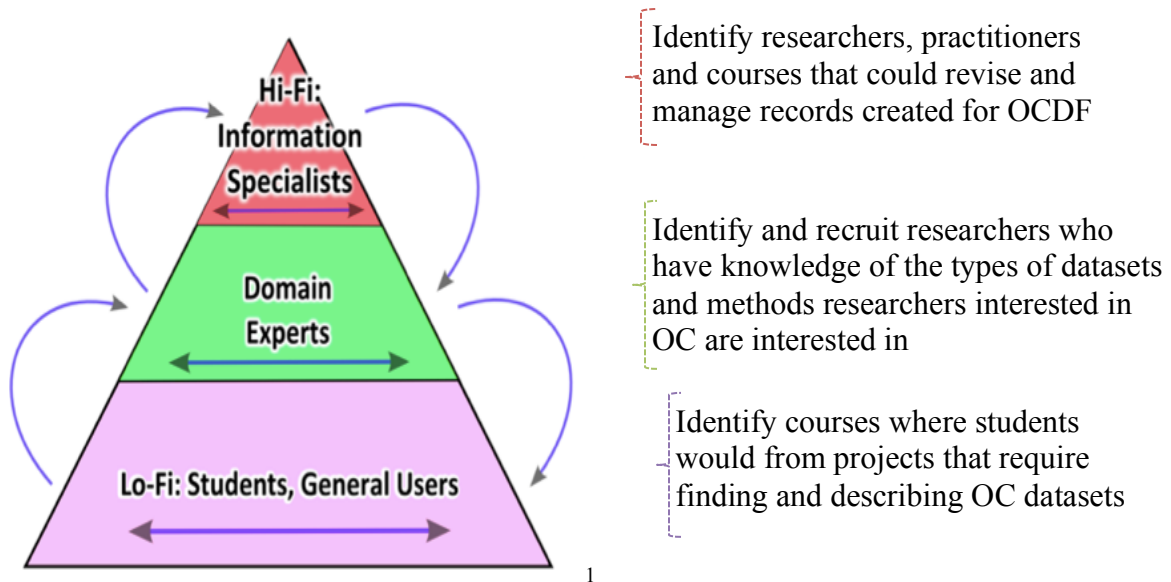


Developing an effective OCDF metadata workflow will make it possible to provide information about existing datasets about online communities. An effective workflow requires developing standards for metadata entry and revision. Developing metadata standards that reflect the interests of researchers from a variety of academic disciplines will enhance the variety and quality of contributions to the OCDF. In general, there are two different goals of establishing a workflow: first, establishing where to collect information about datasets from, and second, outlining best practice guidelines for metadata creation, revision and maintenance. Identifying interests, needs (technical and practical) and resources will make it possible map specific practices onto tools and functions that support the OCDF. Doing this will make it possible to describe how metadata can enhance projects undertaken by other working groups contributing to OCDF and, thus, provide opportunities for enhancing the infrastructure and tools available for recording information about datasets. Also, it will promote further opportunities for courses and research relating to the exploration and analysis of open online communities.

Identifying contributors:

Contributors to OCDF will have a broad range of experience working with metadata, which means that a workflow and model for creating, curating and maintaining metadata is essential. Reasonable and varied expectations among contributors will promote ongoing, diverse modes of participation.

