Cristina I Perez. The RDA's Metadata Standards Directory: Information Gathering. A Master's Paper for the M.S. in L.S. degree. November, 2013. 27 pages. Advisor: Jane Greenberg

The Metadata Standards Directory Working Group that is part of the Research Data Alliance (RDA) is charged with creating a directory of metadata standards. The Digital Curation Centre's (DCC) Disciplinary Metadata Catalogue was selected as a starting point for this effort. A form was created to gather information on eleven aspects of metadata standards, and distributed via email across several interest groups in order to collect information on metadata standards.

The thirty-two responses were evaluated; new and supplemental information was extracted and formatted for uploading onto the DCC's Web page. This resulted in the addition of eleven new standards, ten new extensions, twelve new tools, and twenty-three new use cases. This paper reports the results of this information gathering activity, including a display of the range of metadata standards, extensions, tools, and disciplines received in the responses. Suggestions for the next steps toward an open and collaborative directory are also included.

Headings:

Metadata

Directories

Surveys -- Metadata

THE RDA'S METADATA STANDARDS DIRECTORY: INFORMATION GATHERING

by
Cristina I Perez

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

November 2013

Approved by

_____

Jane Greenberg

INTRODUCTION

The Research Data Alliance (RDA) is an international organization of researchers who strive for the optimization of research data sharing. Members of the RDA formed the Metadata Standards Directory Working Group in 2013 with the goal of implementing an open and collaborative metadata standards directory (Greenberg, Jeffery, & Koskela, 2013). Exploratory work was first pursued by students at the Metadata Research Center, University of North Carolina at Chapel Hill's School of Information and Library Science (UNC- SILS) in the form of a Wiki <http://wiki.dublincore.org/index.php/DCMI_Science_and_Metadata_~_Research_Data_Alliance_(RDA)_Metadata_Directory>. Additionally, the Digital Curation Centre (DCC), a United Kingdom based organization instituted to assist with digital curation and preservation, created a Disciplinary Metadata Catalogue. The Metadata Standards Directory Working Group chose to collaborate with the DCC by using the existing Catalogue as a starting point for their metadata standards directory. This paper documents the data gathering and analysis process for supplementing the DDC Disciplinary Metadata Catalogue.

BACKGROUND AND LITERATURE REVIEW

Implementation of a framework for managing scientific data through metadata requires an understanding of metadata and the challenges associated with scientific data. The first half of this section covers metadata in general and the second half reviews issues relating to metadata for scientific data.

Metadata

The simplest definition of metadata is “data about data,” but in truth metadata can be anything or everything. Hillmann, Marker & Brady (2008) describe five areas of functionality that metadata may fall into: administrative, covering who, when, where, approval, and dates; descriptive, what it is; access/use, rights and restrictions; preservation, ensuring future accessibility; and structural, describing relationships between files. Information professionals create metadata through, harvesting, and author submission.

A metadata standard is the organization of metadata elements into a format including definitions and usage rules (Hirwade, 2011, p. 18). Chan and Zeng (2006) point out in their study of metadata interoperability and standardization that “The rapid growth of Internet resources and digital collections has ben accompanied by a proliferation of metadata schemas, each of which has been designed based on the requirements of particular user communities, intended users, types of materials, subject domains, project

needs, etc." These schemas can also be derived from existing schemas, and are then known as extensions of a source. Controlled vocabularies are another facet of metadata standards, according to Hillmann, Marker, & Brady (2008), "The goals of controlled vocabularies are to: eliminate or reduce ambiguity; control the use of synonyms; establish formal relationships among terms; and test and validate terms" (p. 17). These goals are met through lists, synonym rings, taxonomies, thesauri, and ontologies (Hillmann, Marker, & Brady, 2008, p. 17). The pieces of metadata comprising a standard are developed within the needs of specific disciplines.

Disciplinary metadata allows for the sharing and reuse of data within a particular context. Although development of disciplinary metadata schemes and standards is widespread, Hirwade (2011) found that multiple standards are commonly used in order to fulfill needs. Sharing standards within a discipline is a frequent occurrence, and is referred to as interoperability in the information community. Chan and Zeng (2006) generalize the concept as the exchange of information between systems with minimal effort and no loss of value. Three levels of interoperability agreed upon, as stated in Chan and Zeng (2006), are the schema level, where the focus is on elements, the record level, which is centered on combining and updating records, and the repository level, which revolves around mapping values to elements for cross-collection searching. A key to reuse of metadata standards is visibility; as Hillmann, Marker, & Brady (2008) state, "registries support interoperability by allowing the discovery of available schemes and schemas for description of resources" (p. 19).

