Responding to the Call to Curate

Digital Curation in Practice at Penn State

Patricia Hswe, Michael J. Giarlo, Michael J. Furlough, and Mairéad Martin

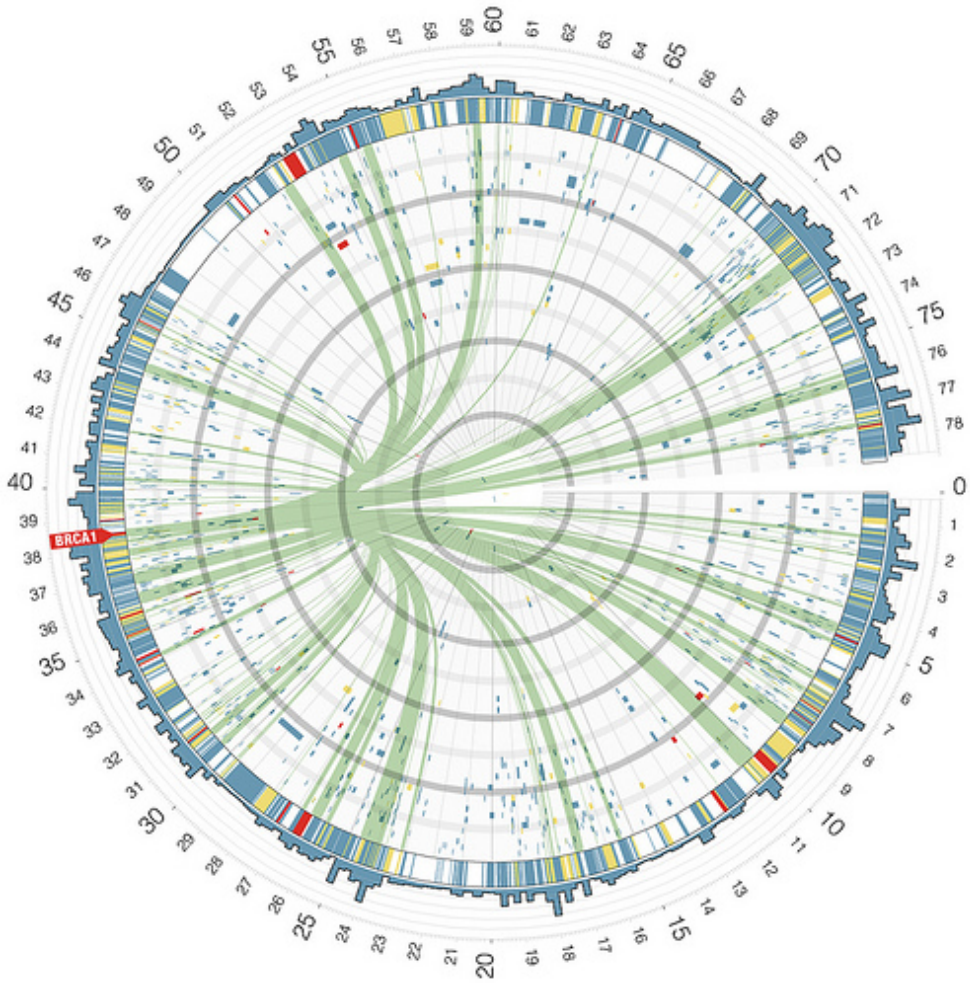
Framing curation in practice at PSU

- Content Stewardship Program at Penn State
- Deploying a Digital Collections Curator and Digital Library Architect
- Organizational contexts for curation in practice
- Current status and prospects

But first: some context

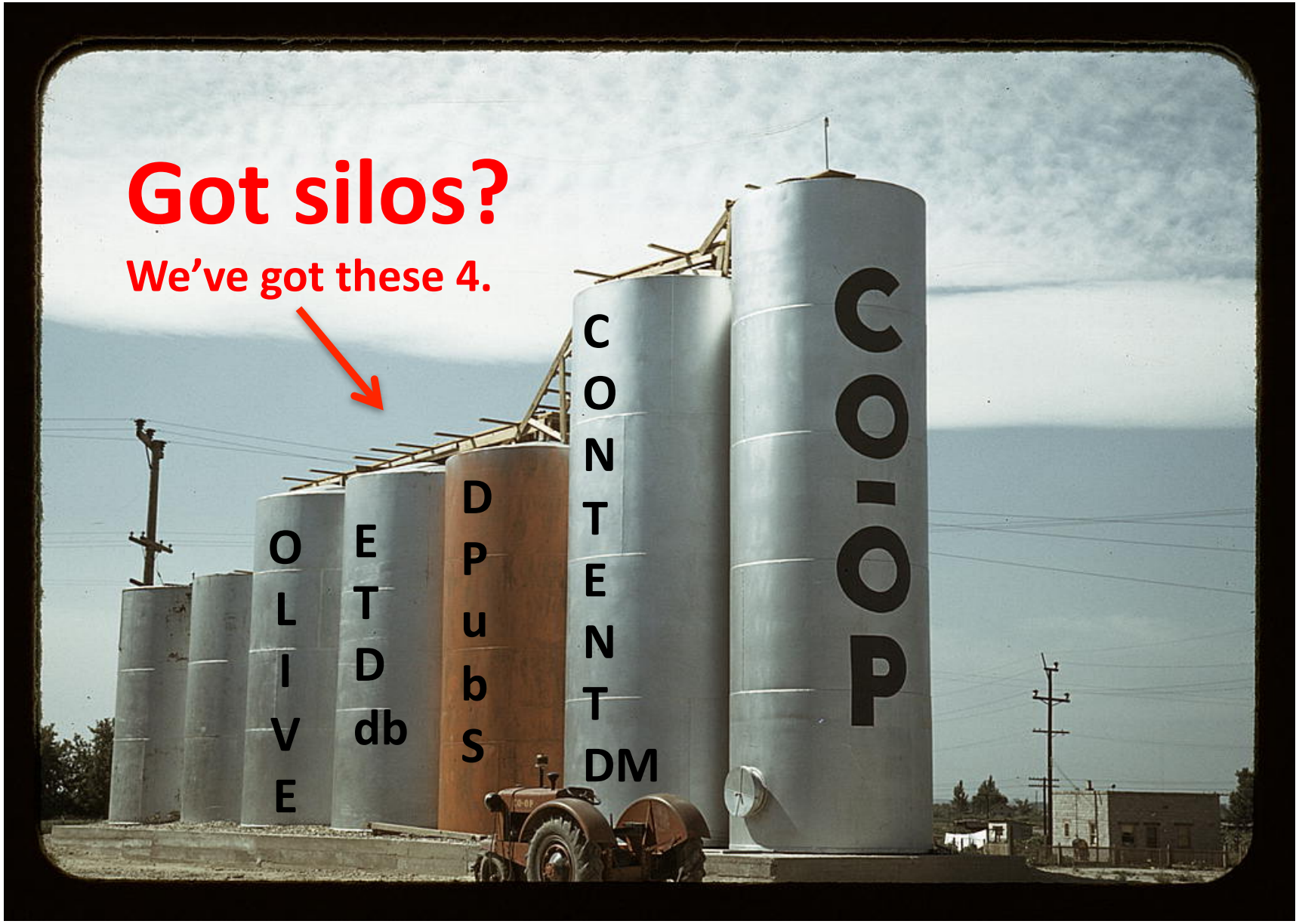
Why Penn State launched the Content Stewardship Program

