

NDSR

BOSTO

NDSA NE - October 30, 2014

University of Massachusetts, Amherst Libraries

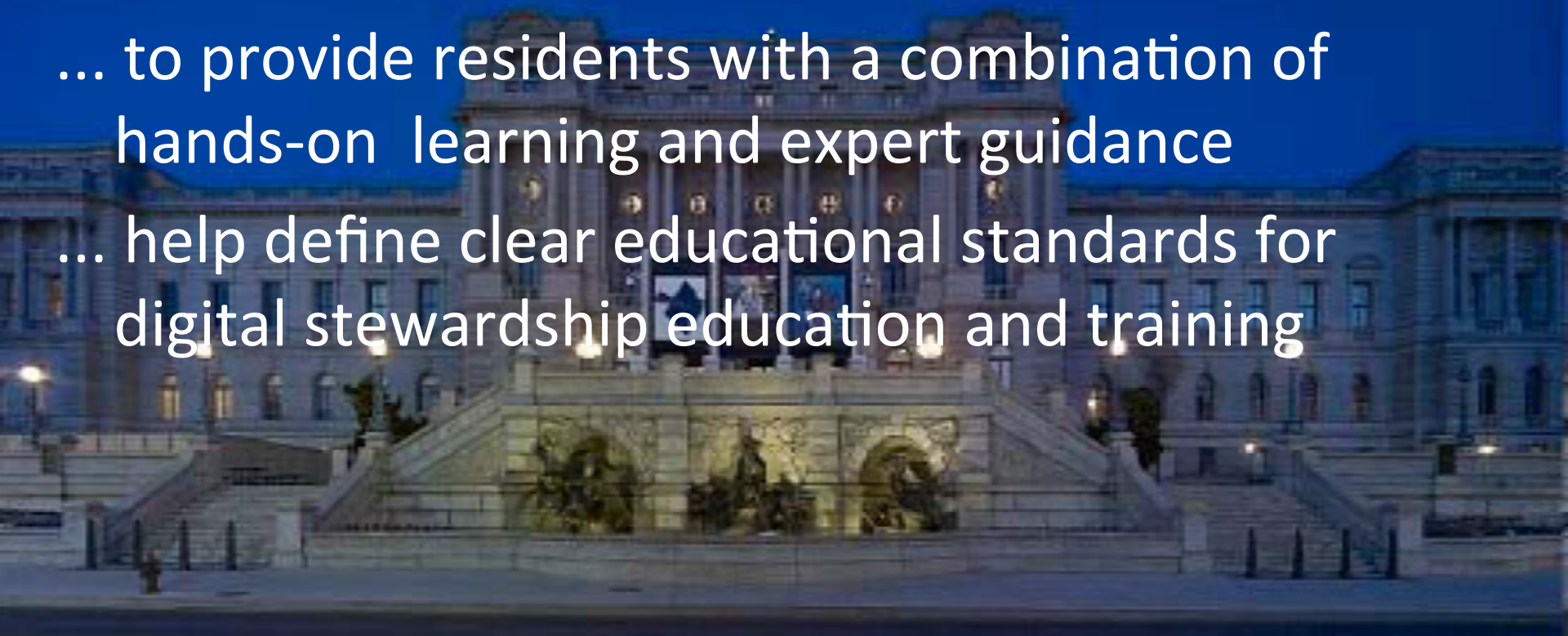
Andrea Goethals, Harvard Library

Overview & Updates

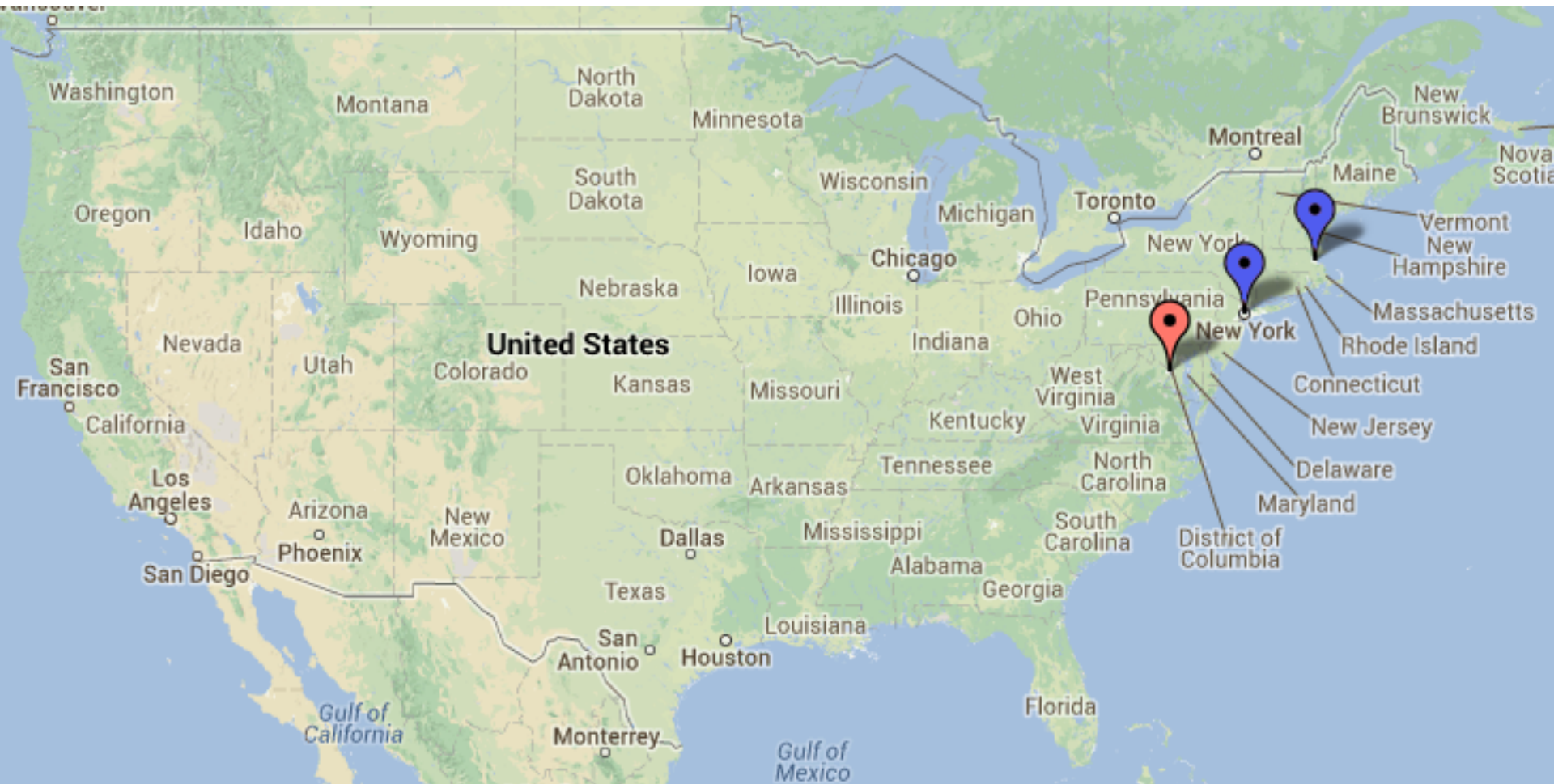


NDSR mission

- ... to develop the next generation of stewards to collect, manage, preserve, and make accessible our nation's digital assets
- ... to provide residents with a combination of hands-on learning and expert guidance
- ... help define clear educational standards for digital stewardship education and training



Replicating & testing the model

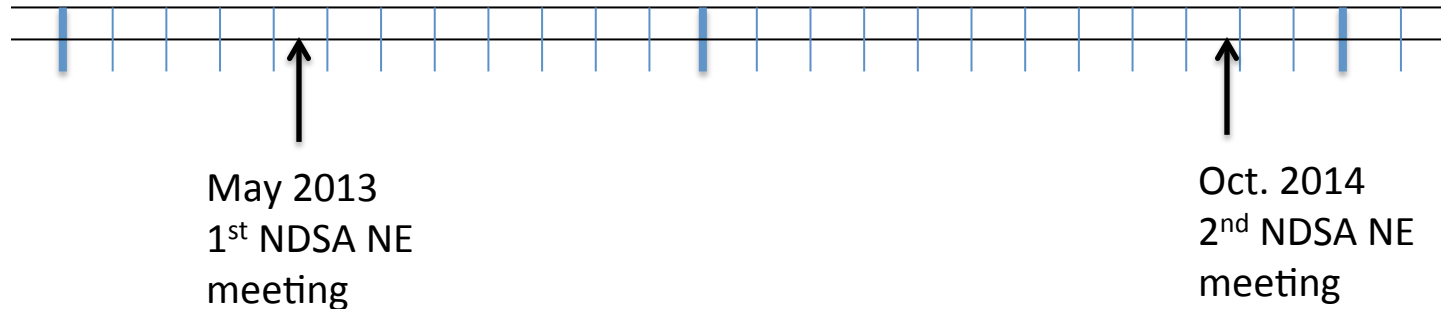


IMLS-funded grants to test model in New York (METRO) and Boston areas (Harvard / MIT)

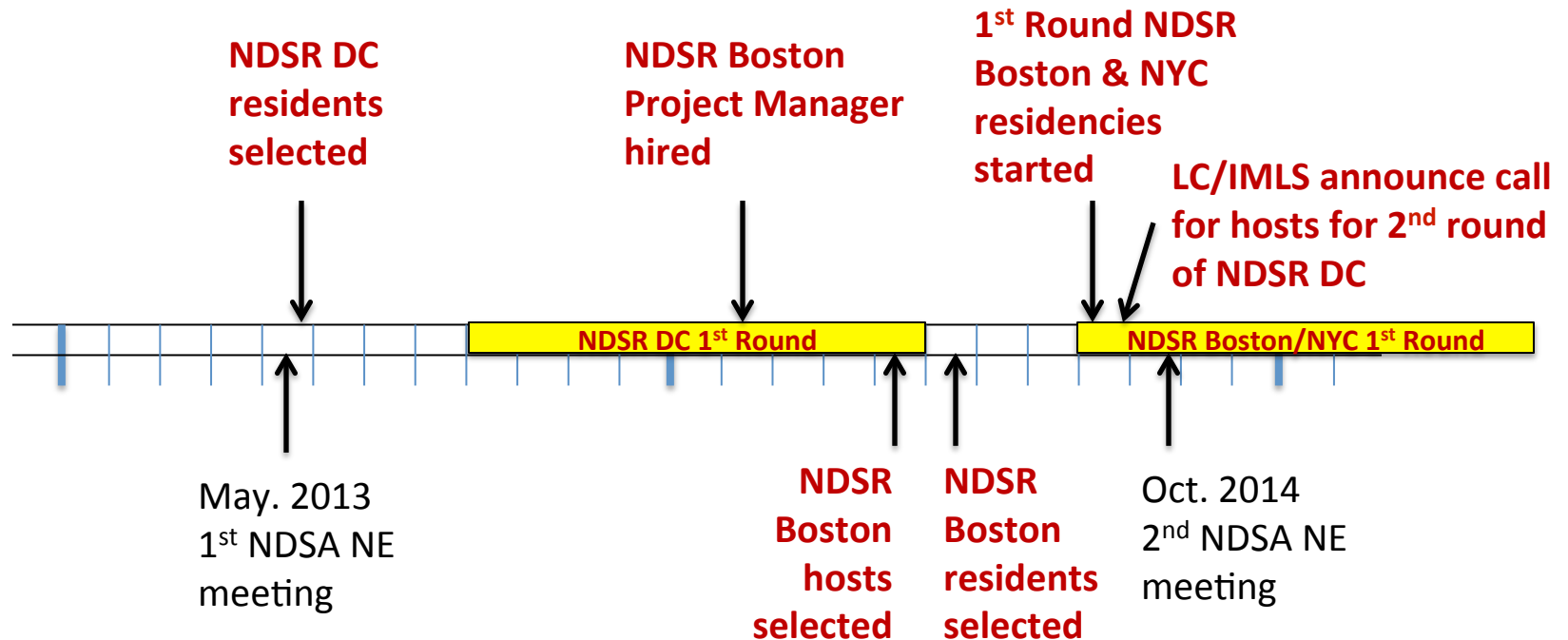
Replicating what exactly?