Chan and Zeng expressed the functions of metadata registries as "registering, publishing, and managing schemas and application profiles, as well as making them

searchable" (2006). Metadata registries, or similarly directories, are available to aid in selecting and accessing standards. There are several existing registries and directories currently in place, most of which are organized by discipline or field of study, selected examples are included in Appendix A.

Metadata for Scientific Data

The amounts and types of data used in scientific research are large, varied, and growing. Scientific metadata supports researchers by resolving common issues such as finding relevant data, extracting information from data sources, and data evaluation (Galhardas, Simon, & Tomasic, 1998, p. 106). Diederich and Milton (1991) found that securing a stable schema in scientific fields of study is a complicated task because features are constantly being added and amended. This explains the existence of the quantity of scientific metadata standards and why so many continue to be created.

Jones, Berkley, Bojilova, & Schildhauer (2002), acknowledge that there are many standards which may be applicable to a particular domain, "…but none are deep and broad enough for effectively documenting biological data" (p. 67). This is one of the reasons behind the prevalence of many scientific metadata standards. Galhardas, Simon, & Tomasic (1998), provide another explanation for the existence of multiple standards, "Even within the environmental community, there is not a common definition [of metadata] and different metadata formats emerged from distinct disciplines. From an analysis of the existing standards, one concludes that there is not and there will never exist a common metadata format" (p. 107). Additional obstacles faced when trying to

combine domain specific metadata are data heterogeneity, data dispersion, and local control (Jones, Berkley, Bojilova, & Schildhauer, 2002, p.60).

This predicament can be found even within an organization. The United States Environmental Protection Agency's Office of Research and Development (ORD) performs research covering several different, and often interrelated, fields of study. Researchers decide on what metadata, and which schemes if any, to use and their work is tracked through the organizational hierarchy of the National Laboratories and Centers. The lack of overarching metadata means that there is no way to track the data across ORD. Collecting metadata about the scientific process through the implementation of a Scientific Data Management Plan is one proposed solution. This undertaking should provide enough information to streamline sharing scientific data that adequately represents the science. An open metadata directory could guide this activity and can facilitate the transparency and reuse of data that exists. With the exception of the DCC's Disciplinary Metadata Catalogue and a few other discipline specific efforts, the availability of registries are limited; and the long term accessibility is unknown. There is a need to study these systems and consider how they may be enhanced and supported over time, which is the goal of this master's paper.

RESEARCH OBJECTIVES

The work presented in this paper addresses three specific objectives.

- To explore the need and applicability of an open directory of metadata standards.
- To gather and analyze data in conjunction with this need, and contribute to the DCC directory.
- To document personal reflections on the outcomes, provide a look into the progress of the effort as well as possibilities for future development

METHODS

The survey method was used in this study to gather information on metadata standards. Sets of tasks were also pursued to explore ways to enhance the existing DCC metadata directory. In this section, several definitions are offered that helped guide the research. The definitions are followed by the procedure.

Definitions

Standard, schema and scheme are often used interchangeably, which can lead to confusion. For the purposes of this project, the definitions for standard, extension, tool, and use case were as follows (A. Ball, personal communication, October 18, 2013):

- Standard means (at least) a set of metadata fields used to describe research data. The standard may or may not define a canonical way of encoding the metadata (i.e. a file format) but it is primarily the set of fields we are interested in.
- An extension is a metadata scheme that is wholly or partly derived from one or more metadata standards, usually to fulfill the needs of a particular repository or data type.
- A tool is any piece of software known to use the metadata standard, while
- A use case is a working implementation (service or repository).

The first step of this process was meeting the team that formed. I first met Sean Chen, a fellow student within the School of Information and Library Science at the University of North Carolina. He was my data gathering and analysis partner for this

project. We were then introduced by Jane Greenberg to the rest of the team, the remaining Metadata Standards Working Group chairs- Keith Jeffery and Rebecca Koskela, and the curator of the DCC Disciplinary Metadata Catalogue, Alex Ball, via email and Skype. Shortly afterwards, decisions were made concerning the process and timeline of the project. A form would be sent out to collect data and the responses then organized and prepared for integration into the DCC's Web pages. A Google Drive folder was created and shared with the team in order to house the documentation of our work.