It started with scientific data management



At the same time . . .

There were other considerations


Got silos?
We've got these 4.
OLIVE
ETDdb
DPubS
CONTENTDM
CO-OP

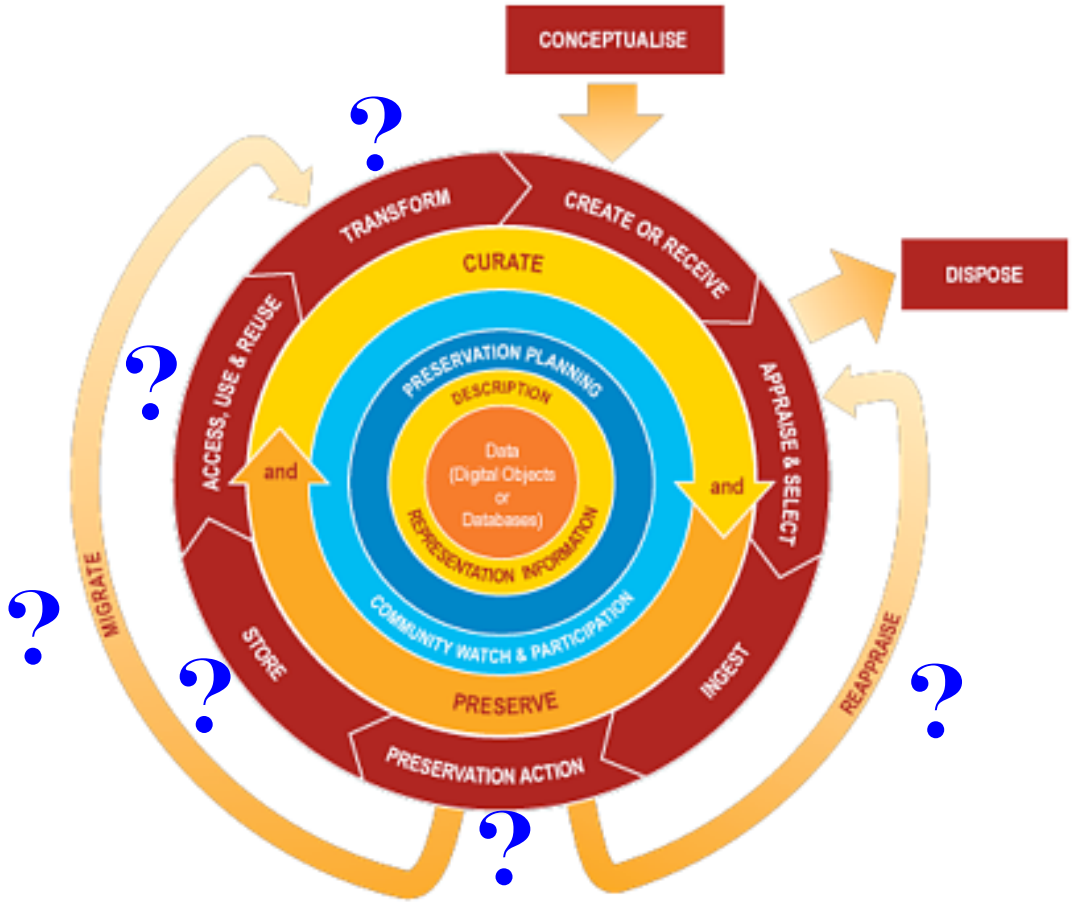
But we did steer clear of IRs

* The Libraries have experimented with Fedora, but there's been no formal instance of it as an institutional repository at PSU.

Emphasis on digitization/production (which we do very well)

Too well?

In mounting collections online, we weren't considering the whole picture of lifecycle management.



PENNSTATE

Content Stewardship @ Penn State

Building tools and partnerships to enrich and manage digital content

Go People | Departments | Penn State

Home | Blog

About

The Content Stewardship program is a institutional initiative to address digital content and data management needs in areas such as digital library collections, scholarly communications, electronic records archiving, and e-science/e-research data management. Building on existing services and infrastructure, the program is developing an interoperable and extensible suite of discovery, preservation, curation, archival, and storage services.

