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 loooking a bit ahead

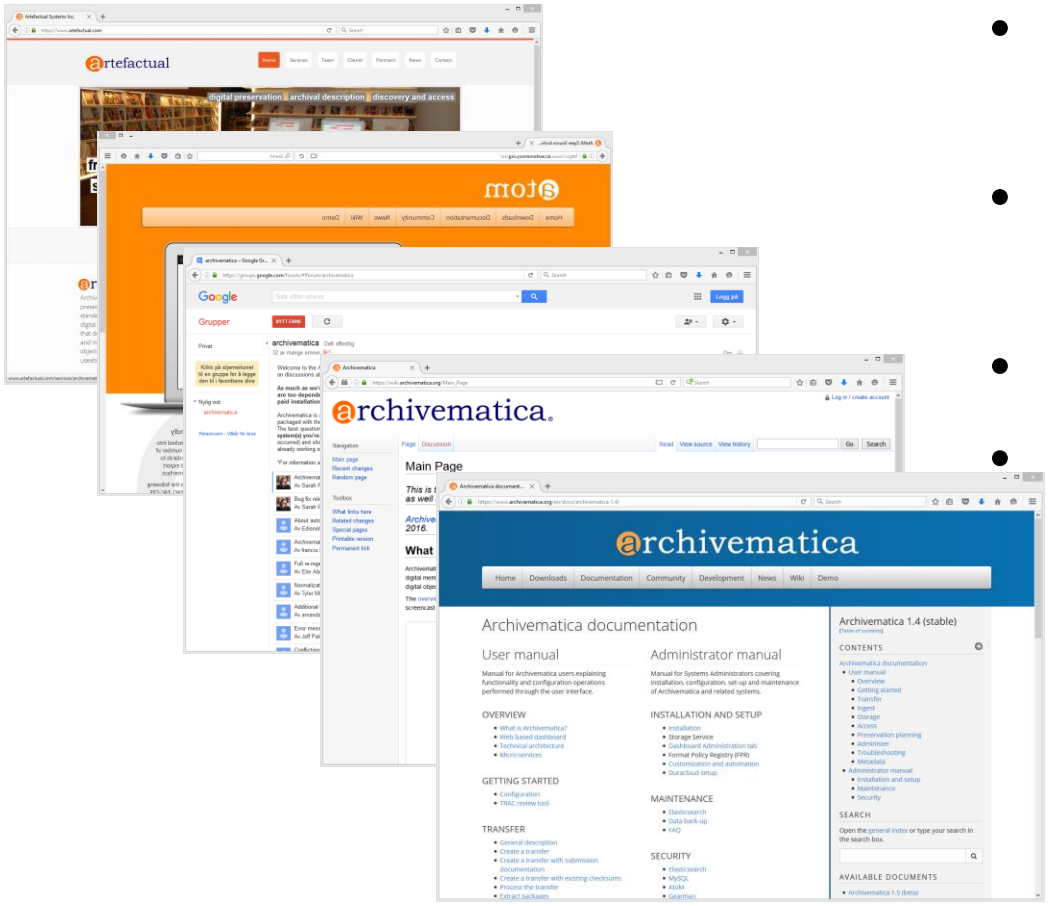
But first… A few words about the work context behind this presentation

- 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

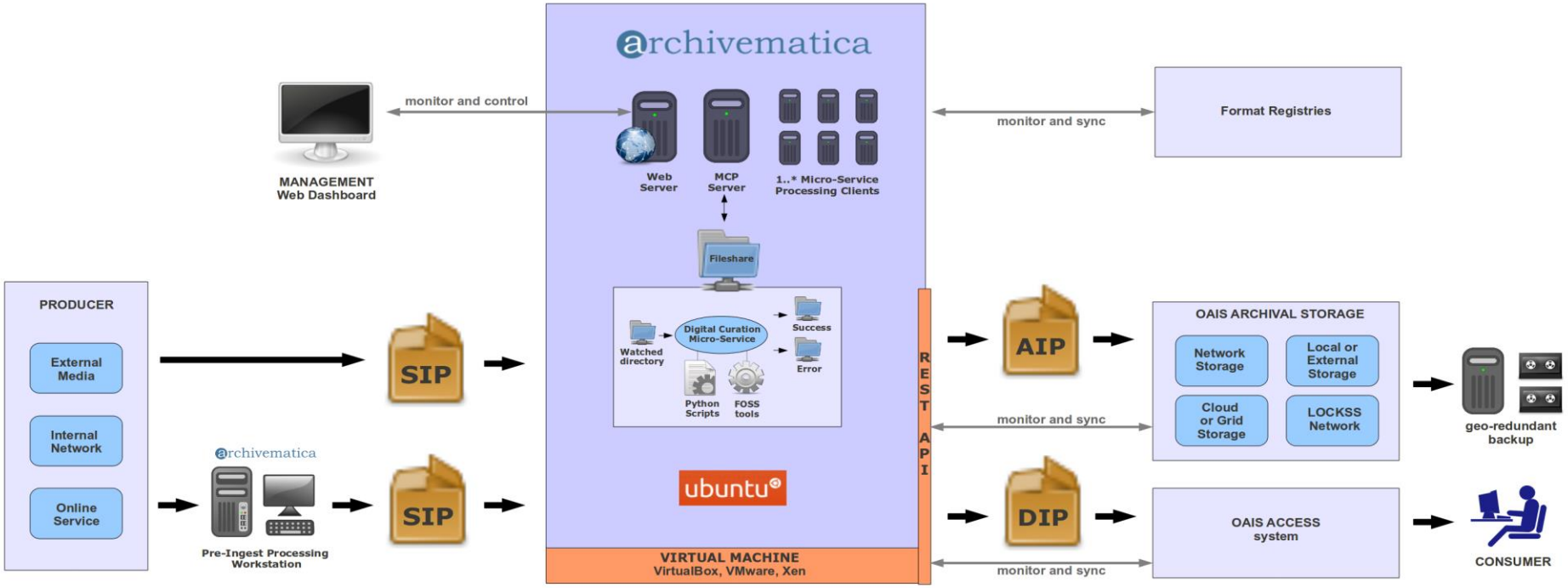
- Archivematica is a free and open-source digital preservation system.
- Designed to maintain standards-based, long-term access to collections of digital objects
- Archivematica is packaged with the web-based content management system AtoM for access to your digital objects
- The vendor is Artefactual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

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



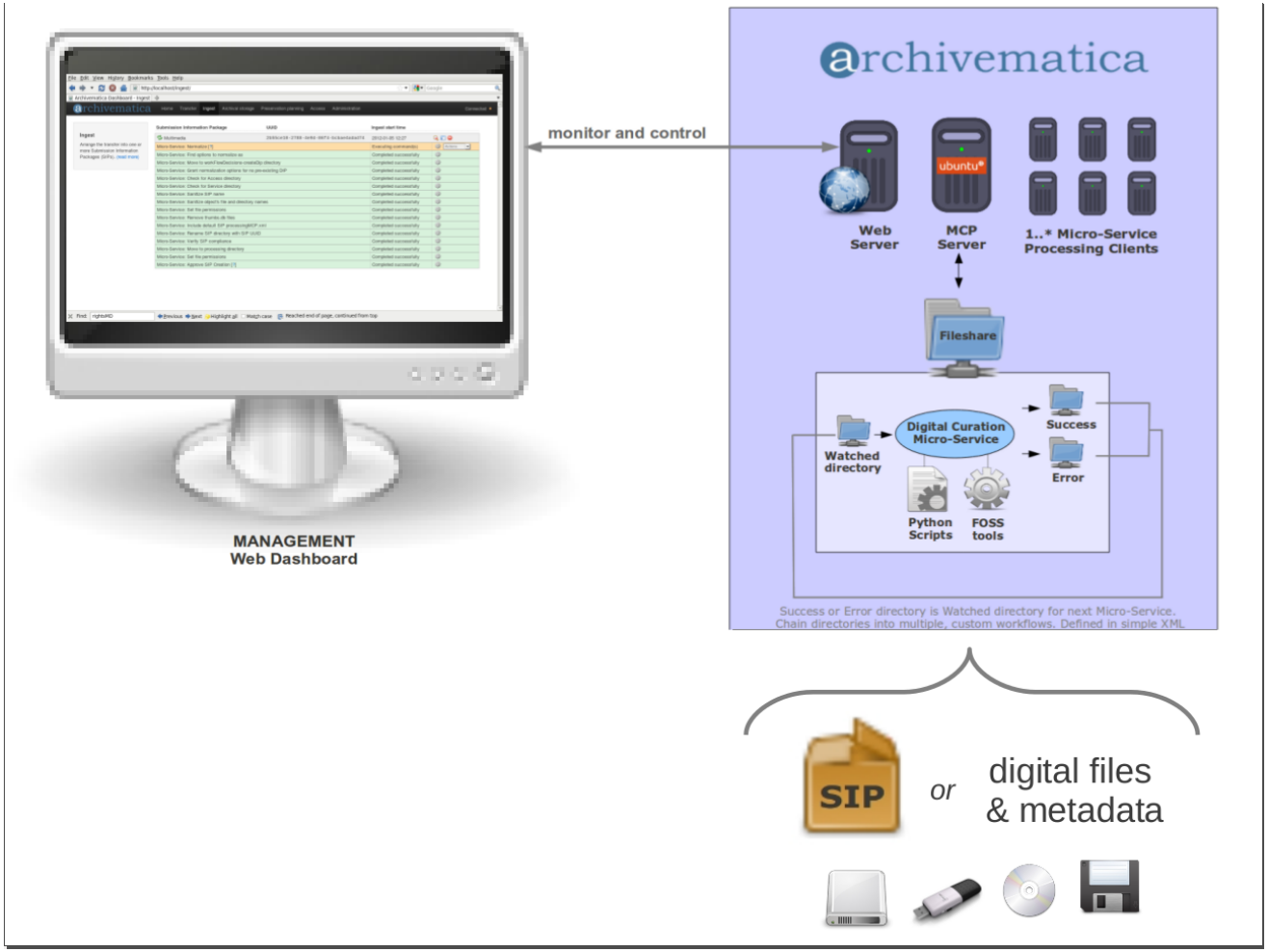
- https://www.artefactual.com/
- https://www.accesstomemory.org
- https://groups.google.com/forum/#!forum/archivematica
- https://wiki.archivematica.org
- https://www.archivematica.org

