Packaging Specification for Simultaneous Deposit of Articles and Data into Multiple Repositories

Sayeed Choudhury and Hanh Vu

Sheridan Libraries, Johns Hopkins University

JOHNS HOPKINS
UNIVERSITY & MEDICINE

Public Access Submission System

Open-Source Software and Hosted service option:

- Simultaneous deposit of articles into institutional repositories and PubMedCentral (used by nine federal agencies including NIH, CDC, FDA, Howard Hughes Medical Institute, NASA)

Harmonizing federal public access and institutional open access policy compliance:

- Thereby reducing researcher burden and raising awareness of open access

1). What current limitations exist to the effective communication of research outputs (publications, data, and code) and how might communications evolve to accelerate public access while advancing the quality of scientific research? What are the barriers to and opportunities for change?

"A possible solution is the creation of more one-stop-shop access points for researchers that integrates grantee and funder operating procedures and requirements. One illustrative example is the PASS System developed by Johns Hopkins University, which is making great strides in simplifying the reporting, sharing, and compliance components of federally funded research."

NSF Grant #1939291

EAGER: Open Infrastructure to Reduce Burden on Researchers and Federal Agencies:

- Development of draft specification for simultaneous deposit of articles (and data) into multiple federal and institutional repositories

Partners include Arizona State University, California Digital Library, Duke, Harvard, MIT, Notre Dame, University of Michigan:

- With input from NLM, NSF ACI, and DOE/OSTI

Starting points for specification development

JOHNS HOPKINS
UNIVERSITY & MEDICINE

Foundational experience

- Examine existing specifications ([BagIt](#), [Data Conservancy Packaging Specification](#), DSpace + METS packaging, etc.)
- Review metadata requirements from various funding agencies
- Review deposit APIs

Major take away and concerns to be addressed:

- Extensibility, sufficient features without being too complex
- Correlation (relating people, awards, funders, scholarly content)
- Metadata
- Trust

Overarching recommendations (1/2)

1. Adopt [BagIt](#) as the packaging mechanism for data exchange
    More detailed recommendations outlined in the [package proposal](#) and [conceptual model](#).
1. Support and implement the ability to unambiguously identify investigators, co-investigators, and authors using ORCIDs in submission metadata and public-facing agency systems or data feeds (e.g. NLM PACM, Federal Exporter).
2. Support and implement the ability to unambiguously identify awards in submission metadata and public-facing agency systems or data feeds.

Overarching recommendations (2/2)

4. Support and implement the ability to include correlation identifiers (supplied by institutions in the package) in public-facing agency systems or data feeds.
5. Develop and implement mechanisms for evaluating trust of the information contained in a package based on submission metadata.

Concerns and feedback

- Identifiers: tracking, correlation
- Use case for metadata-only packages
- Type of support for this specification
- Relationship between SWORD and this specification
- Conceptual Model Feedback

Plan for responses and feedback

Feedback from workshop attendees had been received and to incorporate into the next steps of our work:

- Overview document describing goals, overarching concerns and how we address them

- Draft specification: Created from the content of the spec proposal incorporating relevant feedback

- Updated conceptual model doc to incorporate any relevant feedback from workshop attendees. May create physical model as well to aid with understanding and application of conceptual model

Acknowledgements

JOHNS HOPKINS
UNIVERSITY & MEDICINE

We want to thank:

- Our funder, NSF, and program officer, Beth Plale, for the opportunity to work on this interesting and promising work.
- Kathryn Funk (NIH) and Mary Beth West (OSTI) for offering additional input from funders' perspective at our workshops
- The workshop attendees Debra Hanken Kurtz (ASU), Mahjabeen Yucekul (CDL), Tim McGeary (Duke), Scott Lapinski (Harvard), Jim Ottaviani (UMich), Helen Bailey (MIT), Osman Din (MIT), Sandra Gesing (Notre Dame), Rick Johnson (Notre Dame) for offering invaluable feedback.

Invitation for community feedback

Overarching recommendations:
https://drive.google.com/open?id=1pLrkJWhyQE58PLIlQi2dAM_K7LtKGoFTSdaQIyfCNM4

Approach and proposal for packaging spec:
https://docs.google.com/document/d/1epR1z4gTqxtJUIbrj3lZdwlUj-c9vBbKCTLxHnPMsp4/edit?usp=sharing

Proposed conceptual model:
https://drive.google.com/open?id=1Qqw0JkQ8ze2_ZaSxz_P5zyQCrUg--N2gRdnuzXQmyK4

Comment directly on the documents OR email us at
hanh.vu@jhu.edu or sayeed@jhu.edu