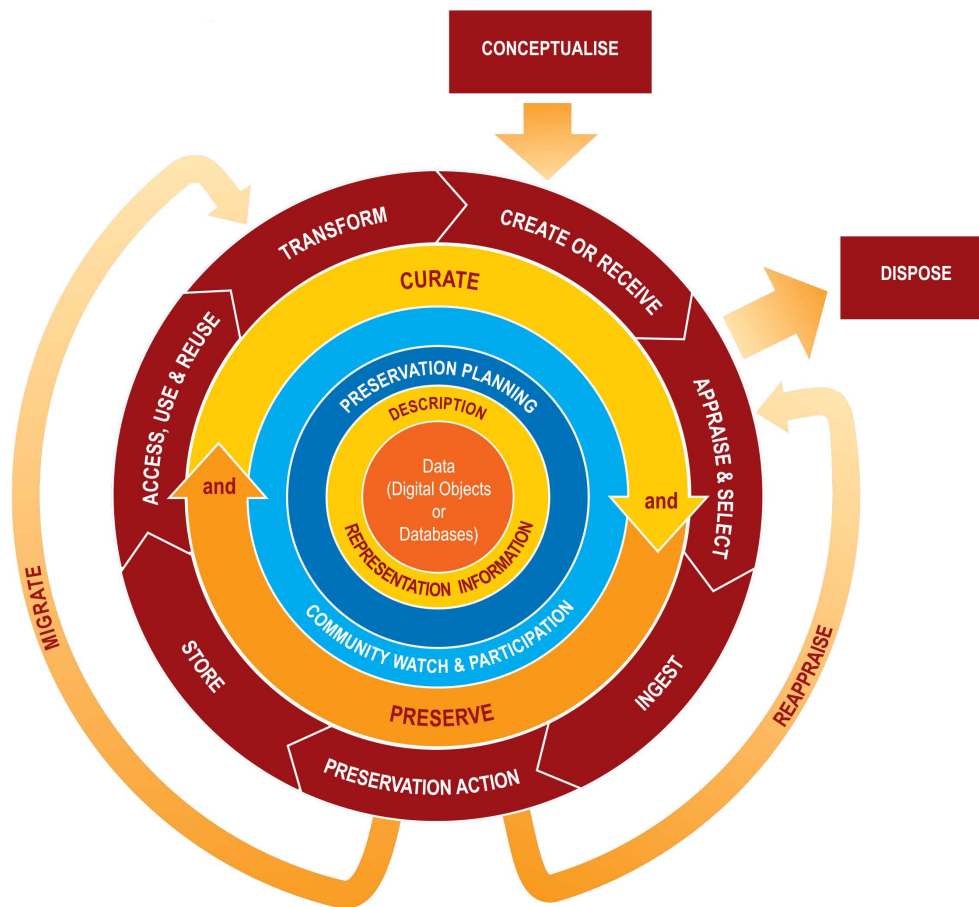


¹ Maron, D.; Missen, C.; McNeirney, K. & Elnora, K.T. (2015). Lo-fi to hi-fi crowd cataloging: Increasing e-resource records and promoting metadata literacy within WiderNet. Poster presented at the iConference.

Identifying expectations for different types of contributors will:

1. Help clarify definitions of what constitutes a dataset about open online communities so that future contributors to the OCDF have a better idea of how to search for, evaluate and describe creators of datasets.
2. Make grading student contributions to the OCDF through the completion of in class projects easier.
3. Establish periods of time that are most suitable for evaluating, revising and indexing records created by students and other contributors to the OCDF.
4. Provide a clearer set of guidelines for identifying datasets that meet the research/ethical requirements for 'quality' datasets
5. Enhance understandings of the OCDF's technical needs (i.e. infrastructure).

Identifying users and the types of contributions they will make to OCDF is a first step towards building a functional and sustainable workflow. The goal of establishing this workflow is to ensure that appropriate datasets are identified and that descriptive records maintained in OCDF are stored and maintained in a manner that allows users to continually access and work with them. The following workflow uses the DCC Curation Lifecycle model to outline a series of iterative tasks that help identify actors, actions and technologies that can contribute to the collection, creation and maintenance of records for datasets about online communities. There are six total rings, each requires a series of action and contributions; however, none are separate from the other and none are one-time only actions or activities.



2

² Jisc (n.d.). DCC curation lifecycle model. Retrieved from: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>

First ring:

Each section of the first ring makes it possible to establish practices that enhance the integrity and coherence of a digital asset. In the case of OCDF this means establishing practices and time frames for identifying relevant datasets, cataloging these datasets and revising metadata when necessary. This portion of the workflow will make it possible to curate a series of records that are accurate, up to date and searchable.

Working groups that can benefit from this:

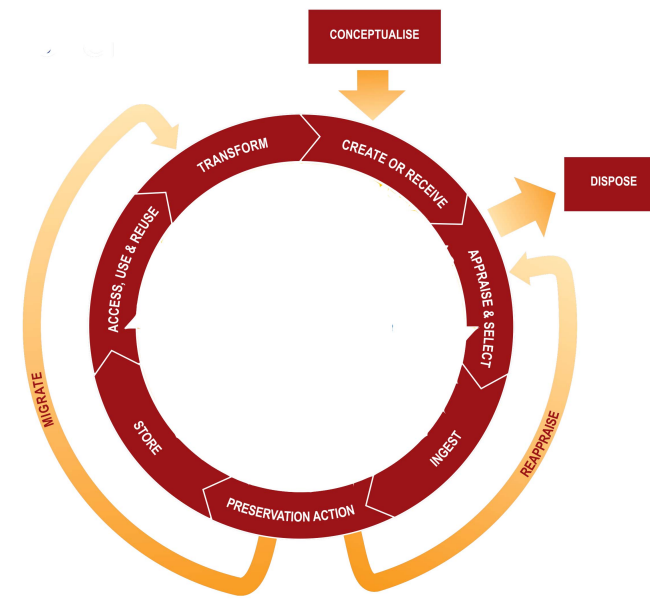
- Leadership

Working group projects that will contribute to understanding this

- Identify possible contributions from working groups
- Establish guidelines for providing feedback on metadata and ontologies
- Begin discussing definitions that will help improve the quality AND quantity of contributions to OCDF
- Begin discussing long term planning and evaluation: Stakeholders, contributors, concerns and areas for improvement.

