وظائف نظم إدارة البيانات البحثية ودورة حياتها

محمد مسعد محمد الزلباني
مكتب مكتبة الكونجرس بالقاهرة
mh_zalabany@hotmail.com
مراجعة وإشراف
أ.د. أسامة السيد محمود
أستاذ المكتبات والمعلومات
كلية الآداب – جامعة القاهرة

المستخلص

تعد البيانات البحثية وادارتها من أهم الموضوعات الساخنة على الساحة العلمية والأكاديمية في الوقت الراهن وقد زاد الاهتمام بها بسبب الثورة الرقمية والتوسع في استخدام شبكة الإنترنت مما سهل عمليات الحفظ والتخزين والمشاركة للبيانات، وأدى ذلك الى تزايد كم البيانات البحثية المنتجة من المؤسسات العلمية والأكاديمية وغيرها من المؤسسات، وتقوم هذه البيانات بدور حيوي في تقييم البحوث علاوة على برهنتها على دقة البحوث هذا بخلاف اعادة استخدام هذه البيانات وما يوفره من وقت وجهد على الباحثين والجهات البحثية،

ولكي تكون البيانات البحثية مفيدة للمجتمع العلمي لابد لها من المرور بعدة مراحل يمكن لنا أن نطلق عليها دورة حياة البيانات البحثية تمر البيانات داخل هذه الدورة في كل مرحلة بالعديد من الخطوات والإجراءات الضرورية للحصول على هذه البيانات في الصيغة النهائية التي يمكن الاستفادة منها بشكل كبير وفي كل مرحلة من هذه المراحل توجد العديد من الادوات والبرمجيات والمعايير التي يجب اتباعها لإدارة البيانات البحثية مرورا بالتخطيط لإدارة البيانات البحثية وجمعها ومعالجتها وحفظها ونشرها ثم اعادة الاستخدام، وتتناول هذه الورقة البحثية وظائف نظم ادارة البيانات البحثية خلال مراحل دورة حياتها.

الكلمات المفتاحية: البيانات البحثية – نظم إدارة البيانات البحثية – دورة حياة البيانات البحثية.

تاريخ استلام البحث: 2023/2/9 **تاريخ إجازة البحث: 2023/3/24**

1/1تمهيد

تُعد عملية إدارة البيانات البحثية عملية معقدة لاشتراك العديد من الأطراف فيها سواء كان الباحثين أو اختصاصي المعلومات أو الجهات المسئولة عن البيانات، ولكنها اذا تمت بشكل منضبط من البداية فإنها ستوفر الكثير من المال والوقت، فعند اعداد بياناتك البحثية لكتابة أطروحة علمية، تأخذ البيانات البحثية صورا متعددة تتراوح ما بين القياسات والأرقام، الصور، الوثائق والمنشورات(**Research data-management-guide**) وجميع البيانات الخام المستخدمة في البحوث وهذه البيانات تحتاج الى العديد من الإجراءات لكي تكون قابلة للنشر من خلال مستودعات البيانات البحثية.

1/2مشكلة الدراسة

تُعد إدارة البيانات البحثية من العمليات ذات التفاصيل الكثيرة ومتعددة الإجراءات لتضمن إدارة وحسن استغلال البيانات البحثية ومن ثم فهذه العمليات كانت تحتاج الي حصر وتوضيح لكافة إجراءاتها وخطواتها وما يتعلق بها من بروتوكولات وأدوات.

1/3أهمية الدراسة

ترجع أهمية الدراسة الى أهمية إدارة البيانات البحثية لتفعيل دورها في مجتمع البحث العلمي والجهات الاكاديمية وتوضيح كيفية إدارة البيانات البحثية والمراحل التي تمر بها

1/4أهداف الدراسة

- توضيح مراحل إدارة البيانات البحثية
- حصر الأدوات المستخدمة في إدارة البيانات البحثية
- معرفة دور مستودعات البيانات البحثية
- معرفة المعايير والبروتوكولات المستخدمة في إدارة البيانات البحثية

1/5تساؤلات الدراسة

- ما هي مراحل إدارة البيانات البحثية
- ماهي الأدوات المستخدمة في إدارة البيانات البحثية

- ماهي معايير البيانات البحثية
- كيف يمكن الإفادة من البيانات البحثية

1/6مصطلحات الدراسة

البيانات البحثية Research Data

هي البيانات التي يتم جمعها وفحصها ومعالجتها واستخدامها كمصادر أولية في الأبحاث لتأكيد صحة نتائج هذه الأبحاث التي يشار اليها كجزء من المخرجات النهائية للمصادر. وتتضمن البيانات البحثية نتائج الملاحظات العلمية والاستبيانات والمقابلات وقوائم المراجعة وغيرها. وقد تكون هذه البيانات خام أو صور تجريبية متاحة للمستفيدين وقد تكون مطبوعة او رقمية او في شكل إحصاءات او صورة أو ملفات صوتية أو فيديو.

إدارة البيانات البحثية(RDM (Research Data Management

تصف إدارة بيانات البحث طريقة لتنظيم وتخزين البيانات التي جمعها مشروع بحثي بأكثر الطرق فعالية وإدارة لهذه البيانات التي تم جمعها طوال عمر المشروع البحثي من خلال اتفاقيات محددة مسبقا علاوة على انها تحدد أنها طريقة مشاركة البيانات والوصول إليها وحفظها والتخلص منها بشكل آمن.

مستودعات البيانات البحثية Research Data repositories

هي مستودعات رقمية تقوم بتخزين البيانات التي تم انشائها من قبل الباحثين على المدي الطويل وبالتالي ادارتها والوصول اليها وثم إعادة استخدامها من اجل القيام بأبحاث جديدة وتحسين نتائج الأبحاث.

1/9الدراسات السابقة

1/1/9الدراسات العربية

إدارة البيانات البحثية وصيانتها في مجال البحوث العلمية: دراسة نظرية /أ. دلال عبدالقادر الأنصاري ماجستير إدارة المعلومات جامعة الملك عبدالعزيز، بحث مقدم إلى مؤتمر المؤتمر العلمي الاكاديمي الدولي التاسع تحت عنوان "الاتجاهات المعاصرة في العلوم الاجتماعية، الانسانية، والطبيعية" تركيا - اسطنبول 8171 تموز– يوليو 71 – 71

وقد تناولت الدراسة أهمية إدارة البيانات البحثية في تطوير البحوث العلمية باعتباره موضوع حديث في الأوساط البحثية ,و تهدف الدراسة إلى التعرف على صيانة البيانات البحثية بشكل رئيس ,و توضيح مفهوم ملف صيانة البيانات البحثية و أهداف إنشاءه و استخداماته ,كما تسعى إلى التعرف على ماهية البيانات البحثية و أهمية إدارتها ,فيما تعرض الدراسة نموذج استمارة مخصصة لدراسة و عمل ملف صيانة البيانات البحثية ,بحيث يمكن للباحثين الاستعانة به لعمل صيانة لبياناتهم البحثية ,أو من خلال تدريبهم و تقديم المساعدة لهم عن طريق أخصائي المكتبات ,و توصلت الدراسة إلى أن البيانات البحثية تتمثل في السجلات الواقعية للبيانات التي يتم جمعها أو ملاحظتها كمصادر أساسية لأغراض التحليل وإنتاج نتائج البحوث الأصلية ,كما توضح أن لإدارة البيانات البحثية في مجال البحوث العلمية دور ,يتمثل أبرزها في سهولة إعادة استخدام البيانات من قبل الباحث نفسه أو باحث غيره و ذلك لمنع ازدواجية الجهود ,و التحقق من صحة نتائج الأبحاث ,و مدى التزام البيانات بالمتطلبات القانونية و الأخلاقية ,وتبين الدراسة أن الأساس في عمل ملف صيانة البيانات يتم لكل باحث ولا يمكن عمل نموذج واحد لكل كلية أو لكل قسم أو تخصص ,فيما توضح الدراسة إلى أن هناك ندرة في المصادر العربية التي تتناول موضوع البيانات البحثية و إدارتها و صيانتها ,حيث توصي الدراسة الباحثين بإجراء العديد من الدراسات العربية حول موضوع البيانات البحثية ؛ لكي يتم الاستفادة من قيمة البيانات المتراكمة و غير الفعالة ,كما توصي بضرورة وعي المستودعات المؤسسية و المكتبات العربية بموضوع إدارة البيانات البحثية و صيانتها , وتوفير دورات تدريبية لأخصائي المكتبات حول عمل و طريقة صيانة البيانات للبحوث العلمية .

دراسة الدكتور عبد الرحمن فراج بعنوان البيانات المفتوحة وإدارتها بالجامعات السعودية دراسة تحليلية، وتصور مفاهيمي لإنشاء مرفق للبيانات البحثية والمنشورة في مجلة دراسات المعلومات والتكنولوجيا لجمعية المكتبات المتخصصة فرع الخليج العربي

وتهدف هذه الدراسة إلى التعرف على الممارسات الحالية للجامعات السعودية في إدارة البيانات المفتوحة، وبصفة خاصة البيانات البحثية منها؛ وذلك من حيث المظان المفضلة إتاحة البيانات، والجهة المسؤولة عن هذه الاتاحة، ومدى الافصاح عن السياسات والرخص القانونية ذات الصلة، وماهية أنماط البيانات المتاحة، وصيغ الملفات التي أتيحت بها. وقد خلصت الدراسة إلى اهتمام الجامعات السعودية بإتاحة البيانات الاحصائية الادارية دون البيانات

البحثية، ولم تأت هذه الاتاحة عبر مستودعات الجامعات أو مرافق خاصة بالبيانات، وإنما وردت على هيئة صفحات عنكبوتية ثابتة دون أي إمكانات للبحث، ودون توفير للخدمات ذات الصلة بالبيانات البحثية، ودون خطط محددة لإدارة البيانات. كما توصلت الدراسة إلى أن هذا الاتجاه هو السائد في المملكة على المستوى الوطني بصفة عامة، وبصفة خاصة بوابة البيانات الحكومية المفتوحة وصفحةِ البيانات بموقع وزارة التعليم. وتنتهي الدراسة، اعتمادا على أسلوب التحليل التطوري للمفهوم، إلى إعداد تصور مفاهيمي إنشاء مرفق للبيانات البحثية وتوفير الخدمات ذات الصلة بالجامعات السعودية.

إدارة وحفظ المجموعات الرقمية بمستودعات البيانات البحثية المفتوحة: دراسة تحليلية لوضعية المستودعات الإفريقية فيRe3data

نابتي، محمد الصالح، عاشوري، نضيرة، وبديس، عتروز. (2018). إدارة وحفظ المجموعات الرقمية بمستودعات البيانات البحثية المفتوحة: دراسة تحليلية لوضعية المستودعات الإفريقية في .Re3data المؤتمر الرابع والعشرون: البيانات الضخمة و آفاق استثمارها: الطريق نحو التكامل المعرفي: جمعية المكتبات المتخصصة فرع الخليج العربي، مسقط: جمعية المكتبات المتخصصة فرع الخليج العربي، 1 - 18. مسترجع من https://search.mandumah.com/Record/870150

جاءت هذه الورقة العلمية بهدف عرض بعض أنظمة إدارة وحفظ البيانات الرقمية داخل مستودعات البيانات البحثية المفتوحة، التي تعد جيلا جديدا من مجموعات المعلومات القائمة على المعالجة المتوازية على نطاق واسع، وتوزيع أنظمة الملفات وقواعد البيانات بالإضافة لأنظمة التخزين وبعدها تقدم الدراسة الحالة الراهنة لمستودعات البيانات البحثية المفتوحة الإفريقية الموجودة بدليل re3dataالتي تعمل على تقديم خدمات معلومات للمستفيدين.

2/1/9 الدراسات الأجنبية

رسالة دكتوراه مقدمة إلى جامعة Syracuse University عام 2015

BEYOND "DATA THRIFTING": AN INVESTIGATION OF FACTORS INFLUENCING RESEARCH DATA REUSE IN THE SOCIAL SCIENCES / Renata Gonçalves Curty.

وقد كانت دراسة مسحية تناولت إعادة استخدام البيانات البحثية من قبل الباحثين وكانت أهداف الدراسة الرئيسية هي 1– ماهي العوامل التي تؤثر في إعادة استخدام الباحثين للبيانات البحثية 2-إلى أي مدى تؤثر هذه العوامل على الباحثين لإعادة استخدام البيانات البحثية 3- إلى أي مدى يستخدم الباحثين البيانات البحثية.

وقام الباحث بإجراء تقصي بين عينات مختلفة من الباحثين باستخدام الاستبيانات لمعرفة مدى استخدام البيانات البحثية وقياس الاستخدام بالإضافة إلى تحليل هذه النتائج لأكثر 743 باحث.

وقد توصل الباحث إلى فهم عميق لإعادة استخدام الباحثين للبيانات البحثية في سياق العلوم المفتوحة open science مع تقديم مجموعه من العوامل المؤثر على قرارات الباحثين في العلوم الاجتماعية في إعادة استخدما البيانات البحثية التي تم جمعها بواسطة آخرين ، وأظهرت الدراسة أيضا المعلومات التي نمتلكها عن سلوك الباحثين تجاه البيانات البحثية علاوة على تقديم نموذج مفاهيمي والذي يمكن عن طريقه تقييم الأبحاث المستقبلية مع تقديم ترشيحات لصناع السياسات ، العلماء ،صناع القرار عن مدى الحاجه إلى مستودعات البيانات البحثية وتحديد الاستراتيجيات والمبادرات الخاصة بإعادة استخدام البيانات البحثية (Curty، 2015)

رسالة دكتوراه مقدمة إلى جامعة Faculty of the Graduate School, University of Maryland عام 2016

DATA SHARING ACROSS RESEARCH AND PUBLIC COMMUNITIES / Yurong He

وتهدف هذه الأطروحة إلى تعزيز ثقافة تبادل البيانات البحثية وصقل إطار عمل مشاركة البيانات البحثية وكذلك توفر رؤى حول الممارسة والتصميم لمشاركة البيانات عبر تطوير إطار جديد لتبادل بيانات البنية التحتية للمعرفة لمستويات مختلفة

كما تحدد البنية التحتية المطلوبة البشرية والتكنولوجية لعمليات مشاركة البيانات عبر المجتمعات المختلفة

وتهدف إلى الإجابة عن تساؤل رئيسي وهو: كيف يتم مشاركة البيانات بشكل فعال عبر البحوث والمجتمعات العامة؟

عبر مجموعة من الأهداف الفرعية وهي:

- معالجة الفجوة المعرفية حول تبادل البيانات البحثية بين مجتمع البحث في البيئات واسعة النطاق على الإنترنت.
- تطوير إطار تبادل البيانات.

