مستودعات البيانات البحثية العامة: دراسة استكشافية تحليلية.

اعداد
د. ريهام عاصم غنيم
أستاذ علم المعلومات المساعد
كلية الآداب - جامعة المنوفية.
أستاذ علم المعلومات المشارك
كلية الآداب والعلوم الإنسانية - جامعة طيبة.
rghoneem@taibahu.edu.sa

الملخص:

يهدف نشر ومشاركة البيانات البحثية إلى تمكين إعادة استخدامها وفقًا لديناميكيات ممارسي العلم المفتوح، في ممارسات نشر البيانات البحثية هناك ثلاث جهات فاعلة رئيسية: الباحثون والناشرون ومستودعات البيانات، تستكشف هذه الدراسة مستودعات البيانات البحثية العامة ، أي المستودعات غير المحددة بموضوع أو بفئة معينة من البيانات أو التي ينحصر استخدامها داخل مؤسسة بعينها حيث أنها مفتوحة للمجالات العلمية كافة ولنشر أي نوع من مجموعات البيانات البحثية، كما حللت هذه الدراسة منهجيات الممارسات الحالية لنشر مجموعات البيانات في هذه المستودعات، وذلك فيما يتعلق بثمانية جوانب رئيسية لنشر البيانات، هي (تنسيق مجموعات البيانات، والتوثيق، والترخيص، وتكاليف النشر، والتحقق من الصحة، والتوافر، وقابلية الاكتشاف والوصول، والاقتباس والاستشهاد) مستخدمة في ذلك كل من المنهج الوصفي التحليلي، ومنهج تحليل المحتوي، ويتضح من هذا التحليل أن مستودعات البيانات البحثية العامة الحالية هي خدمات قابلة للتطبيق لنشر البيانات وتحتاج إلى تطوير وتعزيز ممارساتها لدعمها بشكل أفضل، كما يجب على المستودعات العامة معالجة قضية رئيسة تتعلق بكيفية تحسين المواصفات التي تقدمها المستودعات، مثل التنسيقات التي تدعمها، وإجراءات التحقق المقدمة وإدارة تراخيص البيانات، وذلك للتغلب على القيود المفروضة نتيجة عدم التجانس الشديد لمجموعات البيانات.

الكلمات المفتاحية:

مستودعات البيانات البحثية - research data repositories - نشر البيانات البحثية Publication of research data - مجموعات البيانات -datasets 4TU.ResearchData -Csiro data -Figshare -Dryad Digital Repository -Zenodo -Science Data Bank -access portal Open Science Framework.

أولا: الإطار المنهجي:

١- تمهيد

أدت التطورات الهائلة والمتلاحقة لتكنولوجيا الاتصالات إلى العديد من التغيرات في وسائل وأنماط الاتصال العلمي بين العلماء والباحثين وأضحى من السهولة بمكان جمع وحفظ وإدارة وتوزيع البيانات البحثية، وكذلك مشاركتها، وإعادة استخدامها، وقد سعت بيئات البحث الغنية بالبيانات أن تُفَعل ذلك تعزيزاً لمجالات الدراسة الجديدة، وسعيا لفهم أفضل للتعقيدات البحثية. وعليه قامت وكالات التمويل وصناع السياسات، ومؤسسات البحث، والدوريات العلمية إلى تشجيع الباحثين لنشر بياناتهم، وعمدت إلى توظيف خطط لإدارة البيانات وتحديد متطلبات الإيداع، وربط المقالات العلمية بالبيانات الواردة فيها.

تاريخ استلام البحث : ٢٩ ديسمبر ٢٠٢٢ ، تاريخ القبول : ٧ فبراير ٢٠٢٣

يعد نشر البيانات البحثية منهجاً جديداً تدعو له العديد من دور النشر والمؤسسات الأكاديمية وذلك لدعم مشاركة بيانات البحث وإعادة استخدامها من قبل الآخرين حيث كانت ومازالت البيانات هي المحرك الأول للعلم، ويقصد بمصطلح نشر البيانات البحثية إنشاء مجموعات بيانات يتم التعامل معها بشكل جماعي كوحدة واحدة تمكن الباحثين من نشرها وإعادة استخدامها لأغراض بحثية. وعلى الرغم من ذلك هناك العديد من التحديات التي تواجه الاستفادة من بيانات البحث ،مثل التباين في ممارسات العلماء سواء على مستوى الأبحاث الفردية أو التأليف المشترك والخبرة اللازمة لإدارة البيانات؛ ونقص حوافز نشر البيانات؛ ووجود أنظمة مختلفة للملكية الفكرية؛ والسياسات التنافسية لإصدار البيانات والتحكم فيها؛ وغياب المعرفة بأدوات ومنافذ نشر البيانات البحثية مثل دوريات البيانات ، ومستودعات البيانات؛ كل هذه التحديات تمثل مؤشرات على أن التكنولوجيا وحدها لا يمكن أن تغير ممارسات العلماء؛ ويجب أن تتدخل العوامل الاجتماعية والثقافية الأخرى لتشجيع تبادل البيانات، كما يجب أن تستند سياسات مشاركة البيانات إلى التعريف بكيفية نشر الباحثين للبيانات وكيفية استخدامها ثم مشاركتها مع الآخرين. (١)

تلعب مستودعات البيانات البحثية دوراً رئيساً في دعم مشاركة العلوم وذلك من خلال تنفيذ ممارسات منهجية للإشراف على البيانات من أجل تعزيز جمع مجموعات البيانات العلمية الكافية ومعالجتها وحفظها وضمان توافرها على المدى الطويل، كذلك دعم آليات الوصول والنشر المفتوح مدفوع الثمن لمثل تلك الفئة من البيانات. وتنتشر مستودعات البيانات البحثية وتجد الدعم الكبير في العديد من التخصصات العلمية وخاصة تلك التي تنتج كمية هائلة من البيانات فعلي سبيل المثال نجد مستودع High-Energy Physics data (HEP data) في الفيزياء ،ومستودع GenBank في علم الوراثة ،ومستودع ioChem-BD Computational Chemistry Datasets لإدارة البيانات البحثية في مجال الكيمياء كذلك مستودعKNB: The Knowledge Network for Biocomplexity للبيانات البحثية في مجال علوم الأرض ولا يقف الأمر عند العلوم الطبيعية وعلوم الحياة فقط؛ بل يمتد كذلك إلى العلوم الاجتماعية فنجد على سبيل المثال لا الحصر مستودع Inter-university Consortium for Political and Social Research (ICPSR)في العلوم الاجتماعية والسياسية.

جرى الحديث عن مستودعات البيانات البحثية ضمن نهج جديد للعلوم أطلق عليه الذيل الطويل للعلم "long Tail of Science "LToS ويقصد به العدد الكبير من الباحثين الأفراد والمختبرات الصغيرة الذين ليس لديهم إمكانية الوصول إلى موارد حسابية مخصصة وخدمات عبر الإنترنت لإدارة وتحليل كمية كبيرة من البيانات، كذلك المجالات العلمية التي تعتمد على عدد كبير من المختبرات الصغيرة نسبياً في تنفيذ نشاطاتها العلمية وبواسطة باحثين أفراد ينتجون بشكل جماعي غالبية النتائج العلمية. الذيل الطويل غير مرئي تقريباً ويفتقر معظم أعضائه إلى المعرفة الفنية والخبرة في استخدام تقنيات البنية التحتية الإلكترونية ويعيقهم نقص الموارد المخصصة لتخزين البيانات العلمية وتحليلها ومشاركتها مع محدودية الموارد والخبرات، وعادةً ما يهتم المستخدمون ذو الذيل الطويل بالوصول إلى موارد الحوسبة والتخزين والبيانات الكبيرة واستخدامها على المدى القصير لتنفيذ إجراء محدد، إن مثل هذه السياقات العلمية والبحثية على الرغم من تنوعها وتطورها إلا أنها تفتقر إلى مستودعات مجتمعية مشتركة تدعم الاحتياجات الفعلية للعلوم في مشاركة البيانات البحثية، وعليه وفي السنوات العشر الأخيرة كان غالباً ما يتم اقتراح مستودعات البيانات كأدوات لدعم نشر البيانات البحثية حيث توفر تسهيلات لمختلف المساهمين في عملية النشر وعلى الرغم من الحاجة الملحة لنشر البيانات كممارسة بحثية في ظل العلم المفتوح وضرورة إعادة استخدامها ومشاركتها؛ ليتحقق الهدف الأولي من نشرها وعلى الرغم من أهمية مستودعات البيانات كوسيلة تساهم في دعم تلك الممارسات البحثية؛ إلا أنه لا يوجد فهم مشترك لما يجب أن تقدمه تلك المستودعات لدعم نشر ومشاركة وإعادة استخدام البيانات البحثية وذلك من قبل الأطراف المعنية جميعها من باحثين وناشرين ومراكز المعلومات ومستفيدين. (٢)

وعليه تسعى هذه الدراسة لاستكشاف مستودعات البيانات البحثية العامة غير محددة بموضوع أو بفئة معينة من البيانات أو التي ينحصر استخدامها داخل مؤسسة بعينها حيث إنها مفتوحة للمجالات العلمية كافة ولنشر أي نوع من مجموعات البيانات البحثية، وذلك من خلال استكشاف المستودعات المختارة من حيث البيانات العامة والتغطية الزمنية والموضوعية لعملية نشر مجموعات البيانات بها، كذلك تحليل المستودعات في ثمانية جوانب رئيسة تقف من خلالها الدراسة على منهجيات تعامل مستودعات البيانات البحثية العامة مع مجموعات البيانات وهي (تنسيق مجموعات البيانات – التوثيق - الترخيص- تكاليف النشر- التحقق من الصحة - التوافر- قابلية الاكتشاف والوصول- الاقتباس والاستشهاد المرجعي) ودور تلك الخصائص في دعم عمليات نشر وإتاحة البيانات البحثية.

٢-أهمية الدراسة ومبرراتها:

تأتي أهمية الدراسة من أهمية مستودعات البيانات البحثية، ودورها في مشاركة البيانات كممارسة اتصال علمية ناشئة تسهل تقدم العلم من خلال إتاحة البيانات وإمكانية التحقق منها وإعادة إنتاجها، وفي السنوات الأخيرة تزايدت أعداد مستودعات البيانات البحثية، وإن كان معظمها يميل إلى تغطية مجال محدد ولا يوجد إلا عدد قليل منها يتصف بالعمومية. وعلى الرغم من قلة أعداد مستودعات البيانات البحثية العامة، فإن الطلب على مثل هذه المستودعات يتزايد من كل من المؤسسات الأكاديمية، ووكالات التمويل والنشر وحتى الدوريات العلمية التي تسعى جميعها لإيجاد منهجية لإدارة ومشاركة البيانات البحثية، وفي العموم إن فكرة وجود مستودع بيانات بحثية عام يدعم نشر ومشاركة البيانات البحثية في مختلف المجالات العلمية هو مطلب حيوي خاصة في ظل المجتمعات العلمية التي لا تملك منصات لإتاحة ونشر ومشاركة بياناتها البحثية. وعليه تتمثل الأهمية العلمية لتلك الدراسة في استكشاف مستودعات البيانات البحثية العامة كوسيلة رئيسة لنشر ومشاركة البيانات البحثية.

٣- أهداف الدراسة:

تهدف الدراسة إلى تحقيق هدفين رئيسين:

- **الهدف الأول**: استكشاف مستودعات البيانات البحثية العامة، وتحليل خصائصها والسمات الموضوعية، والزمنية لمجموعات البيانات المتاحة بها.
- **الهدف الثاني**: تبيِّن منهجيات تعامل مستودعات البيانات البحثية العامة مع البيانات والمجموعات البحثية، وتأثير تلك المنهجيات في دعم وتعزيز نشر ومشاركة وإعادة استخدام البيانات وذلك من خلال تحديد ثمانية خصائص رئيسة تشكل دوراً هاماً في توصيف الدعم الذي تقدمه هذه المستودعات لنشر البيانات البحثية ومشاركتها وإعادة استخدامها.

٤- مشكلة الدراسة وتساؤلاتها

تتمثل المشكلة الرئيسة للدراسة في تزايد أعداد مستودعات البيانات البحثية، وتزايد الطلب على مستودعات البيانات البحثية العامة كوسيلة رئيسة ورافد هام من روافد نشر وإتاحة ومشاركة البيانات البحثية، هذا التزايد الذي يصاحبه نقص في الدراسات المتعلقة بهذه الفئة من المستودعات، وبالتالي كان من المناسب العمل على اكتشافها وتحليل المنهجيات التي تتبعها في نشر وإتاحة البيانات البحثية، وعليه تثار مجموعة من التساؤلات تحاول الدراسة الحالية الإجابة عنها من خلال رؤية استكشافية تحليلية.

- ما المقصود بمستودعات البيانات البحثية العامة؟
- ما الخصائص والسمات الأولية لمستودعات البيانات البحثية العامة؟

- ما التوزيع الزمني لنشر مجموعات البيانات في مستودعات البيانات البحثية العامة؟
- ما التوزيع الموضوعي لمجموعات البيانات البحثية المنشورة في مستودعات البيانات البحثية العامة؟
- ما منهجيات نشر وإتاحة مجموعات البيانات في مستودعات البيانات البحثية العامة؟

٥- مجال الدراسة وحدودها:

قامت الدراسة الحالية على استكشاف مستودعات البيانات البحثية العامة، وتحديد خصائصها وسماتها الأولية وتحليل منهجيات نشر وإتاحة مجموعات البيانات بهذه المستودعات، وعلى هذا فقد تم تحديد

- **الحدود الموضوعية** لتقف الدراسة على خصائص مستودعات البيانات البحثية من حيث النشأة والتبعية الإدارية ودولة المنشأ ، كذلك رصد التوزيع الزمني والموضوعي لمجموعات البيانات بها، مع تحليل لمنهجيات نشر وإتاحة مجموعات البيانات في ثمانية جوانب رئيسة تقف من خلالها الدراسة على أسس التعامل مع مجموعات البيانات البحثية وهي (تنسيق مجموعات البيانات – التوثيق - الترخيص- تكاليف النشر- التحقق من الصحة - التوافر- قابلية الاكتشاف والوصول- الاقتباس والاستشهاد المرجعي) وعلى هذا يمكن تقسيم مجال الدراسة وفقا للحدود التالية:
- **الحدود الفئوية:** مستودعات البيانات البحثية العامة أي غير المحددة بموضوع أو بفئة معينة من البيانات أو التي ينحصر استخدامها داخل مؤسسة بعينها حيث إنها مفتوحة للمجالات العلمية كافة ولنشر أي نوع من مجموعات البيانات البحثية.
- **الحدود الزمنية:** اعتمدت الدراسة حد زمني لاستكشاف وتحليل مستودعات البيانات البحثية العامة ينتهي بنهاية ديسمبر ٢٠٢٢، مع حصر للتوزيع الزمني والموضوعي لمجموعات البيانات المتاحة بمستودعات الدراسة في السنوات السبع الأخيرة (٢٠١٥-٢٠٢١)، هذا وقد تم جمع البيانات في الفترة من أغسطس إلى ديسمبر من عام ٢٠٢٢.

٦- منهج الدراسة:

قامت الدراسة الحالية على استكشاف مستودعات البيانات البحثية العامة والوقوف على خصائصها وسماتها الأولية هذا بالإضافة إلى تحليل المنهجيات التي تتبعها تلك المستودعات في نشر وإتاحة مجموعات البيانات؛ وعليه فقد فرضت طبيعة وتوجهات الدراسة الاعتماد على **المنهج الوصفي التحليلي** لانسجامه مع طبيعة هذه الدراسة وأهدافها، وذلك في وصف وتحليل النتائج وبلورة تساؤلات الدراسة واجوبتها بالإضافة إلى تحليل النتائج الكمية وعرضها. هذا وقد استعانت الدراسة **بمنهج تحليل المحتوي** لتحليل ورصد منهجيات نشر واتاحة مجموعات البيانات داخل مستودعات الدراسة.

٧- مصطلحات الدراسة:

- **بيانات البحث Research data:** النطاق الواسع وغير المتجانس من البيانات المنتجة أثناء نشاط البحث.
- **مجموعات البيانات Data sets:** وحدة موضوعية من البيانات المنشورة في نشاط البحث. أو هي حزمة من البيانات الموضوعية المرتبطة بالمنشور.
- **مستودعات البيانات البحثية العامة Generalist research data repositories**: هي المستودعات التي لا ترتبط بموضوع أو نشاط بحثي محدد وتعلن قبول أي مجموعة بيانات يقوم

عليها نشاط علمي أو بحثي بشكل أساسي فلا تفرض قيود على موضوع معين أو على شكل وتنسيق محدد وتسعى إلى توفير مجموعات البيانات دون اعتبار لبروتوكولات مؤسسية محددة.

٨- الدراسات السابقة:

لتحديد الاطر النظرية والعملية للدارسة ولمقارنتها بالأدبيات السابقة في مجالها، وتجنبا لتكرار الجهد وبيان أوجه الاتفاق والاختلاف بين الدراسة الحالية وبين غيرها من الدراسات وكخطوة أولى ورئيسة تم إجراء مسحاً شاملاً للإنتاج الفكري لرصد الدراسات السابقة والمثيلة التي أجريت حول موضوع الدراسة وذلك من خلال استقراء الإنتاج الفكري، ورصد ما كتب عن الموضوع، والبحث في قواعد البيانات المتاحة على الخط المباشر، ومحركات البحث على الانترنت، وعليه فقد تبين أن هناك بعض الدراسات التي تناولت مستودعات البيانات البحثية كواحدة من أهم روافد نشر البيانات البحثية، وذلك سواء على المستوى المفاهيمي للبيانات البحثية بصورة عامة أو من خلال رصد دوافع وأهداف النشر بها، كذلك من خلال دراسات الحالة لمستودعات مؤسسية محددة، أو دراسة القبول الدولي لتلك المستودعات من خلال تحديد وتحليل المستودعات في نطاق جغرافي محدد، وإذا ما حاولنا تصنيف هذه الدراسات نجد أنها تتدرج تحت أربعة فئات رئيسة هي:

- دراسات تناولت مستودعات البيانات في نطاق جغرافي محدد.
- دراسات تناولت الأطر المفاهيمية والتطبيقية لإدارة البيانات البحثية.
- دراسات حالة لمستودعات البيانات البحثية.
- دراسات تناولت دوافع نشر ومشاركة البيانات في مستودعات البيانات البحثية.

٨-١- دراسات تناولت مستودعات البيانات في نطاق جغرافي محدد.

- دراسة (Wadhwa, Joshi,2022)[٣] تناولت هذه الدراسة مستودعات بيانات البحث الهندية المدرجة في أرشيف مستودعات بيانات البحث (re3data) من حيث التغطية الموضوعية، وسياسات الوصول، وتنسيقات البيانات والمعايير والمواصفات المستخدمة للتنفيذ . يفهرس الأرشيف ٢٩٣٨ مستودعاً لبيانات البحث من جميع أنحاء العالم؛ يوجد في الهند (٥٠) مستودعاً لبيانات البحث في فئات نمطية مثل المستودعات المؤسسية والتخصصية. وقد خلصت الدراسة إلى أن المستودعات الهندية تغطي بشكل أساسي مواضيع مثل علوم الحياة والطب والعلوم الطبيعية. وتتكون معظم المستودعات من بيانات علمية وإحصائية .كما يوفر ثمانية وثلاثون مستودعاً وصولاً مفتوحاً إلى البيانات، (٤٢٪) من مستودعات البيانات الهندية تستخدم ترخيص حقوق النشر الذاتية، و(٢٢) موقعاً إلكترونياً لمستودعات البيانات تنشر معلوماتها حول العدد الإجمالي لمجموعات البيانات المتاحة في مستودعات البيانات.

- دراسة (Misgar et.al.2020,)[٤]تهدف هذه الدراسة إلى تحليل واستكشاف دور دول البريكس BRICS- البرازيل وروسيا والهند والصين وجنوب أفريقيا- تجاه مستودعات بيانات البحث ذات الوصول المفتوح والمسجلة في أرشيف مستودعات البيانات البحثية (re3data) وقد خلصت الدراسة إلى أن الهند أكبر مساهم في دول البريكس من حيث عدد مستودعات البحث، حيث توجد غالبية المستودعات في الهند (٤٥,١٢٪)، والصين هي ثاني أكبر مساهم بنسبة (٣٤,١٥٪) وهو الأمر الذي يشير إلى أن حجم الاقتصاد والسكان يمثلان عاملان رئيسيين في تطوير مستودعات البيانات البحثية، كما تتصدر الهند تنسيقات البيانات العلمية والإحصائية بعدد (٦٣) تنسيق. يكشف هذا أن العلماء في هذه البلدان يأخذون هذه المستودعات على محمل الجد ويقومون بتخزين البيانات ومشاركتها من خلال

هذه المنصات، كما بينت الدراسة أن اللغة الإنجليزية هي اللغة الأكثر تفضيلاً المستخدمة للتفاعل، وهذا يتماشى مع التوقعات، مع الأخذ في الاعتبار أن اللغة الإنجليزية هي اللغة الدولية الوحيدة التي يتم التحدث بها على نطاق واسع في جميع أنحاء العالم .العلوم الطبيعية هي الفئة الموضوعية الأكثر تغطية ضمن مستودعات البيانات البحثية (٥٢%)، وتوصي الدراسة بضرورة أن تبدأ دول البريكس تعاوناً بحثياً بين التكتل قد يزيد من أداء البحث ويساعد في تطوير المعرفة وتبادل البيانات البحثية .قد يؤدي تبادل المعرفة بين هذه الدول إلى زيادة تقدم البحث وسيساعد في فهم نقاط القوة والضعف.

٨-٢- دراسات تناولت الأطر المفاهيمية والتطبيقية لإدارة البيانات البحثية.

- دراسة (Hamad,et.al.2019)[٥]سعت هذه الدراسة إلى التحقق من المتطلبات والأدوار والمسؤوليات وكذلك تحديات خدمات إدارة البيانات البحثية في المكتبات الأكاديمية في الأردن .تم تطوير استبانة واستخدامها لجمع البيانات المطلوبة من ٢١ مكتبة أكاديمية في الأردن .تشير النتائج إلى إدراك ووعي عالٍ لأدوار ومسؤوليات المكتبات الخاصة بإدارة البيانات البحثية وكذلك المتطلبات والتحديات في المكتبات الأكاديمية في الأردن لتقديم خدمات البيانات البحثية، كما لوحظ أن عوامل مثل المسمى الوظيفي والخبرة ونوع المكتبة لم يكن لها أي تأثير على النتائج.
- دراسة (Patel, 2016)[٦] قدمت هذه الدراسة إطار مفاهيمي لإدارة بيانات البحث على المستوى المؤسسي، حيث استعرضت الدراسة بوضوح دورة حياة البيانات في مراحلها المختلفة منذ إنشائها، وتخزينها وتنظيمها ومشاركتها، كما حاولت أيضاً معالجة القضايا الحاسمة في ادارة بيانات البحث مثل خصوصية البيانات وأمن البيانات، وحقوق التأليف والنشر والترخيص، وذلك مع اقتراح نموذج للمستودع الوطني لبيانات البحث المفتوحة(NRORD) ، وقد خلصت الدراسة إلى أن الإطار المؤسسي قد يساعد في إدارة دورة حياة بيانات البحث بطريقة أكثر كفاءة وفعالية.
- دراسة (Higman, Pinfield,2015)[٧]هدفت هذه الدراسة إلى التحقق من العلاقة بين إدارة البيانات البحثية (RDM) ومشاركة البيانات في صياغة سياسات وتطوير الممارسات في مؤسسات التعليم العالي، حيث تم ذلك من خلال أولاً تحليل سياسات إدارة البيانات البحثية المؤسسية عبر مؤسسات التعليم العالي في المملكة المتحدة، وثانياً، التركيز على دراسات الحالة لمؤسستين محددتين هما جامعة بريستول، جامعة برونيل لندن، وقد خلصت الدراسة إلى أن تشكيل سياسة ادارة البيانات البحثية المؤسسية وتطوير الخدمات أدى إلى إنشاء مجموعة معقدة من الشبكات داخل وخارج المؤسسات التي تضم مجموعات مهنية مختلفة ومتنوعة على نطاق واسع، حيث تم اعتبار مشاركة البيانات نشاطاً مهماً في سياسات وخدمات مؤسسات التعليم العالي التي تمت دراستها، ولكن يمكن أن تُعزى أهميتها في معظم الحالات إلى المواقف التي يتبناها ممولي الأبحاث.

