

DATA MANAGEMENT TRAINING  
TO SUPPORT FACULTY RESEARCH NEEDS: LESSONS LEARNED  
JUNE 7, 2012, WASHINGTON, DC

Ryan Womack([rwomack@rutgers.edu](mailto:rwomack@rutgers.edu))

and the [RUresearch Data Team](#)

IASSIST 2012

- **RUcore** is the institutional repository of the Rutgers University Libraries (RUL). Based on a Fedora platform, RUL has developed its own interface for entering metadata (the Workflow Management System or WMS) and its own search interface for presenting repository content.
- Beginning in January 2011, the **RUresearch Data Portal** was made public. As RUcore Research Data Manager, I have engaged in outreach and presentations to faculty across Rutgers University.
- Our definition of research data includes anything that the researcher uses to create their scholarship. These can be numeric data sources, but can also include images, audio, video, or any kind of digital file.
- Using the existing strengths of the Fedora system allows us to provide complete context for the research data, linking to documentation, relevant published papers, software code, and so on.
- Rutgers is a large and diverse university (over 58,000 students and 6,700 faculty), and we must be prepared to accept data from any faculty member.
- Open Access policy is being considered by faculty, with a vote in October. Implementation of this policy will include education about the possibility of data deposit. We will need to be able to handle larger volumes of data in this case.

**RUresearch** is the research data component of **RUcore**, the Rutgers University Community Repository. RUresearch provides the following services:

- Consulting on Data Management Plans and data best practices
- Permanently archiving data in the RUresearch data portal
- Work on larger and more complex data needs in grant-funded projects

The [RUresearch Data Team](#) was constituted as the operational group behind RUresearch. This team brings together programmers, developers, data specialists, metadata librarians, and subject liaisons. I coordinate the activities of this team.

- To build the skills of its Data Team, an internal Research Data Management course was developed under the leadership of Grace Agnew, Associate University Librarian for Digital Library Systems.
- In fact, in order to become a member of the Data Team, training via the course is required (with two exceptions).
- According to a rough classification, the team consists of three data specialists, three metadata specialists, four subject liaisons, two programmers, and four cyberinfrastructure developers.
- The team is not a policy-making or deliberative body. Its function is to work on data projects (structuring data, ingesting into RUCore, and ensuring correct presentation of data).

- *Data Management Training to Support Faculty Research Needs* is the course's title, and its primary goal.
- The initial core of the course grew from Grace's experience teaching digital library metadata to Rutgers' School of Communication and Information students.
- Designed to give librarians and staff the tools and contextual knowledge needed to handle data in each of their respective roles.
- The course is team taught. Each of our in-house experts presents on their area of expertise.
- Grace coordinates the metadata portion of the course; Ryan coordinates the data management portion of the course.
- Two hour class sessions plus group homework assignments and discussion.
- Initial plan for at least monthly meetings. Some gaps in meetings due to busy schedules!

# COURSE OUTLINE - TOOLS

- 1 *Developing a data model.* The data model identifies the primary entities that are explored in the research, the attributes of each entity, and their relationships to each other. The data model contributes to a shared high level understanding of the research as well as serving as a blueprint for designing the metadata. (Instructor: Grace Agnew)
- 2 *The Metadata registry.* Codifying the data model into basic metadata. This ensures robustness for the metadata and is used to map the requirements of the data to the available RUCore metadata. (instructor: Grace Agnew)
- 3 *The RUCore Metadata schema.* A metadata schema is Structured metadata conforming to standardized practice. We will tour RUCore and understand concepts such as METS (metadata encoding and transmission standard) and XML (eXtensible markup language) and we will map the needed metadata to the schema. (Instructor: Grace Agnew)
- 4 *Metadata for research.* Now that you are familiar with RUCore, we'll look at two other research standards—DDI (Data Documentation Initiative) and Darwin Core to see how other researchers are applying metadata. (Instructor: Grace Agnew)
- 5 *Controlled vocabularies, ontologies, linked data.* We will look at tools and strategies for developing and codifying the unique information that populates metadata. (Instructors: Grace Agnew and Mary Beth Weber)

Completed 1,2,3 of this section to date.

# COURSE OUTLINE - DATA MANAGEMENT

- 1 *Principles of data preservation.* Canonical data formats. What formats is the researcher providing? What are the issues and concerns with the formats? How can the data best be preserved in RUcore? (Instructor: Isaiah Beard)
- 2 *Data reuse.* Terms of license, access control, IRB issues, public use version (Instructors: Janice Pilch and Ryan Womack)
- 3 *Lifecycle of the Data.* What is the context? How is the project structured, what are its outputs? When will they occur? Is this a grant project, with a documented timeline. What are the milestones that should be codified into events (Instructors: Ryan Womack and Ron Jantz)
- 4 *Use cases.* Does the metadata design support the user? Constructing a use case representing probable use of the resource. (instructors: Ryan Womack and Jane Otto)
- 5 *Designing a Research Portal.* How should the data be searched, retrieved and displayed for a specific community? (Instructors: Chad Mills and Kalaivani Ananthan)
- 6 *Workflow /Project Management.* What will it require for RUresearch to support this project? Assessment, Milestones and Deliverables. Who needs to participate? How will it be sustained over time? (Instructors: John Brennan and Rhonda Marker)

Completed 1,2, 3 of this section to date.