وسيتم الإجابة عن التساؤلات التالية:

- من هم مزودي البيانات؟
- من هم وسطاء تبادل البيانات؟
- ما هي عمليات تبادل البيانات؟

ونستطيع القول بأن نتائج الرسالة هي دراسة حالة لممارسات مشاركة البيانات عبر المجتمعات البحثية والعامة (He, 2016)

أطروحة الدكتوراه المقدمة إلى جامعة ولاية فلوريدا

RESEARCH DATA CURATION PRACTICES IN INSTITUTIONAL REPOSITORIES AND DATA IDENTIFIERS / DONG JOON LEE

وتتناول الدراسة عملة معالجة البيانات في المستودعات الرقمية والمستودعات المؤسسية والتفرقة بينهم وتهدف إلى اختبار ممارسات معالجة البيانات البحثية في المستودعات الرقمية باستخدام توصيفات محدده لهذه البيانات مع ظهور الدفعة القوية في استخدام مستودعات البيانات البحثية وتدور تساؤلات الدراسة حول:

- ما هي أنواع البيانات والأنشطة في المستودعات الرقمية وما هي البنية ومتطلبات البيانات الوصفية لهذه الأنشطة؟
- ما هي الأنواع الرئيسية للبيانات البحثية وكينوناتها داخل المستودعات الرقمية وإلى أي المحددات تشير؟
- ما مدى وعي القائمين على المستودعات الرقمية بالأشكال المختلفة للبيانات الوصفية

- كيف يمكن الحكم على جودة البيانات الوصفية للأبحاث؟

وقد خلصت الدراسة إلى:

- فهم أساليب معالجة البيانات وادارتها ومشاركتها وضمان الوصول إليها وإعادة استخدامها وتقييمها والحفظ على المدى الطويل.
- فهم احتياجات معالجة البيانات البحثية والتي تحتاج الى تواصل بين متخذي القرار .
- تحديد البنية التحتية المطلوبة لمعالجة البيانات البحثية.
- تحديد أنواع البيانات البحثية.

وخلصت بالحاجة المستقبلية إلى إجراء دراسات حول بنية المستودعات الرقمية وتطوير البنية التحتية لها لتواكب طبيعة البيانات البحثية(LEE, 2015)

أطروحة ماجيستير مقدمة لجامعة University of Nevada , Reno, the graduate school

NRDC DATA STEWARD: Repository for Sharing NRDC Research Data and Results / Reddy Ashok Padmaraju

وتناولت الدراسة مركز بيانات أبحاث نيفادا NRDC (Nevada research data center) وهو أحد مراكز الأبحاث التي تعمل على حفظ البيانات البحثية وتتناول الدراسة وضع وانشاء برنامج لإدارة البيانات البحثية في المركز وعمل نموذج prototype لمستودع البيانات البحثية للمركز مع توضيح الخصائص والبنية الخاصة بهذا البرنامج ويعتمد البرنامج محل الدراسة على تقنيات الويب كما يتضمن إمكانات التغذية الراجعة feedback لأشكال البيانات المترابطة (Padmaraju, 2016)

رسالة الدكتوراه المقدمة إلى The University of Wisconsin –Milwaukee عام 2018

THE IMPACT OF RESEARCH DATA SHARING AND REUSE ON DATA CITATION IN STEM FIELDS / Hyoungjoo Park

قامت هذه الدراسة بتوظيف الطرق المختلطة للبحث والتي تجمع بين تقنيات البحث شبه الآلية لتقييم تأثير مشاركة البيانات وإعادة الاستخدام للاستشهادات المرجعية في حقل العلوم والتكنولوجيا والهندسة والرياضيات (STEM) وقام البحث على دراسة أكثر من 500000 من عناصر البيانات البحثية المفتوحة في أكثر من 350 مستودع حول العالم وشملت عينة الدراسة عدد 705 مادة ببليوغرافية و15261 حالة من مشاركة البيانات وإعادة استخدامها ومشاركتها وقام الباحث بقياس ظاهرة الاقتباس الرسمي للبيانات والبيانات التي يتم الاستشهاد بها بشكل متكرر والاستشهاد الذاتي للمؤلف على العديد من المستويات

وكشفت هذه الرسالة مجموعة من النتائج الهامة فيما يتعلق بعملية مشاركة البيانات البحثية في حقل STEM:

- تنوع ممارسات تبادل البيانات باختلاف التخصصات
- تزايد عمليات تبادل البيانات في السنوات الأخيرة
- ضبط عمليات الإفادة من المستودعات وفقا للمجالات الموضوعية
- اختلاف اشكال المستودعات الرقمية ومحتواها
- تفاوت عمليات تخليق البيانات البحثية من مجال لآخر

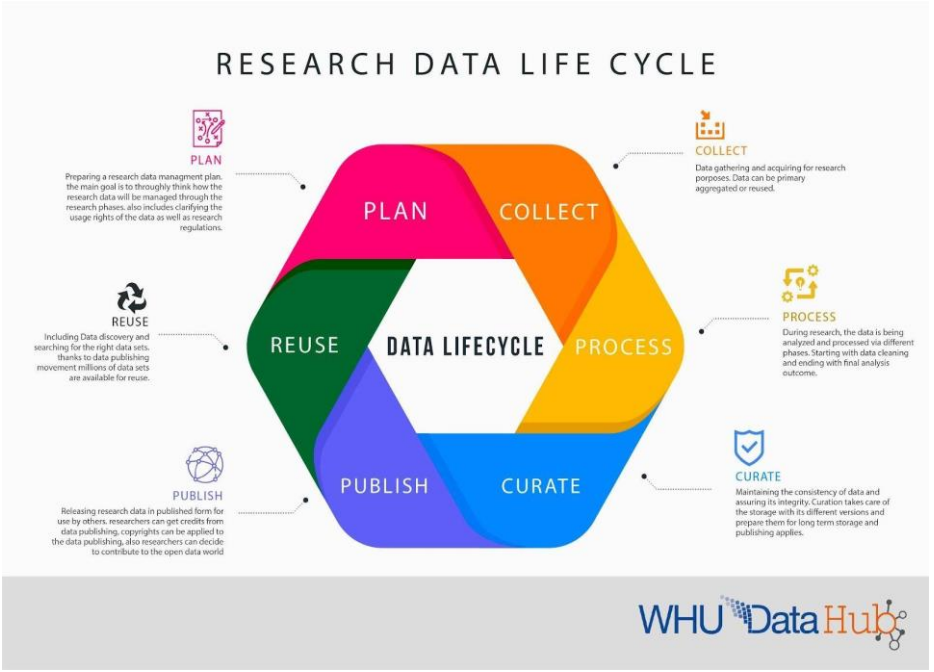
وقد أسهمت الرسالة في نقطتين أساسيتين وهما:

اظهار الأثار العملية لنظم الاقتباس الحالية للباحثين الأفراد وصناع القرار ووكالات التمويل والناشرين

الحد من العمل المطلوب لتحليل النص الكامل للمقالات المرتبطة من اجل تحديد ادلة الاقتباس(Park, 2018)

1/2 مقدمة

بالنظر الى عملية إدارة البيانات البحثية نجد أنها عملية متشعبة تحتوي على العديد من التفاصيل الهامة وتمر إدارة البيانات البحثية بمجموعة من العمليات والمراحل التي تستلزم وجود مجموعة من الأدوات والنظم للقيام بالمهام المتعلقة بإدارة البيانات البحثية ويمكن ان تتلخص هذه المراحل في الرسم التوضيحي التالي:



شكل (1) مراحل ادارة البيانات البحثية

نجد أن هناك العديد من التقسيمات لمراحل إدارة البيانات البحثية ولكن هذه التقسيمة تعتبر الأوقع والأبسط من حيث التنفيذ بالإضافة أنها مستخدمة في العديد من الجهات البحثية والأكاديمية، وهي :-

1. التخطيط
2. الجمع
3. المعالجة
4. الاحتفاظ
5. النشر
6. إعادة الاستخدام

وبالنظر إلى هذه المراحل لإدارة البيانات البحثية نجد أن هذه المراحل قد يستخدم فيها العديد من الأدوات والنظم والتي اما تعمل بشكل منفصل خارج نظم إدارة المستودعات البحثية أو أن تدخل ضمن الوظائف التي تقوم بها نظم إدارة مستودعات البيانات البحثية وسنقوم خلال هذا الفصل باستعراض دورة حياة البيانات البحثية وما تحتاجه إلى أدوات تقنية وبرمجيات سواء كانت متضمنة في نظم مستودعات البيانات البحثية او متاحة كنظم مجانية

أو تجارية منفصلة أو خدمات على شبكة الإنترنت وفقا لدورة حياة البيانات البحثية الخطوات المتعلقة بها

مراحل إدارة نظم إدارة البيانات البحثية

2/2المرحلة الأولى التخطيط والتصميم

تعد خطة إدارة البيانات من اهم مراحل اعداد البحوث العلمية لذلك ينبغي على كل باحث سيقوم بعمل بحث ومشاركة بياناته البحثية بعمل خطة إدارة بيانات البحثية بكافة تفاصيلها.

تعد مرحلة اعداد خطة البيانات البحثية وتنظيمها منذ بداية مشروعات إدارة البيانات البحثية عامل أساسي يساعد في توفير الوقت والتكلفة والمجهود، علاوة على ان المؤسسات المانحة تتطلب هذا النوع من الخطط لتقديم المنح اللازمة لإجراء البحوث.

1/2/2خطة إدارة البيانات (Data Management Plan) DMP

تعد الخطة هي مسار عمل منهجي لتحقيق هدف محدد مسبقًا، وتتضمن الخطة الخطوات والتدابير والاستراتيجيات اللازمة لإدارة دورة حياة البيانات بأكملها، وتتضمن الخطة كيفية تحديد البيانات وجمعها وإعدادها وتنظيمها وتصنيفها ومعالجتها وتحليلها وتخزينها ونشرها وتنظيمها وإعادة استخدامها (Gupta & Müller-Birn، 2018)

وتعد خطة إدارة البيانات البحثية وثيقة رسمية تحدد ما ستفعله ببياناتك أثناء وبعد المشروع البحثي، وقد يطلب هذا النوع من الخطط من العديد من وكالات تمويل البحوث، وخاصة مصادر التمويل الحكومية، وتعد خطة إدارة بيانات جزء من عمليات التقديم الخاصة للحصول على الدعم من هذه الوكالات.

وحتى إذا كنت لا تسعى للحصول على تمويل لبحثك، فإن توثيق خطة البيانات الخاصة بك هي واحدة من أهم الممارسات لتوثيق خطوات البحث وستساعد بشكل كبير على انشاء وتنظيم البيانات المستخدمة في بحثك.

وتعد خطة إدارة البيانات مستند حي قد تحتاج إلى تغييره مع تغير مسار بحثك لذلك فانه يجب عليك المراجعة الدائمة لخطة إدارة بياناتك البحثية لتتأكد من أنها لا تزال تلبي احتياجات بحثك.

2/2/2عناصر خطة ادارة البيانات DMP

وضعت وكالات التمويل، بما في ذلك المؤسسة الوطنية للعلوم (NSF) والمعاهد الوطنية للصحة (NIH) ، معايير محددة لما يجب تضمينه في خطة إدارة البيانات.

نظرًا لأن بعض وكالات التمويل لا تقدم إرشادات محددة، فيما يلي قائمة بعناصر خطة إدارة البيانات النموذجية. علاوة على أنه يجب عليك مراجعة الإرشادات المحددة لتخطيط إدارة البيانات البحثية من قبل وكالة التمويل التي تعمل معها أو تقدم لها خطتك. يمكن إعادة استخدام عناصر خطة إدارة البيانات البحثية (DMP) في البروتوكولات الخاصة بك.

3/2/2عناصر خطة إدارة البيانات البحثية

وتتضمن:-

1- العناصر الأساسية لخطة إدارة البيانات البحثية

أولا بيانات أساسية للمشروع البحثي

- عنوان البحث
- الجهة الممولة
- موضوعات البحث
- تواريخ بدأ وانتهاء البحث

ثانيا المشاركون في البحث

- المشاركون في البحث

2- خطة إدارة البيانات البحثية

مرحلة الجمع (وتهدف للإجابة عن تساؤلين ماهي البيانات التي سيتم جمعها أو انشائها- وكيف سيتم ذلك) وتتضمن هذه العناصر:

- تحديد مصدر البيانات
- تنسيقات البيانات
- أسلوب التعامل مع البيانات وكيفية تعديلها

- كم البيانات المتوقع انتاجها خلال المشروع

مرحلة المعالجة والتوثيق (وتهدف للإجابة عن تساؤل كيف سيتم عمل توثيق وبيانات وصفية للبيانات البحثية) وتتضمن:

- أسلوب التوثيق المتبع لجمع وانشاء البيانات
- طريقة توصيف البيانات البحثية وعناصر توصيفها والمعايير المستخدمة فيه

تحديد المعايير القانونية والأخلاقية المرتبطة بالمشروع (وتهدف للإجابة عن تساؤل كيف سيتم معالجة المسائل القانونية والأخلاقية) وتتضمن:

- الحصول على الموافقة على حفظ البيانات ومشاركتها
- حماية هوية المشاركين إذا لزم الأمر، على سبيل المثال عبر إخفاء الهوية
- التعامل مع البيانات الحساسة لضمان تخزينها ونقلها بشكل آمن
- تحديد من يملك حقوق ملكية البيانات
- كيف سيتم ترخيص البيانات لإعادة الاستخدام؟
- هل توجد أي قيود على إعادة استخدام بيانات الطرف الثالث؟
- هل سيتم تأجيل / تقييد مشاركة البيانات، على سبيل المثال للنشر أو السعي للحصول على براءات الاختراع؟

التخزين والنسخ الاحتياطي (وتتضمن كل ماله علاقة بالتخزين والنسخ الاحتياطي)

- التخطيط لمساحات تخزين البيانات وفي حالة الاحتياج الى تكاليف إضافية قد يتم تحميلها للباحثين
- كيفية نسخ البيانات الاحتياطية وهل ستتم بشكل ألي أم يدوي
- تحديد مسئولية عملية النسخ الاحتياطي والمهام المنوطة بها
- تحديد سيناريو استعادة البيانات الاحتياطية في حالة الكوارث
- تحديد سياسة والية النسخ الاحتياطي بعدد النسخ ومواقعها وكيفية اجراءها

مرحلة الاختيار والحفظ (وتتضمن تحديد البيانات التي يتم الاحتفاظ طويل الأمد او مشاركتها وما هي خطة الحفظ طويل الأمد)

