DELIVERABLE

Project Acronym:	**E-ARK**
Grant Agreement Number:	**620998**
Project Title:	**European Archival Records and Knowledge Preservation**

DELIVERABLE DETAILS

DELIVERABLE REFERENCE NO.	**D5.4**
DELIVERABLE TITLE	**Search, Access and Display Interfaces**
REVISION	1.0

AUTHOR(S) Name(s)	Organisation(s)
Alex Thirifays	Magenta Aps
Andreas Kring	Magenta Aps
Lanre Abiwon	Magenta Aps
Frank Thomsen	Magenta Aps

Mihai Bartha	Austrian Institute of Technology
Bruno Ferreira	KEEP SOLUTIONS, LDA
Zoltán Lux	National Archives of Hungary
Gregor Završnik	National Archives of Slovenia
Anja Paulič	National Archives of Slovenia

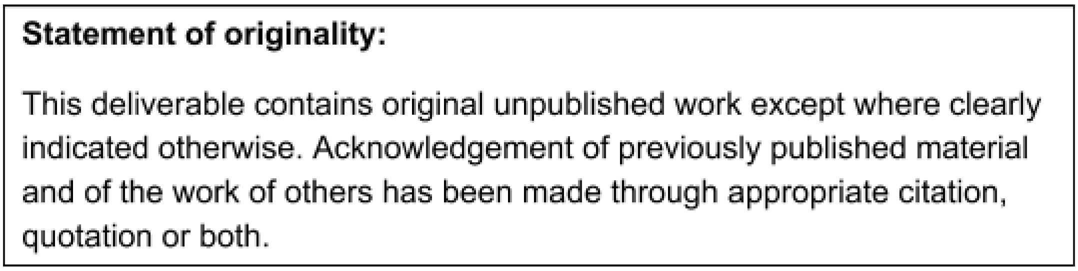
REVIEWER(S)	
Name(s)	Organisation(s)
Kuldar Aas	National Archives of Estonia
Andrew Wilson	University of Brighton

Project co-funded by the European Commission within the ICT Policy Support Programme Dissemination Level		
P	**Public**	x
C	**Confidential, only for members of the Consortium and the Commission Services**	

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Submitted Revisions History

Revision No.	Date	Authors(s)	Organisation	Description
1.0	27/1/2017	Alex Thirifays	Magenta	Original Submitted Version



Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

1 Executive Summary 7

2 Introduction 8

3 Description of tools and tool documentation 10

3.1 The Access Software Platform 11

3.1.1 A common interface 11

3.1.2 User identification and authorisation 12

3.1.3 Access to archival records 12

3.1.4 Integration to data management: E-ARK web and the Access Software Platform 12

3.2 Search Module, Order Management Tool, and IP Viewer 15

3.2.1 User interaction overview 15

3.2.2 System login 16

3.2.3 Global navigation 18

3.2.4 Search Module, and ordering 19

3.2.4.1 Initial search 19

3.2.4.2 Refined search 22

3.2.4.3 Ordering archival records 23

3.2.4.4 Orders overview 26

3.2.4.5 Viewing a single order 27

3.2.5 Order Management Tool 28

3.2.5.1 Listing new and existing orders 28

3.2.5.2 Viewing order details 30

3.2.5.3 Creating a new package from an order 31

3.2.5.4 Deciding to process an order 31

3.2.5.5 Create a pre-DIP 32

3.2.5.6 Copying data and metadata to the DIP 33

3.2.5.6.1 Copying relevant data and metadata to the future DIPu 35

3.2.5.7 Package DIP 38

3.2.6 IP Viewer 39

3.2.6.1 Browsing IPs 39

3.2.6.2 Navigating the IP 39

3.2.6.3 Viewing related metadata 40

3.2.6.4 Viewing individual files 40

3.2.6.5 Searching the IP 43

3.2.7 Code and documentation for Search Module, Order Management Tool, and IP Viewer 43

3.3 The AIP-DIP conversion tool 44

3.3.1 Task execution framework 44

3.3.2 AIP-DIP conversion tasks 44

3.3.3 DIP creation API 46

3.3.3.1 Prepare DIP working area 46

3.3.3.2 Run DIP creation process 47

3.3.3.3 Check job status 47

3.3.4 Index DIP in storage API 47

3.3.4.1 Submitting and indexing job 47

3.3.4.2 Check job status 48

3.3.5 Code and documentation for the AIP-DIP conversion tool 48

3.4 The CMIS Viewer 49

3.4.1 Use case and feature description summary 49

3.4.1.1 Authentication and user authorisation 50

3.4.1.2 Configuring CMIS Viewer repository connection 50

3.4.1.3 Browsing and getting around 52

3.4.1.4 User Management 53

3.4.2 Code and documentation for the CMIS Viewer 54

3.5 Access to specific content information types 54

3.5.1 Access to databases and EDRMS's stored in the SIARD format 55

3.5.1.1 The Database Preservation Toolkit and its GUI: Access via an RDBMS (SQL solution) 55

3.5.1.1.1 The graphical user interface for the Database Preservation Toolkit 56

3.5.1.1.2 Code and documentation for the Database Preservation Toolkit and its GUI 58

3.5.1.2 The Database Visualization Toolkit: Access via Solr (NoSQL-solution) 58

3.5.1.2.1 Code and documentation for the Database Visualization Toolkit 68

3.5.2 Geodata tools: Access to geodata stored in the SMURF format 68

3.5.2.1 Geodata specific search 69

3.5.2.1.1 Example using Peripleo 70

3.5.2.2 Access to geodata 72

3.5.2.2.1 Unstructured files 72

3.5.2.2.1.1 Example using Geotools - Opening a vector dataset in QGIS 72

3.5.2.2.2 Web service 75

3.5.2.2.2.1 Example using Geoserver - providing a link to geodata viewer based on Geoserver 76

3.5.2.2.3 Edited and customised view of unstructured geodata in QGIS 77

3.5.2.2.3.1 Example using QGIS - Creating a unified access QGIS project displaying geodata from multiple AIPs 78

3.5.2.2.4 Reproduction of geodata 80

3.5.2.3 Code and documentation for Geodata tools 81

3.5.3 The SMURF Tool (IP Viewer): Access to Electronic Records management Systems and Simple File-System Based Records 81

3.5.4 The OLAP Tools: OLAP access to information stored in the SIARD format 81

3.5.4.1 Code and documentation for OLAP tools 84

6 Glossary 85

7 References and associated links and documents 90

1 Executive Summary

The aim of this deliverable is to describe the “Search, Access and Display Interfaces” that have been developed in the Access component of the E-ARK project.

The deliverable is mainly a software deliverable and therefore this document provides only underpinning descriptions of and links to the software itself.

The tools that are described and provided allow Consumers (ie. end-users and archivists) to

1. Search and order records (primarily end-users, but also archivists)
2. Manage orders of records and manage the records themselves, including the AIP to DIP conversion (archivists only)
3. Access ordered records as DIPs (primarily end-users, but also archivists)

In addition to the the introductory remarks in chapter 2, the functionality of the tools that allow the Consumers to search, manage, and access records is described in chapter 3. After the description of each tool, links are provided to code and documentation.

2 Introduction

The purpose of this document is to describe the tools for accessing archival material and to provide links to their documentation and the code. The tools are based on specifications that have been partially or fully created in the E-ARK project. The final DIP specifications as well as the relevant access scenarios are described in the E-ARK Final DIP Specification[1].

The Access tools that will be described in this document have been created to support the E-ARK Access workflow, which, together with a number of Access use cases, was identified in the E-ARK DIP Draft Specification[2]. Below is an illustration of the very top level of that workflow, which is further detailed in the General Model[3].

The high-level workflow of the E-ARK Access process encompasses four main steps:

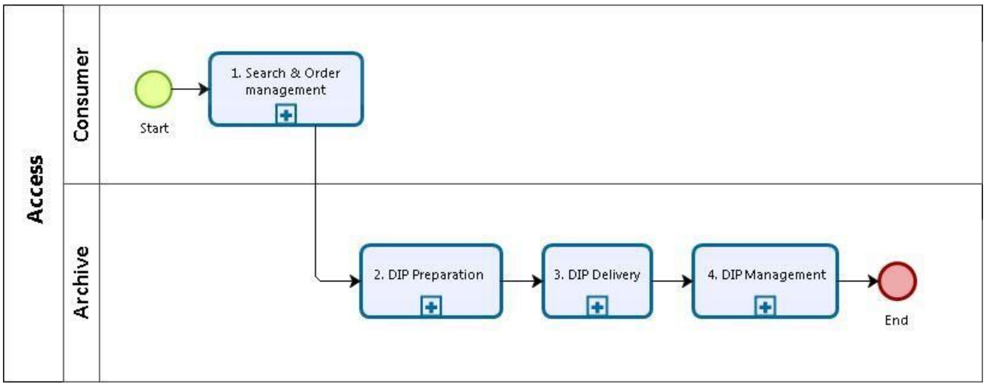


Figure 1 – Overview of the E-ARK Access process

1. "Search & Order Management", where the Consumer[4] can search for, identify, and order information packages of interest, using a Finding Aid[5].
2. "DIP Preparation" where the IP is prepared for the end-user[6], for example by migrating an AIP into a DIP.
3. "DIP Delivery" where the DIP is processed by the appropriate tool and rendered to the end-user via a suitable Graphical User Interface (GUI)[7].
4. "DIP Management" where the DIP is either deleted or sent to a permanent or temporary DIP storage.

[1] Cf. E-ARK Final DIP Specification (http://www.eark-project.com/resources/project-deliverables/91-d532)

[2] D5.2 E-ARK DIP Draft Specification
http://www.eark-project.com/resources/project-deliverables/31-d52

[3] E-ARK General Model
http://www.eark-project.com/resources/general-model

[4] The role played by those persons or client systems, which interact with OAIS services to find preserved information of interest and to access that information in detail. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
In E-ARK "Consumer" is an umbrella term that designates all users of archival holdings, thus both internal users, cf. archivists, and external users, cf. end-user.

[5] A type of Access Aid that allows a user to search for and identify Information Packages of interest. Source: OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf

[6] The end-user designates an external user who seeks content information in archival holdings.

[7] A Graphical User Interface (GUI) is a graphical interface to a program on a computer. It takes advantage of the computer's graphics capabilities to make the program easier to use.

As it is shown later (Figure 2), these four steps have been reduced to three steps, because it is reasonable to include “DIP Management” into “DIP Preparation”, and thus into the tool that caters for the archivist’s work on IPs in general.

It is important to emphasize that the three overall E-ARK Access tools are created to process archival records that comply to the E-ARK IP specifications. As such the present deliverable should be read together with the final DIP specification.

3 Description of tools and tool documentation

This chapter describes the Access Software Platform (ASP) and related tools[8]. It also provides links to the code and the documentation of the software.

Accessing archival records consists of three main steps that allow for the completion of the general Access workflow that is initiated by an end-user when searching for records in an archive.

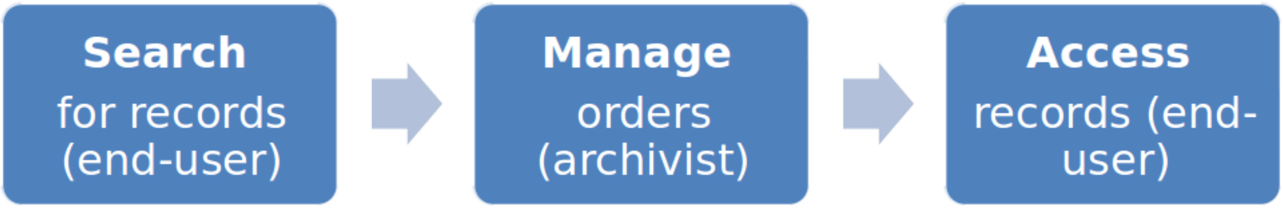


Figure 2 – The three main steps of the Access flow

The first step is covered by the tool used for searching archival records. This tool is called the **Search Module**, and allows the end-user to search and order archival records from the archive.

The second step is covered by the tool used to manage the orders which are sent from the end-users. This tool is called the **Order Management Tool**, and allows for the archivist to process an order, thus retrieving the requested archival records, and creating a Dissemination Information Package (DIP).

The third and last step is covered by the tool that the end-users employ to inspect the ordered DIP. This tool is called the **IP Viewer**, and allows the end-user to browse, search and view DIPs. The IP Viewer is also accessible for the archivist in the Order Management Tool.

All three tools are unified in one platform: The Access Software Platform (ASP).

In addition to these three tools, there are a series of added tools that support accessing archival records:

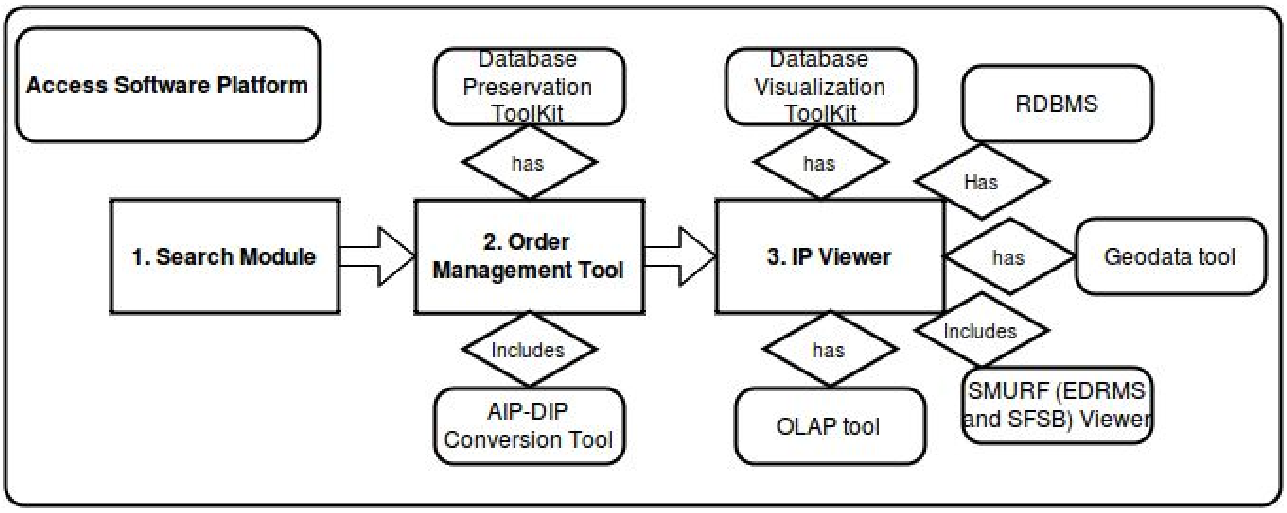


Figure 3 – The Access Software Platform

[8] Referred to as "The Search, Access, and Display Interfaces" in the Description of Work, page 23.

In the illustration above, ‘Has’ means that the tool is part of the workflow, but that it’s not integrated into or invoked from the Access Software Platform. ‘Includes’ means that the tool is integrated into and accessible directly from the Access Software Platform[9].

The tools that are described in this document are:

1. The Access Software Platform (Search Module, Order Management Tool, and IP Viewer), cf. section 3.1 and 3.2
2. The AIP-DIP Conversion Tool, cf. section 3.3
3. The CMIS[10] Viewer, cf. section 3.4. Note that this tool is not a part of the ASP and therefore it is not part of the above illustration.
4. The Database Preservation Toolkit and its GUI, cf. section 3.5.1.1
5. The Database Visualization Toolkit, cf. section 3.5.1.2
6. The Geodata Tools, cf. section 3.5.2
7. The SMURF[11] Tool (IP Viewer): Access to Electronic Records Management Systems and Simple File-System Based Records, cf. section 3.5.3
8. The OLAP[12] tools, cf. section 3.5.4

3.1 The Access Software Platform

3.1.1 A common interface

The interface is common to all three overall steps in the workflow and uses the same web development framework. Because some IP formats (e.g. SIARD, GML) require very special Access software applications (e.g. respectively Database Visualization Toolkit, and QGIS), it has not been possible to integrate all access tools into the Access Software Platform. This is why some IP formats require stand-alone access tools for rendering.

Web technologies have been leveraged to ensure a uniform look and feel for GUI elements across the common interface. An added benefit of using web technologies is that element styles can be defined in a single workflow, which is beneficial to uniformity of the GUI.

The HTML5 technology of choice is AngularJS, which is combined with ordinary HTML/CSS to make single-page javascript GUIs. Design patterns collected and refined by Google have been used[13]. Apart from design patterns, style guides are another crucial element for ensuring uniform design. A style guide is a set of defined elements and their related styles that are created from a central source and reused across a project. In conjunction with AngularJS, Angular Material Design 2[14], which is a collection of reusable widgets and related styles, has been used to ensure uniformity of the GUI.

[9] ‘Has’ and ‘Includes’ are only used to make this distinction in the illustration, but not in the deliverable text in general.

[10] Content Management Interoperability Services (CMIS) is an open standard that allows different content management systems to inter-operate over the Internet. Specifically, CMIS defines an abstraction layer for controlling diverse document management systems and repositories using web protocols, cf. CMIS https://en.wikipedia.org/wiki/Content_Management_Interoperability_Services

[11] Semantically-Marked-Up-Records Format is an E-ARK format that allows the preservation of Single File-Based Records (SFSB) and Electronic Documents and Records Management Systems (EDRMS). D3.3 E-ARK SMURF http://www.eark-project.com/resources/project-deliverables/52-d33smurf

[12] In computing, online analytical processing, or OLAP, is an approach to answering multi-dimensional analytical (MDA) queries swiftly. OLAP is part of the broader category of business intelligence, which also encompasses relational database, report writing and data mining, cf. OLAP https://en.wikipedia.org/wiki/Online_analytical_processing

[13] Material design (Google) https://www.google.com/design/spec/material-design/introduction.html

[14] Angular Material Design 2 https://material.angularjs.org/latest/

3.1.2 User identification and authorisation

E-ARK Web and the Access Software Platform that sits on top of it are to be run in a safe environment since no security framework has been built around it. The Access Software Platform currently has two user roles: The end-user, who has access to limited functionality in the the Search Module and the IP Viewer; and the archivist who has unlimited access to the Search Module, the Order Management Tool, and the IP Viewer.

Most archival institutions will have an already existing user management system. This will often be compatible with LDAP[15] (e.g. Active Directory[16]), so future releases of the Access Software developed within E-ARK should support this protocol. The user database of the access software should be synchronized periodically with the LDAP compatible server within the local archive. With a mechanism for synchronizing users into the database in place, an authentication system can also be implemented, and several options exist to address this issue. One possible solution is to use Kerberos[17], since this is the authentication mechanism for a standard Active Directory. The backend of the E-ARK Access Software Platform is written in Python[18], and there are libraries that handle client-server interactions using the Kerberos protocol.

3.1.3 Access to archival records

The access environment does not set access restrictions. This is managed in the data management layer of the repository, and ultimately access restrictions are managed by the local policies of each archive that determine which archival records are to be indexed and made searchable, and which are not. There are several solutions to this issue. Some rely on logical access restrictions, thus setting these in descriptive metadata (e.g EAD[19]) or in preservation metadata (e.g. PREMIS[20]). Some simply do not physically put restricted records together with the unrestricted ones. The E-ARK search functionality does not make any assumptions regarding access restrictions, and it should not.

3.1.4 Integration to data management: E-ARK web and the Access Software Platform

E-ARK Web, which provide the data, the metadata and the services invoked by the Access Software Platform, provides data management via the Lily open source data management platform[21]. Lily combines big data storage, indexing and search, and unifies Apache HBase[22], Hadoop[23] and Solr[24] into an integrated data platform that includes APIs[25], a high-level data model and schema language, real-time indexing and the powerful search capacities of Apache Solr.

The data management is seen as the technical component for maintaining IPs.

One the one hand, there is the distributed AIP storage component and the dm-file-ingest service (Hadoop-MapReduce-Job)[26] which allows uploading AIPs into an HDFS (Hadoop Distributed File System) storage cluster in order to make these available for distributed file ingest into the Lily repository. Using scheduled processing, new files added to the distributed storage for AIPs are automatically unpackaged,

[15] Lightweight Directory Access Protocol https://en.wikipedia.org/wiki/Lightweight_Directory_Access_Protocol
[16] Active Directory https://en.wikipedia.org/wiki/Active_Directory
[17] Kerberos https://en.wikipedia.org/wiki/Kerberos_(protocol)
[18] Phyton https://en.wikipedia.org/wiki/Python_(programming_language)
[19] Encoded Archival Description. A non-proprietary de facto standard for the encoding of Finding Aids for use in a networked (online) environment.
[20] The PREMIS Data Dictionary for Preservation Metadata is the international standard for metadata to support the preservation of Digital Objects and ensure their long-term usability. (https://www.loc.gov/standards/premis/v3/)
[21] Lily http://www.lilyproject.org/lily/index.html
[22] Apache HBase http://hbase.apache.org/
[23] Hadoop Distributed File System https://hadoop.apache.org/
[24] Apache Solr http://lucene.apache.org/solr/
[25] Aplication Programming Interface https://en.wikipedia.org/wiki/Application_programming_interface
[26] dm-file-ingest https://github.com/eark-project/dm-file-ingest

ingested, and indexed by the dm-file-ingest service. In this way, the Lily repository always has a complete directory of all files contained in the AIPs.

On the other hand, there are various storage areas, such as the Local Storage to upload additional files from other sources by using the local file system, the Working Area, which is used in case an AIP must be transferred into a working environment where it can be unpackaged and modified, and the DIP storage, which can be managed according to local archival policies (deletion of certain DIPs after a certain period of time, etc.).

The AIP to DIP conversion component is part of the E-ARK Web and can be invoked by the Access Software Platform.

The archivist never has to go to E-ARK Web to process an order and complete an order (including the AIP-DIP conversion), but can do everything from the user friendly Access Software Platform.

The E-ARK Web is thoroughly described in another deliverable[27], but the way that the Access Software Platform and the E-ARK Web interact can be seen here:

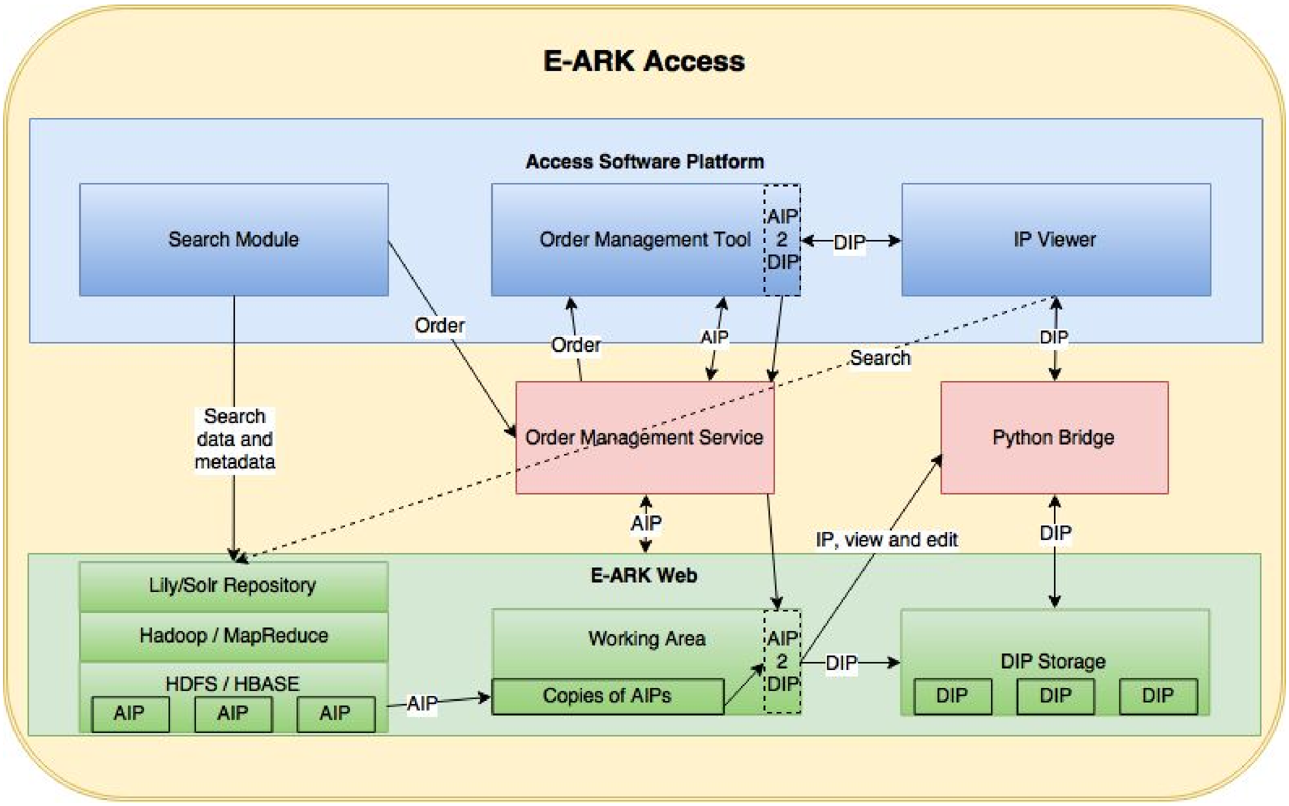


Figure 4 – Overview of the interaction between the Access Software Platform and the E-ARK Web

The **Search Module** is the user interface for searching archival material. End-users **search for metadata and data** that reside in the **Lily Repository** and that have been indexed by **Solr** (which serves as query interface as well; there is no bespoke search API). The underlying storage clusters are part of the Apache **Hadoop** open source software which secures scalable and distributed computing. The name of the computation framework is **MapReduce**.

[27] D6.2 E-ARK Integrated Platform Reference Implementation http://www.eark-project.com/resources/project-deliverables/54-d62intplatformref-1

When the end-user has identified desired archival records, she sends an **order** to the archivist who receives it via the **Order Management Service** when logging into the **Order Management Tool (OMT)**. The archivist then sends a request to retrieve the **AIP** that the end-user ordered. This is also done through the Order Management Service. This service identifies the requested records in the **HDFS / HBASE** systems and copies them onto the **Working Area**. When in the Working Area, the archivist can manipulate the records using the IP Viewer, which is accessible via the OMT, and ultimately create the **DIP** via the **AIP2DIP** component from **E-ARK Web**.

The DIP is now sent to the **DIP Storage**, and presented to the end-user (or archivist) via the **IP Viewer**. These two components communicate via the **Python Bridge** that fulfills several functions, for example the archivist's ability to **view** files and **edit** IPs (create directories; copy/paste and delete directories and files), preparing them for the end-user. The Consumer can also **search** inside the DIP via Solr's search functionality.

Orders and IP metadata are exchanged using the data-interchange format, JSON[28].

3.2 Search Module, Order Management Tool, and IP Viewer

As explained above three tools have been made available to support the overall workflow of accessing archival information.

The tools are:

1. Search Module
2. Order Management Tool
3. IP Viewer

These tools are all available in the same web user interface (UI), and the UI requires login to access. Logged in users will have different access permissions based on their role. End-users can only access IPs through the Search Module and IP Viewer if the archival institution allows it, and never has access to the Order Management Tool. Archivists have unrestricted access to all three tools.

3.2.1 User interaction overview

In a typical use case, an end-user will login to the Access Software Platform and search for available archival records. When the end-user finds one or more interesting archival records, he or she can request full access to them by creating an order for them and submit this order to the archive. He or she can now log off and wait for the order to be processed.

Later, an archivist will login to the Access Software Platform and notice the new order in a list of incoming orders. He or she will review the requested archival records and decide to make them available by creating a DIP containing them. Once the DIP is created and indexed, the original order's status is set to "ready".

As the end-user logs in to check on the progress of an order, he or she will see that the order has a new status, "ready". From the order list, selecting an individual order will display the order details, including a link to the DIP that the archivist created for the user. The end-user can now browse the requested archival records, see their details, preview them, and download them for later use.

[28] JSON, JavaScript Object Notation http://www.json.org/

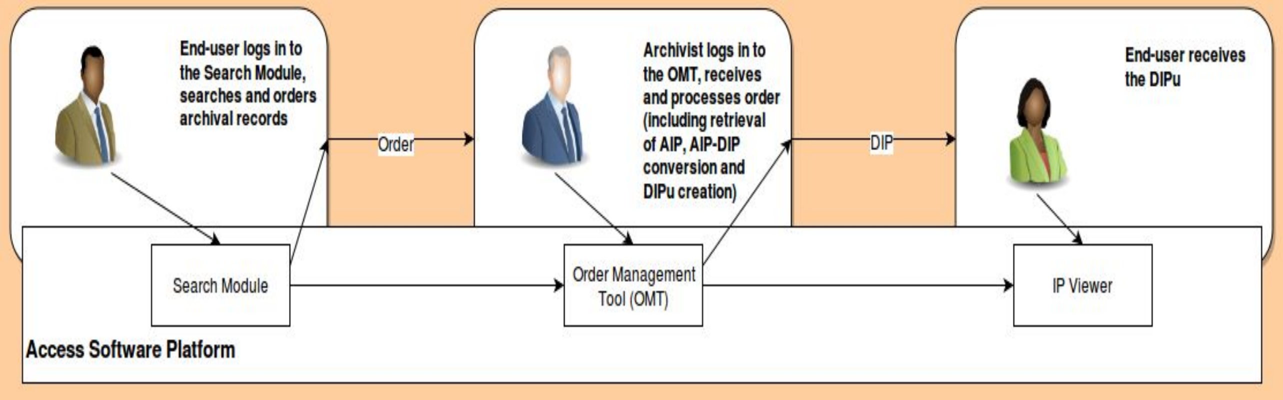

End-user logs in to the Search Module, searches and orders archival records
Order
Archivist logs in to the OMT, receives and processes order (including retrieval of AIP, AIP-DIP conversion and DIPu creation)
DIP
End-user receives the DIPu
Search Module
Order Management Tool (OMT)
IP Viewer
Access Software Platform

Figure 5 – User Interaction Overview

The DIPu is the user-oriented DIP.[29]

The Access Software Platform functionality will be described in greater detail and with accompanying screenshots in the following sections.

3.2.2 System login

First of all, when accessing the system, users will be met with a login screen and asked to type their username and password to proceed. We assume the user has been given those login credentials beforehand. This is something the archival institution needs to cater for, but it can be implemented within the ASP, cf. section 3.1.2 'User identification and authorisation'. Users must enter a username and password and hit the "LOGIN" button to get started.

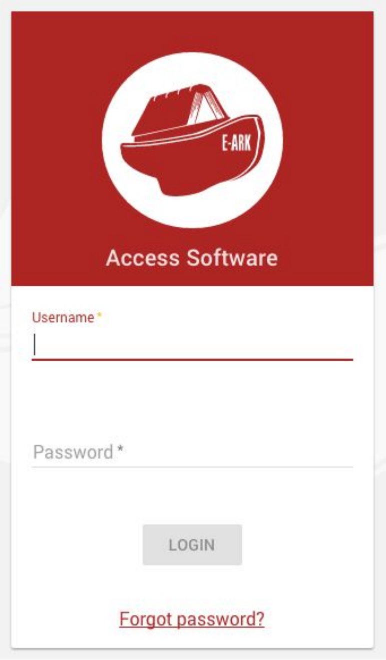
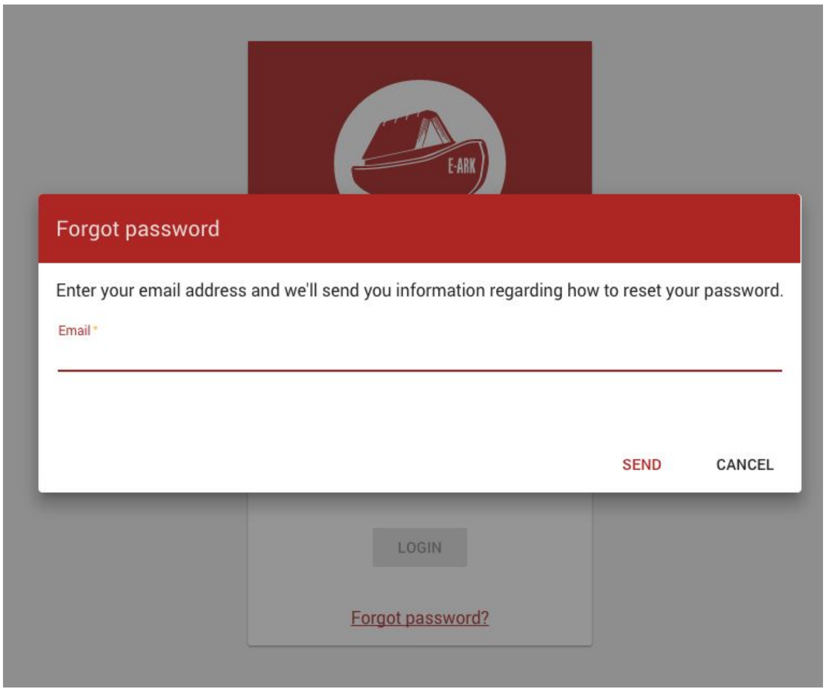


Figure 6 – The login screen

[29] The E-ARK project defines three different DIP statuses: DIP0, which is a provisional Dissemination Information Package directly derived from one or more AIPs, which may or may not be ready for use, according to the user's order and access rights; DIPu, which is a Dissemination Information Package, ready to be accessed, and previously checked against user's order and access rights; and DIPp, which is a permanent Dissemination Information Package, available to be accessed indefinitely by users due to frequent requests for the same data. The DIPp can be available on-line.

If a user forgets his or her password, he or she can click “Forgot password?” and have a new password sent via email.



Enter your email address and we'll send you information regarding how to reset your password.

Figure 7 – The “Forgot password” dialog

Logged in as an end-user, the first thing to be seen is the Search Module.

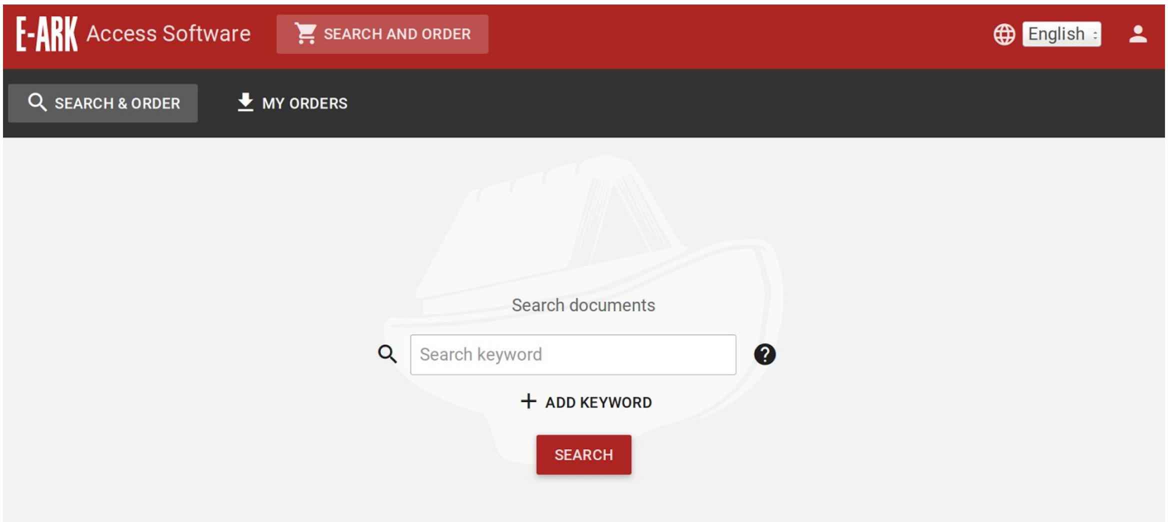
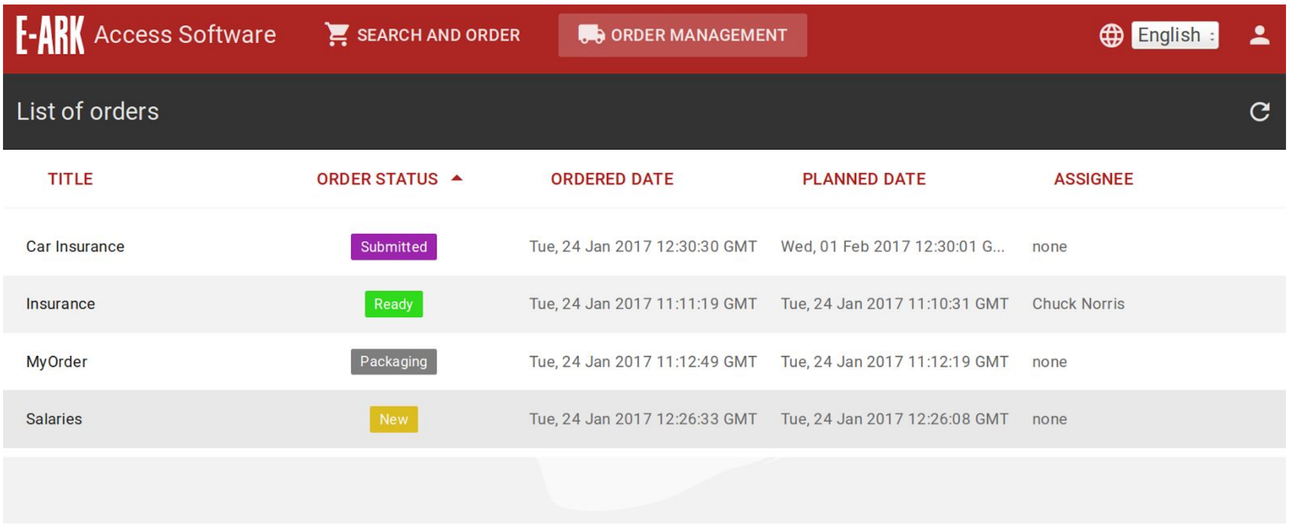


Figure 8 – The search screen

Logged in as an archivist, the first thing to be seen is the list of current orders.



List of orders

TITLE	ORDER STATUS	ORDERED DATE	PLANNED DATE	ASSIGNEE
Car Insurance	Submitted	Tue, 24 Jan 2017 12:30:30 GMT	Wed, 01 Feb 2017 12:30:01 G...	none
Insurance	Ready	Tue, 24 Jan 2017 11:11:19 GMT	Tue, 24 Jan 2017 11:10:31 GMT	Chuck Norris
MyOrder	Packaging	Tue, 24 Jan 2017 11:12:49 GMT	Tue, 24 Jan 2017 11:12:19 GMT	none
Salaries	New	Tue, 24 Jan 2017 12:26:33 GMT	Tue, 24 Jan 2017 12:26:08 GMT	none

Figure 9 – The order list

3.2.3 Global navigation

The top bars on the screen are for navigation and various global tasks. The red bar in Fig. 3 allows for switching between modes respectively for searching and managing archival material. The "SEARCH AND ORDER" will thus lead to the module for searching and ordering, while the "ORDER MANAGEMENT" will lead to the order list and processing tools that are for archivists only. End-users will only have the "SEARCH AND ORDER" link available.

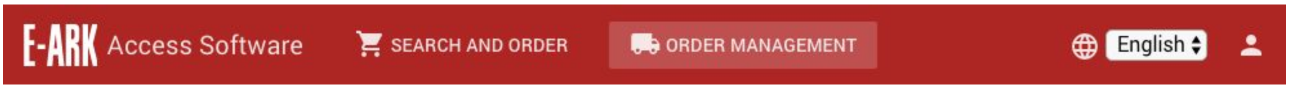


Figure 10 – Top bar for main navigation, logout, and language selection available for archivists

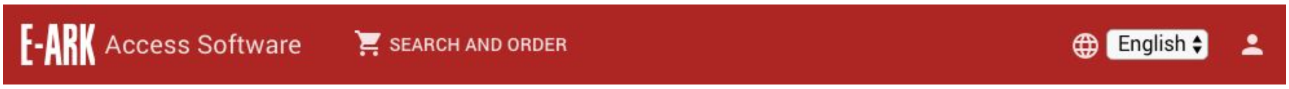


Figure 11 – End-users have fewer options available

The top bar also contains a language selector and a user action button. The language selector is used to select one of the languages installed to be used throughout the UI. Language files are available, and new languages can be added very easily.

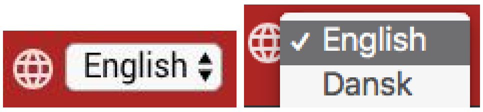


Figure 12 – The language selector

The user actions button at the far right displays the currently logged in user's name, and enables users to logout from the system.

