



ARKIVVERKET
RIKSARKIVET

Looking into Archivematica og AtoM:

Evaluating State of the Art Software for Long-term Preservation and Access

Arne-Kristian Groven,

Senior Advisor,

The National Archives of Norway,

NorDig,

Stockholm, 25.-26.10 2017

This presentation aims at

- Giving a brief introduction to Archivematica and AtoM, and the technologies they are based on
- Showing small examples from an end-user perspective, using screenshots
- Sharing some of the experiences gained so far and looking a bit ahead

But first... A few words about the work context behind this presentation

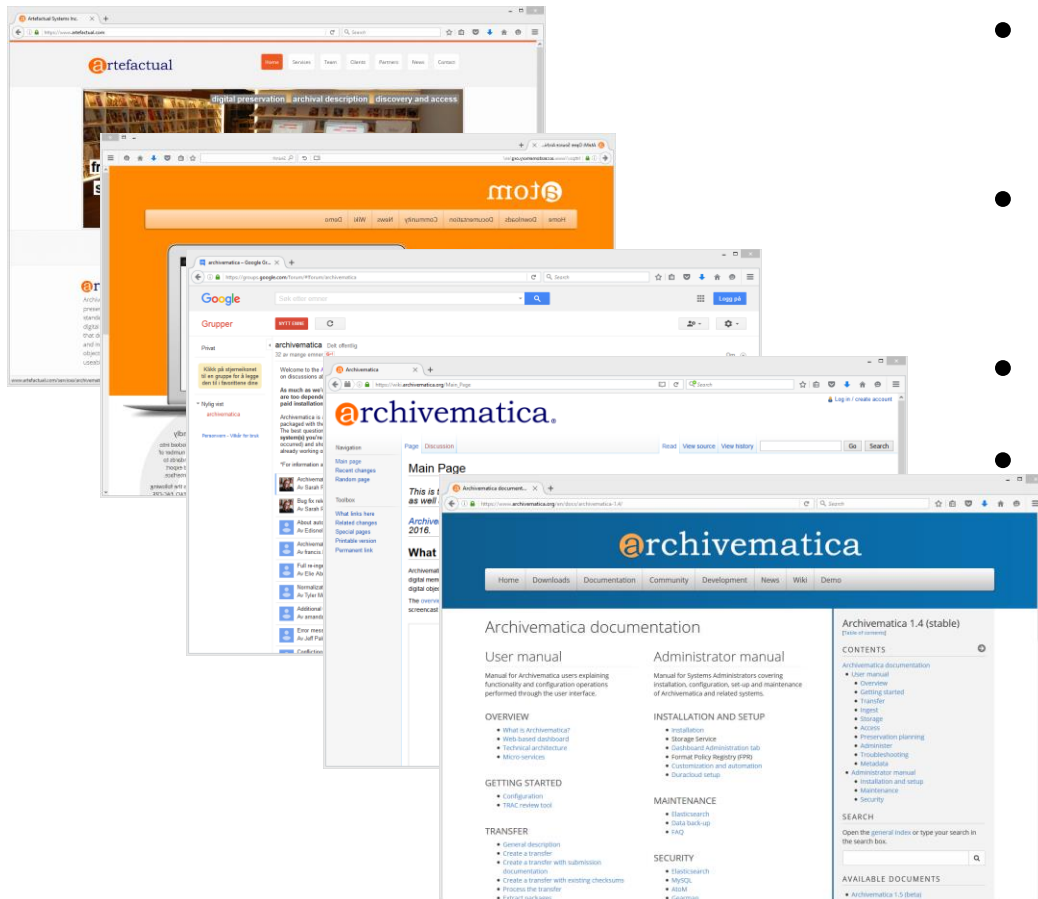
- My work is about evaluating new, innovative technologies and processes
- So this is not a report from the production line at the National Archives of Norway
- But (some of it) might be some day?!
- This work has been performed in cooperation with partners from the Norwegian private and inter-municipal archival sectors

Technologies we are looking into

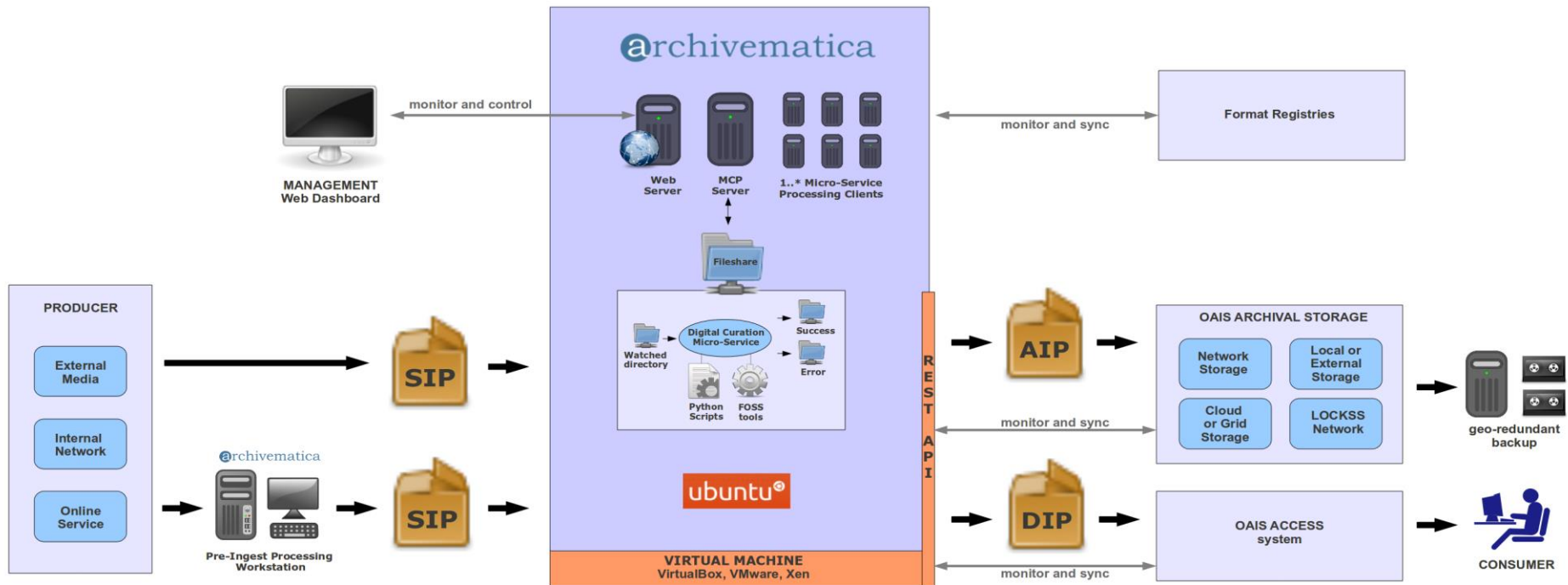
- Archivematica is a free and open-source digital preservation system.
- Designed to maintain standards-based, long-term access to collections of digital objects
- Archivematica is packaged with the web-based content management system AtoM for access to your digital objects
- The vendor is Artefactual Systems Inc.

For in-depth information: Vendor & Product pages, Wiki & Forums

- <https://www.artefactual.com/>
- <https://www.accesstomemory.org>
- <https://groups.google.com/forum/#!forum/archivematica>
- <https://wiki.archivematica.org>
- <https://www.archivematica.org>

