

The Community-Driven Evolution of the Archivemática Project

Peter Van Garderen and Courtney C. Mumma
IPres Toronto, October 4, 2012



Peter Van Garderen, MAS
President / Systems Analyst
@pvangarderen



Evelyn McLellan, MAS
Director,
Consulting Services



David Juhasz
Director,
Technical Services



Courtney Mumma, MAS
Product Manager,
Archivematica



Jessica Bushey, MAS
Product Manager,
ICA-AtoM



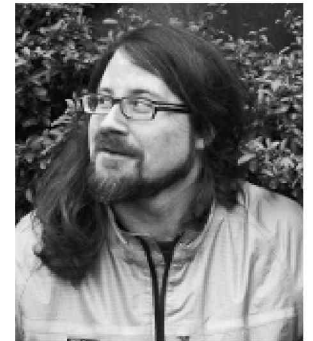
Austin Trask
Systems Engineer



Jesús García Crespo
Software Engineer



Joseph Perry
Software Engineer



Mike Cantelon
Software Engineer



Artefactual Systems is located in Greater Vancouver on Canada's west coast. The easiest way to reach us is via email. We work with [clients](#) worldwide:



[View Larger Map](#)

[Street View](#)

Overview

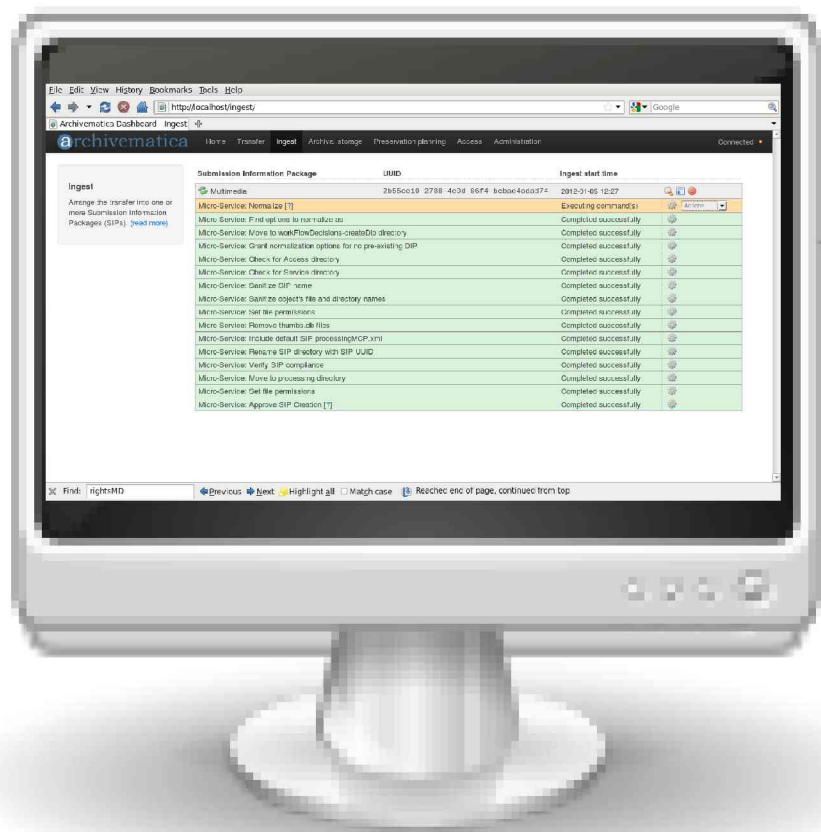
- What is Archivemata?
- How is the Archivemata project managed?
- Community-driven feature development

What is Archivemata?

- free and open-source digital preservation system (AGPLv3)
- designed to maintain standards-based, long-term access to collections of digital objects

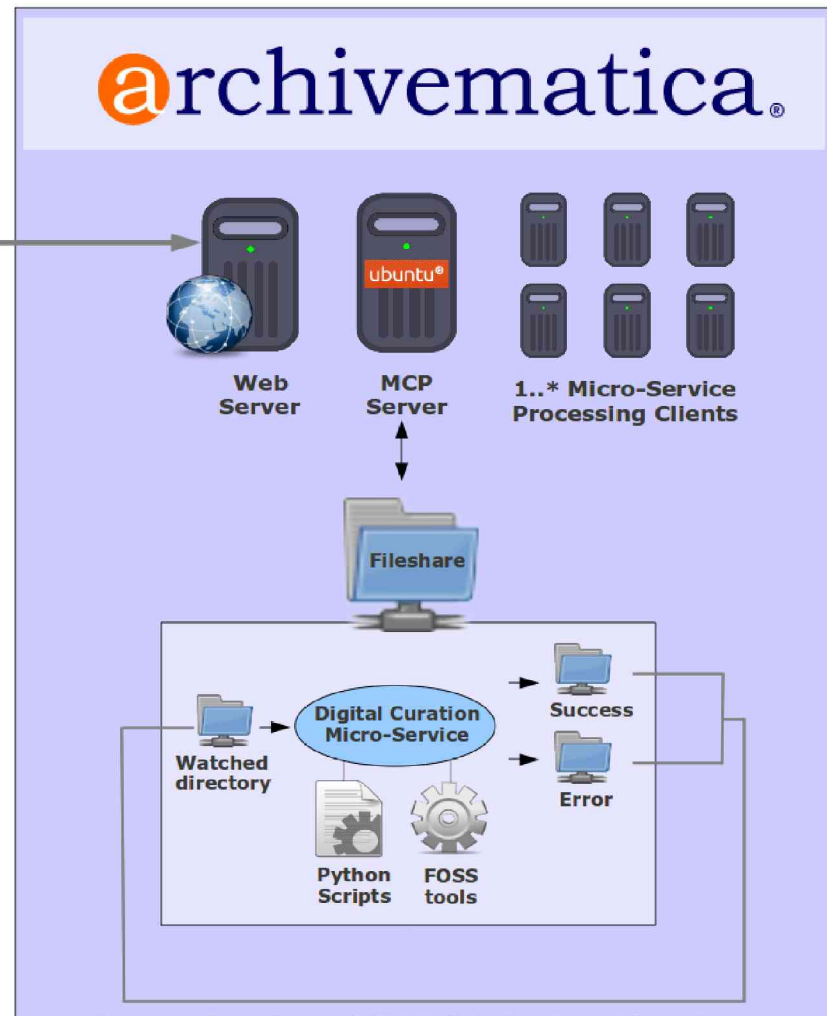
What is Archivemática?

- allows users to process digital objects from ingest to access in conformance with the ISO-OAIS functional model
- Archivemática implements format normalization upon ingest and preserves originals to support emulation and migration strategies



