Digital Content Reviews: Mapping the Landscape of Digital Content at MIT Libraries

Helen Bailey, hbailey@mit.edu

Nancy McGovern, nancymcg@mit.edu

Data-Driven Digital Curation

Sustainable digital curation programs require a means to routinely gather and analyze information about both currently managed and anticipated digital content. To address this community need, MIT Libraries has demonstrated how to extend and enhance a proven digital content review process developed by and for the Digital Preservation Management workshop curriculum and tested at several organizations.

The digital content review (DCR) is a structured process that uses a life cycle framework for information gathering and analysis to produce an overview and report of digital content types, a landscape view of all the digital content in scope for an organization, and a data set for data-driven curation and preservation planning.

Digital Content Review Ontology

- Digital Content Review Data Set
  - Experimental approach to manage data gathered in the digital content review.
  - Dataset contains cumulative information about all digital content types reviewed.
  - Uses a digital content review ontology and linked data format to model data for strategic planning.
  - Interactive web-based data querying and visualization allow for a dynamic approach to working with the data set (see demo).
  - Extensibility of data model offers flexibility for new data types to address future needs.

- Life Cycle Data Management
  - Digital content review process, results, data creation, data visualization, and planning stages are iterative and cumulative.
  - Data about the digital content must be managed along with the content itself.

- Digital Content Review
  - Structured process to complete a gap analysis for a wide variety of digital content types.
  - Uses a template based on eight life cycle stages with questions for each section.
  - Identifies implications for managing digital content types across generations of technology.
  - Results in a digital content report and an overview of each content type.
  - Allows for prioritization of digital content management improvements.

DCR Template/Report Sections: Review of Life Cycle Stages

- Collection Development/Selection
- Rights Management Assessment
- Acquisition/Accessioning
- Preservation Planning
- Intellectual property issues, rights language, processes
- Process, metadata, formats
- Process, metadata, training
- Content/project management, monitoring, costs
- User agreements, confidentiality, capacity

Overview Diagram Conventions

- Scope: Content currently managed by MIT Libraries
- Status: Active
- Shape: Circle, top-down view of a content "bucket"

- Scope: Content currently managed by MIT Libraries
- Status: Active
- Shape: Triangle, could become a bucket

- Scope: Digital content managed elsewhere, with services at MIT
- Status: Monitor
- Shape: Square, not likely to come to MIT

Additional diagram conventions include:

- Size of bucket represents relative amount of content.
- Dotted line indicates digitized material.
- Gray fill represents content that will be acquired in the future.
- Arrows represent relationships between buckets.

Digital Content Review Ontology

- Digital Content Types
  - Major categories or types of content that need to be managed and/or monitored.
  - Unique to an institution, based on types of collections managed and acquired.
  - Identify stakeholders, curators, and information-holders for each content type.
  - Conduct interviews to gather data on the current content and potential future content in each content type.
  - Determine preservation and curation challenges specific to each content type based on workflow, format, anticipated uses, etc.

- Digital Content Overviews
  - Uses information gathered in the digital content review to develop a high-level overview of each content type.
  - Documents current status of workflow and practice for each content type.
  - Identifies content with common characteristics ("buckets").
  - Provides a concise view of each digital content type highlighting areas for improvement.
  - One overview per digital content type.
  - Cumulative overview of digital content.

- Digital Content Landscape
  - Visualizes data about all of the digital content types as a whole.
  - Offers insight into patterns of content and needs for infrastructure and planning.
  - Highlights areas of high priority due to factors such as amount of content, expected acquisition timeline, formats, etc.
  - One of many views onto digital content review data, offering a high-level picture of the entire landscape of digital content.
  - Goal is to display entire overview image, not detail.

Sample Overview Diagram: Research Data