- تحديد البيانات التي يجب الاحتفاظ بها / إتلافها لأغراض تعاقدية أو قانونية أو تنظيمية
- كيف ستقرر ما هي البيانات الأخرى التي يجب الاحتفاظ بها؟
- تحديد الاستخدامات البحثية المتوقعة للبيانات
- تحديد مدد الاحتفاظ بالبيانات والحفاظ عليها
- كيفية إعادة استخدام البيانات، على سبيل المثال للتحقق من صحة نتائج البحث أو إجراء دراسات جديدة أو للتدريس.
- تحديد مكان حفظ البيانات (مستودع أو أرشيف)
- تحديد التكاليف، إن وجدت، التي سيتحملها مستودع البيانات أو الأرشيف المحدد
- تخصيص ميزانية للوقت والجهد لإعداد البيانات للمشاركة / الحفظ؟

مشاركة البيانات (وتتضمن كيفية مشاركة البيانات وماهي القيود المفروضة على مشاركتها

- تحديد طريقة وأسلوب التعريف ونشر البيانات للمستخدمين المحتملين
- تحديد اللذين سيتم مشاركة البيانات معهم وشروط المشاركة
- هل ستشارك البيانات عبر مستودع ، أو تتعامل مع الطلبات مباشرة أو تستخدم آلية أخرى؟
- توقيت توافر البيانات
- هل سيتم حصول البيانات على معرف دائم DOI على سبيل المثال

المسئوليات والمصادر (وتتضمن تحديد مسئوليات إدارة البيانات وما هي المصادر التي سيتم اتاحتها من خلال خطتك)

- تحديد المسئول عن تنفيذ خطة إدارة البيانات وضمان مراجعتها
- تحديد مسئولية نشاط إدارة البيانات

- كيفية تقسيم المسئوليات المتعلقة بإدارة البيانات البحثية
- هل هناك احتياج الى تدريب متخصص لإدارة البيانات
- هل ستحتاج الى أجهزة او برامج إضافية لإدارة البيانات البحثية
- هل سيتم تطبيق أي رسوم من قبل مستودعات البيانات

3- مخرجات البحث

وتشمل مخرجات البحث على تسجيل البيانات النهائية المنتجة من بحثك على اختلاف أنواعها على سبيل المثال (المجموعات – الصور – المواد السمع بصرية -أوراق البيانات – المصادر التفاعلية – عناصر البيانات – الأحداث – الخدمات – المجسمات – البرامج – الصوتيات – النصوص – تدفق العمل – العناصر المادية – غيرها)

وتتضمن عنوان لكل مخرج من المخرجات وتوصيف كامل له بالإضافة الى البيانات الوصفية لكل عنصر من العناصر والتراخيص الصادر تحتها (الإصدارات المختلفة من تراخيص المشاع الإبداعي) ونمط الإتاحة (مفتوح أم مقيد أم مغلق)((**DMP tool**

4/2/2أدوات انشاء خطط إدارة البيانات البحثية

قامت العديد من الجامعات والجهات البحثية بعمل أدوات لإعداد خطة البيانات البحثية لتسهيل عملية انشائها وقد تم نشر ورقة بحثية تناولت تحليلا ومقارنة بين حوالي 14 أداة من أدوات التخطيط للبيانات البحثية وسردت ميزات المنصات المتاحة والفجوات الموجودة بها والتوصيات الخاصة بتحسينها (GAJBE; TIWARI; GOPALJI; SINGH, 2021)

وبما أن اغلب الأدوات متشابهة بشكل كبير فسنستعرض أشهر أداة لإنشاء خطط البيانات البحثية وهي DMPToolالخاصة بجامعة California Digital Library

وتعد هذه الأداة هي أداه مجانية لمساعدة الباحثين والمؤسسات في انشاء خطط إدارة بيانات عالية الجودة تلبي احتياجات ممولي البحوث وسيتم تناول هذه الأداة بالتفصيل خلال الفصل التالي والذي يتناول تطبيقات إدارة البيانات البحثية بشكل متكامل.

3/2المرحلة الثانية الجمع والإنشاء

تعد عملية جمع وانشاء البيانات واحدة من المراحل الأساسية في انشاء البيانات البحثية وتبدأ بعملية تجميع البيانات المختلفة التي تستخدم في البحث وتجميعها والتي تتحول بدورها فيما بعد إلى البيانات البحثية، وتجميع البيانات قد يتم بأكثر من الية سواء كان يدويا من خلال الباحث أو من خلال استخدام أدوات وبرمجيات لتجميع وحصاد البيانات.

وتعد البيانات المجمعة أحد اهم عناصر البيانات المستخدمة في البحوث والتي تحتاج الى عملية ضبط وتوصيف لاستخدامها في البحث او إعادة استخدامها في بحوث أخرى.

ويمكن جمع البيانات وحصادها من خلا أدوات جاهزة أو من خلال واجهة لغة التطبيقاتAPI أو من خلال مواقع متخصصة في جمع البيانات مثل

تأتي بعد ذلك عملية انشاء البيانات وتظهر الحاجة الى تقنين إجراءات إنشاء البيانات على وجه الخصوص عند اشتراك فريق عمل في انشاء البيانات البحثية، وتعد أفضل بداية لهذه المرحلة هي وضع هيكلة مناسبة وتقسيم الأدلة الرئيسةDirectories ووضع اصطلاحات لتسمية الملفات .**naming conventions** مع تحديد نوعية البيانات يوفر تحديد الهيكل التنظيمي لحفظ البيانات في وقت مبكر من العملية أساسًا تنظيميًا لجميع ملفات البيانات والمستندات الخاصة بك.((**Collect & Analyze**

وتشتمل مرحلة إنشاء البيانات بشكل تفصيلي على:

1/3/2انشاء طريقة موحدة لتسمية الملفات Naming Conventions واستخدام هذه التسميات بشكل منتظم لتحسين الوصول إلى الملفات والتسجيلات بشكل كبير

ويجب عند إنشاء طريقة لتسمية الملفات مراعاة التالي:

- تحتوي على مفردات واضحة وقياسية لأسماء الملفات، بحيث يستخدم الجميع لغة مشتركة.
- علامات الترقيم تحديد كيفية استخدام علامات الترقيم ومعناها وكيفية استخدام رموز الترقيم والأحرف الكبيرة والواصلات والمسافات ومتى يتم استخدامها.

- التواريخ: - يجب التوافق على الصيغة المستخدمة في كتابة التواريخ بحيث يتم عرضها بالترتيب الزمني، أي YYYY-MM-DD أو غيرها من الصيغ (يمكن استخدام معيار ISO في كتابة التاريخ)
- الإصدارات: - ويتم تحديد اصدار الملف في حالة ما إذا كان للملف أكثر من اصدار او تعديل يتم عليه وبالتالي يمكن العثور عليها بسهولة
- الترقيم: -يجب تحديد الأرقام التي سيتم استخدامها في الترقيم بحيث يتم سرد الملفات رقميًا، على سبيل المثال 01، 002، إلخ.

2/3/2بناء طريقة لتقسيم الملفات وهيكلتها توضح كيفية تنظيم الملفات وتنظيمها لتعظيم الإفادة منها

من أهم جوانب إدارة البيانات تنظيمها ويتضمن ذلك انشاء الهياكل والعلاقات لتنظيم البيانات، لذلك يُنصح الباحثون بهيكلة مجلداتهم (سواء كانت ورقية أو إلكترونية) لتتوافق مع كيفية إنشاء السجلات ولتكملة مهام سير العمل المقترحة أو الحالية.

تُمكِّن هياكل الملفات عمليات البحث من أن تكون أكثر دقة، وتسهل على المحققين تحديد المكان الذي يجب حفظ الملفات فيه ، وفي النهاية تجعل الاسترجاع والأرشفة أكثر كفاءة.

لذلك يجب وضع خطة لتنظيم وحفظ البيانات وهيكلتها قبل أن تبدأ حتى في جمع البيانات أو العمل معها، وهذا الأسلوب في تنظيم البيانات يسمح لفريق العمل بجمع البيانات وتحليلها

يوضح هيكل الملف أدناه إحدى الطرق التي يمكنك من خلالها التفكير في تنظيم الملفات المرتبطة بمشروع معين:

يجب تسمية المجلدات بشكل مناسب مع وضع اسم المجلدات بعد مجالات العمل التي تتعلق بها وليس بعد الباحثين أو الطلاب الفرديين، يؤدي ذلك إلى تجنب حدوث ارتباك في مساحات العمل المشتركة إذا غادر أحد أعضاء فريق العمل، كما يسهل التنقل في نظام الملفات بحثًا عن أشخاص جدد ينضمون إلى مساحة العمل:

- هيكلة المجلدات بشكل هرمي - ابدأ بعدد محدود من المجلدات للموضوعات الأوسع، ثم قم بإنشاء مجلدات أكثر تحديدًا داخل هذه المجلدات

- افصل بين العمل الجاري والمكتمل: - عندما تبدأ في إنشاء الكثير من المجلدات والملفات، من الجيد أن تبدأ في التفكير في فصل المستندات القديمة عن تلك التي تعمل عليها حاليًا
- حاول الاحتفاظ بمجلد "المستندات" الخاص بك للملفات التي تعمل عليها بنشاط، وكل شهر أو نحو ذلك، انقل الملفات التي لم تعد تعمل عليها إلى مجلد أو موقع مختلف، مثل مجلد على سطح المكتب، أو أرشيف خاص مجلد أو محرك أقراص ثابت خارجي
- النسخ الاحتياطي: - تأكد من نسخ ملفاتك احتياطيًا، سواء كانت موجودة على محرك الأقراص المحلي أو محرك أقراص الشبكة
- مراجعة السجلات - قم بتقييم المواد بانتظام أو في نهاية المشروع لضمان عدم الاحتفاظ بالملفات دون داع. ضع تذكيرًا في التقويم الخاص بك حتى لا تنسى.

3/3/2استخدام أسلوب الإصدارات مع تنظيم وحفظ الملفات versioning لمتابعة التغييرات والتحديثات التي تتم على كل ملف من الملفات

يُعد التحكم في الإصدارات هو طريقة مستخدمة لتتبع تغييرات التي تتم على عناصر البيانات بمرور الوقت بحيث يمكنك استدعاء الإصدارات القديمة في وقت لاحق.

```
A) Organized by File type

Dataset.A
  |- Code
  | |- Step.1
  | |- Step.2
  |- Data
  | |- Processed
  | |- Raw
  |- Results/
  | |- Figure.1
  | |- Figure.2
  | |- Models
  | readme.txt
```

```
B) Organized by Analysis

Dataset.B
  |- Figure.1
  | |- Code
  | |- Data
  | |- Results
  |- Figure.2
  | |- Code
  | |- Data
  | |- Results
  |- Table.1
  | |- Code
  | |- Data
  | |- Results
  | readme.txt
```

(شكل2 طريقة تخزين الملفات الهرمية)

يمكن أن يكون تعيين إصدارات الملفات بسيطًا مثل استخدام اصطلاحات تسمية الملفات مثل استخدام لواحق في التسمية برقم الاصدارvn * ، v2 * ، v1 ، أو يمكنك استخدام برنامج التحكم في الإصدار (version control software(VCS.

وتسمح برامج التحكم في الإصدارات لعدة أشخاص في الفريق بالعمل معًا في نفس المشروع في نفس الوقت. يدير هذا البرنامج التغييرات في جميع أنواع الملفات النصية، مثل البرامج النصية وصفحات الويب بالإضافة إلى بعض التنسيقات الرقمية الخاصة، ويمكن من هلاله تسجيل الأنشطة المختلفة التي تتم (الإضافات والحذف والاستبدال) للملفات الفردية ، ويتتبع التحديثات ، ويسمح بتفرع المشروعات التي قد يتم دمجها لاحقًا في المشروع الأصلي.

ويتميز نظام الإصدارات versioning **بالمميزات التالية:**

التراجع اللانهائي: يسمح للمستخدمين بالعودة إلى إصدار سابق من ملف أو ملف تم تعيينه لاستعادة أو استعادة الحالة السابقة. يسمح لك هذا باستعادة الملفات المحذوفة أو التي تم استبدالها عن طريق الخطأ في مشروع، أو استعادة مجموعات الملفات من نسخة سابقة محفوظة.

التفريع والتجريب: هذا أمر قوي لأنه يسمح لك باختبار ميزات جديدة في كود البرمجة أو التفرع في مسار مختلف دون التأثير على عمل المتعاون في كود الإنتاج. لاحقًا، يمكن دمج فرعك في الإنتاج أو حفظه أو إهماله حسب الضرورة.

التعاون: يمكن للمتعاونين العمل محليًا في الملف بينما يتولى برنامج إدارة الاصدارات مهام دمج التغييرات والحفاظ على مزامنة الملفات وتفريعات الدليل. بالإضافة الى انه يساعد على تتبع من قام بالتغييرات ومتى، لذلك يمكن متابعة أي أسئلة حول التغييرات في الملفات.

- الأدوات والبرمجيات المستخدمة في إدارة الإصدارات
  - Git - GitHub – GitLab -Bitbucket – Subversion
- الأدوات المستخدمة في مشاركة الملفات
  - Dropbox – one drive – google drive – open science frame work

- إنشاء ملفات وصفية Readmefile.txt تقدم وصفا لطريقة التنظيم والتسمية وكافة الإرشادات للمسئولين عن توصيف البيانات وحفظها

4/3/2التوثيق والبيانات الوصفية

ما هو "التوثيق" و"البيانات الوصفية"

يشمل مصطلح "**التوثيق**" جميع المعلومات اللازمة لتفسير وفهم واستخدام مجموعة بيانات معينة أو مجموعة من المستندات. وعادة ما تكون البيانات الوصفية مضمنة في ملف البيانات ونستخدم ملفات للتوثيق بإرفاقها مع ملفات البيانات لتسجيل كل ما له علاقة بإنشاء مجموعات البيانات.

1/4/3/2أولا البيانات الوصفية

هي معلومات مهيكلة تصف وتشرح وتحدد وتجعل من السهل الوصول الى مصادر المعلومات او عناصر البيانات، ودائما ما يطلق عليها البيانات عن البيانات، وتستخدم للتأكيد سياق انشاء البيانات وكيفية تحليلها وتخزينها _National information standards 2004 organization

ويعتمد توصيف البيانات على استخدام البيانات الوصفية المقننة لتسهيل الوصول الى عناصر البيانات وتسهيل مشاركتها وتعد البيانات الوصفية " مجموعة من المعلومات المهيكلة التي تصف أو تشرح أو تحدد موقع أو تسهل استرجاع أو استخدام أو إدارة مصدر معلومات. غالبًا ما تسمى البيانات الوصفية بيانات حول البيانات أو المعلومات حول المعلومات ". إنه يضمن أن يكون السياق الخاص بكيفية إنشاء بياناتك وتحليلها وتخزينها واضحًا ومفصلاً وبالتالي يمكن استنساخه. - الهيئة الوطنية لمعايير المعلومات 2004

البيانات الوصفية الجيدة هي التي تمكنك من فهم واستخدام ومشاركة بياناتك الخاصة حاليا وفي المستقبل وتساعد الباحثين الاخرين على اكتشافها والوصول اليها واستخدامها في بحوث جديدة، والاستشهاد بها وتمكن أيضا من الارشفة طويلة الأمد وحفظ البيانات.