**MANAGEMENT
Web Dashboard**

monitor and control



or

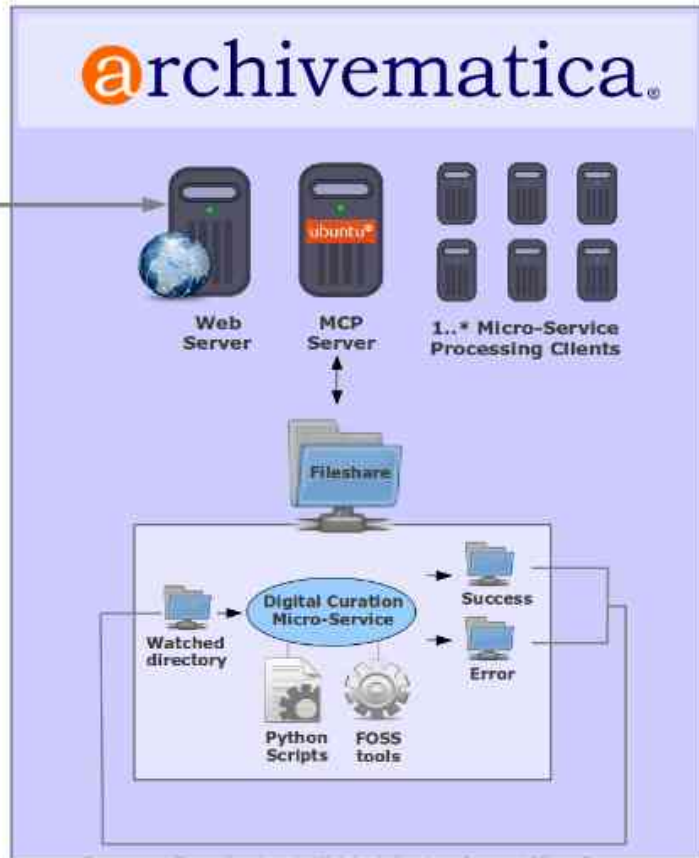
transfer of
digital files
& metadata





MANAGEMENT Web Dashboard

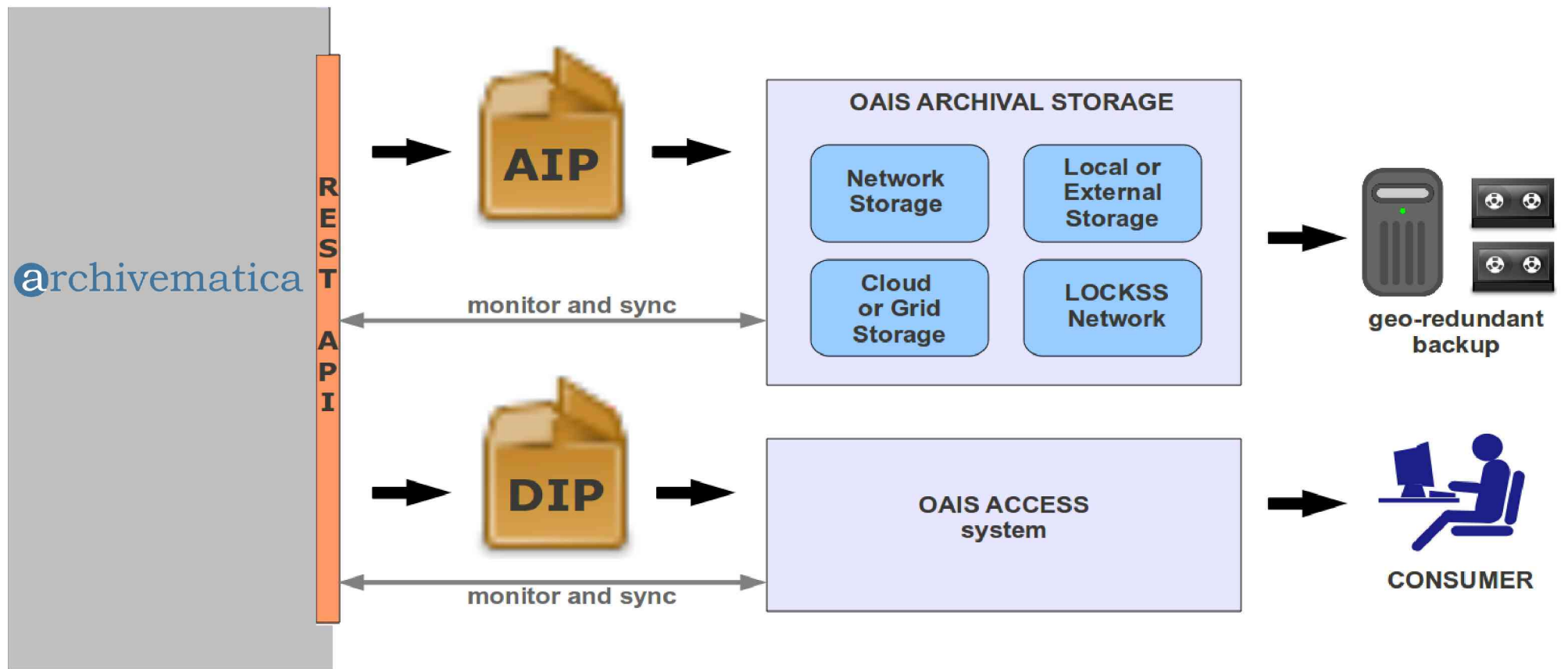
monitor and control



Transfer	UUID	Transfer start time	
Pictures ▶ Micro-service: Approve transfer Job: Approve transfer [?]	ba113e3a-f809-462d-8e4d-34f7fb4f7640	2012-08-15 10:15	
		Awaiting decision	<input type="button" value="Actions"/>
Mail dear ▶ Micro-service: Create SIP from Transfer ▶ Micro-service: Complete transfer ▶ Micro-service: Characterize and extract metadata ▶ Micro-service: Clean up names ▶ Micro-service: Scan for viruses ▶ Micro-service: Extract packages ▶ Micro-service: Generate METS.xml document ▶ Micro-service: Verify transfer checksums ▶ Micro-service: Assign file UUIDs and checksums ▶ Micro-service: Include default Transfer processingMCP.xml ▶ Micro-service: Rename with transfer UUID ▶ Micro-service: Approve transfer	4bb41b9e-1c6b-4cb1-9e41-56fe0ba3dd64	2012-08-14 15:44	
Bad names ▶ Micro-service: Create SIP from Transfer ▶ Micro-service: Complete transfer	747da9a9-46f6-4d77-bf98-caa7d5ee29f8	2012-08-14 14:32	

What is Archivemática?

- Archivemática is a processing pipeline
- Archivemática is designed to output high-quality, standards-compliant Archival Information Packages
 - Bagit, METS, PREMIS



The METS file

<dmdSec> (descriptive metadata)

Dublin Core XML

<amdSec> (administrative metadata)

<techMD>

PREMIS: object

<digiProvMD>

PREMIS: events

PREMIS: agents

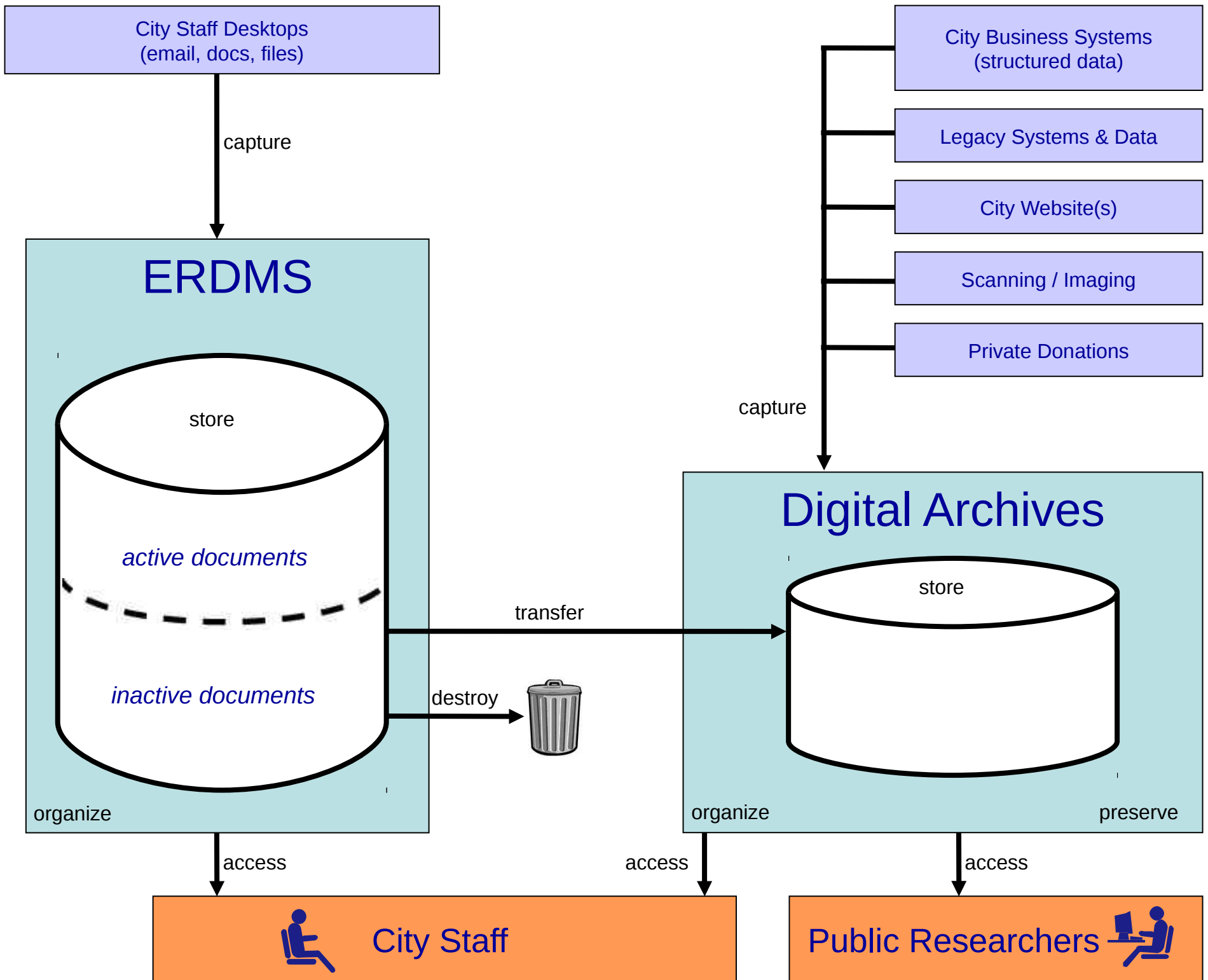
<rightsMD>

PREMIS: rights

<fileSec> (a list of the files and their roles and relationships)