The Archivematica architecture



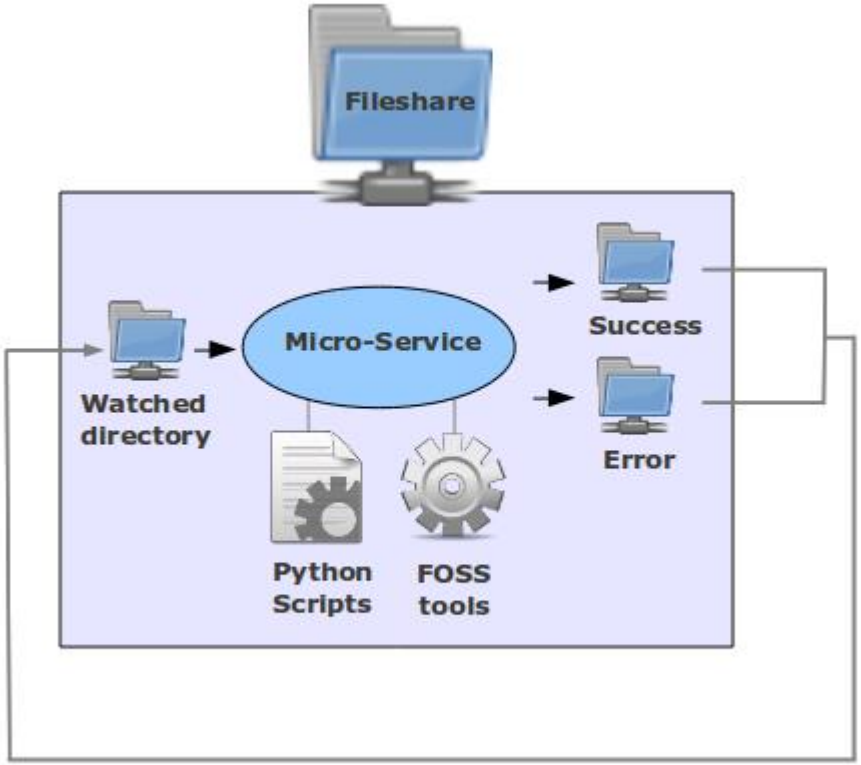
Kilde: https://www.archivematica.org/wiki/File:Archivematica-0.8-beta-architecture.png

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

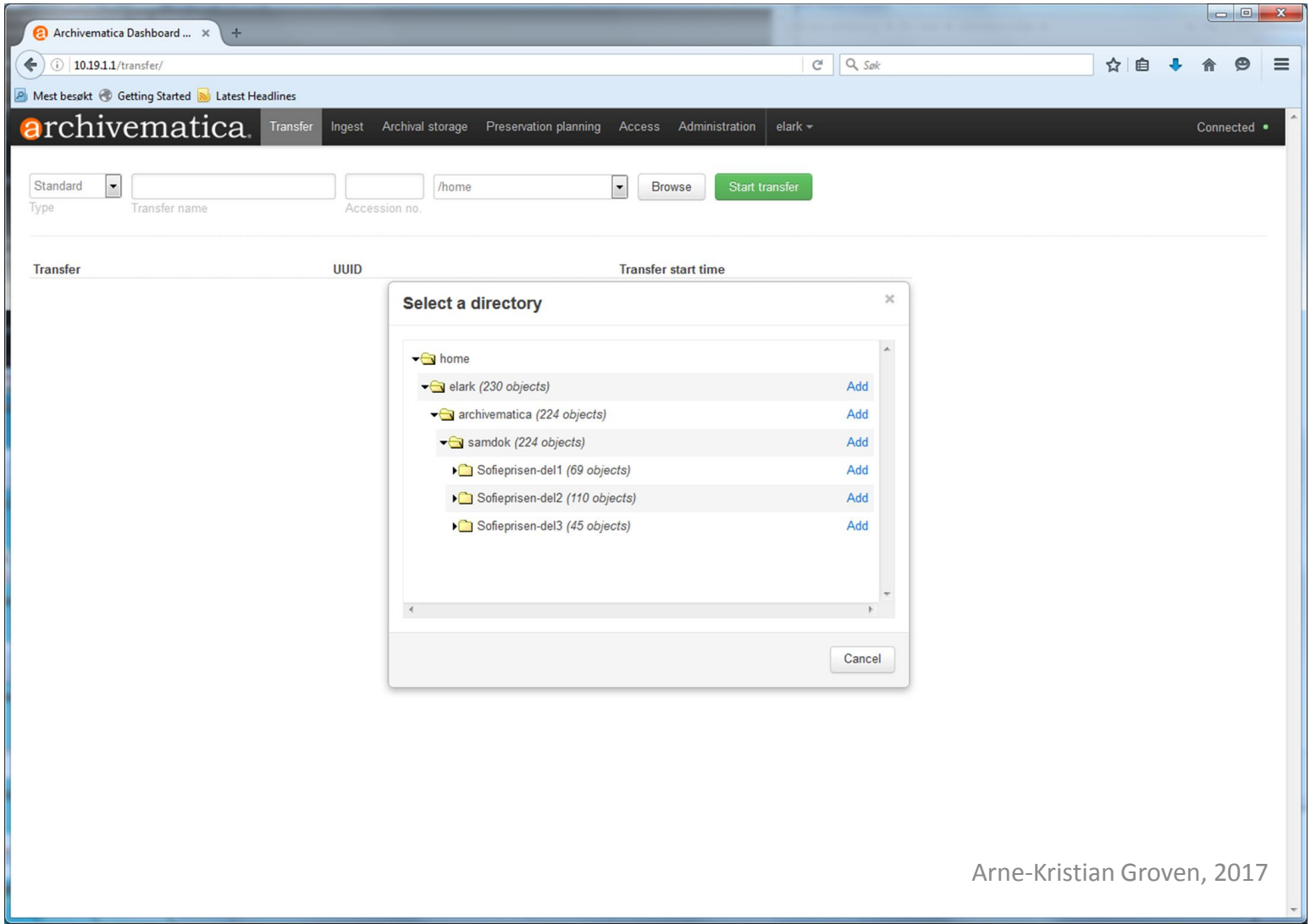


Kilde: https://www.archivematica.org/wiki/File:Archivematica-0.8-beta-architecture.png

