


Case Report

The Fast and the FRDR: Improving Metadata for Data Discovery in Canada

Clara Turp ^{1,*}, Lee Wilson ², Julianne Pascoe ³ and Alex Garnett ⁴

¹ McGill University Library, 3459 rue McTavish, Montreal, QC H3A 0C9, Canada

² Portage, ACENET, 309 Cooper Street, Suite 203, Ottawa, ON K2P 0G5, Canada; lee.wilson@ace-net.ca

³ Library and Archives Canada, 395 Wellington Street, Ottawa, ON K1A 0N4, Canada; julienne.pascoe@canada.ca

⁴ Simon Fraser University Library, 515 West Hastings Street, Vancouver, BC V6B 5K3, Canada; garnett@sfu.ca

* Correspondence: clara.turp@mcgill.ca; Tel.: +1-514-398-6969

Received: 14 February 2020; Accepted: 27 April 2020; Published: 2 May 2020



Abstract: The Federated Research Data Repository (FRDR), developed through a partnership between the Canadian Association of Research Libraries' Portage initiative and the Compute Canada Federation, improves research data discovery in Canada by providing a single search portal for research data stored across Canadian governmental, institutional, and discipline-specific data repositories. While this national discovery layer helps to de-silo Canadian research data, challenges in data discovery remain due to a lack of standardized metadata practices across repositories. In recognition of this challenge, a Portage task group, drawn from a national network of experts, has engaged in a project to map subject keywords to the Online Computer Library Center's (OCLC) Faceted Application of Subject Terminology (FAST) using the open source OpenRefine software. This paper will describe the task group's project, discuss the various approaches undertaken by the group, and explore how this work improves data discovery and may be adopted by other repositories and metadata aggregators to support metadata standardization.

Keywords: discovery; repositories; metadata standardization; OpenRefine; Faceted Application of Subject Terminology

1. Introduction

There is a need for repositories to move beyond stand-alone systems by promoting interoperability. Systems are slowly evolving away from the stand-alone model, whether through metadata aggregation services, discovery layers, or linked open data-based approaches (for examples of national and international aggregation services, see Research Data Australia (<https://researchdata.andis.org.au/>) and the European Union's OpenAIRE Explorer (<https://explore.openaire.eu/>)). These models provide opportunities for the increased discoverability and interconnectivity of cross-domain datasets. Even with such ongoing efforts to de-silo research data, a lack of standardized metadata practices poses significant challenges to normalizing metadata harvested from multiple repositories [1]. This paper will share a community-driven project undertaken by stakeholders from multiple Canadian universities to improve user experience and discoverability for the Federated Research Data Repository (FRDR), a national repository and data discovery service for Canada. This paper will provide: (1) a brief overview of the technologies that enable metadata harvesting and aggregation across platforms, a discussion on using automatic entity resolution techniques to improve metadata, and an introduction to the concept of cooperative cataloging; (2) background information on Portage, the Canadian Research Data Management (RDM) context, and metadata standardization challenges; (3) a presentation of a case study using OpenRefine to reconcile subject keywords to Faceted Application

of Subject Terminology (FAST); and (4) a discussion of challenges and how this work may eventually be adopted by other repositories and metadata aggregators.

2. Literature Review

A cornerstone of the FRDR's metadata discovery project involves harvesting from the open endpoints already deployed at many research data repositories [2]. At the governmental level, these usually implement the Comprehensive Knowledge Archive Network (CKAN) or Socrata application programming interfaces (API); at the library level, they almost all implement the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) API. Historically, the OAI-PMH was used to expose journal and article metadata [3] but is now widely used by institutional repositories that provide access to research outputs of all types. The OAI-PMH is a relatively static API, providing an Extensible Markup Language (XML) endpoint that can be parsed programmatically. There have been past efforts at developing a linked data native OAI-PMH successor that could be more effectively queried on the fly [4,5], but to date, these have not materialized on a sufficient scale. Most recent efforts have focused on building sufficiently large local databases of harvested OAI-PMH endpoints that can then be indexed with ElasticSearch through a unified discovery portal [6]. These portals are usually supported at the national level, with Australia's National Data Service [7,8] as a prominent example.

