

The Rutgers Experience: Building Data Management and Repository Services

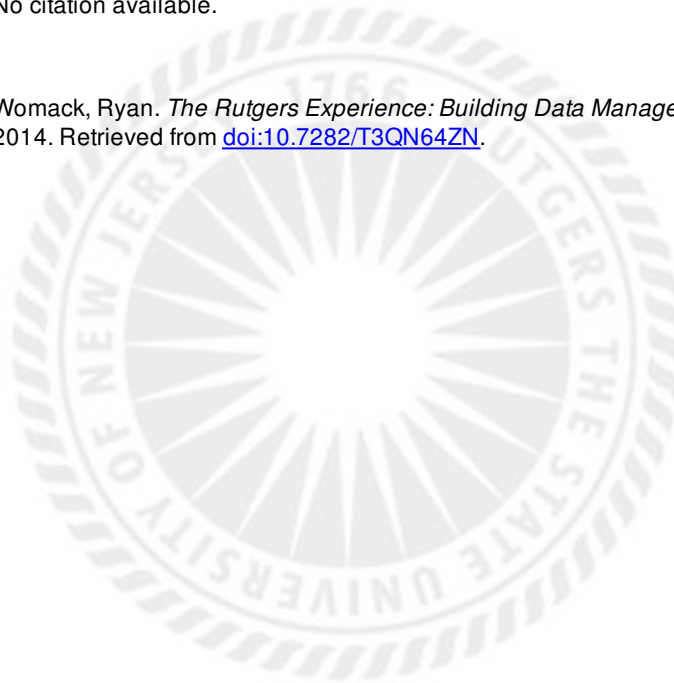
Rutgers University has made this article freely available. Please share how this access benefits you.
Your story matters. [\[https://rucore.libraries.rutgers.edu/rutgers-lib/43374/story/\]](https://rucore.libraries.rutgers.edu/rutgers-lib/43374/story/)

This work is an **UNIDENTIFIED VERSION OF A PUBLISHED WORK**

Citation to Publisher No citation available.

Version:

Citation to *this* Version: Womack, Ryan. *The Rutgers Experience: Building Data Management and Repository Services*, 2014. Retrieved from [doi:10.7282/T3QN64ZN](https://doi.org/10.7282/T3QN64ZN).



Terms of Use: Copyright for scholarly resources published in RUcore is retained by the copyright holder. By virtue of its appearance in this open access medium, you are free to use this resource, with proper attribution, in educational and other non-commercial settings. Other uses, such as reproduction or republication, may require the permission of the copyright holder.

Article begins on next page

THE RUTGERS EXPERIENCE

BUILDING DATA MANAGEMENT AND REPOSITORY SERVICES

Ryan Womack(rwomack@rutgers.edu)
Data Librarian, Rutgers University
Presentation to NN/LM MAR
Research Data Management Symposium

April 28, 2014



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#).

A BIT ABOUT ME

My background [more at <http://www.rci.rutgers.edu/~rwomack>]

- “Ordinary” subject librarian for business and economics to 2008, with major support for financial data resources
- Became Data Librarian in 2008
- Research Data an increasing component of that role from 2009-2013
- Chair of RUresearch Data Team since 2011
- Still responsible for public services side of data
- Bitten by statistics bug

$$\lim_{n \rightarrow \infty} G_n(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

[Central Limit Theorem]

Rutgers University Libraries has a large team that supports data, mostly indirectly

- RUresearch Data Team currently over 25 members, including metadata librarians, software developers, and subject librarians
- Grace Agnew, Associate University Librarian for Digital Library Systems
- Ron Jantz, Digital Library Architect
- Aletia Morgan, Research Data Manager
- Many highly skilled professionals, but only one dedicated exclusively to data

RUcore is the Rutgers University Community Repository
(<http://rucore.libraries.rutgers.edu>)

- Institutional Repository for Rutgers, operational since 2006, based on Fedora platform
- Has expanded from initial faculty documents deposit module to include modules for Video, New Jersey materials, and Electronic Theses and Dissertations.
- The goal of RUcore is to advance research and learning at Rutgers, to foster interdisciplinary collaboration, and to contribute to the development of new knowledge through the archiving, preservation, and presentation of digital resources.
- RUcore Open Source development
 - OpenETD - open source module for management of dissertations.
 - OpenMIC – open source moving image collections cataloging.
 - OpenWMS – open source Workflow Management System

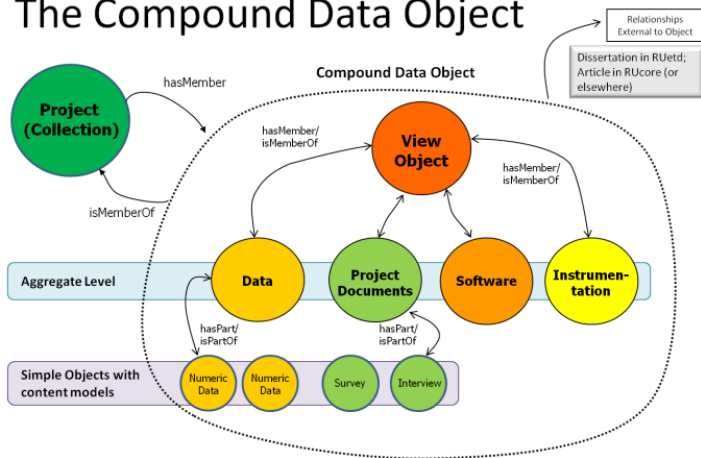
RU Data Working Group formed in Fall 2009

- Small team (5-7 people), chaired by Ron Jantz
- Operated for 2 years
- Designed initial prototype data portal
- Ingested several datasets into the repository
- Ingestion is mediated by librarians

RUCore is Fedora based.

