



D.5.1 GAP report between E-ARK requirements for access, and current access solutions.

DOI: 10.5281/zenodo.1172937

Grant Agreement Number:	620998
Project Title:	European Archival Records and Knowledge Preservation
Release Date:	14 th February 2018
Contributors	
Name	Affiliation
Kathrine Hougaard Edsen Johansen	Danish National Archives
Alex Thirifays	Danish National Archives
Piret Randmäe	National Archives of Estonia
Jože Škofljanec	National Archives of Slovenia
Janet Anderson	University of Brighton
David Anderson	University of Brighton
Kuldar Aas	National Archives of Estonia

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

EXECUTIVE SUMMARY

This report is a GAP analysis between current access services and user requirements for Access. It describes the landscape of access services today and highlights the results of the examinations of user needs for access services. The study targets the producers of the access services (service providers and archives), the clients of these services (archives) and the end-users of them (third-party users, the archives and the content providers themselves).

The purpose of this GAP analysis is first and foremost to be used internally within the E-ARK project to inform the format specification of the *Dissemination Information Packages* and the development of related access tools. Secondly, it can be read by external readers to gain understanding of the current situation and assess their own solution.

The report consists of an introduction; a description of the methodology used; a description of the results of the desktop research and the interviews conducted; an analysis of these results as well as an overview of current access solutions and user requirements for these solutions; a gap analysis; and a series of conclusions. The appendices cover a supplement to the methodology; the survey questions; an assessment of the interviewed stakeholders; the questions from the qualitative interview; and a terminology.

The most important contributions of this report are:

- A description of the landscape of current access services
- A description of user needs today
- A gap analysis comparing the access services to user needs

The landscape of current access services is three-fold: Complete or semi-complete access services offered to archives and developed by service providers, especially during recent years; access services developed in-house by archives that are tailored for the archives' own specific needs; and lastly archives that do not offer any access services, mostly because they have no accessible digital material yet. It should be mentioned, though, that most archives offer access to images, e.g. parish records. Amongst archives that have access services there is a general lack of interoperability of services, which detracts from the user experience and heightens costs. This impediment is being remedied gradually by full-fledged, homogeneous services offered by service providers, as well as solutions that archives are developing themselves.

User needs revolve around the:

- usability of access services (user-friendliness; flexibility; speed; ease-of-use and interoperability) and the
- quality of the search, as well as the quality of the use of *Information Packages* (comprehensive and understandable metadata; state of the art search engines with different entries; meaningful presentation of results; easy access to *Information Packages*; tools that support analyses)

Barriers to the fulfilment of these user requirements can be put in two main categories: Legislation and lack of maturity of archives and/or archival solutions. The former means that users cannot always access all the

Information Packages that they want; they cannot access them as fast as they would like; and procedures of obtaining access can be tedious. The latter means that the tools offered simply do not represent the state of the art – this applies to *Finding Aids*, presentation tools, and the viewing and analytic tools available. It also means that the *Information Packages* that the archives provide access to are not always particularly user-friendly: Either they cannot be easily found or understood (lack of metadata), or they are not delivered in a usable format (lack of formats for *Dissemination Information Packages*) nor with adequate *Representation Information* or other relevant metadata.

TABLE OF CONTENTS

1. TABLES	7
2. FIGURES.....	8
3. INTRODUCTION	9
3.1 Objectives of this deliverable.....	9
3.2 Scope	9
3.3 Background	9
3.4 What is an access service?.....	10
3.5 Who are the users and what are their requirements for access services?.....	12
3.6 Existing research	13
3.7 Legislation and access services	14
4. METHODOLOGY.....	16
4.1 Approach.....	16
4.2 Survey	16
4.3 Interviews	17
5. RESULTS	19
5.1 Survey	19
5.2 Interviews	19
5.3 The survey results	19
6. ANALYSIS OF RESULTS	20
6.1 Respondents	20
6.2 Overview of current access services	22
6.2.1 The indexing, accessibility and presentation of metadata: The Finding Aids	22
6.2.2 The platform for providing access: The presentation tool.....	24
6.2.3 Workflows and technical settings	26
6.3 Users and their needs	27
7. GAPS BETWEEN EXISTING ACCESS SERVICES AND USERS' NEEDS	31
7.1 Evaluation of gaps for identification process.....	31
7.2 Evaluation of gaps for request process.....	33
7.3 Evaluation of gaps for presentation process	34
7.4 Overall assessment of how well needs are met and recommendation for further E-ARK work	36
8. CONCLUSIONS	37

9. REFERENCES	39
Appendix A: Joint description of the cross-task methodology for information gathering	41
Appendix B: Survey Questions for Archives.....	47
Appendix C: Survey Questions for Service Providers.....	49
Appendix D: Assessment of stakeholders for interview from point of view of D5.1	50
Appendix E: Interview questions for Archives	53
Appendix F: Interview questions for Service Providers	55
Appendix G: Terminology.....	56