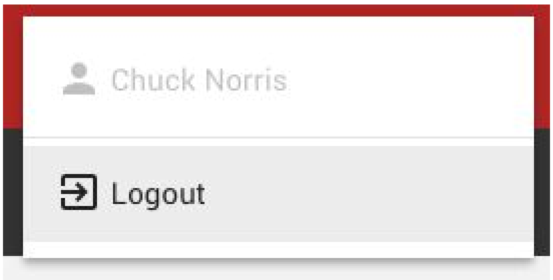


Figure 13 – Logging out

Generally, if there is a button with an arrow facing left in the dark bar below the red bar, it enables the user to go back to the previous page or an overview page related to the current page.

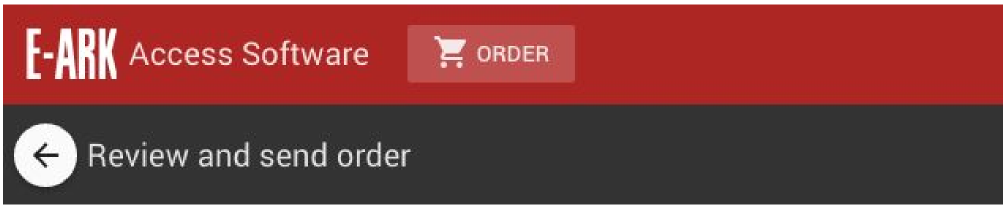


Figure 14 – The "back" button that is displayed in many screens

3.2.4 Search Module, and ordering

The Search Module enables users to search for records that can potentially be made available and order them for later review. Features include:

- Search interface for documents with Solr combined search features (AND, OR, NOT). Users can view available metadata for archival records in the search result list.
- Picking out archival records from a search result and collecting them into an order using a shopping basket feature. Once collected, orders can be sent along with some comments to the archivists.
- A list of current and previous orders. Users are able to see the status of individual orders using this list.

As of now it is possible to identify to which IP an archival record belongs by inspecting its metadata and viewing the pertaining IP's ID. It is also possible to identify its relative place - to which descriptive unit it belongs - in an IP's hierarchy (if applicable) by inspecting the c-level[30] of the archival record in question. The prerequisite is that descriptive metadata is available.

What follows is a step by step introduction to the use of the features listed above.

3.2.4.1 Initial search

End-users are taken directly to the Search Module when logged in. A single search input is the first thing they'll notice. This carries out a full-text search across all indexed content and metadata stored within the repository.

[30] EAD3 <c> http://www.loc.gov/ead/EAD3taglib/#elem-c

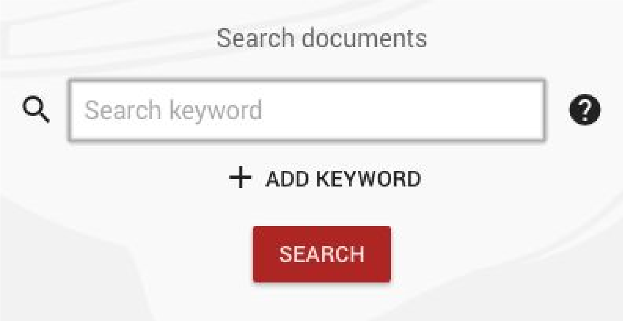
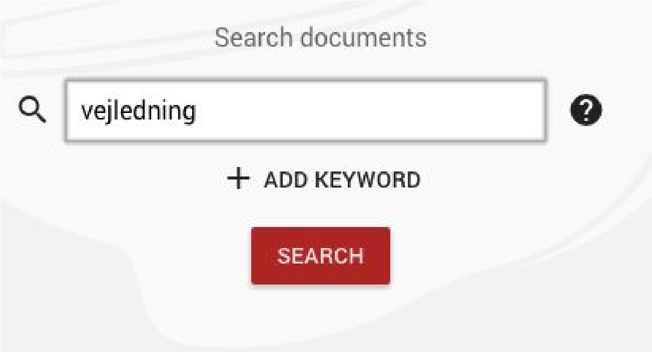


Figure 15 – Search input field

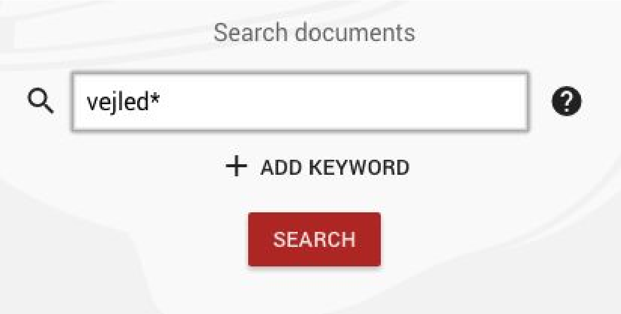
Any keyword can be entered into the field.



vejledning

Figure 16 – Entering keyword into search input field

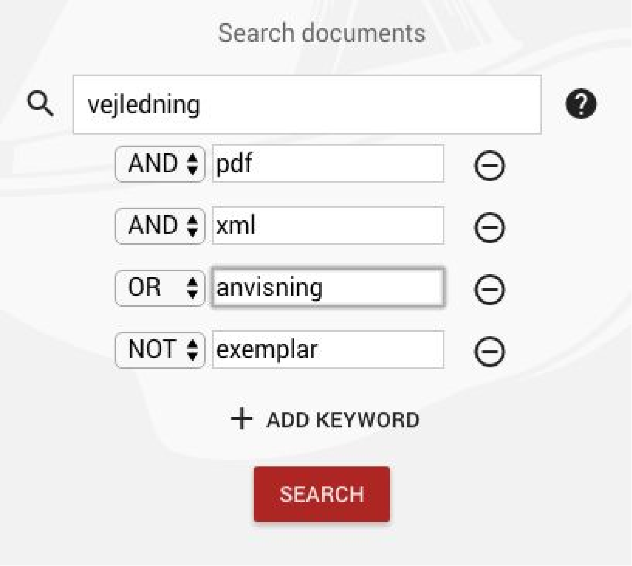
The “*” wildcard can be used when only a part of a keyword is known. Searching for “vejledning” will return results containing the word “vejledning”. Searching for “vejled*” will return results containing either for example “vejledning” or “vejleder”.



vejled*

Figure 17 – Using the * wildcard in search

More keywords can be added using the “add keyword” button. It is possible to refine the search by defining whether extra keywords should be mandatory (AND), optional (OR), or decrease the number of results (NOT).



vejledning

AND pdf

AND xml

OR anvisning

NOT exemplar

Figure 18 – Adding additional search terms

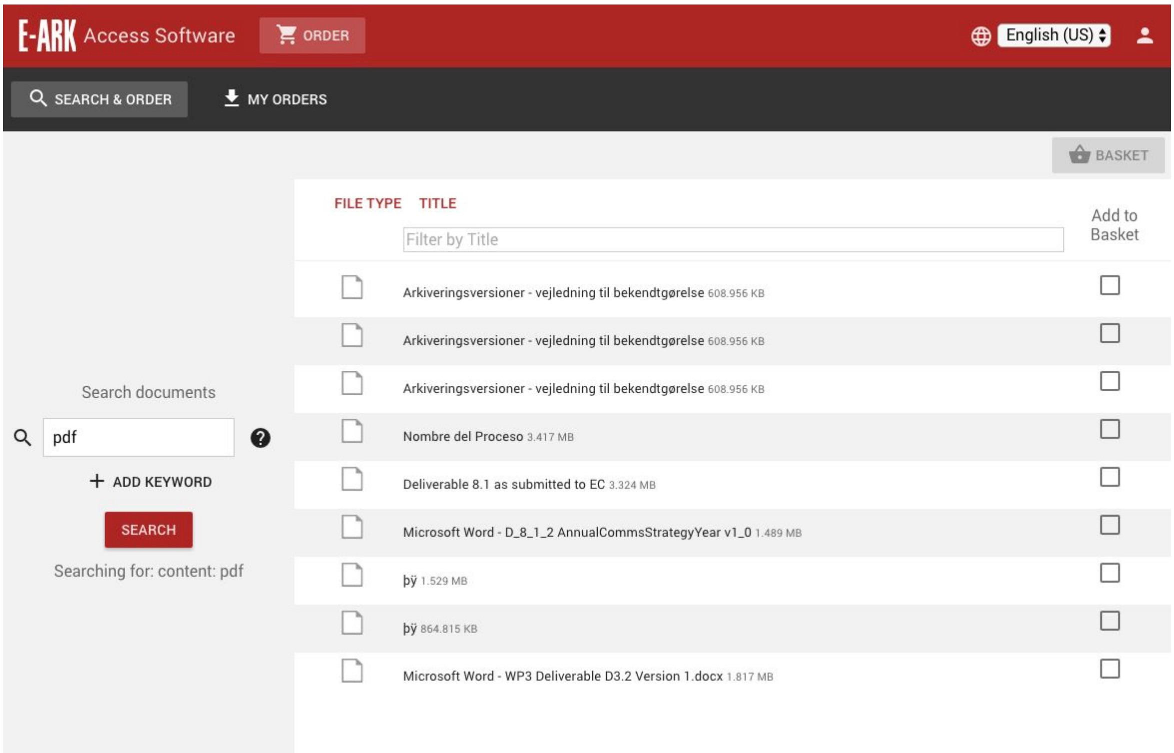
Remove keywords by clicking the "-" icon to the right of each input field.

Clicking the "Search" button afterwards will initiate the search. If the search produces any search results, they will be displayed on the screen.

Queries from the Search Module are made directly through the Solr query API defined and used for E-ARK Web. The back-end is described thoroughly elsewhere[31]. The front-end enables the execution of the back-end's powerful Apache Solr search functionality. Technologically, the same search is provided both in the Search Module and in the IP Viewer, so that the end-user can perform full text searches not only in the archive's repository, but also inside the DIP that he or she has requested.

[31] D6.2 E-ARK Integrated Platform Reference Implementation http://www.eark-project.com/resources/project-deliverables/54-d62intplatformref-1

3.2.4.2 Refined search



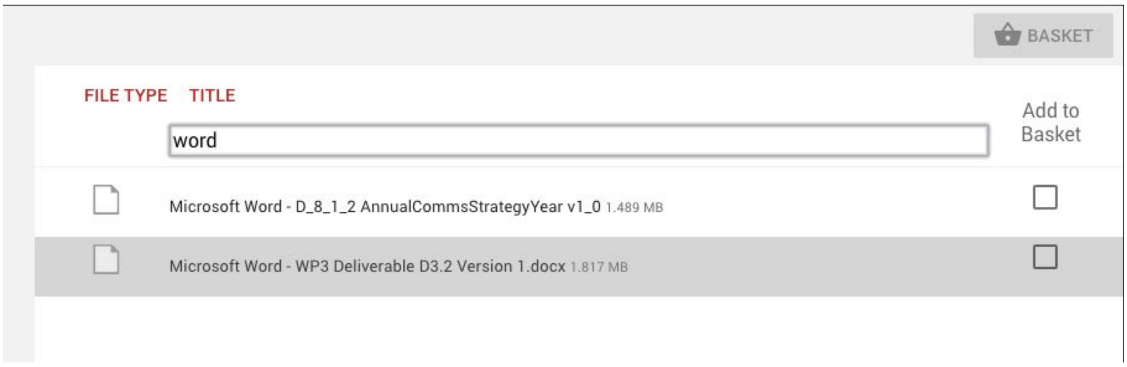
pdf

Searching for: content: pdf

FILE TYPE	TITLE	Add to Basket
	Filter by Title	
	Arkiveringsversioner - vejledning til bekendtgørelse 608.956 KB	□
	Arkiveringsversioner - vejledning til bekendtgørelse 608.956 KB	□
	Arkiveringsversioner - vejledning til bekendtgørelse 608.956 KB	□
	Nombre del Proceso 3.417 MB	□
	Deliverable 8.1 as submitted to EC 3.324 MB	□
	Microsoft Word - D_8_1_2 AnnualCommsStrategyYear v1_0 1.489 MB	□
	þÿ 1.529 MB	□
	þÿ 864.815 KB	□
	Microsoft Word - WP3 Deliverable D3.2 Version 1.docx 1.817 MB	□

Figure 19 – The search result screen

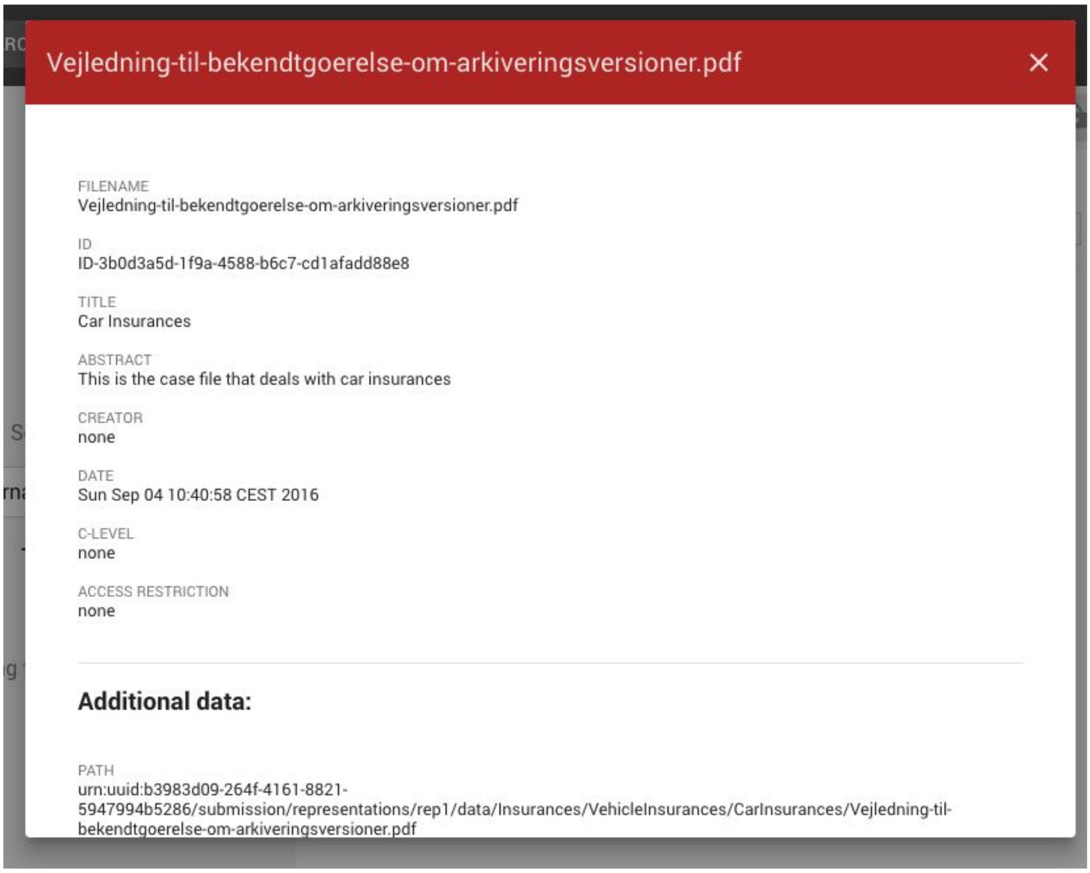
If there is an overwhelming number of search results, the returned result can be filtered using the “filter by title” input.



FILE TYPE	TITLE	Add to Basket
	word	
	Microsoft Word - D_8_1_2 AnnualCommsStrategyYear v1_0 1.489 MB	□
	Microsoft Word - WP3 Deliverable D3.2 Version 1.docx 1.817 MB	□

Figure 20 – Filtering search results

Clicking an item will display additional metadata about the archival record. There is currently no way of viewing a file at this stage. Viewing files is possible in the IP Viewer, which is accessible both for the archivist (from within the Order Management Tool (OMT) and for the end-user. However, just as the IP Viewer is integrated into the OMT, file viewing functionality from the IP Viewer can also be integrated into the Search Module, which has not been enabled yet.



Vejledning-til-bekendtgoerelse-om-arkiveringsversioner.pdf ×

FILENAME
Vejledning-til-bekendtgoerelse-om-arkiveringsversioner.pdf

ID
ID-3b0d3a5d-1f9a-4588-b6c7-cd1afadd88e8

TITLE
Car Insurances

ABSTRACT
This is the case file that deals with car insurances

CREATOR
none

DATE
Sun Sep 04 10:40:58 CEST 2016

C-LEVEL
none

ACCESS RESTRICTION
none

Additional data:

PATH
urn:uuid:b3983d09-264f-4161-8821-5947994b5286/submission/representations/rep1/data/Insurances/VehicleInsurances/CarInsurances/Vejledning-til-bekendtgoerelse-om-arkiveringsversioner.pdf

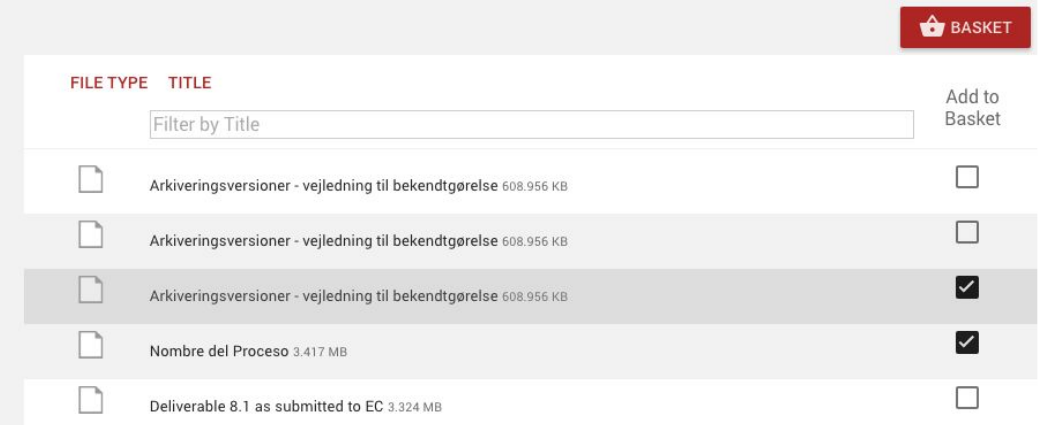
Figure 21 – The document information dialog

Click “x” in the upper right corner or anywhere outside the dialog to close it[32].

Now the user can search through the archival records that are displayed in the result list and get basic information about them before deciding if he or she wishes to order some of them for full access.

3.2.4.3 Ordering archival records

Using the checkboxes on the right hand side of the search results, one or more archival records can be selected for ordering. When items are selected, the “BASKET” button becomes active and lets the user proceed to the ordering step.

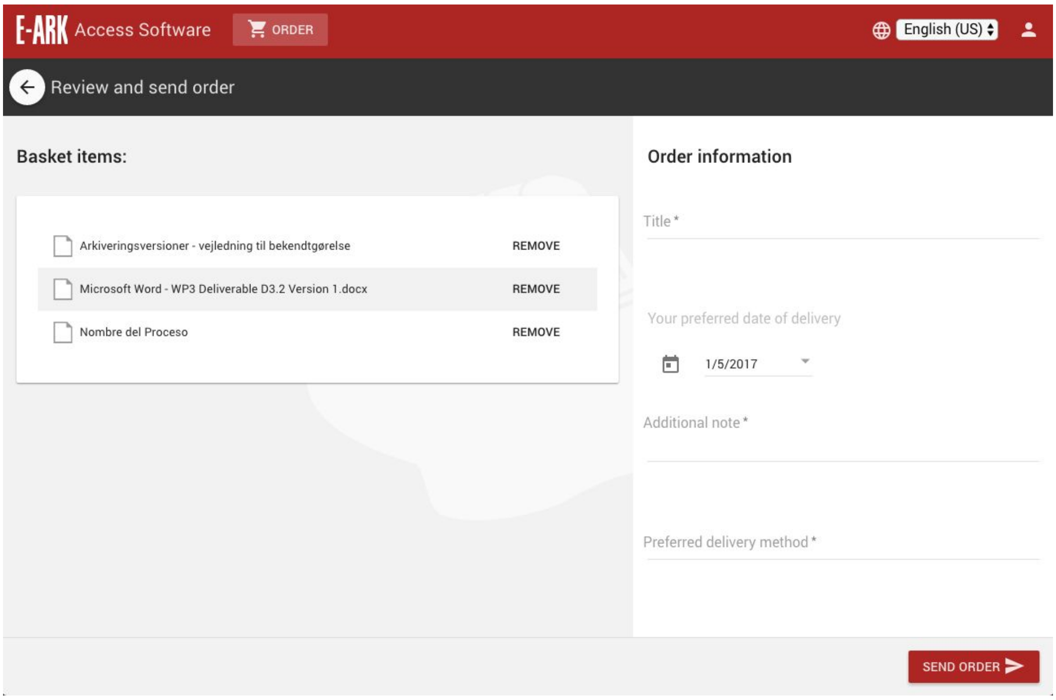


FILE TYPE	TITLE Filter by Title	Add to Basket
	Arkiveringsversioner - vejledning til bekendtgørelse 608.956 KB	□
	Arkiveringsversioner - vejledning til bekendtgørelse 608.956 KB	□
	Arkiveringsversioner - vejledning til bekendtgørelse 608.956 KB	☑
	Nombre del Proceso 3.417 MB	☑
	Deliverable 8.1 as submitted to EC 3.324 MB	□

[32] Note that the ‘ABSTRACT’ mentions ‘case file’ where it should have said ‘file’. This shows the importance of correct descriptive metadata.

Figure 22 – Selecting multiple items for ordering

Clicking the “BASKET” button takes the user to the ordering page.



Basket items:

- Arkiveringsversioner - vejledning til bekendtgørelse REMOVE
- Microsoft Word - WP3 Deliverable D3.2 Version 1.docx REMOVE
- Nombre del Proceso REMOVE

Order information

Title *

Your preferred date of delivery: 1/5/2017

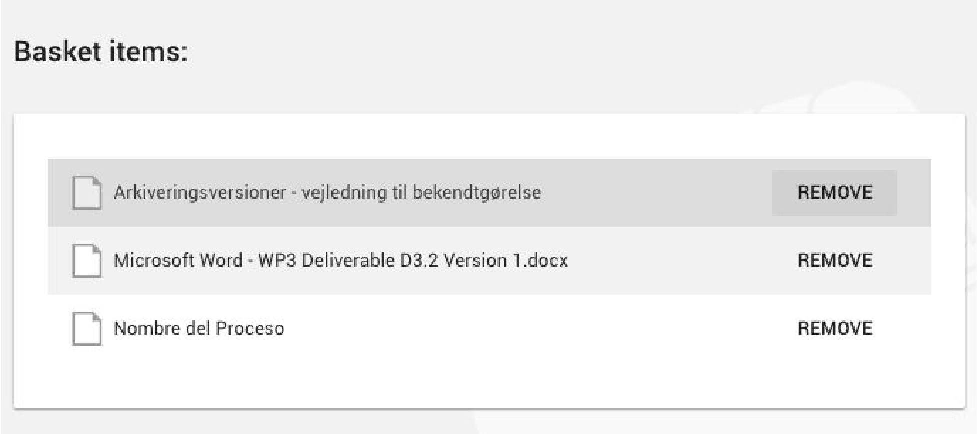
Additional note *

Preferred delivery method *

Figure 23 – The ordering screen (basket)

The ordering page presents an overview of the archival records that are about to be ordered (“Basket items”), some input fields that are required for the order (“Order information”), and a “SEND ORDER” button at the bottom of the screen.

Clicking “REMOVE” to the right of a listed archival records will remove it from the list.


Basket items:

- Arkiveringsversioner - vejledning til bekendtgørelse REMOVE
- Microsoft Word - WP3 Deliverable D3.2 Version 1.docx REMOVE
- Nombre del Proceso REMOVE

Figure 24 – Basket items

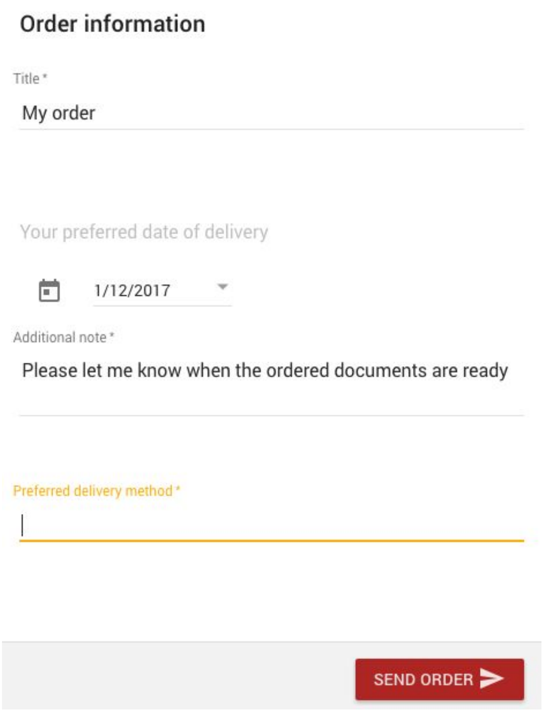
Alternatively, it is possible to use the back arrow in the dark top bar to navigate back to the search results and select (or deselect) additional archival records.

In the order information column, it is possible to enter:

- A title for the order

- A due date for when the end-users would like the order to be ready. This helps the archive organise their work.
- A comment regarding the order. The end-user can state the purpose of the request to obtain access to the archival records
- A comment regarding the preferred way of receiving the ordered archival records.

All the order information fields are mandatory.



Order information

Title *
My order

Your preferred date of delivery
1/12/2017

Additional note *
Please let me know when the ordered documents are ready

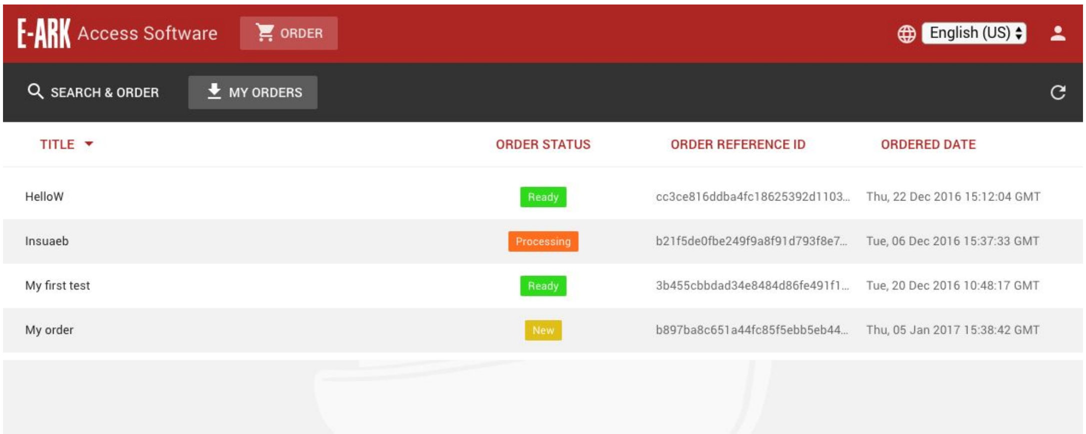
Preferred delivery method *

Figure 25 – Entering ordering information

Clicking the “SEND ORDER” button will notify the end-user that the order was sent and transfer the end-user to the “MY ORDERS” screen. The order just sent will be displayed with a status of “new”.

3.2.4.4 Orders overview

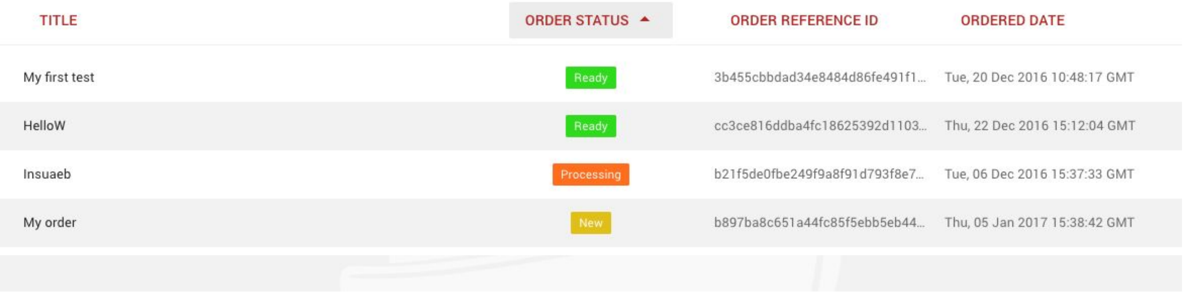
The “My orders” screen displays an overview of orders created by the current user and can be accessed by clicking the “MY ORDERS” link in the dark top bar when in the “ORDER” section.



TITLE	ORDER STATUS	ORDER REFERENCE ID	ORDERED DATE
HelloW	Ready	cc3ce816ddba4fc18625392d1103...	Thu, 22 Dec 2016 15:12:04 GMT
Insuaeb	Processing	b21f5de0fbe249f9a8f91d793f8e7...	Tue, 06 Dec 2016 15:37:33 GMT
My first test	Ready	3b455cbbdad34e8484d86fe491f1...	Tue, 20 Dec 2016 10:48:17 GMT
My order	New	b897ba8c651a44fc85f5ebb5eb44...	Thu, 05 Jan 2017 15:38:42 GMT

Figure 26 – The order list screen ("My orders")

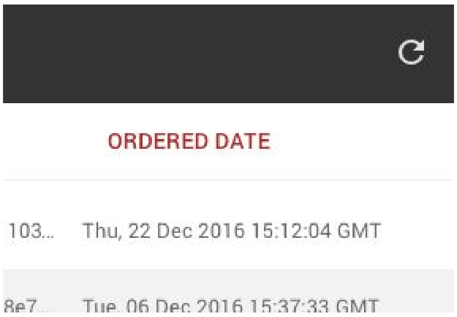
The order list can be sorted in various ways by clicking the column titles. It is helpful to sort the list according to order status.



TITLE	ORDER STATUS ▲	ORDER REFERENCE ID	ORDERED DATE
My first test	Ready	3b455cbbdad34e8484d86fe491f1...	Tue, 20 Dec 2016 10:48:17 GMT
HelloW	Ready	cc3ce816ddba4fc18625392d1103...	Thu, 22 Dec 2016 15:12:04 GMT
Insuaeb	Processing	b21f5de0fbe249f9a8f91d793f8e7...	Tue, 06 Dec 2016 15:37:33 GMT
My order	New	b897ba8c651a44fc85f5ebb5eb44...	Thu, 05 Jan 2017 15:38:42 GMT

Figure 27 – Sorting the order list

There is a refresh button in the right corner of the dark top bar that reloads the list of orders when clicked.



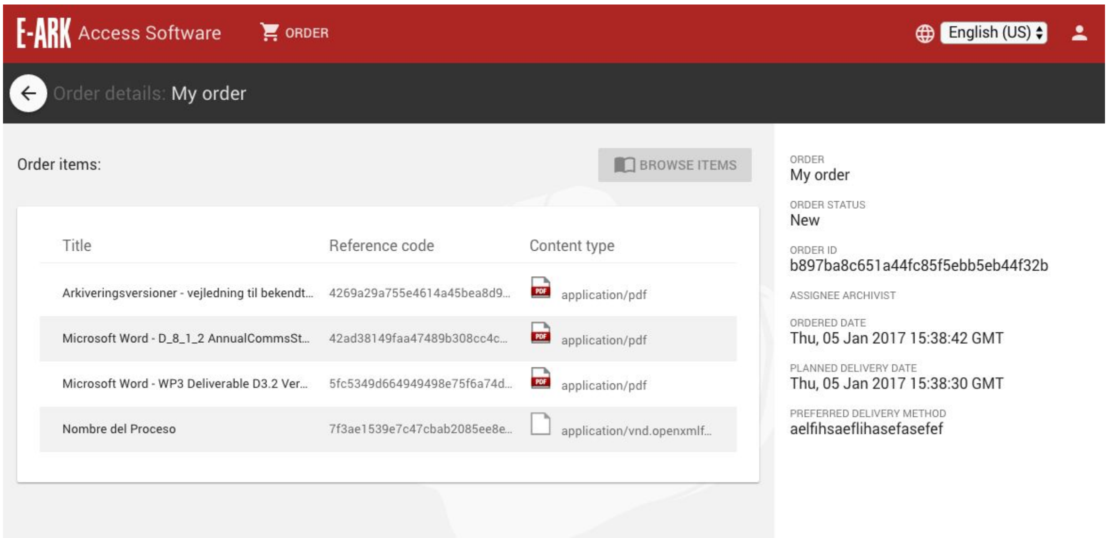
	ORDERED DATE
103...	Thu, 22 Dec 2016 15:12:04 GMT
8e7...	Tue, 06 Dec 2016 15:37:33 GMT

Figure 28 – "Refresh" icon for refreshing the order list

This can be helpful when the end-user wants to quickly check for status changes for his or hers orders.

3.2.4.5 Viewing a single order

Clicking an order in the list leads to the order detail page. This page contains information about the order such as archival records ordered, status and delivery dates. The option to browse the ordered archival records ("Browse items") is disabled as long as the order is still being processed by the archive. When the order is fulfilled, the option is activated and the end-user can start looking at the ordered files.



Order details: My order

Order items:

Title	Reference code	Content type
Arkiveringsversioner - vejledning til bekendt...	4269a29a755e4614a45bea8d9...	application/pdf
Microsoft Word - D_8_1_2 AnnualCommsSt...	42ad38149faa47489b308cc4c...	application/pdf
Microsoft Word - WP3 Deliverable D3.2 Ver...	5fc5349d664949498e75f6a74d...	application/pdf
Nombre del Proceso	7f3ae1539e7c47cbab2085ee8e...	application/vnd.openxmlf...

ORDER
My order

ORDER STATUS
New

ORDER ID
b897ba8c651a44fc85f5ebb5eb44f32b

ASSIGNEE ARCHIVIST

ORDERED DATE
Thu, 05 Jan 2017 15:38:42 GMT

PLANNED DELIVERY DATE
Thu, 05 Jan 2017 15:38:30 GMT

PREFERRED DELIVERY METHOD
aelfihsaeflihasefasefef

Figure 29 – The order details screen

Clicking any row in the list of ordered archival records ("Order items") will display a dialog with more information on the archival record. The 'Additional note' that the end-user added to the original order, cf. section 3.2.4.3 Ordering archival records, is currently not displayed, but should be.

If the end-user wishes to go back to the 'search' page, (s)he can do it either by clicking 'Search & Order' in the top menu, or by clicking the '⟵' and then the 'SEARCH & ORDER' button.

3.2.5 Order Management Tool

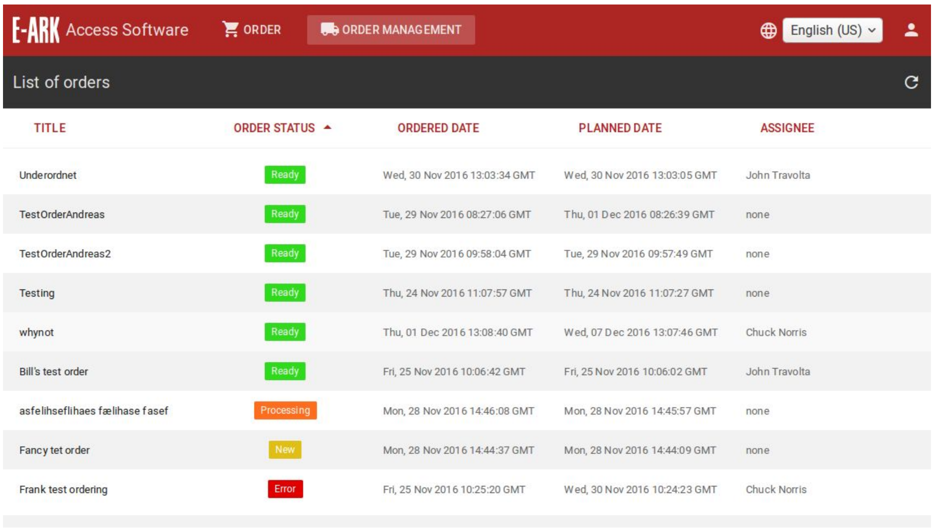
The Order Management Tools sits in the center of the Access workflow, and it is from this tool that the archivist is able to control the flow of IPs and respond to end-user order requests: The Order Management Tool enables archivists to review incoming orders, package IPs containing archival records from those orders, and make the corresponding DIPs available for end-users. Features include:

- A list of orders from end-users. The list displays the status of individual orders and can be sorted in various ways.
- Detailed view of individual orders. Ordered archival records are listed here and archivists can carry out various actions on the order.
- A feature to "process" orders, ie. build and index new IPs containing the requested archival records.
- Once "pre-processed", an order can be manually customised by the archivist by copying and moving files in the IP.
- Automatic update of order status once an DIP is made available to the end-user. This tells the end-user that an order is ready for viewing.

What follows is a step by step introduction to the use of those features.

3.2.5.1 Listing new and existing orders

When a user with "archivist" status logs in, he or she will be presented with the list of current orders in the "Order Management" section of the Access Software Platform's site.



List of orders ↻

TITLE	ORDER STATUS ▲	ORDERED DATE	PLANNED DATE	ASSIGNEE
Underordnet	Ready	Wed, 30 Nov 2016 13:03:34 GMT	Wed, 30 Nov 2016 13:03:05 GMT	John Travolta
TestOrderAndreas	Ready	Tue, 29 Nov 2016 08:27:06 GMT	Thu, 01 Dec 2016 08:26:39 GMT	none
TestOrderAndreas2	Ready	Tue, 29 Nov 2016 09:58:04 GMT	Tue, 29 Nov 2016 09:57:49 GMT	none
Testing	Ready	Thu, 24 Nov 2016 11:07:57 GMT	Thu, 24 Nov 2016 11:07:27 GMT	none
whynot	Ready	Thu, 01 Dec 2016 13:08:40 GMT	Wed, 07 Dec 2016 13:07:46 GMT	Chuck Norris
Bill's test order	Ready	Fri, 25 Nov 2016 10:06:42 GMT	Fri, 25 Nov 2016 10:06:02 GMT	John Travolta
asfelihseflihaes fælihase fasef	Processing	Mon, 28 Nov 2016 14:46:08 GMT	Mon, 28 Nov 2016 14:45:57 GMT	none
Fancy tet order	New	Mon, 28 Nov 2016 14:44:37 GMT	Mon, 28 Nov 2016 14:44:09 GMT	none
Frank test ordering	Error	Fri, 25 Nov 2016 10:25:20 GMT	Wed, 30 Nov 2016 10:24:23 GMT	Chuck Norris

Figure 30 – The order list for archivists

Orders can be sorted by “Title”, “Order status”, or “Assignee” by clicking the column headers. An order might change its status while one is looking at the list but the list does not automatically update. Clicking the “refresh” icon in the right corner of the dark bar will refresh the list so any changes made to orders become visible.

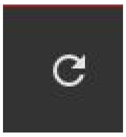


Figure 31 – “Refresh” button

The order statuses help the archivist identify where in the process an order is:

New This status informs the archivist about a newly created order.

Submitted This status is displayed after the archivist hits the ‘PROCESS ORDER” button which initiates a series of automatic steps that run in the background (E-ARK Web). These automatic steps copy the AIP to to Working Area (cf. Figure 4 above) and untar the package so that the archivist can start modifying it. Remember that to reach the next step, “Processing”, it is necessary to manually update the page by clicking the ↻ button.

Processing This status informs the archivist that she can go into the newly copied package and start creating the final DIPu. To see how a DIPu is finalized, refer to section 3.2.5.6 Copying data and metadata to the DIP.

Packaging This status is displayed while the finalized DIPu is being packaged for the end-user. Again click the ↻ button to proceed.

Indexing This status is displayed while the finalized DIPu is being indexed so that it becomes searchable from within the IP Viewer. Again click the ⟳ button to proceed.

Ready When the DIPu is ready to be sent to the end-user, this status will appear.