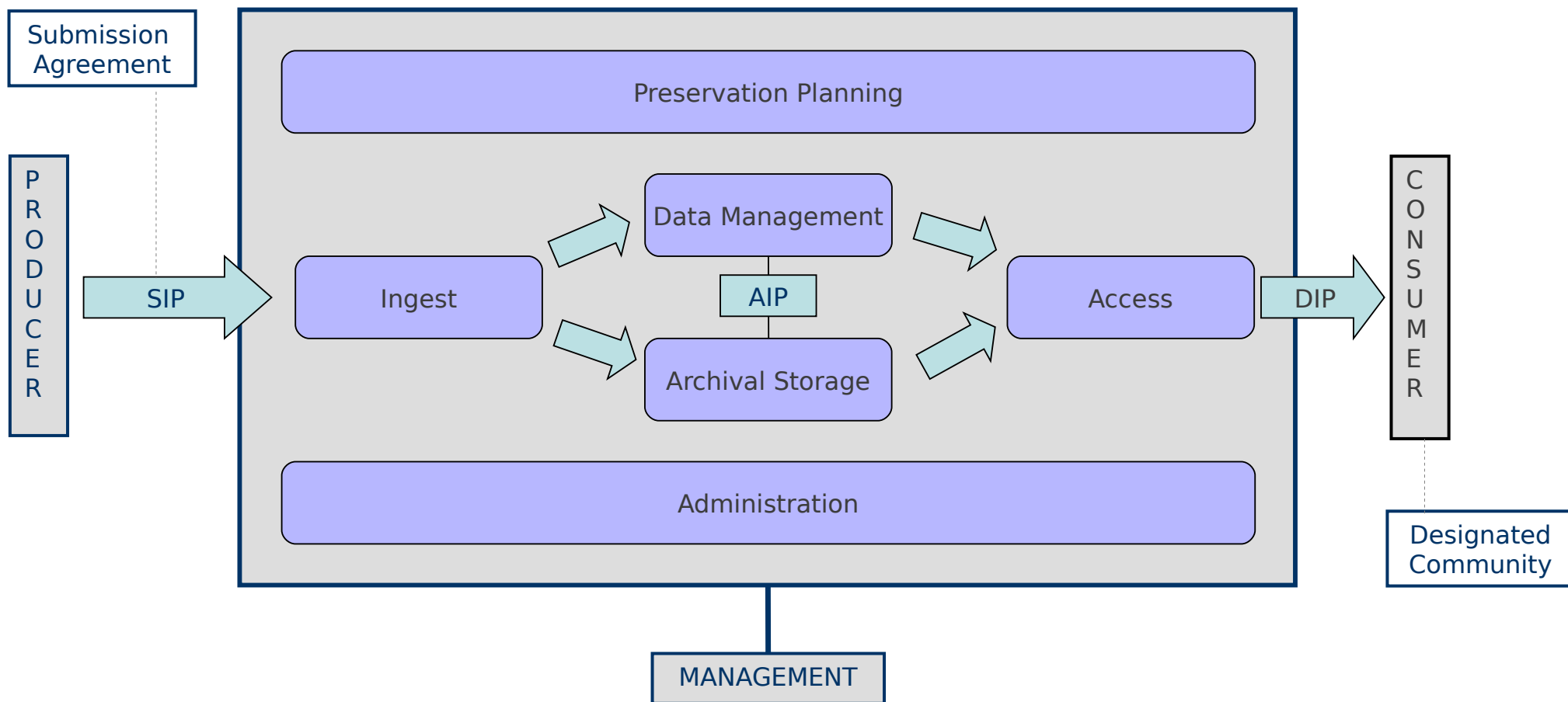
<structMap> (a representation of the physical structure of the AIP)

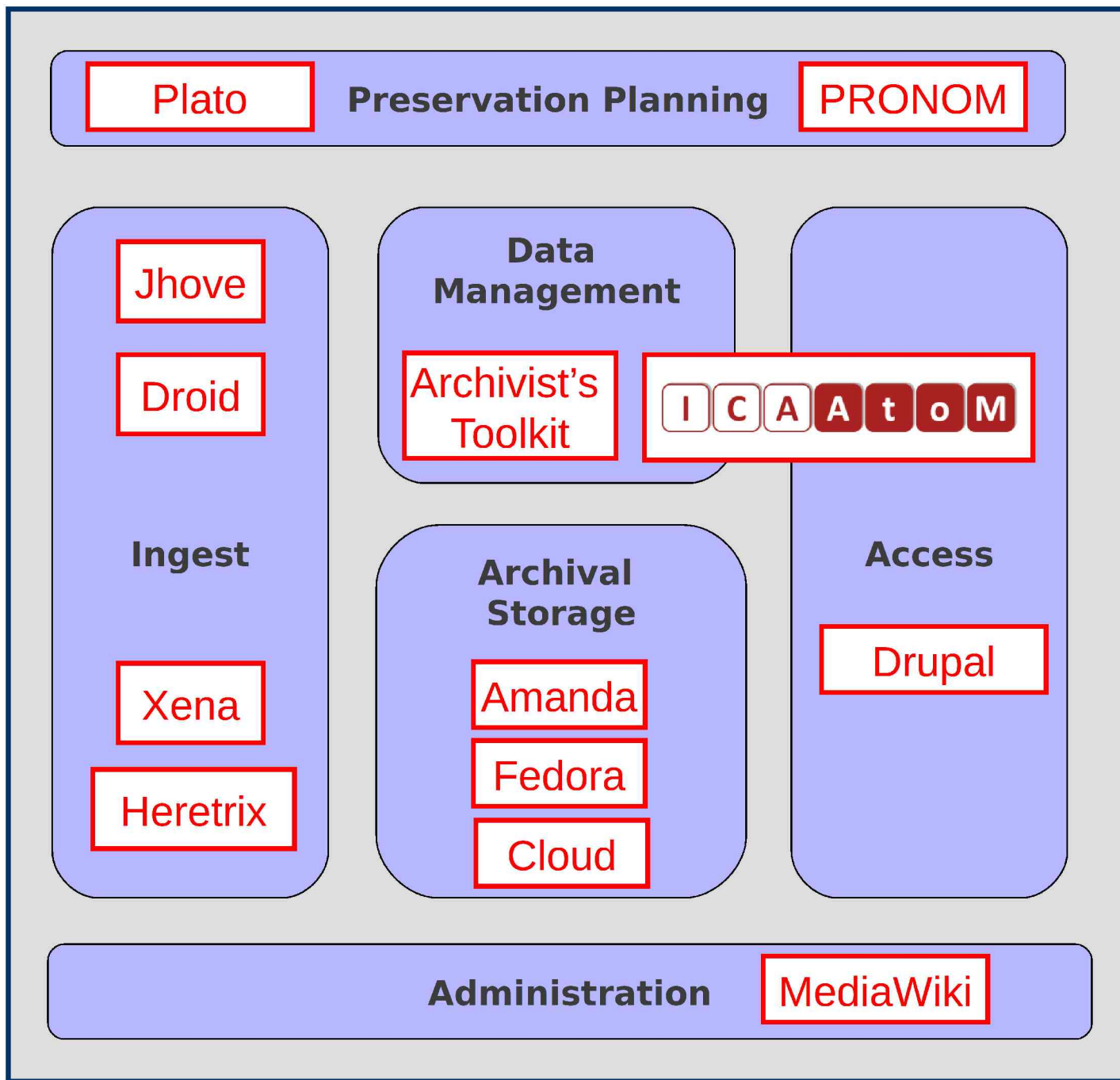
Where did
Archivemata come from?