٨-٣- دراسات حالة لمستودعات البيانات البحثية.

- دراسة (Limani,2020)[٨] التي هدفت إلى فهم ممارسات إدارة البيانات البحثية في المستودع المؤسسي للبيانات بجامعة جنوب شرق أوروبا (SEEU)، وذلك من خلال استبيان موجه لتقييم أنشطة وخدمات إدارة بيانات البحث في بيئة جامعية، وتظهر نتائج المسح وجوداً متزايداً لبيانات البحث؛ بالإضافة إلى الحاجة إلى المستودع المؤسسي لدعم المجتمعات المشاركة. كما يكشف الاستطلاع أن الممارسات الشائعة بين الباحثين تتمثل في حفظ البيانات في جداول البيانات والوثائق النصية وقواعد البيانات العلائقية، والتي يديرونها شخصياً دون أي دعم مؤسسي .يقوم معظم الباحثين بتخزين هذه البيانات على جهاز حاسب شخصي أو محركات تخزين محمولة .على الرغم من ممارسات التخزين هذه، والتي ليست مثالية للمشاركة، يحتاج معظم المشاركين إلى مشاركة بياناتهم، ويقومون بذلك في

الغالب باستخدام مرفقات البريد الإلكتروني، ويكشف الاستطلاع أيضاً عن الميزات التي يجب أن توفرها المستودعات، معظمها أنشطة متعلقة بإنشاء البيانات البحثية ونشرها، ولكنها تغطي أيضاً تلك التي تتعلق بالنسخ الاحتياطي ومنح الوصول.

- دراسة (Anggawira, Mayesti,2020) [٩] تناقش هذه الدراسة مشاركة بيانات البحث من خلال دراسة حالة لممارسات المستودع العلمي الوطني الإندونيسي، الذي يديره مركز البيانات العلمية والتوثيق بالمعهد الإندونيسي للعلوم .الغرض من هذه الدراسة هو وصف عملية مشاركة بيانات البحث وتحديد العوامل الداعمة والعقبات التي واجهتها، استخدمت هذه الدراسة نهجاً نوعياً لدراسة الحالة، حيث تضمنت تقنيات جمع البيانات؛ الملاحظات الميدانية والملاحظات على نظام المستودعات؛ ومقابلات شبه منظمة مع العديد من الباحثين، وكذلك فرق التطوير وأمناء المكتبات؛ وتحليل وثائق السياسة والمبادئ التوجيهية، وخلصت الدراسة أنه على الرغم من تطوير المركز لنظام مستودع DataVerse لتمكين مشاركة بيانات البحث، لا تزال هناك العديد من المشكلات التي تعيق المستودع عن تحقيق الأهداف المؤسسية لزيادة الوصول إلى البيانات فهناك حاجة لمزيد من التدريب للباحثين لتشجيعهم وتحفيزهم على مشاركة بياناتهم البحثية من خلال هذه الخدمة؛ بالإضافة إلى ذلك، يحتاج الموظفون إلى اكتساب الكفاءة في إدارة البيانات ومعالجتها فلا يزال الباحثون وأمناء المكتبات المشاركون في أنشطة تبادل البيانات البحثية يواجهون عقبات مختلفة في مجالات السياسة والإعلان عن الخدمة والترويج لها.

٨-٤- دراسات تناولت دوافع نشر ومشاركة البيانات في مستودعات البيانات البحثية.

- دراسة (Mozersky,etal.,2019)[١٠] تهدف هذه الدراسة إلى الكشف عن دوافع وممارسات مشاركة البيانات النوعية الصحية في مستودع البيانات البحثية للمعهد الوطني للصحة، وذلك من خلال المقابلات المتعمقة شبه المنظمة مع ٣٠ مشاركاً في البحث النوعي، و٣٠ أميناً لمستودع البيانات، و٣٠ باحثاً نوعياً؛ لاستكشاف دوافعهم وخبرتهم ومعرفتهم بإدارة ومشاركة البيانات النوعية ، كذلك الاخلاقيات المرتبطة بمشاركة البيانات الصحية. وتشير النتائج التي توصلت إليها الدراسة إلى أن معظم الباحثين غير ملمين بمفهوم مشاركة البيانات النوعية، فليس جميعهم من ذوي الخبرة في تنظيم مجموعات البيانات النوعية، وأشاروا إلى أنهم يرغبون في الحصول على إرشادات ومعايير خاصة بمشاركة البيانات البحثية، مما قد يترك الباحثين الذين لديهم أقل قدر من المعرفة دون توجيه، كما أشارت الدراسة إلى أن ليس جميع البيانات النوعية مناسبة للمشاركة (على سبيل المثال ، لقطات الفيديو الحساسة التي لا يمكن إخفاء هويتها بشكل كافٍ، والبيانات الحساسة التي تم إخبار المشاركين بأنها لن يتم مشاركتها، والبيانات التي من شأنها أن تعرض المشاركين لضرر جسيم أو تفصح عن هويتهم)

- دراسة (Youngseek, Adler,2015)[١١] والتي هدفت إلى تحديد العوامل الفردية والمؤسسية والموارد التي تؤثر على سلوكيات مشاركة البيانات بين علماء الاجتماع في الولايات المتحدة الأمريكية. تم تطوير نموذج بحث والتحقق من صحته بناءً على نتائج دراسة استقصائية شملت ٣٦١ من علماء الاجتماع؛ لرسم خريطة للدوافع الفردية الأساسية والضغوط المؤسسية وتوافر الموارد التي تسهل مشاركة علماء الاجتماع للبيانات البحثية .وقد خلصت الدراسة إلى أن سلوكيات مشاركة البيانات لعلماء الاجتماع مدفوعة أساساً بالدوافع الشخصية (أي الفوائد والمخاطر المهنية، والجهود المتوقعة، والموقف تجاه مشاركة البيانات)؛ ومع ذلك، فإن الضغوط المؤسسية من قبل وكالات التمويل والدوريات ومستودعات البيانات بحاجة إلى مزيد من التشجيع لدعم سلوكيات علماء الاجتماع في مشاركة البيانات بشكل أفضل.

٨-٥-الدراسات العربية.

ما زالت الأدبيات العربية تتلمس خطواتها الأولى في موضوع مستودعات البيانات البحثية فنجد دراسة دراسة (السيد، وأخ.،٢٠٢٢) [١٢] تستكشف اتجاهات الباحثين في جامعة بني سويف تجاه بياناتهم البحثية وطرق التعامل معها تجميعاً وحفظاً ومعالجةً، وكذلك اتجاهات هؤلاء الباحثين نحو فكرة إنشاء مستودع رقمي للبيانات البحثية بجامعة بني سويف، معتمدة في ذلك على المنهج المسحي الميداني. ومستخدمة الاستبانة كأداة لجمع البيانات، وقد خلصت الدراسة إلى ميل المبحوثين نحو إنشاء مستودع رقمي لبياناتهم البحثية، وموافقتهم على مشاركة تلك البيانات مع غيرهم من الباحثين وخاصة في مجالات العلوم الاجتماعية والعلوم الطبيعية.

ودراسة (السيد، وأخ.، ٢٠٢١) [١٣] وهي واحدة من الدراسات التي تحاول الوقوف على ماهية البيانات البحثية وأنواعها. والتعرف على دورة حياة البيانات البحثية ومستودعاتها، وكذلك التعرف على أدلة رصد مستودعات البيانات البحثية .وذلك بالاعتماد على المنهج النظري الذي يعتمد على تجميع البيانات من مصادر المعلومات التقليدية والإلكترونية المتاحة لرصد الظاهرة موضوع الدراسة ووصفها وتفسيرها، هذا وقد خلصت الدراسة إلى أن البيانات البحثية تتضمن نتائج الملاحظات العلمية والاستبيانات والمقابلات وقوائم المراجعة وغيرهم من أدوات جمع البيانات، وقد تكون هذه البيانات في صورتها الخام أوفي صورة تجريبية. وقد تكون مطبوعة أو رقمية، كذلك قد تتعدد أشكالها ما بين إحصاءات وصور، أو ملفات فيديو. كما خلصت الدراسة أن دورة حياة البيانات البحثية لا تختلف كثيراً عن دورة حياة أي مشروع بحثي آخر حيث تبدأ بالتخطيط وتجميع البيانات ومعالجتها وتحليلها ونشرها ومشاركتها وحفظها وأخيراً إعادة استخدامها، كما عرفت الدراسة مستودعات البيانات البحثية وقسمتها وفقا للجهات القائمة على إعدادها ما بين مستودعات حكومية ومستودعات للبيانات الجامعية والمؤسسية ومستودعات للبيانات الشخصية، وأخيراً مستودعات البيانات الخاصة بوكالات التمويل، وألقت الدراسة الضوء على دليلي Databib،وRe3data كأشهر أدلة لحصر مستودعات البيانات البحثية.

كما قدم (فراج،٢٠٢٠) [١٤] افتتاحية ألقى من خلالها الضوء على مستودعات البيانات البحثية، والأسس التي تنبني عليها، وأنواعها، وسبل الوصول إليها داعيا المجتمعات العلمية العربية لإتاحة البيانات والعمل على توفير مستودعات للبيانات البحثية، سواء على مستوى مؤسسي أو بحسب التخصصات العلمية.

كما نجد دراسة (السيد، وأخ.، ٢٠٢١) [١٥] وهي واحدة من الدراسات التي تحاول الوقوف على ماهية البيانات البحثية وأنواعها. والتعرف على دورة حياة البيانات البحثية ومستودعاتها، وكذلك التعرف على أدلة رصد مستودعات البيانات البحثية .وذلك بالاعتماد على المنهج النظري الذي يعتمد على تجميع البيانات من مصادر المعلومات التقليدية والإلكترونية المتاحة لرصد الظاهرة موضوع الدراسة ووصفها وتفسيرها، هذا وقد خلصت الدراسة إلى أن البيانات البحثية تتضمن نتائج الملاحظات العلمية والاستبيانات والمقابلات وقوائم المراجعة وغيرهم من أدوات جمع البيانات، وقد تكون هذه البيانات في صورتها الخام أوفي صورة تجريبية. وقد تكون مطبوعة أو رقمية، كذلك قد تتعدد أشكالها ما بين إحصاءات وصور، أو ملفات فيديو. كما خلصت الدراسة أن دورة حياة البيانات البحثية لا تختلف كثيراً عن دورة حياة أي مشروع بحثي آخر حيث تبدأ بالتخطيط وتجميع البيانات ومعالجتها وتحليلها ونشرها ومشاركتها وحفظها وأخيراً إعادة استخدامها، كما عرفت الدراسة مستودعات البيانات البحثية وقسمتها وفقا للجهات القائمة على إعدادها ما بين مستودعات حكومية ومستودعات للبيانات الجامعية والمؤسسية ومستودعات للبيانات الشخصية، وأخيراً مستودعات البيانات الخاصة بوكالات التمويل، وألقت الدراسة الضوء على دليلي Databib،وRe3data كأشهر أدلة لحصر مستودعات البيانات البحثية.

هذا بالإضافة الى رسالة دكتوراه مسجلة (الزلباني، قيد الاعداد) [١٦] يسعى الباحث فيها لعمل مسح شامل لمشروعات إدارة البيانات البحثية الناجحة حول العالم ونظم المستودعات الخاصة بها للخروج بمعايير ومواصفات ونموذج تجريبي يمكننا تطبيقه في مؤسساتنا البحثية العربية.

ومن خلال استعراض ما سبق من دراسات، نجد أن الدراسات السابقة قد ركزت على تناول مستودعات البيانات البحثية في نطاق جغرافي محدد، وعلى الجانب الآخر وجدت العديد من الدراسات تناولت الأطر المفاهيمية والتطبيقية لإدارة البيانات البحثية، بالإضافة إلى دراسات حالة للعديد من مستودعات البيانات للوقوف على أنماط إدارة البيانات البحثية بها وفئات هذه البيانات ، كذلك دراسات تتمحور حول فهم الدوافع التي تقود الباحثين إلى استخدام مستودعات البيانات البحثية في مشاركة بياناتهم البحثية، وتستكمل الدراسة الحالية ما بدأته الدراسات السابقة حيث تعمد إلى استكشاف ورصد مستودعات البيانات البحثية العامة أي غير المحددة بتخصص موضوعي أو بفئة معينة من البيانات أو التي ينحصر استخدامها داخل مؤسسة بعينها – وهو مالم تتطرق إليه أيٍ من الدراسات السابقة- وذلك من خلال الوقوف على الخصائص والسمات الأولية لتلك المستودعات والمنهجيات التي تتبعها في نشر وإتاحة مجموعات البيانات، هذا وقد استفادت الدراسة الحالية من هذه الدراسات في جوانب متعددة، وبخاصة من النتائج التي توصلت إليها هذه الدراسات.

ثانيا: الإطار النظري.

١- البيانات البحثية: قضايا عامة:

عالم البيانات البحثية معقد، ويشارك فيه الكثير من الجهات الفاعلة التي تستخدم أنواعاً متعددة من البيانات للعديد من الأغراض العلمية المختلفة، شهدت السنوات الأخيرة ظهور عدد كبير من عمليات جمع البيانات القوية والمرنة، مع السماح بأنواع البيانات غير المتجانسة والبيانات الوصفية المرتبطة التي تم تطويرها لتلبية مجموعة واسعة من متطلبات مجتمعات البحث المتنوعة.

١-١- البيانات البحثية:

نعني بالبيانات البحثية القياسات العلمية أو الفنية والقيم المحسوبة والملاحظات أو الحقائق التي يمكن تمثيلها بالأرقام، أو الجداول، أو الرسوم البيانية، أو النماذج، أو النصوص، أو الرموز والتي تُستخدم كأساس للتفكير أو الحساب. ويتم إنشاؤها بواسطة مختلف الوسائل والأدوات، بما في ذلك الملاحظة أو الحساب أو التجريب. ويعتبر العلماء البيانات بمثابة تمثيلات دقيقة للعالم المادي ودليل لدعم الادعاءات [١٧]

يمكن تمييز البيانات بناءً على أصولها سواء كانت قائمة على الملاحظة أو حسابية أو تجريبية، البيانات القائمة على الملاحظة هي البيانات التي رُصدت من خلال الملاحظات المباشرة، بينما يتم إنتاج البيانات الحسابية عن طريق تنفيذ نموذج حاسوبي أو محاكاة؛ وتتميز بأنه يمكن إعادة إنتاجها .ويتم جمع البيانات التجريبية عن طريق إجراء التجارب؛ من حيث المبدأ يمكن إعادة إنتاج البيانات من التجارب بدقة، ومع ذلك، فمن الناحية العملية، قد لا يكون من الممكن إعادة إنتاج جميع الظروف التجريبية بنفس الدقة [١٨]

كذلك يمكن الإشارة إلى البيانات على أنها أولية أو مشتقة أو تم التحقق منها حيث تتكون البيانات الأولية من ملاحظات أصلية، مثل تلك التي يتم جمعها بواسطة القمر الصناعي وإرسالها إلى الأرض أو تم إنشاؤها بواسطة أداة أو جهاز استشعار أو تم جمعها عن طريق إجراء تجربة؛ بينما يتم إنشاء البيانات المشتقة من خلال أنشطة المعالجة حيث تخضع البيانات الخام في كثير من الأحيان لمراحل لاحقة من الصقل والتحليل، اعتماداً على أهداف البحث، وقد يكون هناك تعاقب من الإصدارات؛ في حين أن البيانات الأولية قد تكون الشكل الأكثر اكتمالاً، فإن البيانات المشتقة قد تكون أكثر قابلية للاستخدام من قبل الآخرين

لأن المعالجة عادة ما تجعل البيانات أكثر قابلية للاستخدام أو مرتبة أو مبسطة، وبالتالي زيادة وضوحها، ويتم إنشاء البيانات التي تم التحقق منها من خلال الأنشطة التنظيمية؛ وبالتالي ضمان جودتها ودقتها. [١٩]

١-٢- مجموعات البيانات:

تنقسم مجموعات البيانات إلى ثلاث فئات وظيفية [٢٠]:

أ. **مجموعات البيانات البحثية** هي نتاج واحدً أو أكثر من المشاريع البحثية المركزة وتحتوي عادةً على بيانات تخضع لمعالجة محدودة. قد تتوافق أو لا تتوافق مع معايير المجتمع البحثي، مثل معايير تنسيقات الملفات وهيكل البيانات الوصفية وسياسات الوصول إلى المحتوى.

ب. **مجموعات بيانات الموارد** تخدم مجتمعاً علمياً أو هندسياً واحداً، وغالبًا ما تنشئ هذه المجموعات الرقمية معايير على مستوى المجتمع البحثي إما عن طريق الاختيار من بين المعايير الموجودة مسبقاً أو عن طريق جمع ودمج مجموعة من المعايير لتطوير معايير جديدة.

ج. **مجموعات البيانات المرجعية** تهدف إلى خدمة شرائح كبيرة من المجتمع العلمي والتعليمي. والسمات المميزة لهذه الفئة من المجموعات الرقمية هي النطاق الواسع والمتنوع لمجتمعات المستخدمين بما في ذلك العلماء والطلاب والمعلمين من مجموعة متنوعة من التخصصات الموضوعية والانتماءات المؤسسية والجغرافية، ويعد التوافق مع المعايير القوية والراسخة والشاملة أمراً ضرورياً، وغالباً ما يكون لاختيار المعايير من خلال المجموعات المرجعية تأثير في إنشاء معيار عالمي.

١-٣- الجهات الفاعلة الرئيسية في عالم البيانات العلمية هي[٢١]

- **مؤلفي البيانات:** وهم أفراد أو فرق تشارك في أنشطة بحثية تنتج بيانات رقمية يتم إيداعها لاحقاً في مجموعة بيانات. تكمن اهتماماتهم في ضمان تمتعهم بفوائد عملهم، بما في ذلك الحصول على التقدير والاعتراف المناسبين، وأن نتائجهم يمكن أن تنتشر على نطاق واسع ويتم أرشفتها بشكل آمن.
- **مستخدمي البيانات:** وهم ممثلي المجتمعات العلمية، وتكمن اهتماماتهم في الحصول على مجموعات بيانات جاهزة ومفهومة ويمكن اكتشافها، أي مُعرفة جيداً وموثقة جيداً .
- **مديري البيانات:** وهم أفراد مسؤولون عن تشغيل وصيانة مجموعات البيانات .
- **علماء البيانات:** وهم علماء معلومات وخبراء حاسب آلي يطورون مفاهيم مبتكرة في تكنولوجيا قواعد البيانات وعلوم المعلومات، بما في ذلك نمذجة البيانات العلمية، واكتشاف البيانات، وتصور البيانات، وتطبيقها على مجالات العلوم ذات الصلة بجمع البيانات.

١-٤- استخدامات البيانات:

تُستخدم البيانات بطرق مختلفة وفقاً لسياقاتها، ويمكن تقسيم عمليات استخدام البيانات إلى فئتين: [٢٢]

- **الاستخدام النهائي:** وهو القدرة على الوصول إلى مجموعة بيانات للتحقق من بعض الحقائق أو أداء بعض المهام المتعلقة بالعمل أو المهام الشخصية .
- **الاستخدام المشتق:** وهو القدرة على البناء على مجموعة بيانات موجودة مسبقاً عن طريق استخراج المعلومات من مجموعة بيانات واحدة أو أكثر من أجل إنشاء مجموعة بيانات جديدة يمكن استخدامها لنفس الغرض أو لغرض مشابه أو مختلف تماماً فيما يتعلق بمجموعة البيانات الأصلية.

إجمالا، يمكننا أن نؤكد أن البيانات البحثية موجودة في العديد من الأنواع والصيغ المختلفة، وتخضع لقيود قانونية وثقافية وعملية مختلفة. وغالباً ما يأتي مؤلفي البيانات والمديرون والمستخدمون من بيئات موضوعية ومهنية وثقافية مختلفة؛ لها احتياجات وتوقعات ومسؤوليات وسلطات وخبرات مختلفة، ويخضع هؤلاء جميعاً على اختلافهم لقيود قانونية، ومادية، وعلمية، وثقافية، وغيرها، وعليه يعد التنوع في البيانات والأفراد والتخصصات والسياقات والثقافات هو التحدي الكبير الذي يواجه الباحثون من أجل تسخير البيانات والمعرفة المتراكمة التي تنتجها مجتمعات البحث وجعلها قابلة لإعادة الاستخدام.

