Updated: Apr 5, 2024
Finalized: Jun 6, 2024

Proposal: GREI Harvard Dataverse Repository and Dataverse Software Metadata Support

IMPORTANT: Jun 6, 2024 This proposal is now finalized. The related GitHub issue has been closed: https://github.com/IQSS/dataverse.harvard.edu/issues/230. Proposal implementation will be tracked here: https://github.com/IQSS/dataverse-pm/issues/127

Purpose

- Define and plan the work needed to ensure that Harvard Dataverse and Dataverse software supports the GREI Metadata and Search Subcommittee Recommendations.

Updated: Apr 5, 2024
Finalized: Jun 6, 2024

Background

The NIH GREI Metadata and Search Subcommittee developed a set of recommendations for metadata that all GREI repositories should strive to support. See GREI Metadata and Search Subcommittee Recommendations from DataCite schema version 4.4.

This proposal aims to describe how Harvard Dataverse repository and Dataverse software can support these metadata recommendations.

Tasks include identifying relevant current Github issues and relevant documentation.

Topics will include specific support for Harvard Dataverse, such as changes to the repository's metadata blocks, training, and support resources, as well as code changes to the Dataverse software that will benefit the wider Dataverse Community.

Related Work

- NIH OTA
  - NIH GREI Aim 2: Increase support for **biomedical and cross-domain metadata standards** and **controlled vocabularies, particularly regarding "expanding descriptive and citation metadata to support funding information and related fields"**
  - NIH GREI Aim 4: Improve **harvesting and packaging** standards to share metadata and data across repositories
  - NIH GREI Aim 5: Standardize **usage metrics** with the repository and across other repositories
- Related documents & presentations
  - How does Harvard Dataverse follow NIH GREI's metadata recommendatio…
  - Support for core PIDs in Dataverse
  - Draft plan for NIH GREI Open Metrics work
  - Harvard Dataverse's comments about NIH GREI metadata subcommittee's…

Approach

Plan

- Call for participants
- Review the GREI Metadata Recommendations

  Identify which recommendations Harvard Dataverse has already satisfied and which are already planned. See

How does Harvard Dataverse follow NIH GREI's metadata recommendations? for if and how Harvard Dataverse already satisfies these recommendations

- Create a plan to implement the recommendations in alignment with Dataverse goals and objectives and Harvard Dataverse users' needs
    - Merging DataCite and OpenAIRE exports
        - Update and continue work proposed in the GitHub issue at https://github.com/IQSS/dataverse/issues/5889 to merge the DataCite and OpenAIRE exports.
        - After the DataCite and OpenAIRE exports are merged, create GitHub issue to track work of sending to DataCite the metadata of datasets already published in Harvard Dataverse
    - Collecting and sending ORCIDs and ROR IDs
        - Organize members of the UX working group to continue research-informed re-designs of the citation metadata block, including:
            - The use of ORCIDs for authors and contributors that are people
            - The use of ROR IDs for authors, contributors, and funders that are organizations
            - The use of ROR IDs for affiliations of authors that are people and for affiliations of contributors that are people
        - This will include evaluation and recommendations to iterate on the external controlled vocabulary functionality
        - This will include close coordination with Jim Myers and his work on the ORCID integration grant
    - Collecting and sending metadata about related research objects
        - Organize members of the UX working group to continue:
            - Research-informed re-designs of the citation metadata block so that Dataverse can record and send PIDs and other metadata about related research objects
            - Discussions about other ways to improve the quality and amount of metadata about related research objects

- Create assessment plan to evaluate how well our implementations satisfy the recommendations

Assessment

Description of how the effort will be assessed for success and later course corrections

- Usability tests of redesigns of the citation metadata block to measure how easily and effectively depositors and curators can include ROR IDs and ORCIDs of people and organizations associated with their deposits
- Usability tests of redesigns of the citation metadata block to measure how easily depositors and curators can include PIDs and other metadata about research objects related to their deposits
- Tracking changes in the amount and quality of certain metadata published in the Harvard Dataverse: ORCIDs, ROR IDs, and metadata about related research objects

Participants

Participant	Role	Notes
Julian Gautier	Lead	
Members of the UX working group		

Assumptions and Dependencies

- This work is related to and may be dependent on grant-funded obligations to improve Dataverse's integration with ORCID
- This work is dependent on discussions related to handling changes to metadata models that result in the removal of metadata fields. See https://github.com/IQSS/dataverse/issues/10431

Requirements and Resources

List of any requirements and resources needed to complete the planned work (e.g., software, subscriptions, money)

