



DELIVERABLE

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

DELIVERABLE DETAILS

DELIVERABLE REFERENCE NO.	D5.2
DELIVERABLE TITLE	E-ARK DIP Draft Specification
REVISION	4.0

AUTHORS	
Names	Organisations
Alex Thirifays (main author) Anders Bo Nielsen Ann-Kristin Egeland Kathrine Hougaard Edsen Johansen Lone S. Jespersen	Statens Arkiver - The Danish National Archives
Anja Paulič Boris Domajnko Jože Škofljanec	Ministrstvo Za Kulturo – The Slovenian National Archives
Luis Faria	KEEP SOLUTIONS LDA
Björn Skog Karin Bredenberg	ES Solutions

Torben Lauritzen	Magenta ApS
Jan Rörden	Universität zu Köln – University of Cologne
István Alföldi Levente Szilágyi Zoltán Lux	National Archives of Hungary
Sven Schlarb	Austrian Institute of Technology GmbH
REVIEWERS	
Names	Organisations
Kuldar Aas	Rahvusarhiiv: The National Archives of Estonia
Dr Andrew Wilson	University of Portsmouth

**Project co-funded by the European Commission within the ICT Policy Support Programme
Dissemination Level**

P	Public	X
----------	---------------	----------

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Submitted Revisions History

Revision No.	Date	Authors(s)	Organisation	Description
4.0	25/4/2015	Alex Thirifays et al	DNA	Final Document

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This deliverable contains original unpublished work except where clearly indicated otherwise.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.

Table of Contents

1	Executive summary.....	7
2	Purpose and Method	8
3	High-Level Illustration and Textual Description	10
3.1	Product Context.....	11
3.2	Assumptions	11
3.3	Dependencies	12
3.4	Illustrations and descriptions.....	12
3.4.1	Search & Order Management.....	16
3.4.2	<i>DIP</i> Preparation.....	19
3.4.3	<i>DIP</i> Delivery.....	21
3.4.4	3.1.4 <i>DIP</i> Management	25
4	Access use cases	27
4.1	Process step 1: Search and Order Management.....	28
4.1.1	Use case 1.1: Search & Select records: Search in descriptive metadata and data.....	28
4.1.2	Use case 1.2: Manage Order.....	32
4.2	Process step 2: <i>DIP</i> Preparation.....	39
4.2.1	Process step 2.1 & 2.2: Checking the existence of appropriate <i>DIP</i>) & Confirmation of identified <i>DIP</i>	39
4.2.2	Use case 2.2: Check <i>DIP</i> availability and suitability.....	39
4.2.3	Use case 2.3: Creating <i>DIP</i> from <i>AIP</i>	42
4.2.4	Process step 2.4: Check if modification is needed	43
4.2.5	Use case 2.5: Finalization of <i>DIP</i>	43
4.3	Process step 3: <i>DIP</i> Delivery.....	48
4.3.1	Use Case 3.1: Prepare <i>DIP</i> delivery	48
4.3.2	Use Case 3.2: Provide <i>Access</i> rights	48
4.3.3	Use Case 3.3: Notify end-user	49
4.3.4	Use case 3.4: Access <i>DIP</i>	49
4.4	Process step 4: <i>DIP</i> Management	66
4.4.1	Use case 4.1: <i>DIP</i> Management	66

5	Requirements for E-ARK DIP format	67
5.1.1	Requirements derived from the E-ARK Common Specification	67
5.1.2	Requirements derived from access use cases	67
6	<i>DIP</i> format.....	71
6.1	<i>DIP</i> data model and folder structure	71
6.2	Metadata in the <i>DIP</i>	72
6.3	<i>DIP</i> specific metadata	73
6.3.1	Rendering information.....	73
6.3.2	Authenticity	74
6.3.3	Access Restrictions.....	76
6.3.4	<i>DIP</i> status	78
6.3.5	Dissemination notes	79
6.4	<i>DIP</i> specification for content data types and rendering scenarios	81
6.5	Access related metadata that will not be in the <i>DIP</i>	81
6.5.1	<i>DIP</i> order	81
7	High-Level component overview	84
7.1	Data Management component.....	86
7.2	<i>AIP</i> to <i>DIP</i> Transformation component.....	86
7.2.1	<i>AIP</i> Selection module	86
7.2.2	<i>DIP</i> Preparation module.....	87
7.3	<i>DIP</i> Delivery.....	88
7.3.1	Access module	88
7.3.2	Maintenance module.....	88
8	Glossary	89
9	References	92

Figures

Figure 1 - E-ARK Access high-level process	14
Figure 2 - Expanded E-ARK Access process	15
Figure 3 - [1] Search & Order Management	16
Figure 4 - [1.2] Manage order.....	17
Figure 5 - [2] <i>DIP</i> Preparation	20
Figure 6 - [3] <i>DIP</i> Delivery	21
Figure 7 - [3.1] Prepare <i>DIP</i> Delivery.....	22

Figure 8 - [3.2] Deliver <i>DIP</i>	22
Figure 9 - [3.3] Access <i>DIP</i>	23
Figure 10 - [4] <i>DIP</i> Management.....	25
Figure 11 - Illustration of the identification of requirements.....	27
Figure 12 - Search in descriptive metadata and data	29
Figure 13 – Issue Order.....	33
Figure 14 - Validate Order	36
Figure 15 - Modify <i>DIP</i> for databases	44
Figure 16 - Modify <i>DIP</i> for single records	47
Figure 17 - Search in database with “Google” functionality.....	50
Figure 18 - Search with existing forms.....	53
Figure 19 - Search in database with SQL / DBMS functionality	55
Figure 20 - Search with a combination of google search and SQL / DBMS functionality.....	58
Figure 21 - Search in single records	62
Figure 22 - Analyze with OLAP.....	65
Figure 23 - Overview of components, modules and storage areas used to prepare and deliver the <i>DIP</i>	85

Tables

Table 1 – [1] Search & Order Management.....	17
Table 2 – [2] <i>DIP</i> Preparation.....	20
Table 3 – [3] <i>DIP</i> Delivery.....	23
Table 4 – [4] <i>DIP</i> Management	25
Table 5 - Search in descriptive metadata and data	29
Table 6 - Issue Order.....	33
Table 7 - Validate Order.....	36
Table 8 - Check <i>DIP</i> availability and suitability.....	40
Table 9 - Modify <i>DIP</i> for databases.....	44
Table 10 - Search in database with “Google” functionality	50
Table 11 - Search with existing forms	53
Table 12 - Search in database with SQL / DBMS functionality.....	55
Table 13 - Search with a combination of google search and SQL / DBMS functionality.....	59
Table 14 - Search in single records	62
Table 15 - Requirements derived from the E-ARK Common Specification	67
Table 16 - Requirements for Search and Select IPs	68
Table 17 - Requirements for Create Initial Order	68
Table 18 – Requirements for Validate Order	68
Table 19 - Requirements for Prepare <i>DIP</i>	69
Table 20 - Requirements for Presentation and Search.....	70
Table 21 - Rendering information metadata elements	73
Table 22 - Authenticity metadata elements	74
Table 23 – Access restrictions metadata elements.....	76

Table 24 – *DIP* status metadata elements 78

Table 25 – Dissemination notes metadata elements 79

Table 26 – *DIP* order elements and their descriptions 82

Table 27 - Glossary..... 89

1 Executive summary

The primary aim of this deliverable¹ is to present the first version of the E-ARK *Dissemination Information Package (DIP)*² format. In order to do so the deliverable describes the workflows and use cases of archival access services and ultimately uses these to present a set of requirements which should be followed when designing a DIP format.

As access to archival records is largely dependent on the used tools and environments the secondary aim of the deliverable is to also go beyond the DIP format and look closely at the tools needed for preparing and using the DIP. Therefore this deliverable is also going to be used in E-ARK as the basis for the following tool development and to assure that the tools and DIP requirements are well aligned.

The purpose is to allow for the subsequent development of E-ARK tools for *access*³.

Chapter 2 Purpose and Method describes the process and methodology for creating the *DIP* format.

Chapter 3 High-Level Illustration and Textual Description define the overall scope of the *access* activities to be supported by the E-ARK project. It provides the product context, identifies assumptions and dependencies, and describes each process step of the High-Level illustration.

Chapter 4 Access use Cases derives use cases from the identified process steps in Chapter 3. It also identifies acceptance criteria where relevant.

Chapter 5 Requirements for E-ARK DIP format identifies the requirements needed to build the E-ARK DIP format, which in turn constitutes the basis for the development of *Access* tools.

Chapter 6 DIP Format provides both an overall description of the *DIP* format and a detailed specification of the metadata elements to be included in the *DIP*.

Chapter 7 High-level Component Overview provides a preliminary technical overview of software components and information flow, displaying the main building blocks of the *access* system.

Chapter 8 Glossary lists key terms and defines them.

Chapter 9 References contains the sources used to complete this deliverable.

¹ D5.2 E-ARK DIP draft specification.

² To the extent possible, OAIS terms are employed. OAIS terms are *italicized*.

³ Cf. Tasks 5.4 and 5.5 of the Description of Work, pp. 21-22.

2 Purpose and Method

The purpose of this deliverable is to define an E-ARK *DIP* format and to enable the development of the *access* software components that are going to process it and be part of the reference implementation^{4 5}. The current document is a *draft* and will serve as the basis for the multiple iterations that will be worked on by the archivists and the developers until the E-ARK *DIP* pilot specification is completed and delivered in Month 27 (April 2016)⁶. The end goal is to define the *access* requirements in sufficient detail to allow developers to program the *access* software components needed to underpin the user needs identified in a previous deliverable⁷.

The targeted audiences of the present deliverable are:

1. The European Commission as it is an official deliverable of the E-ARK-project;
2. The intended users of *Content Information* in E-ARK *DIPs*. This can be both end-users (external users) and archivists (internal users)⁸;
3. The digital preservation practitioners of the archives that need to process the E-ARK *DIP* format. This is the personnel in the archives dealing with *DIPs*;
4. The developers of the E-ARK-project who will develop tools that process the format.

This deliverable is based on existing work and requirements have been identified using both a bottom-up and a top-down approach.

The bottom-up approach identified relevant requirements in the Common Specification⁹; by analyzing best practices¹⁰ and user needs⁷; by examining requirements for the SIP¹¹ and the AIP¹²; requirements from the pilot sites and from *access* supporting tools¹³ were also analyzed; as well as metadata elements from metadata standards¹⁴.

The top-down approach consisted of creating high-level illustrations and descriptions of the generic steps in the whole *access* process. This also contributed to reaching a common understanding between the users (archivists) as well as defining the scope of the *access* activities that need to be underpinned by tools

⁴ The reference implementation is the name for the wall-to-wall digital archival solution that E-ARK will provide. It will be hosted as open source on <https://github.com/eark-project>, enabling users to either download the whole solution or hand-pick the software components that they want to integrate in their running digital archive.

⁵ Terms particularly important in the E-ARK project are underlined and explained in Chapter 8.

⁶ D5.3 E-ARK *DIP* pilot specification (revision of D5.2).

⁷ D5.1 GAP report between requirements for access and current access solutions.

⁸ E-ARK distinguishes users in two: Internal users, cf. archivists, and external users, cf. end-users.

⁹ Internal E-ARK deliverable: Common Specification for IPs in the EARK project.

¹⁰ D3.1 Report on available best practices and D4.1 Report on available formats and restrictions.

¹¹ D3.2 E-ARK SIP draft specification

¹² D4.2 E-ARK AIP draft specification

¹³ D2.1 General pilot model and use case definition.

¹⁴ Internal E-ARK deliverable: E-ARK *DIP* Format Requirements.

developed in E-ARK. Subsequently, the identification of these generic process steps enabled the creation of use cases that have:

1. served as communication platforms between archivists and developers and thus been used to facilitate the creation of a deployment environment facilitating agile development where short feedback cycles quickly rectify potential misconceptions; and
2. completed the identification of the requirements of the *DIP* format.

Where appropriate the use cases were then enriched with acceptance criteria that in essence define quality goals (how will the product satisfy the user?).

The two approaches are complementary and were adopted to ensure that all requirements were taken into account in the development of the E-ARK *DIP* format and that communication between the two main parties (archivists and developers) would run smoothly.

3 High-Level Illustration and Textual Description

This chapter presents a high-level illustration of the *access* flow and its associated high-level description.

Both have come into existence via a series of physical and virtual meetings between archivists and developers of the E-ARK project, and are based on the E-ARK General Model¹⁵.

The purpose of this section is to:

1. Create an *access* process diagram generic enough for all archives to acknowledge¹⁶, and detailed enough to be used for purpose #4, see below;
2. Define the scope of *access* within the E-ARK project;
3. Create a platform for discussion;
4. Identify use cases in order to subsequently specify functional requirements for the E-ARK *DIP* format and access tools.

The E-ARK tools that will be developed need to provide support for the search, preparation, delivery and management of the *information objects* desired by a *consumer*. The main features that the E-ARK tools should provide are:

- ◆ Distinction between different roles (e.g. archivist; end-user)
- ◆ Validation of orders against the E-ARK order.xml schema
- ◆ Present and manage orders with their statuses
- ◆ Interface with the storage system(s) in order to fetch the *IP*
- ◆ Temporarily store the *DIP* before it is delivered to the *consumer*
- ◆ Management of *DIPs* depending on their status
- ◆ Present a list of *DIPs* with their statuses
- ◆ Insight into the structure of *DIP* (list of computer files that make up the *DIP*) with the possibility to *access* and render. General or specific viewers will be available depending on the content data type at hand
- ◆ Preparation of *DIP* by the Presentation tool
- ◆ *DIP access* for the end-user to view, search and navigate the content of the *DIP*.

Furthermore, the end-users should be able to have appropriate tools to be used according to their preferred *access* scenario (*access* in the reading room, full data export, ftp, etc.).

¹⁵ The General Model can be found here:

<http://www.earkadmin.com/WP2/Forms/AllItems.aspx?RootFolder=%2FWP2%2FGeneral%20Model&FolderCTID=0x01200032C6018B55647C4589C1BA5C211FB229&View={866C752C-88FD-4CEC-A859-58528116FC60}>

¹⁶ All archives of the E-ARK project have approved of it.

A *DIP* may be the result of the *transformation* of a part of one *AIP*, one *AIP* or several *AIPs* at the request of an end-user. However, it can also be the result of a decision of the *Archive* to prepare a *DIP* in advance to simplify and facilitate user access to the *information object* in question.

3.1 Product Context

This document is a requirements specification for the E-ARK *Dissemination Information Package*¹⁷ format, just called in short the E-ARK *DIP* format, or the product.

The requirements for the format are being developed by describing the processes of *access*, resulting in both harmonized processes and a common format. Common and harmonized tools will be developed to support this format and the harmonized archival processes.

The *DIP* format is the last in sequence of the three *IP* formats defined in the OAIS reference model. The two others, the E-ARK *SIP* and the E-ARK *AIP* format are being defined in WP3 and WP4. All three formats use the same Common Specification in order to ensure compatibility between *IP* formats.

3.2 Assumptions

The requirements are based on the following assumptions:

- ◆ The order.xml will be defined with information on:
 - Users and their credentials¹⁸;
 - Relationships between *AIP(s)* and the ordered *information objects* (each ordered object is based on a Unit of description ID that is related to a concrete *AIP* ID, which leads to the object in the *AIP* storage);
- ◆ There is an index of existing *DIPs*, which allows the tools to check whether there is a need to retrieve data from *AIP* storage and create a new *DIP*, or if it is possible to re-use one or more of the existing *DIPs*.
- ◆ In different *archives* different *access* policies exist, so there may be different possible statuses of *DIPs*. The tool should therefore support configuration and *DIP status* management during the *access* process. Statuses can be changed during automated or manual processing of data, because processing of data (e.g. checking access restrictions, adjusting *DIP*) can have a direct influence on the status of *DIP* (DIP_o, DIP_u, DIP_p). Of course it depends on records and archival policy.. The list of possible statuses will be prepared and be configured to reflect the exact *DIP* workflow for each *archive*.
- ◆ For widespread types of computer files (e.g. .pdf, .tiff, .mp3, .xml) already available tools are used. For unstructured computer files general viewers are available (e.g. Adobe Reader) and installed at the end-user workstation.
- ◆ For more complex data content types instructions will be provided on how tools have to be configured and used (e.g. QGIS or ERMS).
- ◆ E-ARK will develop a specific Presentation Tool for accessing archived databases

¹⁷ To the extent possible, OAIS terms are employed. OAIS terms are *italicized*.

¹⁸ Credentials may be an integral part of the order or there is a specific list of registered users and their credentials.

- ◆ *DIP* storage depends on local implementation, archival *access* policy and data content types.
- ◆ The predecessor tool of the Order Management Tool is the Archival Catalogue, and the *DIP* will be displayed by the E-ARK Access Tools.
- ◆ The Archival Catalogue provides the order.xml.
- ◆ The Order Management Tool fetches the *DIP* either in the *DIP* storage or by requesting the *transformation* of an *AIP* into a *DIP* by the *AIP-DIP Transformation Tool* using the *AIP* storage system interface.
- ◆ The *DIP* is given to the Access Tools using a pertinent delivery scenario.
- ◆ The Archival Catalogue and Order Management Tool enable easy maintenance of the *DIP* workflow for the administrator and for an end-user

3.3 Dependencies

The dependencies that affect the requirements:











- ◆ Search tools will be partly based on local implementations of cataloguing systems
- ◆ Defined structure of Order (XML schema)
- ◆ Defined structure of *DIP* (XML schema)
- ◆ Defined structure of the list of registered users and their credentials
- ◆ Defined requirements concerning the structure of OLAP Cube
- ◆ Defined manual and automated steps of the *access* procedure (in 2.2.2.4)
- ◆ Defined types of records (DB, ERMS, GEO ...)
- ◆ A functioning tool for ordering, or defined requirements for this tool
- ◆ A functioning tool for transformation of *AIPs* to *DIP*, or defined requirements for this tool
- ◆ Dependencies between software components, cf. Figure 23, page 85.

3.4 Illustrations and descriptions

This illustration and description of the process steps refers to only part of the standard procedure for accessing archived records. It does not describe in details those parts of the process which are related to the ordering, displaying or accessing of data by way of an OLAP cube¹⁹.

¹⁹ Data mining scenarios will be developed at a later stage.

Legend of graphical symbols used in the description below:

Graphical symbol	Description
	Process start event.
	Marks process that involves sub-processes.
	Process step (can be manual, semi-automated or fully automated depending on several different reasons: archival policy, <i>DIP</i> existence etc.).
	Manual process step. The tool supports carrying out the procedure manually in a way that allows manual input of information (i.e. information about activities e.g. change of <i>DIP</i> status).
	Fully automated process step. The tool supports carrying out the procedure in an automated way that does not need any manual input of information (e.g. creating <i>DIP</i> from <i>AIPs</i> metadata in the Order).
	Semi-automated process step. The tool supports the implementation process in a way that part of input data is automatically processed; the final decision is taken and manually entered by the <u>user</u> of tool (e.g. Entering the decision on the validation of a request). ²⁰
	1. Data and/or metadata necessary for carrying out the procedure; accessible to the tool without any further <u>user</u> interference. 2. Data and/or metadata, which are the result of execution of the procedure.
	Data and metadata input or output from the process in the form of a formal document (i.e. request order).
	Decision gateway. Step in the process, in which decision is made (by the <u>user</u> or by the tool)..
	Process end event.

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

1. “**Search & Order Management**” where the *consumer* wants to find information within the *archive*, then orders it;
2. “**DIP Preparation**” where the *DIP* is prepared for the *consumer*, for example by transforming an *AIP* into a *DIP*;
3. “**DIP Delivery**” where the *DIP* is delivered to the *consumer* via a suitable Graphical User Interface (GUI);
4. “**DIP Management**” where the *DIP* is either deleted or sent to a permanent or temporary *DIP* storage.

²⁰ Manual, semi-automated and automated indicators will probably be removed in the next version as this depends on local implementations and policies.

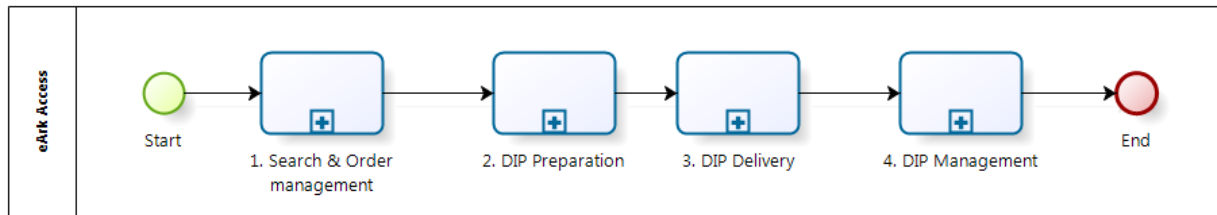
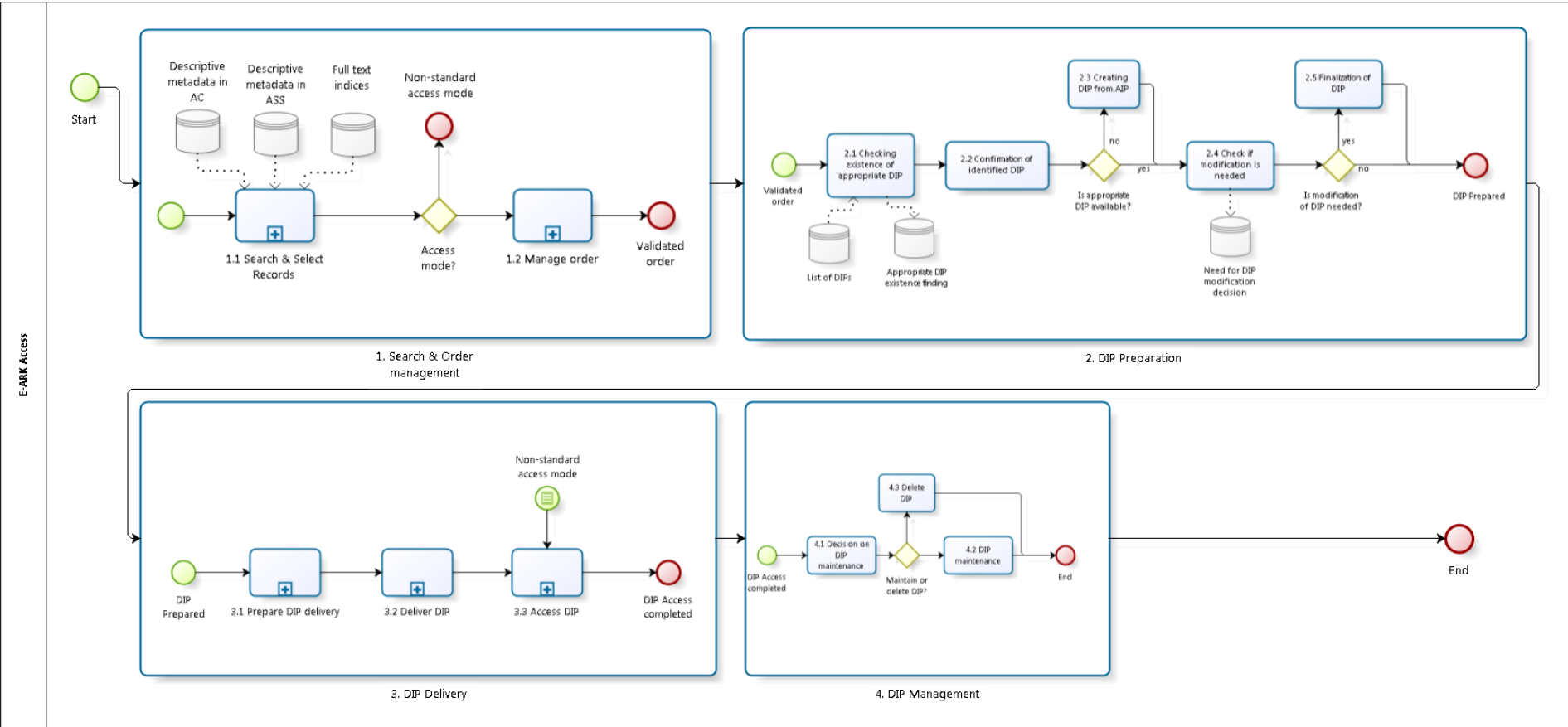


Figure 1 - E-ARK Access high-level process

And an expanded illustration of the E-ARK access process:



3.4.1 Search & Order Management

“Search & Order Management” describes the first step of the E-ARK *access* process where the *consumer* wants to find information within the *archive*, then orders it.