١-٥-الأسس المفاهيمية لقابلية إعادة استخدام البيانات:

إن إعادة استخدام البيانات تعني وجود علاقة بين مؤلف البيانات ومستخدم البيانات؛ ويتطلب ذلك أن تكون البيانات متاحة على مستويات مختلفة من التجريد والتمثيل، ويتطلب أيضاً إثراء البيانات لغوياً من أجل أن تكون قادرة على عبور الحواجز الدلالية دون تشويه دلالي، ولوصف علاقة قابلية اعادة استخدام البيانات بين منتجي البيانات ومستهلكي البيانات، علينا النظر في مجموعة من النقاط الرئيسة أهمها الاختلافات المرتبطة بعلاقة قابلية إعادة الاستخدام بين مؤلف البيانات ومستخدم البيانات؛ حيث سيكون من المهم، تحديد الاختلافات في المعرفة ووجهات نظر مؤلف البيانات ومستخدم البيانات عند العمل في سياق نشاط بحث تعاوني متعدد التخصصات، كذلك من الضرورة عدم التغافل عن التغييرات التي تميز العلاقة، وإدراك أن الاختلافات والتبعيات التي تميز علاقة قابلية اعادة الاستخدام بين منتج البيانات ومستهلك البيانات تتغير بمرور الوقت، ويجب أن نفترض أن تلك العلاقة تخضع لعملية تحسين ومراجعة مستمرة من خلال التفاعل العلمي بين كلاً من مؤلف البيانات، ومستخدم البيانات. (٢٣)

يتم أيضاً إنشاء مجموعة غنية من المعرفة الصريحة والضمنية في جميع الأنشطة البحثية، مع إنتاج البيانات العلمية ومن أجل جعل البيانات العلمية قابلة لإعادة الاستخدام بشكل فعال، يجب أيضاً جعل المعرفة الداعمة الصريحة والضمنية قابلة لإعادة الاستخدام، ويشير مفهوم إعادة استخدام المعرفة إلى عملية نقل وإعادة استخدام قواعد المعرفة الموجودة في سياقات مختلفة بشكل كبير. وعليه فمن الضروري توظيف هذين النوعين من المعرفة كسلعة يمكن استخراجها وتمثيلها وتعبئتها في سياق معين - سياق منتج البيانات- ونقلها وإدراجها بسهولة في سياق آخر - سياق مستخدم البيانات- مما يعني أن كلا النوعين من المعرفة يجب أن يكونا جزءً من عملية نشر البيانات، أي العملية التي يتم من خلالها جعل البيانات العلمية قابلة للمشاركة وقابلة لإعادة الاستخدام. وتتمثل الصعوبة في جعل المعرفة قابلة لإعادة الاستخدام في حقيقة أن ما تم تدوينه في تخصص واحد قد لا يكون مفهوماً للمجالات والتخصصات الأخرى بسبب المحتوى الفكري والخلفية العلمية المطلوبة. (٢٤)

توفر بيانات البحث سرداً لنتائج العمل العلمي، وبالتالي يجب أن تكون مفهومة لمن يرغبون في فهمها أو تدقيقها، لذلك يجب التمييز بين اتصالات البيانات لفئات مختلفة من الجماهير ذات الخلفيات العلمية والمعرفية المختلفة .هذا يعني أن اتصال البيانات الفعال يجب أن يُمَكِن المستفيدين من فحص مجموعة البيانات على مستوى التجريد بما يلائم خلفيتهم العلمية واهتماماتهم البحثية.وهناك نوعان رئيسيان من التجريد: الأنطولوجي والمعرفي ؛ النهج الأنطولوجي للتجريد يهتم بمستويات التنظيم المختلفة لنظام يمكن تحديده وتعريفه، حيث يمكن أن تحتوي قاعدة البيانات على مستويات تنظيمية ومفاهيمية ودلالية ونحوية ومادية، ويهتم النهج المعرفي للتجريد بالمستويات المختلفة للملاحظة أو التفسير التي يمكن من خلالها دراسة النظام بما يمكن من ملاحظة قاعدة البيانات وتحليلها على مستويات مختلفة من التجريد، والتي تتألف من بيانات مرتبطة بالوقت أو المكان أو الأداة أو موضوع الملاحظة.(٢٥)

١-٦- تمثيل البيانات:

يعد تمثيل البيانات ضرورياً لتمكين العلماء من تفسير البيانات بشكل صحيح واستخدامها بشكل مناسب حيث يمكن تمثيل محتوى المعلومات نفسه بشكل مختلف في لغات وصف البيانات المختلفة، وتتمثل المشكلة الرئيسة في عدم وجود نموذج مفاهيمي رسمي مشترك لتمثيل البيانات يكون دقيقاً ومفصلاً بشكل كافٍ لدعم احتياجات البيانات للعلماء الذين ينتمون إلى تخصصات علمية مختلفة، حيث أن نموذج البيانات العلائقية التقليدي غير مناسب لتمثيل احتياجات البيانات لمعظم التخصصات العلمية، فبالنسبة لبعض التخصصات العلمية مثل علم الفلك، يكون نموذج بيانات المصفوفة أكثر ملاءمة، في حين أن بعض التخصصات الأخرى مثل علم الأحياء، وعلم الجينوم تعتبر الرسوم البيانية والتسلسلات أكثر ملاءمة لاحتياجاتهم، هذا ويعد التمثيل التصويري للبيانات ذا أهمية قصوى حيث يجعل عرض البيانات أكثر وضوحاً، ويسمح للمستفيدين برؤية السمات البارزة للبيانات بسهولة، وإبراز الأنماط والاتجاهات الخفية لمجموعات البيانات المعقدة .[٢٦]

١-٧-عوائق إعادة استخدام البيانات:

على الرغم من أهمية إعادة استخدام ومشاركة البيانات إلا أن هناك عدة عقبات قد تعوق تنفيذ هذا الأمر، [٢٧] ومنها:

١. **تمثيلات البيانات غير المتجانسة:** هناك مجموعة متنوعة من نماذج البيانات العلمية وتنسيقاتها والمعلومات العلمية المعبر عنها في شكل واحد ولا يمكن دمجها مباشرة في أشكال أخرى .

٢. **عدم تجانس لغات الاستعلام:** تتم إدارة مجموعات البيانات بواسطة مجموعة متنوعة من الأنظمة التي تدعم لغات الاستعلام المختلفة، ومن الصعب مشاركة البيانات إذا تم ترميزها ببرمجيات مختلفة.

٣. **عدم وجود اتفاقيات اتصال:** لا تتطلب إعادة استخدام البيانات بالضرورة قاعدة بيانات مشتركة .إذا كان بإمكان الأنظمة المنفصلة الاتصال ببعضها البعض، فيمكنها الاستفادة من قواعد بيانات بعضها البعض دون وجود قاعدة بيانات مشتركة .إلا أن هذا النهج غير ممكن بشكل عام في الوقت الحاضر فأنظمة قواعد البيانات العلمية ومستودعات الملفات لا تعتمد على نماذج البيانات الرسمية، وبالتالي، فإن جعلها قابلة للتشغيل البيني أمر صعب للغاية، إذ تفتقر إلى بروتوكول متفق عليه يحدد كيفية قيام هذه الأنظمة بالاستعلام عن بعضها البعض وبأي شكل يتم تسليم الإجابات والنتائج، كما تفتقر إلى البروتوكولات القياسية التي من شأنها أن توفر إمكانية التشغيل البيني بين البنى التحتية لبيانات البحث .

٤. **عدم تطابق المفردات:** عائق آخر لإعادة استخدام البيانات هو عدم وجود مفردات أو مصطلحات مشتركة على مستوى المجال العلمي الواحد.

هذا وتتعدد أساليب نشر وإتاحة البيانات البحثية فتتنوع ما بين إيداع البيانات في دوريات البيانات كذلك في الأرشيفات والمستودعات الرقمية ومجموعات المكتبات، كما يمكن نشرها بالمواقع الشخصية أو مواقع المختبرات والمواقع المؤسسية، هذا بالإضافة إلى تبادلها بصفة شخصية بين الباحثين والعلماء. إن إيداع البيانات في المستودعات الرقمية يعرف بالمعيار الذهبي؛ حيث تصبح البيانات أكثر قابلية للبحث والاسترجاع، كما تتميز بالاستدامة، ويمكن الاستشهاد بها وربطها مع المصادر الأخرى. [٢٨]

٢- مستودعات البيانات البحثية:

يطلق على مستودعات البيانات Data repositories عدة مسميات مثل أرشيفات البيانات Data archives، ومراكز البياناتData centers، وقواعد البيانات العلمية Scientific databases،

ومستودعات البيانات البحثية Research Data Repositories وهي عبارة عن بنى تحتية كبيرة لقواعد بيانات تم إعدادها لإدارة مجموعات بيانات الباحثين ومشاركتها والوصول إليها وأرشفتها .قد تكون المستودعات متخصصة وتنحصر في تجميع البيانات في مجال موضوعي محدد أو أكثر عمومية حيث تغطي مجالات معرفية أكبر، مثل العلوم أو العلوم الاجتماعية .قد تقوم المستودعات البيانات البحثية أيضاً بتجميع بيانات الخبراء عالمياً أو محلياً. كما تُعرف مستودعات البيانات البحثية أيضاً بأنها قواعد بيانات ضخمة أنشأت لإدارة مجموعات بيانات الباحثين بهدف أرشفتها، وتيسير سبل الوصول والاستشهاد بها، وإتاحة المشاركة فيها، وقد يشار إليها على أنها بنية أساسية كبيرة لقاعدة البيانات أو عدة قواعد بيانات تجمع مجموعات البيانات وتديرها وتخزنها للتحليل والمشاركة وإعداد التقارير.

تسمح مستودعات البيانات بفحص نتائج البحث وإثباتها ومراجعتها والتحقق منها من قبل خبراء آخرين، ويتيح وضع بيانات البحث عبر الإنترنت الوصول الفوري من قبل مجموعة متفرقة من الباحثين عالمياً لمشاركة النتائج وفهمها وتوليفها، حيث يوفر هذا التجميع والتوليف فرصة للتبصر والتقدم والسعي البشري الفريد لتحقيق فهم أكبر، كما تسمح مستودعات البيانات أيضاً بنشر البيانات السلبية أو المخفية ، وهي التجارب التي لم تنجح في الأساس؛ بما يتيح للباحثين الآخرين تجنب الطرق المسدودة لأولئك الذين جربوا طريقاً قبلهم ويمهد لإيجاد طريق أفضل نحو منطقة أكثر خصوبة للبحث العلمي.

يتم تصنيف مستودعات البيانات بشكل عام إلى أربعة أنواع رئيسة [٢٩]:

- **مستودع البيانات Data Warehouse**: يعتبر أكبر أنواع المستودعات يتم فيه جمع البيانات من عدة قطاعات أو مصادر أعمال .في هذا المستودع، تُستخدم البيانات المخزنة عموماً للتحليل وإعداد التقارير مما يساعد المستخدمين على اتخاذ القرار الصحيح في أعمالهم أو مشروعاتهم .
- **بحيرة البيانات Data Lake**: في هذا المستودع، يمكن أن تكون البيانات بأي شكل؛ منظم أو شبه منظم أو غير منظم. والسبب الرئيس لوجود بحيرة البيانات هو محدودية مستودعات البيانات ويساعد هذا النوع من مستودعات البيانات في الحصول على إدارة أفضل للبيانات والتحكم الكامل في البيانات الموجودة فيه.
- **سوق البيانات Data Mart**: يمكن اعتبار ذلك النوع من المستودعات مجموعة فرعية من مستودع البيانات حيث يركز على موضوع أو قسم معين أو أي منطقة محددة أخرى نظراً لأننا حصلنا على البيانات المخزنة لمنطقة معينة، ويمكن للمستخدم الوصول بسرعة إلى الإحصاءات دون قضاء الكثير من الوقت في البحث في مستودع بيانات كامل.
- **مكعب البيانات Data Cube**: يحتوي هذا المستودع على أكثر البيانات تعقيداً، ويمكن وصفها على أنها امتدادات متعددة الأبعاد لجداول مختلفة، وعادة ما يتم استخدامها لتمثيل البيانات المعقدة للغاية بحيث لا يمكن وصفها من خلال صفوف وأعمدة الجداول فقط، لذلك يمكن استخدام مكعب البيانات عندما نقوم بتحليل البيانات المتوفرة لدينا والتي تتجاوز الأبعاد الثلاثية.

كما يمكن تصنيف مستودعات البيانات وفقاً للجهات القائمة على إعدادها إلى أربع فئات [٣٠]:

- **المستودعات الشخصية**، وكما يتضح من اسمها هي المستودعات التي تهتم بجمع الإنتاج الفكري والعلمي الخاص بباحث معين قبل نشرها، وذلك لعرضها للبحث والتقييم.
- **مستودعات بيانات الجامعات والمؤسسات البحثية**، وهي باختصار منصة مفتوحة لنشر وأرشفة البيانات البحثية الجامعية.
- **مستودعات البيانات الخاصة بوكالات التمويل المتخصصة**، وهي المستودعات التي تحتوي على البيانات البحثية للمشاريع التي تقوم وكالات التمويل المختلفة بتمويلها.

- **مستودعات البيانات الحكومية**، وهي المستودعات التي يتم تجميع بياناتها من الجهات الحكومية المختلفة؛ بما يضمن الحفاظ على تلك البيانات وضمان توافرها على المدى الطويل.

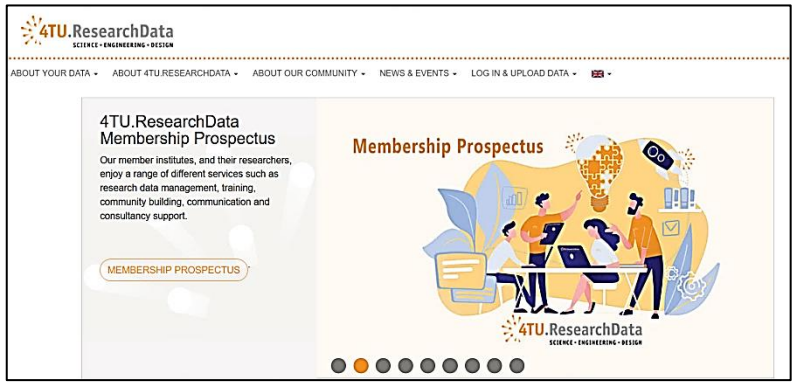
كذلك يمكن تقسيم مستودعات البيانات وفقا للتغطية الموضوعية إلى فئتين:

- **مستودعات البيانات المتخصصة** أي التي تتخصص في مجال موضوعي محدد.
- **مستودعات البيانات العامة** وهي المستودعات التي تسعى إلى تغطية مجموعة متنوعة من المجالات الموضوعية بصرف النظر عن الجهات التابعة لها.

وقد حصرت الدراسة الحالية سبع مستودعات للبيانات البحثية العامة وهم:

٢-١- 4TU.ResearchData [٣١] المعروف سابقاً باسم 3TU.Centre for Research Data ، بدأ في عام ٢٠٠٨ كمبادرة من ثلاثة جامعات تقنية في هولندا - جامعة دلفت للتكنولوجيا وجامعة أيندهوفن للتكنولوجيا وجامعة توينتي وكان الهدف الأول من إنشاء المستودع هو إنشاء منصة وطنية متطورة لتخزين والحفاظ على بيانات البحوث العلمية ولجعل هذه البيانات متاحة للجميع .يعمل المستودع بكامل طاقته منذ عام ٢٠١٠ وتطور ليصبح مستودعاً موثوقاً ومعتمداً للعلوم، وفي الوقت الحالي، يحتوي المستودع على ما يزيد قليلاً عن ٦٨٠٠ مجموعة بيانات، بما يتوافق مع حوالي ٤٢ تيرابايت من البيانات. تشارك 4TU.ResearchData في Research Data Netherlands، وهو تحالف وطني لأرشيف البيانات مهمته تعزيز الأرشفة طويلة المدى وإعادة استخدام بيانات البحث، ويتم استضافة 4TU.ResearchData وإدارتها بواسطة مكتبة TU Delft، حيث يتم تخزين مجموعات البيانات محلياً على خوادم TU Delft، وتحكمها البيانات القانونية لـ TU Delft

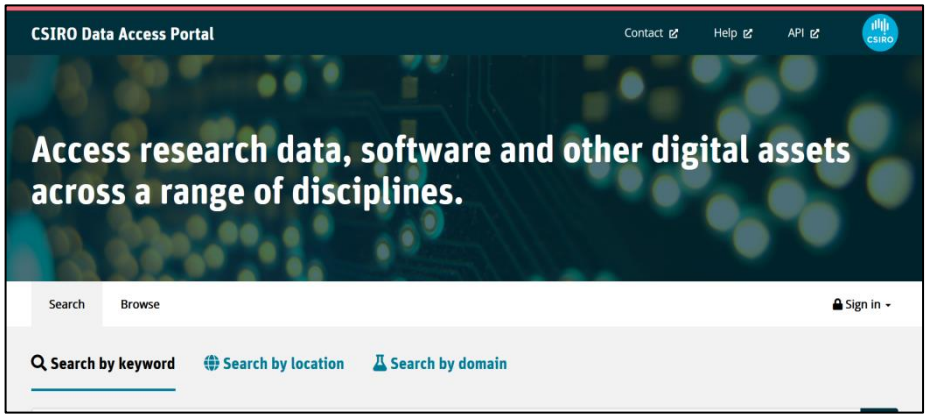
كما يتولى موظفو 4TU.ResearchData مسؤولية إدارة مستودع البيانات وصيانته، بما في ذلك سياسة الحفظ الخاصة به لضمان قابلية استخدام البيانات التي يحتفظ بها على المدى الطويل، على الرغم من الجذور الهولندية للمستودع، تقدم 4TU.ResearchData خدماتها إلى معاهد البحث من جميع أنحاء العالم وترحب بالمزيد من التعاون من خارج هولندا والاتحاد الأوروبي.



شكل رقم (١) صورة ضوئية للصفحة الرئيسية لمستودع 4TU.ResearchData

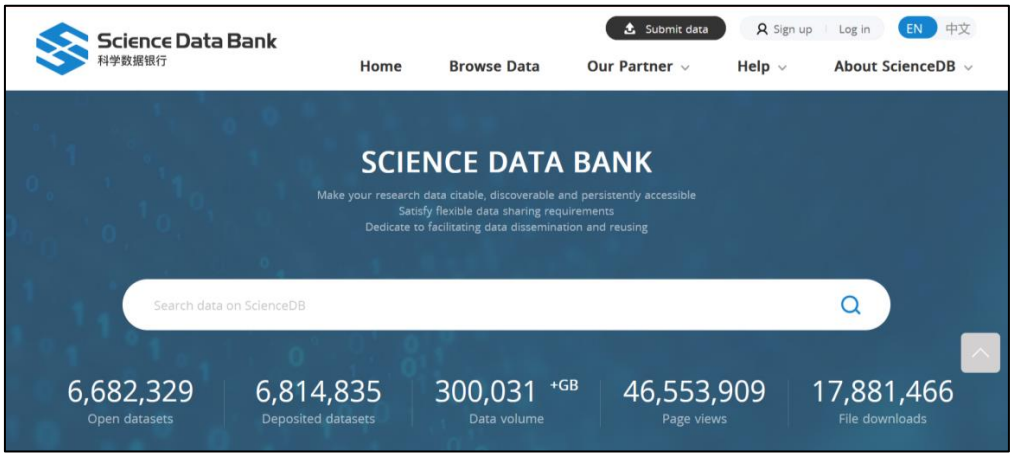
٢-٢- Csiro data access portal [٣٢] تأسس من قبل وكالة العلوم الوطنية الأسترالية CSIRO، يهدف إلى إدارة البيانات واكتشافها ومشاركتها عبر مجالات بحثية مختلفة، وهو جزء من خدمات البيانات تحت مظلة National الأسترالية يوفر Csiro data access portal الوصول إلى بيانات البحث والبرامج والأصول الرقمية الأخرى التي تنشرها CSIRO عبر مجموعة من التخصصات،

وتتمثل مهمة Csiro data access portal في توفير الوصول إلى بيانات أبحاثCSIRO، وتمكين إعادة الاستخدام، وضمان الحفظ والاستمرار على المدى الطويل. تم إصداره للجمهور في عام ٢٠١١ ويتضمن أصول البيانات من مجموعة واسعة من مجالات البحث، ويسعى المستودع للتأكد من أن المنظمة تجمع وتنشر وتدير البيانات الصحيحة لدعم الابتكار والتعاون والنزاهة العلمية بما يضمن تخزين بيانات البحث تحت وصاية CSIRO بشكل آمن، وتحديد موقعها بسهولة، وتوفيرها للآخرين لإعادة استخدامها؛ كذلك توفير القدرة على نشر البيانات لدعم قابلية التكاثر وسلامة البحث؛ وإدارة البيانات في بيئة بحثية تعاونية بشكل متزايد؛ والاستجابة لمتطلبات الحكومة والهيئة الممولة لمشاركة البيانات الناشئة عن الأبحاث الممولة من القطاع العام .المستودع عبارة عن تعاون بين فرق تكنولوجيا إدارة المعلومات بالإضافة إلى شركاء البحث الداخليين لتقديم حل شامل لإدارة بيانات البحث، ويعرف Csiro data access portal عن نفسه بأنه نقطة الالتقاط، وبوابة الاكتشاف، لأصول بيانات البحث الخاصة بـCSIRO .



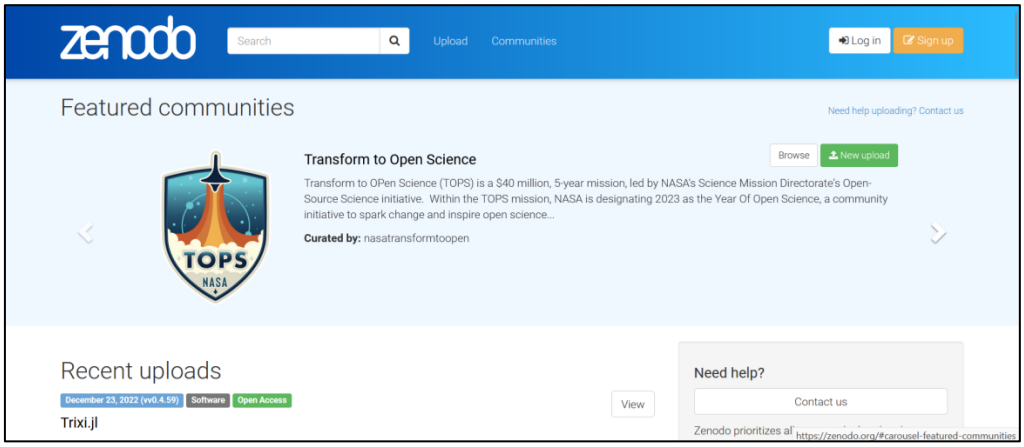
شكل رقم (٢) صورة ضوئية للصفحة الرئيسية لمستودع .Csiro data access portal

٢-٣- Science Data Bank [٣٣] هو مستودع بيانات عام مفتوح تم تطويره وصيانته بواسطة الأكاديمية الصينية لعلوم الحوسبة ومركز معلومات الشبكة (CNIC) وهو مستودع بيانات عام للأغراض العامة يهدف إلى توفير خدمات البيانات؛ مثل الحصول على البيانات، وحفظها على المدى الطويل، والنشر، والمشاركة، والوصول للباحثين ومشاريع وفرق البحث والدوريات والمؤسسات والجامعات، إلخ. يتم بناؤه وتشغيله بواسطة مركز معلومات شبكة الحاسب الصينية، والأكاديمية الصينية لعلوم الحوسبة. وتتمثل مهمته في نشر وتمثيل البيانات العلمية بما يتلاءم مع معايير واتفاقيات البيانات المتفق عليها، لتعزيز إمكانية العثور عليها وإمكانية الوصول إليها وإمكانية التشغيل البيني وإعادة استخدامها، ونشر ثقافة مشاركة البيانات في مجتمع العلوم الصيني .وينشر Science Data Bank مجموعات البيانات بجميع التنسيقات التي تم إنشاؤها من مجموعة كاملة من المجالات التخصصية، حيث يمكن للمستخدمين زيارة وتنزيل جميع موارد البيانات دون قيود كما يتميزScience Data Bank بخدمات آلية ثنائية اللغة، والتي تدعم ربط البيانات والتدفق الحر عبر الحدود اللغوية، بهدف النشر السريع للبيانات المفتوحة.



