

Digital Preservation Update

ACCESS 2011 - Vancouver



“uplifting the whole people”

— HENRY MARSHALL TORY, FOUNDING PRESIDENT, 1908

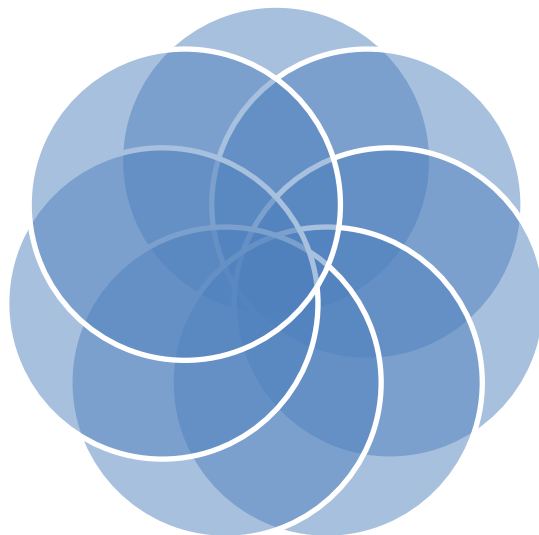
A chicken's view of why data loss matters.



Examples of digital assets include:

**250,000+ IA digitized
books and archival
materials (~70 million
pages)**

**8,000+ prairie monos,
16,000+ images,
1000+ maps**



**100+ newspaper titles
(550,000 pages)**

ejournals (licensed)

**eTheses, OJS, preprint
and other IR content**

**eResearch/eScholarship projects
e.g. Canadian Writing Research
Collaboratory (CWRC);
Editing Modernism in Canada
(EMiC)**

**Research data – e.g. Polar
Year IPY-DACN; Data
Library; census materials**



“uplifting the whole people”

— HENRY MARSHALL TORY, FOUNDING PRESIDENT, 1908

500 year commitment to long term access.

What this means:

182,500 days of ensuring short term access.

Current preservation storage capacity: ~52 TB usable with ~18 TB currently in use – EOL'd

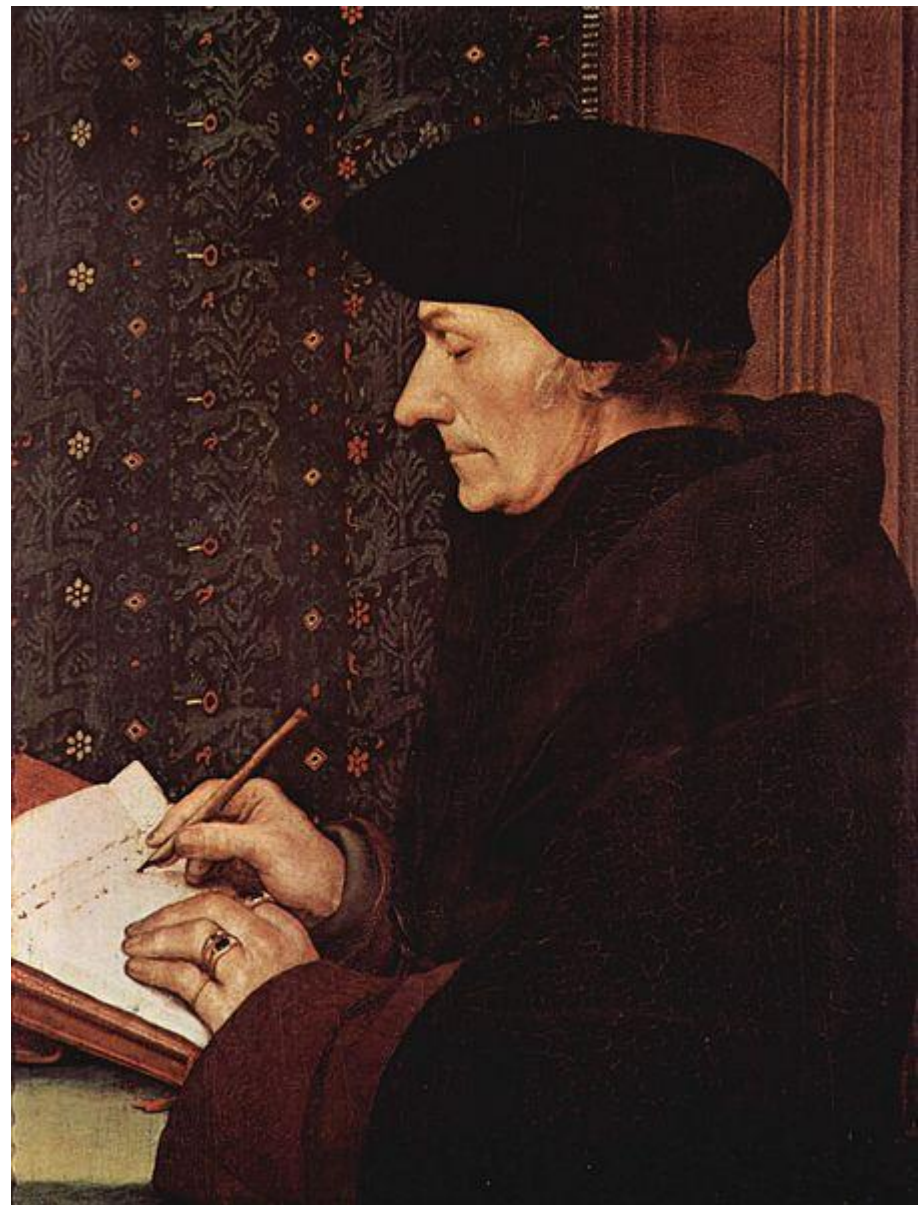
“Only storage vendors with a demonstrated record of 0% data loss since 1511 need apply.”