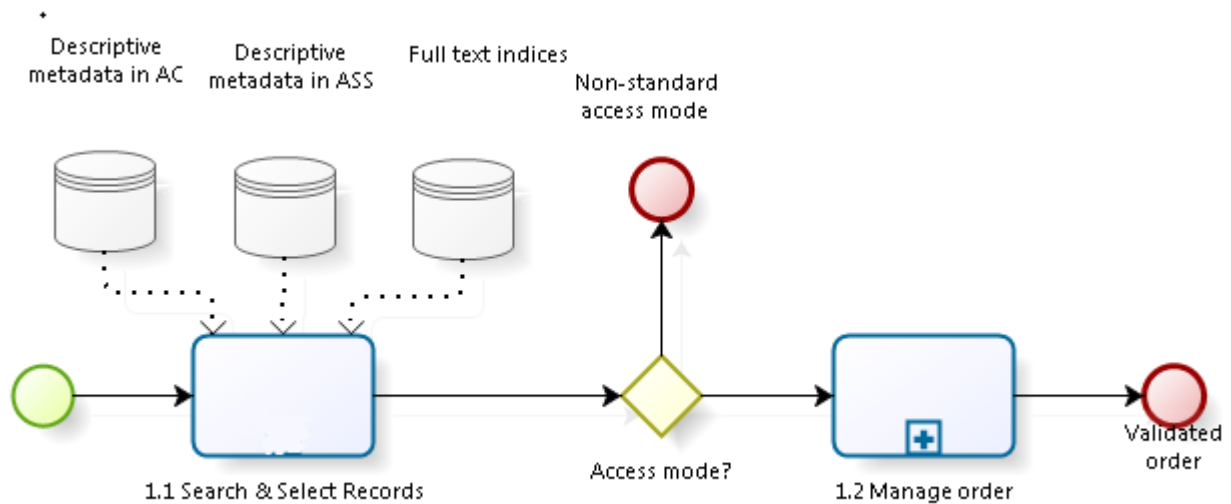


Figure 3 - [1] Search & Order Management

The process “Search & Order Management” consist of two sub-processes: “1.1 Search & Select Records” and “1.2 Manage Order”.

3.4.1.1 Search & Select Records

“1.1 Search & Select Records” is the process during which the end-user consults different sources and search indices to ultimately select a set of desired *information objects*.

3.4.1.2 Manage Order

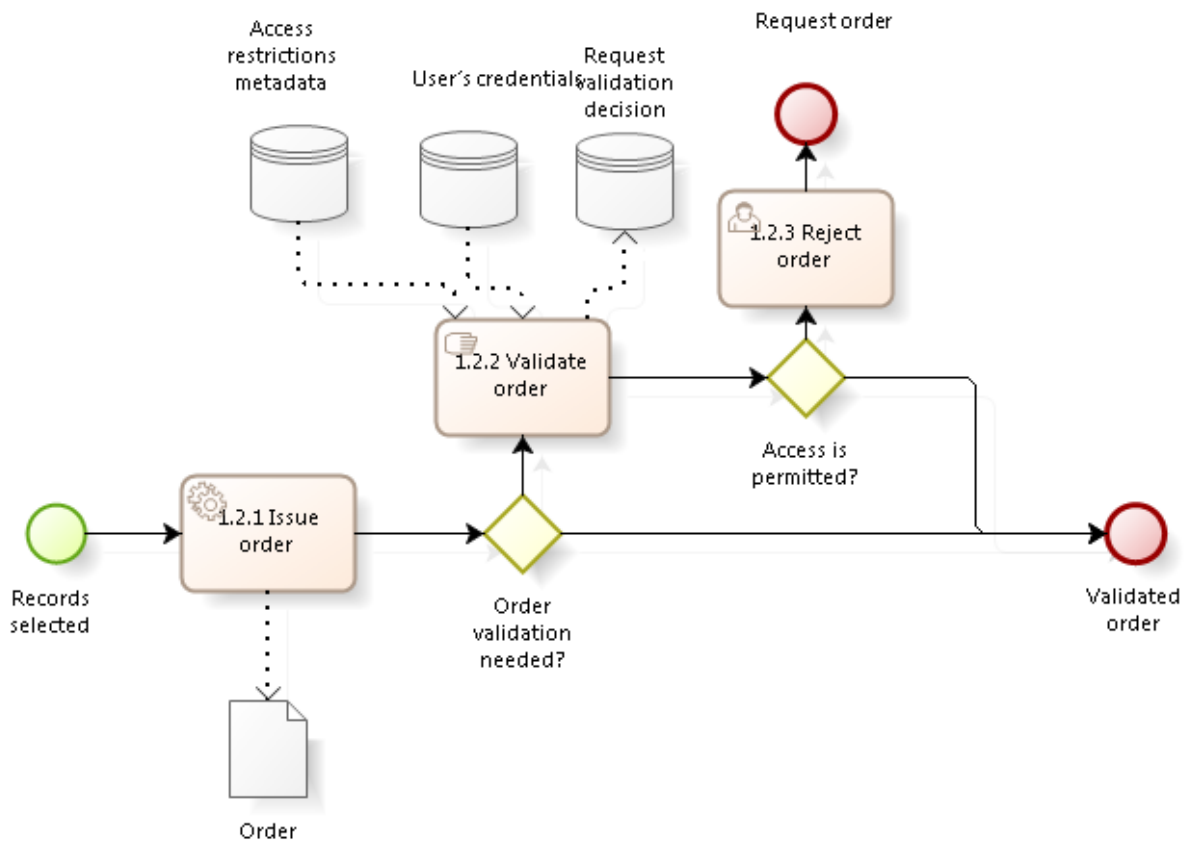


Figure 4 - [1.2] Manage order

All processes are described in the table below.







Table 1 – [1] Search & Order Management

ID	Step	Description	Mandatory/Optional	Tool	In-/Output	Use cases
	Start	Process start event.				
1	Search and Order management	Processes that cover end-user's and archivist's actions regarding search and order				
1.1	Search and Select Records	By searching the <u>descriptive metadata</u> the user selects the records he/she wants to <i>access</i> .	M	AC ²¹	I: <u>Descriptive metadata</u> O: Selected records	
	<u>Descriptive metadata in AC</u>	<u>Descriptive metadata</u> in local module of AIS ²² – ASS ²³ (<i>Finding Aid/catalogue</i>) as an input for this process.	O	AC		
		<u>Descriptive metadata in AIP storage system - ASS</u> as an input for this process.	O	ASS		

²¹ Archival catalogue.







²² Archival Information System.

²³ Access Services or Solutions.

ID	Step	Description	Mandatory/ Optional	Tool	In-/Output	Use cases
	<u>Descriptive metadata</u> in ASS					
	 Access mode?	Depending on the <i>access</i> environment the process has three flow options. <ol style="list-style-type: none"> The standard procedure for <i>access</i> starts with the issue of the order. A second process flow is possible when there is already a pre-created <i>DIP</i> for selected records while accessing data in the <i>DIP</i> is not restricted and available online. In this case, there is no need to issue the order. The third process flow covers activities mostly done by or within a presentation or data mining platform²⁴. 	M			
1.2	 Manage order	Actions needed when <i>DIP</i> is not directly accessible.				
1.2.1	 Issuing the Order	If the selection of records does not enable immediate <i>access</i> to a <i>DIP</i> or <i>access</i> to an <u>OLAP cube</u> , an order must be issued. (Standard procedure for <i>access</i> starts.)	M – for search and ordering tool	<u>AC</u>	I: Selected records, Id of <u>end-user</u> O: Order	
	 Order	<i>Information object</i> created by <u>AC</u> (or <u>ASS</u>) with information on requested records (<i>AIPs</i> or part of <i>AIP</i>), the <u>end-user</u> information (including credentials and skills) and <i>access</i> restrictions on requesting records (if available). Order will be delivered to <u>OMT</u> ²⁵ .	M – information on requested records O – information on <u>user's</u> credentials O – <i>access</i> restrictions on requesting records	<u>AC</u> <u>OMT</u>		
	 Access restrictions metadata	Metadata about <i>access</i> restrictions for the ordered records.	M – if the order does not contain information on requested records <i>access</i> restrictions metadata	<u>ASS</u>		
	 Order validation needed?	The <u>OMT</u> checks ordered records <i>access</i> restrictions for the purpose of validation of the request.	M	<u>OMT</u>	I: Order (list of requested records), <i>Access</i> restrictions metadata O: Decision (Y/N)	

²⁴ Data mining is not inside the scope of this deliverable, but is nevertheless shown in some of the diagrams for purposes of completeness.

²⁵ Order Management Tool.

ID	Step	Description	Mandatory/Optional	Tool	In-/Output	Use cases
	 User's credentials	Information about <u>end-user's access rights</u> (could be based on types of <u>users</u>) provided by <u>UMS</u> ²⁶ have to be available to <u>OMT</u> .	M – if order does not contain information on the <u>user's</u> credentials	<u>UMS</u>		
1.2.2	 Validate order	The <u>OMT</u> validates the request. It checks <u>end-user's</u> credentials against ordered records' <u>access</u> restrictions; then provides the information about the <u>access</u> restrictions assigned to the records in order and <u>end-user's</u> <u>access</u> permissions. An <u>archivist</u> takes a final decision on <u>access</u> permission. (If there is no validation needed, this step can be skipped.)	M – if there are restrictions on <u>access</u> to requested records	<u>OMT</u>	I: Order (ordered records), <u>Access</u> restrictions metadata, <u>User's</u> credentials	
	 Order validation decision	Metadata about order validation decision.		<u>OMT</u>	O: Metadata	
	 Access is permitted?	Decision gateway. Direct process flow according to its predefined condition. (If no validation is needed, this step can be skipped.)	M – if there are restrictions on <u>access</u> to requested records	<u>OMT</u>	O: Decision (Y/N)	
1.2.3	 Reject order	The <u>OMT</u> generates the draft of an <u>access</u> rejection notice and the <u>archivist</u> will finalise it; then the notice is sent to the <u>user</u> . ²⁷ (If there is no validation needed, this step can be skipped.)	M – if <u>access</u> to ordered records is not permitted	<u>OMT</u>	I: Order validation decision	
	 List of <u>DIPs</u>	List of existing <u>DIPs</u> (pre-created or ordered by other <u>users</u>)	M	<u>OMT</u>	O: Metadata	

3.4.2 DIP Preparation

“DIP Preparation” describes the steps in which the DIP is prepared for the consumer, for example by transforming an AIP into a DIP.

²⁶ User Management System.

²⁷ The process can be continued with access negotiations or other comparable procedures, which may result in a change of user credentials. Depending on the archives access policy access process at this stage can be suspended or terminated. (Negotiation process is not part of the scope of the project.)

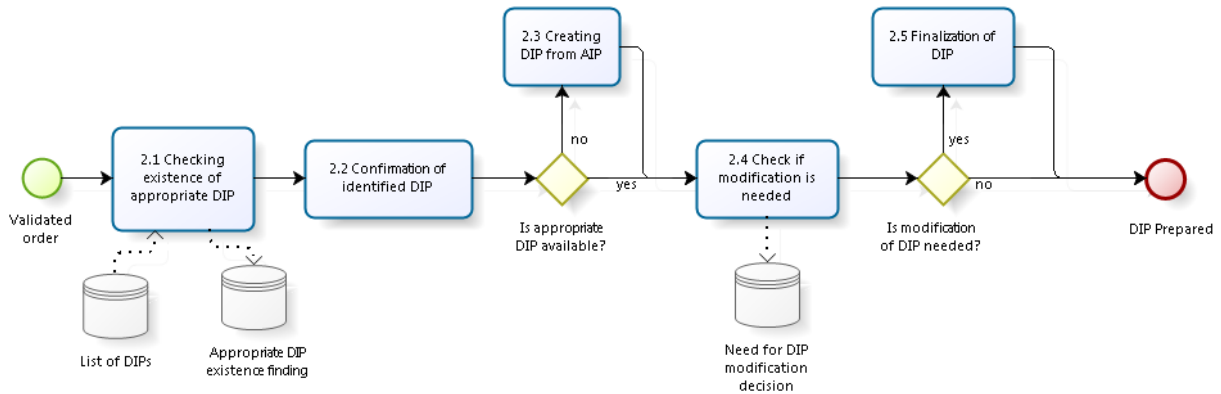







Figure 5 - [2] DIP Preparation

The process “DIP Preparation” does not comprise any sub-processes that need additional breakdown.

All processes are described in the table below.

Table 2 – [2] DIP Preparation

ID	Step	Description	Mandatory/Optional	Tool	In-/Output	Use cases
2	<u>DIP Preparation</u>	Checking for and finding or creating appropriate <i>DIP</i> .				
2.1	Checking existence of appropriate <i>DIP</i>	<u>OMT</u> checks based on the content of the Order and list of existing <i>DIPs</i> (it contains information of contained <i>AIPs</i> or parts of <i>AIP</i> from which <i>DIPs</i> was created) if there is a suitable <i>DIP</i> and thus transformation of data from <i>AIP</i> in <i>DIP</i> is not needed? If appropriate <i>DIP</i> exists, it will be re-used in further processing.	O	<u>OMT</u>	I: Order, <i>DIP</i> O: <i>DIP</i>	
	Finding appropriate <i>DIP</i>	Metadata about finding an existing appropriate <i>DIP</i> .	O	<u>OMT</u>	O: Metadata	
2.2	Confirmation of identified <i>DIP</i>	The <u>archivist</u> additionally checks whether the identified <i>DIP</i> is suitable for re-use in general and by the specific <u>end-user</u> . (If there is no identified <i>DIP</i> , this step can be skipped.)	M – if appropriate <i>DIP</i> exists	<u>OMT</u>	I: Identified appropriate <i>DIP</i> O: <i>DIP</i>	
	Is appropriate <i>DIP</i> available?	Decision gateway. Based on result of previous steps.	O	<u>OMT</u>	I: Order, List of <i>DIPs</i> O: Appointed appropriate <i>DIP</i> (if exists)	

ID	Step	Description	Mandatory/Optional	Tool	In-/Output	Use cases
2.3	 Creating <i>DIP</i> from <i>AIP</i> ²⁸	Creating the <i>Dissemination Information Package (DIP)</i> from the <i>AIP(s)</i> . (If there is already an identified appropriate <i>DIP</i> , this step can be skipped.)	M – if identified appropriate <i>DIP</i> doesn't exist	<u>ASS</u>	I: Order, <i>AIPs</i> O: <i>DIP</i>	
2.4	 Check if modification is needed	The <u>archivist</u> checks if any modifications to the <i>DIP</i> should take place. Modifications may be necessary, for example, in order to set up a specific <i>access</i> tool (e.g. by unpacking a SIARD file), to prevent <i>access</i> to restricted data (e.g. hide sensitive personal data), or to meet the <u>end-user</u> request of gaining <i>access</i> to only a subset of data from the initial <i>DIP</i> .	O	<u>OMT</u>	I: Order, <i>DIP</i> O: Decision (Y/N)	
	 Need for <i>DIP</i> modification decision	Metadata about <i>DIP</i> modification decision (<u>status</u> in list of <i>DIPs</i>).	O	<u>OMT</u>	O: Metadata	
	 Is modification of <i>DIP</i> needed?	Decision gateway. Based on decision in step 2.4.	O	<u>OMT</u>	I: Decision Y/N from 2.4	
2.5	 Finalisation of <i>DIP</i>	The <u>archivist</u> manually modifies the content of <i>DIP</i> . (E.g. hides sensitive data, lowers the resolution of the pictures, extracts a smaller set of data from entire DB, etc.)	O	ATs	I: Order, <i>DIP</i> O: (finalized) <i>DIP</i>	

3.4.3 *DIP* Delivery

“*DIP* Delivery” describes delivery of the *DIP* to the *consumer* via a suitable Graphical User Interface (GUI);

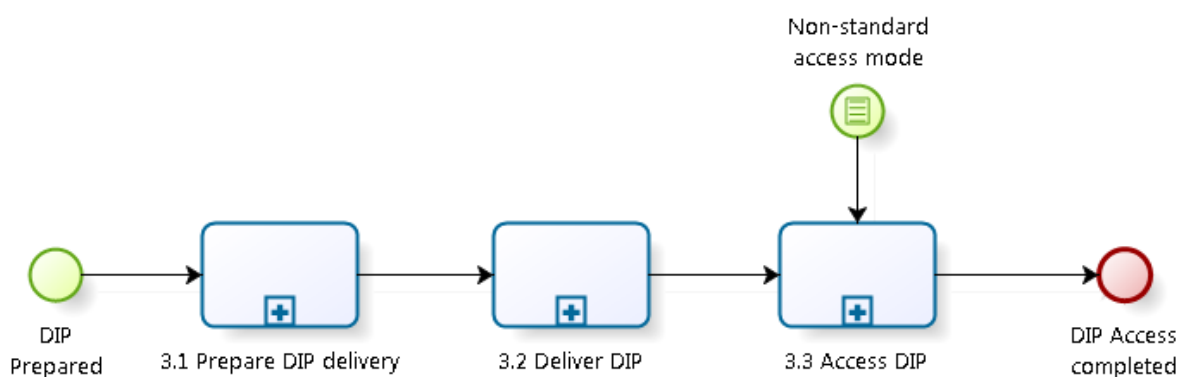


Figure 6 - [3] *DIP* Delivery

The process “*DIP* Delivery” consists of 3 sub-processes: “Prepare *DIP* Delivery”, “Deliver *DIP*” and “Access *DIP*”

²⁸ This step is described in detail at AIP to *DIP* transformation tool.

3.4.3.1 Prepare DIP delivery

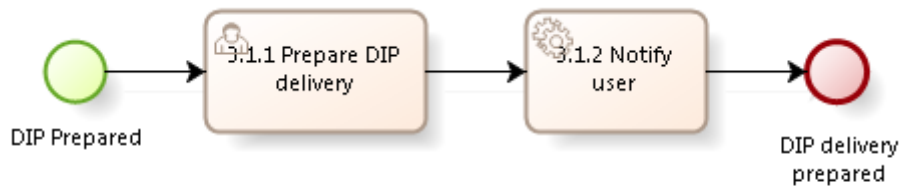


Figure 7 - [3.1] Prepare *DIP* Delivery

3.4.3.2 Deliver *DIP*

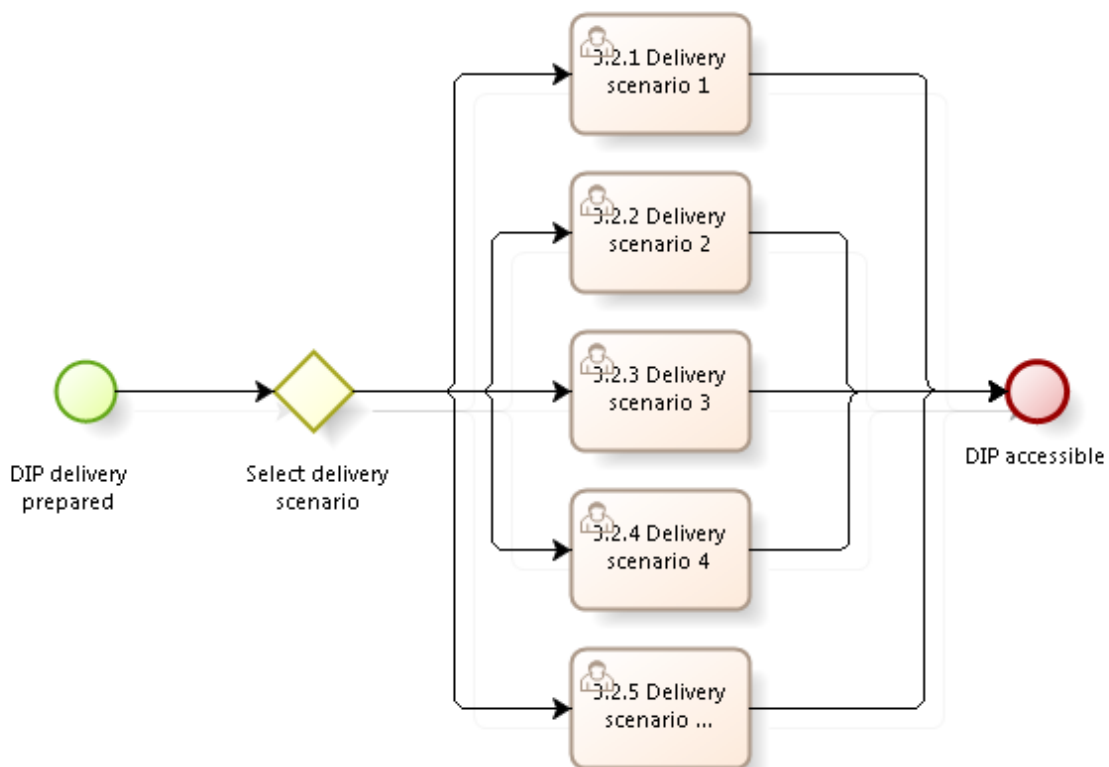


Figure 8 - [3.2] Deliver *DIP*

3.4.3.3 Access DIP

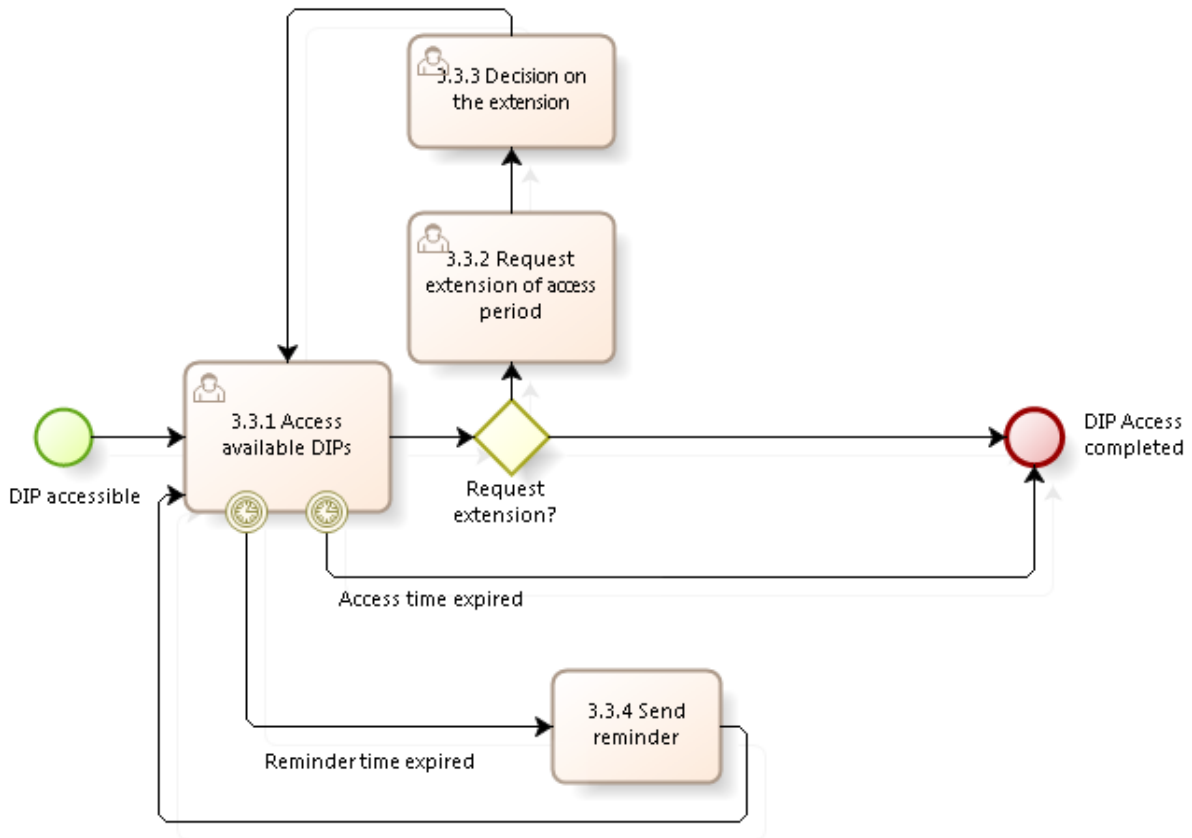















Figure 9 - [3.3] Access DIP

All processes are described in the table below.

