National Digital Stewardship Residency

Kris Nelson
knel@loc.gov
January 26, 2014

#NDSR

NDSR Overview

- 10 Residents
- 10 Organizations
- 2 Weeks of Workshop
- 9 Months of Hands-on Training
- 2 Professional Conferences

NDSR Hosts

- Association of Research Libraries
- Dumbarton Oaks
- Folger Shakespeare Library
- Library of Congress
- National Library of Medicine
- National Security Archive
- Public Broadcasting Service
- Smithsonian Institution Archives
- Maryland Institute for Technology in the Humanities
- World Bank Group

NDSR Projects

…..a few projects:

- ARL – *Developing Policies and Services to Make Digital Assets of Research Libraries Accessible*
- NSA – *The Digital Dissemination Challenge*
- MITH – *Accessing Born-Digital Literary Materials*
- NLM – *Developing a Thematic Web Archive Collection*

2013-2014 NDSR Residents

- Julia Blase, U. Denver – NSA
- Heidi Dowding, Wayne State U. – Dumbarton Oaks
- Maureen Harlow, UNC – NLM
- Jaime McCurry, Long Island U. – Folger Shakespeare
- Lee Nilsson, Eastern Washington U. – LOC
- Margo Padilla, San Jose State U. – MITH
- Emily Reynolds, U. Michigan – World Bank
- Molly Schwartz, U. Maryland – ARL
- Erica Titkemeyer, NYU – Smithsonian Archives
- Lauren Work, U. Washington - PBS

National Digital Stewardship Residency

File Format Action Plans

In Theory and Practice at the Library of Congress

DigitalPreservation.gov

Sustainability of Digital Formats
Planning for Library of Congress Collections

Introduction | Sustainability Factors | Content Categories | Format Descriptions | Contact

Format Descriptions >> Format Description Categories >> Browse Alphabetical List >> Format Descriptions as XML

Format Descriptions

Still Image
- SVG_1_1
- TIFF_6
- All still image format descriptions

Sound
- WAVE
- MP3_FF
- All sound format descriptions

Moving Image
- MPEG-4_FF_2
- AVI
- All moving image format descriptions

Textual
- NITF
- XML
- All text format descriptions

Web Archive
- ARC_IA
- WARC
- All Web archive format descriptions

Datasets
- DBF
- HDF5
- All dataset format descriptions

Geospatial
- ESRI_shape
- GeoPDF_2_2
- All geospatial format descriptions

Generic
- ASF
- RIFF
- All generic format descriptions

About the format descriptions

The digital content format descriptions accessible here provide specific information about individual formats and their characteristics. Each description provides moderately detailed information and citatio structures and encodings, and the mechanisms used to compress files or bitstreams. Inclusion of a description for a format does not imply that the format is preferred or acceptable for Library of Congres information about how to download the XML versions is presented here.

Descriptions are drafted and added to the site individually rather than on a scheduled basis. The compilers encourage persons with specialized knowledge to review and comment on the descriptions. Pl

These descriptions will not only serve as a resource for selection and acquisition at the Library but will also support the preservation of digital content as a resource for implementations of the *Representa* The Format Descriptions are intended to be consistent with and complementary to the proposed Unified Digital Format Registry.

PRONOM Registry

Florida Digital Archive Format Action Plans

Search: Search Terms

Apply

Description	Format Type	Extension	FDA Action Plan and Background Report	Production Date	Confidence Level	Notes
PrimeOCR PRO Format	Application	.PRO	(Action Plan) (Background Report)	05/07/2007	Medium	
Waveform Audio	Audio	.WAV, .WAVE	(Action Plan) (Background Report)	11/01/2005	See Notes	High: PCM Low: compressed
Audio Interchange File Format	Audio	.AIF, .AIFF, .AIFC	(Action Plan) (Background Report)	Action Plan not implemented	See Notes	High: PCM Low: compressed
Portable Document Format 1.2	Document	.PDF	(Action Plan) (Background Report) (PDF Normalization)	12/01/2005	See Notes	Medium: embedded fonts Low: encrypted
Portable Document Format 1.3	Document	.PDF	(Action Plan) (Background Report) (PDF Normalization)	12/01/2005	See Notes	Medium: embedded fonts Low: encrypted

Action Plan Example

Action Plan: WAVE

Implementation Date: December 1, 2005
Last Review: June 2009

1. Ingest Processing
- Identification: Yes.
- Validation: Yes. See the validation scheme described in Section 1.13 of WAVE Action Plan Background Report.
- Characterization: Yes. Extracted technical metadata are described in section 2 of WAVE Action Plan Background Report.
- Normalization: Each audio stream contained in the WAVE file is normalized into an uncompressed PCM(LPCM) audio stream with a sample size of 16 bits/sample.
- Migration: No.

2. Significant Properties
WAVE exhibits the following significant properties[1]. They are the significant properties which the FDA intends to preserve when adopting migration policy in the future.
- Content: Duration and number of channels.
- Context: Title and creating application information.
- Behavior: N/A
- Structure: Channel relationship (channel mapping).
- Appearance (Rendering): Sampling frequency and bit-depth.

3. Long-term Preservation Strategy

File Types at LC

1	File type	File count	
2	tif	50567621	http://en.wikipedia.org/wiki/Tagged_Image_File_Format
3	xml	8688909	http://en.wikipedia.org/wiki/XML
4	jp2	7860497	http://en.wikipedia.org/wiki/JPEG_2000
5	pdf	6742269	http://en.wikipedia.org/wiki/Portable_Document_Format
6	jpg	3544093	http://en.wikipedia.org/wiki/JPEG
7	txt	2764803	http://en.wikipedia.org/wiki/Text_file
8	gz	2694642	http://en.wikipedia.org/wiki/Gzip
9	html	205646	http://en.wikipedia.org/wiki/HTML
10	htm	162323	http://www.sightspecific.com/~mosh/www_faq/ext.html
11	gif	138831	http://en.wikipedia.org/wiki/Graphics_Interchange_Format
12	sid	127353	http://en.wikipedia.org/wiki/MrSID
13	sdw	121414	http://en.wikipedia.org/wiki/StarOffice
14	mp3	102035	http://en.wikipedia.org/wiki/MP3
15	aux	95273	http://whatis.techtarget.com/fileformat/AUX-TeX-LaTeX-Auxilliary-Reference-file
16	pcx	89862	http://en.wikipedia.org/wiki/PCX
17	wav	89367	http://en.wikipedia.org/wiki/WAV
18	doc	88786	http://en.wikipedia.org/wiki/Doc_(computing)
19		83382	No file extention
20	dat	78366	generic file extension for *data* files by various applications with no universal format
21	tfw	77745	http://www.fileinfo.com/extension/tfw
22	lyr	71135	http://www.lib.ncsu.edu/gis/formats.html
23	obj	71012	http://en.wikipedia.org/wiki/Wavefront_.obj_file
24	sum	70838	http://file-extension.paretologic.com/detail.php/File-Extension-sum
25	adf	70163	http://en.wikipedia.org/wiki/Amiga_Disk_File
26	xls	64900	http://en.wikipedia.org/wiki/Microsoft_Excel#File_formats
27	cr2	63076	http://en.wikipedia.org/wiki/Raw_image_format

Crowdsource the Action Plans

FileFormatAction Plans Wiki

On the Wiki   Wiki Content   Community

Wiki Activity   Random page   Videos   Photos

Home   Edit ▼   Talk 0

Description	Type	Extentions	Documentation	Date	Long Term Preservation Prospects	Software	Preservation Action	Notes
Tagged Image File Format	Raster Image Format	Tif, Tiff, Gtiff, Tif8, Btf, Dng	Action Plans , Technical Details (by digitalpreservation.gov)	12/2/2013	High	Adobe Photoshop , MS Paint , Gimp, Aperture (OS X), Almost all image editing suites	Retain	
Quicktime File Format	Multimedia Container File (video)	Mov, Qt, M4p, M4b	Action Plans, Technical Details (by digitalpreservation.gov)	12/2/2013	Medium	Quicktime Pro , VLC Media Player, Handbrake or Windows Live Movie Maker for	Retain	

Format Versions

Quicktime Format Type ✎ Edit ▼ Comments 1

Versions

As of 1/1/2014

The Quicktime File format is a multimedia wrapper which uses various encodings. After identifying your quicktime file or fileset with DROID⌐, the free and opensource characterization software Mediainfo ⌐ may be used to determine which codex was used to create it. While most of the action plans associated with different encodings will be the same. Some will be nessessarly altered if the encoding includes special elements such as DRM, timecodes, captions, or interactivity.