وتنقسم البيانات الوصفية إلى:

نوع البيانات الوصفية	الغرض منها	أمثلة على عناصر البيانات	الإستخدام الأساسي
البيانات الوصفية	لإيجاد وفهم المصادر	العنوان المؤلف الموضوع المنشئ تاريخ النشر	الإكتشاف العرض التكامل
البيانات الوصفية الإدارية			
البيانات الوصفية الفنية	لفك تشفير البيانات وتقديمها	نوع الملف حجم الملف تاريخ ووقت الانشاء أسلوب الضغط	التكامل إدارة حفظ العناصرالرقمية
البيانات الوصفية الخاصة بالحفظ	لإدارة الملفات على المدي الطويل	Checksum Preservation event	التكامل إدارة حفظ العناصرالرقمية
البيانات الوصفية الخاصة بالملكية الفكرية	لتحديد حقوق الملكية الفكرة المرفقة بالمحتوى	حالة الملكية الفكرية شروط الترخيص حاملي حقوق الملكية	التكامل إدارة العناصر الرقمية التكامل
البيانات الوصفية الهيكلية	وتحدد العلاقات بين البيانات ومصادر المعلومات وبعضها البعض	Sequence Place in hierarchy	الإبحار والتصفح
لغات الترميز Markup languages	وتقوم بدمج البيانات الوصفية وادراج ميزات الربط الدلالي داخل المحتوي	Paragraph Heading List Name Date	الإبحار والتصفح التكامل

هناك العديد من الأمثلة على البيانات الوصفية:

البيانات الوصفية الخاصة بعناصر البيانات نفسها والتي تختلف باختلاف نوع العنصر مثلا ان كان عنصر البيانات هو عينات سريرية للكشف عن الأمراض فتكون البيانات (اعداد الخلايا والأجسام المضادة والكواشف الكيميائية المستخدمة).

البيانات الوصفية الفنية: وهي المعلومات التي يتم انشاءها تلقائيا بواسطة أدوات البحث والبرامج المرتبطة بها مثل تاريخ الإنشاء ونوع العنصر وغيرها.

البيانات الوصفية التجريبية : وتحتوي على معلومات عن ظروف اجراء التجارب (مثل نوع الفحص، والمحددات الزمنية، والبروتوكول التجريبي، والمعدات المستخدمة لإنشاء البيانات، وطرق تحليل البيانات، وتفاصيل المخرجات.

البيانات الوصفية على مستوي مجموعة البيانات: وتحتوي على اهداف المشروع البحثي والمحققين والمشاركين والمنشورات ذات الصلة ومصادر التمويل.

2/4/3/2التوثيق Documentation

يهدف توثيق البيانات إلى وصف البيانات التي تم جمعها لتسهيل استخدامها واسترجاعها وإدارتها. يأخذ توثيق البيانات أشكالًا مختلفة ويصف البيانات على مستويات متعددة.

وصف مجموعة البيانات وكائنات البيانات باسم البيانات الوصفية، أي بيانات حول البيانات. تتمثل إحدى طرق إضافة البيانات الوصفية في إرفاق ملف تمهيدي ببياناتك. تقدم ResearchData NL إرشادات لهذا الغرض. قدمت CESSDA إرشادات مفصلة للغاية متاحة لإنشاء الوثائق والبيانات الوصفية لبياناتك.

بالإضافة إلى وصف مجموعات البيانات والأشياء الخاصة بهم ، يمكن للباحثين الرجوع إلى مقترح المشروع حيث يمكن للباحثين الآخرين العثور على معلومات حول البحث ، على سبيل المثال الأهداف والأهداف والمنهجية وجمع البيانات والأشخاص المسؤولين عن المشروع وما إلى ذلك. يؤثر نوع البحث وطبيعة البيانات أيضًا على نوع التوثيق الضروري.

تخضع الأنواع المختلفة من البيانات لمعايير مختلفة، ويجب أخذها في الاعتبار عند توثيق البيانات. تشمل هذه المتطلبات، على سبيل المثال لا الحصر:

- مبادئ بيانات FAIR: مجموعة المبادئ (يمكن العثور عليها ، ويمكن الوصول إليها ، وقابلة للتشغيل البيني ، وقابلة لإعادة الاستخدام) لتبادل البيانات.
- معايير البيانات الوصفية النوعية: مبادئ توجيهية لتوثيق البيانات. يمكن أن يشير هذا إلى وثائق مجموعة البيانات أو وصف الكائن أو كليهما. يمكن لمعايير البيانات الوصفية النوعية توثيق مجموعة البيانات ككل أو ككائن بيانات (انظر الرقم 5).
- وثائق المشروع: وصف المشروع الذي يتضمن جمع البيانات. غالبًا ما تستخدم هذه الوثائق للتحقق من البحث ومنشأه.
- مجموعة البيانات الوصفية: وصف مجموعة البيانات ، وغالبًا ما تُستخدم لاكتشاف مجموعات البيانات داخل المستودع.
- البيانات الوصفية لكائن البيانات: تعريف اسم كائن البيانات ، غالبًا ما يتم إعداده بواسطة الباحث لهيكلة البيانات أو بواسطة مجموعة البحث للتعاون أثناء المشروع.

https://libguides.vu.nl/rdm/data-documentation

3/4/3/2الوثائق الداعمة README file

تعد ملفات README طريقة شائعة لتوثيق المحتويات وهيكلة المجلدات أو مجموعة البيانات والتي تمكن الباحث من تحديد المعلومات التي يحتاجها. يمكن الاحتفاظ بتوثيق البيانات في مجموعة متنوعة من الأشكال.

ويجب عند انشاء هذه الملفات أن تحتوي على وصفا واضحا وموجزا لجميع التفاصيل المرتبطة بعملية جمع البيانات ومعالجتها وتحليلها، وهذه الملفات ستساعد الباحثين الأخرين في تفسير مجموعة البيانات الخاصة بك وإعادة تحليلها مرة أخرى.

ومن أهم استخدامات هذه الملفات

- توثيق التغييرات على الملفات أو أسماء الملفات داخل المجلدات.
- شرح اصطلاحات تسمية الملفات والممارسات الخاصة بذلك.
- تكون مرافقة لعناصر البيانات المخزنة.

من أفضل الممارسات إنشاء ملفات داعمة README لكل مجموعة بيانات بغض النظر عما إذا كان يتم إيداعها في مستودع لأن المستند قد يصبح ضروريًا لاحقًا.

من الممارسات الجيدة للبيانات تخزين ملف readme.txt مع كل مجموعة بيانات مميزة تشرح اصطلاح تسمية الملفات إلى جانب أي اختصارات أو أكواد استخدمتها.

يفضل أن يكون الملف مكتوب باستخدام نص عادي ، وتجنب التنسيقات الاحتكارية ، مثل Microsoft Word ، كلما أمكن ذلك. ومع ذلك ، فإن PDF مقبول عندما يكون التنسيق مهمًا.

إذا قمت بإيداع مجموعات البيانات النهائية في مستودع بيانات ، فقد يطلب منك المستودع تقديم ملف README مع تفاصيل إضافية حول مجموعات البيانات الخاصة بك ، مثل المعلومات المنهجية أو معلومات المشاركة / الوصول. سيساعدك إنشاء ملف README في بداية عملية البحث الخاصة بك ، وتحديثه باستمرار خلال بحثك ، على تجميع ملف README نهائي عندما تكون بياناتك جاهزة للإيداع.

قد يحتوي أيضا هذا النوع من الملفات على شروح أو تعليمات حول السرية واستخدام البيانات أو إعادة استخدامها. تتضمن أمثلة الوثائق الداعمة ما يلي:

- أوراق عمل أو كتب معملية
- استبيانات أو أدلة مقابلة
- تقارير ومنشورات المشروع النهائية
- البيانات الوصفية للكتالوج

يجب تنظيم الوثائق الداعمة، بحيث يمكن استخدامها لتحديد البيانات وتحديد موقعها عبر متصفح الويب أو الكتالوج المستند إلى الويب. عادةً ما يتم تنظيم البيانات الوصفية للفهرس وفقًا لمعيار دولي وترتبط بالبيانات من خلال المستودعات أو مراكز البيانات عند إيداع المواد.

يوفر مركز التنظيم الرقمي أمثلة للبيانات الوصفية الخاصة بالتخصصات، والتي يمكن عرضها هنا.

4/2المرحلة الثالثة مرحلة المعالجة والتحليل

تعد مرحلة المعالجة وتحليل البيانات هي مرحلة فحص وتنقية وتحويل ونمذجة البيانات بهدف اكتشاف معلومات مفيدة واستخدامها في دعم اتخاذ القرار ومن ثم انشاء وحدات بيانات Data sets يمكننا استخدامها في اجراء البحوث العلمية، ويمكن أن تساهم الاختيارات التي تقوم بها أثناء تحليل بياناتك أيضًا في إدارة بيانات البحث الخاصة بك بشكل فعال:

وتتضمن هذه المرحلة:

1/4/2التحقق من الصحة والاكتمال Data Provenance and data Lineage

يشير البرهان على البيانات الى معرفة مصدر انشاء البيانات، وأين تم تجميعها، وبواسطة من، ولأي سبب، والجوانب المماثلة التي تساعد في فهم كيفية جمع البيانات ومعالجتها وتعديلها في الأصل ".

إن توثيق برهنة البيانات وتعديل سجلاتها هو عنصر أساسي للتأكيد على الممارسات العلمية الجيدة وتمكين إعادة إنتاج النتائج

توفر معظم تطبيقات تنظيف البيانات أو معالجات الملفات المجدولة أو أدوات التحليل الإحصائي طريقة آلية لتتبع التغييرات على بياناتك. يمكن لبرمجيات مثل Jupyter تصدير هذه البيانات أيضًا. إذا قررت استخدام OSF ، فيمكنه تنسيق التغييرات وتوثيقها. هناك تنسيقات بيانات وصفية قياسية لتخزين مصدر البيانات مثل DDI.

معيار(مبادرة توثيق البيانات) DDI

هو معيار دولي لوصف البيانات التي تنتجها (على سبيل المثال لا الحصر) المسوحات وطرق المراقبة الأخرى في العلوم الاجتماعية والسلوكية والاقتصادية والعلوم الصحية والإحصاءات الرسمية هو معيار مجاني يستخدم لتوثيق وإدارة المراحل المختلفة في دورة حياة بيانات البحث، مثل وضع المفاهيم، وجمع البيانات، ومعالجتها، والتوزيع / النشر، والاكتشاف، والأرشفة.

توثيق البيانات باستخدام DDI يسهل الفهم والتفسير والاستخدام - من قبل الناس وأنظمة البرامج وشبكات الكمبيوتر.

ويستخدم معيار DDI لأغراض (التوثيق والاكتشاف والتكامل).

ويقدم المعيار:

- هيكل لوصف بيانات البحث والمعلومات المتعلقة بها.
- معيار مفتوح مصمم لدعم مشاركة البيانات وفهمها وإعادة استخدامها.
- آلية لإنشاء وثائق متسقة للبيانات قابلة للتنفيذ.

يستخدم معيار DDIلدعم البحث "نص كامل" عن البيانات الخاصة بك وعمل فهرس كامل لبياناتك ليس فقط البيانات الوصفية.

أسماء الملفات والعناصر الأخرى للبيانات الوصفية غالبا تكون غير كافية تم العثور على العديد من الموضوعات من ضمن البيانات نفسها.

يستخدم أيضا المعيار لمساعدة المستخدمين على تقييم البيانات قبل الوصول إليها بالإضافة إلى تقديم الوثائق القياسية(**DDI : learn**).

عمل وصف متكامل للبيانات لاستخدام (دفتر الرموز ، قاموس البيانات).

إنتاج بنوك الأسئلة وقوائم المتغيرات لاكتشاف البيانات وإعادة استخدامها.

وينقسم المعيار لأكثر من نوع

كتاب الترميز **DDI (DDI - C)**

نسخة خفيفة الوزن من المعيار تهدف إلى توثيق بيانات المسح البسيطة.

دورة حياة **DDI (DDI - L)**

مصمم لتوثيق البيانات وإدارتها عبر دورة الحياة بأكملها من التصور إلى التحليل.

DDI عبر المجال التكامل **(DDI - CDI)**

المواصفات المستندة إلى النموذج والتي يمكن إدارتها وتوسيعها والتعبير عنها بسهولة أكبر في تمثيلات مختلفة.(وهذا المعيار لا يزال قيد المراجعة ولم يصدر)

يتكامل معيار **DDI** مع العديد من المعايير الدولية مثل:

- Dublin Core and MARC
- Generic Statistical Information Model (GSIM)

- ISO/IEC 11179 (Geography)
- ISO 19118 (Geography)
- ISO 17369 SDMX – Statistical Data and Metadata Exchange
- METS and PREMIS
- Structured Data Transformation Language (SDTL)
- Validation and Transformation Language (VTL)
- Others

يمكننا استخدام معيارDDI في:

- إنشاء كتب الشفرات التفاعلية interactive codebooks
- تنفيذ كتالوجات البيانات والبوابات
- نقل البيانات بين حزم التحليل
- بناء بنوك للأسئلة question banks
- إنشاء تعيينات التوافق concordance mappings
- تنسيق البيانات ومقارنتها
- إدارة مجموعات البيانات الطولية longitudinal data sets

لكل خطوة من خطوات تحليل البيانات الخاصة بك، من الضروري توثيق البيانات الجيدة.(**Data Provenance**)

2/4/2تنظيف للبيانات Data cleansing

وتتضمن هذه العملية إصلاح أو حذف البيانات الغير صحيحة أو المعطوبة أو الموضوعة في شكل غير مناسب أو غير مكتملة داخل مجموعة البيانات dataset. عند تجميع أكثر من مصدر للبيانات قد يحدث تكرار أيضا للبيانات الذي يجعل من الضروري مراجعته وحذفه.