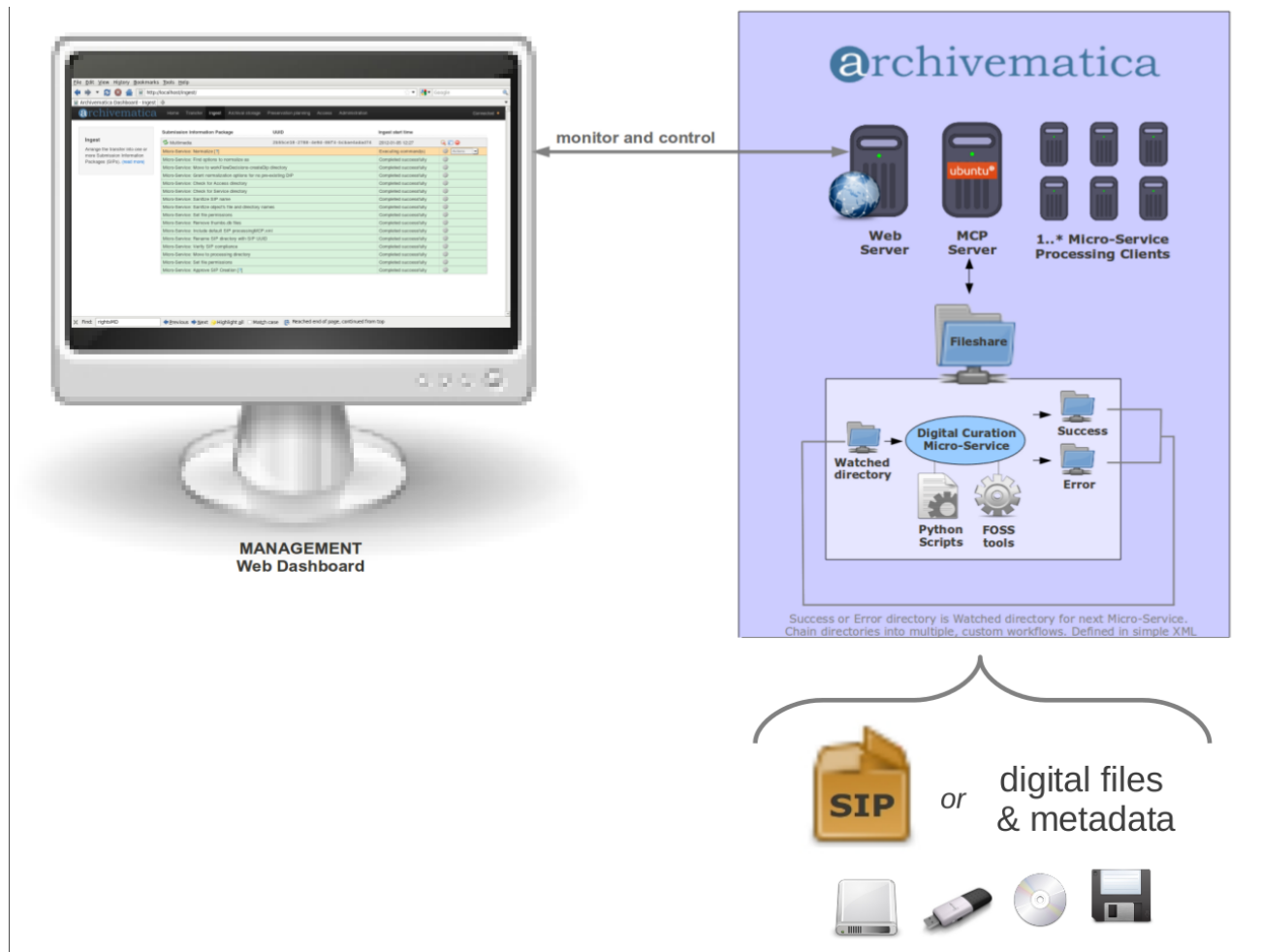


The Archivemata architecture



Kilde: <https://www.archivemata.org/wiki/File:Archivemata-0.8-beta-architecture.png>

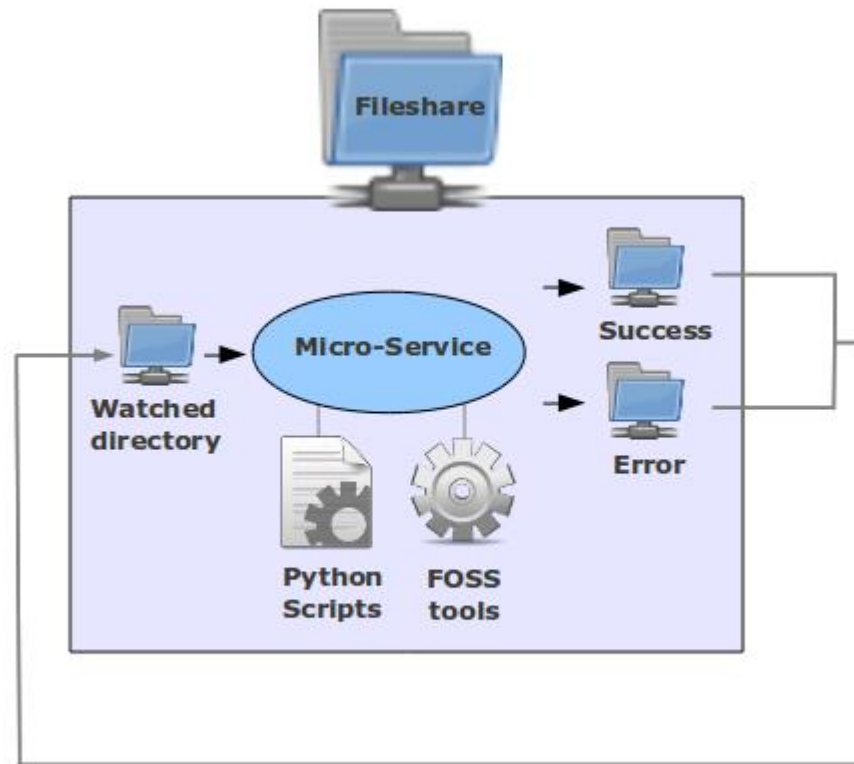
Web access... role-based access control



Kilde: <https://blogs.loc.gov/thesignal/2012/10/archivematica-and-the-open-source-mindset-for-digital-preservation-systems/>

Mico services deep down...

A (semi-) automated factory



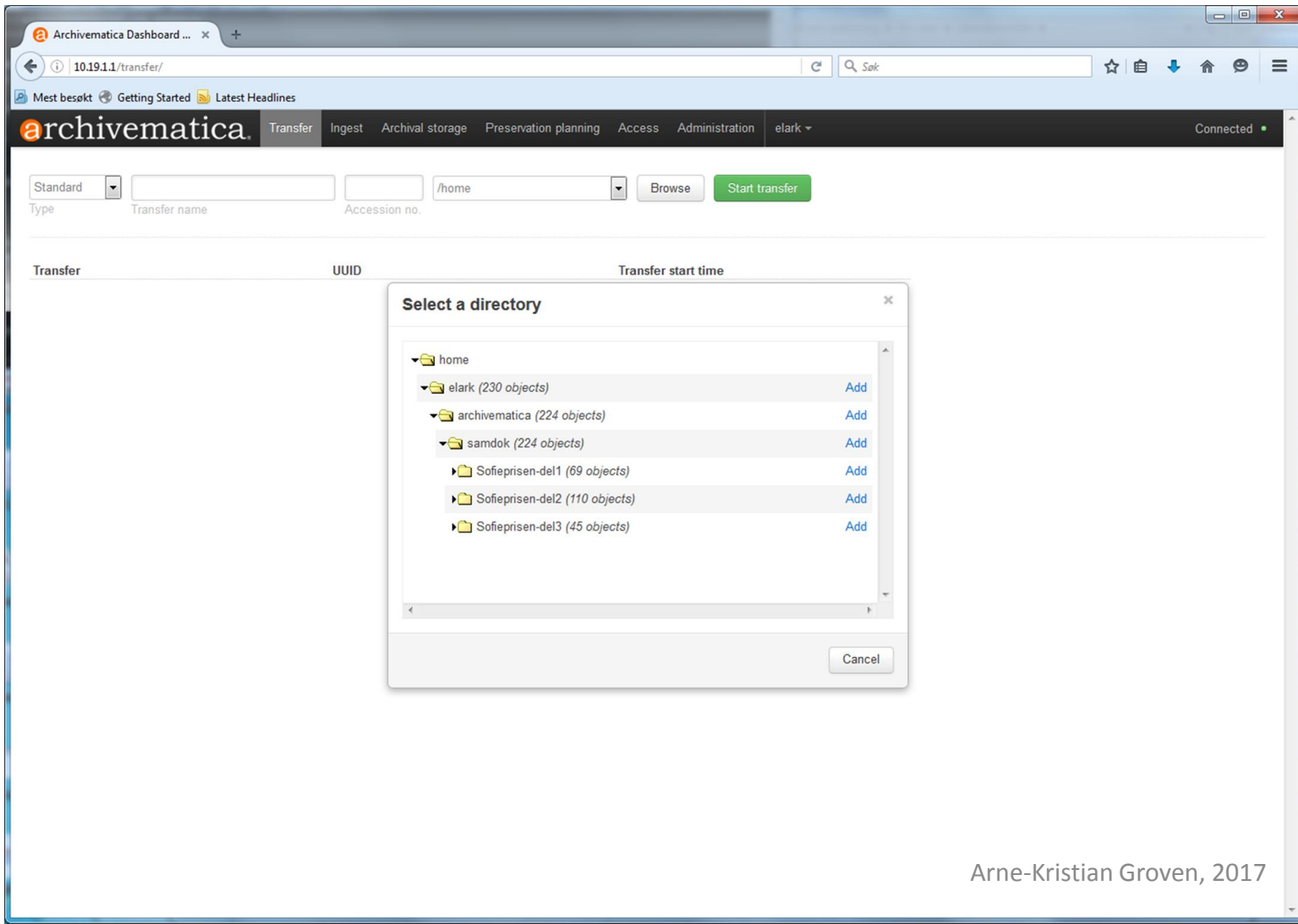
Kilde: <https://www.archivematica.org/wiki/File:Archivematica-0.8-beta-architecture.png>