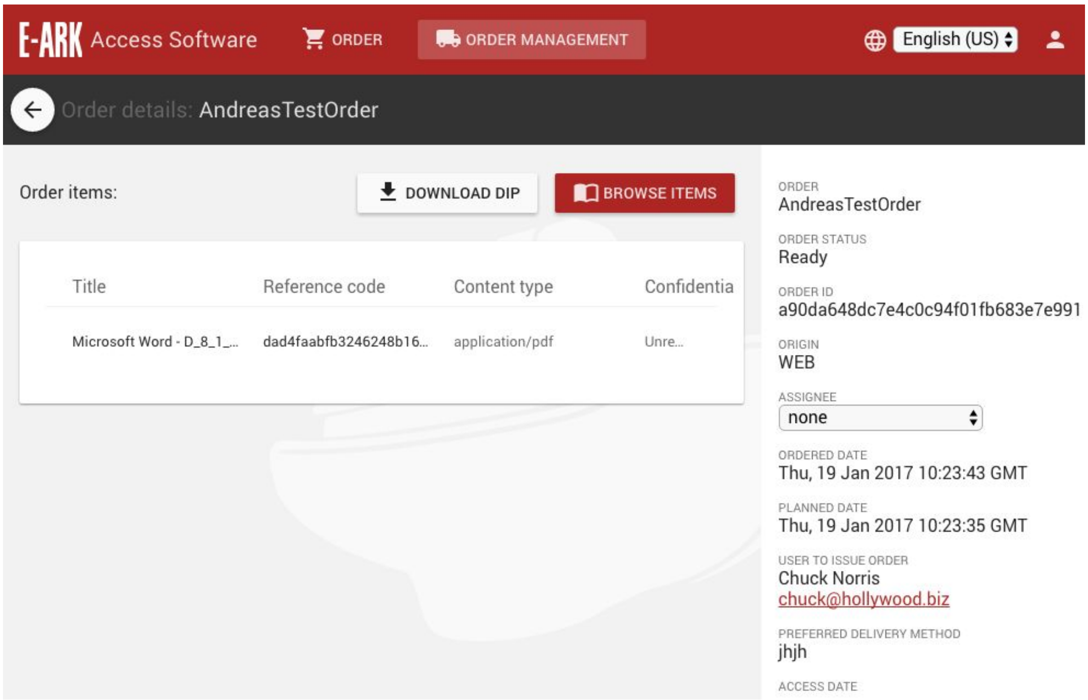
Error This status appears when the system encounters an error. Note also that if the ⟳ button is clicked too quickly, then the error status may also appear. In this case, click the ⟳ button again.

Clicking any row in the list leads to the details page for a specific order.

Archives can chose to assign specific archivists to specific orders. This can help them organize their work. It is not mandatory, however, and should not be as it depends on local policies. Per default all archivists have the same rights and can thus view all the orders regardless of who is assigned to what. It is possible to sort alphabetically to get an overview of the orders assigned to each archivist.

3.2.5.2 Viewing order details

The order details page displays a list of ordered items and some metadata about the order.



Order details: AndreasTestOrder

Order items:

Title	Reference code	Content type	Confidentia
Microsoft Word - D_8_1_...	dad4faabfb3246248b16...	application/pdf	Unre...

ORDER
AndreasTestOrder

ORDER STATUS
Ready

ORDER ID
a90da648dc7e4c0c94f01fb683e7e991

ORIGIN
WEB

ASSIGNEE
none

ORDERED DATE
Thu, 19 Jan 2017 10:23:43 GMT

PLANNED DATE
Thu, 19 Jan 2017 10:23:35 GMT

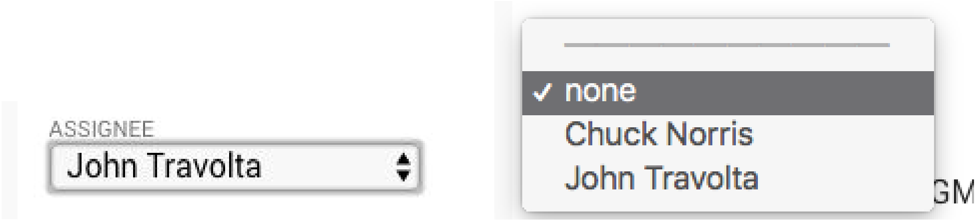
USER TO ISSUE ORDER
Chuck Norris
chuck@hollywood.biz

PREFERRED DELIVERY METHOD
jhjh

ACCESS DATE

Figure 32 – The order detail page

In the metadata section it is possible to choose an assignee to be responsible for handling the order.

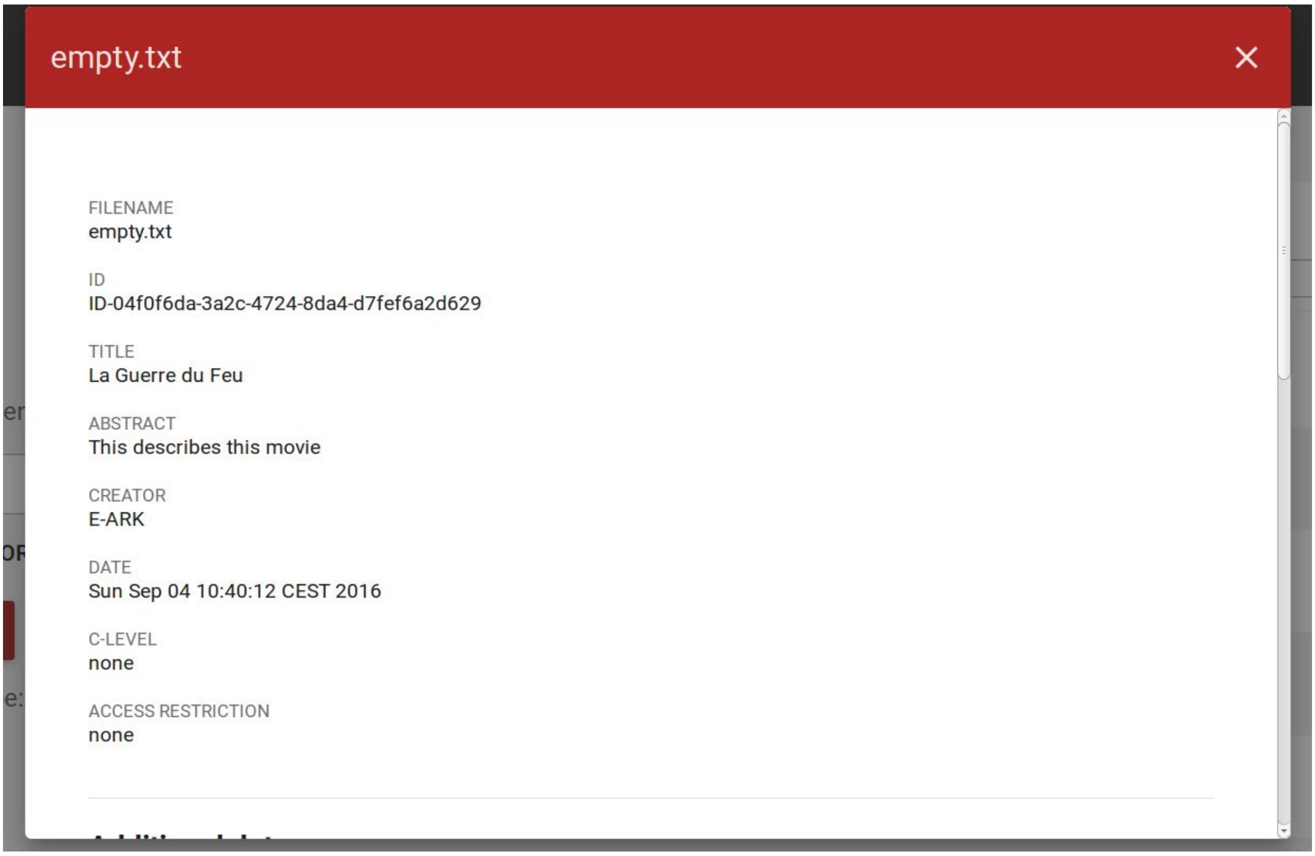


ASSIGNEE
John Travolta

Figure 33 – Changing an order’s assignee

Assigning an assignee does not affect the order in any way. It is a handy tool to signal to other archivists that there is someone already taking care of a specific order. When looking at the order list, the assignees for each order can be seen.

Clicking an item in the “Ordered items” list will display a dialog with some additional information about the requested document.



empty.txt

FILENAME
empty.txt

ID
ID-04f0f6da-3a2c-4724-8da4-d7fef6a2d629

TITLE
La Guerre du Feu

ABSTRACT
This describes this movie

CREATOR
E-ARK

DATE
Sun Sep 04 10:40:12 CEST 2016

C-LEVEL
none

ACCESS RESTRICTION
none

Figure 34 - The item information dialog

3.2.5.3 Creating a new package from an order

When an end-user requests some archival records from the archive, the archivist reviews the order and creates a dissemination information package (DIP) for the end-user to access. Creating a DIP involves a couple of steps for the archivist. The steps involved are:

- Decide to process order
- Create pre-DIP from AIP repository (copy of the original AIP)
- Copying relevant data and metadata to the future DIPu

- Package DIP

3.2.5.4 Deciding to process an order

Glancing over the “Order items” list and the order details on the order detail page gives the archivist an indication whether the order should be processed or not. Some items might be confidential and though end-users might be able to view the item’s metadata, they cannot be given access to the actual item. Make sure by clicking the item and looking up the “Access restriction” attribute.

ACCESS RESTRICTION
Unrestricted

Figure 35 – the data pane will display information regarding access restrictions

In a scenario where all the requested items are restricted or the IP needs further processing because it cannot be immediately viewed (e.g. a database contained in a SIARD file), the order will be processed by an archivist in the Order Management Tool.

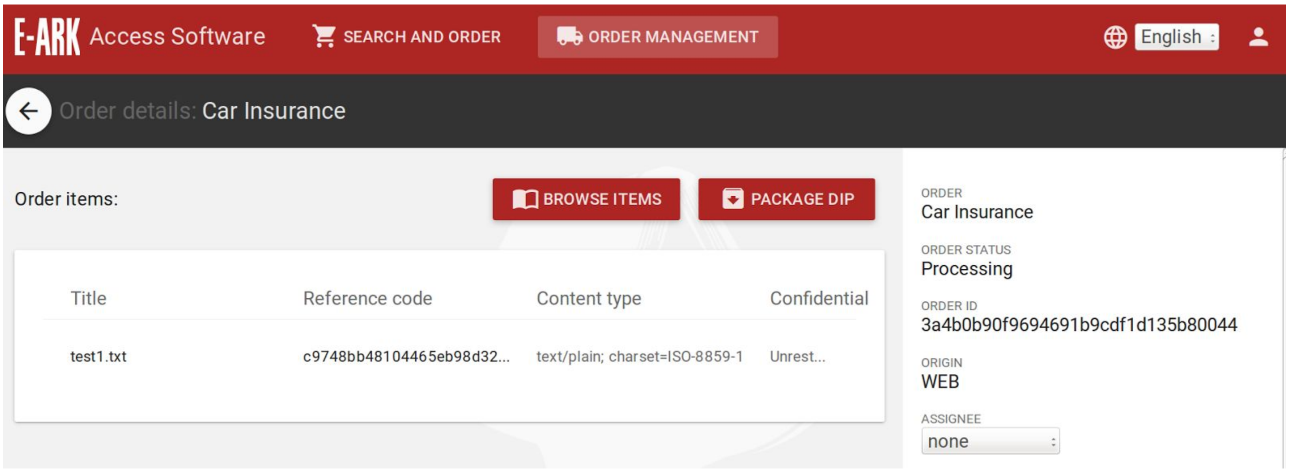
3.2.5.5 Create a pre-DIP

Depending on the current status of the order, action buttons for every step in the package creation process are available from the order detail page.

When an order has status “new”, archivists will have the option of initiating pre-DIP processing by clicking the “PROCESS ORDER” button.

Figure 36 – The “PROCESS ORDER” button

Clicking this button will copy the requested AIP and copy it, creating a pre-DIP ready for customization and packaging. The order’s status will change to “processing” while processing takes place. Within a couple of minutes depending on the size of the order, processing should finish and there will be some new options available for the archivist: the “BROWSE ITEMS” and “PACKAGE DIP” buttons.



Order details: Car Insurance

Order items:

Title	Reference code	Content type	Confidential
test1.txt	c9748bb48104465eb98d32...	text/plain; charset=ISO-8859-1	Unrest...

ORDER
Car Insurance

ORDER STATUS
Processing

ORDER ID
3a4b0b90f9694691b9cdf1d135b80044

ORIGIN
WEB

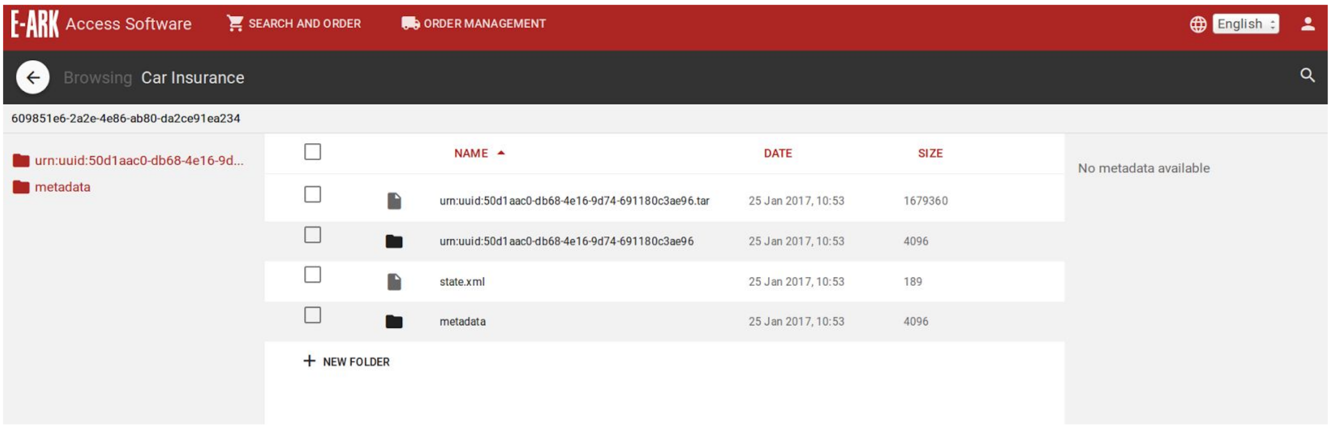
ASSIGNEE
none

Figure 37 – The “BROWSE ITEMS” and “PACKAGE DIP” buttons

When an order is in the “PROCESS ORDER” stage, it also means that the editing features will be available to the archivist when clicking the “Browse items” button. Thus the archivist can manually edit the pre-DIP and create a user-specific DIPu.

3.2.5.6 Copying data and metadata to the DIP

When the initial processing is done, a pre-DIP (copy of the corresponding AIP) will be available. It can be accessed from the order detail page by clicking the “BROWSE ITEMS” button. This will open the IP Viewer in edit mode so the IP can be customized:



609851e6-2a2e-4e86-ab80-da2ce91ea234

□		NAME	DATE	SIZE
□		urn:uuid:50d1aac0-db68-4e16-9d74-691180c3ae96.tar	25 Jan 2017, 10:53	1679360
□		urn:uuid:50d1aac0-db68-4e16-9d74-691180c3ae96	25 Jan 2017, 10:53	4096
□		state.xml	25 Jan 2017, 10:53	189
□		metadata	25 Jan 2017, 10:53	4096

Figure 38 – IP Viewer in edit mode

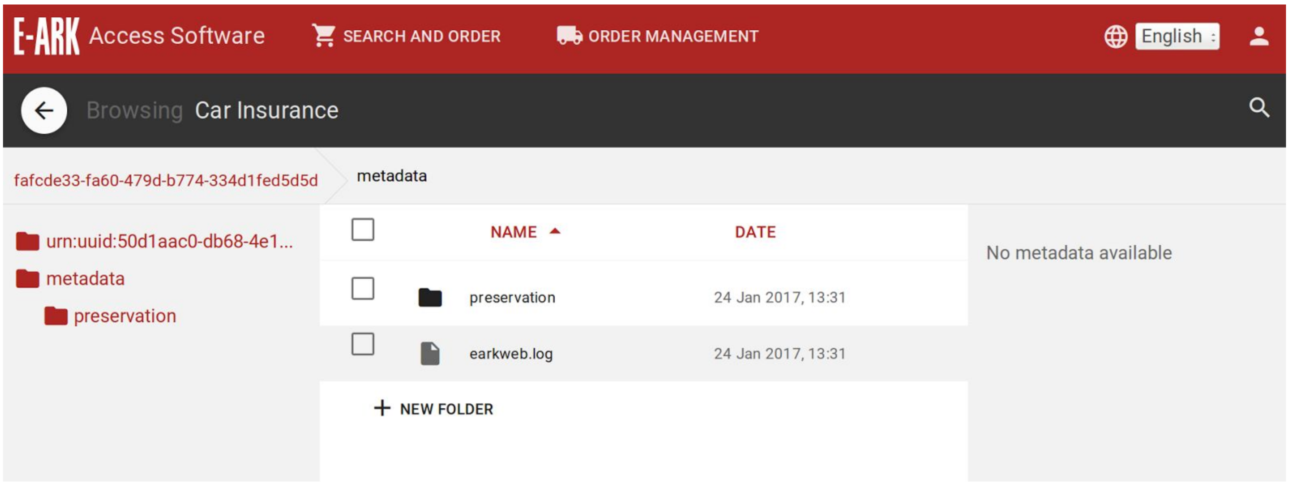
This view represents the archivists view of an IP in the IP Viewer. The view is a ‘technical’ view, and it allows for the archivist to perform specialized tasks that are necessary to create the DIPu. It also allows him to inspect which automatic tasks have been performed on the IP by E-ARK Web. By the end of the DIP creation process, the DIP will be cleansed off of all ‘technicalities’ (i.e. log files, etc.), and the end-user will be presented with relevant information only (cf. section 3.2.6 IP Viewer).

The pre-DIP is delivered as both a TAR file (first item in figure 38, ‘urn:uuid [...]) and in a folder structure (second item), which complies with the E-ARK IP structure. The contents of the two items are identical.

The TAR file can be exported, or simply used as backup. The folder version of the IP (second item) is the one the archivists works with when creating the DIP. We will refer to this as “source IP folder” in the following sections.

The “state.xml” is a log file that indicates which machine actionable task was the last one performed on the IP.

Lastly, the metadata folder includes preservation metadata (PREMIS) that can be viewed, and this folder also includes a log file from the E-ARK Web:



fafcde33-fa60-479d-b774-334d1fed5d5d › metadata

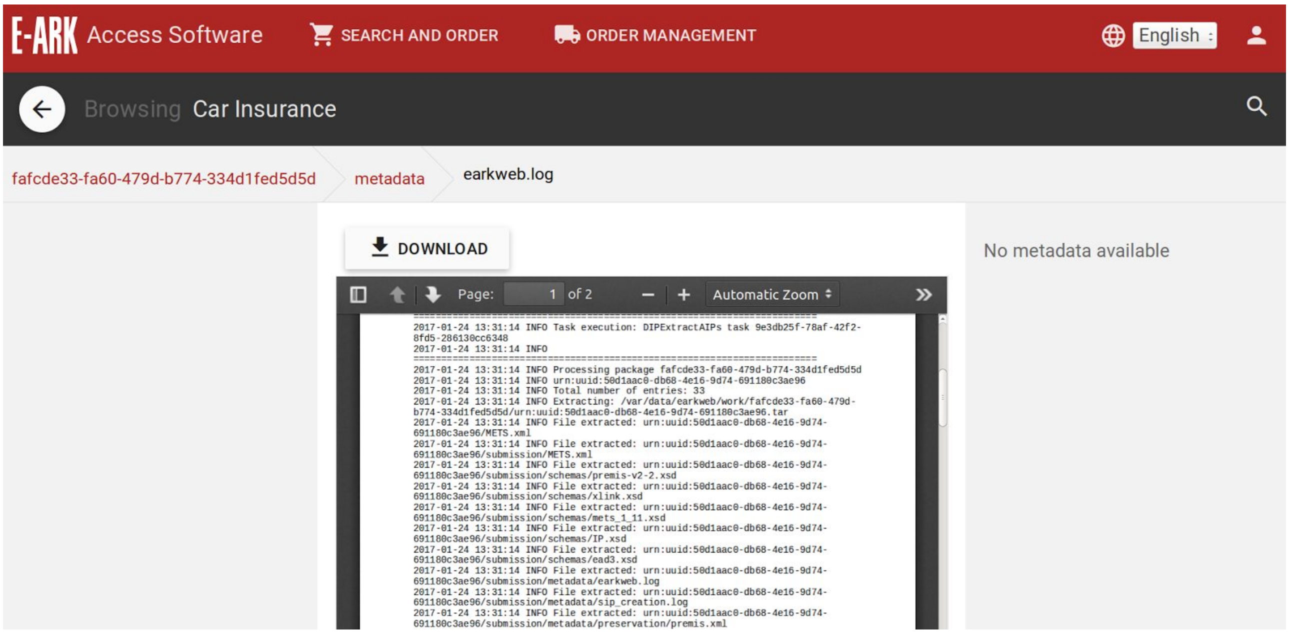
- urn:uuid:50d1aac0-db68-4e1...
- metadata
  - preservation

	NAME ▲	DATE
□	preservation	24 Jan 2017, 13:31
□	earkweb.log	24 Jan 2017, 13:31

No metadata available

Figure 39 – Metadata in the pre-DIP

The log file (‘earkweb.log’) gives an overview of the files that have been copied from the AIP:



fafcde33-fa60-479d-b774-334d1fed5d5d › metadata › earkweb.log

No metadata available

```
===========================================================================
2017-01-24 13:31:14 INFO Task execution: DIPExtractAIPs task 9e3db25f-78af-42f2-
8fd5-286130cc6348
2017-01-24 13:31:14 INFO
===========================================================================
2017-01-24 13:31:14 INFO Processing package fafcde33-fa60-479d-b774-334d1fed5d5d
2017-01-24 13:31:14 INFO urn:uuid:50d1aac0-db68-4e16-9d74-691180c3ae96
2017-01-24 13:31:14 INFO Total number of entries: 33
2017-01-24 13:31:14 INFO Extracting: /var/data/earkweb/work/fafcde33-fa60-479d-
b774-334d1fed5d5d/urn:uuid:50d1aac0-db68-4e16-9d74-691180c3ae96.tar
2017-01-24 13:31:14 INFO File extracted: urn:uuid:50d1aac0-db68-4e16-9d74-
691180c3ae96/METS.xml
2017-01-24 13:31:14 INFO File extracted: urn:uuid:50d1aac0-db68-4e16-9d74-
691180c3ae96/submission/METS.xml
2017-01-24 13:31:14 INFO File extracted: urn:uuid:50d1aac0-db68-4e16-9d74-
691180c3ae96/submission/schemas/premis-v2-2.xsd
2017-01-24 13:31:14 INFO File extracted: urn:uuid:50d1aac0-db68-4e16-9d74-
691180c3ae96/submission/schemas/xlink.xsd
2017-01-24 13:31:14 INFO File extracted: urn:uuid:50d1aac0-db68-4e16-9d74-
691180c3ae96/submission/schemas/mets_1_11.xsd
2017-01-24 13:31:14 INFO File extracted: urn:uuid:50d1aac0-db68-4e16-9d74-
691180c3ae96/submission/schemas/IP.xsd
2017-01-24 13:31:14 INFO File extracted: urn:uuid:50d1aac0-db68-4e16-9d74-
691180c3ae96/submission/schemas/ead3.xsd
2017-01-24 13:31:14 INFO File extracted: urn:uuid:50d1aac0-db68-4e16-9d74-
691180c3ae96/submission/metadata/earkweb.log
2017-01-24 13:31:14 INFO File extracted: urn:uuid:50d1aac0-db68-4e16-9d74-
691180c3ae96/submission/metadata/sip_creation.log
2017-01-24 13:31:14 INFO File extracted: urn:uuid:50d1aac0-db68-4e16-9d74-
691180c3ae96/submission/metadata/preservation/premis.xml
```

Figure 40 – Log file from E-ARK Web

In order to create a DIPu, the archivist will have to perform two steps manually before initiating the automatic task that completes the creation of the DIPu (cf. section 5.2.5.7 Package DIP):

1. Copy relevant data to the future DIPu
2. Copy pertaining descriptive metadata to the future DIPu

3.2.5.6.1 Copying relevant data and metadata to the future DIPu

Before explaining the concrete manual steps necessary to create the DIPu, the editing features enabling archivists to modify the DIP will be described[33].

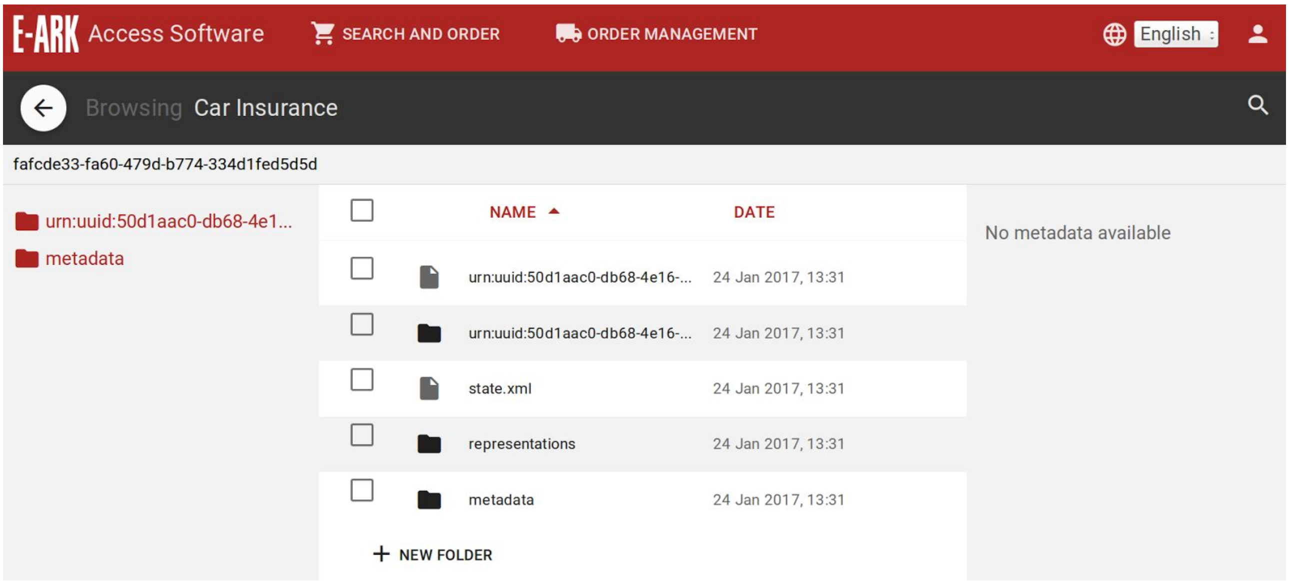
Editing features of the IP Viewer

[33] Using the IP Viewer to navigate folder structures will be more closely covered in section 3.2.6.1 Browsing IPs.

- **Create folders**
  As with the “representations” folder, new directory folders can be created anywhere in the folder hierarchy by clicking the “New folder” button below the files and folders list. Enter a folder name in the dialog that pops up and select “OK” to create these.
- **Item selection**
  The files and folders list has a column of checkboxes. When copying or deleting items, these boxes are checked to select certains items for the requested action. Selecting the topmost checkbox will select all currently visible files and folders.
- **Item copy**
  One or more selected items can be copied by clicking the “Copy” button below the list of files and folders. A clipboard appears at the bottom of the screen displaying the items currently selected for copying. To finalize the copying action, browse to the desired target directory (usually “representations”) and click the “Paste here” button that will be visible below the files and folders list. To cancel a copy action, click the trash bin icon in the clipboard.
  When copying a folder, its subfolders will also be copied to the target location.
- **Item delete**
  Items can be deleted by selecting them and clicking the “Delete” button that is visible below the files and folders list. When deleting a directory, all its subdirectories are deleted along with it.

The two manual steps of DIPu creation

In order to copy relevant data (folders and files) from within source IP folder, a new “representations” folder must be created in the root:



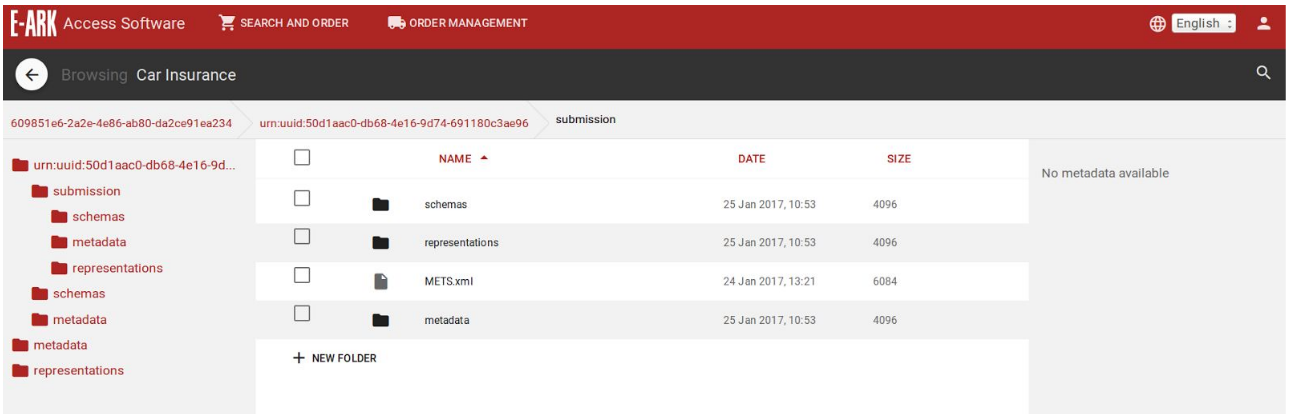
fafcde33-fa60-479d-b774-334d1fed5d5d

□	NAME	DATE
□	urn:uuid:50d1aac0-db68-4e16-...	24 Jan 2017, 13:31
□	urn:uuid:50d1aac0-db68-4e16-...	24 Jan 2017, 13:31
□	state.xml	24 Jan 2017, 13:31
□	representations	24 Jan 2017, 13:31
□	metadata	24 Jan 2017, 13:31

No metadata available

Figure 41 – Creation of the DIPu “representations” folder

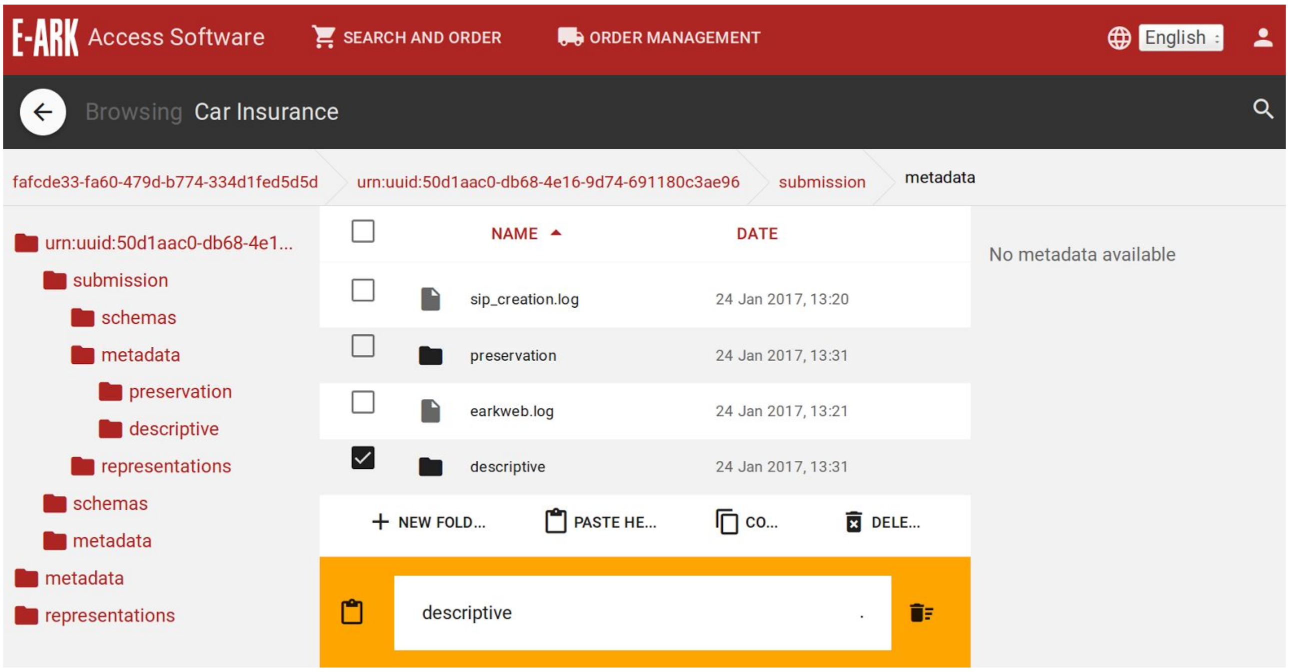
The requested items must be copied into this folder from the “representations” folder of the source IP, which is found inside the ‘“submission” folder:



□		NAME	DATE	SIZE
□		schemas	25 Jan 2017, 10:53	4096
□		representations	25 Jan 2017, 10:53	4096
□		METS.xml	24 Jan 2017, 13:21	6084
□		metadata	25 Jan 2017, 10:53	4096

Figure 42 – Selection of content for the DIPu

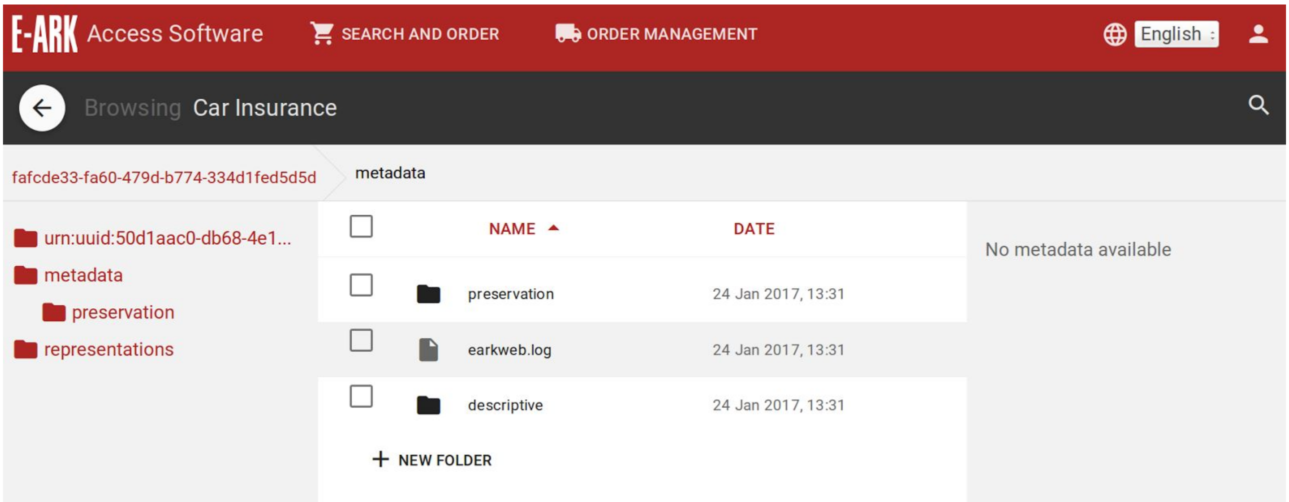
In order to enrich the future DIPu with descriptive metadata, copy the ‘descriptive’ folder that also resides within the ‘submission’ folder of the source IP folder:



□		NAME	DATE
□		sip_creation.log	24 Jan 2017, 13:20
□		preservation	24 Jan 2017, 13:31
□		earkweb.log	24 Jan 2017, 13:21
☑		descriptive	24 Jan 2017, 13:31

Figure 43 – Descriptive metadata enrichment of the DIPu #1

and paste it into the root metadata folder:



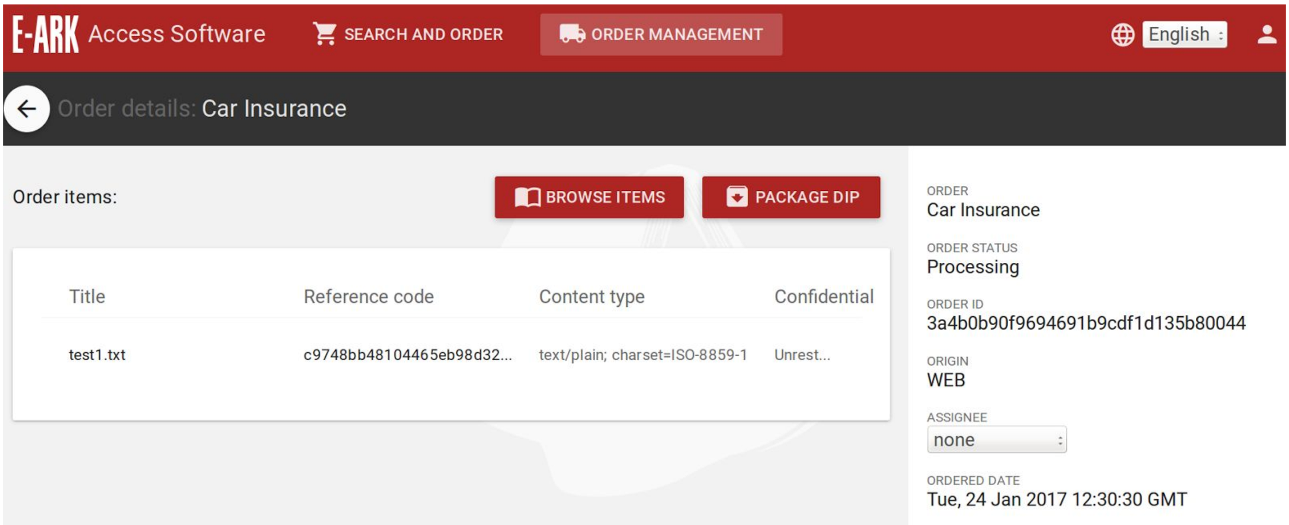
	NAME	DATE
□	preservation	24 Jan 2017, 13:31
□	earkweb.log	24 Jan 2017, 13:31
□	descriptive	24 Jan 2017, 13:31

Figure 44 – Descriptive metadata enrichment of the DIPu #2

Note that the E-ARK project has created an EAD Editor, which has not been integrated into the Order Management Tool, and thus remains a stand-alone tool. If the archivist needs to edit an EAD file, (s)he needs to do this from outside the Order Management Tool.

3.2.5.7 Package DIP

After customizing the pre-DIP, only the automated packaging step remains before the end-user can access the requested items in a finalized DIPu. From the order detail page, click the "PACKAGE DIP" button to initiate the packaging process.



Title	Reference code	Content type	Confidential
test1.txt	c9748bb48104465eb98d32...	text/plain; charset=ISO-8859-1	Unrest...

ORDER
Car Insurance

ORDER STATUS
Processing

ORDER ID
3a4b0b90f9694691b9cdf1d135b80044

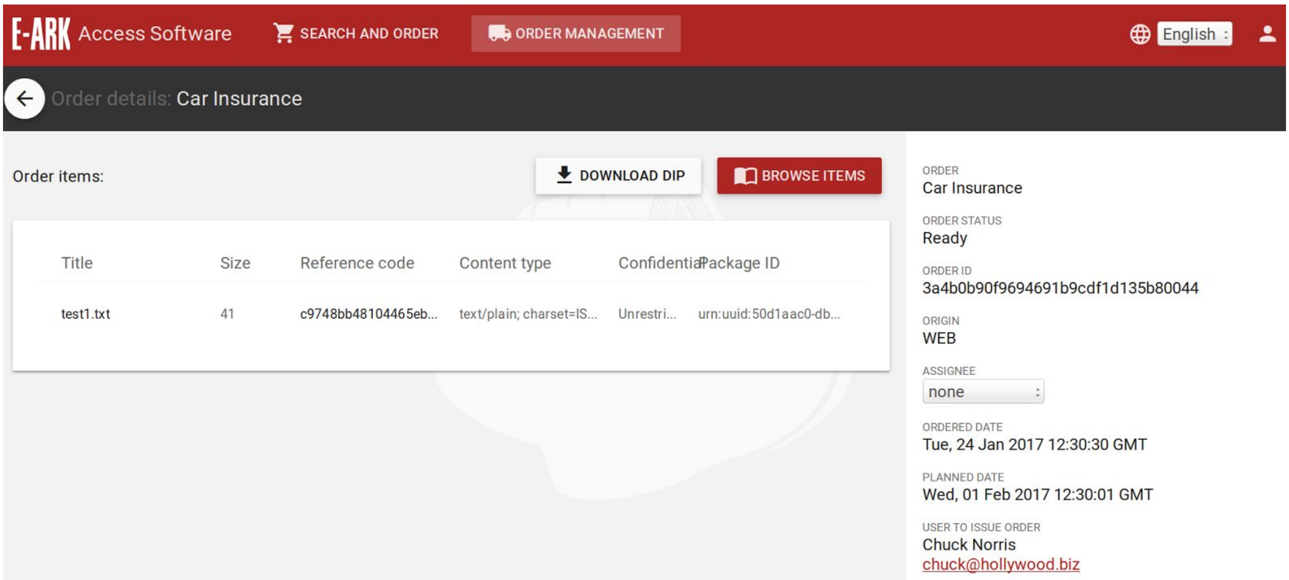
ORIGIN
WEB

ASSIGNEE
none

ORDERED DATE
Tue, 24 Jan 2017 12:30:30 GMT

Figure 45 – Initiation of "Package DIP"

After a while, depending on the size of the documents, the order will change its status from "Packaging" to "Indexing" and finally to "Ready". Clicking the "BROWSE ITEMS" again from the order details page will display the finalized DIP. In addition, a "DOWNLOAD DIP" button will be available for downloading the entire DIP.