Home

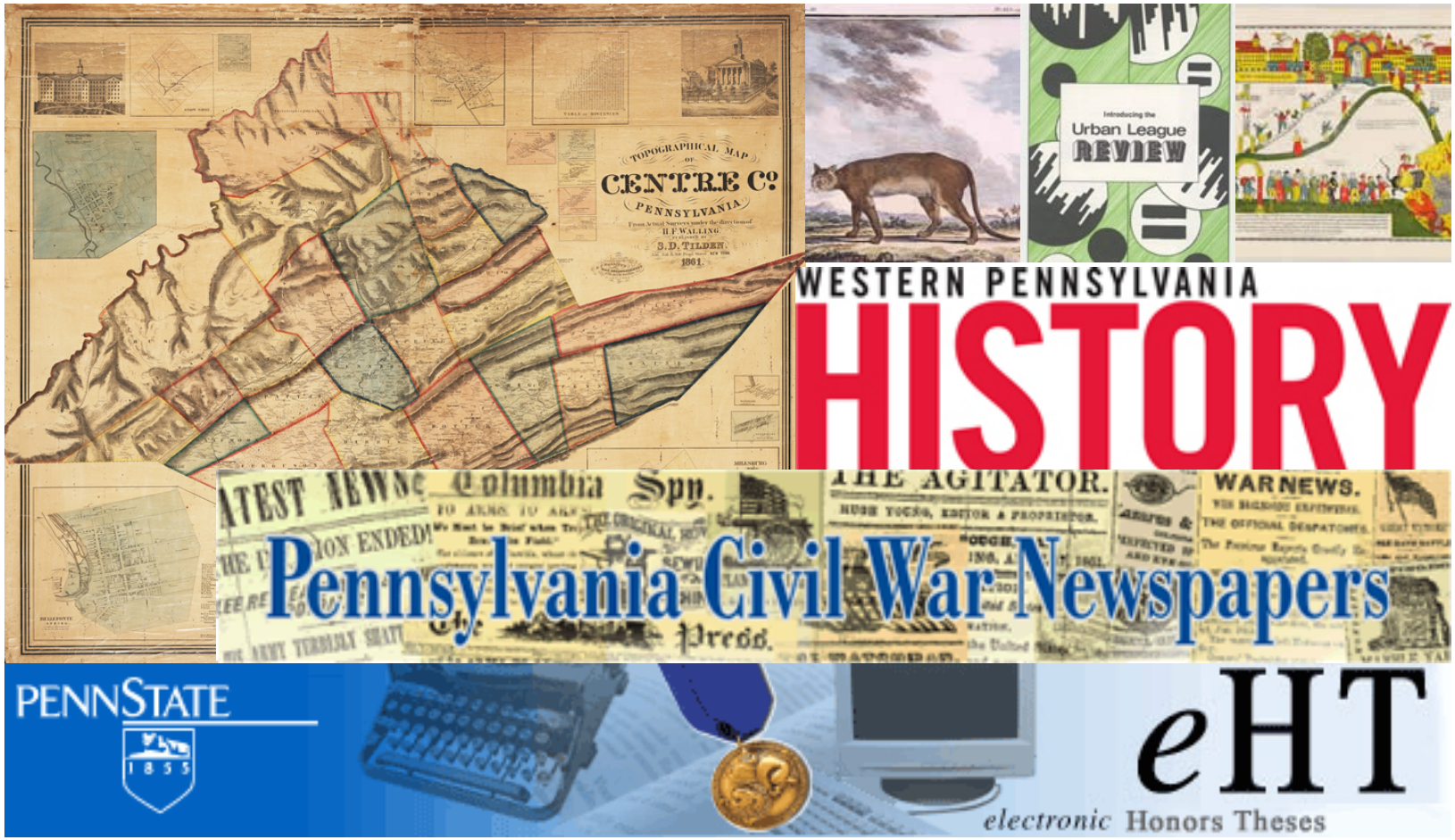
Program's First Year

- National search for Digital Collections Curator and Digital Library Architect
  - Year-long effort
- In parallel, investigation into storage issues
  - eXtensible Access Method (XAM) prototype

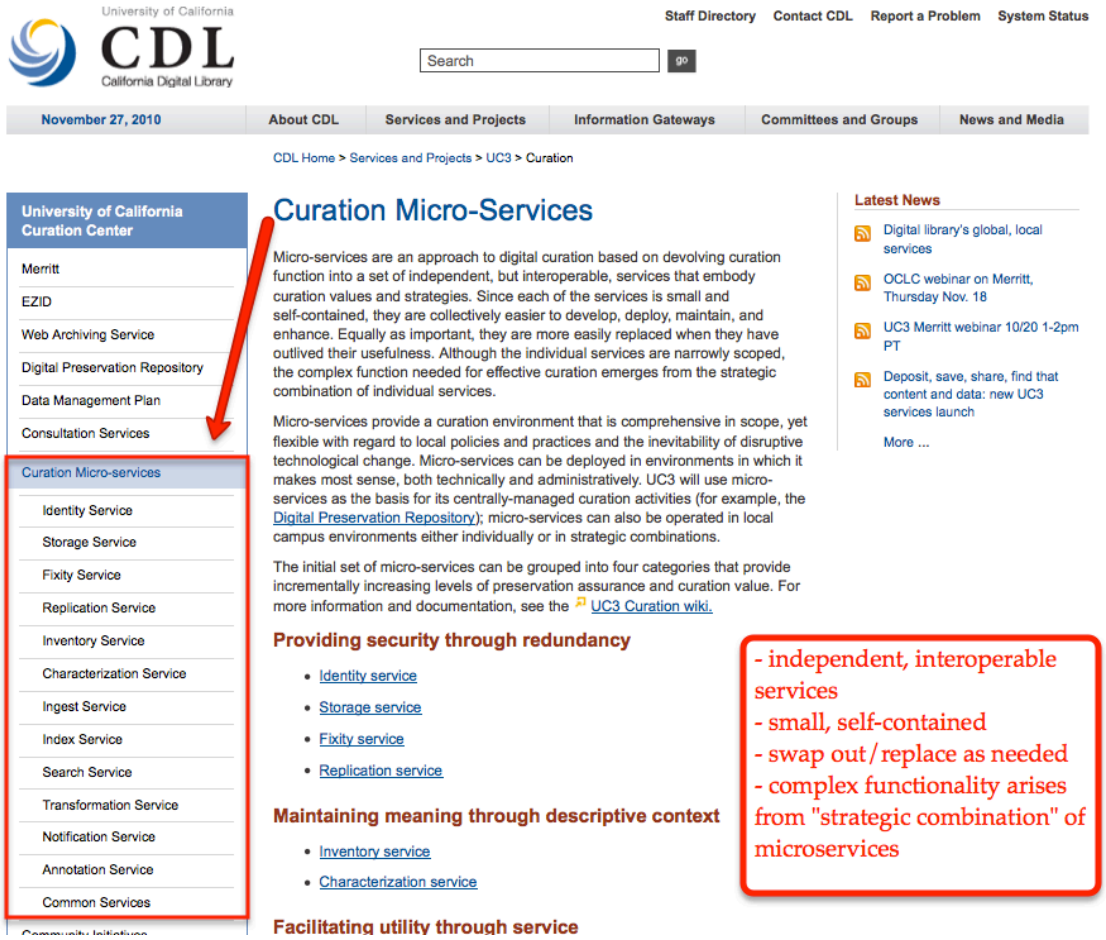
Deploying a Digital Collections Curator and Digital Library Architect

