Binghamton University
The Open Repository @ Binghamton (The ORB)

Library Scholarship | University Libraries

10-2017

They are more like guidelines: reflections on best practices from new professionals

Amy Gay
aegay@binghamton.edu

Joe Carrano

Charlotte Kostelic

Megan Potterbusch

Follow this and additional works at: https://orb.binghamton.edu/librarian_fac

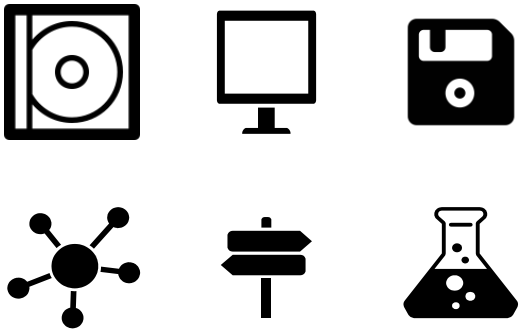
Part of the Library and Information Science Commons

Recommended Citation

Gay, Amy; Carrano, Joe; Kostelic, Charlotte; and Potterbusch, Megan, "They are more like guidelines: reflections on best practices from new professionals" (2017). *Library Scholarship*. 38.
https://orb.binghamton.edu/librarian_fac/38

This Presentation is brought to you for free and open access by the University Libraries at The Open Repository @ Binghamton (The ORB). It has been accepted for inclusion in Library Scholarship by an authorized administrator of The Open Repository @ Binghamton (The ORB). For more information, please contact ORB@binghamton.edu.

They’re More Like Guidelines: Reflections on Best Practices from New Professionals



NDSR DC 2016-2017
Joe Carrano
Elizabeth England
Amy Gay
Charlotte Kostelic
Megan Potterbusch

We've Been Compromised: Revealing Our Digital Preservation (Not So Best) Practices

Joe Carrano | @joecar25
Georgetown University Library

Some Context

- 2016-2017 National Digital Stewardship Residency at Georgetown University Library to design and implement workflows to transfer content to preservation storage.
- Chosen digital preservation storage solution was Academic Preservation Trust (APTrust)

What is APTrust?

- A distributed, dark, digital-preservation repository.
- A consortium of institutions that are a part of the governance and development of the repository.

Where is the data stored?

- APTrust stores member institution data in both Amazon S3 and Glacier cloud storage.
- It is stored both in Northern Virginia and in Oregon, with 3 copies of the data stored in each cloud storage service and location.

Should you trust cloud storage for preservation?

Best practice: Not on its own.

Cloud storage providers aren’t in it for the long term

Lots of Amazon's projects have failed...and that's ok, says Amazon's Andy Jassy

Julie Bort
Oct. 25, 2016, 3:51 PM 3,826

Spoiler: Andy Jassy is the CEO of Amazon Web Services (AWS)

Should you trust cloud storage for preservation?

Our practice: Yes (sometimes).
Why?

- Costs for current local network solutions are too high.
- Staff have little time to manage local content.
- Decide on a case by case basis if need to keep locally.

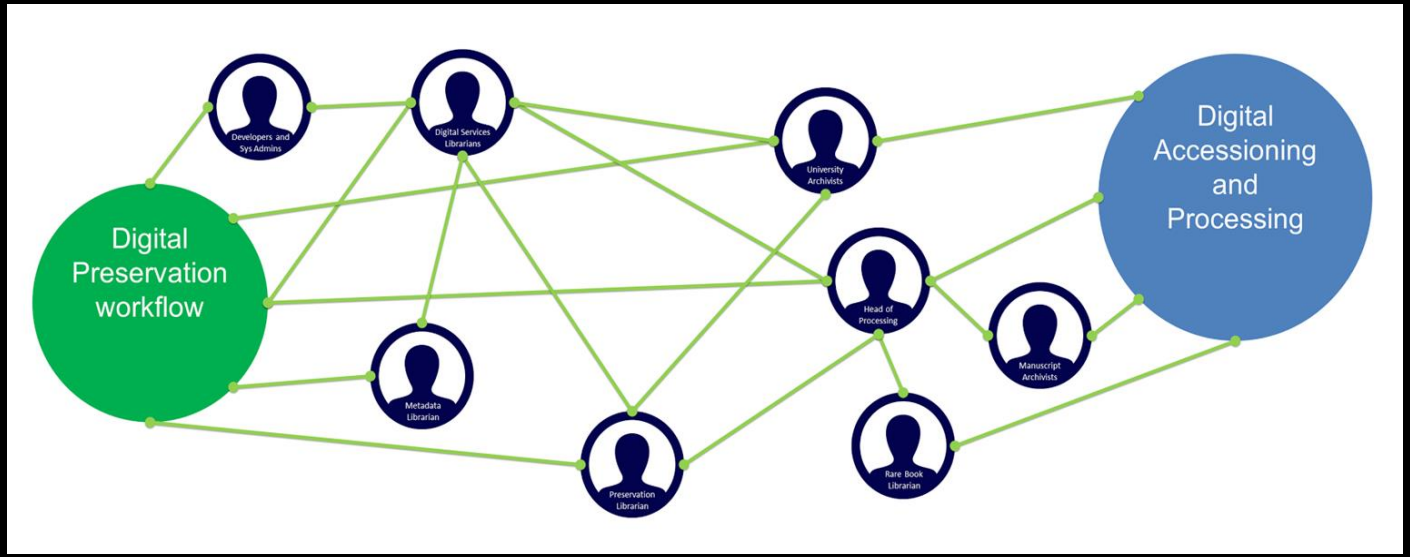
Should there be at least one staff member dedicated full time to digital preservation?

Best practice: Yes.

Should there be at least one staff member dedicated full time to digital preservation?

Our practice: Someday. For now now we need to distribute the work.

Digital preservation as a distributed responsibility



8006 digital objects

3.1 TB of content

12 staff trained

Most of them willingly!

Room to grow

Staff and workflow are ready to scale up

Is your preservation copy always the unaltered original?

Best Practice: Yes.

When to redact AIPs vs DIPs in digital archives workflows?

511 views

+1 vote

Tools like bitcurator are integrating processes for redacting information from disk images. So, not just identifying information that you might want to use to make aprasial decisions about what to keep but also the ability to directly redact information from files. I know some folks are planning on using those tools that auto locate personal info and then redacting files or redacting information from inside files.

Given that we have (at least in theory) Archival Information Packages and Disimenation Information packages we could be redacting, when should archivists be redacting the archival copy and when do we want to just be redacting the access materials? If one has information that is sensitive now but would be useful and non-sensitive in 50 years? Or, is the threat of mantaining sensitive information and multiple copies of materials too significant to warrent such an approach? Or, is this something that is going to depend on the particular issues in a collection. If the answer is "it depends" what is it that it depends on?

redaction | disk-image | aip | dip | workflow

asked **Jun 24, 2014** by **tjowens** (2,350 points)

flag **Answer** comment

Ideally the archivist would never have to redact the archival copy. Disk images are best though of as instances of the original disk--that is, we keep the disk image of a 3.5" floppy disk because we don't want to have to read the data from a floppy disk every time we want to access it and because we're concerned that we'll lose the data due to bit rot, etc. Remember that a disk image that is the full bitstream from the original media isn't a copy, it's the same object but moved to a more sustainable medium. To redact a disk image, then, would be like redacting a paper archive by scratching words out on the original with a sharpie.

To redact a disk image, then, would be like redacting a paper archive by scratching words out on the original with a sharpie. - Porter Olsen

http://qanda.digipres.org/121/when-to-redact-aips-vs-dips-in-digital-archives-workflows?show=156#c156

Is your preservation copy always the unaltered original?

Our Practice: It depends.

Information Security

SECURITY

Equifax data breach: Find out if you were one of 143 million hacked

Library Fends Off DDoS Attack

July 20, 2016 by John Sayers

PRIVACY AND SECURITY

FBI Arrest Chinese National Linked to OPM Data Breach Malware

It isn't a question of whether the bad guys will get in to the swimming pools and barrels of data, and exfiltrate it. It is simply when they will do so, and how long it will take you to find out that they have. - David Rosenthal

http://blog.dshr.org/2017/10/not-whether-but-when.html

Delete Your Account

We’re concerned about information security at Georgetown.

- University InfoSec policy
- Acknowledge our technical limitations
- Ultimately realized it aligns with a commitment to privacy and ethics.

When taking in archival collections with PII, we decided on an approach similar to what Maciej Cegłowski describes in his talk “Haunted by Data.”

http://idlewords.com/talks/haunted_by_data.htm

Don’t Collect It

One step we take is to work with donors.

- Talk with the donor and exclude unnecessary PII from the donation if possible.
- If information is deemed valuable, donor is helpful in identifying where PII is located for later redaction.
- Aligns with the idea of community-driven digital curation and an ethic of care advocated for by Stacie Williams in “Sustainable Digital Scholarship.”

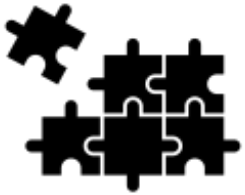
BitCurator

Don’t Store It

The BitCurator Redaction tool is our main solution for disk images.

- Utilize the bulk redaction utility that redacts items surfaced by bulk extractor.
- Known items outside of the that report, use custom config file and dfxml file to redact.

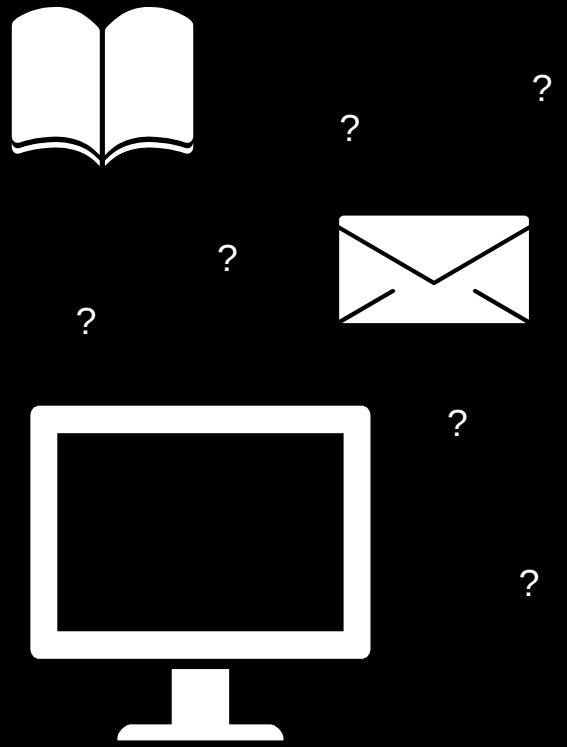
Connecting 18th Century Data for the 21st century: George Washington and George III in the Digital Age



Charlotte Kostelic | @chuckkostelic
Library of Congress/Royal Collection Trust

Georgian Papers Programme

Why am I presenting at DigiPres if I work with 18th century manuscripts?



Interoperability: n.

The ability of computer systems or software to exchange and make use of information

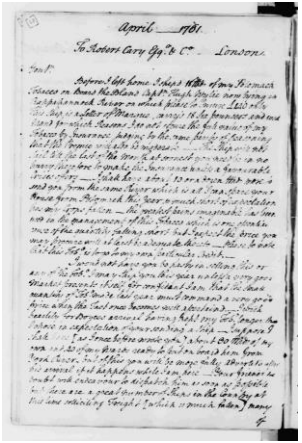
https://en.oxforddictionaries.com/definition/interoperability

Two institutions, with distinct curatorial units, & many applied standards

Varying digitization standards through time

George Washington Papers

[scanned in mid-1990s]



Alexander Hamilton Papers [scanned in 2017]

Interoperability Levels

How can we measure interoperability?
Is there a maturity model for interoperability in digital libraries?

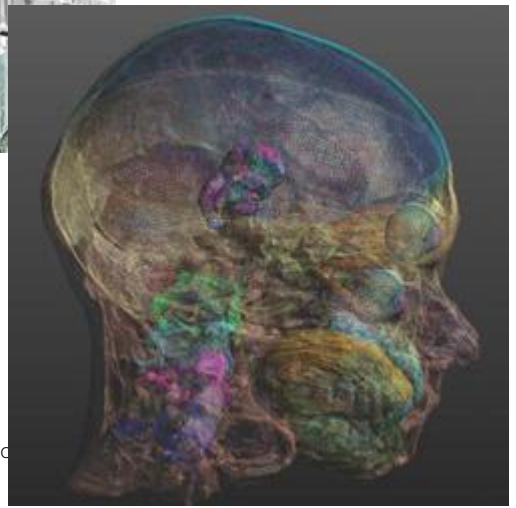
Level	Layer Name	Description of level
L6	Conceptual	Interoperating systems at this level are completely aware of each other's information, processes, contexts, and modeling assumptions.
L5	Dynamic	Interoperating systems are able to re-orient information production and consumption based on understood changes to meaning, due to changing contexts.
L4	Pragmatic	Interoperating systems will be aware of the context (system states and processes) and meaning of information being exchanged.
L3	Semantic	Interoperating systems are exchanging a set of terms that they can semantically parse.
L2	Syntactic	Have an agreed protocol to exchange the right forms of data in the right order, but the meaning of data elements is not established.
L1	Technical	Have technical connection(s) and can exchange data between systems
L0	No	NA

Using Collaboration for the Win … Superheroes do it too

Amy Gay
U.S. Food and Drug Administration

This presentation reflects the views of the author and should not be construed to represent FDA's views or policies

Background of CDRH & OSEL



Center for Devices & Radiological Health
Office of Science & Engineering Laboratories
Mission: To protect and promote Public Health

Images from: https://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDRH/CDRHOffices/uc

What I walked in expecting

- ➔ Everyone wants this and is on board
- ➔ Everyone knows what a data catalog is
- ➔ Public Access Policies and Data Management Plans already common knowledge
- ➔ All projects ready to

Reality Check

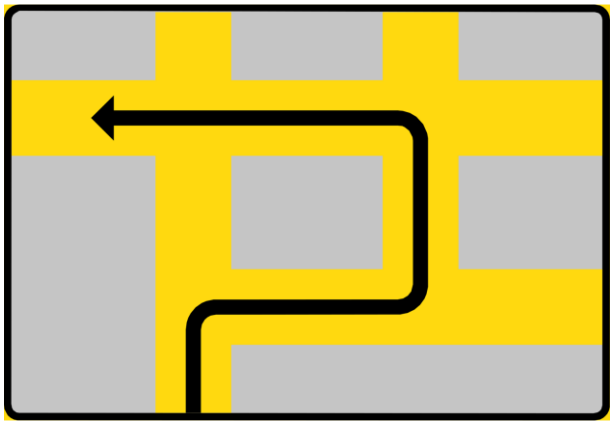
- Federally mandated public access policies
- Research Data Management still a new concept
- No institutional repository
- No current tracking method
- You're a *librarian*?

So About those “Best” Practices…

- Free, Open Source Tools
- Access versus Access
- Approval Time & Process
- Research, Research, Research

Discoveries during User Research

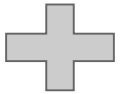
- Data Management Practices
- Public Access
- Data Management Plans
- Policy Implementation



www.fda.gov

Parallel Project Collaboration

OSEL Research Project Inventory



OSEL Science Data Catalog

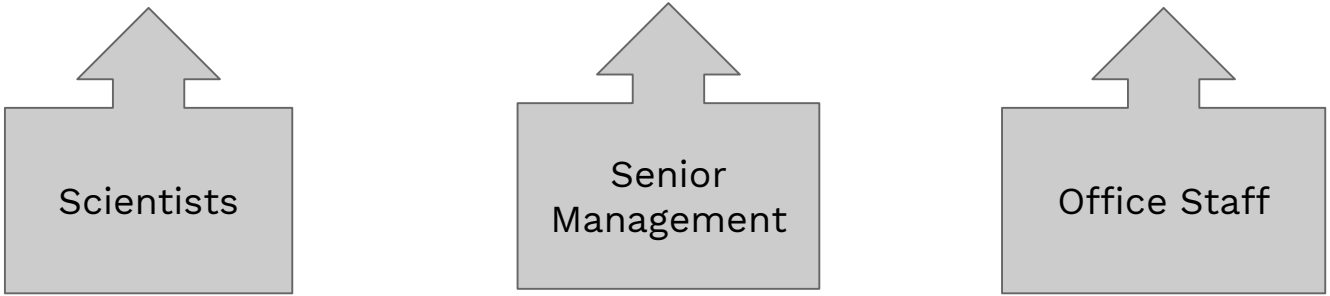
Cross-Agency Virtual Collaboration

Federal Agency Data Curation Toolkit: A Resource for Agencies Managing Data

Micro-Volunteers from 9 different agencies

100% Virtual

OSEL Communications Working Group



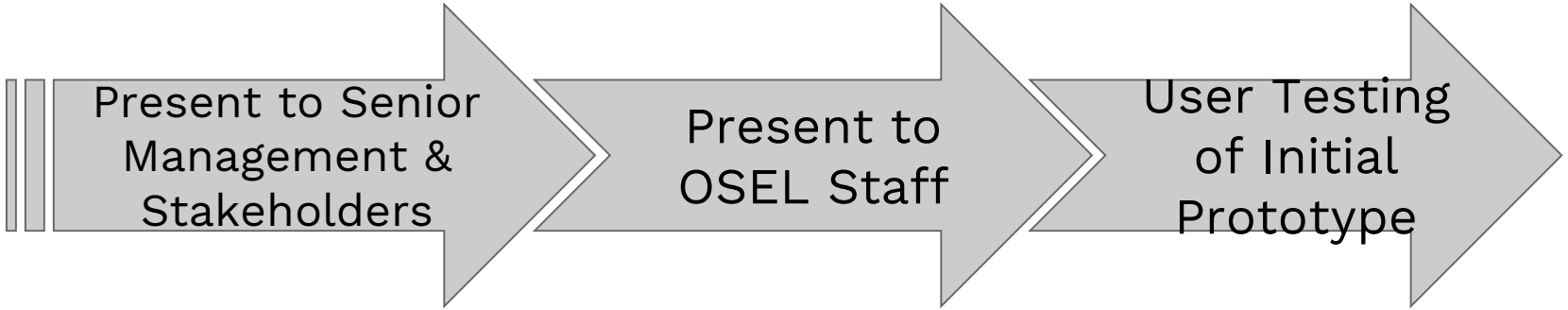
Where We are at Now

SharePoint Landing Page & Internal Database

User Guide for Science Data Catalog

FDA Library Landing Page & Catalog Records

Next Steps


Present to Senior Management & Stakeholders
Present to OSEL Staff
User Testing of Initial Prototype
October 2017
November 2017
December 2016 - February 2017

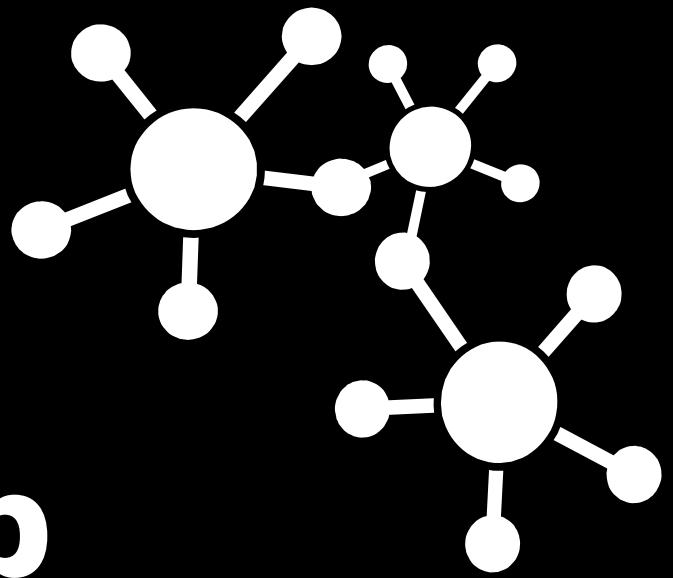
First Things First: Where are you storing your active research?

