A Class Focused Approach to Research Outputs and Policy Literature Metadata

Les Kneebone
Analysis & Policy Observatory

Abstract

Successful research object sharing requires that systems and users understand the structure, semantics and rules that govern a given research object collection.

A number of metadata standards define ontologies and vocabularies for consistent expression of research object semantics. Supporting, clarifying and sometimes extending these standards are metadata application profiles (MAPs). MAPs play a key role defining metadata element cardinality and data types. MAPs may also mandate or recommend controlled vocabularies, where metadata standards have not already mentioned these in formal range declarations, encoding schemes and semantics that are to be consumed by external systems. MAPs also guide design options for in-house systems and workflows. In this paper, development of a draft MAP for grey-literature policy and research collections is discussed. A focus of the discussion is the considerations around selection and adoption of metadata standards given the research data and literature communities in the APO stakeholder map.

This paper presents a work-in-progress version of a Dublin Core Application Profile (DCAP) candidate. The Analysis & Policy Observatory Metadata Application Profile (APO-MAP) takes research object class structure as a starting point and considers class model options, especially given the availability of registry services and Persistent Indenter (PID) systems. The discussion finds that MAP development progresses towards a best fit that balances the need to adopt widely supported standards, local business drivers, and community acceptance.

Received 07 December 2018 ~ *Accepted* 13 August 2019

Correspondence should be addressed to Les Kneebone, 400 Burwood Road, Hawthorn 3122 Australia. Email: lkneebone@apo.org.au

An earlier version of this paper was presented at the 13th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

The Analysis and Policy Observatory (APO) is a grey literature collection, or a repository "comprised of research and information resources produced and disseminated… by organisations, outside of the commercial or scholarly publishing industry focusing on public policy and research" (Lawrence, 2016). APO curates and tags these resources with contextual information, referred to here as metadata. While APO curators are active in selecting and describing resources, APO collection is also developed via user-contributions. APO metadata editing systems, then, are designed to accommodate infrequent and non-expert contributors (with a subset of administrator-reserved tasks).

Whether internally curated or externally contributed, populating resource metadata can be laborious and even error prone. User guides and system-level validation go some way to ensuring a consistent metadata authoring. But APO is meeting a tremendous challenge by managing grey literature, as it does not conform to familiar workflows, standards and rewards system associated with commercial and scholarly publishing, making sourcing, storage and cataloguing tasks more challenging (Lawrence, 2016).

As well as contributing grey literature in the form of policy documents and research reports, APO also allows users to create metadata records for research Organisations and Persons associated with research activities, datasets and publications. To this environment, APO has drafted a provisional object class Project that will include properties that describe attributes of research projects and other named research activities.

To support the expansion of metadata classes in APO repository, a key task is repurposing existing properties, or identifying additional properties needed to describe data in the new classes. Where new properties are needed, the Dublin Core (DC) approach is always scrutinised first. The rationale for preserving DC guidelines, including the Dublin Core Application Profile guideline (DCMI, 2009), is simply that Dublin Core has underpinned the APO approach since it first published exchangeable metadata. Dublin Core is understood internally and recognised externally. An attempt is made at APO to use and reuse metadata properties drawn from the best supported, most familiar and most robust international standards, while aiming for a good fit with APO metadata requirements – more often than not this is able to be solved with DC.

A DCAP can use any terms that are defined on the basis of RDF, combining terms from multiple namespaces as needed (DCMI, 2009). Nonetheless, APO aims to limit the proliferation of adopted ontologies so that its exportable data formats and services are not awash with excessive namespace declarations and inconsistent datatypes and obligations, therefore simplifying data consumption for both APO and client systems.

A key assumption in development of the APO-MAP is that linked-data applications will be able to consume APO metadata. Enabling linked data applications is somewhat implied in the DCAP guideline, which recommends that profile designers include those with "… an understanding of the Semantic Web and the linked data environment (DCMI, 2009). However, APO also needs to interoperate with non-Semantic Web collections and services. Therefore, identifying and characterising communities is an inescapable part of ontology adoption.

 doi:10.2218/ijdc.v14i1.640

APO-MAP

Method

The following criteria for evaluating ontologies have been used:

- **Uptake:** is the ontology in use by strategic partners?
- **Roadmap:** is there evidence that the ontology is maintained or subject to a review cycle?
- **Class scope:** does the ontology class model fit with APO class model requirements?
- **Data structure:** can identifier such as http URIs be given as data values within the ontology?
- **Vocabulary compatibility:** can key properties in the ontology be readily populated with distribute vocabulary services?

