

Edward M. Corrado | [ecorrado@binghamton.edu](mailto:ecorrado@binghamton.edu)  
Binghamton University Libraries | <http://library.binghamton.edu/>

# Preservation and Research Data at Binghamton University Libraries

Presented 14 November 2011 at  
Research Data: Management, Access  
and Control Symposium,  
University at Buffalo



# Binghamton University (SUNY)

*Binghamton University, one of four comprehensive doctoral research universities within the State University of New York, is recognized for stellar academics, an international focus, high graduation rates and overall value*

- Undergraduates: **11,706**
- Graduate students: **3,007**
- Average SAT score for 2011 incoming Freshman: **1305**
- Top 25% of high school class: **85%**
- Students of color: **33.3%**
- International students: **10%**
- **#1 in 2011 as a best value among the nation's public colleges for out-of-state students and #5 overall**  
(Kiplinger's Personal Finance, 2011)
- **Students come from all 50 states and 100 countries**




Binghamton University:  
*"The premier public university in the northeast" and "best buy"*  
(Fiske Guide To Colleges, 2010)

# Collections profile



**2,409,043**  
Volumes  
(print, gov. documents,  
and electronic books)

**93,414**  
Print & e-Journals,  
Serials & Orders



*A Map of the Night*  
Jordan Smith. The Antioch Review. Yellow  
Abstract | Full text | Full Text





**1,869,980**  
Microforms &  
Microfiche

**118,948**  
Sound Recordings  
(phonodiscs, CDs, audio  
cassettes, DVD audio)



**2,951**  
CD-ROMs



**120,959**  
Maps

**3,751**  
DVDs &  
Videocassettes



**234**  
Electronic  
Databases



			Pitoh	sf
it. p.	Time	Time	ave	ft.
ntue	100 feet	100 feet	100 feet	per
	etc	etc		Time

**1,077 linear feet**  
Archives & Manuscripts

# Our facilities

## Three Libraries

**Bartle**

**Science**

**UDC**



**All three contain  
Information Commons  
& circulate wireless  
laptops/netbooks.**

**Bartle is open 24x6  
noon Sunday  
till midnight  
Friday;  
Saturday noon  
till midnight**



**Library Annex@Conklin**  
**Contains 350,000 volumes of infrequently  
used research materials**

**Annex Reading  
Room Hours:  
Monday - Friday  
9AM - 4PM**



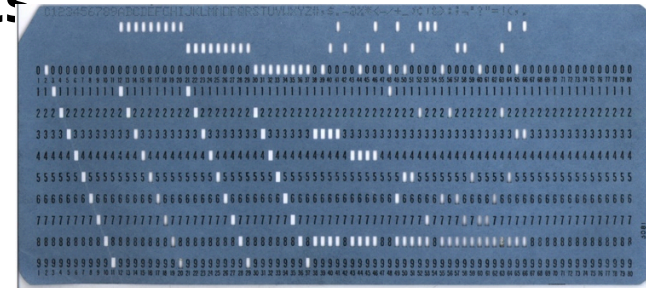
ha's 44 chapters in the United  
is support partners and commu-  
s to provide education to these  
while attempting to instill  
reness about related matters like  
aliness, health and savings.



**Digitizes articles  
on demand and  
sends to faculty  
via email**

# Digital preservation: What isn't digital preservation?

- Backups ≠ Preservation
  - “Actions required to maintain access to digital materials **beyond the limits of media failure** or technological change.” (Digital Preservation Collation, 2009)
  - Backups alone are not sufficient
  - Don't protect against obsolete file formats, software, hardware, etc.
- Providing access ≠ Preservation
  - Digital Asset management systems offer access but not [necessarily] long term preservation



# Digital preservation should not be confused Open Access/Open Data

- While Digital preservation can support Open access and/or Open data, preservation does not and can not always imply Openness
  - Patents and other **legal** issues
  - **Confidential data** such as Blood Serum Collection
  - Researcher/Discipline **Norms**
  - Discipline Specific Repositories such as **arXiv.org** and **Inter-University Consortium for Political and Social Research (ICPSR)**