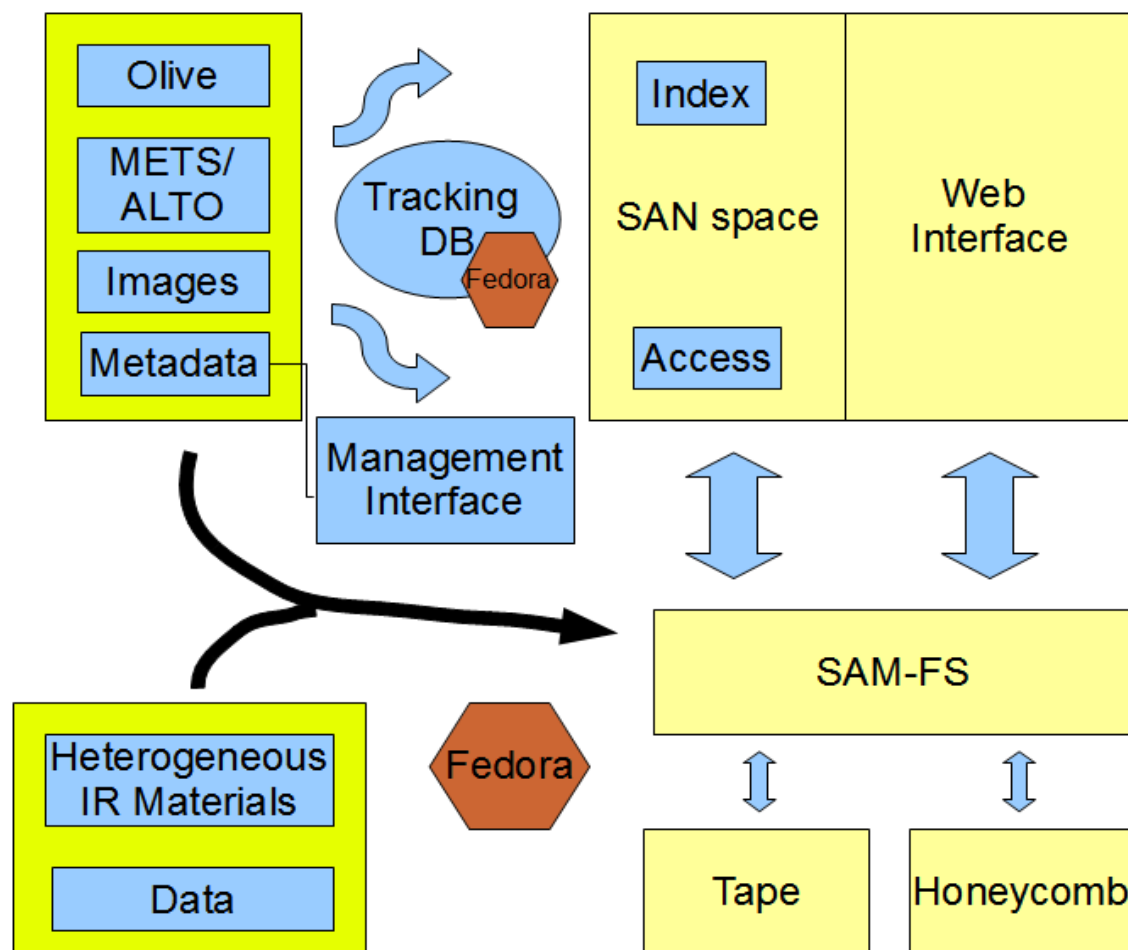
- P. Binkley, 2011

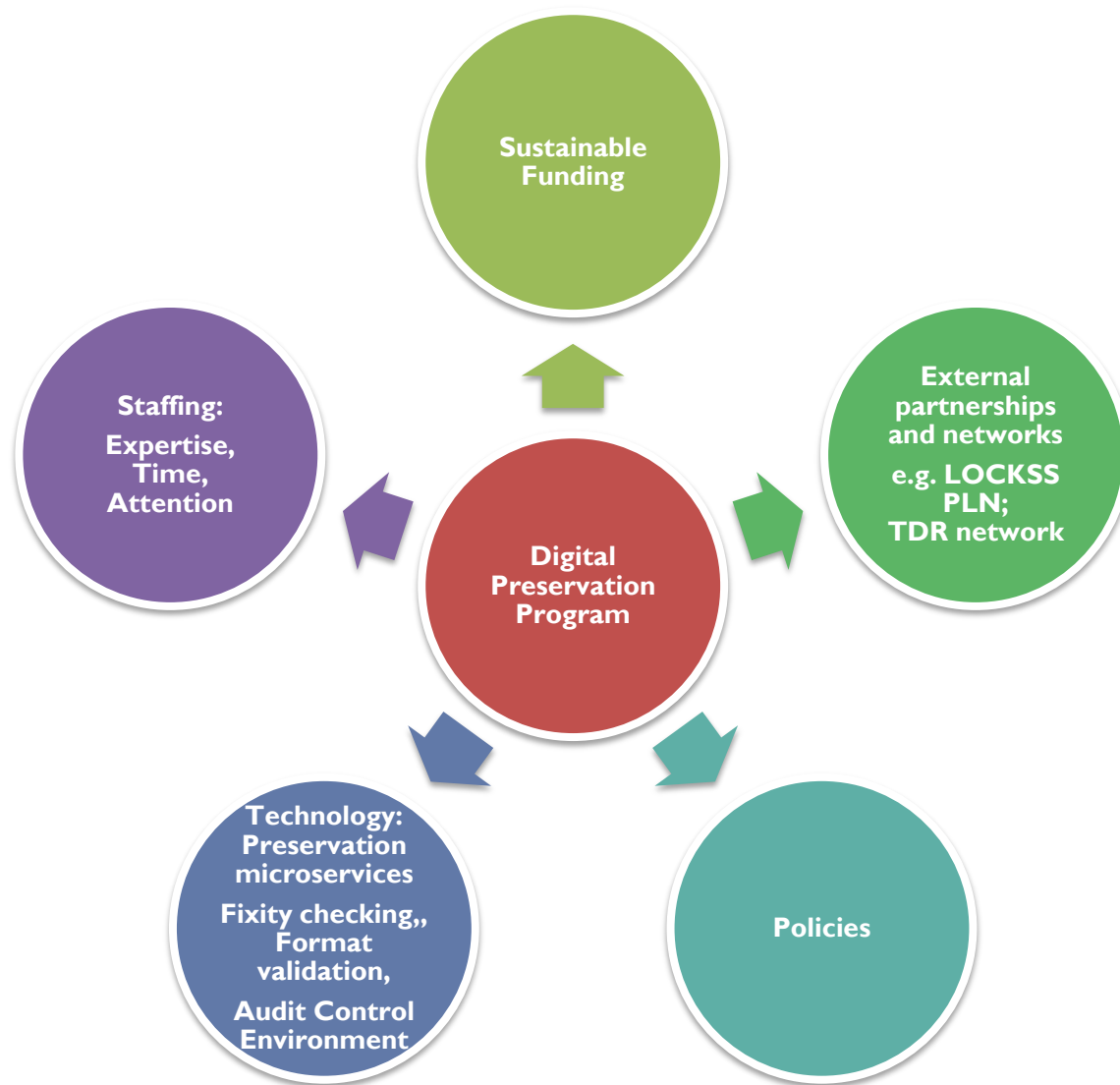
Preservation storage: scalability, predictability and sustainability