Arne-Kristian Groven, 2017

Archivematica: A few screenshots to illustrate the GUI and the user interaction

- Content upload
- DC, Dublin Core, information edit
- File identification user interaction
- SIP (OAIS-Submission Information Package) generation
- File format conversion/migration, called normalization, generating formats used by DIPs and AIPs
- A look into a normalized structure

Inserting the content



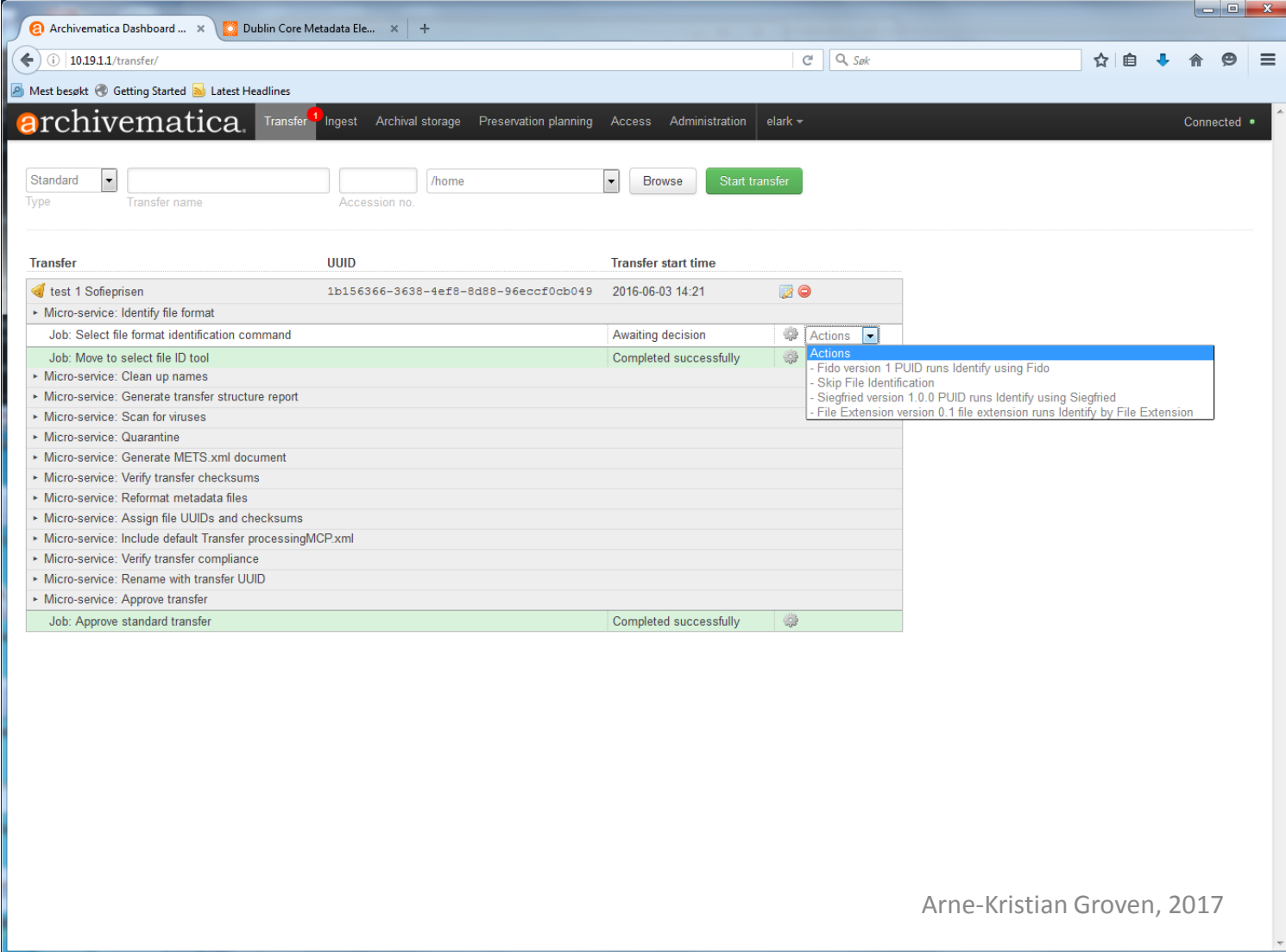
Dublin Core (DC) editing

The screenshot shows the Archivematica Dashboard in a web browser. The browser's address bar displays the URL `10.19.1.1/transfer/1b156366-3638-4ef8-8d88-96eccf0cb049/metadata/add/`. The dashboard's navigation bar includes links for 'Mest besøkt', 'Getting Started', and 'Latest Headlines'. The main navigation menu contains 'Transfer' (highlighted with a red notification badge), 'Ingest', 'Archival storage', 'Preservation planning', 'Access', 'Administration', and 'elark'. A breadcrumb trail at the top of the content area reads 'Transfer / test 1 Sofieprisen / Metadata / Add'. The 'Metadata' section is titled 'test 1 Sofieprisen' and contains several form fields for Dublin Core metadata:

- Title:** Sofieprisen 2010
- Part of AIC:** (empty field)
- Optional:** leave blank if unsure
- Creator:** (empty field)
- Subject:** Prisutdelingen 2010
- Description:** Bilder fra prisutdelingen
- Publisher:** (empty field)
- Contributor:** (empty field)
- Date:** 2010-05-15
- Use ISO 8061:** (YYYY-MM-DD or YYYY-MM-DD/YYYY-MM-DD)
- Format:** (empty field)
- Identifier:** (empty field)
- Source:** (empty field)

Arne-Kristian Groven, 2017

File format identification, either selecting tool options interactively or preconfigured



The screenshot displays the Archivematica Dashboard interface. At the top, there are browser tabs for 'Archivematica Dashboard' and 'Dublin Core Metadata Element'. The address bar shows '10.19.1.1/transfer/'. The navigation bar includes links for 'Transfer', 'Ingest', 'Archival storage', 'Preservation planning', 'Access', 'Administration', and 'elark'. A 'Connected' status indicator is visible on the right.

Below the navigation bar, there is a form for initiating a transfer. It includes a 'Type' dropdown set to 'Standard', a 'Transfer name' field, an 'Accession no.' field, a '/home' directory path, a 'Browse' button, and a green 'Start transfer' button.