إذا كانت البيانات غير صحيحة، فإن النتائج والخوارزميات لن يمكنك الاعتماد عليها، على الرغم من أنها قد تبدو صحيحة. لا توجد طريقة واحدة مطلقة لوصف الخطوات الدقيقة في

عملية تنظيف البيانات لأن العمليات ستختلف من مجموعة بيانات إلى مجموعة بيانات. ولكن من الأهمية بمكان إنشاء نموذج لعملية تنظيف البيانات الخاصة بك حتى تعرف أنك تقوم بذلك بالطريقة الصحيحة في كل مرة.

خطوات عملية تنظيف البيانات

الخطوة الأولى: إزالة البيانات المكررة أو غير ذات الصلة

قم بإزالة البيانات غير المرغوب فيها من مجموعة البيانات الخاصة بك، بما في ذلك الملاحظات المكررة أو الملاحظات غير ذات الصلة. سيتم تكرار البيانات في أغلب الأحيان أثناء عملية جمع البيانات. عندما تقوم بدمج مجموعات البيانات من أماكن متعددة، أو تتلقى بيانات من عملاء أو أقسام متعددة، فهناك فرص لإنشاء بيانات مكررة. يعد إلغاء التكرار أحد أهم النقاط التي يجب مراعاتها في هذه العملية.

البيانات غير ذات الصلة أيضا هي البيانات الغير مرتبطة بالموضوع الذي يجري عليه البحث او مرتبطة. على سبيل المثال، إذا كنت ترغب في تحليل البيانات المتعلقة بالعملاء من جيل الألفية، لكن مجموعة البيانات الخاصة بك تتضمن الأجيال الأكبر سناً، يمكنك إزالة تلك البيانات غير ذات الصلة. يمكن أن يجعل هذا التحليل أكثر كفاءة ويقلل من تشتيت الانتباه عن هدفك الأساسي - بالإضافة إلى إنشاء مجموعة بيانات أكثر قابلية للإدارة وأكثر أداءً.

الخطوة الثانية: إصلاح الأخطاء الهيكلية

تحدث الأخطاء الهيكلية عندما تقوم بقياس البيانات أو نقلها وتلاحظ اصطلاحات تسمية غريبة أو أخطاء مطبعية أو أحرف كبيرة غير صحيحة. يمكن أن تتسبب هذه التناقضات في تسمية فئات أو فئات غير صحيحة. على سبيل المثال ، قد تجد كلاً من "لا ينطبق" و "غير قابل للتطبيق" يظهران ، ولكن يجب تحليلهما على أنهما نفس الفئة.

الخطوة الثالثة: تصفية القيم المتطرفة غير المرغوب فيها

في كثير من الأحيان، ستكون هناك ملاحظات لمرة واحدة حيث لا يبدو أنها تتناسب مع البيانات التي تقوم بتحليلها. إذا كان لديك سبب منطقي لإزالة هذا البيان ، مثل إدخال بيانات غير لائق، فإن القيام بذلك سيساعد في تحسين عملية تحليل البيانات.

الخطوة الرابعة: معالجة البيانات المفقودة

لا يمكنك تجاهل البيانات المفقودة لأن العديد من الخوارزميات لن تقبل القيم المفقودة. هناك طريقتان للتعامل مع البيانات المفقودة. كلاهما ليس هو الأمثل، ولكن يمكن النظر في كليهما.

- كخيار أول ، يمكنك إسقاط الملاحظات التي تحتوي على قيم مفقودة ، لكن القيام بذلك سيؤدي إلى إسقاط المعلومات أو فقدها ، لذا كن على دراية بهذا قبل إزالته.
- كخيار ثانٍ ، يمكنك إدخال القيم المفقودة بناءً على ملاحظات أخرى ؛ مرة أخرى ، هناك فرصة لفقد سلامة البيانات لأنك قد تعمل من الافتراضات وليس الملاحظات الفعلية.
- كخيار ثالث ، يمكنك تغيير طريقة استخدام البيانات للتنقل الفعال بين القيم الخالية.

الخطوة الخامسة: التحقق من الصحة وضمان الجودة

- في نهاية عملية تنظيف البيانات، يجب أن تكون قادرًا على الإجابة على هذه الأسئلة كجزء من عملية التحقق الأساسية:
  - هل البيانات منطقية؟
  - هل تتبع البيانات القواعد المناسبة لمجالها؟
  - هل يثبت أو يدحض نظرية العمل الخاصة بك، أو يسلط الضوء على أي بصيرة؟
  - هل يمكنك العثور على اتجاهات في البيانات لمساعدتك في تكوين نظريتك التالية؟
  - إذا لم يكن كذلك ، فهل هذا بسبب مشكلة جودة البيانات؟

خصائص البيانات الجيدة

- الصلاحية. الدرجة التي تتوافق بها بياناتك مع قواعد أو قيود العمل المحددة.
- الصحة. تأكد من أن بياناتك قريبة من القيم الحقيقية.
- الاكتمال. درجة معرفة جميع البيانات المطلوبة.
- التناسق. تأكد من أن بياناتك متسقة في نفس مجموعة البيانات و / أو عبر مجموعات بيانات متعددة.

- التوحيد. الدرجة التي يتم بها تحديد البيانات باستخدام نفس وحدة القياس.
- مزايا وفوائد تنظيف البيانات
- سيؤدي وجود بيانات نظيفة في النهاية إلى زيادة الإنتاجية الإجمالية والسماح بأعلى جودة للمعلومات في عملية اتخاذ القرار. الفوائد تشمل:
- إزالة الأخطاء عند وجود مصادر متعددة للبيانات.
- تحدث أخطاء أقل للعملاء الأكثر سعادة والموظفين الأقل إحباطًا.
- القدرة على تعيين الوظائف المختلفة وما تهدف بياناتك إلى القيام به.
- مراقبة الأخطاء وإعداد التقارير بشكل أفضل لمعرفة مصدر الأخطاء، مما يسهل إصلاح البيانات غير الصحيحة أو الفاسدة للتطبيقات المستقبلية.
- سيؤدي استخدام أدوات تنظيف البيانات إلى زيادة كفاءة ممارسات الأعمال واتخاذ القرارات بشكل أسرع.
- هناك العديد من البرمجيات والأدوات التي تقوم بعمليات تنظيف البيانات مثل tableau وغيرها من الأدوات.

1- تحليل عناصر البيانات الجاهزة

إن تحليل عناصر البيانات الجاهزة التي تم جمعها ومراجعتها لإنتاج نتائج واضحة ومتسقة وخالية من الأخطاء إلى أقصى حد ممكن، عند العمل في مشروع بحثي، يجب اتخاذ خطوات للتأكد من أن بياناتك آمنة وحقيقية وقابلة للاستخدام.

نظرًا لأن البيانات غالبًا ما تكون فوضوية، مع إدارة البيانات، فإننا نهدف إلى تنظيفها قبل تحليلها. فيما يلي مفاهيم وخطوات لإعداد مجموعات البيانات الجاهزة للتحليل، يقال أن أكثر من 80% من الوقت المنقضي في عمليات تحليل البيانات يكون في التنظيف والتجهيز لهذه البيانات.

على الرغم من أن تحليل البيانات هو عملية مستمرة طوال فترة المشروع البحثي، إلا أننا هنا سنركز على تحليل البيانات بعد جمعها. للتأكد من أن البحث تجريبي وقابل للتحقق، من

الضروري أن يحتفظ الباحثون بسجلات (توثيق البيانات) لكل خطوة يتم إجراؤها أثناء تحليل البيانات.

غالبًا ما يتطلب تحليل البيانات استخدام برامج متخصصة، وتتضمن البرامج التي تقدمها الجامعة وترخصها حاليًا: Stata و SPSS و Atlas.TI. بعض البرامج متاحة للتنزيل على: download.vu.nl. للبرامج المفتوحة ، انظر أدناه.

في بعض الحالات ، يكتب الباحثون نصوصهم الخاصة لتحليل البيانات. في VU ، تتم كتابة معظم البرامج النصية بلغات R و Python و SQL.

إذا كنت ترغب في قراءة تحليل البيانات، فيجب عليك التحقق من مقالات المجلات والكتب التي توفرها مكتبة VU حول هذا الموضوع:

جميع المصادر: تحليل البيانات

- تحليل البيانات الكمية
- تحليل البيانات النوعية
- البيانات الكبيرة
- بيانات التعدين(data-analysis)

للقيام بعملية تحليل بيانات دقيقة يجب مراعاة العديد من النقاط أثناء دورة حياة البيانات البحثية وهي:

3/4/2إنشاء البيانات

يفضل انشاء البيانات في شكل جداول ثنائية البعد لتخزين وعرض وتحليل والتعديل على البيانات.

الممارسات الجيدة لإنشاء عناصر البيانات

1- عنونة جميع البيانات
2- يجب منح كل موضوع تجريبي معرف فريد للدراسة
3- يجب وضع البيانات في شكل ملفات مسطحة flat files
4- يجب أن تمثل الصفوف وحدة التحليل المناسبة

5- يجب أن تمثل الأعمدة السمات الفريدة للصفوف.

6- يجب أن تحتوي ملفات البيانات على نفس عدد الأعمدة في كل صف. تنشأ المشاكل عندما تكون البيانات مفقودة في منتصف الصف.

7- لا ينبغي دمج البيانات المنفصلة داخل عمود واحد

4/4/2تشكيل البيانات formatting

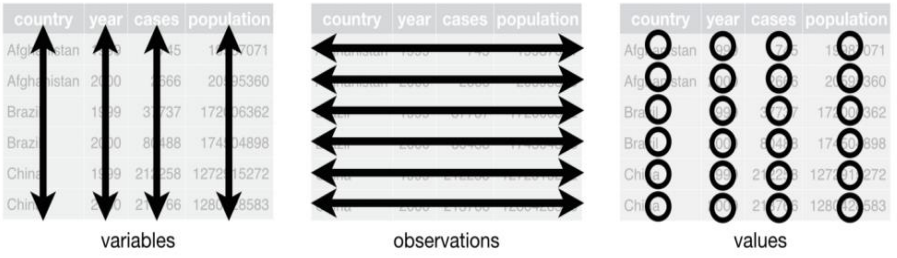
أصبحت Tidy Data على وجه التحديد التنسيق القياسي للعلوم لأنها تتيح للأشخاص بسهولة تحويل جدول البيانات إلى رسوم بيانية وتحليلات ورؤى. صاغ الدكتور هادلي ويكهام ، كبير العلماء في RStudio وأستاذ الإحصاء المساعد في جامعة أوكلاند وستانفورد وجامعة رايس، مصطلح "البيانات المرتبة" لتقليل الجهد المبذول عند إعداد البيانات للتصور والنمذجة الإحصائية. تحتوي "مجموعة البيانات المرتبة" على البنية الافتراضية:

1. يشكل كل متغير عمودًا

2. تشكل كل ملاحظة صفًا

3. تحتوي كل مجموعة بيانات على معلومات عن وحدة تحليل واحدة فقط يمكن تتبعها (مثل العائلات، والمشاركين، وزيارات المشاركين)

بيانات مرتبة من علوم البيانات مع R بواسطة Garrett Grolemund. يتم وضع كل متغير في العمود الخاص به ، وكل ملاحظة في صفها الخاص ، وكل قيمة في الخلية الخاصة بها.



(شكل3 أساليب تحليل البيانات)

5/4/2التحقق validation

يساعد التحقق من صحة البيانات على ضمان جمع البيانات بشكل صحيح.

أفضل الممارسات للتحقق من صحة مجموعات البيانات الخاصة بك:

1. برمجة نطاقات محددة لإدخال البيانات في الحقول عند الإمكانية لذلك(باستخدام القيم المسجلة سابقا وتقليل الكتابة اليدوية).
2. بتطبيق تنسيق البيانات على الحقول مسبقًا لمنع مخاطر التنسيق "التلقائي" غير الدقيق.
3. منع إدخال المسافات البادئة و / أو اللاحقة أو الأحرف الأخرى التي قد تتداخل مع تحليل البيانات.
4. تضمين اجابات (بأخرى) في حالة عدم تضمين الإجابة من الخيارات المتاحة.
5. تضمين قيمة "أفضل عدم الإجابة".

/6/4/2المعيارية standardization

يضمن اتباع المعايير اتساق البيانات ويضمن أن تكون البيانات من نفس النوع والتنسيق لكل عنصر بيانات تقوم بجمعه. كما أنه يساعد في تقليل أخطاء جمع البيانات وتحليلها ويمنع التناقضات.

أفضل الممارسات توحيد البيانات:

1. تكويد البيانات بشكل متناغم ومتناسق
2. توحيد النص الحر في بيانات مقسمة ومحددة
3. تحديد نمط محدد لتسجيل التاريخ والوقت مع تثبيت استخدام هذا النمط طوال الوقت على سبيل المثال ISO 8601 YYYY-MM-DD)

7/4/2التنظيف cleaning

قبل إجراء تحليل لبياناتك، راجع مجموعات البيانات بحثًا عن عدم الدقة أو التناقضات أو البيانات الحساسة. يسمح لك تنظيف بياناتك بتحديد القيم المتطرفة أو الأخطاء قبل تجميع نتائجك.

أفضل الممارسات لتنظيف بياناتك:

1. تحقق من القيم المتطرفة. تأكد من أن جميع عناصر البيانات في التنسيقات والنطاقات الصحيحة.

2. تحقق من البيانات المفقودة. تأكد من عدم وجود عناصر بيانات أو سجلات مفقودة، مما يؤدي إلى إنشاء عناصر خالية. كود ينقص البيانات بشكل مناسب.

3. تأكد من أن بياناتك لا تحتوي على معلومات صحية محمية (PHI). تتطلب HIPAA من الباحثين حماية خصوصية وسرية مرضاهم. يجب عدم تضمين أي معلومات صحية يمكن تحديدها بشكل فردي في مجموعات البيانات الخاصة بك.

1. يقوم برنامج Redcap تلقائيًا بوضع علامة على بيانات التعريف الشخصية.

2. راجع قائمة العناصر المدرجة في المعلومات الصحية المحمية.

8/4/2التوثيق documentation

قبل تحليل بياناتك أو مشاركتها، تأكد من أن لديك الوثائق المناسبة. يسهل التوثيق المناسب فهم البيانات الخاصة بك وتحليلها ومشاركتها وإعادة استخدامها.

أفضل الممارسات لتوثيق بياناتك:

1. يجب تخزين البيانات مع البيانات الوصفية المناسبة.

2. إنشاء واستخدام قاموس البيانات وملفات README.

3. حفظ البيانات كملف ASCII أو UNICODE يمكن قراءته آليًا.

4. اعتماد ممارسات تسمية الملفات المناسبة لاستيعاب إصدارات متعددة من ملفات البيانات(**analysis-ready-datasets**).