Codex/Version	Extentions	Documentation
QTV_Apple⌐	Mov	Action Plan
QTV_Cinepak⌐	Mov	Action Plan
QTV_DV⌐	Mov	Action Plan
QTV_MPEG⌐	Mov	Action Plan
QTV_MJPEG⌐	Mov	Action Plan
QTV_MP4_V⌐	Mov	Action Plan
QTV_MP4_AVC⌐	Mov	Action Plan
QTV_Sorenson⌐	Mov	Action Plan
QTV_SWF⌐	Mov	Action Plan
QT_210⌐	Mov	Action Plan
QT_UYVY⌐	Mov	Action Plan
QT_YUY2⌐	Mov	Action Plan
QTA_AAC⌐	M4p, M4b	Action Plan
QTA_MP3⌐	Mp3	Action Plan

Action Plan

Quicktime Sorenson Codec ✏ Edit ▾   Comments 0

Action Plan: Quicktime File Format

Last Updated 12/05/2013

Description: A multimedia container file which can contain audio, video, effects and subtitles. A propriatary Apple, Inc. format with specifications avalible here ⬈, it is made up "atoms," the basic unit of data comprising a quicktime file. QTFF is the basis for the MPEG-4 format.

Codec: The Sorenson Codec is widely used in video filemaking. It is compressed and often used as a middle state format for lower resolution streaming versions. It has also been observed to contain special features in some versions, such as multiple lines of audio/subtitles and even interactive elements.

Software: Apple Quicktime Pro ⬈ is used for saving or converting individual QTFF files. .Mov files can be played with VLC media player ⬈, iTunes ⬈, Windows Media Player ⬈ and many others. MediaInfo ⬈ is useful for identification of codecs and file characteristics.

Long Term Preservation Plan: The Quicktime File File Format is a propratary non-open source format. While widely used, it has since been eclipsed in popularity by file-types such as its direct successor MPEG-4. Apple has been moving away from .mov and its long-term future is uncertain. Yet the format remains popular and every major browser can support it. Some interactive elements specific to certain Quicktime versions have been observed becoming version obsolete. Migration is not at present recommended, however some files may require older versions of quicktime to play. Check the mediainfo report to determine what version was used initially as a guide. Many unuseable files can be played using VLC media player ⬈, however some interactive elements may be lost.

Division Specific Notes:

American Memory ⬈ : "American Memory has produced QuickTime files for web service since the mid-1990s, typically using the Cinepak codec at 10 fps and 160x120 pixels. For the special videos associated with the Dance Instruction Manuals collection, a second set of higher-resolution QuickTime files was produced, using the Sorenson codec at 15 fps and 320x240 pixels. The audio stream in both types employs IMA (Interactive Multimedia Association) 4:1 compression, based on PCM sampling at 22 kHz. (http://www.digitalpreservation.gov/ ⬈)"

Motion Picture, Broadcasting, and Recorded Sound Division ⬈ : "In 2004...collected a series of nine "enhanced" QuickTime movies produced by BMW (formerly available athttp://usa.bmwfilms.com/), with the main video object in the form of QTV_Sorenson.(http://www.digitalpreservation.gov/ ⬈)"

Thanks!

Lee Nilsson

@preselectlee
leegnilsson@gmail.com
Lnil@Loc.gov

LIBRARY OF
CONGRESS

NDSR at The National Security Archive

Julia Blase

National Digital Stewardship Resident

ALA Midwinter 2014

The National Security Archive

- Non-profit

- Journalists and historians

- Government documents

- FOIA requests and analyst research

- Usable history, human rights, open government
  - Actionable work

Going digital

- Used consultants for physical document management
- Less so with digital document management
- Decision making on an as-needed basis
- Five systems, almost 5 TB of content
- Reconsidering, but where to begin?
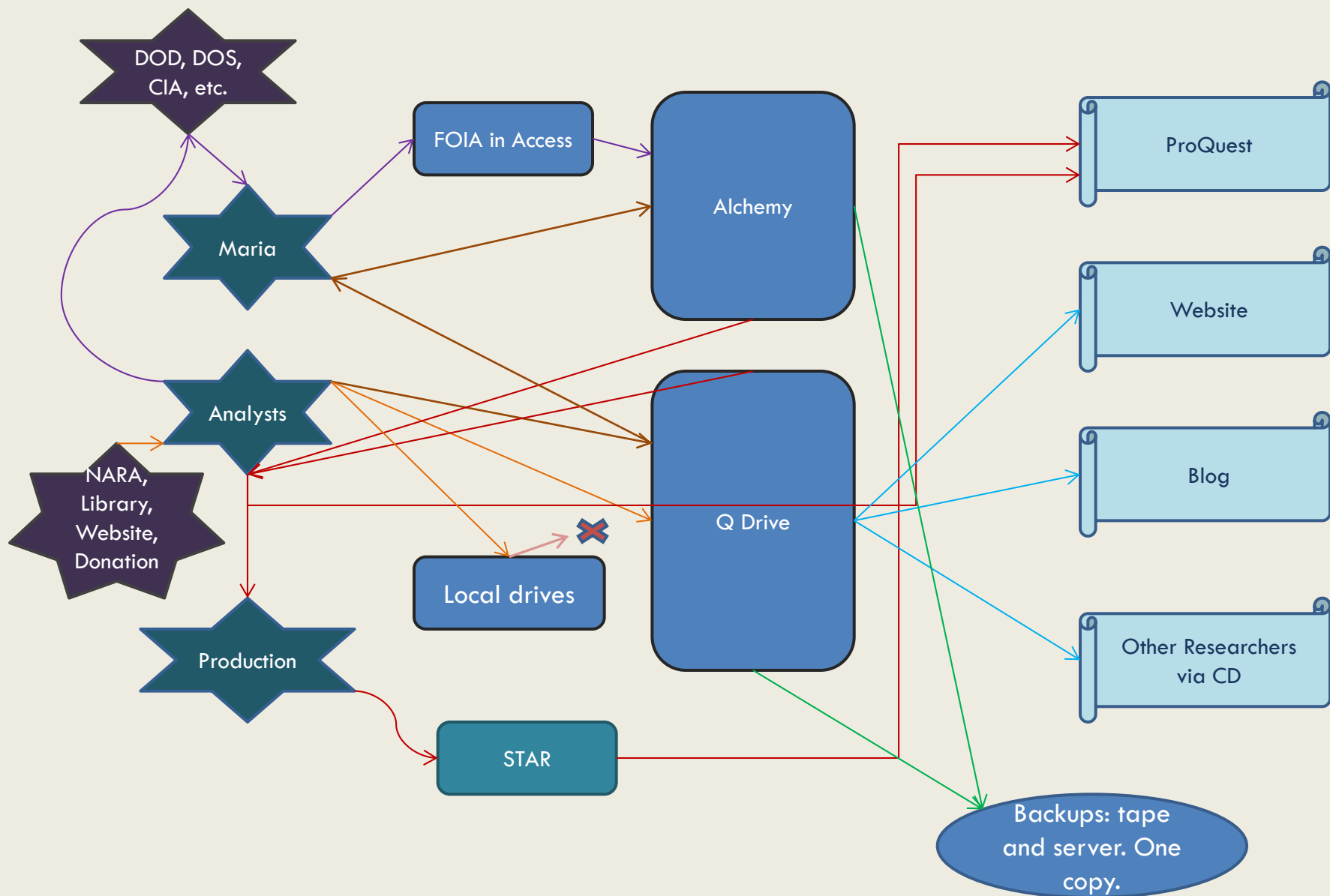
Current systems and workflows map

Crafting a strategy

- Goals of the residency:
  - Create comprehensive collections index
  - Analyze systems, procedures, mission, culture
  - Propose solutions that could improve digital asset management at the N. S. Archive

Current Status

☐ Assessment report completed

☐ Revisions necessary based on director's input

☐ Next steps:

◻ Formal "project management plan"

◻ Policy creation
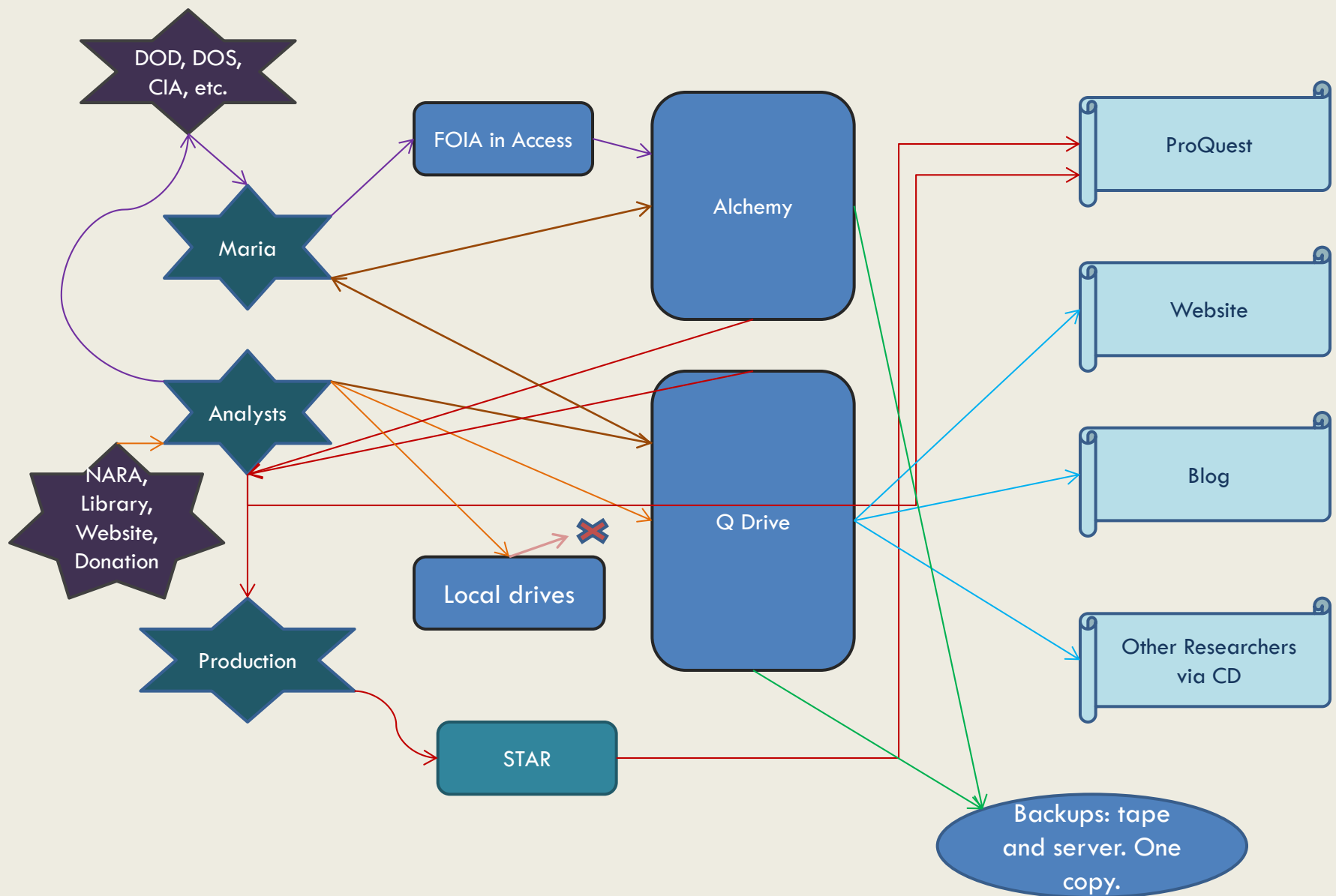
Current systems and workflows
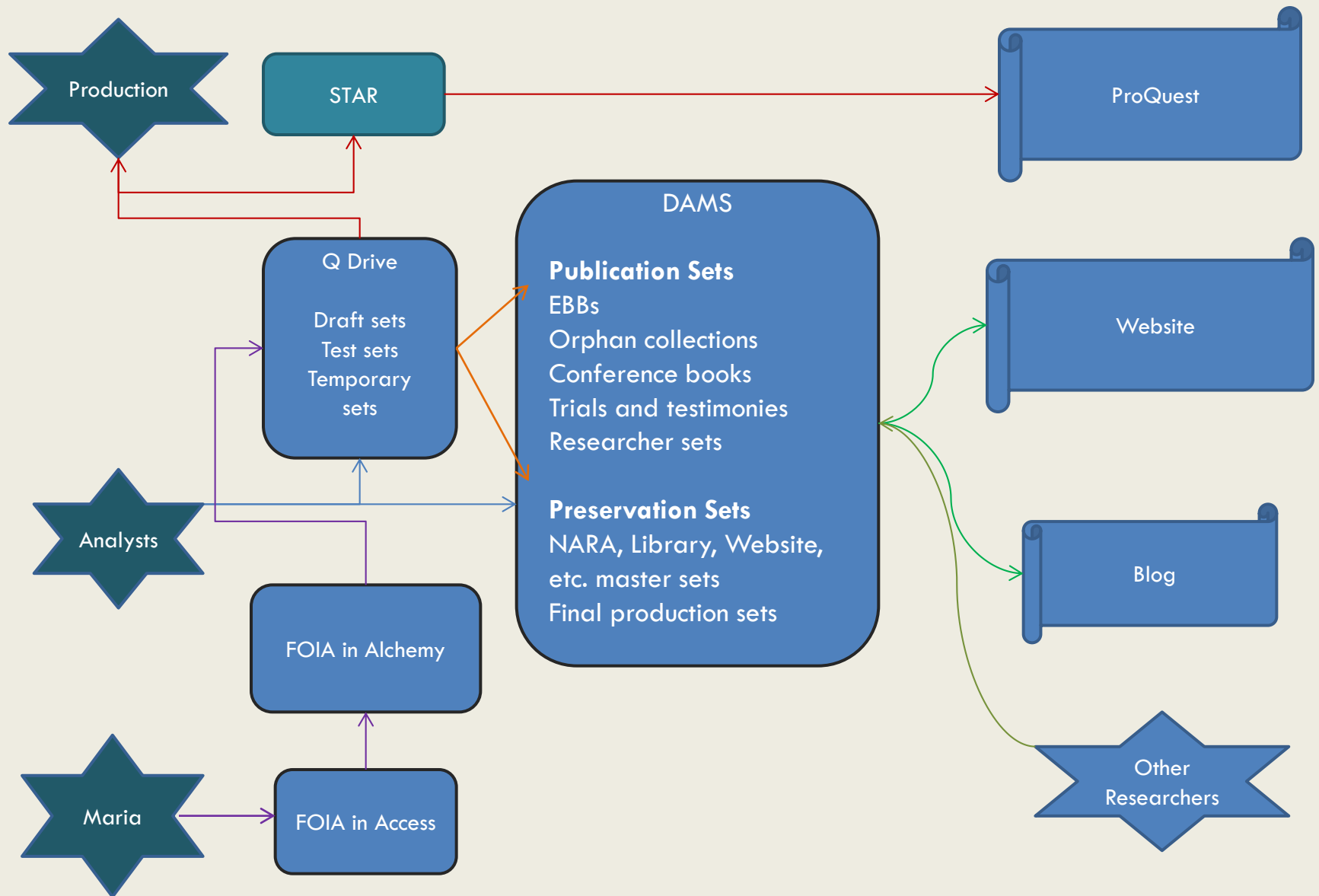
Plan 1 – proposed changes
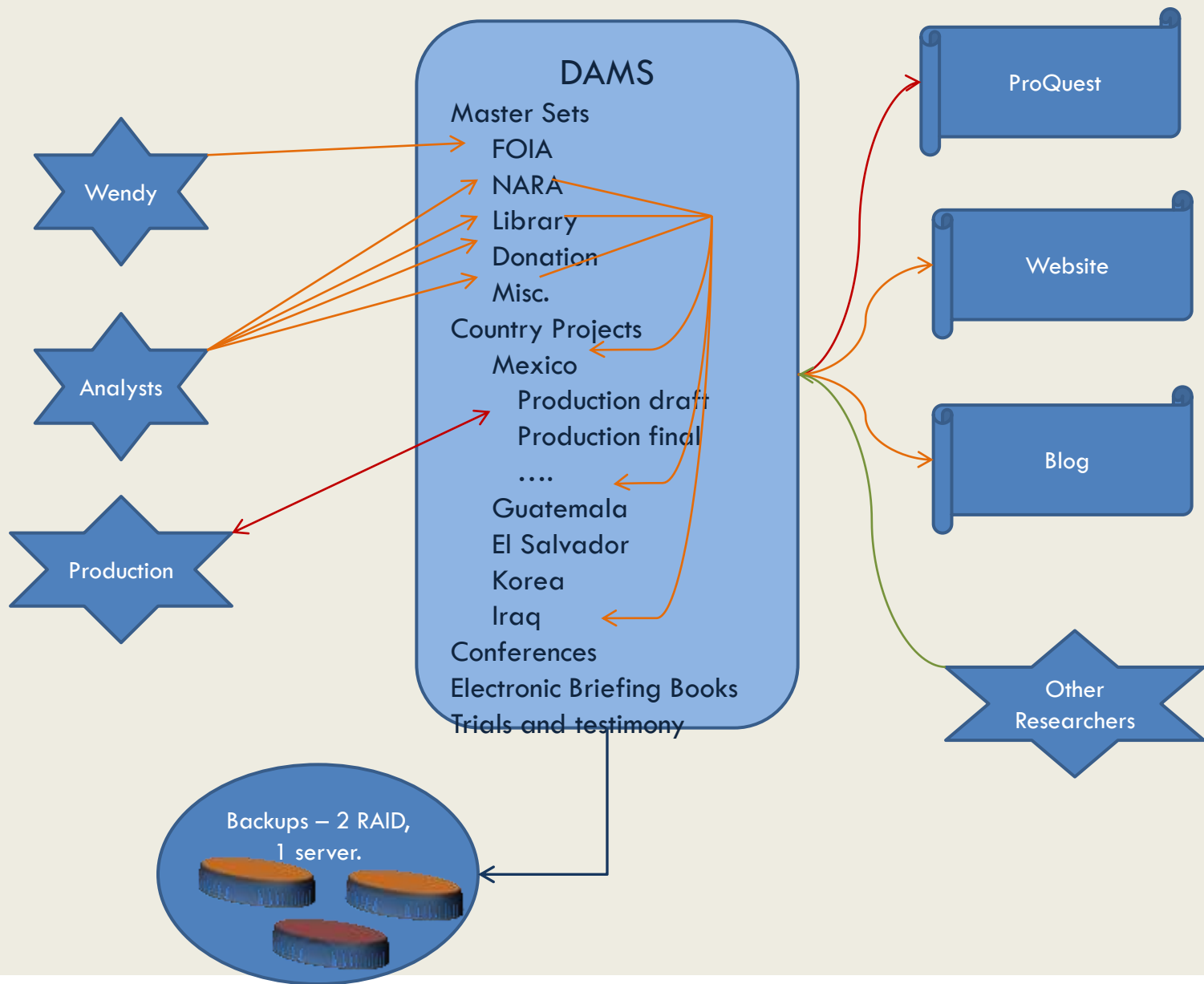
Plan 2 – proposed changes

World Bank eArchives project

Emily Reynolds
National Digital Stewardship Residency

eArchives project

World Bank **Access to Information** policy

International audience of researchers

New forms of **digital** scholarship

Project scope

Managing digitization workflows
Launching finding aid portal
Preserving files long-term

Digitization workflows

Digitization workflows

Automation of processes

Integration with systems

Finding aid portal

Finding aid portal

Branding & UI

Metadata crosswalks

Links to digitized materials

Long-term preservation

Long-term preservation

Evaluating format sustainability

Managing storage systems

Thank you!

@emilyrnlds
ek.rnlds@gmail.com

ACCESS for all

in a

digital
world

@mollyfication | ALA Midwinter 2014 |
The Association of Research Libraries

ACCESS forever

in a

digital world

ACCESS for all

Designing with the user in mind

ACCESS for all

Designing with ~~the~~ user in mind

@mollyfication

ACCESS for all

Designing with any user in mind

@mollyfication

ACCESS for all

Designing with any user in mind

- Universal Design

- Born-accessibility

Facilitate diverse information

consumption experiences.

ACCESSibility Toolkit

For research libraries

1. Establish standards and best practices
2. Provide a roadmap to accessibility
3. Build a community

@mollyfication

ACCESSibility Toolkit

In action

accessibility.arl.org

@mollyfication

Thanks!

Questions?

@mollyfication
molly@arl.org

Avoiding Digital "Woe, destruction, ruin, and decay."

The National Digital Stewardship Residency at the
Folger Shakespeare Library

Jaime McCurry
@Jaime_Ann
ALA Midwinter | 2014

Project Description:

To establish local routines and best practices for archiving and preserving the Folger's born-digital content

Web Archiving

Collections:

FSL Websites and Social Media
Seeds: 32

Shakespeare Festivals and Theatrical Companies
Seeds: 300+

Web Archiving

Refining Quality Assurance Processes

-Local documentation

-Weeding seeds

-Maintaining scope/modifying crawl limits

Enhancing Metadata Records

-Added location information for 290+ individual seeds

-Grouped by location for improved search functionality

Increase Awareness

-Outreach (local and public discussion, blog posts)

-Opportunities Presentation/Interaction

-Crowd-sourcing

New Collections

+ Shakespeare in the Media

+ Prominent (Modern) Shakespearean Actors and Actresses

Digital Asset Management

Digital Asset Management

File Format Inventory

-Utilizing DROID: Digital Record Object Identification Software

- Which will fuel research for a…

Digital Asset Management System

- In the near future!

Immediate Benefits:

-A healthy profile of these assets and their preservation requirements

-Workflows for ingestion into existing information systems

-Applied knowledge: APTV and Theatre can shed light onto organizational "content creator" needs across the board

Questions?

**Jaime McCurry
jmccurry@FOLGER.edu
@Jaime_Ann
ALA Midwinter | 2014**

Introducing Time-Based Media Art to a Trustworthy Digital Repository Environment

Smithsonian Institution Archives

Erica Titkemeyer
TitkemeyerE@si.edu
@ekemeyer

WHAT'S THE PROBLEM?

TBMA

(digital asset + tech/visual/auditory specs
+ artistic intent + behavior
+ married components)

=

Authentic Experience

$\downarrow$

Variable

WHAT'S THE PROBLEM cont.

TDR $\longrightarrow$ **Conservation & Curatorial Workflow**

SMITHSONIAN

- Pan-institutional
- Diverse works
- NPG ≠ SAAM ≠ Hirshhorn

Installation of *Random Access* – Nam June Paik @ SAAM

What is (or can be) sacrificed to serve multiple galleries/museums?

SO, HERE WE ARE….
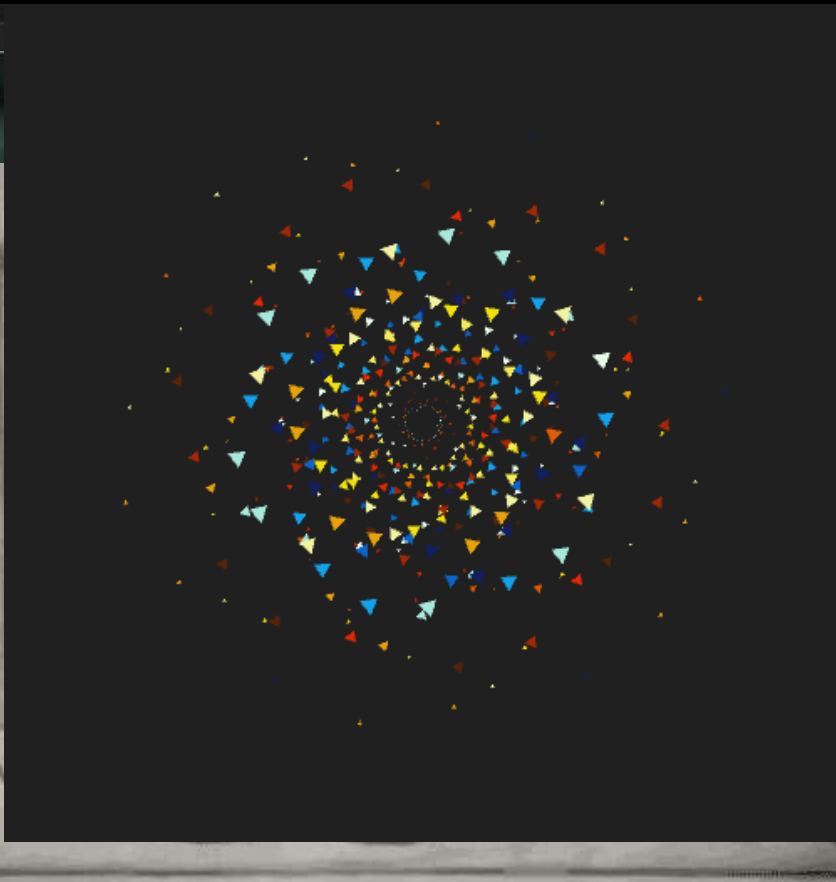
SOME ANSWERS

Variable Media Network

INCCA

Matters in Media Art

DOCAM

Artist Interview

Technical
Specifications

DOCUMENTATION
STRATEGIES

Video
documentation

Preservation Histories

SOME ANSWERS

Behavior

Contained
Installed
Performed
Reproduced
Duplicated
Encoded

CLASSES

Category

Single-Channel Video
Sculptural-based
Software-based
Website
Video Game (online/offline)
Generative

MANS (Richard Rinehart):
Score: didl xml metadata document itself
<DIDL>
Descriptor: descriptive data about score document
<DESCRIPTOR>
    Work: logical media art work or project
    <CONTAINER>
    Descriptor: descriptive data about work or project
    <DESCRIPTOR>
        Version: occurrence/state/ account of work
        <ITEM>
            Part(optional): logical sub-component
            <ITEM>
            Choice(optional): variables affecting configuration
            <CHOICE>
                Resource: physical or digital components
                <RESOURCE>

AIP

PDI   CI   →
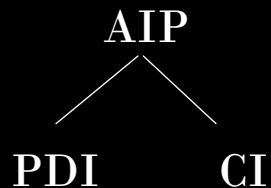
FINALLY...

COLLABORATION

CHANGING ROLES

MIGRATION, EMULATION, RE-INTERPRETATION

Jenny Holzer installation @ Neue Nationalgalerie

Thank You

Erica Titkemeyer
TitkemeyerE@si.edu
@ekemeyer

GLAMorizing Digital Preservation:

Digital Asset Management at Dumbarton Oaks

Heidi Elaine Dowding
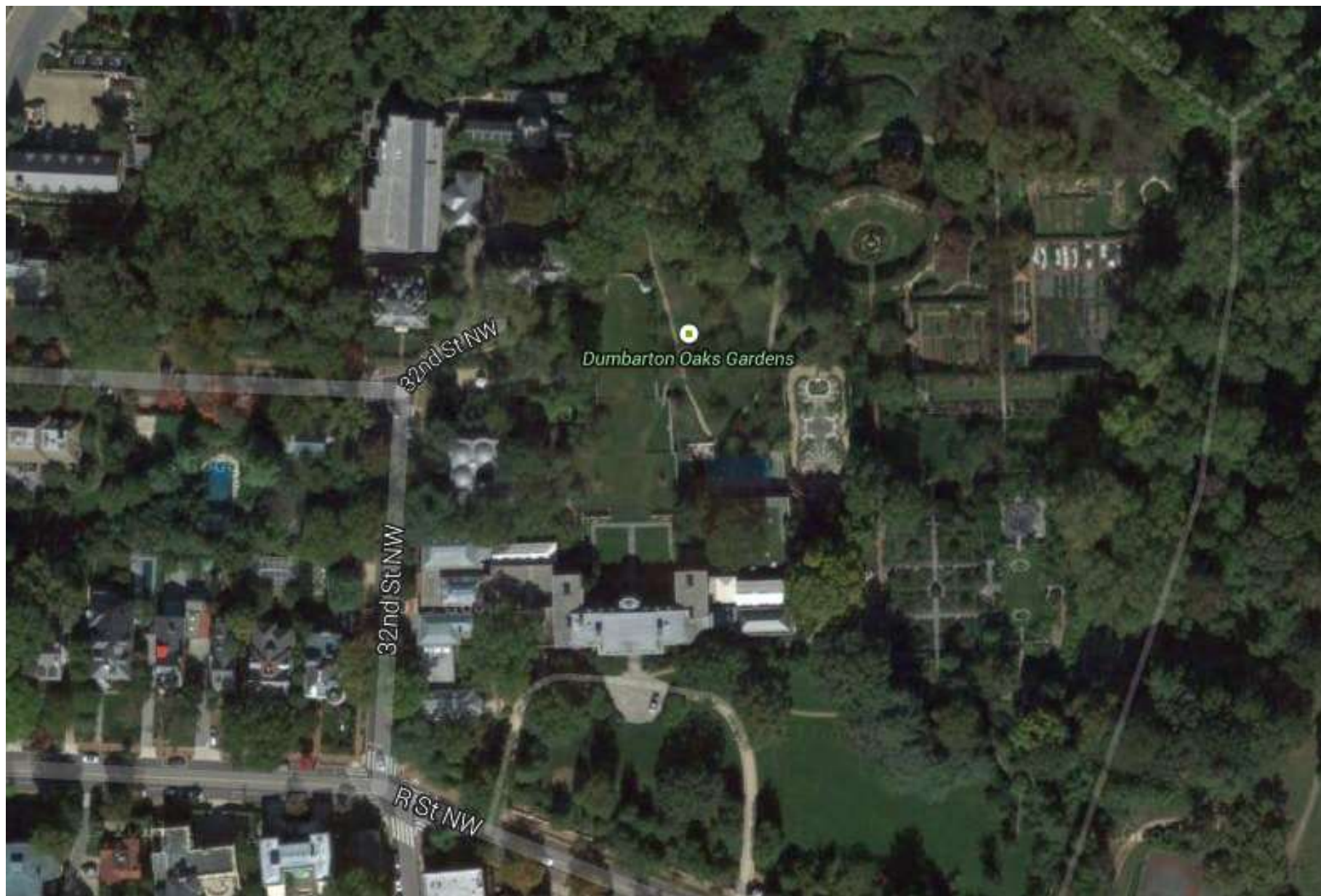
@theglobal_lib

Heidi Elaine Dowding
@theglobal_lib

Heidi Elaine Dowding
@theglobal_lib

There are a lot of physical spaces in which scholars can share their ideas and research…

But no digital equivalent.

Heidi Elaine Dowding

@theglobal_lib

Archival assets are being mixed in with **everyday files**.

Heidi Elaine Dowding

@theglobal_lib

Some great digital assets are getting hidden.

Heidi Elaine Dowding

@theglobal_lib

Ad hoc practices are taking precedent over **best practices**.

jpg					
Profile	**Count**	**Sum**	**Min**	**Max**	**Average**
Untitled-4	62	103026233	4096	2382755	1661713
Profile totals	**62**	**103026233**	**4096**	**2382755**	**1661713**

m4v					
Profile	**Count**	**Sum**	**Min**	**Max**	**Average**
Untitled-4	12	9781258682	4096	2171807947	815104890
Profile totals	**12**	**9781258682**	**4096**	**2171807947**	**815104890**

mov					
Profile	**Count**	**Sum**	**Min**	**Max**	**Average**
Untitled-4	78	667725990437	4096	47377659635	8560589620
Profile totals	**78**	**667725990437**	**4096**	**47377659635**	**8560589620**

mp4					
Profile	**Count**	**Sum**	**Min**	**Max**	**Average**
Untitled-4	40	14676585092	4096	1723401260	366914627
Profile totals	**40**	**14676585092**	**4096**	**1723401260**	**366914627**

Heidi Elaine Dowding
@theglobal_lib

So how can we eat lunch together in the digital realm?

Heidi Elaine Dowding
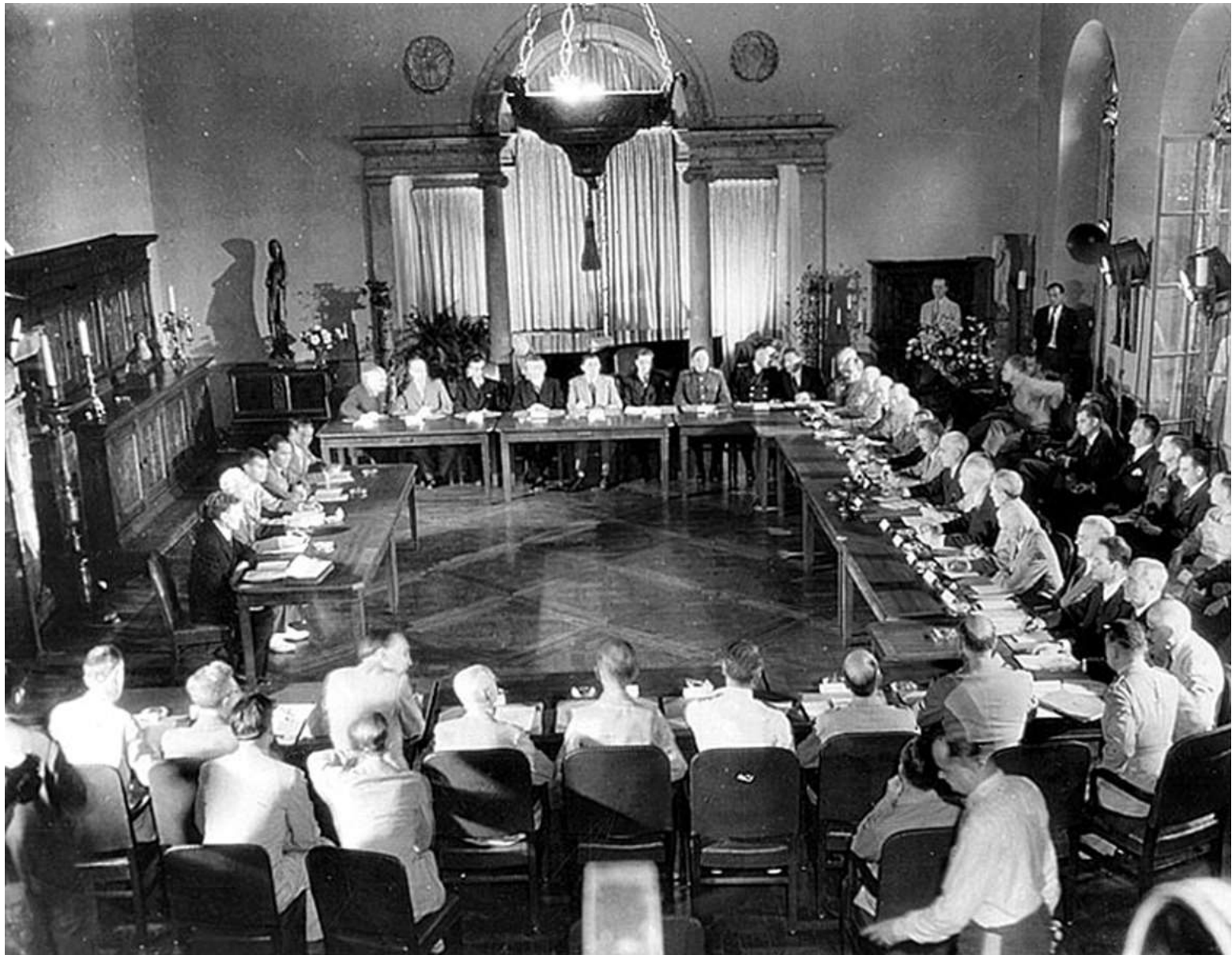
@theglobal_lib

Heidi Elaine Dowding
@theglobal_lib
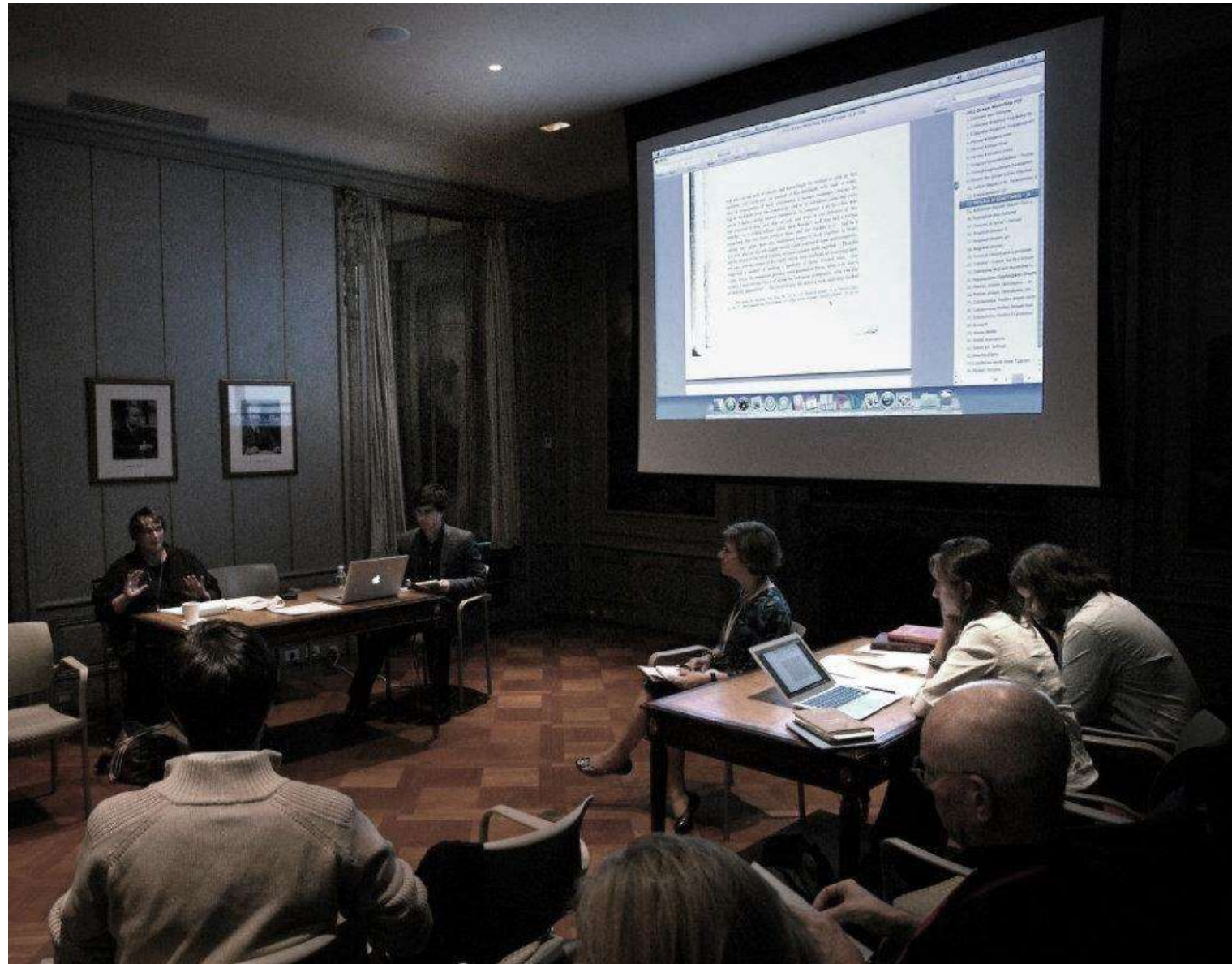
Heidi Elaine Dowding
@theglobal_lib

Thank you.

Heidi Elaine Dowding
@theglobal_lib

Into the [Digital] Wild

Building a Thematic Web Collection at the National Library of Medicine

Maureen McCormick Harlow

Agenda

The Value of Thematic Collections

Using Thematic Collections to Fill Gaps

NDSR@NLM: The Project

NDSR@NLM: The Model

Collecting at NLM's History of Medicine Division

Collection Development Manual

HOME | < PREV | SUBJECTS | FORMATS | NEXT > | INDEX

History of Medicine

Scope and emphasis:

The scope of the Library's collecting effort in the history of medicine is broad, including such topics as the history of the biological sciences; all aspects of the practice of medicine; medical and allied professions; histories of selected hospitals, schools of medicine medical museums, and other health-related institutions; public health; biographies of medical figures; medicine as depicted in art, stamp and coins; medical theories and philosophies; popular and folk medicine; quackery; and the sociology of medicine. NLM also selectively acquires significant works in related subjects to document the sociocultural context of a particular era. Scholarly works are collected without regard to their language or geographical origin.

Selected related works are acquired as needed to provide a cultural context for the study of medical history, as are the reference and bibliographical tools of historical scholarship deemed most important for the NLM mission.

But…

…how can you possibly collect comprehensively on all of those things?

You Can't!!

MT
JEFFERSON

WILDERNESS

WILLAMETTE
National Forest

Using Thematic Collections to Fill Collecting Gaps

NDSR@NLM: The Project

Build and Maintain a thematic Web collection:

"Disorders of the Developing and Aging Brain: Autism and Alzheimer's on the Web"

64 Seeds evenly split, covering 6-7 perspectives:
- –Current Understanding
- –Caretakers (first-person resources, primarily blogs of caretakers)
- –Patients/Sufferers (also first-person, also primarily blogs)
- –Research
- –Causes
- –Treatment
- –Prevention (for Alzheimer's only)

NDSR@NLM: The Model
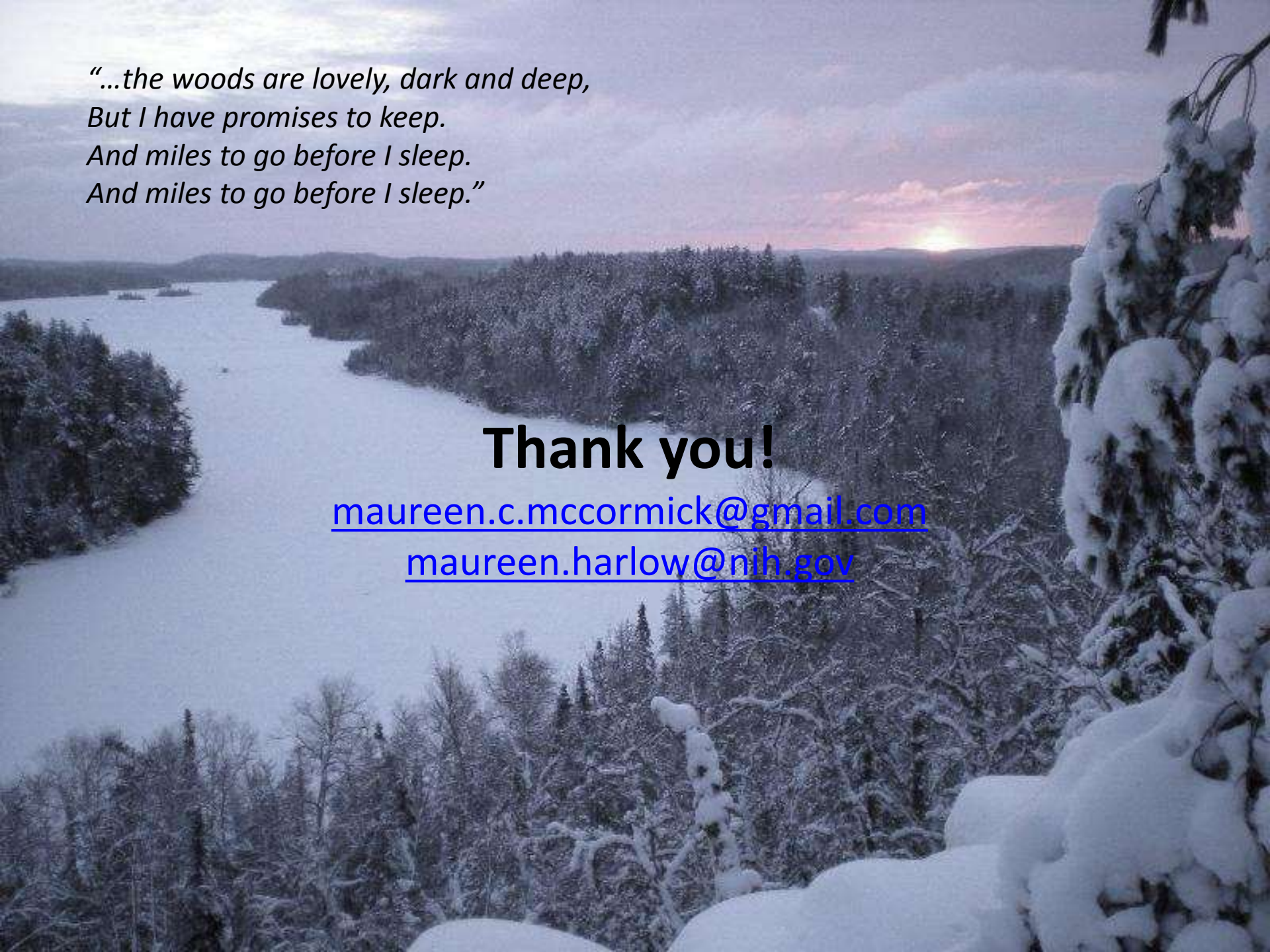
Thematic Collections as "Two Sides of the Same Coin":

Teen pregnancy and infertility   Scoliosis and Osteoarthritis
Diabetes Type 1 and Diabetes Type 2   Eating disorders and obesity

Diagnoses made during development and in old age.

"...the woods are lovely, dark and deep,
But I have promises to keep.
And miles to go before I sleep.
And miles to go before I sleep."

Thank you!

maureen.c.mccormick@gmail.com
maureen.harlow@nih.gov

Developing a Digitization Initiative at PBS
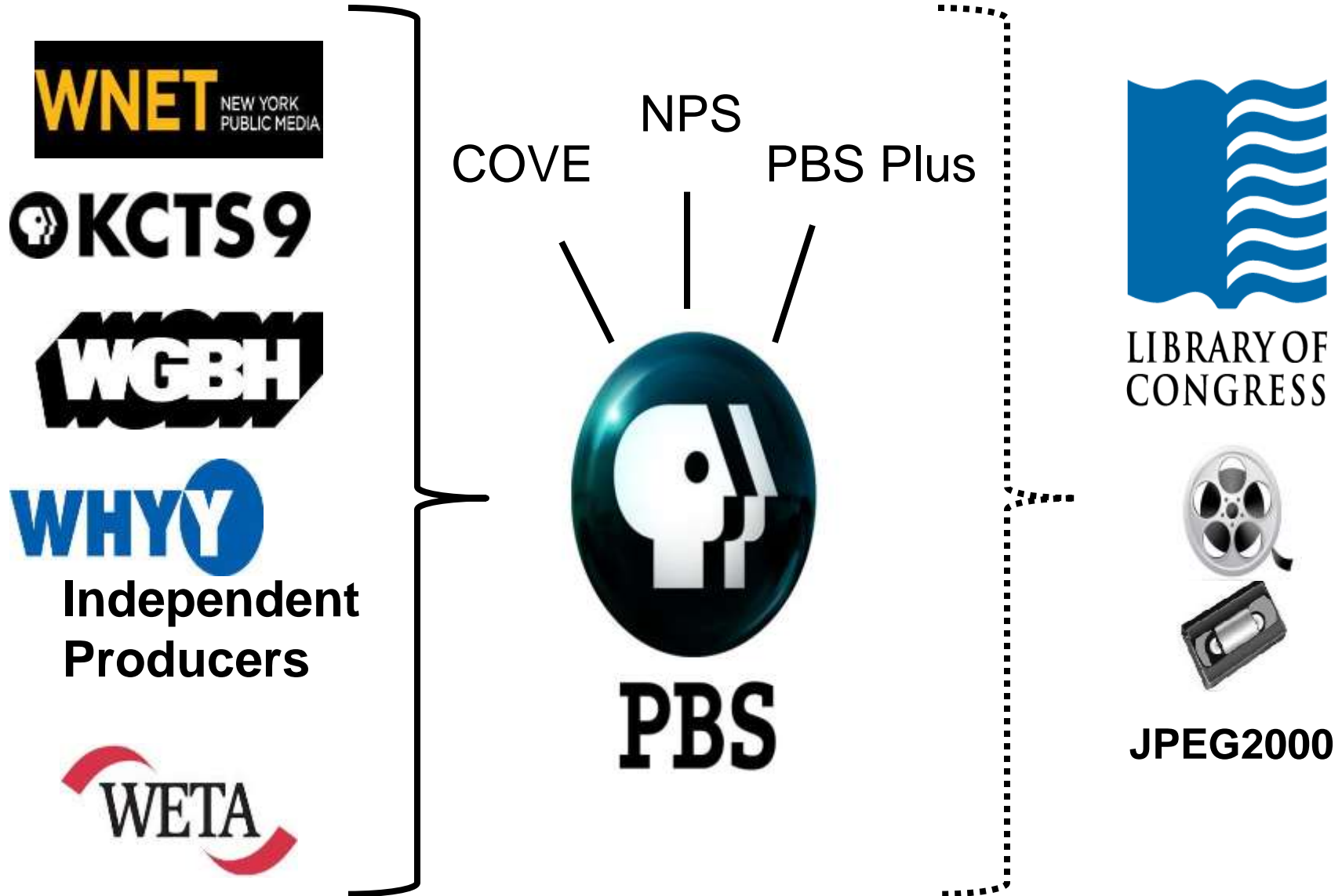
Lauren Work
ALA Midwinter
1/26/2014
@squaredsong | lawork@pbs.org

Produce

WNET NEW YORK PUBLIC MEDIA

KCTS 9

WGBH

WHYY

Independent Producers

WETA

Distribute

NPS

COVE

PBS Plus

PBS

Deposit

LIBRARY OF CONGRESS

JPEG2000

Project Description

1. Develop selection criteria for at-risk media

2. Create digitization workflow for selected media

3. Policy update recommendations

1. Selection Criteria

- Title held on multiple formats
- Unique title (emphasis on orphans)
- Copyright and descriptive metadata available
- Title fits with current PBS/LC deposit policy
- **Title held elsewhere**

Outreach and Collaboration

UMD/PBS Collection overlap

UMD National Public Broadcasting Archives

4%

PBS 1" and 2" Media Collection
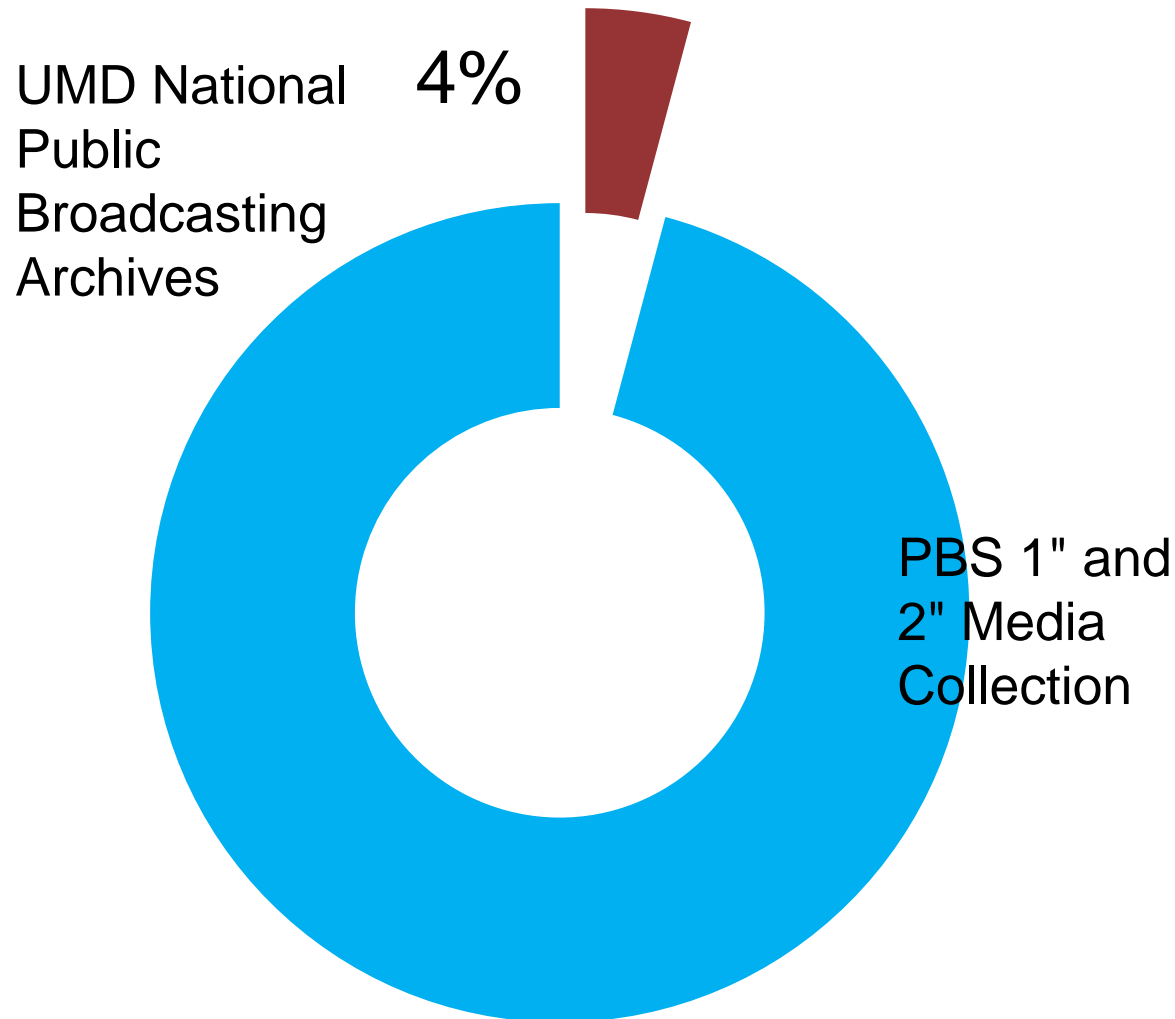
2. Digitization Workflow

-Digitization trial run

-Specifications for digital files & storage

-Access and use cases

Next steps

1. Selection Criteria
2. Workflow
3. Policy

DEVELOPING AN ACCESS MODEL FOR BORN-DIGITAL COLLECTIONS

National Digital Stewardship Residency
University of Maryland Libraries and
Maryland Institute for Technology in the Humanities
ALA Midwinter 2014
Margo Padilla
@margo_padilla

CURRENT METHODS

- Case-by-case

- Onsite access vs. remote access

- Access not provided

- Consider wide access first and deep access second

- Work within existing models

- Documentation, knowledge sharing, transparency

GETTING STARTED

IDENTIFY YOUR COLLECTIONS

- Low hanging fruit

- Prioritize

- Anticipate user interaction

- No archive or library is an island

IDENTIFY YOUR USERS

- Survey

- Interview

- Embedded liaison

CREATE YOUR ACCESS POLICY

- Policy and Reference Manual

- Transparency and Documentation

- Design an experience

THANK YOU

National Digital Stewardship Residency
University of Maryland Libraries and
Maryland Institute for Technology in the Humanities
ALA Midwinter 2014
Margo Padilla
@margo_padilla