The ontologies surveyed were selected as result of an environmental scan. Most evaluated ontologies are considered to be 'research management' ontologies, with some generic ontologies also considered (e.g. DC; FOAF). The ontologies surveyed include:

- CERIF Ontology 1.3 (euroCRIS, 2013)
- DataCite Metadata Scheme v4.1 (DataCite, 2017)
- Data Catalog Vocabulary (DCAT) (W3C, 2014a)
- DCMI Metadata Terms (DCMI, 2012)
- FOAF (W3C, 2014b)
- MODS (Library of Congress, 2018)
- ResearchGraph schema v2.0 (Research Graph, 2018)
- RIF-CS schema 1.6.2 (ANDS, 2017)
- Scholix Metadata Schema 3.0 (Scholix, 2017)
- SKOS (W3C, 2009).

Communities

APO metadata impacts a number of different communities, some of which are named in this paper. Without identifying all collaborations and agreements, these communities can be caricatured as:

- APO's own core web assets
- Research data archives where links from APO Resources can be made

- Linked data, RDF environments that interface with global registry services
- Academic database systems that interface with university libraries
- Monolithic hosts of indexing, analytic, citation and social media services.

While engaging with both semantic-web and traditional library environments, APO finds itself with the somewhat challenging task of nailing down a core approach to ontology adoption. The DC Terms namespace works well in semantic web applications and maps reasonably well with standards such as MARC. But MARC properties, still underpinning many library systems, is easier to transform from a MODS approach. Perhaps more than any other modelling decision, the DCTERMS / MODS juncture illustrates the complexity of the APO stakeholder interface.

Given the somewhat immutable metadata requirements of communities such as Google, Twitter and Facebook, APO has drafted a sub-MAP for managing the interface between APO database and these services (see Meta-Tags section below).

Class Structure

Key considerations for selecting ontologies in the APO-MAP (APO, 2018) include whether object classes accommodate Persistent Identifiers (PIDs); whether class properties can be easily populated via lookup of vocabulary services; whether ontologies are used or favoured by key stakeholders and partners; and whether ontologies align with key international metadata approaches and trends.

We found that a top-down approach, where ontologies are evaluated at a class level, is compatible with a bottom-up approach where each data property is scrutinised against business requirements. Ontologies that facilitate automatic or semi-automatic metadata creation should be a key selection criterion for adoption within metadata schemes.

In APO-MAP, metadata properties are distributed over a number of content classes. The class model is a somewhat pivotal artefact that determines selection of ontological elements from the outset. In addition to a number of administrative classes that drive internal or proprietary operations not discussed here, the APO classes are:

- Bibliographic resources [Resources]
- Persons
- Organizations
- Conferences
- Collections
- Projects
- Concepts

For some of these classes, the APO-MAP is aspirational, especially for Collections and Projects, where elements have as yet not been implemented in metadata systems.

A key decision in the class model level is whether or not to distinguish datasets from other bibliographic resources. The research literature industry has identified, and is meeting, the challenge identified earlier this century to publish metadata about datasets (Brase, 2004; Green, 2009), thus making them persistently citable in research publications (Brase, 2014) and dereferenceable within semantic applications that interlink data and literature (Burton, 2015; Aryani et al., 2018).

Several research metadata schema distinguish datasets from literature/publications in their class models, including Scholix (2017), ResearchGraph (2018), and RIF-CS (ANDS, 2017). Indeed, APO needs these standards in order to consistently establish links between its research objects and datasets in other collections. However, APO content curation workflow does not result in production of a great number of datasets within its own collection. The relatively small number of datasets in the APO collection are effectively sub-classes of the Resources, refined using the DCMI Type vocabulary (DCMI, 2012). APO's core offering is curation of research literature that is derived from analysis of datasets, and its class structure somewhat reflects this and other APO priorities.

As well as the APO business focus on literature curation, a second rationale for leaving datasets out of the class model is that datasets are, conventionally, identified with the same registry and Persistent Identifier (PID) system as research literature. Digital Object Identifiers (DOI). While there are other PID systems reserved for other research object classes (such as ORCID[1] or ResearcherID[2] for Persons; RAiD[3] for research activities; GRID[4] or ISNI[5] for research organisations), datasets are identified with the same system used for literary works. This matters in cases where metadata repositories interact with DOI registry services – if datasets are described as a separate class with a different set of properties from literature, the interoperability challenge is doubled when aligning with registry-familiar standards such as DataCite Metadata Scheme v4.1[6]. APO therefore takes 'PID classes' as a high-level model for defining classes in the APO-MAP. The working assumption is that alignment through open registry systems, PID systems and locally defined class models will better streamline interactions (such as harvesting, sharing, and augmenting) between local repositories and global registries.