كل هذه الممارسات السابقة تنتج لنا بيانات جيدة تصلح لعملية التحليل بشكل كبير ويمكن إدخالها لبرمجيات تحليل البيانات للحصول على المعلومات المطلوبة.

9/4/2برمجيات تحليل البيانات

بيئة التطوير المتكاملةRStudio

وهي منصة تحتوي على وحدة تحكم وتمييز للاكواد البرمجية بالإضافة إلى أدوات التخطيط والتاريخ وتصحيح الأخطاء وإدارة مساحة العمل.

برنامج التحليل الإحصائيSAS

مجموعة البرامج الإحصائية لإدارة البيانات والتحليلات المتقدمة والتحليل متعدد المتغيرات والتحليلات التنبؤية

برنامج وخدمة الإحصاءات SPSS

حزمة البرامج المستخدمة للتحليل الإحصائي التفاعلي أو المجمّع

أداة تحليل البيانات Tableau

يسمح لك بتقديم بياناتك في مجموعة واسعة من التنسيقات المرئية ، والتي يمكن أن توفر غالبًا عرض رسومي لا يكون واضحا باستخدام الجداول التقليدية.

أداة التحليل الكيفية NVivo

هو برنامج برمجي على المستوى المهني للبحث النوعي يستخدم للتعليق التوضيحي للصور (الخرائط ، الصور ، الرسوم البيانية) وتحليل المقابلة (الفيديو ، الصوت ، النص ، الملاحظات الميدانية) وهو متاح بشكل مجاني.

أداة Electronic Lab Notebook

ELN)) هي أداة برمجية تقوم في أبسط أشكالها بتكرار واجهة تشبه إلى حد كبير صفحة في دفتر ورقي. في ، يمكنك إدخال البروتوكولات والملاحظات والبيانات الأخرى باستخدام جهاز الكمبيوتر أو الجهاز المحمول. تقدم ELNs العديد من المزايا مقارنة بأجهزة الكمبيوتر المحمولة التقليدية ؛ وبذلك تسهل ممارسات إدارة البيانات الجيدة ، وتوفر أمان البيانات والتدقيق والتعاون. يمكن لبعض ELNs أيضًا إدارة قوائم جرد العينات والكواشف والإمدادات الأخرى ، بالإضافة إلى تتبع جداول صيانة المعدات والمعدات. بالإضافة إلى ذلك ، توفر بعض ELNs أدوات علمية متخصصة للرسم الكيميائي أو البيولوجيا الجزيئية.

إدارة بيانات الصور Managing Image Data

أدى نمو جمع الصور البحثية عبر العلوم ، إلى جانب الأدوات والأجهزة القوية بشكل متزايد لالتقاط الصور ، إلى خلق تحديات وفرص جديدة لإدارة الصور عبر دورة حياة بيانات البحث. يمكن جمع الصور بعدة طرق مختلفة ، مثل المسح الضوئي الداخلي أو التصوير الفوتوغرافي أو الإنشاء الرقمي أو الشراء من مصادر خارجية. تمامًا مثل أي عملية لجمع البيانات ، لكي تنجح طرق الجمع ، يجب على الباحثين وضع خطة لجمع الصور والتقاطها وتحليلها وتخزينها.

من أهم الأدوات المستخدمة:

- **OMERO**

برنامج خادم العميل لتصور وإدارة وتحليل صور المجهر البيولوجي. من المجهر إلى النشر ، يتعامل OMERO مع جميع صورك في مستودع مركزي آمن.

- **ADOBE bridge**

هو برنامج مجاني لتنظيم الصور محليًا. يعد Bridge مدير أصول إبداعية قويًا يتيح لك معاينة أصول إبداعية متعددة وتنظيمها وتحريرها ونشرها بسرعة وسهولة.

ويقوم بتحرير البيانات الوصفية وإضافة الكلمات الرئيسية والتسميات والتصنيفات إلى الأصول وتنظيم الأصول باستخدام المجموعات والبحث عن الأصول باستخدام عوامل التصفية القوية وميزات البحث في البيانات الوصفية المتقدمة.

- **ImageJ**

هي أداة معالجة وعرض صور مجانية مفتوحة المصدر تستند إلى .Java يمكنه قراءة وكتابة الصور بتنسيقات GIF و JPEG و BMP و PNG و PGM و FITS و ASCII وImagej .TIFF

- Trophy

هو برنامج مجاني ومفتوح المصدر يسمح لك بتنظيم وإدارة ووصف صور المواد البحثية. يختصر Tropy الخطوات في العثور على مصادر أرشيفية حتي الكتابة عنها.

5/2المرحلة الر ابعة الحفظ والأرشفة

تحتاج بيانات البحوث الى ترتيبات معقدة لعمليات التخرين والحفظ ونقل بيانات البحوث علاوة على أنه توجد هناك حاجة إلى تخطيط قوي لخصوصية البيانات وأمنها لحماية خصوصية الموضوعات البحثية وتأمين المعلومات الحساسة والمميزة للهوية الشخصية. ويعد اختيار طريقة التخزين المناسبة طوال دورة حياة البيانات أمرا ضروريا، لضمان بقاء البيانات آمنة والالتزام ببروتوكولات السلامة الموصي بها.

في مرحلة التخطيط والتصميم، يجب تجهيز خطة تخزين البيانات، مع إبراز التفاصيل المحددة المتعلقة بتخزين البيانات في جميع أنحاء المشروع، مثل خيارات التخزين.

عند البدء في جمع مجموعات البيانات وإنشائها، إما عن طريق استيعاب البيانات من مصادر خارجية أو إنشاء مجموعات بيانات جديدة، يمكن معالجة اعتبارات التخزين الإضافية، مثل أمان وسلامة البيانات.

يتطلب إكمال المشروع تحديد استراتيجيات لتقييم البيانات وأرشفتها، لتحديد المكان الذي يجب تخزين البيانات فيه، وما الذي يجب الاحتفاظ به على المدى الطويل، أو إلى أي درجة يجب مشاركة البيانات، لتشجيع إعادة إنتاج البيانات.

1/5/2خيارات التخزين للبيانات

هناك العديد من خيارات التخزين للبيانات يمكن استخدامها مع البيانات البحثية، تعتمد على الظروف الخاصة بالمشروع ونطاق العمل وتعتمد على بعض المحددات وهي :

- هل يشترك في المشروع بين أكثر من مؤسسة أو قسم
- حساسية البيانات: هل تحتوي البيانات على بيانات شخصية أو تخضع لحقوق ملكية فكرية أو بيانات تجارية
- هل توجد أي اطراف تجارية متضمنة
- هل المشروع البحثي بتضمن اكثر من موقع جغرافي (دولي-وطني)
- هل ستكون هناك اجهزة معملية تنتج بيانات تحتاج للتخزين
- ماهو حجم البيانات المراد تخزينها

- هل سيكون هناك تعاملات كثيرة على البيانات باستخدام البرمجيات أو الأدوات الحلول المقترحة لتخزين البيانات
- التخزين المحلي على اجهزة الحاسب الالي والشبكات
- تخزين البيانات من خلال استخدام قاعدة بيانات مشتركة على خادم
- مواقع محددة التي يتم فيها تخزين البيانات والعينات المستخدمة
- استخدام خدمات التخزين السحابية Cloud based storage

ويمكن تقسيم خيارات التخزين إلى خيارين رئيسيين هما:

- الخوادم المحلية

وهذا الخيار يكون مناسبا في حالة استخدام البيانات محليا داخل المؤسسة مع ضمان تأمين هذه الخوادم ووضع صلاحيات للوصول والتعديل على البيانات

- الخوادم السحابية

وتعد الخوادم السحابية هي الخيار الأمثل في حالة المشروعات التي تشترك فيها أكثر من مؤسسة وقد تكون في اكثر من دولة نظرا لسهولة الوصول اليها من خلال الانترنت وتأمينها بصورة كبيرة علاوة على خضوعها لقوانين حماية البيانات المتعارف عليها(**Storage During Research**)

يجب خلال مرحلة تخزين البيانات مراعاة النقاط التالية:

-مراجعة خيارات التخزين للمؤسسة لفهم وتحديد أماكن وكيفية تخزين البيانات بناء على العمليات، والأداء، وإتاحة الوصول اليها، لذا يجب الاحتفاظ بنسخة أصلية من البيانات الخام، مع توثيق أي تعديلات مستقبلية او تحديثات عليها.

- يجب مراجعة سياسات الاحتفاظ بالبيانات لتحديد مدد حفظ البيانات ويجب تخزين البيانات في مواقع تخزين مناسبة بناء على الاستخدام، بعض المؤسسات تفضل الاحتفاظ بالبيانات لمدة ثلاثة أعوام على الأقل.

- تحديد ما إذا كان لديك مساحة تخزين كافية أثناء معالجة البيانات وبناء عليه يجب اتخاذ القرار بخصوص شراء مساحات تخزين.

2/5/2حماية البيانات

1/2/5/2 تأمين البيانات

يجب حماية البيانات التي يفترض بها أن تكون سرية مثل بيانات التدريس والأبحاث وغيرها من الأنشطة وقد تخضع هذه الحماية لسياقات قانونية أو سياسة الجامعة.

يجب خلال هذه المرحلة الإجابة عن هذه التساؤلات:

- أين ستخزن بياناتك؟
- ما هي المخاطر التي يتعرض لها أمن البيانات وكيف ستتم إدارتها؟
- كيف ستتحكم في الوصول للحفاظ على أمان البيانات؟
- كيف ستضمن وصول المتعاونين إلى بياناتك بأمان؟
- هل سيتم الاحتفاظ بالوسائط الخارجية المتعلقة بأبحاثك، مثل دفاتر المختبر الورقي، خزانات مقفلة آمنة مع سجلات وصول وقائمة بالمستخدمين المصرح لهم؟
- كيف ستحمي سلامة بياناتك؟

يتضمن تأمين البيانات أيضا حماية الأجهزة والبرمجيات (على سبيل المثال، برنامج مكافحة الفيروسات، خطة منهجية لتحديث التطبيقات ونظام التشغيل، وجدار الحماية، وبرامج مكافحة التطفل، والوصول المادي المقيّد(data-security)

ويمكن تقسيم مستويات التأمين الخاصة بالبيانات الي 5 مستويات وهي:

1. بيانات متاحة للجمهور وغير مقيدة
2. بيانات بحثية غير حساسة غير منشورة، سواء أكانت قابلة للتحديد أم لا. بيانات البحث النشطة حتى يتم نشرها.
3. البيانات الحساسة: بعض البيانات الخاضعة للتنظيم أو البيانات التي يمكن أن تضر بالمركز المالي للموضوع أو بآفاقه المهنية أو الاقتصادية أو العلاقات الشخصية أو التأمين أو السمعة أو الوصم
4. البيانات الحساسة التي يمكن أن تعرض الشخص لخطر المسؤولية الجنائية أو المدنية الكبيرة أو البيانات التي تتطلب تدابير أمنية أقوى لكل لائحة

5. البيانات الحساسة التي يمكن أن تعرض الشخص لخطر شديد بالضرر أو البيانات مع المتطلبات التعاقدية للتدابير الأمنية الاستثنائية(DONALDSON; KOEPKE, 2022)

سياسة تأمين البيانات المتبعة في جامعة هارفارد

2/2/5/2عمليات النسخ الإحتياطي Backup

تعد عملية النسخ الاحتياطي للبيانات واحدة من أهم طرق تأمين البيانات ويجب تحديد آلية لعمليات النسخ الاحتياطي للبيانات مع متابعتها الدقيقة ووضع سياسة وخطة واضحة لعمليات النسخ الاحتياطي لضمان وجود نسخة من البيانات يمكن استرجاعها في حالة حدوث المشكلات.

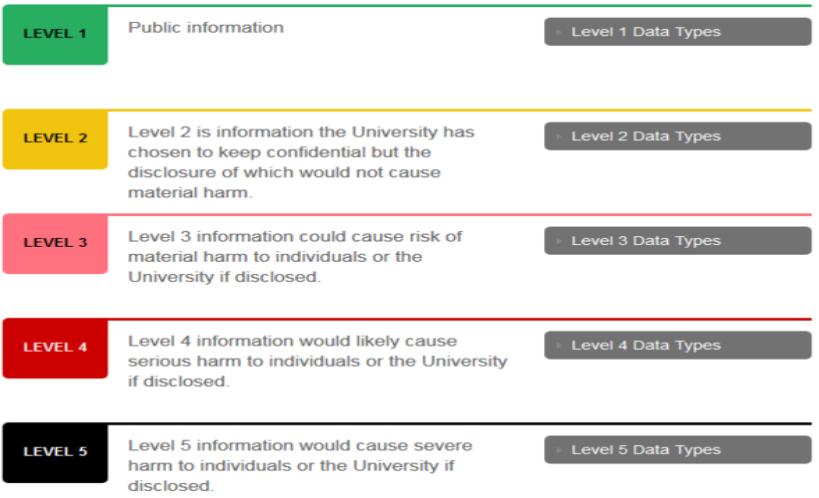
يمكن أن يساعد الاحتفاظ بنسخة احتياطية واحدة على الأقل في منع حوادث البيانات. أفضل طريقة لتجنب فقدان البيانات هي اتباع قاعدة 3-2-1:

•عمل 3 نسخ من البيانات (على سبيل المثال، أصلية + خارجية / محلية + خارجية / بعيدة)

يجب توزيع النسخ جغرافيًا (محلية مقابل بعيدة)(**storage-options**)

لإنشاء خطة أمان البيانات الخاصة بك يجب الإجابة عن هذه التساؤلات:

- كيف سيتم نسخ البيانات احتياطيًا؟
- أين يتم تخزين النسخ الاحتياطية (محركات أقراص الشبكة، التخزين البعيد (Cloud / Harvard Dropbox)؟
- كم مرة تقوم بعمل نسخة احتياطية من بياناتك؟
- كم عدد النسخ التي يتم إجراؤها (نسخ كاملة أو جزئية)؟
- من سيكون مسؤولاً عن النسخ الاحتياطي والاسترداد؟
- كيف سيتم استعادة البيانات في حالة وقوع حادث / كارثة؟
- كيف ستحمي هوية المشاركين أثناء مشاركة البيانات
- كيف سيتم التعامل مع البيانات الحساسة لضمان تخزينها ونقلها / مشاركتها بشكل آمن؟
- كيف يمكن أن تؤثر إدارة المعرفات على قابلية استخدام مجموعة البيانات للتحليل الثانوي؟(**Storage During Research**)



(شكل 4 مستويات سرية البيانات)

3/2/5/2سلامة البيانات

يجب التأكد من سلامة البيانات وفقا لبروتوكولات محددة وتشمل على:

تأمين ارسال واستقبال البيانات، مراجعة خطط إدارة البيانات، اتاحة التفاصيل حول استخدام البيانات، حماية هوية المشاركين، وتحديد الية التعامل مع البيانات الحساسة للتأكد من تخزينها ونقلها بشكل آمن؟

4/2/5/2أرشفة البيانات

وتعني هذه المرحلة بما يخص الاحتفاظ الدائم، أو الأرشفة، والتسكين المستمر للأشكال الإلكترونية للبيانات الملفات، فضلاً عن خدمات الصيانة والوصول إلى السجلات إلى الأبد

وتشتمل هذه المرحلة على :

إدارة المحفوظات والسجلات

تحديد البيانات التي سيتم الاحتفاظ بها بشكل دائم يمكن تحديد نسبة صغيرة من البيانات والسجلات التي تمثل السجل التاريخي للنظام أو المؤسسة أو كملكية فكرية لحفظها بشكل دائم، وتشتمل هذه المرحلة على عدة مراحل فرعية وهي:

تقييم البيانات

لاختيار البيانات التي سيتم الاحتفاظ بها بشكل دائم حيث يمكن تحديد نسبة صغيرة من البيانات والسجلات للتخزين الدائم كجزء من السجل التاريخي للمؤسسة.