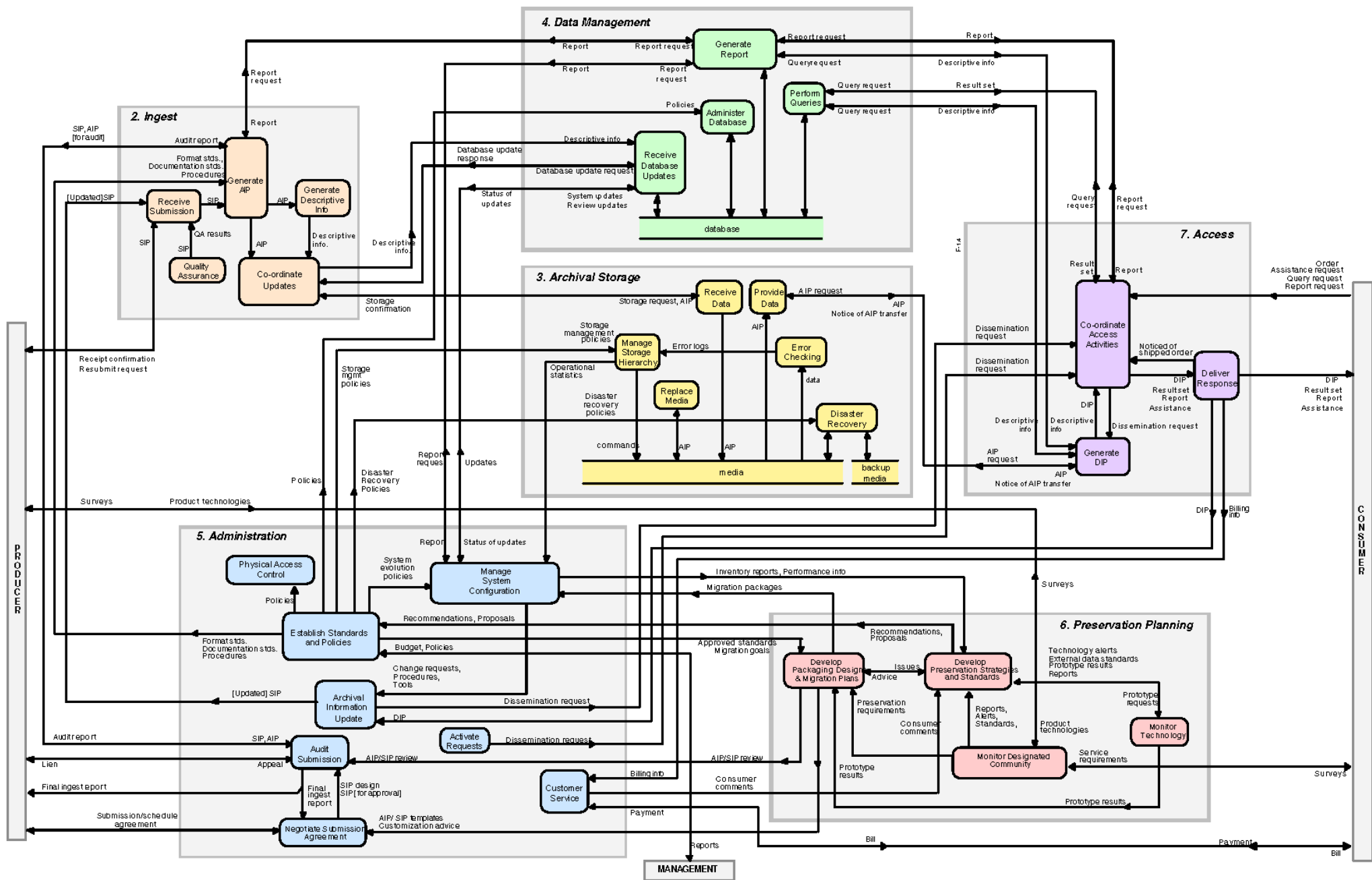


Open Archival Information System

ISO 14721:2003







search

- toolbox
- [What links here](#)
 - [Related changes](#)
 - [Special pages](#)
 - [Printable version](#)
 - [Permanent link](#)

UC-1.1 Receipt of SIP

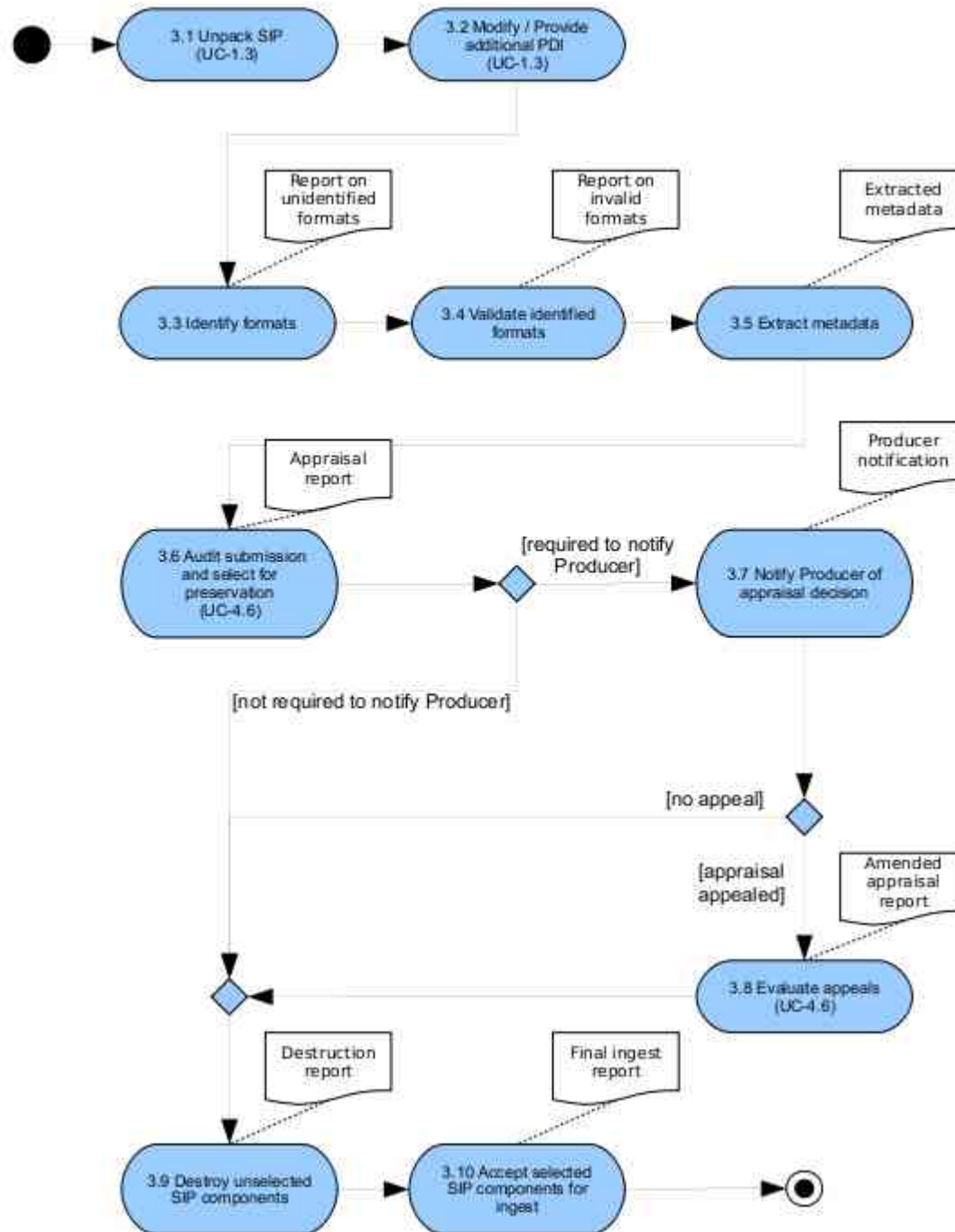
Summary

Number	UC-1.1
Name	Receipt of SIP
Context	UC-1 Ingest
Sub-Use Cases	None
Overview	Producer or Administration>Archival Information Update transfers Submission Information Package (SIP) to Receive Submission . Receive Submission provides the appropriate storage capability or devices to receive a SIP.
Actors	<ul style="list-style-type: none"> ■ Producer ■ Receive Submission ■ Administration>Archival Information Update

