COLUMBIA UNIVERSITY
LIBRARIES

Digital Projects/Public Viewer
Digital Preservation Storage Update
National Preservation Projects

Robert Cartolano, Stephen Davis, Jim Neal

April 23, 2014

Theme: From Local to National

- Local Effort
  - Digital Collections
  - Public Digital Collections Viewer (DCV)
- Local and National
  - Digital Preservation Storage
  - Sustainable Funding
  - Sustainable Technology
- National Preservation Strategy

Digital Collections Projects

Columbia University Libraries / Information Services has made commitments …

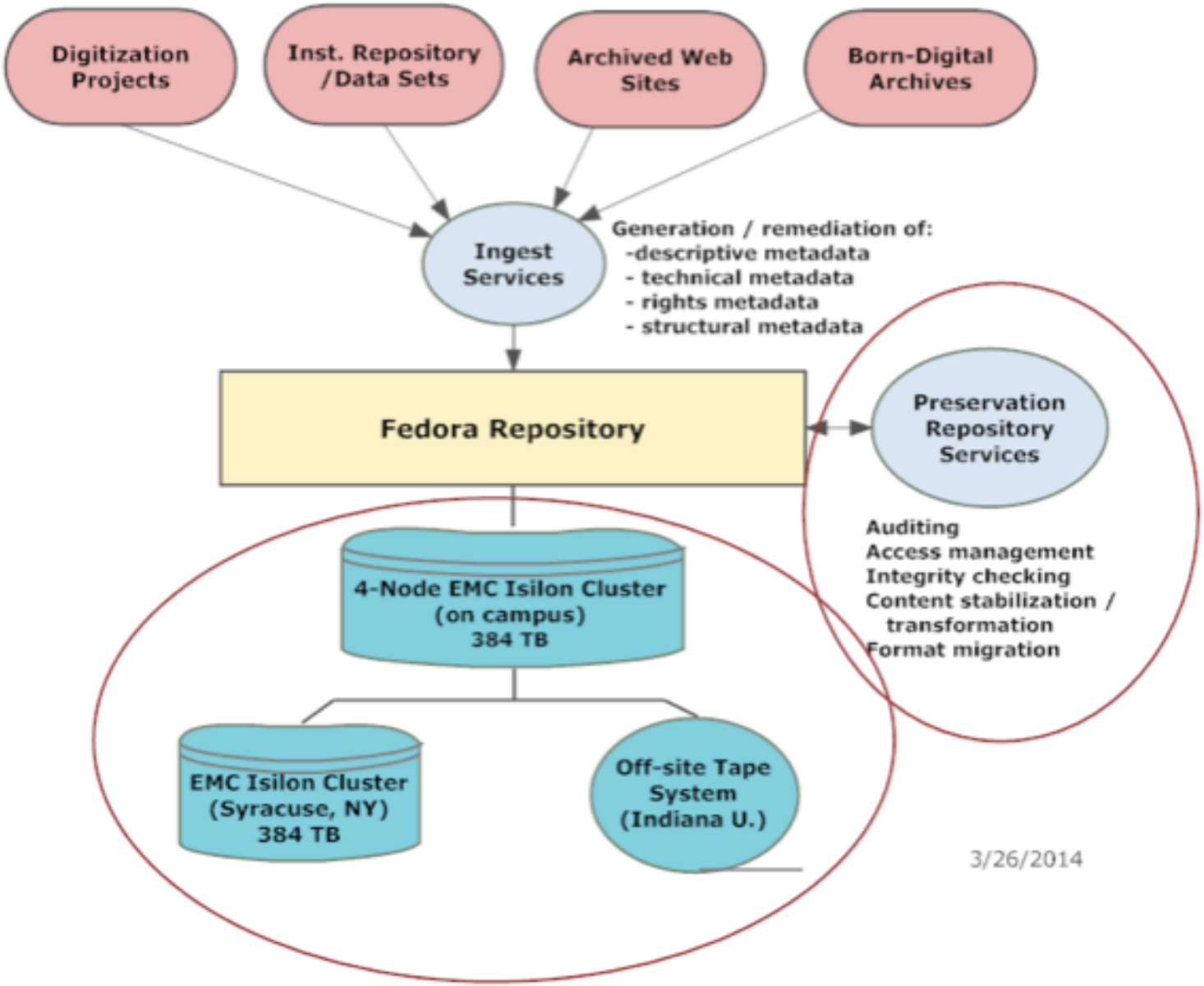
- to granting agencies to provide long-term digital archiving for digital content created with *grant funds*
- to third-party content creators to provide permanent access to *born-digital content* acquired from them
- to continuing to collect and preserve *archival collections*, now partly or wholly born-digital content
- to permanently preserve *University-generated archival and research content*

Digital Collections

We must preserve and provide access to …

- Local Digitization Projects
- Preservation-Related Digitization
- Institutional Repository / Data Sets
- Born Digital Archival Content
- Archived Web Sites
- Super Dark Archives – *highly secure*

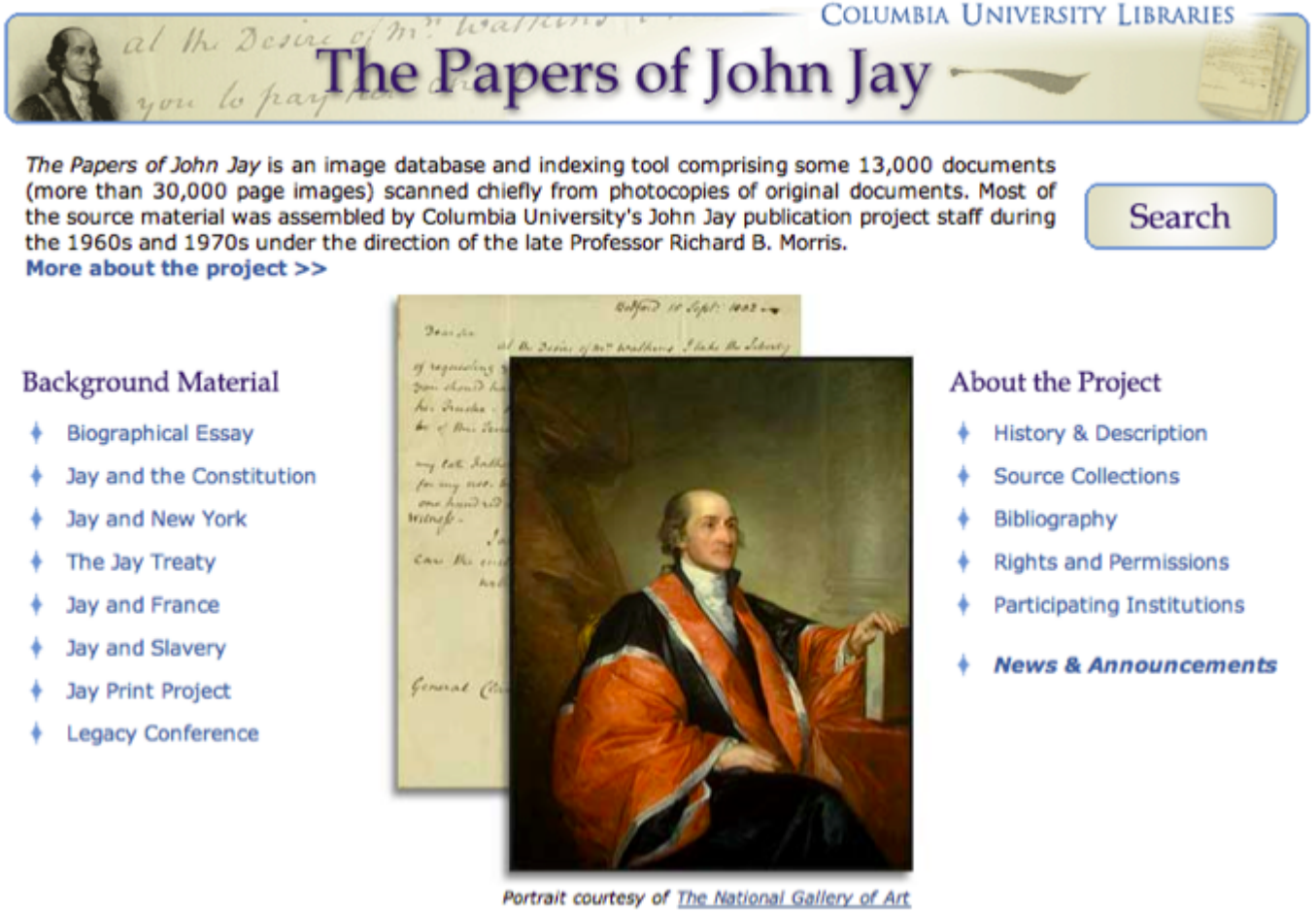
Current Digital Asset/Preservation Architecture



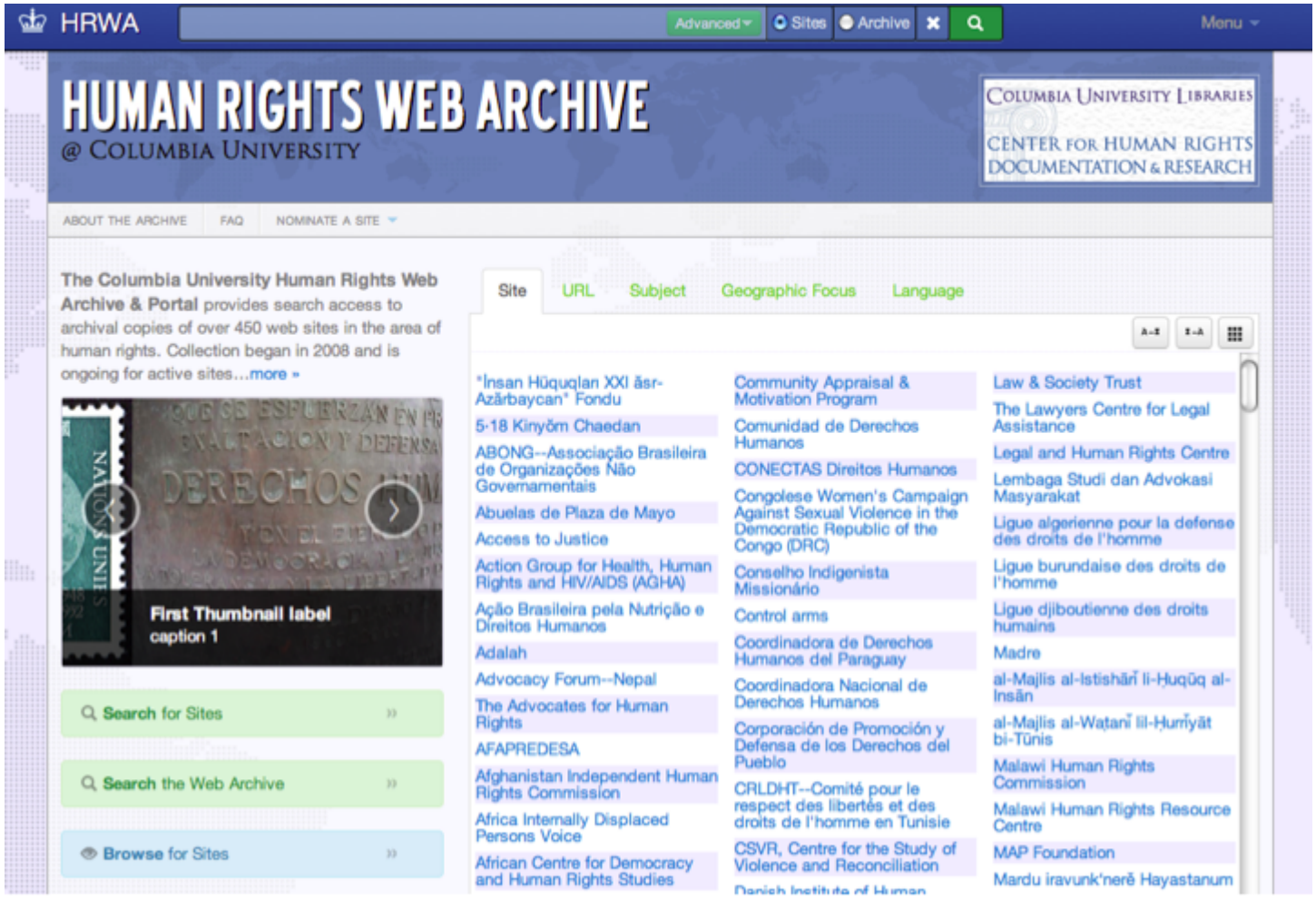
Example: Digital Collection Project

Academic Commons

Older Projects - Data Migration, Refresh



Web Archiving



Public Digital Collections Viewer (DCV)

- A system to provide access to traditional cultural heritage digital collections, audio/video collections, and born digital archival collections
- A platform for integrating existing collections and publishing new collections
- A set of tools for basic viewing / playing content in most media formats
- A framework for bringing to the surface relationships between content in Columbia's collections and content in other related collections at other locations

Who is making this happen?

- Advisory Group: Jane Gorjevsky, Chris Sala, Kate Harcourt, Melanie Wacker, Robbie Blitz, Matthew Baker, Chengzhi Wang, Carolyn Yerkes
- Project Lead: Stephen Davis
- Project Management Support: Leo Stezano
- Technical Staff: Eric O'Hanlon, Ben Armintor, Erik Ryerson

When will this happen?

Phase I – Tentative Timetable

- April-May 2014: Internal prototypes
- May 2014: CUL internal beta release
- June-August 2014: Further development, additional internal releases; user testing
- September 2014: Public beta release
- October-November: User testing, further development
- Winter 2014/2015: Public launch

What will Phase 1 content include?

200,000+ Digital Objects

- https://wiki.cul.columbia.edu/display/PDCV/DCV+Potential+Digital+Content (Potential content)

30+ Digital Collections

- https://wiki.cul.columbia.edu/display/PDCV/Public+Digital+Collections+Viewer+Home (Project wiki with more info and links)

What will it not include?

- Web archives
- Finding aids
  (but finding aids can link to the DCV)
- Collections / content with no ‘suitable’ metadata

What features will be in Phase I?

- Browsable list of collections / projects
- Keyword searching within individual projects or across all projects
- Standard facets across all collections: names, material types, collection names, repositories
- Custom facets where needed: e.g., geographic or chronological content
- Flexible displays for images; zoomable, multiresolution image functionality

Features in Phase I, cont.

- Persistent identifiers
- Function to generate a standard citation for individual objects
- Session-specific “bookbags” / “shopping carts”
- Search engine optimization features
- Full-text searching where supported by the data (e. g., IFP, University Seminars, Harrison)
- Ability to report errors
- Ability to correct errors (in Hyacinth metadata editor)

What about restricted content?

There will actually be two separate digital collection viewers:

- One available publicly, world-readable
- One restricted to on-site access in a specific custodial location (for example, for parts of archival collections, for content that is under copyright but can be made available locally as preservation copies)

Having two separate, parallel viewers will help ensure the privacy and security of restricted content.

DCV Benefits

- Supports three CUL/IS strategic goals
  - Develop Global and Special Collections
  - Digital Collections
  - Effective Interfaces and Improved Access
- Provides integrated access to CUL specialized and archival digital collections
- Aggregates search and discovery
- Reduces need for multiple separate systems
- Enables potential “bento box” discovery via CLIO
- Facilitates making CUL content available externally, e. g., to DPLA, through APIs, consistent SEO
- Ease of improvements, upgrades, sustainability

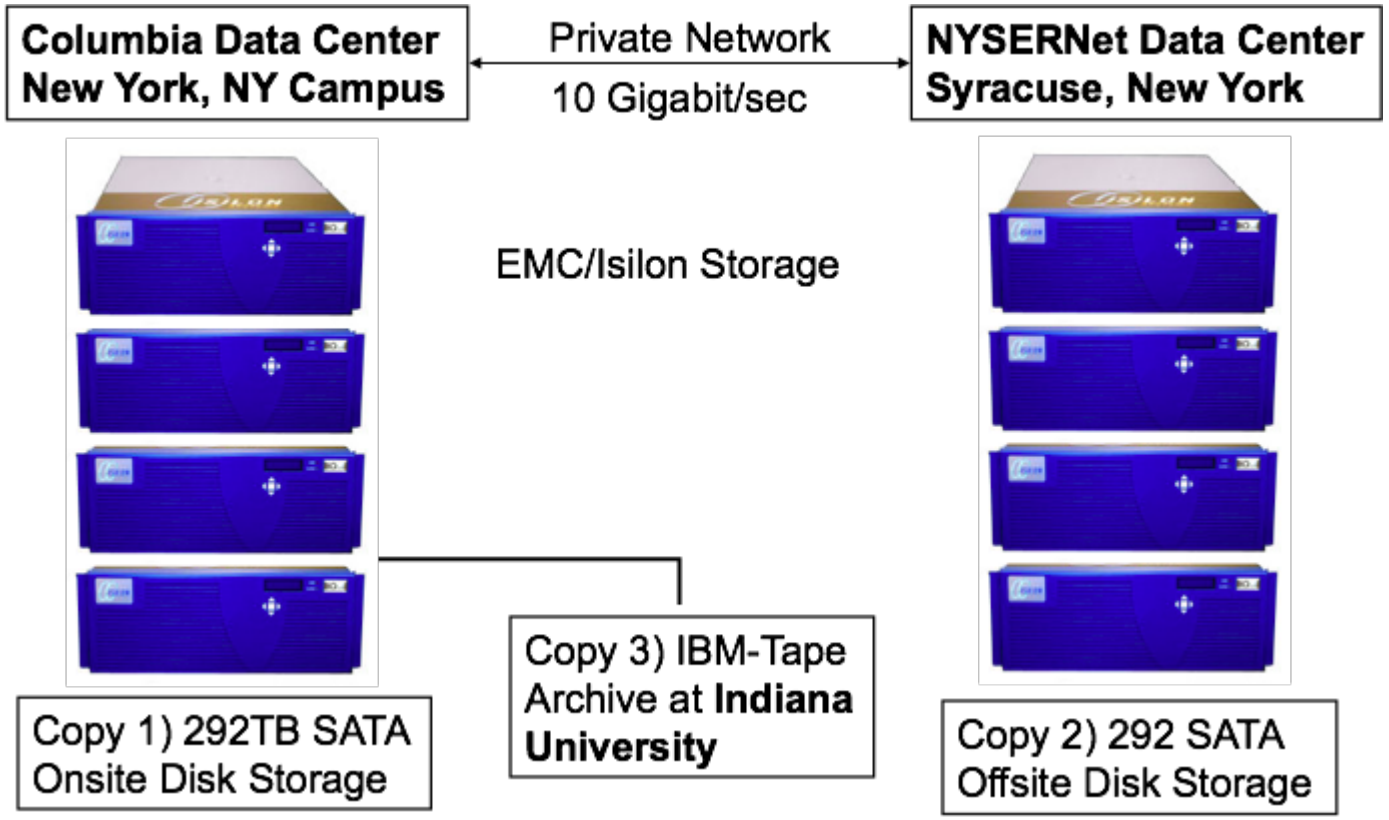
Local Digital Preservation Storage

- Goals
  - Risk-averse technology approach
  - Different storage for different needs
  - Move towards using national systems
- Local Preservation - 2013 thru 2017
  - 300 Terabytes total capacity
  - ca. 140 Terabytes of current, committed projects
- Beyond 2017
  - Local systems - primarily for services, performance
  - National systems - primarily for certified preservation storage (TRAC)

Columbia Preservation Storage

2007 – Purchased and built by Columbia

2013 – Mostly at Columbia, some at Indiana



National Preservation Strategy

- Digital Preservation Network
  - Strategy and implementation plan for national preservation infrastructure
  - 65 member institutions
  - Coordinated with AAU Presidents, Internet 2
- Five DPN Nodes
  - APTrust (University of Virginia) - member
  - Chronopolis
  - HathiTrust (University of Michigan) - member
  - Stanford Digital Repository
  - University of Texas Digital Repository

http://www.dpn.org/

National Preservation Strategy (cont'd)

- DuraSpace
  - Infrastructure for durable, persistent access to data
  - Sustainable software, communities, funding
  - 90+ sponsors

http://duraspace.org/

- Hydra Partners
  - Building sustainable software tools
  - 20+ partner institutions

http://projecthydra.org/

- Research Data Alliance
  - Enable open sharing of data
  - 35+ countries at 3rd plenary conference

https://rd-alliance.org/

Academic Preservation Trust (APTrust)

- Led by University of Virginia, 17 partner institutions
- APTrust is working toward TRAC certification
- First ever repository to be built from the ground up taking TRAC into account.
- Certification Working Group established and will be advising and consulting with the APTrust staff and partners on TRAC objectives.
- Initial development work is proceeding at the level of Digital Object Management and Infrastructure.

http://aptrust.org/

Examples of TRAC Requirements

- “The repository shall have an appropriate succession plan, contingency plans, and/or escrow arrangements in place in case the repository ceases to operate or the governing or funding institution substantially changes its scope.”
- “The repository shall have short- and long-term business planning processes in place to sustain the repository over time.”
- “The repository shall have contracts or deposit agreements which specify and transfer all necessary preservation rights, and those rights transferred shall be documented.”
- “The repository shall have the appropriate number of staff to support all functions and services.”
- “The repository shall have and use a convention that generates persistent, unique identifiers.”

Emerging National Digital Preservation Strategy

- Institution-level funding required
- Local efforts integrated with national systems
- Succession rights planning
- Governance
  - HathiTrust - Bob Wolven
  - DPN - Jim Neal - DPN

Questions