

IT, arkivarens verktøy eller forbannelse?

Arne-Kristian Groven, Riksarkivet

arngro@arkivverket.no

24. mars 2015

Litt om meg

- Mitt navn er Arne-Kristian Groven
- Jeg er inne i mitt fjerde år på Riksarkivet
- Forut for dette har jeg ytt konsulentbistand og vært med i norske og internasjonale forsknings- og innovasjonsprosjekt, innenfor områder som:
 - IT-sikkerhet
 - Forbedring av arbeidsprosesser der IT-inngår
 - Forbedring av programvarekvalitet
 - Digital bevaring, m.m.

Bakgrunnen for denne presentasjonen...

- Jeg har lært Arkiv-Norge godt å kjenne de siste åtte årene.
- Med dets mange utfordringer...
- Jeg har nylig deltatt i en kartlegging som har estimert hvor mange (kilde-)systemer det bør bevares informasjon fra:
 - Tallet er stort!
- Og hvor mange det har blitt bevart informasjon fra:
 - Tallet er forsvinnende lite!!
- Situasjonen er ikke bedre i statlig sektor, i oljesektoren, i helsesektoren, i privat sektor!!!

Bakgrunnen for denne presentasjonen...

- Det er så mange utfordringer og uløste oppgaver at det av mange kan oppfattes som:
 - Problemer som ikke kan løses....
 - Ligninger med for mange ukjente...
- Med min bakgrunn fra ulike forsknings- og innovasjonsprosjekter vet jeg at løsninger finnes der ute i den store verden
 - Dette vil jeg dele litt med dere

Bakgrunnen for denne presentasjonen...

- Det er så mange løse tråder, uløste oppgaver at det av mange kan oppfattes som:
 - Problemer som ikke kan løses....
 - Ligninger med for mange ukjente...
- Men med min bakgrunn fra ulike forsknings- og innovasjonsprosjekter vet jeg at løsninger finnes der ute i den store verden
 - Dette vil jeg dele litt med dere

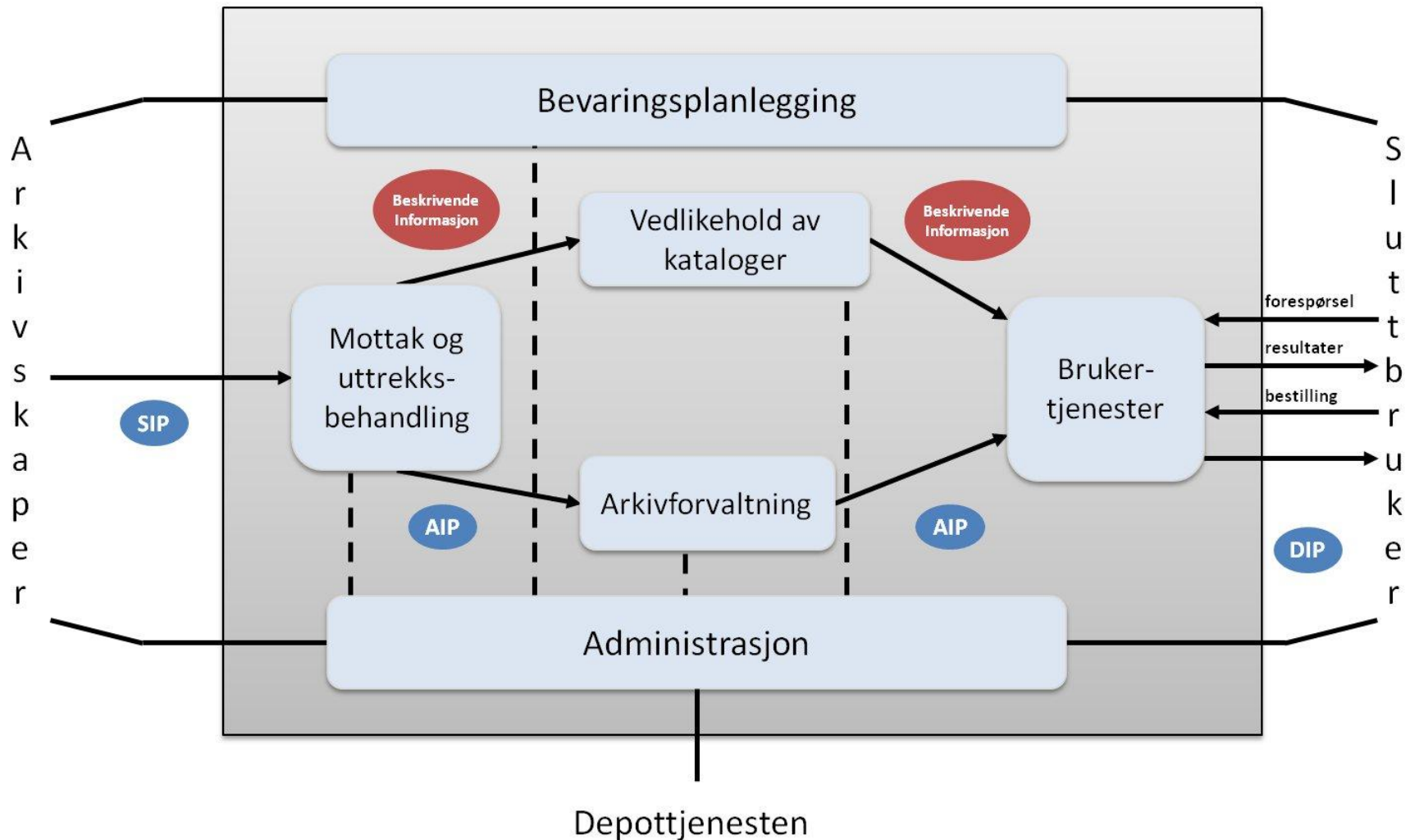
Denne presentasjonen vil først...