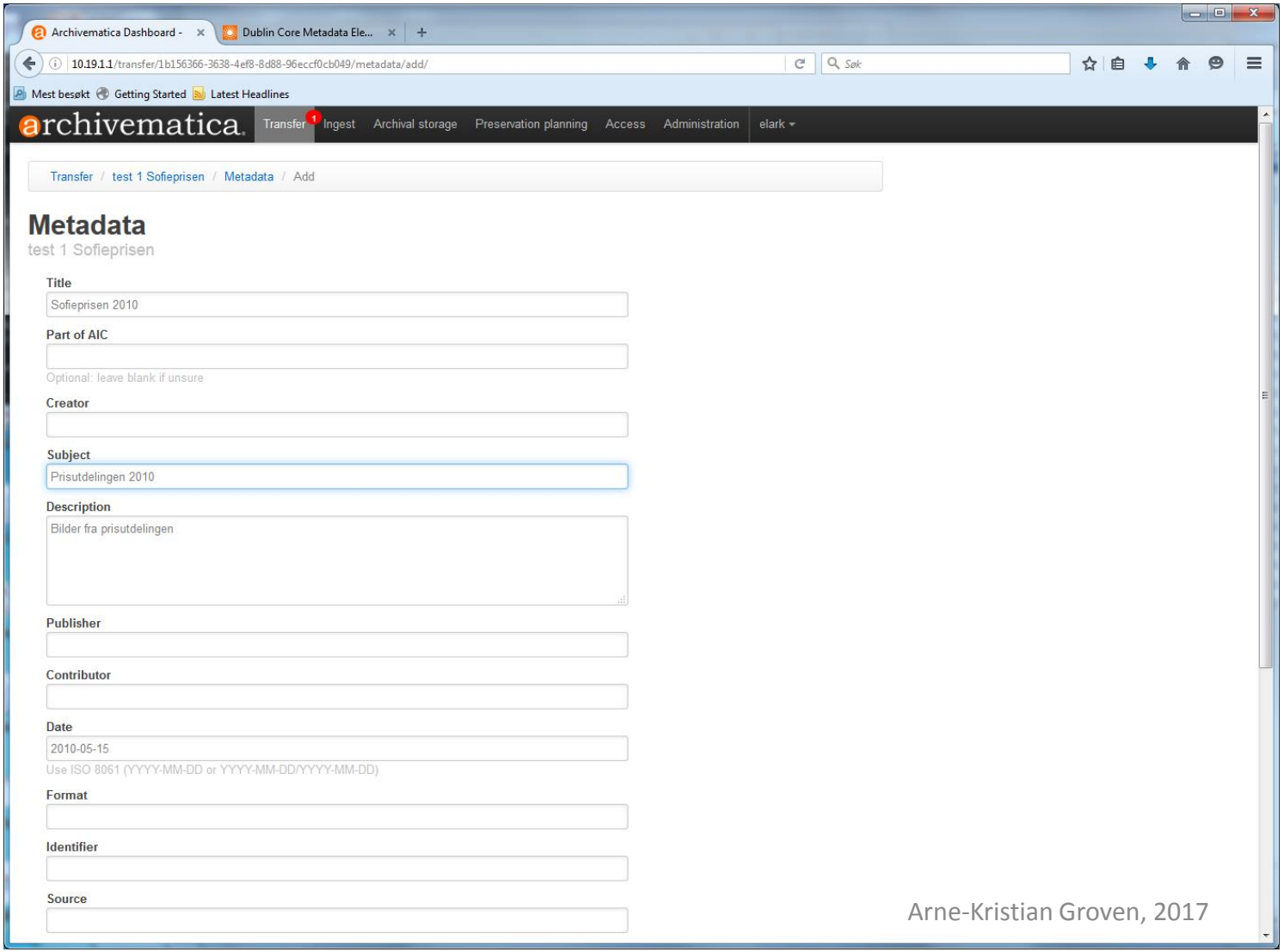
Archivematica: A few screenshots to illustrate the GUI and the user interaction

- 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

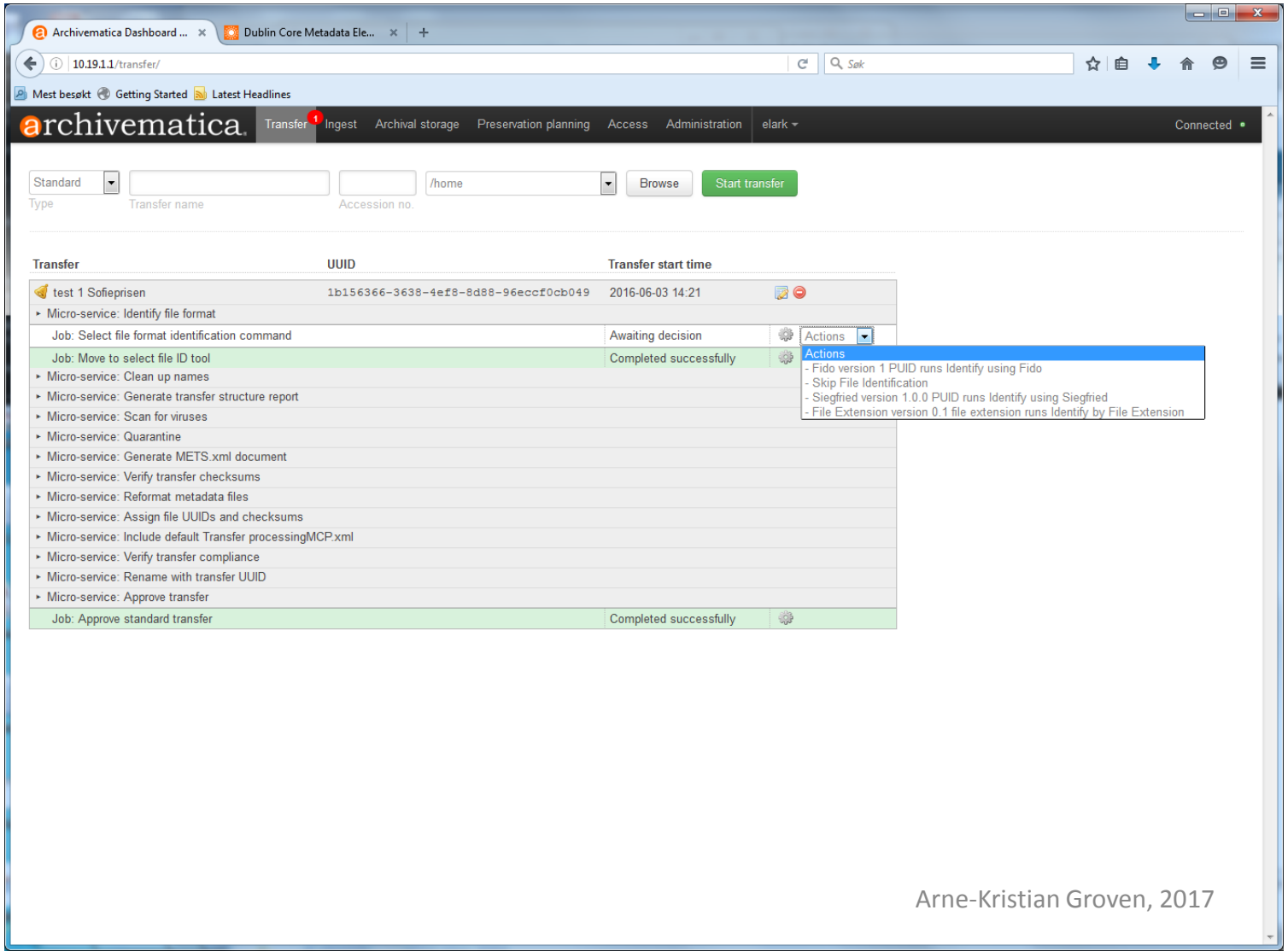


Dublin Core (DC) editing



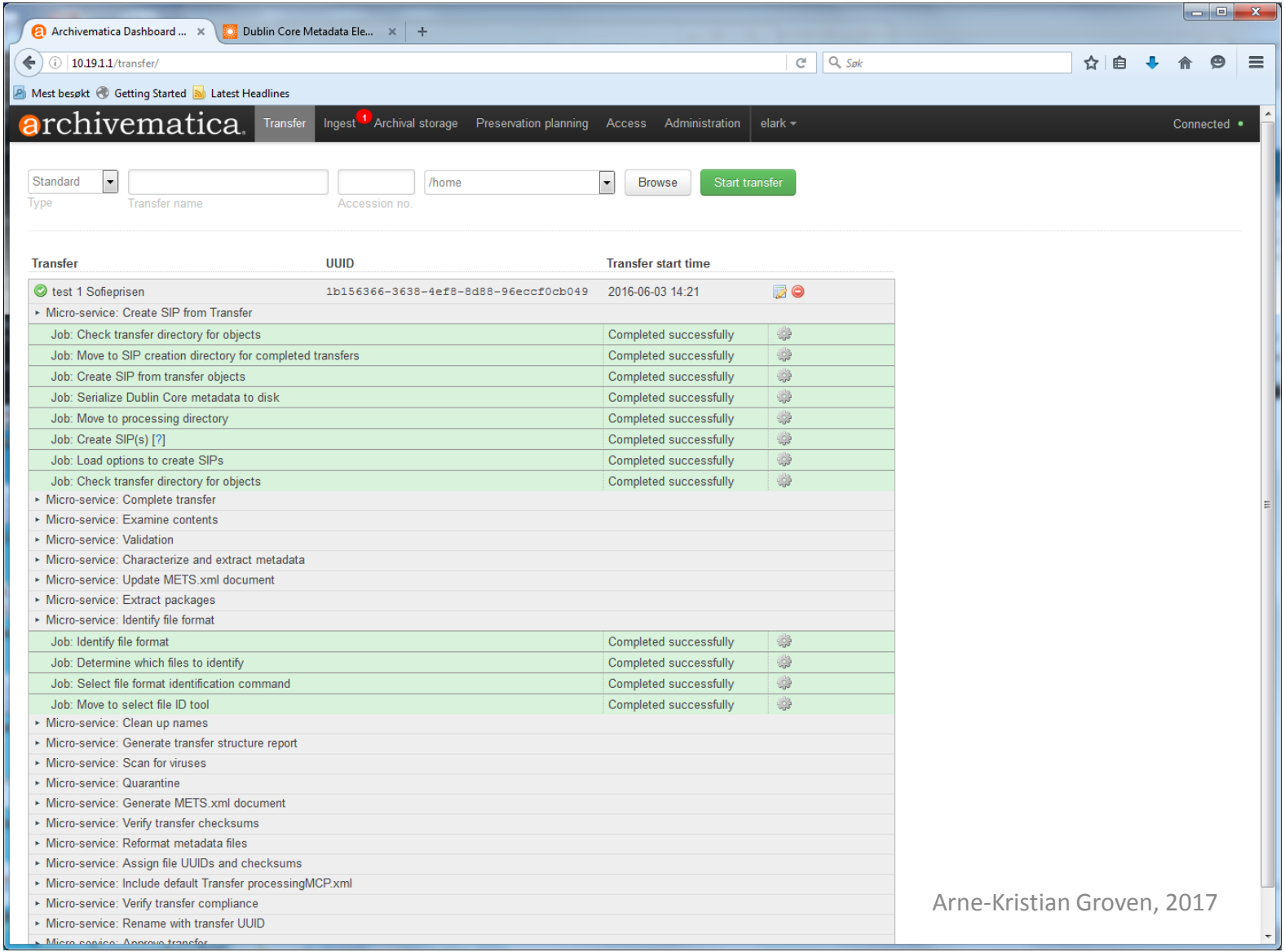
Arne-Kristian Groven, 2017

File format identification, either selecting tool options interactively or preconfigured