The main content area shows a table of transfer jobs. The table has columns for 'Transfer', 'UUID', and 'Transfer start time'. The first job is 'test 1 Sofieprisen' with UUID '1b156366-3638-4ef8-8d88-96eccf0cb049' and start time '2016-06-03 14:21'. The job status is 'Awaiting decision'. A dropdown menu is open next to the job, showing a list of actions:

- Fido version 1 PUID runs Identify using Fido
- Skip File Identification
- Siegfried version 1.0.0 PUID runs Identify using Siegfried
- File Extension version 0.1 file extension runs Identify by File Extension

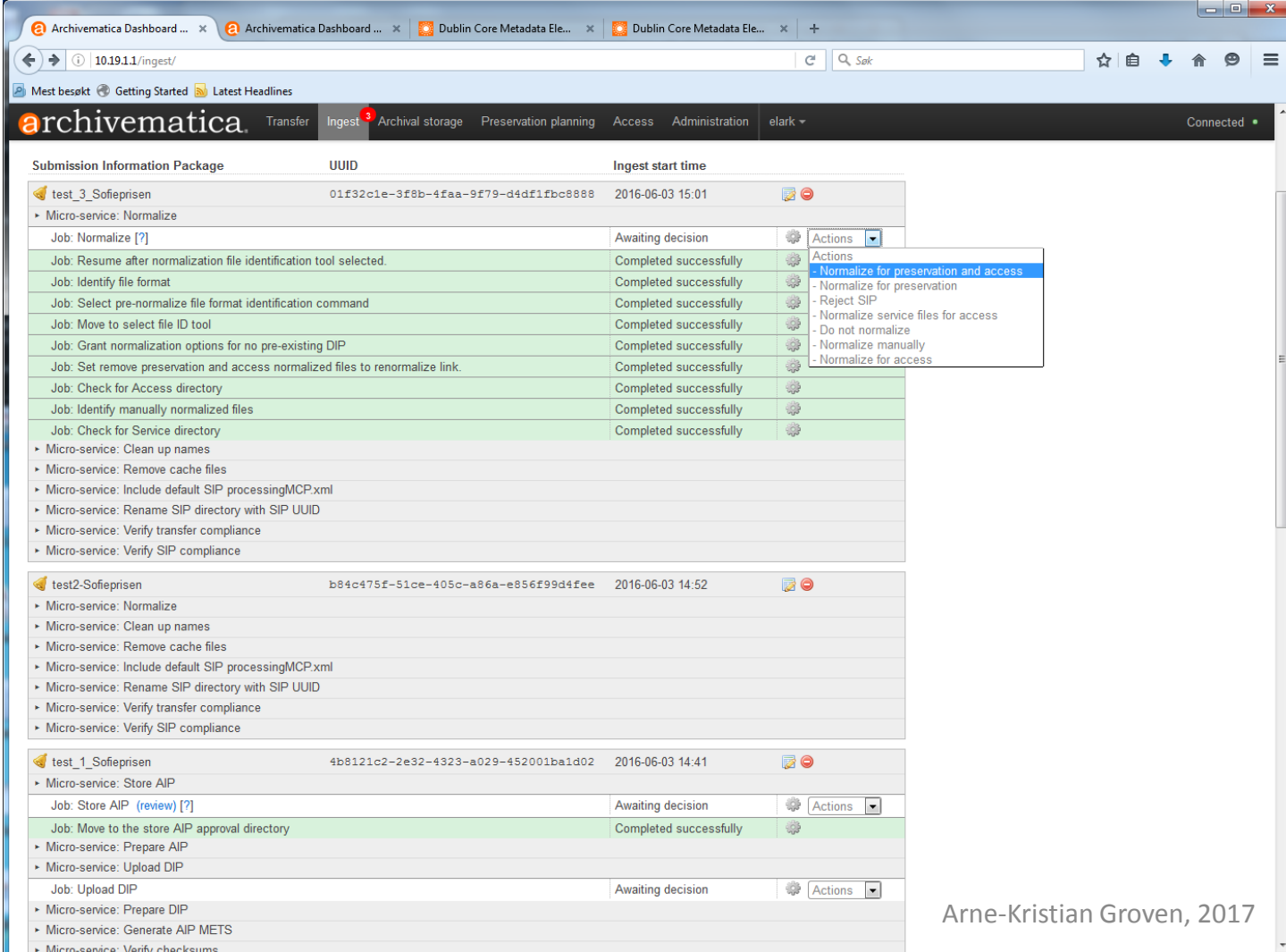
The table also shows a list of micro-services for the transfer, including 'Identify file format', 'Clean up names', 'Generate transfer structure report', 'Scan for viruses', 'Quarantine', 'Generate METS.xml document', 'Verify transfer checksums', 'Reformat metadata files', 'Assign file UUIDs and checksums', 'Include default Transfer processingMCP.xml', 'Verify transfer compliance', 'Rename with transfer UUID', and 'Approve transfer'. The final job in the list is 'Approve standard transfer', which is marked as 'Completed successfully'.

Generating a SIP

The screenshot shows the Archivematica Dashboard interface. The browser address bar displays '10.19.1.1/transfer/'. The dashboard header includes the Archivematica logo and navigation tabs: Transfer, Ingest, Archival storage, Preservation planning, Access, Administration, and elark. The 'Transfer' tab is active. Below the header, there are input fields for 'Type' (set to 'Standard'), 'Transfer name', 'Accession no.', and a 'Browse' button. A green 'Start transfer' button is also present. The main content area displays a table of transfer jobs.

Transfer	UUID	Transfer start time
test 1 Sofieprisen	1b156366-3638-4ef8-8d88-96eccf0cb049	2016-06-03 14:21
<ul style="list-style-type: none"> Micro-service: Create SIP from Transfer <ul style="list-style-type: none"> Job: Check transfer directory for objects Job: Move to SIP creation directory for completed transfers Job: Create SIP from transfer objects Job: Serialize Dublin Core metadata to disk Job: Move to processing directory Job: Create SIP(s) [?] Job: Load options to create SIPs Job: Check transfer directory for objects Micro-service: Complete transfer Micro-service: Examine contents Micro-service: Validation Micro-service: Characterize and extract metadata Micro-service: Update METS.xml document Micro-service: Extract packages Micro-service: Identify file format <ul style="list-style-type: none"> Job: Identify file format Job: Determine which files to identify Job: Select file format identification command Job: Move to select file ID tool Micro-service: Clean up names Micro-service: Generate transfer structure report Micro-service: Scan for viruses Micro-service: Quarantine Micro-service: Generate METS.xml document Micro-service: Verify transfer checksums Micro-service: Reformat metadata files Micro-service: Assign file UUIDs and checksums Micro-service: Include default Transfer processingMCP.xml Micro-service: Verify transfer compliance Micro-service: Rename with transfer UUID Micro-service: Approve transfer 		

Interactive decision points, (Here: normalization/file format conversion)



The screenshot displays the Archivematica Dashboard interface. The browser address bar shows the URL `10.19.11/ingest/`. The dashboard header includes the Archivematica logo and navigation tabs: Transfer, Ingest, Archival storage, Preservation planning, Access, Administration, and elark. The 'Ingest' tab is active, showing a list of submission information packages.

Submission Information Package	UUID	Ingest start time
test_3_Sofieprisen	01f32c1e-3f8b-4faa-9f79-d4df1fbc8888	2016-06-03 15:01
Micro-service: Normalize		
Job: Normalize [?]	Awaiting decision	Actions
Job: Resume after normalization file identification tool selected.	Completed successfully	Actions
Job: Identify file format	Completed successfully	Normalize for preservation and access
Job: Select pre-normalize file format identification command	Completed successfully	Normalize for preservation
Job: Move to select file ID tool	Completed successfully	Reject SIP
Job: Grant normalization options for no pre-existing DIP	Completed successfully	Normalize service files for access
Job: Set remove preservation and access normalized files to renormalize link.	Completed successfully	Do not normalize
Job: Check for Access directory	Completed successfully	Normalize manually
Job: Identify manually normalized files	Completed successfully	Normalize for access
Job: Check for Service directory	Completed successfully	
Micro-service: Clean up names		
Micro-service: Remove cache files		
Micro-service: Include default SIP processingMCP.xml		
Micro-service: Rename SIP directory with SIP UUID		
Micro-service: Verify transfer compliance		
Micro-service: Verify SIP compliance		
test2-Sofieprisen	b84c475f-51ce-405c-a86a-e856f99d4fee	2016-06-03 14:52
Micro-service: Normalize		
Micro-service: Clean up names		
Micro-service: Remove cache files		
Micro-service: Include default SIP processingMCP.xml		
Micro-service: Rename SIP directory with SIP UUID		
Micro-service: Verify transfer compliance		
Micro-service: Verify SIP compliance		
test_1_Sofieprisen	4b8121c2-2e32-4323-a029-452001ba1d02	2016-06-03 14:41
Micro-service: Store AIP		
Job: Store AIP (review) [?]	Awaiting decision	Actions
Job: Move to the store AIP approval directory	Completed successfully	
Micro-service: Prepare AIP		
Micro-service: Upload DIP		
Job: Upload DIP	Awaiting decision	Actions
Micro-service: Prepare DIP		
Micro-service: Generate AIP METS		
Micro-service: Verify checksums		

An action menu is open for the 'Job: Normalize [?]' row, showing options: 'Normalize for preservation and access', 'Normalize for preservation', 'Reject SIP', 'Normalize service files for access', 'Do not normalize', 'Normalize manually', and 'Normalize for access'.