Order details: Car Insurance

Order items:

Title	Size	Reference code	Content type	Confidentia	Package ID
test1.txt	41	c9748bb48104465eb...	text/plain; charset=IS...	Unrestri...	urn:uuid:50d1aac0-db...

ORDER
Car Insurance

ORDER STATUS
Ready

ORDER ID
3a4b0b90f9694691b9cdf1d135b80044

ORIGIN
WEB

ASSIGNEE
none

ORDERED DATE
Tue, 24 Jan 2017 12:30:30 GMT

PLANNED DATE
Wed, 01 Feb 2017 12:30:01 GMT

USER TO ISSUE ORDER
Chuck Norris
chuck@hollywood.biz

Figure 46 – The DIP is ready and available from the order details screen

3.2.6 IP Viewer

The IP Viewer enables archivists and end-users to browse and search DIPu’s and makes it possible to view and download the files that are stored within the DIP. This gives a first overview over a DIP - its structure, its data, and its metadata.

The IP Viewer is used in read-only mode for letting end-users view archival records in finalized DIPu’s and in edit mode for letting archivists prepare AIPs for dissemination from within the Order Management Tool. Features include:

- File tree and breadcrumb path for navigation support.
- Detailed view of individual archival records with option to preview and download files.
- View metadata of selected items (files and folders) in the IP.
- View and download metadata files.
- Search archival records within the IP using Solr search.
- Add folders, copy and delete files in order to create DIPs (edit mode for archivists only).

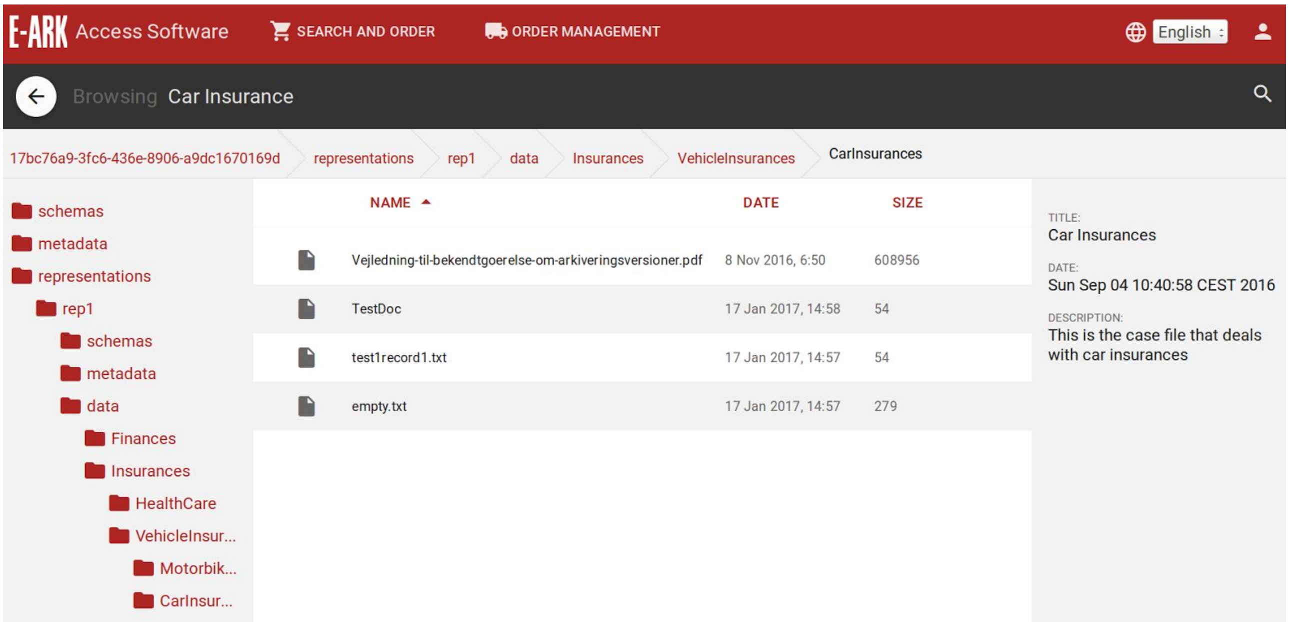
What follows is a step by step introduction to the use of those features.

3.2.6.1 Browsing IPs

When an order has status “Ready”, end-users can access a DIP with the ordered items by going to the order detail page and clicking the “BROWSE ITEMS” button (cf. figure 45 above). This takes them to the IP Viewer with the contents of the DIP loaded into view.

3.2.6.2 Navigating the IP

The central part of the IP Viewer is the files and folders list. Clicking any item in the list will change the center view to display the contents of the file or folder selected.



NAME	DATE	SIZE
Vejledning-til-bekendtgoerelse-om-arkiveringsversioner.pdf	8 Nov 2016, 6:50	608956
TestDoc	17 Jan 2017, 14:58	54
test1record1.txt	17 Jan 2017, 14:57	54
empty.txt	17 Jan 2017, 14:57	279

TITLE:
Car Insurances

DATE:
Sun Sep 04 10:40:58 CEST 2016

DESCRIPTION:
This is the case file that deals with car insurances

Figure 47 – The IP Viewer interface

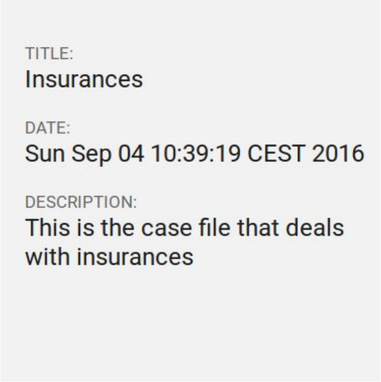
A directory navigation tree is available to the left of the files and folders list. The navigation of the tree is handled similarly to widely used tools (Windows Explorer, etc.). Clicking a folder name in the navigation tree will lead to a view of that folder. Note that the navigation tree only shows subdirectories along the path that are currently being navigated. There can be subdirectories available that are not visible until the parent folder name is clicked.

Above the navigation tree and files/folders list is the breadcrumb navigation. This displays the current path of subdirectories that are currently navigated to. Clicking a folder name in the breadcrumb will lead to a view of the that folder.

Click the “back” arrow in the top dark bar to return to the order details page.

3.2.6.3 Viewing related metadata

Whenever there is metadata available for the file or folder which is currently inspected, these will be displayed in the metadata pane to the right.



TITLE:
Insurances

DATE:
Sun Sep 04 10:39:19 CEST 2016

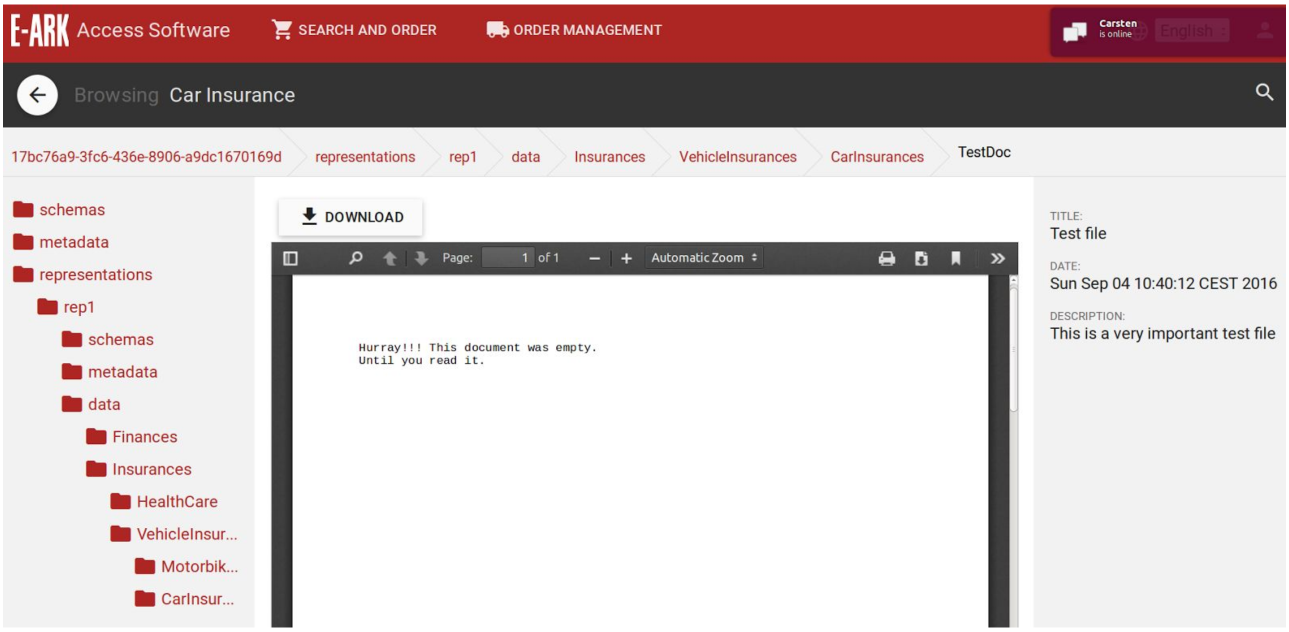
DESCRIPTION:
This is the case file that deals with insurances

Figure 48 – The metadata pane

If a file or folder has no metadata attached to it “No metadata is available” will display. More metadata elements can be displayed.

3.2.6.4 Viewing individual files

Clicking a filename will change the center view to single file view. This displays a preview of the file when available, along with a button to download the file in question. The file that is selected will be converted into PDF. Also, any available descriptive metadata for the file will display in the metadata pane.



```
Hurray!!! This document was empty.
Until you read it.
```

TITLE:
Test file

DATE:
Sun Sep 04 10:40:12 CEST 2016

DESCRIPTION:
This is a very important test file

Figure 49 – Individual file view

The individual file view displays metadata pertaining to the selected file. It also gives a series of options/features (reading from left to right on the top of the file):

- Toggle Sidebar, which offers these possibilities when clicked:
  - Show Thumbnails
  - Show Document Outline
  - Show Attachments

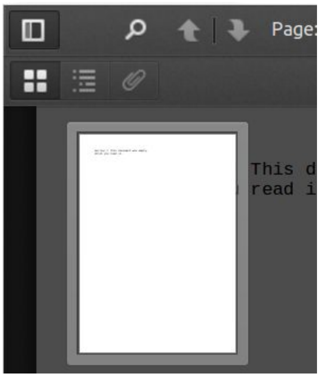
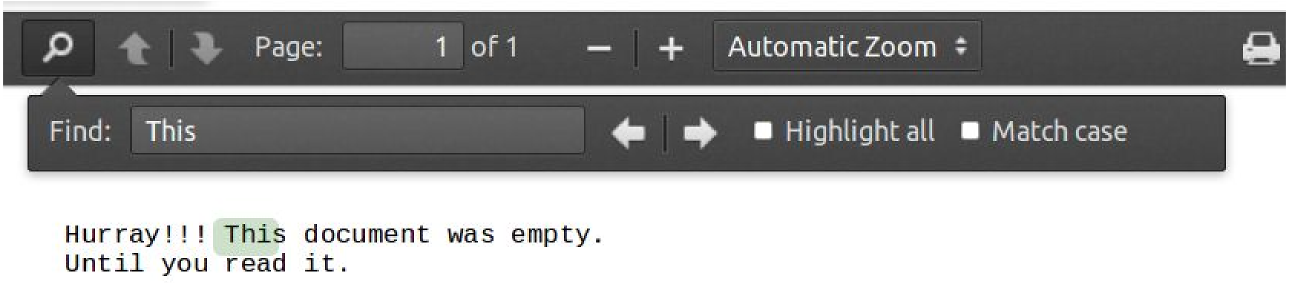


Figure 50 – Toggle Sidebar

- Find in Document



```
Hurray!!! This document was empty.
Until you read it.
```

Figure 51 – Find in Document

- Previous Page / Next Page

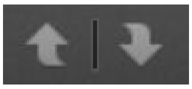


Figure 52 – Previous Page / Next Page

- Page (jump to another page if applicable)

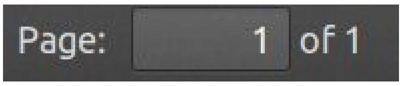


Figure 53 – Jump to Another Page

- Manual and Automatic Zoom

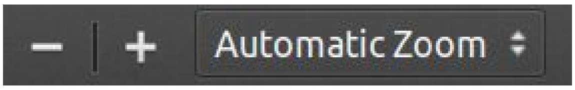


Figure 54 – Manual and Automatic Zoom

- Print, Download and Current view

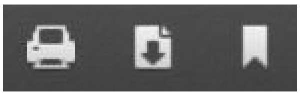


Figure 55 – Print, Download and Current view

- Tools
  - Go to First Page
  - Go to Last Page
  - Rotate Clockwise
  - Rotate Counterclockwise
  - Enable hand tool
  - Document Properties...

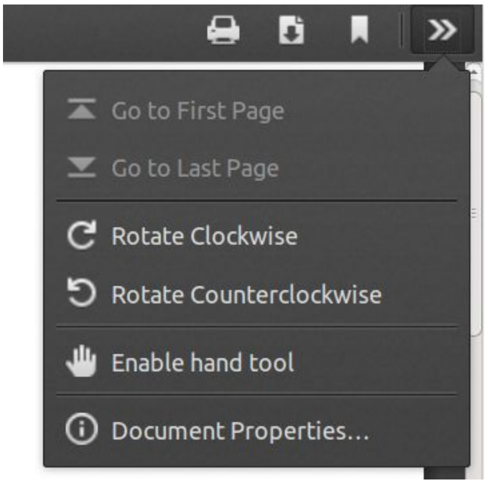


Figure 56 – Tools

3.2.6.5 Searching the IP

IPs can contain a lot of files so a search feature has been added to enable users to search for specific files within an IP. In the IP view, select the search icon in the upper right corner and a small search input field will reveal itself. Enter a search term in the field and hit “Enter” to initiate a search within the IP. The search feature works in the same way as described in “3.2.4.1 Initial search”.

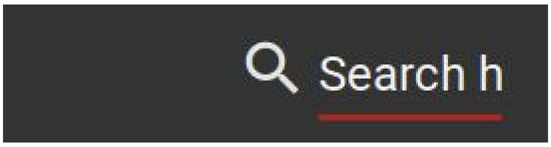


Figure 57 – IP Viewer Search

3.2.7 Code and documentation for Search Module, Order Management Tool, and IP Viewer

What follows are some general source code details for Search Module, Order Management Tool, and IP Viewer (Access Software Platform tools). The software is open source and publicly available via GitHub.

Link to source code for the Access Software Platform:
https://github.com/eark-project/E-Ark-Platform-UI

The project’s README has instructions for installation:
https://github.com/eark-project/E-Ark-Platform-UI/blob/master/README.md

Details about translating the UI can be found in the source code documentation:
https://github.com/eark-project/E-Ark-Platform-UI/blob/master/app/src/i18n/README.md#adding-a-new-translation

Specifically, the source code for the Order Management tool is available from:
https://github.com/eark-project/E-Ark-Platform-UI/tree/master/app/src/order_management

The Order Management Tool communicates with the Order Management Service (OMS) which functions as the backend for the Order Management Tool. The source code for the OMS along with documentation for its API can be found on GitHub: https://github.com/eark-project/OMS

The source code for the IP Viewer is available from: https://github.com/eark-project/E-Ark-Platform-UI/tree/master/app/src/ipview

3.3 The AIP-DIP conversion tool

This tool is a component of the E-ARK Web. The component operates automatically when initiated by the Access Software Platform through the E-ARK Web API.

3.3.1 Task execution framework

The AIP-DIP conversion component consists of a set of individual tasks which are executed in a specific order to convert an E-ARK Archival Information Package (AIP) into the E-ARK Dissemination Information Package (DIP). It is an extensible workflow, which can be adapted by the digital repository administrator for specific needs by inserting new tasks at any point of the workflow. E-ARK Web uses a modular approach for defining atomic tasks, which perform specific transformation steps in the AIP-DIP conversion, such as the extraction of an AIP or the validation of descriptive metadata it contains. However, a specific task does not necessarily execute one single action, but it can initiate a series of tasks or a complete workflow as well.

Each task is implemented as a python class and is available in the python module “workers/tasks.py” of the E-ARK Web application. A task which performs a step of the AIP-DIP conversion must extend the default task class ‘DefaultTask’ defined in the module “workers/default_task.py”.

The default task makes sure that the pre-conditions for executing a task are fulfilled (e.g. the package is not in an error state). The default task also verifies if task execution is allowed given the current state of the package. Each task has a property which defines the list of tasks which are accepted as previously executed tasks. The fact that a task is defined as an “accepted last task” means that if execution was successful, there is the assumption that it produces valid output to be used as input of the current task. For example, to execute the ‘DIPExtractAIP’s task which extracts the contents of the selected AIPs, it is required that the AIP TAR[34] files have been retrieved from the storage by the ‘DIPAcquireAIPs’ task[35].

3.3.2 AIP-DIP conversion tasks

The table below provides an overview of the tasks which together represent the AIP-DIP conversion component. In addition to the tasks listed in the “Accepted inputs” column each task can be executed after itself.

Task name	Accepted inputs	Task description
AIPtoDIPReset	All	Resets the AIP-DIP workflow and removes all data.
DIPAcquireAIPs	AIPtoDIPReset	Retrieves the selected AIPs from the storage and saves them in the DIP working directory.
DIPAcquireDependentAIPs	DIPAcquireAIPs	Retrieves the dependent AIPs in cases where the selected AIPs are segments of a larger archive.

[34] Tar https://en.wikipedia.org/wiki/Tar_(computing)

[35] For more information on the E-ARK Web and its workflow engine please consult D4.4 Final version of SIP-AIP conversion component. (http://www.eark-project.com/resources/project-deliverables/89-d44)

DIPExtractAIPs	DIPAcquireAIPs DIPAcquireDependentAIPs	Extracts the AIP TAR files to the DIP working directory.
DIPImportSIARD	DIPExtractAIPs	Finds SIARD files and imports them in a database for processing. The database server is configured in the E-ARK Web settings.
DIPExportSIARD	DIPImportSIARD	Exports the database into a SIARD file. This is usually done in cases where changes to the database are necessary before creating a DIP.
DIPGMLDataValidation	DIPExtractAIPs DIPImportSIARD DIPExportSIARD	Validates GML[36] data contained in the working folder, if any.
DIPGMLDataConversion	DIPImportSIARD DIPExportSIARD DIPGMLDataValidation	Converts GML data to Peripleo[37] specific format.
DIPPeripleoDeployment	DIPGMLDataConversion	Imports the converted geo data to Peripleo for visualization.
DIPMetadataCreation	DIPExtractAIPs DIPExportSIARD	Creates the METS[38] and additional metadata for the DIP.
DIPIdentifierAssignment	DIPMetadataCreation	Assigns a new identifier for the DIP. This id is used for storing the DIP.
DIPPackaging	DIPIdentifierAssignment	Packaging the DIP as a TAR file.
DIPStore	DIPPackaging	Stores the DIP in the file system in the Pairtree storage. This is the storage area of the standalone software stack. In the cluster software stack this storage area represents the staging

[36] The Geography Mark-up Language: the XML grammar defined by the Open Geospatial Consortium (OGC) to express geographical features. GML serves as a modelling language for geographic systems as well as an open interchange format for geographic transactions on the Internet. (GML, Geography Markup Language https://en.wikipedia.org/wiki/Geography_Markup_Language)

[37] Peripleo https://wiki.digitalclassicist.org/Peripleo

[38] The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. (METS, Metadata Encoding and Transmission Standard http://www.loc.gov/standards/mets/mets-schemadocs.html)

		area holding packages which are going to be uploaded to the Lily Repository[39].
DIPCreateAccessCopy	DIPStore	Copies the DIP from the Pairtree storage to a location that can be accessed through an URL.

Table 1 – AIP to DIP Conversion tasks

3.3.3 DIP creation API

The API is available at "earkweb/search" endpoint and implemented as follows.

3.3.3.1 Prepare DIP working area

In order to prepare the DIP working area a POST call with the process_id has to be issued on the “earkweb/search/prepareDIPWorkingArea” endpoint. This creates the necessary folder and database entries for the new process.

curl -X POST -d '{"process_id": "c1b1c16e-2c00-474f-b99b-42019b3eaeed"}'
http://localhost:8000/earkweb/search/prepareDIPWorkingArea

One of the following response messages is returned:

- 201 : Created - The job was submitted successfully (does not mean successfully finished, though!)
- 412: Precondition failed - No AIPs selected for this DIP creation process
- 400: Bad Request - JSON body malformed or wrong request type
- 404: Not Found - The Process-ID does not exist
- 500: Internal Server Error - Some error occurred (see message)

Response Body (success)

{"message": "DIP preparation job submitted successfully.", "process_id":
"c1b1c16e-2c00-474f-b99b-42019b3eaeed", "success": true, "jobid":
"9b33d6bf-859b-42d5-ac15-d6ce7de45fa0"}

3.3.3.2 Run DIP creation process

In order to execute the AIP to DIP creation process a POST call with the process_id has to be issued on the “/earkweb/search/createDIP” endpoint .

curl -X POST -d '{"process_id": "c1b1c16e-2c00-474f-b99b-42019b3eaeed"}'
http://localhost:8000/earkweb/search/createDIP

One of the following response messages is returned:

- 201 : Created - The job was submitted successfully (does not mean successfully finished though!)
- 412: Precondition failed - No AIPs selected for this DIP creation process
- 400: Bad Request - JSON body malformed or wrong request type
- 404: Not Found - The Process-ID does not exist

[39] The files are uploaded to the Hadoop Distributed File System (HDFS) of the Lily Repository. See deliverable D6.2 E-ARK Integrated Platform Reference Implementation for details about the Lily Repository at http://www.eark-project.com/resources/project-deliverables/54-d62intplatformref-1.

- 500: Internal Server Error - Some error occurred (see message)

Response Body (success)

{"status": "finished", "message": "DIP creation finished successfully.", "process_id": "c1b1c16e-2c00-474f-b99b-42019b3eaeed", "success": true, "download_url": "http://127.0.0.1:8000/static/earkweb/download/gznpyhze/cd5cf9fe-947b-46ae-9b61-cf1337db6b54.tar"}

The download_url in the JSON response body contains the URL where the newly created DIP can be downloaded.

3.3.3.3 Check job status

In order to check the DIP creation status a GET call has to be issued on the “/earkweb/search/jobstatus” endpoint with the process_id appended at the end of the URL.

http://localhost:8000/earkweb/search/jobstatus/210a1870-aad3-442a-bbc9-75438b39e87a

One of the following response messages is returned:

- 200 : OK - Job status request successful
- 400: Bad Request - Wrong request type
- 500: Internal Server Error - Some error occurred (see message)

Response Body (success)

{"message": "DIP creation job submitted successfully.", "process_id": "c1b1c16e-2c00-474f-b99b-42019b3eaeed", "success": true, "jobid": "210a1870-aad3-442a-bbc9-75438b39e87a"}

3.3.4 Index DIP in storage API

3.3.4.1 Submitting and indexing job

In order to enable searching, after successful DIP creation, the package has to be indexed. This can be accomplished by starting an indexing job for a specific package. The indexing job is started by issuing a POST call to the “/earkweb/earkcore/index_local_storage_ip” endpoint with the DIP identifier provided in a JSON message. Upon execution existing index information with the same identifier is removed. The response message after sending the indexing request only means that the job was successfully submitted. Querying for the job status can be done by using the same job status function as for the DIP creation functionality.

curl -X POST -d '{"identifier": "urn:uuid:08e41ebb-bfa1-452c-ad05-7e4cd6809d82"}'[40]
http://localhost:8000/earkweb/earkcore/index_local_storage_ip

One of the following response messages is returned:

- 200 : OK - Job status request successful
- 400: Bad Request - Wrong request type
- 500: Internal Server Error - Some error occurred (see message)

[40] This is the IP identifier.

Response Body (success)

```
{"message": "Indexing job submitted successfully.", "identifier":
"urn:uuid:08e41ebb-bfa1-452c-ad05-7e4cd6809d82", "success": true, "jobid":
"f173caf7-6866-4427-b105-e5c0a14591de"}
```

3.3.4.2 Check job status

In order to check the indexing job status a GET call has to be issued on the “/earkweb/search/jobstatus” endpoint with the process_id appended at the end of the URL.

```
curl -X GET http://localhost:8000/earkweb/search/jobstatus/f173caf7-6866-4427-b105-e5c0a14591de
```

One of the following response messages is returned:

- 200 : OK - Job status request successful
- 400: Bad Request - Wrong request type
- 500: Internal Server Error - Some error occurred (see message)

Response Body (success)

```
{"status": "finished", "message": "IP indexing finished successfully.", "identifier":
"urn:uuid:08e41ebb-bfa1-452c-ad05-7e4cd6809d82", "success": true}
```

3.3.5 Code and documentation for the AIP-DIP conversion tool

1. https://github.com/eark-project/earkweb/blob/master/workers/tasks.py
2. https://github.com/eark-project/earkweb/blob/master/workers/default_task.py

3.4 The CMIS Viewer

The CMIS Viewer is a web based simple tool for browsing a CMIS repository.
The tool is architecturally divided into two components, an angular-js based UI which communicates via a RESTful[41] interface to a Java built backend (CMIS Bridge), as illustrated in the diagram below.

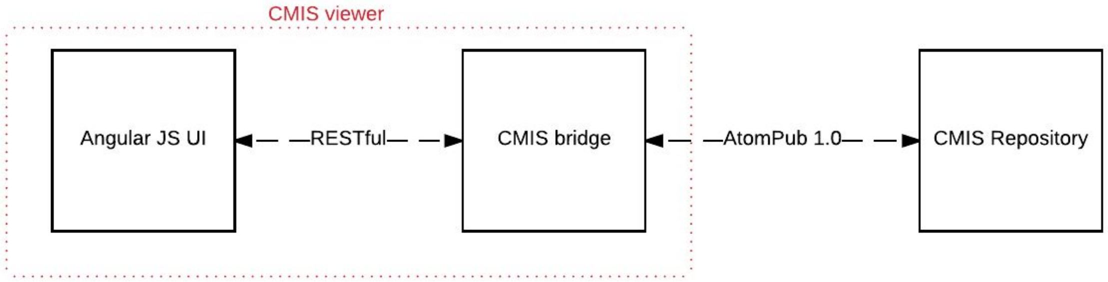


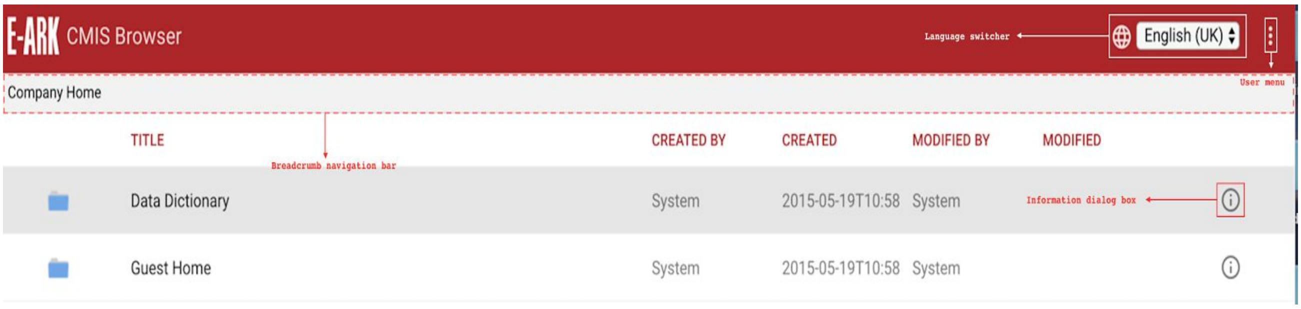
Figure 58 – CMIS Viewer architecture

3.4.1 Use case and feature description summary

The short list of features for the CMIS Viewer is as follows:

- Authenticate users
- Switch interface language
- Configure CMIS repository connection parameters
- Add/create new users to access the Viewer

The simple interface of the browser is depicted in the image below:



	TITLE	CREATED BY	CREATED	MODIFIED BY	MODIFIED
	Data Dictionary	System	2015-05-19T10:58	System	
	Guest Home	System	2015-05-19T10:58	System	

Figure 59 – CMIS Viewer interface

[41] Fielding, Roy Thomas http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm

3.4.1.1 Authentication and user authorisation

Users are authenticated against the CMIS viewer's user database, **NOT** against the CMIS repository itself. In fact the relation between the user and the datasource (the CMIS repository) is two tiered, both in terms of authentication and authorisation and is depicted in the image below.

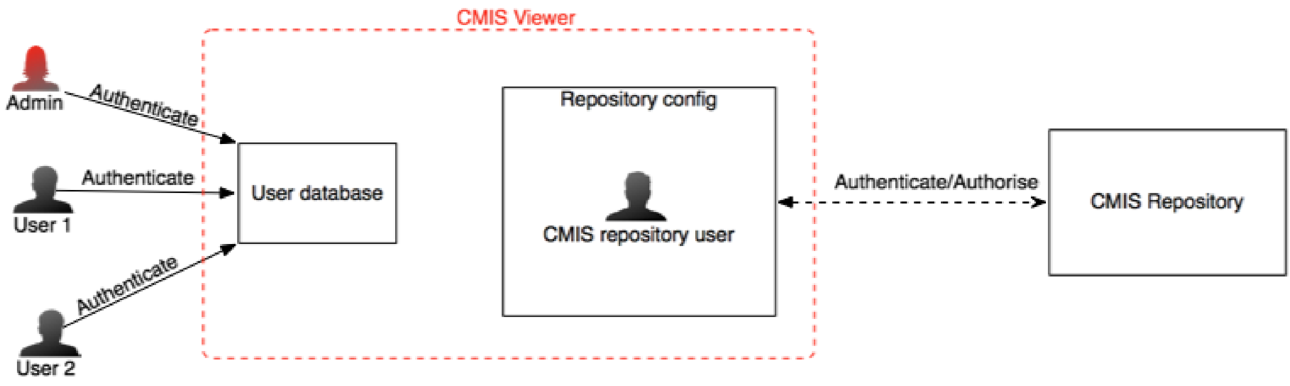


Figure 60 – CMIS Viewer authentication and user authorisation

The CMIS viewer recognises two types of authority, an ADMIN user and a STANDARD user. The difference is simply that additional system configuration menus are exposed to the ADMIN user under the user menu, otherwise user interaction with the repository browser is the same.

The relation between the CMIS viewer and the target CMIS repository is through a valid user that is recognised by the CMIS repository, as such this user must be present within the CMIS repository's own user database. The implication of this relationship between the viewer and the CMIS repository means that all users of the CMIS repository browser are constrained by the authorisation of this user with regards to the user's authorisation on the CMIS repository itself; therefore the CMIS repository browser has no effect whatsoever on the CMIS repository itself. One should think of the CMIS viewer as a simple reader with read only rights on the CMIS repository.

3.4.1.2 Configuring CMIS Viewer repository connection

An administrator account is required to be able to modify the CMIS repository connection configuration. This menu is accessible from the user menu.

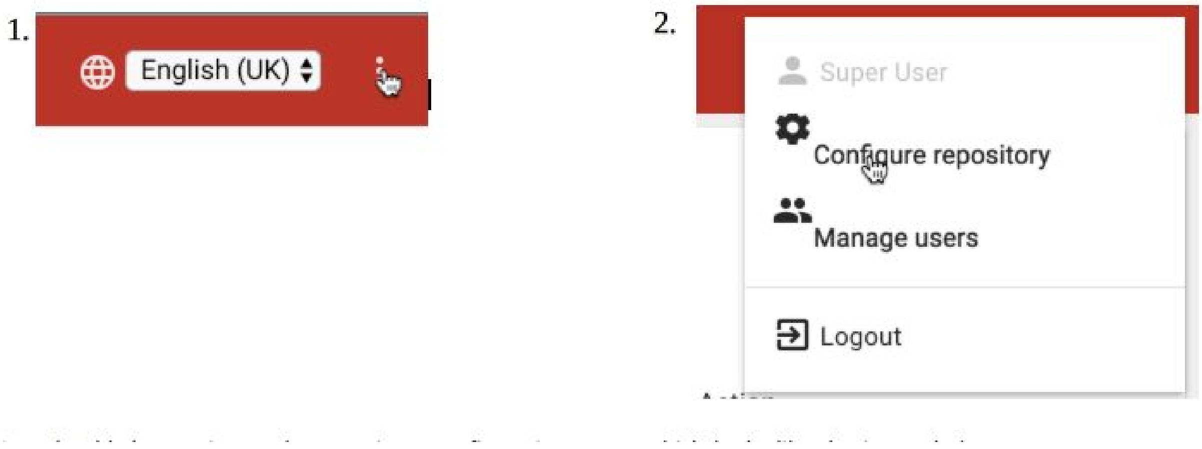
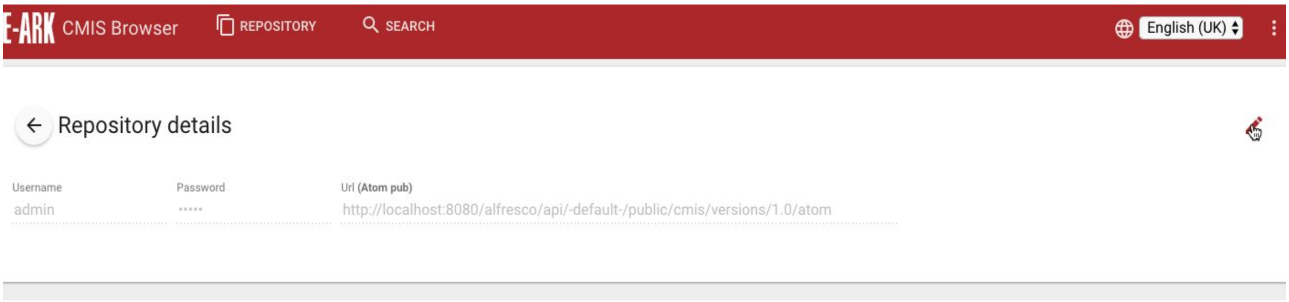


Figure 61 – CMIS Viewer repository connection

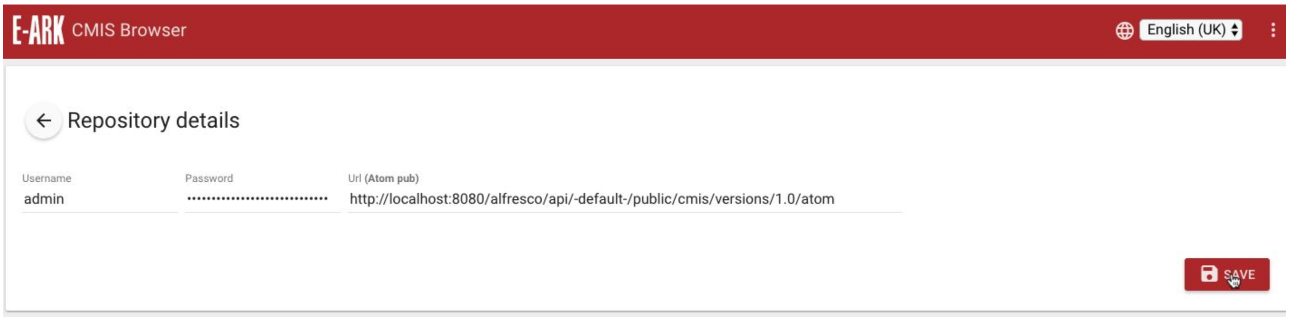
One should then arrive at the repository configuration menu which looks like the image below:



Username: admin
Password: •••••
Url (Atom pub): http://localhost:8080/alfresco/api/-default-/public/cmis/versions/1.0/atom

Figure 62 – CMIS Viewer repository configuration menu

Clicking the pen icon as shown in the picture above will result in the fields becoming editable and a save button appearing on the bottom right corner of the screen as shown in the next picture below:



Username: admin
Password: ••••••••••••••••••••••••••••••
Url (Atom pub): http://localhost:8080/alfresco/api/-default-/public/cmis/versions/1.0/atom

Figure 63 – Edit CMIS Viewer repository

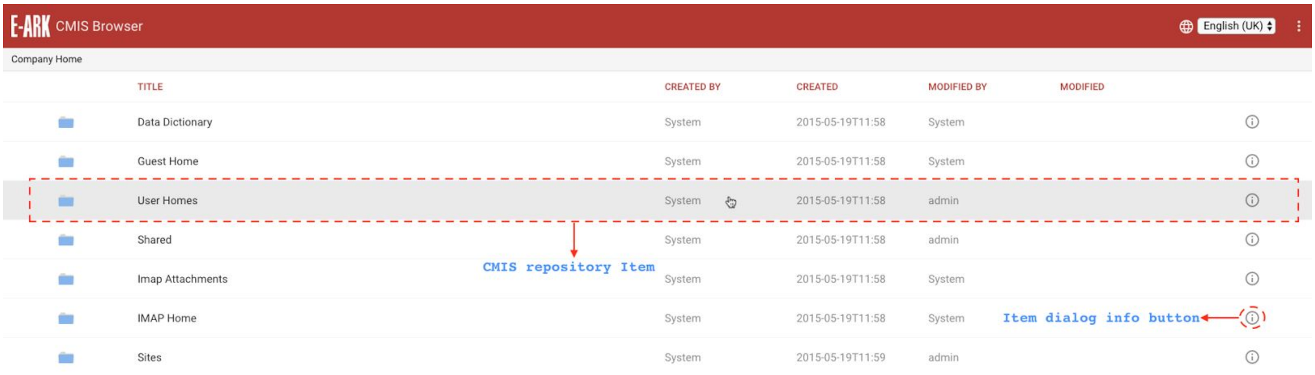
Again one should remember that the “user name” and “password” fields must match an authorised user in the CMIS repository itself; so all users browsing the CMIS repository will do so under authorisation of the user details saved here. The Url field MUST be an Atompub 1.0 url[42].

After saving, one can then click on the “⟵Repository details” button to go back to the repository browsing view

[42] Atom Publishing Protocol 1.0 https://movabletype.org/documentation/developer/api/atompub/

3.4.1.3 Browsing and getting around

After the initial login to the browser, the user is directed to the repository browsing view which should like the image below:



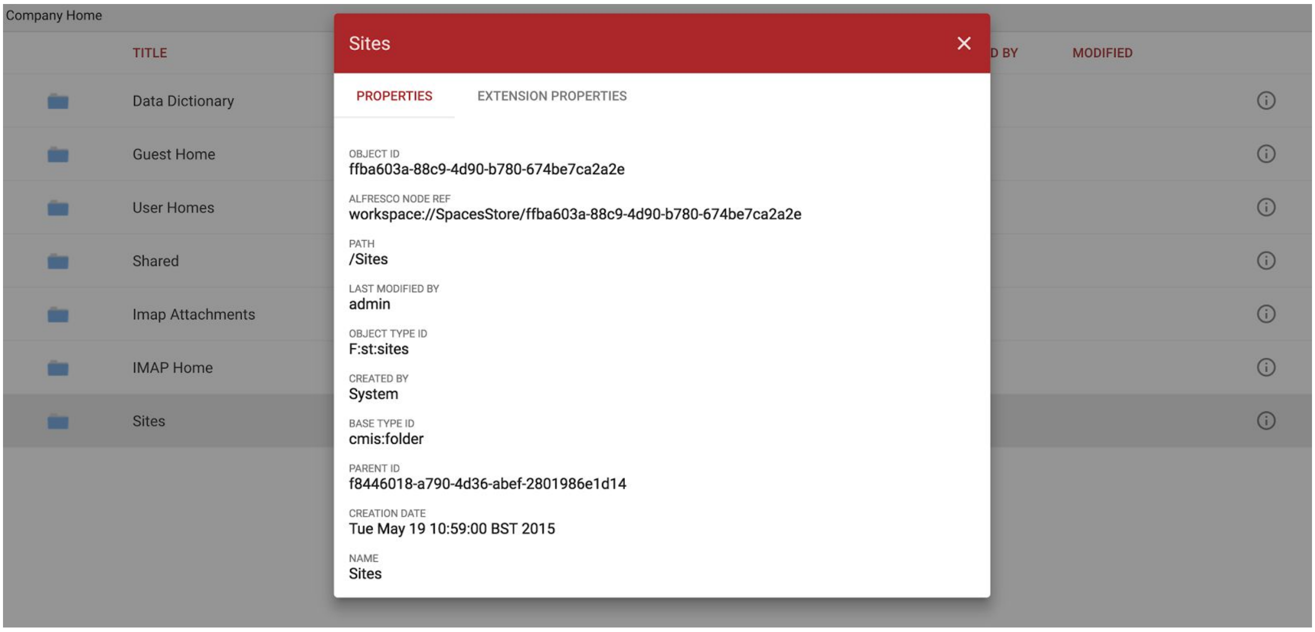
TITLE	CREATED BY	CREATED	MODIFIED BY	MODIFIED
Data Dictionary	System	2015-05-19T11:58	System	
Guest Home	System	2015-05-19T11:58	System	
User Homes	System	2015-05-19T11:58	admin	
Shared	System	2015-05-19T11:58	admin	
Imap Attachments	System	2015-05-19T11:58	System	
IMAP Home	System	2015-05-19T11:58	System	
Sites	System	2015-05-19T11:59	admin	

Figure 64 – Repository browsing

Mousing over a repository item highlights it as the current selection as can be seen in the image above, and clicking anywhere on the line except the item's dialog information button would result in the following actions:

- if the item is a directory, the view will refresh to display its contents.
- If the item is a file then a new window is opened where the user is prompted to download and save the file unto the local system. If the user's browser has a plugin to view the file, for example a pdf, then the file is displayed in the newly opened window instead. The user still has an option to download the file in this previewed window.