- Presentere (depot-) arkivarens verktøykasse:
 - Gi en smakebit på de IT-verktøy som finnes der ute i verden, hva jeg vil kalle «state-of-art» verktøy.
- De utgjør en fleksibel helhet som ikke var der for bare fem år siden, og dekker:
 - uttrekk, mottak, forvaltning av metadata, depottjenester, brukertjenester/aksess/gjenfinning
- Jeg vil spesielt synliggjøre kritiske suksessfaktorer!

...Deretter vil jeg

- Beskrive det jeg vil kalle «forbannelsen»
 - Steder en kan trå feil når en arbeider med langtidsbevaring av digitalt arkivmateriale,
 - og konsekvensen av slike feilsteg
- Avslutningsvis håper jeg vi får tid til en liten diskusjon
 - Eller vi kan ta den under paneldebatten?

Er OAIS kjent? Jeg vil i det følgende beskrive IT-verktøy fra høyre mot venstre...



Verktøy for sluttbruker: Har vi/dere slik programvare i dag?

The screenshot displays the University of Strathclyde Archives website. The search bar at the top contains the term 'geddes' and shows '6 results'. The left sidebar contains navigation menus for LANGUAGE (Unique records: 6, English: 6), LEVEL OF DESCRIPTION (All, Item: 6), ARCHIVAL INSTITUTIONS (All, University of Strathclyde Archives, U...: 6), CREATORS (All, Mears, Norah, 1887-1967, née Geddes: 3), NAMES (All, Mears, Norah, 1887-1967, née Geddes: 3), PLACES, SUBJECTS, and MEDIA TYPES (All, Image: 6). The main content area lists several items:

- Sketches of Chessel's Court by Norah Geddes**
T-GED/7/5/30/19 · Item · 1909
Mears, Norah, 1887-1967, née Geddes
- Sketch of King's Wall Garden, arranged by O[pen] Spaces Committee**
T-GED/7/5/30/14 · Item · No date
Drawn by Norah Geddes.
Mears, Norah, 1887-1967, née Geddes
- Sections and elevations of proposed steps at King's Wall Garden, by Norah Geddes**
T-GED/7/5/30/16 · Item · No date
Mears, Norah, 1887-1967, née Geddes
- Photograph of group of people inside West Port Garden, Edinburgh**
GB 249 T-GED-22-1-British Isles-Scotland-518-4 · Item · No date [c 1910]
Patrick Geddes in the foreground.
- Suggested plan for National Institute of Geography, Edinburgh**
GB 249 T-GED-22-1-British Isles-Scotland-539-1 · Item · 1902
Cover information: 'Suggested Plan for National Institute of Geography According to Designs of Professor Geddes' Scale: 1:400.
- Photograph of children from Castlehill School, Edinburgh going in procession to the ...**
GB 249 T-GED-22-1-British Isles-Scotland-475 · Item · No date [c 1910]
Author: Anon. Label on reverse: 'Photograph of children from Castlehill School going in procession to the

Kilde: <http://strathclyde.ica-atom.org>

Eksemplet er hentet fra et tilfeldig arkiv ute i verden... (Skottland)

University of Strathclyde Archives - A unique learning and research resource open to all

University of Strathclyde Archives, United Kingdom


Holdings Quick search

- Collection T-GED - Patrick Geddes papers
- Section 22 - Maps, plans, photographs, prints and d...
- Series 1 - Topography
- Subseries British Isles - British Isles
- Subseries Scotland - Scotland
- File 518 - Photographs of opening of W...

- Item 1 - View from West Port Garden to...
- Item 2 - View from West Port Garden to...
- Item 3 - View from West Port Garden to...
- Item 4 - Photograph of group of people...**
- Item 5 - Photograph of group of people...
- Item 6 - Photograph of group of people...
- Item 7 - Photograph of group of childre...
- Item 8 - Photograph of group of childre...

Item 4 - Photograph of group of people inside West Port Garden, Edinburgh

Patrick Geddes papers > Maps, plans, photographs, prints and d... > Topography > British Isles > Scotland > Photographs of opening of West Port G... > Photograph of group of people inside ...