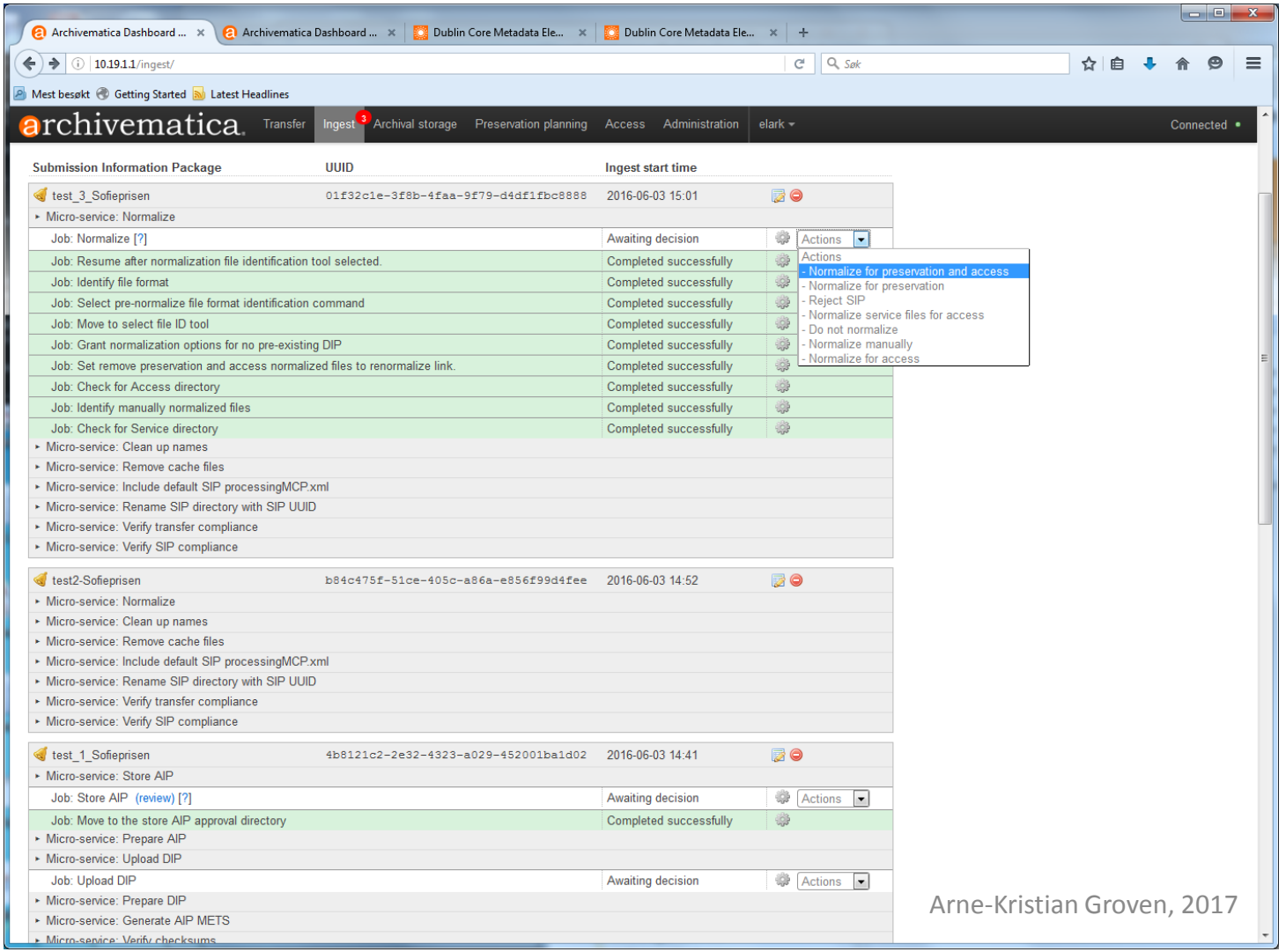
Arne-Kristian Groven, 2017

Generating a SIP



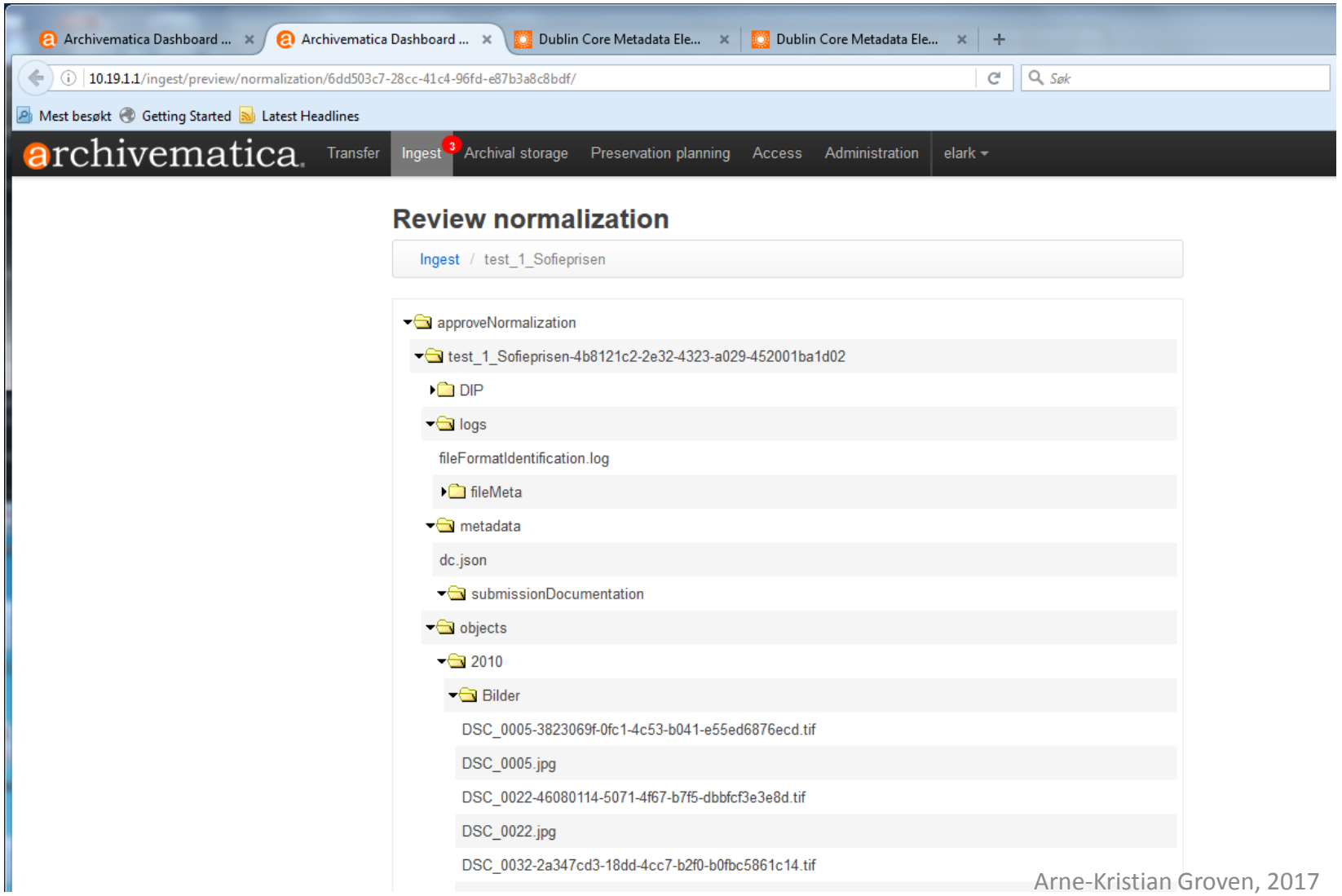
Arne-Kristian Groven, 2017

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



Arne-Kristian Groven, 2017

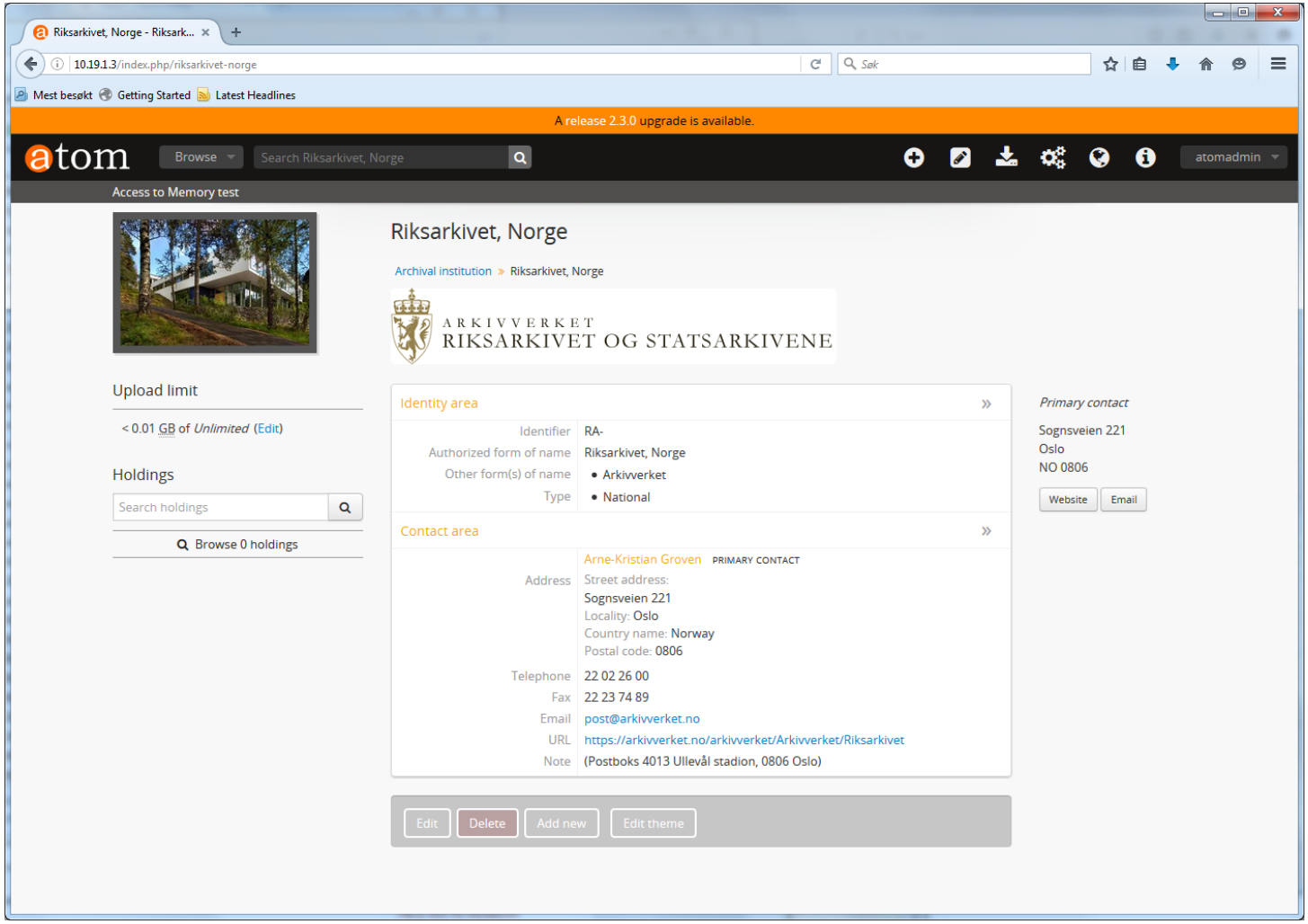
Resulting structure, ready for review



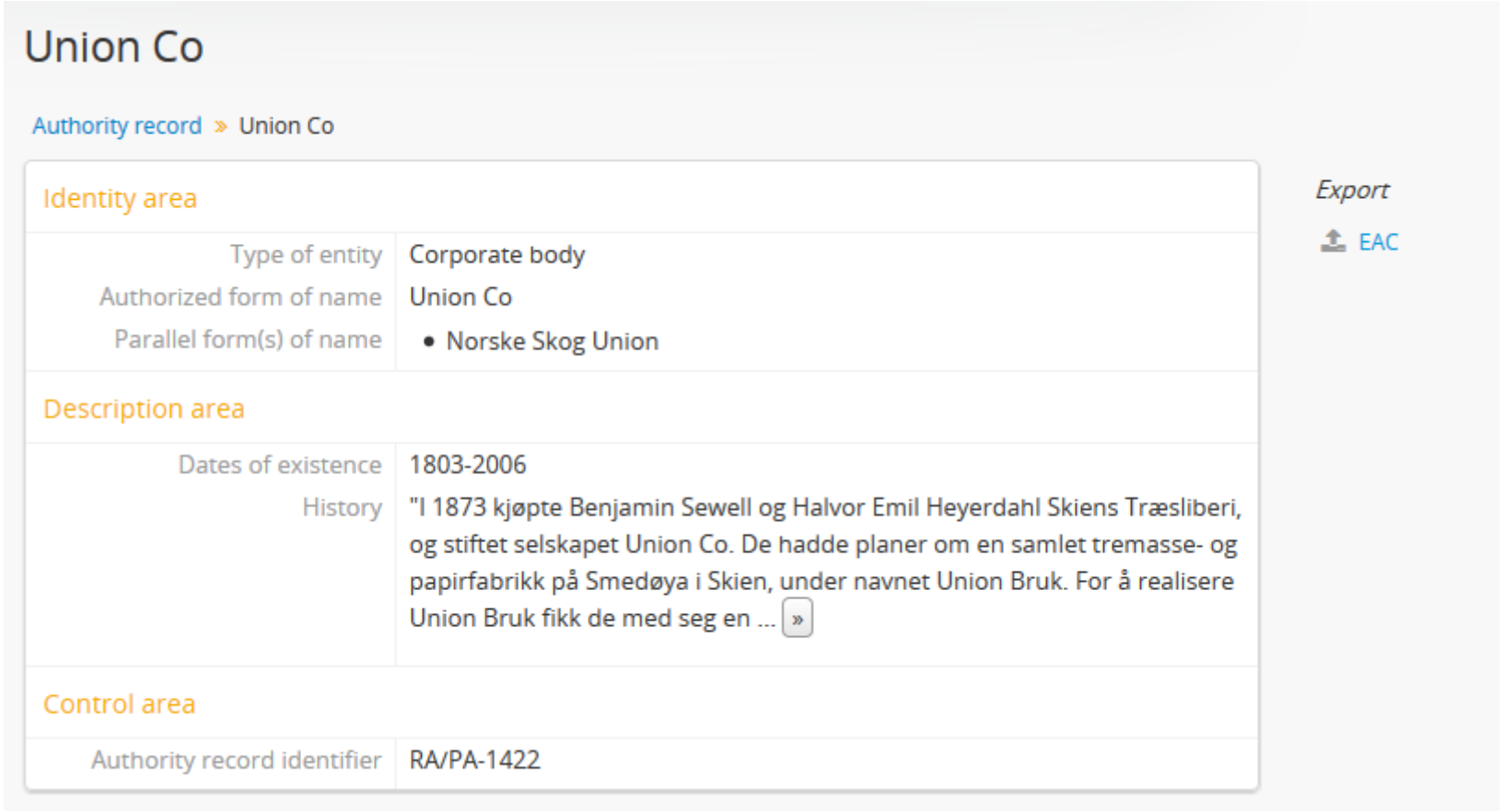
AtoM (Access to Memory): A few screen shots to illustrate the GUI and the user interaction

- 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



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

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 ... »

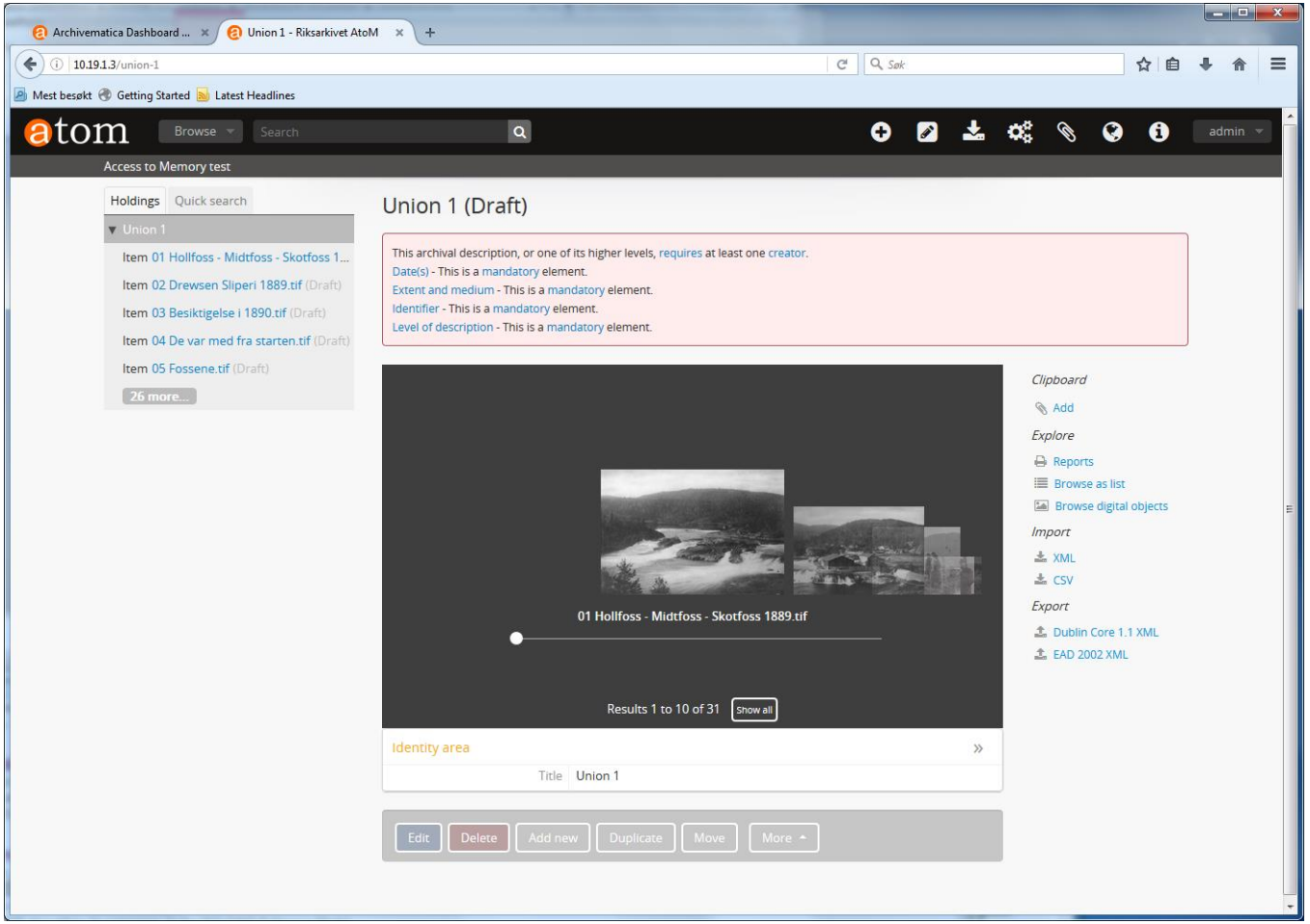
Control area

Authority record identifier	RA/PA-1422

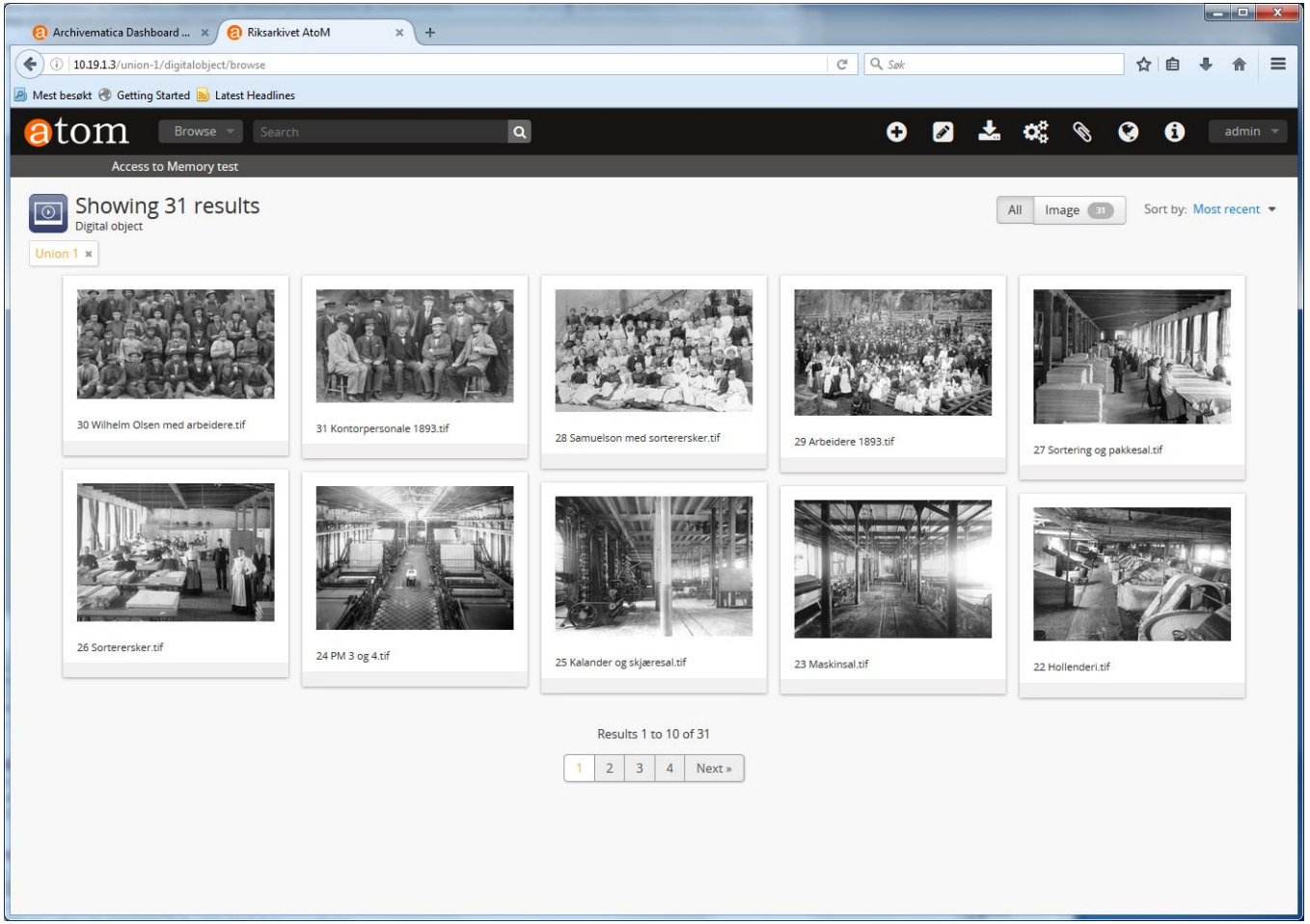
Export

EAC

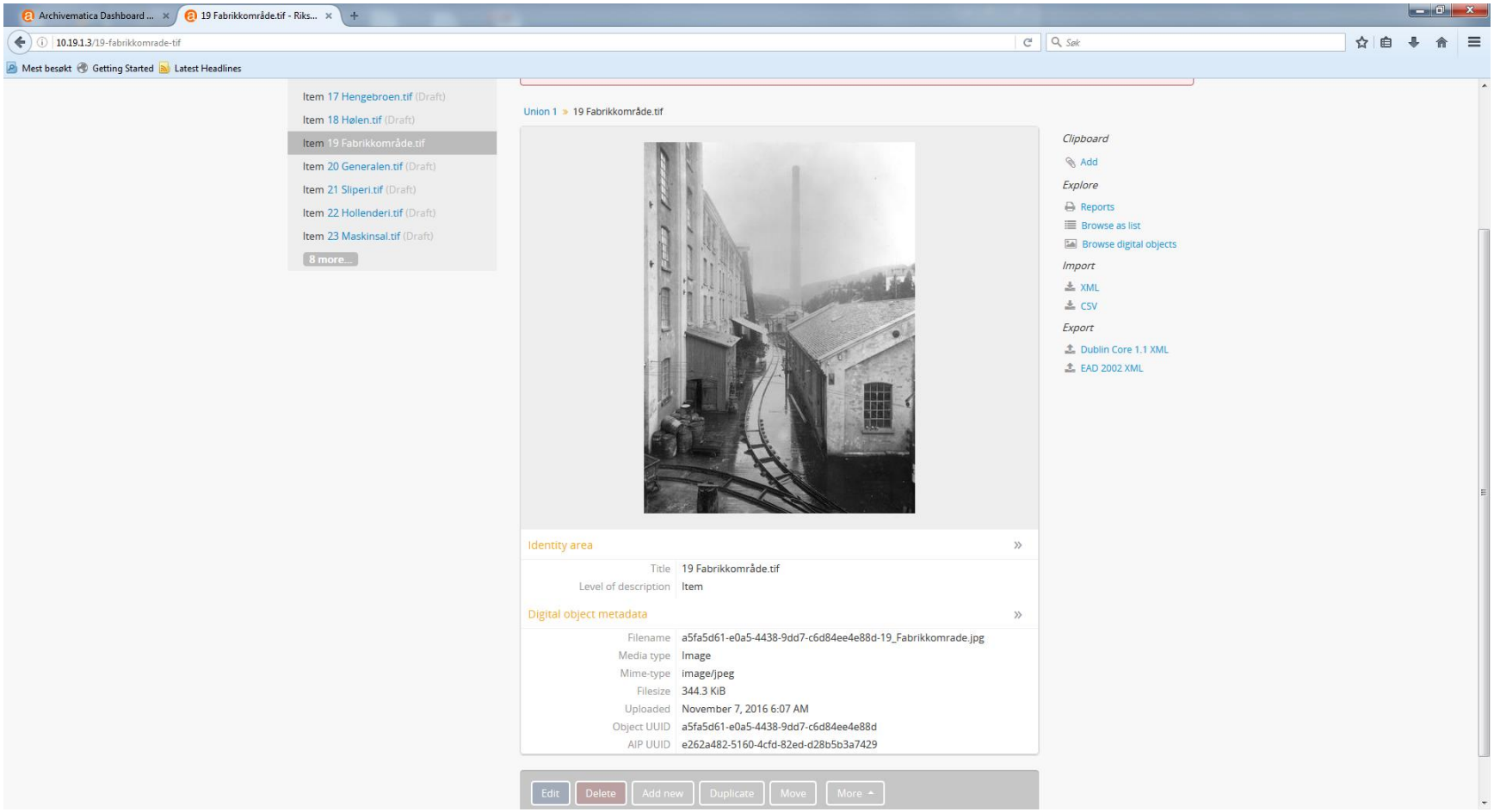
Overview of the content



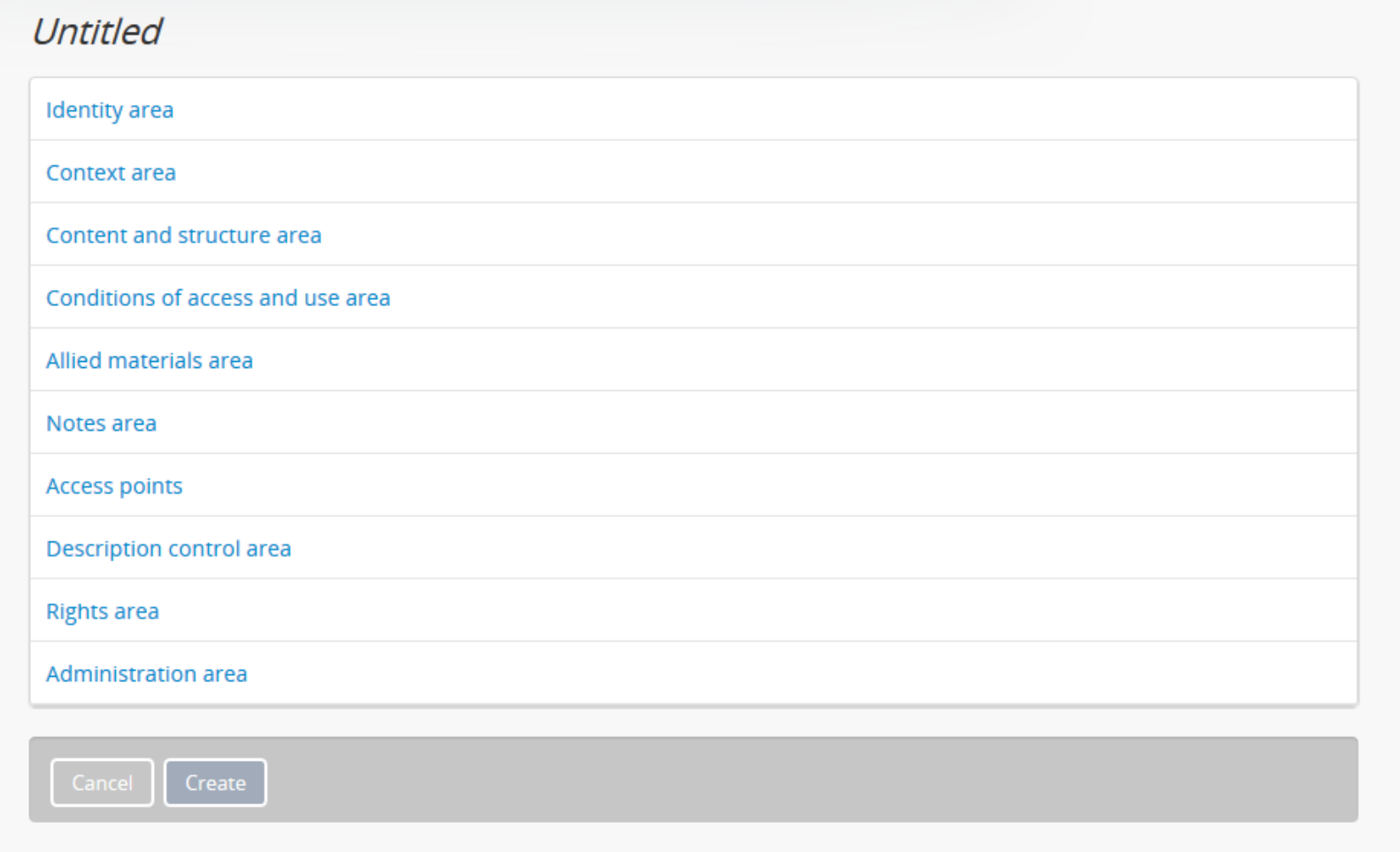
Thumbnails,
(this example is a pure photo collection)



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