Updated: Apr 5, 2024
Finalized: Jun 6, 2024

Requirements

-

Resources

-

Timeline

Start, end, and milestones

See

Milestone	Start	End
GREI Metadata Recommendations reviewed. See How does Harvard Dataverse follow NIH GREI's metad…	2023/05/09	2024/03/22
This proposal is drafted	2023/09/25	2024/0#/##
This draft proposal is shared with Dataverse Project team	2024/0#/##	2024/0#/##
Proposal for merging DataCite and OpenAIRE exports	2019/05/28	2024/04/08
DataCite and OpenAIRE exports are merged. See #5889	2024/??/??	
DataCite is sent new metadata (in merged DataCite/OpenAIRE export) of deposits published in Harvard Dataverse (See #5144)	2024/??/??	
Propose plan, including assessment plan, for collecting and sending ORCIDs and ROR IDs	2024/04/??	
Benchmarks of the amount of ORCIDs and ROR IDs in Harvard Dataverse metadata	2024/04/??	
Citation metadata block is redesigned and external vocabulary scripts are adjusted in order to improve collection of ORCIDs and ROR IDs		
Propose plan, including assessment plan, for collecting and sending metadata about related research objects		
Benchmarks of the amount of related research object metadata published in Harvard Dataverse		

Cycle of redesign and assessment of the citation metadata block		
Code for creating DataCite exports is adjusted to include the ORCIDs, ROR IDs, and metadata about related research objects that are collected using the redesigned citation metadata block and external vocabulary scripts		
Redesigned citation metadata block and adjusted DataCite export code are applied to Harvard Dataverse		February 2024 (end of GREI year 3)
Metadata of deposits published before DataCite export was adjusted is sent to DataCite (See #5144)		February 2024 (end of GREI year 3)
A year after Harvard Dataverse's citation metadata block is redesigned: ● Conduct benchmarks of the amount of ORCIDs, ROR IDs, and related research object metadata published in Harvard Dataverse ● Review cases where depositors entered organizational names and Dataverse did not include ROR IDs, and work with ROR to add entries for organizations that their database is missing ● Conduct usability tests of the citation metadata block, distribute results and make recommendations informed by them		A year after Harvard Dataverse's citation metadata block is redesigned

Deliverables

Specific deliverables that are expected at the end of the proposed work. May include documents, code improvements, and events or training.

- Adjustments to code that creates the DataCite export (after the DataCite and OpenAIRE exports are merged)
- Updates of the Dataverse 4+ Metadata Crosswalk to reflect changes to the DataCite export after the DataCite and OpenAIRE exports are merged
- Scripts or other methods for sending to DataCite the updated metadata of the DataCite exports of deposits in the Harvard Dataverse and other Dataverse installations

- Plans to assess the effectiveness of changes to the citation metadata block and DataCite exports:
    - Members of the UX working group will plan for testing the usability of changes to fields in the citation metadata block, such as usability testing scripts, who will moderate tests, which depositors and curators will participate in testing, who will review and share testing results, and who will create recommendations based on the results
    - Members of the UX working group will plan for benchmarking the amount of certain metadata in deposits published in Harvard Dataverse and other Dataverse repositories, such as the count of ORCIDs and RORs and the amount of PIDs and other metadata of related research objects
    - Members of the UX working group will plan for measuring changes in the amount and quality of certain metadata published by Harvard Dataverse and other repositories that have been affected by changes to the citation metadata block and DataCite exports for a certain period of time
- Redesign of the citation metadata block, including its TSV and .properties files
- Possible adjustment of scripts used for adding external vocabularies to fields in the citation metadata block
- Further adjustments to the code used to create the DataCite export (whose metadata is available through the UI, API and in OAI-PMH feeds)

Outcomes

What you expect to achieve at the end of the project, beyond deliverables. For example: "Improved communication with stakeholders" or "Fewer tickets related to spam"

- ROR IDs and ORCIDs are more easily and more often included in the metadata of deposits published in Harvard Dataverse and other Dataverse repositories and in the metadata they send to DataCite
- PIDs and other metadata about related research objects are more easily and more often included in the metadata of deposits published in Harvard Dataverse and other Dataverse repositories and in the metadata they send to DataCite

Benefits

List of benefits for the proposed work

- People, including people from organizations that fund research, can more easily and effectively find data and code published in Harvard Dataverse and other Dataverse repositories
- People can more easily and effectively find research related to the data and code published in Harvard Dataverse and other Dataverse repositories
- As Harvard Dataverse and other Dataverse repositories send to DataCite more metadata about the relationships between the deposits they publish and other research objects, people can more easily measure the influence of data and code

Risks

Which resource constraints may limit the ability of team members to implement the plan

- Limitations of the technology used to create UI form interactions may constrain the redesign of the citation metadata block and the external controlled vocabulary functionality, especially if stakeholders decide that some design options shouldn't be implemented in Dataverse's current infrastructure

Risk Mitigation

Plans for mitigating the identified risks

- Redesigns may be more feasible by working as closely as possible with developers who are familiar with the opportunities and constraints of the technology used to create UI form interactions, including technologies used in Dataverse's current architecture and the SPA.

End document.