Reports
Export
Dublin Core 1.1 XML
EAD 2002 XML

Identity area

Reference code	GB 249 T-GED/22/1/British Isles/Scotland/518/4
Title	Photograph of group of people inside West Port Garden, Edinburgh
Date(s)	<ul style="list-style-type: none">No date [c 1910] (Creation)
Level of description	Item
Extent and medium	1 photograph mounted on cardboard: black and white; 14.6 x 10.3 cm on 19.4 x 16.1 cm backing

Context area

Name of creator	Inglis, Alexander A., 1881-1913, photographer (1881-1913)
Repository	University of Strathclyde Archives, United Kingdom

Content and structure area

Scope and content	Patrick Geddes in the foreground.
-------------------	-----------------------------------

Ulike innganger til arkivmaterialet

- Level of description
 - Item, File, Series, Section, Collection, Subseries
- Archival institutions
- Creators
- Names
- Items
- Places
- Subjects
 - Biology, Urban planning, Government, Navigation, Mechanics, Public health, Housing, Chronology, Mineralogy, Zoology...
- Media types
- (Pluss fritekstsøk)

Et (lite) antall assosierte metadata

Identity area	
Reference code	GB 249 T-GED/7/5/30/19
Title	Sketches of Chessel's Court by Norah Geddes
Date(s)	<ul style="list-style-type: none">1909 (Creation)
Level of description	Item
Extent and medium	2 sketches: pencil and wash
Context area	
Name of creator	Mears, Norah, 1887-1967, née Geddes
Repository	University of Strathclyde Archives, United Kingdom
Conditions of access and use area	
Conditions governing access	Open
Language of material	<ul style="list-style-type: none">English
Access points	
Name access points	<ul style="list-style-type: none">Mears, Norah, 1887-1967, née Geddes (Creator)
Digital object metadata	
Filename	T_GED_7_5_30_19_atom.jpg
Media type	Image
Mime-type	image/jpeg
Filesize	674 KiB
Uploaded	April 17, 2013 5:34 AM

Kilde: <http://strathclyde.ica-atom.org>

Merk dataeksportmulighetene!

University of Strathclyde Archives, United Kingdom

Search University of Strathclyde Archives, Unite Browse

University of Strathclyde Archives - A unique learning and research resource open to all

Item 19 - Sketches of Chessel's Court by Norah Geddes

Patrick Geddes papers > Edinburgh and the Outlook Tower > Outlook Tower: Collection, Activities, Ev... > Sketch plans and drawings en

Sketches of Chessel's Court by Norah G...

Holdings Quick search

- Collection T-GED - Patrick Geddes papers
- Section 7 - Edinburgh and the Outlook ...
- Series 5 - Outlook Tower: Collection, Ac...
- File 30 - Sketch plans and drawings enti...

Item 15 - Outline plan of Portsburgh Ga...

Item 16 - Sections and elevations of pro...

Item 17 - Block plan of garden near Cas...

Item 18 - Plan of St. John's Garden, [St. J...]

Item 19 - Sketches of Chessel's Court b...

Item 20 - Sketch of garden open space, ...

Item 21 - Sketch plan of alternative layo...

Item 22 - Letter from S. & R. Slater (Buil...

Item 23 - Letter from S. & R. Slater (Buil...

Related people organization

Mears, Nora
née Geddes

Reports

Export

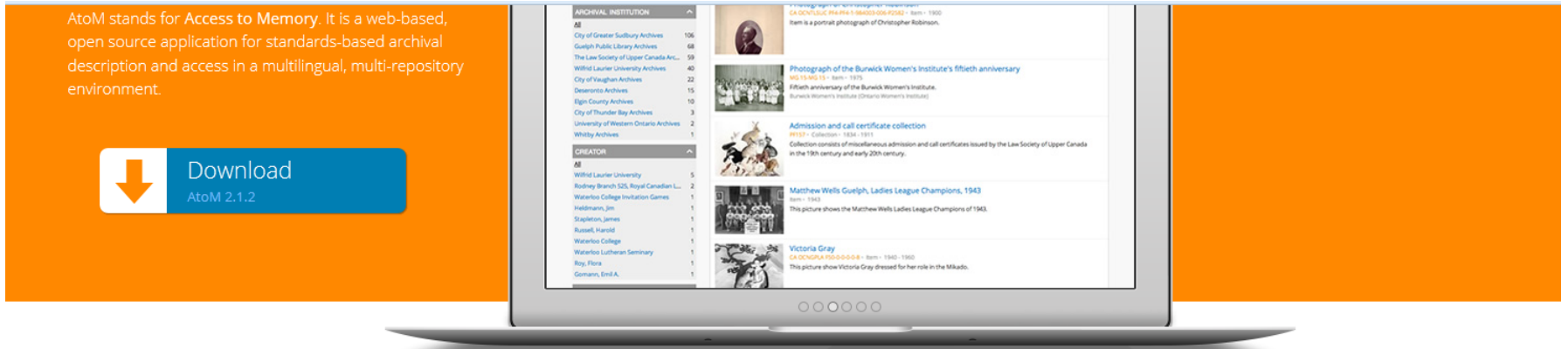
- Dublin Core 1.1 XML
- EAD 2002 XML

Identity area

Reference code	GB 249 T-GED/7/5/30/19
Title	Sketches of Chessel's Court by Norah Geddes
Date(s)	<ul style="list-style-type: none">1909 (Creation)
Level of description	Item
Extent and medium	2 sketches: pencil and wash

Context area

Programmet som muliggjør dette er fritt tilgjengelig for alle...