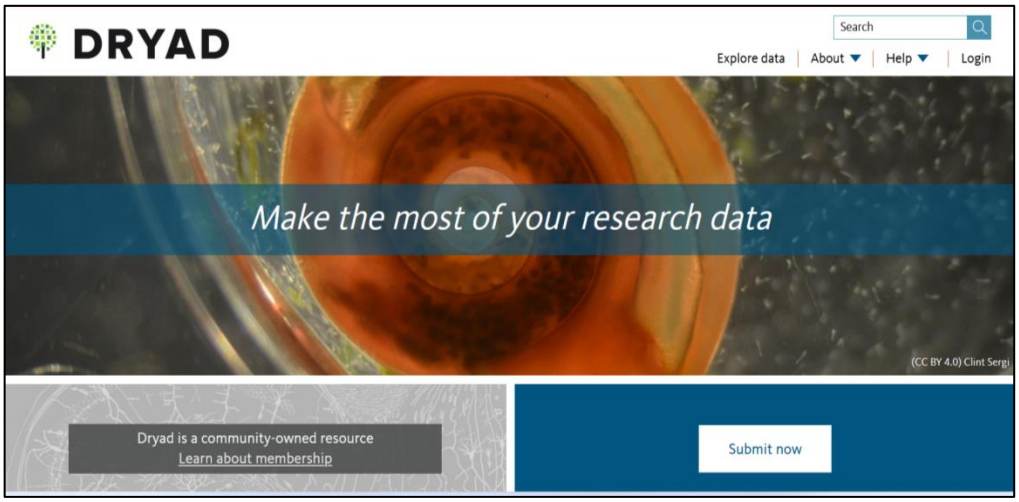
شكل رقم (٣) صورة ضوئية للصفحة الرئيسية لمستودع.Science Data Bank

٢-٤- Zenodo [٣٤] هو مستودع مفتوح للأغراض العامة تم إطلاقه ضمن مشروع OpenAIREplus التابع للاتحاد الأوروبي كجزء من بنية تحتية بحثية على مستوى أوروبا .يهدف إلى تمكين الباحثين من حفظ ومشاركة أي نوع من مخرجات البحث، مع التركيز بشكل خاص على تلك المنتجة في سياق الذيل الطويل للعلم .ويتم تمويل المستودع بشكل مشترك من قبل المفوضية الأوروبية عبر OpenAIRE وتستضيفه CERN. يسمح Zenodo للباحثين بإيداع الأوراق البحثية ومجموعات البيانات وبرامج البحث والتقارير وأي مواد رقمية أخرى متعلقة بالبحوث مع تحديد معرّف كائن رقمي دائم (DOI) لكل عملية إرسال، مما يجعل العناصر المخزنة قابلة للاستدلال عليها بسهولة. وقد تم إطلاق Zenodoفي الثامن من مايو ٢٠١٣ كخلف لمستودع OpenAIRE Orphan Repository للسماح للباحثين في أي مجال بالامتثال لأي شرط متعلق بإيداع العلوم المفتوحة في غياب مستودع مؤسسي. وأعيد إطلاقه باسم Zenodo في عام ٢٠١٥ لتوفير مساحة للباحثين لإيداع مجموعات البيانات.



شكل رقم (٤) صورة ضوئية للصفحة الرئيسية لمستودع.Zenodo

٢-٥- Dryad Digital Repository [٣٥] مستودع منظم للأغراض العامة يجعل البيانات قابلة للاكتشاف، ويمكن إعادة استخدامها بحرية، حيث يمكن إرسال معظم أنواع الملفات (مثل النصوص وجداول البيانات والفيديو والصور ورموز البرامج) بما في ذلك الأرشيفات المضغوطة لملفات متعددة وذلك نظراً لأن المبدأ التوجيهي لـ Dryad هو ضمان وصول طويل الأمد لمحتوياتها دون تكلفة للباحثين أو المعلمين أو الطلاب، بغض النظر عن الجنسية أو الانتماء المؤسسي. نشأ Dryad Digital Repository من مبادرة بين مجموعة من الدوريات الرائدة والجمعيات العلمية لاعتماد سياسة أرشفة بيانات مشتركة (JDAP) لمنشوراتهم، والاعتراف ببيانات مفتوحة وسهلة الاستخدام وغير هادفة للربح يحكمها المجتمع حيث انبثق Dryad من ورشة عمل "حفظ البيانات الرقمية ومشاركتها واكتشافها: تحديات مجتمعات العلوم الصغيرة في العصر الرقمي" في مايو ٢٠٠٧. يهدف Dryad Digital Repository إلى السماح للباحثين بالتحقق من صحة النتائج المنشورة، واستكشاف منهجيات تحليل جديدة، وإعادة تحديد الغرض من البيانات لأسئلة البحث التي لم يتوقعها المؤلفون الأصليون، وإجراء دراسات تركيبية، ويعمل Dryad كمستودع للجداول، وجداول البيانات، والملفات المسطحة، وجميع الأنواع الأخرى من البيانات المنشورة التي لا توجد مستودعات متخصصة لها بالفعل.

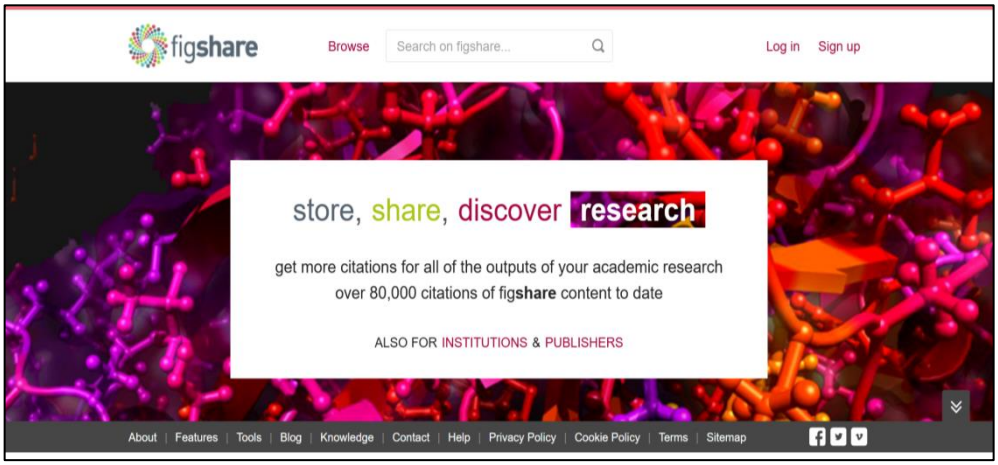


شكل رقم (٥) صورة ضوئية للصفحة الرئيسية لمستودع Dryad Digital Repository.

٢-٦- Figshare [٣٦] هو مستودع بيانات مفتوح الوصول حيث يمكن للباحثين نشر مخرجات أبحاثهم مع ضمان أرشفتها والحفاظ عليها، مثل مجموعات البيانات والصور ومقاطع الفيديو وجعلها قابلة للاكتشاف .يسمح Figshare للباحثين بتحميل أي تنسيق ملف وتعيين معرف كائن رقمي (DOI) للاستشهادات

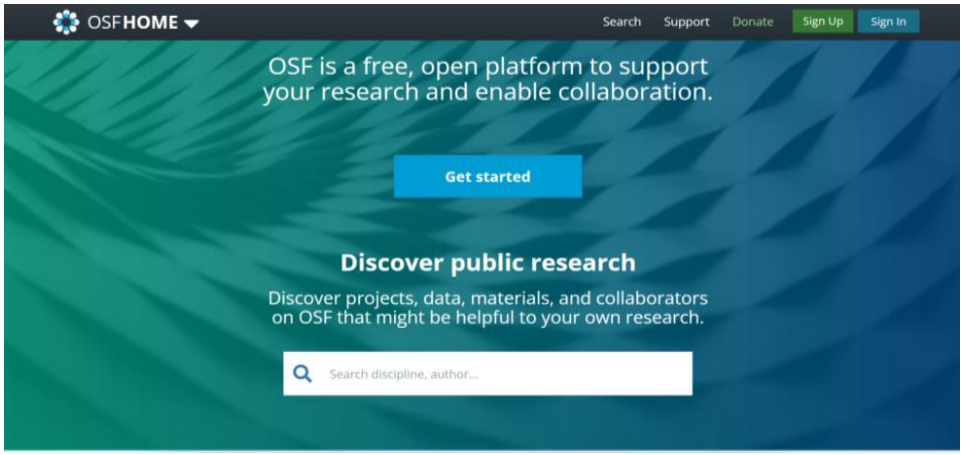
أطلق Mark Hahnel مستودع Figshare في يناير ٢٠١١، كوسيلة لتخزين وإدارة ونشر أي نوع من مخرجات البحث بسهولة.ويتم تمويله من قبل قسم التكنولوجيا العالمي في Macmillan Science & Education. وقد استثمر هذا القسم في سلسلة من المبادرات التي تهدف إلى توفير أدوات البرمجيات والتكنولوجيا للعلماء والباحثين، على سبيل المثال، خدمة Altmetric إحدى الخدمات التي تتعامل مع قياس أثر البحث حسب المقاييس غير التقليدية. وطور Hahnel لأول مرة المنصة كأداة شخصية لتنظيم ونشر

مخرجات الدكتوراه في بيولوجيا الخلايا الجذعية .وتستخدمه أكثر من ٥٠ مؤسسة الآن، وتصدر Figshare حالة البيانات المفتوحة كل عام لتقييم المشهد الأكاديمي المتغير حول البحث المفتوح.



شكل رقم (٦) صورة ضوئية للصفحة الرئيسية لمستودع . Figshare

٢-٧- Open Science Framework [٣٧] يُعد (OSF) مستودع مفتوح للتخصصات العلمية عامة وهو أداة تعزز تدفقات العمل المفتوحة والمركزية من خلال تمكين تجميع الجوانب والمنتجات المختلفة لدورة حياة البحث، وهو أداة تعاون مفتوحة المصدر مصممة لمساعدة الباحثين على توثيق دورة حياة مشروعهم البحثي؛ بما في ذلك تطوير فكرة البحث وتصميم ودراسة وتخزين وتحليل البيانات المجمعة، وكتابة التقارير، وأوراق البيانات، ونشرها. وتم تطويره وصيانته من قبل مركز العلوم المفتوحة (COS) وهو منظمة غير ربحية تأسست في ٢٠١٣ وتهتم بالممارسات العلمية، وتبني وتدعم مجتمعات البحث العلمي، وتطور أدوات البحث والبنية التحتية لتمكين إدارة البحوث وأرشفتها .



شكل رقم (٧) صورة ضوئية للصفحة الرئيسية لمستودع .Open Science Framework

ثالثا: تحليل بيانات الدراسة

١- جمع البيانات

تتعدد وتتزايد مستودعات البيانات وخاصة المتخصصة منها، وذلك على الرغم من الدعم الذي تقدمه كلا من الحكومات ووكالات التمويل والناشرين لدعم المستودعات العامة، والدفع بالمجتمع البحثي نحو إيجاد منهجية لنشر ومشاركة البيانات البحثية، تركز الدراسة الحالية على مستودعات البيانات البحثية العامة وللتحقق من تلك المستودعات تم البحث في أرشيف مستودعات البيانات البحثية Registry of Research Data Repositories (Re3data)[٣٨]، كذلك البحث من خلال محركات البحث وقواعد البيانات المختلفة عن الأدبيات المتعلقة بها، مع استكمال التحقق في صفحات الدوريات والروابط ذات الصلة، وفى الدراسة الحالية تم حصر سبعة مستودعات للبيانات البحثية العامة، ويقصد بالعمومية هنا عدم اقتصار تلك المستودعات على تخصص موضوعي محدد بل تمتد إلى تغطية مجموعة متنوعة من المجالات والتخصصات العلمية، دون التقيد بحدود مؤسسية أو دولية أو شكلية، هذا وقد تم جمع البيانات وفقا لمجموعة من المراحل التتابعية.

المرحلة الأولى (مستودعات البيانات البحثية العامة)

- تم تعريف مستودعات البيانات البحثية العامة بأنها المستودعات التي تعمد إلى نشر وإتاحة البيانات البحثية في مجموعة من التخصصات العلمية دون التقيد بموضوع أو بفئة محدد من البيانات، ودون أن ينحصر استخدامها في مؤسسة محددة أو لخدمة حكومة بعينها، هي مستودعات مفتوحة التخصص تنشر أي نوع من مجموعات البيانات البحثية.
- باستخدام الادبيات العلمية ذات الصلة وبالبحث في أرشيف مستودعات البيانات البحثية (Re3data) ومحركات البحث وقواعد البيانات على الخط المباشر وصفحات الروابط ذات الصلة مثل دوريات البيانات المختلفة؛ تم تحديد سبعة من مستودعات البيانات البحثية العامة.
- تم فحص النتائج وتصفيتها واستبعاد النتائج غير المطابقة وإلغاء التكرارات.

المرحلة الثانية (تحليل النتائج)

حددت الدراسة سبع مستودعات للبيانات البحثية العامة، وهي (4TU.ResearchData، و Csiro data access portal، وScience Data Bank،وZenodo، وDryad Digital Repository، وFigshare، وOpen Science Framework) وقد عمدت الدراسة في هذه المرحلة إلى تحليل الخصائص والسمات النوعية والفئوية لتلك المستودعات – محل الدراسة - بالإضافة إلى تحليل التغطية الزمنية والموضوعية لها.

المرحلة الثالثة (تحليل المحتوى)

أوضحت الدراسة في هذه المرحلة منهجية تعامل مستودعات البيانات البحثية العامة مع البيانات والمجموعات البحثية، وتأثير تلك المنهجيات في دعم وتعزيز نشر ومشاركة وإعادة استخدام البيانات وذلك من خلال تحديد ثمانية خصائص رئيسة؛ ترى الباحثة أن تلك الخصائص تشكل دوراً هاماً في توصيف الدعم الذي تقدمه هذه المستودعات لنشر البيانات البحثية ومشاركتها وإعادة استخدامها.

٢- نتائج الدراسة:

٢ -١- مستودعات البيانات البحثية العامة: الخصائص والسمات الأولية.

يقصد بخصائص وسمات مستودعات البيانات البحثية العامة التعريف بالانتماءات والتبعية الإدارية والدولية لتلك المستودعات وتاريخ نشأة وانطلاق تلك المستودعات، بما يسمح بإلقاء نظرة عامة على مستودعات البيانات محل الدراسة، ويستعرض الجدول التالي (جدول رقم١) تلك الخصائص الأولية؛ حيث يوضح تنوع مستودعات البيانات البحثية المتاحة على الخط المباشر ما بين مستودعات حكومية تخضع لمؤسسات الدولة، ومستودعات بيانات تابعة للجامعات والمؤسسات البحثية، ومستودعات البيانات التي تتبع وكالات تمويل متخصصة؛ أثنين من المستودعات العامة محل الدراسة (٢٨,٥%) تخضعا لإشراف أكاديمي وهما(4TU.ResearchData) حيث يشرف عليه ثلاثة جامعات هولندية هم جامعة دلفت للتكنولوجيا وجامعة أيندهوفن للتكنولوجيا وجامعة توينتي، و(Science Data Bank) الذي يخضع للمتابعة والاشراف الأكاديمي من قبل الأكاديمية الصينية لعلوم الحوسبة و شبكة الحاسب الصينية، كما أن أثنين من مستودعات الدراسة(٢٨,٥%) يخضعا لإشراف الجهات الحكومية وهم(Csiro data access portal) الذي تشرف عليه وكالة العلوم الوطنية الاسترالية، و(Zenodo) الذي يخضع لإشراف OpenAIREplus التابع للاتحاد الأوروبي، وأخيراً ثلاثة مستودعات من مستودعات الدراسة(٤٢,٨%) يتبعوا وكالات تمويل غير ربحية وهم (Dryad Digital Repository) تشرف علية مجموعة من الدوريات والجمعيات العلمية، و(Figshare) الذي يشرف عليه .Macmillan Science & Education ، و(Open Science Framework) التابع لمركز العلوم المفتوحة.

جدول رقم (١) الخصائص والسمات الأولية لمستودعات البيانات البحثية العامة – محل الدراسة-

المستودع	التبعية الادارية	الدولة	النشأة
4TU.ResearchData	أكاديمية	هولندا	2008
Csiro data access portal	حكومية	استراليا	2011
Science Data Bank	أكاديمية	الصين	2015
Zenodo	حكومية	الصين	2013
Dryad Digital Repository	وكالة تمويل	امريكا	2008
Figshare	وكالة تمويل	ألمانيا	2011
Open Science Framework	وكالة تمويل	الصين	2011

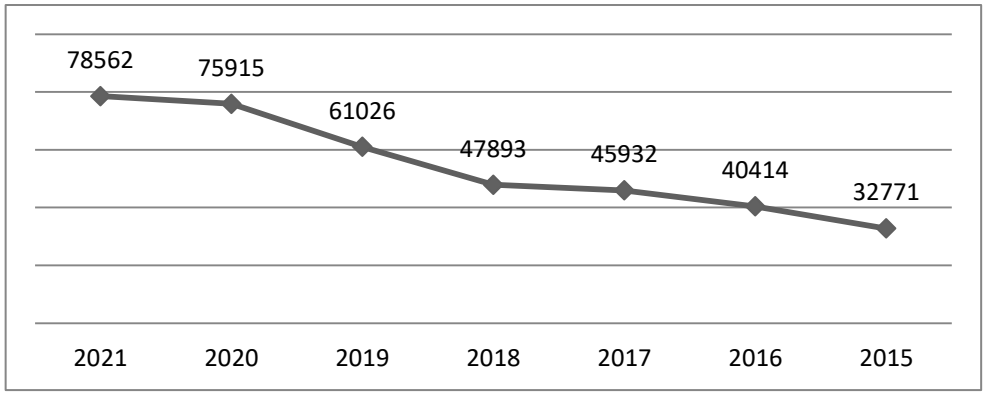
يتبين من الجدول السابق (جدول رقم١) أن أعمار المستودعات محل الدراسة تتراوح بين ست سنوات لمستودع(Science Data Bank) وهو أصغرهم عمراً، وثلاثة عشر عاماً لمستودعي (4TU.ResearchData) ،و (Dryad Digital Repository) وهما أكبر المستودعات محل الدراسة عمراً ، كذلك تتنوع التبعية الدولية لمستودعات الدراسة ما بين جمهورية الصين الشعبية؛ مستودع (Open Science Framework)، ومستودع(Zenodo)، ومستودع(Science Data Bank)، وهولندا (4TU.ResearchData)، واستراليا (Csiro data access portal)، والولايات المتحدة الامريكية (Dryad Digital Repository)، وألمانيا(Figshare) مثل هذا التنوع الإداري والدولي وحتى العمري يضمن بصورة أو بأخرى تنوع التغطية الموضوعية على الرغم من عموميتها ويضمن تنوع واختلاف الممارسات وفقاً لطبيعة ونشاط المؤسسات الإدارية المشرفة على المستودعات محل الدراسة.

٢-٢-التوزيع الزمني لنشر مجموعات البيانات في مستودعات البيانات البحثية العامة – محل الدراسة.

يستعرض الجدول والشكل التاليين (جدول رقم٢)، و(شكل رقم٨) التوزيع الزمني لنشر مجموعات البيانات في مستودعات البيانات البحثية محل الدراسة ويتبين منهما الارتفاع الطردي لمجموعات البيانات البحثية المتاحة في مستودعات الدراسة خلال السنوات السبع الأخيرة من ٢٠١٥ إلى ديسمبر ٢٠٢١. ويؤكد ذلك على قبول المجتمع العلمي لهذا الشكل من البيانات البحثية، وقبول الباحثين لفكرة مشاركة البيانات وإعادة استخدامها.

جدول رقم (٢) التوزيع الزمني لنشر مجموعات البيانات في مستودعات البيانات البحثية العامة

المستودع	2015	2016	2017	2018	2019	2020	2021	مجموع
4TU.ResearchData	118	109	269	446	602	689	918	3151
Csiro data access portal	222	416	418	493	720	749	853	3871
Science Data Bank	230	321	438	454	620	711	901	3675
Zenodo	878	1200	1609	1990	1997	2424	2687	12785
Dryad Digital Repository	7918	8110	9676	10424	10713	11306	11971	70118
Figshare	22803	29321	32415	32818	44971	58221	58623	279172
Open Science Framework	602	937	1107	1268	1403	1815	2609	9741
مجموع	32771	40414	45932	47893	61026	75915	78562	382513
%	8.5673	10.565	12.01	12.52	15.95	19.85	20.54	100.00



شكل رقم (٨) التوزيع الزمني لنشر مجموعات البيانات في مستودعات البيانات البحثية العامة.

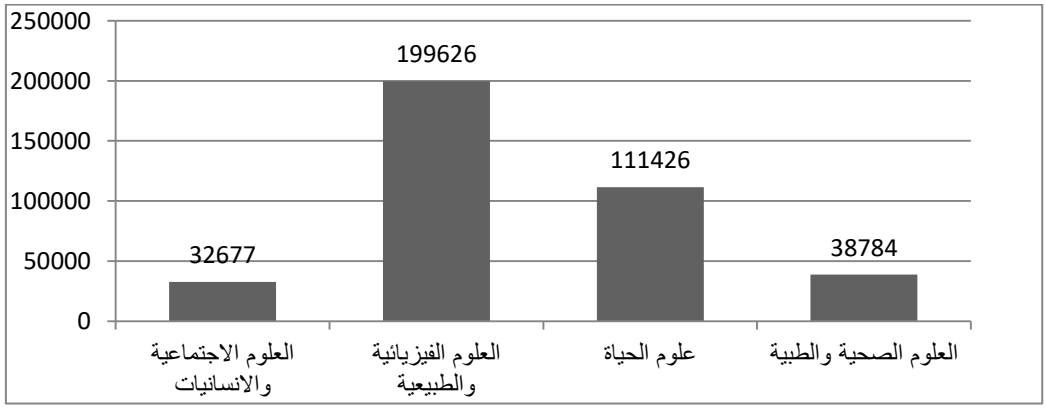
٢-٣- التوزيع الموضوعي لمجموعات البيانات البحثية المنشورة في مستودعات البيانات البحثية العامة- محل الدراسة.

قسمت الدراسة مجموعات البيانات المتاحة والمنشورة في المستودعات البيانات البحثية العامة وفقا لتصنيف Scopus للعلوم في أربعة قطاعات موضوعية عريضة هي (العلوم الاجتماعية والانسانيات وتشمل إدارة الأعمال، والمحاسبة، والسياسة، والاقتصاد، وعلم النفس، والتعليم، والفلسفة، واللغات، والآداب، والديانات، والفنون الجميلة، وفنون الأداء، والتاريخ، والجغرافيا، وعلوم المعلومات) و(العلوم الفيزيائية والطبيعية وتشمل الكيمياء الهندسية، والكيمياء، والحاسب الآلي، وعلوم الأرض والكواكب، وعلوم الطاقة والهندسة، والعلوم البيئية، والرياضيات، والفيزياء، والفلك) و(علوم الحياة وتشمل علوم النبات، والحيوان، والعلوم الزراعية، الجيولوجيا، والحفريات، وعلم الوراثة، والأحياء الدقيقة) و(العلوم

الصحية والطبية وتشمل الطب، والتمريض، والطب البيطري، وطب الأسنان، والمهن الصحية، وعلوم الصيدلة والدواء)

جدول رقم (٣) التوزيع الموضوعي لمجموعات البيانات البحثية المنشورة في مستودعات البيانات البحثية العامة

المستودع	العلوم الاجتماعية والانسانيات	العلوم الفيزيائية والطبيعية	علوم الحياة	العلوم الصحية والطبية	مجموع
4TU.ResearchData	318	1512	980	341	3151
data access portal Csiro	504	1293	1143	931	3871
Science Data Bank	340	1631	1296	408	3675
Zenodo	1108	5210	4390	2077	12785
Dryad Digital Repository	9790	28669	22577	9082	70118
Figshare	19709	157399	78400	23664	279172
Open Science Framework	908	3912	2640	2281	9741
مجموع	32677	199626	111426	38784	382513
%	8.54	52.19	29.13	10.14	100.00



شكل رقم (٩) التوزيع الموضوعي لمجموعات البيانات البحثية المنشورة في مستودعات البيانات البحثية العامة

يتضح من الجدول والشكل السابقين؛ (جدول رقم ٣) و(شكل رقم ٩) أن: العلوم الفيزيائية والطبيعية هي العلوم الأكثر تمثيلاً في مجموعات البيانات المنشورة في مستودعات الدراسة بنسبة (٥٢,١٩%) تليها علوم الحياة بنسبة (٢٩,١٣%) ثم العلوم الصحية والطبية بنسبة (١٠,١٤%) وأخيراً العلوم الاجتماعية والانسانيات بنسبة (٨,٥٤%) وغالبا ما تعود التغطية الموضوعية لمجموعات البيانات إلى الطبيعة التخصصية لجهات الاشراف على مستودعات البيانات محل الدراسة أنفسهم حيث ترى الباحثة أن هذه الأرقام متحيزة جزئياً، كذلك من المعلوم أن تخصصات موضوعية مثل العلوم الصحية والطبية تجنح إلى نشر البيانات البحثية في دوريات البيانات المتخصصة[٤٠] وكذلك الحال في تخصصات علوم الحياة وهو الأمر الذي ربما يفسر التطرف في التوزيع الموضوعي لمجموعات البيانات البحثية بمستودعات الدراسة.