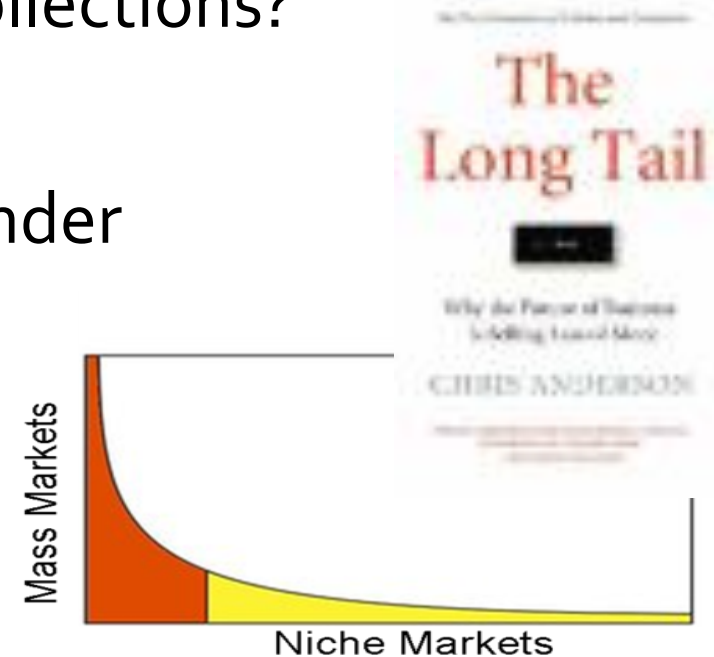


# Digital preservation: What does it involve?

- JISC Beginner's Guide to Digital Preservation elaborates:
  - **Managed:** Digital preservation is a Management problem.
  - **Activities:** The policy needs to filter down to a list of processes: tasks that can take place at specified times and in specified ways.
  - **Necessary:** What needs to be done. How long do you want to preserve the objects for? Discussions about the activities needed to achieve a level of preservation are necessary.
  - **Continued Access:** Access is the key here. Most objects in the public sphere are preserved to support access and retrieval.
  - **Digital Materials:** Digital materials, digital objects, call them what you will. This is the stuff you are preserving. Different objects require different processes.

# Why digital preservation?

- Local Content as the Future of [Academic] Libraries?
  - At least in regards to Physical Collections?
  - Google Books, HathiTrust
  - To a large degree the material under the “Bell Curve” (journals, gov’t docs, etc.) is already being “managed” outside of libraries
- The University is a collection of **Niche Markets** (John Meador, Jr.)





# Why digital preservation? (part 2)

- Why Libraries?
  - Libraries have been preserving information for centuries
  - Furthers **the role of libraries** to the digital world
  - Not a new idea, a **new format**
  - Majority of new material is published in digital format (Scholarly Articles, Campus newsletters, Course catalogs, Web sites...)



# Why digital preservation? (part 3)

*“Digitization for Access is Not Preservation;  
Without Preservation, There is No Access”*



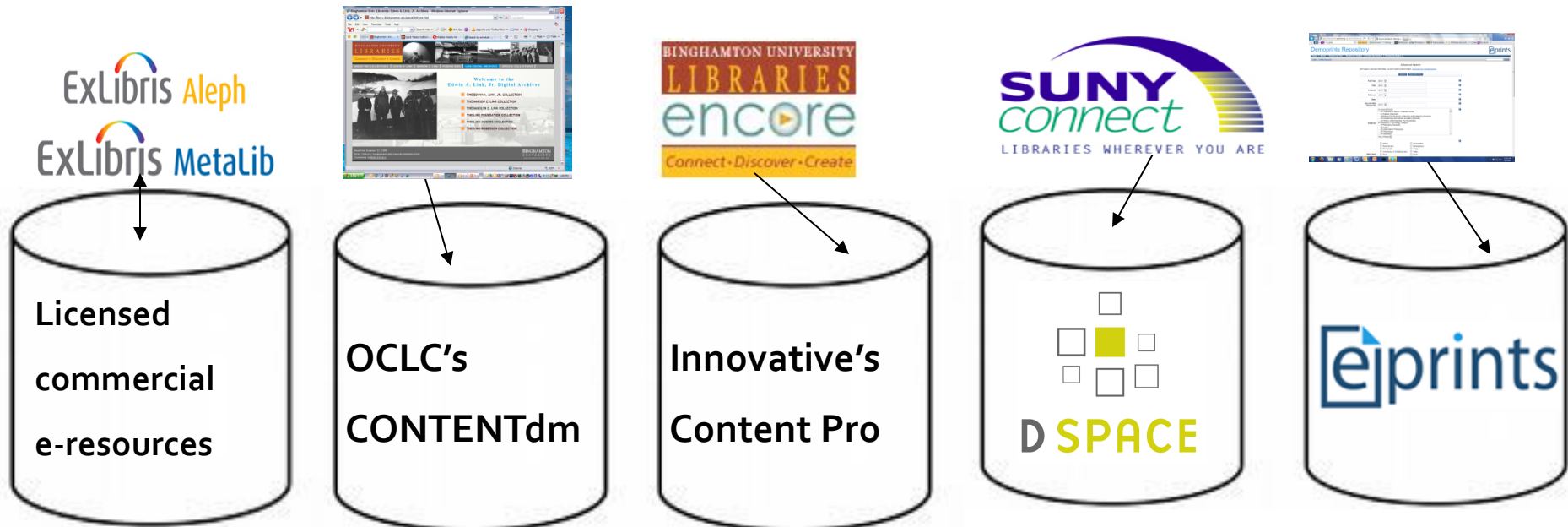
**Sustainable Economics for a Digital Planet:  
Ensuring Long-Term Access to Digital Information**

*Final Report of the Blue Ribbon Task Force on  
Sustainable Digital Preservation and Access*

[Click to download](#)

The image shows a promotional graphic for a report. On the left is a small thumbnail of the report cover, which features a blue and green abstract design with a globe and binary code. To the right of the thumbnail, the title and subtitle of the report are displayed in white text against a dark blue background with binary code and a large, glowing green circular graphic. A yellow 'Click to download' link is positioned below the subtitle.