The communities of participants chosen as recipients of our data gathering form were: the RDA general listerv, the RDA Metadata Standards Directory Working Group, the RDA Metadata Interest Group, euroCRIS, European Plate Observing System (EPOS), Dublin Core general list, ASIST- Research Data and Preservation list, DataONE, Earth Science Information Partners (ESIP), Science & Technology Facilities Council (STFC), and those who attended the RDA Second Plenary interest meeting.

The original Google form created was sent out as a week long pilot to five individuals, as well as the team. The feedback received led to changing the wording of the questions and the addition of the submitter information disclosure section (see Appendix B). Once the revisions were made, we wrote the text for the cover letter that would accompany the link to the form, as documented in Appendix C. By the end of the week it was officially sent out to the participant lists. The following week, Sean and I sent out personal emails to those who attended the interest meeting at the RDA Second Plenary. We received responses throughout the two-week window established for data collection. After the closing date, we sent out emails thanking the respondents and followed up with questions and comments received during the process.

A cursory review of the responses allowed the team to gauge the breadth and depth of data gathered. In the meantime, we were granted access to the DCC page, which allowed us to edit information in the Catalogue. Several meetings followed to discuss the next steps and check for quality of data. It was during one of these meetings that we decided to separate the work by task to ensure continuity in the data, rather than splitting the work in half as we had for the preceding steps.

My task was data analysis and curation; I began by transferring the responses received during our collection period into an Excel spreadsheet in order to not affect the responses from after the closing date (the form is still open and gathering additional data).

The steps for analysis were:

1. Check for duplicates in responses- consolidate and mark any present
2. Check if standards exist in the DCC and mark (yes/no) accordingly
    a. Check existing standards for new supplemental information (tools, extensions, use cases)
    b. New standards – format fields for DCC

The formatting process was as follows:

1. Find official name and/or acronym
2. Test links provided
3. Check for readability of free text, edit if necessary
4. Note additional (written in) domains. Brought up in discussion of expanding disciplines on the DCC page
5. Supply summary and/or description if none present
6. Check that keywords are in the HESA JACS3 <http://www.hesa.ac.uk/content/view/1787/281/>
7. Input ‘cleaned’ data into shared spreadsheet on Google Drive for upload to DCC

RESULTS

Tabular

Response Type	Applicable to New Entries	Related to Existing Entries	Total
Standards	11	4	15
Extensions	5	5	10
Tools	7	5	12
Use Cases	19	4	23
Total Responses **32**		Total Updates to DCC Catalogue	**60**

Table 1: Summary of Findings Tabular

Name	Domain	Source URL
DCAT	General Research Data	http://www.w3.org/TR/vocab-dcat/
FITS- Flexible Image Transport System	Physical Science	http://fits.gsfc.nasa.gov; specification site: http://fits.gsfc.nasa.gov/fits_standard.html
Genome Metadata	Biology	http://enews.patricbrc.org/faqs/genome-metadata-faqs/
O&M - Observations and Measurements	Earth Science; Physical Science; General Research Data	http://www.opengeospatial.org/standards/om
OAI-ORE - Open Archives Initiative Object Reuse and Exchange	Social Sciences	http://www.openarchives.org/ore/
Observ- OM	Biology	http://www.molgenis.org/wiki/ObservStart
Protocol Data Element Definitions (DRAFT)	Biology	http://prsinfo.clinicaltrials.gov/definitions.html
PROV	General Research Data	http://www.w3.org/2001/sw/wiki/PROV
QuDEx - Qualitative Data Exchange Format	Social Sciences	http://www.data-archive.ac.uk/create-manage/projects/qudex?index=1
Resource Metadata for the Virtual Observatory	Physical Science	http://www.ivoa.net/documents/latest/RM.html
The RDF Data Cube Vocabulary	General Research Data	http://www.w3.org/TR/vocab-data-cube/