- Residency structure
- Resident & host cohort model
- Resident eligibility & application requirements
- Single institution administering program
- Distributed hosts within the same metropolitan area
- Core curriculum

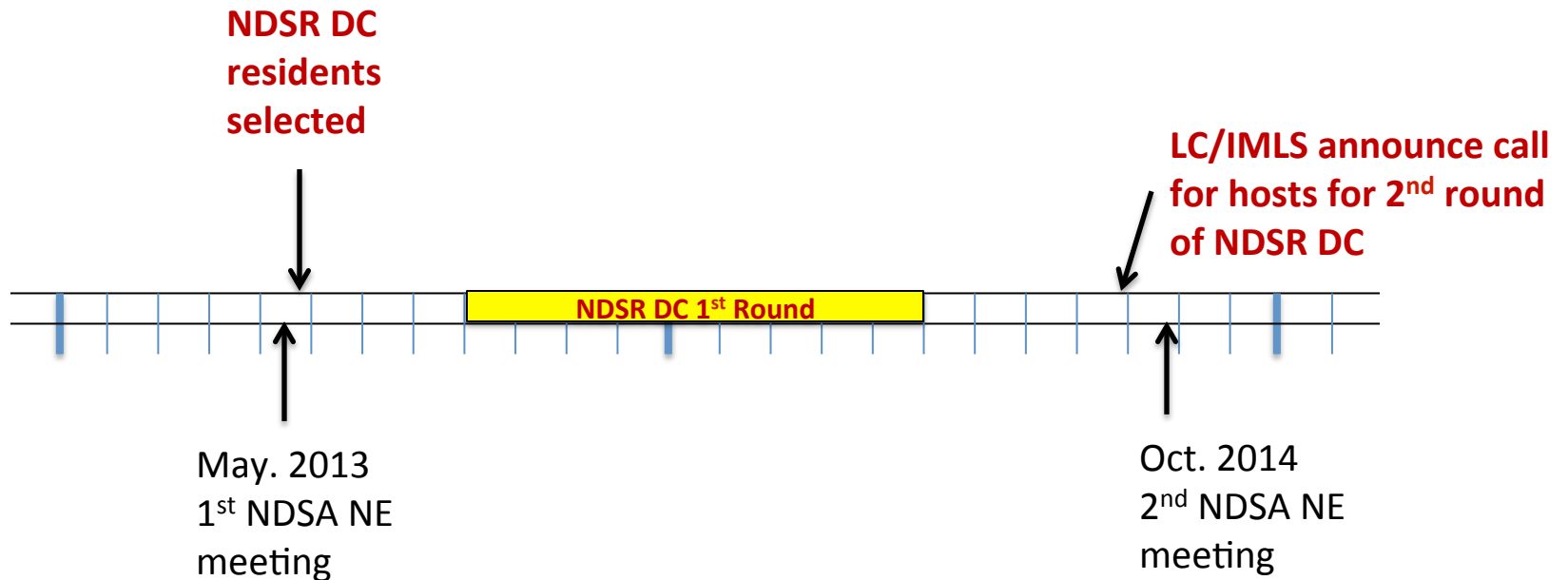
Since we last met...



A lot has happened!



In DC



DC pilot: positive benefits

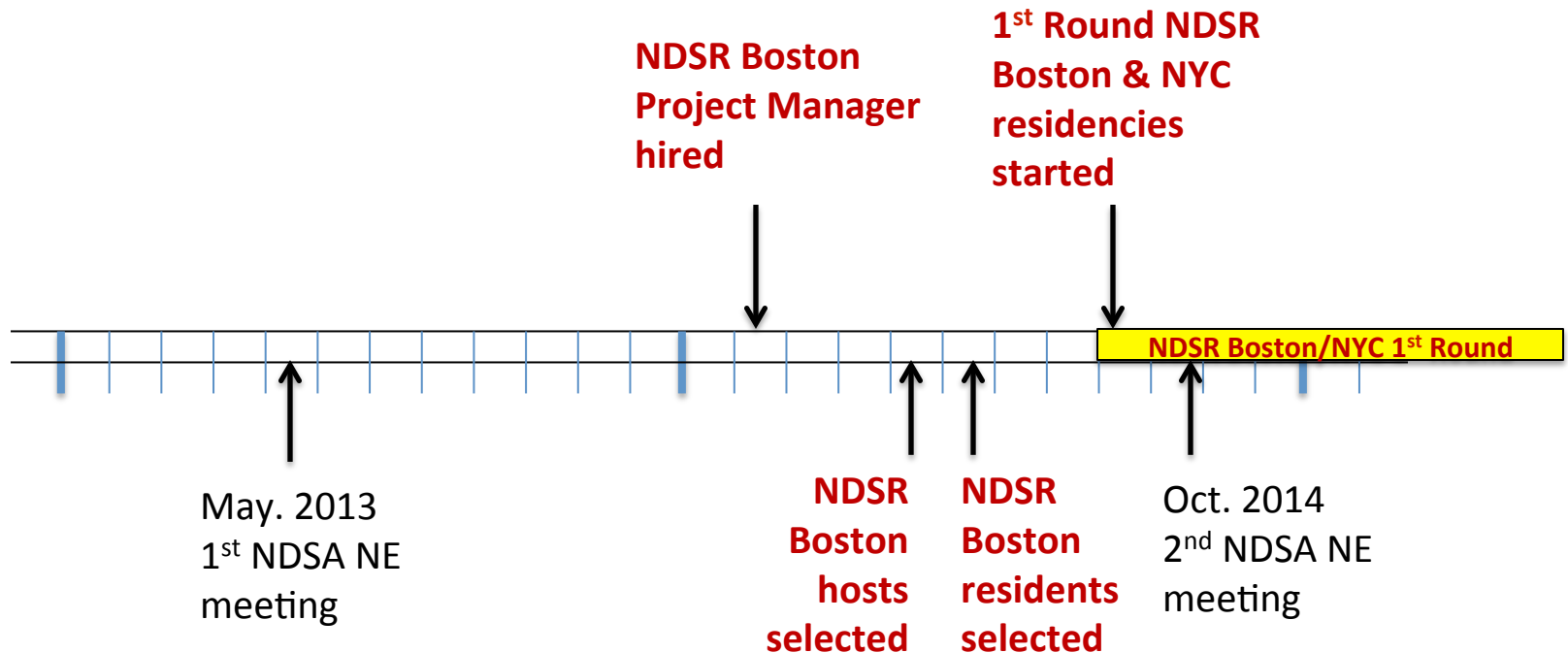
100% of the residents said the program:

- Increased their knowledge and skills in digital stewardship
- Strengthened their professional networks
- Made them more competitive in the job market

DC pilot: areas to improve

- Program logistics (communication, management, etc.)
- Host engagement
- Curriculum beyond the immersion week
- Building a national program out of the regional NDSR programs

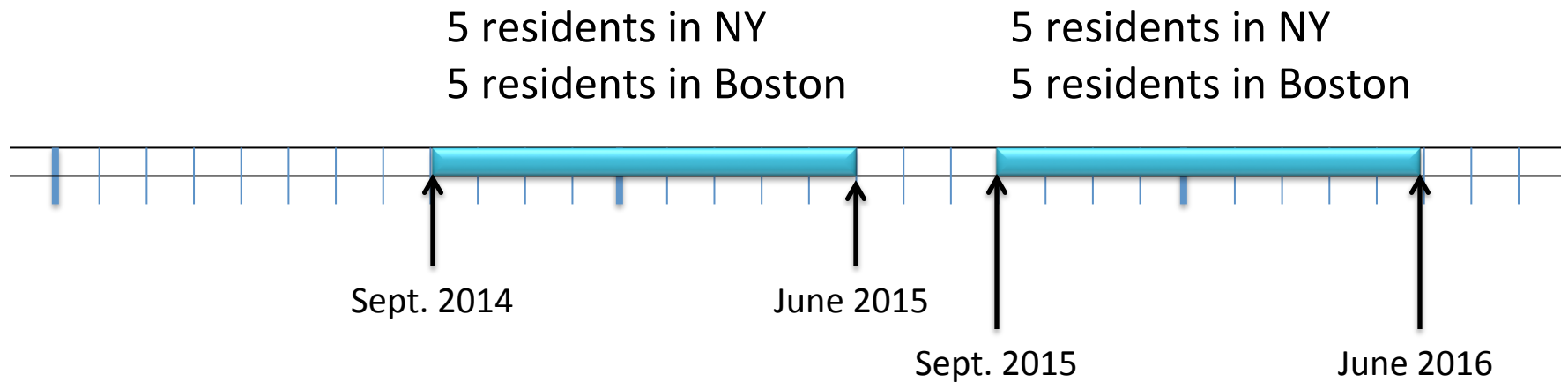
In Boston (and NYC)