# Results of our initial approach to digitization: Unlinked digital collections & interfaces with no preservation



# What we Learned; The need to:

- Adhere to International Standards
  - "Librarians can take over the world." (Dr. Barry Smith)  
But we need to use tools that have been proven - not building new ontologies
- Capture the locally born digital objects that are replacing titles formerly found in our print archives
- Ensure Digital Curation & Preservation
- Provide Cross-Collection Search
- Demonstrate proof of concept before soliciting faculty research

# Digital preservation: Binghamton's solution

- Experimented with various “Digital Content” systems including Content Pro, CONTENTdm, DSpace, EPrints
- None of these have preservation “built-in”
- Building our own was not practical
  - Staffing levels
  - Lack of programmers
  - Mission creep?
  - Sustainability?
- Rosetta by Ex Libris

Rosetta

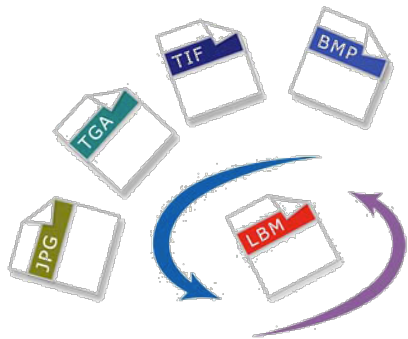
# Rosetta

- Scalable
- Expandable
- Flexible
- Accessable
- Standards-based
  - “Based on the Open Archival Information System (OAIS) model and conforming to trusted digital repository (TDR) requirements.”

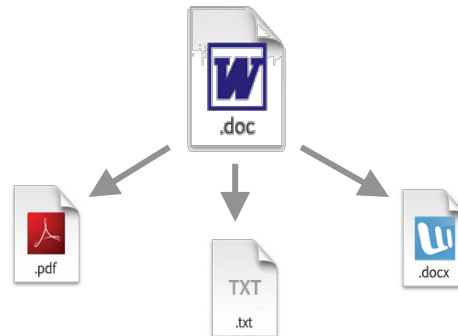
Rosetta

# Rosetta: A digital preservation solution

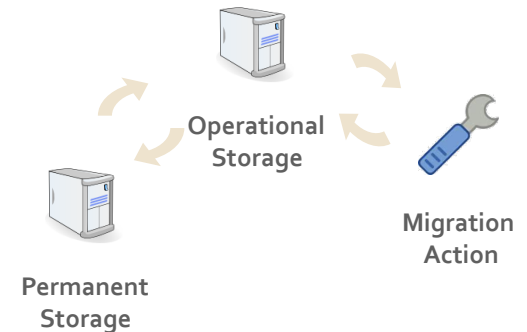
Complete preservation solution allowing collection, archiving and preservation of digital materials of any type. Rosetta ensures data integrity and provides access over-time to digital materials.



Identify Risks



Evaluate Alternatives



Execute Preservation  
Actions

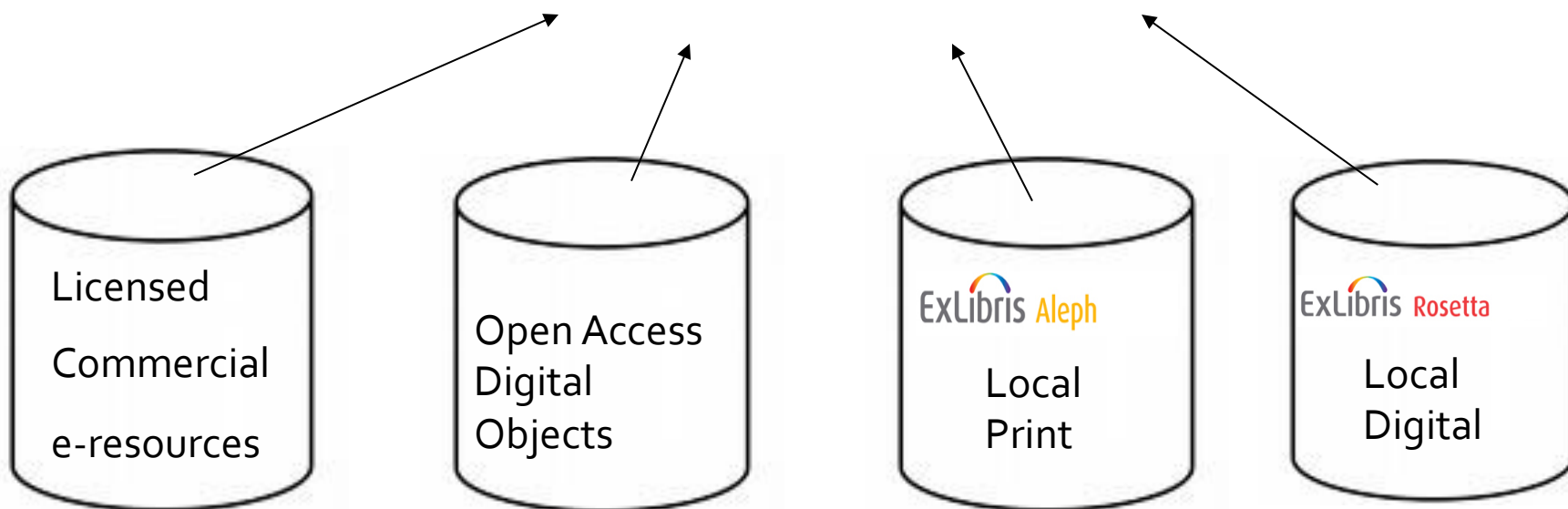
# Rosetta and discovery

- No preservation systems is useful if there is no access (especially for a University Library)
- Rosetta does not have a public discovery layer
- Rosetta's Digital Publishing System is flexible so there are options
- Primo for discovery
  - First University to use Primo with Rosetta
  - Works well with other library systems such as Aleph and Primo Central
  - One stop shopping

