Digital Initiatives Symposium

May 2nd, 2:45 PM - 3:25 PM

Research Data and the Duke Digital Repository: Building a Service-Driven Data Curation Workflow

Mara Sedlins
Duke University, mara.sedlins@duke.edu

Research Data and the Duke Digital Repository: Building a Service-Driven Data Curation Workflow

Session Type
Poster

Abstract

This year, Duke University Libraries launched the Duke Digital Repository (DDR), a service that supports the activities of the University's faculty, researchers, students, and library staff by preserving, securing, and providing access to digital resources. The DDR is built on three program areas: research data, scholarly publications, and library collections. The research data area of the DDR reflects a new focus on research data management services at Duke, which includes preservation-level storage in the DDR, as well as consulting and instruction. The new program of services is supported by four new staff members (two research data management consultants and two digital repository content analysts), as well as a CLIR postdoctoral fellow. Along with two subject specialist librarians and the head of Data and Visualization Services, we form the Research Data Working Group at Duke Libraries, which provides policy, procedure, and technical recommendations on data curation and management issues related to the library's repository program. We are currently piloting new services, policies, and workflows to promote data sharing and curation, quality metadata, and reproducible research, offering consultations and workshops in research data management best practices. This poster will report on the organizational structures, data sharing policies, metadata requirements, and workflows for data ingest that are currently in place.

Location
Rotunda / Garden of the Sea

Keywords
institutional repository, data curation, research data management, preservation

Research Data and the Duke Digital Repository: Building a Service-Driven Data Curation Workflow

Research Data Working Group at Duke Libraries: Jennifer Darragh, Moira Downey, Elena Feinstein, Arianne Hartsell-Gundy, Joel Herndon, Susan Ivey, Sophia Lafferty-Hess, & Mara Sedlins

DUKE UNIVERSITY LIBRARIES

Duke Digital Repository

The Duke Digital Repository (DDR) is a service of Duke University Libraries that supports the activities of the University's faculty, researchers, students, and library staff by preserving, securing, and providing access to digital resources.

Collection Areas

Research Data
Data management and curation for research at Duke

Scholarly Publications
Open access for Duke authors

Library Collections
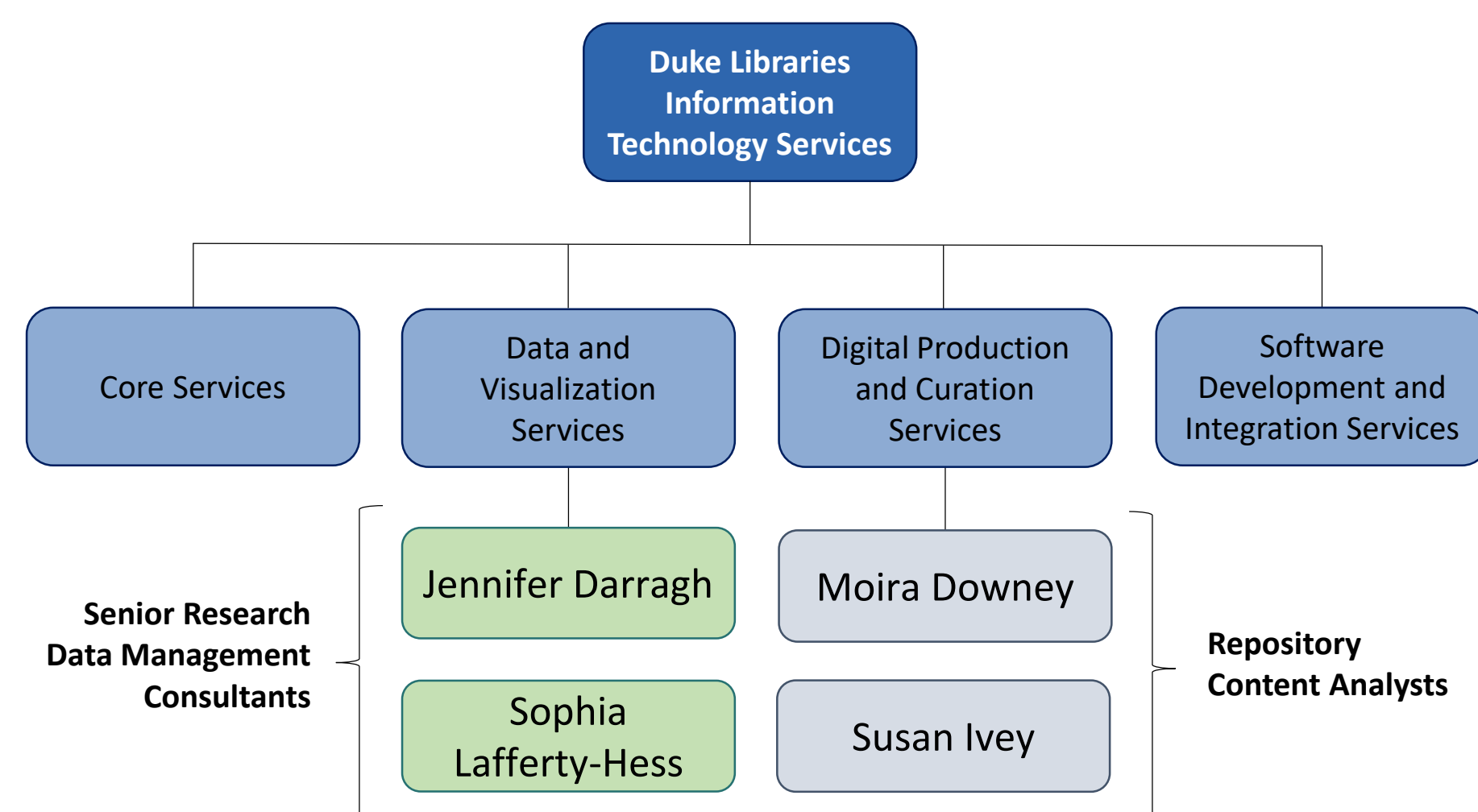Digital items owned or stewarded by Duke University Libraries

