Coalition Publica Metadata Working Group

Technical Report: Metadata Feedback for Coalition Publica

Working Group Members

- Lise Brin | Canadian Association of Research Libraries
- Haiyun Cao | York University
- Jessica Clark | Coalition Publica
- Bart Kawula | Scholars Portal
- Inba Kehoe | University of Victoria
- Pierre Lasou | Université Laval
- Tomasz Mrozewski | York University
- Mike Nason (Chair) | University of New Brunswick
- Mathieu Pigeon | Érudit, Université de Montréal
- Brianne Selman | University of Winnipeg
- Sarah Severson | University of Alberta

Mandate/Terms of Reference

1. Provide guidance and advice on best practices for metadata quality in library-hosted scholarly journals to ensure increased discoverability, especially for Coalition Publica's dissemination service
2. Develop procedures and documentation to guide librarians and editors in assisting journals in their creation of high-quality metadata
3. Provide recommendations to Coalition Publica's Operations Committee towards improving and encouraging better metadata within PKP's and Érudit's platforms.

Introduction

In the Winter of 2019, a Metadata Working Group (MWG) was struck by Coalition Publica, a collaborative effort between the Public Knowledge Project (PKP) and the Érudit consortium. At the same time, a preliminary process invited a small group of invested parties to convey their existing concerns with metadata fidelity in both Open Journals Systems (OJS) and the Érudit platform. This document, which represents concerns identified by the working group as well as feedback provided by the invited groups, contains three specific sections related to metadata and the Coalition Publica project:

1. A review of metadata ingestion processes utilized by libraries;
2. A review of emergent issues in metadata regarding equity, inclusion, and culture;
3. A review of metadata issues that need to be addressed by PKP in OJS and/or Érudit on the erudit.org platform.

Since the MWG was created there have been a number of significant improvements to both OJS and the workflows associated with Coalition Publica. Some of the initial concerns identified in this group's Interim Report have since been rectified. For example, OJS 3.2 customized and simplified the process by which journals could choose required, additional metadata fields. It also introduced new author-name metadata conventions that resolve a number of the problems in recording name metadata in the first place. It has, through careful UI/UX testing, injected a little more intention into the publication process.

These changes to the workflow and the smoothness of content transitions between OJS and Érudit represent a maturation of the project.

Mike Nason
Scholarly Communications and Publishing Librarian, UNB Libraries
Chair, Coalition Publica Metadata Working Group

Libraries and Metadata

The Libraries and Metadata subgroup was tasked with exploring how libraries handle the metadata of journals that use PKP's OJS platform in their production workflow. Libraries hold a unique position in open scholarly publishing in that they often fulfill the role of a publisher or provide hosting services, in addition to their roles in acquisition, discovery, and preservation. In many Canadian academic libraries, OJS hosting and support services account for a large portion of a library's scholarly communication efforts, but even with that relationship on the production side of publishing, there are quite a few challenges in integrating the content into library collections. This report aims at identifying some of the shortcomings that inhibit the flow of metadata from the OJS platform to library collection systems and proposes a way for librarians to assess the quality and validity of the metadata coming out of OJS.

The Need for Indexing

To gather information for this report, members of this group consulted with colleagues in various departments within their own institutions and surveyed colleagues at other institutions to better understand how both journal-level and article-level metadata from OJS journals were handled.

The general consensus is that libraries with publishing programs will facilitate getting OJS journals into their library catalogues. Typically, a journal-level entry is created when it first begins publishing or joins the library publishing platform. This usually consists of manually creating a skeletal serials record with the title, ISSN, subject headings and an 856 field that links to the journal's home page. Such records are rarely updated. No libraries we spoke to independently index OJS journals at the article-level.

Most libraries, however, use or are moving towards adopting single interface library services platforms (LSPs) that incorporate outside catalogue records and various indexes with article-level entries. This means that getting indexed at the article-level has become critical for both the journal that wants to maximize its visibility and the library that wants to make it discoverable in their LSP. A number of library publishers provide consulting services to journals that help them identify appropriate indexes, advise on what they need to meet indexing services' requirements and assist with applications. This level of indexing

support is not yet widespread or formalized at all institutions, but is commonly requested by journals.

More established journals have experience working with third-party commercial indexing services, such as JSTOR, ProQuest, EBSCO, or Web of Science. Most often journals are approached by these services to join and the criteria for inclusion is not transparent. These commercial indexing services create subject-specific indexes that may or may not be picked up in an LSP's central index. In addition, these commercial indexes are often subscription-based, so journals participating in these will sometimes receive royalties, but it is often a very small amount, and there is little transparency in how royalties are calculated. This type of indexing might contribute to a few journals being searchable from a library's discovery layer, but these opaque processes mean that a lot of journals fall through the cracks.

Given the importance of article-level indexing for journal discovery in library catalogues, how do we circulate article-level metadata of newer, less established, esoteric, or even defunct journals? These are the gaps that libraries often find themselves trying to fill. Fortunately, most of the technical infrastructure necessary for filling this gap is already built into OJS via the OAI-PMH API.

JATS

Ingestion procedures for each indexing service vary with some requiring journals to manually submit new content, while others will automatically harvest from an OAI-PMH API. Commercial indexes will have a team of content analysts that will ensure metadata conforms to their standards and even enhance it in ways that might increase discoverability within their indexes. Some LSPs like those offered by ProQuest allow for automated local indexing via OAI-PMH and recommend using JATS [https://jats.nlm.nih.gov/] as the preferred metadata format for ingestion because it ensures the most description.