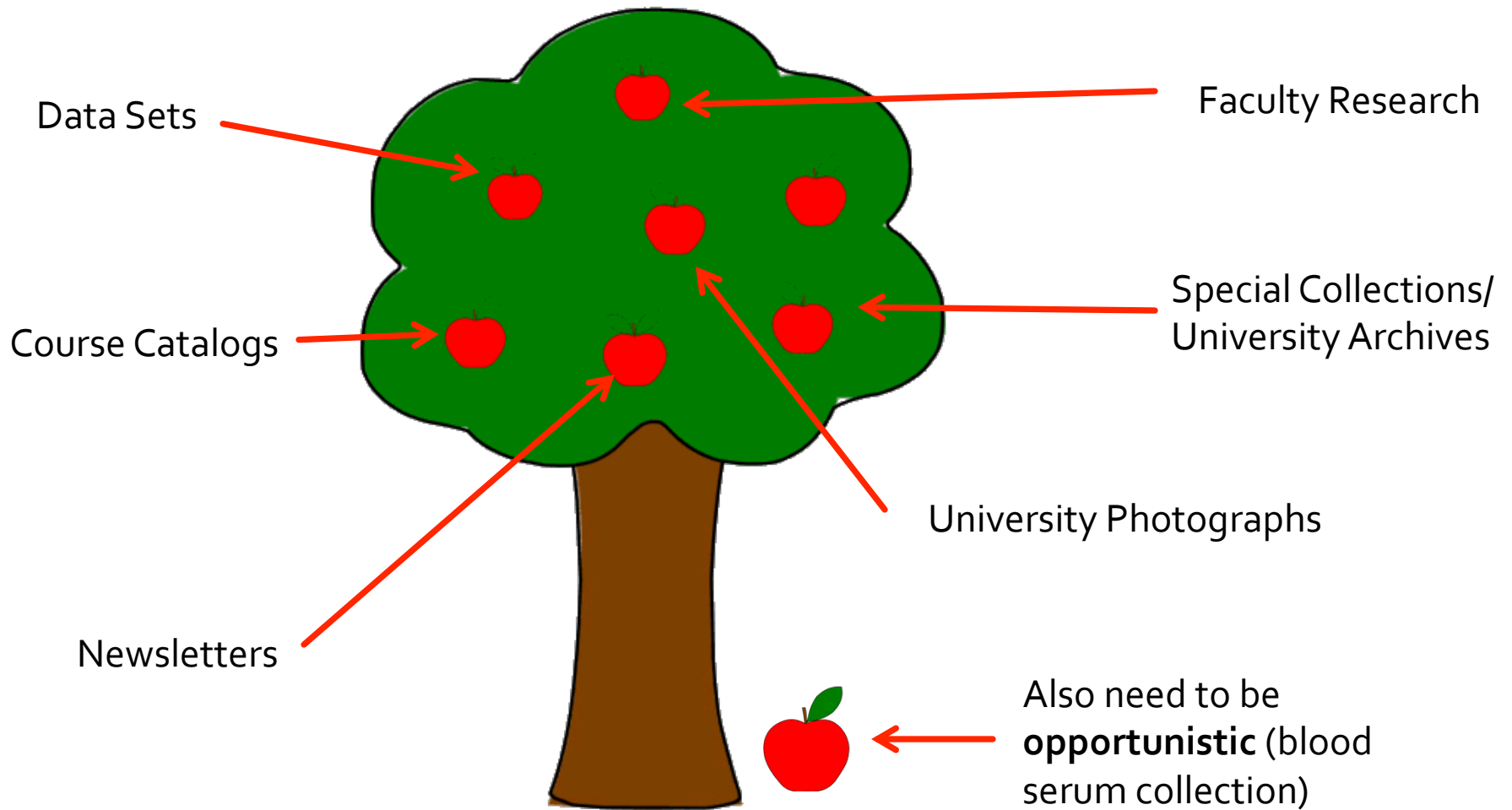


# Foundation of our future digital library: Unified access to digital & print assets with local preservation

ExLibris Primo



# Digital content at Binghamton University



# At-risk materials



Image cc-by-nc-sa 2.0: <http://www.flickr.com/photos/yakibah/3512735385/>

# Overall staffing

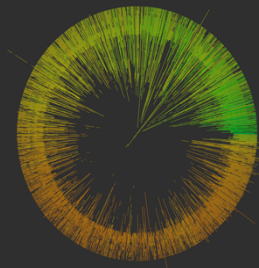
- Systems: 1 person (~0.5 FTE)
  - Project Management
  - Systems/Technical
- Metadata/Cataloging: 3 people (~1.0 FTE)
- User Interface: Part of Web Services Librarian's Time
- Special Collections: Not directly involved with implementation, but relied on heavily for collection level expertise

# Metadata/cataloging staffing

- Metadata Librarians are Project Managers
  - Decide on appropriate descriptive metadata fields
  - Create the metadata forms
  - Provide training
  - Develop and/or provide specialized terminology (such as LCSH, TGM, TGN)
  - Review submissions as appropriate
  - DO NOT typically create the metadata (student workers or other staff will create metadata)