Arne-Kristian Groven, 2017

Resulting structure, ready for review

Archivematica Dashboard ... x Archivematica Dashboard ... x Dublin Core Metadata Ele... x Dublin Core Metadata Ele... x +

10.19.1.1/ingest/preview/normalization/6dd503c7-28cc-41c4-96fd-e87b3a8c8bdf/

Mest besøkt Getting Started Latest Headlines

archivematica Transfer Ingest ³ Archival storage Preservation planning Access Administration elark v

Review normalization

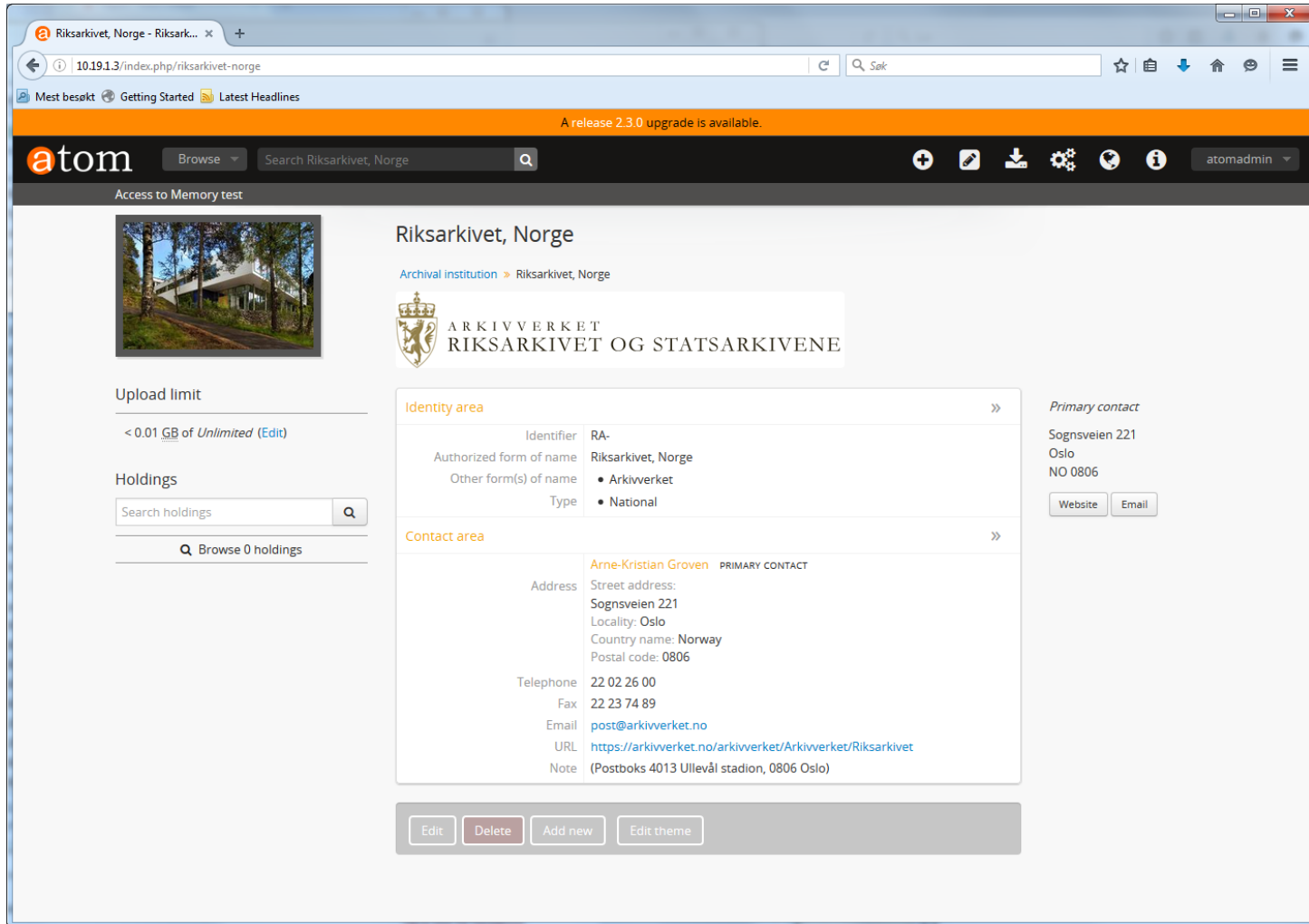
Ingest / test_1_Sofieprisen

- ▼ approveNormalization
 - ▼ test_1_Sofieprisen-4b8121c2-2e32-4323-a029-452001ba1d02
 - ▼ DIP
 - ▼ logs
 - fileFormatIdentification.log
 - ▼ fileMeta
 - ▼ metadata
 - dc.json
 - ▼ submissionDocumentation
 - ▼ objects
 - ▼ 2010
 - ▼ Bilder
 - DSC_0005-3823069f-0fc1-4c53-b041-e55ed6876ecd.tif
 - DSC_0005.jpg
 - DSC_0022-46080114-5071-4f67-b7f5-dbbfcf3e3e8d.tif
 - DSC_0022.jpg
 - DSC_0032-2a347cd3-18dd-4cc7-b2f0-b0fbc5861c14.tif

AtoM (Access to Memory): A few screen shots to illustrate the GUI and the user interaction

- Description of archival institutions, ISDIAH
- Authority records, ISAAR (CPF)
- Photo collection example
- Archival descriptions, ISAD(G)
- High resolution presentation

Description of archival institutions, using the ISDIAH standard



The screenshot shows a web browser window displaying the Riksarkivet, Norge website. The browser's address bar shows the URL `10.19.1.3/index.php/riksarkivet-norge`. The website's header includes the 'atom' logo, a 'Browse' dropdown, a search bar with the text 'Search Riksarkivet, Norge', and a notification for a 'release 2.3.0 upgrade is available.' The main content area is titled 'Riksarkivet, Norge' and includes a photo of a building. Below the photo, there are sections for 'Upload limit' (showing '< 0.01 GB of Unlimited (Edit)') and 'Holdings' (with a search bar and 'Browse 0 holdings' link). The central part of the page displays the institution's name 'Riksarkivet, Norge' and its logo 'ARKIVVERKET RIKSARKIVET OG STATSARKIVENE'. To the right, there are two main sections: 'Identity area' and 'Contact area'. The 'Identity area' contains fields for 'Identifier' (RA-), 'Authorized form of name' (Riksarkivet, Norge), 'Other form(s) of name' (Arkivverket, National), and 'Type' (National). The 'Contact area' contains fields for 'Address' (Sognsveien 221, Oslo, Norway, 0806), 'Telephone' (22 02 26 00), 'Fax' (22 23 74 89), 'Email' (post@arkivverket.no), 'URL' (https://arkivverket.no/arkivverket/Arkivverket/Riksarkivet), and 'Note' (Postboks 4013 Ullevål stadion, 0806 Oslo). A 'Primary contact' section on the right lists 'Sognsveien 221 Oslo NO 0806' and includes 'Website' and 'Email' buttons. At the bottom, there are buttons for 'Edit', 'Delete', 'Add new', and 'Edit theme'.

Riksarkivet, Norge

Archival institution > Riksarkivet, Norge

ARKIVVERKET
RIKSARKIVET OG STATSARKIVENE

Upload limit

< 0.01 GB of Unlimited (Edit)

Holdings

Search holdings

Browse 0 holdings

Identity area

Identifier	RA-
Authorized form of name	Riksarkivet, Norge
Other form(s) of name	• Arkivverket
Type	• National

Contact area

Address	Arne-Kristian Groven PRIMARY CONTACT
Street address:	Sognsveien 221
Locality:	Oslo
Country name:	Norway
Postal code:	0806
Telephone	22 02 26 00
Fax	22 23 74 89
Email	post@arkivverket.no
URL	https://arkivverket.no/arkivverket/Arkivverket/Riksarkivet
Note	(Postboks 4013 Ullevål stadion, 0806 Oslo)

Primary contact

Sognsveien 221
Oslo
NO 0806

Website Email

Edit Delete Add new Edit theme

Authority records, describing the producer, ISAAR(CPF) standard

Union Co

[Authority record](#) » Union Co


Identity area

Type of entity	Corporate body
Authorized form of name	Union Co
Parallel form(s) of name	• Norske Skog Union

Export

 [EAC](#)