Table 2: New Standards

Name	Associated Standard	Source URL
ADMS	DCAT	http://www.w3.org/TR/vocab-adms/
GSIM -Generic Statistical Information Model	DDI	http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model
isaconfig-diXa	ISA- Tab	https://bitbucket.org/kanterae/isaconfig-dixa
PROV Extension list	PROV	http://www.w3.org/TR/prov-implementations/#prov-extensions
Resource Metadata for the Virtual Observatory	Dublin Core	http://www.ivoa.net/documents/latest/RM.html
USGIN ISO profile	ISO 19115	http://repository.stategeothermaldata.org/repository/resource/98ddf901b9782a25982e01af3b0bda50/
VarioML	Observ- OM	http://www.ncbi.nlm.nih.gov/pubmed/23031277
WaterML	O&M	http://www.opengeospatial.org/standards/waterml
WCS - World Coordinate System	FITS	http://fits.gsfc.nasa.gov/fits_wcs.html
WMO Core Metadata Profile	ISO 19115	http://wis.wmo.int/2006/metadata/WMO%20Core%20Metadata%20Profile%20%28October%202006%29/documentation.htm

Table 3: New Extensions

Name	Associated Standard	Source URL
Danish DdiEditor	DDI	http://code.google.com/p/ddieditor/
DataCite Metadata Store API	DataCite	https://mds.datacite.org/static/apidoc
Fiji	OME- XML	http://fiji.sc/Fiji
FITS Image Software Packages	FITS	http://fits.gsfc.nasa.gov/fits_viewer.html
Linked Data Cubes Explorer	RDF Data Cube	http://www.ldcx.linked-data-cubes.org:8000/ldcx-trunk/ldcx/ld-cubes-explorer.html
Metacat	EML	http://knb.ecoinformatics.org/knb/docs/
MOLGENIS	Observ- OM	http://www.molgenis.org/wiki/WikiStart
PATRIC Download Tool	Genome Metadata	http://patricbrc.org/portal/portal/patric/Downloads?cType=taxon&cId=
PROV Implementation Report	PROV	http://www.w3.org/TR/prov-implementations/
ProvToolbox	PROV	https://github.com/lucmoreau/ProvToolbox/
SOS -Sensor Observation Service	O&M	http://www.opengeospatial.org/standards/sos
Stat/Transfer	DDI	http://www.stattransfer.com/

Table 4: New Tools

Name	Associated Standard	Source URL
GRIIDC - The Gulf of Mexico Research Initiative Information and Data Cooperative	ISO 19115-2	https://data.gulfresearchinitiative.org/
GBIF Global Biodiversity Information Facility	ABCD	http://www-old.gbif.org/informatics/standards-and-tools/publishing-data/data-standards
CKAN	DCAT	http://ckan.org/
STEREO Science Center	FITS	http://stereo-ssc.nascom.nasa.gov/
HEASARC	FITS	https://heasarc.gsfc.nasa.gov/
Australian Ocean Data Network Portal	ISO 19115	http://portal.aodn.org.au/aodn/
Integrated Marine Observing System Portal	ISO 19115	http://imos.aodn.org.au/imos/
ESA - European Space Agency	FITS	http://www.esa.int/ESA
BAV - Biblioteca Apoltolica Vaticana	FITS	http://www.vatlib.it/home.php
JAXA - Japan Aerospace Exploration Agency	FITS	http://www.jaxa.jp/index_e.html
Virtual Solar Observatory	FITS	http://sdac.virtualsolar.org/cgi/search
IVOA - International Virtual Observatory Alliance	FITS	http://www.ivoa.net/
List of RDF Data Cube Vocabulary Implementations	RDF Data Cube	http://www.w3.org/2011/gld/wiki/Data_Cube_Implementations
WormQTL	Observ- OM	http://www.wormqtl.org/
International dystrophic eb Patient Registry	Observ- OM	http://www.deb-central.org/
CHD7 Database	Observ- OM	http://www.chd7.org/
WormQTL-HD	Observ- OM	http://www.wormqtl-hd.org/
MVID Patient Registry	Observ- OM	http://www.mvid-central.org/
INSPIRE - Infrastructure for Spatial Information in Europe	O&M	http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.9_O&M_Guidelines_v2.0rc3.pdf
ODIP -Ocean Data Interoperability Platform	O&M	http://www.odip.org/welcome.asp
Open Archives Initiative	OAI-ORE	http://www.openarchives.org/
ProvBench	PROV	https://sites.google.com/site/provbench/
UK Data Service	QuDEx	http://ukdataservice.ac.uk/

Table 5: New Use Cases

Table 1 reports the data received in responses separated by type and relationship to the existing entries in the DCC Catalogue, these numbers demonstrate the total items added to the DCC site. Table 2 lists out the new standards gathered from reviewing the responses along with their domains (from a list determined by the DCC). Tables 3, 4, and 5, display the new extensions, tools, and use cases, respectively; these tables show the name of the new elements and their associated standard. The information about each of these entries added into the DCC's page also included a brief summary; new standards included more detailed description and keywords.