2012 = will add 300 TB

2014 = expecting to grow to 500 TB or more depending on projects and partnerships







Held multi-day planning workshop with CRL on TDR criteria and accreditation process

Program planning underway > goal is to learn from focused implementation and scale outwards

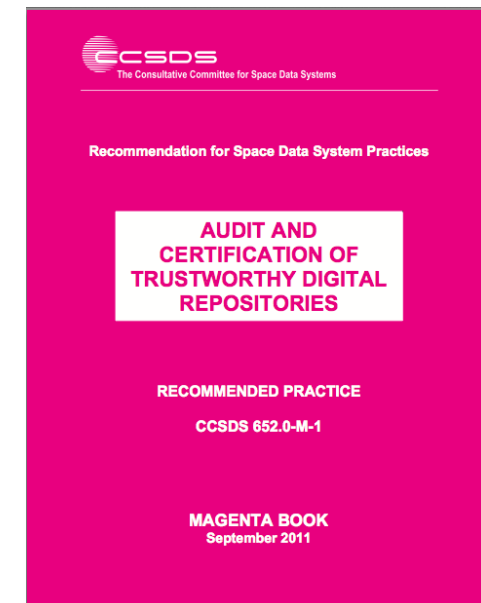
UAL Preservation Program includes TDR but also gold, silver, bronze(?) level preservation efforts

(Gold = must survive)

(Silver = most will survive)

(Bronze = we won't go out of our way to destroy it)

Created new position: Digital Preservation Officer
– policy creation, audit, documentation, adherence, and herding of cats.



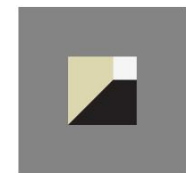
- **Preservation planning is always front and centre as we become involved in digital projects.**
- **Use OAIS reference model terminology to ensure our SIPs, AIPs and DIPs do what they need to do.**
- **Apply preservation actions and microservices where it makes sense to do so.**
- **Multiple preservation strategies/streams employed to reduce risk and the issue of single points of failure.**
- **Preserve what matters (preservables vs derivatives; can't keep everything...)**

Electronic journal content (incl. OJS and other materials) preserved via:

- **LOCKSS**
- **COPPUL LOCKSS PLN**
- **CLOCKSS**
- **LOCKSS-USDOCS**
- **Portico**



Archived In



PORTICO



Example: International Polar Year – Data Archive Canadian Network

iRODS preservation backbone

- High speed **CANARIE** dedicated fiber between **UAlberta** and **OCUL/UofT**



Need addressed: Preservation replication service for grant funded **Canadian polar research**

- Governance and sustainability (economic and otherwise) – multiple nodes
- Messy data makes for messy work
 - Bit level preservation is achievable
 - Preserving meaning and usefulness is a whole other challenge...
- Gaining experience with eScience programs and future research data preservation projects