Megan Potterbusch @librarpotter
Association of Research Libraries

Project Hosts

- Association of Research Libraries
- George Washington University
- Center for Open Science

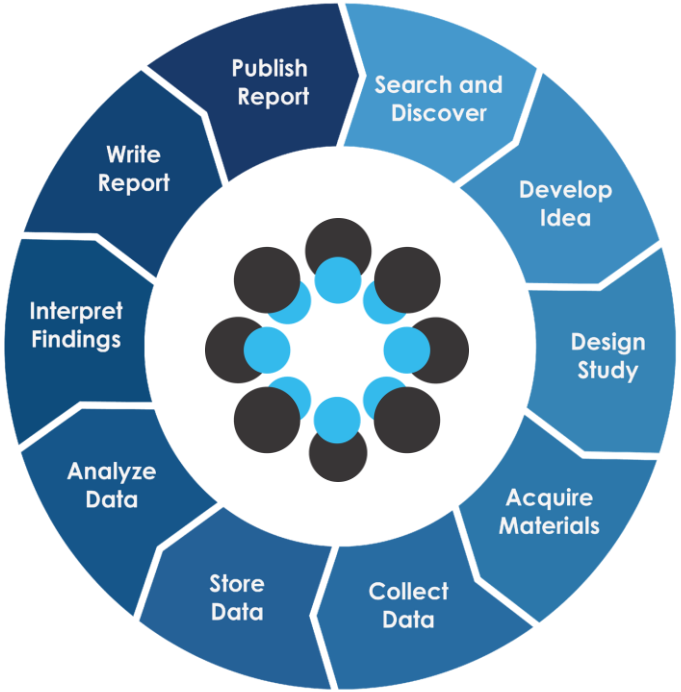
Supporting Digital Stewardship



Get involved with research workflows in order to support better stewardship practices

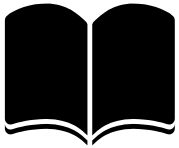
Stewardship is an ethic that embodies the responsible planning and management of resources.

Active research lifecycle



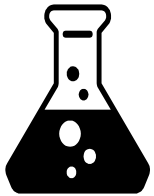
Designed to support entire research lifecycle
“Registrations” feature bundles workspace to “respect des fonds” of research

Open Science Framework (OSF)



Humanities

A documentary history project



Life Sciences

Biomedical teams

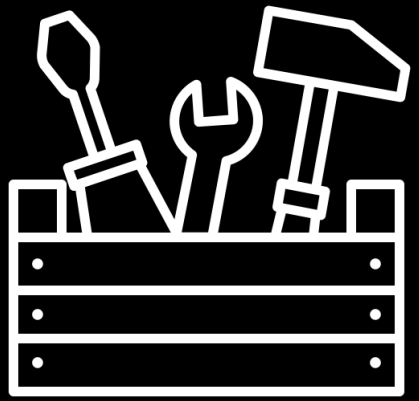
Team Science

Interdisciplinary Think Tanks

Created by David García
from Noun Project

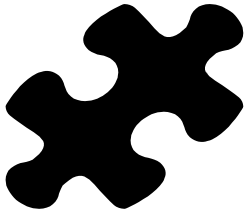
OSF Toolkit

Documentation, presentations, lessons learned, API code, example projects, etc.



Created by Maria Kislitsina
from Noun Project

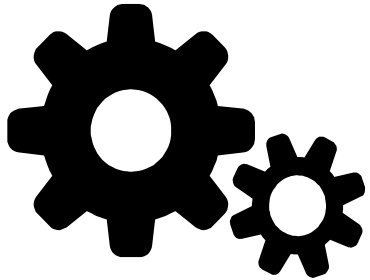
OSF Toolkit for Digital Scholarship Support



- Resources for working with researchers
- Scripts and Code
- Presentations, workshops, and outreach materials
- Partnering Across Research Services
- Find similar projects
- Example projects

https://osf.io/ubzve/

Resource for librarians using OSF



- Add. reuse, reshape, share
- Learn from each other
- Link to new resources
- Supplement and point to resources

https://osf.io/ubzve/

Community of Practice

Share knowledge, develop materials, and contribute to a common resource.

Created by Creative Stall
from Noun Project

Thank you!

Questions?