Respondent Location and Organizational Affiliation

Respondents were located in Australia, Europe, and North America, representing various institutions including*:

- Syracuse University
- Purdue University Libraries
- National Center for Ecological Analysis and Synthesis
- DICE Center, UNC-CH
- National Snow and Ice Data Center, Univ. of Colorado
- ResearXis-Discinnet
- University Medical Center Groningen
- Commonwealth Scientific Industrial Research Organisation
- Open Microscopy Environment
- Rensselaer Polytechnic Institute
- Newcastle University
- UK Data Archive
- US Virtual Astronomical Observatory / Space Telescope Science Institute
- EDINA, The University of Edinburgh
- U. S. Geoscience Information Network
- Met Office, UK
- IMOS/AODN Integrated Marine Observing System/Australian Ocean Data Network
- NRCan/GeoConnections

*Organizations listed are from responses where the submitter selected to share their organizational connection

Members of the team retained access to the Google Drive folders, the contents of which track the steps of data collection. Additional files were created to document the comments received, recommended steps for responses received after our data collection period ended, and to summarize the process and reflect our stopping point.

DISCUSSION

Reflections

The amount of responses (thirty-two) we received met our expectations formed at the opening of the data collection. Four of the responses were not processed because of insufficient data; this left twenty-eight functional responses that created sixty new entries. These numbers illustrate how each response produced multiple additions to the DCC.

Several factors affected the data gathering efforts in both positive and negative ways. The United States Government shut down on the first day of our pilot and reopened the last day of data collection. This limited the respondents, as any Federal employees were not able to check their work emails. Positive influences on our data collection included the timeliness of the RDA Second Plenary in Washington, D.C. on September 16-18, 2013, and the fact that the information community is predisposed to sharing.

If we had unlimited time and resources, it would have been beneficial to have a standalone survey that allowed for formatting of questions to better represent the types of data being asked for, in hopes of reducing the amount of manual curation needed to determine the highest category of standard. It would have been advantageous to involve others in the data collection, as having more minds behind the effort would allow for delving into responses to glean as much information as possible about the standards and related resources. Continuation to the next steps in the Metadata Standards Directory

Working Group's plan would have been possible, and the data could be ingested into a stand-alone system for the directory.

The form will remain open for continuous data collection; it is available on the RDA Metadata Standards Directory Working Group page. It will be sent out at intervals and the link included with presentations associated with the Working Group. Future UNC- SILS Master's students may also continue the efforts in place, whether it is to replicate the process or expand on it. Any future expansion should plan to move into the next phase of the proposed directory.

Brainstorming

The implementation of the directory utilizing a social media configuration would foster connectivity by providing space for researchers to share their standards while being able to search and view those of their peers. Giving researchers a medium to connect in a professional, yet social and controlled, environment creates a new channel for interaction and making connections.

The system can be pre-populated with data from existing established registries of metadata standards. Providing initial information for each standard's page guarantees that the foundational information in the system, because of the curation efforts in place for the initial ingest, is reliable. Drawing from existing data sources for pre-population and maintenance is not intended to replace these tools, but rather to create a more user friendly interface for viewing the data as a whole.

The interfaces of social networking sites are simple enough that learning how to insert information takes minimal effort. Users can add in information about their

standards and work that was not covered by the automatically ingested data. The search feature and the ability to click through pages will provide links between people, research, and organizations that might not be documented elsewhere.

Rating and review systems present in sites that rely on social interactions from their customer base, such as Amazon, Netflix, and Goodreads, would be helpful for the metadata standards directory. A star rating system allowing the users to vote on each standard provides numerical and visual representations of approval, while a text review option provides for more detailed reasoning behind ratings. Social media outlets link to each other, adopting this ability to 'share' standards as is done on Facebook, Google +, Twitter, and blogs, translates to more visibility for both the standards and the directory.