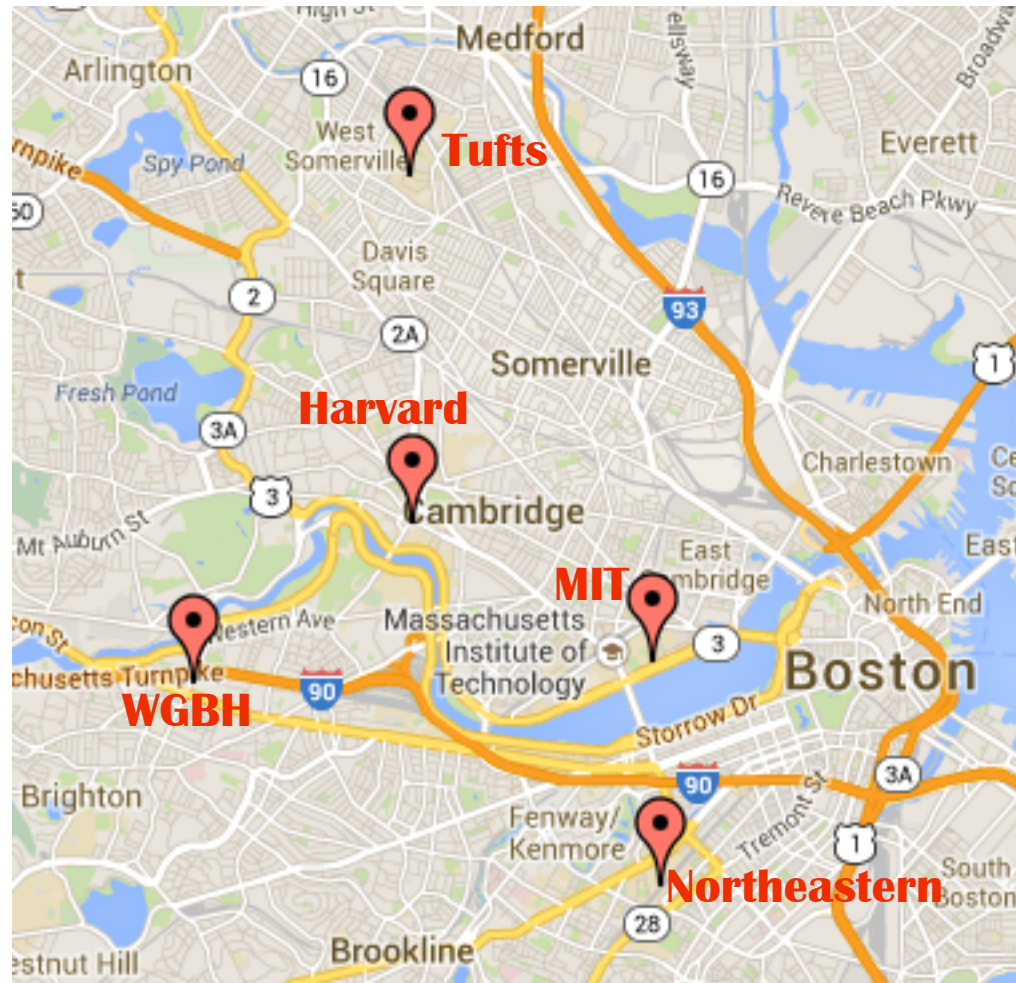
Boston model refinement

- Extended curriculum
- Formal development plans, residency requirements, progress tracking
- More host involvement
- Preservation-specific focus for resident projects
- Additional roles for instructors and community