Access & Dissemination

IPY-DACN

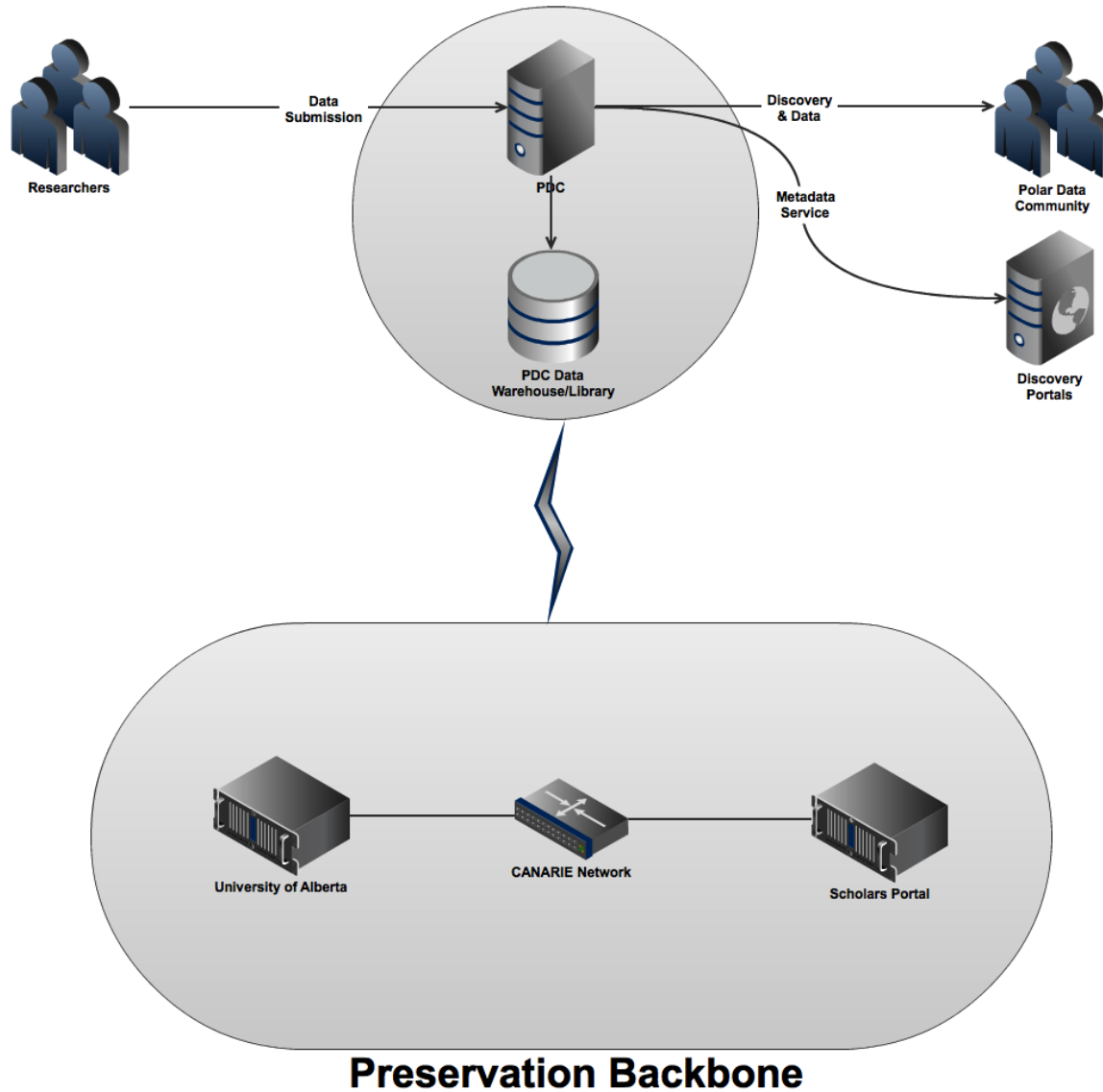


Diagram courtesy of Steve Marks, OCUL

Web Archiving

Using Archive-It subscription



Plans include:

- Developing further collections in areas related to W. Canada, Circumpolar, Gov Docs, current/significant events
- Discovery strategy – metadata; access points
- Locally preserve (and possibly host) data in 2012
 - LOCKSS plug-in for Archive-It
 - Local version of Heretrix

Current project:

1. Inventory collections

2. Determine research or other value - criteria

3. Determine storage/access media migration strategy:

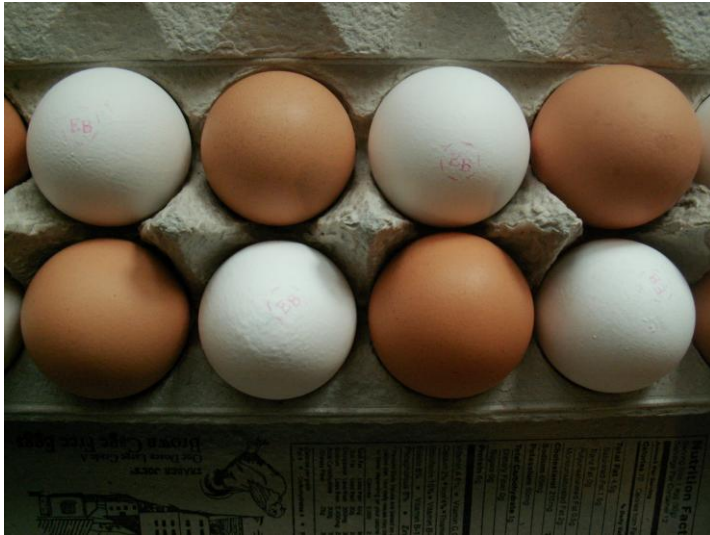
- Avoid moving from obsolete to soon-to-be obsolete

•Considerations:

- bit level preservation and emulation
- file format migration
- cost
- IP issues



No eggs were harmed (yet) in the making of this preservation plan.



geoffrey.harder@ualberta.ca

www.library.ualberta.ca/digitization/preservation

Digital Preservation @ SFU

Mark Jordan

Just getting started

- Starting to think systematically about digital preservation
- Digital preservation and digital curation are in the Library's new 5-year plan
- We're not starting with a platform, we're trying to start with policy and practice

Library can't do it alone

- IT Services
- SFU Archives
- VP Research
- SFU Library

University needs to think strategically

- Value of institutional history
- Direct \$ value of digital content (e.g. Library or other digitization that is not institutional 'records')
- Risk management (e.g. research data management)
- Economies of scale and economies of collaboration

Existing activity

- Bit-level preservation
- Institutional Repository
- Theses Management System
- LOCKSS
- Workflow integration proof of concept
- SFU Archives' pilot

Emerging collaborations

- Archives' pilot
 - Goal is to preserve ex-President Stevenson's email plus some departmental records
 - IT Services and Library are involved
- New IR (Summit)
 - User, not Library focus
 - Agile platform for integrating with preservation infrastructure
- CARL CFI data management proposal

Current areas of interest

- How to handle video
 - popular and big
- How to handle data
 - also popular and big, with interesting use cases
- Preservation metadata
- Digital-only theses
 - And associated content like video and data

Digital Preservation On The East Coast

Memorial University of Newfoundland

Slavko Manojlovich

Associate University Librarian (IT) / Manager, Digital Archives Initiative
Memorial University of Newfoundland

ACCESS
CONFERENCE 2011

OCTOBER 19-22
VANCOUVER, BC
THE LIBRARY IS OPEN

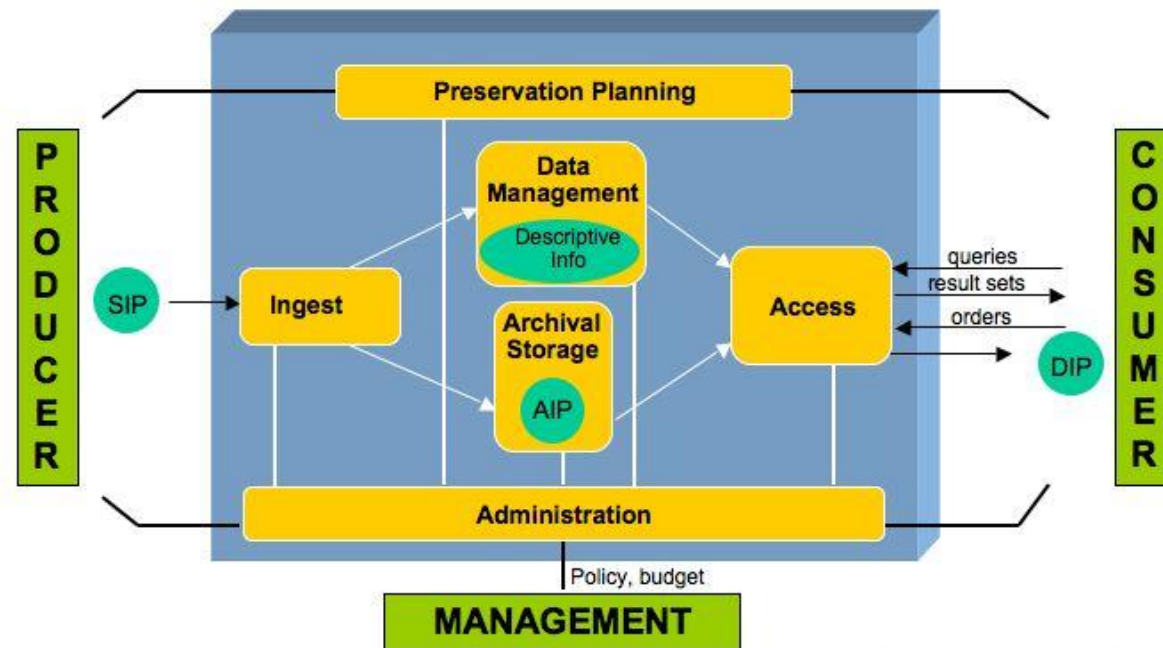
Current Environment [October, 2011]