Web-based

Access your AtoM installation from anywhere you have an internet connection. All core AtoM functions take place via a web browser, with minimal assumptions about end-user requirements for access. No more syncing multiple installations on a per-machine basis – install AtoM once, and access it from anywhere.

Multilingual

All user interface elements and database content can be translated into multiple languages, using our built-in translation interface. Our translations

Open source

All AtoM code is released under a GNU Affero General Public License (A-GPL 3.0) – giving you the freedom to study, modify, improve, and distribute it. We believe that an important part of access is accessibility, and that everyone should have access to the tools they need to preserve cultural heritage materials. AtoM code is always freely available, and our documentation is also released under a Creative Commons Share-alike license. Check out our [code repository](#)!

Multirepository

Built for use by a single institution for its own descriptions, or as a multi-repository “union list” (network, portal) accepting descriptions from any

Standards-based

AtoM was originally built with support from the [International Council on Archives](#), to encourage broader international standards adoption. We've built standards-compliance into the core of AtoM, and offer easy-to-use, web-based edit templates that conform to a wide variety of international and national standards.

Constantly improving

AtoM is an active, dynamic open-source project with a broad user base. We're constantly working with our community to improve the application, and all

Import/export friendly

Your data will never be locked into AtoM – we implement a number of metadata exchange standards to support easy import and export through the AtoM user interface.

Currently AtoM supports the following import/export formats: [EAD](#), [EAC-CPF](#), [CSV](#) and [SKOS](#). Find out more in our [user manual](#).

...og kan motta arkivpakker skapt i et annet verktøyet, også det fritt tilgjengelig

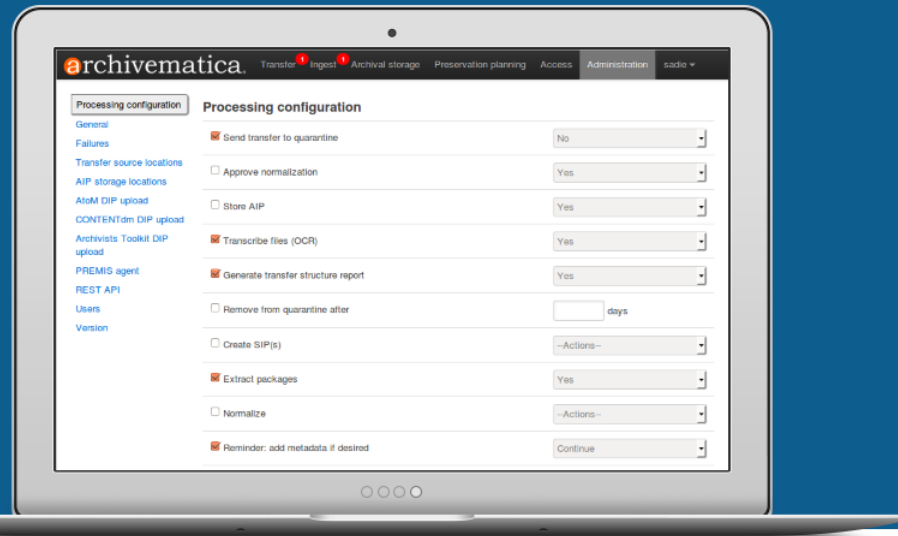
archivematica

Home Downloads Documentation Community Development News Wiki Demo

Preserving memory since 2009

Archivematica is a web- and standards-based, open-source application which allows your institution to preserve long-term access to trustworthy, authentic and reliable digital content.

 Download



Standards-based


Archivematica uses a [micro-services](#) design pattern to provide an integrated suite of software tools that allows users to process digital objects from ingest to access in compliance with the ISO-OAIS functional model. Users monitor and control the micro-services via a web-based dashboard. Archivematica uses METS, PREMIS (events, agents, rights and restrictions), Dublin Core, the Library of Congress BagIt specification and other best practice standards and practices to provide **trustworthy, authentic, reliable, and interoperable** archival packages (AIPs) for storage in your preferred repository.

La oss se på noen kritiske suksessfaktorer....