Description

Preconditions	Establish a Submission Agreement with the Producer . Create a virtual Submission Agreement if necessary. (see OAIS, page 2-9, section 2.3.2)
Trigger	A Data Submission Session is started by the Producer , OAIS or Management. (see OAIS, page 2-9, section 2.3.2)
Successful outcome	Receive Submission sends SIP to Quality Assurance
Steps	<ol style="list-style-type: none"> 1. Producer (or Administration>Archival Information Update) transfers SIP to Receive Submission (OAIS Page 4-5, Section 4.1.1.2) 2. Receive Submission provides the appropriate storage capability or devices to receive a SIP from Producer (OAIS Page 4-5, Section 4.1.1.2) 3. Receive Submission transmits SIP to Quality Assurance (OAIS Figure 4-2:Functions of Ingest Page 4-5)

OAIS: Accept SIP for Ingest



MEMORY OF THE WORLD

Towards an Open Source Repository and Preservation System

Recommendations on the Implementation of an Open Source
Digital Archival and Preservation System and on Related
Software Development



By

Kevin Bradley

National Library of Australia

UNESCO Memory of the World Sub-Committee on Technology

Junran Lei,

Australian Partnership for Sustainable Repositories,

Chris Blackall,

Australian Partnership for Sustainable Repositories



Free Beer!

“They’ll never take our freedom”



© 1995 Paramount Pictures & 20th Century Fox
See fair use rationale: http://en.wikipedia.org/wiki/File:Brave_mel.jpg





hosting
installation
integration
software development
tech support
training
system analysis
strategy

annual maintenance program

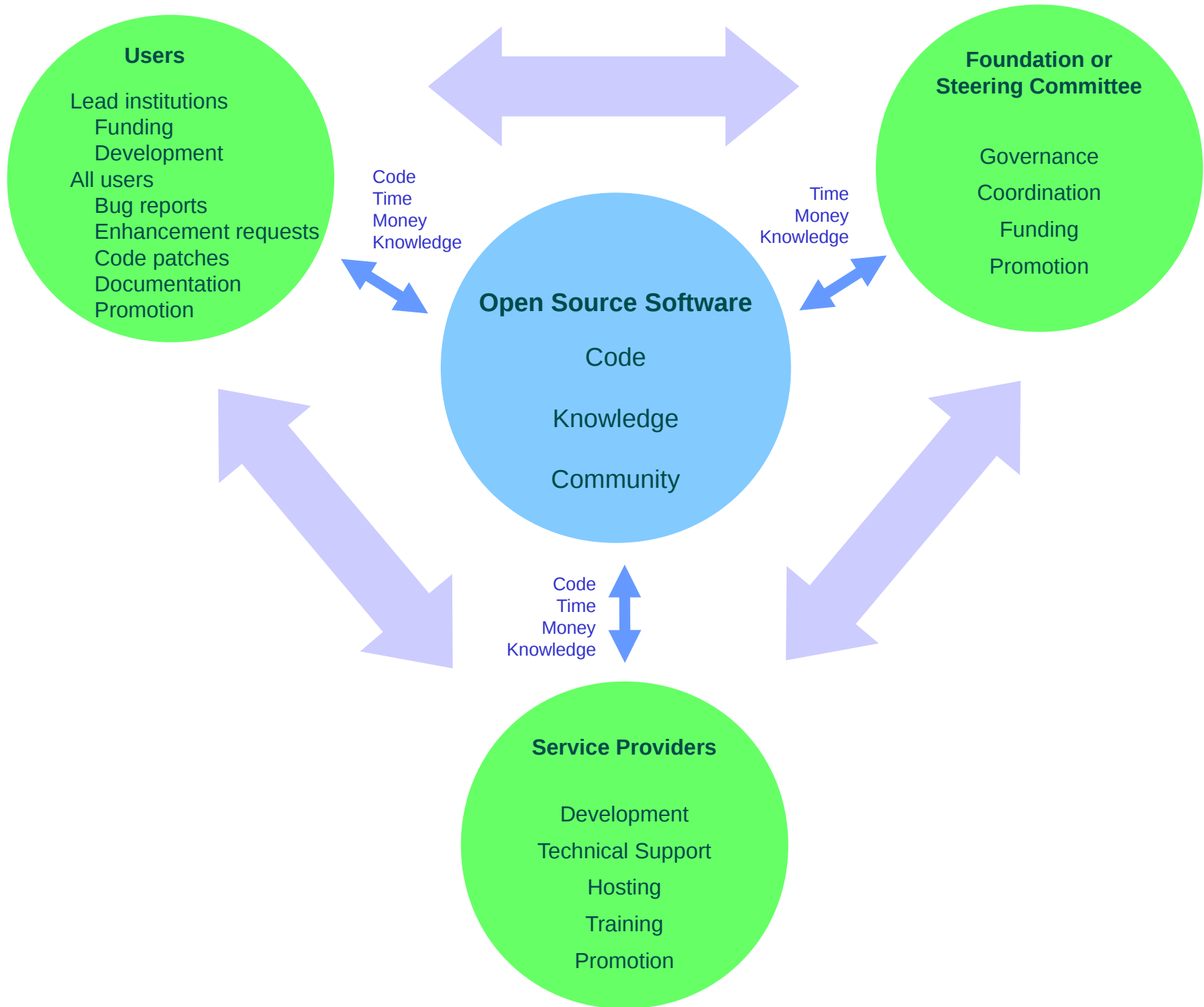
Community Support

We will try to answer fairly straight-forward questions from the open source community about installing and configuring our software. We are client contract service provider and we do what we can between client deadlines.

Due to our limited resources, community forum questions cannot become too specific, in-depth, or time-consuming for our support team. When we think a particular query is beyond these free support parameters we will inform the user that it may be necessary to address it as paid, commercial support.

Commercial Support

Our software is always free and open source, but with our hosting and support services, the Artefactual development team will assist a client with more in-depth questions to get the software installed and operating as required, whether on one of our servers or their own.



Agile development

- Individuals and interactions over processes and tools
- Working software over comprehensive documentation
- Customer collaboration over contract negotiation
- Responding to change over following a plan

Beck, Kent; et al. (2001). "Manifesto for Agile Software Development". Agile Alliance. Retrieved 14 June 2010.

Agile development model

- “Release early, release often”
 - Feb 2009: Release 0.1-alpha
 - ...
 - January 2012: Release 0.8-alpha
 - August 2012: Release 0.9-beta
 - February 2013: Release 1.0
- Each iteration leads to updated and improved:
 - Requirements
 - Software
 - Documentation
 - Development resources

archivematica.org/roadmap

The screenshot shows a web browser window displaying the Archivematica development roadmap. The browser's address bar shows the URL https://www.archivematica.org/wiki/Development_roadmap:_Archivematica_1.0. The page features the Archivematica logo at the top left and a navigation menu on the left side. The main content area is titled "Development roadmap: Archivematica 1.0" and includes a "Contents" section with a list of sub-sections. Below the contents, the "Release 1.0" section is visible, listing key dates: "Feature freeze: Friday, 30 Nov 2012" and "Code freeze: Monday, 14 Jan 2013".