Table 3 – [3] DIP Delivery

ID	Step	Description	Mandatory/ Optional	Tool	In-/Output	Use cases
3	DIP Delivery	Process that includes actions from preparing the <i>DIP</i> to end of use of the <i>DIP</i> .				
3.1	Prepare <i>DIP</i> delivery	Enabling <i>access</i> to the <i>DIP</i> and <i>user</i> notification.				
3.1.1	Enable <i>DIP</i> delivery	Based on the <i>user</i> requirements in the order, the ordered record's types, and in the context of the archival <i>access</i> policy the <i>archivist</i> enables <i>access</i> (change <i>status</i> in list of <i>DIPs</i>) and defines the <i>access</i> scenario.	M	OMT	I: Order, List of <i>DIPs</i> O: List of <i>DIPs</i> , defined scenario type	
3.1.2	Notify <i>user</i>	Based on the decisions and activities in the previous steps (modification is needed / finalisation) the <i>status</i> of the <i>DIP</i> is automatically changed into "accessible" and "notify <i>end-user</i> ". The <i>DIP</i> is on the	M	OMT	I: List of <i>DIPs</i> , defined scenario type O: <i>DIP</i> in	

ID	Step	Description	Mandatory/Optional	Tool	In-/Output	Use cases
		list of <i>DIPs</i> available for display.			<u>status</u> "accessible"	
	 Selecting delivery scenario	Decision gateway.	M	<u>OMT</u>	I: List of <i>DIPs</i> , defined scenario type	
3.2	Deliver <i>DIP</i>	Executing a delivery scenario to make the <i>DIP</i> accessible.				
3.2.1	 Delivery scenario 1	Simple unstructured records (e.g. .pdf, .tiff, .mp3, .xml): According to the archival policy, the <i>DIP</i> (ordered records with accompanying metadata) can be sent out by e-mail, put on ftp folder, or be available in the Archival reading room. For rendering, the general viewers installed at the <u>end-user</u> workstation (e.g. Adobe Reader) will be used.	M	ATs	I: <i>DIP</i>	
3.2.2	 Delivery scenario 2	Geo-data: According to the archival policy, the <i>DIP</i> (ordered records with accompanying metadata) with a <u>user</u> guide for <i>access</i> , including information on the specific tools for rendering, can be sent out by e-mail or put on a ftp folder. The <i>DIP</i> with accompanying <u>user</u> guides can be retrieved also in an archival reading room.	M	ATs	I: <i>DIP</i> , <u>user</u> guide	
3.2.3	 Delivery scenario 3	<u>Databases</u> (ERMS included): The <i>DIP</i> will be loaded into a Relational Database Management System (RDBMS) and <i>accessed</i> via a GUI that sits on top of it. <i>Access</i> to the data will be done via SQL queries.	M	ATs	I: <i>DIP</i> , <u>user</u> guide,, manuals	
3.2.4	 Delivery scenario 4	Use of specific ATs in the reading room: According to the archival policy, the <i>DIP</i> or its parts can be retrieved only by specific ATs in an archival reading room. In that scenario, the <u>user</u> will get <u>user</u> guides and <i>access</i> permission for a specific AT.	M	ATs	I: <i>DIP</i> , <u>user</u> guide, manuals	
3.2.5	 Delivery scenario ...	If needed, additional scenarios will be added.				
	...					
3.3	 Access <i>DIP</i>	Managing the <i>DIP</i> during the <i>access</i> period.				
3.3.1	 Access available <i>DIPs</i>	Ordered <i>DIPs</i> show on <u>user's</u> list of <i>DIPs</i> . The <u>user</u> can view <i>DIPs</i> , request an extension of the <i>access</i> period and complete usage of the <i>DIPs</i> .	M	<u>OMT</u>	I: List of <i>DIPs</i>	
	 Request extension?	Decision gateway. According to the policies of archival institutions, the <i>DIP</i> can be accessed only for a limited period of time. In that case the <u>user</u> can	O	<u>OMT</u>		

ID	Step	Description	Mandatory/Optional	Tool	In-/Output	Use cases
		request an extension of the <i>access</i> period.				
3.3.2	 Request extension of <i>access</i> period	The <i>end-user</i> can check how long his/her <i>DIPs</i> are available for him/her. The <i>end-user</i> can request an extension of the <i>access</i> period for each <i>DIP</i> .	O	OMT	I: List of <i>DIP</i> 's O: Request for extension of that <i>access</i> period	
3.3.3	 Decision on the extension	An <i>archivist</i> may approve the extension request or not. <i>Access</i> period data in the list of <i>DIPs</i> shall be amended in accordance with the decision and the <u>OMT</u> should notify the <i>end-user</i> .	O	OMT		
3.3.4	 Send reminder	<u>OMT</u> automatically sends a reminder that the <i>access</i> period will expire shortly.	O	OMT		
3.3.5	 Completion of <i>access</i>	After the completion of <i>access</i> , the <i>end-user</i> changes the <i>status</i> of the <i>DIP</i> in the list of <i>DIPs</i> .	M	OMT		

3.4.4 3.1.4 DIP Management

“*DIP* Management” is where the *DIP* is either deleted or sent to a permanent or temporary *DIP* storage

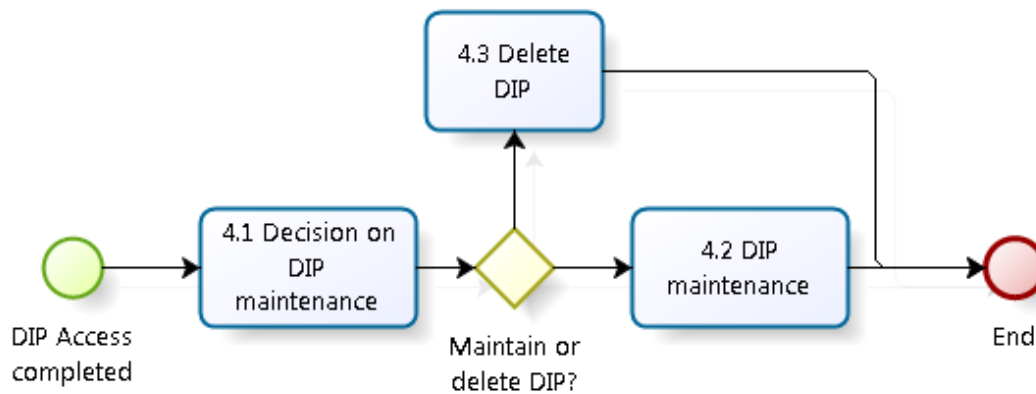









Figure 10 - [4] *DIP* Management

All processes are described in the table below.

Table 4 – [4] *DIP* Management

ID	Step	Description	Mandatory/Optional	Tool	In-/Output	Use cases
4	 <i>DIP</i> Management	Actions applied to the <i>DIP</i> after the completion of the <i>access</i> period.				

ID	Step	Description	Mandatory/ Optional	Tool	In-/Output	Use cases
4.1	 Decision on <i>DIP</i> maintenance	According to the archival policy and the content of <i>DIP</i> , the <u>archivist</u> takes a decision on whether to maintain or delete the <i>DIP</i> . The decision is entered into the <u>OMT</u> .	M	<u>OMT</u>		
	 Maintain or delete <i>DIP</i> ?	Decision gateway. Based on decision in previous step.	M	<u>OMT</u>	I: <u>Archivist's</u> decision	
4.2	 <i>DIP</i> maintenance	<i>DIP</i> <u>status</u> is changed to "persistent".	M	<u>OMT</u>		
4.3	 Deletion of <i>DIP</i>	The <i>DIP</i> (with its entire content) is automatically deleted from the repository of <i>DIPs</i> .	M	<u>OMT</u>		
	 Deletion of "persistent" <i>DIP</i>	If needed, the <u>archivist</u> can always delete a "persistent" <i>DIP</i> .	M	<u>OMT</u>		
	 End	Process end event.				

4 Access use cases

This chapter presents the use cases that have been identified for the E-ARK Access process as outlined in chapter 3.

Each of the sections below represents a process step from the high-level workflow. Each process step (1) is associated with one or more use cases (2). Where appropriate the use cases have a series of acceptance criteria attached to them, and have been used to deduct requirements (3) presented in Chapter 5 Requirements for E-ARK DIP format.

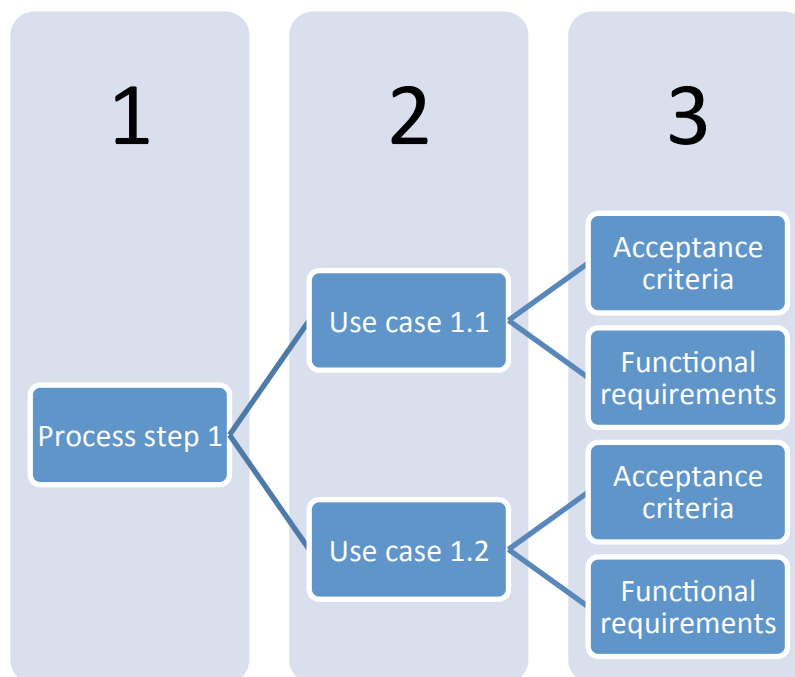


Figure 11 - Illustration of the identification of requirements

The following provides an overview of all use cases and the structure of the present chapter:

1. Process step 1: Search and Order Management
 - 1.1. Search & Select records
 - Search in descriptive metadata and data
 - 1.2. Manage Order
 - 1.2.1. Issue Order
 - 1.2.2. Validate Order
2. Process step 2: *DIP* Preparation
 - 2.1. Checking existence of appropriate *DIP*
 - 2.2. Confirmation of identified *DIP* (2.1 and 2.2 are merged in the use cases)
 - 2.3. Creating *DIP* from *AIP*
 - 2.4. Check if modification is needed
 - 2.5. Finalization of *DIP* (2.3 and 2.4 are merged in the use cases)
 - 2.5.1. Modify *DIP* for databases

2.5.2. Modify *DIP* for single record

2.5.3. *DIP* - GIS content

3. Process step 3: *DIP* Delivery

3.1. Prepare *DIP* delivery

3.2. Provide *Access* rights

3.3. Notify end-user

3.4. Access *DIP*

3.4.1. Database & EDRM-system

3.4.1.1. Search with google functionality

3.4.1.2. Search with existing search forms

3.4.1.3. Search with SQL / DBMS functionality

3.4.1.4. Search with combination of google functionality and SQL / DBMS functionality

3.4.2. Search in single records

3.4.3. Search in GIS data

3.4.4. Analyze with OLAP

4. Process step 7: *DIP* Management

4.1 Process step 1: Search and Order Management

4.1.1 Use case 1.1: Search & Select records: Search in descriptive metadata and data

This use case illustrates an end-user wanting to identify where (and how) (s)he can get information about archival records. The search can be run by using the Archive's *Finding Aid* or an EARK-tool, dependent on the Archive's choices.

It is the local Archive's policy that determines when a user is required to log in, but it will probably be required from the moment an end-user wants to search data. The Order basket is well known from web-shops, but it has low priority here. The use case ends when the user has identified desired *information objects* or an item which can be automatically accessed. No order creation is needed in the latter case.

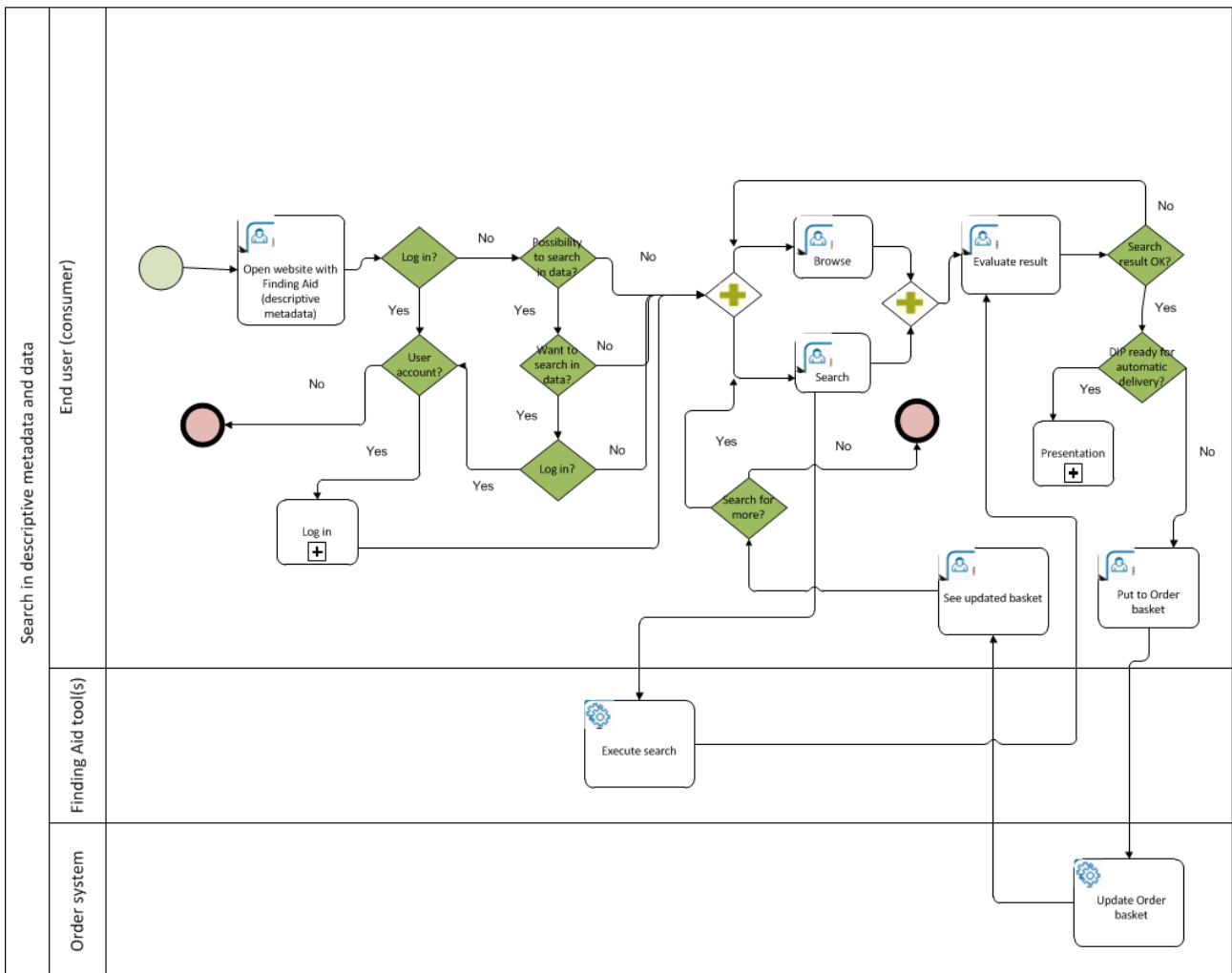









Figure 12 - Search in descriptive metadata and data

The use case illustration is described in the table below.











Table 5 - Search in descriptive metadata and data




Step	Description	Remarks
Start		
Open website with Finding Aid	The user opens the archives general <i>Finding Aid</i> (catalogue) or <i>Finding Aid(s)</i> for specific collection(s). Normally these <i>Finding Aids</i> can be found on the Internet	An example of an online <i>Finding Aid</i> can be seen at: http://www.ra.ee/fotis/index.php?lang=en In some cases the <i>Finding Aid</i> or parts of it are only accessible in reading rooms. E-ARK assumes that most <i>Archives</i> already have online <i>Finding Aids</i> in place and continue using these. However,

		<p>simplistic <i>Finding Aid</i> functionality should be developed by E-ARK, possibly based on Lily²⁹ or any other available out-of-the-box open-source solution.</p>
 Log in?	<p>It depends on the specific archival policy whether a <u>user</u> has to log in to be able to search in the <i>Finding Aid</i></p>	<p>It might be required for statistical, security or restriction reasons.</p>
 <u>User</u> account?	<p>If a <u>user</u> doesn't have a <u>user</u> account, (s)he is prompted to apply for a <u>user</u> account (with a link)</p>	<p>Applying for a <u>user</u> account and set-up of the account on the archive are not in the scope of the E-ARK access framework. It is assumed that the specific archive implementing E-ARK DIP and tools will have a solution in place.</p>
 Log in	<p>Log in according to the specific archival policy and log in tool</p>	<p>It is assumed that the specific archive has a solution for log in.</p>
 End		
 Possibility to search in data?	<p>It is assumed that search in (part of) <u>descriptive metadata</u> is always possible, but it might also be possible to search in data, e.g. inside the content of <u>databases</u> or documents</p>	<p>The answer depends on archival policy and technical capabilities of the system. Only part of data may be indexed. It may only be possible to search in data in reading rooms. E-ARK will implement the functionality specifically on top of the Hadoop³⁰ infrastructure.</p>
 Want to search in data?	<p>The <u>user</u> has the possibility to choose search only in <u>descriptive metadata</u> or also in data</p>	
 Log in?	<p>If the <u>user</u> is not already logged in, log in might be requested for searching in data – as there are often special <i>access</i> rules due to e.g. information about persons</p>	

²⁹ Lily is a data management platform combining planet-sized data storage, indexing and search with on-line, real-time usage tracking, audience analytics and content recommendations: <http://www.lilyproject.org/lily/index.html>.

³⁰ The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models: <https://hadoop.apache.org/>.

		
 Browse	Browse through the hierarchy of the <i>Finding Aid</i>	
 Search	“Google” search and faceted search.	Search functionality for E-ARK could conduct searches in the <i>AIP</i> repository by using the Hadoop / Lily tool.
		
 Execute search		
 Evaluate result	Result with code, title, dates, etc.	
 Search result OK?		
 <i>DIP</i> ready for automatic delivery?	<i>DIP</i> may be immediately available with no restrictions, in which case the <u>user</u> does not need to submit an order	<p>Even if the <i>DIP</i> location is not immediately known to the <i>Finding Aid</i> there could be a quick and automated workflow which asks for the <i>DIP</i>/binary files and presents these next to the object’s metadata.</p> <p>For example: when the <u>user</u> clicks on an object in the search results’ list the system makes an automated query “Is there a <i>DIP</i> available” and the return might be “Yes, at URL www.xxx.yy” or “No”.</p> <p>According to the result the <u>user</u> interface shows the object visually/inline, offers file download possibilities, etc.</p>
 Presentation	The <u>user</u> opens the <i>DIP</i> , for example by clicking at the metadata, and the <i>DIP</i> is presented in the relevant tool, described in the use cases presented in section 4.3.4, Use case 3.4: Access <i>DIP</i> , page 49.	Single records, but may also be <u>databases</u> .
 Send to	Similar to online shopping deposit of desired objects	The primary objective of E-ARK is to implement the automated <i>access</i> workflow without the need for an

Order basket		order basket. Therefore this functionality is of low priority.
 Update Order basket		
 View updated Order basket		In real-life implementation this could be integrated with a module for payment. However, it is beyond the scope of E-ARK.
Search for more?		
 End		

4.1.1.1 Acceptance criteria

Given that an end-user wants to search for information about something, and given that the end-user expects that the desired information can be found in the Archive:

- ◆ When the end-user opens the *Finding Aid*, (s)he gets a message about possibilities and need for a user account if that is required.
- ◆ When the end-user opens the page for search, (s)he can:
 - perform free-text search in all metadata elements with one search field (“Google search”)
 - search on (user) specified metadata elements with use of operators.
- ◆ Given that the end-user is allowed to search in data and documents, when the end-user opens the page for search, (s)he can:
 - search for a value in all data with one search field, including the possibility to use operators
 - search for word(s) or part of word(s) in the content of digital documents.
- ◆ When the end-user gets the result of a search, (s)he can browse metadata concerning the identified *IPs* or parts of *IPs*.
- ◆ When the end-user browses the result list, (s)he can see if the *IP* is automatically accessible or an order is not needed.
- ◆ When the end-user sees that a *DIP* is automatically accessible, (s)he can go to the *DIP*, which is represented in the relevant presentation tool.

4.1.2 Use case 1.2: Manage Order

4.1.2.1 Use case 1.2.1: Issue Order

This use case starts when the end-user has identified the *IPs* with the required information and the *IP* cannot be accessed automatically from the Search tool / *Finding Aid*. The use case illustrates a process with an Order basket similar to a web shop, but this has low priority. Instead the end-user can be asked to complete an order form with data about the user, *IP* and special wishes. The most important part of this use case is the output: an xml-file (order.xml) with all data necessary to process the order.

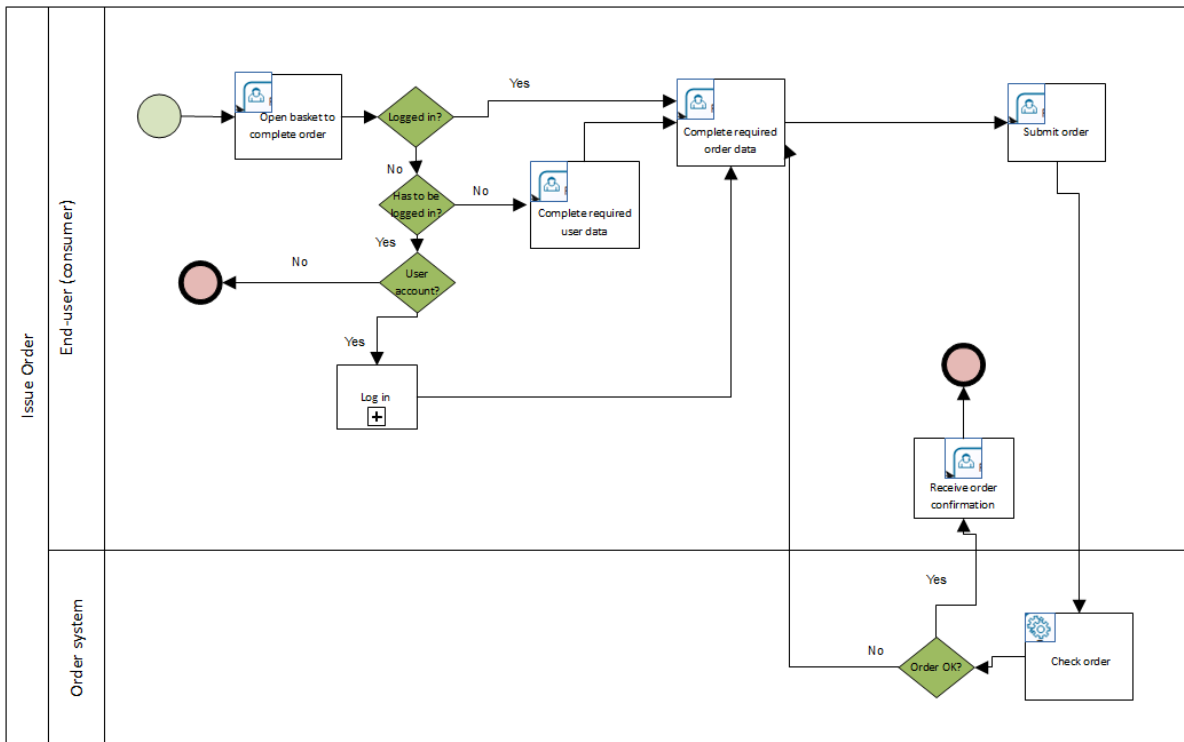















Figure 13 – Issue Order

The use case illustration is described in the table below.