Åpen kildekode/
fri programvare

Open Source

All Archivematica code is released under a GNU Affero General Public License (A-GPL 3.0) – giving you the freedom to study, modify, improve, and distribute it. We believe that an important part of preservation is transparency, and that memory institutions should be able to demonstrate at every stage what happens when they process cultural heritage materials for preservation. Archivematica code is always freely available, and our documentation is also released under a Creative Commons Share-alike license.



Check out our code repositories



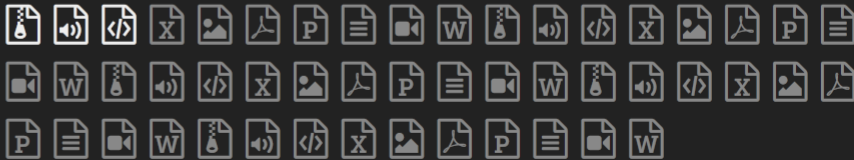
Flexible and customizable

Archivematica provides several decision points that give the user control over choices about format identification tools, printing the original order of the directories ingested, examining contents for private and personal information, extracting contents of packages and for disk images, transcribing content, and more. Users may also preconfigure most of these options for seamless ingest to archival storage and access. Archivematica offers many ingest workflows: metadata and submission documentation import, zipped and unzipped Bag ingest, digital forensic image processing, SIP arrangement, manual normalization, and dataset management.

Tåler stor
variasjon i
input

Compatible with hundreds of formats

In the Format Policy Registry (FPR), Archivematica implements its default format policies based on an analysis of the significant characteristics of file formats. The FPR also offers an editable, flexible framework for format identification, package extraction, transcription and normalization for preservation and access. Your institution can update tools, rules and commands in your local FPR from the Artefactual-managed FPR server. You can also add your own, local policies to your internal FPR. The FPR is integrated with PRONOM.



Learn more about the FPR service

Fleksibel løsning for
konvertering til arkiv-
filformat

Og enda noen flere kritiske suksessfaktorer....



Advanced search and storage management

You can easily search your backlog and your archival storage from within your Archivematica web-based dashboard. This means you can download stored AIPs as complete packages, individual objects, or every package in an AIC. You can also manage your storage and processing locations using the Archivematica Storage Service, including a two-step deletion process that requires justification and approval to eliminate a stored AIP.

[Learn more about the storage service](#)



Integrated with third-party systems

Memory institutions have dedicated voluminous resources over the past couple of decades to implement various software platforms to manage digital objects. For this reason, we believe in leveraging the strength of other tools and integrating with them wherever possible.