There are two exceptional cases in the APO-MAP that break these assumptions somewhat; Collections and Conferences are managed with a separate class without specialised PID systems. These cases are discussed further below as special issues for each class are elaborated.

Resources

The APO is a repository of research objects that are mostly located towards the end of the research data management lifecycle. That is, APO is mostly a collection of policy documents and research reports that are derived from analysis and distillation of research data and activities. Such a collection can be characterised as a repository of 'bibliographic resources' (Resources hereafter), which is a Dublin Core (DC) class of information resources – defined by DC as "book[s], article[s] or other documentary resource[s]". Therefore, the predominant metadata approach has aligned with the DC Terms ontology (DCMI, 2012) that is a key theoretical system that underpins the APO database structure.

The metadata requirements for APO Resources exceed the scope and purpose of the DCTERMS namespace. Given the wide range of publishing workflows and lifecycles characteristic of a grey literature and research repository, a number of metadata elements and vocabularies have been introduced to formally express Resource

1 ORCID: https://orcid.org/
2 ResearcherID: http://www.researcherid.com/
3 RAiD: https://www.raid.org.au/
4 GRID: https://www.grid.ac/
5 ISNI: http://www.isni.org/
6 DataCite Metadata Schema v4.1: https://schema.datacite.org/meta/kernel-4.1/

semantics. Work is underway to, where possible, adopt elements from published standards used within research literature communities. Some of these elements are presented and discussed below.

Article body

While APO is best described as a metadata repository, APO occasionally stores the full text of an article within its metadata. APO has identified emerging requirements from content providers to perform an archiving function – that is, beyond storage of surrogate information about an article, the article needs to be stored and rendered in a similar manor to its original hosted environment. Perhaps more than any other activity, this use case breaks the metadata / content divide.

Full article text should be distinguished from abstracts and summary descriptions so that the latter can be used to fulfil the user tasks Find and Identify Resources from within a search and search result context (International Federation of Library Associations, 1998). An Article Body may contain a mix of datatypes and document types, including text, hyperlinks and images, interactive graphs and other rich 'embedded' content. APO has selected `schema.articleBody` from the schema.org system to express instances where full content is captured in its system.

Principle investigator

Dublin Core Terms provide a formal means of identifying a *creator*, or the agent who is primarily responsible for the intellectual content of a Resource, as well as those who have made a secondary contribution (*Contributor* element).

Research reports often attribute a chief investigator or principle researcher role to a contributor. Chief Investigators named within research grants are likewise credited in journal articles and distinguished from co-investigators. Outside of the academic contexts, similar roles such as *Principle Researcher* are credited within research publications such as *RMIT ABC Factcheck*[7] articles.

The `rifcs:principleInvestigator` was taken from the Registry Interchange Format – Collections and Services (RIF-CS) schema 1.6.2[8]. The RIF-CS can be used to describe research objects in a format required by the Research Data Australia (RDA) Registry[9]. Within the RIF-CS standard, the domain for the Principle Investigator element is an Activity (research activity, or research project). APO is, therefore, testing the semantics of this element, which is intended to describe research activities, rather than the outputs of those activities (such as articles). This property is a good, or perhaps better fit with the provisional Project class in APO-MAP.

Content association

Another challenge in the APO collection relates to attribution of research organizations. In Dublin Core, and within traditional cataloguing systems, a single agent is attributed as publisher. It would be unwise to break this model; dereferencing a single, unambiguous source responsible for issuing a Resource is critical in preserving the provenance of a work. However, and within the research literature context, publisher information is often insufficient in capturing the inputs from multiple research institutions. Indeed, within the direct publishing model context, when a research institution is arbitrarily selected as a publisher they may receive an uneven attribution share.

7 RMIT ABC Factcheck: https://www.abc.net.au/news/factcheck/
8 RIF-CS Schema: https://www.ands.org.au/online-services/rif-cs-schema
9 RDA: https://researchdata.ands.org.au/

Therefore APO records all institutions involved in a research publication in a locally defined element. The `apo:contentAssociation` element is taken from a provisional namespace system that APO does not promulgate – indeed, APO prefers to reuse elements from well-known published ontologies.

Persons

While Resources make up a great proportion of the APO collection, the wider research data management endeavour is concerned with other kinds of research objects as well as with bibliographic resources (Resources). APO already manages two agent classes that are pivotal in joining up research outputs: *research organisations* and *researchers*. These agent classes are both named within APO as *Organisations* and *Persons* respectively. These object classes require properties that are either not readily available in DC or are available and require some shoe-horning with locally-defined value spaces (custom taxonomies) that refine and qualify property semantics.

A `foaf:person` is described in APO with a subset of properties from FOAF[10] vocabulary. APO extends the name, first name and last name property set with `apo:formerName` and `apo:alternativeName`.

An important development in FOAF is the collaboration with OpenID. The `foaf:openID` property allows expression of an *indirect identifier*, as described in *Architecture of the World Wide Web, Volume One* (W3C, 2004). APO intends to use `foaf:openID` to express ORCID and scopusID URIs associated with Person records.

APO also classifies elements in the `foaf:Person` class against a the Protective marking vocabulary, part of the *Protective Security Policy Framework* (Australian Government Attorney-General's Department[11]). The Protective marking attribute can be used to define rules for sharing personal attributes within APO and partner systems. The Protective marking attribute is used only within the `foaf:Person` class as it breaks the tabular structure of the APO-MAP – it is, nevertheless, a step towards declaring what APO will do with personal information, in keeping with adopted compliance regimes such as GDPR[12].

Organisations

Organisations can be authors or publishers in the APO collection. Given the APO collection focus on direct-published materials, the contributing organisations vary in structure and purpose. Some are sub-units of parent organisations, such as centres, faculties or schools in universities or departments or statutory authorities within government. Research objects are published by corporate entities beyond traditional research contexts. Therefore, APO uses properties to describe owning relationships between Organizations, and the purpose of each Organization.

APO is seeking the best way to express these relationships. The Organization Ontology is a W3C Recommendation and includes properties for interrelating organisations and organisational units. The GRID registry provides a similar approach. It is tempting to go with the GRID approach, as this achieves aforementioned alignment between registry, PID system and local repository. However, the GRID registry, a

10 FOAF: http://xmlns.com/foaf/spec/

11 See: https://www.protectivesecurity.gov.au/information/sensitive-classified-information/Pages/default.aspx

12 GDPR: https://eugdpr.org/

somewhat ambitious endeavour, holds only a small fraction of records that correspond with organisation records in APO.

Conferences

In APO database, Conferences are a third party type in addition to Persons and Organisations. APO is seeking a standardised approach to Conference definition and properties. Provisionally, a Conference is classed as a `dcterms:Event`. This approach works well with some of APO's activities, where for example call for submissions are advertised and promoted. These activities focus on conference instances – time bound, spatially located events.

However, Conferences may also be publishers. There are many different patterns in conference publishing workflow, from direct publishing, publishing by underpinning association, to third-party commercial publishing of conference papers and proceedings. To the extent that a Conference is primarily responsible for issuing Resources, it may be better to treat them as a sub-class of `foaf:agent`.

While `foaf:group` is a logical candidate class that would mean a FOAF-based approach to all agent classes in APO-MAP, APO is also watching another development that may shed some light on Conference class development. The PIDs for Conferences and Projects Working Group (Crossref, 2018) will hopefully shed some light on conference characterises and property range. A dedicated PID system, and possibly registry of conferences could go some way to validating the conference-as-agent model.

Projects

In addition to Resources, Organisations and Persons, APO is considering the object class *Project*. A Project means a research projects or research activity, which can be captured as a discrete research object. Further upstream from Resources in the research data management cycle, Projects are the source of research outputs such as dataset and articles. By publishing Projects as records, APO could alert its audience about current research activities where there may be current opportunities for collaboration (Resources do this too, but after the fact).

Some of the benefits of deploying a Project class relate to attribution issues, as discussed above for Resources. ResearchGraph, Scholix, RIF-CS and CERIF (euroCRIS, 2013) metadata standards that articulate a research project / activity class are more or less in agreement that Projects are where principle investigators and partnering organisations should be formally declared.