To further the value of data stored in the Metadata Standards Directory can apply the Semantic Web to the underlying data of the directory. Berners-Lee, Hendler, & Lassila (2001) describe the Semantic Web as an expansion of the Web as we know it to include meaning and trust in the information which allows for work between computers and people. With the use of Semantic Web principles such as ontologies and Resource Description Framework (RDF), the data would be machine readable and interoperable.

An ontology applied to the metadata about the metadata standards prepares the data to be stored in RDF, this specification of concepts and relationships between them allows for inferences to be made. The structure of the data in the underlying database then permits querying the data as well as the searching through function that is integrated into the system. This format for the metadata standards directory will simplify searches and lead to more discoveries based on deductions made from data that previously did not exist.

These interpretations of the data from the metadata standards directory are also machine readable. Semantic Web agents query resources using in order to find ways to meet user needs; they can access the metadata standards directory and interpret the data in terms of suitability for the user. Connected with other Semantic Web applications, researchers would have an aid in selecting the appropriate standard that meets their needs. Agents communicate with one another; they can refine the process of searching for metadata standards by inferring additional connections between researchers and their data.

Cardoso's (2009) study provides evidence of the Semantic Web shifting from conceptual to existing in practice. Although it takes a more widespread buy-in to deploy the Semantic Web's full potential, it would not hurt to have the data ready for when the possibility becomes a reality.

CONCLUSION

The Metadata Standards Directory Working Group aims to produce a community driven metadata standards directory; it was this goal that served as a driving force for the work documented in this paper. The purpose of this paper was to explore the need and applicability of an open directory of metadata standards through addressing the issues associated with metadata standards used in research. It includes a report detailing the information gathering and data analysis performed, the results of which were incorporated into the DCC directory. Thirty-two responses yielded sixty updates to the DCC catalogue, receiving this feedback over a two week period exemplifies the willingness to share within the community of information professionals. This positive reception coupled with the discussions launched from the progress on this endeavor demonstrates there is an understanding of the need for an open, collaborative metadata standards directory and that people are ready and willing to make it happen.

References

Berners-Lee, T., Hendler, J., and Lassila, O. (2001) The Semantic Web. Scientific American 284, 5, 34–43.

Cardoso, J. (2009). The Semantic Web: A mythical story or a solid reality?. In *Metadata and Semantics* (pp. 253-257). Springer US.

Chan, L. M., & Zeng, M. L. (2006). Metadata interoperability and standardization-A study of methodology, Part I. *D-Lib Magazine*, *12*(6), 3.

Diederich, J., & Milton, J. (1991). Creating domain specific metadata for scientific data and knowledge bases. *Knowledge and Data Engineering, IEEE Transactions on*, *3*(4), 421-434.

Galhardas, H., Simon, E., & Tomasic, A. (1998, August). A framework for classifying scientific metadata. In *AAAI'98 Workshop on AI and Information Integration*.

Greenberg, J., Jeffery, K., and Koskela, R. (2013). Metadata Standards Working Group Case Statement. V0.02. Research Data Alliance: https://www.rd-alliance.org/filedepot_download/418/100.

Hillmann, D. I., Marker, R., & Brady, C. (2008). Metadata standards and applications. *The Serials Librarian*, *54*(1-2), 7-21.

Hirwade, M. A. (2011). A study of metadata standards. *Library Hi Tech News*, *28*(7), 18-25.

Jones, M. B., Berkley, C., Bojilova, J., & Schildhauer, M. (2001). Managing scientific metadata. *Internet Computing, IEEE*, *5*(5), 59-68.

Appendix A – List of existing efforts

Dublin Core Metadata Initiative and Research Data Alliance (2012). Metadata Directory Wiki: http://wiki.dublincore.org/index.php/DCMI_Science_and_Metadata_~_Research_Data_Alliance_(RDA)_Metadata_Directory .

Digital Curation Centre (2013). Disciplinary Metadata: http://www.dcc.ac.uk/resources/metadata-standards .

Riley, J. and Becker, D. (2009 – 2010). Glossary of Metadata Standards: http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards_glossary_pamphlet.pdf.

Qin, J., Small, R., and D'Ignazio, J. (2008). The Science Data Literacy Project's List of Metadata Standards: http://sdl.syr.edu/?page_id=32.

BioSharing (2009-2013). Standards sharing: http://www.biosharing.org/standards.