JATS is vastly more descriptive than the alternate metadata formats offered via the OJS OAI-PMH API and as a result far better suited for the needs of indexing and preservation. Unlike Dublin Core or MARC, JATS is a schema that includes tags for describing top-level metadata elements like ISSNs, contributors (types and their affiliations), source information that demarcates issue, volume and page numbers, etc. It also allows for the description of the journal itself, supplementary content, internationalization, and most importantly the body of the text. Although the JATS plugins in OJS do not necessarily include the body text in properly formatted JATS, having the text extracted from the PDF

and included in the body tags greatly improves the potential relevancy ranking if the indexing service is using the full text of the article. All of the extra descriptive elements in JATS are what make the metadata far more valuable for discovery and offers the most complete picture of a journal and its content.

Dublin Core Metadata (oai_dc)

Title	Argument Explanation Complementarity and the Structure of Informal Reasoning
Author or Creator	Mayes, Gregory Randolph
Subject and Keywords	Keywords
Subject and Keywords	argument
Subject and Keywords	evidence
Subject and Keywords	explanation
Subject and Keywords	cause
Subject and Keywords	complementarity
Subject and Keywords	editorial.
Description	Abstract: Argument and explanation are distinct forms of reasoning with an underappreciated complementary relationship. In this essay I define these terms precisely, identify the mischief that results from conflating them, elucidate their complementary relationship and employ this relationship to provide a fruitful approach to analyzing the logical structure of the common editorial. Keywords: argument, evidence, explanation, cause, complementarity, editorial.
Publisher	University of Windsor
Date	2010-03-19
Resource Type	info:eu-repo/semantics/article
Resource Type	info:eu-repo/semantics/publishedVersion
Resource Type	addendum
Resource Type	Conceptual analysis and teaching studies/advice
Format	application/pdf
Format	application/pdf
Resource Identifier	https://informallogic.ca/index.php/informal_logic/article/view/419
Resource Identifier	10.22329/il.v30i1.419
Source	Informal Logic; Vol 30 No 1 (2010); 92-111
Source	2293-734X
Source	0824-2577
Language	eng
Relation	URL *URL not shown as it is very long*.
Relation	URL *URL not shown as it is very long*.
Rights Management	Copyright (c) 2014 Informal Logic

An article record represented in Dublin Core

```
<title-group>
    <article-title xml:lang="en" >Argument Explanation Complementarity and the Structure of Informal Reasoning</article-title>
</title-group>
<contrib-group content-type="author" >
    <contrib corresp="yes" >
        <name name-style="western" >
            <surname>Mayes</surname>
            <given-names>Gregory Randolph</given-names>
        </name>
        <xref ref-type="aff" rid="aff-1" ></xref>
        <email>mayesgr@csus.edu</email>
        <uri>www.csus.edu/indiv/m/mayesgr</uri>
    </contrib>
</contrib-group>
<aff id="aff-1" >
    <institution content-type="orgname" >California State University Sacramento</institution>
</aff>
<pub-date date-type="collection" >
    <year>2010</year>
</pub-date>
<pub-date date-type="pub" publication-format="epub" >
    <day>19</day>
    <month>03</month>
    <year>2010</year>
</pub-date>
<volume seq="3" >30</volume>
<issue>1</issue>
<issue-id>353</issue-id>
<fpage>92</fpage>
<lpage>111</lpage>
<permissions>
    <copyright-statement>Copyright (c) 2014 Informal Logic</copyright-statement>
    <copyright-year>2014</copyright-year>
</permissions>
<self-uri xlink:href="https://informallogic.ca/index.php/informal_logic/article/view/419/4395" content-type="application/pdf" ></self-uri>
<self-uri xlink:href="https://informallogic.ca/index.php/informal_logic/article/view/419/2364" content-type="application/pdf" ></self-uri>
<self-uri xlink:href="https://informallogic.ca/index.php/informal_logic/article/view/419" ></self-uri>
<abstract>Abstract: Argument and explanation are distinct forms of reasoning with an underappreciated complementary relationship. In this essay I define these terms precisely, identify the mischief that results from conflating them, elucidate their complementary relationship and employ this relationship to provide a fruitful approach to analyzing the logical structure of the common editorial. Keywords: argument, evidence, explanation, cause, complementarity, editorial.</abstract>
<kwd-group xml:lang="en" >
    <title>Keywords</title>
    <kwd>Keywords</kwd>
    <kwd>argument</kwd>
    <kwd>evidence</kwd>
    <kwd>explanation</kwd>
    <kwd>cause</kwd>
    <kwd>complementarity</kwd>
    <kwd>editorial.</kwd>
</kwd-group>
<counts>
    <page-count count="20" ></page-count>
</counts>
</article-meta>
</front>
<body>
    <p>Argument-Explanation Complementarity and the Structure of Informal Reasoning Abstract: Argument and explanation are distinct forms of reasoning with an underappreciated complementary relationship. In this essay I define these terms precisely, identify the mischief that results from conflating them, elucidate their complementary relationship and employ this relationship to provide a fruitful approach to analyzing the logical structure of the common editorial. Keywords: argument, evidence, explanation, cause, complementarity, editorial. 1. Introduction Competent inquiry requires an intuitive grasp of the difference between argument and explanation, two forms of reasoning which are easily conflated. Mastering this difference also makes it possible to see how arguments and explanations are related. In natural language these forms of reasoning tend to occur together and they normally exhibit an interesting complementarity. This fact is not widely appreciated, nor is the reason for it. In this paper I will briefly summarize the difference between argument and explanation, and then illustrate their complementary relationship by showing how it is modeled in ordinary reasoning contexts, notably the common editorial. Ultimately my point is to show that when analyzing ordinary reasoning, it is a good idea to do so with the expectation of having to disentangle logically distinct but complementary arguments and explanations. When this expectation is not met, it will sometimes point to a fundamental weakness in the reasoning provided. 2. The difference between argument and explanation Argument and explanation are two different forms of reasoning, so let's begin by being clear on what we mean by that. Reasoning is just the process of making certain statements, which we call reasons, in support of other statements, which we call conclusions. This relationship may be visualized as follows: Reason Conclusion We are called upon to support our
```

An article record represented in JATS.