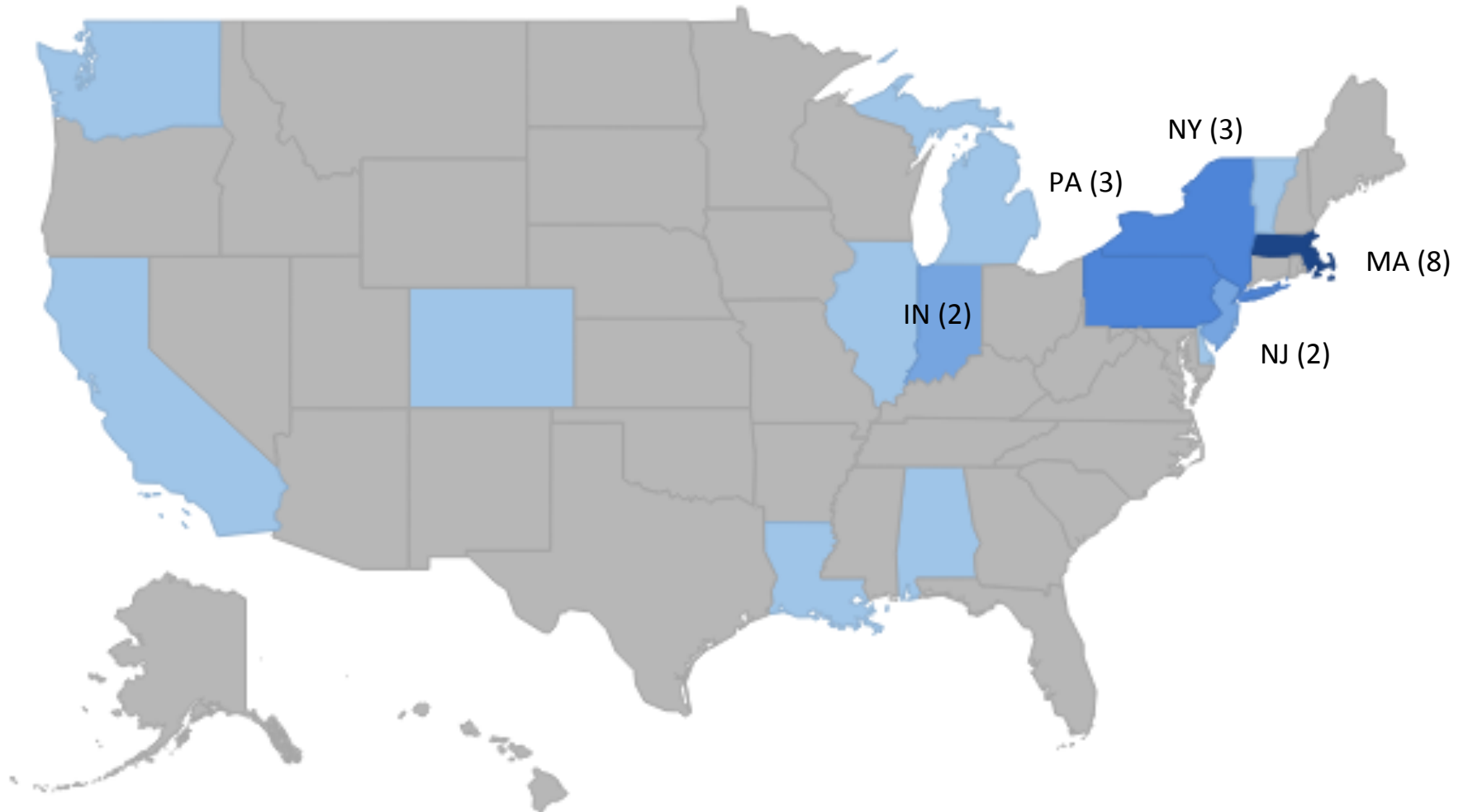
Timeline



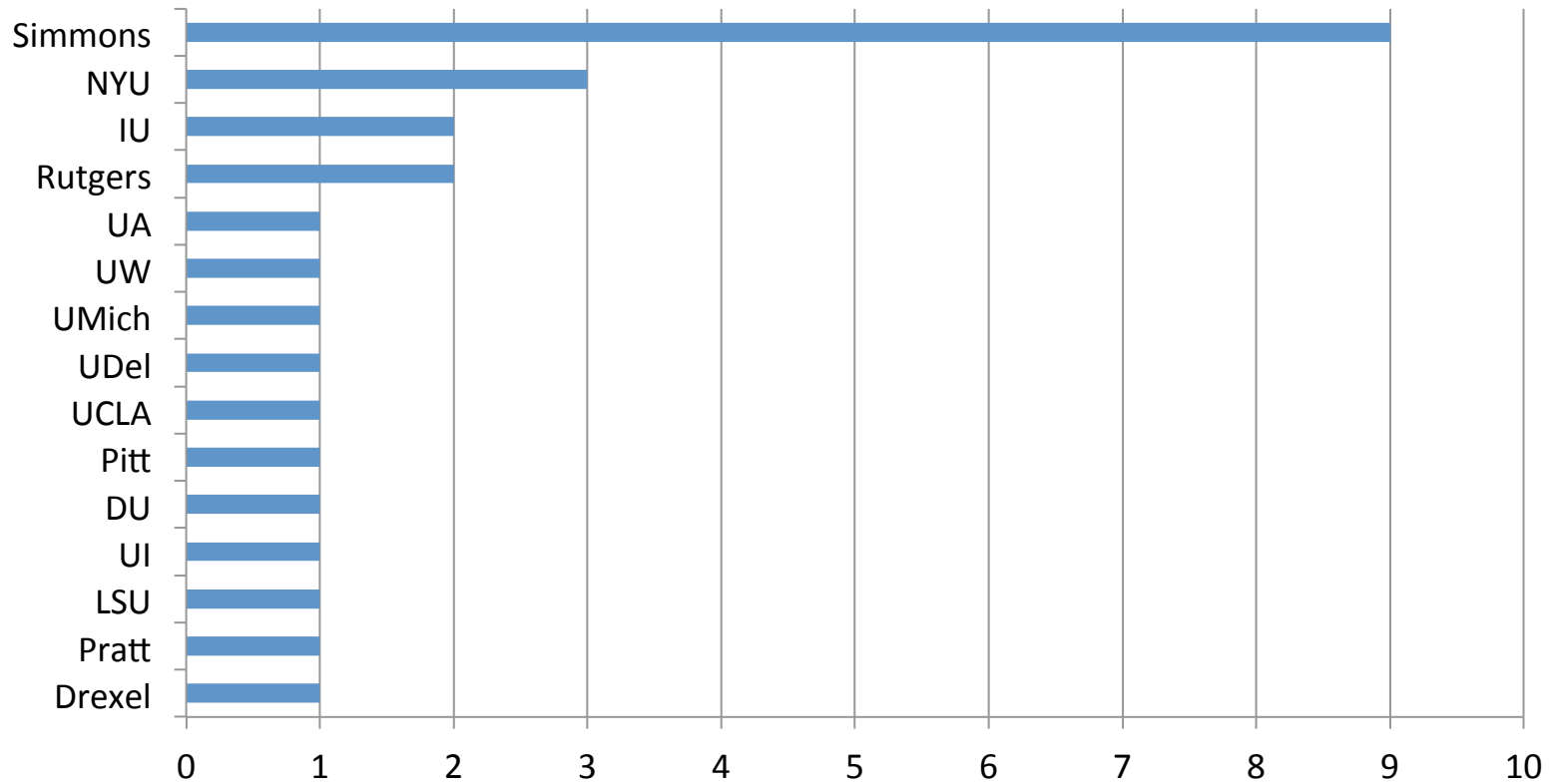
Boston hosts 2014-15



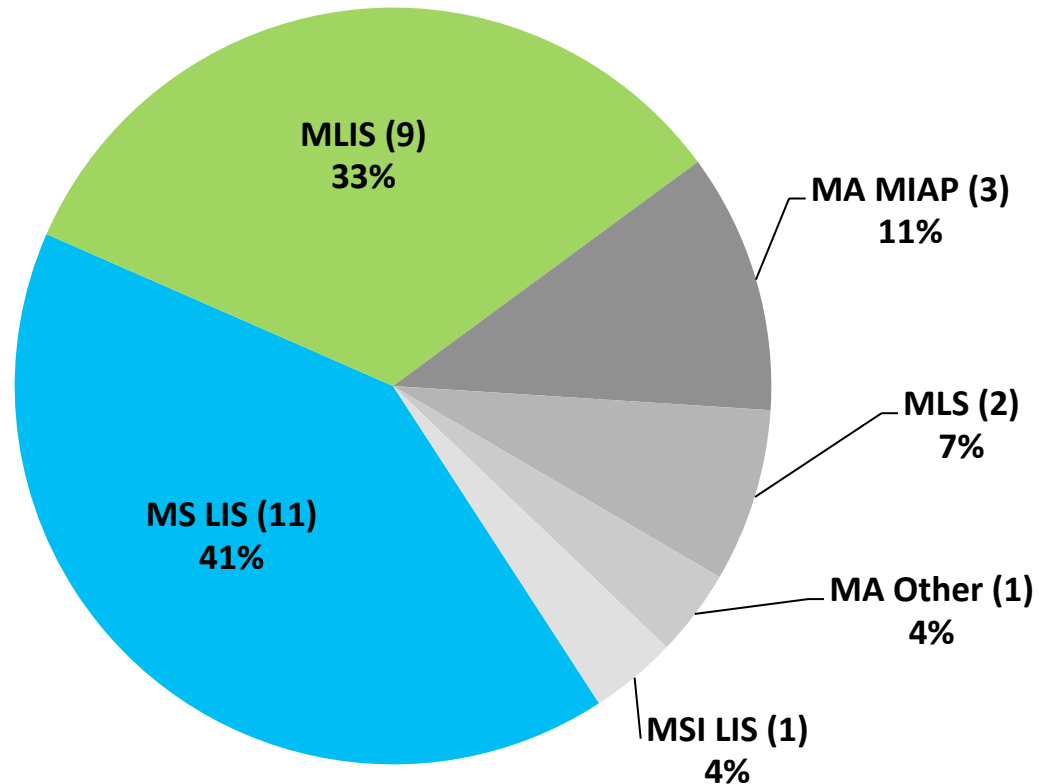
Resident applicants - state



Boston applicants - school



Boston applicants - degrees



Boston Residents 2014-15



UNIVERSITY of
DENVER



Jen LaBarbera
Host: Northeastern
MLIS



Tricia Patterson
Host: MIT
MS LIS



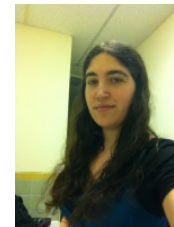
Samantha DeWitt
Host: Tufts
MS LIS



NEW YORK UNIVERSITY



Joey Heinen
Host: Harvard
MA MIAP



Rebecca Fraimow
Host: WGBH
MA MIAP

Nancy McGovern, MIT Libraries

NDSR Curriculum



Curriculum

- Immersive week to provide a landscape view of digital stewardship
- Scheduled events that build on pre-existing community events whenever possible
- Group activities that extend experiences and build community
- Individual residency requirements tailored to projects, skills, and interests

Immersion week

- Launched the residencies
- Morning instruction
 - DPOE*: Identify, Select, Store, Protect, Manage, Provide
 - Interactive with examples pertaining to projects
- Afternoon demonstration and hands-on
 - Demo relevant tools for projects
 - Review workflow scenarios and practical examples
- Discussion and review
- Participation by hosts, instructors

*DPOE = Digital Preservation Outreach and Education

Resident development plans

- Developed by the resident in consultancy with their mentor
- Documents:
 - **Project-specific** topics, skills & tools to learn so that the project goals can be achieved
 - **Professional development** goals – topics & tools to advance resident's knowledge and skills

Residency events

- Scheduled curriculum events
 - In-person presentations, demonstrations, etc.
 - On-line activities, e.g. webinars plus discussion
- Group activities
 - Events at host institutions
 - Participation in regional events
 - Tours, visits and networking
 - Collaborate on organizing events
- Capstone event

September residency events

- Optical Media, NDSA Standards & Practices Working Group Call
- Launch of residents' blog (ndsrbboston2014.wordpress.com)
- The Information: A History, A Theory, A Flood by James Gleick, NEASIST Program Committee
- MIT Fellows event
- Open house, MIT Digital Sustainability Lab

Some upcoming events

- Site visits to NEDCC, JFK library
- NYC & Boston Google Hangout
- Jan. 2015 – Open mid-year event
- Monthly events at NDSR Boston hosts
- Mar. 2015: NDSR DC/Boston/NYC presentation at NEA/MARAC
- May 2015: Open NDSR Boston capstone event

Samantha DeWitt, Host: Tufts University

Institutional Knowledge of Research Data at Tufts University



NDSR

• Boston

Samantha DeWitt

Tufts University

Tisch Library

Medford, MA

Institutional Knowledge of Research Data

NDSR

• Boston

Tufts University

10 Schools:

School of Arts and Sciences • School of Engineering • School of Medicine
School of Dental Medicine • Sackler School of Graduate Biomedical Sciences
Friedman School of Nutrition Science and Policy • Cummings School of
Veterinary Medicine • Fletcher School of Law and Diplomacy
College of Special Studies • Tisch College of Citizenship and Public Service

- Tufts was first named a Research Category I university by the Carnegie Foundation in 1989
 - Today Tufts is rated as a university with "very high research activity"
-

NDSR

• Boston

Tufts University Project Goals:

- Help Tufts gain a more complete understanding of the research data produced by its faculty, research staff, post docs and graduate students.
 - Investigate strategies for producing metadata objects for the Fedora-based digital repository. These objects would represent Tufts-created datasets, which would be either referenced (more likely) or stored in the repository (less likely).
-

NDSR Boston

Research Data



XLMiner : Multiple Linear Regression

Row ID	Predicted Value	Actual Value	Residual	CRM	ZN	INDUS	CHAS	NOX	RM	AGE
1	30.24656555	34	4.04500055	0.00932	18	2.31	0	0.439	8.575	65.2
4	39.87652272	33.4	4.783471282	0.02237	0	2.18	0	0.458	9.598	45.8
5	27.78434086	36.2	8.43669135	0.09095	0	2.18	0	0.458	7.147	54.2
6	25.8204032								6.43	58.7
9	11.54583087								5.831	100
10	19.12586187								6.004	85.9
12	21.56655773								6.009	82.9
17	20.80054159								5.005	29.3
18	16.24622552								5.99	81.7

Row ID	AD	AP
7854	53	4830
7855	7878	4819
7856	7254	4915
7857	8354	5357
7858	1748	9658
7859	2533	7083
7860	2040	3650
7861	6598	9940
7862	4008	5324
7863	4910	2799
7864	9239	1305
7865	8898	811
7866	9584	1727
7867	2281	3780
7868	476	1576
7869	9818	8029
7870	3064	7237
7871	3979	5525
7872	1247	8358
7873	2819	2903
7874	1765	1032

Region	Product	Subcategory	Store	SALES	Inventory	Revenue
United States	Slipper	New York	24	1242.118	81.201.51	81.201.51
United States	Slipper	New York	13	807.191	8141.193	8141.193
United States	Women's C	New York	10	875.612	9338.602	9338.602
United States	Women's B	New York	10	1205.412	95.123.45	95.123.45
United States	Boat	Seattle	10	1705.749	9225.071	9225.071
United States	Pen & Ink	Seattle	6	188.842	9114.038	9114.038
United States	Men's Sh	Seattle	10	1232.287	9274.181	9274.181
United States	Shoe	Seattle	14	86.498	138.194	138.194
United States	Slipper	Seattle	14	8118.358	1893.459	1893.459
United States	Slipper	Seattle	28	883.112	1119.291	1119.291
United States	Slipper	Seattle	2	187.210	128.610	128.610
United States	Women's C	Seattle	25	910.436	8139.237	8139.237
United States	Boat	Copenhagen	2	81.688	81.437	81.437
Western Europe	Boat	Copenhagen	2	81.688	81.437	81.437
Western Europe	Boat	Copenhagen	2	81.688	81.437	81.437
Western Europe	Boat	Copenhagen	2	81.688	81.437	81.437
Western Europe	Boat	Copenhagen	2	81.688	81.437	81.437
Western Europe	Boat	Copenhagen	2	81.688	81.437	81.437
Western Europe	Boat	Copenhagen	2	81.688	81.437	81.437
Western Europe	Boat	Copenhagen	2	81.688	81.437	81.437
Western Europe	Boat	Copenhagen	2	81.688	81.437	81.437
Western Europe	Boat	Copenhagen	2	81.688	81.437	81.437

Data Set	Cases	Class	Features	Inputs	Outputs	Hidden	Epochs
breast-cancer-w	699	2	9	9	1	5	30
credit-a	690	2	6	9	1	10	35
credit-g	1000	2	7	13	63	1	10
diabetes	768	2	9	9	1	5	30
glass	214	6	9	9	9	1	10
heart-cleveland	303	2	8	8	13	1	5
hepatitis	185	2	6	13	32	1	10
svm-svm-84	435	2	16	16	1	5	40
lypo	3972	5	7	22	85	5	15
ionosphere	351	2	34	34	1	10	40
lin	135	3	4	4	3	5	80
lin-vs-kg	3196	2	36	36	1	15	30
labor	89	2	8	8	1	10	80
letter	20000	26	16	16	16	20	30
svm-svm-936	936	2	16	16	1	40	30
svm-svm-1679	1679	2	16	16	1	40	30

	Body Length (1998)	Age (1999)	Weight (1999)	Body Length (1999)	Age (2000)	Weight (2000)	Body Length (2000)
8	6.8	2	8.5	8.7	3	8.52	9.1
6	15.2	3	9.57	17.8	4	9.23	19.1
1	13.9	4	6.2	15.6	5	6.97	16.3
2	8.2	2	8.7	12	3	8.5	14
9	15.7	4	13.2	15.75	5	13.5	16.2
3	9.1	4	9.7	16	5	10.3	17
6	8.1	1	10.1	15.1	2	10.2	17
6	8.1	7	8.3	16.5	8	8.4	16

Storage



NDSR

• Boston

Q:

Why is it important for Tufts to understand the kind of research data it produces and how does this project relate to the missions of the Tufts libraries?

A:

Understanding the scope of present and future datasets enables Tufts to better understand its data management and stewardship obligations.

Tisch Library Mission Statement

“...The library acts as a gateway and a contributor to networked information in and outside of Tufts.”