Marine Metadata Initiative (2013). Content Standards References: https://marinemetadata.org/conventions/content-standards.

Open Knowledge Foundation (2013). LOV- Linked Open Vocabularies: http://lov.okfn.org/dataset/lov/.

Group on Earth Observations (2013) GEOSS Standards Registry: http://seabass.ieee.org/groups/geoss/.

Appendix B- Survey used for gathering data – Online survey can be found at <https://docs.google.com/forms/d/1kzV3-Ri7LkYq1_fPo811CSKcu8nn6bBYCEm1A73KWFQ/viewform>

Metadata Directory Collection

In cooperation with the UNC SILS Metadata Research Center and the Research Data Alliance students at UNC SILS are gathering information about metadata standards that apply to scientific data. The aim of this is achieve a short-term goal of the Research Data Alliance's Metadata Standards Directory Working Group to:

Develop a prototype wiki-based directory (RDA Metadata Directory) listing metadata standards applicable to scientific data. The initial emphasis will be on widely used and domain community- endorsed metadata standards and schemas with significant interoperation / re-use capability.

Information submitted to this project will be integrated with similar information from the Digital Curation Center to prototype and build a sustainable platform for sharing and exposing information about metadata standards.

Please submit one standard per form. At the end of the questionnaire a link to another form submission is available.

* Required

Standard Name *
[TEXT ENTRY BOX]

Description of the Metadata Standard
May be a URL to about page or free text summary.
[TEXT ENTRY BOX]

Metadata Standard URL *
URL for the standard's home page
[TEXT ENTRY BOX]

Domain of the Metadata Standard *
Domains where the standard is in use

- ☐ Biology
- ☐ Earth Science
- ☐ Physical Science
- ☐ Nuclear and Particle Physics
- ☐ Social Sciences
- ☐ General Research Data
- ☐ Economics

Other: [TEXT ENTRY BOX]

Discipline Keywords
Keywords related to the use of the standard and the area of work done; comma separated
[TEXT ENTRY BOX]

Related Resources

Provide a URL if known (multiple URLs should be separated by commas) or text description if none is available.

Tools for using the Standard
[TEXT ENTRY BOX]

Extensions of the Standard
[TEXT ENTRY BOX]

Repositories, data portals or organizations using the standard
[TEXT ENTRY BOX]

Submitter Information

The following fields are optional; the information gathered will be used to make inferences concerning the environments in which each standard is used.

Submitter Name
[TEXT ENTRY BOX]

Submitter Organizational Affiliation
[TEXT ENTRY BOX]

Submitter Contact Email
[TEXT ENTRY BOX]

Do you consent to the information from the "Submitter" section being made public?
Answering 'Yes' means that your contribution to this effort as well as personal and organizational affiliations may be made public

- ○ Yes
- ○ No

Appendix C – Cover letter explaining the purpose of data gathering

As you may remember the Research Data Alliance's Metadata Standards Directory Working Group is working on developing a robust and usable directory of metadata standards used in the scientific data contexts.

We are beginning our work by asking for contributions from the community.

Contributions are simply submitted through a web form and will be incorporated into the information that the Digital Curation Centre has complied in an effort to document metadata standards.

The Working Group is looking for information about metadata standards; the tools and use cases associated with them, and additional information that shows where and how scientists use them worldwide.

By participating, you will be contributing information that can enhance the Digital Curation Centre's metadata directory (http://www.dcc.ac.uk/resources/metadata-standards). A long-term goal is to contribute to a community sustainable directory of metadata standards.

The form can be found here: Metadata Directory Information Collection or directly at: http://bit.ly/1fToaqd

Direct questions and feedback to: Sean Chen <schen@law.duke.edu> and/or Cristina Perez <Perez.C1988@gmail.com>

Thank you for your interest and time,

Sean Chen
Digital Resources Librarian
J. Michael Goodson Law Library
Duke University School of Law

Cristina Perez
MSLS Candidate 2013, UNC Chapel Hill
Student Services Contractor
Information Management Support Division | Office of Science Information Management | Office of Research and Development | US EPA


DCMI Science and Metadata Community


RDA
RESEARCH DATA ALLIANCE
Research Data Sharing
without barriers


DCC


DataONE


euroCRIS
Current Research Information Systems


UNC
SCHOOL OF INFORMATION
AND LIBRARY SCIENCE
Metadata Research Center <MRC>