Activities in the first year

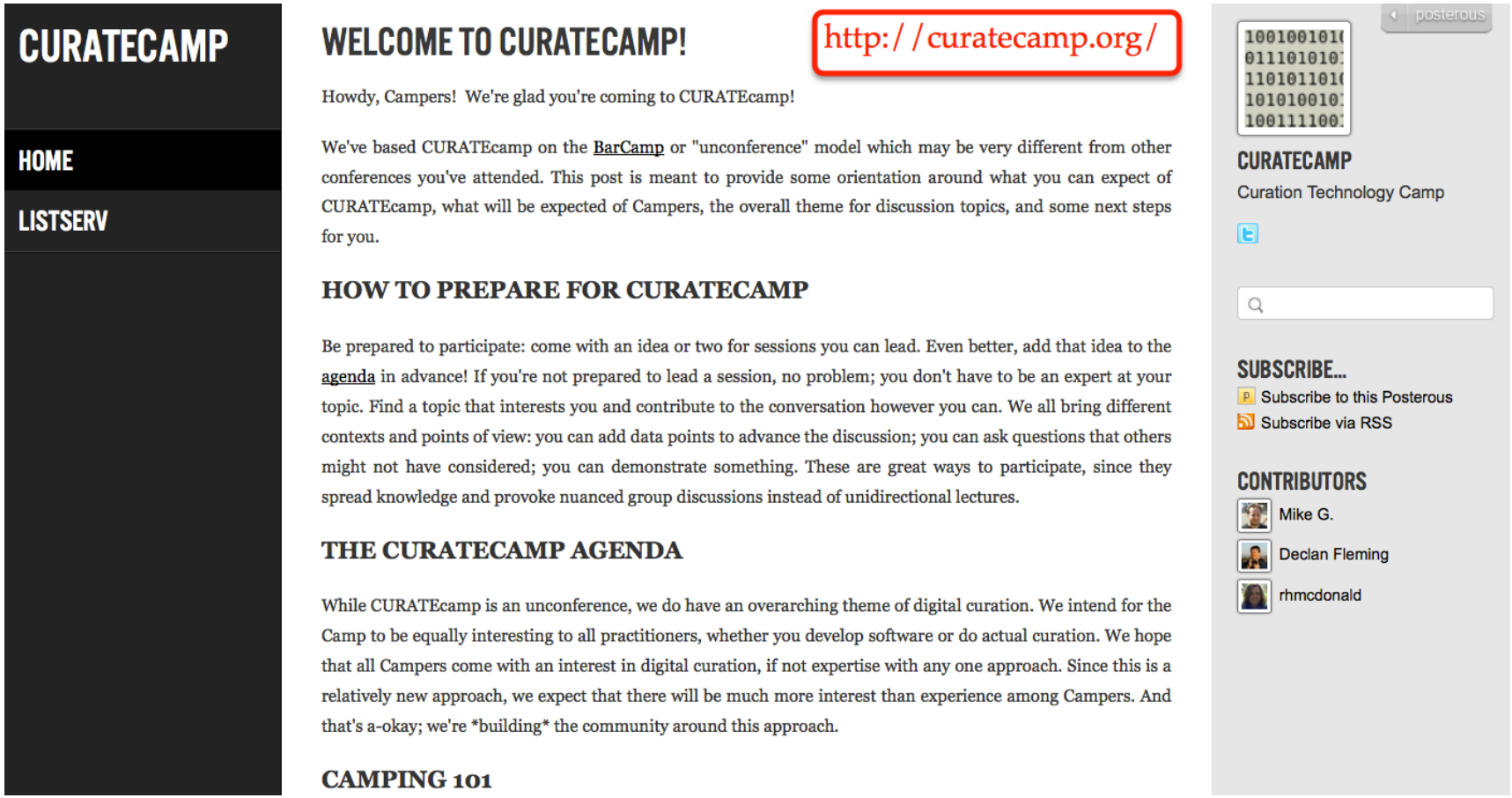
Platform Review



Curation Microservices Pilot Project



Building a digital curation community - 1



CURATECAMP

HOME

LISTSERV

WELCOME TO CURATECAMP!

http://curatecamp.org/

Howdy, Campers! We're glad you're coming to CURATEcamp!

We've based CURATEcamp on the BarCamp or "unconference" model which may be very different from other conferences you've attended. This post is meant to provide some orientation around what you can expect of CURATEcamp, what will be expected of Campers, the overall theme for discussion topics, and some next steps for you.

HOW TO PREPARE FOR CURATECAMP

Be prepared to participate: come with an idea or two for sessions you can lead. Even better, add that idea to the agenda in advance! If you're not prepared to lead a session, no problem; you don't have to be an expert at your topic. Find a topic that interests you and contribute to the conversation however you can. We all bring different contexts and points of view: you can add data points to advance the discussion; you can ask questions that others might not have considered; you can demonstrate something. These are great ways to participate, since they spread knowledge and provoke nuanced group discussions instead of unidirectional lectures.

THE CURATECAMP AGENDA

While CURATEcamp is an unconference, we do have an overarching theme of digital curation. We intend for the Camp to be equally interesting to all practitioners, whether you develop software or do actual curation. We hope that all Campers come with an interest in digital curation, if not expertise with any one approach. Since this is a relatively new approach, we expect that there will be much more interest than experience among Campers. And that's a-okay; we're *building* the community around this approach.

CAMPING 101

posterous

CURATECAMP

Curation Technology Camp

SUBSCRIBE...

Subscribe to this Posterous

Subscribe via RSS

CONTRIBUTORS

Mike G.