NDSR

• Boston

Office of Research Administration (ORA)

Office of Proposal Development (OPD)

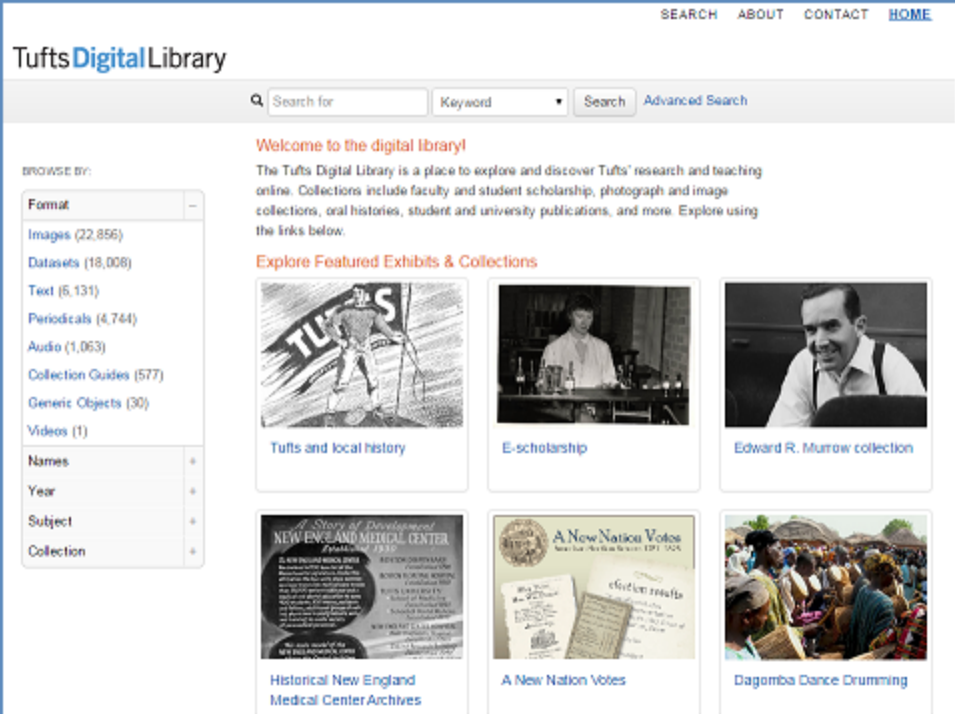
In providing assistance in research data management, Tufts libraries augment the work of the university's research offices by helping to:

- Assure that Tufts is in compliance with funders' data management mandates.
 - Create an environment that encourages and facilitates research.
-

NDSR Boston

The collections of the Tufts Digital Library support teaching and research at Tufts and feature resources of enduring value, created by the community's members.

<http://dl.tufts.edu/>



The screenshot shows the homepage of the Tufts Digital Library. At the top right, there are links for SEARCH, ABOUT, CONTACT, and HOME. The main header reads "Tufts Digital Library". Below this is a search bar with a magnifying glass icon, a "Search for" input field, a "Keyword" dropdown menu, and "Search" and "Advanced Search" buttons. On the left side, under "BROWSE BY:", there is a list of categories with counts: Format (Images: 22,856; Datasets: 18,008; Text: 6,131; Periodicals: 4,744; Audio: 1,063; Collection Guides: 577; Generic Objects: 30; Videos: 1), Names, Year, Subject, and Collection. The main content area features a "Welcome to the digital library!" message, followed by a paragraph describing the library's purpose and a list of featured exhibits and collections. These include "Tufts and local history", "E-scholarship", "Edward R. Murrow collection", "Historical New England Medical Center Archives", "A New Nation Votes", and "Dagomba Dance Drumming".

SEARCH ABOUT CONTACT HOME

Tufts Digital Library

Search for Keyword Search Advanced Search

BROWSE BY:

Format

- Images (22,856)
- Datasets (18,008)
- Text (6,131)
- Periodicals (4,744)
- Audio (1,063)
- Collection Guides (577)
- Generic Objects (30)
- Videos (1)

Names

Year

Subject

Collection

Welcome to the digital library!

The Tufts Digital Library is a place to explore and discover Tufts' research and teaching online. Collections include faculty and student scholarship, photograph and image collections, oral histories, student and university publications, and more. Explore using the links below.

Explore Featured Exhibits & Collections

Tufts and local history

E-scholarship

Edward R. Murrow collection

Historical New England Medical Center Archives

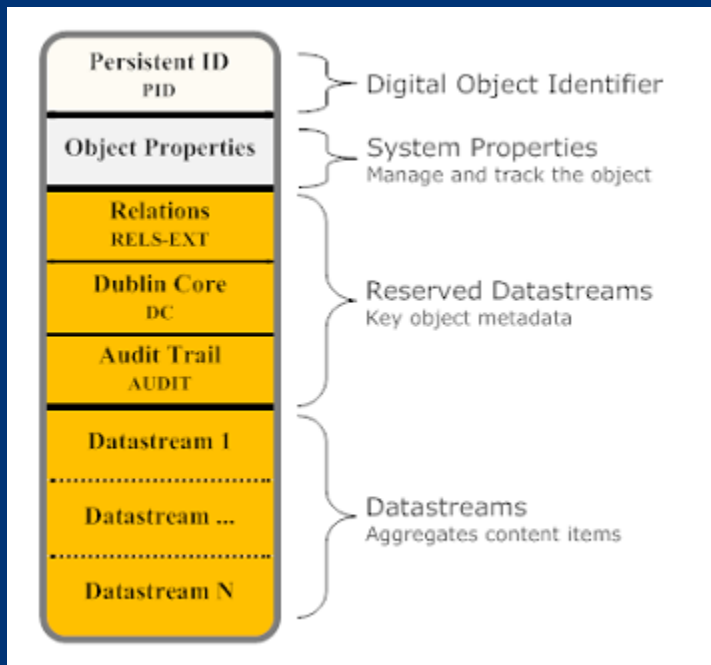
A New Nation Votes

Dagomba Dance Drumming

NDSR

• Boston

Fedora Object Model



Control Groups:

- **Internal XML Content**
- **Managed Content**
- **Externally Referenced Content**
- **Redirect Referenced Content**

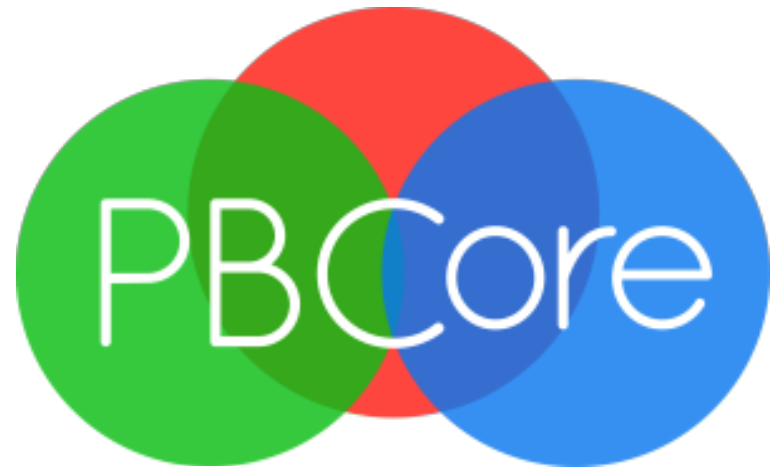
Rebecca Fraimow, Host: WGBH

WGBH Digital Media Preservation Project



NDSR BOSTON

Major initiatives...



the American
Archive
of public broadcasting

...and daily operations

- **Stock sales**
- **Production assistance**
- **Researcher assistance**
- **Administrative metadata management**
- **Descriptive metadata management**
- **Rights metadata management**
- **Ingest into DAM**
- **LTO backup and maintenance**

FRONTLINE



**MASTERPIECE
THEATRE®**

Phase One



FRONTLINE



**MASTERPIECE
THEATRE®**

Phase Two



the American
Archive
of public broadcasting



Phase Three



basic **Black**



Phase Four



the American
Archive
of public broadcasting



Rebecca Fraimow

rebecca_fraimow@wgbh.org

ndsrboston2014.wordpress.com

@rhfraim



NDSR BOSTON

Joey Heinen, Host: Harvard Library

Format Migration Plans and Framework for Harvard Library



NDSR BOSTON

Joey Heinen

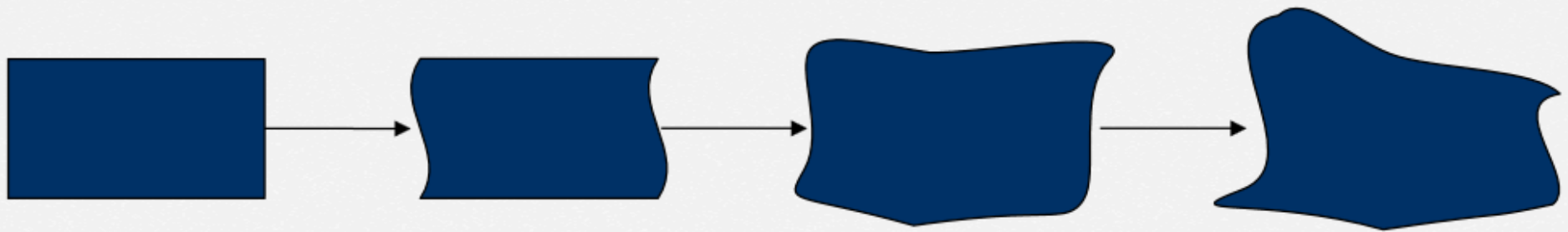
National Digital Stewardship Resident
Format Migration Plans and Framework for Harvard Library -
Kodak PhotoCD, SMIL playlists, RealAudio

HARVARD
LIBRARY



NDSR BOSTON

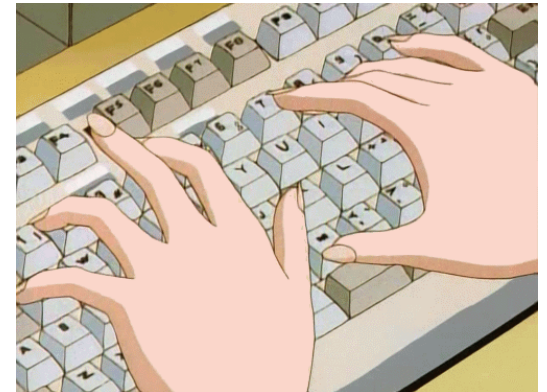
File Formats Migration – What are the real hazards