When the dialog information button is clicked, a dialog pops up to display the item's metadata similar to the picture below:



OBJECT ID
ffba603a-88c9-4d90-b780-674be7ca2a2e

ALFRESCO NODE REF
workspace://SpacesStore/ffba603a-88c9-4d90-b780-674be7ca2a2e

PATH
/Sites

LAST MODIFIED BY
admin

OBJECT TYPE ID
F:st:sites

CREATED BY
System

BASE TYPE ID
cmis:folder

PARENT ID
f8446018-a790-4d36-abef-2801986e1d14

CREATION DATE
Tue May 19 10:59:00 BST 2015

NAME
Sites

Figure 65 – Metadata view

The item metadata dialog is split into two panes:

1. Properties pane which displays the standard CMIS properties of the item and
2. Extension properties pane which displays the extension properties of the item; if any.

3.4.1.4 User Management

Managing users involves creating, deleting and editing a user's details in the viewer. This is the only other feature that is available to an administrator aside from the "Configure repository" feature.

Just like the repository configuration menu, this menu is also accessible from the user menu:

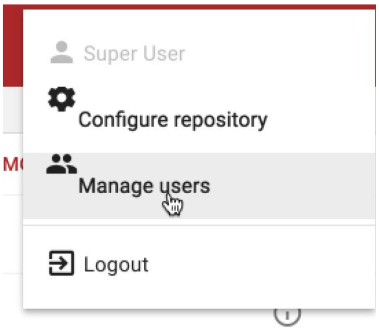
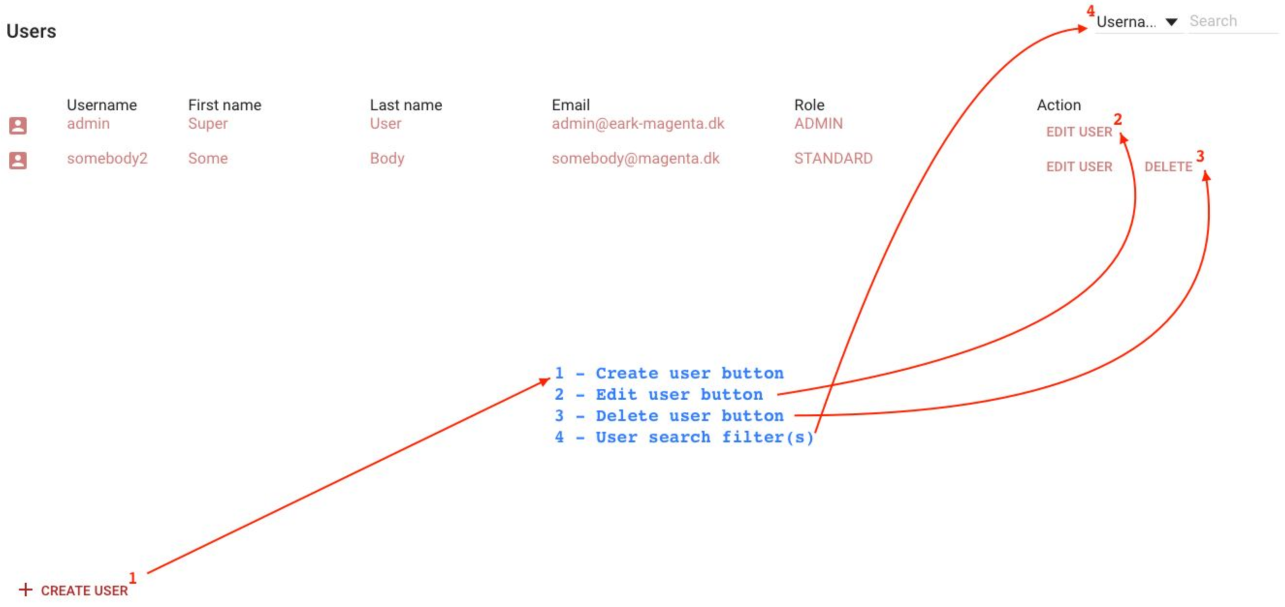


Figure 66 – CMIS Viewer user management

The user management screen should look similar to the picture below (minus the illustrations):

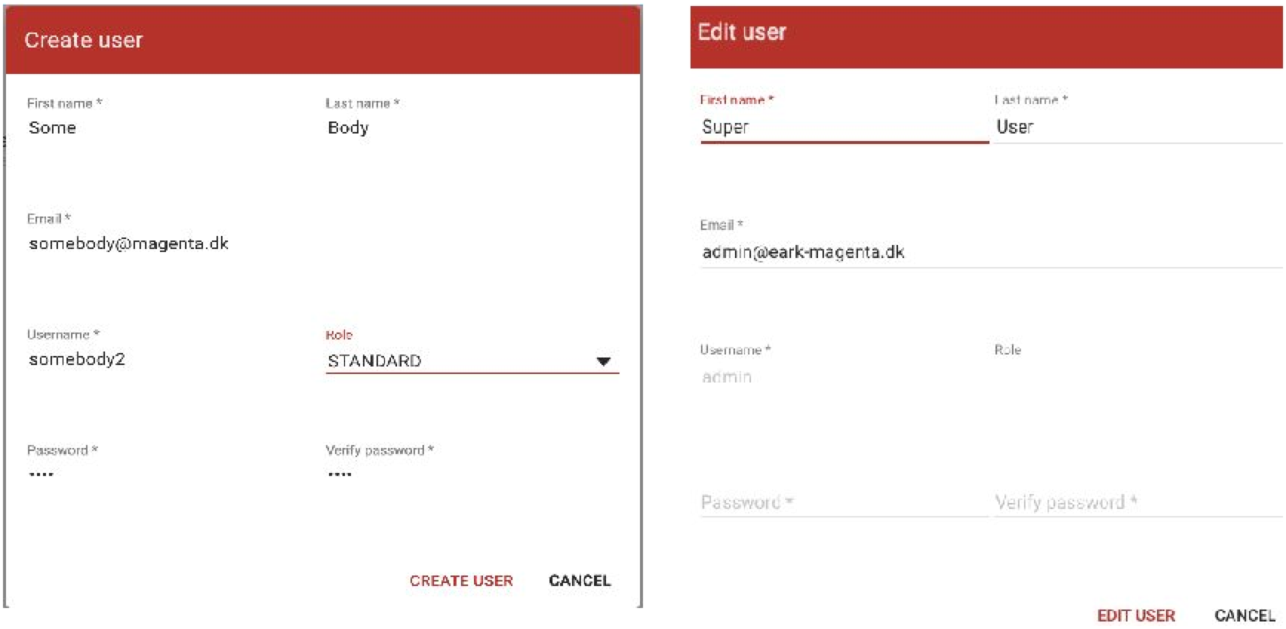


	Username	First name	Last name	Email	Role	Action
	admin	Super	User	admin@eark-magenta.dk	ADMIN	EDIT USER
	somebody2	Some	Body	somebody@magenta.dk	STANDARD	EDIT USER DELETE

```
1 - Create user button
2 - Edit user button
3 - Delete user button
4 - User search filter(s)
```

Figure 67 – CMIS Viewer user management screen

Management of a user happens interactively through a dialog box and is self explanatory. Creation and editing of users is enabled through the use of dialog boxes for both operations:



Create user

First name * Some

Last name * Body

Email * somebody@magenta.dk

Username * somebody2

Role STANDARD

Password * ••••

Verify password * ••••

Edit user

First name * Super

Last name * User

Email * admin@eark-magenta.dk

Username * admin

Role

Password *

Verify password *

Figure 68 – CMIS Viewer user management dialog boxes

When editing a user, the Username field is non modifiable, and in the case of the primary admin user (i.e. the user with the “admin” username), both the username and the roles are non-modifiable.

With regards to deleting a user, all users except the primary admin user are deletable.

3.4.2 Code and documentation for the CMIS Viewer

The code repository for both the UI[43] and the cmis bridge[44] can be found here: https://github.com/eark-project/E-Ark-CMIS-Viewer

3.5 Access to specific content information types

This section describes the end-user’s access to requested records of a specific content information type.

When access is given to an end-user by the archivist, the end-user receives access to the DIP via the IP Viewer, which is described above.

The IP Viewer allows for the end-user to browse the DIP, search it using Solr search, and view files and metadata.

However, as stated before, not all of the E-ARK content information types can be rendered automatically within the IP Viewer, but need specialised tools because of their complexity. These are opened with special E-ARK Viewers.

- Access to an IP is given via the IP Viewer (allows users to browse and view metadata)
- Access to databases and EDRMS stored in the SIARD format either via
    - an SQL access solution (the Database Preservation Toolkit loading the SIARD file into an RDBMS), or via
    - a NoSQL solution (using the Database Visualization ToolKit (DBVTK))
- Access to geodata stored in the SMURF format via QGIS/Geoserver.

[43] E-ARK CMIS Viewer https://github.com/eark-project/E-Ark-CMIS-Viewer/tree/master/frontend

[44] E-ARK CMIS Viewer https://github.com/eark-project/E-Ark-CMIS-Viewer/tree/master/bridge

- Access to EDRMS and unstructured records (SFSB[45]) stored in the SMURF format via the IP Viewer itself.
- OLAP access to information stored in SIARD via Oracle.

We repeat that not all rendering tools are integrated into the IP Viewer. However, in all of the scenarios described below, the IP can still be provided using the IP Viewer, but in some of the IPs (see bulleted list above) *the content file* (e.g. SIARD) is extracted from the IP and viewed by a bespoke tool.

3.5.1 Access to databases and EDRMS's stored in the SIARD format

The E-ARK project offers the ability to render SIARD files in two ways: via an SQL solution (Database Preservation Toolkit); or via a NoSQL solution (Database Visualisation Toolkit).

3.5.1.1 The Database Preservation Toolkit and its GUI: Access via an RDBMS (SQL solution)

The Database Preservation Toolkit (DBPTK) allows conversion between database formats, including connection to live systems, for purposes of digitally preserving databases. The toolkit allows the conversion of live or backed-up databases into preservation formats, such as SIARD. The toolkit also allows the conversion of the preservation formats back into live systems to allow the full functionality of databases, like querying and data analysis tasks.

The tool can be used either as a command line tool, or one can choose to plug its graphical user interface on top, for less technical users (see section below).

The toolkit is created as a platform that uses input and output modules. Each module supports read or write to a particular database format or live system. New modules can easily be added, providing the ability to convert to or from a new database format or live system.

The currently supported RDBMSs are:

- MySQL/MariaDB
- PostgreSQL
- Oracle
- Microsoft SQL Server
- Microsoft Access (only as input)
- And other databases (using JDBC)

Using this tool, any database present in one of these systems can be converted to SIARD. The same tool is also used to load the SIARD database into a live system (from the above list), providing access to the database using the capabilities of the database system, such as SQL querying and data analysis.

This tool is a command line application, so IT personnel can execute it on servers efficiently, without the overhead of a Graphical User Interface.

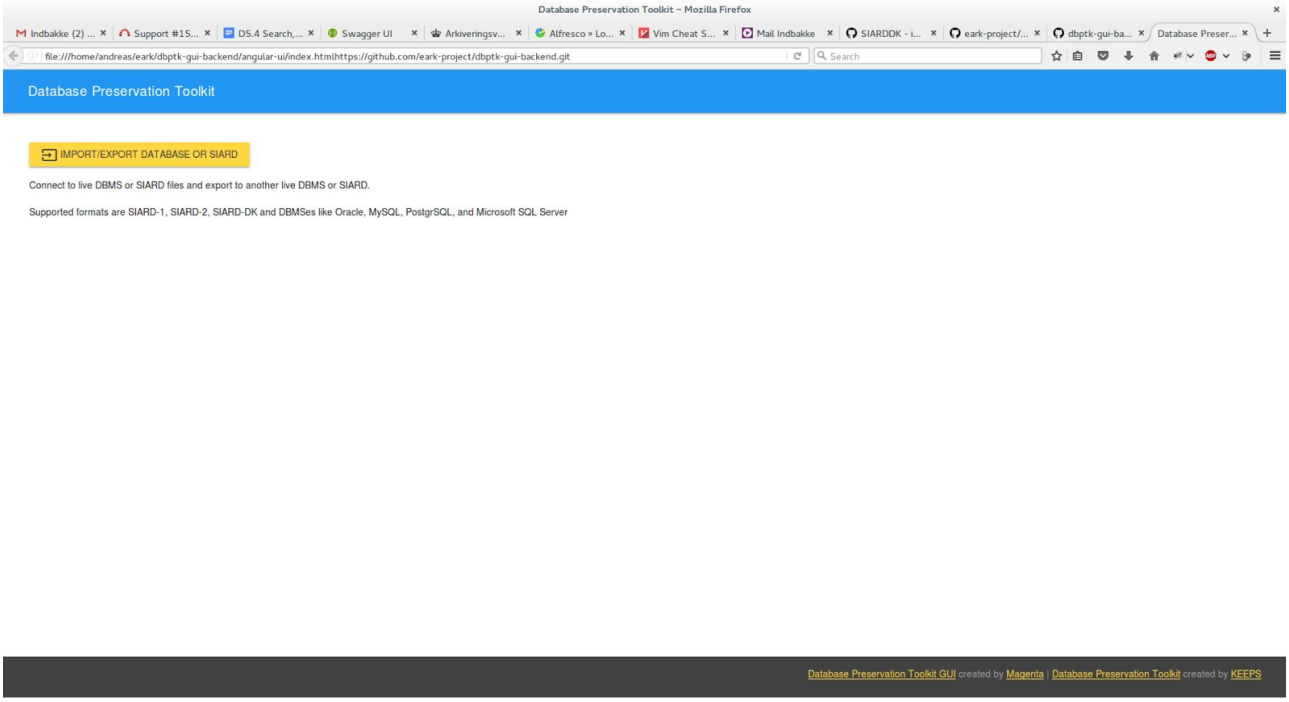
[45] Simple File-System Based records: records that contain simple file-system based folders or files, including those originating from content and data management systems, such as SharePoint, that are not based on true file systems. They address the submission of computer files or folders from the file Producers rather than from an ERMS. They require manual enrichment with additional descriptive metadata.

3.5.1.1.1 The graphical user interface for the Database Preservation Toolkit

A web-based GUI for the DBPTK has been written in Python (backend) and Material Design Lite (frontend). The frontend for the GUI communicates with the Python backend which initiates the DBPTK by making system calls to the DBPTK jar-file using Java.

The GUI works as follows.

First, the user is presented with a page like the one shown below:



file:///home/andreas/eark/dbptk-gui-backend/angular-ui/index.htmlhttps://github.com/eark-project/dbptk-gui-backend.git

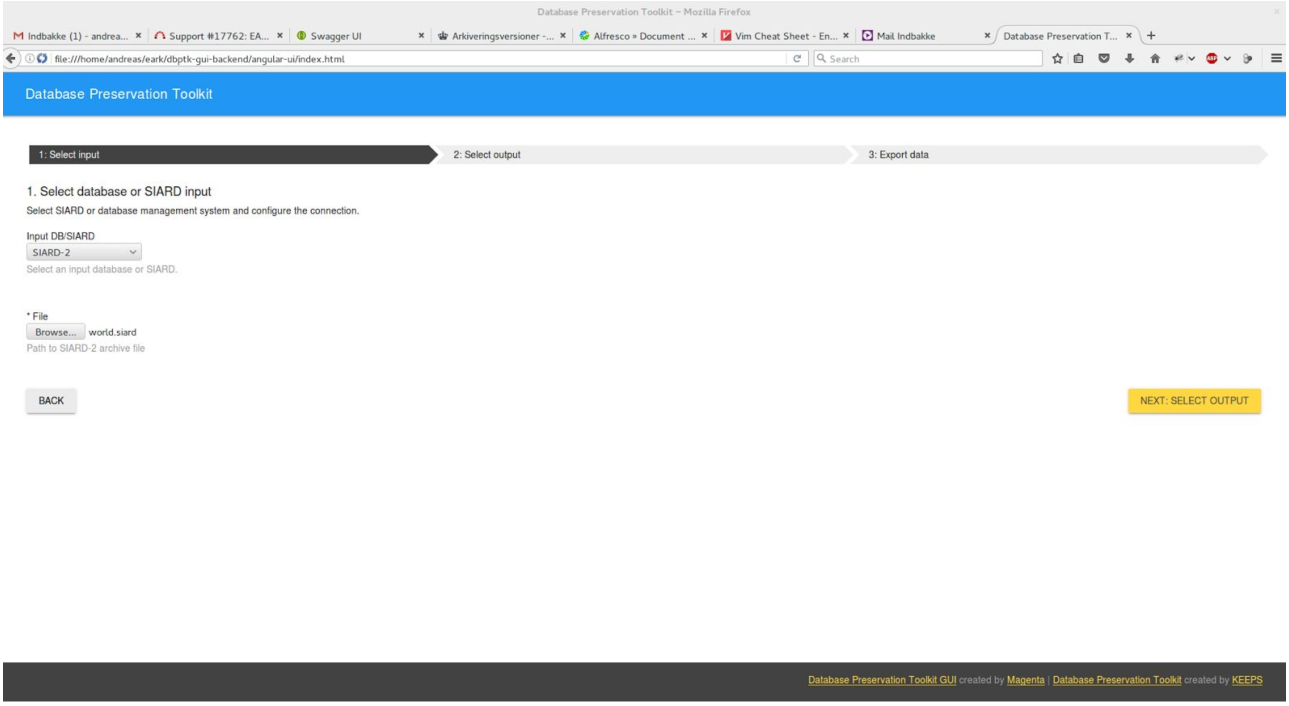
Connect to live DBMS or SIARD files and export to another live DBMS or SIARD.

Supported formats are SIARD-1, SIARD-2, SIARD-DK and DBMSes like Oracle, MySQL, PostgrSQL, and Microsoft SQL Server

Database Preservation Toolkit GUI created by Magenta | Database Preservation Toolkit created by KEEPS

Figure 69 – Database Preservation Toolkit GUI

Clicking the yellow button will load a page that enables the user to specify the import settings:


Database Preservation Toolkit – Mozilla Firefox
file:///home/andreas/eark/dbptk-gui-backend/angular-ui/index.html
Database Preservation Toolkit
1: Select input
2: Select output
3: Export data
BACK
NEXT: SELECT OUTPUT
Database Preservation Toolkit GUI created by Magenta | Database Preservation Toolkit created by KEEPS

1. Select database or SIARD input

Select SIARD or database management system and configure the connection.

Input DB/SIARD: SIARD-2

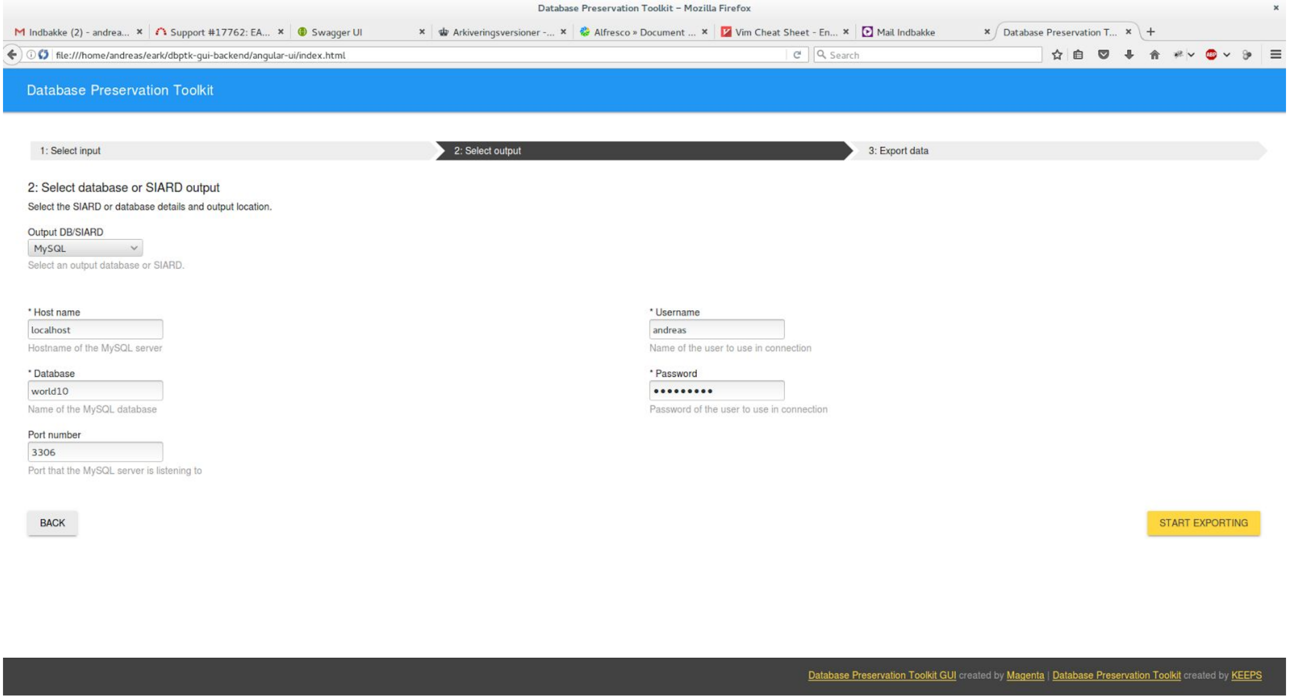
Select an input database or SIARD.

* File: Browse... world.siard

Path to SIARD-2 archive file

Figure 70 – DBPTK GUI import settings

The user can select where to import data from. As seen in the screenshot, the user has in this case selected to import data from a SIARD-2 archive file called “world.siard”. When clicking the “Next” button, the user will be directed to the export page, where (s)he can specify the export settings:


Database Preservation Toolkit – Mozilla Firefox
file:///home/andreas/eark/dbptk-gui-backend/angular-ui/index.html
Database Preservation Toolkit
1: Select input
2: Select output
3: Export data
BACK
START EXPORTING
Database Preservation Toolkit GUI created by Magenta | Database Preservation Toolkit created by KEEPS

2: Select database or SIARD output

Select the SIARD or database details and output location.

Output DB/SIARD: MySQL

Select an output database or SIARD.

* Host name: localhost

Hostname of the MySQL server

* Database: world10

Name of the MySQL database

Port number: 3306

Port that the MySQL server is listening to

* Username: andreas

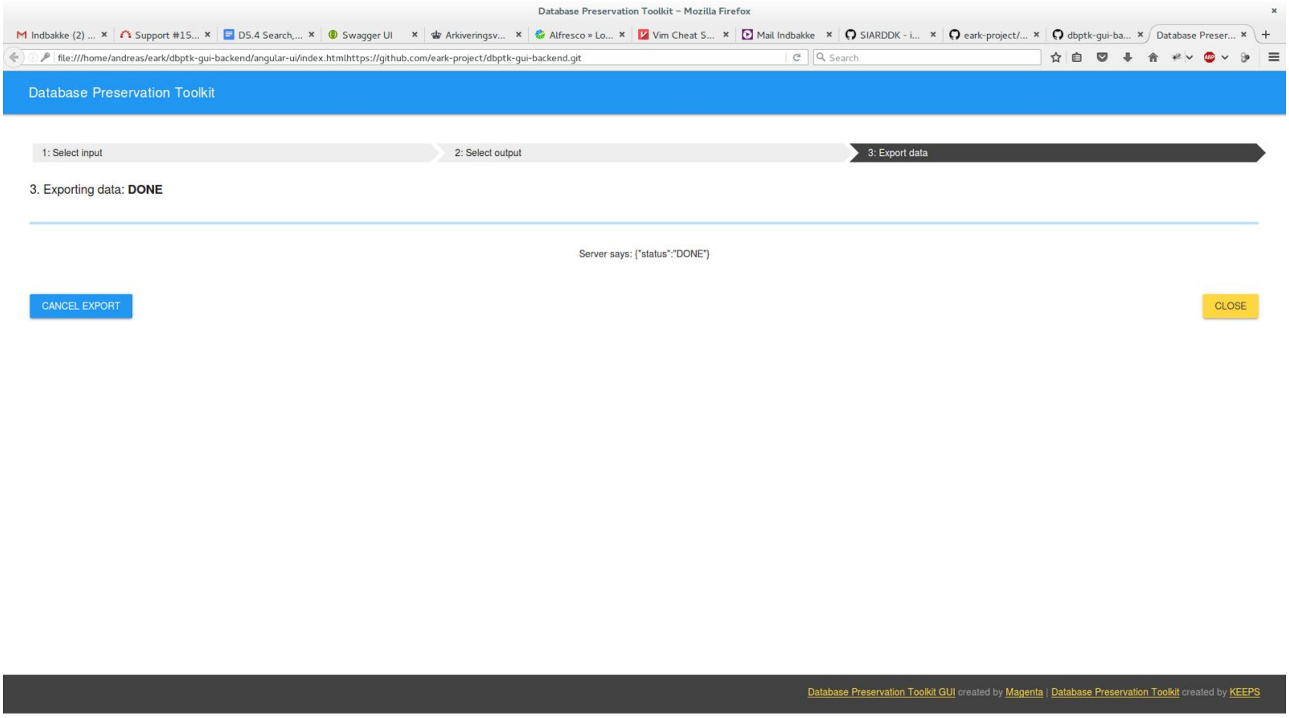
Name of the user to use in connection

* Password: •••••••••

Password of the user to use in connection

Figure 71 – DBPTK GUI export settings

In this example, the user has chosen to export to a MySQL DBMS and that the resulting database should be named “world10”. To start the export, the user clicks the “Start Exporting” button and when the export process has finished the user will be notified of this, as shown here:


Database Preservation Toolkit – Mozilla Firefox
Database Preservation Toolkit
1: Select input
2: Select output
3: Export data
CANCEL EXPORT
CLOSE

3. Exporting data: **DONE**

Server says: {"status":"DONE"}

Database Preservation Toolkit GUI created by Magenta | Database Preservation Toolkit created by KEEPS

Figure 72 – DBPTK GUI export notification

3.5.1.1.2 Code and documentation for the Database Preservation Toolkit and its GUI

The DBPTK is available at http://www.database-preservation.com/ along with general information, user guidelines and video tutorials on the tool. The source code is available at https://github.com/keeps/db-preservation-toolkit.

The code and documentation for the DBPTK GUI can be found at https://github.com/eark-project/dbptk-gui-backend

3.5.1.2 The Database Visualization Toolkit: Access via Solr (NoSQL-solution)

The Database Visualization Toolkit (DBVTK) was developed to provide a way for the designated community to access the archived databases without the need to go through the complex process of setting up a RDBMS and loading a SIARD file into it. This tool allows archivists and consumers to preview, explore and retrieve information from preserved databases. The software is aimed at end-users with little or no experience with SQL, and its main goal is to enable non-technical users to quickly find data of interest and provide a means to export and print these data.

The DBVTK is designed to be a scalable web-service that is able to serve multiple archived databases at once. It is optimized to provide almost instantaneous responses to searches on millions of database records. It uses a client-server approach to provide access to databases. Its three main components are illustrated in figure 3.7.1.2-1:

1. The web interface, with which the user interacts to access the database;
2. The server-side application, providing a business logic layer between the web interface and the data layer;
3. The data layer, where database information is stored and indexed.

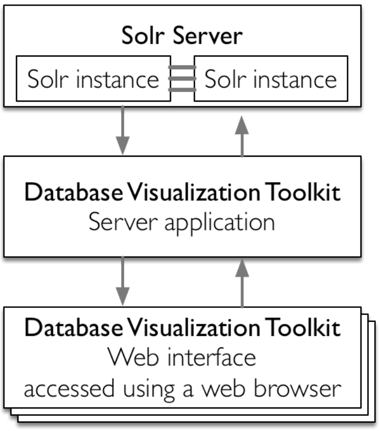

Solr Server
Solr instance
Solr instance
Database Visualization Toolkit
Server application
Database Visualization Toolkit
Web interface
accessed using a web browser

Figure 73 – Main components of the Database Visualization Toolkit

It would be difficult to build a scalable data layer using RDBMS technologies (e.g. MySQL), because they are not capable of handling tens of databases containing millions of records each. Therefore, Apache Solr (a NoSQL technology and indexing system) emerged as the chosen data layer platform. Apache Solr is an open source enterprise search platform. It is built for versatility, scalability, and ability to provide almost instantaneous responses to searching and filtering queries on millions of records.

The Solr platform is used to index preserved database records and provide searching and filtering functionality on those records. To handle the migration of databases in SIARD2 format to Solr, so they can be displayed by the Database Visualization Toolkit, the most fitting strategy was found to be the development of a Database Preservation Toolkit (DBPTK) Solr export module capable of loading databases into the Solr server.

For more advanced uses, like SQL queries and OLAP, the database can be exported back into a live and full-featured RDBMS, access to which can be provided using standard tools.

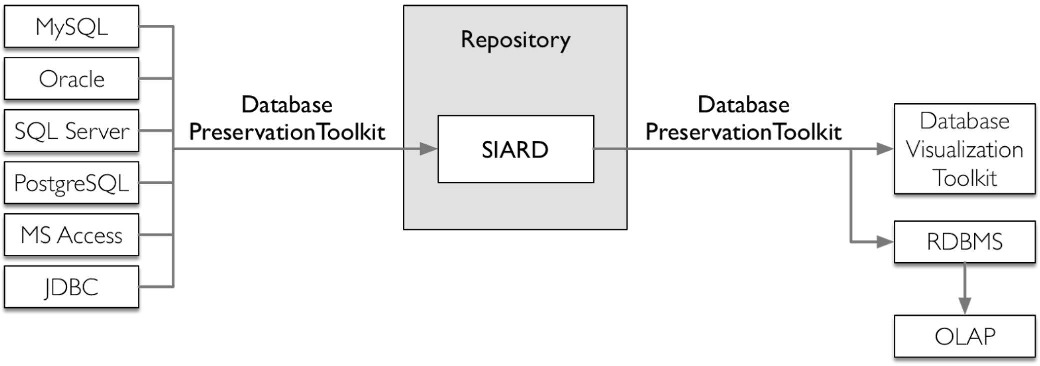
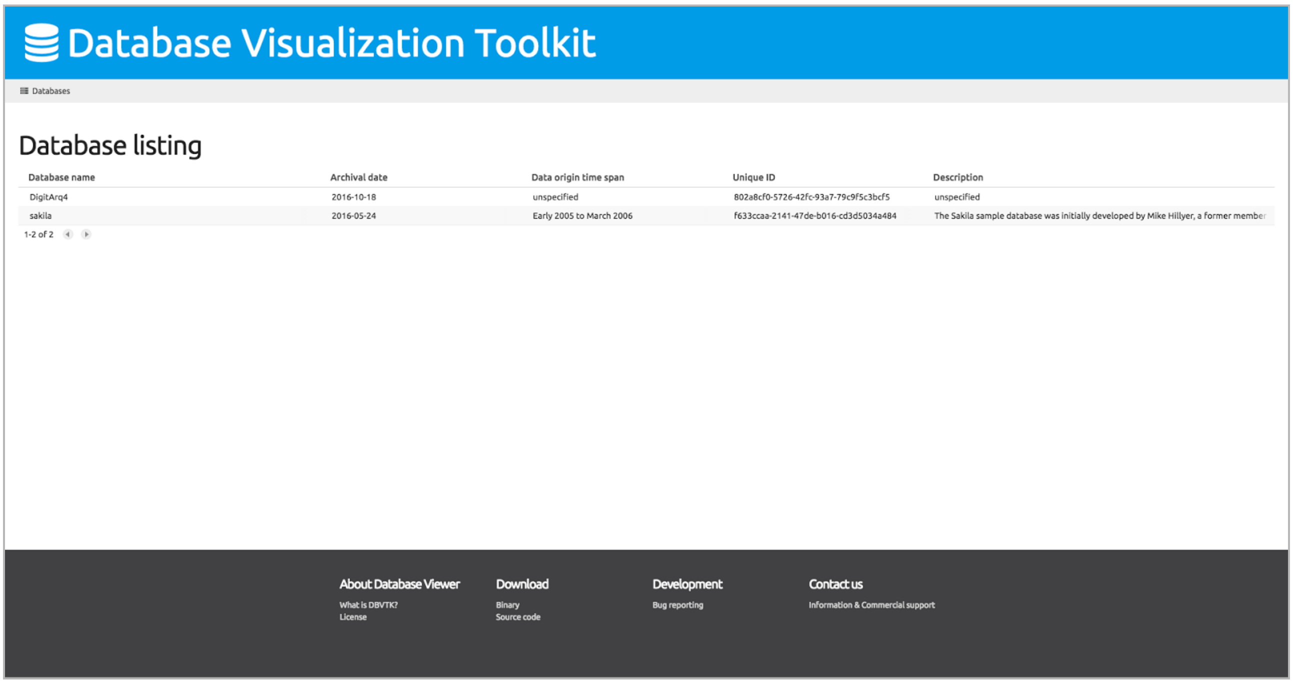

MySQL
Oracle
SQL Server
PostgreSQL
MS Access
JDBC
Database PreservationToolkit
Repository
SIARD
Database PreservationToolkit
Database Visualization Toolkit
RDBMS
OLAP

Figure 74 – Usage of the DBPTK and DBVTK in a repository

The DBVTK is then responsible for retrieving and rendering the information in the web interface.

Figure 72 shows the list of databases that are currently loaded in the system and ready to be accessed.



Database Visualization Toolkit

Database listing

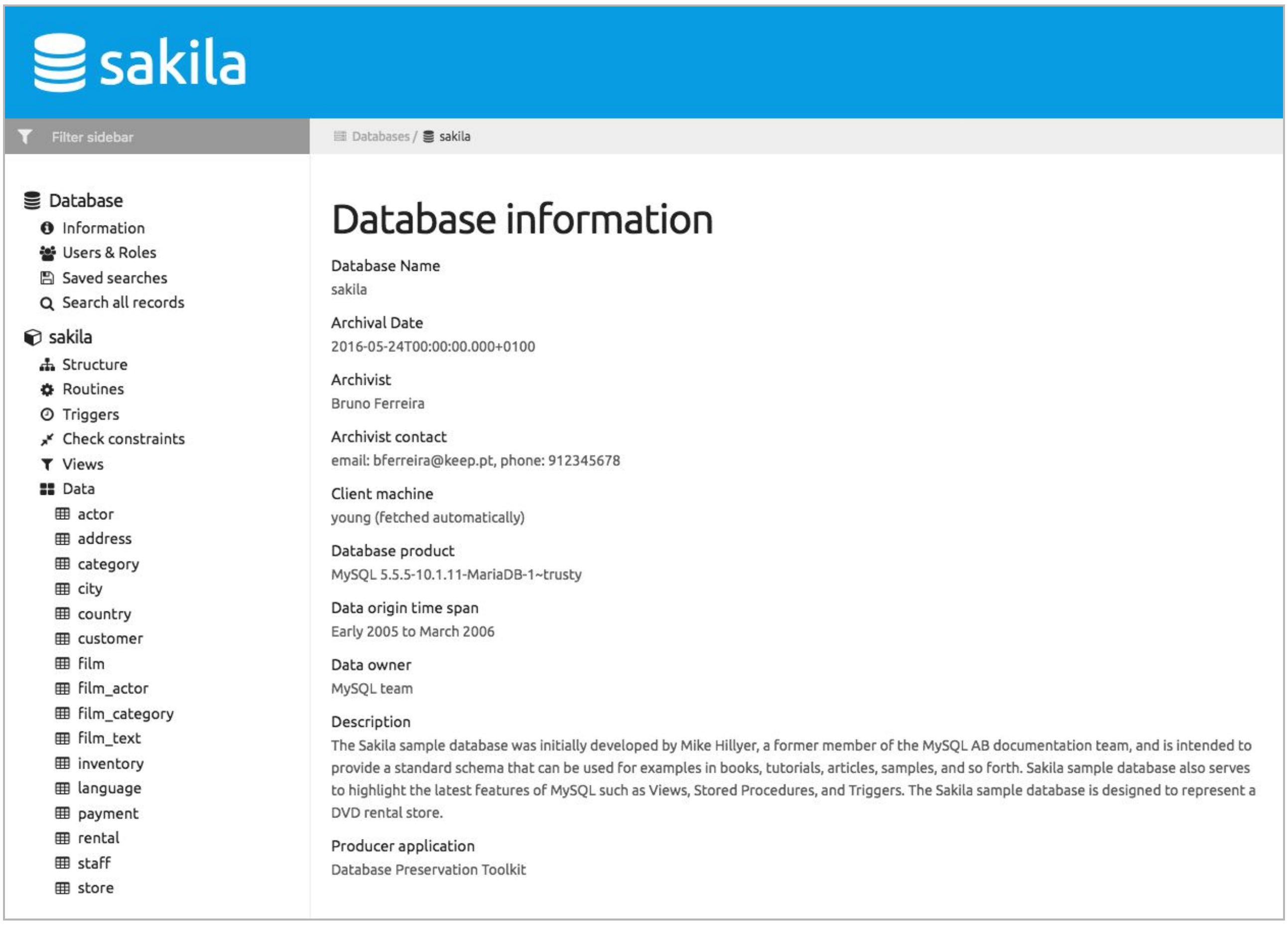
Database name	Archival date	Data origin time span	Unique ID	Description
DigitArq4	2016-10-18	unspecified	802a8cf0-5726-42fc-93a7-79c9f5c3bcf5	unspecified
sakila	2016-05-24	Early 2005 to March 2006	f633ccaa-2141-47de-b016-cd3d5034a484	The Sakila sample database was initially developed by Mike Hillyer, a former member

1-2 of 2

Figure 75 – List of databases

Clicking an item in that list shows descriptive metadata about the whole database, while a sidebar appears on the left side of the page with hyperlinks that allow the consumer to access information about other database elements. The hyperlinks direct the user to pages where metadata for different database elements is shown, because showing all the database metadata in a single web page would make the page bulky and difficult to read.

Figure 76 shows the view of full metadata about the database, which opens after clicking the hyperlink of a specific database in the list.