Student participation:

Currently the OCDF uses Semantic MediaWiki to host records for datasets. This accounts for a form of creating/receiving information about datasets and supports the evaluation of the Semantic MediaWiki's overall functionality as research tool and resource.



Second ring:

The second ring makes it possible to conceptualize the function and purpose of datasets, which makes it possible to evaluate contributions, timelines for contributions as well as methods of identifying future contributors.

During the actions in the second ring it is time to evaluate and revise metadata standards for the data census. Seeking out feedback and evaluations of the process of creating metadata will be necessary and essential.

Working groups that can benefit from this:

Metadata and ontologies

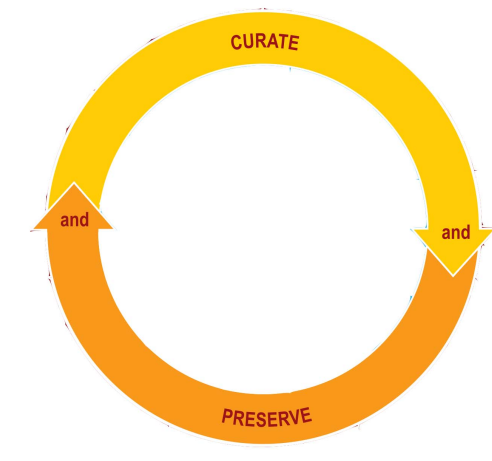
Infrastructure

Working group projects that will contribute to understanding this

- Establishing a content standard for data entry
- Establishing guidelines for evaluating metadata records
- Working with OCDF participants to design courses that encourage more in-depth/long term student participation
- Evaluating technologies and systems for metadata entry and review: what can be automated, what needs changed?

Student participation:

- Courses in data management, database management, information retrieval, metadata etc. can help evaluate the quality and functionality of metadata records, which will help curate information about datasets dealing with OOC.
- Digital stewardship, digital preservation etc. can help evaluate the practices used to maintain versions of metadata records.



Third ring:

Community watch and participation guarantee that the metadata standard used to catalog datasets is relevant and complete. Establishing acceptance of a metadata standard will ensure that metadata created for datasets included in OCDF is accurate, relevant and useful. This process can take many forms: online evaluation forms, research on the value of participating in a data census, a review of the types of datasets cataloged in the data census etc. This is an opportunity to critique practices and the workflow thus far.

Working groups that can benefit from this:

- Leadership
- Ethics

Working group projects that will contribute to understanding this

- Encouraging further contribution from working groups
- Recruiting new participants and projects
- Building relationships with stake holders
- Evaluating and discussing projects, issues, concerns and emerging best practice guidelines

Student participation:

- Assignments that require students contribute to the OCDF has the potential to evolve into a survey of existing datasets and repositories for datasets, which presents an opportunity for projects (student and otherwise) to require students consider the ethical implications of OOC research.
- Having specific metadata fields that highlight modes of dataset creation and curation has the potential to engage students in a process of evaluating datasets, which presents an opportunity to begin establishing criteria for inclusion in the OCDF.



Fourth ring:

OCDF is not a repository for datasets. So, instead of planning on methods and technologies for preserving datasets it will be necessary to establish practices, procedures and technologies for managing practices for validating records and keeping back-up versions of valid records.

Preservation planning is an opportunity for reflecting on the implications of editing and versioning metadata standards.

Working groups that can benefit from this:

- Metadata and ontologies
- Infrastructure
- Ethics

Working group projects that will contribute to understanding this

- Metadata schema: Establishing a timeframe for reviewing records
- Infrastructure: Establishing automated systems for validating links and other content entered into records by contributors
- Establishing criteria for evaluating datasets to ensure they meet standards and best practices outlined by researchers.

Student participation:

- Student evaluation's of OCDF metadata forms can support planning for revisions and improvements to the schema and content standard.
- Identifying assignments and grading periods in classes can help establish a timeframe for evaluating and revising metadata records.



Fifth ring:

Documenting and sharing preservation strategies and practices will make it possible to demonstrate the overall function and value of the data census. Sharing metadata standards and workflows will make it possible to better describe how metadata enhances the accessibility and re-usability of datasets and the methods used for creating them.

Working groups that can benefit from this:

- Metadata and ontologies
- Infrastructure

Working group projects that will contribute to understanding this

- Metadata schema: Evaluating terms and content standard
- Infrastructure: Usability and sustainability of the system, and identifying future technical needs

Student participation:

- Class assignments will structure ways for students to identify and describe OOC datasets. This will help develop a broader knowledge of institutions, repositories and projects with goals similar to OCDF, or potential stakeholders who could contribute to OCDF.



Sixth ring:

The goal of the data census is to make datasets available so that research methods can be replicated. This requires providing consistent access to datasets and information about these datasets. Each ring of the DCC Curation lifecycle model provides an opportunity to begin conceptualizing stakeholders and activities that will make this possible.