Research – A Manic Process

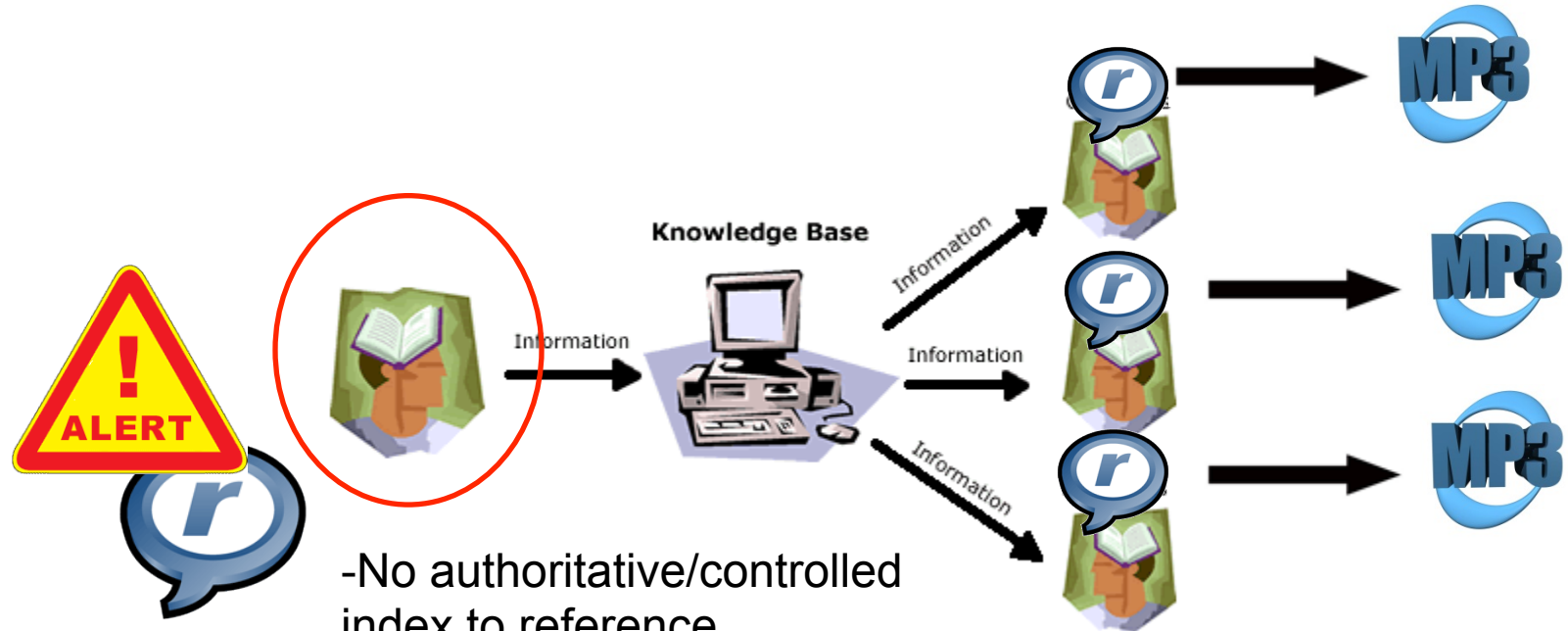
Compiled Bibliography, reviewed, and culled to 30+ materials

- Data Migration Projects (large data centers)
- Format-specific case studies
- Workflow design and integrated systems (some tools)
- Development of new tools
- Understanding formats



Tools vs. Experiments – A few red herrings

Examples – AONS (Automatic Obsolescence Notification System)
GREAT concept, but major pieces missing...



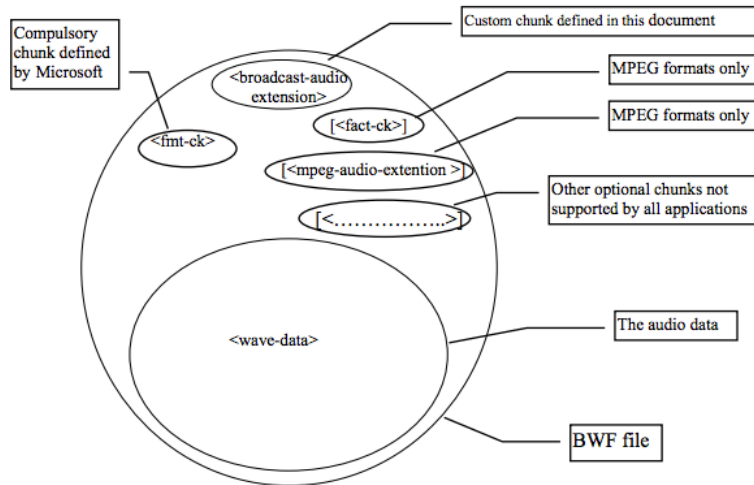
- No authoritative/controlled index to reference
- Lack of robustness to characterize objects and identify inherent risks

Characterization – More Missing Pieces

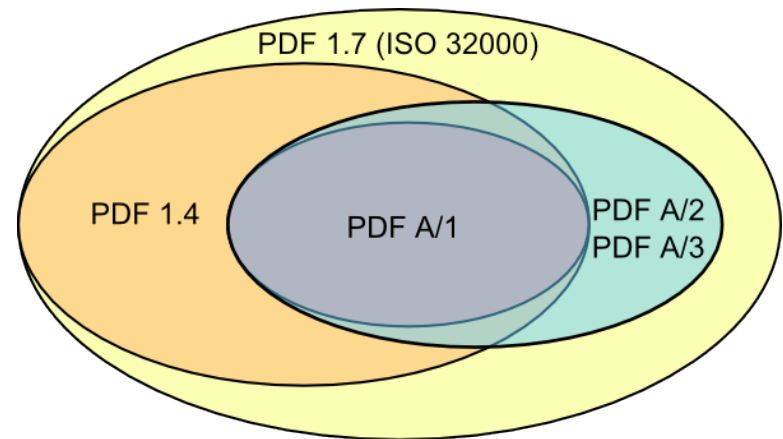
Even within formats there can be derivations in how that object is delivered and represented:

Formats can vary in their supported fonts, types of embedded content, how technical information is formatted and where it is located in the file, etc.

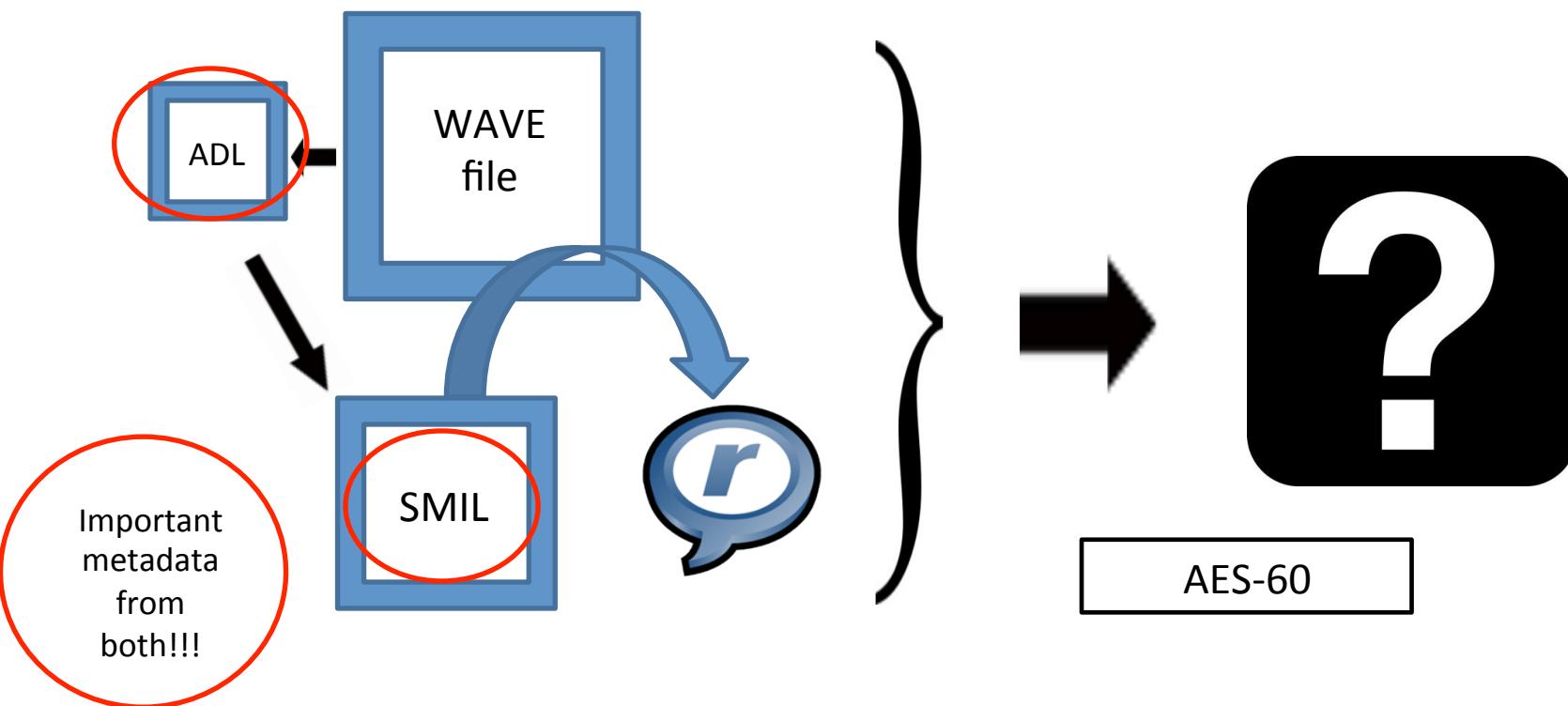
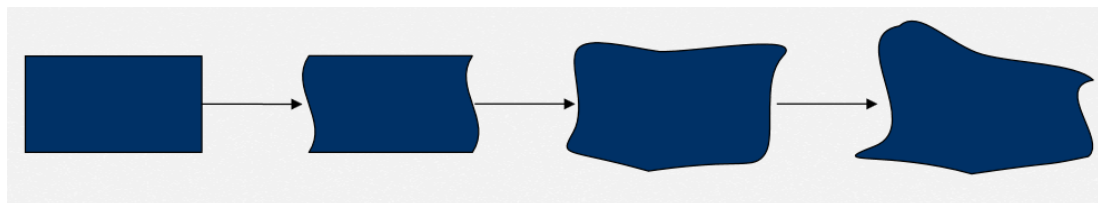
Data “chunks” in Broadcast WAVE



Attribute Overlap in PDF



From Source to Target – Not a One-Hop Journey!



Conclusions Drawn:

-Characterization is an essential step, content models should be tested to determine adequacy. All aspects of the workflow need to have a shared understanding of format complexity.

-Digital “objects” may be multivalent – external relationships are part of characterization and need to be considered in tandem with migration.

Next Steps:

-Map research/bibliography to migration workflow drafts: overall framework, format and deliverables (artifacts), context/stakeholders.

-Deep exploration of the Digital Repository Services (DRS) and how the workflow interfaces with this system.

-Into the Belly of the Beast: Intensive research on the three formats.