Database information

Database Name
sakila

Archival Date
2016-05-24T00:00:00.000+0100

Archivist
Bruno Ferreira

Archivist contact
email: bferreira@keep.pt, phone: 912345678

Client machine
young (fetched automatically)

Database product
MySQL 5.5.5-10.1.11-MariaDB-1~trusty

Data origin time span
Early 2005 to March 2006

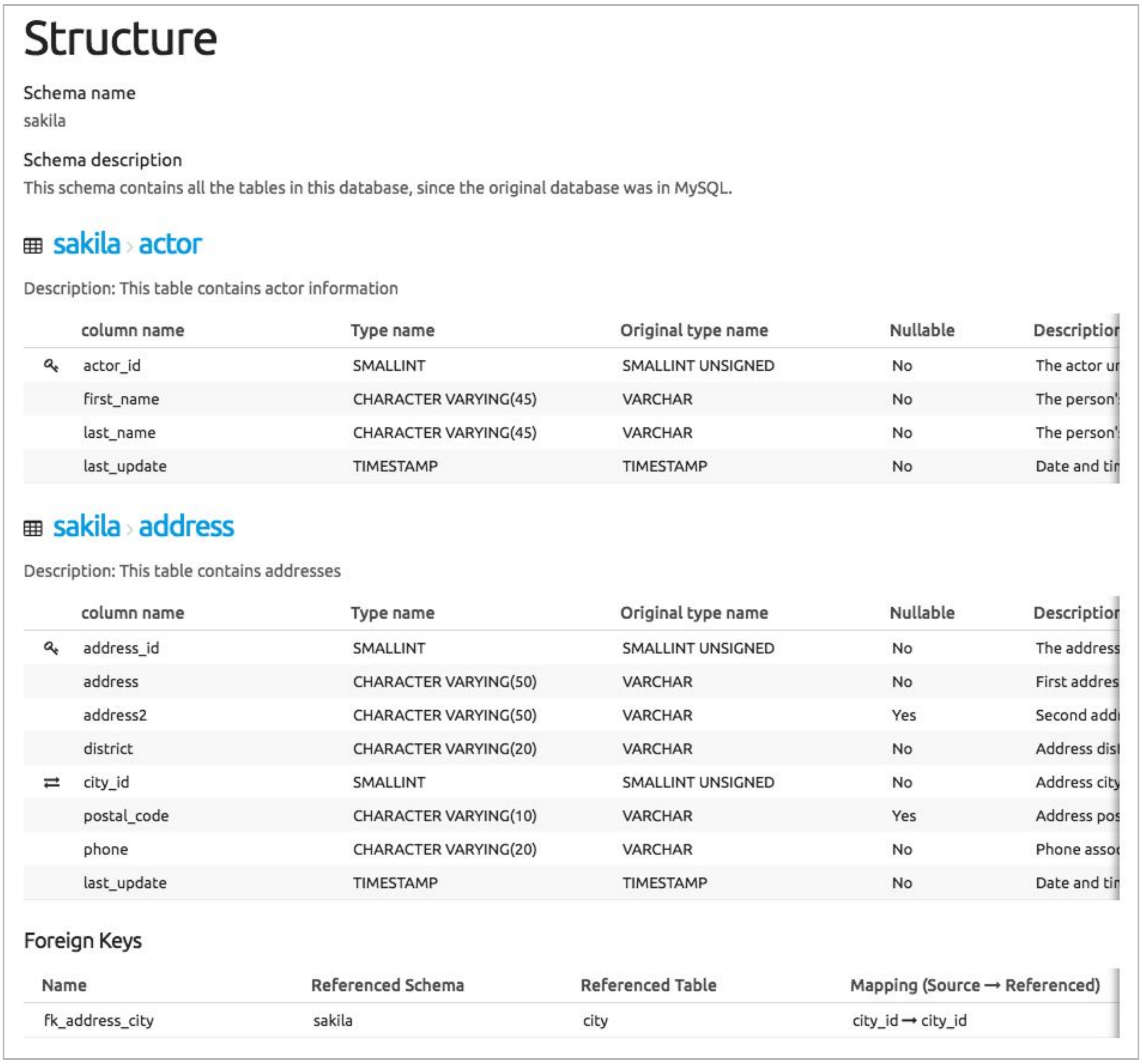
Data owner
MySQL team

Description
The Sakila sample database was initially developed by Mike Hillyer, a former member of the MySQL AB documentation team, and is intended to provide a standard schema that can be used for examples in books, tutorials, articles, samples, and so forth. Sakila sample database also serves to highlight the latest features of MySQL such as Views, Stored Procedures, and Triggers. The Sakila sample database is designed to represent a DVD rental store.

Producer application
Database Preservation Toolkit

Figure 76 – Descriptive metadata about a database

The following screenshot displays the database structure as shown in the DBVTK.



Structure

Schema name

sakila

Schema description

This schema contains all the tables in this database, since the original database was in MySQL.

sakila › actor

Description: This table contains actor information

	column name	Type name	Original type name	Nullable	Descriptio
🔑	actor_id	SMALLINT	SMALLINT UNSIGNED	No	The actor ur
	first_name	CHARACTER VARYING(45)	VARCHAR	No	The person'
	last_name	CHARACTER VARYING(45)	VARCHAR	No	The person'
	last_update	TIMESTAMP	TIMESTAMP	No	Date and tir

sakila › address

Description: This table contains addresses

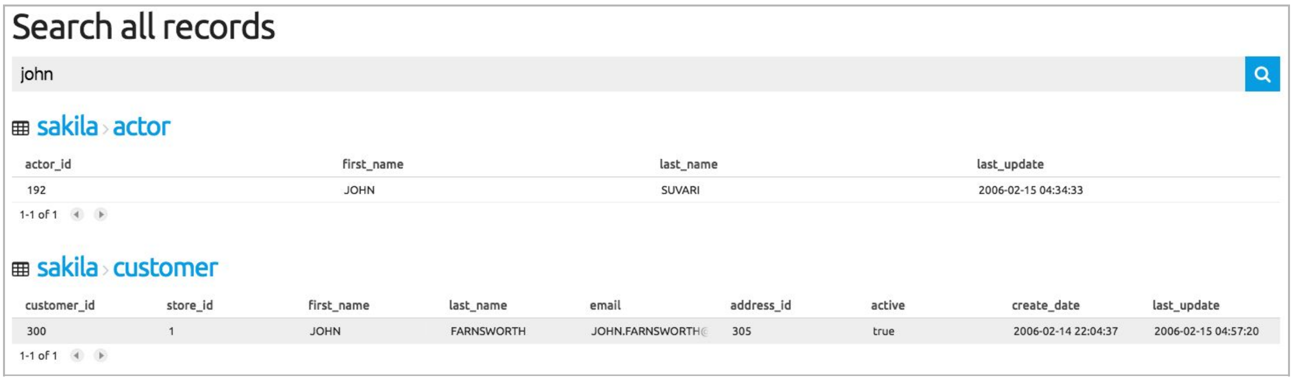
	column name	Type name	Original type name	Nullable	Descriptio
🔑	address_id	SMALLINT	SMALLINT UNSIGNED	No	The address
	address	CHARACTER VARYING(50)	VARCHAR	No	First addres
	address2	CHARACTER VARYING(50)	VARCHAR	Yes	Second addi
	district	CHARACTER VARYING(20)	VARCHAR	No	Address dis
⇄	city_id	SMALLINT	SMALLINT UNSIGNED	No	Address city
	postal_code	CHARACTER VARYING(10)	VARCHAR	Yes	Address pos
	phone	CHARACTER VARYING(20)	VARCHAR	No	Phone assoc
	last_update	TIMESTAMP	TIMESTAMP	No	Date and tir

Foreign Keys

Name	Referenced Schema	Referenced Table	Mapping (Source → Referenced)
fk_address_city	sakila	city	city_id → city_id

Figure 77 – Database structure metadata

Using the searching and filtering capabilities provided by Solr, and loading the database information in a way that makes the most of those capabilities, the DBVTK allows a consumer to search the database. The consumer can search the whole database, and doing so will show data grouped according to the table in which they appear (figure 78); or the consumer can search a specific table (figure 79).



Search all records

john

sakila › actor

actor_id	first_name	last_name	last_update
192	JOHN	SUVARI	2006-02-15 04:34:33

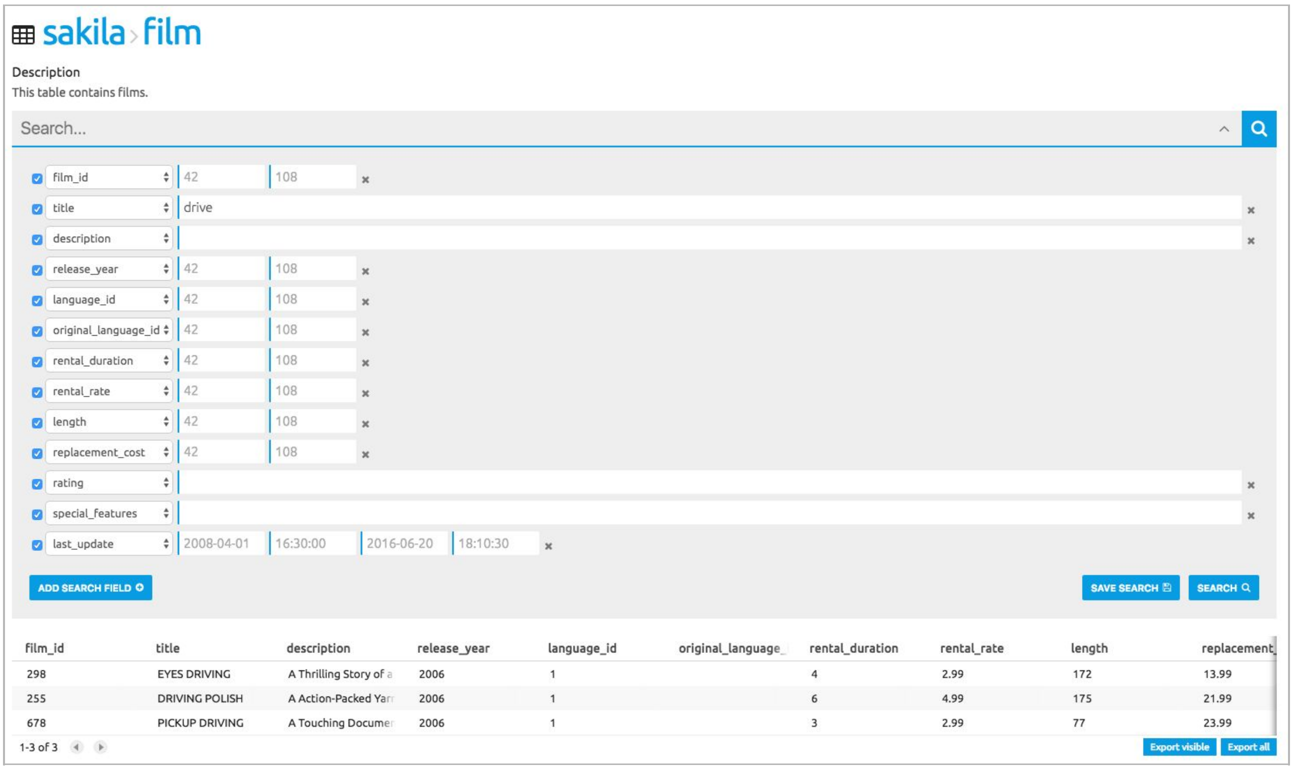
1-1 of 1

sakila › customer

customer_id	store_id	first_name	last_name	email	address_id	active	create_date	last_update
300	1	JOHN	FARNSWORTH	JOHN.FARNSWORTH	305	true	2006-02-14 22:04:37	2006-02-15 04:57:20

1-1 of 1

Figure 78 – Searching records in all tables of a database at once



sakila › film

Description

This table contains films.

Search...

☑ film_id	42	108		
☑ title	drive			
☑ description				
☑ release_year	42	108		
☑ language_id	42	108		
☑ original_language_id	42	108		
☑ rental_duration	42	108		
☑ rental_rate	42	108		
☑ length	42	108		
☑ replacement_cost	42	108		
☑ rating				
☑ special_features				
☑ last_update	2008-04-01	16:30:00	2016-06-20	18:10:30

film_id	title	description	release_year	language_id	original_language_	rental_duration	rental_rate	length	replacement_
298	EYES DRIVING	A Thrilling Story of a	2006	1		4	2.99	172	13.99
255	DRIVING POLISH	A Action-Packed Yarr	2006	1		6	4.99	175	21.99
678	PICKUP DRIVING	A Touching Documer	2006	1		3	2.99	77	23.99

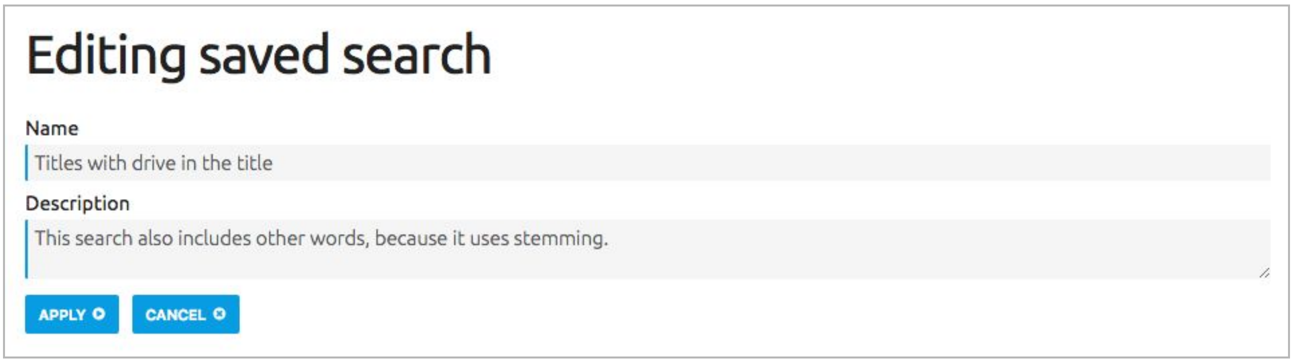
1-3 of 3

Figure 79 – Advanced search on a single table

The web interface also supports searching values in specific columns, either by providing an exact value or by defining a range of accepted values, as depicted in the figure above (79). This functionality is referred to as the advanced search.

The tables shown in the previous screenshots (78 and 79) also support sorting the results using the values of a column. Clicking the column header once sorts the values in ascending order and clicking it again sorts the values in descending order.

The advanced search (in figure 79) includes a button to save the search for later re-use. Upon saving the search, the user is shown a form to add a name and a description to the saved search. Upon submitting a name and description, the query will be displayed in a list containing the saved queries for the current database. Clicking an item on this list will open the URI that identifies this saved search, display the page to search a specific table and execute the search. This page shows the search results along with the original search parameters, allowing the consumer to change some of the search parameters and search again. This functionality is shown in figures 80 and 81 below.



Editing saved search

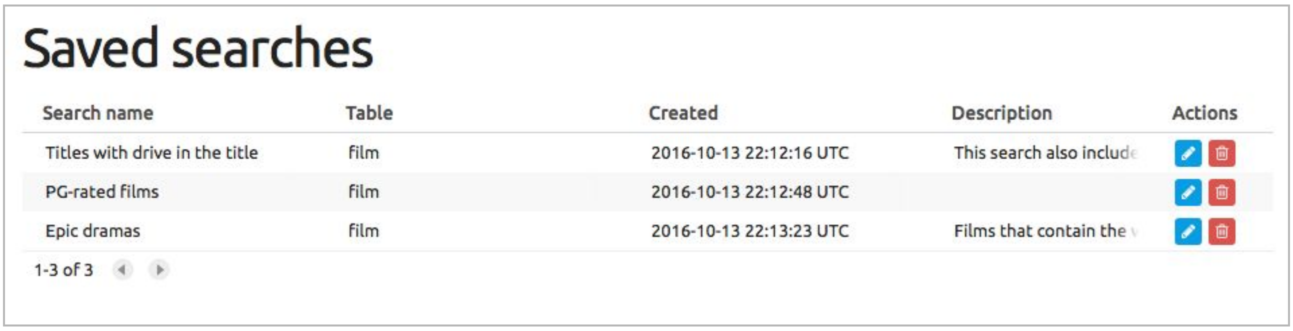
Name

Titles with drive in the title

Description

This search also includes other words, because it uses stemming.

Figure 80 – Creating or editing a saved search



Saved searches

Search name	Table	Created	Description	Actions
Titles with drive in the title	film	2016-10-13 22:12:16 UTC	This search also include	✎ 🗑
PG-rated films	film	2016-10-13 22:12:48 UTC		✎ 🗑
Epic dramas	film	2016-10-13 22:13:23 UTC	Films that contain the	✎ 🗑

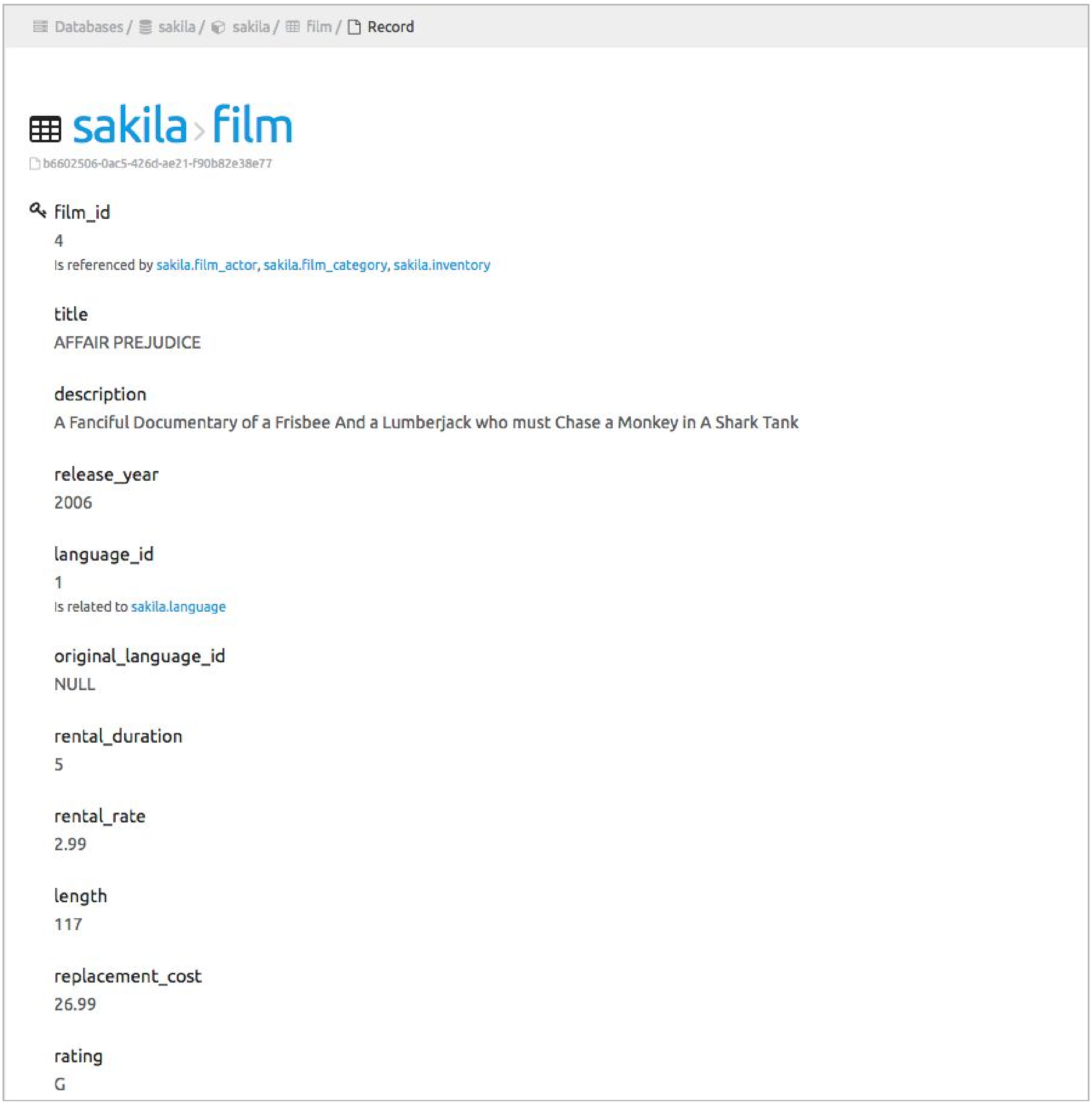
1-3 of 3

Figure 81 – Listing saved searches for a database

Below the search results, in figure 79, there are two buttons that provide the functionality to export the search results to CSV. There are two buttons because the search results are paginated, which means that only a subset of the results are shown and a button must be clicked to advance to the "next page" and view the next subset of results. One of the buttons exports the currently visible result subset to CSV and the other exports all the search results.

The search results pagination avoids overloading the server, by ensuring that just a few records are loaded at a time. Even when sending the CSV export, data is obtained and sent in chunks, to avoid loading all the data at once.

The single record page (figure 80) displays information in a list, to effectively show cell contents of various lengths. If the cell is related to other cells via a foreign key, by pointing to other records or by having other records point to it a link appears below the cell contents that allows the navigation to the related records via a specific foreign key. Clicking this link directs the user to the list of related records, or to a single record if there is only one related record. This functionality is depicted in figures 82 and 83.



sakila › film

b6602506-0ac5-426d-ae21-f90b82e38e77

film_id
4
Is referenced by sakila.film_actor, sakila.film_category, sakila.inventory

title
AFFAIR PREJUDICE

description
A Fanciful Documentary of a Frisbee And a Lumberjack who must Chase a Monkey in A Shark Tank

release_year
2006

language_id
1
Is related to sakila.language

original_language_id
NULL

rental_duration
5

rental_rate
2.99

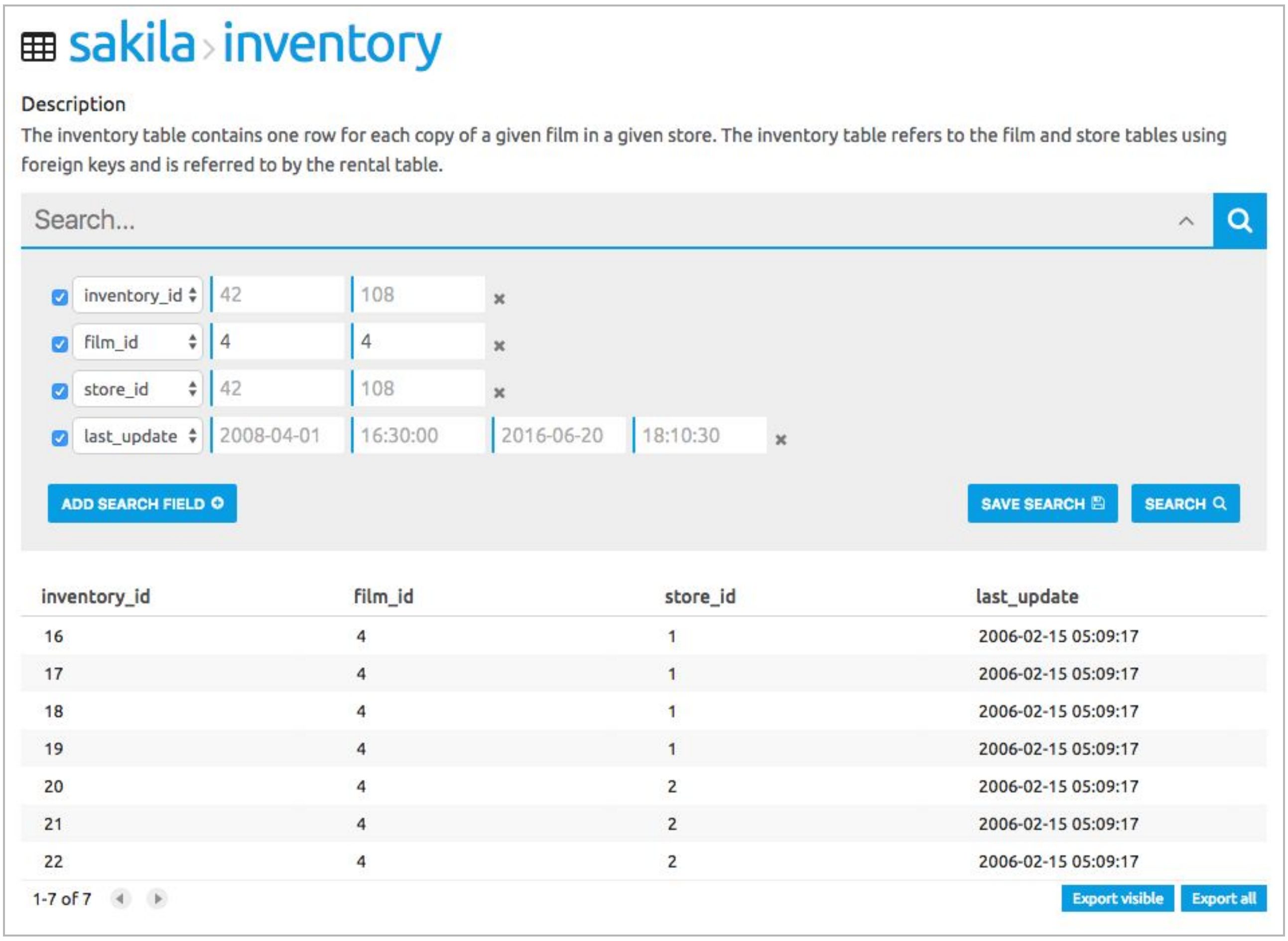
length
117

replacement_cost
26.99

rating
G

Figure 82 – Displaying a single row

The following screenshot shows a list of related records. It was obtained by clicking the relation in the column `film_id` to the table `sakila.inventory`, this performed a search for a specific value in the table `sakila.inventory`.



sakila › inventory

Description

The inventory table contains one row for each copy of a given film in a given store. The inventory table refers to the film and store tables using foreign keys and is referred to by the rental table.

Search...

☑ inventory_id 42 108

☑ film_id 4 4

☑ store_id 42 108

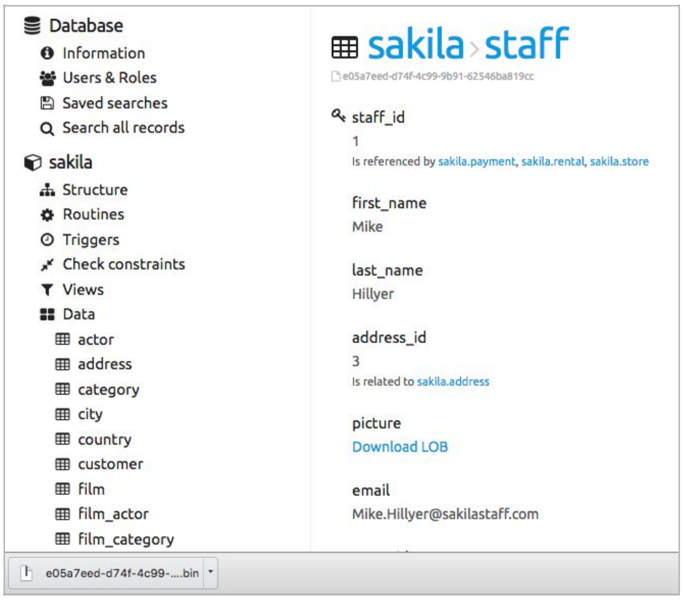
☑ last_update 2008-04-01 16:30:00 2016-06-20 18:10:30

inventory_id	film_id	store_id	last_update
16	4	1	2006-02-15 05:09:17
17	4	1	2006-02-15 05:09:17
18	4	1	2006-02-15 05:09:17
19	4	1	2006-02-15 05:09:17
20	4	2	2006-02-15 05:09:17
21	4	2	2006-02-15 05:09:17
22	4	2	2006-02-15 05:09:17

1-7 of 7

Figure 83 – Rows related to the row in figure 79 above

In the DBVTK web interface, the LOB[46] cells are displayed as a link that the consumer can click to download the corresponding LOB file. In the screenshot below (figure 84), the LOB present in column `picture` has been downloaded by clicking the "Download LOB" link.



sakila › staff

e05a7eed-d74f-4c99-9b91-62546ba819cc

staff_id
1
Is referenced by sakila.payment, sakila.rental, sakila.store

first_name
Mike

last_name
Hillyer

address_id
3
Is related to sakila.address

picture
Download LOB

email
Mike.Hillyer@sakilastaff.com

[46] LOB stands for Large Object, and is a data type for storing large objects. (LOB, Large object https://docs.oracle.com/cd/B10501_01/appdev.920/a97269/pc_16lob.htm)

Figure 84 – Downloading a LOB

Some databases that are submitted to an archive already include a data dictionary and an entity-relationship diagram, but it may not be explicitly stated in those documents which tables are most important in a database. The DBVTK is able to dynamically create and display a simple diagram about database tables and their relations.

The application dynamically generates this diagram, drawing each table as a circle, and connecting circles with arrows, representing the foreign key relations. To simplify identifying the most important tables, two highlighting methods are used: colour variation, and size variation. Tables containing more rows and columns are represented by larger circles, and tables with more relations are coloured darker. As an example, if the diagram contains a very big and dark circle, it probably corresponds to the main table in the database.

The highlighting methods adapt to the database information, mapping the provided database values (number of rows, columns and foreign key relations) to specific sizes and colours. The table with the most relations is coloured darkest, the table with the lowest number of relations is coloured lightest, and the other tables' colours are linearly distributed between the darkest and lightest colour according to their number of relations. To determine the size, the number of rows and columns for each table is obtained and modelled through a function that increases the size difference of the biggest and smallest tables in comparison to other tables, making them easier to distinguish from the "average sized" tables.

An example of this diagram is shown in the following screenshot:

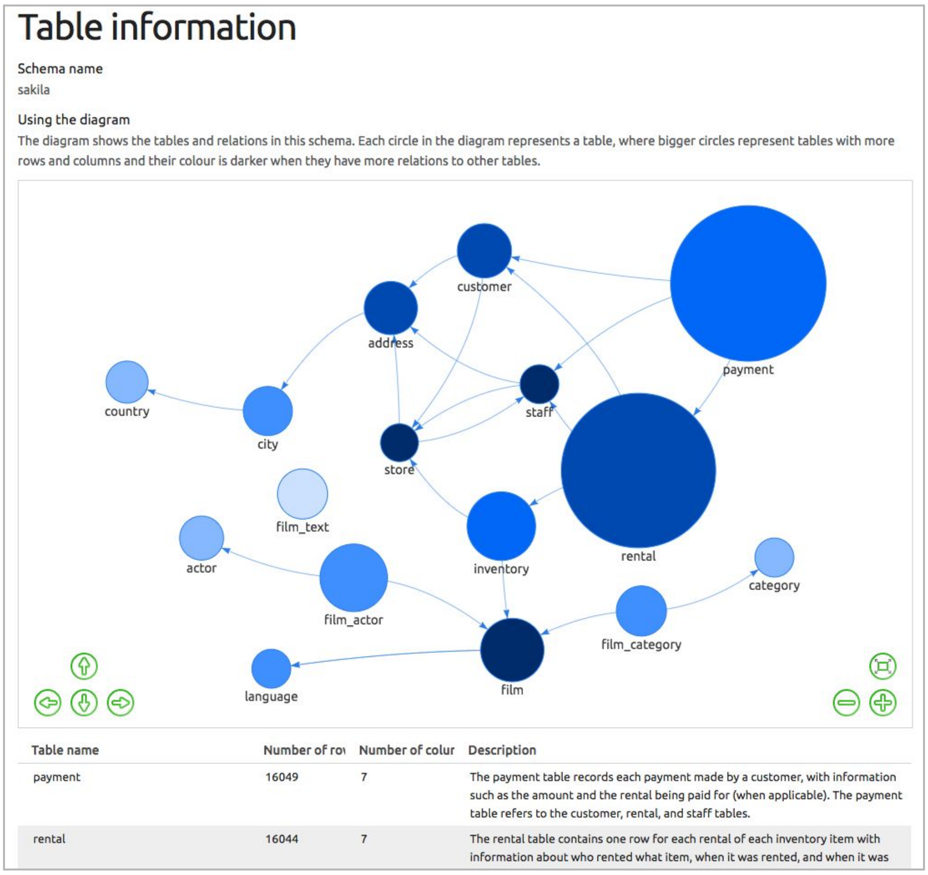


Table information

Schema name

sakila

Using the diagram

The diagram shows the tables and relations in this schema. Each circle in the diagram represents a table, where bigger circles represent tables with more rows and columns and their colour is darker when they have more relations to other tables.

Table name	Number of ro\	Number of colur	Description
payment	16049	7	The payment table records each payment made by a customer, with information such as the amount and the rental being paid for (when applicable). The payment table refers to the customer, rental, and staff tables.
rental	16044	7	The rental table contains one row for each rental of each inventory item with information about who rented what item, when it was rented, and when it was

Figure 85 – Diagram showing the most important tables and relations in a database

By looking at this diagram about the Sakila sample database, it is possible to identify that the tables "rental", "payment", "film", "staff" and "store" are the most important, and since the database has meaningful table names (i.e. they are not encoded) it is possible to guess that the database is about "a film rental store", which is indeed the purpose of the database.

3.5.1.2.1 Code and documentation for the Database Visualization Toolkit

The DBVTK is available at http://visualization.database-preservation.com/ along with general information on the tool. The source code is available at https://github.com/keeps/db-visualization-toolkit.

3.5.2 Geodata tools: Access to geodata stored in the SMURF format

Another content information type that the E-ARK project proposes tools for is geodata. All of the geodata tools that pertain to Access are already available open source tools that have been customized in order to fulfil E-ARK requirements.

The E-ARK geodata tools allow you to:

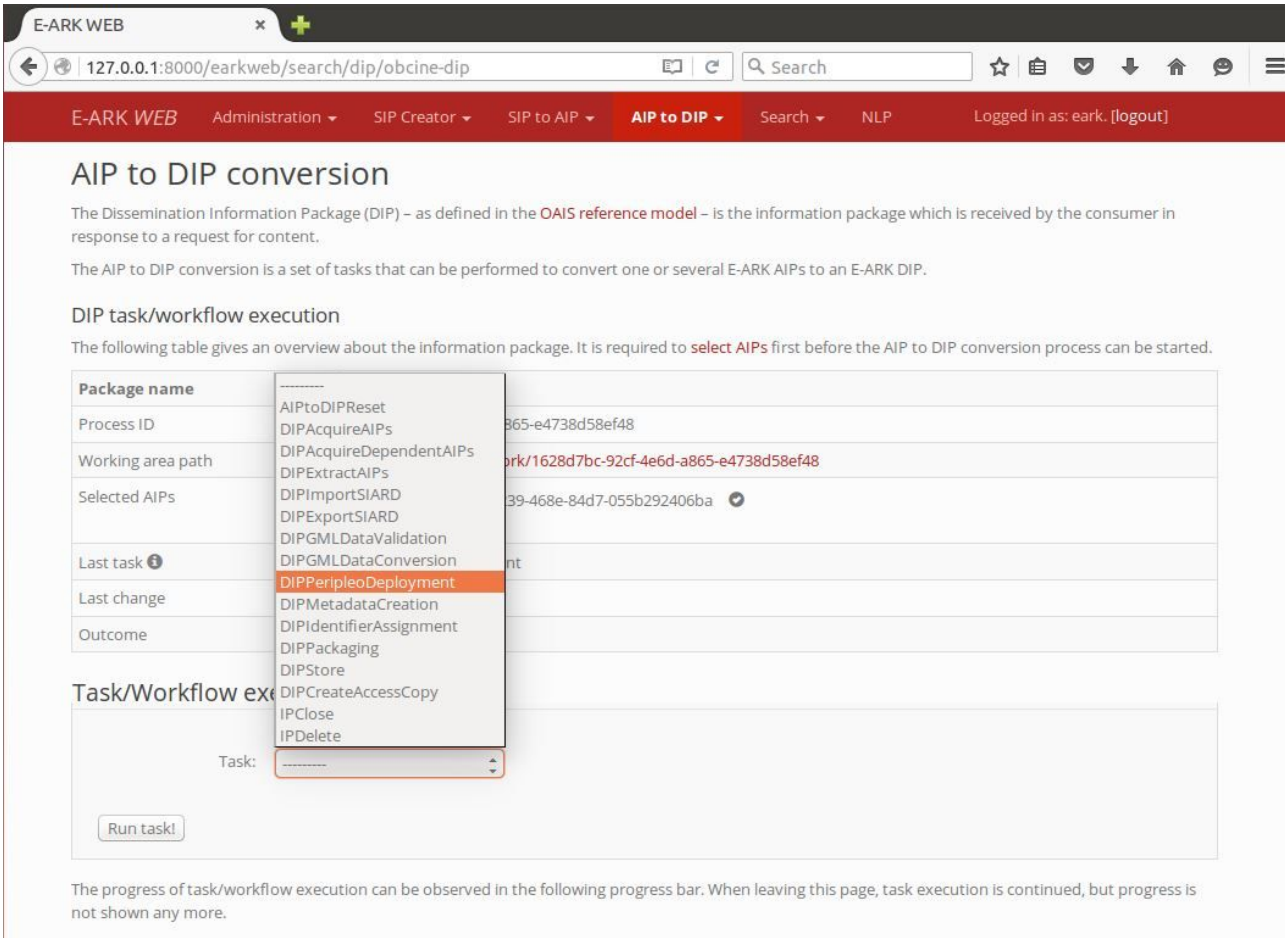
- Search geodata and display it on a map, filter the results according to time and/or topics (Peripleo[47])
- Transform geodata into from older to newer formats, or perform other preservation actions (QGIS[48])
- Display and query rendered geodata in a form of unstructured data or as a web service (QGIS and GeoServer[49])
- Manipulate data in order to replicate the content of an official document based on geodata (QGIS and GeoServer)

3.5.2.1 Geodata specific search

Peripleo is the E-ARK tool that enables geodata specific search. In order to search within Peripleo, it is necessary to create a separate spatial index that is the basis for executing geo-search functions. The spatial index can be created from within E-ARK Web.

To add a spatial index, select the “Active SIP to AIP conversion processes” option from the AIP to DIP section in E-ARK Web and in this section, select the IP package that you want to index.

[47] Cf. Glossary and Peripleo https://wiki.digitalclassicist.org/Peripleo
[48] Cf. Glossary and QGIS http://www.qgis.org
[49] Cf. Glossary and GeoServer http://geoserver.org/



AIP to DIP conversion

The Dissemination Information Package (DIP) – as defined in the OAIS reference model – is the information package which is received by the consumer in response to a request for content.

The AIP to DIP conversion is a set of tasks that can be performed to convert one or several E-ARK AIPs to an E-ARK DIP.

DIP task/workflow execution

The following table gives an overview about the information package. It is required to select AIPs first before the AIP to DIP conversion process can be started.

Package name	
Process ID	865-e4738d58ef48
Working area path	ork/1628d7bc-92cf-4e6d-a865-e4738d58ef48
Selected AIPs	39-468e-84d7-055b292406ba
Last task	nt
Last change	
Outcome	

Task/Workflow exe

Task: ---------

The progress of task/workflow execution can be observed in the following progress bar. When leaving this page, task execution is continued, but progress is not shown any more.

Figure 86 – Adding a spatial index to Peripleo from E-ARK web

In this window go to the "Task/Workflow execution" section and select the task named DIPPeripleoDeployment.

3.5.2.1.1 Example using Peripleo

Peripleo is a tool for rendering and searching geodata. It enables end-user to make a full-text search of the data, refining the results with temporal and spatial component. Results are rendered over a baselayer map, which can be selected by the end-user (three options are given). Peripleo's specialty is rendering an area not only as a point on a map, but also as a line or polygon feature[50].

Search is executed by the end-user entering a keyword into a search field. All search results are shown and the end-user can intuitively select those, he or she is interested in.

[50] Simon R. et al. http://journal.code4lib.org/articles/11144

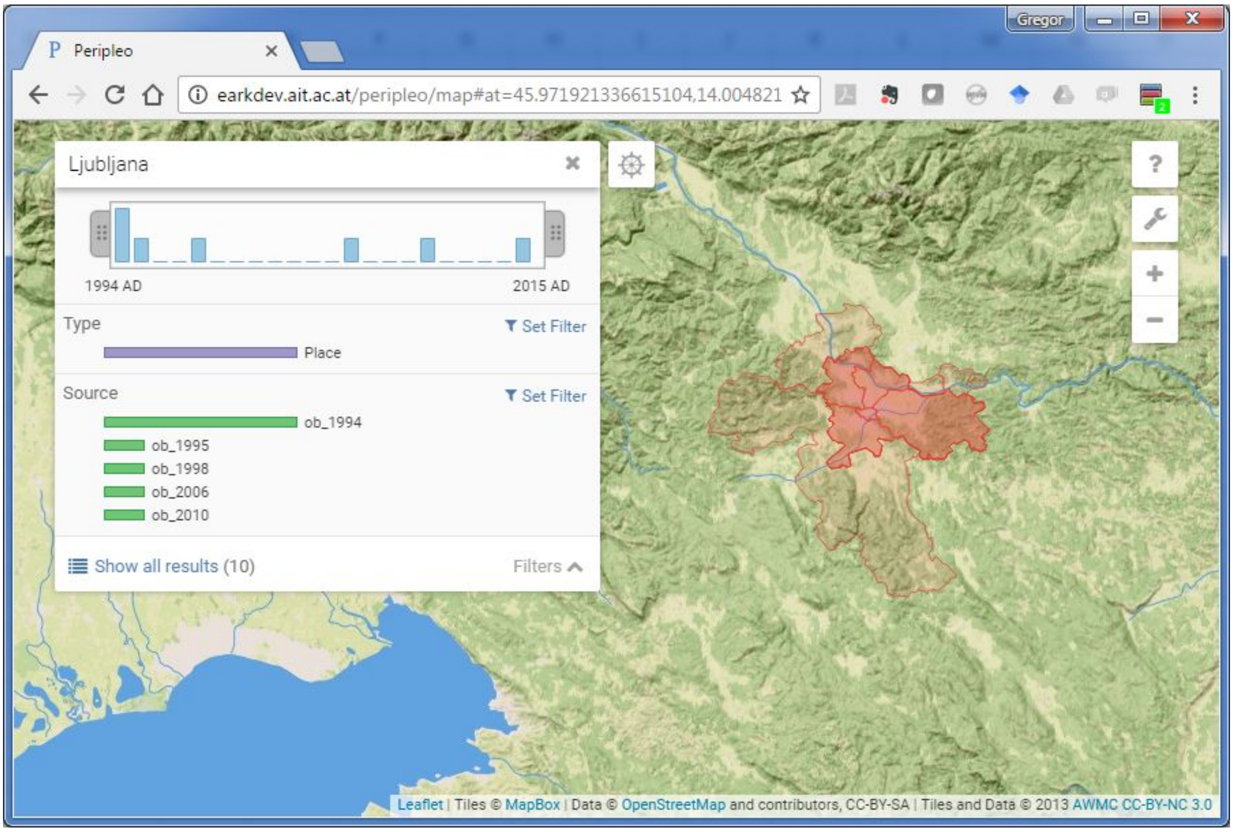

Peripleo
earkdev.ait.ac.at/peripleo/map#at=45.971921336615104,14.004821
Gregor
Ljubljana
1994 AD
2015 AD
Type
Set Filter
Place
Source
Set Filter
ob_1994
ob_1995
ob_1998
ob_2006
ob_2010
Show all results (10)
Filters
Leaflet | Tiles © MapBox | Data © OpenStreetMap and contributors, CC-BY-SA | Tiles and Data © 2013 AWMC CC-BY-NC 3.0

Figure 87 – Peripleo interface with search result for keyword Ljubljana

The result page offers several refining methods. In the present search case the end-user can refine search results by selecting specific time period, select a type or select from available sources listed in the results.