- Software: ContentDM, EPrints, OJS, [Archivematica]
- OS: Windows, Ubuntu, MAC OS
- Collections: 150
- Content: theses, videos, audio, maps, journals, monographs, newspapers, images, handwritten diaries with transcripts, 3D models, EADs, [research data]
- No. of digital objects: 1.4 million
- Disk storage: 7+ TB
- Preservation: bit-level, system agnostic [BagIt], logical
- Logical Preservation ➔ Media Type Preservation Plan

OAIS Reference Model



OAIS Functional Entities



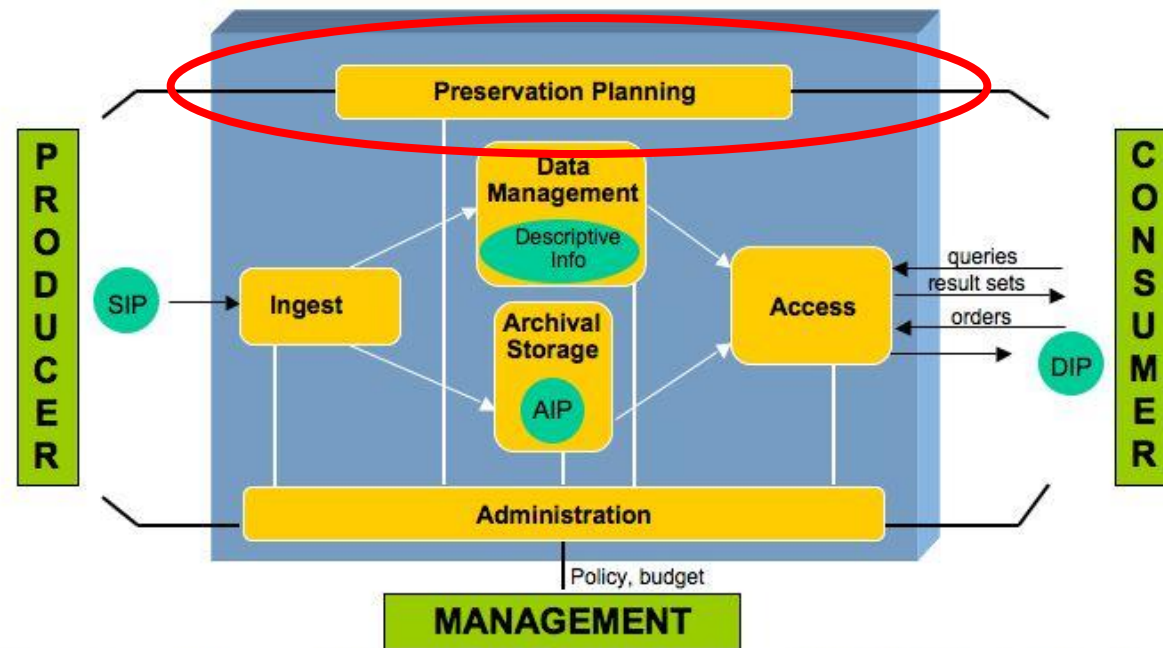
SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package

The OAIS Environment
from 10,000 ft

OAIS Reference Model



OAIS Functional Entities



SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package

The OAIS Environment
from 10,000 ft

DROID Inventory of File Types

Formats/Risks

Request File Type Recount

Format Breakdown

Portable Document Format (Version 1.4)	17
Portable Document Format (Version 1.5)	14
ZIP Format	8
Portable Document Format (Version 1.3)	7
Microsoft Office Open XML (Version 2007)	2
Portable Document Format (Version 1.6)	2
Microsoft Powerpoint Presentation (Version 97-2002)	2
Microsoft Word for Windows Document (Version 97-2003)	1
Portable Document Format (Version 1.2)	1
Acrobat PDF/A - Portable Document Format (Version 1)	1
OLE2 Compound Document Format	1

Last Update: Tue Oct 11 17:47:27 2011

Media Type Preservation Plan

Media type	Supported Ingest File Format Extensions	Long-Term Preservation Format(s)	Access Format(s)	Normalization tool
Audio	MP3, WMA, WAV	WAVE (LPCM)	MP3 and WMA	Switch Sound File Converter version 4
Portable Document Format	PDF, PDF/A	PDF/A searchable	PDF/A searchable	PDF/A Manager (PDFTRON)
Presentation files	PPT	PDF/A searchable Original file with embedded fonts. ODF	PDF/A searchable, original file with embedded fonts	Unoconv/OpenOffice, PowerPoint or Adobe Pro

http://archivematica.org/wiki/index.php?title=Media_type_preservation_plans

Preservation Planning

- Monitor designated community
- Monitor technology
- Develop preservation strategies and standards
- Develop packaging designs and migration Plans

[Digital Preservation Management: Implementing Short-term Solutions for Long-term Problems](#)

Monitor Technology

Cross-Platform Access Video Format

- 2005: wmv (Windows Media Video) format using Windows Media Player (or other players) for Windows and Flip4MAC Quicktime extension for Macintosh.
- 2005 – 2009: swf (Adobe Flash) format with Adobe flash plug-ins available for Windows and Macintosh browsers becomes the flavour of the day for web delivery of video content.


Monitor Technology

Cross-Platform Access Video Format

- Fast forward to April, 2010: mp4 (H.264) format with players/support for Windows, Macintosh and **IPAD**.
- IPAD does not support wmv or swf video formats.
- Video conversion history: wmv→swf→mp4 from original DVD vobs.
- DVD vob files are being preserved with a goal of converting them to MXF Motion JPEG 2000 for long-term preservation.