Joey Heinen
joseph_heinen@harvard.edu
ndsrboston2014.wordpress.com
@joey_heinen #ndsr



Jen LaBarbera, Host: Northeastern University

Channeling Streams of Archival Records: Northeastern University



Project Goals

- 1) Born digital materials: Create a workflow for ingesting born-digital content to new digital repository service (*in progress*)
- 2) Develop digital preservation plan for NU Libraries (*beginning November/December 2014*)
- 3) Legacy digital materials: Ingest obsolete legacy digital materials (e.g. floppy disks) to DRS (*beginning early spring 2015*)



Ingesting born-digital materials

Goal: Transfer all materials from current states / platforms (Omeka, external hard drive, local server, Google drive, etc) to new Fedora-based repository

Challenges:

- metadata currently in Dublin Core, needs to be in MODS
- managed by doctoral student who has minimal time to dedicate
- agreements are scattered
- new DRS is still in pre soft launch stage
- others TBD

Progress - Literature Review

Literature Review - list of resources

Digital Preservation and the Digital Humanities

<http://www.trevorowens.org/2014/03/digital-preservations-place-in-the-future-of-the-digital-humanities/>

Born Digital: Guidance for Donors, Dealers, and Archival Repositories

<http://www.clir.org/pubs/reports/pub159/pub159.pdf>

Digital Forensics and Born-Digital Content in Cultural Heritage Collections

<http://www.clir.org/pubs/reports/reports/pub149/pub149.pdf>

A Survey of Digital Cultural Heritage Initiatives

<http://www.clir.org/pubs/reports/reports/pub118/pub118.pdf>

Digital Humanities and the Library (link roundup)

<http://miriamposner.com/blog/digital-humanities-and-the-library/#datapres>

Digital Humanities SPEC Kit (ARL)

<http://www.arl.org/storage/documents/publications/spec-326-web.pdf>

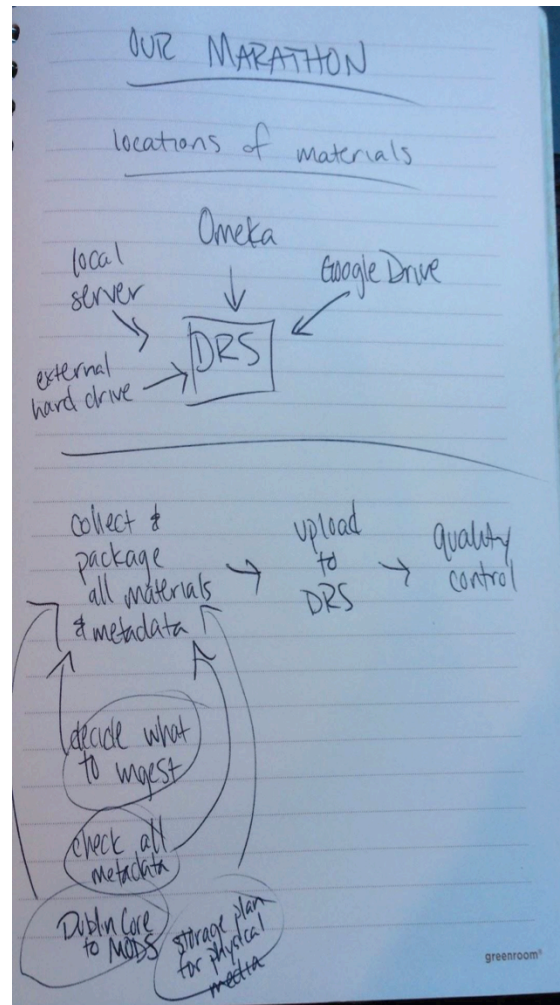
Checking your Digital Content: What is Fixity, and When Should I Be Checking It?

http://digitalpreservation.gov/ndsa/working_groups/documents/NDSA-Fixity-Guidance-Report-final100214.pdf

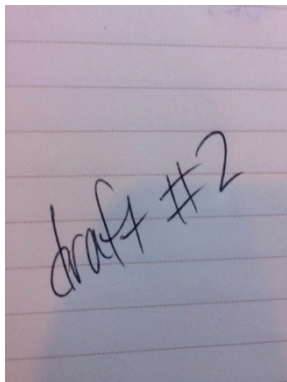
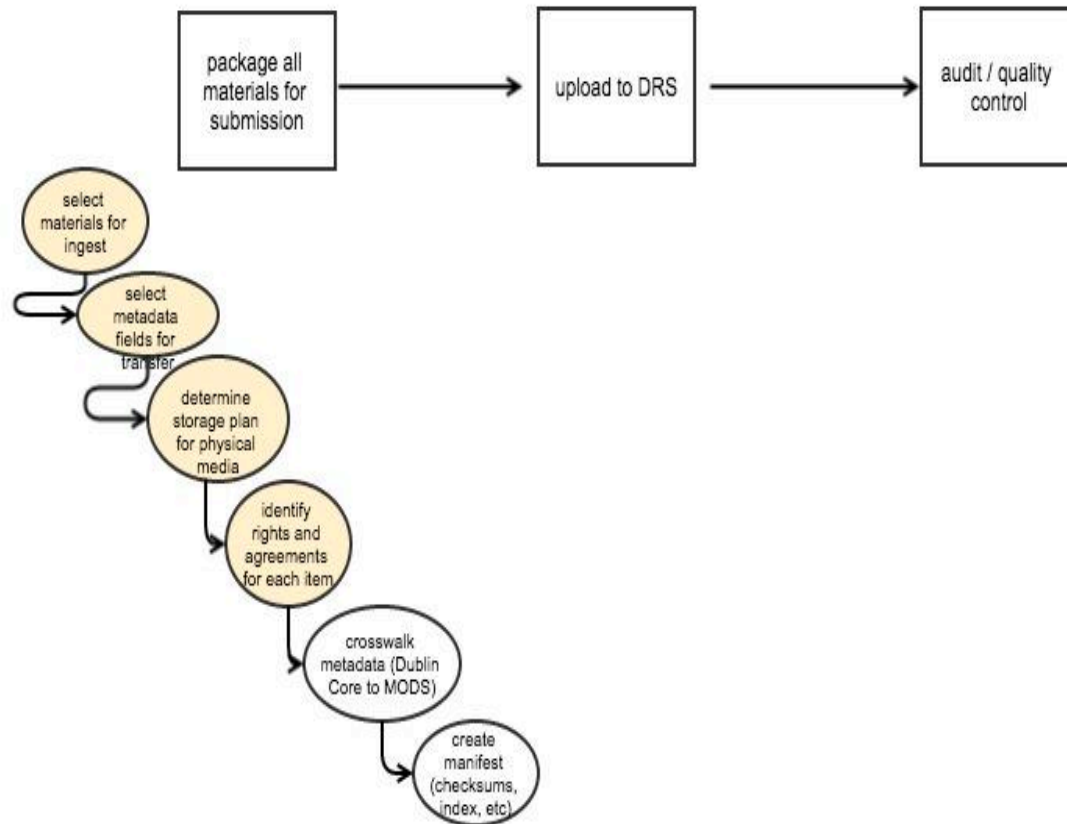
Repository to Repository Transfer of Enriched Archival Information Packages

Progress - Workflow

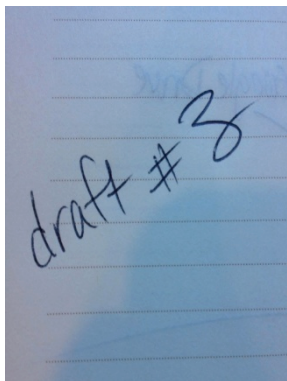
draft #1



Progress - Workflow



Progress - Workflow



Next Steps

- Complete workflow for born-digital materials
- Test born-digital workflow on Our Marathon, revise, test again, revise...
 - *(in the background)* DRS soft launch (November), user testing (Nov-Dec), DRS full launch (January)
- Next (concurrent) project: Digital Preservation Plan

Tricia Patterson, MIT Libraries

Making Music Last: Preservation Planning for 'Music at MIT' Digital Audio Content



NDSR BOSTON

MAKING MUSIC LAST

PRESERVATION PLANNING FOR 'MUSIC AT MIT'

TRICIA PATTERSON
MIT LIBRARIES

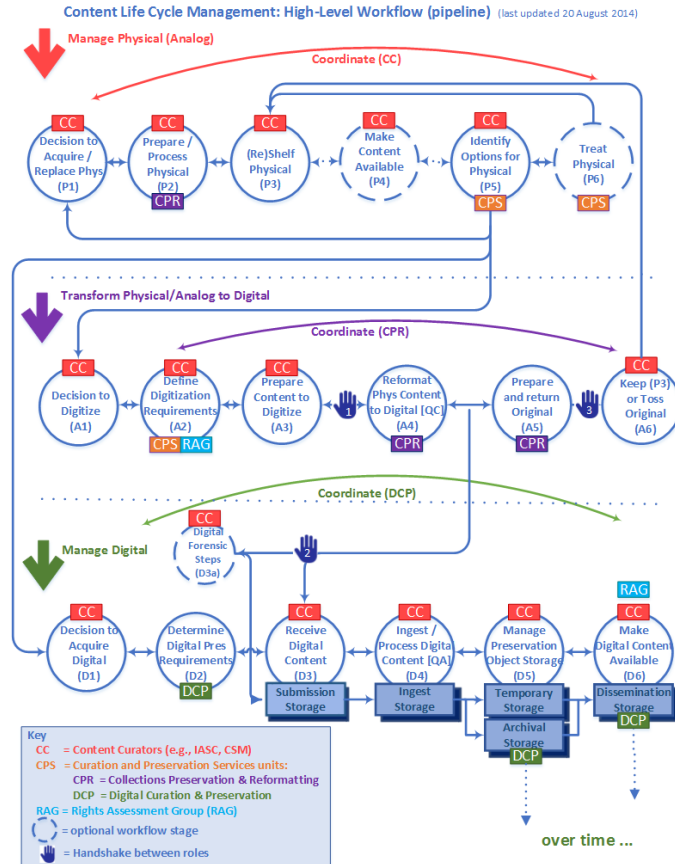
THE CONTEXT



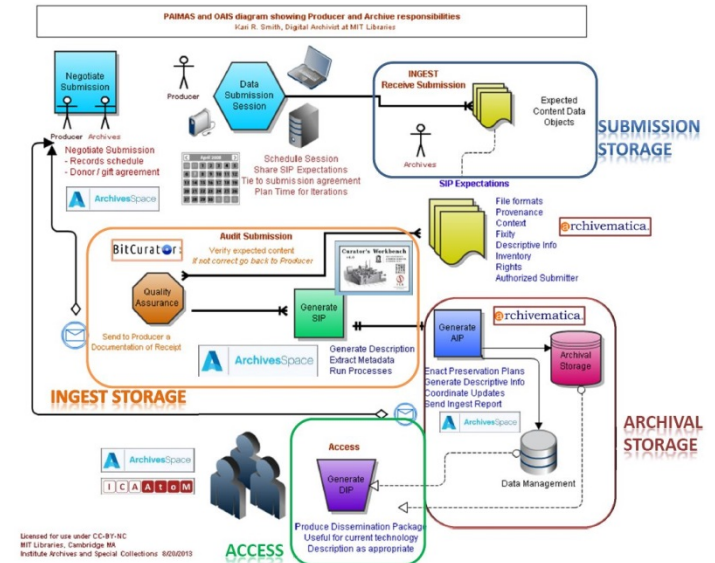
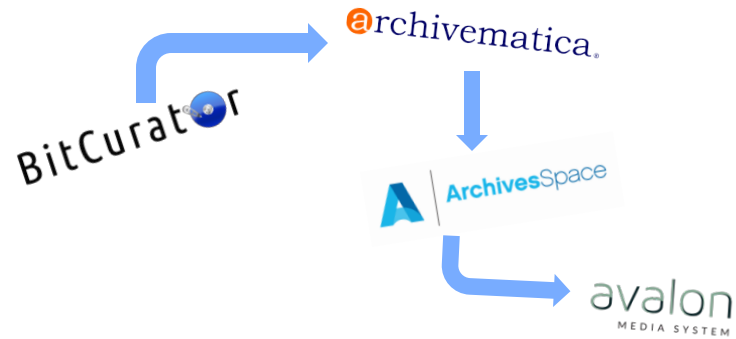
- ♪ Music is the 2nd most popular minor at MIT
- ♪ Lewis Music Library is a subject-specific library (obviously)
- ♪ Increased support for digital content management at MIT libraries
- ♪ Ongoing project to inventory, digitize, preserve, and facilitate access to music content
- ♪ Workflow from this can be applied to other types of digital content

THE PROJECT

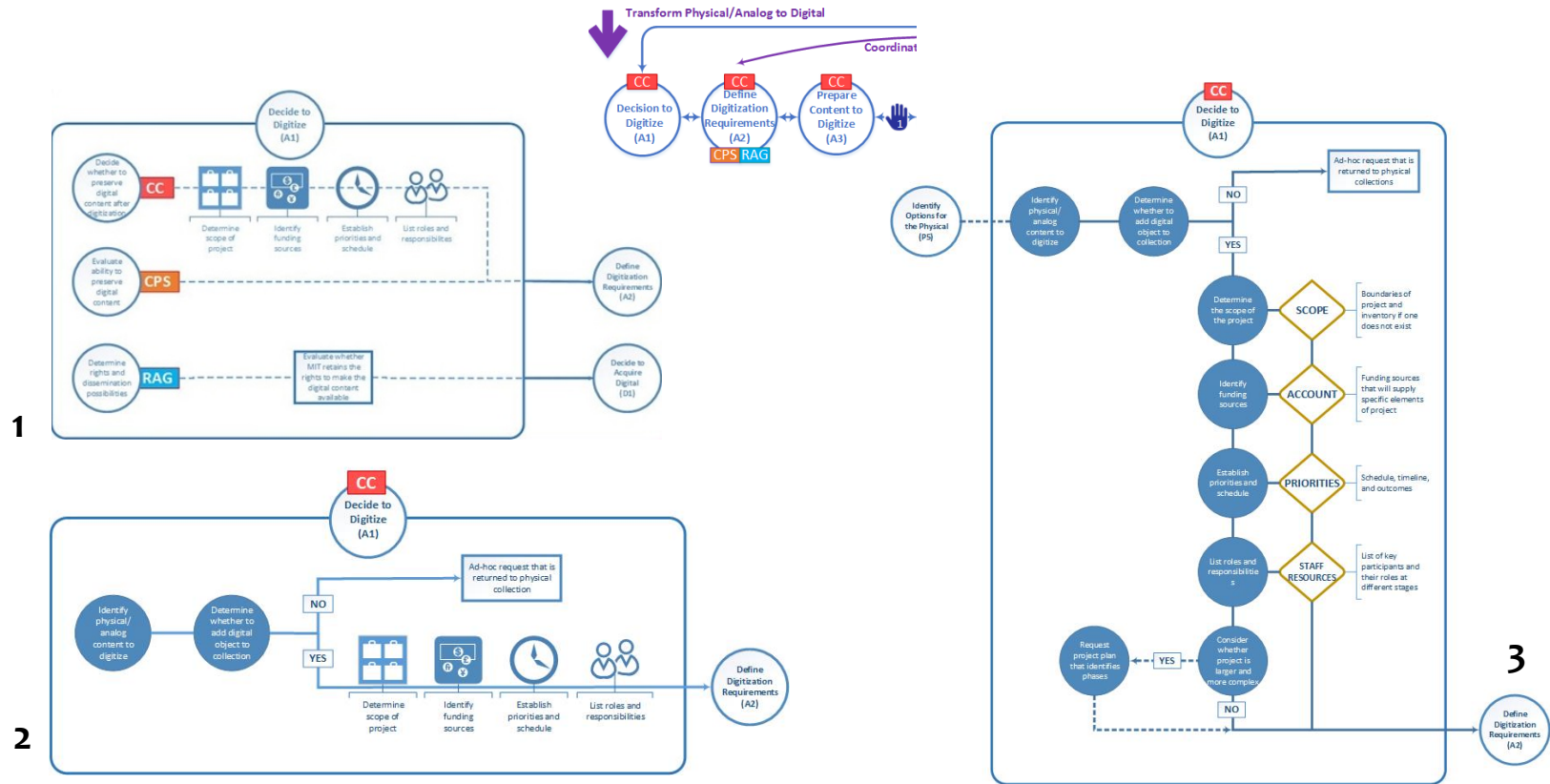
- ♪ Gap analysis of workflow documentation
- ♪ Test and enhance the life cycle workflow
- ♪ Create general digital content management workflow
- ♪ Create digital audio use case
- ♪ Help evaluate dissemination options for audio
- ♪ Consider extensibility of access platform



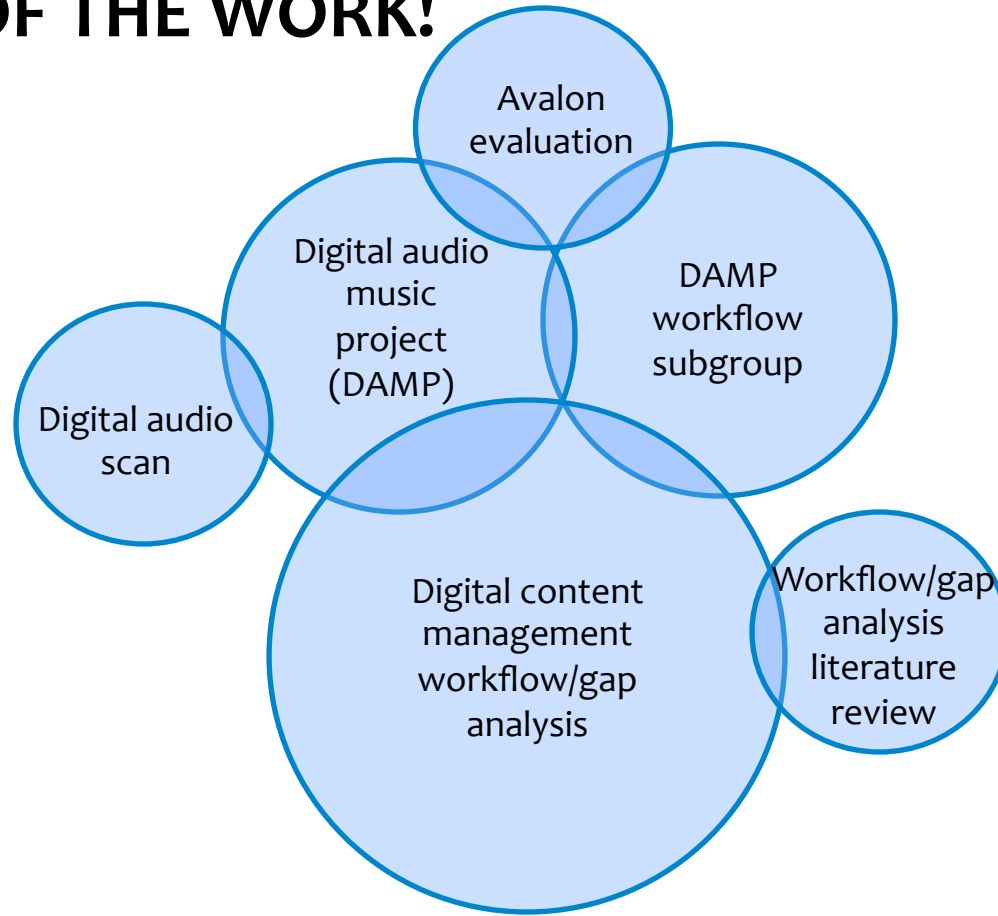
WHAT IS A WORKFLOW?



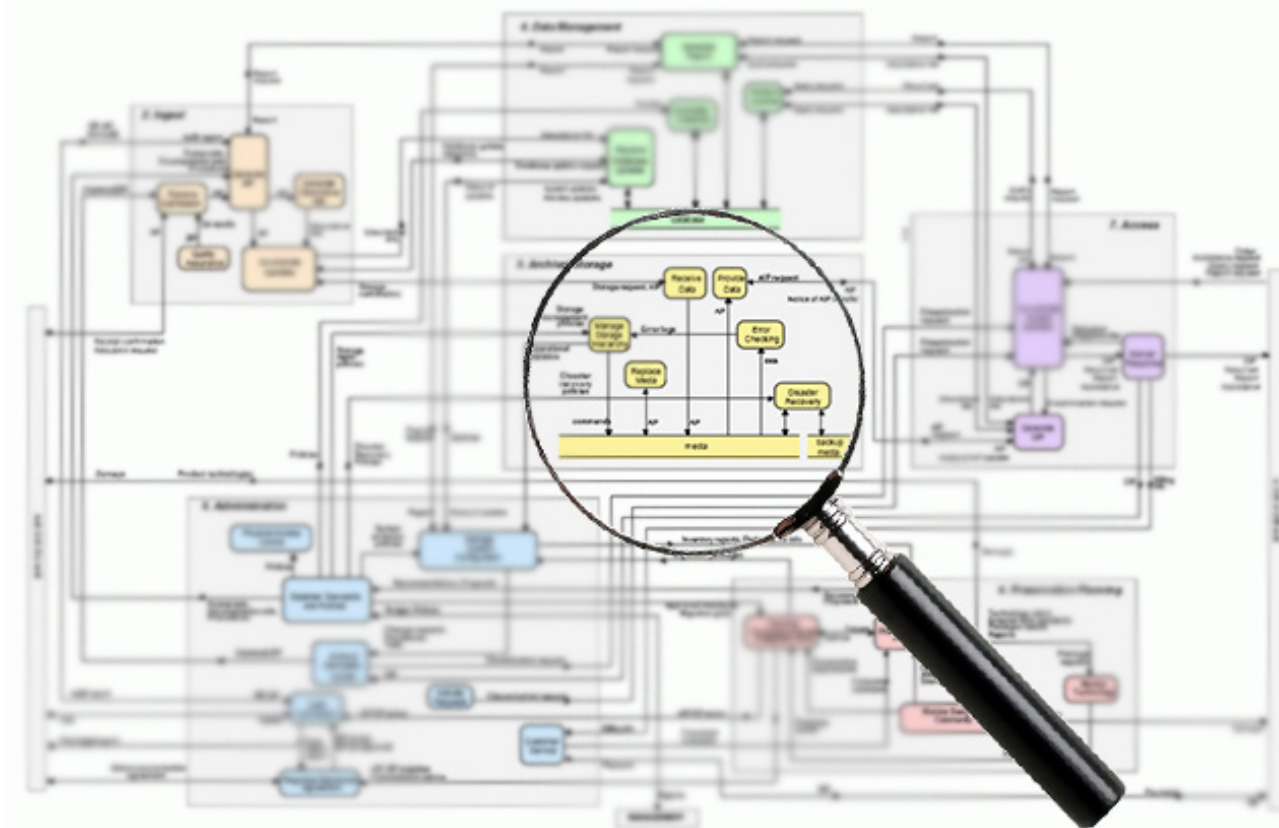
THE WORK!



MORE OF THE WORK!



THE END



Thank You!

Questions?