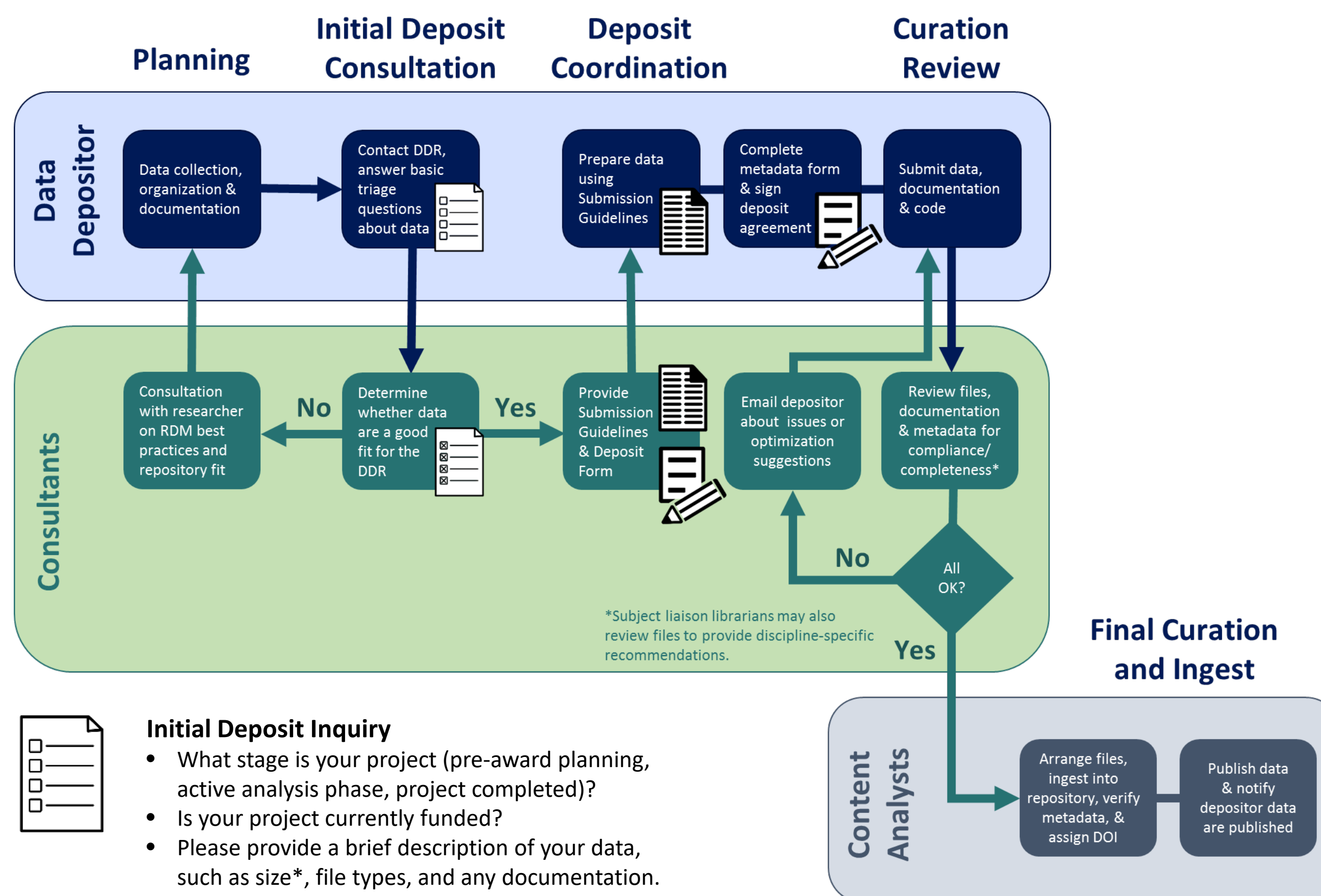## Research Data Curation Staffing Model

Fall 2015: Digital Research Faculty Working Group convened, made recommendations to expand research data services[1]

Fall 2016: CLIR Fellow in Data Curation co-sponsored by Duke Libraries and the Duke Social Science Research Institute

Winter 2017: Four new positions hired to support research data curation at Duke Libraries

Duke Libraries Information Technology Services

Core Services | Data and Visualization Services | Digital Production and Curation Services | Software Development and Integration Services

Senior Research Data Management Consultants: Jennifer Darragh, Sophia Lafferty-Hess

Repository Content Analysts: Moira Downey, Susan Ivey

repository.duke.edu

Research Data Curation Workflow

Planning — Initial Deposit Consultation — Deposit Coordination — Curation Review

Data Depositor:
- Data collection, organization & documentation
- Contact DDR, answer basic triage questions about data
- Prepare data using Submission Guidelines
- Complete metadata form & sign deposit agreement
- Submit data, documentation & code

Consultants:
- Consultation with researcher on RDM best practices and repository fit
- Determine whether data are a good fit for the DDR
- Provide Submission Guidelines & Deposit Form
- Email depositor about issues or optimization suggestions
- Review files, documentation & metadata for compliance/completeness*

All OK? No / Yes

*Subject liaison librarians may also review files to provide discipline-specific recommendations.

Final Curation and Ingest

Content Analysts:
- Arrange files, ingest into repository, verify metadata, & assign DOI
- Publish data & notify depositor data are published

Initial Deposit Inquiry
- What stage is your project (pre-award planning, active analysis phase, project completed)?
- Is your project currently funded?
- Please provide a brief description of your data, such as size*, file types, and any documentation.
- Is any of this data protected or sensitive? If so, please explain.

Submission Guidelines
- Prepare to describe your dataset in order for it to be discoverable in the repository (minimum: title, author list, description).
- Save your files in a preservation-friendly format.
- Provide appropriate documentation that clearly describes your data so that others can interpret and use your files (README, codebooks, etc.).
- Organize your files in a logical order, avoiding heavily nested folders.
- Decide on the appropriate license for your data.
- Leave files uncompressed.

Data Deposit Form
- Provide basic metadata – at minimum a title, author list, and a description
- Select a license – default is CC0, but depositors may specify another CC license
- Functions as a deposit agreement

Curation Review Checklist
- Review the Data Deposit Form to ensure metadata is clear and understandable.
- Perform a visual inspection of data files for the presence of PII or PHI. The liability for the presence of any sensitive information will remain with the depositor.
- Check that the data package includes adequate documentation to allow for others to interpret and use the data files.
- Check to see if files are stored in preservation friendly formats.
- *Optional*: Have subject liaison librarians review the data package to provide any other data documentation suggestions.

*Storage Policy
- 2.5 GB per researcher
- 10 GB for funded projects
- Additional storage available at cost

[1] Duke Digital Research Data Services and Support: A recommendation from the Digital Research Faculty Working group. January 29, 2016.

Check Form by Randomhero from the Noun Project
Document by Sandra from the Noun Project
Letter by Stock Image Folio from the Noun Project

Definition of Research Data

Research data are the original sources or materials (born digital or converted to digital) that were created or gathered in the process of your research. They serve as the foundation from which you draw conclusions and produce results/findings (test hypotheses, study trends, provide evidence, refute claims). They may be numeric or qualitative, structured or unstructured. Among many possible forms, data may take the form of notebooks, statistical or spatial data tables, audio or visual recordings, photographs or models. The end-results/findings of your research such as monographs, articles, white papers, or presentations are not considered research data.

Training and Outreach Initiatives

Fall 2016: CLIR Fellow conducted informal interviews with social science NSF grant recipients about data management planning

Winter 2017: Repository launch - presentations at campus Research Computing Symposium

New Research Data Management workshops piloted:
- Data Management Fundamentals
- Documenting Data in Excel with Colectica
- Data Management and Reproducibility: Enabling Open and Transparent Research Through Data Sharing
- Consent, Data Sharing, and Data Reuse
- Data Management Tools: The Dataverse Project
- Research Data Management for the Social Sciences
- Data Management and Grants: Complying with Mandates

Ongoing: RDM Consultant meetings with various campus stakeholders