Declan Fleming

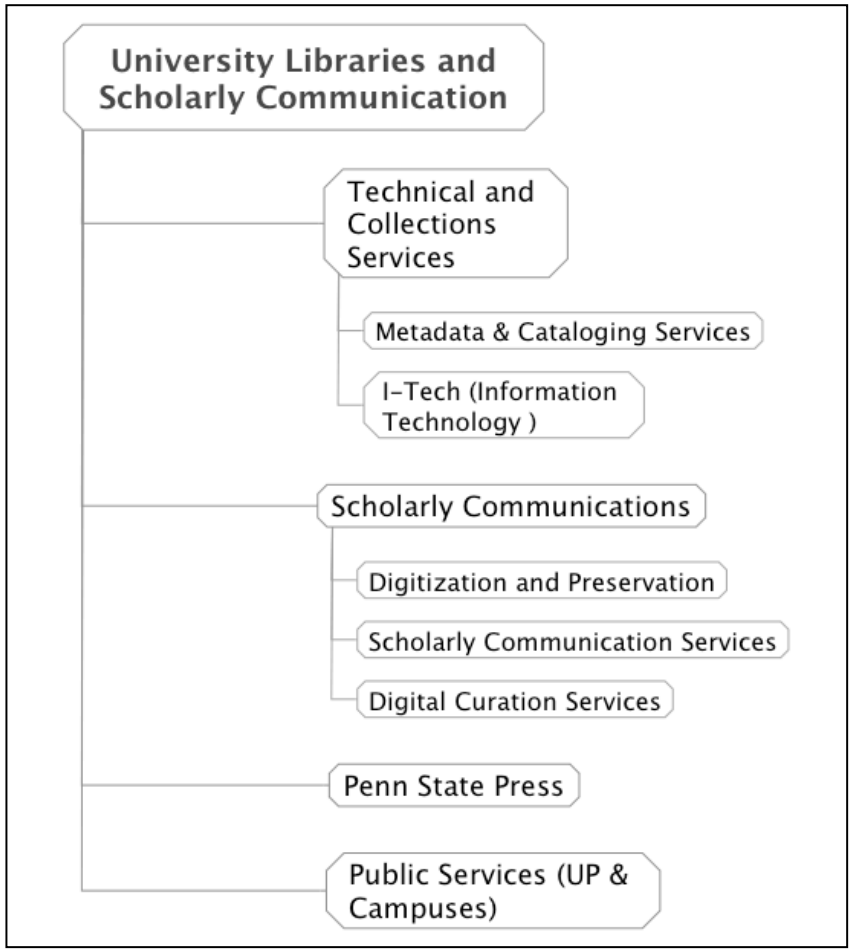
rhmcdonald

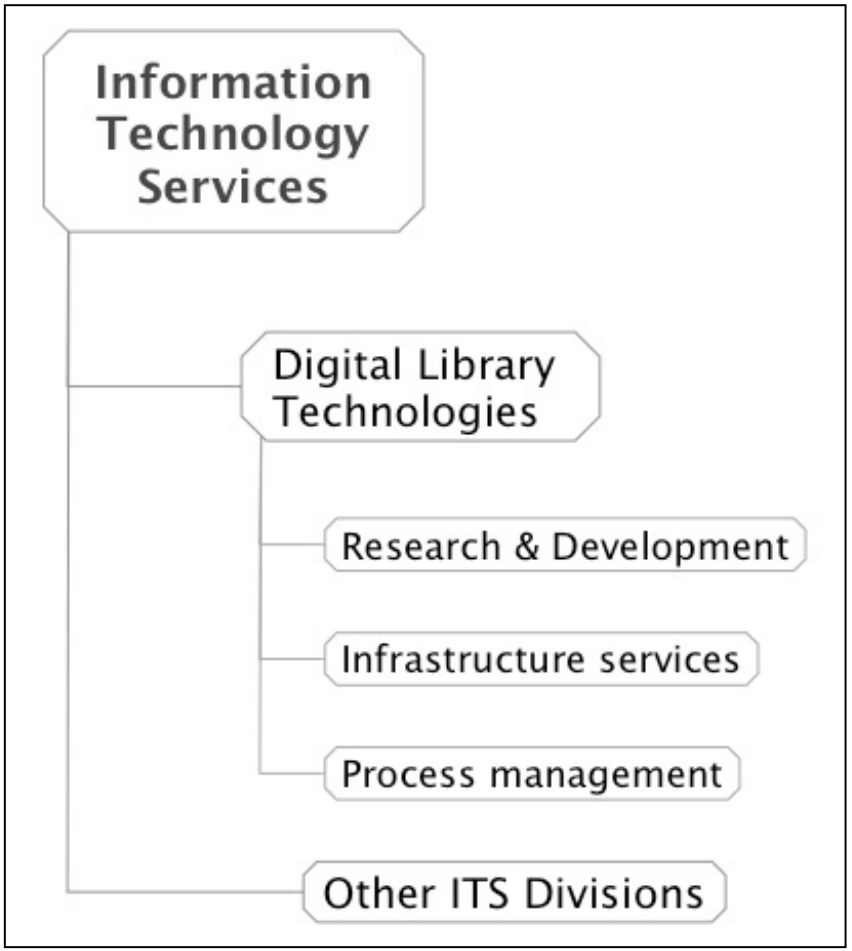
Building a digital curation community - 2

- Listserv – Digital Curation (Google Group)
  - http://groups.google.com/group/digital-curation
- Blog – Content Stewardship @ Penn State
  - http://stewardship.psu.edu/
- Facebook – CURATEcamp Group
- Conferences
  - *DLF Fall 2010 Forum, IDCC 2010*
  - *Code4Lib 2011*