Collections

Another object class defined in APO-MAP is the Collections class. APO already publishes Collections, or Resources aggregated around some theme, via its website[13]. Collections are a significant activity by which APO adds value to the grey literature publication cycle. It is due to the significant part that Collections play in APO collection development and stakeholder engagement that they are, in the APO-MAP, elevated Class status when they would elsewhere be considered a *type of Resource*, as per the DCMI Type vocabulary. Given that collection types are defined in the same way as datasets within DC, it is worth considering if the range of the Collection class should

13 See: http://apo.org.au/collections

include the same PIDs as Resources. APO is investigating the appropriateness, conceptual fit and community interest in assigning DOIs to Collection objects.

Concepts

APO uses a combination of third-party and locally built vocabularies, or taxonomies, to populate metadata fields. The taxonomies populate fields in Resources, Collections and Organization records. The taxonomies provide basis for site navigation, augmenting search indexes and enabling reporting and analytic operations.

A requirement to expose APO vocabularies as stand-alone objects has been identified, especially given projects where APO content is shared with external metadata systems. By exposing taxonomies in their entirety, APO content partners can validate taxonomy terms supplied in APO records. Full taxonomy files and portal also provide context, including relationship between terms (hierarchy; associations) and alternative terms (synonyms). As APO is developing thesaurus relationships in its taxonomies, as described in Z39.19 (ANSI/NISO, 2010), `skos:concept` has been selected as the RDF class for managing taxonomy terms. *Simple Knowledge Organization System* (W3C, 2009) provides properties that express all of the key thesaurus relationships in Z39.19. In the interim, APO taxonomies can be looked up within the draft APO-MAP.

SKOS is widely used, including within international vocabulary API and query services hosted by Basel Register of Thesauri, Ontologies and Classifications (BARTOC) Skosmos Browser[14], and within Research Vocabularies Australia linked data API[15]. SKOS is also used to define elements within FAST (Faceted Application of Subject Terminology), which APO uses within subject taxonomies. The case for the SKOS ontology is strong considering both community standards and uptake, and local business requirements for expressing and using taxonomy data.

There is no registry service for individual concepts, nor mandated PID systems for concept identifiers, although registries of whole vocabularies exist (e.g. BARTOC and Taxonomy Warehouse[16]). Any URI can identify a `skos:concept`. Therefore assumptions about PIDs and classes is not relevant to APO taxonomies.

Meta-Tags

The APO-MAP is an ongoing negotiation between widely used standards, community requirements and local business drivers. However, not all applications that serve APO audience are run by accessible communities. Increasingly, APO is interfacing with web applications run by faceless 'tech giants' responsible for services such as indexing, citations and social media. A different metadata response is needed to meeting the requirements of these communities.

The DCAP guideline states that DCAP needs to be both fit for purpose and interoperable – that is, designed with particular (known) applications in mind, but also conforming to recognised standards and approaches so that unanticipated applications may consume resulting data. It is tempting, therefore, to assume that a single DCAP should be sufficient for any given enterprise or community. APO found, however that a second derivative profile was needed for managing its webpage source code. 'Meta tags' in the APO website page HTML were found to be the locus of many requirements

14 BARTOCK Skosmos Browser: https://bartoc-skosmos.unibas.ch/en/
15 Research Vocabularies Austalia Linked Data API: https://documentation.ands.org.au/display/DOC/Linked+Data+API
16 Taxonomy Warehouse: http://www.taxonomywarehouse.com/

linked to web applications, including Google Scholar, Google Analytics and a number of citation, metrics and social media services.

Social media web applications, such as Facebook and Twitter, harvest structured data from webpages such as titles, descriptions and images. Citation services will harvest further detail, such as publisher information or volume, issue and pagination details. Search indexing services will demand even more elements, such as subject keywords and document types. Taken together, APO has found that a great many elements need to be published in source html to enable these services.

Furthermore, web applications dictate use of specific HTML meta-tags as a precondition for resources to be resolvable, searchable or trackable within proprietary environments. Therefore the number of meta-tags needed is a factor of the number of element functions times the number of ontologies needed. This is a somewhat inescapable fact – there is no negotiating with a community of stakeholders when serving web applications, no opportunity to compromise on preferred ontologies or to design cross-walks or element mappings.

A challenge for Meta-Tag profiling is serving the needs of multiple applications with as little element duplication as possible.

It is worth thinking about the HTML source code as an API itself – once the rules are set for how the source code is structured, it is a development effort to make even minor changes. And changes have to be carefully planned in order to reduce duplication, redundancy and clutter.

Therefore, a sub-profile has been drafted for managing APO Meta-Tags. There are opportunities for confusion, both internally and with stakeholders about the presence of two MAPs and so APO has structured the second application profile so that it:

- is derived from the APO-MAP,
- does not extend the domain model (does not introduce new classes),
- authorises element-to-element mapping from APO-MAP and APO Meta-Tags, and
- is a reference only and not a public consultation draft.

The APO Meta-Tags maps elements back to APO-MAP elements – a many-to-one relationship, effectively grouping Meta-Tags into semantically similar categories.

Conclusion

APO is working towards greater standardisation of research and policy grey literature. Towards this aim, a number of assumptions drive the selection, adoption and extension of well-known metadata approaches. Not all assumptions work in all cases. We have pointed to cases where metadata communities of concern work significantly with different metadata standards. Metadata standards are themselves constructed with different object class structures. Relationships between PID systems and object classes are one-to-one, one-to-many or irrelevant. And global registries relevant to research outputs are in varying stages of evolution and relevance to local collection scope. Given the complexity of these arrangements, APO sees ongoing discussion about metadata approaches as a key activity towards finding a best-fit approach. Releasing the draft APO-MAP is key locus for that discussion.

References

ANDS. (2017). *RIF-CS schema 1.6.2*. Retrieved from https://www.ands.org.au/online-services/rif-cs-schema

ANSI/NISO. (2010). *Z39.19-2005 (R2010) Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Retrieved from https://www.niso.org/publications/ansiniso-z3919-2005-r2010

APO. (2018). APO-MAP. Retrieved from https://github.com/leskneebone/APO-MAP

Aryani, A., Poblet, M., Unsworth, K., Wang, J., Evans, B., Devaraju, A., & Kaplun, S. (2018). A research graph dataset for connecting research data repositories using RD-Switchboard. *Scientific Data, 5,* 180099.

Brase, J. (2004). Using digital library techniques – registration of scientific primary data. In R. Heery & L. Lyon (Eds) Research and Advanced Technology for Digital Libraries. ECDL 2004. *Lecture Notes in Computer Science,* 3232. Springer, Berlin, Heidelberg

Brase, J. (2014). Making data citeable: DataCite. In S. Bartling & S. Friesike (Eds) Opening Science. Springer, Cham.

Burton, A. et al. (2015). On bridging data centers and publishers: The data-literature interlinking service. In E. Garoufallou, R. Hartley & P. Gaitanou (Eds) Metadata and Semantics Research. MTSR 2015. *Communications in Computer and Information Science*, 544. Springer, Cham.

Crossref. (2018). PIDs for conferences and projects. Retrieved from https://www.crossref.org/working-groups/conferences-projects/

DCMI. (2009). *Guidelines for Dublin Core Application Profiles (working draft)*. Retrieved from http://dublincore.org/documents/profile-guidelines/

DCMI. (2012). *DCMI Metadata Terms.* Retrieved from http://dublincore.org/documents/dcmi-terms/

euroCRIS. (2013). *Common European Research Information Format.* Retrieved from https://www.eurocris.org/cerif/main-features-cerif

DataCite. (2017). *DataCite Metadata Scheme v4.1*. Retrieved from https://schema.datacite.org/meta/kernel-4.1/

Green, T. (2009). We need publishing standards for datasets and data tables. *Learned Publishing*, *22*(4), 325-327.

International Federation of Library Associations, IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998). *Functional requirements for bibliographic records: Final report*. Retrieved from www.ifla.org/VII/s13/frbr/frbr.pdf

Lawrence, A. (2016). *Digital curation of public policy resources: Discovery, access and management for policy and practice.* Paper presented at Vala Conference 2016. doi:10.4225/50/570B2F6678DFA

Library of Congress. (2018). Metadata Object Description Schema. Retrieved from http://www.loc.gov/standards/mods/

Research Graph. (2018). *ResearchGraph schema v2.0*. Retrieved from http://researchgraph.org/schema/

Scholix. (2017). *Scholix Metadata Schema 3.0*. Retrieved from http://www.scholix.org/schema

W3C. (2004). *Architecture of the World Wide Web, Volume One*. Retrieved from https://www.w3.org/TR/webarch/

W3C. (2009). *SKOS Simple Knowledge Organization System Reference*. Retrieved from https://www.w3.org/TR/skos-reference/

W3C. (2014a). *Data Catalog Vocabulary (DCAT)*. Retrieved from https://www.w3.org/TR/vocab-dcat/

W3C. (2014b). *FOAF Vocabulary Specification 0.99*. Retrieved from http://xmlns.com/foaf/spec/