٢-٤- مستودعات البيانات البحثية العامة: منهجيات نشر وإتاحة مجموعات البيانات.

تم تحليل مستودعات البيانات البحثية العامة من خلال فحص كيفية تعامل المستودعات محل الدراسة مع ثماني خصائص تشكل دوراً رئيساً في توصيف الدعم الذي تقدمه هذه المستودعات لنشر مجموعات البيانات البحثية ومشاركتها وإعادة استخدامها. وهذه الخصائص هي:

١. التنسيق (Formatting)، وتعني ترتيب مجموعة البيانات وفقاً لتنسيق معين لضمان سهولة استخدام مجموعة البيانات على المدى الطويل.

٢. التوثيق (Documentation)، أي مدى استكمال مجموعة البيانات بالمعلومات لتسهيل استرجاعها وفهمها وإعادة استخدامها من قبل كل من البشر والآلات.

٣. الترخيص (Licensing)، تصف سياسات وتراخيص نشر البيانات وإعادة استخدامها بما في ذلك تراخيص وحقوق الوصول إلى تلك البيانات.

٤. تكاليف النشر(Publication costs)، أي مبلغ التمويل الذي سينفقه المؤلف لنشر مجموعة البيانات، لضمان الوصول العادل والمفتوح إليها.

٥. التحقق (Validation)، ويقصد بها إجراءات تقييم جودة أو صحة وسلامة البيانات المنشورة، بما يتلاءم مع الغرض من نشرها.

٦. التوافر (Availability)، أي الدرجة التي تكون بها مجموعات البيانات المنشورة تحت تصرف المستفيد، لضمان الوصول إلى مجموعات البيانات المنشورة بمرور الوقت.

٧. قابلية الاكتشاف والوصول (Discovery and Access)، أي التسهيلات التي تمكّن المستفيدين من تحديد مجموعات البيانات المحتملة الفائدة والحصول عليها.

٨. الاستشهاد المرجعي والاقتباس (Citation)، ويقصد بها توفير مرجع لمجموعة أو مجموعات البيانات التي تصف خصائص البيانات التي تتيح الاعتمادات والإسناد، والاكتشاف، والربط، والوصول.

هذه الخصائص الثمانية تغطي المرحلتين الرئيستين اللتين تمر بهما عملية نشر البيانات البحثية، وهما (١) مرحلة التحضير والتقديم وفيها يتم تنسيق مجموعات البيانات وتوثيقها وترخيصها وتحديد تكلفة النشر، و(٢) مرحلة الاستهلاك وفيها يجب أن تكون مجموعات البيانات قابلة للتقييم والاتاحة والوصول؛ وعليه يتم مناقشة تلك الخصائص من خلال رصد وتحليل المنهجيات التي تطبقها المستودعات محل الدراسة.

٢-٤-١- التنسيق (Formatting)

في مستودعات البيانات، هناك على الأقل مفهومان للتنسيق: أولاً تنسيق الملف؛ أي الطريقة التي يتم بها تشفير مجموعة البيانات في واحد أو أكثر من الملفات، وثانياً تنسيق المحتوى؛ أي الطريقة التي يتم بها تنظيم محتوى مجموعة البيانات بالفعل. على سبيل المثال، عند التعامل مع البيانات الجدولية يمكن أن يكون تنسيق الملف هو csv بينما يمكن أن يكون تنسيق المحتوى وفقا لطريقة تنظيم البيانات في الجدول. وجود البيانات المنسقة بشكل مناسب هو شرط مسبق لأي استخدام لها حيث يؤثر تنسيق الملف على قدرة البرامج الحالية والمستقبلية على استيراد محتوى مجموعة البيانات؛ بينما يؤثر تنسيق المحتوى على تفسير وفهم مجموعة البيانات.[٤١]

تتعامل جميع المستودعات محل الدراسة مع الملفات وتنسيقاتها حيث يُطلب من مؤلفي ومالكي البيانات تزويد المستودعات بالملفات التي تشتمل على مجموعات البيانات وكذلك مجموعة من البيانات التوثيقية التي يتم تزويد المستفيدين بها عند حصولهم على مجموعات البيانات، أما بالنسبة إلى تنسيق المحتوى، فقد أهملت كافة المستودعات محل الدراسة الإشارة إلى كيفية تنظيم مجموعة البيانات، وعليه يركز التحليل الحالي على تنسيق الملفات فقط.

تقبل جميع المستودعات محل الدراسة مجموعات البيانات المخزنة في ملفات بأي تنسيق، فلا توجد قيود معينة على التنسيقات المسموح بها .ومع ذلك، غالباً ما يتم تشجيع مؤلفي ومالكي البيانات على إرسال البيانات بتنسيقات قياسية أو مفتوحة حيث تكون مناسبة للحفظ وإعادة الاستخدام .على سبيل المثال، يوفر 4TU.ResearchData للمستخدمين جدولاً يتضمن التنسيقات المفضلة لأي نوع من الملفات، وبالتحديد التنسيقات المثلى المستخدمة لحفظ البيانات على المدى الطويل والتي تضمن تهجير ونقل البيانات وتحويلها وتطبيعها وكذلك عرضها بتنسيقها الأصلي. ويوصي Dryad بتنسيقات csv و XMLو JSON كما يعلن Csiro data access portal عن التنسيقات المخصصة المدعومة لمجموعات بيانات محددة على سبيل المثال، جميع الملفات التي تؤلف مجموعات بيانات ATNF Astronomy Observations تتوافق مع معيار PSRFITS.

4TU.ResearchData
SCIENCE · ENGINEERING · DESIGN

Preferred File Formats

The choice of file format is of essential importance in order to ensure that the research data will remain usable in the future. 4TU.ResearchData therefore strongly encourages the use of standard, exchangeable or open file formats. For the preferred formats, 4TU.ResearchData guarantees that the research data will remain accessible and that they will be migrated or converted if necessary.

Preferred (sustainable) file formats that receive full preservation	
Text	Plain text, XML, HTML, PDF (PDF/A-1), JSON, PDB (Protein Data Bank), XYZ *(all formats should be encoded in UTF-8)*
Spreadsheets	CSV (Comma-separated values), Tab-delimited values.
Images	JPEG, TIFF, PNG, SVG
Geospatial	GML (Geographical Mark-up Language), KML (Keyhole Mark-up Language), ESRI Shapefile, Geo-referenced TIFF
Numerical	NetCDF, CSV, JSON
Video	*No sustainable format established*
Audio	Waveform Audio File Format (WAVE)
Databases	Delimited Flat File w/DDL
Archives	ZIP, TAR, GZIP, 7Z

شكل رقم (١٠) صورة ضوئية لملف التنسيقات المفضلة في مستودع *4TU.ResearchData* [٤٢]

جدول رقم (٤) تنسيقات الملفات التي تستخدمها مجموعات البيانات المنشورة في المستودعات الرقمية محل الدراسة.

Format	4TU.ResearchData	Csiro data access portal	Science Data Bank	Zenodo	Dryad Digital Repository	Figshare	Open Science Framework
adf	✓	✓	✓	✓	✓	✓	✓
av	✓	✓	✓	✓	✓	✓	
cif	✓		✓	✓	✓	✓	✓
csv	✓	✓	✓	✓		✓	✓
dat	✓	✓	✓	✓	✓	✓	✓
doc	✓		✓	✓	✓	✓	✓
docx	✓	✓		✓	✓	✓	
Excel	✓	✓	✓	✓	✓	✓	✓
ext/plain	✓	✓	✓	✓	✓	✓	✓
fits		✓	✓	✓		✓	✓
FTp	✓		✓	✓	✓	✓	✓
gz	✓		✓	✓	✓	✓	
itp	✓	✓	✓	✓	✓	✓	✓
jpg		✓	✓	✓	✓	✓	✓
mpeg	✓	✓	✓	✓	✓	✓	✓
nit	✓	✓	✓	✓	✓	✓	✓
octet-stream	✓	✓			✓	✓	✓
ods	✓	✓	✓	✓	✓	✓	✓
pdf	✓	✓	✓	✓	✓	✓	✓
png	✓	✓	✓		✓	✓	✓
text/plain	✓	✓	✓	✓	✓	✓	
tiff	✓	✓	✓	✓	✓	✓	✓
txt	✓	✓		✓		✓	✓
x-fasta	✓	✓	✓	✓	✓	✓	✓
x-gzip	✓		✓		✓	✓	✓
x-hdf5	✓	✓	✓	✓	✓	✓	
xlsx	✓	✓	✓	✓		✓	✓
x-msvideo	✓	✓	✓	✓	✓	✓	✓
x-netcdf	✓	✓	✓	✓		✓	✓
zip	✓	✓	✓	✓	✓	✓	✓

يعرض الجدول السابق (جدول رقم٤) تنسيقات الملفات الأكثر شيوعاً التي تستخدمها مجموعات البيانات المنشورة في المستودعات الرقمية محل الدراسة بشكل عام، يمكن ملاحظة مجموعة كبيرة ومتنوعة من التنسيقات بما في ذلك الملفات المضغوطة ZIP مما يجعل من الصعب معرفة عدد الملفات التي تحتوي عليها بالفعل وما هي تنسيقاتها، والتنسيقات العامة تتراوح من pdf إلى تنسيقات الجداول مثل csvو Excel ، هذا بالإضافة إلى تنسيقات خاصة بالمجال، على سبيل المثال، في Csiro data access portalهناك استخدام كبير للصيغ الشائعة في علم الفلك، وبصرف النظر عن تنسيقات الملفات، فإن المستودعات لها بعض القيود على أحجام الملفات المسموح بها؛ إذ تميل المستودعات إلى تحديد حد أقصى لأحجام الملفات يمكن تجاوزه بتكاليف إضافية حيث يسمح Dryad بتحميل ما لا يزيد عن ١٠ جيجابايت من المواد لمنشور واحد؛ ويسمح 4TU.ResearchData بتحميل مجموعات البيانات حتى ٤ جيجابايت؛ كما يقبل Zenodo ملفات تصل إلى ١٠جيجابايت؛ ويتيح Figshare للمستخدمين تخزين ما يصل إلى ١ جيجابايت من البيانات كذلك يسمح Science Data Bank بتحميل ما يزيد عن ٥ جيجابايت من البيانات، ويُمَكن Csiro data access portal المستخدمين من تخزين ما يصل إلى ١٠ جيجا بايت من البيانات في مساحتهم الخاصة مع ملفات يصل حجم كل منها إلى ٢٥٠ ميجا بايت. وأخيراً يتيح Open Science Framework مساحة تحميل لا تزيد عن ٤ جيجابايت.

اجمالاً ينبغي الإشارة إلى أن مستودعات البيانات البحثية العامة– محل الدراسة – تتبع نهج عام ومحايد لتسهيل نشر وإعادة استخدام البيانات؛ حيث لا تضع قيوداً محددة على ملف البيانات وتنسيقات المحتوى المطلوب من المؤلفين ومالكي البيانات، وعلى الرغم من ذلك هناك اتجاه آخر يتعلق بتحديد وتعزيز تنسيقات مجموعات البيانات بحيث يتبنى كل مستودع تنسيقات محددة لنشر وإعادة استخدام البيانات، كذلك تحديد تنسيق خاص لكل مجموعة من التخصصات أو لتخصصات محددة بعينها بما يتناسب مع الطبيعة العملية والدراسية لهذا التخصصات. ومن أمثلة الجهود الجارية في هذا الاتجاه ما قدمها منتدى Open Grid[٤٣] الذي يعمل على تعريف لغة وصف تنسيق البيانات لوصف النصوص، وتنسيقات البيانات الثنائية بطريقة توضيحية وذلك من خلال تقديم Cimphony Orchestra وهو تطبيق أصلي لإدارة بيانات سطح المكتب وتحريرها وتحليلها من Open Grid Systems لبياناتCIM حيث توفر خدمات لإدارة البيانات وتحريرها التحقق من صحة البيانات، وتحويل البيانات القائم على النموذج. كذلك ما تقدمه مجموعة عمل Data Type Registries [٤٤] حيث تناقش مجموعة العمل هذه كيفية ربط تنسيق واضح بمجموعة البيانات؛ وتستهدف فكرة توحيد التنسيق التي تروج لها المجموعة إلى توصيف مجموعات البيانات على مستويات متعددة من الدقة بداية من ملفات البيانات الفردية وأوراق العمل حتى مجموعات البيانات بأكملها، وتقترح المجموعة استخدام ما يعرف باتحاد السجلات لتمكين مستخدمي البيانات من اكتشاف وصف دقيق لتنسيق محدد بالإضافة إلى تحديد معلومات إضافية لإدارة مثل هذا النوع من التنسيقات. توافر اتحاد السجلات هذا للتنسيقات من شأنه أن يسهل على المستودعات توسيع مجموعات التنسيقات بما يمكن المستفيدين من الوصول السريع إلى مجموعات البيانات.

٢-٤-٢- التوثيق (Documentation)

الهدف النهائي لنشر البيانات هو إتاحتها للتحقق منها وإعادة استخدامها داخل التخصصات الأصلية والمجتمع الأوسع، ولتحقيق هذا الهدف بشكل كامل، يجب تزويد مجموعة البيانات التي ينتجها مجتمع الممارسة ببيانات مساعدة توفر معلومات سياقية حول مجموعة البيانات مثل البيانات الوصفية Metadata والتخصص الموضوعي ومصدر هذه البيانات وتوقيت الحصول عليها وغيرها من المعلومات التي تساعد على تحقيق الإثراء الدلالي للبيانات بما يجعلها مفهومة وقابلة للمشاركة وإعادة الاستخدام. [٤٥]

في الوقت الحالي، تهدف البيانات الوصفية المرتبطة بمجموعات البيانات إلى توفير وصف ثابت لمجموعات البيانات، أي متى، ومن، وأين وكيف تم جمعها أو إنتاجها. وقد تم اعتماد طرق مختلفة لتجهيز مجموعات البيانات بعضها من خلال اتاحة وثائق تشتمل على شرح توضيحي لمجموعة البيانات مع البيانات الوصفية المناسبة المحفوظة في مستودع مجموعة البيانات نفسه، أو من خلال نشر ورقة بيانات محددة حول مجموعة البيانات يحتوي على رابط إلى مكان إيداع مجموعة البيانات، إلا أنه من الصعب إنتاج وثائق فعالة وشاملة بسبب سيناريوهات إعادة الاستخدام المفتوحة فمن المستحيل توقع دوافع واحتياجات كل إعادة استخدام لمجموعات البيانات وبالتالي تحديد ما يناسب كل مجتمع علمي. (٤٦)

تطلب جميع المستودعات- محل الدراسة- تحديد البيانات الوصفية الأساسية عند إرسال مجموعة البيانات؛ ولا يطلب أي منها صراحة ربط مجموعات البيانات المودعة بورقة بيانات، عند تحليل الوثائق المتاحة للجمهور، ولم يتم العثور على أي مواصفات للبيانات الوصفية المدعومة داخلياً من قبل المستودعات المحددة. ومن أجل تحديد هذه البيانات الوصفية، تم تحليل البيانات الوصفية المطلوبة لتحميل البيانات على المستودعات والبيانات الوصفية التي تم عرضها لمستخدمي المستودع، وعليه تم تحديد إحدى عشر سمة لهذه البيانات:

١. التوفر (Availability): تمكين الوصول إلى مجموعة البيانات ومحتواها، على سبيل المثال، اضافة معرف الكيان الرقمي DOI أو عنوان المصدر الموحد URI.

٢. البيانات الببليوجرافية الأساسية (Basic bibliographic data): مثل العنوان، المؤلفين، الملخص...الخ.

٣. التغطية (Coverage): وصف امتداد مجموعة البيانات، بما في ذلك التغطية المكانية والزمانية والموضوعية.

٤. التاريخ (Date): تقديم معلومات حول توقيت إنشاء مجموعة البيانات وتقديمها ونشرها، بما في ذلك فترات المنع أو حظر الاستخدام.

٥. التنسيق (Format): تمييز مجموعة البيانات من منظور التنسيق أي تنسيق الملف بما في ذلك الحجم.

٦. الترخيص (License): يصف السياسات التي تحكم إعادة استخدام مجموعة البيانات، بما في ذلك حقوق الوصول والتراخيص .

٧. المرجع (reference) توفير المرجع أو المراجع للمنشورات ذات الصلة، بما في ذلك معرف الكيان الرقمي DOIأو عنوان المصدر الموحد URI.

٨. المشروع (Project): يصف المبادرة المؤدية إلى إنتاج مجموعة البيانات، بما في ذلك أهداف البحث، نوع البحث ومصادر التمويل.

٩. المصدر (Provenance): تحديد المنهجيات التي تؤدي إلى إنتاج مجموعة البيانات، بما في ذلك المصادر والأدوات والبرمجيات الأصلية المستخدمة لإنشاء ملفات البيانات.

١٠. الموضوعات (Subjects): بما في ذلك الكلمات المفتاحية وأكواد التصنيف ورؤوس الموضوعات.

١١. بيانات القياسات البديلة (Altmetric data): تقديم تقارير عن إحصاءات مجموعة البيانات بما في ذلك عدد التحميلات والمشاهدات.

يقدم الجدول التالي (جدول رقم٥) صورة تلخيصية لسمات البيانات الوصفية التي يدعمها كل مستودع. هناك عدم تجانس ملحوظ بين المستودعات محل الدراسة في تطبيق هذه السمات، سواء كانت السمة إلزامية أم لا، على سبيل المثال، تطبق جميع المستودعات – محل الدراسة- معلومات عن المرجع أو المراجع للمنشورات ذات الصلة إلا أن هذه المعلومات إلزامية فقط لـDryad يتطلب Dryad أيضاً سمات محددة جداً حول المقالات المرتبطة بمجموعة البيانات، بما في ذلك العنوان، المؤلفين والمجلات والملخص والكلمات الرئيسية والتغطية؛ بدلاً من ذلك، يقترح Zenodo فقط تقديم المعلومات حول اسم المجلة والمجلد والصفحات. عند تحميل ملفات البيانات، تطلب بعض المستودعات من مؤلفي ومالكي البيانات تحميل أنواع أخرى من المستندات الداعمة .وذلك مثل مستودع Csiro data access portal ومستودع Dryad حيث يتم تشجيع المؤلفين على تقديم توثيق إضافي في شكل ملفات ReadMe للمساعدة في شرح طبيعة وخصائص البيانات وطرق واساليب وتوقيت جمعها بما يضمن مشاركتها وإعادة استخدامها ، كما يوصي Dryad بأن يكون الملف التمهيدي عبارة عن ملف نصي عادي يحتوي على وصفاً موجزاً لمجموعات البيانات المضمنة وذلك لكل ملف من مجموعة البيانات ؛كما يشتمل على تعريفات عناوين الأعمدة، وعلامات الصفوف، وأكواد البيانات، ووحدات القياس؛ كذلك أي خطوات في معالجة البيانات قد تؤثر في تفسير النتائج؛ هذا بالإضافة إلى وصف لمجموعات البيانات المرتبطة المخزنة في مكان آخر- إن وجدت – وأخيراً معلومات الاتصال لمزيد من للأسئلة.

جدول رقم (٥) سمات البيانات الوصفية التي يدعمها كل مستودع من مستودعات الدراسة

سمات البيانات الوصفية	4TU.ResearchData	Csiro data access portal	Science Data Bank	Zenodo	Dryad Digital Repository	Figshare	Open Science Framework
التوفر	✓	✓	✓	✓	✓	✓	✓
البيانات الببليوجرافية الأساسية	✓	✓	✓	✓	✓	✓	✓
التغطية	✓	✓	✓		✓		✓
التاريخ	✓	✓	✓	✓	✓	✓	✓
التنسيق	✓		✓		✓	✓	✓
الترخيص		✓		✓	✓	✓	
المرجع	✓	✓	✓	✓	✓	✓	✓
المشروع		✓		✓		✓	
المصدر	✓			✓	✓	✓	✓
الموضوعات	✓	✓	✓	✓	✓	✓	✓
بيانات القياسات البديلة		✓			✓	✓	

مما لا شك فيه أن استخدام البيانات الوصفية لوصف مجموعات البيانات له أهمية قصوى لتحقيق مشاركة وإعادة استخدام تلك البيانات. ومع ذلك، فإن سمات البيانات الوصفية الحالية التي تدعمها المستودعات العامة- محل الدراسة- محدودة وذلك نظراً لأن هذه المستودعات تهدف إلى خدمة العديد من المجتمعات العلمية غير المحددة. وتشير الدراسات أن هناك مجموعة كبيرة ومتنوعة من خطط البيانات الوصفية قيد الاستخدام كما أن هناك قواسم مشتركة في الخطط الحالية عبر التخصصات وأنواع البيانات،

وهناك حاجة لمزيد من البحث المتعلق بالبيانات الوصفية لتحقيق المشاركة الفعلية وإعادة استخدام بيانات البحث. [٤٧] للتغلب على القيود التي تنتج بالضرورة عن استخدام سمات عامة للبيانات الوصفية، وقد بدأ مؤلفي ومالكي البيانات في ربط ورقة العمل بمجموعة البيانات بما يمكن من تزويد المستفيد بالتفاصيل التي تسهل فهم وإمكانية إعادة استخدام البيانات.

يمكن تحسين عمليات اكتشاف وإعادة استخدام البيانات من خلال توفير أكثر من نمط للبيانات الوصفية. وذلك لنفس مجموعة البيانات؛ بحيث يكون كل نمط من البيانات الوصفية موجه للاحتياجات جمهور مستهدف ومحدد من مستخدمي البيانات. وقد يتم توفير هذه البيانات الوصفية بواسطة مؤلفي أو مالكي مجموعات البيانات أو يتم إنشاؤها تلقائيا باستخدام البرامج المختلفة، وبالمثل يمكن إعداد أوراق إرشادية متعددة يقدم من خلالها عروض تقديمية مركزة تتناول مجموعات البيانات المشار إليها ومعلومات عنها وإرشادات لكيفية استخدامها، وعليه يمكن للمستفيدين من البيانات الاختيار من بين تلك البدائل المتعددة؛ إما بيانات وصفية تفصيلية متعددة الأنماط، أو أوراق إرشادية، أو عروض تقديمية، أو أي مستند آخر يلبي احتياجاتهم على أفضل وجه. [٤٨]

إن إثراء البيانات الوصفية لمجموعة البيانات بروابط إلى الأوراق البحثية ذات الصلة هو مجرد مثال على اتجاه مهم تزايد في السنوات الماضية حيث يتم تقديم الربط بشكل تدريجي لإضافة المزيد من المعلومات السياقية لنتائج البحث، وبالتالي تسهيل فهمها وإعادة استخدامها، وينبغي أن نشير إلى أن التطورات التكنولوجية على وجه الخصوص في تقنيات التفتيش عن البيانات، وتوافر البيانات، وتطوير البنية التحتية للبيانات التي تقدم خدمات متخصصة تسهل اليوم إلى حد كبير الإنشاء التلقائي لمثل هذه الروابط. [٤٩]

٢-٤-٣- الترخيص (Licensing)

الترخيص عبارة عن صك قانوني لمالك مجموعة البيانات ليوضح شروط استخدام مجموعة البيانات المنشورة بين الأطراف المعنية (المؤلف – الناشر – المستفيد) [٥٠] من الصعب تحديد موقف قانوني افتراضي حول كيفية استخدام مجموعة البيانات حيث إنها تنطوي على العديد من العوامل ذات الصلة بسياق مجموعة البيانات وقد تعتمد أيضاً على المؤثرات الخارجية مثل قوانين حقوق التأليف والنشر، إن عملية نشر مجموعات البيانات من خلال المستودعات الرقمية محل الدراسة يخضع لنوعين من التراخيص، الأول متفق عليه بين المستودع ومالك البيانات، والثاني متفق عليه بين المستودع ومستهلكي البيانات. يتم الحصول على هذين الترخيصين جزئياً بواسطة شروط الخدمات أو سياسات المستودع، أي أنها جزء من القواعد التي يجب على مستخدم المستودع الموافقة على قبولها عند استخدام خدمة المستودع . تدعم المستودعات عادةً ترخيص إيداع واحد يحدد الحقوق المتوقع منحها للمستفيدين من قبل مالك مجموعة البيانات، متطلبات حقوق الملكية والاستخدام والتحكم في الاستغلال التجاري لمجموعة البيانات. واحدة من أولى الخطوات التي يُطلب من مالك البيانات الاتفاق عليها عند إيداع مجموعات بيانات في المستودع. فيما يتعلق بنوع التراخيص، يوجد في الوقت الحاضر توجه تدريجي، لا سيما في سياق الأبحاث الممولة من القطاع الحكومي والأكاديمي، للانتقال من استخدام تراخيص الملكية لصالح تراخيص المشاع الإبداعي. [٥١]

بالنسبة لتراخيص الإيداع، تمارس المستودعات مجموعة متنوعة من السياسات حيث يطلب من مالكي البيانات التسجيل في المستودع وبمجرد إرسال أو تحميل مجموعات البيانات فإنهم بهذا يوافقون ضمنيا على سياسات المستودع. هذه السياسات غير متجانسة تماما بين المستودعات محل الدراسة.

- **4TU.ResearchData**: يمنح مالكي ومؤلفي البيانات ترخيصاً غير حصري للمستودع تمكنه من تخزين مجموعة البيانات وإتاحتها للمستخدمين.
- **Csiro data access portal**: يمنح مالكي البيانات المستودع الحق في إتاحة مجموعات البيانات بشكل غير حصري مع إرفاق مجموعة البيانات بشروط الاستخدام في وقت التنزيل.
- **Science Data Bank** : يوفر المستودع ١١ اتفاقية ترخيص بيانات (٨ اتفاقيات ترخيص عامة، و٣ اتفاقيات ترخيص برمجيات) ويعطي الحق لمالكي البيانات في اختيار الترخيص الملائم.
- **Zenodo**: يجب على مالكي البيانات تحديد الترخيص الذي يحكم محتوى مجموعة البيانات من خلال الاختيار من بين مجموعة واسعة و متنوعة من التراخيص التي يتيحها المستودع نفسه.
- **Dryad Digital Repository**: يمنح مالكي البيانات إذن للمستودع لإتاحة مجموعة البيانات للجمهور بموجب تنازل CC0 (على الرغم من السماح بشروط الترخيص الأخرى)
- **Figshare**: يوافق مالكي البيانات على أن أي مواد مخزنة للجمهور بما في ذلك مجموعات البيانات متاح مع تراخيص المشاع الإبداعيCC-BY .
- **Open Science Framework**: لا يقدم المستودع اقتراحات بشأن أفضل ترخيص للاستخدام، بل يعرض مجموعة من التراخيص مع شروح مبسطة لها ويمنح الحرية لمالكي البيانات في اختيار الترخيص.

جميع المستودعات محل الدراسة تقوم بتحديد فترة حظر؛ أي تحديد فترة زمنية يتم فيها تقييد الوصول إلى مجموعة البيانات، وبالإضافة إلى فترات الحظر بعض المستودعات مثل Zenodo، وFigshare تسمح بتحميل مجموعات البيانات لتظل خاصة privet (يمكن اعتبار هذا الاستخدام للمستودع نوعاً من مرحلة ما قبل النشر حيث تتاح فقط لمالكي البيانات أنفسهم دون اتاحتها لسائر المستفيدين) في جميع الحالات لا يؤدي تحميل مجموعات البيانات إلى حدوث تغييرات في الملكية.

الشروط العامة للاستخدام التي وضعها 4TU.ResearchData تشبه رخص المشاع الإبداعي غير التجارية (CC BY-NC)، والتي تنص على أن أي مستخدم يحق له إعادة استخدام مجموعة بيانات شرط أن تستخدم لأغراض تجارية، وبالمثل، يمنح ترخيص بيانات Csiro data access portal المستخدم ترخيصاً خالياً من حقوق الملكية وغير قابل للتحويل لاستخدام مجموعة البيانات للأغراض غير التجارية فقط، وفي حالة Dryad، و Figshare ، و Zenodo الخيار الافتراضي هو CC0 ومع ذلك، يستخدم Figshare ترخيص CC-BY لمجموعات البيانات المخزنة كمجموعات ملفات بينما يتيح Open Science Framework ،و Science Data Bank،و Zenodo لمالكي البيانات ربط الترخيص الذي يفضلونه عن طريق الاختيار من قائمة يتيحها المستودع ، ومن هذا التحليل، يتبين أن عدد التراخيص المستخدمة بالفعل كبير علاوة على ذلك، فإن استخدام تراخيص المشاع الإبداعي منتشر للغاية.

تؤثر تراخيص استخدام مجموعات البيانات على جوانب مختلفة من عمليات مشاركة وإعادة استخدام تلك المجموعات تتراوح من الاستشهاد إلى الاستغلال التجاري. يجب أن تعزز المستودعات الدعم الذي يقدمه مالك البيانات في وقت الإيداع من خلال توضيح تأثير التراخيص المتنوعة على استخدامات مجموعة البيانات، على سبيل المثال، يجب الإشارة بوضوح إلى الاختلاف الناجم عن ترخيص CC0 فيما يتعلق بترخيص CC-BY عند الاستشهاد. حيث يجب أن توفر واجهة استخدام لمقارنة التراخيص، إن مناهج المستودعات العامة- محل الدراسة- محدودة النطاق فيما يتعلق بترخيص مجموعات البيانات وهذا يجعل من الصعب إعادة الاستخدام الفعال للبيانات، إن قضية تحديد تراخيص الاستخدام لا تزال تتطلب

الكثير من العمل لتحديد الحلول الواضحة والقياسية بما يضمن تجنب مشاكل التشغيل البيني. حيث يجب على المستودعات أن توفر مجموعة واسعة من خيارات الاستخدام والتحكم، فلا توجد حلول قياسية واحدة تناسب الجميع. (٥٢)

٢-٤-٤- تكاليف النشر(Publication costs)

التكلفة هي أحد العوامل الرئيسية التي تعيق مشاركة واعادة نشر واستخدام البيانات، وتتضمن الجهد المطلوب لإعداد مجموعات البيانات كتنسيقها وتوثيقها بطريقة تمكن الآخرين من الاستفادة منها، والوقت اللازم للتعرف على الخدمة واستخدامها، هذا بالإضافة إلى التكلفة المالية لأرشفة مجموعة البيانات في مستودع موثوق به يوفر أيضاً إمكانية الوصول إلى محتوى مجموعة البيانات. تركز الدراسة الحالية على التكلفة النقدية لنشر مجموعات البيانات في المستودعات - إن وجدت- إن تقديم المستودعات لخدماتها لهو أمر مكلف، وعليه تتجه جميع المستودعات محل الدراسة إلى فرض رسوم على مالكي وناشري البيانات، بدلا من فرض تلك الرسوم على مستخدمي أو مستهلكي البيانات.

يشبه مستودع Dryad إلى حد كبير نموذج الوصول المفتوح المعتمد من قبل العديد من الدوريات حيث يتيح مجموعات البيانات علنا لجمهور المستخدمين، ولكن يلزم دائماً دفع رسوم في وقت تقديم البيانات، وتحديداً يطلب المستودع من المالك دفع رسوم نشر البيانات، ما لم يكن مالك البيانات ومقدم الطلب في بلد مصنف على أنه اقتصاد منخفض أو متوسط الدخل من قبل البنك الدولي، أو أن الدورية التي تنشر مقالة البحث المرتبطة بمجموعة البيانات لديها تعاقد بالفعل مع Dryad، ويقدم Dryad سلسلة من الخطط التي توفر خصماً على الحجم وذلك مثل: ١- خطة القسيمة a voucher plan حيث تقدم الدفع مقابل نشر عدد من حزم البيانات، و٢- خطط الاشتراك subscription plans، على سبيل المثال ، دفع رسوم سنوية؛ حيث تبلغ تكلفة حزمة بيانات واحدة يصل حجمها إلى ١٠ جيجابايت نحو ٨٠ دولاراً، ولحزم البيانات التي يتجاوز حجمها ١٠ جيجا بايت يتم تحصيل ١٥ دولاراً أمريكياً لأول جيجابايت و١٠ دولارات أمريكية لكل جيجابايت إضافية أو جزء منها. علاوة على ذلك، في الدوريات التي لا تستخدم خدمة تقديم البيانات المتكاملة التي يقدمها المستودع، يتعين على مالكي البيانات دفع مبلغ رسوم إضافية بقيمة ١٠ دولارات في وقت التقديم لتغطية تكاليف التنظيم الإضافية. يعد Dryad هو المستودع الوحيد في مستودعات الدراسة- الذي يفرض على مقدمي ومالكي البيانات دفع رسوم بشكل مستقل عن حجم الملفات. كل المستودعات تقدم مساحة تخزين محددة حيث يمكن للمستخدمين النشر مجاناً.

يقدم Figshare خطة مجانية تماماً، تتيح للمستخدمين تخزين بيانات تصل إلى ١ جيجابايت مجاناً، يشمل الدفع فقط التخزين الخاص، حيث إن مساحة نشر البيانات العامة دائماً ما تكون غير محدودة لكل خطة، وتتطلب الخطط الأخرى من المستخدمين دفع مبلغ شهري اعتماداً على حجم مساحة التخزين الخاصة وحجم الملفات، على سبيل المثال، رسوم ٨ دولارات شهرياً تضمن مساحة خاصة تبلغ ١٠ جيجا بايت.

تقدم Zenodo خدمتها مجاناً، ولكنها تفرض حالياً قيوداً على الحجم تبلغ ٢ جيجا بايت لكل ملف ومع ذلك، لا يستبعد المستودع مجموعات البيانات الأكبر حجماً، لذا فهو يخطط لوضع سقف للمساحة التي يمكن تقديمها بلا مقابل، وإدخال خطط دفع لحجم أكبر من البيانات.

أما 4TU.ResearchData لا يتطلب دفع أي مقابل مجموعات البيانات التي تصل إلى ٤ جيجابايت. بالنسبة لمجموعات البيانات الأكبر حجماً يجب على المستخدمين الاتصال بموظفي المستودع لترتيب تحميل مخصص.

في حين لا يتطلب مستودعCsiro data access portal دفع أي تكاليف مادية مقابل مجموعات البيانات والتي تصل إلى نحو عشرة جيجا بايت.

لا يفرض مستودع Science Data Bank رسوماً على المودعين مقابل نشر البيانات في الوقت الحالي، فقط في حالة تجاوز حجم البيانات ٥٠٠ ميجابايت يتم التواصل مع الموظفين لترتيب تحميل مخصص للبيانات.

يقدم مستودع Open Science Framework مجموعة من الخطط لتحميل وإتاحة مجموعات البيانات، GitHub Free هي الخطة الأساسية التي تم إنشاؤها للأفراد والفرق الصغيرة ؛ وتسمح تلك الخطة بتحميل ٥٠٠ ميجابايت مجانا، أما خطة GitHub Pro تتيح استخدام وظائف إضافية للتعاون المتقدم في حسابات المستخدمين الفردية، وتتيح التعامل مع ٢ جيجا بايت في مقابل ٤٤ دولار، وأخيراً يقدم GitHub Enterprise وظائف إضافية للتعاون المتقدم عبر المستودعات للفرق المتزايدة بما يسمح بتحميل ٥٠ جيجا بايت في مقابل ٢٣١ دولار.

التكاليف الإضافية والحوافز الهادفة إلى زيادة حجم البيانات المنشورة لها تأثير محدود إذا لم تكن مصحوبة بتدابير تتناول الحد من تكاليف النشر، حتى لا يتم تثبيط الباحثين عن نشر مجموعات البيانات الخاصة بهم؛ ويمكن تحقيق ذلك من خلال ربط التكاليف صراحةً بالخدمات المقدمة وجودتها وامتداد هذه الخدمات لتغطي عدد أكبر من السنوات، إن قصر التكلفة على نشر المجموعات فقط من شأنه أن يسبب قصور شديد في تقديم وتطوير سائر الخدمات الأخرى.

من الواضح أن جودة هذه الخدمات تؤثر بشكل كبير على تكاليف النشر، المستودعات محل الدراسة لا تحدد بالتفصيل جودة الخدمة المقدمة لأولئك الذين يقررون نشر مجموعات البيانات الخاصة بهم، وعليه فمن الصعب فهم الدوافع الكامنة وراء التكاليف، ومن شأن الخدمات أيضاً تسهيل التوزيع العادل للتكاليف بين الجهات الفاعلة المشاركة في عملية نشر البيانات حيث لا يجب بالضرورة تغطية تكلفة خدمات المستودع بالكامل بواسطة الباحثين ومالكي البيانات، وينبغي في هذا الصدد أن نشير إلى الاستراتيجيات والجهود التي يتم تنفيذها حالياً لتعزيز الوصول المفتوح إلى بيانات البحث والاستراتيجيات والجهود التي يتطلبها الاعتراف العام بالفوائد الناتجة عن ذلك ؛ حيث يمكن اقامة شراكات مع كل من القطاع الخاص أو العام والدوريات العلمية. على سبيل المثال، بدأت Figshare بالفعل شراكات مع بعض الناشرين لدعم المؤلفين الذين يرغبون في مشاركة مجموعات البيانات والمواد التكميلية لمقالاتهم بشكل علني [٥٣]، الاتجاهات الحالية لتقليل التكلفة التشغيلية للمستودع مكرسة بشكل أساسي للاعتماد على الطرف الثالث – الناشرين والموردين .... الخ، على سبيل المثال، الحصول على الموارد اللازمة من مقدمي الخدمات السحابية بدلاً من بذل الجهد في تشغيل هذه الموارد بواسطة المستودعات، وهناك أيضاً العديد من المحاولات لاستغلال التقنيات التي تقلل من تكاليف المعالجة عن طريق الاستخراج التلقائي لجزء من المعلومات الضرورية من خلال تحليل المصادر ذات الصلة. [٥٤]

٢-٤-٥-التحقق (Validation)

يعد التحقق من صحة مجموعة البيانات مرحلة أساسية في عملية نشر البيانات حيث يساهم في تقييم جودة مجموعة البيانات .ونعني بالتحقق أي عملية تهدف إلى تقييم صحة أو سلامة البيانات المنشورة .

تقوم مستودعات البيانات – محل الدراسة- بالتحقق من صحة مجموعات البيانات من خلال مجموعة من الممارسات والخدمات لمرحلتي ما قبل النشر وما بعد النشر؛ مرحلة ما قبل النشر؛ هي بمثابة تقييم وقائي يهدف إلى تجنب نشر مجموعات بيانات ضعيفة أو ذات جودة رديئة ويعد ذلك نوعا من ضمان الجودة الذي يشمل كل من مالك البيانات، والبيانات نفسها، والمستودع الذي يقوم بدور الناشر. أما مرحلة

ما بعد النشر؛ فهي بمثابة التحقق من ردود الفعل الواردة من المستخدمين الذين يحاولون بالفعل إعادة استخدام مجموعة البيانات المنشورة، سواء كانت تلك الردود إيجابية أو سلبية.

تدعم المستودعات -محل الدراسة- مرحلة ما قبل النشر باستخدام التقنيات التي تهدف إلى التحقق من أن مجموعات البيانات ليست تالفة وأن البيانات الوصفية المرتبطة بها صحيحة ومكتملة من الناحية التركيبية، إلا أنه لا توجد رقابة لإثبات السلامة العلمية لمجموعة البيانات مثل الصلاحية العلمية، أو الدقة أو الاكتمال.

تقوم جميع المستودعات – محل الدراسة- باستثناءFigshare ، بإجراء عملية التحقق من صحة كل من مجموعات البيانات والبيانات الوصفية لها. ومع ذلك، فإن Dryad فقط هو الذي يعطي معلومات دقيقة حول هذه العملية في شروط الخدمة Terms of Service؛ حيث يقوم موظفي Dryadبسلسلة من الفحوصات التي تتراوح بين الفحوصات الفنية مثل امكانية فتح الملفات، أو عدم تلفها، أو عدم احتوائها على فيروسات كذلك مجموعة من الفحوصات الإدارية، على سبيل المثال، البيانات الوصفية صحيحة من الناحية الفنية، بيانات الورقة ذات الصلة موجودة، كما قد يقوم Dryad بمراجعة المحتوى بما في ذلك وجود معلومات غير ملائمة وبيانات حقوق النشر غير المتوافقة مع CC0 . في جميع الأحوال لا يتحقق Dryadمن مجموعة البيانات من منظور علمي أو يعدل المحتوى. وإلى جانب ضوابط ما قبل النشر، تقوم جميع المستودعات – محل الدراسة- بإجراء عمليات تحقق تهدف إلى ضمان بقاء محتوى مجموعات البيانات كما هوحيث يتم تخزين مجموعات البيانات مع اختبار ومراجعة لمحتواها. على سبيل المثال، يتم تخزين ملفات البيانات المرسلة باستخدام MD5 checksum [٥٥] للتحقق من سلامة التثبيت والتنزيل.

للتحقق من صحة البيانات في مرحلة ما بعد النشر، تميل المستودعات- محل الدراسة- إلى تقديم أهم الإحصاءات المتعلقة بتنزيل مجموعات البيانات أو الاستخدام حيث يوفر Figshare إمكانية معرفة عدد المرات التي تمت فيها مشاركة أو عرض كل مجموعة بيانات من خلال خيار البحث أو التصفح، حيث يقدم المستودع قائمة مرتبة بجميع مجموعات البيانات ويمكن فرز تلك القائمة حسب الأكثر مشاهدة أو الأكثر مشاركة ، كما يسجل Dryad عدد التنزيلات و يجعل من الممكن تصفح التنزيلات الأكثر شعبية ، وينشر 4TU.ResearchData الإحصائيات المجمعة عن مجموعات البيانات التي تم تنزيلها كنوع من التحقق من صحة المستودع نفسه.

يعد التحقق من صحة مجموعة البيانات عملية معقدة لا تزال بعيدة عن التوصيف الكامل، وهناك العديد من الأسباب لهذا التعقيد يمكن ادراجها فيما يلي:

- أنها مسؤولية مشتركة؛ فلا يوجد فاعل واحد في سيناريو نشر البيانات العلمية يمكنه تحمل مسؤولية تقييم صلاحية مجموعة البيانات.
- أنها عملية مستمرة؛ حيث لا يقتصر استخدام مجموعات البيانات المنشورة على مجتمع معين أو حالة استخدام فقط، وقد يتطلب أي استخدام لمجموعة بيانات منشورة إجراء تقييم للتحقق من صحتها.
- أنها قضية متعددة الجوانب؛ تتعلق بالجوانب الفنية والعلمية والتنظيمية حيث تتطلب جميعها خبرة متنوعة في المجال.
- لا يتعلق الأمر بشهادة المستودع؛ إذ يكاد يكون من المستحيل تصور نظام اعتماد يضمن نشر مجموعات بيانات سليمة علمياً بواسطة المستودع.

٢-٤-٦- التوافر (Availability)

توافر مجموعة البيانات هي الميزة التي تهدف إلى ضمان أن تكون مجموعات البيانات المنشورة تحت تصرف المستفيد، وعادة ما ترتبط هذه الميزة بمصداقية المستودع، إنها واحدة من الميزات الرئيسة التي تستخدمها مستودعات البيانات البحثية لدعم إعدادات نشر مجموعات البيانات حيث تقدم المستودعات آليات للتوافر الحالي؛ أي مجموعات البيانات المتاحة حاليا للوصول، كذلك التوافر المستقبلي؛ أي مجموعات البيانات المتاحة بمرور الوقت. تنفذ المستودعات- محل الدراسة- نهجين رئيسين لضمان التوافر الحالي والمستقبلي، أي أرشفة البيانات بطريقة آمنة وتطبيق آلية لحفظها.

للأرشفة الآمنة للبيانات، جميع المستودعات – محل الدراسة- تخزن نسخ متعددة من مجموعات البيانات إما في قواعد بيانات خاصة بها أو من خلال مزودي الخدمة فيما يعرف بالطرف الثالث ، على سبيل المثال يستخدم Figshare روابط Amazon Facili ، وتستخدم Zenodo مركز بيانات CERN، كما يُدعم Open Science Framework من قبل Center for Open Science (COS) بالإضافة إلى ذلك، دخل كل من Dryad و Figshare في شراكة مع منظمة CLOCKSS ، وهي شبكة موزعة جغرافياً وسياسياً مكونة من ١٢ أرشيف لـ ١٢ مكتبة بحثية رئيسية حول العالم .كما يحتفظ Dryad بنسخة احتياطية لجميع المحتويات المنشورة به في أرشيفCLOCKSS ، لذلك إذا لم يعد بإمكانه الحفاظ على ملف الخدمات النشطة، سيتم تحديث جميع المعرفات DOIs المسجلة في Dryad للوصول إلى النسخة الموجودة في أرشيفCLOCKSS ، والتي ستستمر في توفير الوصول إلى المحتوى بموجب نفس شروط الترخيص ويتم تنفيذ إجراء مشابه بواسطةFigshare .

لحفظ البيانات، فبجانب تخزينها في نسخ متعددة، تميل المستودعات محل الدراسة إلى استخدام التنسيقات الخاصة بتهجير البياناتData migration، ويعد تنسيق التهجير ميزة صعبة يجب ضمانها، حيث يمكن إنشاء إصدار جديد من ملف بتنسيق التهجير وإضافته إلى مجموعة البيانات الأصلية متى تم إنشاء ملف المستودع وهذا من شأنه أن يسهل الحفظ. وقد لا تحتوي الملفات التي تم تهجيرها على جميع المعلومات المتوفرة بتنسيق الملف الأصلي، لكن المستودع يحاول تقليل فقدان المعلومات الناجم عن تهجير الملف؛ وعلى أي حال، لا يتم تعديل محتوى المعلومات الخاص بالملف الأصلي أبداً، ويؤكد 4TU.ResearchData بشدة على أهمية اتخاذ التدابير المناسبة لضمان إمكانية الوصول إلى البيانات في المستقبل، وتنص اتفاقية الترخيص بين 4TU.ResearchData ومقدمي ومالكي البيانات صراحة على أن المستودع يضمن بأفضل ما لديه من قدرات وموارد أن مجموعة البيانات المودعة ستظل مقروءة ويمكن الوصول إليها؛ كما يجب الحفاظ على مجموعة البيانات دون تغيير في شكلها الأصلي، مع مراعاة التقنيات الحالية وتكاليف التنفيذ؛ وأن للمستودع الحق في تعديل شكل مجموعة البيانات إذا كان ذلك ضرورياً من أجل تسهيل الاستدامة الرقمية أو التوزيع أو إعادة استخدام مجموعة البيانات. كما يوفر المستودع للمستخدمين جدولاً يتضمن تنسيقات الملفات المثلى المستخدمة على المدى الطويل للحفاظ على البيانات، بالإضافة إلى أن عدد التنسيقات المدعومة محدود من أجل تسهيل التحويل في المستقبل إلى تنسيقات أخرى، وإن كان يتم تحديث الجدول بانتظام بتنسيقات جديدة.

يمكن تقسيم مستويات الدعم للحفظ على المدى الطويل إلى ثلاثة مستويات رئيسة: (٥٦)

- **المستوى الأول:** يتم اتخاذ جميع الإجراءات اللازمة لتسهيل الاستخدام؛ قد تتضمن الإجراءات تنسيق التهجير أو التطبيع أو التحويل.
- **المستوى الثاني:** يتم اتخاذ خطوات محدودة للحفاظ على سهولة الاستخدام؛ يمكن تحويل تنسيقات الملفات بشكل فعال من تنسيق واحد إلى آخر للتخفيف من تقادم التنسيق.

- **المستوى الثالث**: يضمن الوصول إلى البيانات في ملف التنسيق الأصلي.

لضمان التوفر الحالي والمستقبلي لمجموعات البيانات المنشورة يجب على المستودعات معالجة كل من التحديات الفنية والمالية. من المنظور التقني لا يعد الحفاظ على المعلومات الرقمية على المدى الطويل مصدر قلق، ومع ذلك فإن الحفاظ على مجموعة البيانات في سياق المستودعات العامة لديه تحديات محددة ناتجة بشكل رئيس عن أنماط مجموعات البيانات شبه المفتوحة التي يجب إدارتها، ومواجهة هذه التحديات له بالضرورة تأثير مالي على المستودعات، فعلى سبيل المثال إذا تم إنشاء مجموعات للبيانات تتطلب توفر برامج نادرة أو خاصة، ومن ثم قد يكون تنفيذ آليات نسخ وتهجير البيانات مكلفاً للغاية، وبالتالي، فإن اعتماد سياسة توافر محددة هو بالضرورة الحل المثالي أو القابل للتنفيذ المتوافق مع الموارد المتاحة بما في ذلك الموارد المالية للمستودع.[٥٧] يمكن أن تقرر المستودعات بعد ذلك اتخاذ خيارات مختلفة وتقديم مستويات مختلفة من التوفر. هذا المستوى قد يتغير خلال عمر المستودع، مع الأسف سياسة التوفر لا توصف صراحة من قبل المستودع، وهذا يجعل من الصعب على المستخدمين - سواء كانوا مالكي ومؤلفي مجموعات البيانات أو المستفيدين- اختيار السياسة التي تناسب احتياجاتهم على أفضل وجه .

٢-٤-٧- قابلية الاكتشاف والوصول (Discovery and Access)

إن اكتشاف مجموعات البيانات والوصول إليها هو الوسيلة التي تمكن المستفيدين من إدراك وجود مجموعات بيانات معينة والقدرة على الوصول إلى محتواها، وما يرتبط بها من توثيق، إنها السمة والميزة الرئيسة التي من أجلها انشأت مستودعات البيانات البحثية، حيث تضمن هذه الميزة الوصول إلى البيانات العلمية وتقديمها للمستفيدين في سياق عمليات النشر العلمي، قد تتضمن هذه الميزة وظائف يقوم بها المستخدم user-driven functions، مثل البحث والتصفح والنقر للتنزيل، بالإضافة إلى الوظائف شبه الآلية semi-automatic functions، مثل الإشعارات والتوصيات .

تقدم المستودعات – محل الدراسة مجموعة من التسهيلات اللازمة لاكتشاف مجموعات البيانات التي توفرها المستودعات؛ حيث تقدم جميعها الأساليب المعروفة التالية لاكتشاف مجموعات البيانات:

- البحث البسيط المباشر القائم على الكلمات المفتاحية حيث يمكن للمستخدمين تحديد المعلومات التي يحتاجونها من خلال مجموعة من الكلمات الرئيسية.
- البحث بالحقول حيث يمكن للمستخدمين تحديد احتياجاتهم من المعلومات من خلال مجموعة من الحقول القائمة على معايير تصفية.
- يتم توفير استعراض للمستخدمين مع قائمة بمجموعات البيانات لتصفحها .يمكن أن تكون هذه القائمة إما القائمة الكاملة لـ مجموعات البيانات التي ينشرها المستودع أو قائمة مجموعات البيانات الناتجة عن معيار اختيار معين، على سبيل المثال، الكلمات الأساسية، والموضوع، والنوع، والشكل، والسنة، والمنشئ .

يجب التمييز بين تلك المستودعات التي تقدم التصفح من خلال نظام تصنيف وأولئك الذين لا يفعلون ذلك، أولهما CSIROوFigshare، حيث لا يسمحان للمستخدمين فقط بتصفح مجموعات البيانات المصنفة ضمن مجال البحث الذي يهتمون به، ولكن يمنحاهم أيضاً القدرة على عرض تفاصيل وصف مجموعات البيانات والقدرة على التعبير عن المزيد من عمليات البحث عن الكلمات الرئيسية بناءً على المصطلحات لضمان نتائج إيجابية .

يقدم CSIRO أيضاًعمليتي بحث محددتين: (أ) البحث حسب الموقع، أي يمكن للمستخدمين تحديد احتياجاتهم المعلوماتية باستخدام خريطة للإشارة إلى مجال اهتمامهم بالإضافة إلى الكلمات الرئيسية

لتضييق نطاق البحث. (ب) بحث خاص بالمجموعة، على سبيل المثال، يتم تزويد المستخدمين بنماذج تمكنهم من تحديد احتياجاتهم المعلوماتية الدقيقة حول مجموعات بيانات محددة.

تسمح 4TU.ResearchData بالتصفح حسب الموقع الجغرافي لتحديد مجموعات البيانات التي يكون تحديد الموقع الجغرافي ضرورياً للوصول اليها.

توفر جميع المستودعات–محل الدراسة- إمكانية تنزيل مجموعات البيانات ككل، بالإضافة إلى تنزيل ملفات فردية واحداً تلو الآخر من خلال الروابط المناسبة المعروضة على صفحة الويب الخاصة بمجموعة البيانات. على سبيل المثال، يعرض Open Science Framework كل مجموعة بيانات كهيكل ملف، مع مربعات اختيار لاختيار ملفات معينة لتنزيلها.

بصرف النظر عن تسهيلات التنزيل الرئيسة، توفر المستودعات طرائق بديلة ومخصصة للوصول إلى مجموعات البيانات، على سبيل المثال، يوفر CSIRO الوصول إلى مجموعات البيانات عبر بروتوكولات مثل WebDAV و SFTPكما يسمح للمستخدمين المسجلين بتنزيل البيانات على أجهزة محددة والوصول إلى البيانات مباشرة على هذه الأجهزة، وفي بعض الحالات، يتم السماح للمستخدمين بمعاينة مجموعة البيانات، على سبيل المثال، يتيح CSIRO تصفح الصور الموجودة في مجموعة البيانات، ويقدم Figshare معاينة لمحتوى مجموعة البيانات في المتصفح.

إلى جانب هذه التسهيلات الموجهة نحو البشر، تدعم المستودعات الوصول البرمجي إلى محتواها، ويتم تحقيق ذلك من خلال دعم البروتوكولات القياسية، مثل OAI-PMH [٥٨]، بالإضافة إلى واجهات برمجة APIsالقائمة على الويب .هذه التسهيلات هي في صميم البنى التحتية التجميعية، ويعتمد OpenAIREعلى خدمة OAI-PMH لجمع مجموعات البيانات المنشورة بواسطة Zenodo.

تقوم مستودعات البيانات – محل الدراسة- بتقديم تسهيلات اكتشاف البيانات على غرار التسهيلات التقليدية التي تقدمها مستودعات المقالات منذ سنوات عديدة، إلا أن المعايير التي يرغب الباحثون في صياغتها عند البحث عن مجموعات البيانات، غالباً ما تكون مختلفة تماماً عن تلك التي يستخدمونها في البحث عن المقالات العلمية، حيث يتم البحث في مجموعات البيانات بشكل أساسي لإجراء التحقق وإعادة الاستخدام في الأنشطة العلمية .لدعم هذه الأنشطة قد يتطلب الأمر معايير لاكتشاف إمكانية تحليل مجموعات البيانات باستخدام أدوات معينة تم إعادة استخدامها في سياقات بحث مشابهة لتلك الخاصة بمجموعة بيانات معينة، أو حتى التي تمت الموافقة عليها بواسطة زملاء العمل أو الممارسين، وقد ترتبط معايير أخرى بمحتويات مجموعة البيانات، على سبيل المثال، قد يرغب المستخدمين في اكتشاف مجموعة بيانات ذات صلة بمنطقة جغرافية معينة، ربما يكون المؤشر الوحيد الذي يتعين عليهم اكتشاف مجموعة البيانات فيه هو عنوان الورقة التي تحتوي على النتائج التي تم الحصول عليها من خلال استخدام مجموعة البيانات.

قابلية الاكتشاف والوصول إلى مجموعات البيانات عبارة عن خدمة يمكن أيضاً تقديمها بواسطة محركات البحث، والبروتوكولات وواجهات برمجة التطبيقات لتجميع محتوى البيانات الوصفية metadata. تميل المستودعات إلى دعم OAI-PMH ومع ذلك يجب عليهم تعزيز نطاق البروتوكولات والتسهيلات التي يقدمونها برمجياً للوصول إلى المحتوى الخاص بهم، على سبيل المثال، عن طريق عرض محتواهم بتنسيقات أخرى غير HTML مثلSchema.org والبيانات المرتبطة، من خلال دعم البروتوكولات القياسية مثل OpenSearch وSRU . وتعتمد الخدمات التي ينفذها موفرو الطرف الثالث إلى حد كبير على البيانات الوصفية metadata، وبالتالي يمكننا القول إيجازاً أن جودة البيانات الوصفية يؤثر بشدة أيضاً على تنفيذ هذه الميزة.

٢-٤-٨- الاستشهاد المرجعي والاقتباس (Citation)

الاستشهاد المرجعي والاقتباس هو ممارسة تقديم مرجع لمجموعة أو مجموعات البيانات، والمقصود بها وصف خصائص البيانات بما يمكن من الاستشهاد، واكتشاف، وربط، والوصول إلى مجموعات البيانات، إنها آلية رئيسة في نشر البيانات البحثية لأنها تمكن مالكي البيانات من الحصول على القبول المناسب لنشر مجموعات البيانات الخاصة بهم، وتمكن مستهلكي البيانات للإشارة صراحة إلى مجموعات البيانات التي يتم إعادة استخدامها في أبحاثهم.

مستودعات البيانات البحثية -محل الدراسة- تدعم كل من المالكين والمستهلكين، حيث تقدم المستودعات خيارات متنوعة لمعالجة اقتباس البيانات.

سلسلة اقتباس تتيح للمستخدمين الحصول على بيان استشهاد يمكن نسخه ولصقه ببساطة في وثائقهم، وتحتوي سلسلة الاقتباس على تنسيق عام مثل اسلوب DataCite والذي يتكون من المؤلفين وسنة النشر والعنوان والناشر واسم المستودع ومعرف الكيان الرقمي DOI.

- **خيار التصدير** يسمح للمستخدمين بتصدير الاقتباس مباشرة إلى مجموعة البيانات في مجموعة متنوعة من الأنماط العامة، أكثرها شيوعا هو نمط RIS المتوافق مع برامج إدارة المراجع مثل EndNoteو Reference Managerو ProCiteوRefWorks، ونمط BibTexالمتوافق مع برامج مثل LaTeX وBibDesk، بالإضافة إلى DataCiteو Dublin Core و NLM و MARCXML.
- **خيار التضمين** يسمح للمستخدمين بتضمين ارتباط بالبيانات في كود مصدر HTML الخاص بمواقعهم الشخصية وصفحاتهم الرقمية.
- **خيار المشاركة** يسمح للمستخدمين بمشاركة رابط إلى مجموعة البيانات مباشرة عبر البريد الإلكتروني، أو من خلال مجموعة متنوعة من الشبكات الاجتماعية، مثل Mendeley، وFacebook، وTwitter، وResearchGate.

تجدر الإشارة إلى أن كل هذه الأساليب تعتمد على معرف الكيان الرقمي DOI المخصص لكل مجموعة بيانات في وقت النشر.

يخضع الاستشهاد بالبيانات حاليًا لنشاط بحثي مكثف، حيث ناقش فريق عمل Data Citation Synthesis Working Group معايير وممارسات الاستشهاد بالبيانات وأصدر ثمانية مبادئ حول الاقتباس المناسب للبيانات [٥٩] تغطي هذه المبادئ الغرض والوظيفة والسمات الخاصة باستشهادات البيانات حيث يمكن تلخيصها فيما يلي:

١. **الأهمية**: يجب اعتبار البيانات منتجات بحثية مشروعة وقابلة للاقتباس، حيث يجب إعطاء الاستشهادات بالبيانات نفس الأهمية في السجل العلمي مثل الاستشهادات من مواد بحثية أخرى، كالمنشورات.

٢. **الائتمان والاسناد**: يجب أن تسهل الاستشهادات بالبيانات منح المصداقية العلمية والإسناد المعياري والقانوني لجميع المساهمين في البيانات، مع الاعتراف بأن نمطاً واحداً أو آلية واحدة للاستشهاد قد لا تكون قابلة للتطبيق على جميع البيانات.

٣. **الدليل**: في الأدبيات العلمية متى وأينما تطلب الامر الاعتماد على البيانات، يجب الاستشهاد بالبيانات المناظرة.

٤. **تعريف فريد**: يجب أن يتضمن الاقتباس من البيانات طريقة ثابتة للتعريف تكون قابلة للتنفيذ آلياً، وفريدة من نوعها عالمياً ومستخدمة على نطاق واسع من قبل المجتمع العلمي.

٥. **الوصول**: يجب أن تسهل الاستشهادات بالبيانات الوصول إلى البيانات نفسها وإلى البيانات الوصفية والوثائق والرموز وغيرها من المواد ذات الصلة، حسب الضرورة لكل من البشر والآلات للاستفادة المستنيرة من البيانات المرجعية.

٦. **الثبات**: يجب أن تستمر المعرفات الفريدة DOIsوالبيانات الوصفية Metadataالتي تصف البيانات وترتيبها حتى بعد انقضاء عمر البيانات الموصوفة.

٧. **الخصوصية وقابلية التحقق**: يجب أن تسهل الاستشهادات بالبيانات تحديد البيانات اللازمة لدعم المطالبة والوصول إليها والتحقق منها. كما يجب أن تتضمن الاقتباسات أو البيانات الوصفية للاقتباسات معلومات حول المصدر، وأن تكون كافية لتسهيل التحقق من أن الشريحة الزمنية المحددة والإصدار من البيانات المسترجعة لاحقاً هو نفسه الذي تم الاستشهاد به في الأصل.

٨. **قابلية التشغيل البيني والمرونة**: يجب أن تكون طرق الاستشهاد بالبيانات مرنة بما يكفي لاستيعاب الممارسات المتغيرة بين المجتمعات العلمية.

يمكن التأكيد على أن مستودعات البيانات – محل الدراسة- تلبي تلك المبادئ إلى حد ما، ولكن لا يزال هناك الكثير مما يتعين القيام به لتلبية المتطلبات التفصيلية التي يفرضها العلماء عند نشر مجموعات البيانات أو استخدامها. على سبيل المثال، تساهم المستودعات في توفير الائتمان للمساهمين بالبيانات، إلا أنها لم تطور بعد أي تسهيل موجه نحو الاستشهاد الجزئي الذي يهدف إلى تسليط الضوء على الأدوار التفصيلية للمساهمين - من فعل ماذا؟ غالباً ما يكون للمجتمعات العلمية طريقتها الخاصة في الاستشهاد ببياناتها، وسيكون من المفيد جداً أن تقدم المستودعات تسهيلات تمكن مجتمعات معينة من تخصيص الطريقة التي ينبغي بها الاستشهاد ببياناتهم الخاصة مع الحفاظ على قابلية التشغيل البيني في نفس الوقت.

يتم الحصول على مجموعات البيانات المستخدمة في التجارب والدراسات في كثير من الأحيان عن طريق إصدار استفسارات محددة لمصادر ديناميكية مثل قواعد البيانات. مطالبة العلماء بتحميل لقطات ثابتة أو أوصاف نصية لمجموعات البيانات الفرعية كما تفعل معظم مستودعات البيانات البحثية الحالية، لا يكفي لتمكين التحديد الدقيق للجزء المحدد.

ظهرت في الآونة الأخيرة مقترحات تطالب مصادر البيانات أن تعزز تسهيلات الاستعلام الخاصة بها لتمكين إصدار البيانات، أي ضمان إمكانية استرداد النماذج السابقة لمجموعة البيانات، والطابع الزمني، بما يعني تمييز العمليات على مجموعات البيانات بطابع زمني. يمكن أن تساهم المستودعات التي تتعامل مع نشر البيانات في تنفيذ النهج المقترح من خلال: (أ) تحديد مجموعة البيانات باستمرار مثل، التعامل مع تخزين الاستعلام الذي يحدد بشكل موحد جزء مجموعة البيانات والبيانات الوصفية النسبية ، وتعيين المعرف الثابت Persistent Identifier (PID) لهذا الاستعلام الغني، و(ب) حل المعرّفات الثابتة (PIDs) واسترداد جزء البيانات فعلياً مثل، إنتاج صفحة مقصودة يمكن قراءتها بواسطة الإنسان وقابلة للتنفيذ على الآلة للوصول إلى جزء من مجموعة البيانات عبر تنفيذ الاستعلام والبيانات الوصفية. [٦٠]

يجب أن تأخذ مستودعات البيانات البحثية العامة في الاعتبار الاتجاه الناشئ لتوثيق مجموعات البيانات المنتجة من خلال أوراق البيانات، حيث تحتوي أوراق البيانات على معلومات حول مجموعات

البيانات وكيفية إنتاجها، مما يسهل فهمها بشكل أفضل، و يساهم هذا إلى حد كبير في تبسيط وتحسين إعادة استخدام مجموعات البيانات، كما يجب أن تحافظ المستودعات أولاً وقبل كل شيء على رابط بين مجموعة البيانات وورقة ، أو أوراق البيانات الخاصة بها - إن وجدت- وتشجع مستخدمي البيانات الراغبين في الاستشهاد بمجموعات البيانات على الاستشهاد فعلياً بورقة ، أو أوراق البيانات، بالإضافة إلى ذلك ومن أجل الحصول على فهم أكثر اكتمالاً ، يجب أن تحافظ المستودعات أيضاً على رابط لجميع الأوراق التي تشير إلى مجموعة البيانات نظرا لأن العلاقة بين مجموعات البيانات والأوراق كثيرة ومتعددة الأشكال، لتحقيق هذا المقترح يجب على الناشرين الموافقة على إرشادات مفصلة وتحرير النسخ وتطوير ممارسات لضمان أن أوراق البيانات التي ينشرونها تحتوي على اقتباس واضح لمجموعة أو مجموعات البيانات التي تدور حولها كل ورقة بيانات، ويجب أن تتضمن هذه الاقتباسات إشارة إلى المستودع حيث يتم تخزين البيانات بالفعل.

٣- الخاتمة والتوصيات.

٣-١- الخاتمة:

قامت الدراسة الحالية باستكشاف مستودعات البيانات البحثية العامة، أي المستودعات التي تقبل نشر أي نوع من مجموعات البيانات، في أي تخصص موضوعي بصرف النظر عن الانتماء المؤسسي أو الجغرافي للمستودع، وبصرف النظر عن الطبيعة الشكلية للبيانات؛ وعند تطبيق معايير الاختيار هذه، استبعدنا بالضرورة المستودعات المتخصصة موضوعيا أو مؤسسيا ومستودعات الجامعات والمؤسسات البحثية والتي تلعب بالتأكيد دوراً رئيساً في الإشراف على البيانات البحثية وذلك من خلال التعرف على البيانات الأولية للمستودعات و مسح الممارسات والأساليب الخاصة بنشر البيانات البحثية التي تنتهجها تلك الفئة من المستودعات.

قدمت الدراسة مستودعات البيانات البحثية العامة، واستعرضت الممارسات الحالية للمستودعات وناقشت قضاياها المفتوحة فيما يتعلق بثماني ميزات رئيسة لنشر البيانات، هي تنسيق مجموعة البيانات، والوثائق، والترخيص، وتكاليف النشر، والتحقق، والتوافر، والاكتشاف، والوصول، والاستشهاد .

تتراوح أعمار المستودعات محل الدراسة بين ست سنوات لأحدث المستودعات وثلاثة عشر عاما لأقدم المستودعات وهو عمر إلى حد ما متوسط لقياس تأثير تلك المستودعات على الساحة المعلوماتية ودورها في دعم وتعزيز نشر البيانات البحثية.

كذلك تتنوع التبعية الدولية لمستودعات الدراسة ما بين جمهورية الصين الشعبية، وهولندا، واستراليا، والولايات المتحدة الامريكية، وألمانيا مثل هذا التنوع الإداري والدولي وحتى العمري يضمن بصورة أو بأخرى تنوع التغطية الموضوعية على الرغم من عموميتها ويضمن تنوع واختلاف الممارسات وفقا لطبيعة ونشاط المؤسسات الإدارية المشرفة على المستودعات محل الدراسة.

يتبين من التوزيع الزمني لنشر مجموعات البيانات في مستودعات البيانات البحثية محل الدراسة الارتفاع الطردي لمجموعات البيانات البحثية المتاحة في مستودعات الدراسة خلال السنوات السبعة الأخيرة من ٢٠١٥ إلى ديسمبر ٢٠٢١. ويؤكد ذلك على قبول المجتمع العلمي لهذا الرافد من روافد نشر البيانات البحثية.

يوضح التوزيع الموضوعي لمجموعات البيانات البحثية المنشورة والمتاحة في مستودعات الدراسة أن العلوم الفيزيائية والطبيعية هي العلوم الأكثر تمثيلاً في مجموعات البيانات المنشورة في مستودعات الدراسة بنسبة (٥٢,١٩%)، وأن العلوم الاجتماعية والانسانيات هي العلوم الأقل تمثيلاً بنسبة (٨,٥٤%)

حيث غالبا ما تعود التغطية الموضوعية لمجموعات البيانات للطبيعة التخصصية لجهات الاشراف على المستودعات البيانات محل الدراسة.

لتحليل منهجية تعامل مستودعات البيانات البحثية العامة مع مجموعات البيانات، وتأثير تلك المنهجيات في دعم وتعزيز نشر ومشاركة وإعادة استخدام البيانات، تم تحديد ثمانية خصائص رئيسة؛ ترى الباحثة أن تلك الخصائص تشكل دوراً هاما في توصيف الدعم الذي تقدمه هذه المستودعات لنشر البيانات البحثية ومشاركتها وإعادة استخدامها،وهي (تنسيق مجموعات البيانات – التوثيق - الترخيص- تكاليف النشر- التحقق من الصحة - التوافر- قابلية الاكتشاف والوصول- الاقتباس والاستشهاد المرجعي) ودور تلك الخصائص في دعم عمليات نشر وإتاحة البيانات البحثية.