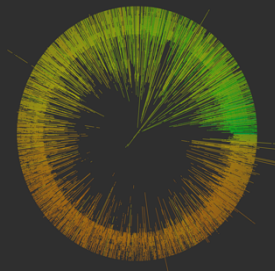
# Preserving research



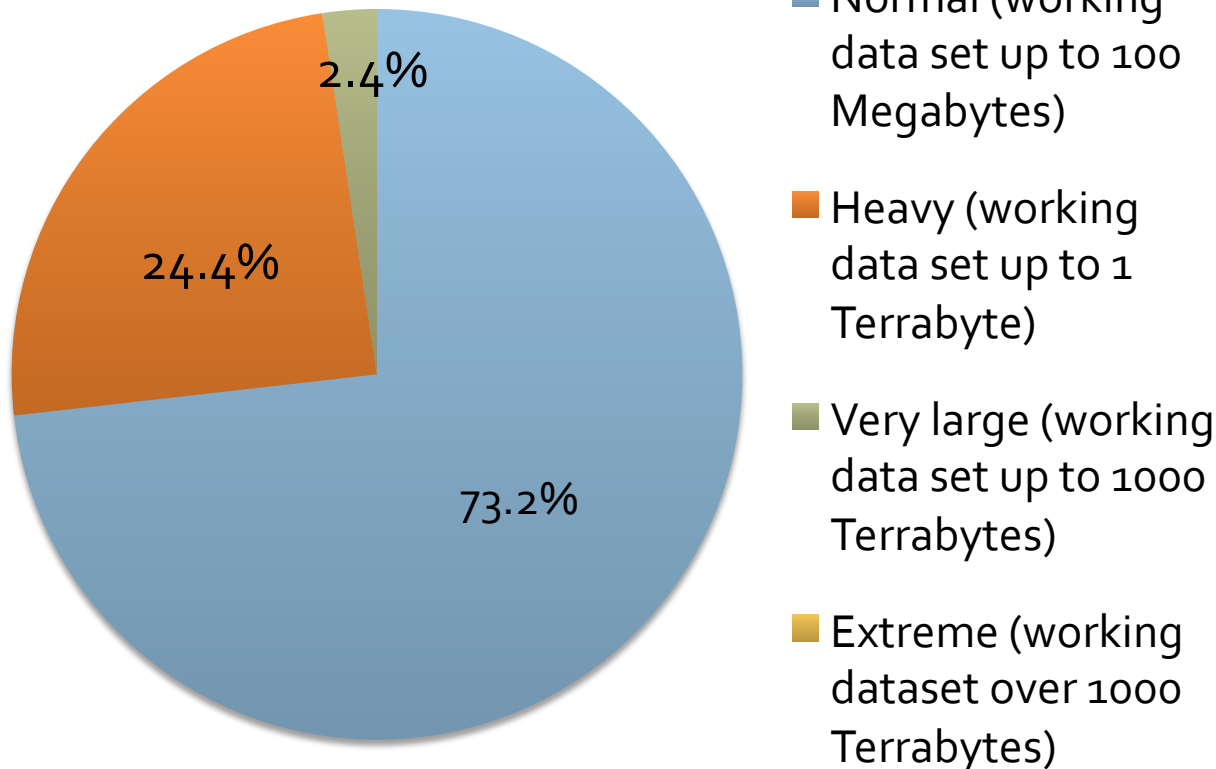
- In the preliminary planning stages
- Need to demonstrate we can do what we say
- Scholarly output
  - Articles, proceedings, etc.
  - Research data
  - Related material including grey literature, research notes, correspondence, etc.