The fullness of JATS also increases the complexity of the metadata and as a result the potential for problems and omissions. For the descriptive power of JATS to be utilized there needs to be an effort to ensure that journal metadata is entered properly. Within the context of the work done by librarians who deal with metadata issues post-publication, there need to be mechanisms in place where problems can be easily identified retroactively.

Given that harvesting metadata from OJS is often an automated process, spotting issues is often impossible until they show up on a different platform. Libraries often lack resources like teams of content analysts that can correct any incoming or outgoing data, but fortunately, a lot of this analysis could be easily automated and generated into a tabular report.

Completeness

What we mean by completeness is, "does the metadata produced by a journal and retrieved from OJS meet a certain threshold in order to achieve maximum impact within

an index"? What should the basic requirements be? These will differ depending on the content type and a lot of the fields like the journal-title will be populated by default, but for a typical article the presence and accuracy of essential fields, or the lack thereof, need to be reported. Some of these fields may be obvious and mentioned elsewhere in this report while others might not be essential, but knowing about their presence and value is still important. Below is a list of JATS fields that should be considered *from the perspective of library management systems and cataloguing*.

Journal-level Metadata

JATS Element	**Requirement**	**Notes/possible points of friction**
journal-title	Yes	This is a default field populated automatically, but is it correctly translated if the journal is multilingual?
abbrev-title	Yes	Is the abbreviated title entered, along with its translations?
issn	Recommended	Strongly recommended for any active journal.
publisher-name	Yes	Who is the publishing body?
publisher-loc	Recommended	Where is the publishing body located?
self-uri	Recommended	The journal's URL. Useful for linking back to the journal's homepage. This is currently not included in the OJS JATS crosswalk.
notes	Recommended	A brief description of the journal can be included in the <notes> field. Not included in the OJS JATS crosswalk.

Article-level Metadata

JATS Element	Requirement	Note
article-title	Yes	Are translations included if published in more than one language?
contributors	Yes	Are the contributor names and their roles displaying correctly?
aff	Recommended	Affiliation information is also very valuable metadata that should not be overlooked.
abstract	Yes	Abstract is an essential field for indexing. Are applicable translations included?
kwd	Yes	Keywords are also essential for indexing. Are applicable translations included? These often get omitted.
volume	Recommended	Is all issue-level metadata accessible for each issue? This is necessary for aggregating content into volume / issues, but it is not always displayed.
issue	Recommended	Is all issue-level metadata accessible for each issue? This is necessary for aggregating content into volume / issues, but it is not always displayed.
pub-date	Yes	This should at least include a year.

fpage & lpage	Recommended	Pagination should be included where possible.
self-uri	Yes	OJS articles will have at least two <self-uri> elements. One for the article page and one for the galley file, such as a PDF. Does the galley file URI work? Does it have the correct content-type attribute?
supplementary-material	Yes, if applicable	If the journal has supplementary material (data, figures, research instruments), do all of the corresponding elements display correctly? Does the <self-uri> link work?
body	Yes	In the context of OJS, this is the extracted full-text of the article. This field will not be included if PDF indexing is not enabled on the OJS instance. Full-text indexing will contribute to higher relevancy ranking.
ref-list	Recommended	A lot of indexes will cross link on references, so having a reference list can greatly improve relevancy ranking.
license	Recommended	Is the article open access? Are there any specific licenses assigned?

Consistency

One of the bigger struggles in managing multiple journals as a library service provider is consistency across titles. As it stands, there is no easy way to get a birds-eye view of existing metadata in the way that you might in repository software like Dspace. Seeing a report that lists all of a journal's metadata in one place would make it easier to follow or change the conventions a journal uses. Likewise, editors, journal managers, or librarians could quickly identify problems and correct them.

There are quite a few journals that have been using OJS for nearly two decades and editorial boards and journal managers change often, sometimes even with each issue. The time that faculty have to devote to editorial work can be quite limited so they might not notice or care that an issue published ten years ago has broken or missing galley files.

Another common problem is that issue and volume naming conventions are not consistent. The benefit of consistency here is that it makes it easier to aggregate journal data into volumes and issues in the same way that it was published. Consistent conventions make it a lot easier to link to or find a referenced article within a particular index. Tools to help get a sense of consistency would be a significant help to service providers.

Libraries and Metadata Conclusion

Like publishers, libraries also rely on article indexes for making journals discoverable to researchers and the fuller the metadata the more visibility they can achieve. Of course, as the requirements for indexing become more complex, the more difficult it will be to ensure that a journal's metadata is meeting the basic requirements of the indexing service.

The lack of understanding of how indexes operate and the downstream effects of poor metadata practices will inevitably have a detrimental effect on how and where an article appears in the results of a search. Having a bird's-eye view of how each article meets a basic set of requirements would save time for librarians working in journal publishing and encourage better article findability.

Inclusion and Equity in Metadata

The Equity and Metadata subgroup was tasked with identifying known and potential issues related to equity, diversity and inclusion in publication metadata. It is clear that both libraries and the publishing industry have an increasing need to review problematic or absent metadata fields with regards to identity and cultural expression. While it is well beyond the scope of Coalition Publica (let alone this working group) to shape these narratives internationally, it is vital to acknowledge them.

Subgroup members were instructed to create a report that would be relatively short and identify – with some context – shortcomings and sensitive issues in publishing metadata. Suggested issues (this list was not intended to be definitive) were:

- support for Indigenous place names in location metadata
- support for non-Western naming conventions
- support for multilingual metadata
- support for alternative or multiple names
- support for pronoun statements

The group was explicitly told that “these are not problems this working group needs to solve”.