Automatic entity resolution, wherein a term or terms are matched against a canonical vocabulary and assigned a ranking and numerical score for how accurately they correspond to terms from the original corpus, is widely implemented by many lookup APIs, but can often provide only a relative confidence measurement of its own accuracy, rather than a necessarily usable result [9]. Libraries, having both a high degree of confidence in their own data and a high expectation of the labor involved in producing these data, have nevertheless been slow to adopt automatic entity resolution for bibliographic control systems, preferring, wherever possible, to perform this canonical lookup manually. However, a significant amount of labor can be saved by using lookup APIs as a starting point, provided that the interface for reconciling automatic and manual curation is good enough; this is the implicit position of many "supervised" as opposed to "unsupervised" machine learning classifiers, which encourage users to bias automated results along desirable lines in order to improve an API's output, and it is also reflected in the design of Google's OpenRefine platform [10]. OpenRefine can be particularly useful for library work under an ideal set of circumstances, namely, when (1) an authority source is available, (2) an interface to perform supervised reconciliation is necessary for data quality, and (3) the users have some amount of familiarity with text parsing. "Cooperative" cataloging is when the cataloging labor is distributed across multiple institutions or stakeholders that may have their own norms and expectations, but which is expected to produce a mutually useful result [11]. This cataloging practice has already been theorized, put into practice, developed into successful business models (most obviously those of Online Computer Library Center (OCLC), and been accepted as a norm. However, it has not to our knowledge been used in a project of our particular scope.

3. Background

In Canada, RDM activities are undertaken by a diverse group of stakeholders operating on various local, regional, and national scales [12]. The drivers for RDM include both top-down requirements from journals and publishers and bottom-up approaches that highlight the importance of good data management practices and an increasing recognition of data as a primary research output [13]. In 2016, the Canadian Tri-Agencies released a Statement of Principles on Digital Data Management that outlined funder expectations for RDM, and the responsibilities of researchers, research communities, institutions, and funders in meeting these expectations [14]. Since then, the Tri-Agencies have been engaged in community consultation around an RDM policy that is expected to provide a mandate for institutional strategies for RDM, requirements around data management planning at the researcher level, and expectations for the deposition of data that directly support the conclusions of research outputs into a

repository [15]. The policy, which is planned to be released in 2020, is expected to increase the demand for RDM services and has helped to encourage national coordination around RDM activities.

The Portage Network (https://github.com/axfelix/frdr_harvest), an initiative of the Canadian Association of Research Libraries (CARL), was launched in 2015 with the goal of building capacity for RDM activities in Canada through its national Network of Experts and partnerships with service provider organizations and other RDM stakeholders. Portage's Network of Experts comprises several ongoing Expert Groups and Working Groups, many of which also contain temporary, project-based task groups. Expert and Working Groups are largely composed of memberships drawn from Canadian academic libraries, but also include researchers, research administrators, non-library-based data management experts, and representatives from various government departments and agencies. The FAST and the FRDR task group was led by the Data Discovery Expert Group through its FRDR Discovery Service Working Group. The Data Discovery Expert Group supports data creators and curators in planning, producing, and managing descriptive metadata to enable the effective discovery and reuse of research data across a wide range of disciplines. In addition, the group researches and promotes standards for metadata and data that support both machine-to-machine and human-to-machine discovery activities. The Data Discovery Expert Group is also involved in the development of the FRDR.