Table 6 - Issue Order

Step	Description	Remarks
 Start		
 Open basket to complete order	The <u>end-user</u> can open the basket to review the list of ordered items and additional data needed for fulfilling the order.	The <u>Order basket</u> is “could”. It is “must” that order data is submitted to the system. If the Order basket is not implemented the <u>user</u> is required to make a new order for each individual item.
 Logged in?	The <u>end-user</u> may already be logged in.	

 Has to be logged in?	Requirement for log in and <u>user</u> account to make an order depends on the Archive's policy	
 <u>User</u> account?	If the <u>end-user</u> doesn't have a <u>user</u> account, (s)he is prompted to apply for a <u>user</u> account (with a link)	Applying for a <u>user</u> account and set-up of the account on the archive are not part of E-ARK.
 Log in	Whether it is necessary for a <u>end-user</u> to log in depends on the specific archival policy and log in tool	It is assumed that the specific archive has a solution for log in
 Complete required <u>user</u> data	If the <u>user</u> does not have an account, (s)he has to complete the order form with <u>user</u> data	
 Complete required order data	<i>DIP</i> data has been automatically transmitted from the <i>Finding Aid</i> , but specific data about the order has to be completed	The <u>user</u> may add: specific wishes about the format of the object; specific wishes for processing the object before delivery; delivery method: <ul style="list-style-type: none"> ◆ by e-mail ◆ personal pick-up on external media (optical) / postal service ◆ use in reading room ◆ online use contact details of the <u>user</u> ; specific wishes / requests / incl. authorization and restrictions management
 Submit order	Instead of an <u>Order</u> basket and a <u>user</u> account, there may be a digital order form which the <u>user</u> has to be complete with all data about the <u>user</u> , the <i>IP(s)</i> and specific requirements	
 Check order	The <u>end-user</u> has the chance to check the order before submitting it	
 Order OK?	Decision: if the order is OK the <u>end-user</u> can submit it. Otherwise amend the data in the order.	
 Receive order	The <u>end-user</u> should receive a confirmation that the order has been submitted	

confirmation	successfully.	
 End		

A specific sub use case: Use of Content Management Interoperability Services (CMIS)³¹

The digital archives are accessed by an external catalogue (as an example: agencies' records management system). The external system only knows about the IDs of the objects and has some metadata, but can call the repository using CMIS messages.

This scenario is especially helpful in situations where specific agencies want to have access to archived content integrated into their live systems. The workflow in this scenario is as follows:

- ◆ original content from the source system is prepared and transferred to the digital repository system
- ◆ the source system receives an ID of the content in the storage
- ◆ the source system keeps a minimal amount of metadata about the archived content, including the ID
- ◆ in case access is wanted the source system forwards the request for the *IP* by means of the ID
- ◆ the digital repository automatically fetches the *IP* and, if necessary and implemented, carries out the *AIP* to *DIP* conversion

As such the whole workflow can be seen as a subset of the usual *access* workflow with only automated tasks being included and the possibility to access only a single *IP* (and not parts of it or multiple ones) at once.

4.1.2.1.1 Acceptance criteria

Given that it is the end-user alone who has to complete the order form:

- ◆ When (s)he has located a desired *IP*, (s)he is offered a link to the order form in order to complete it with data about the *IP(s)* as well as user data.

Given that the Order basket is implemented:

- ◆ When the user has located all *IPs* (s)he wants to order, (s)he can complete the order and submit it, as in a web shop.

Given that the end-user has submitted an order:

- ◆ When the order is passed on for automatic or manual processing, the order data is in a predefined xml-format (order.xml).

³¹ http://en.wikipedia.org/wiki/Content_Management_Interoperability_Services

4.1.2.2 Use case 1.2.2: Validate Order

This use case starts when an order has been submitted from an end-user to the archive. The main purpose is to evaluate the user's access rights. It is difficult to outline a generic process and to automate it because national legislations and Archives' policies differ. It is assumed that this process is succeeded by the process "Prepare *DIP*", but it can just as well be the other way round; or the processes may occur simultaneously.

The main point is that the xml-file with order data is updated with the user's authorization to *access* the ordered *IP(s)*.

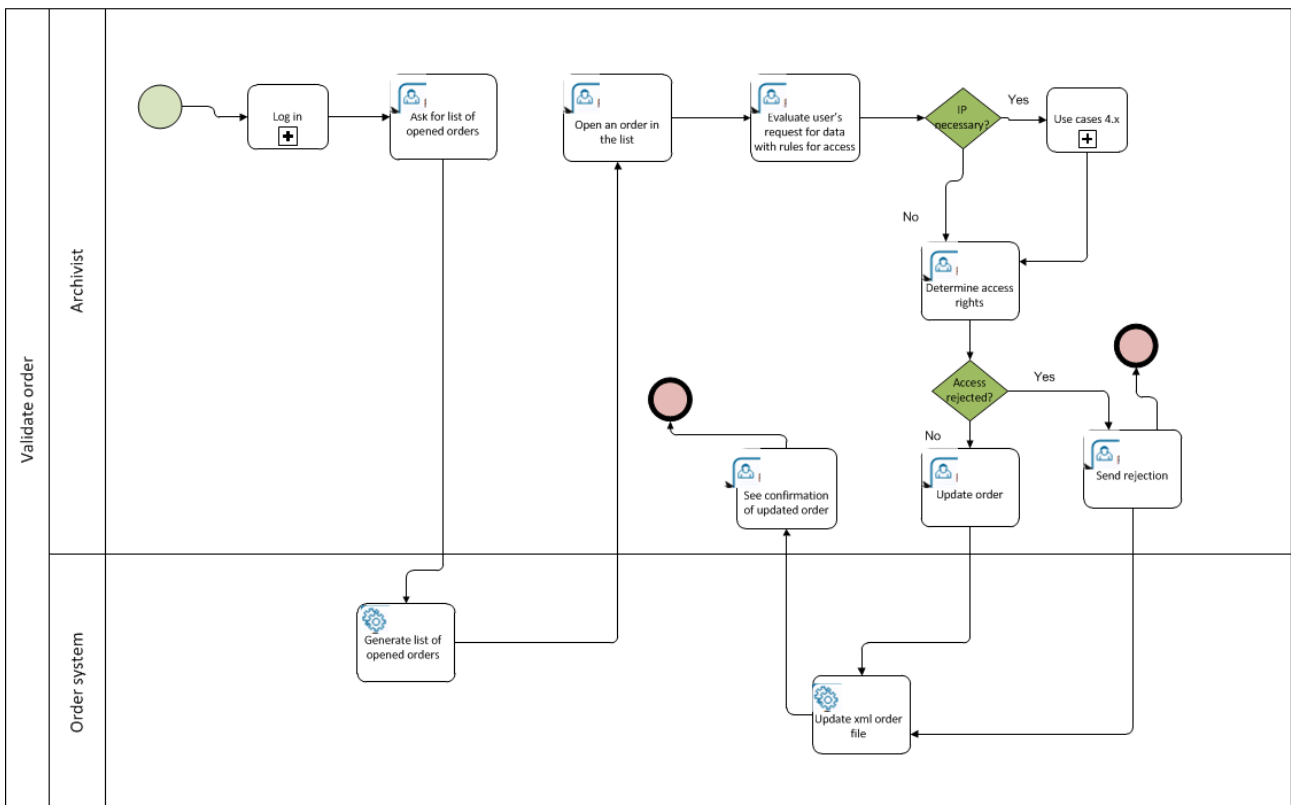

















Figure 14 - Validate Order

The use case illustration is described in the table below.

Table 7 - Validate Order

Step	Description	Remarks
 Start		
 Log in	The <u>archivist</u> logs into the order management system according to the specific <i>Archive</i> policies by using the local login tool	It is supposed that the specific <i>Archive</i> has a solution for log in

 Ask for list of opened orders	<p>The <u>archivist</u> can ask the system to show a list of open orders (as an example by clicking on the appropriate tab or command in the GUI)</p>	
 Generate list of opened orders	<p>For each order the list shows information about when it was submitted, who submitted it, and the deadline for processing</p>	<p>If there is a maximum level of automation in the order system, for each order is shown:</p> <ul style="list-style-type: none"> ◆ restricted material requested (a red flag shown when some of the IPs contain restricted content; yellow flag when the restriction is unknown, green flag when everything requested is open) ◆ DIPs available (a green flag shown when DIPs are already available for all or some of the IPs; yellow flag when DIPs can be generated automatically from AIPs; red flag when the DIP cannot be generated automatically or the status is unknown); ◆ modification requested (a green flag when none; red flag when yes).
 Open an order in the list	<p>The <u>archivist</u> can open a detailed view of an individual order out of the list.</p>	
 Evaluate <u>user's</u> request for data with rules for access	<p>The <u>archivist</u> can double-check the information in the order and, if necessary, initiate an evaluation based on the content and metadata on permitting access.</p>	<p>Access rules differ for <i>Archives</i> and for countries</p>
 IP check necessary?	<p>The <u>archivist</u> may need to take a look at the IP before Access rights can be determined</p>	
 Use cases 4	<p>If the <u>archivist</u> needs to have a look at data as they will be presented to the <u>user</u>, (s)he goes to use cases described in section 4.3.4, Use case 3.4: Access <i>DIP</i>, page 49.</p>	<p>Another possibility may be to go to the use case in section 4.1.1, Use case 1.1: Search & Select records: Search in descriptive metadata and data, page 28.</p>

 Determine access rights	An important issue is to determine if any modification is to be made due to access restrictions	
 Access rejected?	The <u>archivist</u> makes the decision about granting access or not	
 Send rejection	If the order is rejected, the order flow terminates here, and a rejection notice is sent to the <u>end-user</u> .	If the order is approved, it is not necessary to send a message to the <u>user</u> before the DIP is ready for delivery, unless the archive's policy demands that the <u>user</u> is notified
 Update order	Order is updated with information about whether it has been approved or rejected. It is also important that the order is updated with information about modifications needed to be carried out in order to ensure access.	
 Update xml order file		
 See confirmation of updated order		
 End		

4.1.2.2.1 Acceptance criteria

Given that one or more orders are submitted from end-users, and there is low level automation of the order management system:

- ◆ When the archivist starts the management system, (s)he is prompted to log in.
- ◆ When the archivist has logged in, (s)he gets a list of open orders with statuses and deadlines.
- ◆ When the archivist has evaluated the user's access rights, (s)he updates the order.xml file OR

Given that one or more orders are submitted from users, and there is maximum level automation of the order management system:

- ◆ When the archivist starts the management system, (s)he is prompted to log in.
- ◆ When the archivist has logged in, (s)he gets a list of open orders with statuses, deadlines and automatically generated flags for manual evaluation of access rights yes/no and for manual modification of content yes/no.

- ◆ When the archivist opens an order, (s)he gets all necessary information about the order, the user's previous authorisations for *access* to restricted *DIPs* and the possibility of viewing the content of the *DIP*.
- ◆ When the archivist has evaluated the user's *access* rights, (s)he updates the order.xml file

4.2 Process step 2: *DIP* Preparation

4.2.1 Process step 2.1 & 2.2: Checking the existence of appropriate *DIP*) & Confirmation of identified *DIP*

These two process steps have been merged into one use case: Check *DIP* availability and suitability:

4.2.2 Use case 2.2: Check *DIP* availability and suitability

This use case starts when an order has been submitted from an end-user to the archive, and the main purpose is to check if an appropriate *DIP* already exists. It is assumed that this process is preceded by the 'Validate Order' process (Figure 14 - Validate Order), but it can just as well be the other way round; or the processes may occur simultaneously.

The process may be automated to a certain degree. It depends on archival policies, local repositories, the level of access to these, the correctness and level of detail in the end-user's order, and on how complex the ordered *DIPs* are.

The main point is that the order.xml is updated with information about availability of a *DIP* (yes/no), identification of the *DIP*, and information about need for further modification. If an appropriate *DIP* exists, the process continues with the Deliver *DIP* use case seen in Figure 8 - [3.2] Deliver *DIP*. If a *DIP* exists, but needs further modification, the process continues with the use case described in section 4.2.5 Use case 2.5: Finalization of *DIP*, page 43. If there is no *DIP*, the process continues with the use case described in section 4.2.3 Use case 2.3: Creating *DIP* from *AIP*, page 42.

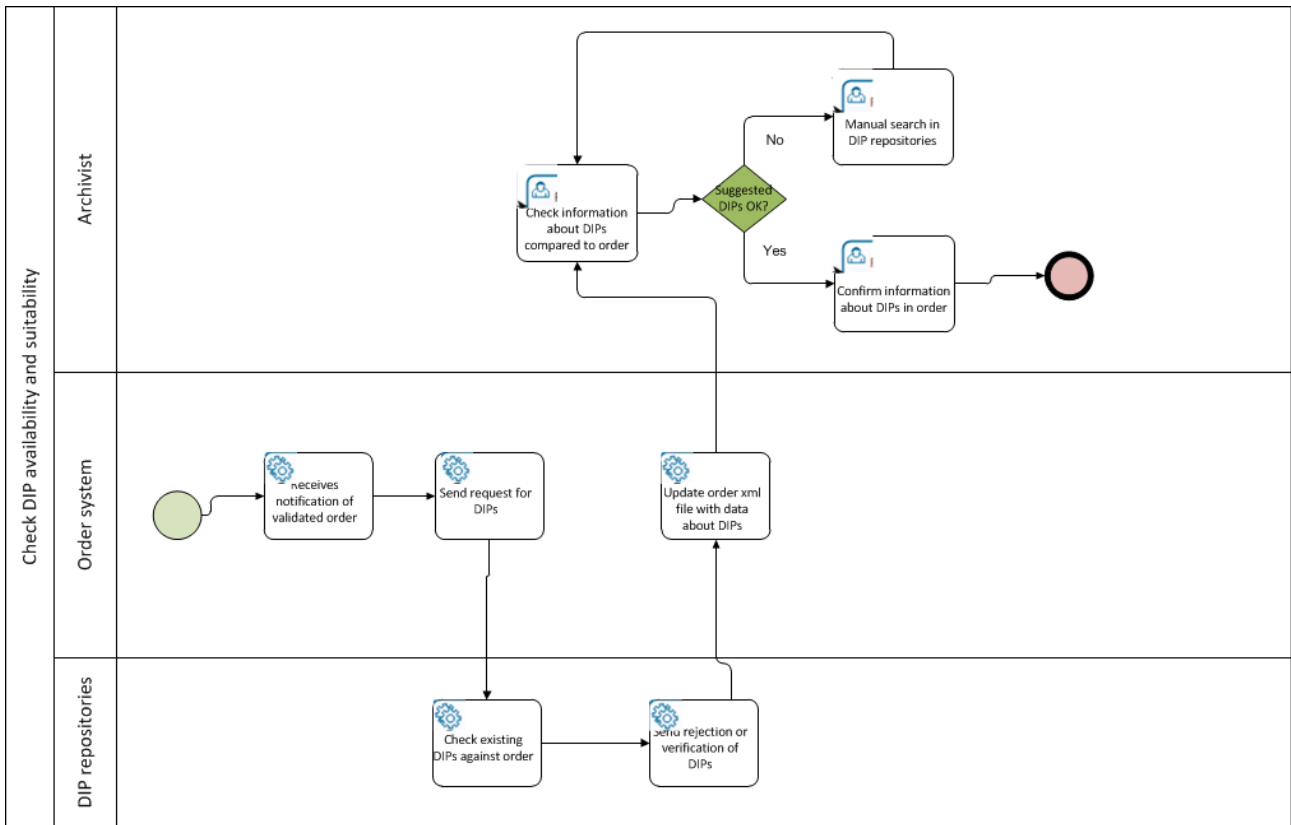









Figure 1 - Check DIP availability and suitability

The use case illustration is described in the table below.

Table 8 - Check DIP availability and suitability

Step	Description	Remarks
Start		
Receives notification of validated order	The <u>end-user</u> receives a notification of the validation by, e.g. by e-mail	Choice of automatic or manual process in an <i>Archive</i> depends on many factors, for instance frequency and complexity of orders, accessibility of <i>DIP</i> repositories, <i>Archive's</i> policies, cost/benefit evaluation.
Send request for <i>DIPs</i>		
Check existing <i>DIPs</i> against order	Precondition that the order is clearly specified, and that the <i>DIP</i> repositories have (or have access to) complete metadata	There might be: <ul style="list-style-type: none"> ◆ no <i>DIPs</i> available for a requested <i>IP</i>

		<ul style="list-style-type: none"> ◆ 1 DIP available for a requested IP ◆ multiple DIPs available for a requested IP
 Send rejection or verification of <i>DIPs</i>	If <i>DIPs</i> exist, appropriate metadata are returned, for instance, format, location, speed of delivery, whether modifications have been done to the <i>DIP</i> (compared to the <i>AIP</i>), whether it has been anonymised or not, etc.	
 Update <u>order.xml</u> file with data about <i>DIPs</i>	The information about existing <i>DIP(s)</i> is carried into the <u>order.xml</u> file.	
 Check information about <i>DIP</i> compared to order	The information about the existing <i>DIP</i> is compared against the wishes and authorisations expressed in the order.	
 Suggested <i>DIPs</i> OK?	The archivist can decide whether available <i>DIPs</i> are suitable for the user or not.	In some cases the search for <i>DIPs</i> might also return multiple results for a single <i>IP</i> (i.e. an anonymised <i>DIP</i> and a “full” <i>DIP</i>). In these situations the archivist shall decide manually which one to deliver to the user.
 Manual search in <i>DIP</i> repositories	Instead of or in addition to the automatic <i>DIP</i> availability check the <u>archivist</u> can also do a manual search in <i>DIP</i> repositories.	
 Confirm information about <i>DIPs</i> in order	The <u>archivist</u> can confirm or decline found <i>DIPs</i> for delivery to the user or for further modification.	
 End		

4.2.2.1.1 Acceptance criteria

Given that one or more orders are submitted from end-users and there is low level automation of the order management system:

- ◆ It is possible for the archivist to handle the process in this use case before or simultaneously with Use case 1.2.2: Validate Order.

- ◆ When the archivist asks to check *DIP* availability, (s)he gets a list of open orders with status and deadlines.
- ◆ When the archivist has opened an order, (s)he checks lists of *DIP* repositories.
- ◆ When the archivist has identified a *DIP* that may be appropriate, (s)he opens the *DIP* to check metadata and data together with the order.
- ◆ When the archivist has identified appropriate *DIP*(s), (s)he updates the order.xml file with data about the *DIP* and modifications to be done, if needed.
- ◆ When the order.xml file is updated, it is passed on to Use case 2.3: Creating *DIP* from *AIP*, the use cases attached to Process step 2.4: Check if modification is needed or those associated with Process step 3: *DIP* Delivery.

OR

Given that one or more orders are submitted from end-users, and there is maximum level automation of the order management system:

- ◆ The management system can handle the process in this use case before or simultaneously with Use case 1.2.2: Validate Order.
- ◆ When the management system has received an order, it asks *DIP* repositories for appropriate *DIP*s.
- ◆ When the management systems have received data about appropriate *DIP*s, the order.xml file is updated with identification of *DIP*s and modifications to be done, if needed.
- ◆ When the management system has received a negative answer, the order.xml file is updated.
- ◆ When the management system is not capable of evaluating if existing *DIP*s are appropriate, the archivist is prompted to make the evaluation.
- ◆ When the order.xml file is updated, it is passed on to Use case 2.3: Creating *DIP* from *AIP*, the use cases attached to Process step 2.4: Check if modification is needed or those associated with Process step 3: *DIP* Delivery.

4.2.3 Use case 2.3: Creating *DIP* from *AIP*

Normal behaviour: If *DIP*s are not available or not suitable the process of creating *DIP*s from *AIP*s can be initiated either:

- ◆ automatically by the order management system (given that no existing *DIP* is available and authorisations do not need to be checked)
- ◆ manually by the archivist following a suitability check and/or authorisation check.

The use case continues with one of the following two alternatives.

1. Technical alternative one: the order system passes the request “create *DIP* from *AIP* X in format Y” to the preservation system. The preservation system has an *access* workflow implemented which takes the original *AIP* X, normalises the content/object into format Y, packages it as an E-ARK *DIP* and returns it to the order system.

Technically the AIP2DIP Tool is used as a workflow step in the preservation system.

2. Technical alternative two: the order system asks the archival management system to “give me *AIP* X”. The preservation system returns the *AIP* and the order system (automatically) or the archivist (manually) reformats the content/object and packages it into the *DIP* format.

Technically the AIP2DIP Tool is used as a stand-alone tool or component of the order management system.

4.2.4 Process step 2.4: Check if modification is needed

This process step has been merged into the use case below.

4.2.5 Use case 2.5: Finalization of *DIP*

This use case is part of the order management process, and preconditions are that a suitable *DIP* exists (Process step 2.1 & 2.2: Checking the existence of appropriate *DIP*) & Confirmation of identified *DIP*) or has been created (Use case 2.3: Creating *DIP* from *AIP*), and that it has been recorded in the order.xml file that modification is needed, either to fulfil the end-user's specific wishes or to meet the conditions for the end-user's access authorisation. Result of the use case is a *DIP* ready to be delivered to the end-user.

4.2.5.1 Use case 2.5.1: Modify *DIP* for databases

This use case is complicated because content and structure has to be analysed in order to create search forms that can be used to retrieve the desired information in the database - see Use Case 3.4.1: Database & EDRM-system. If a tool for a scenario as described in Use case 3.4.4: Analyze with OLAP is developed, the archivist will probably use that for the required modification.

It is difficult to foresee the kind of modification needed. If the *DIP* is modified to fulfil end-users' wishes, it may be necessary to create one or more search forms, or to extract specific data according to the order. If the *DIP* is modified to meet the conditions of access, it may be necessary to reduce the dataset by removing database records, by removing tables, or by removing fields in tables. It may also be necessary to depersonalise records.

If the *DIP* consists of data from more than one *AIP*, it can be time consuming to merge the databases and the documents.

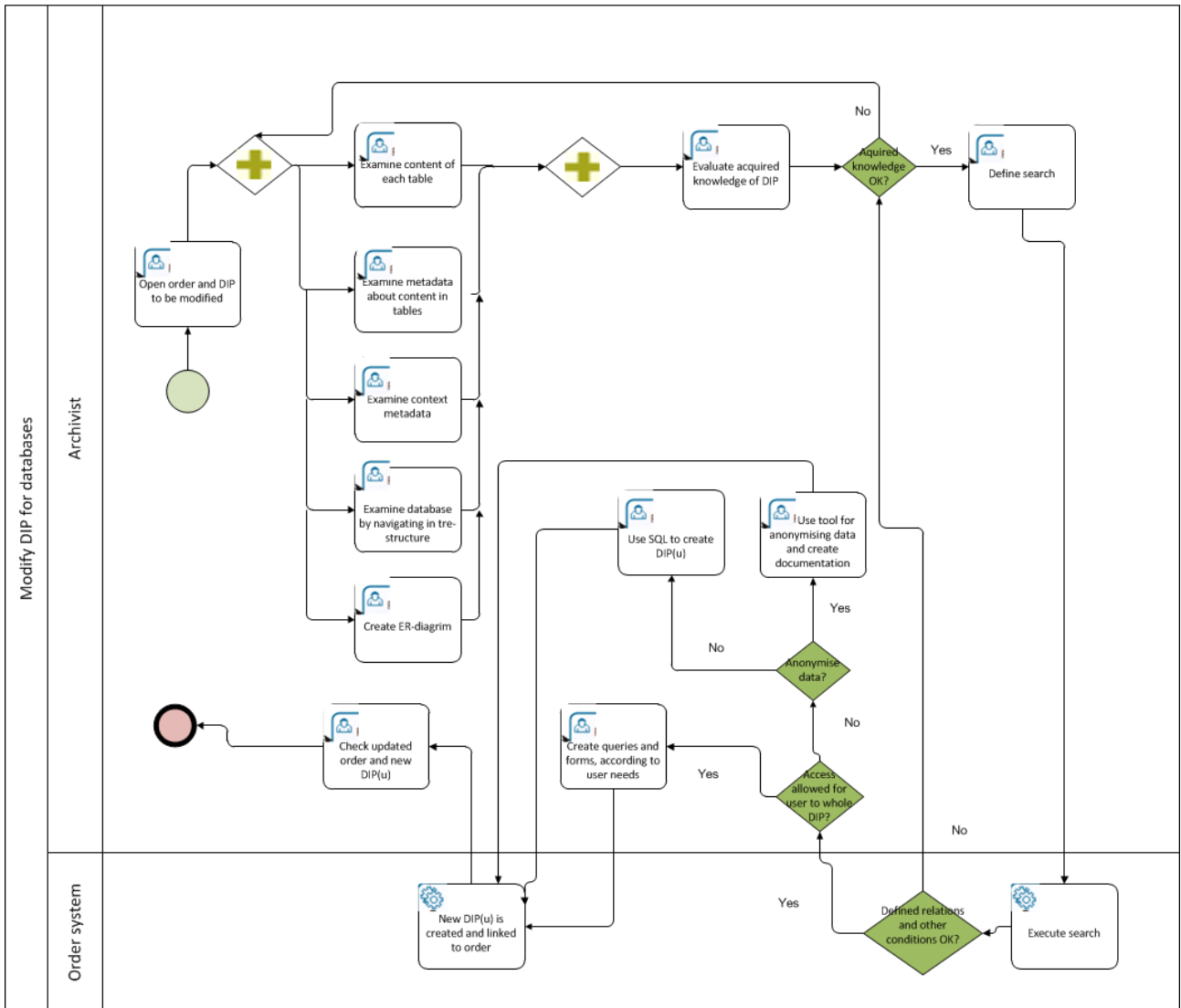
























Figure 15 - Modify DIP for databases

The use case illustration is described in the table below.