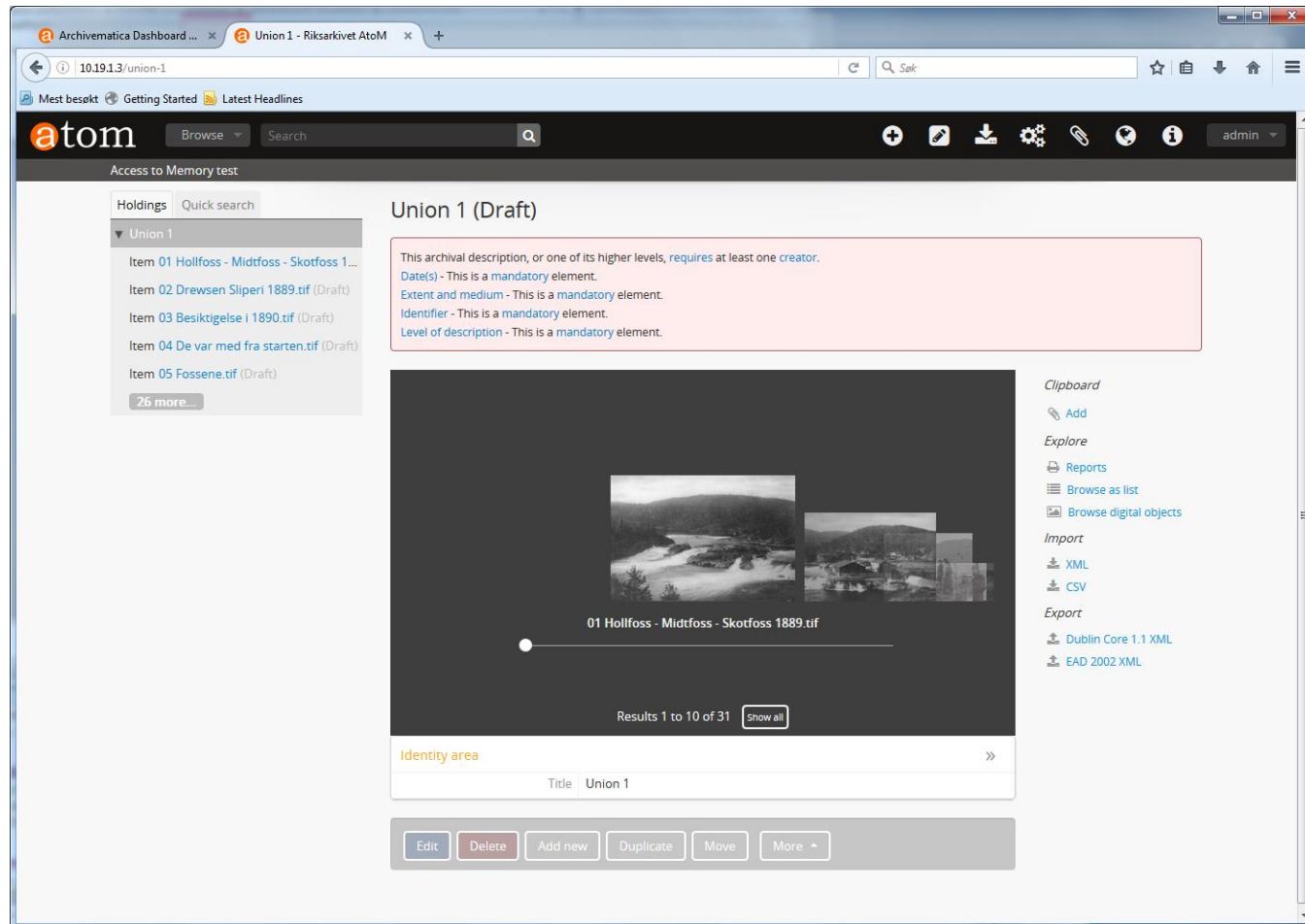
Description area

Dates of existence	1803-2006
History	"I 1873 kjøpte Benjamin Sewell og Halvor Emil Heyerdahl Skiens Træsliberi, og stiftet selskapet Union Co. De hadde planer om en samlet tremasse- og papirfabrikk på Smedøya i Skien, under navnet Union Bruk. For å realisere Union Bruk fikk de med seg en ... 

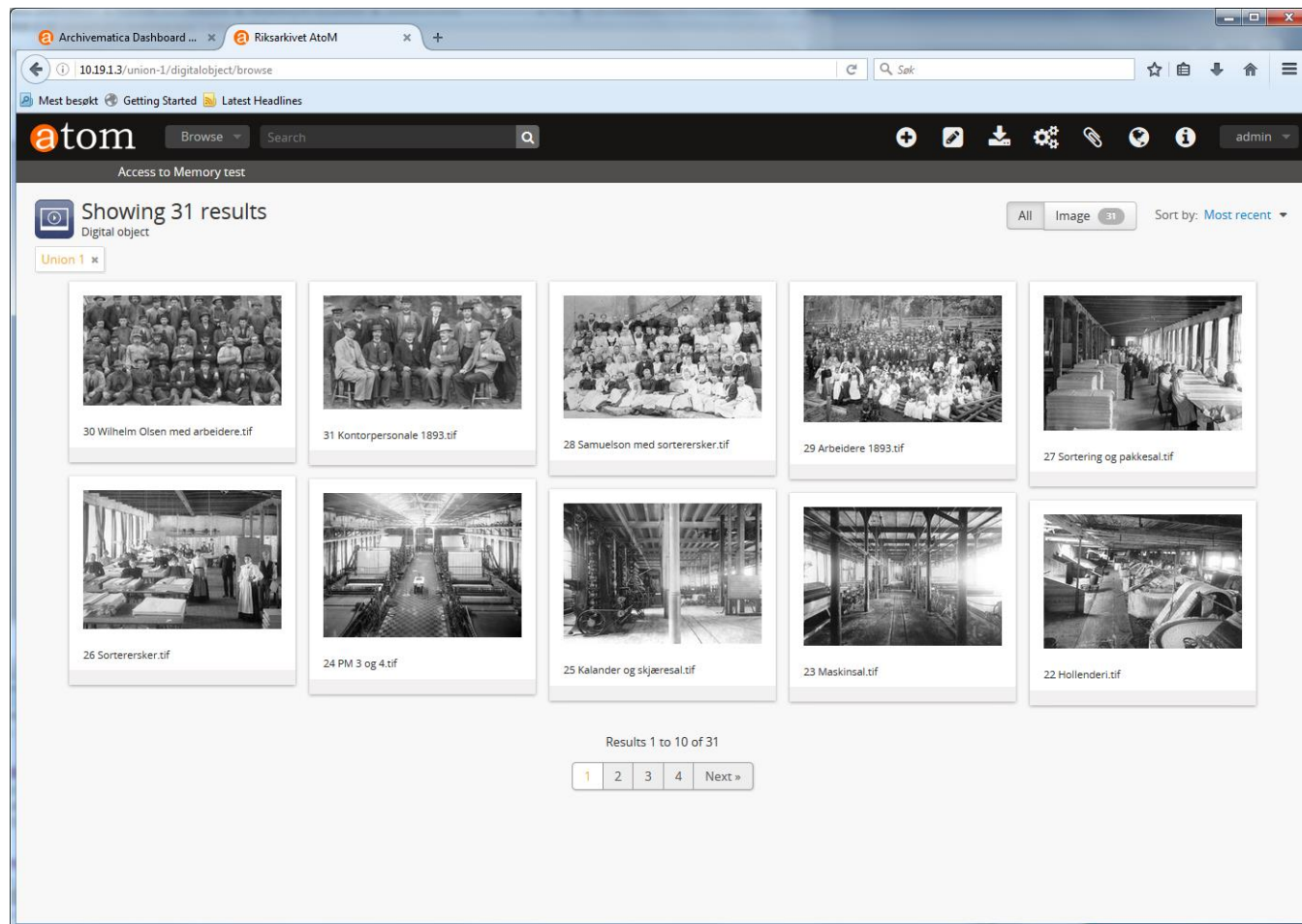
Control area

Authority record identifier	RA/PA-1422
-----------------------------	------------

Overview of the content



Thumbnails, (this example is a pure photo collection)



Digital object and associated meta data, in ISAD(G)


Archivematica Dashboard ... x 19 Fabrikkområde.tif - Riks... x

10.19.1.3/19-fabrikkomrade.tif

Mest besøkt Getting Started Latest Headlines

- Item 17 Hengebroen.tif (Draft)
- Item 18 Hølen.tif (Draft)
- Item 19 Fabrikkområde.tif
- Item 20 Generelen.tif (Draft)
- Item 21 Sliperi.tif (Draft)
- Item 22 Hollenderi.tif (Draft)
- Item 23 Maskinsal.tif (Draft)
- 8 more...