Inbox - courtney.mumma@... x Development roadmap: Arc... x

← https://www.archivematica.org/wiki/Development_roadmap:_Archivematica_1.0 ☆ Google

Artefactual Projects Blogs Tools Archivematica Artefactual Records Open Planets Foun... Follow-up ndsa

Log in / create account

archivematica®

Navigation

- Main page
- Recent changes
- Random page

Toolbox

- What links here
- Related changes
- Special pages
- Printable version
- Permanent link

Page Discussion Read Edit View history Go Search

Development roadmap: Archivematica 1.0

Main Page > Development > Development roadmap: Archivematica 1.0

Contents [hide]

- 1 Release 1.0
 - 1.1 Features and enhancements
 - 1.1.1 Format Policy Registry (FPR)
 - 1.1.2 Normalization
 - 1.1.3 Dashboard
 - 1.1.4 Workflow
 - 1.1.5 Data management
 - 1.1.6 AIP management
 - 1.1.7 Interfaces
 - 1.1.8 Deployment
 - 1.2 Bug fixes
- 2 Release 1.1
 - 2.1 Features and enhancements
 - 2.1.1 Dashboard
 - 2.1.2 Deployment

Release 1.0

[edit]

- Feature freeze: Friday, 30 Nov 2012
- Code freeze: Monday, 14 Jan 2013

How we decide what's next

- In-house research
 - evolving best practices
 - new tools
- User community input
 - Email list
 - Issues list
- Client requirements

User email list

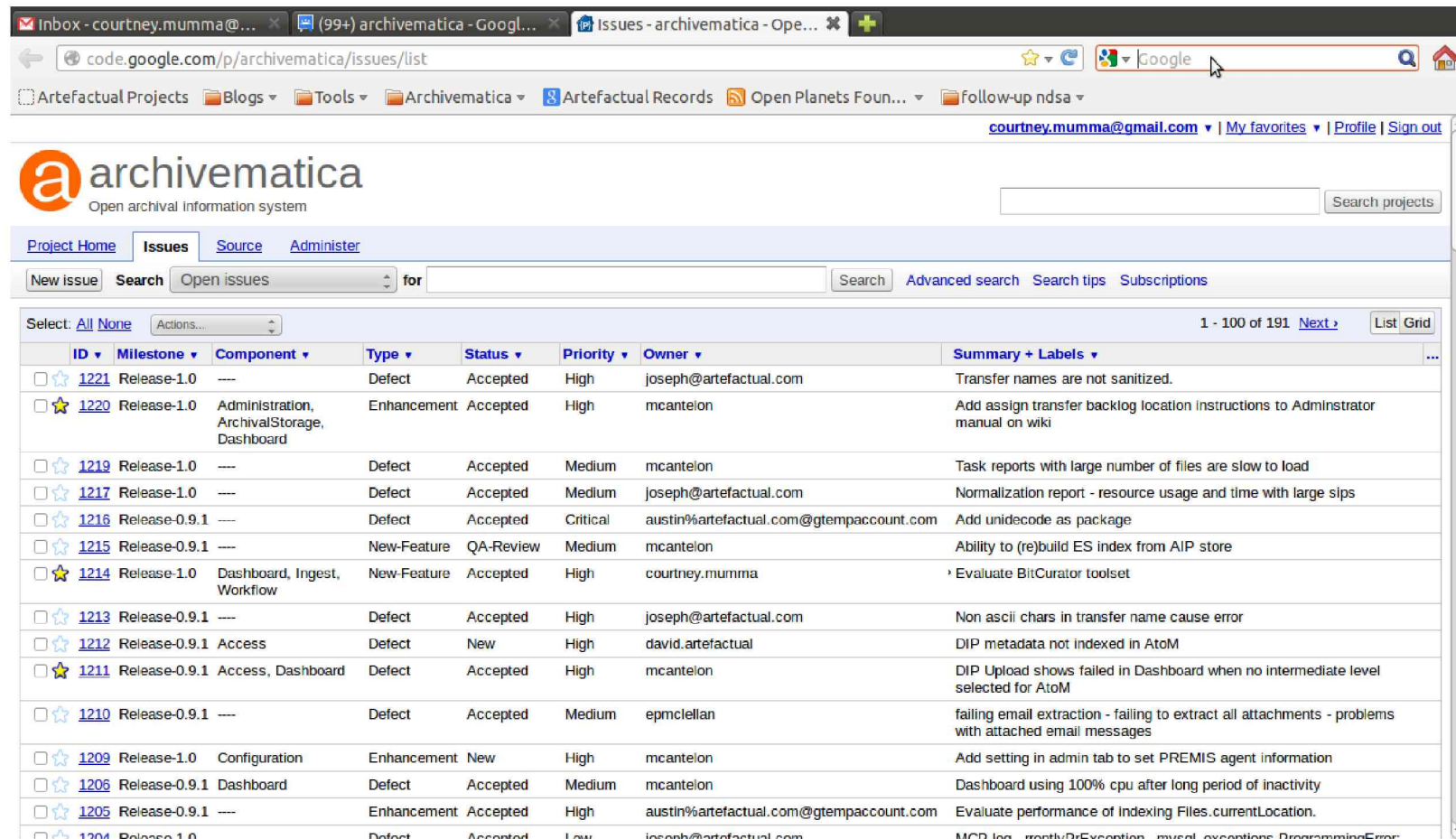
The screenshot shows a web browser window displaying a Google Groups forum page. The browser's address bar shows the URL: <https://groups.google.com/forum/?fromgroups#!forum/archivemata>. The page title is "archivemata" and it shows "Showing 31 of 239 topics (99+ unread)".

The forum page includes a navigation bar with "Groups", "NEW TOPIC", "Mark all as read", "Filters", and "Members" buttons. The main content area lists several topics:

- Archivemata 0.9 beta release (1)**
By me - 1 post - 20 views - updated Sep 5
- Technical documentation. (6)**
By Stefan Gutten - 6 posts - 5 views - updated Sep 19 (1 day ago)
- Backing up (and moving?) AIPs and DIPs (3)**
By Paul James - ARCW Digital Preservation - 3 posts - 4 views - updated Sep 14 (6 days ago)
- Question about desktop computer type (3)**
By Lisa Snider - 3 posts - 6 views - updated Sep 14 (6 days ago)
- Re: Archivemata Tutorial Issue (1)**
By Courtney Mumma - 1 post - 5 views - updated Sep 12 (8 days ago)
- Session presenters for SAA 2013 annual meeting (1)**
By Abby R. Adams - 1 post - 7 views - updated Sep 12 (8 days ago)
- Crashing problems with v. 0.9-beta on Oracle VB**
By Kari Smith - 3 posts - 6 views - updated Sep 10 (10 days ago)
- Periodic verification of AIP checksums (4)**
By Mark Jordan - 4 posts - 3 views - updated Sep 5

The left sidebar contains navigation links: "My groups", "Home", "Starred", "Announcements", "Recently viewed", "fits-users", "archivemata", "BitCurator Users", "Digital Curation", "archivemata issues", and "Recent searches".

Issues list



code.google.com/p/archivematica/issues/list

Search projects

Project Home Issues Source Administer

New issue Search Open issues for Search Advanced search Search tips Subscriptions

Select: All None Actions... 1 - 100 of 191 Next List Grid

ID	Milestone	Component	Type	Status	Priority	Owner	Summary + Labels
<input type="checkbox"/> 1221	Release-1.0	---	Defect	Accepted	High	joseph@artefactual.com	Transfer names are not sanitized.
<input type="checkbox"/> 1220	Release-1.0	Administration, ArchivalStorage, Dashboard	Enhancement	Accepted	High	mcantelon	Add assign transfer backlog location instructions to Administrator manual on wiki
<input type="checkbox"/> 1219	Release-1.0	---	Defect	Accepted	Medium	mcantelon	Task reports with large number of files are slow to load
<input type="checkbox"/> 1217	Release-1.0	---	Defect	Accepted	Medium	joseph@artefactual.com	Normalization report - resource usage and time with large sips
<input type="checkbox"/> 1216	Release-0.9.1	---	Defect	Accepted	Critical	austin%artefactual.com@tempaccount.com	Add unicode as package
<input type="checkbox"/> 1215	Release-0.9.1	---	New-Feature	QA-Review	Medium	mcantelon	Ability to (re)build ES index from AIP store
<input type="checkbox"/> 1214	Release-1.0	Dashboard, Ingest, Workflow	New-Feature	Accepted	High	courtney.mumma	Evaluate BitCurator toolset
<input type="checkbox"/> 1213	Release-0.9.1	---	Defect	Accepted	High	joseph@artefactual.com	Non ascii chars in transfer name cause error
<input type="checkbox"/> 1212	Release-0.9.1	Access	Defect	New	High	david.artefactual	DIP metadata not indexed in AtOM
<input type="checkbox"/> 1211	Release-0.9.1	Access, Dashboard	Defect	Accepted	High	mcantelon	DIP Upload shows failed in Dashboard when no intermediate level selected for AtOM
<input type="checkbox"/> 1210	Release-0.9.1	---	Defect	Accepted	Medium	epmclellan	failing email extraction - failing to extract all attachments - problems with attached email messages
<input type="checkbox"/> 1209	Release-1.0	Configuration	Enhancement	New	High	mcantelon	Add setting in admin tab to set PREMIS agent information
<input type="checkbox"/> 1206	Release-0.9.1	Dashboard	Defect	Accepted	Medium	mcantelon	Dashboard using 100% cpu after long period of inactivity
<input type="checkbox"/> 1205	Release-0.9.1	---	Enhancement	Accepted	High	austin%artefactual.com@tempaccount.com	Evaluate performance of indexing Files.currentLocation.
<input type="checkbox"/> 1204	Release-1.0	---	Defect	Accepted	Low	isaiah@artefactual.com	MCP log: java.lang.RuntimeException: must, exceptions: ProgrammingError

WE NEED YOU!