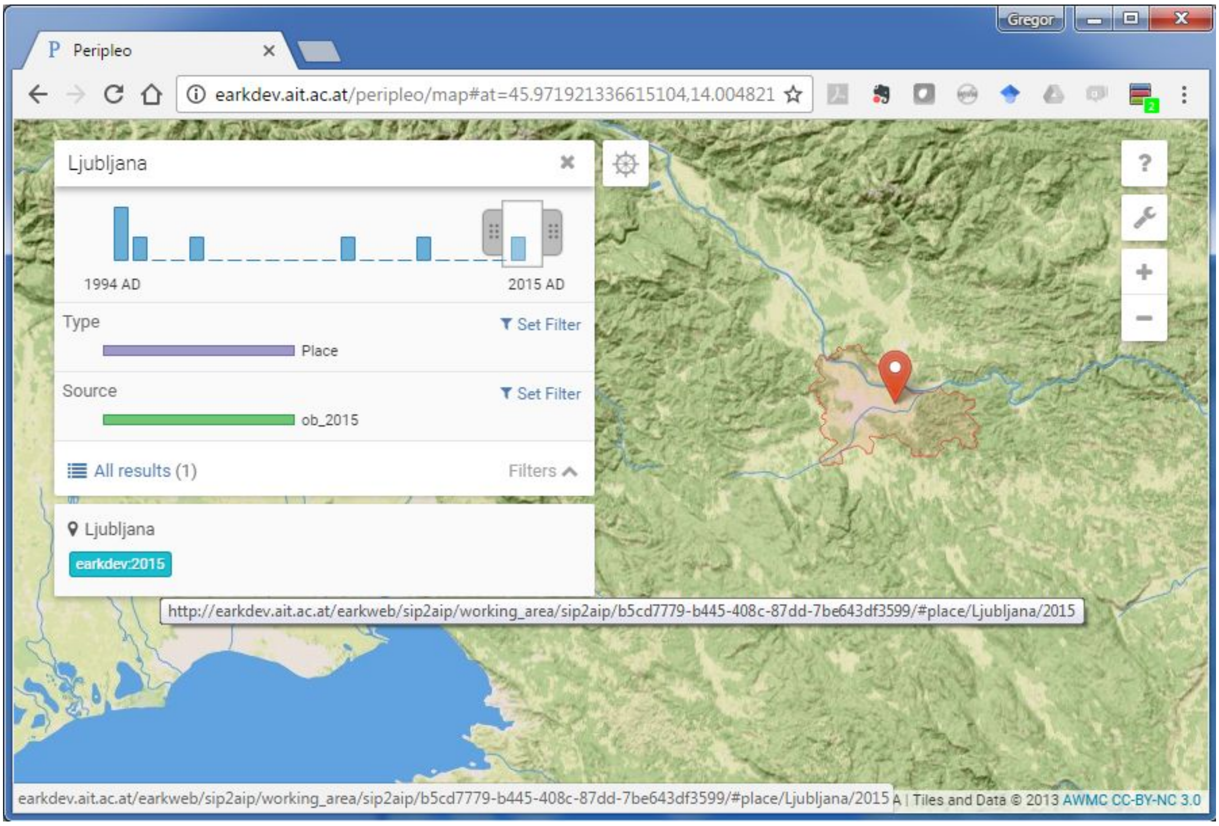
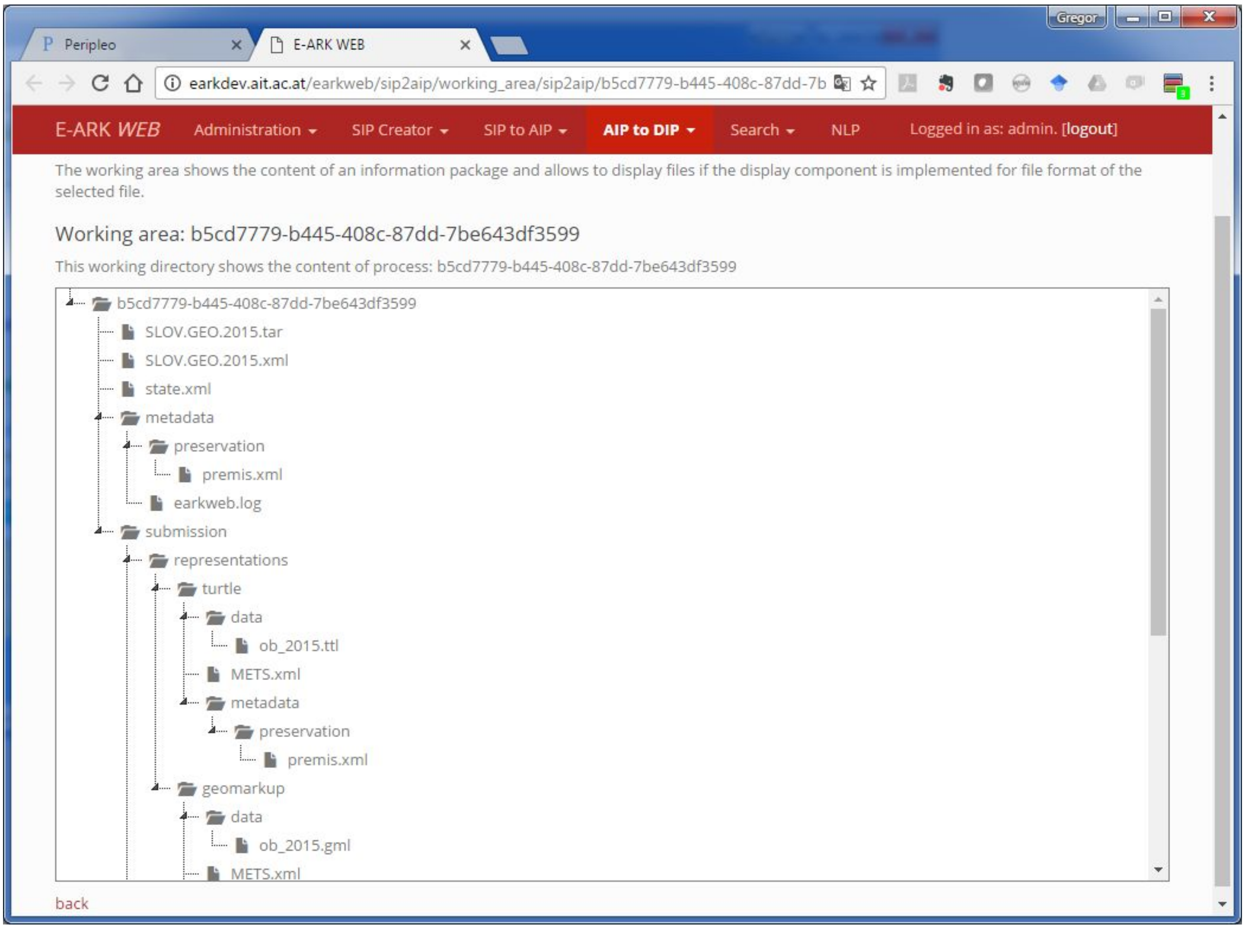

Peripleo
earkdev.ait.ac.at/peripleo/map#at=45.971921336615104,14.004821
Gregor
Ljubljana
1994 AD
2015 AD
Type
Set Filter
Place
Source
Set Filter
ob_2015
All results (1)
Filters
Ljubljana
earkdev:2015
http://earkdev.ait.ac.at/earkweb/sip2aip/working_area/sip2aip/b5cd7779-b445-408c-87dd-7be643df3599/#place/Ljubljana/2015
earkdev.ait.ac.at/earkweb/sip2aip/working_area/sip2aip/b5cd7779-b445-408c-87dd-7be643df3599/#place/Ljubljana/2015
Tiles and Data © 2013 AWMC CC-BY-NC 3.0

Figure 88 – Peripleo interface with refined search results for the year 2015 and the link to archival package

Once filters are set, e.g. a specific time range, specific results are shown in the bottom of the search window (in our case: "Ljubljana - earkdev:2015"). By clicking on the result the link leads to the indexed AIPs and/or DIPs available in E-ARK Web. Currently Peripleo is configured to work as a tool embedded within E-ARK Web, however it can be configured to work with other systems. The full integration with the Access Software Platform (ASP) is not developed at the moment, however results from E-ARK Web can be used to identify the archival package in the ASP.

In general this tool is now used to search and browse existing DIP_u and DIP_p[51] packages, however if the archival package containing geodata is indexed right after the AIP creation in E-ARK Web, than the end-user can browse those packages too.


Gregor
Peripleo
E-ARK WEB
earkdev.ait.ac.at/earkweb/sip2aip/working_area/sip2aip/b5cd7779-b445-408c-87dd-7b
E-ARK WEB
Administration
SIP Creator
SIP to AIP
AIP to DIP
Search
NLP
Logged in as: admin. [logout]
back

The working area shows the content of an information package and allows to display files if the display component is implemented for file format of the selected file.

Working area: b5cd7779-b445-408c-87dd-7be643df3599

This working directory shows the content of process: b5cd7779-b445-408c-87dd-7be643df3599

```
b5cd7779-b445-408c-87dd-7be643df3599
    SLOV.GEO.2015.tar
    SLOV.GEO.2015.xml
    state.xml
    metadata
        preservation
            premis.xml
        earkweb.log
    submission
        representations
            turtle
                data
                    ob_2015.ttl
                METS.xml
                metadata
                    preservation
                        premis.xml
            geomarkup
                data
                    ob_2015.gml
                METS.xml
```

Figure 89 – E-ARK Web interface showing the DIP in the working area of E-ARK Web, selected in Peripleo

3.5.2.2 Access to geodata

Archives store geodata in a form of unstructured files. Since some users are not experienced enough to understand and use dedicated GIS software to view and analyse this information, archivists can render geodata as a web service and offer access via web application. Four specific scenarios for accessing geodata are described below. They vary depending on use experience and complexity of the requested geodata.

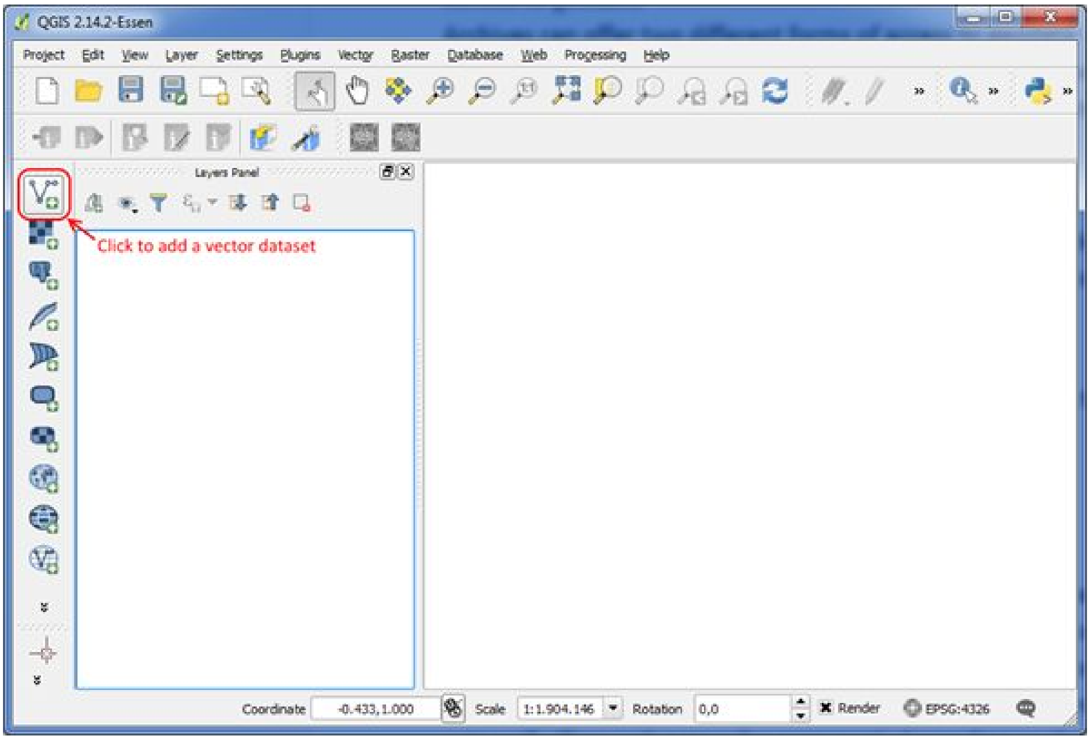
[51] The DIPu ('u' for user) being the DIP prepared for end-user consultation, as opposed to the DIP0 ('0' for 'not prepared'), which is an unprocessed copy of an AIP, and as opposed to the DIPp ('p' for permanent), which is a DIP that is stored in DIP storage for further consultation.

3.5.2.2.1 Unstructured files

Advanced end-users such as geodata experts can work with raw data using QGIS[52]. They can work with QGIS in the reading room using the archive's infrastructure or order a copy of a specific set of data to use within their own tools outside the reading room, according to the archive's policy.

3.5.2.2.1.1 Example using Geotools - Opening a vector dataset in QGIS

After the user has found and ordered the package containing geodata, and the order is prepared for download or access in the user area, the user can view and use geodata in QGIS, the user must run the application first. After opening the QGIS application, an empty project is opened. The user must then add a dataset to the project in QGIS. To add data into the project, users need to select the appropriate button on the left side of the window. There are many possibilities ranging from adding geodata in vector or raster format, forms of various web services or adding data from databases (like PostGIS). In order to add a GML vector dataset, users choose the first button as shown in the image below:

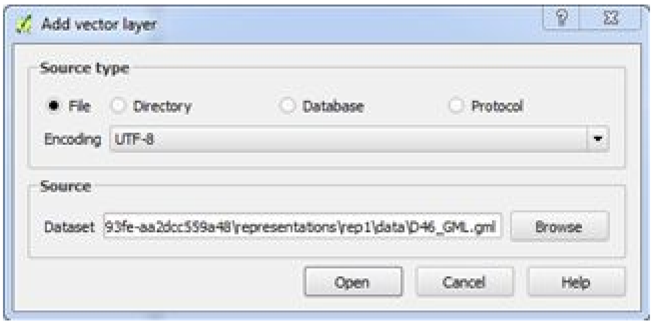


Coordinate -0.433,1.000 Scale 1:1.904.146 Rotation 0,0 ☒ Render EPSG:4326

Figure 90 – QGIS interface, showing a button for adding a vector layer to the map

After this click a new window opens in which users enter the source type (file) and source of the dataset. Users also need to choose the encoding codepage (in this case UTF-8).

[52] QGIS is an opensource GIS application used for rendering and analysing geodata - more on QGIS http://qgis.org/en/site/about/index.html



Add vector layer

Source type

● File ○ Directory ○ Database ○ Protocol

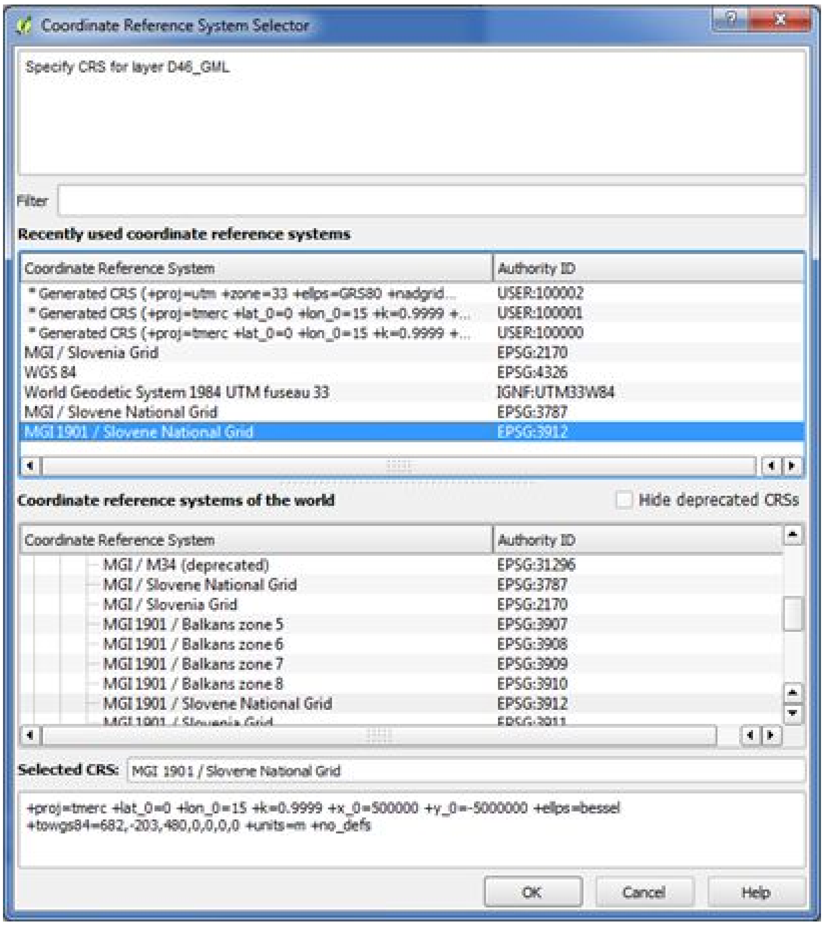
Encoding: UTF-8

Source

Dataset: 93fe-aa2dcc559a48\representations\rep1\data\D46_GML.gml

Figure 91 – The Add vector layer window in QGIS

Next step is to assign the coordinate reference system for this file (EPSG:3912[53] here).



Coordinate Reference System Selector

Specify CRS for layer D46_GML

Filter

Recently used coordinate reference systems

Coordinate Reference System	Authority ID
* Generated CRS (+proj=utm +zone=33 +ellps=GRS80 +nadgrid...	USER:100002
* Generated CRS (+proj=tmerc +lat_0=0 +lon_0=15 +k=0.9999 +...	USER:100001
* Generated CRS (+proj=tmerc +lat_0=0 +lon_0=15 +k=0.9999 +...	USER:100000
MGI / Slovenia Grid	EPSG:2170
WGS 84	EPSG:4326
World Geodetic System 1984 UTM fuseau 33	IGNF:UTM33W84
MGI / Slovene National Grid	EPSG:3787
MGI 1901 / Slovene National Grid	EPSG:3912

Coordinate reference systems of the world ☐ Hide deprecated CRSs

Coordinate Reference System	Authority ID
MGI / M34 (deprecated)	EPSG:31296
MGI / Slovene National Grid	EPSG:3787
MGI / Slovenia Grid	EPSG:2170
MGI 1901 / Balkans zone 5	EPSG:3907
MGI 1901 / Balkans zone 6	EPSG:3908
MGI 1901 / Balkans zone 7	EPSG:3909
MGI 1901 / Balkans zone 8	EPSG:3910
MGI 1901 / Slovene National Grid	EPSG:3912
MGI 1901 / Slovenia Grid	EPSG:3911

Selected CRS: MGI 1901 / Slovene National Grid

```
+proj=tmerc +lat_0=0 +lon_0=15 +k=0.9999 +x_0=500000 +y_0=-5000000 +ellps=bessel
+towgs84=682,-203,480,0,0,0,0 +units=m +no_defs
```

Figure 92 – Coordinate Reference System selector window in QGIS

[53] The coordinate reference system is defined within the GML file and within metadata. The user can inspect the GML file prior to adding it into the GIS Viewer.

After clicking OK, the GML vector dataset is added to the project and displayed in QGIS.

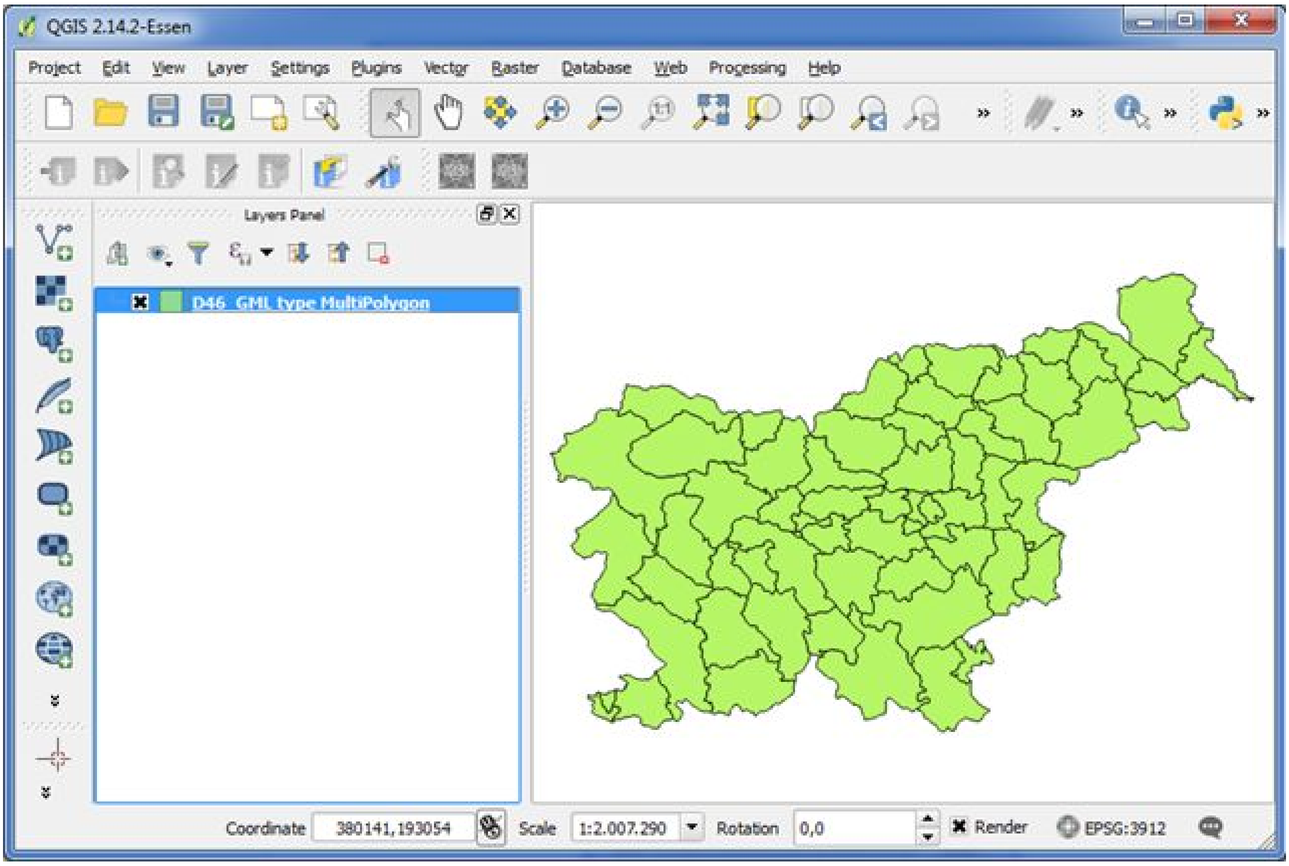

QGIS 2.14.2-Essen
Project Edit View Layer Settings Plugins Vector Raster Database Web Processing Help
Layers Panel
D46_GML type MultiPolygon
Coordinate 380141,193054
Scale 1:2.007.290
Rotation 0,0
Render
EPSG:3912

Figure 93 – QGIS interface showing the added layer in the map view

More information on working with vector data is available from online QGIS documentation (http://docs.qgis.org/2.14/en/docs/training_manual/basic_map/vector_data.html)

3.5.2.2.2 Web service

Serving geodata through OGC[54] web services, though technically more demanding, enables a broader access to geodata and enforces greater control over how data is accessed and manipulated; it can also manage access for reuse. It brings geodata to those without specific knowledge of working with geodata. Access can be made possible via a Web or mobile application (a login or some sort of authentication is needed) or externally (free web access for everyone), depending on the archival policy.

As a possible implementation, archives could set up a GeoServer[55] as the baseline architecture (or any other commercial GIS server). If users order geodata, the appropriate data is loaded into the GeoServer, and the appropriate link to the web service is forwarded to the user.

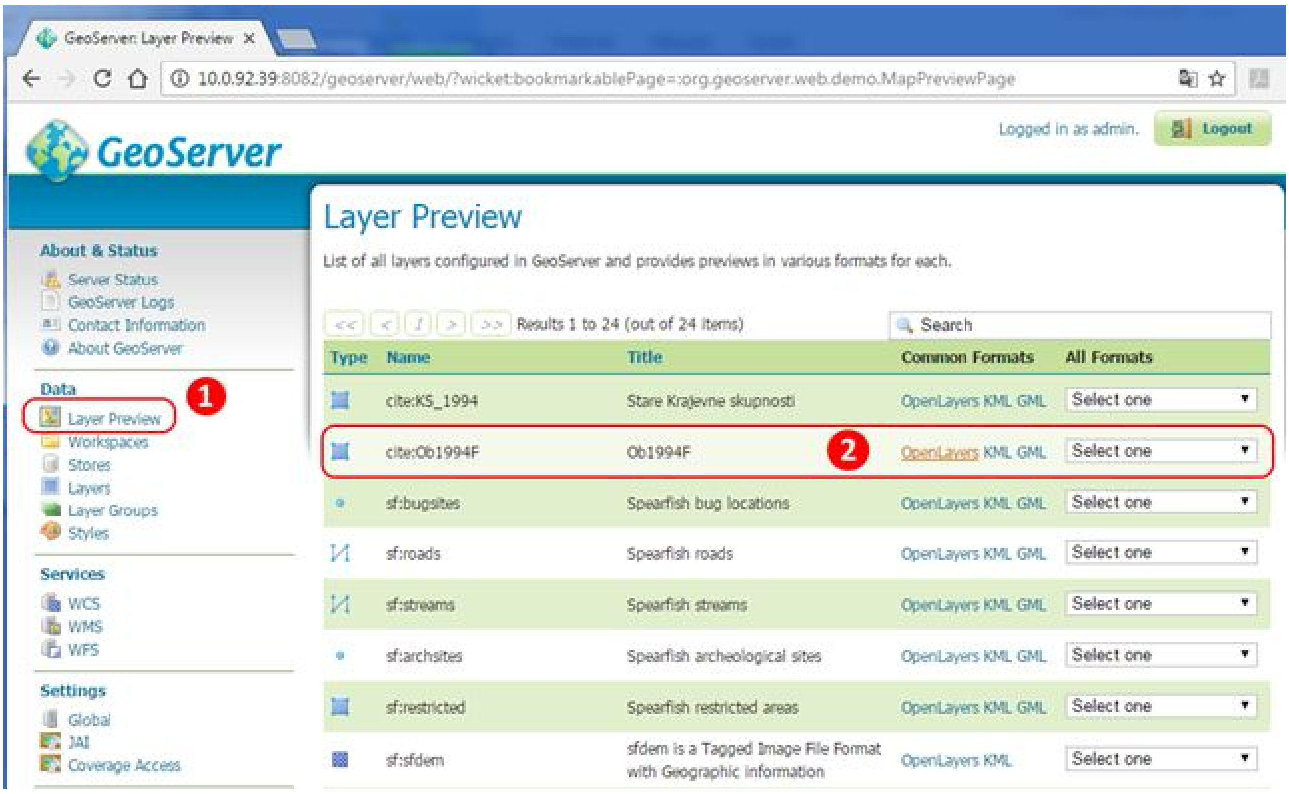
The web service allows the user to view layers of geodata, and might allow to even recreate maps from multiple layers of geodata and add query access - all without the need to set up a complex access environment him-/herself.

[54] OGC, Open Geospatial Consortium http://www.opengeospatial.org/

[55] GeoServer http://geoserver.org/

3.5.2.2.2.1 Example using Geoserver - providing a link to geodata viewer based on Geoserver

After a geodata package has been ordered, within Order Management, the archivist extracts the GML file from the DIP0 and publishes it as a layer within GeoServer. Than a link can be retrieved from the Geoserver management console as shown in the image below.



Layer Preview

List of all layers configured in GeoServer and provides previews in various formats for each.

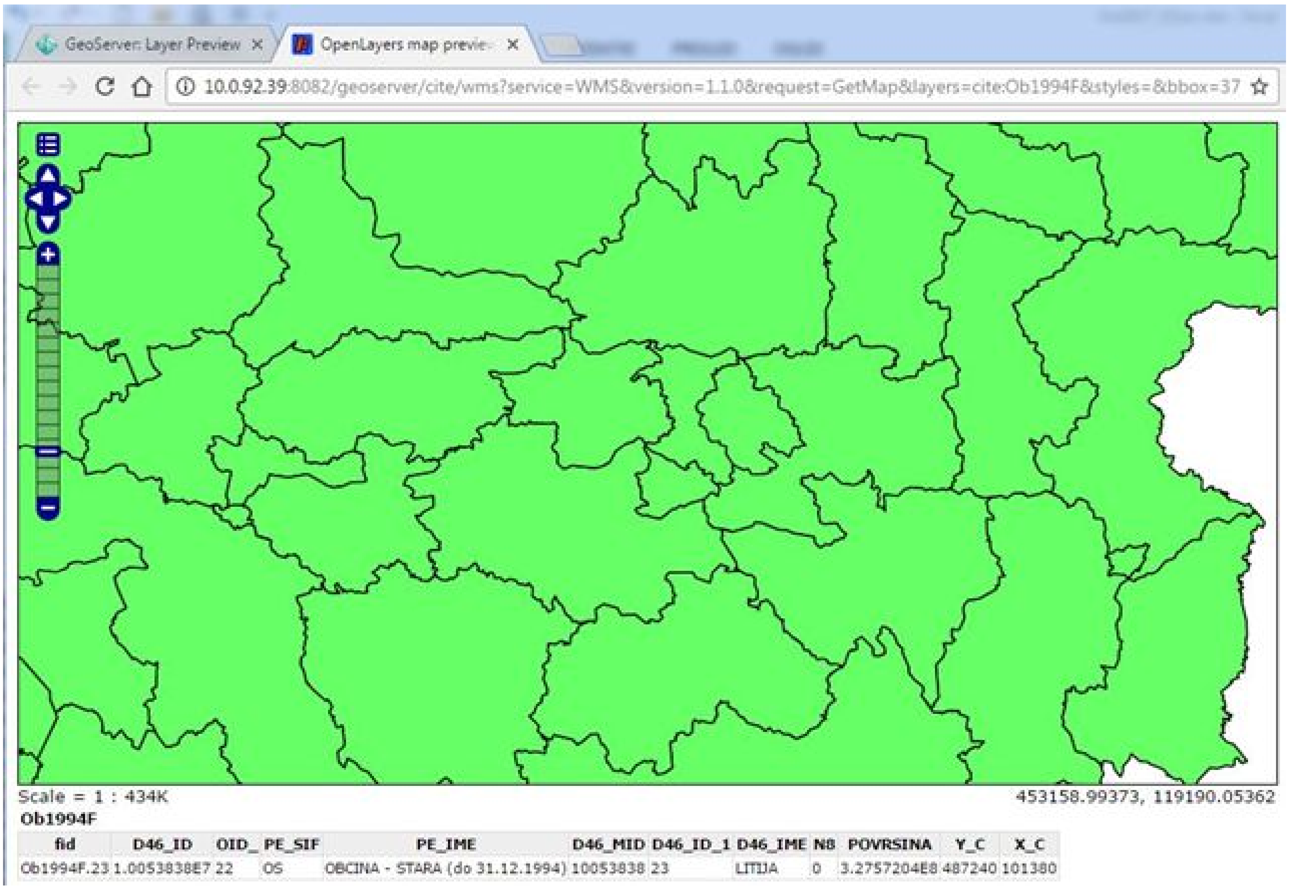
Results 1 to 24 (out of 24 items)

Type	Name	Title	Common Formats	All Formats
▦	cite:KS_1994	Stare Krajevne skupnosti	OpenLayers KML GML	Select one ▾
▦	cite:Ob1994F	Ob1994F	OpenLayers KML GML	Select one ▾
•	sf:bugsites	Spearfish bug locations	OpenLayers KML GML	Select one ▾
/	sf:roads	Spearfish roads	OpenLayers KML GML	Select one ▾
/	sf:streams	Spearfish streams	OpenLayers KML GML	Select one ▾
•	sf:archsites	Spearfish archeological sites	OpenLayers KML GML	Select one ▾
▦	sf:restricted	Spearfish restricted areas	OpenLayers KML GML	Select one ▾
▦	sf:sfdem	sfdem is a Tagged Image File Format with Geographic information	OpenLayers KML	Select one ▾

Figure 94 – Geoserver management console – Layer Preview view

When the »OpenLayers« link is clicked, a simple web application opens that enables an overview of the geospatial layer. In order to get this link the archivist can right click the “OpenLayers” link under Common Formats column of the chosen geospatial layer. The link is then sent to the user to view the ordered geodata.

Figure 92 shows how the geodata layer is served as a web service and can be viewed in a simple web application. When the archivist wants to provide a user simple access to geodata, they can simply send them the link to this application, which is automatically generated within Geoserver. The application enables the user to zoom and pan the layer. By clicking the object, the application displays its attribute information below the map.



10.0.92.39:8082/geoserver/cite/wms?service=WMS&version=1.1.0&request=GetMap&layers=cite:Ob1994F&styles=&bbox=37

Scale = 1 : 434K

453158.99373, 119190.05362

Ob1994F

fid	D46_ID	OID_	PE_SIF	PE_IME	D46_MID	D46_ID_1	D46_IME	N8	POVRSINA	Y_C	X_C
Ob1994F.23	1.0053838E7	22	OS	OBCINA - STARA (do 31.12.1994)	10053838	23	LITIJA	0	3.2757204E8	487240	101380

Figure 95 – OpenLayers web application showing the Geodata layer served as a web service

More information on managing geodata within Geoserver, can be found in Geoserver documentation. A link is provided in chapter 5.7.2.7 Code and documentation for Geodata tools.

3.5.2.2.3 Edited and customised view of unstructured geodata in QGIS

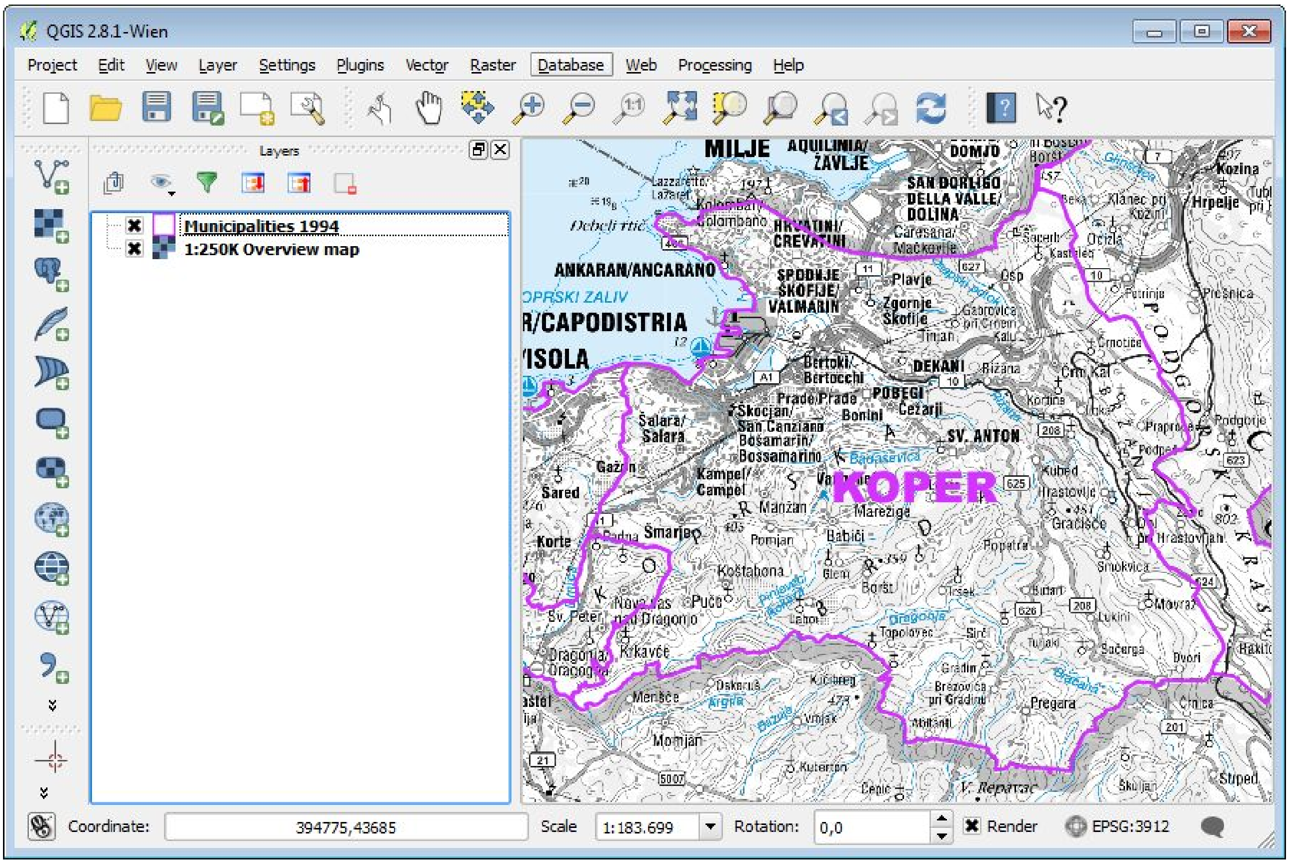
In case the user orders a geodata set, that contains restricted information, the archivist can use QGIS to omit the restricted information. The Archivist uses the geo tools in QGIS to distort, delete or generalise the restricted information. From this point on, the scenario is the same as the first one.

1. The user with full access (an archivist – the keeper of the archived records – and the user in the reading room with allowance) can edit, manipulate and view a geodata DIP$_u$ in whatever way they see fit. Within QGIS they can perform several types of action, like simplification, selection of elements, transformation. With access to documentation, they can recreate the visualization and can recreate GIS projects.
2. If the geodata has to be anonymised, an archivist or employee with knowledge of geodata manipulation modifies the DIP$_0$. The DIP$_0$ is then transformed into a DIP$_u$ and is made ready for the end-user for reading room use or reproduced for outside use.

The end-user with restricted access can work with a geodata DIP$_u$ which has been modified by an archivist or employee with geodata knowledge. (S)he can manipulate data within QGIS in the same way as users with full access, only the data-sets are different.

3.5.2.2.3.1 Example using QGIS - Creating a unified access QGIS project displaying geodata from multiple AIPs

Sometimes the user can order an archival package, containing geodata, that was initially rendered in a specific way and combined with basemaps from other fonds in order to produce a result (a building permit or a confirmation of a restricted status, etc). In that case, in order to reproduce a copy of a legal document, or to provide the proper context for understanding a geospatial dataset, data from multiple fonds needs to be combined or at least from different description units and render it according to documentation. In that case all the data can be combined in one map view within QGIS and this view can be saved as a QGIS project file (*.qgs). The benefit of this is that the user will only run this file and QGIS will load all the required data as they have been prepared by the archivist.