dSpace

CONTENTdm

Islandora

LOCKSS

AtoM

DuraCloud

OpenStack

Archivists' Toolkit

Integrasjon med mange ulike depotløsninger

Mange aktive brukere

Active community

Og enda noen flere kritiske suksessfaktorer

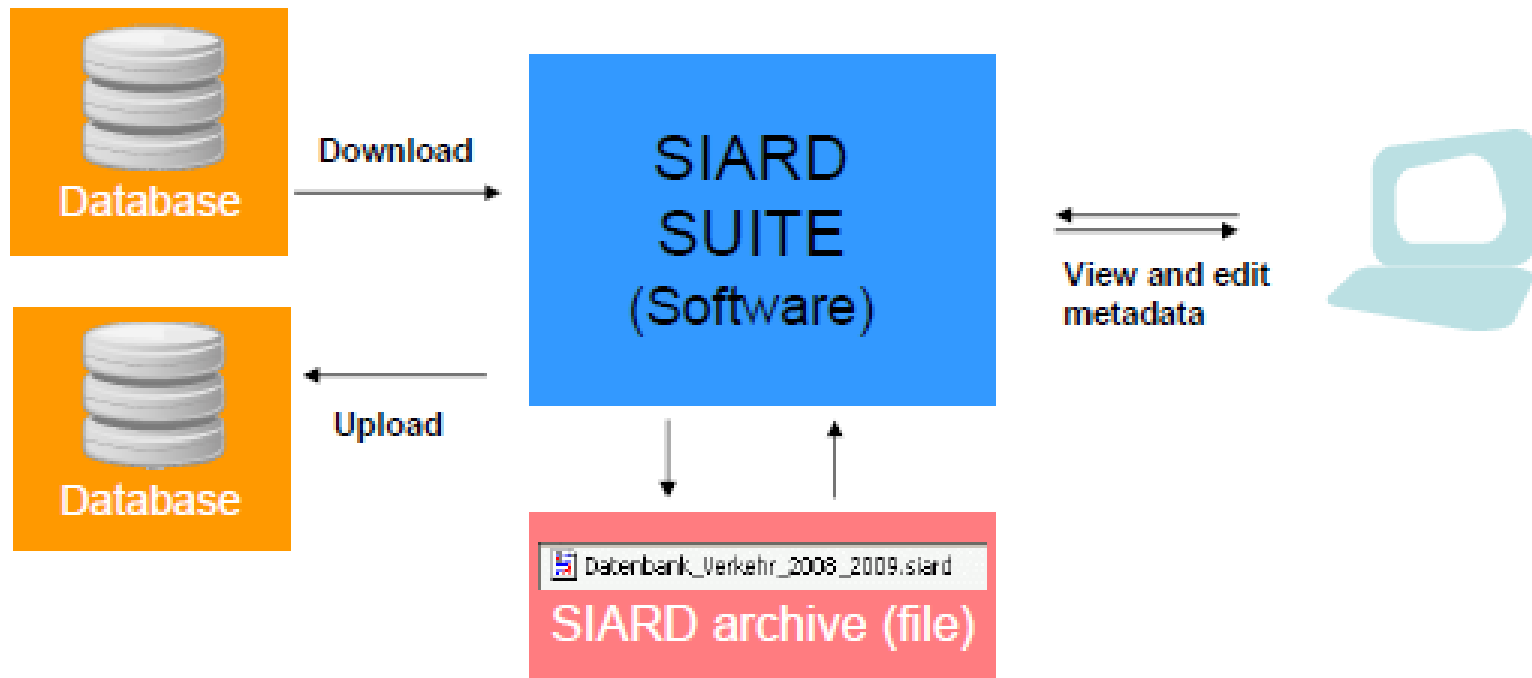


Standards-based

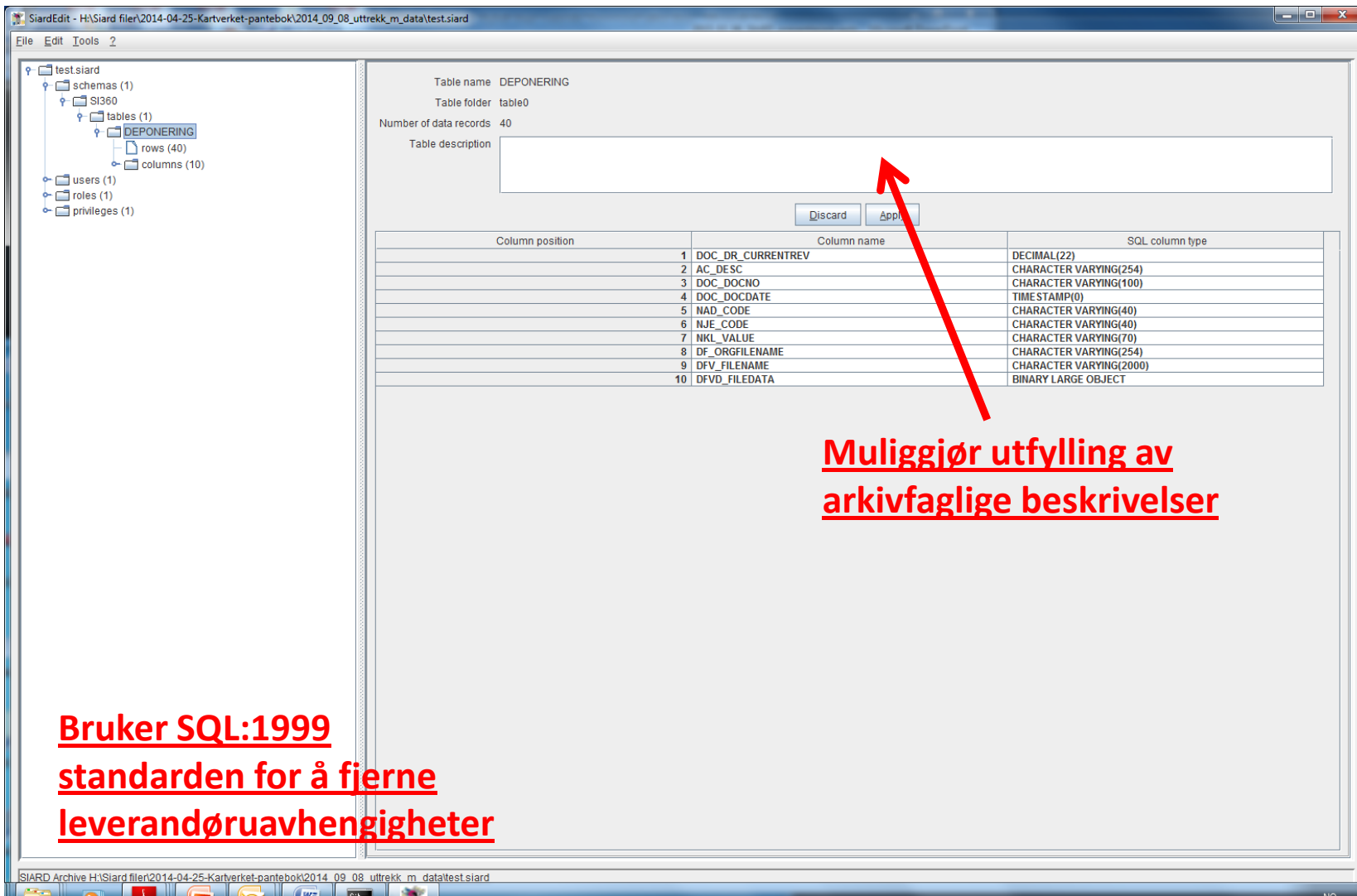
Archivematica uses a [micro-services](#) design pattern to provide an integrated suite of software tools that allows users to process digital objects from ingest to access in compliance with the ISO-OAIS functional model. Users monitor and control the micro-services via a web-based dashboard. Archivematica uses METS, PREMIS (events, agents, rights and restrictions), Dublin Core, the Library of Congress BagIt specification and other best practice standards and practices to provide **trustworthy, authentic, reliable, and interoperable** archival packages (AIPs) for storage in your preferred repository.