Monitor Technology


Google Drops H.264 in Favour of WebM and Theora Codecs (Jan 11, 2011)

**VCC**
VIDEO-COMMERCE.ORG

The Video Commerce Consortium
advancing video in e-commerce

VCC Blog Jobs About the VCC *Members-Only Discussion Forum*

Google Chrome To Drop H.264 – How Should Video Marketers Prepare?



THE SEARCH ENGINE WHO SHAGGED ME

Google announced this week that they are dropping support for the H.264 video codec in their Chrome Web Browser, making the blogosphere buzz with debate and speculation as to what this means for the online video industry. Who will gain? Who will lose? And more importantly, what should web video marketers do to prepare for this new development?

[Source](#)

Monitor Technology

Microsoft Adds H.264 Support (Feb 2, 2011)



The screenshot shows the PCWorld website interface. At the top is a red navigation bar with the PCWorld logo and links for News, Reviews, How-To's, Downloads, Shop & Compare, Apps, and Business Center. A search bar and links for Sign in, Join Us, and Newsletters are also present. Below the navigation bar is a dark banner with a PCWorld magazine subscription offer on the left and a Cisco advertisement for business networks on the right. The main content area features a breadcrumb trail 'PCWorld » Blogs » Geek Tech' and social media sharing options (Tweet, Digg, Like, Comments, etc.). The article title is 'Microsoft's Got H.264's Back, Releases Plug-in for Chrome Users' by Nick Mediati, dated Feb 2, 2011 1:09 PM. The article text discusses Google's decision to discontinue H.264 support in Chrome and Microsoft's response with a plug-in. A Google Chrome logo is shown next to the text. To the right of the article are two sidebar widgets: 'REMOTE MANAGEMENT' and 'LAPTOP SHOWCASE'. At the bottom right is a large Cisco Webex advertisement with the text 'make meetings WEBEX' and a 'TRY IT FREE' button.

PCWorld » Blogs » Geek Tech

Tweet 1 2 Digg Like 1 0 Comments +2 Recommendations Email Print

GEEK TECH

Microsoft's Got H.264's Back, Releases Plug-in for Chrome Users

By Nick Mediati, PCWorld Feb 2, 2011 1:09 PM

Google made waves in the Web geek community last month when it announced that it would be [discontinuing support for H.264 video](#) in its Chrome Web browser. While some supported the move, others were concerned that it [would hold back HTML5 adoption](#). But Microsoft has [released a plug-in for Google Chrome](#) that will give you H.264 video support even after Google pulls the plug.

The [Windows Media Player HTML5 Extension for Chrome](#), as Microsoft calls it, will work with Chrome on Windows 7, though this means that users of older versions of Windows, or Chrome on Linux or Mac OS X, will still be out of luck, but it's better than nothing.

If you're game, hit up Microsoft's site and [download the plug-in](#).

The HTML5 video format situation is still a mess, and it'll be a while until it's all sorted out. But hopefully the main browser vendors can get together and agree on a single format sooner rather than later.

REMOTE MANAGEMENT
Manage remote devices, workers and offices.
[Learn how.](#)

LAPTOP SHOWCASE
Balance Performance and Portability with Lenovo IdeaPads and Thinkpads.
[See special offers.](#)

make meetings
WEBEX
[TRY IT FREE](#)
Cisco webex together we are the human network. CISCO

[Source](#)

Monitor Technology

“That puts Google at odds with Microsoft, which has [publicly declared its support for H.264](#) as the default video codec in IE9. More importantly, it puts Apple between a rock and a hard place. Google still publicly supports Adobe Flash, which offers a supported path for developers to deliver H.264 encoded content in Google’s browser. But Apple’s bitter public feud with Adobe means it has banned Flash from all iOS devices, leaving H.264 as the only supported codec.”

[Source](#)

Digital Preservation Strategy Pilot Program Implementation

Bronwen Sprout, Digital Initiatives Coordinator, UBC Library
With input from Artefactual Systems
and Sarah Romkey, Rare Books and Special Collections Archivist, UBC Library



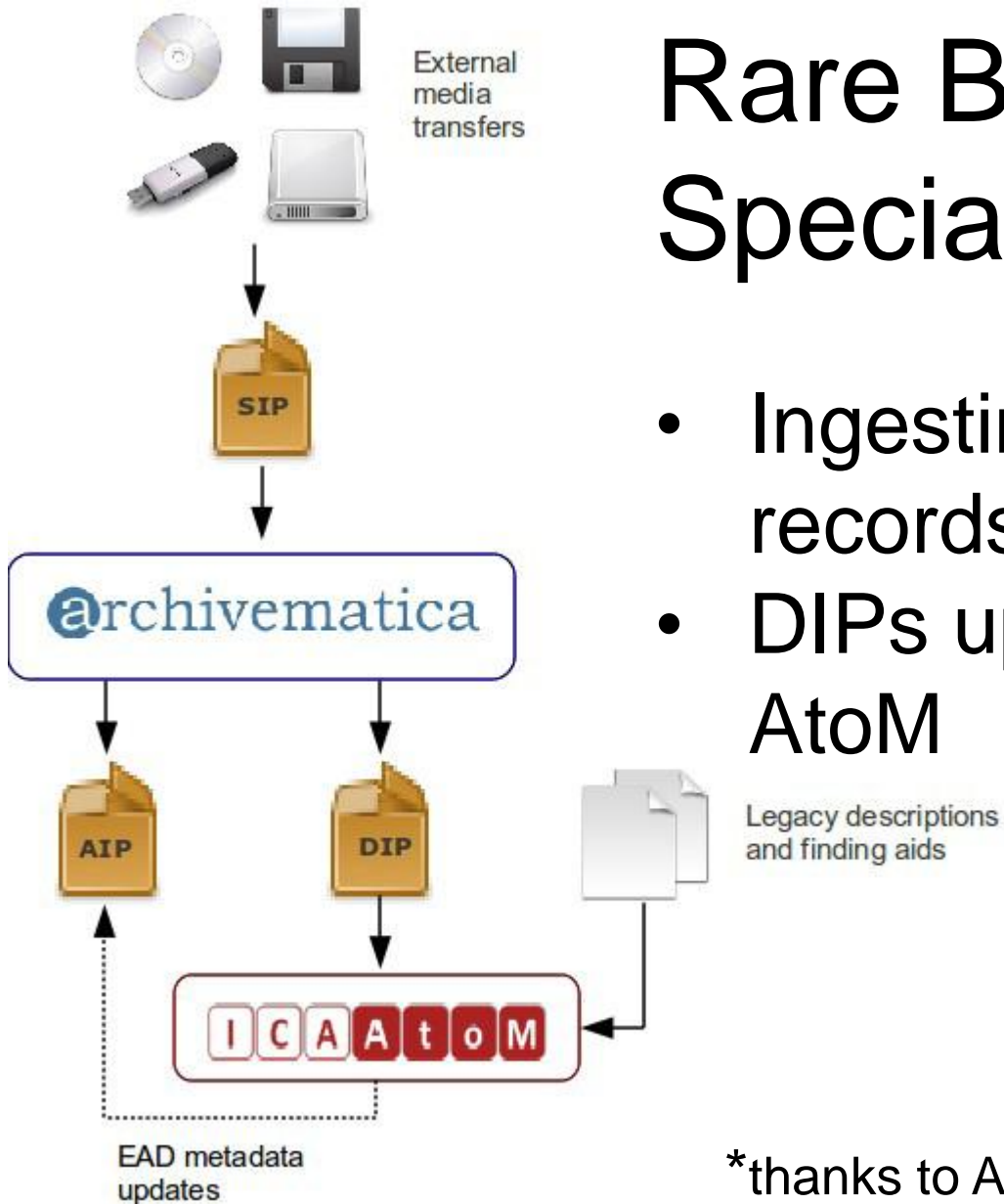
a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA
Library