- In other collections, each object in the repository has been treated as a single item in a top level collection
- Data is more complex
 - Need to associate multiple data files, multiple formats, and related documentation and software within a single object
 - Categories for Data, Documents, Software, Instrumentation
- For data, added relationship mapping, additional object types, and a revised interface design to handle more complex relationships
- METS/MODS basic metadata framework can be extended to include other metadata standards as necessary
- Metadata profile for research data has been developed to standardize terminology

The Compound Data Object



See “[Designing Data Services for the Institutional Repository](#)”, IASSIST Annual Conference, Cornell University, Ithaca, New York, June 2, 2010, with Ron Jantz, for more details.

- Major campus outreach in Spring and Fall 2011 via a series of presentations (over 100 attendees, mostly faculty, through various sessions)
- Most of our projects developed out of this outreach
- Data Management guide (<http://libguides.rutgers.edu/datamanagement>) developed as part of this [good for its time, but no longer state of the art]
- Faculty needs will be best met by providing working data storage alongside an easy path to data archiving

The Working Group, having designed the prototype portal, now named RUresearch, morphed into the **RUresearch Data Team** in July 2011 (<http://rucore.libraries.rutgers.edu/research/about/team.php>)

- Membership expanded to include more metadata librarians and subject specialists
- Internal training course developed for all team members - 12 sessions covering all aspects of the data lifecycle
- Details on the Training Course can be found in “[Data Management Training to Support Faculty Research Needs: Lessons Learned](#)”, IASSIST Annual Conference, Washington, DC, June 7, 2012

IMPORTANCE OF PARTNERSHIPS

Attended ARL E-Science in Fall 2011 [Angew, Morgan, and Womack]

- Because we were already doing outreach, had a repository, and were ready to accept data, did not follow through with building a strategic plan
- Because we already had a promising service, we did not form strong ties with Research Office, Computing, and Faculty Stakeholders beyond initial discussions
- Partly because of the above, never adopted DMPTool or other systematic promotional opportunities
- No governing or advising body outside of the Libraries
- **MISTAKE!** [we'll return to this later]

We have done workshops and presentations on DMP writing, along with general data handling advice

- Some open workshops, some by request with departments
- Continue to receive individual requests for advice, which we handle
- This activity is not distinctive or large scale in comparison to other institutions
- Eventually hope to have more subject librarian involvement

SOME PROJECTS

Much of 2012 was spent working on projects recruited during the initial intensive outreach. Outreach continued in 2012, but at a slower rate. Currently have 9 ingested data projects. Here will review 3 projects that are

- Relatively complete
- Make available unique data that would otherwise be unavailable
- Represent unique metadata and technical work

Research conducted by Vorsi, Polashock, and Georgi at the Philip E. Marucci Center for Blueberry and Cranberry Research and Extension of Rutgers

- Data is genomic sequencing of a unique cranberry variety
- 40 GB in 4 files, requiring special tweaking of Fedora installation and our processes to handle large files [Fedora 4 has taken care of this]
- Learning about gene sequencing formats and instrumentation
- see <https://rucore.libraries.rutgers.edu/research/cranberry/> and <http://hdl.rutgers.edu/1782.1/rucore00000002131.Dataset.m000003398>

Rob Scott, Dept. of Anthropology at Rutgers, studies dental wear on various fossil primate teeth

- Tooth scans on 22 species from the article, Scott, R. S., Teaford, M. F. and Ungar, P. S. (2012), Dental microwear texture and anthropoid diets. *Am. J. Phys. Anthropol.*, 147: 551–579. doi: 10.1002/ajpa.22007
- RUresearch record links to both article and grant award (see <http://dx.doi.org/doi:10.7282/T3CJ8BF1>)
- Each species has 60-150 individual scans for various teeth at different angles
- Process involved understanding the proprietary format and structure of the data

Carolyn Moehling and Anne Piehl, Dept. of Economics at Rutgers, in the article “Immigrant assimilation into US prisons, 1900–1930” [Journal of Population Economics, January 2014, Volume 27, Issue 1, pp 173-200]

- Used Census records to study characteristics of prison inmates in relation to their immigration status
- Codebook prepared by the researchers specifically for this data release
- Privacy concerns led to redaction of names
- (see <http://dx.doi.org/doi:10.7282/T3TH8JPB>)

Technical development has continued, with notable and innovative features (thanks to our developers!)

- Portal Redesign (<http://rucore.libraries.rutgers.edu/research>)
- DOIs
- Structure Map Ingest to capture an entire directory and its relationships
- Some lessons, or maybe just one lesson:
 - ① Technical development is hard and takes time
 - ② Do not promise more than you can deliver (due to #1)
 - ③ However long you think it will take, it will take longer (#1 again)

Integration of librarians from UMDNJ into Rutgers and interest from other subject librarians has led to:

- Expansion of the Data Team to over 25 members
- A new round of training (going on right now)
- New potential for outreach to departments, especially in the biomedical sciences

GETTING TO “PRODUCTION”

In 2013, efforts were made to formalize the data deposit services offered:

- To make the service less “boutique” and more “production”, something that all faculty could rely on
- To define the parameters of the kinds of data we would promise to accept
- To specify necessary copyright and licensing arrangements

HOW SOON IS NOW?

“...See, I’ve already waited too long...”

- A very slow process
- Still struggling to define extent of services
- No external interest groups to push for a solution
- Lack of long-term goals and planning and lack of external involvement = lack of “deep roots” to overcome difficulties