السجلات المؤهلة للاحتفاظ الدائم ربما هي:

- توثيق الاختراق
- البيانات التي يتم إنشاؤها بواسطة معمل أو فرد كان له تأثير كبير على المجال
- البيانات التي تكون قابلة لإعادة الاستخدام بدرجة كبيرة في مجال بحث معين.

ويمكن لهذه الأسئلة أيضا تحديد المواد التي يتم الاحتفاظ لها من عدمه

- ما هي البيانات التي يجب الاحتفاظ بها / إتلافها لأغراض تعاقدية أو قانونية أو تنظيمية؟
- ما هي السجلات الأساسية المطلوبة لفهم بيانات هذا البحث والمشروع؟
- ما هو تأثير هذا البحث على تخصصه؟
- ما هو تأثير الباحث في مجاله؟
- هل بيانات البحث قابلة للتكرار؟
- هل يوجد فهرس للبيانات؟ كيف يفهم الباحثون المستقبليون البحث؟
- هل تم نشر هذا البحث؟ أين؟
- هل تم حفظ البيانات في مستودع للأبحاث؟
- هل هناك سجلات إضافية تتعلق بالبيانات؟
- هل هناك أية مشكلات تتعلق بالأمان أو الوصول؟
- هل يمتلك شخص آخر البيانات؟
- هل هناك أي سجلات ذات صلة يجب أن تصاحب البيانات؟

الاحتفاظ بالبيانات (أرشفة البيانات)

يجب الاحتفاظ بسجلات البحث، بشكل عام، لمدة لا تقل عن سبع (7) سنوات بعد نهاية مشروع أو نشاط بحثي.

يسعى التخزين والحفظ على المدى الطويل إلى التأكد من أن بيانات البحث ستكون متاحة لأولئك الذين يبحثون عنها ويمكن الوصول إليها لفترة زمنية محددة يحددها الممول والمؤسسة الأم، وغالبًا ما يكون الاحتفاظ الدائم، أو الأرشفة، استثمارًا مهمًا للمؤسسة، لأنه يعني الانتقال من عملية الحفظ المحدد بوقت للبيانات الى الحفظ المستمر لها مما يستلزم استعدادات خاصة متعلقة بعمليات التخزين وخدمات الرعاية والصيانة والوصول إلى السجلات إلى الأبد(**archives-and-records-management**).

وهناك بعض الأسئلة التي يجب طرحها في حالة الرغبة في أرشفة البيانات:

- ما هي الاستخدامات البحثية المتوقعة للبيانات؟
- ما هي السجلات الأساسية المطلوبة لفهم بيانات هذا البحث والمشروع؟
- هل بيانات البحث قابلة للتكرار؟ هل تم نشر البحث؟

تتضمن بيانات ومواد البحث معلومات بحثية مسجلة أو ملموسة أو غير ملموسة، بغض النظر عن الشكل أو الوسائط التي قد يتم تسجيلها عليها، والتي يتم إنشاؤها أو جمعها في عملية إجراء البحث، سواء كانت مدعومة من موارد الجامعة أو من قبل ممولين خارجيين.

تشمل بيانات ومواد البحث، على سبيل المثال لا الحصر:

- برامج الحاسب الآلي (برامج الحاسب الآلي، قواعد بيانات الحاسب الآلي ، وتوثيقها).
- مواد مثل العينات البيولوجية المعدلة وغير المعدلة
- الكيانات الكيميائية الجديدة أو المعدلة
- دفاتر المختبر
- ملاحظات من أي نوع
- المواد المقدمة إلى و / أو المعتمدة من قبل IRB أو IACUC أو لجان الإشراف البحثي الأخرى (على سبيل المثال، التطبيقات، مواد التوعية / الإعلان، نماذج الموافقة، إجراءات الاستبيان / الاستبيانات ونصوص استخلاص المعلومات)
- الصور الفوتوغرافية والأفلام والتسجيلات الصوتية والصور الرقمية
- العينات البيولوجية والبيئية الأصلية أو المعدلة
- المواد الهلامية، والأطياف، وخطوط الخلايا، والكواشف

- البروتوكولات والخوارزميات والرسوم البيانية والرسوم البيانية
- النتائج التجريبية العددية الخام
- مخرجات مفيدة
- المنجزات الأخرى بموجب الاتفاقات التي ترعاها؛ البيانات غير الملموسة مثل الإحصاءات والنتائج والاستنتاجات والمخرجات الأخرى بموجب اتفاقية برعاية؛ وأي سجلات أخرى، أو بأي شكل يمكن استخدامه، لإعادة بناء وتقييم نتائج البحث التي تم الإبلاغ عنها أو المنشورة بأي شكل آخر.

سجلات البحوث الأساسية

سجلات البحث الأساسية هي تلك السجلات البحثية التي تعد جزءًا لا يتجزأ من:

- إثبات طلبات المنح أو إثبات الامتثال للشروط التعاقدية، إذا تمت رعاية البحث
- إثبات البحوث المنشورة وبراءات الاختراع، سواء أكان البحث برعاية أم لا
- إثبات البحث الموصوف في مقترحات المنح وطلبات التمويل الأخرى
- أو السجلات التي يتم النظر في حفظها والوصول إليها بشكل دائم بواسطة برنامج إدارة المحفوظات والسجلات في مركز تاريخ الطب.
- المبادئ التوجيهية المعمول بها

تتضمن سجلات البحث الأساسية أيضًا أي بيانات أو مواد بحثية تم تحديدها على أنها ضرورية من قبل المدارس، بما يتوافق مع أفضل الممارسات للتخصص ذي الصلة.

5/2/5/2تدمير البيانات

- قد يكون تحديد البيانات والسجلات التي يمكن التخلص منها بأمان أمرًا معقدًا.
- إلا أنه يجب اتباع سياسة محددة لإهلاك البيانات من خلال اعداد وثيقة (سياسة تدمير البيانات) ويجب تحديد موقف البيانات الحساسة

6/2المرحلة الخامسة المشاركة والنشر

1/6/2مشاركة البيانات

تعد مشاركة البيانات ضرورية لترجمة نتائج الأبحاث إلى معرفة ومنتجات وإجراءات لتحسين البحث العلمي بشكل كبير وتطبيق هذه الأبحاث في الحياة العملية.

في العقد الماضي، أصبح من الشائع بشكل متزايد أن يقوم الباحثون بإتاحة بياناتهم للآخرين عند إكمالهم للدراسة. يشار إلى هذا عادة باسم مشاركة البيانات أو نشر البيانات. تتزايد مشاركة البيانات في الغالب بسبب سياسات البيانات الحديثة من المجلات والممولين ونظرا لوجود متطلبات مشاركة البيانات من خلال مستودعات البيانات التي يرغب الباحثون في نشر بياناتهم عليها لاكتساب السمعة الأكاديمية والمساهمة في المجتمع العلمي بشكل فعال من خلال النشر والمشاركة وتلقي التعليقات ومن ثم الوصول الى طرق جديدة في الاتصال الأكاديمي في المطبوعات والنشر.

يرتبط بعملية مشاركة البيانات مجموعة من النقاط الرئيسية وهي:

متطلبات النشر في المجلات العلمية

- تتطلب العديد من المجلات أن تكون المقالات المنشورة مصحوبة ببيانات البحث الأساسية
- غالبًا ما توجد سياسات مشاركة البيانات في إرشادات المؤلفين

السرية

- يجب ألا تحتوي البيانات المتاحة للجمهور على معلومات من شأنها المخاطرة بسرية المشاركين فيها.

حقوق الملكية الفكرية

- مشاركة البيانات التي أنتجتها أو جمعتها بنفسك
- مشاركة البيانات التي جمعتها من مصادر أخرى
- تطبيق التراخيص المناسبة على عملك المشترك

اتفاقيات استخدام البيانات

- يجب أن تكون اتفاقيات استخدام البيانات معلنة ومحددة مسبقا لتحديد كيفية ونمط استخدام البيانات ولضمان عدم استغلالها بشكل خاطئ

سياسات المشاركة والوصول الحر للبيانات

- يجب تحديد سياسات المشاركة الخاصة بالبيانات والاتفاقية الخاضعة لها اذا ما كانت خاضعة لأحد اتفاقيات الإتاحة العامة والوصول الحر

تعد مشاركة البيانات ضرورية للترجمة السريعة لنتائج البحث إلى معرفة ومنتجات وإجراءات لتحسين صحة الإنسان.

في العقد الماضي، أصبح من الشائع بشكل متزايد أن يقوم الباحثون بإتاحة بياناتهم للآخرين عند إكمالهم للدراسة. يشار إلى هذا عادة باسم مشاركة البيانات أو نشر البيانات. تتزايد مشاركة البيانات في الغالب بسبب سياسات البيانات الحديثة من المجلات والممولين.

وتشتمل هذه المرحلة على تبادل ومشاركة البيانات حيث تساهم عملية مشاركة البيانات في البحث عن بياناتك الخاصة بعد سنوات من الانتهاء من المشروع بالإضافة إلى تمكين الآخرين من إعادة استنساخ عملك وتمكين الآخرين من إجراء تحليلات جديدة باستخدام بياناتك(Share and Manage Your Research Data, 2019).

2/6/2العناصر الأساسية لمرحلة المشاركة والنشر

1/2/6/2اتفاقية استخدام البياناتDUA)Data Usage Agreement)

اتفاقية استخدام البيانات (DUA) هي عقد ملزم يحكم الوصول إلى البيانات غير العامة أو الحساسة التي يقدمها أحد الأطراف ("موفر البيانات") إلى طرف آخر ("المستلم") ومعالجتها. غالبًا ما تكون DUA مطلوبة من قبل الأطراف الخارجية وقد تكون ضرورية أيضًا للكشف عن البيانات البحثية إلى مؤسسات أخرى. تختلف شروط وأحكام DUA تبعًا للقوانين واللوائح التي تحكم النوع المحدد من البيانات التي سيتم مشاركتها ، بالإضافة إلى سياسات و / أو متطلبات المزود والمتلقي. تساعد الاتفاقيات الرسمية في هذه الحالات على تجنب سوء الفهم والنزاعات

حول استخدام البيانات وتخزينها ، والوصول المناسب وتدابير الأمان ، وغيرها من الأمور المهمة، بما في ذلك حقوق النشر وملكية النتائج.

2/2/6/2الملكية الفكرية

تشير الملكية الفكرية (IP) إلى أنواع مختلفة من التعبيرات غير الملموسة (مثل العمل الفني والأدبي والاكتشافات والاختراعات والكلمات والرموز والتصاميم) التي يتم الاعتراف بحقوق احتكار محددة لها بموجب قوانين محددة.

بشكل عام، تعتبر البيانات الأولية من تلقاء نفسها حقائق وبالتالي لا يمكن أن تكون محمية بحقوق الطبع والنشر. ومع ذلك، فإن البيانات التي يتم تجميعها معًا بطريقة فريدة ومبتكرة، مثل قواعد البيانات، يمكن أن تكون محمية بحقوق الطبع والنشر أو مرخصة. من المهم فهم ترخيص البيانات من منظور كل من مستخدم البيانات ومنشئ البيانات.

البيانات قد تكون غير محمية بحقوق النشر. ولكن يمكن أن تكون الأشكال المختلفة من البيانات، مثل جدول في كتاب ما، محمية بحقوق الطبع والنشر المستخدمة مع المنشورات التقليدية.

لذا يجب عليك تعيزز المشاركة والاستخدام غير المحدود لبياناتك من خلال إتاحتها بموجب ترخيص مناسب لضمان إعادة الاستخدام والإسناد بشكل صحيح، وتتوفر العديد من التراخيص التي تضمن الحقوق لمنشئ البيانات والمرخص له. هناك خياران لتقديم التراخيص المفتوحة لبيانات البحث وهما:

ترخيص Open Data Commons: مجموعة من الأدوات والتراخيص القانونية لمساعدتك في نشر البيانات المفتوحة وتوفيرها واستخدامها.

رخصة المشاع الإبداعي: توفر طريقة معيارية لمنح الإذن العام لاستخدام أعمالهم الإبداعية بموجب قانون حقوق النشر.

مشاركة البيانات التي جمعتها من مصادر أخرى يمكن أن يكون للبيانات المرخصة قيود في طريقة استخدامها أو مشاركتها، لذا عند إعادة استخدام البيانات الحالية، تأكد من توضيح الملكية، والحصول على الأذونات إذا لزم الأمر، وفهم الحدود التي تحددها التراخيص، ويجب عليك التأكد من تقديم الإسناد والاقتباس المناسبين.

يجب تحديد التالي فيما يخص البيانات:

• هل هناك أي قيود على إعادة استخدام بيانات الطرف الثالث؟

• من سيمتلك حقوق الطبع والنشر وحقوق الملكية الفكرية لأي بيانات ستجمعها أو تنشئها، إلى جانب الترخيص (التراخيص) لاستخدامها وإعادة استخدامها؟ (بالنسبة للمشاريع متعددة الشركاء، قد تكون حقوق الملكية الفكرية راجعة لاتفاقية بين الشركاء لتضمن حقوقهم)

3/2/6/2إستخدام المعرفات الثابتة .(persistent identifiers (PIDs

وتعد المعرفات الثابتة واحدة من الطرق الهامة التي تساعد على نشر بياناتك واتاحتها ويعد معرّف الكائن الرقمي (DOI) أحد أكثر المعرّفات شيوعًا على الرغم من وجود العديد من المعرفات الأخرى المستخدمة مع البحوث والباحثين والتي تساعد في نشر المواد وتكوين السمعة الأكاديمية للباحثين على نطاق واسع.

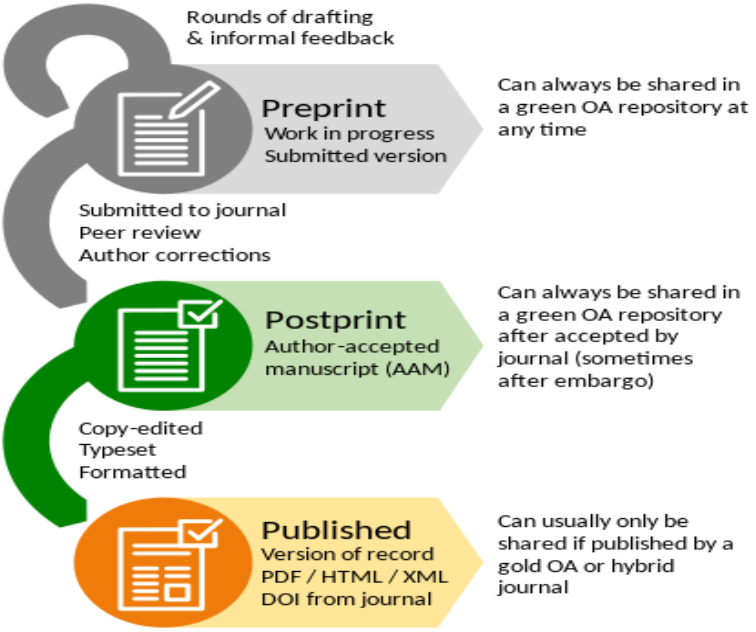
إذا كنت تخطط للنشر، فمن المهم إنشاء معرف فريد للمؤلف / الباحث وتعد أوركيد ORCID هي مبادرة عالمية مفتوحة وغير هادفة للربح ومتعددة التخصصات توفر معرفًا رقميًا ثابتًا يميزك عن أي باحث آخر.

من خلال التكامل في مهام سير العمل البحثية الرئيسية مثل تقديم المخطوطات والمنح ، يدعم أوركيد الروابط الآلية بينك وبين أنشطتك المهنية لضمان الاعتراف بعملك، ويمكن أيضًا ربط معرفات ORCID بمعرف Scopus أو ResearcherID

قنوات النشر للبيانات

المجلات العلمية المحكمة

مستودعات البيانات البحثية



(شكل5 قنوات النشر العلمي)

يمكن وصف مستودعات بيانات البحث على أنها نوع فرعي من البنية التحتية للمعلومات المستدامة التي توفر التخزين على المدى الطويل والوصول إلى بيانات البحث." - مشهد مستودعات بيانات البحث 2015: تحليل إعادة بيانات

لا يقتصر إدارة البيانات على إتاحة المقالات فحسب ، بل يشمل أيضًا البيانات والرموز والمواد المستخدمة لإجراء هذا البحث. تعد مستودعات البيانات موقعا مركزيًا لحفظ البيانات وإتاحة البيانات للاستخدام وتنظيم البيانات بطريقة منطقية.

"يمكن وصف مستودعات بيانات البحث على أنها نوع فرعي من البنية التحتية للمعلومات المستدامة التي توفر تخزينًا طويل المدى وإمكانية الوصول إلى بيانات البحث."

The Landscape of Research Data Repositories in 2015: A re3data Analysis

تساعدك مستودعات البيانات على:

- إدارة البيانات الخاصة بك
- تنظيم وإيداع البيانات الخاصة بك
- استشهد ببياناتك من خلال توفير معرّف دائم
- تسهيل اكتشاف البيانات الخاصة بك
- اجعل بياناتك أكثر قيمة للبحث الحالي والمستقبلي
- الحفاظ على بياناتك على المدى الطويل

استضاف مكتب إستراتيجية علوم البيانات في المعاهد الوطنية للصحة (NIH) والمكتبة الوطنية للطب ورشة عمل حول دور المستودعات العامة لتعزيز إمكانية اكتشاف البيانات وإعادة استخدامها في فبراير 2020. تتوفر تسجيلات الفيديو والشرائح(**Share & Publish**).

- Dryad
- figshare
- GigaScience
- Harvard Dataverse
- Mendeley Data
- Open Science Framework
- Vivli
- Zenodo
- NIH and NCBI Repositories
- Manuscript Repositories

https://datamanagement.hms.harvard.edu/share-publish

4/2/6/2مستودعات بيانات ما قبل النشر

توفير مستودعات ما قبل الطباعة فرصة لمشاركة المخطوطات وأوراق العمل قبل نشر المجلات.

تسمح مستودعات ما قبل الطباعة للباحثين بإيداع واكتشاف ونشر المنح الدراسية في المراحل الأولى من عملية البحث، لم تخضع المخطوطات المطبوعة مسبقًا لنظام المراجعة التقليدية القائم على الناشر.

تتضمن مستودعات ما قبل الطباعة الرئيسية منتدى ملاحظات يسمح للعلماء بتقديم التعليقات والمراجعات وتقييم مخطوطات ما قبل الطباعة بشفافية بشكل جماعي عبر الإنترنت

الصورة: توماس شافي. "سير عمل النشر النموذجي لمقالة في مجلة أكاديمية." مشتركة بموجب ترخيص CC-BY.

مستودعات ما قبل الطباعة المختارة لبحوث العلوم الصحية والطب الحيوي والعلوم الاجتماعية:

bioRxiv: خدمة أرشفة وتوزيع مجانية عبر الإنترنت للطبعات المسبقة في علوم الحياة ويديرها مختبر كولد سبرينغ هاربور.

medRxiv: خادم أرشفة وتوزيع مجاني عبر الإنترنت للمخطوطات الكاملة ولكن غير المنشورة في العلوم الصحية الطبية والسريرية وذات الصلة.

arXiv: تحتوي على مقالات علمية في مجالات الفيزياء والرياضيات وعلوم الكمبيوتر وعلم الأحياء الكمي والتمويل الكمي والإحصاء والهندسة الكهربائية وعلوم الأنظمة والاقتصاد.

SSRN: مستودع ما قبل الطباعة يمثل التخصصات عبر طيف البحث الكامل ، بما في ذلك العلوم التطبيقية والعلوم الصحية والعلوم الإنسانية وعلوم الحياة والعلوم الفيزيائية والعلوم الاجتماعية.

5/2/6/2أدوات التعاون والبرمجيات المستخدمة في المشاركة

يعمل العلماء بشكل متزايد في مشاريع بحثية تعاونية. غالبًا ما تجمع المشاريع التعاونية شركاء عبر التخصصات والمؤسسات والقطاعات. تقدم هذه المشاريع فرصًا للابتكار ولكنها تثير

أيضًا تحديات لتطوير تدفقات عمل تتسم بالكفاءة والفعالية وإدارة البيانات، هناك العديد من الأدوات المتاحة التي يمكن استخدامها لدعم هذه الأنواع من المشاريع.

- أدوات التعاون
- أداة إدارة المشروعات Open Science Framework

أداة مجانية ومفتوحة المصدر لإدارة المشاريع تدعم الباحثين طوال دورة حياة مشروعهم بالكامل.

- طريقة تحسين الابحاث**Protocols.io**

عبارة عن نظام يمكنك من إنشاء بروتوكولات مفصلة وتفاعلية وديناميكية خطوة بخطوة يمكن تشغيلها على الهاتف المحمول أو الويب.

المنتجات العلمية

- RCID
- IDs
- Preprints
- Datasets
- Code
- Protocols
- Reagents
- Training Materials
- المطبوعات والنشر

7/2المرحلة السادسة إعادة الإستخدام

وتتضمن هذه المرحلة تحديد من لديه حق الوصول إلى البيانات وكيف يسمح لهم باستخدام البيانات وكيفية إعادة الإستخدام مع تحديد أحكام وقواعد اعادة الاستخدام وإنشاء المشتقات

لذا فإنه من المهم أن تتمتع البيانات التي تنتجها بالاستدامة وقابلية إعادة الإنتاج أحد المسارات التي يؤكد المجتمع العلمي من خلالها على صحة اكتشاف علمي جديد هو تكرار البحث الذي أنتجه.

1/7/2قائمة مراجعة RDM Offboarding

توفر إدارة بيانات البحث فرصة للباحثين لوضع خطة تضمن تنظيم البيانات ومشاركتها مع الآخرين. عندما يغادر الموظفون ، يأخذون مهاراتهم ومعرفتهم المؤسسية معهم. من المهم تسجيل المعلومات الإعلامية الأساسية المتعلقة بالمشاريع ومجموعات البيانات لضمان نجاح المستخدمين في المستقبل.

وتقدم قائمة المراجعة توثيق كامل لعمليات ادارة البيانات البحثية علاوة على احتوائها على مسرد للمصطلحات المستخدمة في العمل.

2/7/2ملف نقل المعرفة RDM

يجب على الباحثين توثيق المعلومات الأساسية المتعلقة بالمشاريع ومجموعات البيانات في ملف نقل المعرفة. وهذا يضمن استمرار المشاريع بعد مغادرة الباحثين ويعزز اتساق البحث وكفاءته. يجب استخدام ملف نقل المعرفة جنبًا إلى جنب مع RDM Offboarding Checklist ، ولكن يمكن تقديمه في بداية المشروع أو عندما يدخل الباحث إلى معمل جديد.

3/7/2الوصول المفتوح

"البيانات المفتوحة" هي البيانات التي يمكن استخدامها بحرية وإعادة استخدامها وإعادة توزيعها من قبل أي شخص - تخضع فقط ، على الأكثر ، لمتطلبات الإسناد والمشاركة على حد سواء (كتيب البيانات المفتوحة). تركز البيانات العلمية المفتوحة على بيانات البحث الأولية المنشورة داخل الأوراق البحثية أو جنبًا إلى جنب معها.

4/7/2المصادر المفتوحة Open access

"الوصول المفتوح" هو وصول مجاني غير مقيد عبر الإنترنت إلى البحث العلمي والأكاديمي. هناك نوعان من الوسائل الأساسية لتوفير الوصول المفتوح: المجلات المفتوحة، والأرشيفات أو المستودعات ذات الوصول المفتوح.

5/7/2البيانات المفتوحة Open Data

"البيانات المفتوحة" هي البيانات التي يمكن استخدامها بحرية وإعادة استخدامها وإعادة توزيعها من قبل أي شخص - تخضع فقط، على الأكثر ، لمتطلبات الإسناد والمشاركة على حد سواء (Open data handbook). تركز البيانات العلمية المفتوحة على بيانات البحث الأولية المنشورة داخل الأوراق البحثية أو جنبًا إلى جنب معها.

هناك العديد من الخيارات لمشاركة مجموعات البيانات علنًا كشرط للنشر، بما في ذلك المستودعات التي ترعاها الحكومة والمستودعات التأديبية ومستودعات الطرف الثالث و Harvard Dataverse ، وهو مستودع بيانات مجاني مفتوح لجميع الباحثين في جامعة هارفارد.

1/3نتائج وتوصيات الدراسة

- **الاستقرارعلى تقسيم مراحل إدارة البيانات البحثية وهي:**

التخطيط – الجمع والإنشاء – المعالجة والتحليل – الحفظ والأرشفة – المشاركة والنشر – إعادة الأستخدام

- ضرورة اتباع الارشادات الخاصة بكل مرحلة من مراحل إدارة البيانات للوصول لأفضل النتائج.
- وجود العديد من الأدوات الجاهزة والمجانية التي يمكن استخدامها في مراحل إدارة البيانات البحثية تسهل عملية الإدارة.
- يجب اتباع المعايير والبروتوكولات الخاصة بإدارة البيانات لتسهيل عمليات الحفظ والإدارة.
- عمليات التوثيق لخطوات العمل من العمليات الهامة التي لا غني عنها لاستمرارية المشروعات.

1/4 قائمة المصادروالمراجع

1. analysis-ready-datasets. Available in: https://datamanagement.hms.harvard.edu/analyze/analysis-ready-datasets.
2. archives-and-records-management. Available in: https://datamanagement.hms.harvard.edu/evaluate/archives-and-records-management.
3. Collect & Analyze. Available in: https://datamanagement.hms.harvard.edu/collect-analyze.
4. Data Provenance. Available in: https://libguides.vu.nl/rdm/data-provenance. Access at: 01-01-2023.
5. data-analysis. Available in: https://libguides.vu.nl/rdm/data-analysis.
6. data-security. Harvard university.
7. DDI : learn. Available in: https://ddialliance.org/learn.
8. DMP tool. Available in: https://dmptool.org.
9. DONALDSON, D. R.; KOEPKE, J. W. A focus groups study on data sharing and research data management. 9 2022.
10. GAJBE, S. B.; TIWARI, A.; GOPALJI; SINGH, R. K. Evaluation and analysis of Data Management Plan tools: A parametric approach. **Information Processing & Management**, 58, n. 3, p. 102480, 2021/05/01/ 2021.
11. Guide To Data Cleaning: Definition, Benefits, Components, And How To Clean Your Data. Available in: https://www.tableau.com/learn/articles/what-is-data-cleaning. Access at: 2023.
12. https://port.sas.ac.uk/mod/book/view.php?id=1220&chapterid=726
13. Research data-management-guide. Available in: https://www.data.cam.ac.uk/data-management-guide.
14. Share & Publish. Available in: https://datamanagement.hms.harvard.edu/share-publish. Access at: 5-2-2023.
15. Share and Manage Your Research Data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, 12, n. 12, p. 5404-5404, 2019. Article.
16. Storage During Research. Available in: https://libguides.vu.nl/rdm/data-storage.
17. storage-options. Available in: https://datamanagement.hms.harvard.edu/store/storage-options.

Research Data management systems functions and Life cycle

Mohamed Mosaad Mohamed Elzlabany

Library of congress Cairo office

mh_zalabany@hotmail.com

Review and supervision

Usama El-Sayed Mahmoud

Professor of Library and Information Science

Cairo University - Faculty of Arts

Abstract

Research data and its management is one of the most important hot topics in the scientific and academic arena at the present time, and interest in it has increased due to the digital revolution and the expansion of the use of the Internet, which facilitated the preservation, storage and sharing of data. This led to an increase in the amount of research data produced by scientific, academic and other institutions, this type of data plays a vital role in evaluating the research, in addition to proving the accuracy of the scientific research. For the research data to be useful to the scientific community, it must go through several stages that we can call the research data life cycle. There are many tools, software, and standards that must be followed to manage research data in every stage starting from planning, collecting, processing, preserving, publishing, and re-use of the research data.

This paper deals with the functions of Research data management systems functions during the stages of their life cycle.

key words : Research data- Research Data management systems – Research Data Life cycle.