Editing in ISAD(G)



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 consument 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) risiks associated with software aqusitions, 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

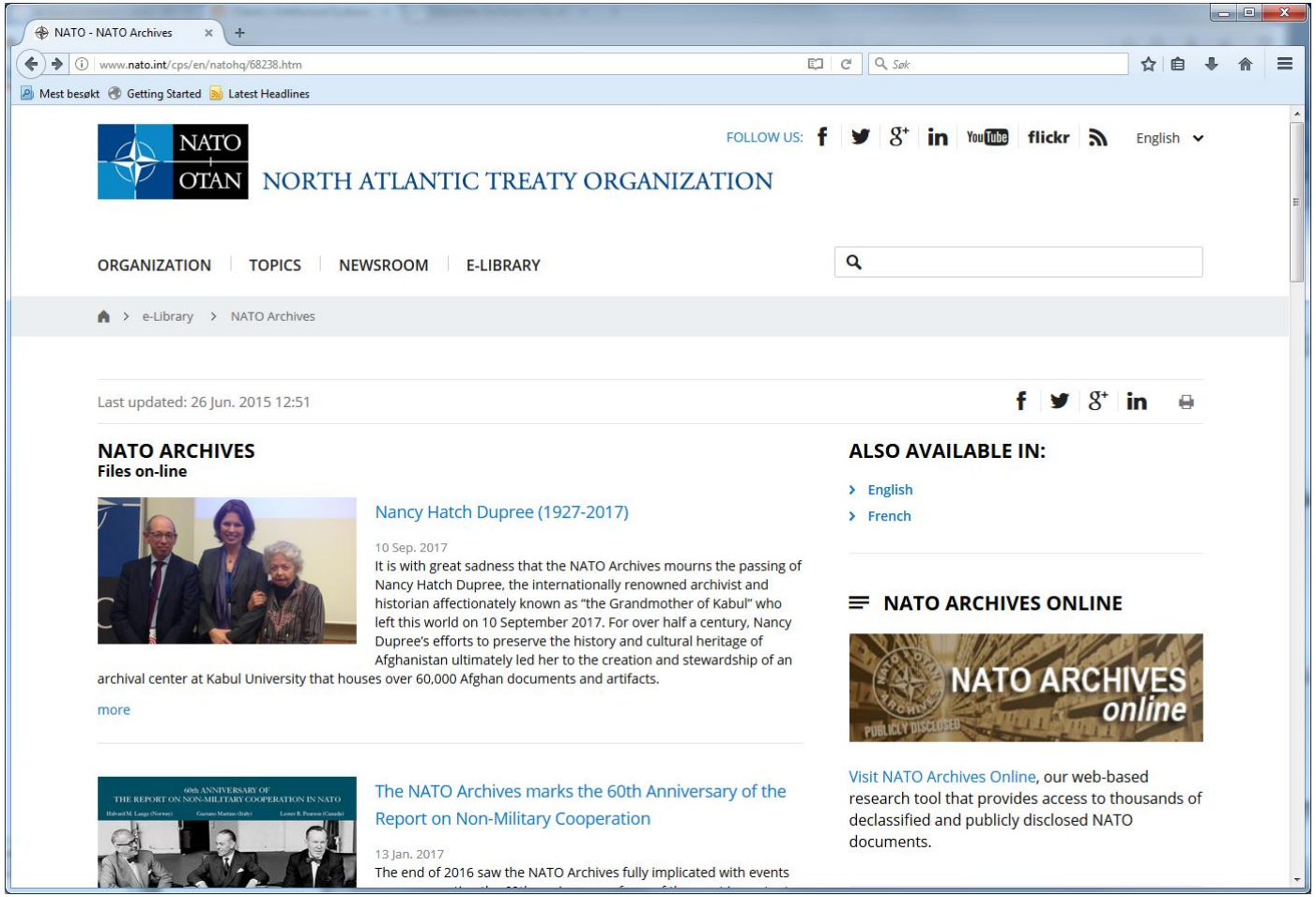
For Archivematica and AtoM the status is as follows, as of 13.10 2017