The FRDR is a scalable data discovery, deposit, and access platform developed through a partnership between CARL Portage and the Compute Canada Federation, Canada's high performance computing provider. The FRDR leverages Canada's digital research infrastructure and is available to all Canadian researchers. Its range of repository services helps researchers publish their research data, regardless of file size or the number of files; curate and prepare their research data for long-term preservation; and comply with emerging funder and journal policies for data discovery and access. Data repositories in Canada operate on national, regional, and local or institutional scales. They are many and varied, with some estimates suggesting that there are roughly 170 in operation across Canada [2], but they can be categorized broadly as discipline-specific, government, and generalist [16]. Metadata and curation practices also vary widely across repository platforms, with services ranging from fully mediated and curated deposits, to self-deposit models. The aim of the FRDR's federated discovery layer is to harvest and, as necessary, crosswalk metadata from individual repositories across Canada to provide a single search portal for Canadian research data (Figure 1). At the time of writing, the FRDR has ingested more than 100,000 metadata records from over 40 Canadian governmental, institutional, and discipline-specific data repositories. The FRDR's discovery layer thereby greatly increases the discoverability of Canadian research data and possibilities for re-use.

While the FRDR has been successful in facilitating discovery across a wide array of otherwise separate data repositories, inconsistent metadata practices and quality, as well as non-standard methods for exposing metadata across platforms, has presented challenges. Recognizing these challenges from the outset, the FRDR development team created their own metadata harvester (https://github.com/axfelix/frdr_harvest) that can parse feeds from several internationally recognized standards, including the OAI-PMH, CKAN, CSW, and Marklogic APIs, with support for new standards being added on an ad hoc basis. The FRDR's default metadata uses Dublin Core and DataCite; crosswalks were therefore incorporated to the harvester to link the different metadata schemas (such as DDI, FGDC, and CSW) to the FRDR's metadata scheme. However, even with crosswalks, specific metadata fields still pose many challenges.

The "subject" field is one that poses interesting challenges. While various standards (loosely) agree on what an author or a publication date are, they do not agree at all on which controlled vocabularies to use, if any. Even beyond controlled vocabularies, what represents a keyword greatly varies from a single word to full sentences, and from very general to hyper-specific terms. The FRDR uses facets (subject, date, author, and source) to improve both the discoverability and browsing capacities of search results pages, by allowing end-users to refine their searches. All harvested keywords populate the subject facet in the interface. When looking through the "subject" facet, the user can find spelling

mistakes, separated near-synonyms or related terms, and many inconsistencies. In the interface, subjects are sorted based on the related total number of documents. The inconsistencies therefore become evident when you go beyond the top results. Moving away from free-text keywords by using controlled vocabularies is one possible solution to these challenges. However, reconciling keywords with a controlled vocabulary and enforcing the assignment of controlled subject headings in the future is not a simple or straightforward task.

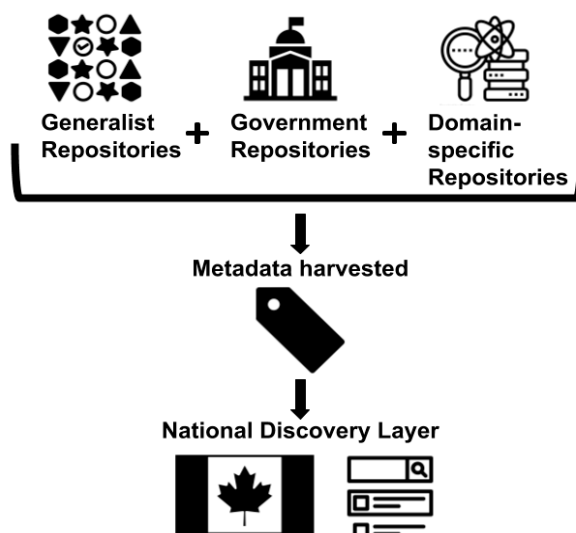


Figure 1. The Federated Research Data Repository (FRDR) harvests metadata from a variety of data repositories to create a national discovery layer for Canadian research data.

4. Case Presentation

4.1. Rationale

In recognition of this challenge, the FRDR Discovery Service Working Group, in support of the work of the Data Discovery Expert Group, engaged in a project to map harvested keywords to OCLC's FAST vocabulary. FAST was derived from the Library of Congress Subject Headings (LCSH) to meet a perceived need for a general-use subject terminology scheme that would retain the rich vocabulary of LCSH while making the schema easier to understand, control, apply, and use [17,18]. The working group chose FAST because it rests on the Library of Congress authorities, which is a mature base with sufficiently generic terms that cover a wide range of domains. FAST also has good open APIs, the potential for linked data applications, and compatibility with OpenRefine. The working group decided to use OpenRefine to do a part-automated part-manual process to map harvested keywords to FAST terms. OpenRefine is an open-source desktop application that has an online spreadsheet-type interface with the functionality of a database. It was originally started by Freebase as Gridworks and was then acquired by Google; it is now fully supported by its own open-source community. The application provides the ability to transform and normalize datasets as well as reconcile to specific Linked Open Data vocabularies such as FAST and Wikidata. This work was facilitated by the use of code originally written by Christina Harlow to provide a turnkey gateway between OpenRefine and the Library of Congress FAST APIs [19]. OpenRefine was chosen in part due to its suitability for this reconciliation task as well as the built in provenance functionality that allowed us to export a log of the cleaning and reconciliation decisions made, allowing us to deploy the results upstream and making this work reproducible by others.

4.2. Procedure

The first phase of the project experimented with OpenRefine and FAST to identify the main issues and provide a proof of concept. A collaborative infrastructure was built using an OpenRefine instance hosted on a shared server to allow multiple concurrent users to contribute to a vocabulary crosswalk through a relatively simple web interface. Eight thousand English keywords from the FRDR index were passed through OpenRefine's reconciliation service, which uses the FAST APIs to automatically match a selected column to FAST terms. The software uses a numerical score to indicate the possible reliability of the mapping. Reconciled terms with high enough scores, i.e., scores that are close to 100%, are automatically mapped, whereas terms with lower scores offer the user a selection of possible keywords for reconciliation. The first pass through OpenRefine yielded only 22% automatic matches. This low figure meant that a significant amount of manual work remained in order to reconcile the harvested keywords that had multiple suggested FAST terms. Based on the initial test, it was decided that a semi-automated process was best, at least for the first round of mapping. However, some major challenges remained, not least of which was motivating volunteers to engage in a relatively large-scale and ongoing joint cataloging effort.

With a proof of concept established, a new task group was established in May 2019. The 14 members of this task group included catalogers, research data management experts, and individuals with programming skills across multiple Canadian institutions. These institutions include nine different Canadian universities, an oceans-based research project supporting marine acoustic data, Canada's national advanced research computing provider, and Canada's national archives. This combined expertise allows the group to tackle both the manual processing and the creation of sustainable long-term workflows.

The group's initial work involved discussions on the best method for approaching the task. Different approaches were considered. First, we considered a more automatic approach that would separate the keywords into multiple categories (e.g., topics, names, and geographical terms) using regular expressions, name entity recognition, geographical APIs, and natural language processing. These different categories could then be cleaned up and reconciled with specific API configurations that would search only topical keywords, for example. The benefits to this approach were its reproducibility and the fact it relies on scripts rather than manual work. However, the separation process had a relatively high error rate, which could impact data quality. The second option we considered would use the full list of keywords and separate them alphabetically between the members of the group. This approach would ensure data quality, but it relies heavily on manual labor, and so is not a sustainable long-term process. There was also a debate about how much clean-up should be done before the mapping starts, e.g., whether basic deduplication and spell-checking would be sufficient, or if more substantial cleaning steps should be applied initially. The decision was made to approach the task manually, while a subset of the group explored ways to automate the process using the manual work. Tracking the manual work and keeping statistics on the automatic reconciliation will, eventually, inform how this work can be automated while still retaining high quality data. A detailed procedure was created to ensure everyone in the task group works in a similar systematic way within the shared OpenRefine instance.

The reconciliation work is currently being manually undertaken by the task group. This process includes (1) comparing the automatic term or suggested terms with the original keyword, (2) using additional information from the datasets and FAST to map the keyword correctly, and (3) selecting or entering the appropriate FAST term when possible. For the terms that were left untouched by the automatic reconciliation process, the task group will search FAST for an appropriate term.

Once the remediation work is completed, there will be three remaining phases to this project.

1. **Integration with the FRDR Interface:** A major outcome of this project is to integrate the FAST terms into the FRDR's existing search facets. One solution is to have a subject field and a keyword field. Original and unmatched keywords would therefore remain in the keyword field while

the remediation work would display in the subject field, which would be used in the filters. Additional work will need to be done to ensure that whatever implementation we decide on integrates smoothly with the existing user experience.

2. **Integration with the Harvester:** The next phase of this project will be to integrate the work of the task group with the FRDR's harvesting process. The most important outcome from this phase will be a sustainable workflow that leverages the manual work done by the task group and automates the remediation process as much as possible while still ensuring data quality. One possibility would be to automatically match keywords that were matched through the initial manual process and have a high percentage of certainty based on the FAST API. The other keywords would be added to a bank of keywords that would eventually need to be manually matched. A possible pitfall of this approach is that it continues to rely heavily on manual remediation. Future work will explore more automated workflows for improved sustainability.
3. **Internationalization:** As a national discovery platform, the FRDR must be available in both of Canada's official languages: English and French. While the FAST terms are not available in French, work has been done by the University of Laval to map the *Répertoire de vedettes-matière* (<https://rvmweb.bibl.ulaval.ca/>), a French controlled vocabulary, to the FAST terms. This work will be highly valuable as it will do a primary mapping between the English keywords already assigned to their French equivalents, while also allowing the task group to map the French keywords harvested by the FRDR.

5. Discussion

The next steps aim to build a workflow that will incorporate scripting and automation. Currently, the workflow shared in this paper requires significant human resources. The goal is to use this methodology to develop a more sustainable, long-term approach. The manual work is a critical component that allows the group to assess and strategize guidelines for large data reconciliation projects. The group is also gathering data about the remediation process. Capturing different categories and the level of satisfaction with automatic matches, and noting different roadblocks to the reconciliation process builds a strong foundation on which to develop a workflow that will be more sustainable and can be shared with the community. Ultimately, our goal is twofold, to reconcile and map subject keywords to FAST in order to improve the discoverability of the FRDR datasets, and to develop and share a reproducible procedure and workflow using open source software for similar work in the field that is focused on reconciling and normalizing aggregated datasets. This project is the first iteration of a process that will have to run on a regular basis, with new harvested databases and uploaded datasets. Any workflow that relies primarily on manual labor is not realistic. Our goal is to leverage this work to develop a more automated workflow that relies less on human interaction while still producing high quality matches. However, if the challenges encountered do not allow us to build such a workflow, the procedure and guidelines for the bulk reconciliation work would be shared, as we believe that other large cataloging and reconciliation projects would benefit from having an example of these procedures.

5.1. Challenges

As we have identified, manual work management and consistency in cataloging decisions represent significant challenges for such large-scale data-driven projects. Additional challenges that require consideration include how to provide equitable access to bilingual data, build sustainable workflows, and incorporate the work into the FRDR in a way that minimizes the duplication of work. Bilingual systems present at least two important challenges: (1) providing metadata in the language of the data and (2) improving discovery by creating bilingual crosswalks. The current harvested keywords contain both English and French terms. These terms must both be reconciled with a chosen controlled vocabulary; however, ideally, the reconciled terms would also be mapped to the other language. Users searching in French, for example, might be interested in searching across all the datasets, regardless of which language is selected in the interface. The Portage Network hopes to

utilize the mapping currently being developed between FAST and the *Répertoire de vedettes-matière* to simplify this work. However, similar to the challenges of incorporating the reconciliation work manually done with English keywords, the working group must create a workflow to reconcile the French keywords and make all keywords equally discoverable. As mentioned previously, ensuring the long-term viability of matching terms to authorities in the least manual method possible will be key to the sustainability of this project. Developing and documenting workflows, automating where possible, and ensuring efforts are not duplicated through the ongoing tracking and management of reconciliation work will be the cornerstone of tackling these challenges—challenges that are common to any large-scale work on improving the discoverability of data aggregations.

5.2. Beyond FAST and the FRDR

This project has cross-domain relevancy and broader applications for working with aggregated datasets that require reconciliation and mapping to linked data terms and headings. Cultural heritage collections, in particular, are an area that often faces overwhelming challenges in terms of unstructured, non-standardized metadata that requires normalization at scale. Some of these challenges are highlighted in the handbook, *Linked Data for Libraries, Archives and Museums*: “The implementation of successive technologies over decades has scattered the metadata of our libraries, archives, and museums across multiple databases, spreadsheets and even unstructured word processing documents . . . At the same time, budget cuts and fast-growing collections are currently obliging information providers to explore automated methods to provide access to resources. We are expected to gain more value out of the metadata patrimony we have been building up over the decades.” [5]. There is an increasing recognition of and need for standardizing datasets in order to develop interoperability and in order to “de-silo” Galleries, Libraries, Archives, Museums (GLAM) collections. In addition, there are challenges in terms of the resources that are required in order to clean up and map datasets. Efforts to provide aggregated access to heritage collections can be seen at various levels, both nationally (DPLA, Canadiana), internationally (Europeana), and locally (for instance across diverse repositories within one’s own institution) but encounter similar challenges in reconciling diverse metadata collections. To different degrees, all three organizations (DPLA, Canadiana, and Europeana) have created metadata profiles and working groups dedicated to tackling core aggregated data issues such as multilingualism, data quality, reconciliation and normalization. For example, Europeana has developed user personas in an effort to enhance discoverability of its diverse datasets through user-centered requirements and, with a focus on metadata, create a series of user scenarios reflecting information seeking behavior (see: <https://pro.europeana.eu/project/data-quality-committee>). Work that demonstrates the use of open-source tools and data normalization and reconciliation strategies for multiple languages, as well as community-driven participation, provides invaluable opportunities for GLAMs to adopt scale-driven data clean-up techniques with a low entry barrier. Using an open-source tool such as OpenRefine and documenting the strategies, code, and working group arrangements allows for the procedures developed in this project to be replicated at scale and across different domains. In addition, further developing this project for different data collections and communities will, in return, yield customization and input for the implementation of data aggregation tools and technology across a broad spectrum of uses. For a case example of the use of OpenRefine in a cultural heritage context, see the recent ALA presentation: Lower the Barrier & Be Empowered: Creating and Including Linked Data Vocabularies for Digital Collections, (Sai (Sophia) Deng, University of Central Florida). This is a project that uses OpenRefine and linked data vocabularies to improve cataloging workflows for the Library’s digital repository. The presentation highlights the high barrier catalogers face in implementing linked data for their collections and the strategies they employed using OpenRefine.

6. Conclusions

The challenges encountered by this project are far from unique. Finding a balance between long-term sustainability and data quality is often difficult, especially in large scale projects. Additionally,

the process of automating manual work is a common challenge across library services. Portage's experience of harvesting repositories, building crosswalks, and developing a bilingual discovery interface can help to inform the work of other discovery and aggregation services. The lessons learned from this project will benefit GLAMs looking into metadata quality and building procedures for distributed, group reconciliation work. However, these lessons learned can also be applicable on smaller scale projects where the reconciliation work is overwhelming compared to the available resources.

Many end-users access repositories in the hope of discovering research data to support new research. Despite technological advances and an increase in metadata aggregation services, significant challenges remain in normalizing the metadata harvested from repositories due to a lack of standardized metadata practices. The results of this project will have a direct impact on the ease with which our users can search for and discover research data through Canada's national discovery service. In leveraging open source tools to respond to this widespread challenge, as well as sharing and outlining our work and procedures, we hope the lessons learned from the Canadian experience will inform similar work being undertaken internationally.

Author Contributions: Conceptualization, C.T., L.W., A.G.; writing-original draft preparation, C.T., A.G., J.P., L.W.; writing-review and editing, L.W., J.P., C.T., A.G.; project administration, C.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shreeves, S.L.; Knutson, E.M.; Stvilia, B.; Palmer, C.L.; Twidale, M.B.; Cole, T.W. Is Quality Metadata Shareable Metadata? The Implications of Local Metadata Practices for Federated Collections. In Proceedings of the Twelfth National Conference of the Association of College and Research Libraries, Minneapolis, MN, USA, 7–10 April 2005; Thompson, H.A., Ed.; Association of College and Research Libraries: Chicago, IL, USA, 2005; pp. 223–237.
2. Vejvoda, B.; Ambi, A.; Barsky, E.; Lindstrom, K.; MacDonald, H.; Matthews, K.; Moosberger, M.; O'Hara, L.; Powelson, S.; Silk, K.; et al. *Portage Data Discovery Expert Group—Collections Development Working Group: Phase One Report*; Portage: Ottawa, ON, Canada, 2017.
3. Van de Sompel, H.; Nelson, M.L.; Lagoze, C.; Warner, S. Resource Harvesting within the OAI-PMH Framework. *D-Lib Mag.* **2004**, *10*, 2–19. [[CrossRef](#)]
4. Haslhofer, B.; Schandl, B. The OAI2LOD Server: Exposing OAI-PMH Metadata as Linked Data. In Proceedings of the International Workshop on Linked Data on the Web (LDOW2008), Beijing, China, 22 April 2008.
5. van Hooland, S.; Verborgh, R. *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish your Metadata*; US-Ed.; Neal-Schuman: Chicago, IL, USA, 2014; ISBN 978-0-8389-1251-5.
6. Smart, R. What Is an Institutional Repository to Do? Implementing Open Access Harvesting Workflows. *Publications* **2019**, *7*, 37. [[CrossRef](#)]
7. Kethers, S.; Shen, X.; Treloar, A.E.; Wilkinson, R.G. Discovering Australia's research data. In Proceedings of the 10th Annual Joint Conference on Digital libraries—JCDL'10, Gold Coast, Australia, 21–25 June 2010; ACM Press: Gold Coast, Australia, 2010; p. 345.
8. Treloar, A.; Wilkinson, R. Access to Data for eResearch: Designing the Australian National Data Service Discovery Services. *IJDC* **2008**, *3*, 151–158. [[CrossRef](#)]
9. Delpuch, A. A survey of OpenRefine reconciliation services. *arXiv* **2019**, arXiv:1906.08092.
10. Tillman, R.K. Extracting, Augmenting, and Updating Metadata in Fedora 3 and 4 Using a Local OpenRefine Reconciliation Service. *Code4Lib J.* **2016**.
11. Baldwin, P.E. Bibliographic Control Problems and Organizational Change Issues Posed by the Implementation of a Computer Based Cooperative Cataloguing Support System. Ph.D. Thesis, Simon Fraser University, Burnaby, BC, Canada, 1976.
12. Baker, D.; Bourne-Tyson, D.; Gerlitz, L.; Haigh, S.; Khair, S.; Leggott, M.; Moon, J.; Ridsdale, C.; Tourangeau, R.; Martha, W. *Research Data Management in Canada: A Background*; Zenodo: Genève, Switzerland, 2019.
13. Shearer, K. *Comprehensive Brief on Research Data Management Policies*; Portage: Ottawa, ON, Canada, 2015.

14. Government of Canada. Tri-Agency Statement of Principles on Digital Data Management. Available online: http://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html (accessed on 25 March 2020).
15. Government of Canada. DRAFT Tri-Agency Research Data Management Policy for Consultation. Available online: https://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html (accessed on 25 March 2020).
16. Portage. *Scholars Portal Repository Options in Canada: A Portage Guide*; Portage: Ottawa, ON, Canada, 2019.
17. OCLC FAST (Faceted Application of Subject Terminology). Available online: <https://www.oclc.org/research/themes/data-science/fast.html> (accessed on 11 February 2020).
18. O'Neill, E.T.; Mai Chan, L. FAST (Faceted Application of Subject Terminology): A simplified vocabulary based on the Library of Congress Subject Headings. *IFLA J.* **2003**, *29*, 336–342. [[CrossRef](#)]
19. Harlow, C. cmharlow/fast-reconcile. Available online: <https://github.com/cmharlow/fast-reconcile> (accessed on 9 December 2019).



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).