Digital Curation Workflows Incorporating Forensics Tools and Methods

Cal Lee, University of North Carolina

Tackling Real-World Collection Challenges with Digital Forensics Tools and Methods

Chapel Hill, NC

June 3-5, 2013

BitCurator

UNC
SCHOOL OF INFORMATION
AND LIBRARY SCIENCE

Work Flow – The Thing to be Represented

“the sequence of processes through which a piece of work passes from initiation to completion”
(Oxford English Dictionary, Second Edition, 1989)

Work Flows as Models – Representations of the Thing

- Explicit, symbolic representation of the workflow
- Usually inspired by new system design or attempts to reengineer a process
- There are many different ways to model a workflow
- But the basic components tend to be similar

Parts of a Workflow

- Entities/Stages – where something happens (e.g. data are transformed, someone makes a decision, data are captured)
- Input(s) – control and/or information that flows into an entity/stage
- Output(s) – control and/or information that flow out of an entity/stage

Digital Resources - Levels of Representation

Level	Label	Explanation
8	Aggregation of objects	Set of objects that form an aggregation that is meaningful encountered as an entity
7	Object or package	Object composed of multiple files, each of which could also be encountered as individual files
6	In-application rendering	As rendered and encountered within a specific application
5	File through filesystem	Files encountered as discrete set of items with associate paths and file names
4	File as “raw” bitstream	Bitstream encountered as a continuous series of binary values
3	Sub-file data structure	Discrete “chunk” of data that is part of a larger file
2	Bitstream through I/O equipment	Series of 1s and 0s as accessed from the storage media using input/output hardware and software (e.g. controllers, drivers, ports, connectors)
1	Raw signal stream through I/O equipment	Stream of magnetic flux transitions or other analog electronic output read from the drive without yet interpreting the signal stream as a set of discrete values (i.e. not treated as a digital bitstream that can be directly read by the host computer)
0	Bitstream on physical medium	Physical properties of the storage medium that are interpreted as bitstreams at Level 1

Digital Resources - Levels of Representation

Level	Label	Explanation
8	Aggregation of objects	Set of objects that form an aggregation that is meaningful encountered as an entity
7	Object or package	Obje also
6	In-application rendering	As re
5	File through filesystem	Files paths
4	File as “raw” bitstream	Bitstr value
3	Sub-file data structure	Discr
2	Bitstream through I/O equipment	Serie using contr
1	Raw signal stream through I/O equipment	Stream of magnetic flux transitions or other analog electronic output read from the drive without yet interpreting the signal stream as a set of discrete values (i.e. not treated as a digital bitstream that can be directly read by the host computer)
0	Bitstream on physical medium	Physical properties of the storage medium that are interpreted as bitstreams at Level 1

Levels where digital forensics methods and tools can provide a lot of assistance

Storage Media Acquisition and Handling Profile for Digital Repositories*

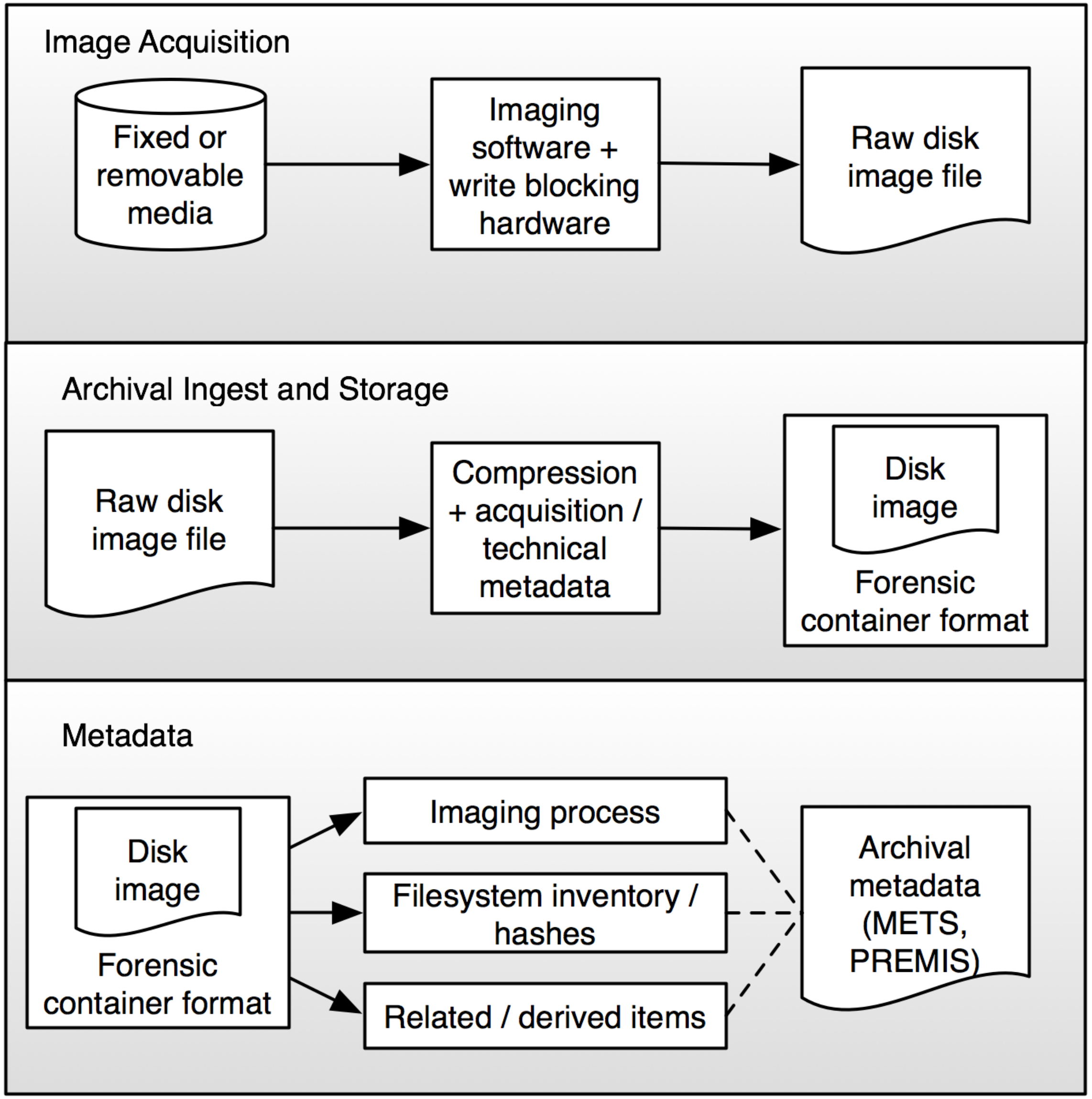
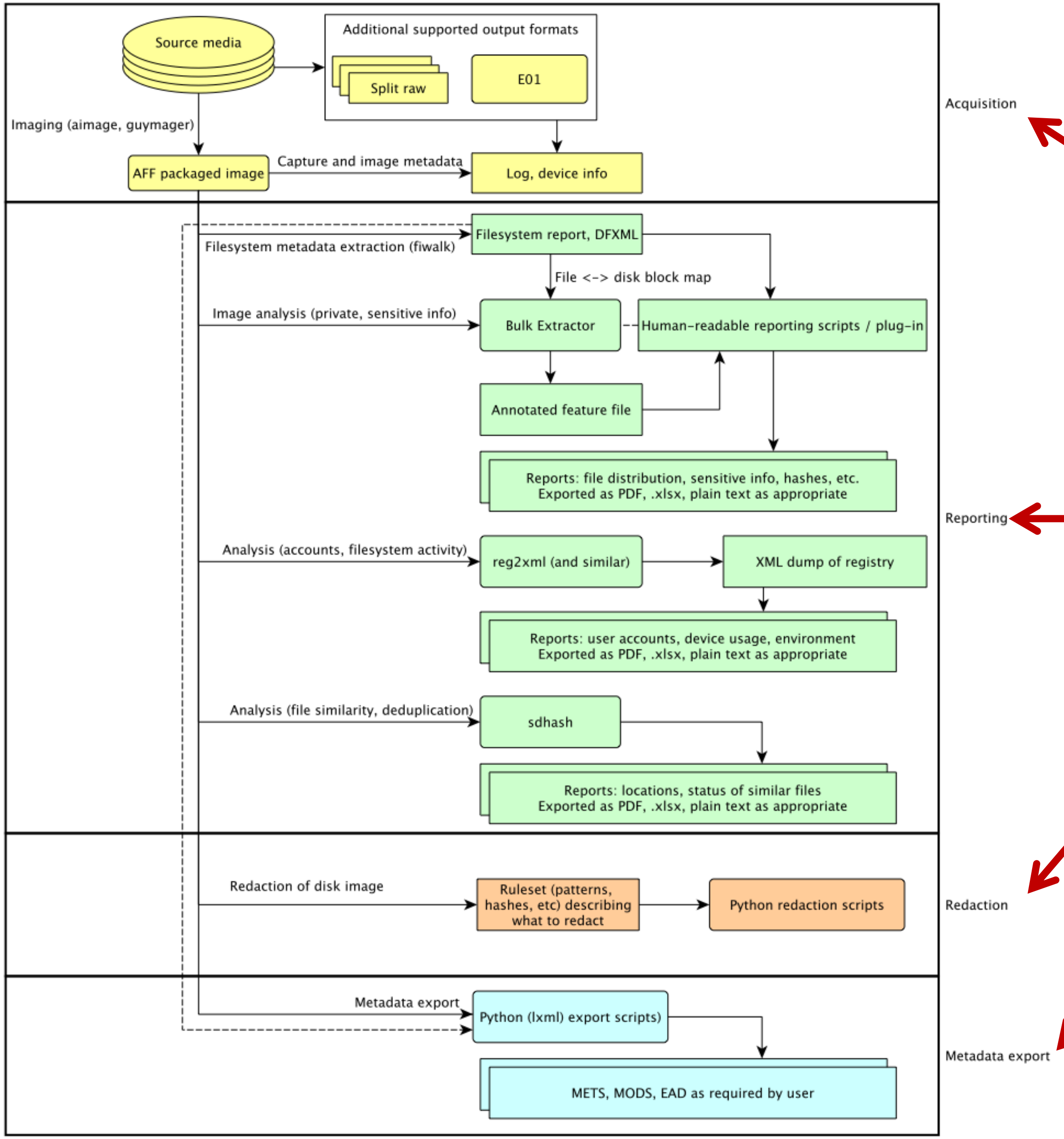


Image Acquisition
Fixed or removable media
Imaging software + write blocking hardware
Raw disk image file
Archival Ingest and Storage
Raw disk image file
Compression + acquisition / technical metadata
Disk image
Forensic container format
Metadata
Disk image
Forensic container format
Imaging process
Filesystem inventory / hashes
Related / derived items
Archival metadata (METS, PREMIS)

*Woods, Kam, Christopher A. Lee, and Simson Garfinkel. “Extending Digital Repository Architectures to Support Disk Image Preservation and Access.” In *JCDL '11: Proceeding of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, 57-66. New York, NY: ACM Press, 2011.

