

Rethinking Repositories

*CNI Spring
Meeting*

*April 3
2017*

Jason Casden
*Head, Software
Development*

Julie Rudder
*Repository Program
Librarian*

Paolo Mangiafico
*Coordinator, Scholarly
Communications
Technology*

Will Sexton
*Head, Digital Curation and
Production*

UNC University Libraries

Duke University
Libraries

Clifford A. Lynch, “Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age,” ARL Bimonthly Report 226 (February 2003), 1-7. ([PDF](#))

“In the fall of 2002, something extraordinary occurred in the continuing networked information revolution.... The development of institutional repositories emerged as a new strategy....”

What goes into the
repository?

What goes into the repository? 1/2

Lynch, 2003:

- “intellectual works of faculty and students - both teaching and research materials”
- “documentation of the activities of the institution itself in the form of records of events and performance(s) and of the ongoing intellectual life of the institution”
- “experimental and observational data captured by members of the institution that support their scholarly activities”

What goes into the repository? 2/2

DDR Mission Statement:

“...research data, scholarly output, digital collections, archival records, and other digital materials that provide enduring value for intellectual inquiry and documentation of University activities.”

Library Collections

Digital Archives
ETDs, e-records,
born-digital, patron
requests

Digital Collections
New & legacy, access
UI, AV, images, MSI

**Gillispie,
UA**

Bragg, DDC

Open Access
Faculty pubs,
DukeSpace, VIVO/
Scholars@Duke

**Mangiafico,
ScholComm**

**Herndon,
DVS**

Campus Services

Faculty Research Data
Provost's mandate,
data management
planning, compliance

DDR Program Committee

Maggie Dickson
Metadata Architect

Jim Tuttle
Repository Architect

**Metadata
Advisory
Group**

Will Sexton
DCAPS
Chair

Ginny Boyer
SDIS

**Preservation
Advisory
Group**



Our Mission

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

[Learn More](#)
[Contact Us](#)

The Duke Digital Repository has three program areas. Follow the links below for more information.

Research Data

Data management planning and curation for research at Duke

Learn more about our services, and access data from Duke researchers

Scholarly Publications

Open access for Duke authors

Learn more about Open Access at Duke, and find scholarship by Duke authors

Library Collections

Digital items owned or stewarded by Duke University Libraries

Learn more about our collections and programs

DDR Program Areas

- Research Data
- Scholarly Publications
- Library Collections

Each program area has its own collections policy and working group



OUR TEAM

The Duke Digital Repository brings together the expertise and effort of staff from many different departments in Duke University Libraries.

Oversight and policy for the DDR are provided by the [DDR Program Committee](#).

CONNECT WITH US

[Bitstreams Blog](#)
[Data and Visualization Services Blog](#)
[Scholarly Communications @ Duke](#)

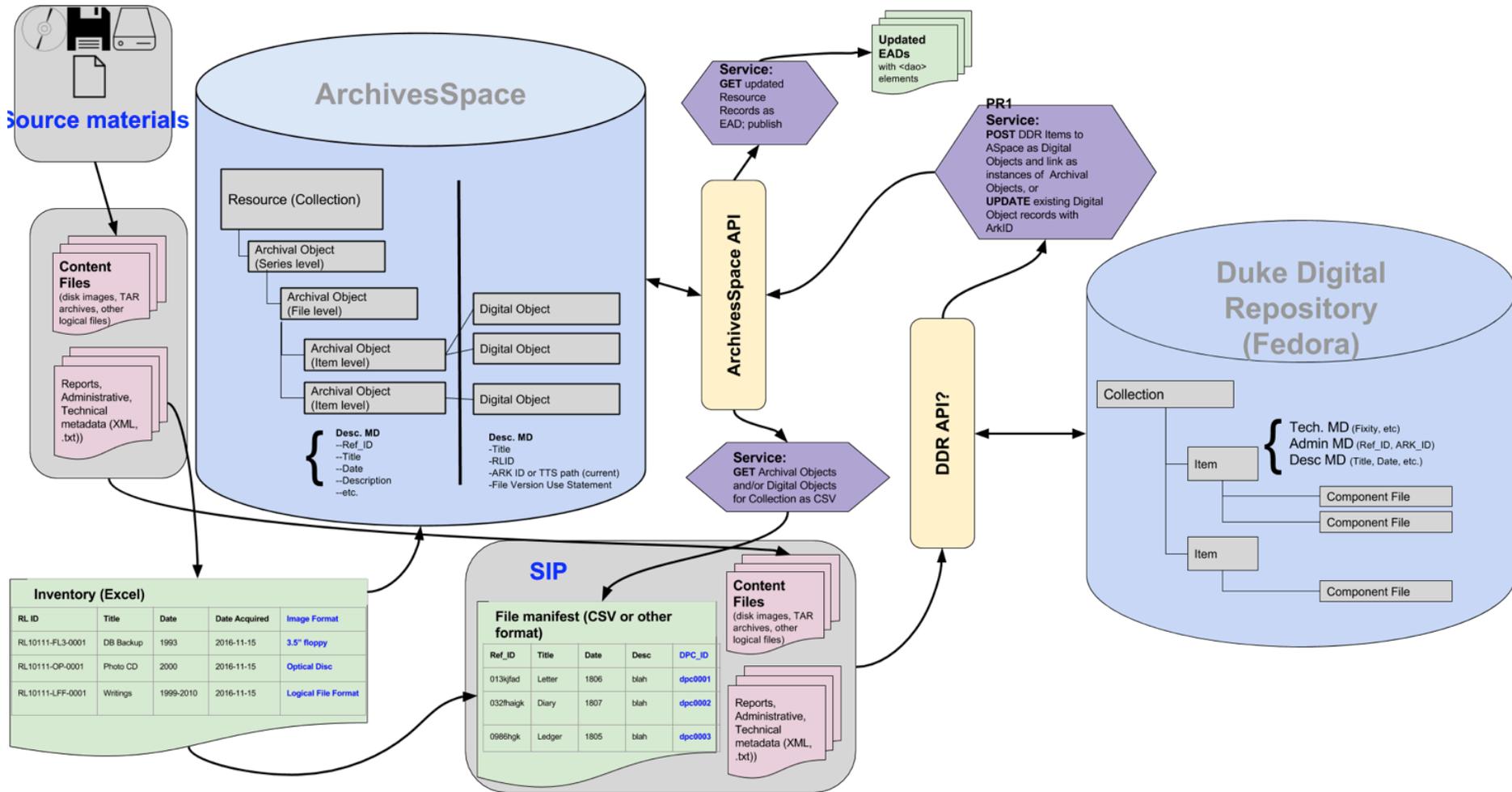
RESOURCES

[About the DDR](#)
[Public Documentation for the DDR](#)

Duke Digital Repository

The DDR is an umbrella service providing preservation, management, access, and discovery for five collecting or depositing areas that currently have staff with dedicated time and established (or developing) processes:

DDR Program Area	DUL Collecting/Depositing Area	Departmental Owner(s)
Research Data	Duke Faculty Research Data	Data & Visualization Services, Digital Curation & Production
Scholarly Publications	Electronic Theses & Dissertations	University Archives
	Open Access Publications	Scholarly Communications
Library Collections	Digital Collections	Digital Curation & Production
	RL Born-Digital	Rubenstein Library



Duke Elements

1. New (or old) publication added to Elements
 - Author (or delegate) confirms publication info harvested from bibliographic source *OR*
 - Author (or delegate) enters publication information
2. Author (or delegate) reviews publisher policy and Duke recommendations regarding open access to full text
3. Author (or delegate) uploads full text PDF

[The Trim39 ubiquitin ligase inhibits APC/CCdh1-mediated degradation of](#)
[get it @Duke](#) [doi](#)

Journal article

Summary Duke users (1) History (0) Data sources (6) Full text (1)

Huang NJ, Zhang L, Tang W, Chen C, Yang CS, Kornbluth S
 J Cell Biol 197(3):361-367 30 Apr 2012 [Author URL](#) [DOI](#) [Repository public URL](#)

4. PDF of full text sent to DukeSpace for open access.

DukeSpace

DukeSpace Home Duke Scholarly Works Scholarly Articles View Item

The Trim39 ubiquitin ligase inhibits APC/C-Cdh1-mediated degradation of the Bax activator MOAP-1

Huang, N-J ; Zhang, L ; Tang, W ; Chen, C ; Yang, C-S ; Kornbluth, S
2012-04-30

Download Full Text

Citation:
JOURNAL OF CELL BIOLOGY, 2012, 197 (3), pp. 361 - 367

Citable Link to This Page: <http://hdl.handle.net/10161/6382>

5. Permanent link generated by DukeSpace goes back to Elements (handle.net link shown as "Repository public URL" in Elements).

7. Open Access link and other publication information is indexed and findable via various external sources.

6. Publication links sent by Elements to Scholars@Duke. Repository link appears there as "Open Access Copy" button and DOI appears as "Full Text" button.

SCHOLARS@DUKE

Huang, NJ; Zhang, L; Tang, W; Chen, C; Yang, CS; Kornbluth, S. [The Trim39 ubiquitin ligase inhibits APC/CCdh1-mediated degradation of the Bax activator MOAP-1](#). *The Journal of Cell Biology*. 197:361-367. 2012 [Full Text](#) [Open Access Copy](#) (Journal Article, Research Support, N.I.H., Extramural, Research Support, Non-U.S. Gov't)

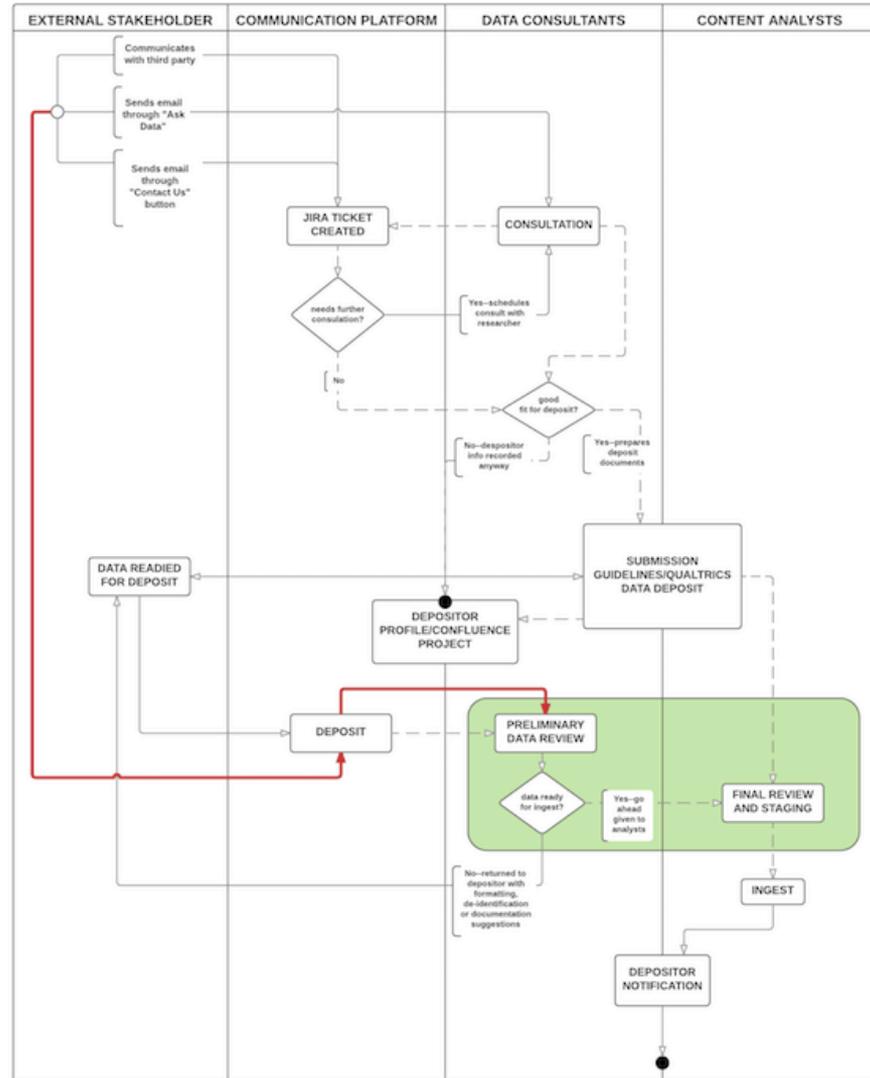
Google Scholar

Google

Duke web sites

VIVO network

More information about this process can be found in the Elements help documentation and more specifically, at <https://scholarworks.duke.edu/elements/help/uploading-full-text/>



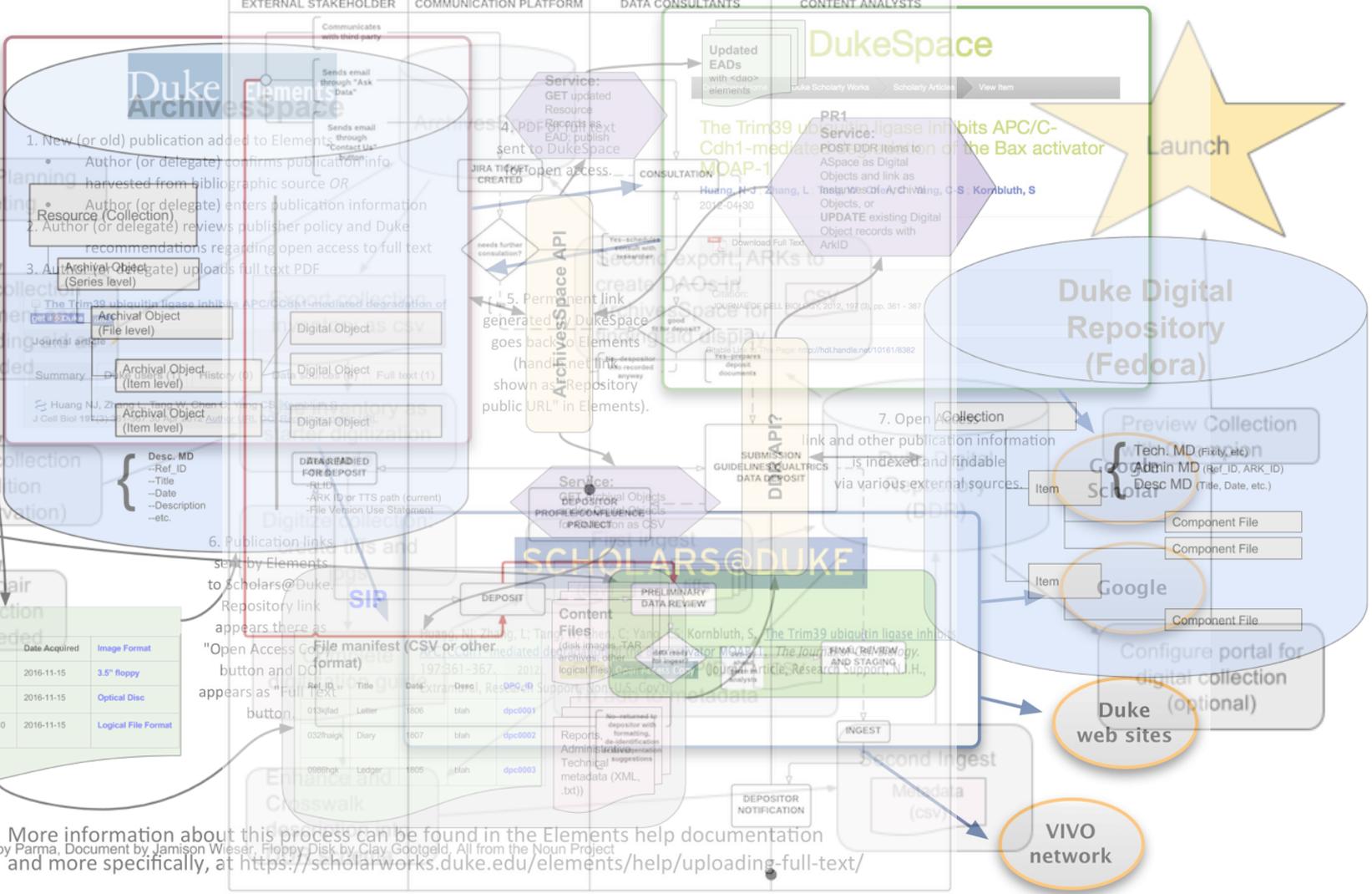
Source materials

Content Files
(disk images, TAR archives, other logical files)

Reports, Administrative, Technical metadata (XML, .bxt)

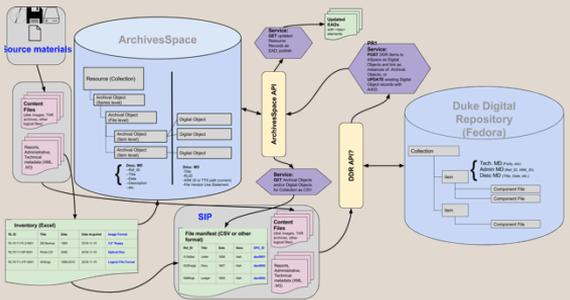
Inventory (Excel)

RL ID	Title	Date	Date Acquired	Image Format
RL10111-FL3-0001	DB Backup	1993	2016-11-15	3.5" floppy
RL10111-OP-0001	Photo CD	2000	2016-11-15	Optical Disc
RL10111-LFF-0001	Writings	1999-2010	2016-11-15	Logical File Format



More information about this process can be found in the Elements help documentation and more specifically, at <https://scholarworks.duke.edu/elements/help/uploading-full-text/>

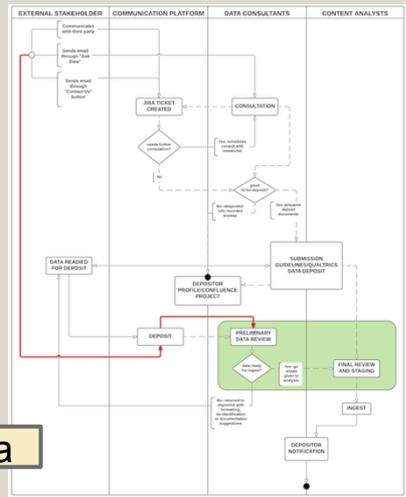
Server by Danil Polshin, CD by Parma, Document by Jamison Wieser, Floppy Disk by Clay Googold, All from the Noun Project



Source: David Pridwin. CC by Perms. Document by James Viner. Ffipsy Data by Clay Gouglard. All from the Moon Project

Library general collections

Scholarly products not data or peer-reviewed



Born-digital

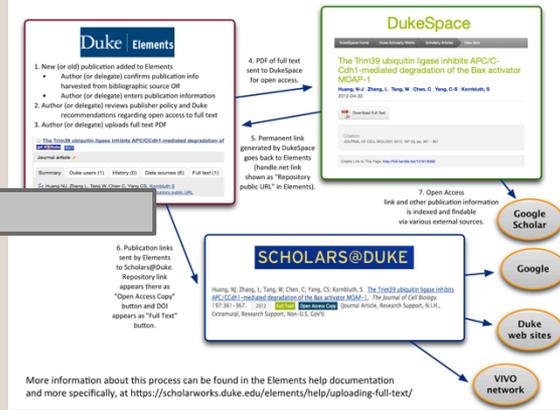
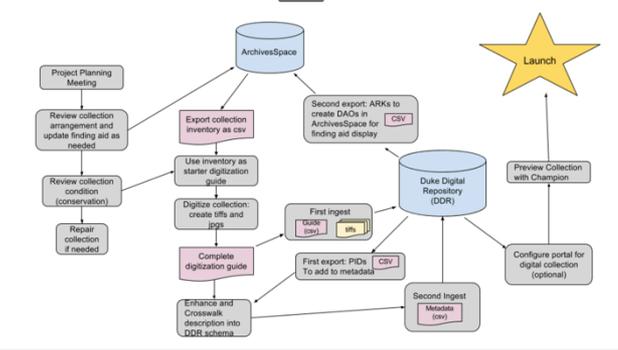
Digital collections

Research Data

Duke Digital Repository

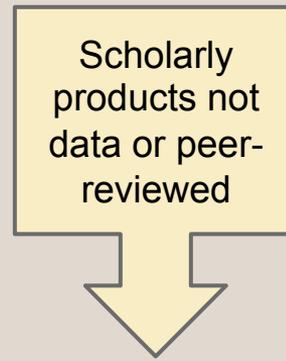
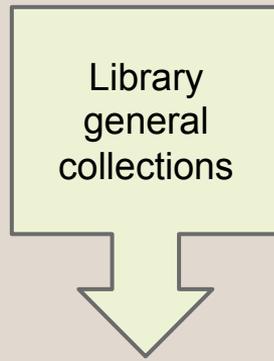
Electronic Theses & Dissertations

Scholarly publications



More information about this process can be found in the Elements help documentation and more specifically, at <https://scholarworks.duke.edu/elements/help/uploading-full-text/>

“proposed additions to the DDR that originate with the collecting and curatorial activities of staff in the Libraries”



DUL gets requests from campus to use the repository for a variety of items that our current policies don't cover

- Materials purchased or licensed by librarians
- Patrons may only use if we provide access

- Student scholarship
- Non peer-reviewed publications
- Archiving web sites for academic programs
- Digital objects related to exhibits
- Spawn of interdisciplinary centers

Is anyone doing these things?

Are we building collections for our libraries, are we providing services for members of our campus communities, are we doing both ... or what?

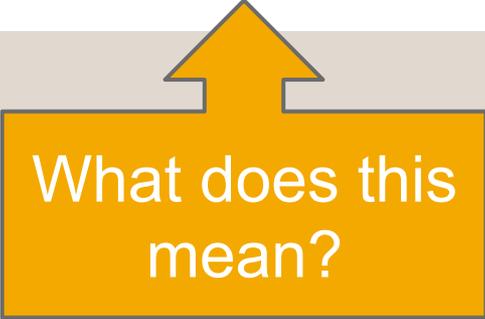
From <https://purr.purdue.edu/start>
“Get Started” @ Purdue University Research Repository

4

Publish Your Data

At any point during your project, you can publish datasets out of your project and make them available to the world from the [PURR repository](#).

- ↻ Each data publication has a title, authors, an abstract, associated documentation, and a unique [Digital Object Identifier \(DOI\)](#) that can be used to reference the data in journal and conference papers.
- ↻ Data publications are maintained for at least 10 years after the date of publication (see [space allocation](#) page for details). [After 10 years, datasets are curated by Purdue Libraries and maintained as a part of the Purdue Libraries' collections.](#)



What does this mean?

UNC at Chapel Hill Libraries

Jason Casden & Julie Rudder

Rethinking repositories

IVANOV'S



DETECTIVE

ROMANS

**VOLLEDIG
VERHAAL**

VERSCHIJNT
WEKELIJKS

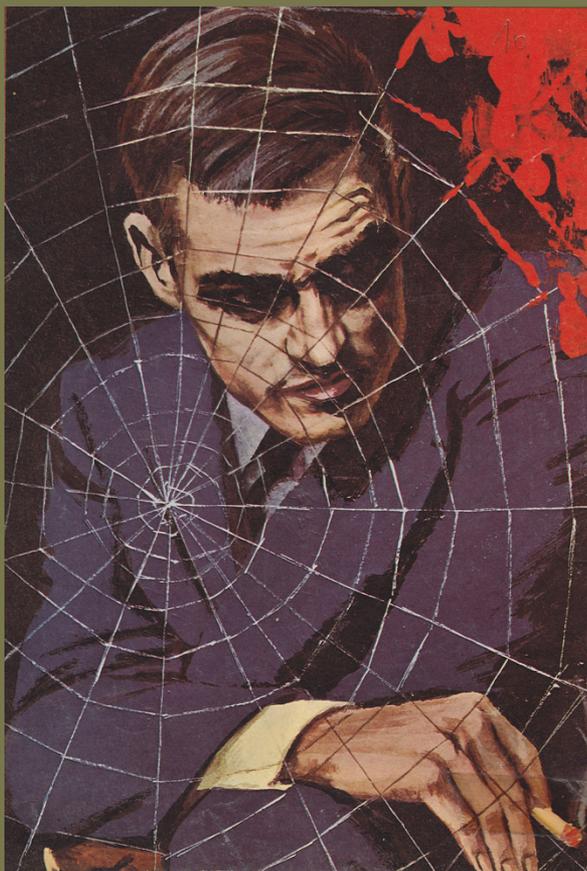


•
**DE MAN
ACHTER
HET WEB**
•

**HELLO
AMBOSS**

1

BELGIË 9 FR
NEDERLAND 65 CT



Rethinking repositories



**A Progress and Recommendations Report
from the Digital Curation/Institutional
Repository Committee, 2005-07:**

**Informing a Successful Institutional Repository Deployment at
the University of North Carolina at Chapel Hill**

Final
February 25, 2008

Carolyn Hank
On behalf of the DC/IRC

Preserve, share, and promote your scholarly and creative work.

Share Your Work



Open Access
Articles



Student
Papers



Posters &
Presentations



Open Access
Datasets

History

Governance increasingly centered at the library

Digital Curation & Institutional Repository Committee (DC/IRC) formed (2005)

Partnership between School of Information and Library Science, Campus IT, and Libraries

Focus on specifying and developing a pilot preservation repository

Carolina Digital Library and Archives (CDLA) formed (2007)

DC/IRC reports out (2007-2008)

Carolina Digital Repository pilots launched (Spring 2009)

Carolina Digital Repository production launch (Fall 2009)

Shifting scope

“The purpose of the UNC-CH Pilot IR is to demonstrate a system for the capture, storage, and dissemination of **digital assets created and/or maintained by the UNC-CH community members**. These digital assets represent the scholarly, intellectual, and creative contributions of UNC-CH community members, as well as represent valuable digital assets required for stewardship in accordance with the **North Carolina Public Records Law**.”

UNC-CH Pilot IR Service Definition Policy from DC/IRC Report (2007-2008)

Shifting scope

“The mission of the UNC Institutional Repository is to collect, preserve, and ensure continuing access to **digital content of enduring value to the University**”

Draft mission statement from DC/IRC Report (2007-2008)

Shifting scope

“An IR is intended as a solution for managing the diverse, unique digital assets produced at the University. ... **Such a system is intended to benefit the digital asset management needs and deficiencies of the University community in its entirety**, composed of numerous stakeholder groups of administrators, faculty, researchers, staff and students, across a multitude of disciplines, and to allow for managed stewardship so that these assets remain accessible in the long-term. Further, an IR is intended **to serve as a rich resource of University-produced and/or managed scholarly materials**, and to make these assets discoverable over time by University community members and the academic and research community at-large.”

Executive summary from DC/IRC Report (2007-2008)

Shifting scope

“The Carolina Digital Repository (CDR) is a digital archives for materials produced by members of the University of North Carolina at Chapel Hill community. The main goal of the CDR is to **keep UNC digital scholarly output safe and accessible for as long as needed**. It also serves as a repository of **historical materials that broadly support the University’s academic mission**. More specifically, the CDR aims to acquire UNC digital material and ensure it is accessible, searchable and safe from alteration.”

“About the Repository” - <https://blogs.lib.unc.edu/cdr/>

Our current infrastructure

ContentDM for digitized materials (~1,500,000 objects)

Carolina Digital Repository for at-risk materials and IR (~650,000 objects)

Simultaneously, we manage ContentDM for digitized collections (access focus),
CDR for born digital and faculty deposit (preservation focus)

Filesystem storage for most digitized preservation masters

Thriving

Large, varied collection

Sophisticated customized preservation workflows around Fedora

Self-deposit features

Developed a community of researcher and departmental users of IR services

ETDs, Master's papers, Honors Theses

Data sets

A/V materials

Publications

Satisficing

CDR has gradually become our de facto preservation repository

“At-risk digital materials” from Special Collections

All born-digital materials

A subset of digitized materials

Research outputs of the institution

Student papers

University Archives

...and also a staging area for unprocessed born digital materials

Prompted by a storage crisis

Custom infrastructure

As complexity increases, we rarely fully benefit from common international efforts.

Services

“While guidance, training and engagement activities are a critical factor in achieving IR support, participation, use, and deposits, with the latter a major obstacle reported by many operational IRs, it is premature to act on such strategic plans unless a system is in place, or nearing completion, for demonstration and use”

Recommendations from DC/IRC Report (2007-2008)

UNC Open Access Policy

“Each Faculty member of the University of North Carolina at Chapel Hill grants to the University a nonexclusive, noncommercial, irrevocable, worldwide license to exercise, and to authorize others to exercise, any and all rights under copyright relating to each of his or her scholarly articles, in any medium, for the purpose of making those articles freely and widely available in an open access repository.”