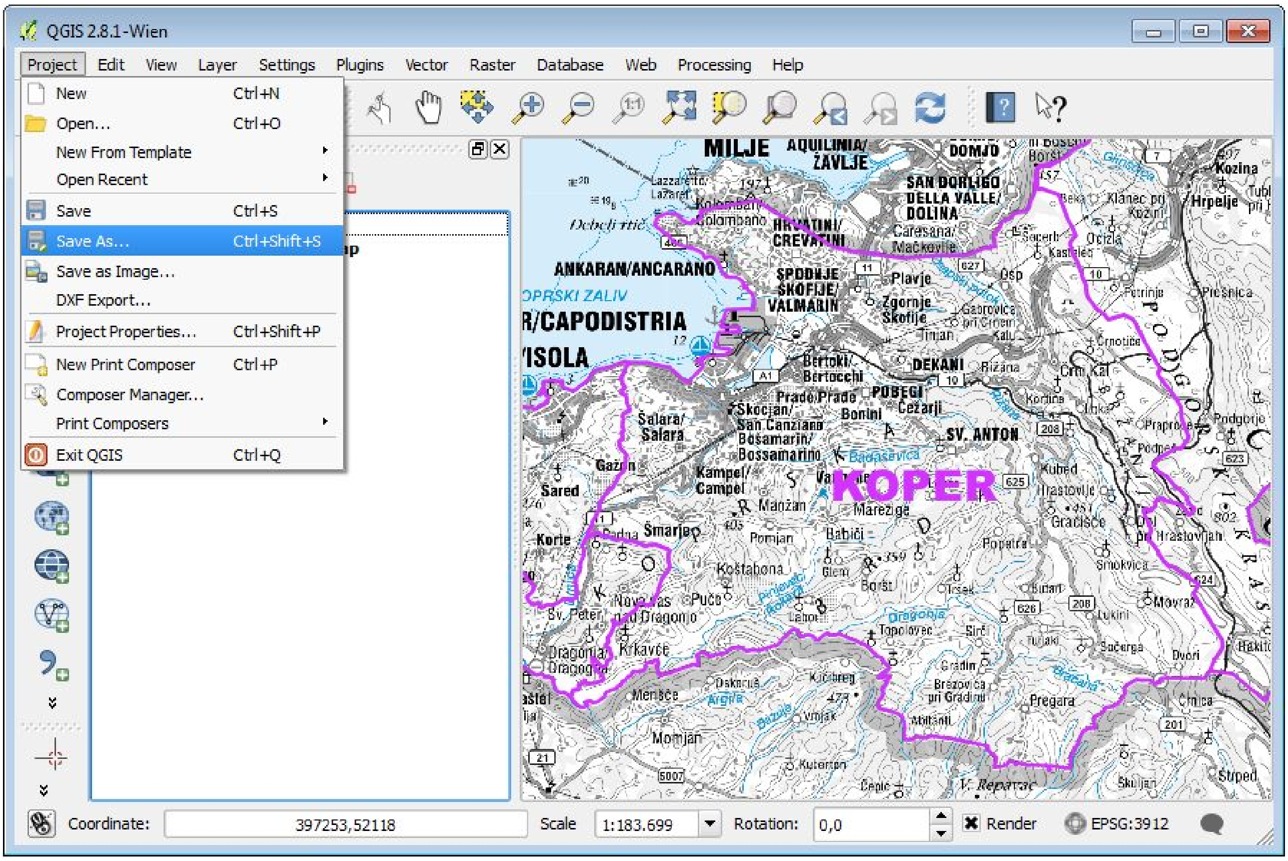
Municipalities 1994
1:250K Overview map

Coordinate: 394775,43685
Scale 1:183.699
Rotation: 0,0
☒ Render
EPSG:3912

Figure 96 – QGIS interface with geodata from multiple sources

In the Figure 93 there is a QGIS map view in which we have 2 datasets, a vector dataset and a basemap providing more context. When comparing the window on the left, where layers are listed, we can see that layer names have been customised to provide a better understanding of the data itself. The vector layer of administrative units was also labeled using the attributes of spatial objects (the name of the Admin. unit) and coloured in a way that it shows the border but doesn't cover the raster background map.

When the archivist has prepared the layers in the map view, they can now save this setup as a project. They open the "Project" dropdown menu from the Menu bar and select "Save As".



Coordinate: 397253,52118

Scale 1:183.699

Rotation: 0,0

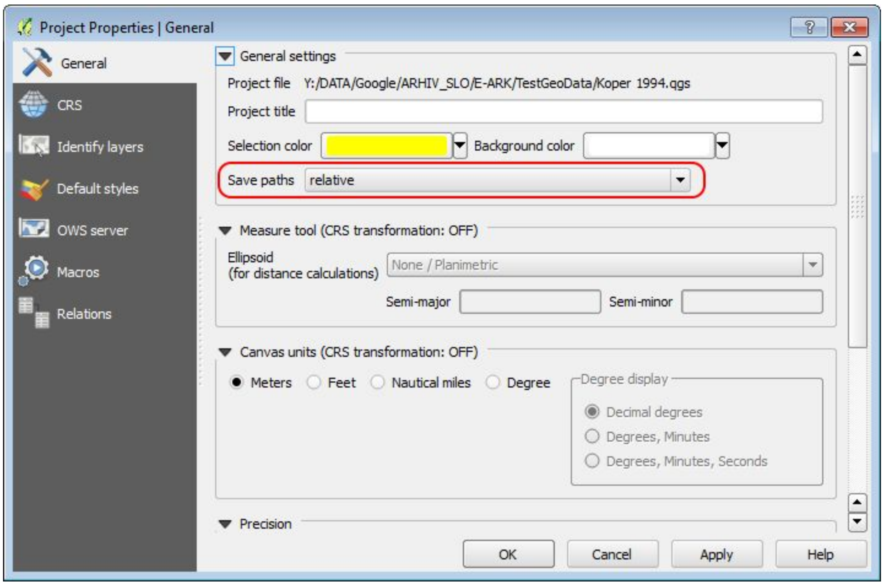
☒ Render

EPSG:3912

Figure 97 – Saving the QGIS project

The important thing when saving the project is that the file is saved in the root of the folder containing all the geodata and that, in the “Project Properties”, the variable “Save paths” is set to relative. This way all the data will be accessible within QGIS even if the DIP folder is copied to a different machine - if the layers are file based.

When preparing a QGIS project containing layers, that were added from a web service or from a spatially enabled RDBMS (PostGIS), the archivist preparing the document, needs to inform the end-user of the access limitations (i.e. the *.qgs not visible if viewed outside of archival network)



Project file Y:/DATA/Google/ARHIV_SLO/E-ARK/TestGeoData/Koper 1994.qgs

Project title

Selection color

Background color

Save paths relative

Ellipsoid (for distance calculations) None / Planimetric

Semi-major

Semi-minor

● Meters ○ Feet ○ Nautical miles ○ Degree

Degree display

● Decimal degrees

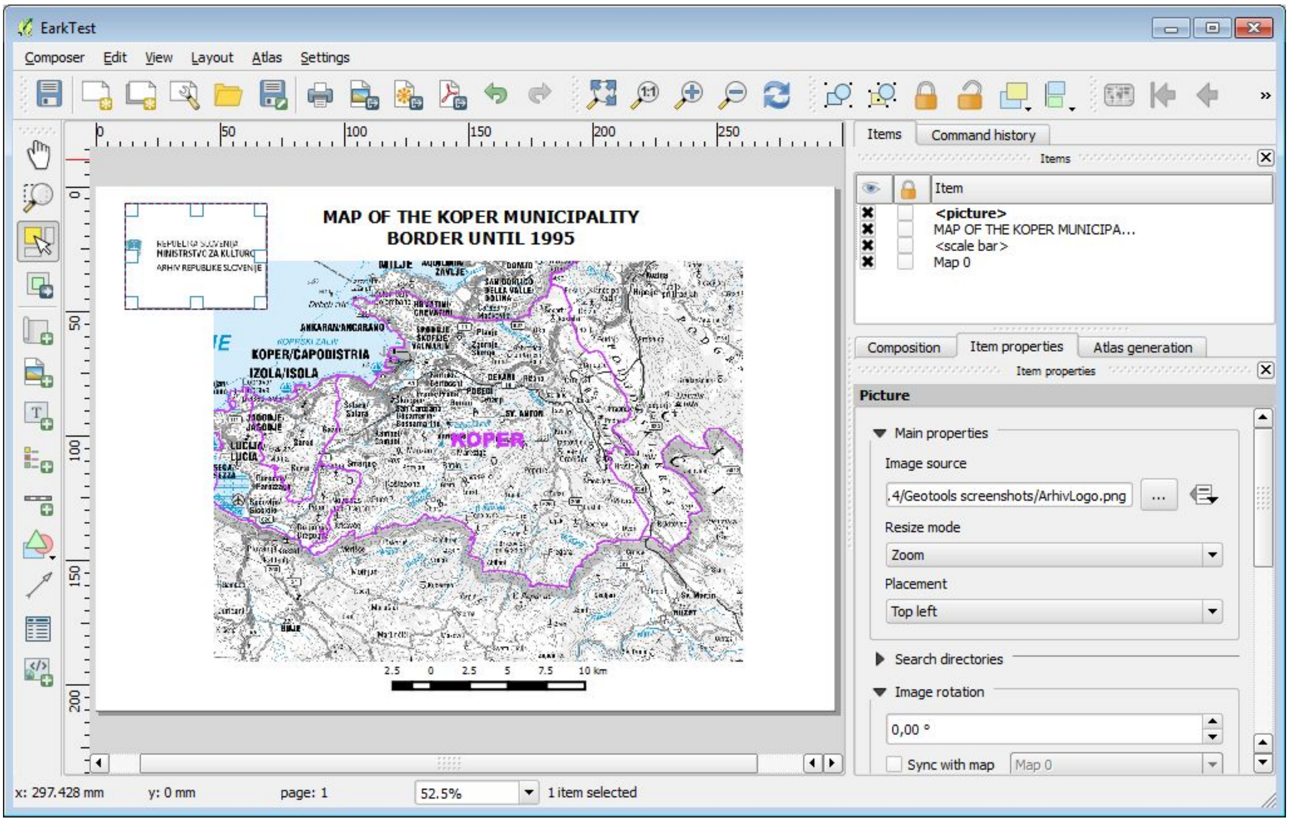
○ Degrees, Minutes

○ Degrees, Minutes, Seconds

Figure 98 – QGIS Project Properties window - showing where to set “Save paths” to relative

3.5.2.2.4 Reproduction of geodata

The end-user can order a reproduction of a set of geodata. They can order an electronic copy or create maps in QGIS, which can be printed. From the “Project” drop down menu, users choose “New Print Composer” or click **Ctrl+P** and a new window opens in which a map can be created.



Items:

- <picture>
- MAP OF THE KOPER MUNICIPA…
- <scale bar>
- Map 0

Image source .4/Geotools screenshots/ArhivLogo.png

Resize mode Zoom

Placement Top left

Image rotation 0,00 °

□ Sync with map Map 0

x: 297.428 mm y: 0 mm page: 1 52.5% 1 item selected

Figure 99 – QGIS tool “Print Composer” in which a map is created

From this tool the map can be printed to the printer, or saved as a PDF file, for the user to print themselves. More information on how to use the Print Composer is available in QGIS tutorials (a link is provided in the next chapter).

3.5.2.3 Code and documentation for Geodata tools

QGIS

- Download the latest version at http://www.qgis.org
- Documentation for any version is here: http://www.qgis.org/en/docs/index.html
- Training tutorials for the latest version are here: http://docs.qgis.org/2.14/en/docs/training_manual/
- More information on working with vector data is available from online QGIS documentation (http://docs.qgis.org/2.14/en/docs/training_manual/basic_map/vector_data.html)

Geoserver

- Download at http://geoserver.org/
- Documentation for any version is here: http://docs.geoserver.org/
- Training tutorials for the latest maintenance version are here: http://docs.geoserver.org/maintain/en/user/tutorials/index.html

Peripleo

- Download and API documentation at http://github.com/pelagios/peripleo
- Introductory article at http://journal.code4lib.org/articles/11144

3.5.3 The SMURF Tool (IP Viewer): Access to Electronic Records management Systems and Simple File-System Based Records

The tool that renders data in SMURF format has already been described in section 3.2.6 IP Viewer. The access scenarios pertaining to the SMURF format are detailed in the final E-ARK DIP Specification[56].

Note that EDRMS stored in the SIARD format suppose different access tools, cf. 3.7.1 Access to databases and EDRMS stored in the SIARD format.

3.5.4 The OLAP Tools: OLAP access to information stored in the SIARD format

The OLAP Tools that are listed below have been used to conduct the E-ARK data mining showcase within the context of the E-ARK pilot 7[57], which is described elsewhere[58].

[56] E-ARK Final DIP specification (http://www.eark-project.com/resources/project-deliverables/91-d532)

[57] https://github.com/eark-project/Data-Warehouse-and-OLAP/blob/master/XT_MNL_EARK_D7_Tools_Presentation_v1.2.docx

[58] D6.3 Data Mining Showcase (http://www.eark-project.com/resources/project-deliverables/90-d63)

The documentation and installation procedures are too extensive to be repeated here, so please consult the appropriate documents[59]. Also, a detailed description of the OLAP E-ARK Access scenarios can be consulted in the final DIP specification[60].

For these reasons this chapter only very succinctly describes the OLAP Access tools used within the E-ARK project. Being a showcase, the use of OLAP in an E-ARK environment only represents a 'minor' task compared to the other content information type specific tasks described above. Lastly, the E-ARK 'OLAP experience' is more an approach than it is a tool development scenario like the others, and as such it cannot be completely replicated in other institutions. The tools used are vendor specific, because this serves as inspiration for other archives interested in data mining.

The tools used in pilot 7 are either briefly described below or links are provided to websites that describe them more fully.

1. ORACLE database (12.C Release 1): http://www.oracle.com/technetwork/database/enterprise-edition/overview/index.html

2. Oracle Application Express (APEX) 5.1: http://www.oracle.com/technetwork/developer-tools/apex/downloads/index.html (E-ARK used the version 5.0.3)

   Oracle Application Express (APEX) is a web-based software development environment that runs on an Oracle database. It is fully supported and comes standard (at no additional cost) with all Oracle Database editions and, starting with Oracle 11g, is installed by default as part of the core database install.

   APEX can be used to build complex web applications which can be used in most modern web browsers. The APEX development environment is also browser-based.

   Project sources: https://github.com/eark-project/Data-Warehouse-and-OLAP/tree/master/Software/DB

Oracle Data Integrator

3. Oracle Data Integrator (ODI) 12.2.1: http://www.oracle.com/technetwork/middleware/data-integrator/overview/index.html

   Oracle Data Integrator (ODI) is a comprehensive data integration platform that covers all data integration requirements: from high-volume, high-performance batch loads, to event-driven integration processes, to SOA-enabled data services.

[59] Detailed documentation of the data warehouse and OLAP related scenarios/tasks of E-ARK pilot 7 can be found here: https://github.com/eark-project/Data-Warehouse-and-OLAP/blob/master/XT_MNL_EARK_D7_Tools_Presentation_v1.2.docx
Documentation of the installation of Oracle components used in E-ARK Pilot 7: https://github.com/eark-project/Data-Warehouse-and-OLAP/tree/master/Documentation

[60] E-ARK Final DIP Specification (http://www.eark-project.com/resources/project-deliverables/91-d532)

In a data warehouse environment it is a crucial step selecting the right integration tool. A well-designed data warehouse with a declarative integration tool can prepare and transform the data from many different data sources for any analytical needs.

Project sources:
https://github.com/eark-project/Data-Warehouse-and-OLAP/tree/master/Software/ODI

Oracle Business Intelligence

4. Oracle Business Intelligence (BI) 12.2.1:
http://www.oracle.com/technetwork/middleware/bi-enterprise-edition/overview/index.html

   Oracle Business Intelligence is a platform that enables customers to make business decisions by offering visual analytics and self-service discovery together with enterprise analytics.

   Project sources:
https://github.com/eark-project/Data-Warehouse-and-OLAP/tree/master/Software/BI

   Oracle Fusion Middleware Infrastructure 12.2.1:
http://www.oracle.com/technetwork/middleware/fusion-middleware/downloads/index.html

   Oracle Fusion Middleware (FMW) consists of several software products from Oracle Corporation. FMW spans multiple services, including Java EE and developer tools, integration services, business intelligence, collaboration, and content management. FMW depends on open standards such as BPEL, SOAP, XML and JMS.

   Oracle Fusion Middleware provides software for the development, deployment, and management of service-oriented architecture (SOA). It includes what Oracle calls "hot-pluggable" architecture, designed to facilitate integration with existing applications and systems from other software vendors.

5. Oracle Rest Data Services (ORDS) 3.0.9:
http://www.oracle.com/technetwork/developer-tools/rest-data-services/overview/index.html
(E-ARK used the version 3.0.4)

   Oracle REST Data Services (ORDS) is a powerful tool that enables developers with SQL and other database skills to build enterprise class, data access APIs to Oracle Databases.

   Oracle REST Data Services (ORDS) makes it easy to develop modern REST interfaces for relational data in the Oracle Database. ORDS is responsible for receiving and translating requests to the APEX engine, and returning the results to the requester browser.

3.5.4.1 Code and documentation for OLAP tools

1. Overall: https://github.com/eark-project/Data-Warehouse-and-OLAP
2. Detailed documentation of the data warehouse and OLAP related scenarios/tasks of E-ARK pilot 7 can be found here:
https://github.com/eark-project/Data-Warehouse-and-OLAP/blob/master/XT_MNL_EARK_D7_Tools_Presentation_v1.2.docx

3. Documentation of the installation of Oracle components used in E-ARK Pilot 7: https://github.com/eark-project/Data-Warehouse-and-OLAP/tree/master/Documentation

6 Glossary

Access	Access refers to the funtional entity from the OAIS reference model https://public.ccsds.org/pubs/650x0m2.pdf
Access Aid	A software program or document that allows Consumers to locate, analyse, order or retrieve information from an OAIS. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
Access Functional Entity	The OAIS functional entity that contains the services and functions which make the archival information holdings and related services visible to Consumers. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
Access Rights Information	The information that identifies the access restrictions pertaining to the content information, including the legal framework, licensing terms, and access control. It contains the access and distribution conditions stated within the Submission Agreement, related to both preservation (by the OAIS) and final usage (by the Consumer). It also includes the specifications for the application of rights enforcement measures. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
Access scenarios	Access scenario is used to describe the environment, the DIP and the Access Software which altogether are used to render content information and associated metadata.
Access Software	A type of software that presents part of or all of the information content of an Information Object in forms understandable to humans or systems. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
Access Software Platform (ASP)	The Access Software Platform provides a uniform interface that allows for the end-user to search, order and access archival records. It allows for the archivist to do those same things as well as manipulate the archival records (e.g. do a AIP-DIP conversion).
AIP-DIP conversion Tool	The AIP-DIP conversion component consists of a set of individual tasks which are executed in a specific order to convert an E-ARK Archival Information Package (AIP) into the E-ARK Dissemination Information Package (DIP). The component is an integrated part of the E-ARK Web. The Access Software Platform calls this AIP-DIP functionality with a single click from its GUI.
Archival Information Package	An Archival Information Package, consisting of the content information and the associated Preservation Description Information (PDI), which is preserved within an OAIS. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
Archival Catalogue	See Finding Aid.
Archival record	Materials created or received by a person, family, or organization, public or private, in the conduct of their affairs that are preserved because of the enduring value contained in the information they contain or as evidence of the functions and responsibilities of their creator. Source Society of American Archivists: http://www2.archivists.org/glossary/terms/a/archival-records#.VyB5VXqd9iN
Authenticity	The degree to which a person (or system) regards an object as what it is purported to be. Authenticity is judged on the basis of evidence. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf

CMIS	Content Management Interoperability Services (CMIS) is an open standard that allows different content management systems to inter-operate over the Internet. Specifically, CMIS defines an abstraction layer for controlling diverse document management systems and repositories using web protocols, cf. https://en.wikipedia.org/wiki/Content_Management_Interoperability_Services
CMIS Viewer	The CMIS viewer is a web based simple tool for browsing a CMIS compliant repository.
Common Specification	The common IP specification for E-ARK IPs conceived as a common basis for the E-ARK SIP, AIP and DIP Specifications. http://www.eark-project.com/resources/specificationdocs/67-e-ark-draft-common-specification-ver-017
Compound Object	A Digital Object composed of multiple Files: for example, a Web Page composed of text and image Files.
Consumer	The role played by those persons or client systems, which interact with OAIS services to find preserved information of interest and to access that information in detail. This can include other OAIS's, as well as internal OAIS persons or systems. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf In E-ARK "Consumer" is an umbrella term that designates all users of archival holdings, thus both internal users, cf. archivists, and external users, cf. end-users.
Content Data Object	The Data Object that together with associated Representation Information comprises the Content Information. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
Content Information	A set of information that is the original target of preservation or that includes part or all of that information. It is an Information Object composed of its Content Data Object and its Representation Information. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
Content Information Type	The data types for which format specifications have been created, cf. Electronic Management Systems (ERMS), Simple File-Based System Records (SFBS), databases, and geo-data.
Database	A database is an organised collection of data. It is the collection of schemas, tables, queries, reports, views and other objects. Source: Wikipedia: https://en.wikipedia.org/wiki/Database
Database Preservation ToolKit (DBPTK)	The Database Preservation Tool Kit is a piece of software which, from an Access perspective, enables the loading of a SIARD file into an RDBMS http://keeps.github.io/db-preservation-toolkit/. It is developed by KEEP SOLUTIONS which is a partner of the E-ARK project http://www.keep.pt/en
Database Visualization ToolKit (DBVTK)	The Database Visualization Toolkit (DBVTK) allows end-user access to archived databases without the need to go through the complex process of setting up a RDBMS and loading a SIARD file into it. This tool allows archivists and consumers to preview, explore and retrieve information from preserved databases. The software is aimed at end-users with little or no experience with SQL, and its main goal is to enable non-technical users to quickly find data of interest and provide means to export and print these data.
Data warehouse	In computing, a data warehouse (DW or DWH), also known as an enterprise data warehouse (EDW), is a system used for reporting and data analysis, and is considered

	as a core component of Business Intelligence [1] environment. DWs are central repositories of integrated data from one or more disparate sources. They store current and historical data and are used for creating analytical reports for knowledge workers throughout an enterprise. Examples of reports could range from annual and quarterly comparisons and trends to detailed daily sales analysis.
DB Viewer	A GUI conceived by the E-ARK project to view and analyse databases.
Descriptive metadata	Also named Descriptive Information in OAIS: The set of information, consisting primarily of Package Descriptions, which is provided to Data Management to support the finding, ordering, and retrieving of OAIS information holdings by Consumers. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf The standard that E-ARK recommends for descriptive metadata is EAD.
Digital Object	An object composed of a set of bit sequences. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
Digital Provenance	Documentation of processes in a Digital Object's life cycle. Digital provenance typically describes Agents responsible for the custody and stewardship of Digital Objects, key Events that occur over the course of the Digital Object's life cycle, and other information associated with the Digital Object's creation, management, use, and preservation. Source PREMIS: http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf
DIP$_0$	A provisional Dissemination Information Package directly derived from one or more AIPs, which may or may not be ready for use, according to the user's order and access rights.
DIP$_p$	A permanent Dissemination Information Package, available to be accessed indefinitely by users due to frequent requests for the same data. The DIPP can be available on-line.
DIP$_u$	A Dissemination Information Package, ready to be accessed, and previously checked against user's order and access rights.
DIP reference format	Refers to the E-ARK container format which is conceived to store the content information and its associated metadata.
DIP Representation Formats	The DIP representation formats are content specific implementations of the DIP reference format and offer examples of content information type specific scenarios.
Dissemination Information Package (DIP)	Dissemination Information Package: an Information Package, derived from one or more AIPs, and sent by archives to the Consumer in response to a request to the OAIS. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
EAD	Encoded Archival Description. A non-proprietary de facto standard for the encoding of Finding Aids for use in a networked (online) environment. Finding Aids are inventories, indexes, or guides that are created by archival and manuscript repositories to provide information about specific collections. While the Finding Aids may vary somewhat in style, their common purpose is to provide detailed description of the content and intellectual organization of collections of archival materials. EAD allows the standardization of collection information in Finding Aids within and across repositories. http://www.loc.gov/ead/eadabout.html

Electronic Records Management System (ERMS)	Electronic Records Management System is a type of content management system and refers to the combined technologies of document management and records management systems as an integrated system.
End-User	The end-user designates an external user who seeks content information in archival holdings.
ERMS Viewer	A GUI conceived by the E-ARK project to view ERMS systems.
Exchange	Refers to the DIP as an exchange format, and as such it is essential that it is possible to transfer DIPs, for example between a repository and various Access environments.
Finding Aid	A type of Access Aid that allows a user to search for and identify Information Packages of interest. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
Geodata	Geodata is information about geographic locations that is stored in a format that can be used with a geographic information system (GIS). Geodata can be stored in a database, geodatabase, shapefile, coverage, raster image, or even a dbf table or Microsoft Excel spreadsheet.
Geodata Tool	The Geodata Tool is a name for the ensemble of tools which have been used in the E-ARK project to process geodata.
GeoServer	The GeoServer is an open source server for sharing geospatial data. http://geoserver.org/
GeoTIFF	GeoTIFF is a public domain metadata standard which allows georeferencing information to be embedded within a TIFF file. The potential additional information includes map projection, coordinate systems, ellipsoids, datums, and everything else necessary to establish the exact spatial reference for the file.
GML	The Geography Mark-up Language: the XML grammar defined by the Open Geospatial Consortium (OGC) to express geographical features. GML serves as a modelling language for geographic systems as well as an open interchange format for geographic transactions on the Internet.
Graphical user interface (GUI)	A Graphical user interface (GUI) is a graphical interface to a program on a computer. It takes advantage of the computer's graphics capabilities to make the program easier to use.
Information Object	A Data Object together with its Representation Information. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
Information Package	A logical container composed of optional content information and optional associated Preservation Description Information. Associated with this Information Package is Packaging Information used to delimit and identify the content information and Package Description information used to facilitate searches for the content information. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
Intellectual Entity	A set of content that is considered a single intellectual unit for purposes of management and description: for example, a particular book, map, photograph, or database. An Intellectual Entity can include other Intellectual Entities; for example, a Web site can include a Web page; a Web page can include an image. An Intellectual

	Entity may have one or more digital representations. Source PREMIS http://www.digitizationguidelines.gov/term.php?term=intellectualentity
IP Viewer	Is part of the Access Software Platform and allows the Consumer to browse and view DIPs.
METS	The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation. Source http://www.loc.gov/standards/mets/
MultiDimensional DBMS	A MultiDimensional DBMS is a particular kind of RDBMS that is specifically geared towards OLAP (in fact MDDBMS is often used co-terminously with OLAP).
Normalisation	The term is used with two meanings: firstly, in the sense in which the digital preservation community is employing the word: on Ingest, content data objects are transformed into long-term friendly formats. Secondly, in database normalisation where columns and tables are organised in order to reduce redundancy.
NoSQL solution	A NoSQL database provides a mechanism for storage and retrieval of data which is modeled in means other than the tabular relations used in relational databases, cf. https://en.wikipedia.org/wiki/NoSQL.
OAIS	The Open Archival Information System is an archive (and a standard: ISO 14721:2003), consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
OLAP	In computing, online analytical processing, or OLAP, is an approach to answering multi-dimensional analytical (MDA) queries swiftly. OLAP is part of the broader category of business intelligence, which also encompasses relational database, report writing and data mining. Typical applications of OLAP include business reporting for sales, marketing, management reporting, business process management (BPM), budgeting and forecasting, financial reporting and similar areas, with new applications coming up, such as agriculture. Source Wikipedia https://en.wikipedia.org/wiki/Online_analytical_processing
OLAP Cube	An OLAP cube is an array of data understood in terms of its 0 or more dimensions. OLAP is a computer-based technique for analysing business data in the search for business intelligence.
OLAP Tool	The OLAP Tools is the name for the ensemble of tools used to conduct the E-ARK pilot 7.
Order Management Tool (OMT)	Is part of the Access Software Platform and allows for the archivist to process an order, thus retrieving the requested archival records and create a Dissemination Information Package (DIP).
order.xml	The xml-file that specifies an order in the E-ARK Access system.
Packaging Information	The information that is used to bind and identify the components of an Information Package. For example, it may be the ISO 9660 volume and directory information used on a CD-ROM to provide the content of several files containing content information

	and Preservation Description Information. Source: OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
Peripleo	Peripleo is a map-based search engine for exploring data annotated by the Pelagios community. Its user interface allows for free browsing as well as keyword and full-text search, while offering filtering options based on time, data source and object type. https://wiki.digitalclassicist.org/Peripleo
PREMIS	The PREMIS Data Dictionary for Preservation Metadata is the international standard for metadata to support the preservation of Digital Objects and ensure their long-term usability. Developed by an international team of experts, PREMIS is implemented in digital preservation projects around the world, and support for PREMIS is incorporated into a number of commercial and open-source digital preservation tools and systems. The PREMIS Editorial Committee coordinates revisions and implementation of the standard, which consists of the Data Dictionary, an XML schema, and supporting documentation. Source: http://www.loc.gov/standards/premis/
Preservation metadata	Preservation metadata is an essential component of most digital preservation strategies. As an increasing proportion of the world's information output shifts from analog to digital form, it is necessary to develop new strategies to preserve this information for the long-term. Preservation metadata is information that supports and documents the digital preservation process. Preservation metadata is sometimes considered a subset of technical or administrative metadata. Source https://en.wikipedia.org/wiki/Preservation_metadata The standard that E-ARK recommends for preservation metadata is PREMIS.
Producer	The role played by those persons or client systems that provide the information to be preserved. This can include other OAISs or internal OAIS persons or systems. Source OAIS: http://public.ccsds.org/publications/archive/650x0m2.pdf
QGIS	A Free and Open Source Geographic Information System. http://www.qgis.org/en/site/
Record	Any 'information created, received and maintained as evidence and information by an organisation or person, in pursuance of legal obligations or in the transaction of business' (ISO 15489-1:2001, 3.15). In MoReq2010®, a record may be further characterised as follows. ● It has an extensible set of metadata that describe it. ● It has one or more components that represent its content. ● It is classified with a business classification. ● It has a disposal schedule that describes explicitly if, how and when it will be disposed of or destroyed. ● It belongs to an aggregation of records. ● Access to it is controlled and limited to authorised users. ● Its destruction may be prevented by a disposal hold. ● It may be exported to another MCRS while retaining all of the characteristics listed above. [MoReq 2010, v 1.1]
Relational Database Management System (RDBMS)	A relational database management system (RDBMS) is a computer software application that interacts with the user, other applications, and the database itself to capture and analyse data. A general-purpose RDBMS is designed to allow the definition, creation, querying, update, and administration of databases.

Representation	The set of files, including structural metadata, needed for a complete and reasonable rendering of an Intellectual Entity. For example, a journal article may be complete in one PDF file; this single file constitutes the representation. Another journal article may consist of one SGML file and two image files; these three files constitute the representation. A third article may be represented by one TIFF image for each of 12 pages plus an XML file of structural metadata showing the order of the pages; these 13 files constitute the representation. Source PREMIS: http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf , p.8
Representation Information	Representation Information is metadata that that transforms a Digital Object into an Information Object and thereby making it understandable by a human being. It consists of Semantic and Structure Information. Source OAIS: http://public.ccsds.org/publications/archive/650x0m2.pdf
Search Module	Part of the Access Software Platform which allows the Consumer to search and order archival records from the archive.
Semantically marked up records formats (SMURF)	The SMURF is an IP format for ERMS systems and SFSB (simple file-system based records) conceived by the E-ARK project.
SFSB Viewer	GUI conceived by the E-ARK project to view Simple File-System Based Records.
SIARD	IP format for databases. Currently there three versions exist: SIARD1.0, SIARDDK and SIARD2.0.
SIARD 1.0	SIARD1.0 is the original SIARD format developed by the Swiss Federal Archives (SFA). Available at: http://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0165&documentVersion=1.0
SIARD 2.0	SIARD2.0 was developed by E-ARK in collaboration with the Swiss Federal Archives (SFA), and is based on the original SIARD format developed by SFA. Available at: http://www.eark-project.com/resources/specificationdocs/32-specification-for-siard-format-v20
SIARDDK	SIARDDK is a format used in Denmark since 2010, and is a variation of SIARD1.0.
Simple File-System Based Records (SFSB)	Simple file-system based records (SFSB) are records that contain simple file-system based folders or files, including those originating from content and data management systems, such as SharePoint, that are not based on true file systems. They address the submission of computer files or folders from the file Producers rather than from an ERMS. They require manual enrichment with additional descriptive metadata.
SMURF	SMURF (Semantically-Marked-Up-Records Format) is an E-ARK format that allows the preservation of Single File Based Records (SFSB) and (Electronic Documents and Records Management Systems (EDRMS). http://www.eark-project.com/resources/project-deliverables/52-d33smurf
SMURF Tool	Is another name for the IP Viewer, cf. IP viewer above.
Structural metadata	Structural metadata describes the physical and/or logical structure of digital resources; it expresses the intellectual boundaries of complex objects and can be used to describe relationships between an object's component parts. Structural metadata is commonly used to facilitate navigation and presentation of complex

	items by defining structural characteristics such as pagination and sequence. And, like METS, can be used to aggregate related metadata. Source http://www.library.illinois.edu/dcc/bestpractices/chapter_11_structuralmetadata.html The standard that E-ARK recommends for structural metadata is METS
Submission Information Package (SIP)	An Information Package that is delivered by the Producer to the OAIS for use in the construction or update of one or more AIPs and/or the associated Descriptive Information. Source OAIS http://public.ccsds.org/publications/archive/650x0m2.pdf
Views (SQL)	In database theory, a view is the result set of a stored query on the data, which the database users can query just as they would in a persistent database collection object. This pre-established query command is kept in the database dictionary. Unlike ordinary base tables in a relational database, a view does not form part of the physical schema: as a result set, it is a virtual table computed or collated dynamically from data in the database when access to that view is requested. Changes applied to the data in a relevant underlying table are reflected in the data shown in subsequent invocations of the view. In some NoSQL databases, views are the only way to query data. Source Wikipedia https://en.wikipedia.org/wiki/View_(SQL)

Table 2 - Glossary

7 References and associated links and documents

Angular Material Design 2 https://material.angularjs.org/latest/

Active Directory https://en.wikipedia.org/wiki/Active_Directory

Apache HBase http://hbase.apache.org/

Apache Solr http://lucene.apache.org/solr/

Application programming interface https://en.wikipedia.org/wiki/Application_programming_interface

Atom Publishing Protocol 1.0 https://movabletype.org/documentation/developer/api/atompub/

Common Specification, draft
http://www.eark-project.com/resources/specificationdocs/67-e-ark-draft-common-specification-ver-017

CMIS, Content Management Interoperability Service
https://en.wikipedia.org/wiki/Content_Management_Interoperability_Services

D2.1 General pilot model and use case definition
http://www.eark-project.com/resources/project-deliverables/5-d21-e-ark-general-pilot-model-and-use-case-definition

D2.2 Legal Issues Report: European Cultural Preservation in a Changing Legislative Landscape
http://www.eark-project.com/resources/project-deliverables/33-d22-legal-issues-report-european-cultural-preservation-in-a-changing-legislative-landscape

D2.3 Detailed Pilots Specification
http://www.eark-project.com/resources/project-deliverables/60-23pilotsspec

D3.1 Report on available best practices
http://www.eark-project.com/resources/project-deliverables/6-d31-e-ark-report-on-available-best-practices

D3.3 E-ARK SIP Pilot Specification
http://www.eark-project.com/resources/project-deliverables/51-d33pilotspec

D3.3 E-ARK SMURF http://www.eark-project.com/resources/project-deliverables/52-d33smurf

D4.3 E-ARK AIP Specification
http://www.eark-project.com/resources/project-deliverables/53-d43earkaipspec-1

D4.4 Final version of SIP-AIP conversion component

D5.1 GAP report between requirements for access and current access solutions
http://www.eark-project.com/resources/project-deliverables/3-d51-e-ark-gap-report

D5.2 E-ARK DIP Draft Specification http://www.eark-project.com/resources/project-deliverables/31-d52

D5.3 E-ARK DIP Pilot Specification
http://eark-project.com/resources/project-deliverables/61-d53-pilot-dip-specification

D6.1 Faceted Query Interface and API
http://www.eark-project.com/resources/project-deliverables/34-d61-faceted-query-interface-and-api

D6.2 Integrated Platform Reference Implementation
http://www.eark-project.com/resources/project-deliverables/54-d62intplatformref-1

D6.3 Data Mining Showcase

Dappert, A., Peyraud, S., Delve, J., Chou, C. Describing and Preserving Digital Object Environments, New Review of Information Networking, 2013, ISSN 1361-4576, 106-173
https://www.researchgate.net/profile/Janet_Delve/publication/262280940_Describing_and_Preserving_Digital_Object_Environments/links/56c60e1b08ae03b93dd9f74f.pdf

Data Warehouse and OLAP
https://github.com/eark-project/Data-Warehouse-and-OLAP/blob/master/XT_MNL_EARK_D7_Tools_Presentation_v1.2.docx

Data Warehouse and OLAP documentation
https://github.com/eark-project/Data-Warehouse-and-OLAP/tree/master/Documentation

Description of Work: Grant agreement for CIP-Pilot actions no. 620998, Annex I - “Description of Work”, Version date 2014-01-17

DIP, ERMS and SFSB Viewer http://178.62.194.129/ipviewer/

dm-file-ingest https://github.com/eark-project/dm-file-ingest

EAD3 https://www.loc.gov/ead/

EAD3 <c> http://www.loc.gov/ead/EAD3taglib/#elem-c

E-ARK Final DIP Specification

E-ARK Web https://earkdev.ait.ac.at:8443/cas/login?service

E-ARK CMIS Viewer https://github.com/eark-project/E-Ark-CMIS-Viewer/tree/master/frontend and https://github.com/eark-project/E-Ark-CMIS-Viewer/tree/master/bridge

Fielding, Roy Thomas. Architectural Styles and the Design of Network-based Software Architectures. Doctoral dissertation, University of California, Irvine, 2000.
http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm

Ferreira B., Faria L., Ramalho J. C., Ferreira M. Database Preservation Toolkit - A relational database conversion and normalization tool, iPRES Conference 2016,
http://repositorium.sdum.uminho.pt/bitstream/1822/43479/1/dbptk-ipres16.pdf

ESSArch Preservation Platform (EPP) http://epp.essarch.org/

General Model - http://www.eark-project.com/resources/general-model

GeoServer http://geoserver.org/

GML, Geography Markup Language https://en.wikipedia.org/wiki/Geography_Markup_Language

Hadoop https://hadoop.apache.org/

JSON, JavaScript Object Notation http://www.json.org/

Kerberos https://en.wikipedia.org/wiki/Kerberos_(protocol)

Lily http://www.lilyproject.org/lily/index.html

Lightweight Directory Access Protocol https://en.wikipedia.org/wiki/Lightweight_Directory_Access_Protocol

LOB, Large object https://docs.oracle.com/cd/B10501_01/appdev.920/a97269/pc_16lob.htm

Material design (Google) https://www.google.com/design/spec/material-design/introduction.html

METS, Metadata Encoding and Transmission Standard http://www.loc.gov/standards/mets/mets-schemadocs.html

OAIS, Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model, ISO 14721:2012 http://public.ccsds.org/publications/archive/650x0m2.pdf

OGC, Open Geospatial Consortium http://www.opengeospatial.org/

OLAP, Online analytical processing https://en.wikipedia.org/wiki/Online_analytical_processing

Peripleo https://wiki.digitalclassicist.org/Peripleo

PREMIS 3.0 https://www.loc.gov/standards/premis/v3/

Python https://en.wikipedia.org/wiki/Python_(programming_language)

QGIS http://www.qgis.org/en/site/ and http://qgis.org/en/site/about/index.html

Redmine https://e-ark-redmine.magenta-aps.dk/

Repository of Authentic Digital Objects (RODA) http://www.roda-community.org/

SIARD2.0 http://www.eark-project.com/resources/specificationdocs/32-specification-for-siard-format-v20

Simon R., Isaksen L., Barker E., de Soto Cañamares P. Peripleo: a Tool for Exploring Heterogeneous Data through the Dimensions of Space and Time, Code4Lib Issue 31, 2016-01-28. ISSN 1940-5758 http://journal.code4lib.org/articles/11144

Tar https://en.wikipedia.org/wiki/Tar_(computing)