**Bruker internasjonale
standarder for å skape tillit til
materialet og for å være i
stand til å utveksle data og
metadata**

Ser vi på databasebevaring så finnes det også her tilgjengelige IT-verktøy



...Slik ser databasebevarings- verktøyet SIARD Suite ut



The screenshot shows the SIARD Suite interface. On the left is a tree view of the database structure. The main area displays details for the table 'DEPONERING', including the number of records (40) and a table description field. Below this is a table of column definitions. A red arrow points from the 'Apply' button to the table description field.

Column position	Column name	SQL column type
1	DOC_DR_CURRENTREV	DECIMAL(22)
2	AC_DESC	CHARACTER VARYING(254)
3	DOC_DOCNO	CHARACTER VARYING(100)
4	DOC_DOCDATE	TIME STAMP(0)
5	NAD_CODE	CHARACTER VARYING(40)
6	NJE_CODE	CHARACTER VARYING(40)
7	NKL_VALUE	CHARACTER VARYING(70)
8	DF_ORGFILENAME	CHARACTER VARYING(254)
9	DFV_FILENAME	CHARACTER VARYING(2000)
10	DFVD_FILEDATA	BINARY LARGE OBJECT

**Muliggjør utfylling av
arkivfaglige beskrivelser**

**Bruker SQL:1999
standarden for å fjerne
leverandøruavhengigheter**

Kritiske suksessfaktorer for digital bevaring

- Bevare tillit til arkivmaterialet:
 - Relatert til dets opprinnelse, ekthet/autentisitet
 - Og dets integritet
 - Mest kritisk her er endringskontroll, enten det er ved transformasjoner/konvertering/migrering av struktur og innhold, eller manipulering av metadata/beskrivelser
- Sørge for presis gjenfinning og korrekt visning

Konvertering/migrering: Hvordan bevare tillit gjennom transformasjoner

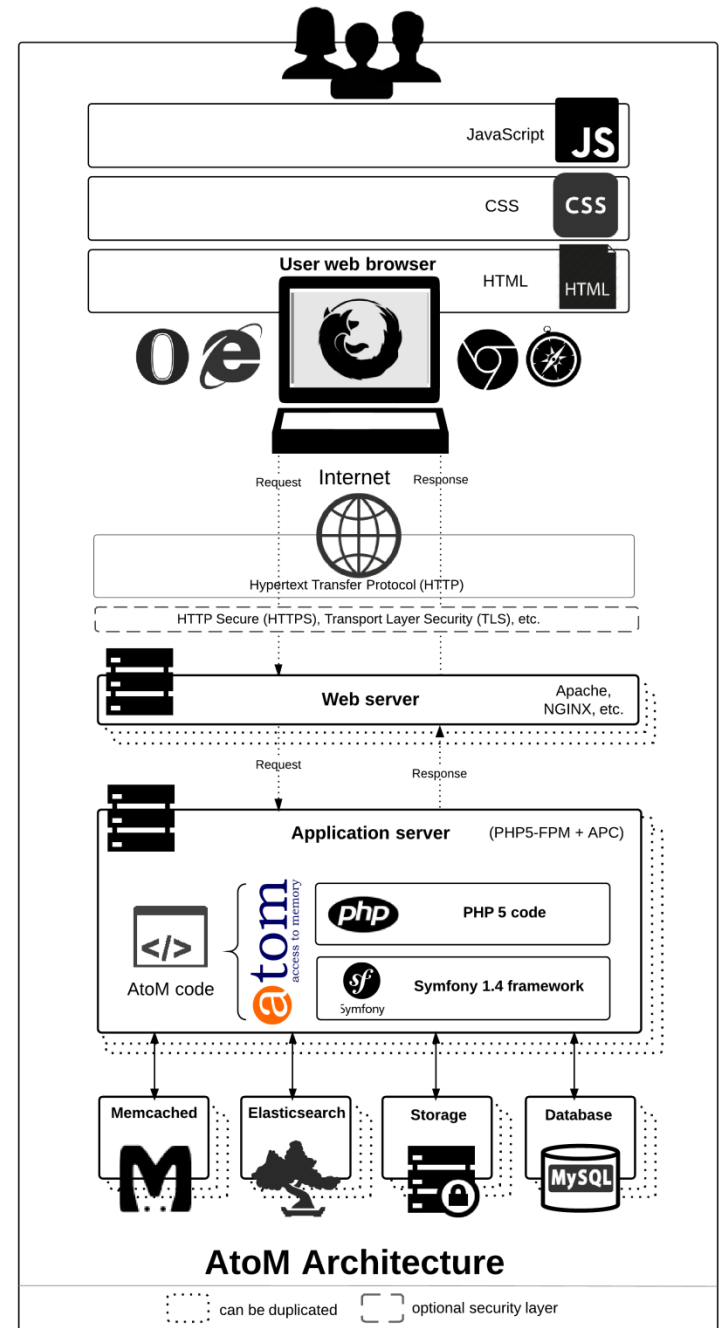
- Mest mulig automatisert og sporbart (autologging)
- Minst mulig manuell manipulasjon
- Manipulasjon kun via sporbare IT-verktøy
- Prosessen må kunne gjentas, med samme resultat hver gang
- Logger, så vel som data/metadata må sikres (IT-sikkerhet)
- Alt dette underbygger hvorfor «riktige» IT-verktøy er nødvendig!!