# Survey of research faculty at Binghamton University

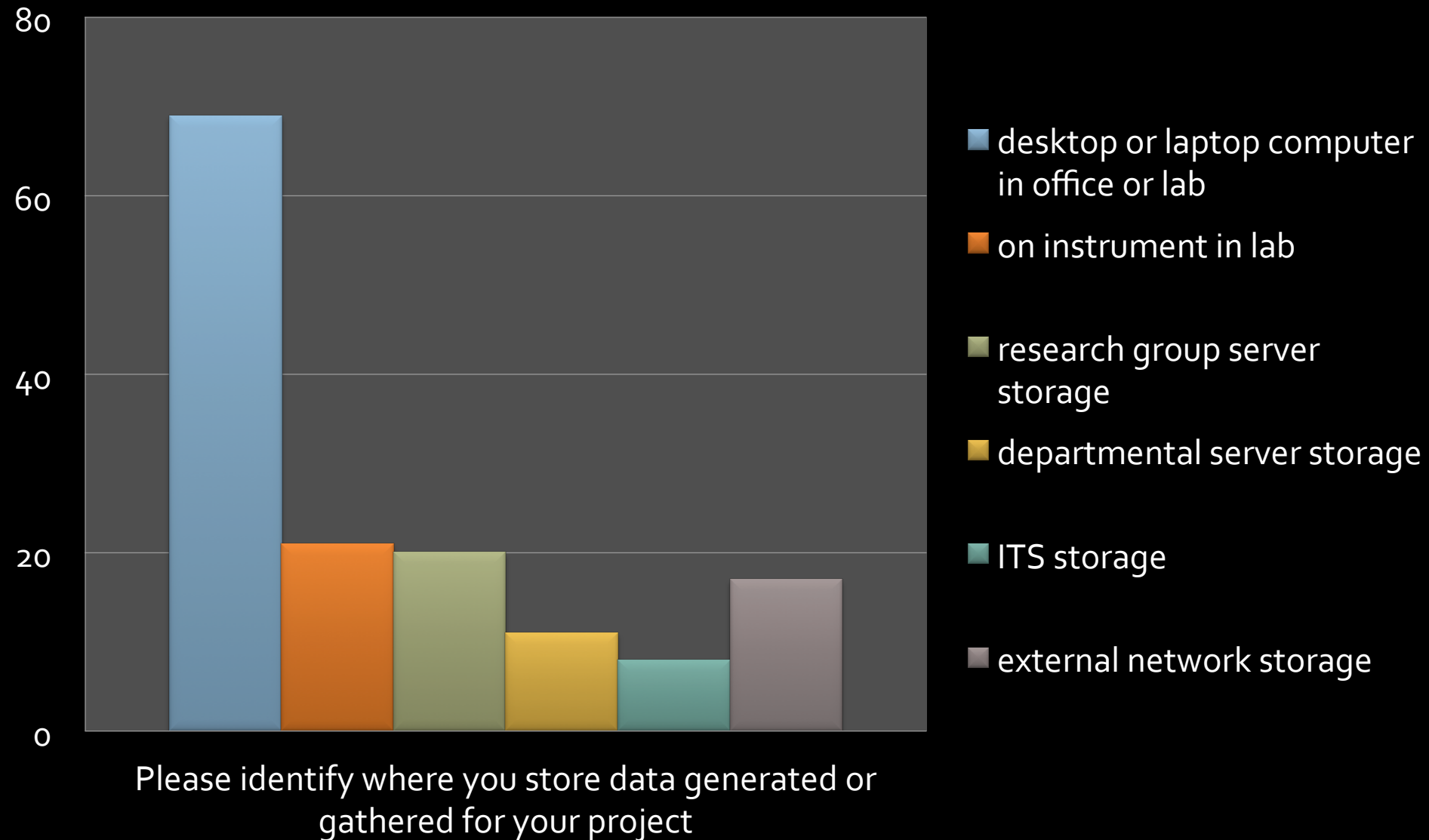
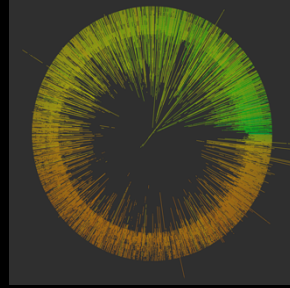


Please characterize your research in terms of data intensity for your analysis run (n=91)



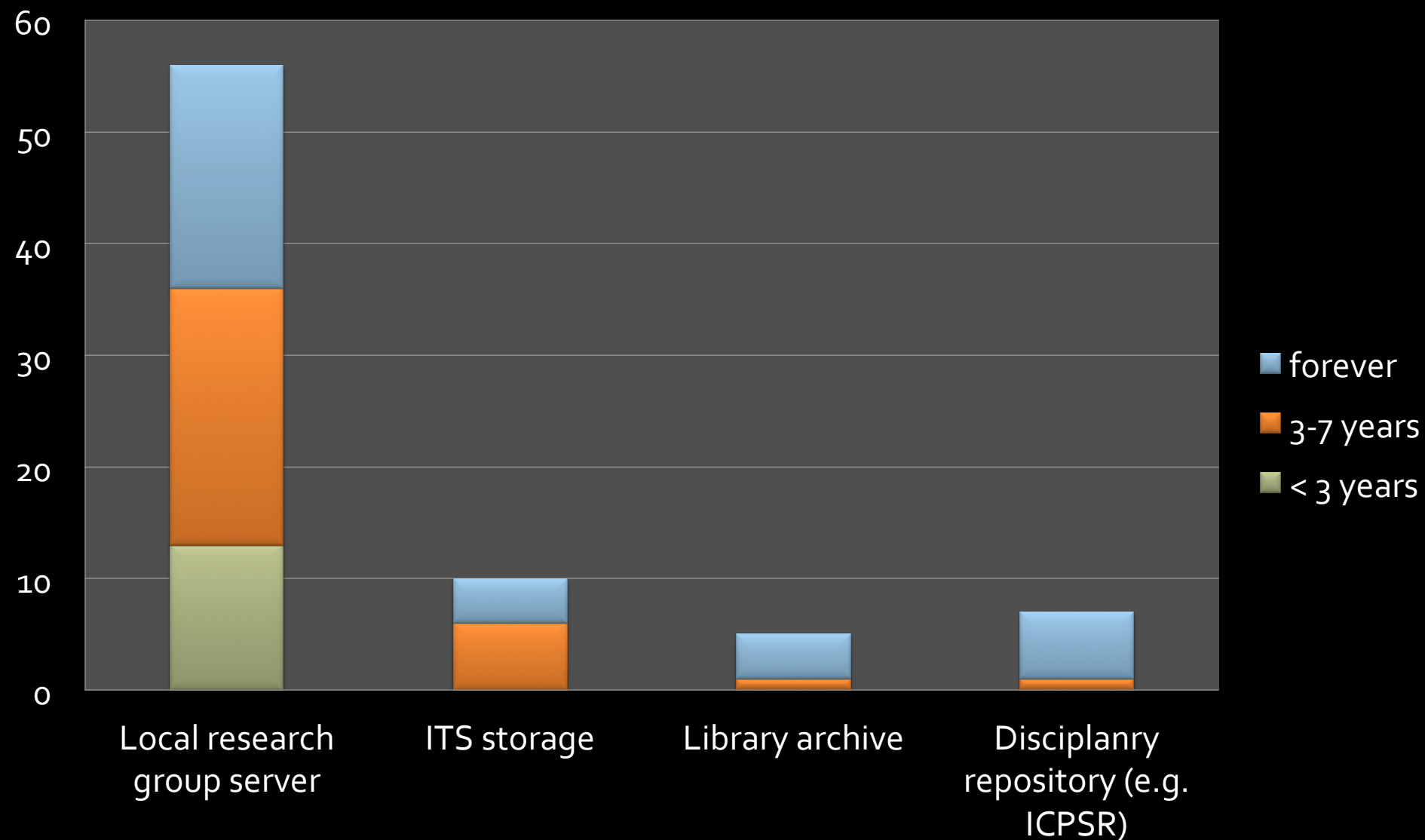
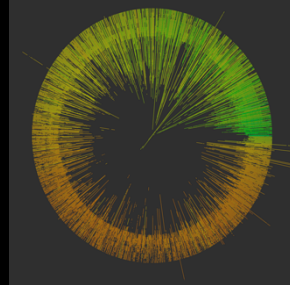
308 individuals who either had an externally sponsored project since 2009 or who had submitted a proposal during that time period where asked to take the survey. By June 15, 2011 91 respondents complete the survey. (Conducted by Jim Wolf, retired Director of Academic Computing)