Project Components

- archivematica install and training
- Gap analysis
- Strategy and systems architecture
- Findings will be openly accessible when complete

Comprehensive and flexible

- Rare Books and Special Collections
- University Archives
- cIRcle digital repository (DSpace)
- Digitization projects (CONTENTdm)
- Website archiving
- Data preservation

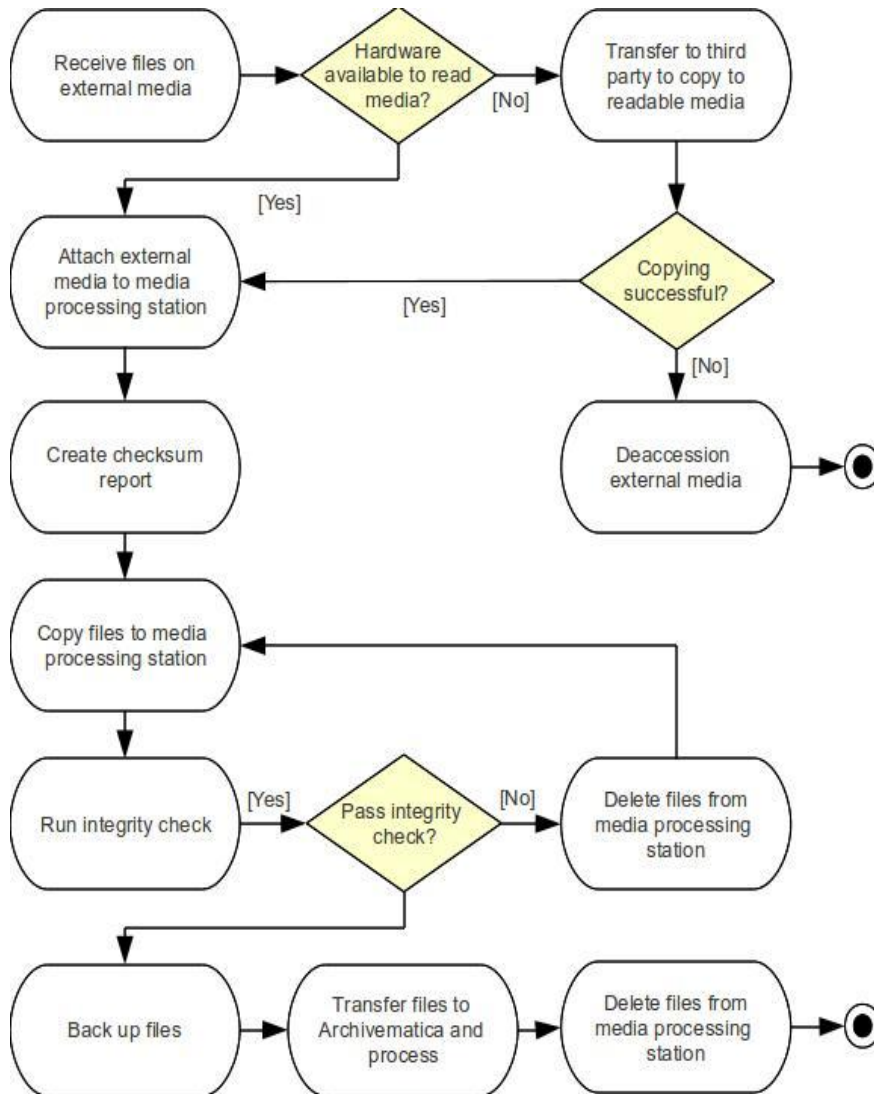


Rare Books and Special Collections

- Ingesting born digital records into archivematica
- DIPs uploaded into ICA-AtoM