Programvareforbannelsen...

- Programvare er «survival of the fittest», Darwin i praksis
 - Enten gjennom markedsmakt (Apple, Microsoft m.fl.)
 - Eller gjennom produktkvalitet, helt nødvendig for at fri programvare (program med åpen kildekode) skal lykkes!
 - Eller gjennom supportkvalitet
- For hvert programmeringsprosjekt som lykkes, mislykkes hundre!
 - Dette bør man tenke på hver gang man starter et egenutvikling av programvare

Det er mye teknologi «under panseret»

- Her er arkitekturen til innsyns- og depotløsningen jeg nettopp viste
- Egenutviklede løsninger må ta høyde for alt dette, for å lykkes!



Programvareforbannelsen...

- Programvaren jeg viste dere er (stort sett) fri programvare
- Programvarekvalitet for fri programvare er lett målbart
 - Mange brukere er et kvalitetsmål
 - Kontinuerlig videreutvikling er et annet... osv.
- Programvaren jeg nettopp viste signaliserer at den går i retning av å være «the fittest»

Programvareforbannelsen...

- I Arkiv-Norge i dag finnes det knapt lokalt utviklede løsninger som kan overleve, slik jeg ser det(!)
- Markedsmakt er via lov- og forskrifter gitt de leverandørene av sak-arkivsystem som blir Noark-godkjente, men
 - Kvaliteten er veldig varierende
 - Dekningsgraden er relativt liten: Det finnes kanskje 1000 ulike IT-systemer som inneholder bevaringsverdig informasjon. Av disse er kanskje 10-20 Noark-godkjente system.

Programvareforbannelsen

- Jeg har dessuten til gode å treffe på en organisasjon som driver troverdig forvaltning og support av IT-verktøy for arkivsektoren i Norge, med:
 - God brukerkommunikasjon
 - Rask respons ved feilrapportering
 - God videreutvikling

Forvaltning av metadata og beskrivelser: Hvordan sikre tillit og presis gjenfinning

- Mest mulig automatisert og sporbart (autologging)
- Manuell manipulasjon kun via sporbare IT-verktøy
- Logger, så vel som data/metadata må sikres (IT-sikkerhet)
- Registrere «riktige» metadata og beskrivelser
- Bruke velegnede formater for metadata og beskrivelser
 - Som kan utveksles mellom IT-systemer
 - Som kan brukes av ulike gjenfinnings-verktøy

Metadataforbannelsen

- Standardiserte metadata-formater, også her gjelder «survival of the fittest»
- De som benyttes av flest IT-verktøy har vunnet!

Metadataforbannelsen

- De standardiserte metadataformatene som er brukt i IT-verktøyene jeg nettopp viste er:
 - METS, for å beskrive struktur
 - PREMIS, for å (auto-)logging
 - EAD, for å beskrive bestand
 - Dublin Core, for å beskrive arkivskaper
 - Bag-IT, for arkivpakker m.m.
- Vil ADDML eller Noark noen gang kunne føyes på en slik liste?

Metadataforbannelsen

- Hva gjelder hva som er «riktige» metadata så kan den diskusjonen fylle en hel forelesning
- Men det er viktig å merke seg at løsningene jeg nettopp viste holder seg på et minimalt antall metadata
- Dette er veldig langt fra Noark-4 sine nesten 300 metadata

Noen betraktninger til slutt

- IT-verktøyene jeg nettopp presenterte tillater stor variasjon i input fra kildesystemene
- Vi har et etterslep i Arkiv-Norge, fra hundretalls eller tusentalls ulike kildesystem
- Skal vi lykkes må variasjon tillates, og et lite antall metadata fra hver kilde aksepteres
- Suksess eller fiasko avhenger ikke minst av de suksesskriterier jeg nå har skissert

TAKK FOR MEG!