How to contribute

Discussion: ask questions and provide opinions/suggestions on project list

Support: answer questions on the discussion list

Dissemination: blog, tweet, present, train

Documentation: manual updates, wiki gardening

Testing: report bugs, request features

Development: fix bugs, contribute patches, write plugins

Maintenance: provide technical services

Client Partners

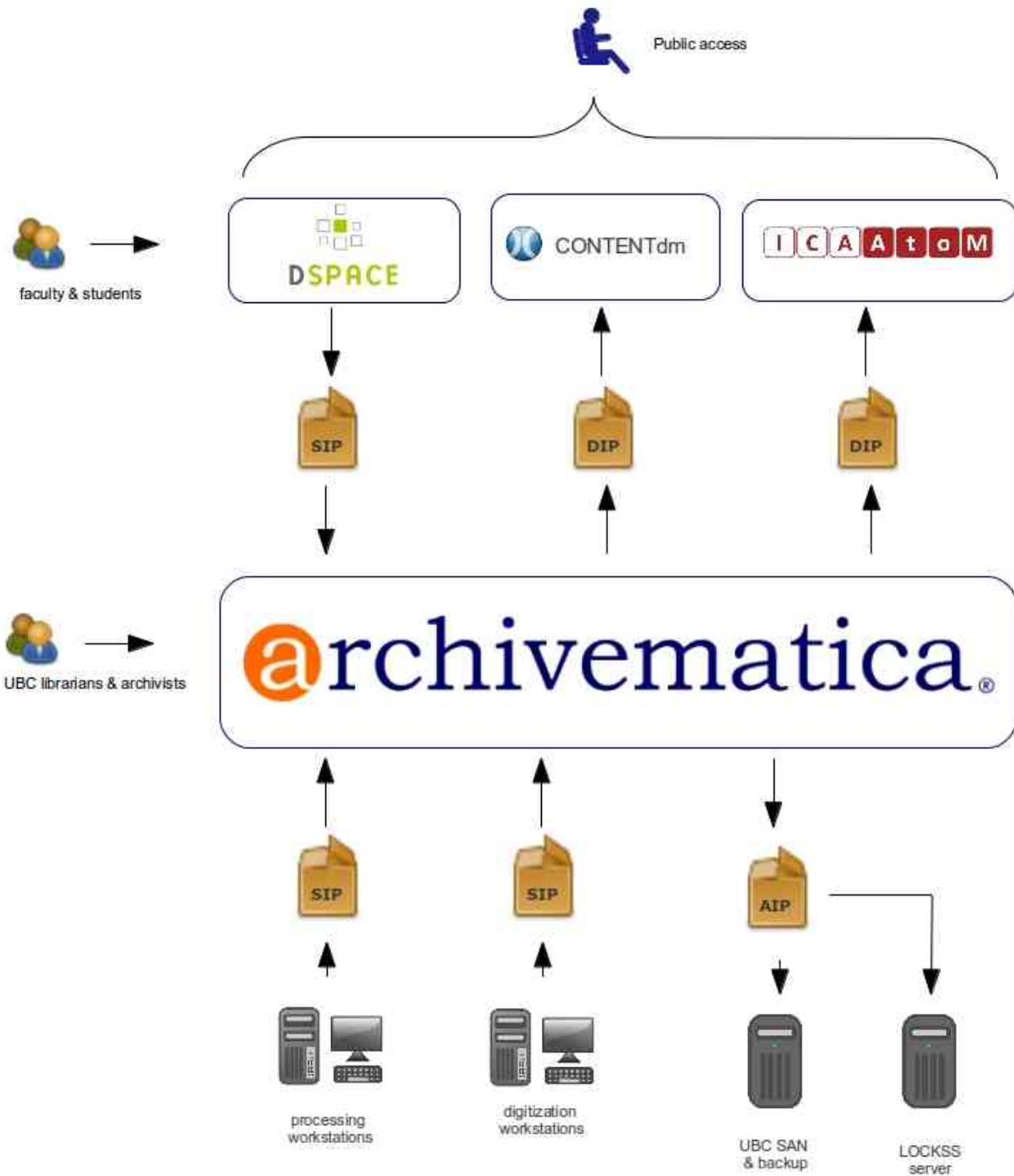


United Nations
Educational, Scientific and
Cultural Organization



SIMON FRASER UNIVERSITY
ARCHIVES AND RECORDS MANAGEMENT

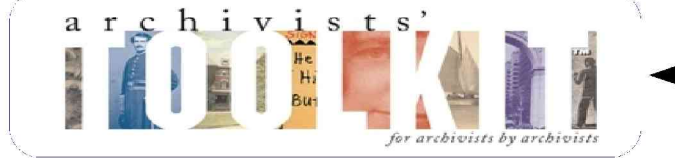




born digital

digitized

archivematica®



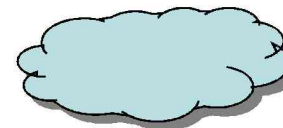
analogue

closed collections

open collections



staff



public

<http://opensourcearchiving.org/content/archivemata-city-vancouver-archives>

Archivemata at the City of Vancouver Archives

Submitted by [Sue Bigelow](#) [Sue Bigelow's blog](#) [Add new comment](#)

in [archivemata](#) [BitCurator](#) [CollectiveAccess](#) [development](#) [digital preservation](#) [DSpace](#) [software tools](#) -

The City of Vancouver Archives has been contributing to the development of the Archivemata digital preservation system for the past several years and we have just started using the 0.8 alpha release for production. This is an overview of why we got involved and where we are now.

Who we are

We're a municipal archives, part of the City of Vancouver government. We follow the Canadian [total archives](#) concept of acquiring the records of both government and the private sector, and a wide variety of media. Our holdings include textual records, still and moving images, audio, maps and plans, and documentary art. We also have born-digital materials, the most significant being the records of the Vancouver Organizing Committee for the 2010 Olympic and Paralympic Winter Games (VANOC). We have the first [Games records](#) that are mostly digital, with more than 25TB acquired. We had been looking for a digital preservation system for many years, and in 2008, thanks to funding from the Olympic Legacy Reserve, we were able to take action.

What we were looking for

Create SIP(s)

Enter Accession # or Transfer Name, or use Browse to locate transfers or entire accessions

Original Order - Select transfers for arrangement Arrangement - Arrange selected records

search Browse

Transfer	Type	Size	Creation Date	Transfer Type	Name	Size
▶ 2009_Flyers	folder	15.6MB	2009-01-12	Generic	AM1260 - Festival of Cats fonds	
▼ Cats_Parade	folder	8.8MB	2002-01-09	Digitized	Series 1 - Publicity files	28.4 GB
Floats	folder	4.3MB	2003-02-23		Flyers	12 GB
GiantSiamese.gif	GIF	1.2MB	2003-03-14		Press releases	16.4 GB
GingerRibbons.jpg	JPEG	1.9MB	2006-05-17		Series 2 - Events files	15 GB
Minneke_Poes.jpg	JPEG	1.2MB	2008-09-21		Opening Party	9.1 GB
Permits					Cat Parade	5.9 GB
Events					Photographs	4 GB
▶ 2006_Meetings	folder				Volunteers	2.9 GB
Minutes					Series 3 - Sponsor files	14 GB
Agenda					Series 4 - Correspondence files	1 GB
▶ Volunteers					Series 5 - Cat Festival Committee files	19 GB
ParadeCleanup					Meetings	9 GB
Cat Toss					Agenda	10 GB
Plushy donations						
▶ Photographs						
2003						