- 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

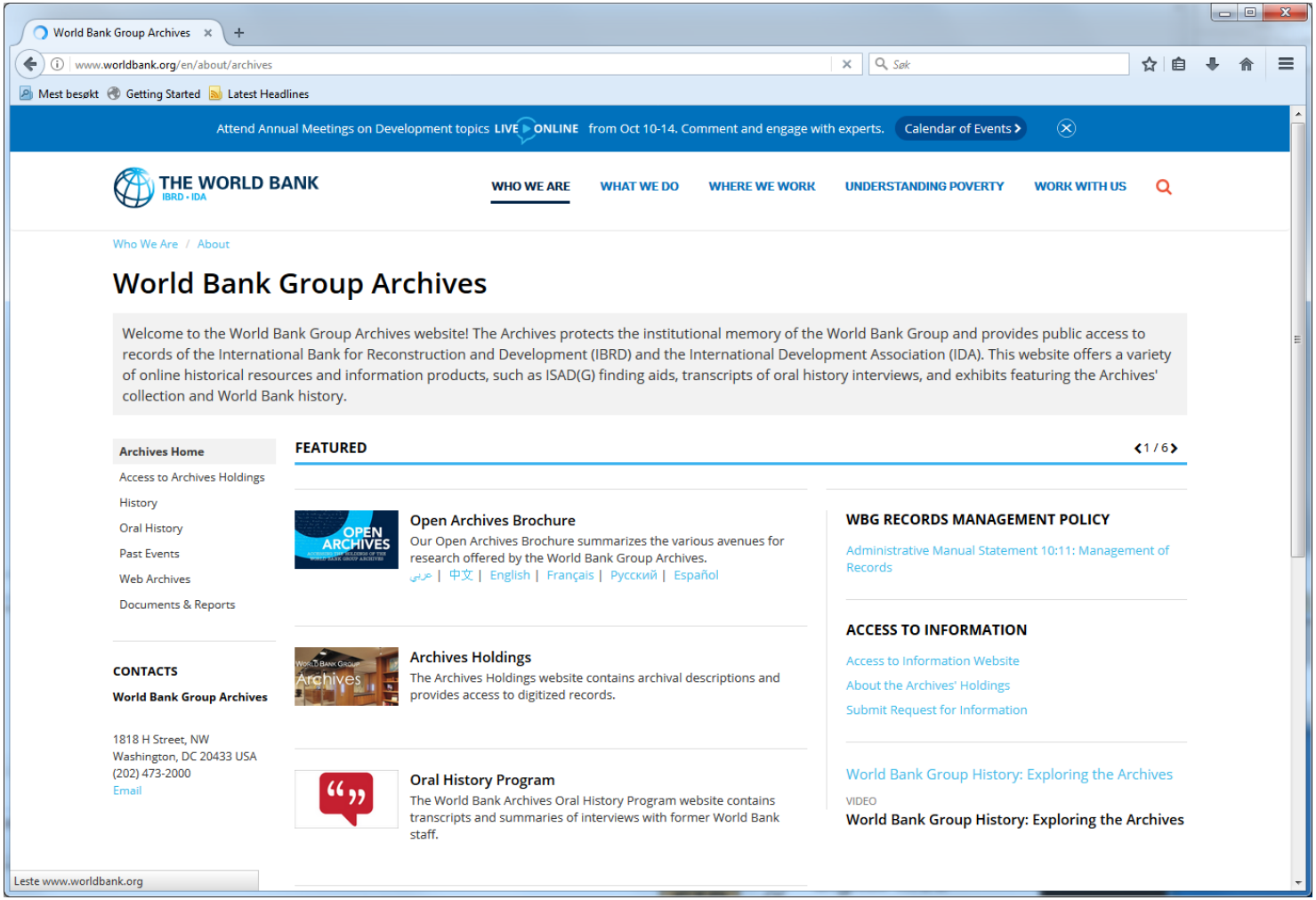
Closed solutions lack transparency: An example from the digital preservation world

- 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 wold be very helpful if you could ask on the customer portal!»
  - «I'm not sure if … had a chance to respond yet?»…

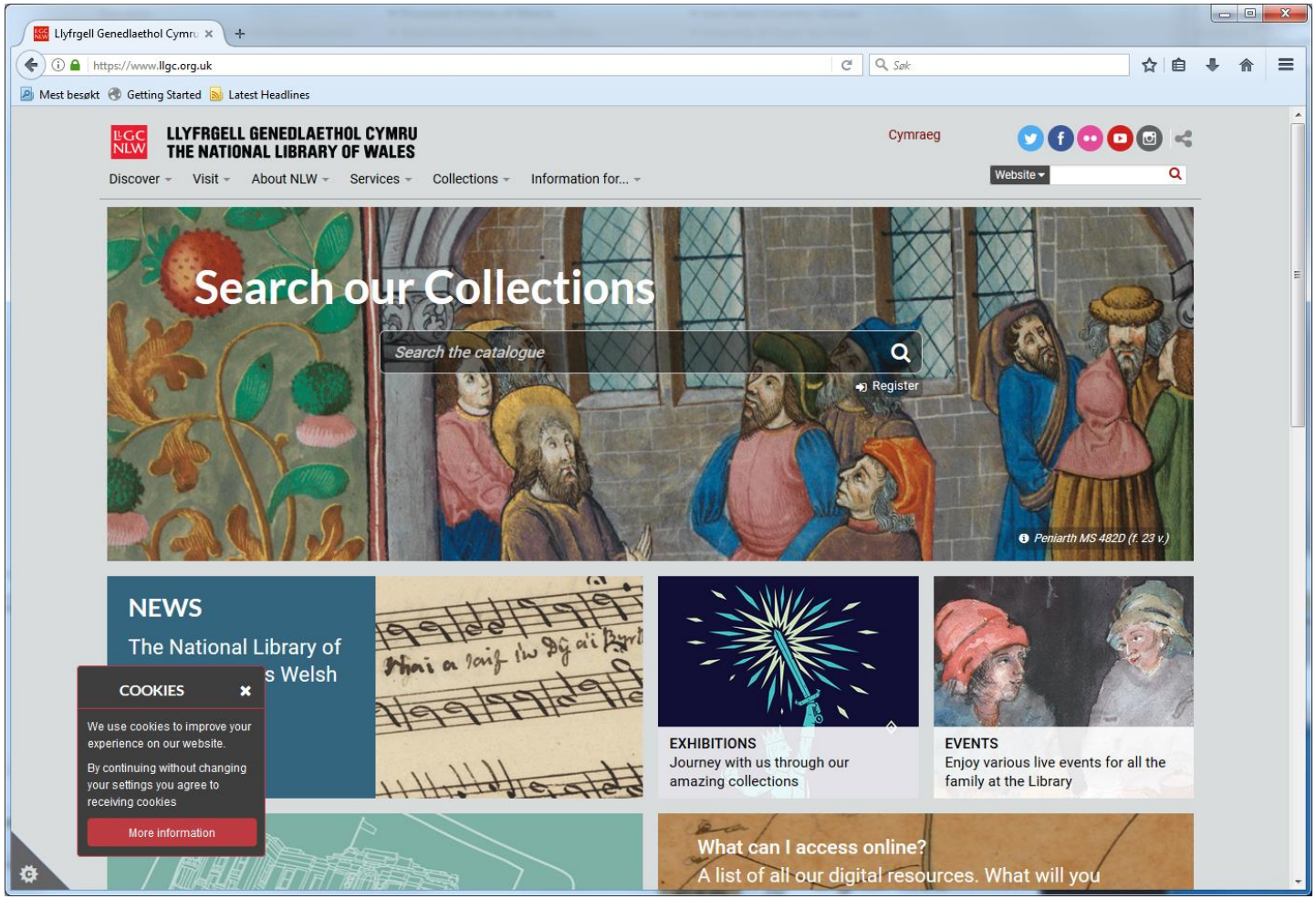
Examples of Archives using Arcivematica and AtoM…



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