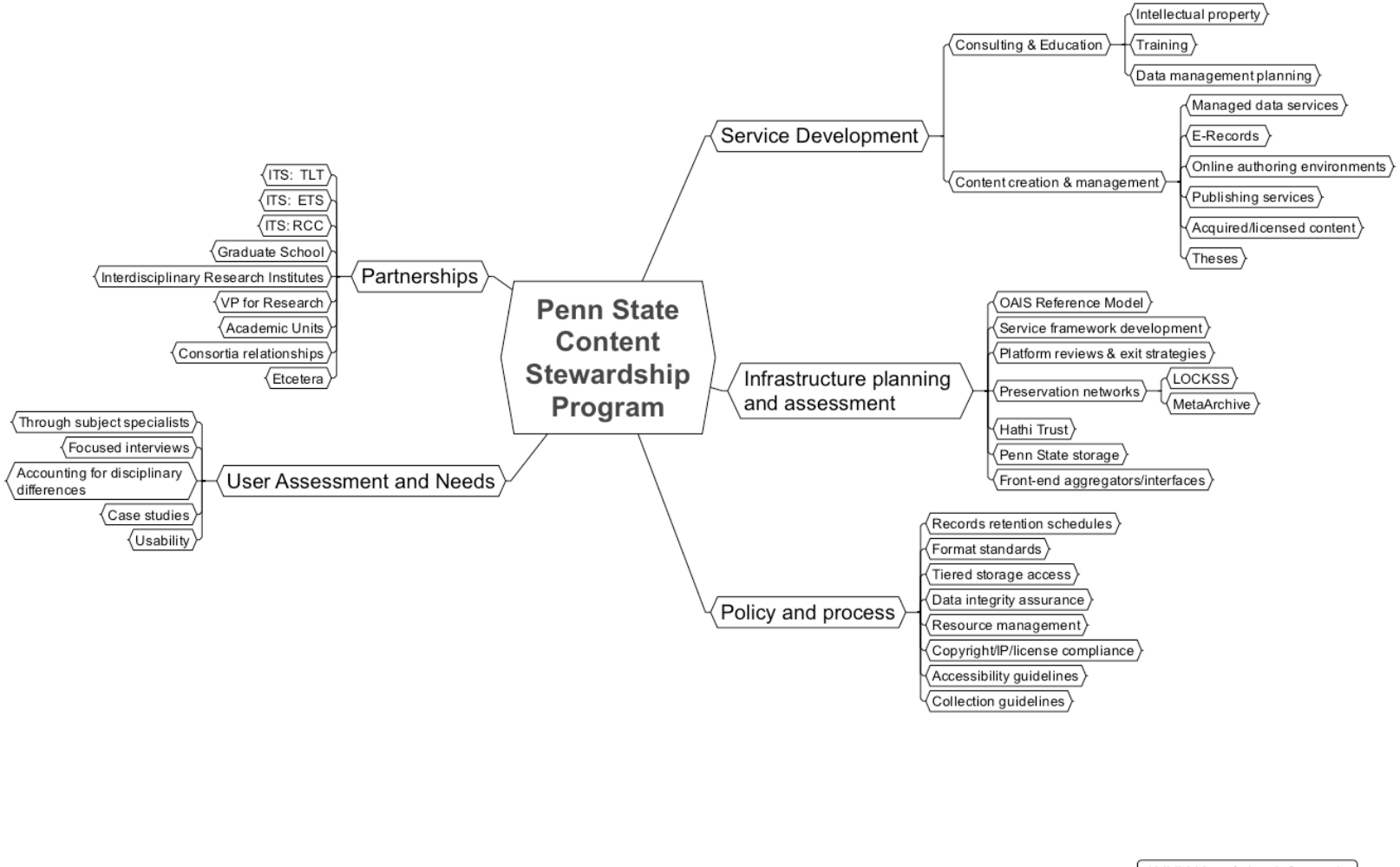
Organizational Contexts for Curation in Practice