BitCurator-Supported Workflow

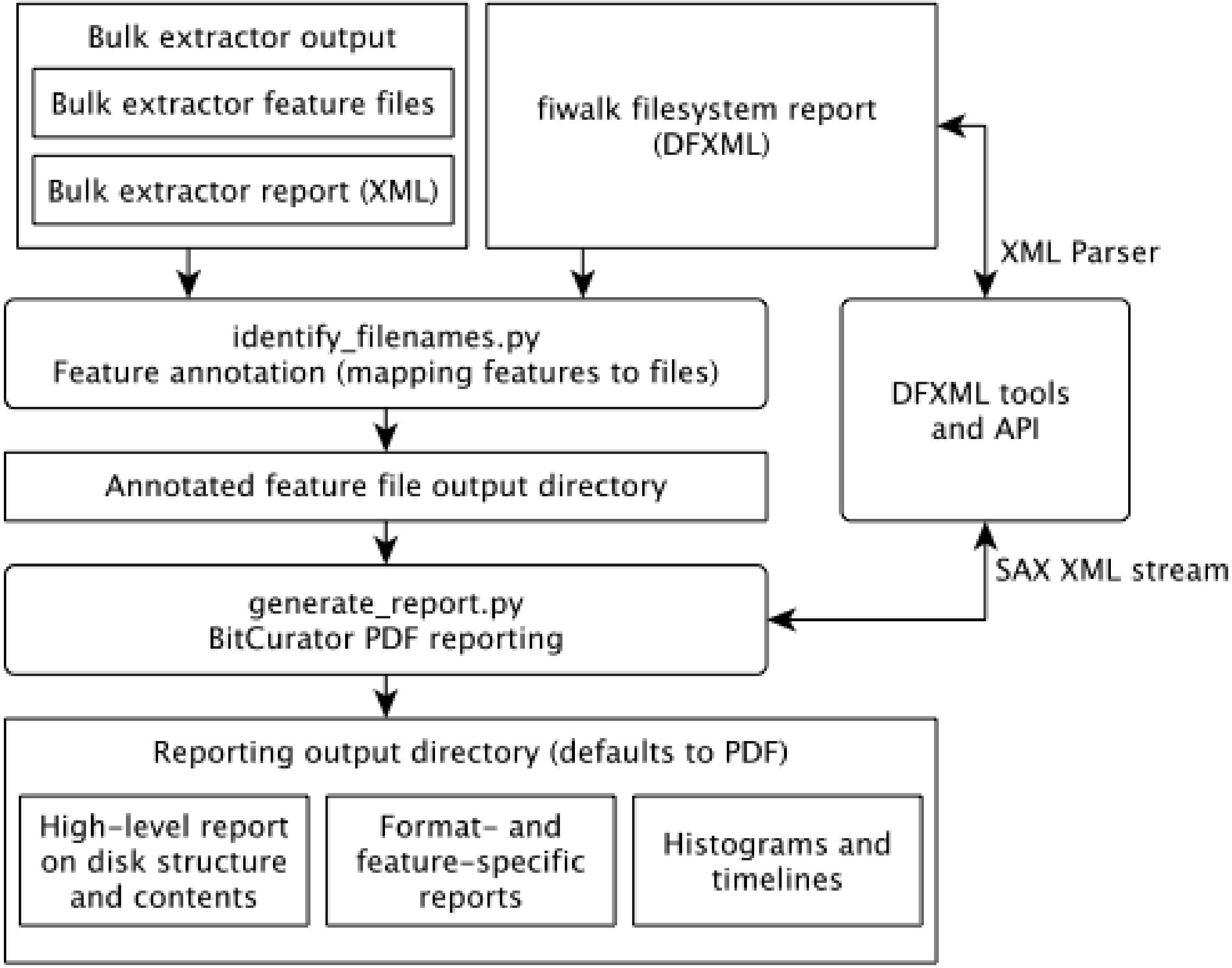

Source media
Additional supported output formats
Split raw
E01
Imaging (aimage, guymager)
AFF packaged image
Capture and image metadata
Log, device info
Acquisition
Filesystem report, DFXML
Filesystem metadata extraction (fiwalk)
File <-> disk block map
Image analysis (private, sensitive info)
Bulk Extractor
Human-readable reporting scripts / plug-in
Annotated feature file
Reports: file distribution, sensitive info, hashes, etc.
Exported as PDF, .xlsx, plain text as appropriate
Reporting
Analysis (accounts, filesystem activity)
reg2xml (and similar)
XML dump of registry
Reports: user accounts, device usage, environment
Exported as PDF, .xlsx, plain text as appropriate
Analysis (file similarity, deduplication)
sdhash
Reports: locations, status of similar files
Exported as PDF, .xlsx, plain text as appropriate
Redaction of disk image
Ruleset (patterns, hashes, etc) describing what to redact
Python redaction scripts
Redaction
Metadata export
Python (lxml) export scripts)
METS, MODS, EAD as required by user
Metadata export

- **Acquisition**
- **Reporting**
- **Redaction**
- **Metadata Export**

See: http://bitcurator.net

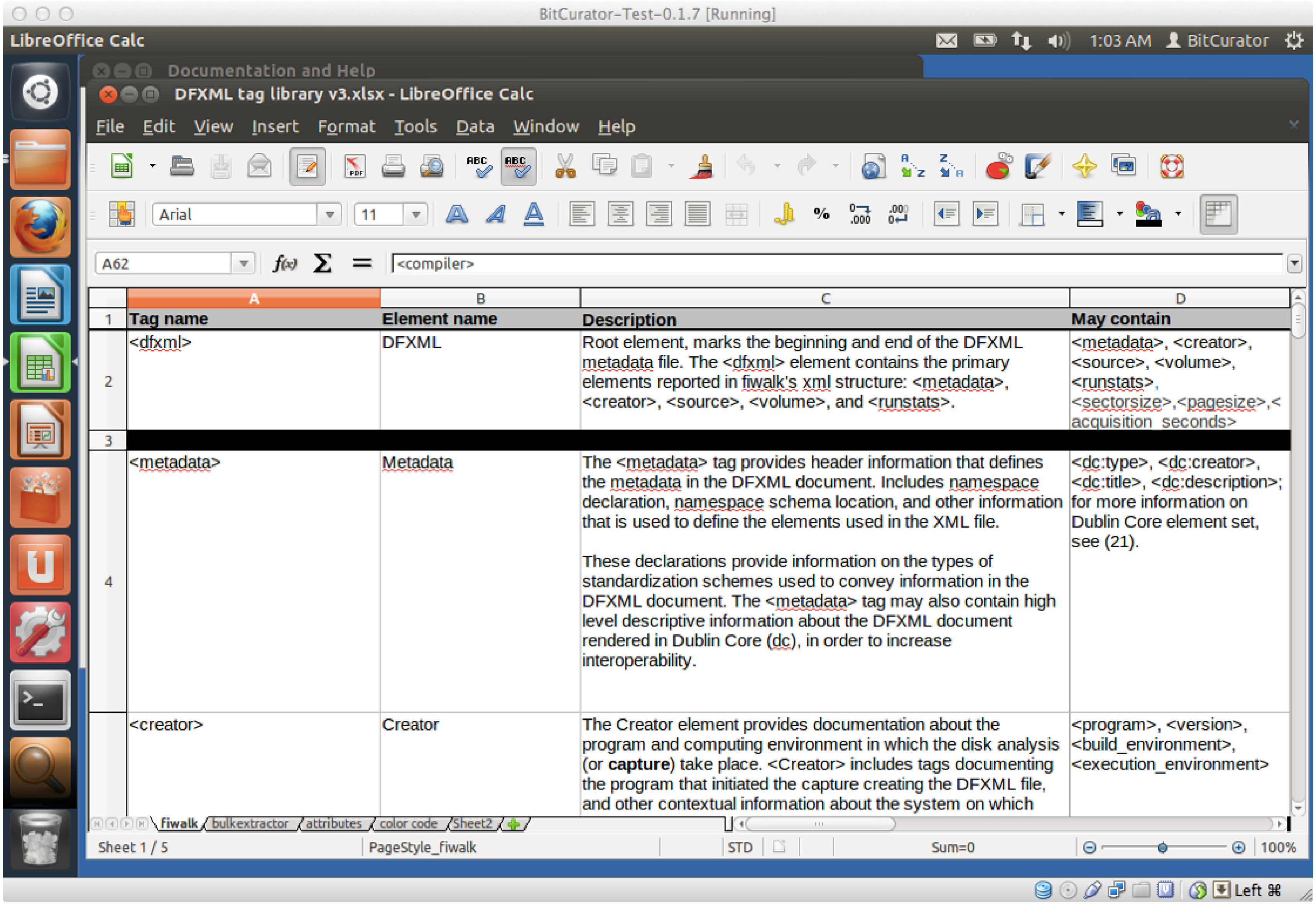
Metadata Generated by Forensics Software

Metadata Generation and Reporting


Bulk extractor output
Bulk extractor feature files
Bulk extractor report (XML)
fiwalk filesystem report (DFXML)
XML Parser
identify_filenames.py
Feature annotation (mapping features to files)
DFXML tools and API
Annotated feature file output directory
SAX XML stream
generate_report.py
BitCurator PDF reporting
Reporting output directory (defaults to PDF)
High-level report on disk structure and contents
Format- and feature-specific reports
Histograms and timelines

See: Woods, Kam, Christopher Lee, and Sunitha Misra. “Automated Analysis and Visualization of Disk Images and File Systems for Preservation.” In *Proceedings of Archiving 2013* (Springfield, VA: Society for Imaging Science and Technology, 2013), 239-244.

Documentation of Digital Forensics XML (DFXML) Elements

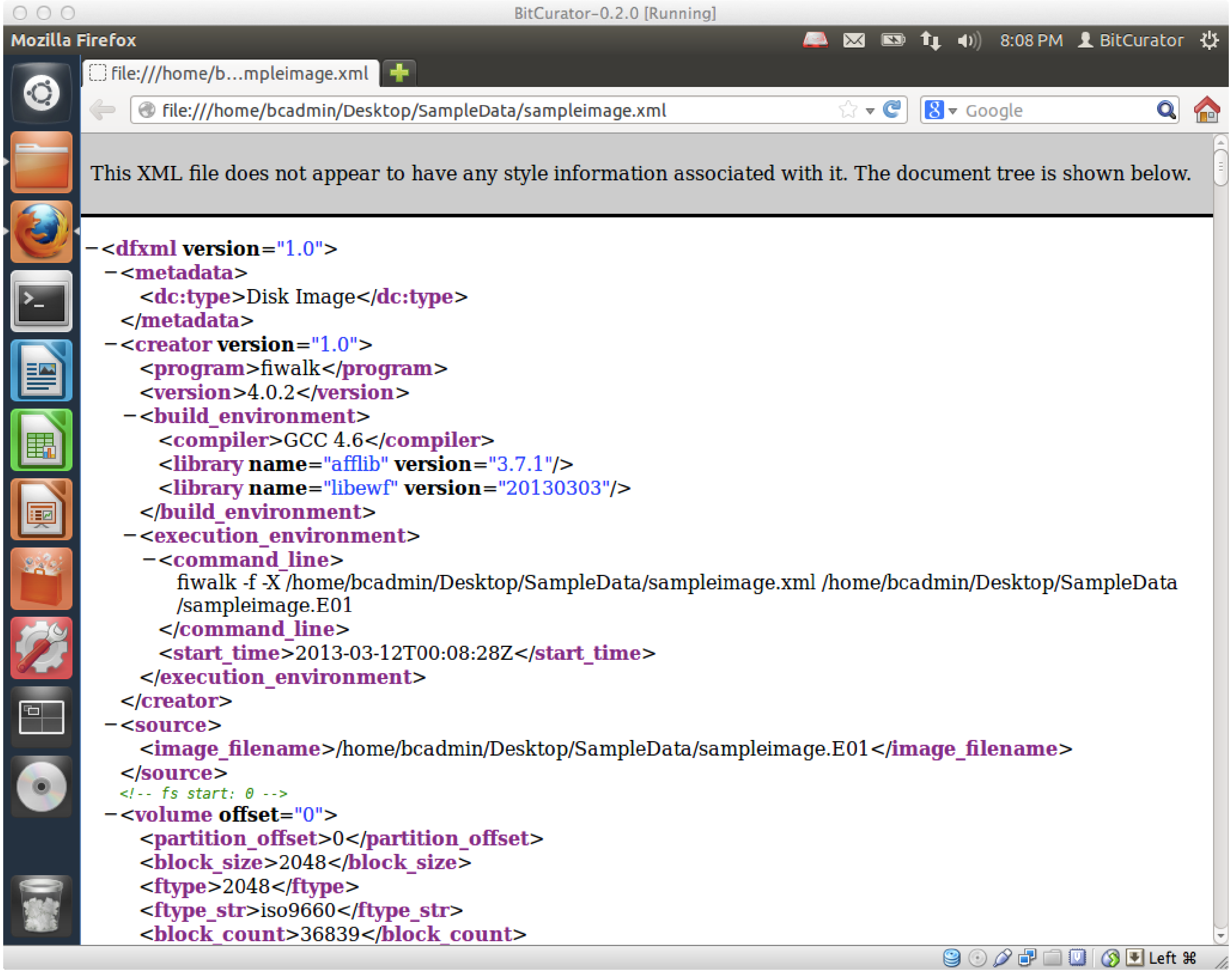


A62 <compiler>

Tag name	Element name	Description	May contain
<dfxml>	DFXML	Root element, marks the beginning and end of the DFXML metadata file. The <dfxml> element contains the primary elements reported in fiwalk's xml structure: <metadata>, <creator>, <source>, <volume>, and <runstats>.	<metadata>, <creator>, <source>, <volume>, <runstats>, <sectorsize>,<pagesize>,<acquisition_seconds>
<metadata>	Metadata	The <metadata> tag provides header information that defines the metadata in the DFXML document. Includes namespace declaration, namespace schema location, and other information that is used to define the elements used in the XML file. These declarations provide information on the types of standardization schemes used to convey information in the DFXML document. The <metadata> tag may also contain high level descriptive information about the DFXML document rendered in Dublin Core (dc), in order to increase interoperability.	<dc:type>, <dc:creator>, <dc:title>, <dc:description>; for more information on Dublin Core element set, see (21).
<creator>	Creator	The Creator element provides documentation about the program and computing environment in which the disk analysis (or **capture**) take place. <Creator> includes tags documenting the program that initiated the capture creating the DFXML file, and other contextual information about the system on which	<program>, <version>, <build_environment>, <execution_environment>

http://www.bitcurator.net/2013/02/06/dfxml-tag-library/

You want provenance? We've got provenance.



BitCurator-0.2.0 [Running]
Mozilla Firefox
8:08 PM
BitCurator
file:///home/b...mpleimage.xml
file:///home/bcadmin/Desktop/SampleData/sampleimage.xml
Google
Left ⌘

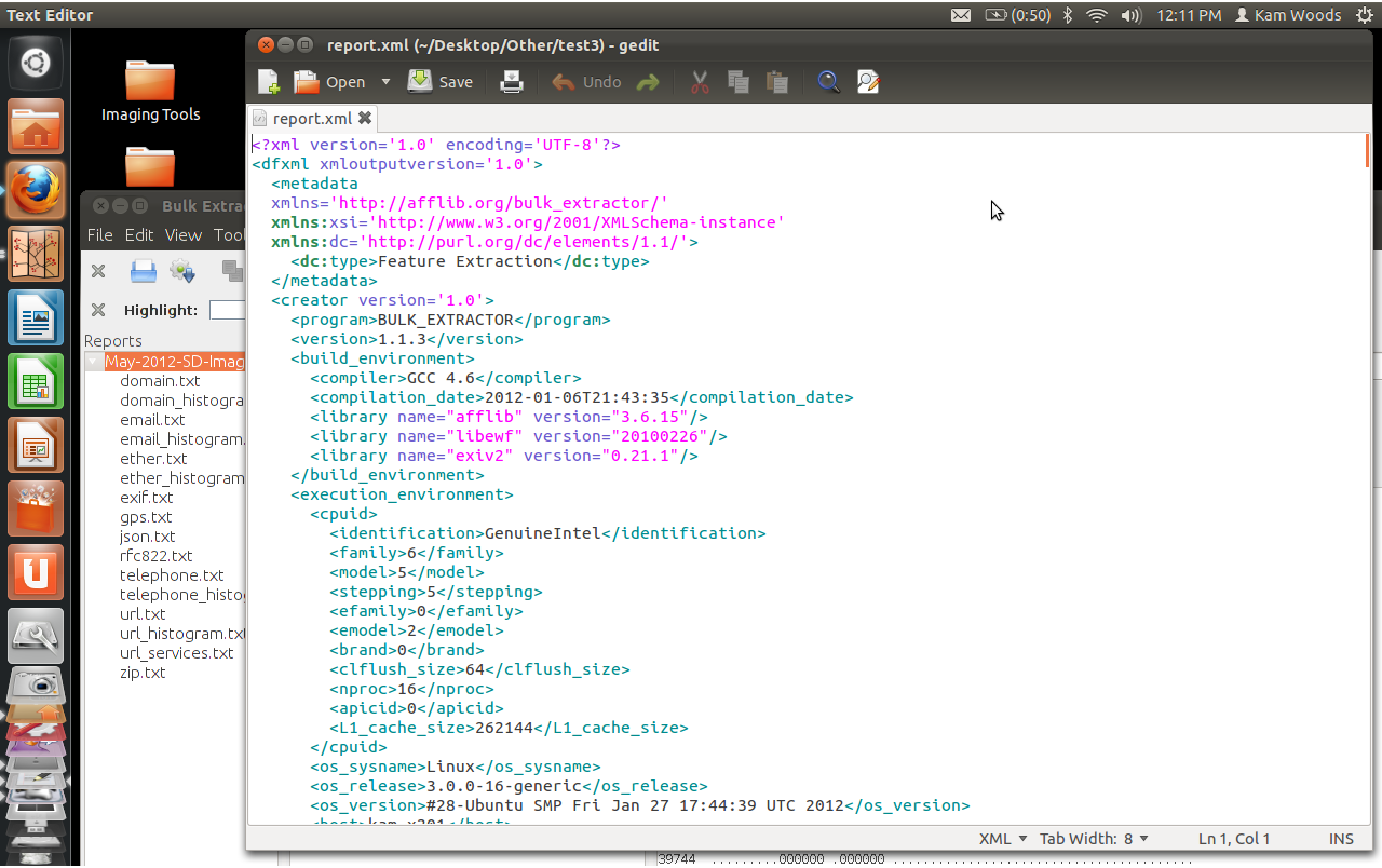
This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
−<dfxml version="1.0">
  −<metadata>
     <dc:type>Disk Image</dc:type>
   </metadata>
  −<creator version="1.0">
     <program>fiwalk</program>
     <version>4.0.2</version>
    −<build_environment>
       <compiler>GCC 4.6</compiler>
       <library name="afflib" version="3.7.1"/>
       <library name="libewf" version="20130303"/>
     </build_environment>
    −<execution_environment>
      −<command_line>
         fiwalk -f -X /home/bcadmin/Desktop/SampleData/sampleimage.xml /home/bcadmin/Desktop/SampleData
         /sampleimage.E01
       </command_line>
       <start_time>2013-03-12T00:08:28Z</start_time>
     </execution_environment>
   </creator>
  −<source>
     <image_filename>/home/bcadmin/Desktop/SampleData/sampleimage.E01</image_filename>
   </source>
   <!-- fs start: 0 -->
  −<volume offset="0">
     <partition_offset>0</partition_offset>
     <block_size>2048</block_size>
     <ftype>2048</ftype>
     <ftype_str>iso9660</ftype_str>
     <block_count>36839</block_count>
```

Exporting Filesystem Metadata - Output from fiwalk (XML)

```
<fileobject>
      <filename>Documents and Settings/All Users/Documents/
                 My Pictures/Sample Pictures/Blue hills.jpg
      </filename>
      ...
      <filesize>28521</filesize>
      <alloc>1</alloc>
      <used>1</used>
      <inode>6245</inode>
      ...
      <uid>0</uid>
      <gid>0</gid>
      <mtime>1208174400</mtime>
      <ctime>1257729636</ctime>
      <atime>1257729636</atime>
      <crtime>1257729636</crtime>
      <seq>2</seq>
      <libmagic>JPEG image data, JFIF standard 1.02</libmagic>
      <byte_runs>
       <run file_offset='0' fs_offset='0' img_offset='363200512'
         len='0'/>
      </byte_runs>
      <hashdigest type='MD5'>
          6fb2a38dc107eacb41cf1656e899cf70
      </hashdigest>
      <hashdigest type='SHA1'>
          4eee44b18576e84de7b163142b537d2fe6231845
      </hashdigest>
</fileobject>
```

Technical Metadata (about the System Used to do the Capture) in a Bulk Extractor Report



```
<?xml version='1.0' encoding='UTF-8'?>
<dfxml xmloutputversion='1.0'>
  <metadata
  xmlns='http://afflib.org/bulk_extractor/'
  xmlns:xsi='http://www.w3.org/2001/XMLSchema-instance'
  xmlns:dc='http://purl.org/dc/elements/1.1/'>
    <dc:type>Feature Extraction</dc:type>
  </metadata>
  <creator version='1.0'>
    <program>BULK_EXTRACTOR</program>
    <version>1.1.3</version>
    <build_environment>
      <compiler>GCC 4.6</compiler>
      <compilation_date>2012-01-06T21:43:35</compilation_date>
      <library name="afflib" version="3.6.15"/>
      <library name="libewf" version="20100226"/>
      <library name="exiv2" version="0.21.1"/>
    </build_environment>
    <execution_environment>
      <cpuid>
        <identification>GenuineIntel</identification>
        <family>6</family>
        <model>5</model>
        <stepping>5</stepping>
        <efamily>0</efamily>
        <emodel>2</emodel>
        <brand>0</brand>
        <clflush_size>64</clflush_size>
        <nproc>16</nproc>
        <apicid>0</apicid>
        <L1_cache_size>262144</L1_cache_size>
      </cpuid>
      <os_sysname>Linux</os_sysname>
      <os_release>3.0.0-16-generic</os_release>
      <os_version>#28-Ubuntu SMP Fri Jan 27 17:44:39 UTC 2012</os_version>
      <host>kam-x201</host>
```

Bulk Extractor Output*

File	Description
aes_keys.txt	AES encryption keys
alerts.txt	Processing errors
ccn.txt	Credit card numbers
ccn_track2.txt	Credit card “track 2” information, which has previously been found in some bank fraud cases
domain.txt	Internet domains found on the drive, including dotted-quad addresses found in text
email.txt	Email addresses
ether.txt	Ethernet MAC addresses found through IP packet carving of swap files and compressed system hibernation files and fragments
exif.txt	EXIF data from JPEG images and video segments
find.txt	Results of specific regular expression searches
gps.txt	Extracted GSP coordinates from Garmin XML and GPS-enabled JPEG files
ip.txt	IP addresses found through IP packet carving
json.txt	Extracted and validated JavaScript Object Notation fragments
kml.txt	Extracted KML files

*See http://afflib.org/archives/tag/bulk_extractor

Bulk Extractor Output (continued)*

File	Description
report.txt	DFMXL file that explains what happened
rfc822.txt	Email message headers including Date:, Subject:, and Message-ID: fields
tcp.txt	TCP flow information found through IP packet carving
telephone.txt	Phone numbers (US and other countries)
url.txt	URLs, typically found in browser caches, email messages, and pre-compiled into executables
url_searches.txt	Histogram of terms used in Internet searches
url_services.txt	Histogram of the domain name portion of all URLs found on the media
winpefect.txt	Windows prefetch files and fragments, recorded as XML
wordlist.txt	A list of all “words” extracted from the disk, useful for password cracking
wordlist_*.txt	The wordlist with duplicates removed, formatted to be imported into a popular password-cracking program
zip.txt	Information about ZIP file components found on media (including compound files such as MS Office documents)

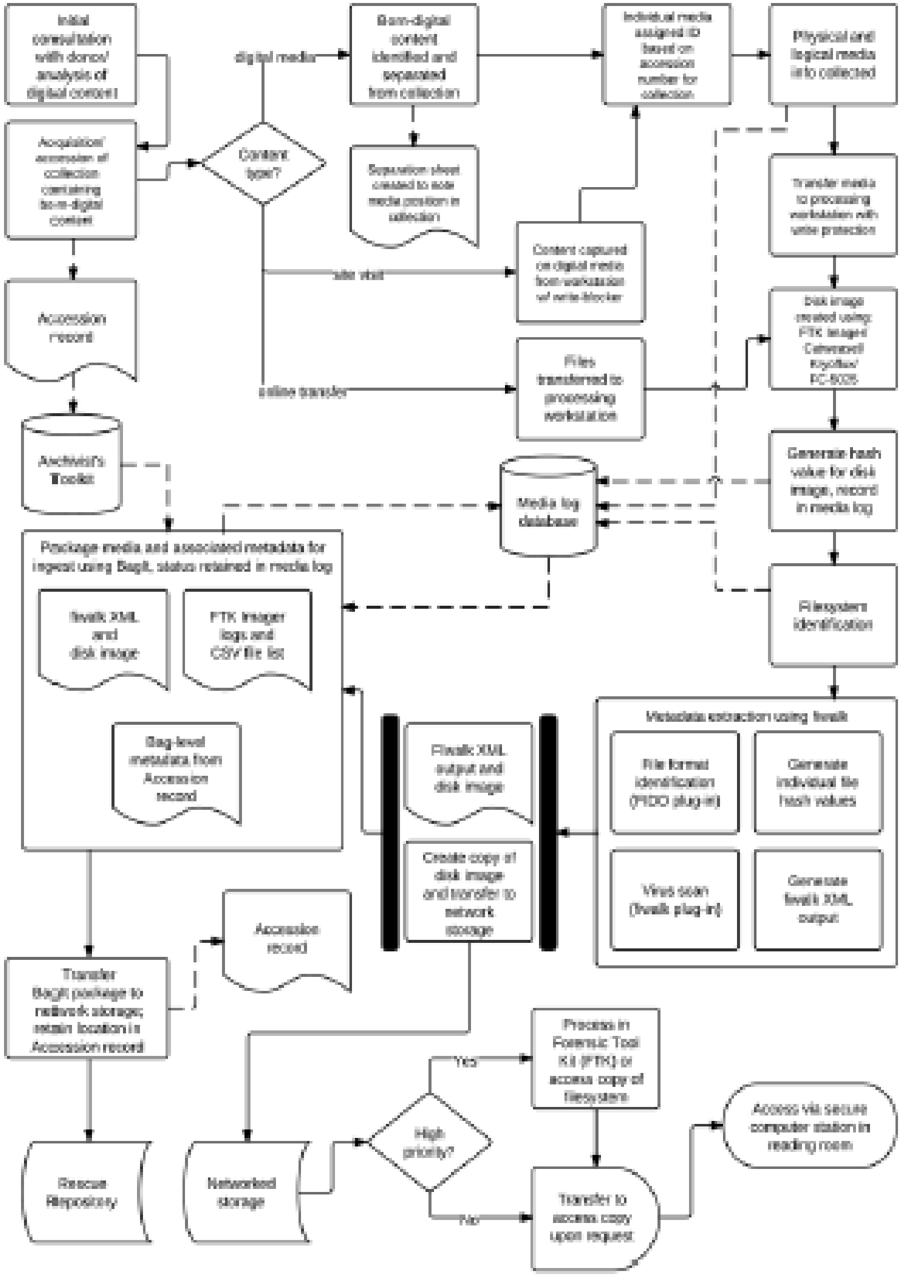
*See http://afflib.org/archives/tag/bulk_extractor

Two Sources of Workflow Examples