January 1, 2016 - <http://policies.unc.edu/policies/open-access/>

UNC Open Access Policy

“The Scholarly Communications Office of the University Library, or other office designated by the Provost, will be responsible for interpreting this policy, resolving disputes concerning its interpretation and application, and recommending changes from time to time. The Scholarly Communications Office shall also create and maintain the repository terms of use governing public access to and use of the scholarly articles licensed under this policy.”

January 1, 2016 - <http://policies.unc.edu/policies/open-access/>

Institutional support for library services

Three new provost-funded positions

Institutional Repository Librarian

Open Access Librarian

Software Developer

One-time funds to support outreach and internal training

Renewing our focus

Redouble scholarly service development efforts

Reevaluate the benefits of community-developed software

Redefine the scope of our preservation infrastructure

Rethink our definition of success

Some of what we set out to do:

Learn from other universities.

Understand the local UNC community and needs.

Involve as many library staff as possible.

What we heard from faculty:

Care about getting content from behind paywalls.

Don't care as much about getting already open-content into the repository.

Care about author's rights (suspect this is the main things their colleagues will care about).

The majority of faculty don't yet know about the policy.

There are some basic IR-type features that are lacking from the current CDR.

The policy absolutely, positively can not be a great burden to faculty.

When we talked to library staff and administration:

Having lots of content in the repository.

Populating faculty profiles.

Showing the impact and the volume of UNC research.

A desire to really understand the policy and strategies to talk to faculty about the benefit of OA.

“Perhaps the most compelling arguments for open access policies, however, have nothing at all to do with the numbers but, instead, are driven by the anecdotal stories we hear of how open research makes a difference to people around the world.”

~Catherine Mitchell

<http://osc.universityofcalifornia.edu/2016/10/does-the-uc-open-access-policy-miss-the-mark-depends-on-which-mark/>

“...in our preoccupation with low compliance rates for open access policies, we often forget to highlight the transformative rights declarations of these policies, which are purely independent of compliance and apply to all authors.”

~Catherine Mitchell

<http://osc.universityofcalifornia.edu/2016/10/does-the-uc-open-access-policy-miss-the-mark-depends-on-which-mark/>

We decided to focus on four areas (3 year plan)

Harvest open content: Use automated tools, invest in sustainable community efforts, integrate with campus faculty activity tracker and profile system.

Paywalled content recovery: Create systematic way to find UNC publications and make them open. IR Librarian's focus.

Open access education and outreach: Launch major campaign (hire marketing firm) around authors' rights. Create a robust outreach and education program. OA's focus.

Cultivate the value of openness (goes beyond policy): Prioritize work that makes our collections more open and usable (rights statements, better metadata and discovery). Clarify and increase support for data and other non-article content.

How does the OA policy relate to the scholarly record and to the IR as a whole?



Data, Publications,
Presentations, Student Work,
ETDs.

**Hyrax (formerly Sufia)- a
Hydra application**

Library Managed Collections

Born-digital collections.
At-risk digitization.

**Current CDR technology
stack.**

Why two systems?

- Very different users with different needs for ingest, management and access of content.
- Need for focused direction governance.
- Distinct user groups and service management needs between use cases.
- Need for different policies
 - Preservation policies
 - Deletion/editing policies
 - Collection development policies

○ Constant collection ingest and churn for Library Managed Collections

Near term work:

- Increasing our data curation services.
- Hire marketing consultants to launch campaign for open access.
- Create better metadata for EVERYTHING.
- Add right statements to EVERYTHING.
 - Train and invest in library staff.
 - ORCID implementation.
 - Needs assessment for Digital Collections.
 - Launch and customize Hyrax

Some of our open questions:

- What level of preservation do we need for IR content?
- How does the heavy preservation focus affect what we accept into an IR?
- How will policies change for the new system?
- Governance model for 2 systems?
- What faculty activity tracking system will UNC adopt? We know this is the future.
- Can we use the current CDR system for all library managed content?
- Where are other institutions doing a good job of inserting in the research life-cycle.