بالنسبة للتنسيق تقبل جميع المستودعات محل الدراسة مجموعات البيانات المخزنة في ملفات بأي تنسيق، فلا توجد قيود معينة على التنسيقات المسموح بها .ومع ذلك، غالباً ما يتم تشجيع مؤلفي ومالكي البيانات على إرسال البيانات بتنسيقات قياسية أو مفتوحة حيث تكون مناسبة للحفظ وإعادة الاستخدام.

فيما يتعلق بالتوثيق؛ تطلب جميع المستودعات- محل الدراسة- تحديد البيانات الوصفية الأساسية عند إرسال مجموعة البيانات؛ ولا يطلب أي منها صراحة ربط مجموعات البيانات المودعة بورقة بيانات، عند تحليل الوثائق المتاحة للجمهور، ولم يتم العثور على أي مواصفات للبيانات الوصفية المدعومة داخلياً من قبل المستودعات المحددة.

بالنسبة لتراخيص الإيداع، تمارس المستودعات مجموعة متنوعة من السياسات حيث يطلب من مالكي البيانات التسجيل في المستودع وبمجرد إرسال أو تحميل مجموعات البيانات فإنهم بهذا يوافقون ضمنيا على سياسات المستودع. هذه السياسات غير متجانسة تماما بين المستودعات محل الدراسة.

فيما يتعلق بتكاليف النشر يعد Dryad المستودع الوحيد في مستودعات الدراسة- الذي يفرض على مقدمي ومالكي البيانات دفع رسوم بشكل مستقل عن حجم الملفات. كل المستودعات تقدم مساحة تخزين محددة حيث يمكن للمستخدمين النشر مجاناً.

للتحقق من صحة مجموعات البيانات تقوم مستودعات البيانات – محل الدراسة- من خلال مجموعة من الممارسات والخدمات، حيث تدعم المستودعات مرحلة ما قبل النشر باستخدام التقنيات التي تهدف إلى التحقق من أن مجموعات البيانات ليست تالفة وأن البيانات الوصفية المرتبطة بها صحيحة ومكتملة من الناحية التركيبية، وللتحقق من صحة البيانات في مرحلة ما بعد النشر، تميل المستودعات- محل الدراسة- إلى تقديم أهم الإحصائيات المتعلقة بتنزيل مجموعات البيانات أو الاستخدام.

بالنسبة للتوافر تنفذ المستودعات- محل الدراسة- نهجين رئيسين لضمان التوافر الحالي والمستقبلي للبيانات، أي أرشفة البيانات بطريقة آمنة وتطبيق آلية لحفظها. للأرشفة الآمنة للبيانات، جميع المستودعات –محل الدراسة- تخزن نسخ متعددة من مجموعات البيانات إما في قواعد بيانات خاصة بها أو من خلال مزودي الخدمة، ولحفظ البيانات، تميل المستودعات محل الدراسة إلى استخدام التنسيقات الخاصة بتهجير البياناتData migration، إلى بجانب تخزينها في نسخ متعددة.

فيما يتعلق بقابلية الاكتشاف والوصول، تقوم مستودعات البيانات – محل الدراسة- بتقديم تسهيلات اكتشاف البيانات على غرار التسهيلات التقليدية التي تقدمها مستودعات المقالات منذ سنوات عديدة، إلا أن المعايير التي يرغب الباحثون في صياغتها عند البحث عن مجموعات البيانات، غالباً ما تكون مختلفة تماماً عن تلك التي يستخدمونها في البحث عن المقالات العلمية، حيث يتم البحث في مجموعات البيانات بشكل أساسي لإجراء التحقق وإعادة الاستخدام في الأنشطة العلمية .

بالنسبة للاستشهاد المرجعي والاقتباس، تدعم مستودعات البيانات البحثية -محل الدراسة- كل من المالكين والمستهلكين، حيث تقدم خيارات متنوعة لمعالجة اقتباس البيانات (سلسلة اقتباس - خيار التصدير- خيار التضمين - خيار المشاركة) وتعتمد كل هذه الأساليب على معرف الكيان الرقمي DOI المخصص لكل مجموعة بيانات في وقت النشر.

الدعم الذي تقدمه مستودعات البيانات العامة حالياً يشبه بشكل أساسي الممارسات المعمول بها في نشر الأدبيات، تم إعادة التفكير في عدد قليل جداً من المرافق للتعامل مع خصائص مجموعات البيانات واستخداماتها المختلفة. على سبيل المثال، الجوانب ذات الصلة المتعلقة بنطاق البيانات، مثل التغطية والمعلومات السياقية؛ كالمصدر والاستشهاد، غالباً لا يحظيان بالاهتمام اللازم. يصبح هذا التناقض أكثر وضوحا عند الأخذ في الاعتبار الممارسات العلمية الجديدة الناجمة عن التقدم التكنولوجي في إدارة البيانات، على سبيل المثال، تفتح البيئات التعاونية القائمة على الويب الطريق لإنشاء منتجات البيانات من خلال الجهد التعاوني للعديد من العلماء الذين ينتجون وينشرون هذه المنتجات في مراحل مختلفة، وبالتالي في العديد من الإصدارات. لا تنعكس هذه الحاجة في الميزات التي توفرها المستودعات التي تم تحليلها والتي لا تشدد كثيراً على قضايا الإصدار، فتركيزهم الرئيسي يقتصر في الواقع على مجموعات البيانات النهائية فقط، وبالمثل، تتيح الأنظمة التي تدعم التعليقات التوضيحية والتصنيف الآن ربط كمية متزايدة من المعلومات السياقية بمجموعة البيانات المنشورة. كما لم يتم تصميم المستودعات بعد لجمع المعلومات حول مجموعات البيانات وإتاحتها واستخدامها بشكل كامل لتوفير خدمة اكتشاف وتحقق أفضل.

في الوقت الحالي لعل أكثر المشكلات التي تواجه مستودعات البيانات البحثية العامة هي التنوع الهائل في تنسيقات وانماط مجموعات البيانات التي يتم التعامل معها، وعدم وجود مجتمع علمي محدد وذلك نظراً لتعدد التخصصات الموضوعية واختلاف الدوافع والأهداف لدى كل من مالكي البيانات ومستهلكي البيانات، واخيراً الافتقار إلى الممارسات الموحدة والمشتركة.

وبالتالي يمكننا أن نستنتج أن مستودعات البيانات البحثية العامة الحالية هي خدمات قابلة للتطبيق لنشر البيانات ومع ذلك تحتاج إلى تطوير وتعزيز ممارساتها لدعمها بشكل أفضل، ويجب على المستودعات العامة معالجة قضية رئيسية للتغلب على القيود المفروضة من خلال عدم التجانس الشديد لمجموعات البيانات، هي كيفية تحسين المواصفات التي تقدمها المستودعات، مثل التنسيقات التي تدعمها، وإجراءات التحقق المقدمة وإدارة تراخيص البيانات.

٣-٢-التوصيات

ينبغي أن نشير إلى أن نتائج هذه الدراسة يمكنها أن تساعد في فهم واستكشاف مستودعات البيانات البحثية العامة، وبناءً على ما تقدم يمكن التوصية بما يلي:

١. إيجاد آلية تمكن مستخدمي البيانات من المساهمة في مجموعات البيانات والتفاعل الآني بينهم وبين مالكي البيانات من خلال المستودعات، وقد يتشكل هذا التفاعل بإنتاج نسخة جديدة أو عن طريق إثراء المحتوى وتصحيح الأخطاء أو الإبلاغ عن تجاربهم الملموسة الناجحة وغير الناجحة في استخدام مجموعات البيانات في تحقيق معين.

٢. المستودعات العامة لا تهدف منفردة لنشر البيانات البحثية، إنها جزء من نظام متكامل حيث يتم استدعاء الجهات الفاعلة المتنوعة من باحثين، وناشرين، وموردين .... لبذل الجهود وتحديد السياسات والممارسات والأدوار المشتركة والاتفاق عليها لتوجيه نشر البيانات المناسب.

٣. من المهم جدا أن تقوم المكتبات الأكاديمية بدورها في حث العلماء على إيداع مجموعات البيانات الخاصة بهم في مستودعات البيانات البحثية مع توفير تعريف تام بالخدمات التي يمكن للمستودع تقديمها.

٤. من الضروري إعادة النظر في تراخيص النشر بحيث تدعم المستودعات نشر إصدارات مختلفة لمجموعة البيانات؛ بحيث يتميز كل منها بترخيص متنوع مما يسهل مشاركة البيانات إلى أقصى حد ممكن.

٥. ضرورة صياغة معايير لنشر البيانات البحثية تتفق مع أصحاب المصلحة من مؤلفين وناشرين وموردين ومستخدمين لمجموعات البيانات، بحيث يصبح تنفيذ خدمات المستودعات أكثر ضبطا.

٦. قد يكون من المفيد استغلال التطورات الناتجة عن أبحاث المقاييس البديلة في وضع وتنفيذ آليات للتحقق من جودة مجموعات البيانات.

٧. من أجل دعم نشر مجموعات البيانات بشكل أفضل، على المستودعات إعادة تصميم وتطوير سياسات التوافر الخاصة بهم والإعلان عنها وجعلها جزء لا يتجزأ من شروط الخدمة، وربما يكون من الأفضل أن تخصص سياسة للنشر على مستوى القطاعات الموضوعية المختلفة.

٨. إجراء دراسات أكثر عمقا في خدمات اكتشاف والوصول إلى مجموعات البيانات التي تقدمها المستودعات؛ فما زال الأمر يتطلب تحليلا متعمقا لأنماط الاكتشاف الأكثر شيوعا عبر المجالات الموضوعية، ودراسات لفهم جدوى وفاعلية البيانات الوصفية ودورها في اكتشاف مجموعات البيانات والوثائق المرتبطة بها.

٩. من الضروري تطوير مبادرات وممارسات وأدوات للاستشهاد مفهومة من قبل الإنسان وقابلة للتنفيذ آليا.

قائمة المصادر والمراجع:

1 4TU.ResearchData(2022) Upload your data in our data repository https://data.4tu.nl/info/en/use/publish-cite/upload-your-data-in-our-data-repository

2 4TU.ResearchData.- available online< https://data.4tu.nl/info/en/>

3 Ball, A, Chen, S, Greenberg, J, Perez, C, Jeffery, K and Koskela, R (2014) Building a disciplinary metadata standards directory. International Journal of Digital Curation, 9(1): 142–151. DOI: http://dx.doi. org/10.2218/ijdc.v9i1.308< https://www.semanticscholar.org/paper/Building-a-Disciplinary-Metadata-Standards-Ball-Chen/929941ab6447b0cc69f7135d9ed9a2abd7a3947e>

4 Broeder, D and Lannom, L (2017) Data type registries: A research data alliance working group. D-Lib Magazine, 20(1/2). DOI: http://dx.doi.org/10.1045/january2014-broeder

5 Burda, D and Teuteberg, F (2018) Sustaining accessibility of information through digital preservation: A literature review. Journal of Information Science, 39(4): 442–458. DOI: http://dx.doi. org/10.1177/0165551513480107< https://journals.sagepub.com/doi/10.1177/0165551513480107>

6 Campbell, J (2015) Access to scientific data in the 21st century: Rationale and illustrative usage rights review. Data Science Journal, 13: 203–230. DOI: http://dx.doi.org/10.2481/dsj.14-043< https://www.researchgate.net/publication/273297647_Access_to_Scientific_Data_in_the_21st_Century _Rationale_and_Illustrative_Usage_Rights_Review>

7 Cimphony Orchestra(2022) Open Grid Systems < https://www.opengrid.com/content/>

8 Costantino Thanos.(2017) Research Data Reusability: Conceptual Foundations, Barriers and Enabling Technologies.- doi:10.3390/publications5010002< https://www.mdpi.com/2304-6775/5/1/2>

9 Creative Commons licenses are 20(2022) Creative Commons .- available online< https://creativecommons.org/>

10 Csiro data access portal .- available online< https://data.csiro.au/ >

11 Data Access And Licensing(2022) THE WORLD BANK . Data Catalog.- available online< https://datacatalog.worldbank.org/public-licenses>

12 Data Citation Synthesis Working Group.- (2022) The Future of Research Communications and e-Scholarship.< https://force11.org/group/data-citation-synthesis-working-group/>

13 Deka Anggawira, Nina Mayesti(2020) The Indonesian National Scientific Repository: A Case Study of Research Data Sharing.- Preservation, Digital Technology & Culture.- June 9, 2020.- - avilable online<https://www.degruyter.com/document/doi/10.1515/pdtc-2019-0015/html>

14 https://datadryad.org/stash> Repository.- available online< Dryad Digital

15 Eschenfelder, K R and Johnson, A(2016) Managing the data commons: Controlled sharing of scholarly data. Journal of the Association for Information Science and Technology, 65(9): 1757–1774. DOI: http://dx.doi. org/10.1002/asi.23086< https://minds.wisconsin.edu/handle/1793/68564>

16 European Commission. (2018) Commission Recommendation on Access to and Preservation of Scientific Information.. Available online:< https://ec.europa.eu/research/science-society/document_ library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf.>

17 European Commission. (2020)Directorate-General for Research & Innovation. H2020 Programme, Guidelines on FAIR Data Management in Horizon 2020. 2016. Available online: http://ec.europa.eu/research/ participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf .

18 figshare (2022) figshare partners with open access mega journal publisher PLOS. Figshare Blog. < http://figshare.com/blog/figshare_partners_with_Open_Access_mega_journal_ publisher_PLOS/68>

19 Figshare.- available online< https://figshare.com/ >

20 Floridi, L.; Sanders, J.(2014) Levellism and the Method of Abstraction; Research Report 22.11.04; Information Ethics Group (Oxford University and University of Bari): Oxford, UK, 2014. Available online:< Available online: https://www.nsf.gov/nsb/documents/2014/LLDDC_report.pdf>

21 Hamad, F., Al-Fadel, M., & Al-Soub, A.(2019). Awareness of Research Data Management Services at Academic Libraries in Jordan:Roles, Responsibilities and Challenges. New Review of Academic Librarianship. -avilable online(under request) < https://www.researchgate.net/profile/Faten-Hamad >

22 and Pinfield, S. (2015), "Research data management and Higman, R. openness: The role of data sharing in developing institutional policies and practices", Program: electronic library and information systems, Vol. 49 No. 4, pp. 364-381. https://doi.org/10.1108/PROG-01-2015-0005>

23 inayak Wadhwa, Manoj Kumar Joshi(2022) Status of Indian Research Data Repositories: A Study based on Research Data Registry.- ENVISIONING DIGITAL TRANSFORMATION IN LIBRARIES FOR NEXTGEN ACADEMIC LANDSCAPE.- available online< https://ir.inflibnet.ac.in/handle/1944/2402 >

24 JISK,(2021) Data Centers: their use, value and impact. A research Information Network Report, September 2021. Available online:< http://www.rin.ac.uk/system/files/attachments/Data_Centres_Report.pdf>

25 Limani, F., Hajra, A., Ferati, M., Radevski, V. (2020) Requirements and Recommendations for University Research Data Repository: A Case Study. - Piet Kommers, Boyan Bontchev and Pedro Isaías (ed.), Proceedings of the 18th International Conference e-Society, 2-4 april, 2020 (pp. 51-58). IADIS Press. - avilable online < http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-96914>

26 Manghi, P and Mannocci, A (2018) Data Searcher: Preliminary Analysis of Data Sources Interlinking. In: Research and Advanced Technology for Digital Libraries. Vol. 8092 of Lecture Notes in Computer Science. pp. 458–461< https://link.springer.com/book/10.1007/978-3-642-40501-3>

27 Moreau, L.; Freire, J.; Futrelle, J.; Mcgrath, R.E.; Myers, J.; Paulson, P.(2018) The Open Provenance Model: An Overview. In IPAW 2018: Provenance and Annotation of Data and Processes; Springer: Berlin, Germany, 2018; Volume 5272. Available online: < https://link.springer.com/chapter/10.1007/978-3-540-89965-5_31>

28 Mozersky, Jessica; Walsh, Heidi; Parsons, Meredith; McIntosh, Tristan; Baldwin, Kari; DuBois, James M. (2019) Are we ready to share qualitative research data? Knowledge and preparedness among qualitative researchers, IRB Members, and data repository curators, IASSIST Quarterly 43(4), pp. 1-23.- .- available online https://pubmed.ncbi.nlm.nih.gov/32205903/

29 </https://osf.io >online available -. Open Science Framework

30 Osterlund, C.; Carlile, P.(2015) Relations in practice: Sorting through practice theories on knowledge sharing in complex organizations. Inf. Soc. 2015, 21, 91–107. Available online: <https://www.researchgate.net/publication/220175059_Relations_in_Practice_Sorting_Through_Practice_Theories_on_Knowledge_Sharing_in_Complex_Organizations>

31 Palmer, C L, Weber, N M and Cragin, M H (2017) The analytic potential of scientific data: Understanding reuse value. Proceedings of the American Society for Information Science and Technology, 48(1): 1–10. DOI: http://dx.doi.org/10.1002/meet.2011.14504801174 < https://core.ac.uk >

32 Parsons, M A, Godoy, LeDrew, E, de Bruin, T F, Danis, B, Tomlinson, S and Carlson, D (2011) A conceptual framework for managing very diverse data for complex, interdisciplinary science. Journal of Information Science, 37(6): 555–569. DOI: http://dx.doi.org/10.1177/0165551511412705

33 Patel, D. (2016), "Research data management:aconceptual framework", *Library Review*, Vol. 65 No. 4/5, pp.226-241.available online < https://doi.org/10.1108/LR-01-2016-0001 >

34 Peters, D (2018) Wiley partnership with figshare enables data sharing. Press release. <http:// eu.wiley.com/WileyCDA/PressRelease/ pressReleaseId-119082.html.>

35 Radhika Makhecha, Ashok Singh (2021) Data Repository: What it is, Types and Guide.- questionpro. Available online: < https://www.questionpro.com/blog/data-repository/>

36 Rauber, A, Asmi, A, van Uytvanck, D and Proll, S (2020) Data citation of evolving data. RDA working group on Data Citation: Making Dynamic Data cite able Recommendations< https://rd-alliance. org/system/files/documents/RDA-DC-Recommendations_150924.pdf.>

37 Re3data. Registry of Research Data Repositories. - available online < https://www.re3data.org/>

38 Wani(2020) A study of Safat Mushtaq Misgar, Ajra Bhat, Zahid Ashraf BRICS countries.- Open Access research data repositories developed by available -.Digital Library Perspectives Vol. 38 No. 1, 2022 pp. 45-5 <https://doi.org/10.1108/DLP-02-2020-0012 >online

39 > https://www.scidb.cn/en >online available -.Science Data Bank

40 Serving the long tail.- DIGITAL INFRASTRUCTURES for RESEARCH 2018< https://www.digitalinfrastructures.eu/content/serving-long-tail>

41 Stonebraker, M.; Becla, J.; Dewitt, D.J.; Lim, K.T.; Maier, D.; Ratzesberger, O.; Zdonik, S.B(2019). Requirements for Science Data Bases and SciDB. In Proceedings of the CIDR 2009, Fourth Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, 4–7 January 2019. Available online: < https://dblp.org/db/conf/cidr/cidr2019.html>

42 Strang, T.; Linnhoff-Poppien, C. A (2018) Context Modeling Survey. In Proceedings of the First International Workshop on Advanced Context Modeling, Reasoning and Management Associated with the Sixth International Conference on Ubiquitous Computing, Nottingham, UK, 7 September 2018. Available online: < https://link.springer.com/chapter/10.1007/978-3-642-17881-8_33>

43 Thanos, C. (2015) Scientific Data (Re)Usability: Concepts, Impediments, and Enabling Technologies. In Proceedings of the International Conference on Digital Presentation and Preservation of Cultural and Scientific Heritage, Veliko Tarnovo, Bulgaria, 28–30 September 2015. Available online: <https://openportal.isti.cnr.it/doc?id=people::318e0b573311265f5828c78989a62a9a>

44 Wallis JC, Rolando E, Borgman CL (2013) If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. PLoS ONE 8(7): e67332. doi:10.1371/journal.pone.0067332 <https://www.researchgate.net/publication/255736892_If_We_Share_Data_Will_Anyone_Use_Them_Data_Sharing_and_Reuse_in_the_Long_Tail_of_Science_and_Technology>

45 Wickett, K.; Sacchi, S.; Dubin, D.; Renear, A.(2021) Identifying Content and Levels of Representation in Scientific Data. In Proceedings of the American Society for Information Science and Technology, Baltimore, MD, USA, 28–31 October 2021. Available online: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.14504901199>

46 Youngseek Kim, Melissa Adler(2015) Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories.- International Journal of Information Management.- Volume 35, Issue 4, August 2015, Pages 408-418.- available online https://www.sciencedirect.com/science/article/abs/pii/S0268401215000432

47 < /https://zenodo.org > available online -.Zenodo

48 Zimmerman(2013), A. Data Sharing and Secondary Use of Scientific Data: Experiences of Ecologists. Thesis, Degree of Doctor of Pholosophy Information and Library Studies, University of Michigan, Ann Arbor, MI, USA, 2013. Available online: <https://deepblue.lib.umich.edu/bitstream/handle/2027.42/39373/ann_zimmerman_dissertation_2003.pdf?sequence=2 >

٤٩ الزلباني، محمد سعد، علي، أسامة السيد – مشرف- (٢٠٢٠) – مخطط بحث- مستودعات إدارة البيانات البحثية: دراسة مسحية للنظم الدولية ووضع نموذج لتنفيذها في المؤسسات العلمية العربية.- البوابة العربية للمكتبات والمعلومات.

٥٠ السيد، أميرة محمد، آمال طه محمد، سهير عبدالباسط عيد (٢٠٢٢).- اتجاهات الباحثين في جامعة بني سويف نحو إنشاء مستودع للبيانات البحثية: دراسة ميدانية .- المجلة المصرية لعلوم

المعلومات.- مج٩, ع١.

٥١ السيد، أميرة محمد، آمال طه محمد، سهير عبدالباسط عيد(٢٠٢١) .- البيانات البحثية ومستودعاتها في الجامعات: دراسة نظرية .- المجلة المصرية لعلوم المعلومات.- مج٨, ع٢.

٥٢ غنيم، ريهام عاصم (٢٠٢٢) دوريات البيانات ودورها في تعزيز نشر ومشاركة البيانات وإعادة استخدامها: دراسة استكشافية تحليلية. المجلة الدولية لعلوم المكتبات والمعلومات مجلد ٩ – عدد٣ https://ijlis.journals.ekb.eg/article_192472_57ee84bd59225e626ed2a64c2>b66be54.pdf

٥٣ فراج، عبد الرحمن. (٢٠٢٠) مستودعات البيانات البحثية. أحوال المعرفة. س٢٤، ٩٨. ص ص ٦٦-٦٩.- متاح على الخط المباشرhttps://zenodo.org/record/3818106#.Y7h9ARXP25c

٥٤ نابتي، محمد الصالح، بديس، عتروز، وعاشوري، نضيرة. (٢٠١٨) إدارة وحفظ المجموعات الرقمية بمستودعات البيانات البحثية المفتوحة: دراسة تحليلية لوضعية المستودعات الإفريقية في. Re3dataا.- لمؤتمر الرابع والعشرون: البيانات الضخمة وآفاق استثمارها: الطريق نحو التكامل المعرفي، مسقط: جمعية المكتبات المتخصصة فرع الخليج العربي، ١ - ١٨ .متاح على الخط المباشر< Record/com.mandumah.search://htt/٨٧٠١٥٠ >