Background and General Issues

It is now recognized within libraries that metadata is a subjective practice that is in need of ongoing review and decolonization. Farnel notes that “the application of ethical principles to all aspects of information work (Brody, 2002), emerged within the library and information science literature in the late 1980s (Samek, 2007)” but that this concern for ethical practices (which extends to metadata) intersects with another area of inquiry, “the growing recognition of the subjective nature of metadata practice.” (Farnel, 2018, p.2) More specifically, the “important role played by context and culture (Alemu & Stevens, 2015; Mai, 2013; Srinivasan, 2017), the biases inherent in many of the widely adopted metadata tools (Bowker & Starr, 1999; Olson, 1998, 2002), and the ability of metadata to reinforce power imbalances and perpetuate structural problems such as sexism, racism, and colonialism (Drabinski, 2013; Furner, 2007; Olson, 1999; Yeh, 1971) are now well documented and for the most part accepted within the discipline.” (Farnel, 2018, p. 2)

Who "holds the pen" when it comes to metadata is in itself an important question. Duarte and Belarde-Lewis (2015) discuss the idea that cataloguing's power is largely ascribed to the "power to name". (p. 681) So perhaps we need to consider not just the practices around metadata, but with whom lies the "power to name" or ascribe metadata. Perhaps accountability in metadata needs to be considered as well?

Decolonizing Classification

The Canadian Federation of Library Associations (CFLA)'s *Truth & Reconciliation Committee Report and Recommendations* includes the following as one of its ten recommendations: (#5) "Decolonize Access and Classification by **addressing the structural biases** in existing schemes of knowledge organization and information retrieval arising from colonialism by **committing to integrating Indigenous epistemologies** into cataloguing praxis and knowledge management." (Canadian Federation of Library Associations, p. 6, emphasis added)

CFLA's *Truth & Reconciliation Committee Report and Recommendations* further recommends the following larger actions in regards to decolonizing classification processes:

- **Acknowledging the structural biases and inadequacies in existing schemes of knowledge organization** and information retrieval arising from colonialism;
- **Adopting an ethic based upon the commitment to integrating Indigenous and Western knowledges into access, arrangement, description, classification and cataloguing praxis**;
- **Engaging with their user communities, particularly Indigenous communities**, in integrating regionally-relevant Indigenous knowledges into their cataloguing practice, arrangement, description, etc. including descriptive metadata;
- **Providing staff training on culturally responsive access praxis**, including the incorporation of Indigenous knowledges into library subject guides, archival finding guides, special collections and digital infrastructure.
- **Ensuring that these efforts occur at the local, regional, provincial, national and international levels**. (Canadian Federation of Library Associations, p. 28, emphasis added)

One area in which libraries have made meaningful strides in recent years is in integrating Indigenous place names into cataloguing records, as most mainstream place names in Canada reflect a history of colonialism and erasures of Indigenous place names. This history is also a history of erasure of different epistemic knowledges about place. However, as very few journals contain metadata to describe place names, this particular aspect is less relevant for Coalition Publica.

However, such information can be included in subject headers, which is relevant for Coalition Publica. For example, the Inuvialuit Digital Library saw the benefit of implementing **faceted and more detailed geographic locations**, moving "from Folk songs, Inuvialuit--Canada--Northwest Territories--Inuvialuit Settlement Region (ISR) to separate entries for Folk songs, Inuvialuit, and each of the geographic subdivisions." (Farnel & Shiri, p.11) The Inuvialuit Digital Library had to develop its own list of place names (both traditional and westernized), capturing **alternative spellings** and **dialect variations**. (Farnel & Shiri, p.11)

In reviewing their metadata practices, it would be useful for Coalition Publica's leadership, as well as journal publishers and editors to familiarize themselves with the First Nations Principles of OCAP (Ownership, Control, Access, and Possession), which are "a set of standards that establish how First Nations data should be collected, protected, used, or shared. They are the de facto standard for how to conduct research with First Nations." (OCAP website).

Another practice that has emerged when working with Indigenous communities and content is the use of Traditional Knowledge (TK) Labels. These have been used mostly to supplement metadata in libraries' and archives' digital collections records. Working collaboratively with the Indigenous communities where the content originated, community-specific tags are adopted that allow the records to incorporate recognition for aspects not typically found in Western metadata conventions.

In cases where an article stems from research involving First Nations communities or intellectual property, there is a duty to appropriately acknowledge and name contributions that are not typically done in Western research. Examples of types of contributor roles that could be considered essential by Indigenous communities could be: translators, storytellers, illustrators, drummers and dancers. (Farnel & Shiri, p. 11)

It is important to remember that many Indigenous communities consider that knowledge is community-owned, not owned by an individual. This needs to be taken into consideration if individuals are being asked to declare copyright ownership over documents or objects that may include Indigenous community-owned content. Going forward, if following OCAP principles, it will be important to reference community protocols and access limitations, and add new types of ownership or copyright statements. It is therefore likely that similar practices which are taking place in cataloguing and in digital collections metadata will become adopted by journals that feature content about and produced through collaboration with Indigenous communities.

Recommendations

- Encourage the use of regionally-relevant Indigenous place names in addition to colonially ascribed names.
- Include any alternative spellings and dialect variations.
- Ensure that location elements allow for multiple names at multiple levels.
- Allow flexibility for additional creator roles beyond author – these might include storyteller, translator, dancer, and communities.
- Follow community knowledge protocols to ensure Indigenous knowledge that is included in publicly facing metadata is appropriate for public display.
- Support for potentially unique copyright for Indigenous community-owned content.

Support for Non-Western Naming Conventions

The inclusion of non-Roman characters and non-Western naming conventions allows for metadata to be expressed in culturally appropriate ways, particularly for personal names and locations. Additionally, many names are not expressed in a "GivenName FamilyName" convention, which can mean incorrect name display and citations if the metadata structure assumes Western naming conventions. Options to include additional languages for certain elements, and *specify the relationship between the languages*, may help with some of these issues.