# Please identify where you store data generated or gathered for your project

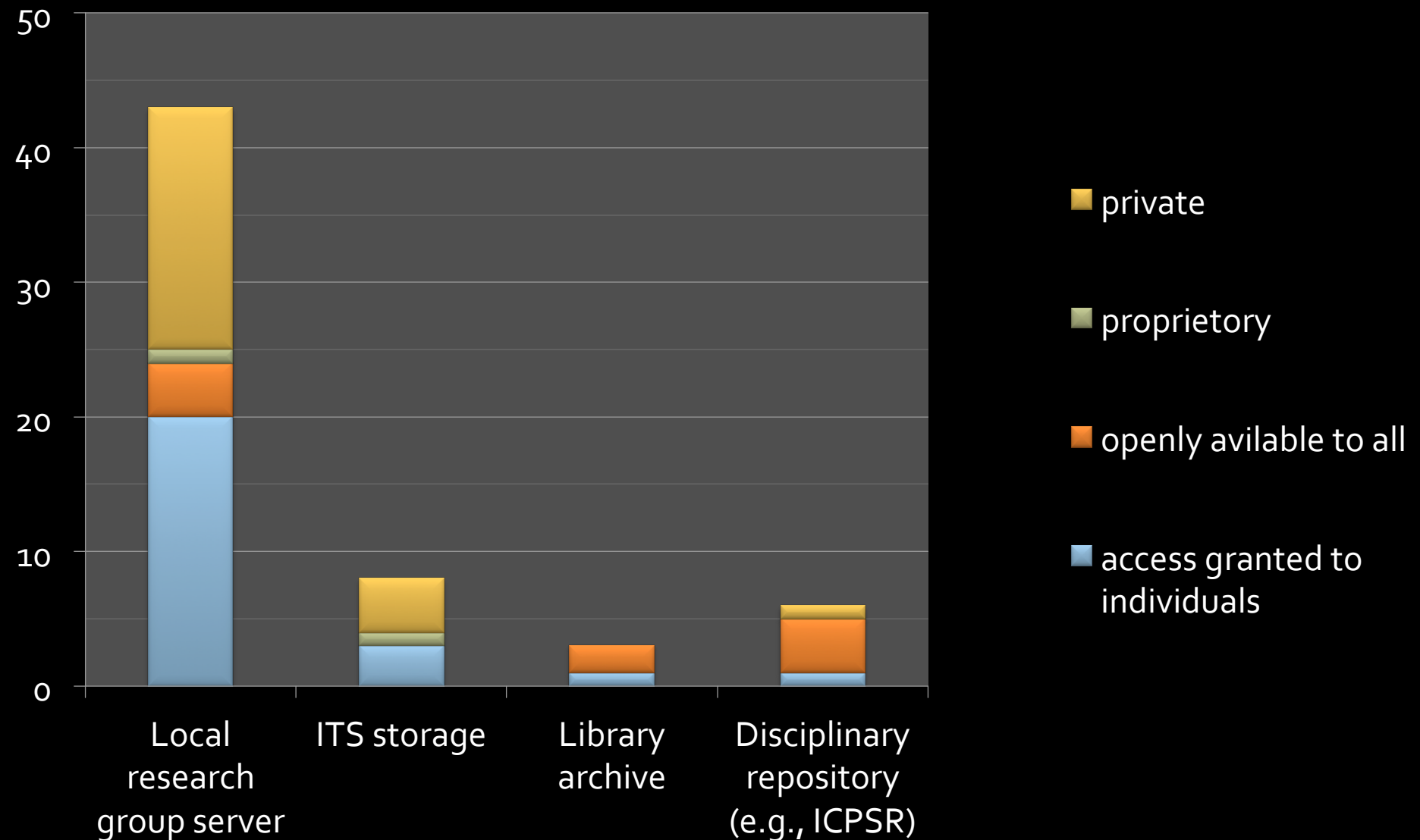
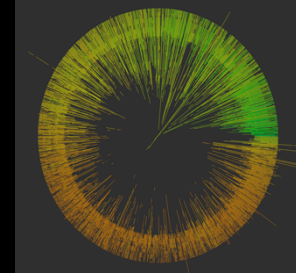