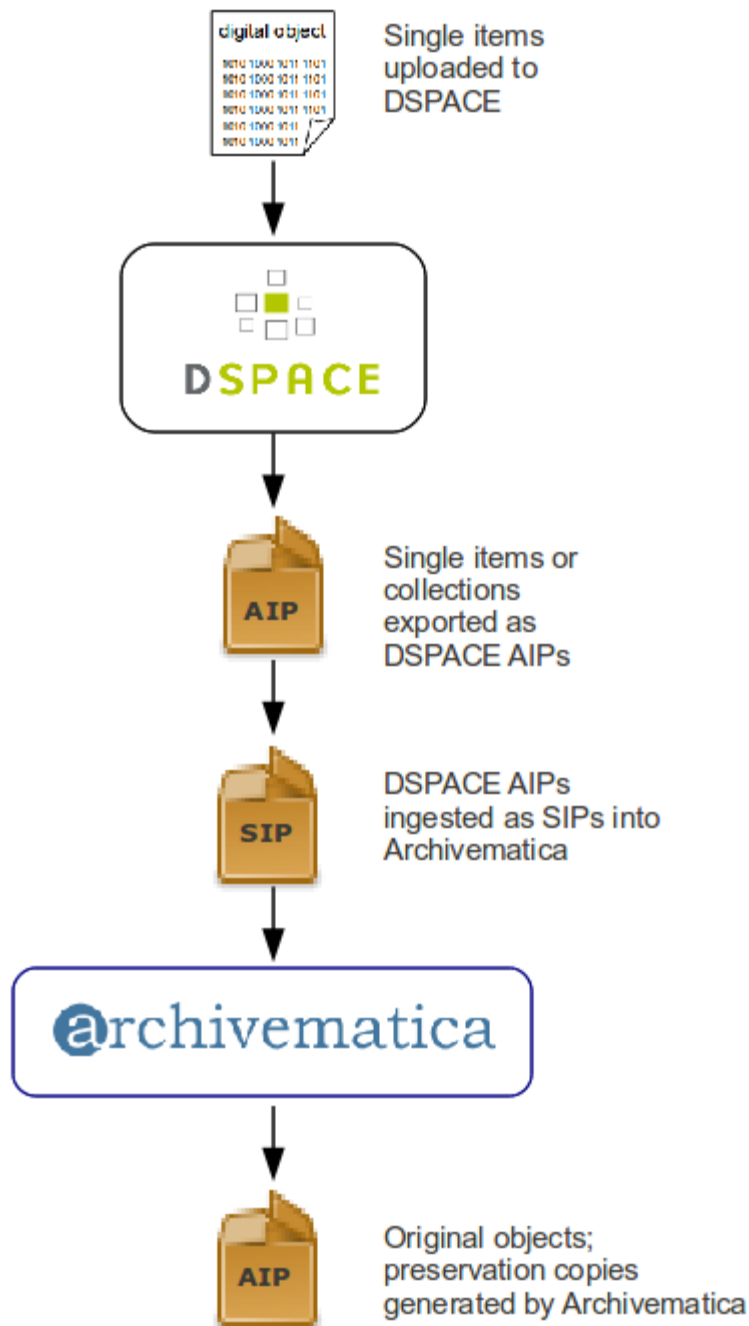
*thanks to Artefactual for all diagrams

External media station workflow



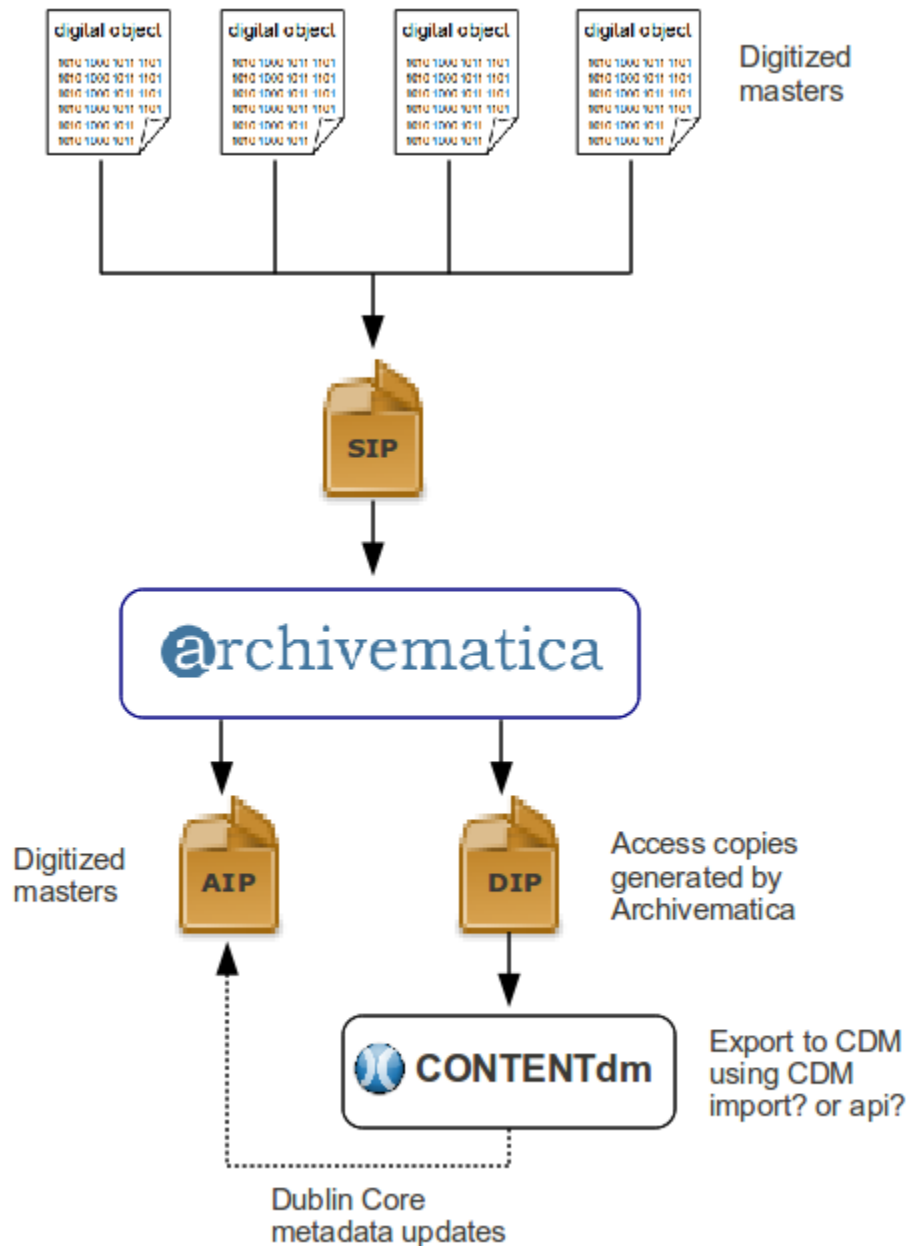
RBSC findings

- Technical problems
 - File failures
- Intellectual issues
 - Appraisal
 - Accessibility
 - Arrangement



clRcle

- Receive SIPs from DSpace



CONTENTdm

- Integrate DIPs from Archivematica

Some of the next steps

- Complete pilot projects – DSpace and CONTENTdm
- Undertake migration projects
- Website archiving - using Heritrix and Wayback Machine
- Research data
- LOCKSS PLN – how does it fit?
- TDR certification prep