challenges." Database Systems Journal, 2013: 31-40.

[6] Jewani, Kajal J et al. "Data Archival using Hadoop Distributed File system." (IJCSIT) International Journal of Computer Science and Information Technologies 6 (2015): 4974 -4976.

[7] Haralampus, Lisa. "Big Data Issues for Federal Records Managers." ARMA Metro Conference. New York City: ARMA International, 2017. 1-24.

[8] السلمي، خلود خالد؛ سليمان ناصر، هاجر. "واقع استخدام البيانات الضخمة في هيئة الوثائق والمحفوظات الوطنية: دراسة استطلاعية". المؤتمر 24 لجمعية المكتبات المتخصصة فرع الخليج العربي: البيانات الضخمة، وآفاق استثمارها: الطريق إلى التكامل المعرفي. مسقط: جمعية المكتبات المتخصصة، 2018م. 1-18.

[9] مقناني، صبرينة؛ شبيلة، مقدم. "دور البيانات الضخمة في دعم التنمية المستدامة بالدول العربية". Journal of Information Studies & Technology (JIS&T), no. 1(2019). available at: https://doi.org/10.5339/jist.2019.4(accessed March 5, 2020)

[10] TOLE, "Big Data Challenges", 33.op.cit.

[11] Saky, Kevin Taylori. "Big Data: Understanding Big Data." researchgate. January 2016. available at: https://www.researchgate.net/publication/291229189 (accessed March 5, 2020)

[12] Ibid.
[13] Khan, M. Ali-ud-din. "Seven V's of Big Data understanding Big Data to extract value." Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education. Bridgeport, CT: IEEE, 2014 available at: Https://ieeexplore.ieee.org/document/6820689/references#references (accessed December 3, 2020).(
[14] Tocatlian, J. "Bibliotheca Alexandrina: From Dream to Reality". The International Information & Library Review.35, no.1(2003): 39–50.
[15] BIBLIOTHECA ALEXANDRINA. About the BA, 2018. https://www.bibalex.org/en/Page/About (accessed June 30, 2020) .(
[16] Ibid.
[17] BIBLIOTHECA ALEXANDRINA. projects& activities: Internet Archive. 2018. available at: https://www.bibalex.org/en/project/details?documentid=283 (accessed June 30, 2020(
[18] International School of Information Science (ISIS) of BA. "BA Internet (Web) Archive." The International School of Information Science (ISIS). 2019. available at: https://www.bibalex.org/isis/frontend/archive/archive_web.aspx (accessed November 4, 2020).
[19] The International School of Information Science (ISIS) of BA. Encyclopedia of life. 2018. available at: http://eol.bibalex.org/ (accessed November 16,2020).
[20] Ibid.
[21] نظام هادوب للملفات الموزعة (Hadoop **Distributed File System (HDFS)**: هو نظام لتوزيع الملفات، والبيانات على الحاسبات الآلية / الخوادم؛ لمعالجتها. للمزيد من المعلومات أنظر الرابط الآتي: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html
[22] تتكون العقدة node من Hardware and software معًا، وتشير node -عادة – في عملية معالجة البيانات الضخمة إلى خادم server، أو الخوادم المسؤولة عن معالجة ملف من ملفات البيانات الضخمة. للمزيد من المعلومات يمكن زيارة الرابط الآتي: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html#NameNode+and+DataNodes
[23] MapReduce: هو نموذج البرمجة (programming model) ؛ المسئول عن تحديد العمليات التي يجب أن تخضع لها البيانات تلقائيًا في أثناء معالجتها؛ ويضمن تطبقها آليًا بشكل متوازن parallel على مجموعات كبيرة من البيانات في الوقت نفسه. للمزيد من المعلومات يمكن زيارة الرابط الآتي: https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html
[24] Aprache Pig : هي منصة platform تختص بتحليل البيانات ذات الكميات الكبيرة، وتتميز بقدرتها على التعامل مع بنيات البيانات المختلفة من: المهيكلة، وغير المهيكلة، وشبه المهيكلة. للمزيد من المعلومات يمكن زيارة الرابط الآتي: https://pig.apache.org/
[25] قواعد البيانات NoSQL: هي قواعد بيانات غير علائقية non-relational database – لا تعتمد على الجداول في حفظ البيانات-وتتميز بحفظ كميات كبيرة من البيانات وبتكلفة منخفضة. للمزيد من المعلومات يمكن زيارة الرابط الآتي:-https://www.geeksforgeeks.org/introduction-to-nosql/
[26] HBase: قاعد بيانات غير علائقية تختص ببرنامج هادوب؛ مسئولة عن حفظ البيانات، والعلاقات فيما بينها؛ للمزيد من المعلومات يمكن زيارة الرابط الآتي: https://hbase.apache.org/
[27] MongoDB: هي قاعدة وثائق غير علائقية تعكس العلاقات بين الوثائق– والتي يقصد بها أي وسيط يحوي بيانات؛ بصرف النظر عن نوعه؛ فقد تكون الوثائق صفحات ويب، أو مستندات رسمية...إلخ - التي تحوى بيانات ذات الصلة. للمزيد من المعلومات يمكن زيارة الرابط الآتي: https://www.mongodb.com/2
[28] Neo4j: هي قاعدة بيانات غير علائقية بيانية graph تختص بحفظ البيانات، والروابط فيما بينها في شكل مخططات بيانية رسومية. للمزيد من المعلومات يمكن زيارة الرابط الآتي: https://neo4j.com/
[29] الدكر، يوسف (مدير تكنولوجيا المعلومات والاتصالات بمكتبة الإسكندرية)، أجرى المقابلة نرمين اللبان. مشروعات البيانات الضخمة في مكتبة الإسكندرية (3 فبراير, 2020).
[30] نفس المصدر السابق
[31] اقتصرت الباحثة في هذه النقطة على معالجة الوثائق رقمية النشأة فقط.
[32] قيحة، داليا (أخصائي بمكتبة الإسكندرية)، أجرى المقابلة نرمين اللبان. إدارة الوثائق، ونظام المسار track system في مكتبة الإسكندرية (19ديسمبر, 2019).
[33] الجبلاوي، سارة (رئيس وحدة بمكتبة الإسكندرية)، أجرى المقابلة نرمين اللبان. حفظ الوثائق في مكتبة الإسكندرية(11فبراير، 2020).
[34] على، وائل (أخصائي بمكتبة الإسكندرية)، أجرى المقابلة نرمين اللبان. إدارة الوثائق، وأرشفتها، في مكتبة الإسكندرية (11أغسطس, 2020).
[35] نفس المصدر السابق.
[36] الدكر، يوسف. مشروعات البيانات الضخمة في مكتبة الإسكندرية. - المصدر السابق
[37] للمزيد من المعلومات يمكن زيارة صفحة مكتبة الإسكندرية المتاحة عبر موقع الفيس بوك: https://www.facebook.com/bibalexOfficial
[38] الدكر، يوسف. مشروعات البيانات الضخمة في مكتبة الإسكندرية. - المصدر السابق.
[39] اللبان، نرمين إبراهيم على." البيانات الضخمة، وإدارة الوثائق". الملتقى الثاني للأرشيفيين في العالم العربي: ما بعدا الرقمنة في الأرشيفيات". الإسكندرية: منظمة التنمية الإدارية العربية،1.2018-19.