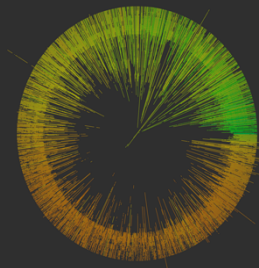
# Desired preservation timeframe



# Desired types of access



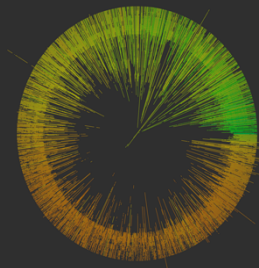
# Blood serum archive: A partnership



- Library provides:
  - Digital platform (Rosetta)
  - Metadata consulting
  - Metadata training
  - Ongoing preservation
- Blood serum archive provides:
  - Subject expertise
  - Digitization
  - Metadata creation

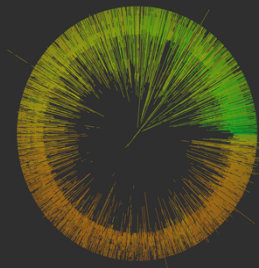


# Blood serum archive: Lessons learned



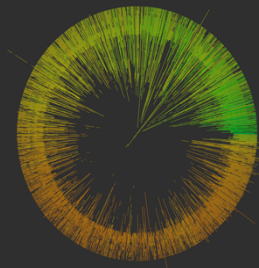
- Bring everyone on board
- Set priorities
- Review metadata and digital objects often (at least in the beginning)
- Metadata may contain confidential and/or legally protected information
  - Will metadata librarians need human subjects/IRB approval?
  - Need for separate discovery mechanisms

# Research data: Some thoughts



- Enlist subject librarians to help make connections
  - A few subject librarians have identified some possible data needing preservation and are going to meet with faculty for preliminary discussions
- Work with faculty on data management plans
  - Many granting agencies such as NSF are requiring data management plans
  - Get involved early
  - Assist with submission requirements for research

# Research data: Some more thoughts



- Provide preservation; offer dissemination
  - Don't confuse preservation with open access
  - Faculty don't always want or can not make data open
    - Dark archive if desired
  - Do not need to replace or replicate current data dissemination methods (unless researchers desire)
- Not all research data is "Big Data"
  - Don't let the challenges of "Big Data" scare you away from all data.

I would like to thank Sandy Card, Rachel Jaffe, Ashley Marcin, John Meador, Jr., & Molly O'Brien for their contributions to this presentation.

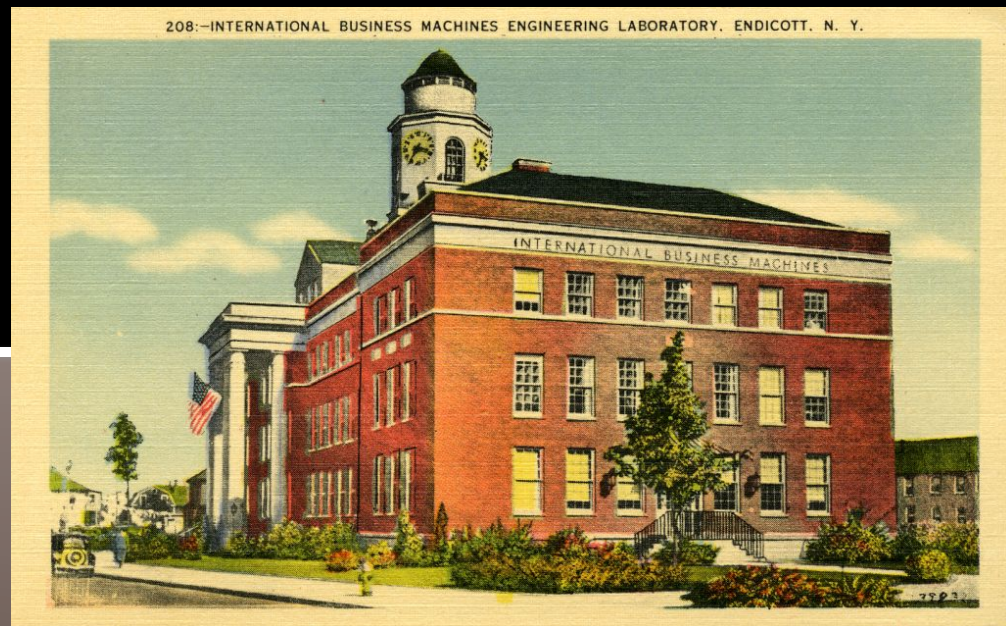
Presenter & Contact Info:

Edward M. Corrado

[ecorrado@binghamton.edu](mailto:ecorrado@binghamton.edu)

@ecorrado

<http://library.binghamton.edu>



Reference:

Digital Preservation Collation (2009). Digital Preservation Handbook. <http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts>