Metadata 2020: Updates and Plans

Juliane Schneider, Metadata 2020 Librarian Group Chair

RDA, Berlin
Mar 2018

METADATA 2020

What is Metadata 2020?

Metadata 2020 is a **collaboration** that advocates richer, connected, and reusable, open metadata for all research outputs, which will advance scholarly pursuits for the benefit of society.

Who is involved?

- A team of 19 advisors initiated and helped launch
- A core team of 12 individuals drive structural next steps
- Over **70 individuals** in community working groups: Publisher; Librarian; Service Provider / Platforms and Tools; Data Publisher and Repositories; Funders
- So far around 35 individuals signed up to projects

Defining the issues

METADATA 2020

Group Work

- Each group has met 2 - 3 times
- They have defined their community problem statements, outlining challenges and opportunities
- Ideas that arose from multiple meetings are now resulting in specific **cross-community projects**
- Shortly, the project teams will be created and additional individuals will be invited (or can volunteer) to contribute

Community Groups

- Publishers
- Librarians
- Service Providers & Platforms and Tools
- Data Publishers and Repositories
- Researchers
- Funders

Publisher Challenges

- Metadata management is **enormously costly** for publishers
- The community **lacks an effective metadata distribution** model
- The author community does not yet understand the rationale for supplying full metadata and see deposit as an inconvenience
- Metadata **versioning is a problem**; metadata goes through a number of manipulations which create challenges around version of record
- Vendor submission systems have certain requirements that **publishers cannot control**
- Attaching appropriate metadata to previously published material would be a huge challenge

Publisher Opportunities

- Potential for **collaboration to define a consistent vocabulary** around metadata for researchers
- Potential to work with submission system platform providers to facilitate more **efficient systems**
- **Develop business cases** for improved metadata that directly results in optimized discoverability; and examples of, or pilots for incentive structure for optimized metadata internal to organizations
- Thoroughly **research customer needs** regarding information required
- **Map metadata**, and collectively work with service providers to define interoperability requirements

Service Provider/Platforms and Tools

Challenges

- Metadata creators make **assumptions that don't travel** with the data
- **Lack of consistency of metadata schema**, requirements, and entry; and inconsistency in dates and special characters
- **Records can disappear** without transparency as to where or why
- **No followed community standards about metadata vocabulary**
- Inadequate understanding of the **context of metadata** and its uses lead to improper choices of schema
- Metadata is often **out-of-date** and updates are not supplied by the metadata creators

Service Provider/Platforms and Tools

Opportunities

- Collaborate with other communities to form and distribute **best practice guidelines**
- Collaborate with other communities to **map metadata**, build awareness about metadata uses, and detect inefficiencies
- Build **awareness of metadata** uses among other communities through use cases, metadata mapping, and community engagement
- Develop and integrate **metadata evaluation tools** for metadata creators in consultation with others in schol comms
- Develop and integrate **new metadata tools** to increase interoperability across systems and avoid duplication of metadata entry

Conclusion

Communities have **similar problems** and **similar solutions** available if they **collaborate.**

Breaking down silos and getting stuff done

METADATA 2020

2018

With all of this opportunity for improvement, this year will be year of ACTION for Metadata 2020.

Project 1: Researcher Communications

Exploring ways to align efforts between communities who aim to increase the impact and consistency of communication with researchers about metadata.

CHALLENGES include:

- Motivating authors and editors to provide accurate and complete metadata due to time constraints
- Authors are largely unaware of the downstream effects and benefits of high-quality metadata
- Researchers and readers have different needs and requirements depending on their field of work

POSSIBLE SOLUTIONS TO EXPLORE include:

- Develop a suite of use cases, relevant to different fields
- Align with other projects
- Create a multi-channel outreach campaign for consistent communication
- Work closely with funders to communicate new mandated metadata deposit

Project 2: Metadata Recommendations and Element Mappings

To converge communities and publishers towards a shared set of recommended metadata concepts with related mappings between those recommended concepts and elements in important dialects.

CHALLENGES include:

- There are many different ways that metadata is created, vetted, used and distributed; and the complexity of this makes finding new efficiencies and systems implementation difficult
- Most groups face interoperability challenges with systems and processes
- There are silos within organizations themselves, making communications challenging

POSSIBLE SOLUTIONS TO EXPLORE include:

- Identify concepts included in relevant community and publisher recommendations
- Identify concepts shared across recommendations
- Mapping can reveal the complexity of metadata use to researchers
- Identify inefficiencies, breakages, etc. to address interoperability problems

Project 3: Defining the Terms We Use About Metadata

In order to communicate effectively about anything, a common language must be acknowledged, tacitly or purposefully. In the metadata space, there is not agreement on what words like property, term, concept, schema, title refer to. This project will develop a glossary of words associated with metadata, both for core concepts and disciplinary areas.

CHALLENGES include:

- Inconsistent definitions describing metadata across scholarly communications
- Librarians, publishers, and researchers use different vocabulary
- Metadata vocabulary is particularly important in regards to accuracy

POSSIBLE SOLUTIONS TO EXPLORE include:

- List current definitions used and map to find most common uses
- Define core metadata glossary of terms e.g. “concept”, “schema”, “title”
- Evolve consistent core metadata glossary to speak to different research fields

Project 4: Incentives for Improving Metadata Quality

To highlight downstream applications and value of metadata for all parts of the community, telling real stories as evidence of how better metadata will meet their goals.

CHALLENGES include:

- Difficulty in describing the importance of metadata conceptually and practically
- Many organizations have one or two metadata ‘evangelists’ but they often lack the decision-making authority on budget or development roadmaps
- Responsibility for metadata quality is often outsourced and not a strategic priority

POSSIBLE SOLUTIONS TO EXPLORE include:

- Gather value stories for each community group, understand their motivations and incentives
- Highlight and tell the stories of downstream metadata applications
- Identify compelling channels through which to deliver these stories and work with communications project to disseminate incentive stories in a compelling way

Project 5: Shared Best Practice and Principles

To build a set of high level best practices for using metadata across the scholarly communication cycle, in order to facilitate interoperability and easier exchange of information and data across the stakeholders in the process.

CHALLENGES include:

- Lack of central core principles, best practices, and consistent guidance
- Researchers are defining their own standards due to lack of direction
- For publishers, big obstacle in synchronizing capture is backlog of

POSSIBLE SOLUTIONS TO EXPLORE include:

- Similar to FAIR principles, define core principles for metadata around scholarly communications, created and disseminated in easily digestible ways for different groups
- Discuss metadata ownership and governance

Project 6: Metadata Evaluation and Guidance

To identify and compare existing metadata evaluation tools and mechanisms for connecting the results of those evaluations to clear, cross-community guidance.

CHALLENGES include:

- Difficulty in evaluating completeness, consistency and accuracy of metadata deposited across multiple systems
- Researchers do not have a good way of assessing the completeness of their metadata
- Publishers do not have a clear idea of the compliance of metadata to multiple standards

POSSIBLE SOLUTIONS TO EXPLORE include:

- Creation of a spirals-like metadata evaluation system for integration with submission systems, assessing the quality and completeness of metadata against standards
- Identify simple quantitative metrics that can be used to measure and monitor completeness, consistency and accuracy of metadata
- Catalog of metadata quality tools and further resources

Can you help?

- Contribute a books perspective to Metadata 2020 projects! Email Clare Dean at cdean@metadata2020.org for details
- Help promote our efforts to the wider community through your organizations, word of mouth, and social media
- Find us on @Metadata2020 Twitter, Facebook, LinkedIn, and at metadata2020.org

Questions

METADATA
2020

Thank you!

Metadata2020.org
@metadata2020
info@metadata2020.org

METADATA 2020