[40] اعتمدت الباحثة على عدد من الدراسات العلمية المتخصصة في مجال البيانات الضخمة؛ لتحديد التحيات الإحدى عشر الواردة في بحثها " البيانات الضخمة، وإدارة الوثائق "؛ ومن هذه الدراسات:

-IEEE Metadata Standard for Supporting Big Data Management, available at: https://bigdata.ieee.org/standards?highlight=WyJiaWciLCJkYXRhIiwiZGF0YSdzIiwiZGF0YScuIiwic3RhbmRhcmRzIiwi YmlnIGRhdGEiLCJkYXRhIHN0YW5kYXJkcyJd

-Analyst Report: Rethinking the enterprise data archive for big data analytics and regulatory compliance, available at: https://gigaom.com/report/rethinking-the-enterprise-data-archive-for-big-data-analytics-and-regulatorycompliance

[41] طبقت الباحثة عنصري " الملكية والخصوصية "و " مهنية موظفي المؤسسة " على فئة وثائق المكتبة للجيل الثاني فقط؛ لأن المكتبة تستخدم تطبيقات الويب الثاني web2.0applications بشكل رسمي؛ للتواصل مع الجمهور؛ ومن ثم ينتج عنها محتوي رسمي ذات صلة بمجال العمل؛ لذلك ينبغي للمكتبة أن تضمن إمكانية استخدامه من دون الإضرار بمصلحة متابعيها عبر هذه التطبيقات، وأن يتمتع هذا المحتوى بالموضوعية، والشفافية، والمهنية اللازمة لاتخاذ قرارات سليمة وفقًا له.

[42] فودة، محمود (ممثل دار الوثائق القومية)، أجرى المقابلة نرمين اللبان. التقييم، ومدد الحفظ في مكتبة الإسكندرية (10ديسمبر، 2019).

[43] Kibe, Lucy. "Application of big data in records management in public organizations in Kenya." Researchergate, 2016. available at: https://www.researchgate.net/publication/298746312_Application_of_big_data_in_records_management_in_public_organizations_in_Kenya (accessed December 3, 2019)

[44] هذه النتيجة توصلت لها الباحثة عن طريق المقابلات الشخصية المقننة مع عدد من موظفي المكتبة المتعاملين مع الوثائق.

[45] الجبلاوي، سارة. حفظ الوثائق في مكتبة الإسكندرية. المصدر السابق.

[46] إدارة الوثائق الجيل الثاني Records management 2,0؛ والتي يقصد بها: كل العمليات، والإجراءات المختصة بضبط وثائق تطبيقات الويب 2,0 المهيكلة وغير المهيكلة خلال دورة حياتها. للمزيد من المعلومات يمكن الاطلاع الرابط الآتي:- https://www.archives.gov/records-mgmt/bulletins/2014/2014-02.html

[47] علي، وائل. إدارة الوثائق، وأرشفتها، في مكتبة الإسكندرية. المصدر السابق.

[48] نفس المصدر السابق

[49] فودة، محمود. التقييم، ومدد الحفظ في مكتبة الإسكندرية. المصدر السابق.

[50] الجبلاوي، سارة. حفظ الوثائق في مكتبة الإسكندرية. المصدر السابق.

[51] نفس المصدر السابق.

[52] الدكر، يوسف. مشروعات البيانات الضخمة في مكتبة الإسكندرية. - المصدر السابق

[53] National Archives and Records Administration (NARA). "Report on Federal Web2.0 Use and Record Valuer." National Archives. 2010. Available at: https://www.archives.gov/records-mgmt/resources/web2.0-use.pdf (accessed August 22, 2019).(

[54] Ibid

[55] National Archives and Records Administration. Records Management Self-Assessment Report. Washington: National Archives, 2011 .