Table 9 - Modify DIP for databases

Step	Description	Remarks
 Start		
 Open order and DIP to be modified	The order is needed for wishes from the <u>user</u> and for any access restrictions	

	The <u>archivist</u> has to gain knowledge of the structure and content of the <u>database</u> , and that can be done in various ways	
 Examine data in each table		
 Examine metadata about content		
 Examine context metadata		
 Examine metadata by navigating in tree structure		
 Create ER-diagram		
		
 Evaluate acquired knowledge of <i>DIP</i>		
		
 Define search		
 Execute search		
 Defined relations and other conditions OK?	There can be problems with the SQL query which prevent it from being executed. The more informative the error message, the	

	easier for the <u>archivist</u>	
 Access allowed for <u>end-user</u> to whole <i>DIP</i> ?	The <u>archivist</u> checks order and whether access restrictions apply and hence if limiting the content of the <i>DIP</i> is necessary.	
 Create queries and forms according to <u>user</u> needs	There are no access restrictions, so the only modifications to be done are according to the <u>user</u> needs as specified by the order	
 Anonymise data?	The access restrictions can normally mean that the <u>end-user</u> is allowed to get only part of data, or that data about persons have to be anonymised	
 Use tool for anonymising data and create documentation		Documentation of anonymising may be important, especially when the <u>end-user</u> wants data for statistical research.
 Use SQL to extract data according to access restrictions	The <u>archivist</u> uses SQL queries to extract and crop data from the original <i>DIP</i>	
 New DIP_u is created and linked to order		
 Check updated order and new DIP_u		
 End		

4.2.5.2 Use case 2.5.2: Modify *DIP* for single records

The modification of *DIPs* with single records is in principle the same as for databases, but far less complicated. It is assumed that the modification is done with external tools like Acrobat Reader for pdf-files. Modification can consist in for example anonymization. For this reason, no description table of the below figure has been created.

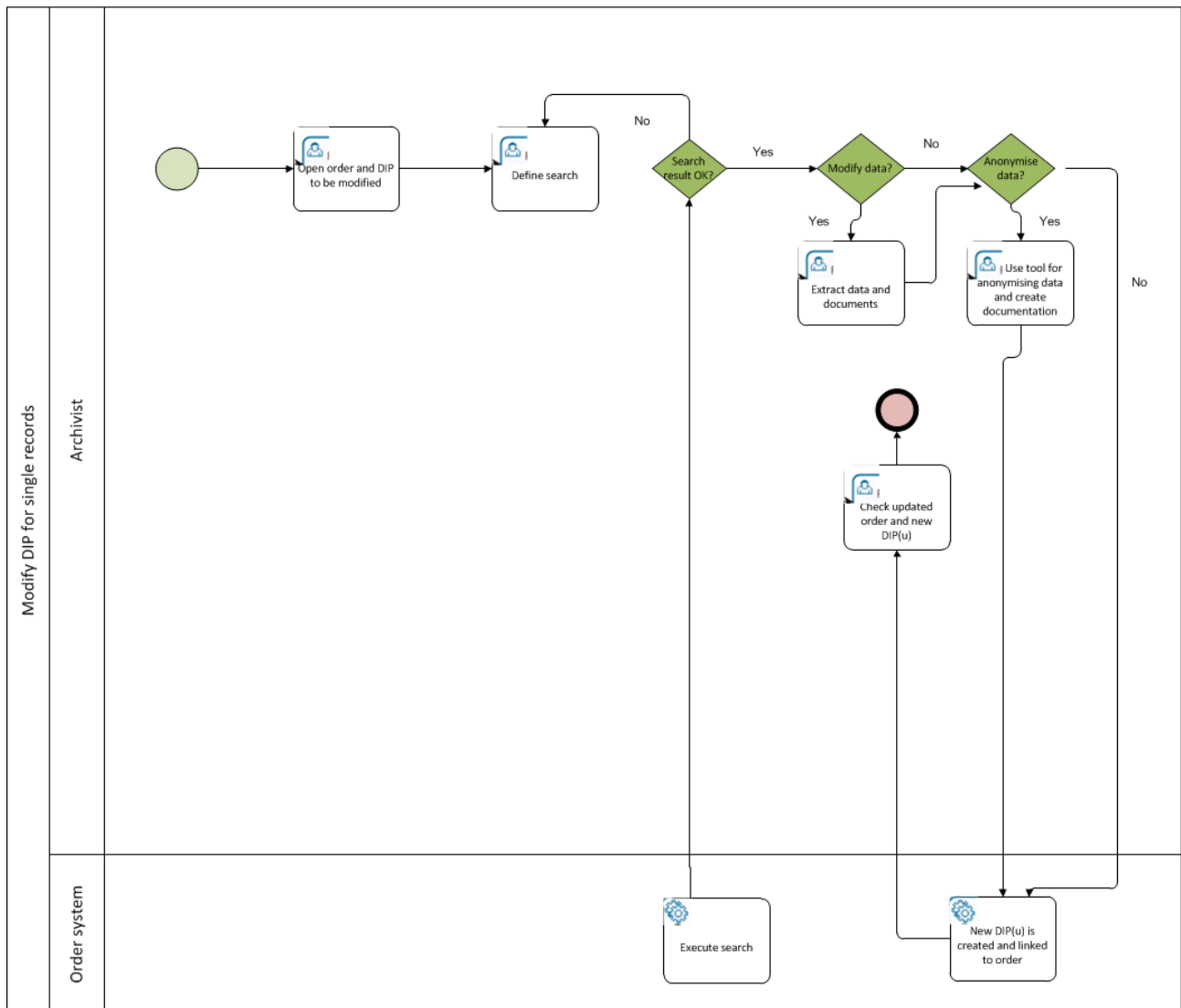


Figure 16 - Modify DIP for single records

4.2.5.3 Use case 2.5.3: DIP - GIS content

This use case must be seen in connection with the Use case 3.4.3: Search in GIS data for presentation of GIS content to the user.

Version 1: Removal of contents is needed:

- ◆ end-user orders certain layers and a DIP_0 is created from AIP
- ◆ part of the DIP_0 content is removed by the archivist; the event is logged in DIP PREMIS metadata
- ◆ the archivist confirms that DIP_u is ready

Version 2: Modification of GML file is requested:

- ◆ end-user orders only part of the layer, which contains e.g. data sets for an entire country
- ◆ the DIP_0 is created from AIP

- ◆ part of the DIP₀ content is extracted, modified by an IT/GIS specialist using a special tool, and inserted again into the *DIP*; all modifications should be back-traced (log file is created and added to the *DIP*; the fact that *DIP* content was changed is also recorded in PREMIS, title of changed computer files is also not the same as title in DIP₀)
- ◆ a DIP_u is created, the end-user will also get the reference for the (log) file, in which changes are logged; the user will thus be aware that the content has been modified
- ◆ the archivist confirms that the DIP_u is ready

4.2.5.4 Acceptance criteria

These acceptance criteria are common for the sub-use cases presented above (4.2.5.1, 4.2.5.2 and 4.2.5.3).

Given that the order order.xml file has been updated with information about *access* restrictions to be implemented and identification of the *DIP* to be modified:

- ◆ When the archivist opens the order, the *DIP* to be modified is also opened.
- ◆ When the *DIP* is opened, it is opened in a tool where the archivist can search and navigate all content.
- ◆ When the archivist is modifying a database, (s)he has a tool for creating SQLs and search forms (if required in the order).
- ◆ When the archivist has to anonymise data, (s)he has tools for retouching or substituting personalized IDs with new IDs.
- ◆ When the archivist has extracted data (database records, tables, fields in tables, or documents), (s)he can save the new DIP_u.
- ◆ When the new DIP_u is saved, it is automatically assigned an ID, the order.xml file is updated, (and the *DIP* repository is updated)

4.3 Process step 3: *DIP* Delivery

4.3.1 Use Case 3.1: Prepare *DIP* delivery

Based on the end-user request the *DIP* is:

- ◆ sent to the end-user by e-mail (or any other similar delivery method, for example, copying *DIP* on external media and posting it per regular mail), this finishes the step;
- ◆ copied onto external media which the user is supposed to pick up at the *Archives* reading rooms / information desk
- ◆ copied onto an FTP site which the user can access
- ◆ set up in an individual “working area” which is either available in the reading rooms or through the web. Note that this kind of working area can come in many different flavours, starting from simply putting the *DIP* into a folder on a local computer in the reading room to having highly customisable web environments with various virtualisation, emulation and support software options available.

4.3.2 Use Case 3.2: Provide *Access* rights

1. The “working area” is set up and a user name / password is created which allows the user to see the *DIP*s which (s)he has ordered.
2. The archivist creates an unique URL and user name/password for the FTP site to download the *DIP*(s).

4.3.3 Use Case 3.3: Notify end-user

The end-user is notified by e-mail (or any other means of communication) about the availability of the *DIP* with appropriate information about how to access / download the *DIP*.

4.3.4 Use case 3.4: Access *DIP*

Access *DIP* covers steps which are carried out using various file format / content type specific applications. These applications are subject to change over time as content data types evolve. The common denominator is that generic functionality provided by E-ARK is in place for extracting the file content and validating the whole *DIP* before using specific tools for specific content data types.

4.3.4.1 Use Case 3.4.1: Database & EDRM-system

The four sub use cases illustrate scenarios where an end-user wants to find information in a database. It also includes access to EDRM-systems which have been archived as a whole (i.e. technically as a database). The main difference between the “pure database” and the EDRM-system is that the latter usually includes much more binary content (i.e. unstructured documents and records). Therefore also the size of the archived EDRM-system is usually much larger and poses additional scalability issues.

The sub use cases are based on the experiences with the Sofia tool at the Danish National Archives (in use since 2008). It is assumed that the database is wrapped in the SIARD format and has been loaded into a standard RDBMS.

Focus is on search in data, but user friendly navigation is also important, and has to be taken into account when making requirements and designing the system.

Saving, exporting and printing functionality should also be available. However, if data are restricted there is a need for preventing access to these functionalities.

4.3.4.1.1 Use case 3.4.1.1: Search in database with “Google” functionality

This sub use case illustrates search in a database using one search field. It is something that E-ARK may develop, but it is prioritised as ‘could’. The problem with this approach to searching in databases is that the search results aren’t always pertinent. However performing the search itself is very user-friendly (writing a search term in a field as opposed to defining SQL queries).

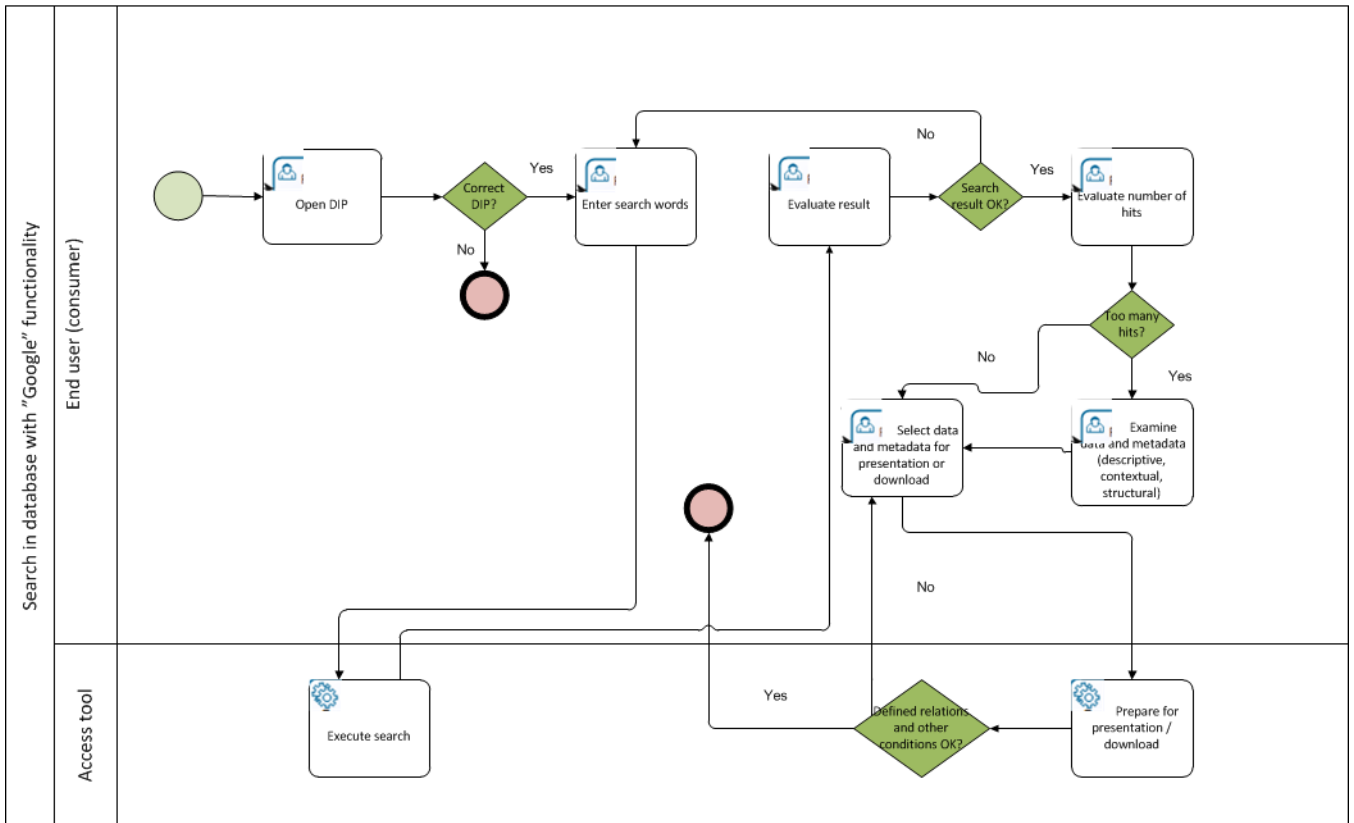












Figure 17 - Search in database with “Google” functionality

The use case illustration is described in the table below.

Table 10 - Search in database with “Google” functionality

Step	Description	Remarks
Start		
Open DIP	<u>End-user</u> opens <i>DIPs</i> and confirms that it is the right <i>DIP</i> and that it contains relevant data	
Correct DIP?		
Enter search words	<u>End-user</u> enters search words in one search box. Default is search in all content, but the <u>end-user</u> can choose to search in data, in documents, in <u>descriptive metadata</u> , in contextual metadata and / or in	Possibility for Advanced search with operators

	structural metadata	
 Execute search		
 Evaluate result	<u>End-user</u> evaluates whether the search has given desired results	
 Search result OK?		
 Evaluate number of hits	<u>End-user</u> evaluates whether there are too many hits. If too many, the <u>end-user</u> has to make a new search	The result must be displayed in a way that makes it easy for the <u>end-user</u> to go through the list and determine which results are relevant. Ranking of results according to different criteria.
 Too many hits?		
 Examine data and metadata	The <u>end-user</u> examines data, documents, contextual, descriptive and structural metadata to find relevant data, documents and metadata	<p>Experience shows that it can be very difficult to evaluate the content of a database because of potentially many tables and columns (at the Danish National Archives the average is 51 tables and 352 columns. There is a <i>DIP</i> with 417 tables and one with 104.638 columns).</p> <p>It can be a help to see:</p> <ul style="list-style-type: none"> ◆ the content of each table where the search word is found ◆ <u>descriptive metadata</u> about tables and fields (<description> elements in SIARD) ◆ contextual metadata ◆ structural metadata (relations between tables) <p>The challenge is to show all this in a way that is clear and understandable for a <u>user</u> without profound knowledge of relational <u>databases</u> and <i>DIP</i> structures.</p> <p>If there are hits referring to documents, the <u>user</u> shouldn't worry about which viewer to use. The relevant one should be chosen automatically by the system</p>
 Select	The <u>end-user</u> selects data and	

data and metadata	metadata of interest, and <ul style="list-style-type: none"> ◆ prints it ◆ saves it ◆ exports it in a format suitable for statistical tools or ◆ sends it in a mail 	
 Prepare for presentation / download		<p>It may be impossible for the <u>user</u> to save, print, or export due to <i>access</i> regulations</p> <p>The <u>user</u> might select fields that cannot be combined in an SQL-query due to lack of relations between tables. The <u>user</u> should receive feedback on this matter and guidance on how to solve the problem. A solution could also be that only the possible fields are shown, and the <u>user</u> is told what is missing and why.</p> <p>There could be a tool destined to alter the automatically generated SQL-query, as well as the relations defined in the structural metadata. Experience at the Danish National Archives shows that there are always inconsistencies no matter how thoroughly the <i>SIP</i> has been tested.</p>
 Defined relations OK?		
 End		

4.3.4.1.2 Use case 3.4.1.2: Search with existing forms

This sub use case illustrates the use of a predefined SQL search with a form that is understandable for the end-user. It limits the end-users' search options to the predefined SQL queries. This is especially necessary when the end-user isn't proficient in SQL and would have limited capabilities of creating such queries and forms him-/herself.

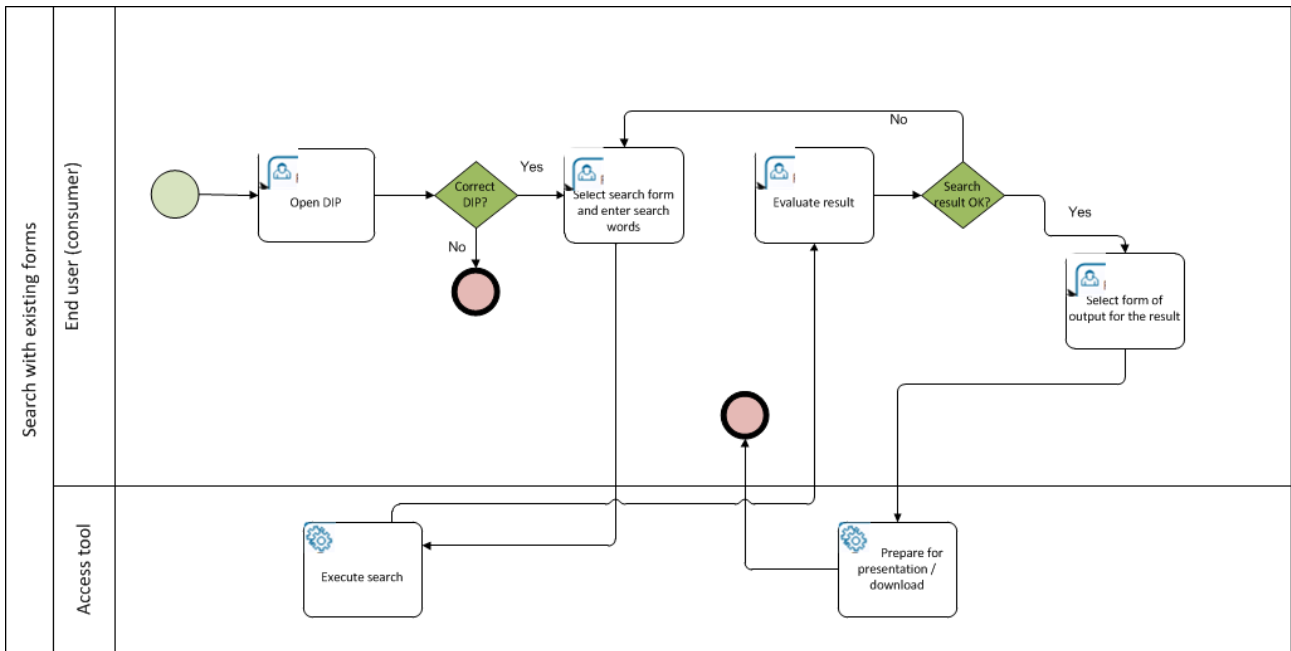








Figure 18 - Search with existing forms

The use case illustration is described in the table below.

Table 11 - Search with existing forms

Step	Description	Remarks
Start		
Open <i>DIP</i>	End-user opens <i>DIPs</i> and confirms that it is the right <i>DIP</i> and that it contains relevant data	
Correct <i>DIP</i> ?		
Select search form and enter search words	<p>End-user selects a search form and enters search words in one or more search fields where each field represents a column in the <u>database</u>.</p> <p>The <u>end-user</u> can only search in these columns and not in the rest of the content.</p>	<p>Possibility for Advanced search with operators.</p> <p>If SQL-queries for typical searches into the IT system are defined by the creator as part of the <i>SIP</i> and tested during delivery to make sure they can be executed, these could be transformed automatically into search forms.</p> <p>It has to be possible for the <u>end-user</u> to create other SQL queries.</p>

 Execute search		
 Evaluate result	<u>End-user</u> evaluates whether the result of the search is OK.	If the result includes documents, the <u>end-user</u> shall not worry about which viewer to use. The relevant one is chosen by the system.
 Search result OK?		
 Select form of output for the result	The <u>end-user</u> selects data and metadata of interest, and <ul style="list-style-type: none"> ◆ prints it ◆ saves it ◆ exports it in a format suitable for statistical tools or ◆ sends it in a mail 	
 Prepare for presentation / download		It may be impossible for the <u>end-user</u> to save, print, or export due to <i>access</i> regulations.
 End		

4.3.4.1.3 Use case 3.4.1.3: Search in database with SQL / DBMS functionality

This sub use case illustrates the end-user working directly on the database using SQL, maybe with a tool like MS Access. This requires knowledge about SIARD, relational databases and SQL that is not common for most end-users. Therefore this sub use case should be rather seen as an exception next to the previous one (Use case 3.4.1.2: Search with existing forms).

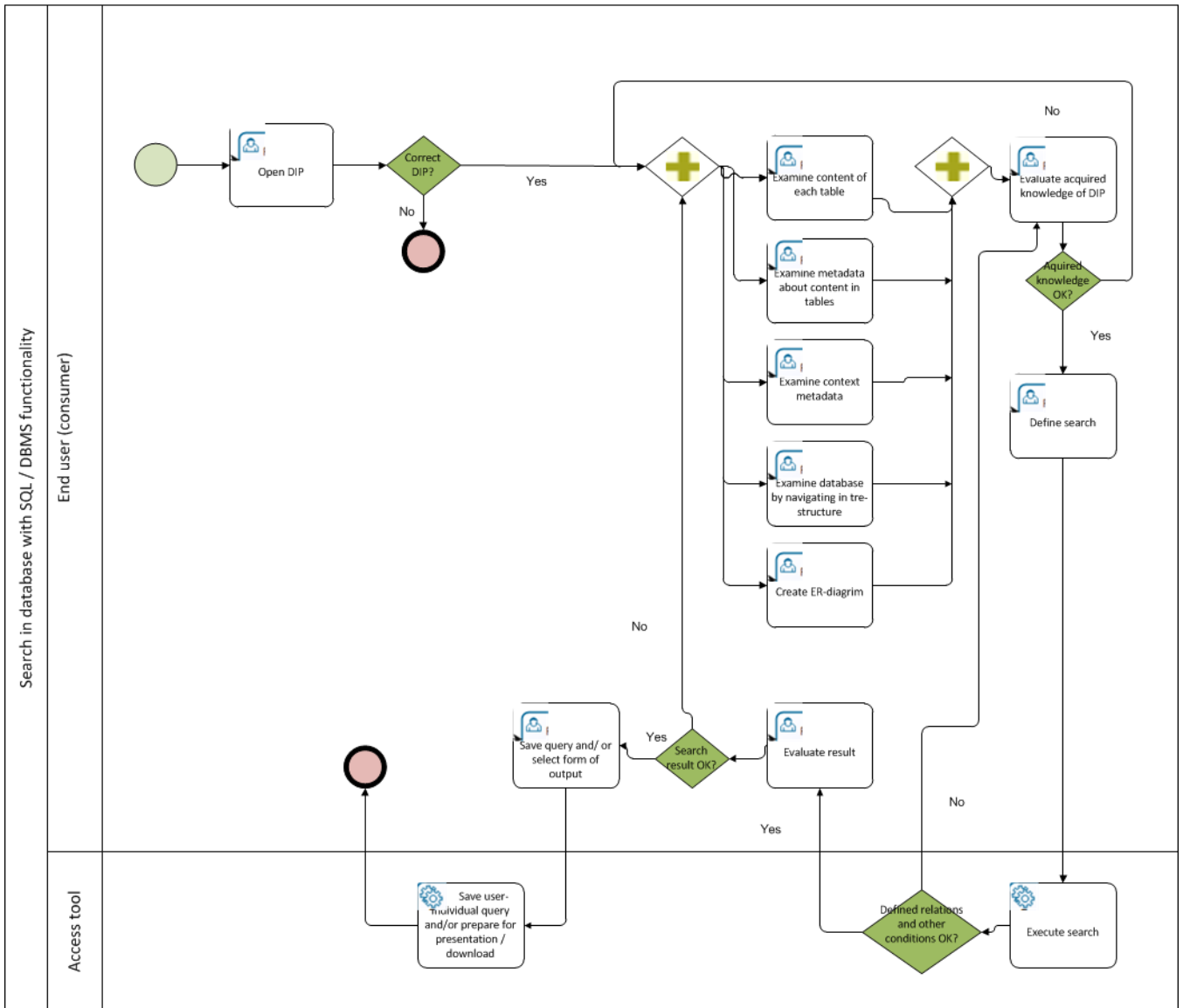











Figure 19 - Search in database with SQL / DBMS functionality









The use case illustration is described in the table below.

Table 12 - Search in database with SQL / DBMS functionality

Step	Description	Remarks
Start		
Open <i>DIP</i>	End-user opens <i>DIP</i> s and confirms that it is the right <i>DIP</i> and that it contains data	
Correct <i>DIP</i> ?		

	<p><u>End-user</u> has to examine data and metadata to identify which SQL query to make in order to search for the desired data.</p> <p>The exact way to do this will probably differ from <u>user</u> to <u>user</u></p>	<p>Examine possibilities of making SQL guidelines for the <u>end-user</u></p>
 Examine content of each table	<p><u>End-user</u> browses content to get an impression of data content in each table</p>	<p>Combining this with “Google” search as described in Use case 3.4.1.4: Search with a combination of google search and SQL / DBMS functionality will make it easier for the <u>end-user</u> to find the relevant tables</p>
 Examine metadata about content in tables	<p><u>End-user</u> browses metadata, shown alongside the data.</p> <p>Metadata can be descriptive (<description> elements in SIARD) and structural (relations between tables)</p>	
 Examine context metadata	<p><u>End-user</u> browses context metadata and other kinds of documentation, e.g. documents from the creator about which data were captured and recorded in the system.</p>	<p>This information can be helpful if browsing content and metadata about content does not provide the needed information to understand the content. It can also help that it is written in plain language</p>
 Examine <u>database</u> by navigating in tree structure	<p><u>End-user</u> navigates in a visualization of the <u>database</u> as a tree structure to explore and get an overview</p>	<p>Integration of this possibility with possibility to browse content and metadata about content</p>
 Create ER-diagram ³²	<p>To visualize the structure <u>end-user</u> creates an ER-diagram by dragging tables from the tree structure to a space</p>	<p>If there are many tables, it is difficult to get an overview, especially of the relations between tables</p>
		
 Evaluate acquired knowledge of <i>DIP</i>		
 Acquired knowledge OK?		

³² entity–relationship.

 Define search	<p>User defines a SQL-query by writing it directly or by using drag and drop from the tree structure to automatically generate a SQL-query (which then can be altered manually)</p>	
 Execute search		
 Defined relations and other conditions OK?	<p>The system carries out an automated check on the technical validity of the defined relations and conditions. If the result of the check is negative, the <u>end-user</u> probably has to go back to examine data and metadata. If the problem is due to a syntax error, a user-friendly error message will be of great help</p>	
 Evaluate result	<p><u>End-user</u> evaluates whether the result of the search is OK</p>	<p>If the result includes documents the <u>end-user</u> shall not worry about which viewer to use. The relevant one is chosen by the system</p>
 Search result OK?		
 Save query and/or select form of output	<p>The <u>end-user</u> selects data and metadata of interest, and</p> <ul style="list-style-type: none"> ◆ prints it ◆ saves it ◆ exports it in a format suitable for statistical tools or ◆ sends it in a mail 	
 Save query and/or prepare for presentation / download		<p>The <u>end-user</u> might get the message that (s)he can't save, print, export, due to <i>access</i> regulations</p>
 End		

4.3.4.1.4 Use case 3.4.1.4: Search with a combination of google search and SQL / DBMS functionality

This sub use case is an attempt to illustrate a combination of Use case 3.4.1.1: Search in database with “Google” functionality and Use case 3.4.1.3: Search in database with SQL / DBMS functionality. In the first one it is questioned whether results will be reliable because of incorrect (semantic) relations between tables and in the latter it is questioned whether the user has sufficient knowledge to create SQL-queries.

However, a combination does not seem to simplify the use case. The idea is that the end-user can choose to use a google search to limit data before defining a search (if it is necessary). The end-user can also use the google search for the result of a SQL search.

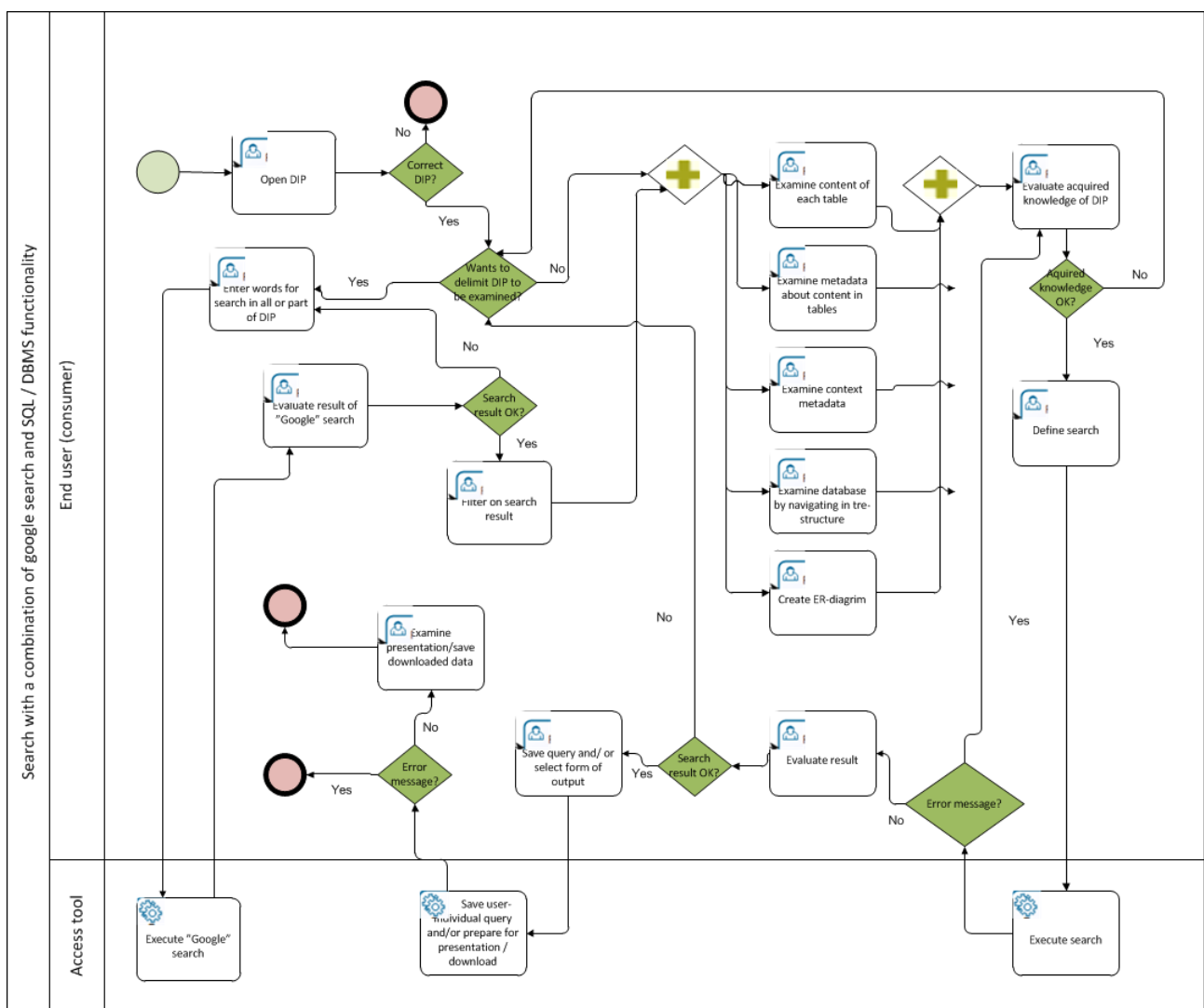


























Figure 20 - Search with a combination of google search and SQL / DBMS functionality

The use case illustration is described in the table below.

Table 13 - Search with a combination of google search and SQL / DBMS functionality

Step	Description	Remarks
 Start		
 Open DIP		
 Right DIP?		
 Wants to delimit DIP to be examined?	The google part of the tool may be used to locate where the requested data are and to delimit the DIP	
 Enter words for search in all or part of DIP		As many options as possible, for instance to search only metadata or to search with delimiters
 Execute "google" search		
 Evaluate result of "google" search		The result may be overwhelming, and the presentation of a search results is important
 Search result OK?		
 Filter on search result		
	The <u>end-user</u> has to acquaint herself with the structure and content of the database, and that can be done in various ways	
 Examine data in each table		

 Examine metadata about content		
 Examine context metadata		
 Examine metadata by navigating in tree structure		
 Create ER-diagram		
		
 Evaluate acquired knowledge of DIP		
 Acquired knowledge OK?		
 Define search		
 Execute search		
 Error message?	There can be problems with the SQL query which prevent it from being executed. The more informative the error message, the easier for the <u>archivist</u>	
 Evaluate result		
 Search result OK?		
 Save query and / or		

select form of output		
 Save <u>user</u> individual query and / or prepare for print / download		
 Error message?		
 Examine print or download		
 End		

4.3.4.1.4.1 Acceptance criteria

The following acceptance criteria cover the four sub use cases for databases described above.

Given that an end user wants to find specific information in a database, and given that the user has *access* to the *DIP*:

- ◆ When the end-user opens the *DIP*, information is displayed confirming that it is the right *DIP*.
- ◆ When the end-user wants to examine descriptive, contextual or structural metadata, (s)he has the option to view it.
- ◆ When the end-user has selected to examine metadata, (s)he can navigate through metadata, and they are presented in an understandable manner.
- ◆ When the end-user wants to examine documentation in documents (if they are included in the *DIP*), (s)he has the option to view it.
- ◆ When the end-user has retrieved a search result, (s)he can navigate it intuitively
- ◆ When the end-user has retrieved a search result including documents, (s)he can open a document without caring about viewers.
- ◆ When the end-user has opened a document, (s)he can view metadata concerning the document.
- ◆ When the end-user has retrieved a search result, (s)he has the option to limit refine this search by applying filters
- ◆ When the end-user has located relevant metadata, documents or extraction of data (from a database), (s)he has an option to mark-up interesting parts.
- ◆ When the end-user has marked up interesting parts, (s)he has options to select print, save or export
- ◆ When the end-user has chosen to print, save or export, (s)he may get the message that (s)he is not allowed to print, save or export with some supplementary explanation.

Given that an Archive is implementing the appropriate tool(s), the Archive can set-up / change default set-up for:

- ◆ (parts of) user interface, labels, menus, etc.

- ◆ viewers
- ◆ formats for export

4.3.4.2 Use case 3.4.2: Search in single records

This use case illustrates an end-user accessing a *DIP* with single records. If the *DIP* contains metadata on files / documents, the user can navigate a list and/or search. If there are no metadata (or they are embedded in the documents) the end-user can only search.

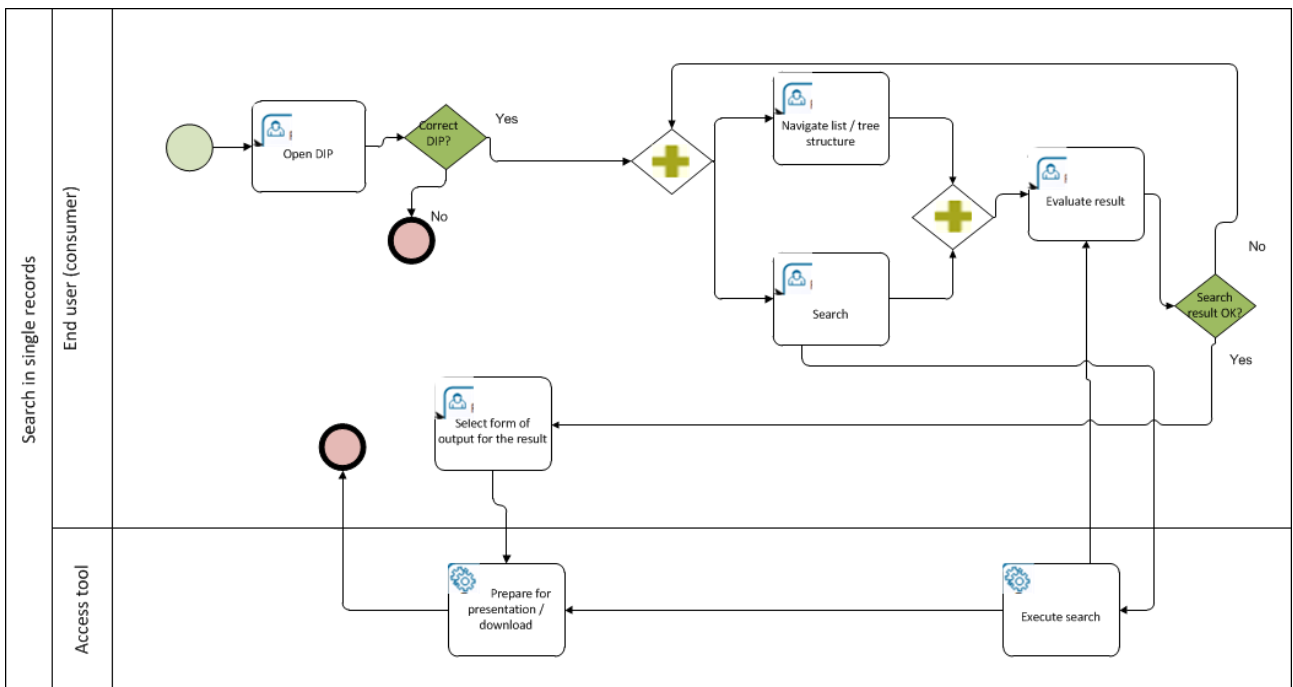















Figure 21 - Search in single records

The use case illustration is described in the table below.

Table 14 - Search in single records

Step	Description	Remarks
 Start		
 Open <i>DIP</i>	<u>End-user</u> opens <i>DIP</i> s and confirms that it is the right <i>DIP</i> and that it contains relevant data	
 Correct <i>DIP</i> ?		

		
 Navigate	<u>End-user</u> navigates a list or a tree structure (metadata)	
 Search	<u>End-user</u> searches in metadata or/and documents	
		
 Execute search		
 Evaluate result	<u>End-user</u> evaluates whether the result of the search is OK	<u>End-user</u> shall not worry about which viewer to use for documents. The relevant one is chosen by the system
 Search result OK?		
 Select form of output for the result	The <u>end-user</u> selects data and metadata of interest, and <ul style="list-style-type: none"> ◆ Opens it with the appropriate viewer ◆ prints it ◆ saves it ◆ exports it in a format suitable for statistical tools or ◆ sends it in a mail 	
 Prepare for presentation / download		It may be impossible for the <u>end-user</u> to save, print, or export due to <i>access</i> regulations
 End		

4.3.4.3 Use case 3.4.3: Search in GIS data

QGIS is chosen as the viewer for viewing and working with geo-data.

The use cases below present different types of use cases (1-4) depending on the chosen DIP₀ format.

It is assumed that a user ordering geo-data is familiar with working with it, and has experience using a GIS-software like QGIS (1-3). If this is not the case a user might be satisfied with only viewing the geo-data on map in a picture file (4).

4.3.4.3.1 View the GIS data

1. The *DIP*: In QGIS the end-user identifies and selects the appropriate GML-files in the *DIP*. The end-user is also responsible for finding the appropriate map (s)he needs, e.g. a map of Denmark.
2. A collection of relevant files, e.g. the GML files (incl. attributes), additional attribute files and a map: the user clicks on the files and views the result with an appropriate tool.
3. A shape-file with the geo-data shown on a map: the end-user views the geo-data in own GIS-system.
4. A picture file with the geo-data shown on a map: the end-user clicks on the picture file and views it on a document viewer.

4.3.4.3.2 Work with the geo-data

1. The *DIP*: In QGIS the end-user can work with the geo-data and make calculations.
2. A collection of relevant files, e.g. the GML files (incl. attributes), additional attribute files and a map. In QGIS the user can work with the geo-data and make calculations.
3. A shape-file with the geo-data shown on a map: the end-user works with the geo-data in own GIS-system.
4. A picture file with the geo-data shown on a map: no manipulation or calculation is possible.

4.3.4.3.3 Add information/attribute tables from *AIP*

1. The *DIP*: it is not possible to view the information/attributes not contained in the GML-file in QGIS. The end-user must open the presentation tool for databases and search for the additional data here, e.g. additional information about a place, building, etc., linked to the geo-data.
2. A collection of relevant files, e.g. the GML files (incl. attributes), additional attribute files and a map. The Archive can deliver the desired additional attributes from the *AIP* as a csv-file and the end-user can then import this attribute table into QGIS.
3. A shape-file with the geo-data shown on a map: the end-user receives a cvs-file with additional attribute data from the *AIP* and works with it in his/her own GIS-system.
4. A picture file with the geo-data shown on a map: The Archive producing the picture file with geo-data on a map (print map from QGIS) can add attribute information to the map. However only a little information can be shown on a map or else it will not be very readable.

4.3.4.3.4 Print the geo-data

1. The *DIP*: the end-user designs and prints a map from QGIS.
2. A collection of relevant files, e.g. the GML files (incl. attributes), additional attribute files and a map. The end-user designs and prints the map from QGIS.
3. A shape-file with the geo-data shown on a map: the end-user prints from own GIS-system.
4. A picture file with the geo-data shown on a map: the end-user can print from document viewer.

4.3.4.4 Use case 3.4.4: Analyze with OLAP

This use case is an attempt to illustrate analysing a set of archived of content with OLAP. It is supposed that the *DIP* is a database, and that the end-user will face the same challenges as in Use Case 3.4.1: Database &

EDRM-system. If the OLAP tool can manage that automatically, the first half of the use case can be removed.

The work on analyses with OLAP has not started yet and is part of the work of Work Package 6 on *Archival Storage, Services, and Integration*. Therefore, no work has been undertaken to describe this use-case in-depth in this deliverable. Thus, there is no table describing the illustration below.

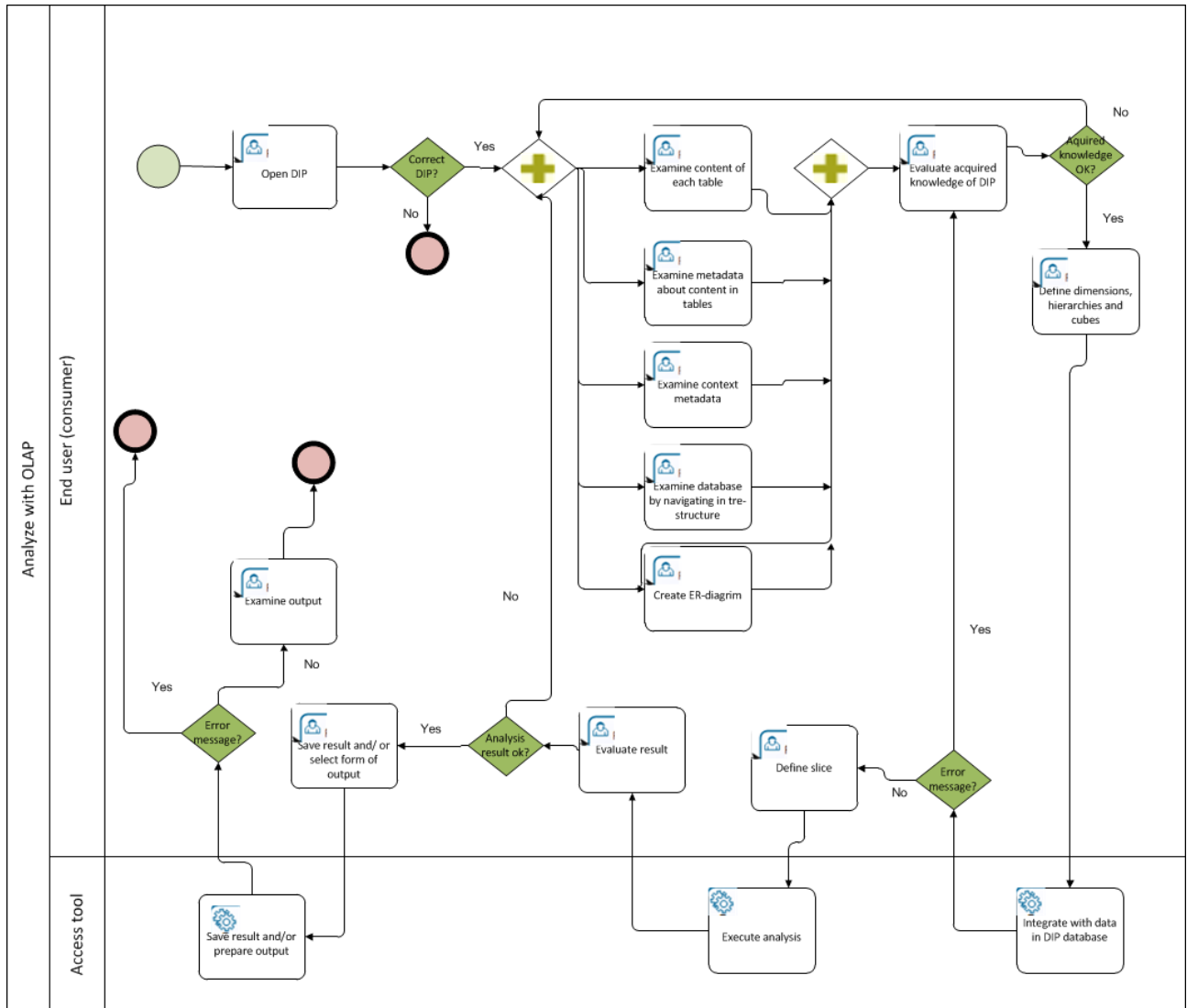


Figure 22 - Analyze with OLAP

4.4 Process step 4: *DIP* Management

4.4.1 Use case 4.1: *DIP* Management

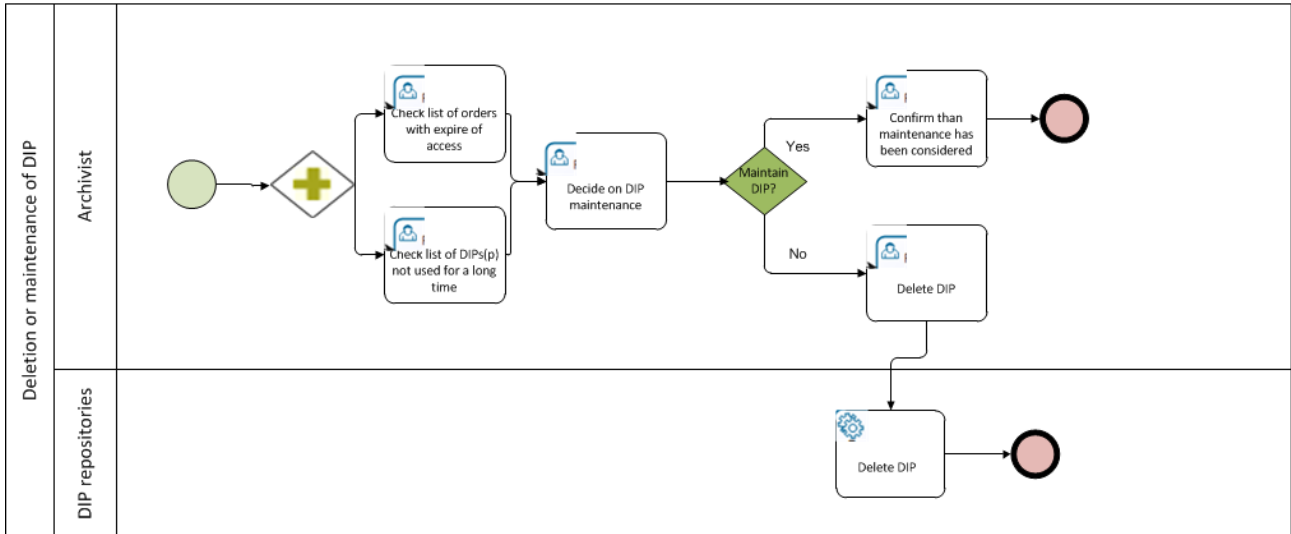


Figure 2 - Deletion or maintenance of *DIP*

The *DIP* Management is seen as low priority, and therefore no description has been made. If deemed necessary, an in-depth description of this use case will be provided in the E-ARK *DIP* pilot specification (D5.3).

5 Requirements for E-ARK DIP format

This chapter presents the requirements which will guide the development of the E-ARK DIP format. The requirements have been mainly derived from two sources:

- ◆ the E-ARK Common Specification which presents generic requirements for all E-ARK *IP* specifications
- ◆ the access use cases which present needs for more detailed and specifically access related requirements

The identified requirements will be transferred to Redmine³³.

5.1.1 Requirements derived from the E-ARK Common Specification

Table 15 - Requirements derived from the E-ARK Common Specification

Req no	Requirement description	MoSCoW
1	It must be possible to validate the structure and content of the <i>IP</i> : <ul style="list-style-type: none"> ◆ integrity, fixity (are all folders and files intact and at the right place) ◆ syntax (is the format of the information content correct) ◆ semantics (manual or semi-automatically) 	M
2	The <i>IP</i> format should allow efficient processing. <ul style="list-style-type: none"> ◆ huge information content must be able to be split and be spanned over many <i>IPs</i> ◆ The <i>IP</i> size is limited due to limitations on: <ol style="list-style-type: none"> a) capacity on removable media (also applies for archival storage on single media) b) efficient network transfer c) file system limitations, especially for number of files per folder 	S
3	The <i>IP</i> format should be expandable for the addition of new metadata standards or data formats (i.e. without forcing migration to a new <i>IP</i> format).	S
4	The <i>IP</i> format should be suitable for migrating the information content into new data formats	S
5	The <i>IP</i> format should include metadata about the formats of information content	S
6	The <i>IP</i> format must be as simple as possible	M
7	Data and metadata should be separated to ease the change of data formats or metadata formats	S
8	Metadata could be separated into several files to ease the change of metadata formats and to limit the time spent on handling metadata	C

5.1.2 Requirements derived from access use cases

5.1.2.1 Process step 1: Search and Select *IPs*

³³ Redmine is the tool which will be used to monitor requirements and the development of Access tools: <https://e-ark-redmine.magenta-aps.dk/>

Table 16 - Requirements for Search and Select IPs

Req no	Requirement description	Use Case	MoSCoW
9	The <i>DIP</i> format could allow for the inclusion of information about the search terms and process, which was used to find the object	UC1.1	C

5.1.2.2 Process step 2: Create Initial Order

Please note that this and the next process step do not pose requirements for the *DIP* format itself but to the “order” (*order.xml*) which is seen by E-ARK as a separate entity logging and describing the process of negotiating access and delivering the *DIP* to the user.

Table 17 - Requirements for Create Initial Order

Req no	Requirement description	Use Case	MoSCoW
10	The order must include a list of all <i>IPs</i> ordered by the <u>user</u>	UC2.1	M
11	The order should include information about the <u>user</u> placing the order	UC2.1	S
12	The order should include information about all known restrictions on the ordered data, including: <ul style="list-style-type: none"> ◆ Information about the sensitivity of the data (i.e. personal data, business oriented restrictions etc.) ◆ Information about copyright restrictions ◆ Information about specific reuse conditions (as an example if the use of the content is only allowed in specific physical locations for a limited time period) ◆ Information about any additional restrictions 	UC2.1	S
13	The order should include information about known needs for altering the content of the <i>IP</i> : <ul style="list-style-type: none"> ◆ For meeting special requests of the <u>user</u> ◆ For ensuring that restricted content is handled in an appropriate manner 	UC2.1	S
14	The order could include information about the usual time period during which the <i>DIP</i> needs to be delivered to the <u>user</u>	UC2.1	C
15	The order should be created automatically whenever possible by reusing information available inside <i>IPs</i> or <i>Finding Aids</i>	UC2.1	S
16	The order could include information about the contact details of the <u>user</u> and his/her preferred delivery or access methods (<i>must</i> if multiple access methods exist or when at least one <i>IP</i> contained in the order is potentially restricted)	UC2.1	C / M

5.1.2.3 Process step 3: Validate Order

Table 18 – Requirements for Validate Order

Req no	Requirement description	Use Case	MoSCoW
17	The order must allow for automated validation: <ul style="list-style-type: none"> ◆ Integrity (is all necessary information available) 	UC3.1	M

	<ul style="list-style-type: none"> ◆ Content validity (is all information valid when compared to other systems in the institution which is validating the order); ◆ Completeness and trust (is the information in the order complete and can be trusted, or should additional manual validation be carried out, for example in regard to potential restrictions to the content) 		
18	The order must support automated decision making (i.e. refusal of delivery, immediate delivery, need for manual check)	UC3.1	M
19	It should be possible to add / amend information about restrictions on the ordered items by the validating agency (especially in the case when descriptions of the restrictions are not available during the creation of the initial order)	UC3.1	S
20	Information about the <u>user</u> should be delivered in a way that the validating agency can uniquely identify the <u>user</u> and check for the order history and any available access clearances	UC3.1	S
21	Information about the <u>user</u> should allow for contacting the <u>user</u> in case additional information or discussions are necessary	UC3.1	S
22	All changes to the order carried out to during the validation could be logged in an understandable manner and included to the order	UC3.1	C

5.1.2.4 Process step 4: Prepare DIP

Table 19 - Requirements for Prepare DIP

Req no	Requirement description	Use Case	MoSCoW
23	The <i>DIP</i> must allow for the inclusion of any descriptive metadata from the <i>AIP</i>	UC4.2	M
24	The <i>DIP</i> must allow for the inclusion of any relevant descriptions of access conditions and restrictions	UC4.2	M
25	The <i>DIP</i> must allow for the inclusion of any relevant technical metadata about its content	UC4.2	M
26	The <i>DIP</i> must allow to use any relevant metadata standards within it	UC4.2	M
27	The <i>DIP</i> must include the date and time of the creation	UC4.2	M
28	The <i>DIP</i> must allow to include data in any type or format within it	UC4.2	M
29	The <i>DIP</i> must include information which allows its validation and authentication by the <u>user</u>	UC4.2	M
30	The <i>DIP</i> should include relevant information about the context and provenance of the package (i.e. the position in the archival hierarchy, reference to the creator and archives)	UC4.2	S
31	The <i>DIP</i> should allow for including / logging information about any changes done to the <i>IP</i> during ingest (<i>SIP</i>), preservation (<i>AIP</i>) or access preparation (<i>DIP</i>)	UC4.3	S
32	The <i>DIP</i> should include information about its current status in the <i>DIP</i> preparation workflow (as an example, whether the <i>DIP</i> is ready for delivery or still being modified)	UC4.3	S

5.1.2.5 Process step 5: Deliver DIP

Req no	Requirement description	Use Case	MoSCoW
33	It must be possible to deliver the <i>DIP</i> by any means supported by the archive and requested by the <u>end-user</u> , including (but not limited to): <ul style="list-style-type: none"> ◆ As a single file / package (for example for FTP download); ◆ Split into multiple parts (for example for restricted capacity media or e-mail delivery) ◆ As a set of individual files (for presenting the <i>DIP</i> in an online environment or dedicated research space in the reading room) 	UC5.1	M
34	Regardless of the means of delivery the <i>DIP</i> must ensure a means for checking the completeness of the <i>DIP</i> and all of its components	UC5.1	M
35	The <i>DIP</i> could include information which allows the <u>user</u> and the Archives to verify the delivery if necessary	UC5.3	C

5.1.2.6 Process step 6: Presentation and Search

Table 20 - Requirements for Presentation and Search

Req no	Requirement description	Use Case	MoSCoW
36	The <i>DIP</i> must allow for the validation of the package after delivery	6.1 – 6.3	M
37	The <i>DIP</i> as a whole must be human understandable in case no dedicated tools are available to process it	6.1 – 6.3	M
38	The <i>DIP</i> should include a simple overview of the contents of the package	6.1 – 6.3	S
39	It must be possible to process the content of the <i>DIP</i> effectively and automatically to allow for the creation of simple visualisation and search tools for the package	6.1 – 6.3	M
40	The different sections / components of the <i>DIP</i> must be easy to recognize (including descriptive metadata, technical metadata, information about restrictions and reuse, etc.)	6.1 – 6.3	M
41	The <i>DIP</i> should allow for the inclusion of representation information (i.e. information which describes how to use and represent the content) if necessary. This requirement is a <i>must</i> if the data in the <i>DIP</i> is in rare or complex formats for which the use can be assumed to be outside the usual capabilities and knowledge of the <u>user</u>	6.1 – 6.3	S

6 DIP format

The purpose of this chapter is to present the *DIP* format which is as standardized as possible, meets the requirements presented in the previous chapter and the additional technical, legal, user and other identified requirements identified previously in WP5³⁴.

6.1 *DIP* data model and folder structure

The first requirement for the E-ARK *DIP* format is that it must comply with the principles outlined in the E-ARK Common Specification⁹. The common *IP* data model which is derived from that specification is seen in the figure below. It consists of “Content” (“Data” and “Documentation”), and “Metadata”, which describe both the *DIP* package itself and the content as a whole. “Content” is not touched upon in this deliverable because it is being specified in Work Package 3, “Transfer of Records to Archives” (e.g. the SIARD-E format for databases and Moreq format for ERMS).

“Metadata” describes the package itself and the content as a whole. “Metadata” comprises both structured metadata, including, for example, administrative, preservation, and descriptive metadata in XML-format, and unstructured metadata, which could be scanned documents such as a user manual giving instructions about how the archived system was used when in production, classification schemas or filing plans, etc.

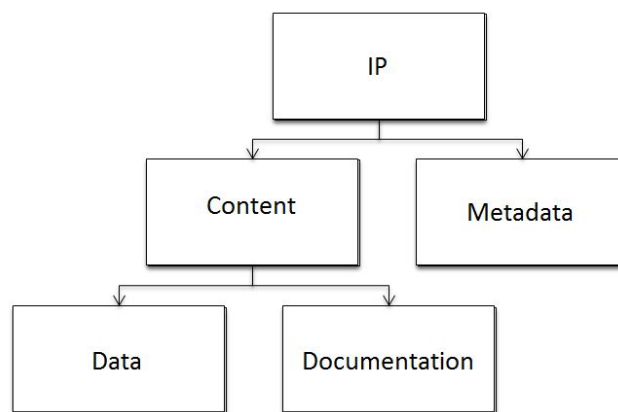


Figure 13 – Common data model for all IPs in EARK

E-ARK prescribes the use of a folder hierarchy within an *IP*. The *DIP* folder hierarchy is the same as for *SIP* (presented in D3.2) and *AIP* (presented in D4.2). The package is described by a METS file (i.e. metadata wrapper), which is to be found in the root folder. Additional package level metadata are placed in the folder “Metadata” in separate metadata files (EAD, PREMIS, etc.) that are referenced from the METS file. Any unstructured metadata (i.e. scanned documents as described above) are also placed in the “Metadata” folder as individual files. The *DIP* folder structure is illustrated below:

³⁴ Internal E-ARK deliverable: “Defining DIP requirements”.

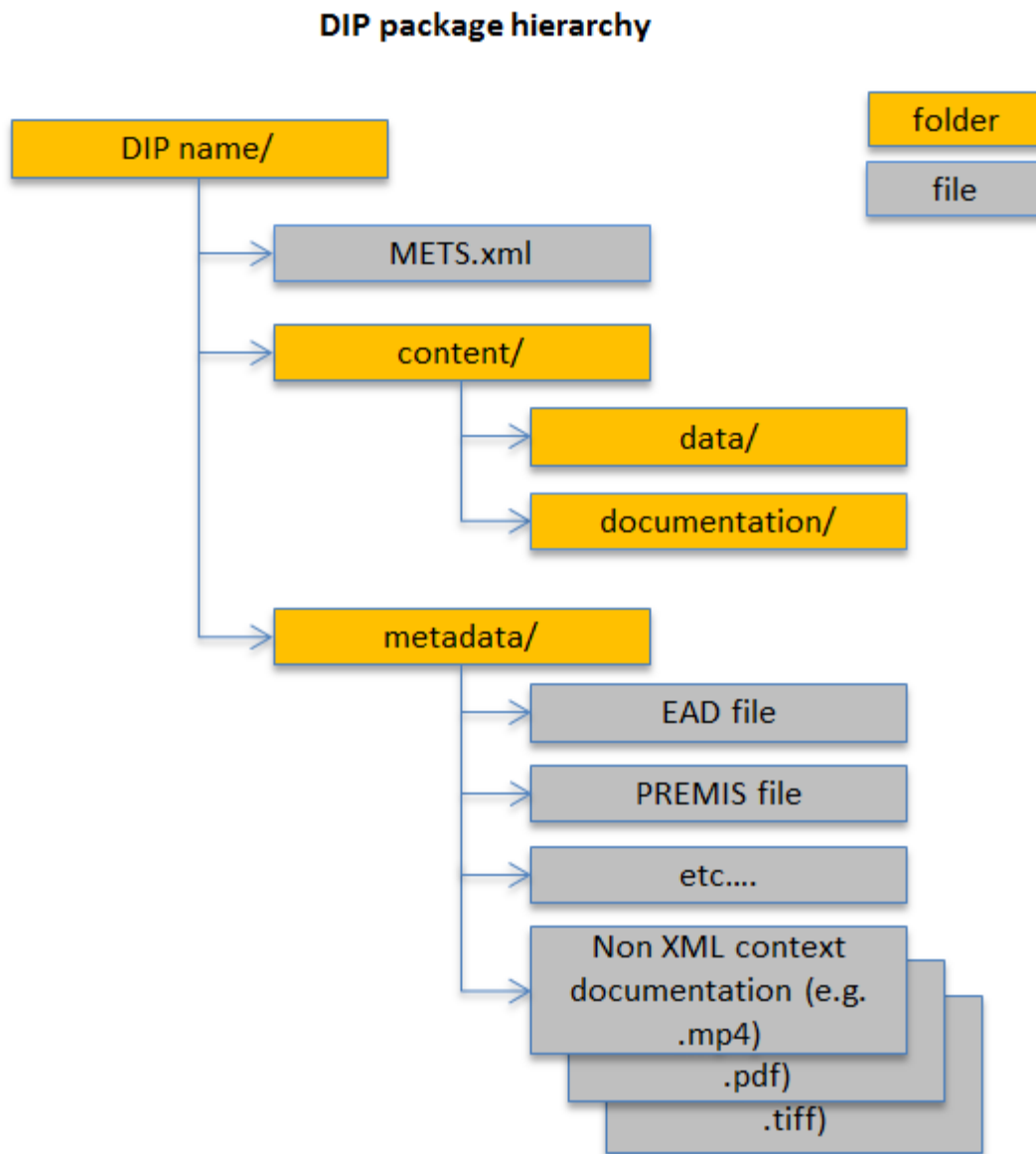


Figure 12 - DIP folder hierarchy

6.2 Metadata in the DIP

The *DIP* inherits metadata from the *AIP* and builds on existing metadata files. In addition the *DIP* must contain *DIP* specific metadata which could be, for example, information about how to display the *DIP* to the user, i.e. “Rendering information”. These metadata are to the extent possible embedded in the existing metadata files, such as METS and PREMIS. Where it is not possible, they are placed in a separate metadata file and referenced from the METS file. Any XSD-schemas will be placed along with the XML-files inside “Metadata”.

The METS file in the *DIP* will use the profile developed for the Common Specification, but will be a *DIP* specific version of the profile that defines what elements will be present in METS files for *DIPs*. This means that if the “IP type” is set to “*DIP*” in the METS file, the elements in the file will accordingly be the ones relevant for *access*.

6.3 *DIP* specific metadata

The *DIP* specific metadata that are included in the *DIP* are divided into the following categories:

- ◆ Rendering information
- ◆ Authenticity
- ◆ Access restrictions
- ◆ *DIP* status
- ◆ Dissemination notes

For each category a list of specific metadata elements that are needed in the *DIP* is identified (see tables below). This will serve as input for a revision of the common metadata profiles so that these will cover the needs for metadata in the *DIP*. The identification and definition of metadata needed in the *DIP* is work in progress and what is seen below is – like the deliverable – a draft.

6.3.1 Rendering information

Rendering information specifies how content in the *DIP* should be rendered (e.g. that geo-data is to be rendered by QGIS³⁵) and contains information about the technical environment and dependencies that should be considered when rendering the content. The specific pieces of information that should be part of the *DIP* metadata are specified in the table below.

While information on which content data type is to be placed in the METS file, detailed rendering information will be in PREMIS.

Table 21 - Rendering information metadata elements

Element name	Element description	Datatype	Occurrence	Mandatory	Notes
content data type	Description of the type of data of the content object. There are 4 possible values for this: <ul style="list-style-type: none"> ◆ single records; ◆ <u>database</u>; ◆ GIS; ◆ dataset [for the data mining <i>access</i> case] 	List	1	M	Details about the categorization and naming of <u>content data types</u> are still being discussed. This will be aligned with the rest of the project when final agreement is reached.

³⁵ <http://www.qgis.org/en/site/n>

size	Information about the size of the content object in bytes	Numeric	1	M	
format name	Information about the format(s) of the content objects	Text	0..1	O	
format version	Information about the version of the object format	Text	0..1	P	
Format ID	Machine readable format ID	Numeric	0..1	O	PRONOM identifiers to be considered
rendering software	Information about recommended software tool(s) to use and/or view the content object	List	0..n	O	
rendering software version	Information about minimum required version of the rendering software, if necessary	Text	0..1	O	
rendering software dependencies	Information about any dependent software required to render the object and details about e.g. code lists, xsd, non-embedded font definitions, etc.	Text	0..1	O	
rendering software notes	Information about other possible rendering software and/or how to obtain it	text	0..n	O	

6.3.2 Authenticity

The *DIP* will inherit authenticity metadata from the *AIP(s)* from which data in the *DIP* come. In addition to this metadata the *DIP* will include the following information to document and ensure authenticity. The specific pieces of information that should be part of the *DIP* metadata are specified in the table below. This supplementary metadata information will most likely be placed in the METS file and separate log file(s).

Table 22 - Authenticity metadata elements

Element name	Element description	Datatype	Occurrence	Mandatory / Optional (M/O)	Notes
<i>AIP</i> ID	ID of the <i>AIP(s)</i> from which data in the <i>DIP</i> comes	Text	1..n	M	
<i>AIP</i> ID_ <i>AIP</i> type	Refers to the content data type stored in the <i>AIP</i>	Text	1	M	This element and the next four sub-

					elements are to be filled separately for each <i>AIP</i> used in the <i>DIP</i>
<i>AIP ID_AIP</i> name	Information about the name of the <i>AIP</i> from which data in the <i>DIP</i> comes	Text	0..1	O	This information is related to a particular <i>AIP</i> . It can be repeated for each <i>AIP</i> from which data in the <i>DIP</i> comes
<i>AIP ID_AIP</i> version	Information about the version of the <i>AIP</i> from which data in the <i>DIP</i> comes. This would in most cases state that the latest version is used	Text	0..1	O	This information is related to a particular <i>AIP</i> . It can be repeated for each <i>AIP</i> from which data in the <i>DIP</i> comes
<i>AIP ID_AIP</i> creation date	Information about the date the <i>AIP</i> , from which data in the <i>DIP</i> come, was created.	Date	0..1	O	This information is related to a particular <i>AIP</i> . It can be repeated for each <i>AIP</i> from which data in the <i>DIP</i> comes
<i>AIP ID_AIP</i> log	Information that tracks the history of the <i>Information Package</i> .	Text	0..n	O	This information is related to a particular <i>IP</i> . It can be repeated for each <i>AIP</i> from which data in the <i>DIP</i> comes
<i>AIP-DIP</i> log	Log from the <i>AIP-DIP</i> conversion including information about any changes, amendments or choices made as part of the conversion	Reference	0..1	O	Reference to a file containing the full log.
<i>DIP</i> creation date	Information about the date the <i>DIP</i> was created	date	0..1	O	It is still to be decided if this information will be included in one of the metadata files or in the <i>AIP-DIP</i> log
<i>DIP</i> creator	Information about the organisation that created the <i>DIP</i>	Text	0..1	O	It is still to be decided if this information will be included in one of the metadata files or in the <i>AIP-DIP</i> log
<i>DIP</i> responsible	Information about the person/unit responsible for		0..1	O	It is still to be decided if this

	the <i>DIP</i> creation				information will be included in one of the metadata files or in the <i>AIP-DIP</i> log
<i>DIP</i> modifications	Details about any changes or amendments made to the <i>DIP</i> after its creation.	Reference	0..n	O	This information can be either a separate document or included as part of the <i>AIP-DIP</i> log. No final decision has been made yet.
<i>DIP</i> checksum			0..1	O	
<i>DIP</i> checksum type			0..1	O	

6.3.3 Access Restrictions

Access restrictions cover both restrictions on obtaining *access* to content and restrictions on how the content can be used and for what purposes. Simple high-level information about *access* restrictions present will be placed in the METS file. Detailed information about *access* right and restrictions on use will most likely be placed in PREMIS.

Table 23 – Access restrictions metadata elements

Element name	Element description	Datatype	Occurrence	Mandatory / Optional (M/O)	Notes
Access restrictions	Information about whether or not there are any restrictions on <i>access</i> to the content of the <i>DIP</i>	Yes/No	1	M	
Restricted unit	ID of the units to which the restrictions relate. It may be e.g. the whole IP, only one record, or a set of files in the IP.	Numeric	1..n	M	
Period of restriction	Information about the length of time a restriction applies. Examples of time periods are 70 years, or 55 years after the death of a person.	Text	0..1	O	
Restriction end date	The date on which a restriction ends (if the date is known)	Date	0..1	O	

Type of restriction	Information about the type of restriction. Possible choices include: <ul style="list-style-type: none"> ◆ Protection of personal data; ◆ Protection of classified data; ◆ Protection of trade secret; ◆ copyright; ◆ obligation of professional secrecy 	List	0..n	O	The list of restriction types has not been finalized. The list shown here is not exhaustive.
Legal basis	The legal basis for restricting <i>access</i> to the selected unit. Quoted should be an exact provision of a regulation or other legal act (e.g. Submission Agreement), which is the basis for restricting <i>access</i> .	Text	0..n	O	This information may be repeated for each type of restriction.
Scope of copyright	The scope of protection of copyright can only refer to the terms and conditions of reuse (eg. for each public presentation written permission must be obtained) or on the conditions of <i>access</i> (eg. Access is permitted only in the archives).	List	0..n	O	
Person authorizing <i>access</i>	Information about the person(s), unit(s) or organization(s) that have the right to allow <i>access</i> to restricted units. This may be designated by the deliverer - usually for records of private provenance; or an official body or a high-level employee of the archive - usually for records of public provenance.	Text	0..n	O	
Reproduction restrictions	Information about restrictions on reproduction of content in the <i>DIP</i> .	Yes/No	1	M	If reproduction restrictions are present the information in the next row should be

					mandatory
Reproduction restrictions-Citation obligation	Information about any citation obligations. Such demand is frequently found in copyrighted works, especially for the protection of moral rights.	Text	0..1	O (m)	This information should be mandatory if reproduction restrictions are present
Publication restrictions	Information about authorized or unauthorized forms of publishing the content of <i>DIP</i> .	Text	0..n	O (m)	This information should be mandatory if reproduction restrictions are present
Publication restrictions-Period of restriction	The date on which a reproduction restriction ends.	Date	0..1	O (m)	This information should be mandatory if reproduction restrictions are present
Publication restrictions-Person authorizing publishing	Information about the person(s), unit(s) or organization that have the right to allow publishing rights to restricted units. This may be designated by the deliverer - usually for records of private provenance; or an official body or a high-level employee of the archive - usually for records of public provenance.	Text	0..1	O (m)	This information should be mandatory if reproduction restrictions are present

6.3.4 *DIP* status

The *DIP* status indicates where in the dissemination process the *DIP* is. There are three possible statuses: One for when the *DIP* has been created (*DIP*₀); one when it has been prepared for the user (*DIP*_u); one for when it has been assigned to the *DIP* permanent storage (*DIP*_p). Information about the status and version of a *DIP* will most likely be placed in the METS file.

Table 24 – *DIP* status metadata elements

Element name	Element description	Datatype	Occurrence	Mandatory / Optional (M/O)	Notes
<i>DIP</i> <u>status</u>	Information about the <u>status</u> of a <i>DIP</i> . There are three	List	0..1	O	It has still to be decided what exact

	<p>possible <u>statuses</u>:</p> <ul style="list-style-type: none"> ◆ <u>DIP₀</u> which is the initial <i>DIP</i> created from one or more <i>AIPs</i>; ◆ <u>DIP_u</u> which is the <i>DIP</i> a <u>user</u> gets <i>access</i> to. This may identical to the <i>DIP₀</i> but may also be different; ◆ <u>DIP_p</u> which is a <i>DIP</i> that is considered to be permanent and thus is stored in a <i>DIP</i> storage area. 				terms will be used for the <u>statuses</u> .
--	--	--	--	--	--

6.3.5 Dissemination notes

In the *AIP-DIP* creation and dissemination process metadata may be generated and it may be necessary or beneficial to include part of this metadata in the *DIP*. This could be, for example, manually generated search forms or notes about the content. The specific pieces of information that should be part of the *DIP* metadata are specified in the table below.

Table 25 – Dissemination notes metadata elements

Element name	Element description	Datatype	Occurrence	Mandatory / Optional (M/O)	Notes
Notes_public	Free text string element. Placeholder for notes or comments about a <i>DIP</i> or its use that should be accessible for <u>users</u> .	Text	0..1	O	
Notes_internal	Free text string element. Placeholder for internal notes and comments that should only be accessible by personnel involved in the access process e.g. <u>archivists</u> who create/amend <i>DIPs</i> , data managers etc. Internal notes should not be in <u>DIP_p</u> s	Text	0..1	O	Presumably, notes can be restricted by <u>user-roles</u> and permissions.
Search form_ID	The ID of a search form created as part of the <i>Access</i> process / <i>DIP</i> generation	Text	0..n	O	The "search forms" referred to here are built on the Danish practice and the Sofia Access Tool. Here search forms are manually created queries that

					are intended to help <u>users</u> search data. They are generally created to reflect the searches that <u>users</u> are most like to perform. For example in a <u>database</u> containing information about fruits in Denmark search forms would be created on types of fruit (e.g. apples), fruits names (e.g. Ingrid Marie), Plant/species (e.g. Malus domestica), etc.
Search form_query	The actual query underlying the search form.	Text	1	M	This information is related to a particular search form. The information can be repeated for each search form.
Search form_title	Human readable title of a search form.	Text	1	M	This information is related to a particular search form. The information can be repeated for each search form.
Search form_description	Textual description of the query and the information it returns.	Text	1	M	This information is related to a particular search form. The information can be repeated for each search form.
Search form_notes	Free text element for any additional information about a search form. It can contain e.g. notes, search tips or suggestions for search terms that can be used in a given query.	Text	1	M	This information is related to a particular search form. The information can be repeated for each search form.
Search form_interface	Information about how to display the search form in the presentation tool.	Text	1	M	This information is related to a particular search

					form. The information can be repeated for each search form.
--	--	--	--	--	---

For each of the above categories specific metadata elements have been identified using the method briefly described in Chapter 2 Purpose and Method.

6.4 *DIP* specification for content data types and rendering scenarios

The data content type categories with which the E-ARK-project operates contribute to provide a clearer understanding about which tools and *IP* specifications are necessary to underpin them. A final decision on which data content types to use has still to be made in unison by E-ARK WP3, 4 and 5 and will be specified in other tasks and published in separate reports. For now the content data types in scope of E-ARK³⁶ are:

1. Single records, e.g. from ERMS (e.g. PDF, TIFF)
2. Databases (in SIARD-E format)
3. Geo-data (in GML format)
4. Datasets for data mining (in OLAP cubes)

After a *DIP* has been created and exported to the staging area, it needs to be rendered to a viewer. Detailed rendering scenarios will be defined and described for each content data type but this is not part of the *DIP* format specification and will not be part of this deliverable.

6.5 Access related metadata that will not be in the *DIP*

Not all *access* related metadata should be included in the *DIP*. The dissemination process will depend on and generate other metadata than are inside the *DIP*. This can for example be metadata that the archives use to administer the *DIPs* and the dissemination process, or information about who has accessed a *DIP* and when. This also applies for orders and *access* requests in the E-ARK order.xml format. Orders are essential in the dissemination process but should not be part of the *DIP*.

Though all of this this metadata does not belong inside the *DIP* archives may choose to keep it for example for statistical purposes. It is up to each local archive to make policies for what to do with these pieces of information addressing questions like e.g. should it be kept? For how long should it be kept? Where should it be stored?

6.5.1 *DIP* order

The order will be a separate XML file. It is not to be included in the *DIP*, but it is important for the development of the tool that processes the orders. The elements to be included in the order.xml are specified below:

³⁶ This is only a preliminary division of content data types that may be changed.

Table 26 – DIP order elements and their descriptions

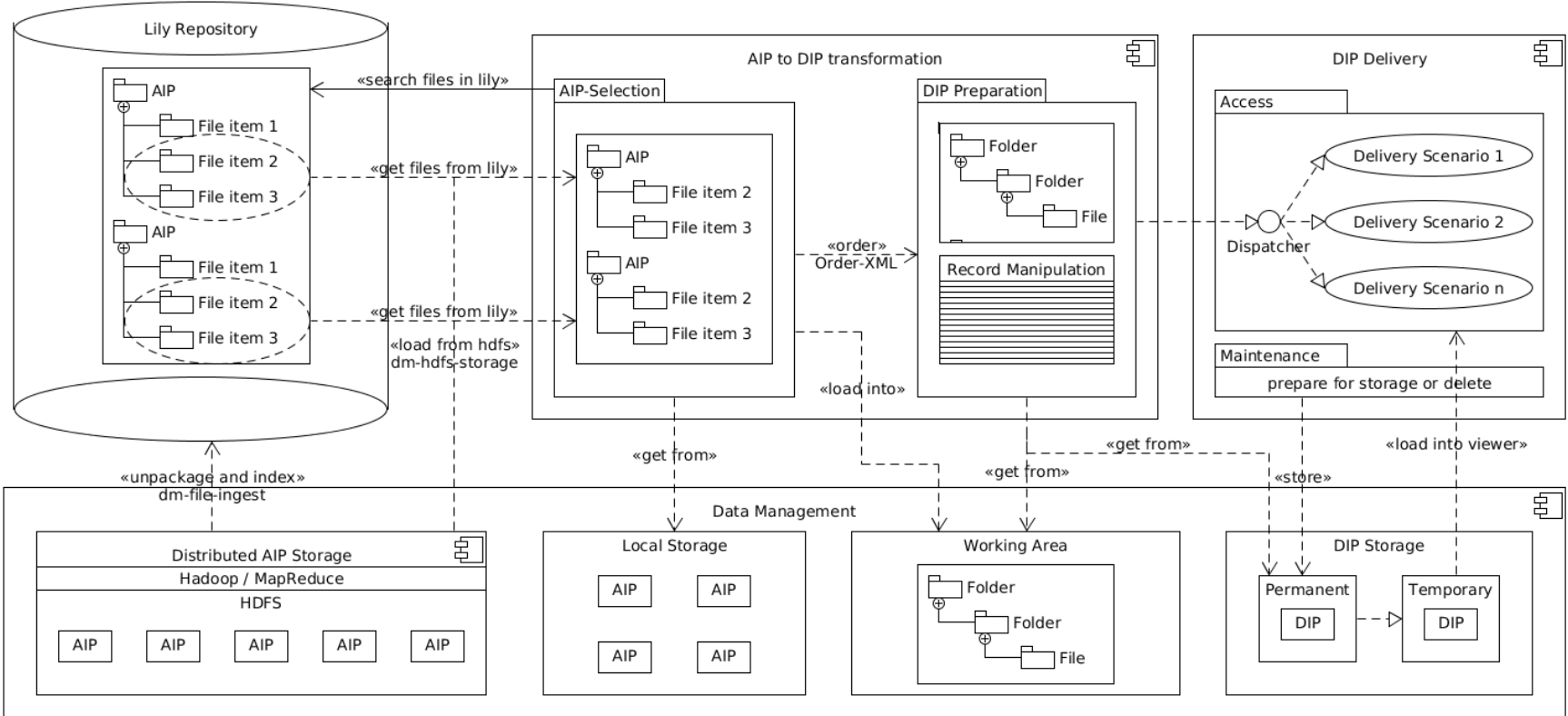
Element name	Element description	Datatype	Occurrence	Mandatory / Optional (M/O)	Notes
Order UniqueID	Each order has a unique ID.	Text	1	M	
Order Title	Title of ordered UD. Human friendly identification.	Text	1	M	
Ordered Item Reference Code	Reference code of ordered UD. One order can contain more units of description (items)	Text	1..n	M	
Level of description of ordered item	Level of description indicates, whether <u>user</u> is ordering entire fond, unit of description etc.	Text	0..n	O	
AIP URI	ID of ordered AIP or ordered computer file. One order can contain data from one or more AIPs or computer files.	Text	1..n	M	
Order Origin	Means of placing order (email, telephone, reading room etc.)	list	0..1	O	
Order Notes	Additional notes for the order added by end <u>user</u> .	Text	0..1	O	
Order Date	Date, on which end <u>user</u> issues the order. Can be added manually. Could be legal requirement.	Date	1	M	
Dossier of orders (of the <u>user</u>)	All orders from each <u>user</u> are grouped in a dossier.	Text	0..1	O	
Order Validation Date	Date, on which order was accepted and confirmed in the archives.	date	1	M	
Order Planned Date	Date, on which end <u>user</u> wants to access the records.	Date	0..1	O	
Access Date	Date, on which order is expected to be ready for the end <u>user</u> . Usually it is a date, but it could also be a text message.	Text	1	M	
End user UniqueID	<u>User's</u> ID number	Text	1	M	The end-user UniqueID gives access to personal <u>user</u> information needed for contact, e.g. the

					<u>user's</u> e-mail address
Responsible person UniqueID	The <u>archivist</u> who checked the order and finalised it.	Text	1	M	The responsible person's UniqueID gives access to personal <u>user</u> information needed, e.g. the <u>archivist's</u> e-mail address

7 High-Level component overview

While the use cases and requirements are more generic and take into account a wider range of needs the present chapter is about the specific implementation intended in E-ARK to demonstrate one way of implementing the use of the E-ARK DIP. The High-Level component overview is work in progress, and the planned iterations will most probably incur changes to the diagram shown below in Figure 23.

Figure 23 shows a diagram with the components used to prepare, create and present a *DIP*. The diagram shows three main components: The *Data Management* component for distributed and local storage, with a division into several storage areas, the *AIP to DIP Transformation* component to find and create an order for the required *AIPs* and to prepare the *DIP*, and the *DIP Delivery* component to give *access* to the *DIP* and maintain it after use.



7.1 Data Management component

Data Management is seen as the technical component for maintaining *IPs*. In E-ARK “Search and Select records” is not being considered as a part of *Data Management* but rather of the “*AIP* to *DIP* transformation”.

The *Data Management* component consists of a sub-component and several storage areas, with corresponding backend-modules and service interfaces for different types of storage.

On the one hand, there is the Distributed *AIP* Storage component and the dm-file-ingest service³⁷ (Hadoop-MapReduce-Job) which allows uploading *AIPs* into an HDFS (Hadoop 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, 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 is divided into a Temporary *DIP* Storage area, where files are being made available to the *consumer*, and the Permanent *DIP* Storage area, where files are stored permanently for re-use and future *access*.

7.2 *AIP* to *DIP* Transformation component

The *AIP* to *DIP* Transformation component allows selecting *AIPs* from various sources and preparing *DIPs* for different types of scenarios. It has two sub-modules, one for selecting the *AIPs* and another one for preparing the *DIP*.

7.2.1 *AIP* Selection module

From the perspective of the *AIP* to *DIP* transformation component it is assumed that the *AIPs* have been copied into HDFS and that the individual files contained in the *AIPs* have been indexed by the Lily repository. The *AIP* Selection module then provides a graphical user interface that allows searching for *AIPs* based on results related to individual file items contained in the packages.

The generic search interface translates a keyword search query, optionally with additional constraints, into an http-request to the Lily repository. In turn it receives a response with meta-information about files contained in a package or about the package itself.

Criteria for each item contained in the result list are:

- ◆ Identifier of the file or package
- ◆ Package reference identifier
- ◆ Content-type
- ◆ Size
- ◆ http-access-URL

³⁷ <https://github.com/eark-project/dm-file-ingest>

- ◆ Optionally: Degree of relevance (rational number between 0 and 1)

If a result item is a file, it must have a package reference identifier which indicates to which package the file belongs. If the item is a package itself, the package reference identifier and the identifier are equal. The 'Content-Type' property allows delegating the rendering to the appropriate viewer module. The 'Size' property allows optimization of the rendering, depending on the file size, e.g. using partial rendering for very large files. And the 'http-access-URL' allows the reading of the content of the file. Optionally, the list can have an internal relevance parameter to allow sorting of the result list.

Because in most cases not all *AIPs* are stored in the Hadoop-based Distributed Storage Layer (HDFS/HBase)³⁸ of the Lily repository, there is also the possibility of adding packages manually, or defining an additional search source that implements the search interface of the *AIP* Selection module (as an example the online *Finding Aid* of the archive).

For loading files into the *AIP* to *DIP* conversion component, there are two options. The first option is to retrieve the list of files contained within the *AIPs* and subsequently *access* these files directly through the Lily repository. The second option is to copy the corresponding *AIPs* from the Distributed Storage into a working area, then unpack them and use the files for further processing.

Based on a selection of files from one or several *AIPs*, the creation of a *DIP* is started. In **Error! Reference source not found.**²⁵ this is represented by the two dotted ellipses to illustrate the selection of file items 2 and 3 each from the first and the second *AIP* in the Lily Repository.

7.2.2 *DIP* Preparation module

The *DIP Preparation* module receives a result document (*order.xml*) from the *AIP* Selection Module (see «*order*» action in **Error! Reference source not found.**²⁵). Next, the existence of an appropriate *DIP* Package for the order is verified. The *user* preparing the *DIP* can decide to load the existing *DIP* into the working directory for doing additional manipulations or to directly forward the existing *DIP* to the *DIP Delivery component* and its *Access* module.

The result document contains identifiers referencing selected *AIPs* which are transferred to the *Working Area* so that file manipulations can operate on a copy of the selected *AIPs*. This allows, for example, removing parts of the data which the *user* requesting the information (*OAIS-consumer*) is not allowed to see or enriching the *Information Package* with additional information.

The *archivist* has various options to do the transformation:

1. Manual *DIP* creation
2. Workflow-based *DIP* creation
 - a. Selecting *DIP*-transformation workflows explicitly
 - b. Rule-based workflow transformation using content-type rules

For the rule-based workflow transformation it is necessary to define the rules for workflow selection that relate to the distribution of content-types in the selected *AIPs*.

³⁸ Part of D6.2 "Integrated Platform Reference Implementation"

The rules can refer to the content-type property of the items in a package, or to using content-type based inspection to decide upon the behaviour by looking for properties in files. An example is XML files which can be detected as SIARD files in order to apply the “Database-DIP-creation workflow” automatically for this type of package.

7.3 DIP Delivery

The *DIP Delivery* component comprises the *Access* and *Maintenance* modules.

7.3.1 Access module

The *Access* module provides a *Dispatcher* service which allows forwarding of a *DIP* to the appropriate *DIP* viewer module. The Dispatcher uses a similar rule-based approach to re-direct the request for *DIP* rendering to a *DIP* viewer or rendering environment.

7.3.2 Maintenance module

The maintenance module allows for defining additional steps after the *access* period of a *DIP* has expired. In particular, it provides for decisions about whether *DIPs* are moved to the permanent storage area or deleted after use.

8 Glossary

Table 27 - Glossary

Access Functional Entity	The OAIS functional entity that contains the services and functions which make the archival information holdings and related services visible to <i>Consumers</i> .
Access restrictions management	Procedures conceived to protect confidential Archival records.
Access Services or Solutions (ASS)	Services and solutions developed in order to underpin the processes that give <i>Consumers</i> access to Archival records.
Access Software (AS)	A type of software that presents part of or all of the information content of an <i>Information object</i> in forms understandable to humans or systems.
AIP	OAIS: An <i>Archival Information Package</i> , consisting of the <i>Content Information</i> and the associated <i>Preservation Description Information (PDI)</i> , which is preserved within an OAIS.
Archival Information System (AIS)	Archival Information System: It consists of several modules that support individual phases of processing <i>Information Packages</i> or parts of them in the archival environment (e.g. submission module, cataloguing module, ordering module, digital storage module, access module). The number of modules, their features and relationships depend on the local implementation.
Archival Catalogue (AC)	See <i>Finding Aid</i> .
Archival records	A document whose long term value justifies its permanent retention.
Archivist	The <u>archivist</u> designates an internal <u>user</u> who administrates archival holdings. See also <u>user</u> and <u>end-ser</u> .
Common Specification	The common <i>IP</i> specification for E-ARK <i>IPs</i> .
Consumer	The role played by those persons or client systems, which interact with OAIS services to find preserved information of interest and to <i>access</i> that information in detail. This can include other OAISs, as well as internal OAIS persons or systems.
Content Data Type	This refers to different content data types, e.g. geo-data, spreadsheets, <u>databases</u> , image files.
Database (DB)	<u>Database</u> .
Descriptive metadata	Metadata that describes the data content.
DIP	<i>Dissemination Information Package</i> : An <i>Information Package</i> , derived from one or more <i>AIPs</i> , received by the <i>Consumer</i> in response to a request to the OAIS.
DIP₀	A provisional <i>Dissemination Information Package</i> directly derived from one or more <i>AIPs</i> , which may or may not be ready for use, according to the <u>user's</u> order and <i>access</i> rights.
DIP_p	A permanent <i>Dissemination Information Package</i> , available to be accessed indefinitely by <u>users</u> due to frequent requests for the same data. The <i>DIP_p</i> can be available on-line.
DIP_u	A <i>Dissemination Information Package</i> , ready to be accessed, and previously checked against <u>user's</u> order and <i>access</i> rights.
DIP Delivery	Represents the third process step of the E-ARK High-Level process described in Chapter 3

	where the <i>DIP</i> is delivered to the <i>consumer</i> via a suitable Graphical <u>User</u> Interface (GUI)
DIP Management	Represents the fourth process step of the E-ARK High-Level process described in Chapter 3 where the <i>DIP</i> is either deleted or sent to a permanent or temporary <i>DIP</i> storage.
DIP Preparation	Represents the second process step of the E-ARK High-Level process described in Chapter 3 where the <i>DIP</i> is prepared for the <i>consumer</i> , for example by transforming an <i>AIP</i> into a <i>DIP</i> .
Dissemination Information Package (DIP)	<i>Dissemination Information Package: an Information Package</i> , derived from one or more <i>AIPs</i> , and sent by Archives to the <i>Consumer</i> in response to a request to the OAIS.
EDRMS or ERMS	Electronic (Documents) and 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 <u>end-user</u> designates an external <u>user</u> who seeks <i>Content Information</i> in archival holdings. See also <u>Archivist</u> and <u>User</u> .
Finding Aid	A type of <i>Access Aid</i> that allows a <u>user</u> to search for and identify <i>Information Packages</i> of interest.
General Model	The <u>General Model</u> is an E-ARK deliverable (D2.1), which breaks down complex processes into conceptual level activities, written in plain language, with minimal technical details, so that stakeholders (record managers, <u>archivists</u> , system designers, programmers) have a common understanding of the given examples.
GML	The Geography Mark-up Language: the XML grammar defined by the Open Geospatial Consortium (OGC) to express geographical features. <u>GML</u> serves as a modelling language for geographic systems as well as an open interchange format for geographic transactions on the Internet.
Information Package	A logical container composed of optional <i>Content Information</i> and optional associated Preservation Description Information. Associated with this <i>Information Package</i> is Packaging Information used to delimit and identify the <i>Content Information</i> and Package Description information used to facilitate searches for the Content Information.
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.
OLAP Cube	OLAP is an acronym for online analytical processing. An <u>OLAP cube</u> 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.
Order basket	Very similar to online shopping, it is the place where the <u>user</u> temporarily stores desired <i>Information Objects</i> .
Order Management Tool (OMT)	The E-ARK tool that manages orders created in the E-ARK <i>access</i> system.
order.xml	The xml-file that specifies an order in the E-ARK <i>access</i> system
Presentation tool	The tool and functionalities that provide <i>access</i> to Archival records.
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.
QGIS	A Free and Open Source Geographic Information System.
Search & Order Management	Represents the first process step of the E-ARK High-Level process described in Chapter 3 where the <i>consumer</i> wants to find information within the archive, then orders it.
SIARD-E	The E-ARK format for <u>databases</u> . Developed in collaboration with the Swiss Federal

	Archives (SFA), and based on the original SIARD format developed by SFA.
Status	The E-ARK <i>DIP</i> can have three statuses: See <i>DIP₀</i> , <i>DIP_u</i> and <i>DIP_p</i> .
Submission Information Package (SIP)	An <i>Information Package</i> that is delivered by the <i>Producer</i> to the OAIS for use in the construction or update of one or more <i>AIPs</i> and/or the associated Descriptive Information.
Transformation Tool	A tool that transform one type of IP into another, e.g. <i>AIP</i> to <i>DIP</i> .
User	The <u>user</u> is an umbrella term that designates all <u>users</u> of archival holdings, thus both internal <u>users</u> , cf. <u>archivists</u> , and external <u>users</u> , cf. <u>end-users</u> ,
User Management System (UMT)	The E-ARK tool that manages <u>users</u> in the E-ARK <i>access</i> system.

9 References

D2.1 General pilot model and use case definition

D3.1 Report on available best practices

D3.2 E-ARK SIP draft specification

D4.1 Report on available formats and restrictions

D4.2 E-ARK AIP draft specification

D5.1 GAP report between requirements for access and current access solutions

D6.2 Integrated Platform Reference Implementation (parts of it; not delivered yet)

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

Hadoop: <https://hadoop.apache.org/>

Internal E-ARK deliverable: Common Specification for IPs in the EARK project

Internal E-ARK deliverable: Defining DIP requirements

lily: <http://www.lilyproject.org>

OAIS, Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model, ISO 14721:2012: http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284

QCIS: <http://www.qgis.org/en/site/>

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