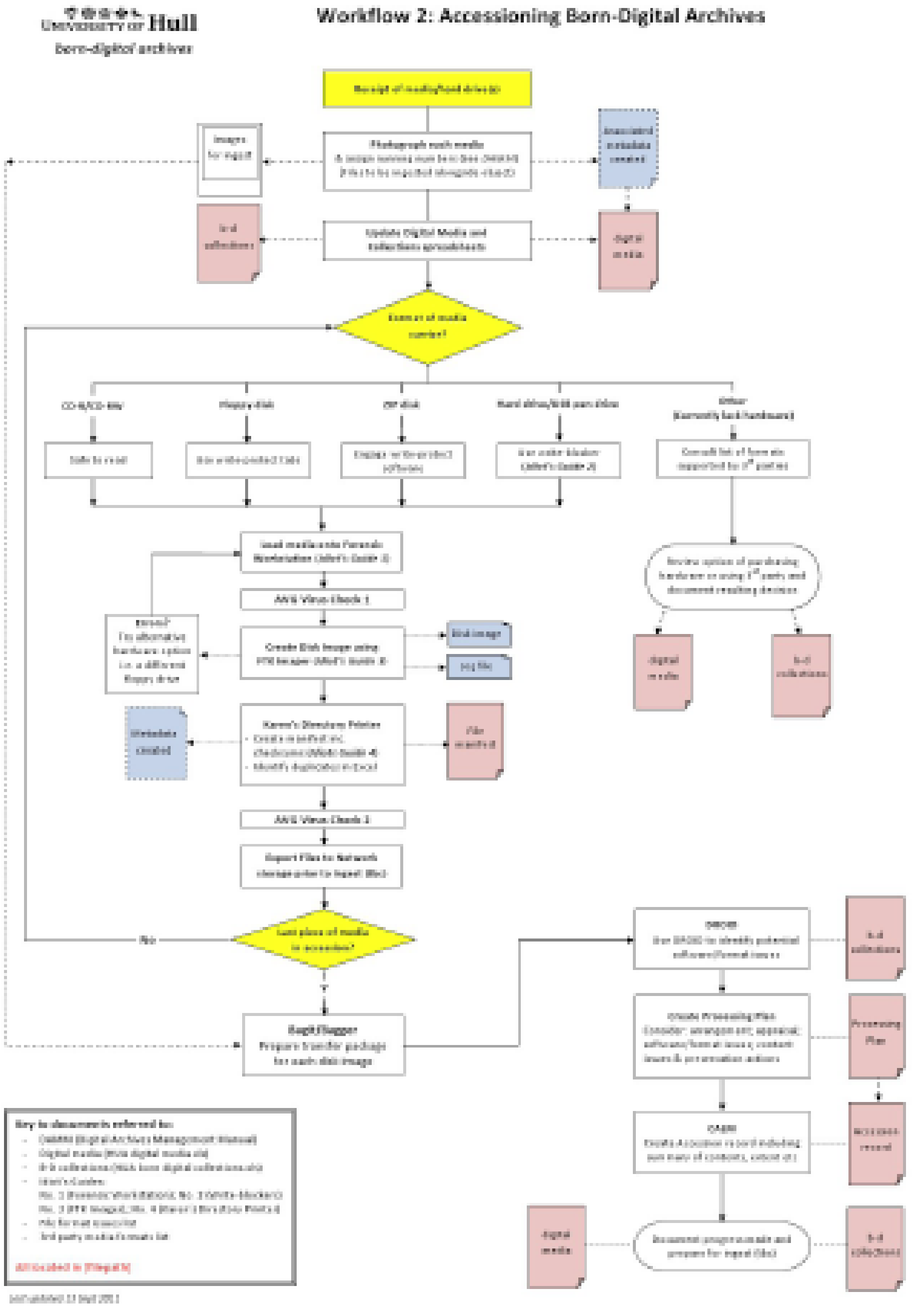
Martin J. Gengenbach, "'The Way We Do it Here': Mapping Digital Forensics Workflows in Collecting Institutions," A Master's Paper for the M.S. in L.S degree. August 2012. http://digitalcurationexchange.org/system/files/gengenbach-forensic-workflows-2012.pdf

AIMS Work Group, "AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship," January 2012. http://www2.lib.virginia.edu/aims/whitepaper/AIMS_final.pdf

Figure 1. Beinecke Rare Book and Manuscript Library, Yale University


Initial consultation with donor/ analysis of digital content
Acquisition/ accession of collection containing born-digital content
Accession record
Archivist's Toolkit
Content type?
digital media
site visit
online transfer
Born-digital content identified and separated from collection
Separation sheet created to note media position in collection
Content captured on digital media from workstation w/ write-blocker
Files transferred to processing workstation
Individual media assigned ID based on accession number for collection
Physical and logical media info collected
Transfer media to processing workstation with write protection
Disk image created using: FTK Imager/ Catweasel/ Kryoflux/ FC-5025
Generate hash value for disk image, record in media log
Filesystem identification
Media log database
Package media and associated metadata for ingest using BagIt, status retained in media log
fiwalk XML and disk image
FTK Imager logs and CSV file list
Bag-level metadata from Accession record
Fiwalk XML output and disk image
Create copy of disk image and transfer to network storage
Metadata extraction using fiwalk
File format identification (FIDO plug-in)
Generate individual file hash values
Virus scan (fiwalk plug-in)
Generate fiwalk XML output
Transfer BagIt package to network storage; retain location in Accession record
Accession record
Rescue Repository
Networked storage
High priority?
Yes
No
Process in Forensic Tool Kit (FTK) or access copy of filesystem
Transfer to access copy upon request
Access via secure computer station in reading room

Martin J. Gengenbach, "'The Way We Do it Here': Mapping Digital Forensics Workflows in Collecting Institutions," A Master's Paper for the M.S. in L.S degree. August, 2012.



AIMS Work Group, “AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship,” January 2012.

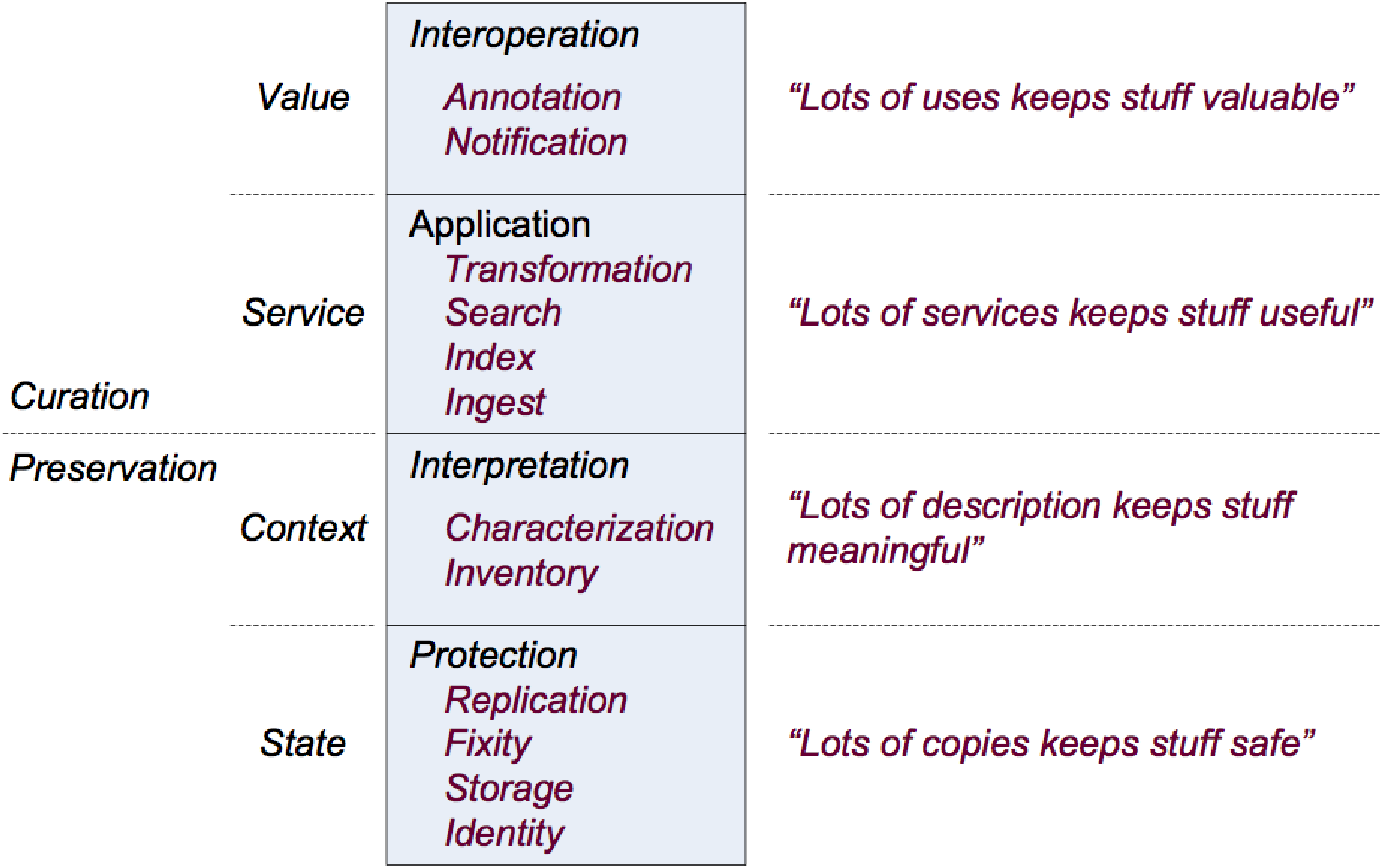
Other Workflow Examples

- Elford, Douglas, Nicholas Del Pozo, Snezana Mihajlovic, David Pearson, Gerard Clifton, and Colin Webb. "Media Matters: Developing Processes for Preserving Digital Objects on Physical Carriers at the National Library of Australia." Paper presented at the 74th IFLA General Conference and Council, Québec, Canada, August 10-14, 2008. http://archive.ifla.org/IV/ifla74/papers/084-Webb-en.pdf
- Glick, Kevin, and Eliot Wilczek. "Ingest Guide." Tufts University and Yale University, 2006. http://dca.lib.tufts.edu/features/nhprc/reports/ingest/index.html
- Klett, Fanny, Ann Hägerfors, and Kuldar Aas. "State-of-the-Art, Stakeholder Needs, Application Scenarios." PROTAGE Consortium, 2008. http://www.protage.eu/files/D1%201-State-of-the-art-Needs-Scenarios%20ver%201%200.pdf [For presentation of workflow, see especially p.49-71, 80-87]
- Mitchell, Marilyn, ed. *Library Workflow Redesign: Six Case Studies*. Washington, DC: Council on Library and Information Resources, 2007. http://www.clir.org/pubs/reports/pub139/pub139.pdf
- Morris, Steven P. and James Tuttle. “Curation and Preservation of Complex Data: The North Carolina Geospatial Data Archiving Project” Paper presented at DigCCurr2007: An International Symposium on Digital Curation, Chapel Hill, NC, April 18-20, 2007. http://ils.unc.edu/digccurr2007/papers/tuttle_paper_4-3.pdf [See also conference presentation: http://ils.unc.edu/digccurr2007/slides/tuttle_slides_4-3.pdf]
- Müller, Eva, Uwe Klosa, Peter Hansson, and Stefan Andersson. "Archiving Workflow between a Local Repository and the National Archive Experiences from the DiVA Project." Paper presented at the Third ECDL Workshop on Web Archives, Trondheim, Norway, August 21, 2003. http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=muller
- Owens, Evan. "Automated Workflow for the Ingest and Preservation of Electronic Journals." In *Archiving 2006: Final Program and Proceedings, May 23-26, 2006, Ottawa, Canada*, edited by Stephen Chapman and Scott A. Stovall, 109-12. Springfield, VA: Society for Imaging Science and Technology, 2006. http://www.portico.org/news/Archiving2006-Owens.pdf
- Underwood, W.E. and S.L. Laib. “PERPOS: An Electronic Records Repository and Archival Processing System.” Paper presented at DigCCurr2007: An International Symposium on Digital Curation, Chapel Hill, NC, April 18-20, 2007. http://ils.unc.edu/digccurr2007/papers/underwood_paper_6-3.pdf [See also conference presentation: http://ils.unc.edu/digccurr2007/slides/underwood_slides_6-3.pdf]
- Vardigan, Mary, and Cole Whiteman. "OAIS Meets ICPSR: Applying the OAIS Reference Model to the Social Science Archive Context." *Archival Science* 7. No. 1 (2007): 73–87. http://www.springerlink.com.libproxy.lib.unc.edu/content/50746212r6g21326/fulltext.pdf

A Big (Common) Idea:

Micro-Services

Merritt - California Digital Library



Value

Interoperation

Annotation
Notification

"Lots of uses keeps stuff valuable"

Service

Application

Transformation
Search
Index
Ingest

"Lots of services keeps stuff useful"

Curation

Preservation

Context

Interpretation

Characterization
Inventory

"Lots of description keeps stuff meaningful"

State

Protection

Replication
Fixity
Storage
Identity

"Lots of copies keeps stuff safe"

Figure 8 – Merritt micro-services

http://www.wf4ever-project.org/wiki/display/docs/RO%2Bpreservation%2Bservices

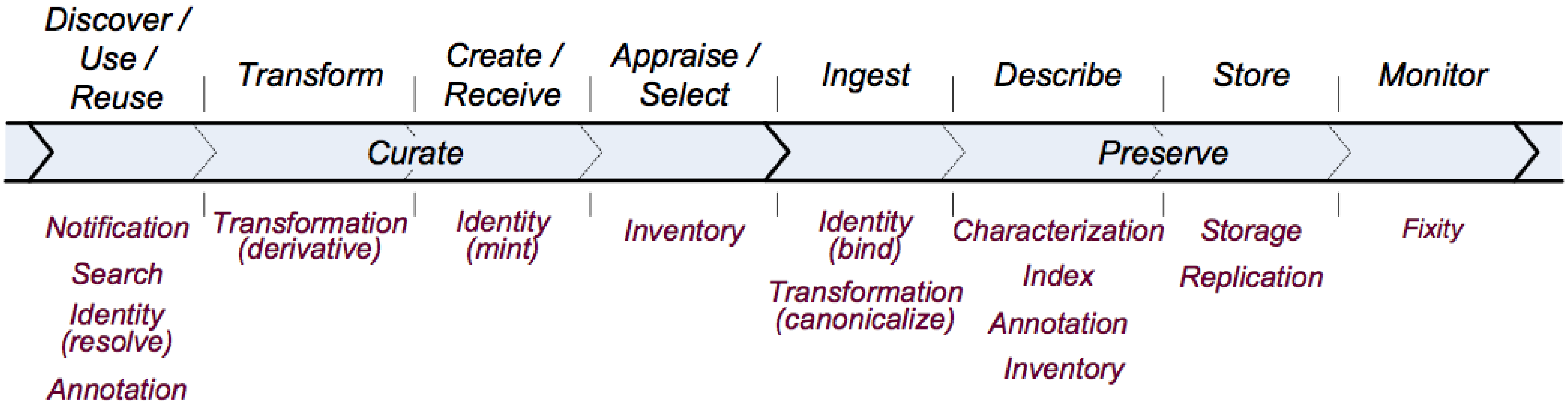
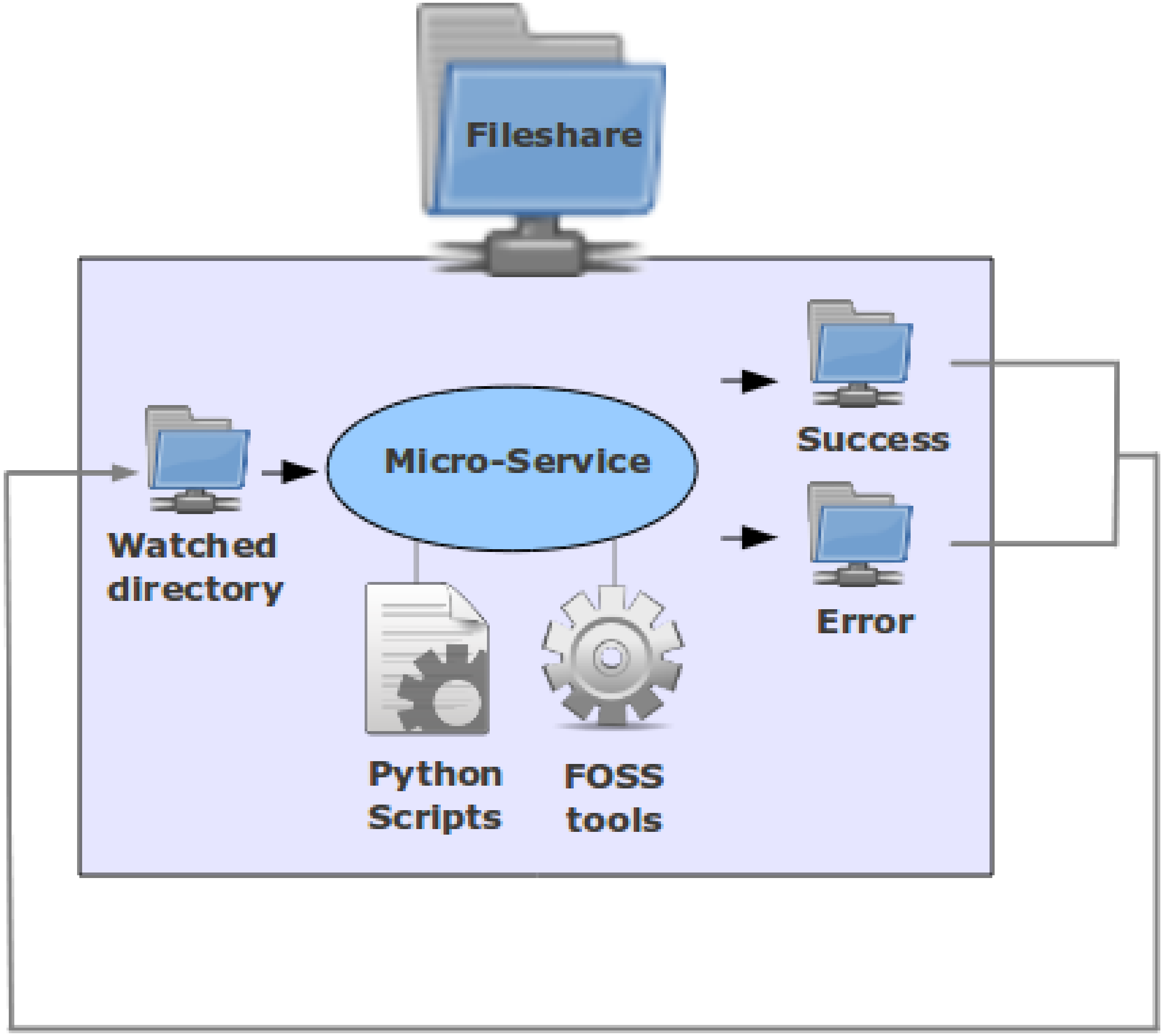


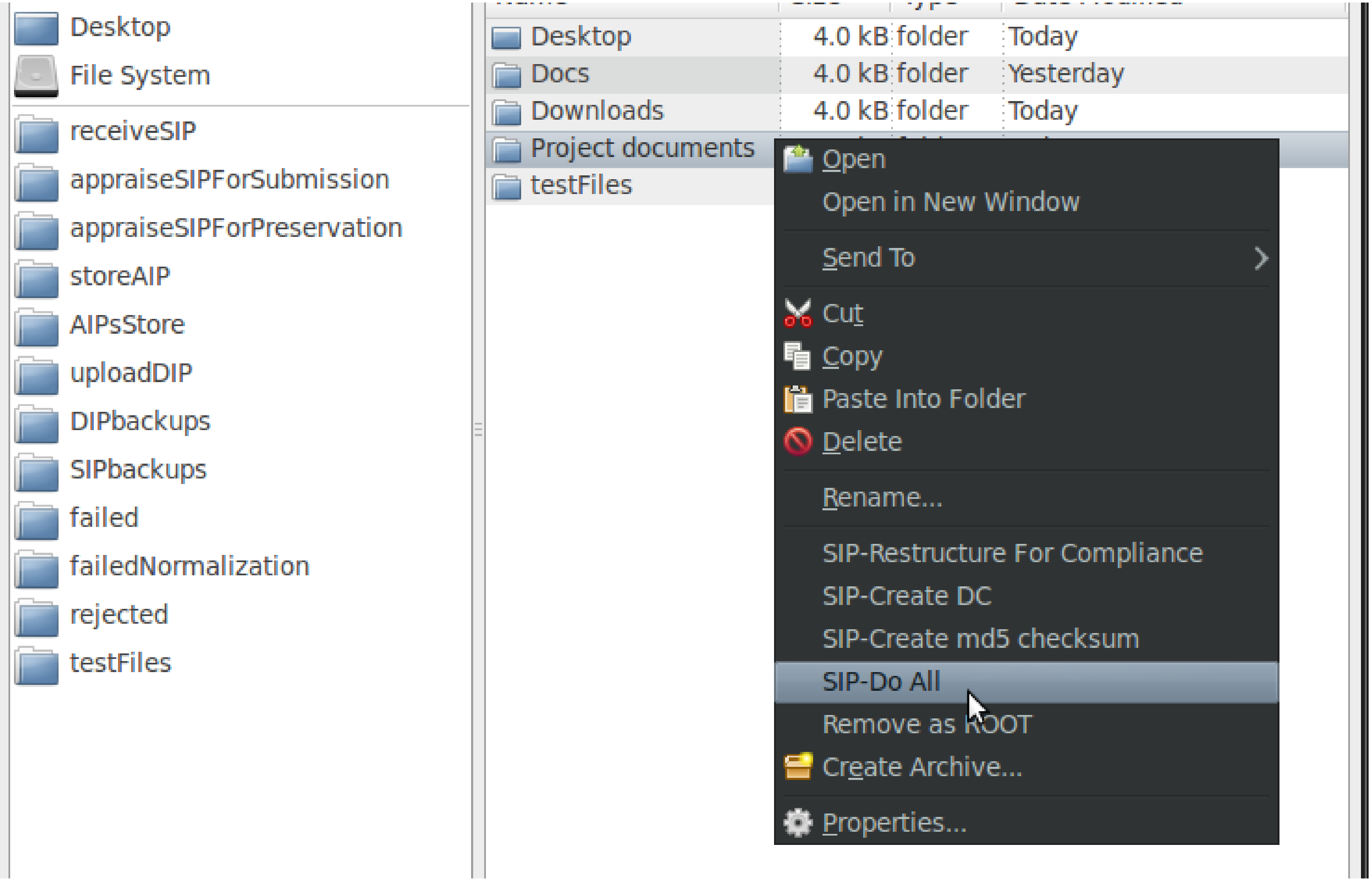
Figure 7 – Micro-service applicability throughout the curation lifecycle
[Adapted from Higgins]

http://www.wf4ever-project.org/wiki/display/docs/RO%2Bpreservation%2Bservices

Archivematica - Artefactual Systems

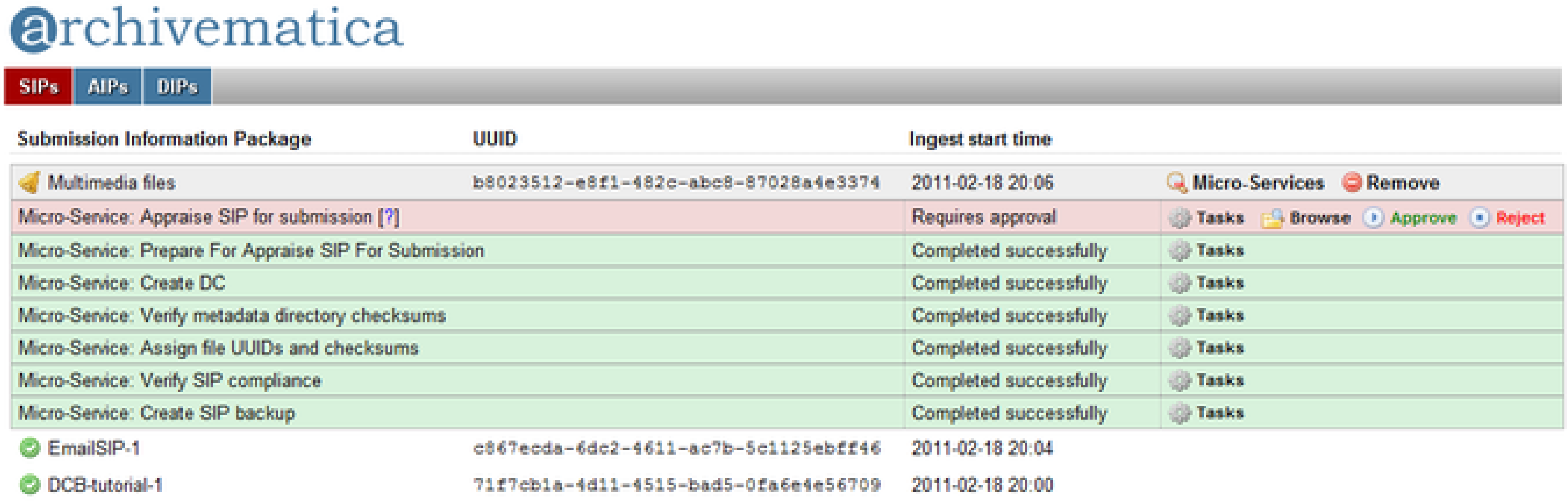


http://www.archivematica.org/wiki/index.php?title=File:Micro-service.png



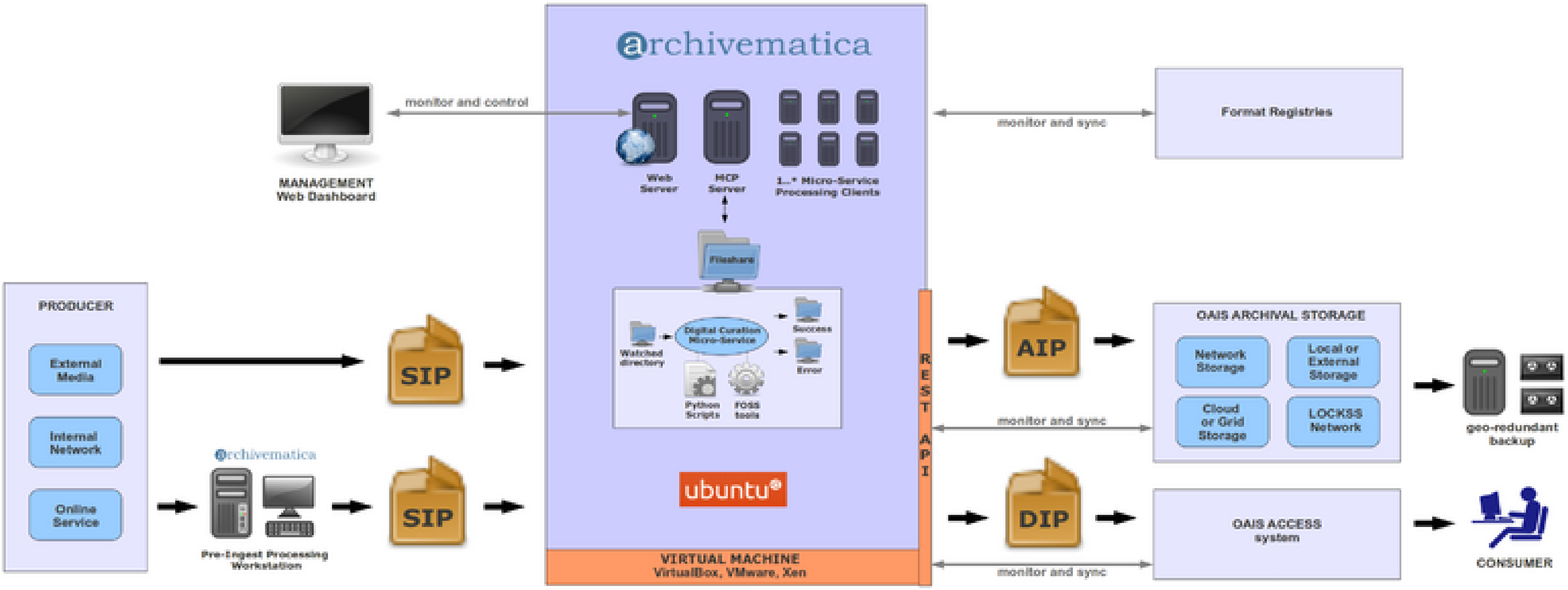
Desktop	4.0 kB	folder	Today
Docs	4.0 kB	folder	Yesterday
Downloads	4.0 kB	folder	Today
Project documents			
testFiles			

http://www.archivematica.org/wiki/index.php?title=File:Pre-ingest-1.png



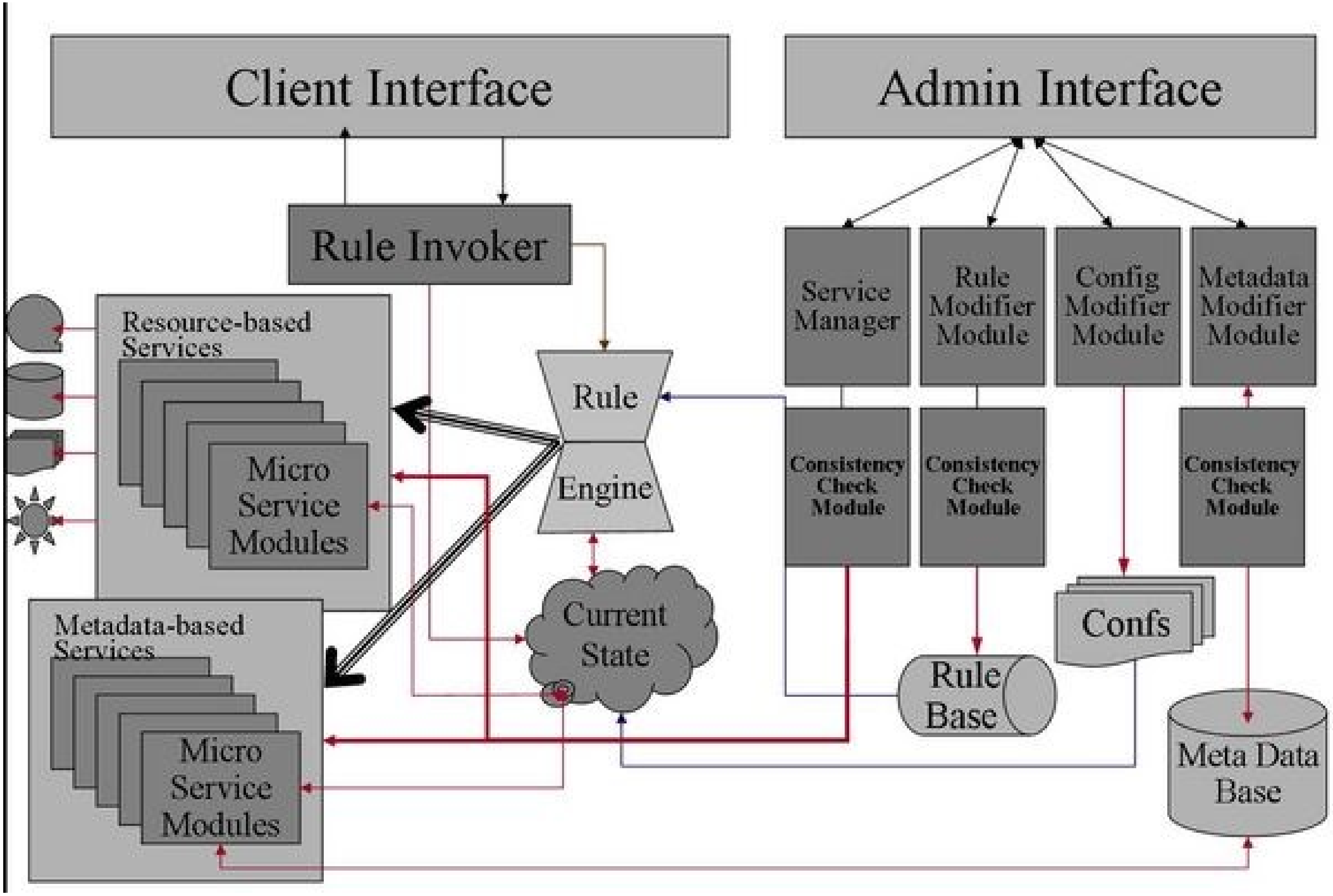
Submission Information Package	UUID	Ingest start time	
Multimedia files	b8023512-e8f1-482c-abc8-87028a4e3374	2011-02-18 20:06	Micro-Services Remove
Micro-Service: Appraise SIP for submission [?]		Requires approval	Tasks Browse Approve Reject
Micro-Service: Prepare For Appraise SIP For Submission		Completed successfully	Tasks
Micro-Service: Create DC		Completed successfully	Tasks
Micro-Service: Verify metadata directory checksums		Completed successfully	Tasks
Micro-Service: Assign file UUIDs and checksums		Completed successfully	Tasks
Micro-Service: Verify SIP compliance		Completed successfully	Tasks
Micro-Service: Create SIP backup		Completed successfully	Tasks
EmailSIP-1	c867ecda-6dc2-4611-ac7b-5c1125ebff46	2011-02-18 20:04	
DCB-tutorial-1	71f7cb1a-4d11-4515-bad5-0fa6e4e56709	2011-02-18 20:00	

http://www.archivematica.org/wiki/index.php?title=File:Dashboard-0.7.png

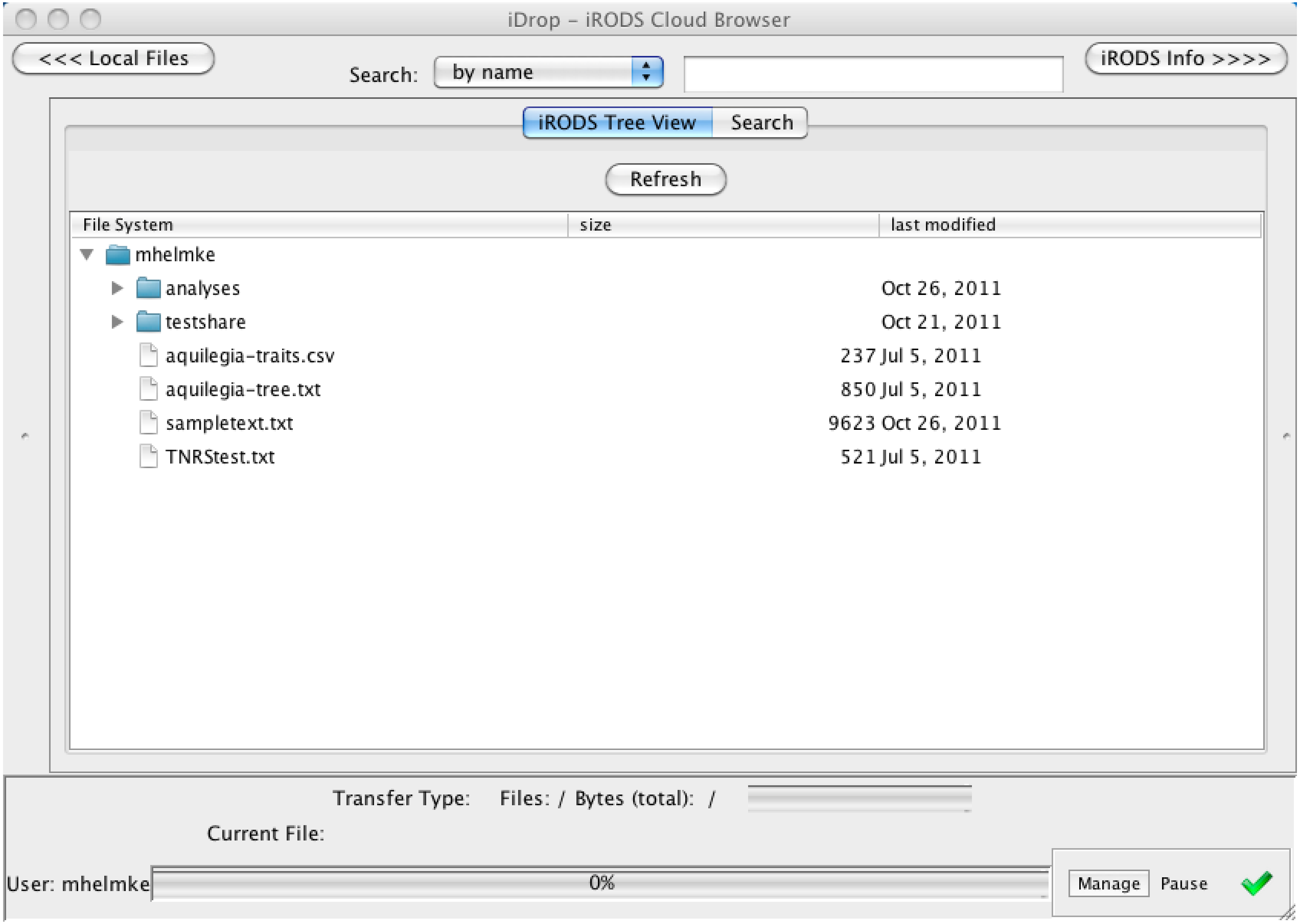


http://www.archivematica.org/wiki/index.php?title=File:Archivematica-0.8-beta-architecture.png

Integrated Rule-Oriented Data System (iRODS)



https://www.irods.org/images/thumb/5/5c/irodsArch.jpg/600px-irodsArch.jpg



Search: by name

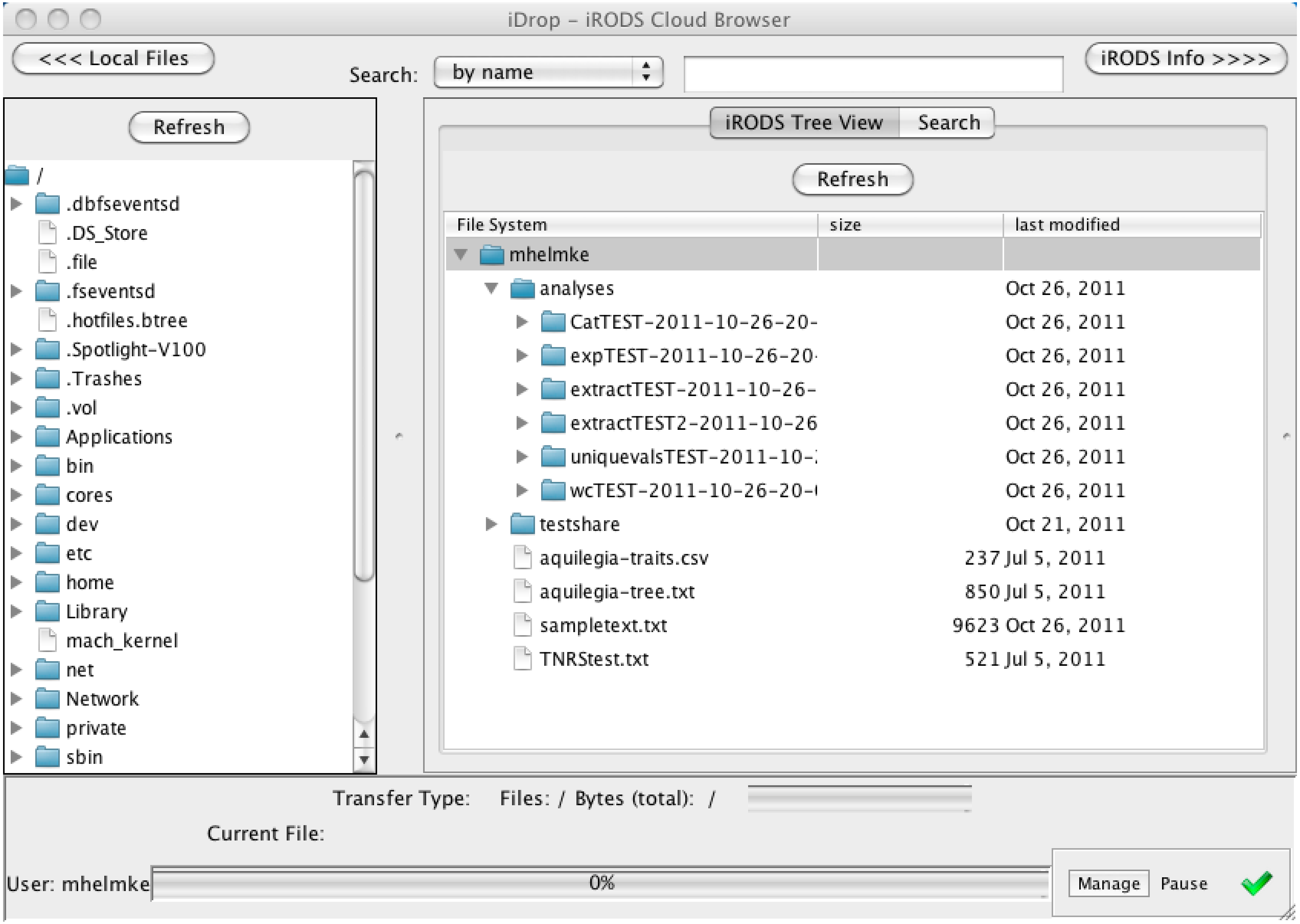
File System	size	last modified
mhelmke		
analyses		Oct 26, 2011
testshare		Oct 21, 2011
aquilegia-traits.csv	237	Jul 5, 2011
aquilegia-tree.txt	850	Jul 5, 2011
sampletext.txt	9623	Oct 26, 2011
TNRStest.txt	521	Jul 5, 2011

Transfer Type: Files: / Bytes (total): /

Current File:

User: mhelmke 0%

https://pods.iplantcollaborative.org/wiki/download/thumbnails/7519648/2011-11-02_1349.png



/
.dbfseventsd
.DS_Store
.file
.fseventsd
.hotfiles.btree
.Spotlight-V100
.Trashes
.vol
Applications
bin
cores
dev
etc
home
Library
mach_kernel
net
Network
private
sbin

File System	size	last modified
mhelmke		
analyses		Oct 26, 2011
CatTEST-2011-10-26-20-		Oct 26, 2011
expTEST-2011-10-26-20-		Oct 26, 2011
extractTEST-2011-10-26-		Oct 26, 2011
extractTEST2-2011-10-26		Oct 26, 2011
uniquevalsTEST-2011-10-		Oct 26, 2011
wcTEST-2011-10-26-20-		Oct 26, 2011
testshare		Oct 21, 2011
aquilegia-traits.csv	237	Jul 5, 2011
aquilegia-tree.txt	850	Jul 5, 2011
sampletext.txt	9623	Oct 26, 2011
TNRStest.txt	521	Jul 5, 2011

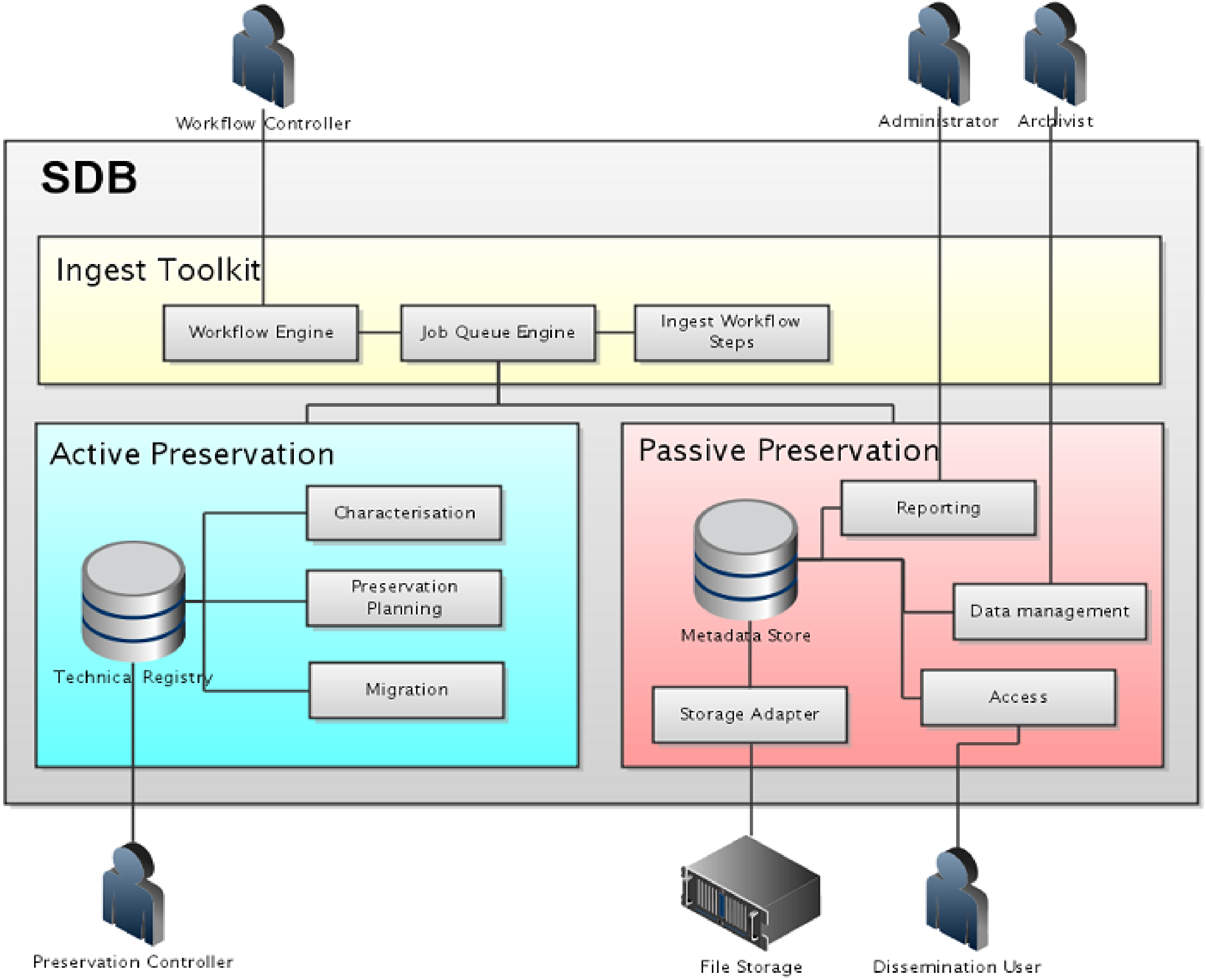
Transfer Type: Files: / Bytes (total): /

Current File:

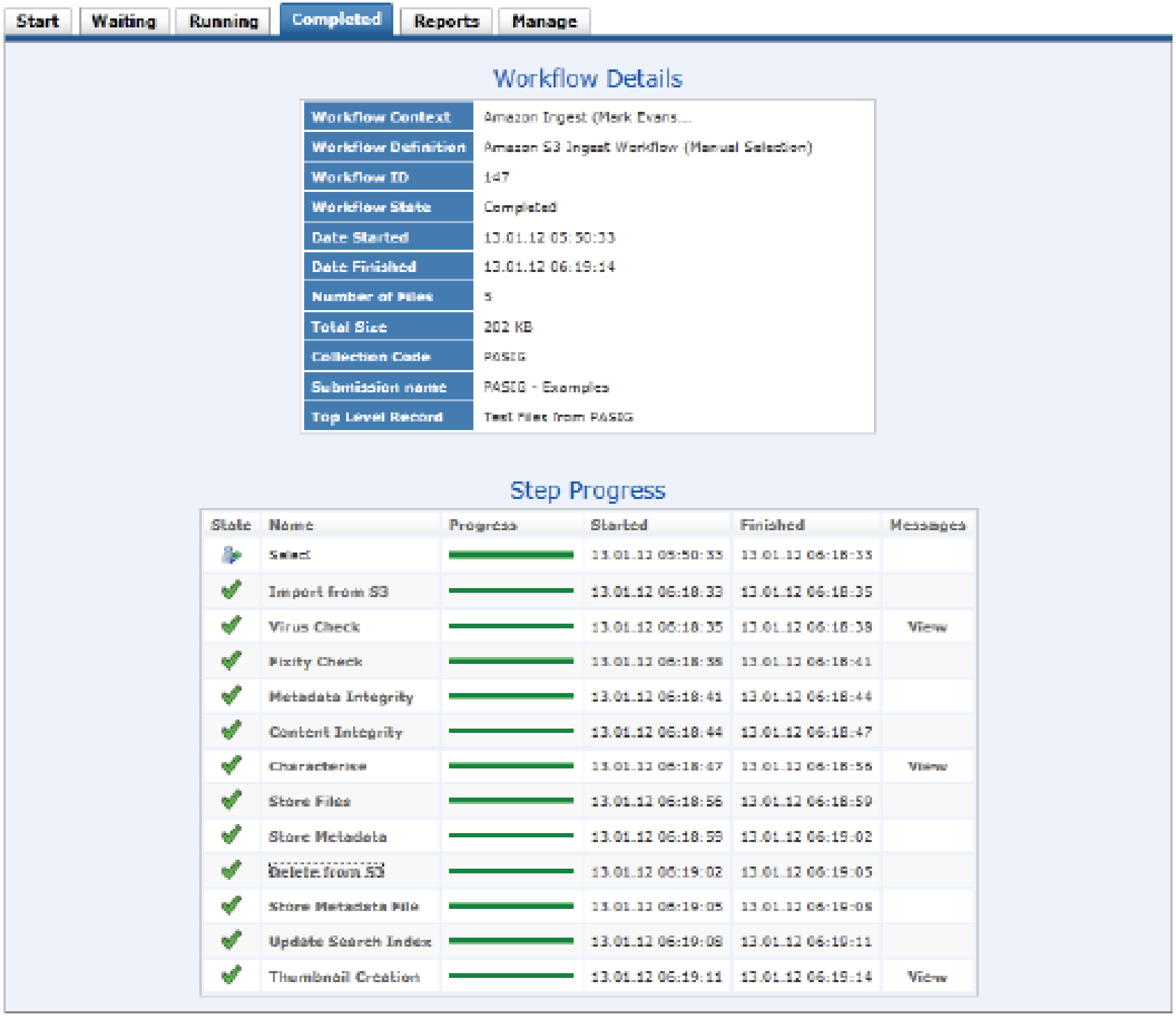
User: mhelmke 0%

https://pods.iplantcollaborative.org/wiki/download/thumbnails/7519648/2011-11-02_1353.png

Safety Deposit Box (SDB) - Tessella



http://lib.stanford.edu/files/PASIG-DC.ppt



Workflow Details

Workflow Context	Amazon Ingest (Mark Evans...
Workflow Definition	Amazon S3 Ingest Workflow (Manual Selection)
Workflow ID	147
Workflow State	Completed
Date Started	13.01.12 05:50:33
Date Finished	13.01.12 06:19:14
Number of Files	5
Total Size	202 KB
Collection Code	PASIG
Submission name	PASIG - Examples
Top Level Record	Test Files from PASIG

Step Progress

State	Name	Progress	Started	Finished	Messages
➔	Select	🟩	13.01.12 05:50:33	13.01.12 06:18:33	
✓	Import from S3	🟩	13.01.12 06:18:33	13.01.12 06:18:35	
✓	Virus Check	🟩	13.01.12 06:18:35	13.01.12 06:18:38	View
✓	Fixity Check	🟩	13.01.12 06:18:38	13.01.12 06:18:41	
✓	Metadata Integrity	🟩	13.01.12 06:18:41	13.01.12 06:18:44	