Greyed out transfers have been added to a SIP in the arrangement pane

In a new tab, search indexed selections for subject terms, confidentiality statements, CC and SIN #s, etc., and create visualizations of contents

Examine contents Examine contents Add child Add sibling Create SIP

Open file in viewer - greyed out if not supported in browser

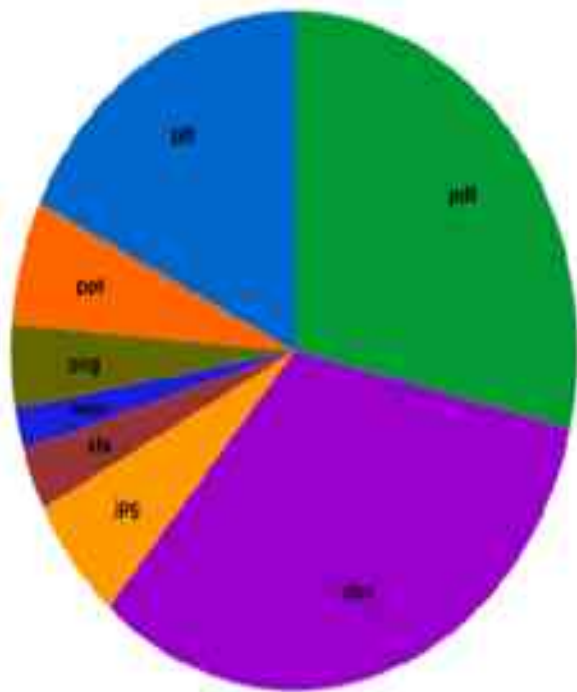
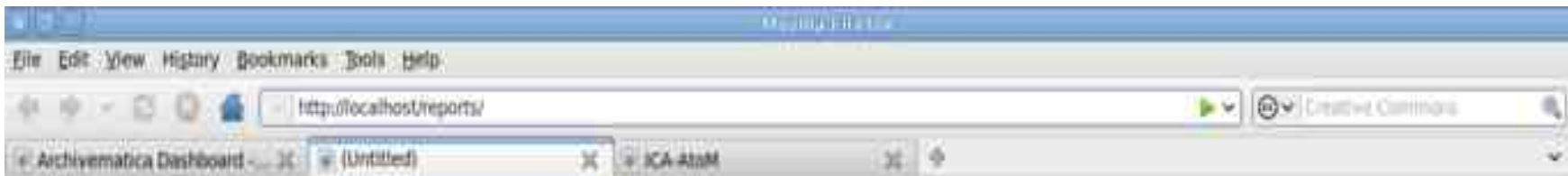
Create SIP from selection(s)

Add metadata to SIP

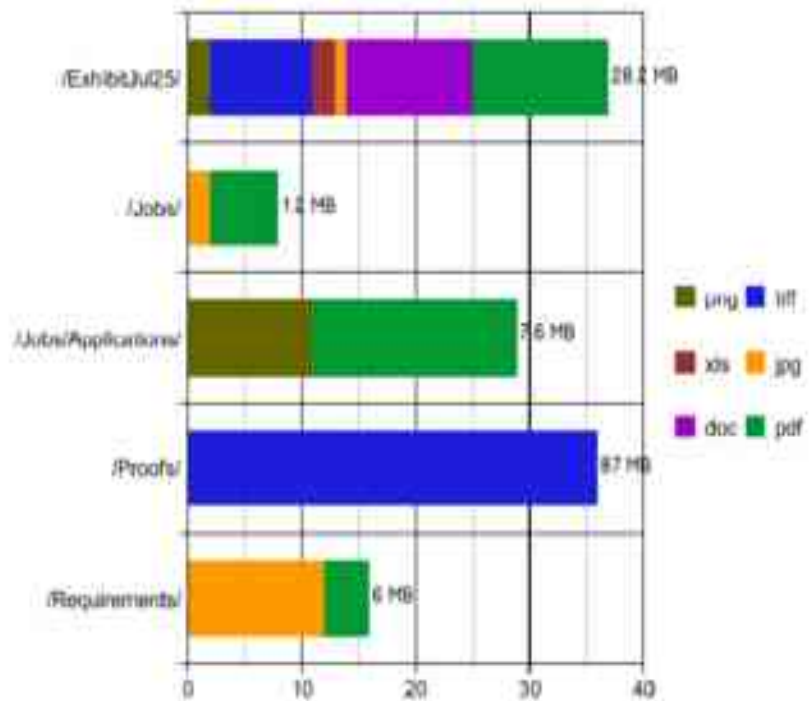
SIP(s)

AM1260-S2-Photographs	<input checked="" type="checkbox"/>	
AM1260-S3-2-Busters-pets	<input checked="" type="checkbox"/>	
AM1260-S1-Flyers	<input checked="" type="checkbox"/>	
AM1260-S1-Press-releases	<input checked="" type="checkbox"/>	

Accept SIP and continue processing in dashboard



File extension breakdown (by number of files) in transferObjects/

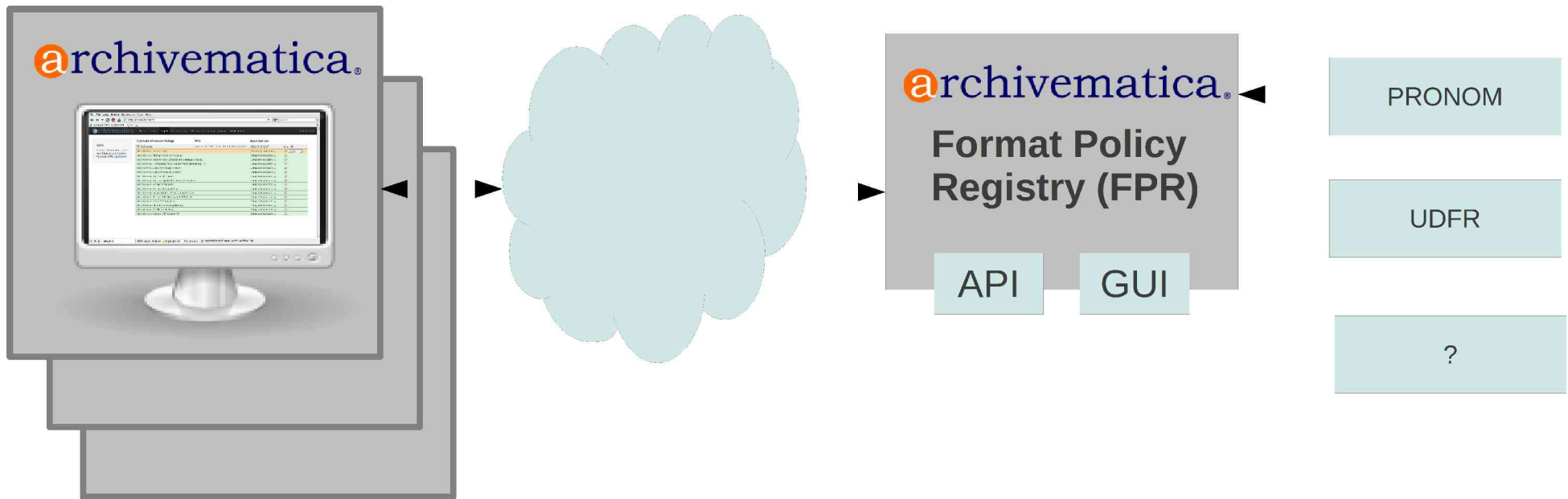


File extensions by folder in transferObjects/ Total size of objects per folder



Email preservation

- Maildir (preservation)
 - OfflineIMAP
- Mbox (access)
 - Muse
- https://www.archivemata.org/wiki/Email_preservation



Home Transfer Ingest ⁴ Archival storage Preservation planning Access Administration

Media type	Extension	Normalization description	Command	Purpose
Audio	ac3	Transcoding to mp3 with ffmpeg	Show	access
	ac3	Transcoding to wav with ffmpeg	Show	preservation
	aif	Transcoding to mp3 with ffmpeg	Show	access
	aif	Transcoding to wav with ffmpeg	Hide	preservation
	mp3	Transcoding to mp3 with ffmpeg	Show	access

Show advanced details

```
ffmpeg -i "%fileFullName%" -ac `ffprobe "%fileFullName%" -show_streams 2>&1 | grep "c`
```

archivematica.org