Our organizational context

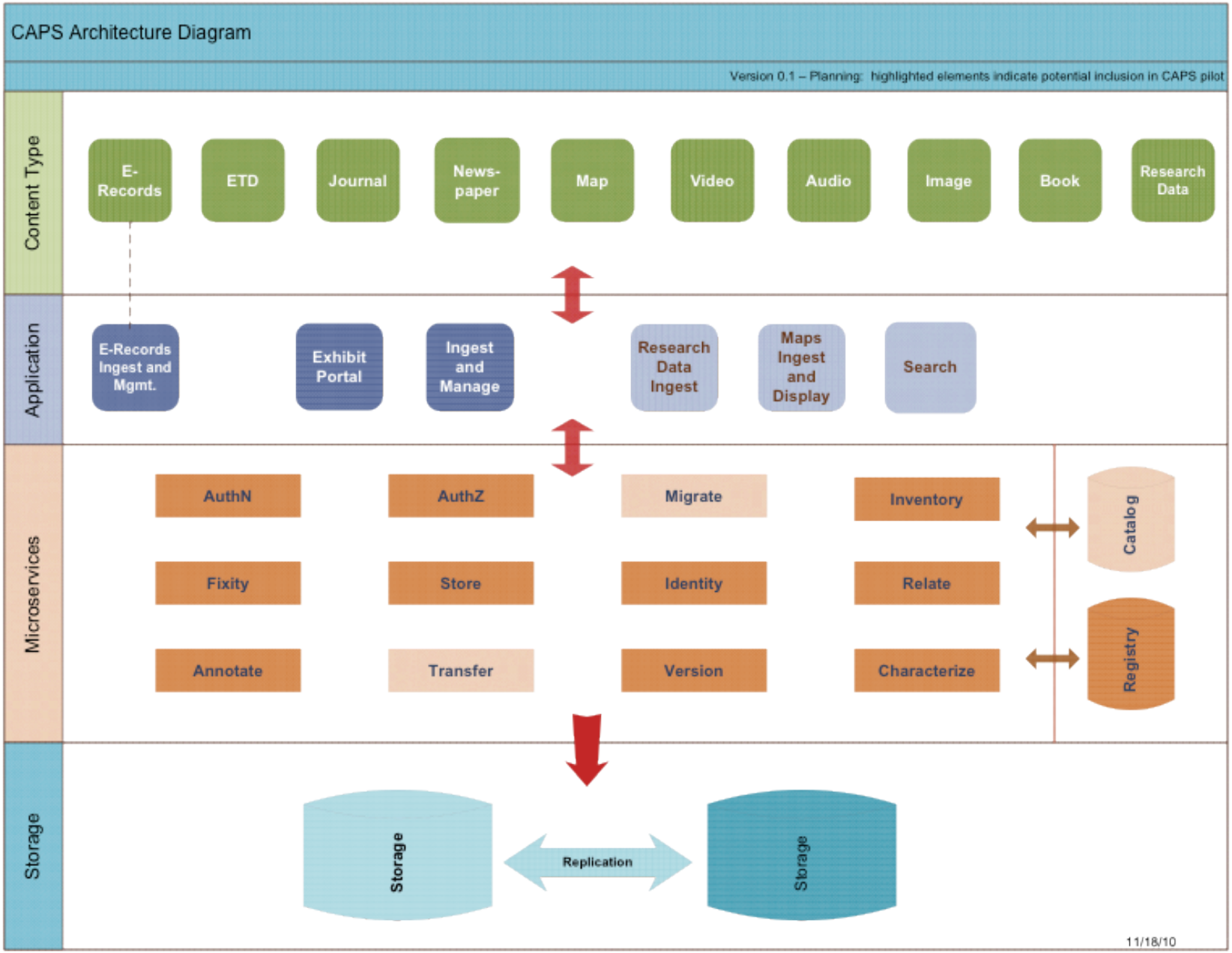




Curation in practice



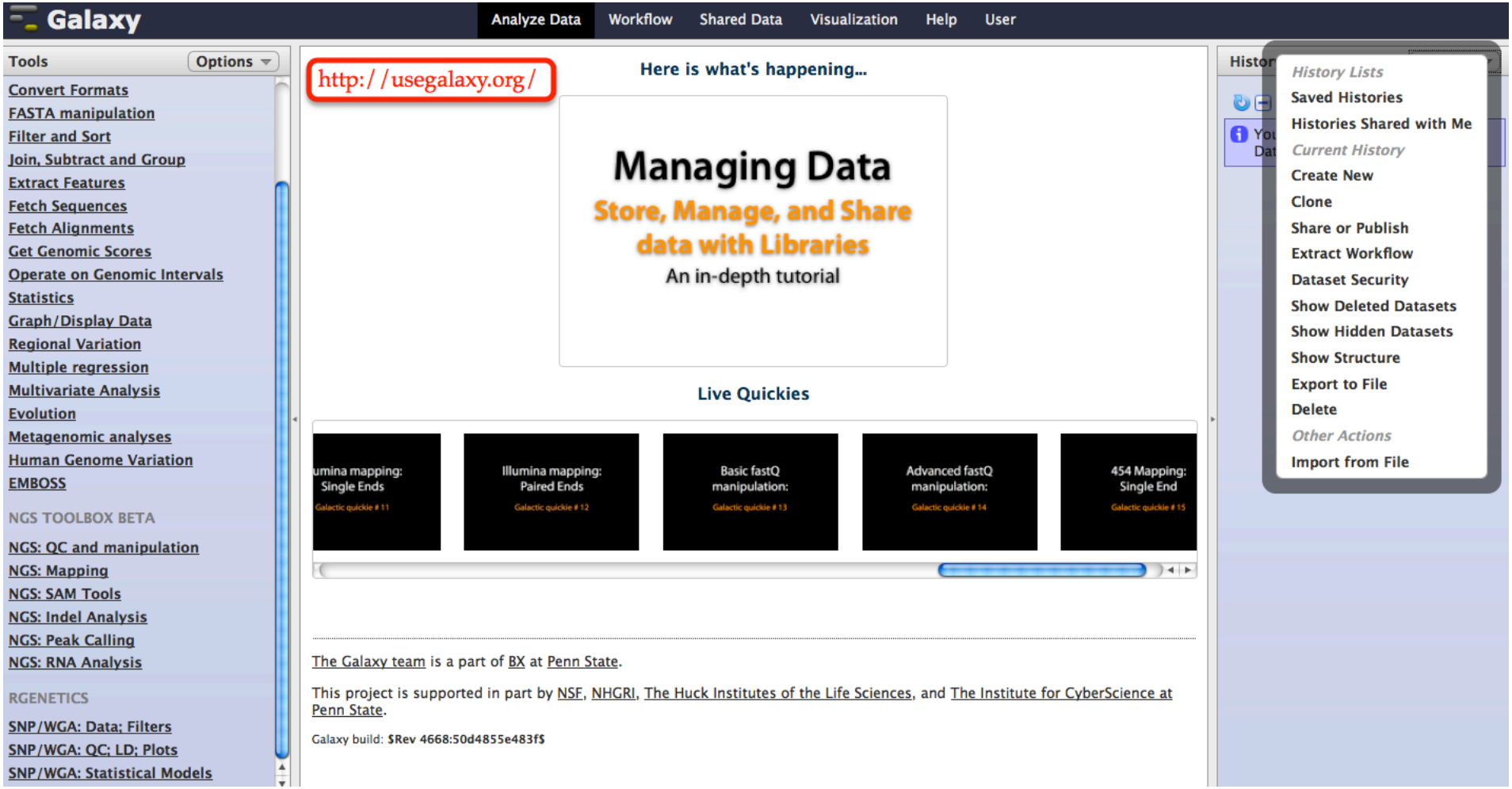
Current Status and Prospects


CAPS Architecture Diagram
Version 0.1 – Planning: highlighted elements indicate potential inclusion in CAPS pilot
Content Type
E-Records
ETD
Journal
News-paper
Map
Video
Audio
Image
Book
Research Data
Application
E-Records Ingest and Mgmt.
Exhibit Portal
Ingest and Manage
Research Data Ingest
Maps Ingest and Display
Search
Microservices
AuthN
AuthZ
Migrate
Inventory
Catalog
Fixity
Store
Identity
Relate
Registry
Annotate
Transfer
Version
Characterize
Storage
Storage
Replication
Storage
11/18/10

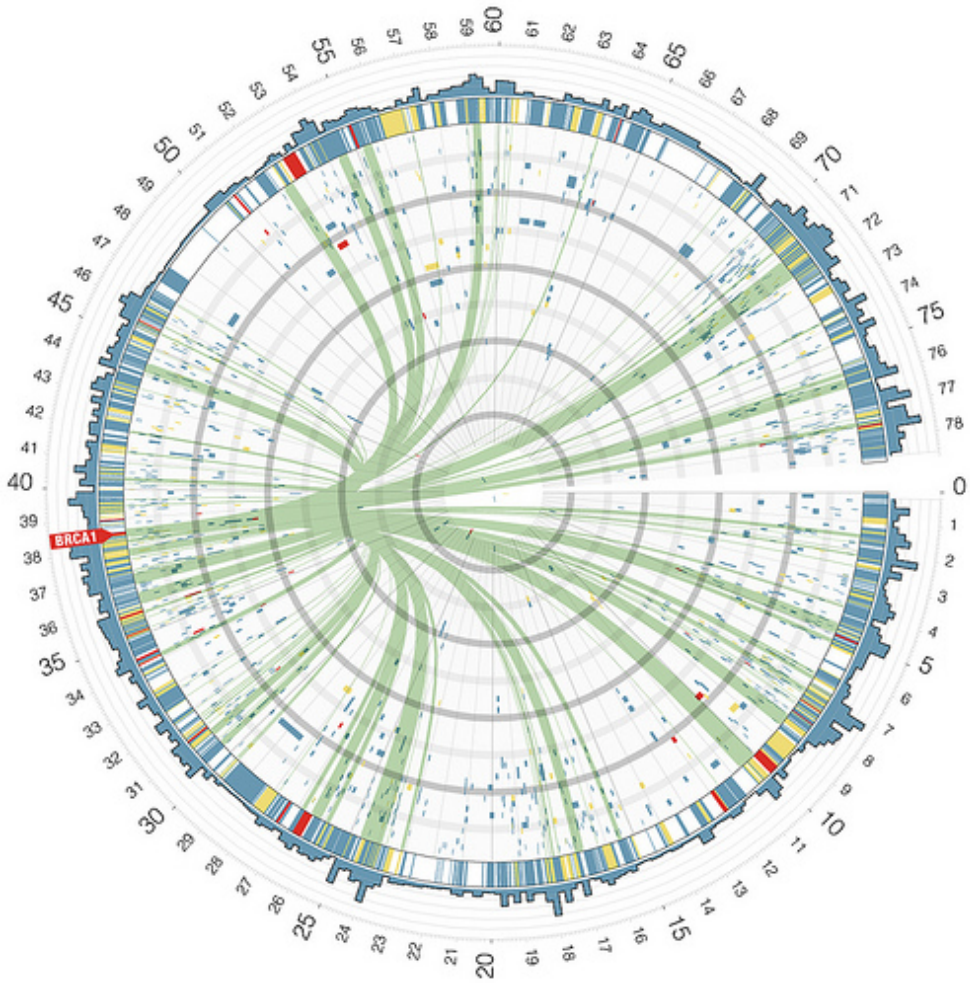
Microservices	DCC Lifecycle Actions
Authenticate	Access, Use & Reuse
Migrate	Preservation Action Store Transform Access, Use & Reuse
Fixity	Preservation Action
Store	Store (part of the Migration process)
Identity	Preservation Action Ingest
Annotate	Description and Representation Information
Version	Ingest Preservation Action Store Access, Use & Reuse
Characterize	Description and Representation Information
Authenticity/Audit	Ingest Preservation Action Store Access, Use & Reuse

These actions make up the Migration process in the DCC Lifecycle Model.


Log on
http://genomebiology.com/2010/11/8/R86/abstract
BioMed Central
Journals
Gateways
Genome Biology
IMPACT FACTOR 6.63
Search Genome Biology for
Go
Advanced Search
Home
Articles
Authors
Reviewers
About this journal
My Genome Biology
Subscriptions
Software
Highly accessed
Open access
Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences
Jeremy Goecks[1], Anton Nekrutenko[2]*, James Taylor[1]* and The Galaxy Team
* Correponding authors: Anton Nekrutenko anton@bx.psu.edu - James Taylor james.taylor@emory.edu
Author Affiliations
For all author emails, please log on.
Genome Biology 2010, 11:R86 doi:10.1186/gb-2010-11-8-r86
Published: 25 August 2010
Abstract
Increased reliance on computational approaches in the life sciences has revealed grave concerns about how accessible and reproducible computation-reliant results truly are. Galaxy http://usegalaxy.org webcite, an open web-based platform for genomic research, addresses these problems. Galaxy automatically tracks and manages data provenance and provides support for capturing the context and intent of computational methods. Galaxy Pages are interactive, web-based documents that provide users with a medium to communicate a complete computational analysis.
Genome Biology
Volume 11
Issue 8
Viewing options
Abstract
Full text
PDF (2.6MB)
Associated material
PubMed record
About this article
Readers' comments
Related literature
Articles citing this article
on Google Scholar
on BioMed Central
on PubMed Central
Other articles by authors
on Google Scholar
on arXiv
on PubMed


Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Tools
Options
Convert Formats
FASTA manipulation
Filter and Sort
Join, Subtract and Group
Extract Features
Fetch Sequences
Fetch Alignments
Get Genomic Scores
Operate on Genomic Intervals
Statistics
Graph/Display Data
Regional Variation
Multiple regression
Multivariate Analysis
Evolution
Metagenomic analyses
Human Genome Variation
EMBOSS
NGS TOOLBOX BETA
NGS: QC and manipulation
NGS: Mapping
NGS: SAM Tools
NGS: Indel Analysis
NGS: Peak Calling
NGS: RNA Analysis
RGENETICS
SNP/WGA: Data; Filters
SNP/WGA: QC; LD; Plots
SNP/WGA: Statistical Models
http://usegalaxy.org/
Here is what's happening...
Managing Data
Store, Manage, and Share data with Libraries
An in-depth tutorial
Live Quickies
umina mapping: Single Ends
Galactic quickie # 11
Illumina mapping: Paired Ends
Galactic quickie # 12
Basic fastQ manipulation:
Galactic quickie # 13
Advanced fastQ manipulation:
Galactic quickie # 14
454 Mapping: Single End
Galactic quickie # 15
The Galaxy team is a part of BX at Penn State.
This project is supported in part by NSF, NHGRI, The Huck Institutes of the Life Sciences, and The Institute for CyberScience at Penn State.
Galaxy build: $Rev 4668:50d4855e483f$
History Lists
Saved Histories
Histories Shared with Me
Current History
Create New
Clone
Share or Publish
Extract Workflow
Dataset Security
Show Deleted Datasets
Show Hidden Datasets
Show Structure
Export to File
Delete
Other Actions
Import from File

Full-circle: doing scientific data management



References

- Slides 4 & 26 - http://www.flickr.com/photos/ethanhein/2272885283/
- Slide 6 - http://www.flickr.com/photos/library_of_congress/2178285893/
- Slide 8 - http://www.flickr.com/photos/carowallis1/2373804335/
- Slide 9 & 22 - http://www.dcc.ac.uk/resources/curation-lifecycle-model
- Slide 24 - http://genomebiology.com/2010/11/8/R86/abstract
- Slide 25 - http://usegalaxy.org/
- For more about Nekrutenko Lab at Penn State: http://www.bx.psu.edu/~anton/, http://www.rps.psu.edu/indepth/galaxy.html