Identified Issues

- **Non-Roman character sets** include Arabic, Chinese, Cyrillic (Russian, Serbian, etc.), Greek, Hebrew, Japanese, Korean, Tamil and Thai script.
- **Library of Congress (LC) name authorities** provide non-Roman references for non-Roman entries for personal names, corporate bodies and locations, which could be the source for adding standardized non-Roman entries in OJS.

Recommendations for OJS Metadata Schema

- Option to enter a second language for author name.
    - Include descriptors that identify each version of the author's name.

Support for Multilingual Metadata

Multilingual metadata deals more with information access and retrieval than UX/UI concerns (though these are also important in the Coalition Publica context). Multilingual metadata should be distinguished from a multilingual interface or multilingual software. It ultimately aims at allowing users to search and retrieve materials using their own language. Several models exist, which "vary from human translation of metadata records and the development of multilingual-controlled vocabularies to automated methods, such as machine translation (MT) and other cross-language information retrieval methods (CLIR)." (Matusiak et. al., p. 136)

Automating multilingual metadata requires advanced IT development. Manually providing multilingual metadata requires significant human resources. This report does not seek to find answers to the complex multilingual challenges but will point out possible actions that could be small steps towards multilingualism:

- The following metadata elements must be prioritized for multilingualism: titles, personal name, corporate name, abstracts, keywords/subject headings
- The xml:lang attribute must be defined in relevant metadata, for the prioritized metadata elements at a minimum.
- For subject headings, when possible allow use of Uniform Resource Identifiers (URI). Some controlled vocabularies are expressed as linked data ontologies and are multilingual (i.e. SKOSS RVM [in development], COAR Controlled Vocabularies).
- NISO JATS Tag Sets provide a way to implement multilingual metadata that can be used as a model, though "1) identification of the language using the xml:lang attribute, 2) repetition of some structures to enable these structures to be present in more than one language, and 3) enclosing some repeated structures in a single wrapper element." (Lapeyre)
- Consider making the EruditArticle XML schema available in languages other than French, particularly English, or move to a non-proprietary, internationally supported open standard, such as JATS (which may also need to be encouraged to support additional languages).

While the above points do not take into account the UI/UX integration of multilingualism for in OJS and on Érudit, this is an important point for further consideration. Do these platforms allow journals to display multilingual metadata in a way that makes sense for their readership? For instance, could an element be displayed in the language of the

interface with the possibility to switch languages at the element level? There may be other scenarios to consider as well, for example a journal may wish to keep their OJS UI in a single language, but would still like to display translated metadata, such as abstracts.

Support for Alternative or Multiple Names

There are numerous reasons for authors to have multiple names either concurrently or throughout their lives – marriage/ divorce, "deadnames"[1], "government names" vs other chosen or given names, titles that reflect clan roles, traditional or kinship names, stage names or pseudonyms, and many other name changes. In some circumstances, authors may wish to have works under multiple names linked; in other cases, they may not.

These issues are, in large part, the focus of author ID services like ORCID. An ORCID iD is an identifier that aims to disambiguate different authors with similar names, to clarify which work is theirs across borders, disciplines, and time. While OJS does support ORCID integration, it is worthwhile to explore how the software may accommodate multiple names without relying on an ORCID membership.

Identified Issues

- Pseudonyms have established practices for Name Authority Records: https://resourcedescriptionandaccess.blogspot.com/2015/03/rda-cataloging-rules-for-pseudonyms.html, https://www.loc.gov/catdir/cpso/pseud.pdf
- JATS provides some flexibility, however:
    - "alternative name" is limited in scope (https://jats.nlm.nih.gov/archiving/tag-library/1.2/element/name-alternatives.html); and
    - "name-style" is limited to Western, Eastern, Given-only, and Islensk (Icelandic) configurations (https://jats.nlm.nih.gov/archiving/tag-library/1.3d1/attribute/name-style.html).
    - While these may be the most common, they are not the totality of naming convention options.

[1] "Deadnaming is the use of the birth or other former name of a transgender or non-binary person without their consent." Wikipedia.

- ORCID uses a single-string for names, and this may be the best-case scenario, however, this does not address machine parsing of names for outside services like reference/citation managers.
- The Library of Congress (LC) practice of assigning names dates [particularly for trans people] (eg, "X" 1954-1974, "Y" 1975-) may not match the person's lived experience of their name. (Billey A. et al.)
- Married names are an example that suggests indexing using multiple author names, with cross-references. Abbreviated names are a hindrance to this. (Pellack & Kappmeyer)
- There are projects that allow author identification across systems – e.g., the OCLC Cooperative Identities Hub and ResearcherID. Authors can self-manage their record by including former names on new publications, updating these projects. (ibid)
- Re: "Deadnames" for trans people – journals should accommodate requests to retroactively change names, as discussed in the article *Is there a way to edit author name after publication for trans people, (at least online)?* Also, the Committee on Publication Ethics (COPE) released the following vision statement in January 2021: https://publicationethics.org/news/vision-more-trans-inclusive-publishing-world
- Name Authority Records that record deadnames "out" people may be harmful. (Billey & Drabinski)

Recommendations

- The JATS alternative name schema could be useful, but could be expanded to allow conventions beyond the four identified "name-styles"
- Single-string names allow names to be entered as desired, but without additional information may not be easily parsed by reference and citation managers.
- Alternative name fields should be unlimited, and should not be used without the author's express permission
- Integration with persistent identifier services may help all works by the same author be claimed despite differences in their name
- It is recommended that journals accommodate requests to retroactively change names and systems should easily accommodate this.

Support for Pronoun Statements

Many people use pronoun statements to identify their gender unambiguously. This can help to ensure everyone is referred to in an appropriate way. Gender cannot be assumed

from names or appearances. However, identifying the gender of authors may also cause harm and be an unnecessary disclosure.

Identified Issues

- There are many different gender expressions, identities, and pronouns, and complete "lists" may not be desirable. There are difficulties with gender essentialism, definition of "concepts" (but also difficulty with practically implementing a non-concept based schema). (Fox, Keilty)
- There is an inherent difficulty with leaving gender determination to cataloguers. There is a MARC tag (375) and RDA standard for gender, which implies that LC is comfortable with cataloguers assigning this based on inference from appearance or name. However, gender can change, and cannot be assumed from names or appearances: "Critical to queer theory is a resistance to social practices that freeze identities in time and universalize them, erasing the real differences that accompany same-sex sexuality on the scales of time and place." (Billey, A. et al.)
- Using gender/pronouns as a searchable field may allow for data "discovery" that is unnecessary. (ibid)
- Sometimes, in languages with gendered nouns, requested gender is a way to prevent mis-gendering elsewhere in the system (eg. rédacteur, rédactrice).
- Pronoun Island is a useful tool for self-definition of pronouns: https://pronoun.is/
- ORCID iD and ResearcherID don't specifically record gender or pronouns, just variations in name.

Recommendations

Currently we do not recommend identifying authors' genders. An optional author's prefix field could be considered, if it included numerous gender variations (eg, Mx). Alternatively, because pronouns are not – to our knowledge – commonly recorded metadata, it might be best for authors wishing to self-identify do so in their biography fields in OJS.

Other Considerations

Names are a common element of metadata, but they are not always personal names. Corporate names, affiliations/organizations, and unnamed entities could also be the progenitor of works.

Recommendations

- It should be possible to have an authorless publication.
- Corporate or organization names should be explicitly supported so that they can be accurately attributed in downstream metadata (like Crossref or Google Scholar).

Equity and Metadata Conclusion

The subjective nature of metadata, and the high likelihood of bias in structures that describe people, places, and things means that metadata is a natural site of struggle. We need to find a balance between enabling equity-seeking efforts via technological structures and metadata recommendations, while still allowing for some degree of structure and machine interoperability. Technological constraints should never be an excuse to diminish someone's personhood, or inaccurately reflect their identity.

Recommendations

What follows is a summary of all currently (as of Winter 2021) relevant recommendations based on the work of the two subgroups above as well as the issues identified in the interim report from 2019 that have yet to be addressed.

Mechanisms for Metadata Review

Harvesting metadata from OJS is often an automated process, so identifying problems is often impossible until they show up on a different platform or are reported by users. Libraries often lack resources to adequately correct any incoming/outgoing data, but a lot of this analysis could be easily automated and generated into a tabular report for each review. (See: "Completeness")

Broader, Optional Adoption of More Granular JATS Elements

Acknowledging that increased granularity is also likely to increase the likelihood of user error, optional (non-mandatory) metadata could be added to better map to useful JATS elements not currently recorded, such as: body (article metadata), self-uri (title-level), and notes (title-level), and season (issue-level). (See: "Completeness").

Season, in particular, is used in both JATS and Crossref's schema. It is also one of the leading causes of abuse in the issue-level "title" field.

Review of "title" Element for Issue Metadata

One of the more common issues with journal metadata is that issue-level metadata is either duplicated or abused for display purposes. "Title" at the issue-level is not well defined within OJS. Most of the time, this field is used as a way to circumvent basic display of volume and issue metadata (for example, replacing Vol. 1 with Volume 1). Because the metadata for volume and issue is already stored, you end up with issue-level metadata both in discrete elements *and* as a single string in a title field.

The applications of this "title" field should be made more clear to users in OJS documentation, or within OJS itself. Possibly with in-line examples for things like "special issues". Another useful addition would be the aforementioned "season" element.

Support for Decolonizing Classification

Acknowledging that OJS does not offer a high volume of location metadata, it is nevertheless important to allow for what *does* exist to facilitate indigenous or alternate metadata fields as outlined in the Truth and Reconciliation Commission guidelines. (See: "Decolonizing Classification")

Support for Broader Author Roles

Indigenous roles in works may well extend beyond author and translator to include roles like: storyteller, illustrator, drummer or dancer. We have also identified in issues related to book review metadata that overall metadata fidelity may improve with a clear delineation between "review author" and "book author." Custom or additional author roles may adequately address these issues. Additionally, these roles should be adequately displayed in article metadata. Currently, for example, there is no distinction in OJS between a translator and an author on article displays and in downstream metadata. (See: (See: "Decolonizing Classification")

Support for Multiple/Alternative Name Fields

OJS 3.2 does have support for a preferred name field, but it only allows *one*. It is possible for authors to have more than one preferred name and this field should be repeatable. This may well be related to a potential solution for Support for multilingual "author" fields below. (See: Support for alternative or multiple names)

Clear Guidelines for Multilingual Journals

Given the variety of approaches to multilingualism and translation in publication metadata, it would be useful for Coalition Publica to develop explicit multilingual guidelines for journals. Additionally, resources should be made available to help journal editors meet these guidelines (eg. documentation, support, translation and consultation).

Reevaluation of Locale Options/Settings

One of the issues identified by working group members is a tension between the options available for multilingual journals and their ability to deliver multilingual content and metadata reliably. A journal may wish to support french users by enabling french UI, but

only have the resources to write journal copy in english. This can be frustrating to french users who only see small portions of a website translated.

One potential solution could be to break up the UI display options to stipulate "journal view" with "dashboard". A multilingual dashboard might be more useful (and more reliably translated via locales) than a multilingual page view. It's clear that the available options may actually be more frustrating for users looking to find content outside of their primary language.

Support for Multilingual "author" Fields

Authors who work in multiple languages may wish to identify themselves in more than one language or script. For example, a name in Chinese script would differ from a name in Roman script. OJS language/locale options can/should extend to the author name field. (See: Support for alternative or multiple names)

Simultaneous Display of Multilingual Metadata

One of the most frequent issues with metadata is that users will enter more than one language into a single field because they want both to display. If users could selectively display more than one language for *some* metadata elements, we would certainly see a decrease in metadata abuse. Abstracts are an excellent example. It is very common to see two languages in a single abstract field because a journal wants to show both abstracts on an article landing page. For example, an English-language journal may have very little additional metadata or website copy available in French, but they want to provide French abstracts for English articles so that their Francophone readers can decide if they want to take the time to read an article in English. This example could be equally true for a French-language journal, or for other metadata fields such as title and keywords.

Promotion of Self-identification in Biography Statements

The subgroup looking at equity issues does not recommend fields for self-identifying pronouns or gender at this time. However, highlighting the possibility of including pronouns in the copy for a biography statement would be a great gesture. (See: Support for pronoun statements)

Biographical Statements Pushed to JATS

Currently, biographical statements are not pushed to JATS through OJS's crosswalk, but there is support for this in the JATS schema.

Expansion of Name Types/Author Options

OJS does not distinguish between corporate/organizational authors and individuals. While it *does* allow for a single preferred name or given name, downstream metadata schema support author types in attributes. OJS needs to allow for a corporate or organization author type. Additionally, it is absolutely necessary to allow for empty or null author metadata, for any text that does not have or does not need an author, such as errata, anonymously authored texts, announcements etc. Any required field without real, accurate metadata promotes the creation of fake or inaccurate metadata. (See: Any other considerations, Interim Report)

Inclusion of Editor(s) in Issue-level Metadata

All editorial records in OJS are ephemeral and unassociated with issue-level and title-level metadata. Encoding editorial roles in issue-level metadata would preserve editorial information over time and this sort of metadata is common in library catalogues and other indexing services.

This is further complicated by the conflation of required roles in the OJS workflow that may not match real-world roles held by specific users. For example, many people may be assigned the role of "journal manager" because they need the corresponding level of access. As such, this issue-level metadata should not be tied to roles assigned within OJS. (See: Interim Report)

Divorce Submission Mechanisms from Article Metadata

The most obvious example of this need is with required email addresses. Understandably, email addresses have been necessary for the way OJS operates. But, it is *very common* for users to fabricate fake email addresses to clear a screen. If article metadata could be (even optionally) unhooked from user profile metadata, it would allow for more fulsome and accurate recording of *actual metadata*. (See: Interim Report)

More Flexibility for Small Changes to Metadata

OJS 3.2 introduced a feature of locking metadata edits once material is published. The intent of this change makes good sense, but it should be possible to make small changes to article metadata without creating a new galley/version or unpublishing content.

Alternate Journal Title

It's not uncommon for journals to modify the metadata of their title for identity purposes that, ultimately, cause problems with indexing or harvesting. A good example of this is journals that want to put both English and French titles in one title for prominent display. JATS has a field for alternate title that could prevent this sort of issue, much in the way "preferred name" is meant to override given/last: https://jats.nlm.nih.gov/publishing/tag-library/1.2/element/alt-title.html. (see: Interim Report)

Addition of Explicit "Journal History" Field

It is important that the metadata of the journal be up to date. Unfortunately, this means that important information about the history of the journal may be lost. Many journals keep no track of:

- The former ISSNs of the journal and the period during which they were in use
- Changes in the title of the journal
- Changes in the license used by the journal
- Changes in the editorial team
- Changes in the publisher
- Etc.

This is technically possible to include in the "About the Journal" section but proactively including the field may encourage this important record keeping for journals that may not think to include it.

Addition of "season" Element to Issue Metadata

Both JATS and Crossref schema account for a seasonal element in issue-level metadata. Most OJS users put this information in the issue title, but when you combine seasons with the existing issue-level metadata, the display can be confusing or unintuitive for users. The

addition of a season element would likely circumvent many of the concerns editors have in this space that lead to abusing the title field.

Accessibility Review of Metadata Workflows

A UI/UX review of metadata forms is already well on PKP's radar but it was noted that the only indication in the metadata UI for multilingual completeness is a coloured logo. It is green if fields are filled out and fuschia if not. But, with an identical logo, this might provide confusion for colourblind users. It is recommended to disambiguate these statuses in a way beyond just colour.

Update QuickSubmit Plugin to Match Current Metadata Requirements

While OJS has made a number of significant and appropriate changes to the metadata collected with submissions, the Quicksubmit plugin remains quite out of date. It is missing, for example, the preferred name field. The Quicksubmit plugin should be updated to match the standards of quality elsewhere in the application.

Promotion of Consistent Journal-level Metadata

Last but not least, the single most obvious problem in all OJS is metadata is one of consistency. This is in *so many ways* not the fault of the software so much as it is a reality of the time and resources of editors and authors. More attempts should be made to promote metadata literacy within the software itself, and go beyond the scope of Coalition Publica's work with Canadian journals. This could be with more explicit copy or with better documentation. There are, of course, limits to this recommendation. The metadata working group acknowledges that PKP can only do so much, and often the issue of inconsistent metadata is well out of their hands.

References

Bantin, P.C. (Ed.). (2016). *Building Trustworthy Digital Repositories: Theory and Implementation*. Lanham, MD: Rowman & Littlefield.

Barnett, J., Lovins, D., Novak, A., Riley, C., & Suzuki, K. (2010). *Investigating Multilingual, Multi-script Support in Lucene/Solr Library Applications.* http://hdl.handle.net/2451/38726

Berkeley Library, University of California. (2016, December 7). Vernacular Parallel Script Standards. https://asktico.lib.berkeley.edu/vernacular-parallel-script-standards/

Billey, A., & Drabinski, E. (2017). *Cataloging, Gender, and RDA Rule 9.7*. ALCTS webinar. Retrieved October 22, 2020, from http://www.ala.org/alcts/confevents/upcoming/webinar/031517

Billey, A., Drabinski, E., & Roberto, K. R. (2014). What's Gender Got to Do with It? A Critique of RDA 9.7. *Cataloging & Classification Quarterly*, 52(4), 412-421. https://doi.org/10.1080/01639374.2014.882465

Billings, L., Llamas, N.A., Snyder, B.E., & Sung, Y.. (2017). Many Languages, Many Workflows: Mapping and Analyzing Technical Services Processes for East Asian and International Studies Materials. *Cataloging & Classification Quarterly*, 55(7-8), 606–629. https://doi.org/10.1080/01639374.2017.1356783

Canadian Federation of Library Association. (2018, October). *CFLA Truth & Reconciliation Committee Report and Recommendations*. Retrieved March 31, 2020, from http://cfla-fcab.ca

Duarte, M. E., & Belarde-Lewis, M. (2015). *Imagining: Creating spaces for Indigenous ontologies*. *Cataloging & Classification Quarterly*, 53(5/6), 677-702. http://dx.doi.org/10.1080/01639374.2015.1018396

EL-Sherbini, M. (2018). *Improve Discoverability of Non-Roman Materials.* ALA Webinar. Retrieved October 22, 2020, from http://www.ala.org

Érudit. (2019, October). *Publishing Schema Documentation*. https://erudit-ps-documentation.readthedocs.io/en/latest/

Farnel, S., & Shiri, A. *Community-Driven Knowledge Organization for Cultural Heritage Digital Libraries: The Case of the Inuvialuit Settlement Region*. Advances in Classification Research Online. http://dx.doi.org/10.7152/acro.v29i1.15453

Farnel, S. (2018). Metadata as Data: Exploring Ethical Metadata Sharing and Access for Indigenous Resources Through OCAP Principles. *Proceedings of the Annual Conference of CAIS*. https://doi.org/10.29173/cais974

FNIGC/CGIPN. *OCAP (Ownership, Control, Access, and Possession) framework*. Retrieved October 22, 2020, from https://fnigc.ca/ocap

Fox, M.J. (2011). Prototype theory: An alternative concept theory for categorizing sex and gender? In Smiraglia, Richard P. (Ed.) *Proceedings from North American Symposium on Knowledge Organization*. http://dx.doi.org/10.7152/nasko.v3i1.12799

Keilty, P. (2012). *Sexual Boundaries and Subcultural Discipline*. Knowledge Organization 39.6, 417-431. http://hdl.handle.net/1807/78146

Kirschner, J. et al. Support for Multilingual Journals using Open Journal Systems. Retrieved February 1, 2021 from http://d-scholarship.pitt.edu/36615/1/LPF18_SupportMultilingualJournals.pdf

Is there a way to edit author name after publication for trans people (at least online)? (2020). Academia. Retrieved October 22, 2020, from https://academia.stackexchange.com/questions/144926/is-there-a-way-to-edit-author-name-after-publication-for-trans-people-at-least

Lapeyre DA, Usdin BT. (2011) Introduction to Multi-language Documents in NISO JATS. *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2011* [Internet]. Bethesda (MD): National Center for Biotechnology Information (US). Available from: https://www.ncbi.nlm.nih.gov/books/NBK62175/

Library of Congress. (2015). *FAQ – LC/PCC practice for creating NARs for persons who use pseudonyms*. Retrieved October 22, 2020, from https://www.loc.gov/catdir/cpso/pseud.pdf

Matusiak, Krystyna K. (2015) Multilingual Metadata for Cultural Heritage Materials. *The Electronic Library*, 33(1), 136-151. https://doi.org/10.1108/EL-08-2013-0141

Mellon, M. (2014). T*he Use of Cyrillic Metadata for Enhancing Discovery of Russian Digital Collection Items: A Case Study of the Bowman Gray World War I Postcards Digital Collection*. (Unpublished master's thesis). University of North Carolina at Chapel Hill. https://doi.org/10.17615/ma5n-m806

<name-alternatives>Name Alternatives. *Journal Archiving and Interchange Tag Library NISO JATS Version 1.2 (ANSI/NISO Z39.96-2019)*. Retrieved October 22, 2020, from https://jats.nlm.nih.gov/archiving/tag-library/1.2/element/name-alternatives.html

@name-style Name Style. *Journal Archiving and Interchange Tag Library NISO JATS Version 1.2 (ANSI/NISO Z39.96-2019)*. Retrieved October 22, 2020, from https://jats.nlm.nih.gov/archiving/tag-library/1.3d1/attribute/name-style.html

National Indigenous Knowledge and Language Alliance. (2019, June 21) *First Nations, Metis and Inuit – Indigenous Ontologies (FNMIIO)*. Retrieved March 31, 2020, from https://nationalindigenousknowledgeandlanguagealliance.home.blog/

Niininen, S., Nykyri, S. and Suominen, O. (2017). The future of metadata: open, linked, and multilingual – the YSO case. *Journal of Documentation*, 73(3), 451-465.

Pellack, L.J., & Kappmeyer, L.O. (2011). The Ripple Effect of Women's Name Changes in Indexing, Citation, and Authority Control. *Reference and Instruction Publications and Papers*. 15. Retrieved October 22, 2020, from https://lib.dr.iastate.edu/refinst_pubs/15

Pseudonyms - RDA Cataloging Rules with MARC 21 Examples of Name Authority Records. (2020). *Librarianship Studies & Information Technology*. Retrieved October 22, 2020, from https://www.librarianshipstudies.com/2020/07/pseudonyms-rda-cataloging-rules-marc-21-examples-name-authority-records.html

Public Knowledge Project. Using JATS with OJS. Open Journals Systems 3 Documentation. https://docs.pkp.sfu.ca/coalition-publica/en/using-jats

Smith-Yoshimura, K. (2018, November 7). Creating metadata for equity, diversity, and inclusion. *Hanging Together, OCLC Research Blog*. Retrieved March 31, 2020, from https://hangingtogether.org/?p=6833