✓	Content Integrity	🟩	13.01.12 06:18:44	13.01.12 06:18:47	
✓	Characterise	🟩	13.01.12 06:18:47	13.01.12 06:18:56	View
✓	Store Files	🟩	13.01.12 06:18:56	13.01.12 06:18:59	
✓	Store Metadata	🟩	13.01.12 06:18:59	13.01.12 06:19:02	
✓	Delete from S3	🟩	13.01.12 06:19:02	13.01.12 06:19:05	
✓	Store Metadata File	🟩	13.01.12 06:19:05	13.01.12 06:19:08	
✓	Update Search Index	🟩	13.01.12 06:19:08	13.01.12 06:19:11	
✓	Thumbnail Creation	🟩	13.01.12 06:19:11	13.01.12 06:19:14	View

http://lib.stanford.edu/files/pasig-jan2012/13J4%20Evans%20PASIG%20CloudSolution-Mark%20Evans.pdf

Describing what you want to get done (process modeling)

Identifying a Process*

- Name it
  - *Verb-noun* – e.g. generate AIP, harvest web site
  - *Verb-qualifier-noun* – e.g. generate descriptive information, develop preservation strategy
  - *Verb-noun-noun* – e.g. assign file permissions, verify object integrity
- Ensure there is a clearly intended result
  - Test: *noun-is-verbed* form (e.g. AIP is generated, web site is harvested, object integrity is verified

*Sharp, Alec, and Patrick McDermott. *Wokflow Modeling: Tools for Process Improvement and Applications Development*. 2nd ed. Boston, MA: Artech House, 2009. p.40

Criteria for Identified Result*

1. *Discrete and identifiable* – “you can differentiate individual instances of the result, and it makes sense to talk about 'one of them‘”
2. *Countable* – “you can count how many of that result you've produced in an hour, a day, or a week”
3. *Essential* – “fundamentally necessary to the operation of the enterprise, not just a consequence of the current implementation,” i.e. “must focus on 'what, not who or how‘”

*Sharp, Alec, and Patrick McDermott. *Wokflow Modeling: Tools for Process Improvement and Applications Development*. 2nd ed. Boston, MA: Artech House, 2009. p.40-41

Exercise

- Consider a part of your total workflow and identify 5 to 10 sub-processes that are directly related to your process.
  - Remember the guidelines from Sharp and Dermott regarding naming processes and sub-processes
    - Name it: *Verb-noun, Verb-qualifier-noun* or V*erb-noun-noun*
    - Ensure that there is a clearly intended result - Test: *noun is verbed* form
- Write each sub-process on a sticky note
- Arrange the sticky notes into a workflow, using arrows to connect them on the large papers
- When possible, label the arrows between the sticky notes to clarify how the sub-processes are linked

Post-Mortem Discussion

Software to Support your Workflow

- Did you identify specific tools to support parts of your workflow?
- Did you identify any gaps (no tool support) or overlaps (multiple tools to support)?

Selection and Evaluation of Tools

- How would you decide which tools to adopt?
- What criteria would you use to evaluate the tools you’ve chosen?

For Further Consideration – The “Three R’s”

- Roles (who are the actors who complete steps in the process?)
- Responsibilities (what are the individual steps that each actor performs?)
- Routes (what are the flows and decisions that connect the steps and define the path?

*Sharp, Alec, and Patrick McDermott. *Workflow Modeling: Tools for Process Improvement and Applications Development*. 2nd ed. Boston, MA: Artech House, 2009. p.203