- Wanted to meet approximately monthly.
- Content was being created for the first time by busy librarians and staff, people were working other jobs (surprise!), so schedule was hard to stick to.
- Schedule had to be rearranged to accommodate the instructors.
- Homework was assigned, but some modules that would have benefited from more hands-on work did not have homework.
- The lack of ongoing data projects as a source of examples that the team could tackle was a problem.



- All but two course participants were interviewed prior to the presentation, according to an interview protocol approved by Rutgers IRB.
- Questions elicited reactions to each module, in addition to overall impressions of the course.
- In general, participants indicated that their understanding of each topic was increased by each course module.
- Exceptions occurred in areas where certain people had very extensive expertise. For example, some metadata librarians with many years of experience found the sessions on metadata elementary, but others said that they now saw the research data context for metadata.
- Copyright and licensing issues were formidable for many - so much to deal with, worry about. Concern about “managing risk as opposed to being risk-averse”
- The Data Model module was another exception.

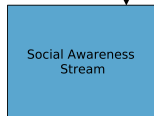
- The data model allows the major data stores and their connections to be visualized. This kind of planning identifies the critical features that the metadata and the structure of the data must support. The data is viewed as an integral part of the desired research process, not simply a final output.

**FIGURE:**

**membership**



- Name
- Location
- Demographics
- Social Networking Use



- Name
- Membership Requirement
- Type of Interactivity

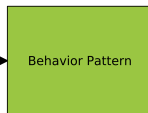
Draft Data Model for NSF Award  
105477. "Novel approaches for Reasoning  
About Local Communities from Social  
Awareness Streams Data." P.I. M. Namaan.

**occurring in**



- Collection Period
- Membership Duration

**resulting in**



- Type of Pattern
- Strength of Pattern
- Duration of Pattern

## THE DATA MODEL, CONTINUED

The Data Model component elicited the strongest reactions of anything in the course.

- Three of 13 interviewees did not learn from this part of the course (“the only part of the course that really bothered me”). Many had studied other versions of the “data model” that conflicted with the representation here.
- Seven of 13 felt that this was the best part of the course (including some who did not feel their understanding increased).
- Data model was the most challenging material, but most liked to be challenged.
- The group assignment tested understanding and brought together those of very different backgrounds. Their conversation and discussion was very valuable.
- For both those who learned and did not learn, there was concern about their ability to master the application of these skills, and how to know if they had been mastered.
- Perhaps not the best module to begin with!

Bringing together groups that rarely worked closely together was universally viewed as positive.

- Those from technical services felt their work would be better understood by others, and liked to hear public services perspectives.
- Those from public services were happy and surprised to learn of the technical capabilities of RUL, and were now able to make referrals.
- Connections made would save time and enable the “graduates” of the course to move more quickly to build new services and capacities (“a group of people that can speak a common language”)
- Several felt that this would prepare them to talk with faculty and others in the university, both about specific projects, and larger policy issues relating to data.
- “RUcore is not the silo that it once was.”

# COURSE STRUCTURE

In spite of busy schedules and other responsibilities, participants liked the work of being in a class.

- Wanted to meet more frequently (even though “finding the time” was one of the most challenging aspects of the course)
- Wanted more examples, more pre-class prep
- Wanted to review and reinforce concepts (“I need a lot of structure to do things”)
- Many liked doing homework best, wanted more (“more homework is a good thing”)
- Wanted a regular, fixed schedule with classes closer together - every two weeks, or even a “boot camp”
- Possibility of repeating classes, or sitting in again if course is offered in the future.
- Differing opinions on Friday afternoon meetings: Fridays were hard - already tired by that time of the week. Fridays were good - separation from workday routine and no to-do list in the head.

Several participants felt that the individual topics were somewhat detached from an overall framework that would provide greater meaning for why we were doing this and how it was to be applied.

Suggestions included:

- Frame the course around the interview questions for faculty with data. What do we need from them to successfully execute a data project?
- Frame the course around requirements for a trusted repository. What are we trying to build?
- Provide guidance for the implementation of topics. For example, why are we collecting/generating metadata? How do we determine the appropriate level of metadata for each project?
- How does the participant evaluate whether they have sufficiently mastered a skill? What is presented just for familiarity/background vs. core knowledge?

# FINAL REACTIONS

- The course was very new for most participants. Few knew what to expect going in (“I don’t think any of us took a course like this in library school”). Even though backgrounds varied, almost all felt sufficiently prepared or motivated to take the course.
- The course sensitized participants to the special issues of research data (“data is different”)
- “It changed the way I look at all of my work.”
- The course demanded work, but this was satisfying for participants (“You will get out of it what you put into it”)
- Almost no one liked the idea of a short version of the course. Taking the time is required to develop the knowledge.
- Some were surprised by the depth of knowledge that was present already in the libraries.
- “It was less dry than I thought it was going to be.”
- “I was expecting great things, and they happened!”



# CONCLUSIONS

In the future, several things are important to improve the learning experience in the course:

- Provide more structure and a tighter schedule
- Integrate more hands-on work and examples with the more abstract content
- Provide room for refreshers, review and discussion in the schedule
- Motivate the content with more explicit goals and framing
- Provide explicit expectations and guidance for student outcomes (self-evaluation?)
- The course will be repeated, primarily to expand the number of subject liaisons trained to work with data. Allow drop-in repeats of lessons as the course is offered again in the future.
- We hope to write up a book or a manual based on the course material.