Union 1 19 Fabrikkområde.tif



Clipboard

- Add

Explore

- Reports
- Browse as list
- Browse digital objects

Import

- XML
- CSV

Export

- Dublin Core 1.1 XML
- EAD 2002 XML

Identity area

Title	19 Fabrikkområde.tif
Level of description	Item

Digital object metadata

Filename	a5fa5d61-e0a5-4438-9dd7-c6d84ee4e88d-19_Fabrikkomrade.jpg
Media type	Image
Mime-type	image/jpeg
Filesize	344.3 KiB
Uploaded	November 7, 2016 6:07 AM
Object UUID	a5fa5d61-e0a5-4438-9dd7-c6d84ee4e88d
AIP UUID	e262a482-5160-4cfd-82ed-d28b5b3a7429

Edit Delete Add new Duplicate Move More

Editing in ISAD(G)

Untitled

Identity area
Context area
Content and structure area
Conditions of access and use area
Allied materials area
Notes area
Access points
Description control area
Rights area
Administration area

The data object, here a high resolution picture



Our little story regarding installation

- Both tools runs (only) on Ubuntu Linux, so knowledge on Linux in general an Ubuntu specifically is needed
- We managed to intall Archivematica and we managed to install AtoM
- But we didn't manage to make them work together, due to our lack of knowledge and as an exception, lack of good documentation
- But a couple of hours support, from support personell in Spain, fixed the whole thing

Experiences with using Archivematica and AtoM

- Both are complex web applications, but with an easy to use graphical interface (GUI)
- Both tools are, generally, well documented
- It is easy to communicate with the vendor who is located in Vancouver, Canada, having some developers/technicians in Europe
- User forums are active, and information about future releases are easy to access
- Support from vendor went smooth

Archivematica

- Archivematica was relatively easy to use, with a graphical user interface, GUI, following the OAIS standard
- The level of automation can be configured
- No requirements needed regarding file formats to be ingested
 - Files formats are identified automatically
 - Rule-based file format conversion

Most importantly, Archivematica is securing

- Effectiveness, that (a complete set of) the right Ingest actions are performed.
- Efficiency, a dramatic reduction in time consumption

Access to Memory (AtoM), archivist perspective

- Work bench for archivists, to add and modify archival descriptions
- Archival descriptions according to international standards: ISDIAH, ISAD(G), ISAAR-CPF, ISF, DC
- Local standards for archival descriptions can be added and mapped (once) against the available international standards,
 - hence enabling switching between local and international standards
- Role based access control

Access to Memory (AtoM), information consumer perspective

- Modern search engine, Elasticsearch, make searches easy
- Digital information can be accessed through the archival descriptions: ISDIAH, ISAD(G), ISAAR-CPF, ISF, DC, or other standards
- Fine-grained access control, role based
- Films, audio, video, documents, and other file formats can be presented together with the archival descriptions in the same GUI

A few words about software risks...

- In general, (negative) risks associated with software acquisitions, are high...
- If the software has few users
- If few people have deep knowledge about and are able to maintain the software

Software risks continued

- Additionally, for long-term preservation and access software, where data need to undergo many format transformations over time
 - Proprietary software will make quality assurance difficult due to lack of transparency, while
 - Free and open source software will reduce these negative risks, through transparency (open code/open community)
- This is the main reason for many leading archival institutions worldwide to use free and open source software as a principle

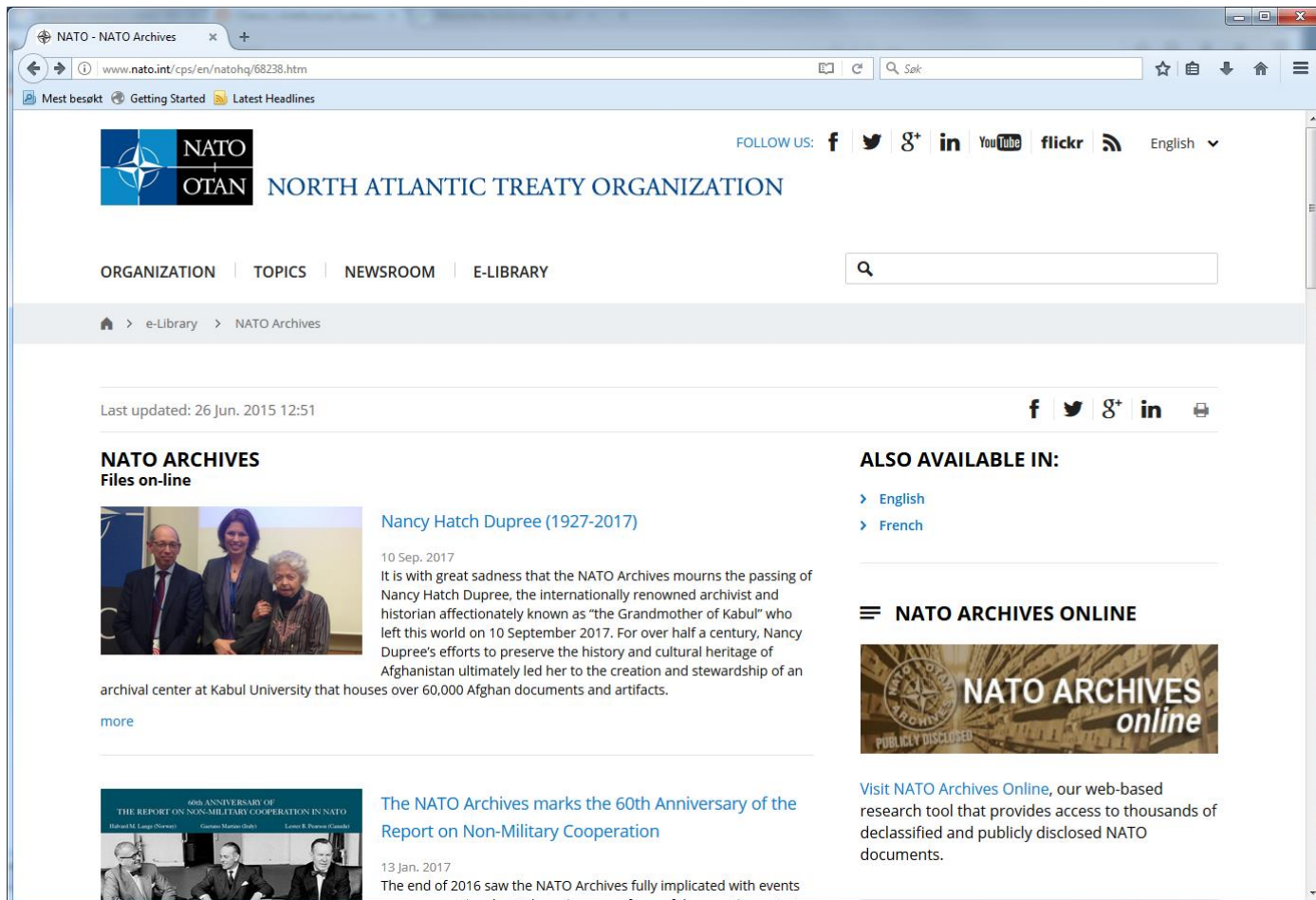
For Archivematica and AtoM the status is as follows, as of 13.10 2017

- 87 archival institutions are using the technology, one or both, and are cooperating with/getting support from the vendor
- 212 themes have been discussed on the user forum since June
- 43 code forks, but most are associated with developers working for/with the vendor, merging the code back again
- All of these are sign of a living product with an active user base, with a certain probability for the tools to survive in the near future

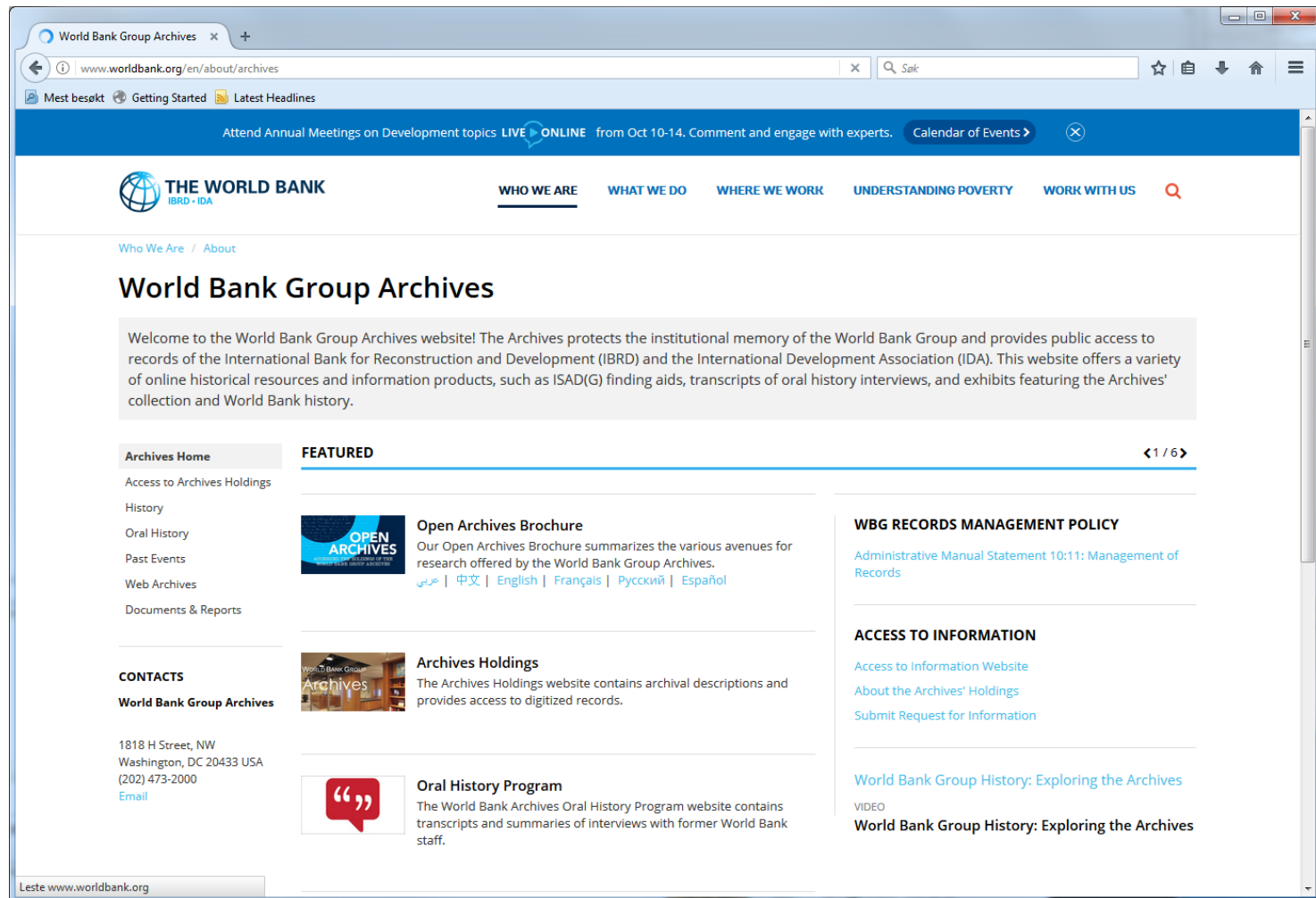
Closed solutions lack transparency: An example from the digital preservation world

- Here is a conversation I found on the internet:
 - «Just a matter of interest is there a ... schema?»
 - «There is documentation customers have. Not sure if published?»
 - «Yeah the documentation isn't published, but this would be a good question to ask on the customer portal. I can ask.»
 - «Yes that would be very helpful if you could ask on the customer portal!»
 - «I'm not sure if ... had a chance to respond yet?»...

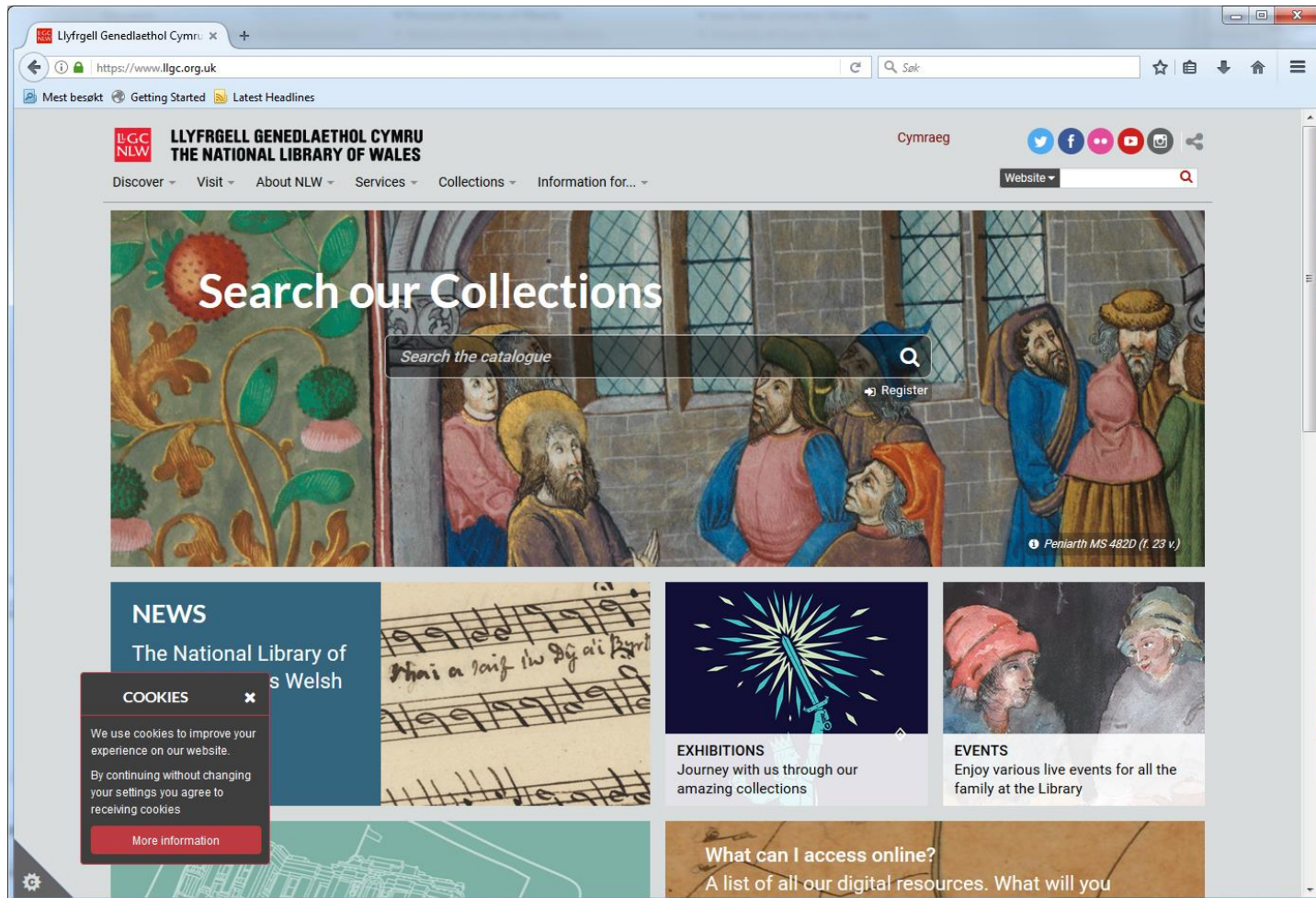
Examples of Archives using Arcivematica and AtoM...



Examples of Archives using Arcivematica and AtoM...



Examples of Archives using Arcivematica and AtoM



To conclude...

- The first impression of the tools are good, even though we still miss this and that
- The tools are also parts of a more holistic (OAIS) toolset, integrated with various 3.party software:
 - Arkivum, ArchivesSpace, Archivist' Toolkit, OpenStack, DuraCloud, LOCKSS, Islandora, CONTENTdm, sSpace
- I believe the tools can survive, at least over the next 10-20 years, due to:
 - Open source code, active communities, ok documentation, and a software vendor in charge of the development

To conclude

- We are currently benchmarking the tools with locally developed technologies we are using, trying to identify what is needed to co-exist
- Scalability should be looked more into. We are not currently able to utilize all the potential lying in all the configuration options
- Visualization plugins/modifications in AtoM more adapted to our content, should be looked more into