1. TABLES

Table 1: List of interviewed stakeholders	18
Table 2: Overview of evaluation of GAPs in the identification process	31
Table 3: Overview of evaluation of GAPs in the request process	33
Table 4: Overview of evaluation of GAPs in the presentation process.....	34

2. FIGURES

Figure 1: Conceptual overview of Access services.....	10
Figure 2: Typical workflow for access	11
Figure 3: Distribution of respondents across countries.....	21
Figure 4: Three necessary elements of the <i>Finding Aids</i>	23
Figure 5: Archives allowing metadata search across Information Packages	23
Figure 6: Platforms used to provide access to content at archives	24
Figure 7: Current users of archives' access services	28
Figure 8: The percentage of Archives who have studied their users' needs	29

3. INTRODUCTION

3.1 Objectives of this deliverable

The present work was conducted from February 2014 to June 2014 within Task 5.1, Access and Presentation Requirements, as part of Work Package 5 on Archival Records Access Services in the E-ARK project.

The purpose of this deliverable is to identify and analyse existing solutions for providing access to digital material, study users' needs and requirements for access services, and carry-out a gap analysis between the two in order to identify gaps that must be bridged by future access services.

The results will feed into the onward work of E-ARK. In particular, specifying a common *DIP* format (T5.3) and developing common access tools (T5.5) will benefit from and use the results. The results will also feed into Milestone MS01 "Best practice overview" that will combine information about best practices for *Ingest, Archival Storage and Access*¹ (cf. OAIS²) which are identified by Work Packages 3, 4 and 5 respectively and presented individually in Deliverables D3.1, D4.1 and D5.1. The results from this report can also be used by institutions to assess how well their own access services meet the identified user needs.

3.2 Scope

The possibilities when studying best practice for access services for digital material are wide. Archives, libraries, museums, government organisations and a wide range of online portals set examples for how Organisations can provide access to digital material. However, the Description of Work determines the scope of this report to be National Archives. To gain a comprehensive study, the scope was further extended to also include other archives and private companies (hereafter service providers) that provide archival solutions to archives. The focus is mainly on Europe, but to the extent possible stakeholders from across the world are also included. Although the work takes into account all kinds of digital material in archival collections, the primary focus is on born-digital material.

3.3 Background

Archives across Europe - and all over the world - are working towards solving the same problems of archiving digital material in the long term, but different solutions and practices for preserving digital material are employed. The differences concern the whole lifecycle including how data are ingested into the archive; how they are stored and preserved in the archives; and how they are disseminated, accessed and used by end-users.

¹ OAIS terms are used to the extent possible and written with first letter in capital and the whole word in italics.

² Open Archival Information System, ISO 14721:2012, <http://public.ccsds.org/publications/archive/650x0m2.pdf>

It is evident that there are multiple reasons for having different solutions. Factors like national legislation, regulations, cultures, traditions, finances and practices, etc., all influence the choice of solution. The consequences of using different solutions are as many. One of the negative impacts is the lack of interoperability, as every institution is developing and maintaining individual tools and workflows. From the point of view of *Access* one of the main downsides is that it is difficult to use data across country borders and institutions.

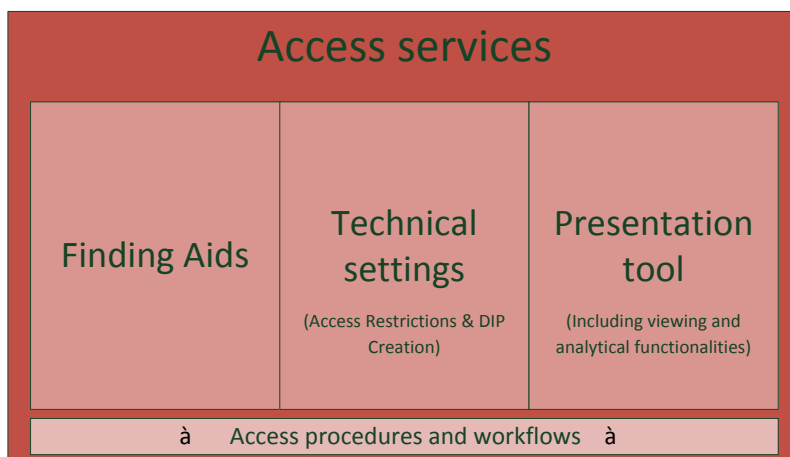
3.4 What is an access service?

Access is the bridge between an archive and its end-users where collections are made visible and accessible. The OAIS reference model defines the *Access* function as the “functional entity that contains the services and functions which make the archival information holdings and related services visible to Consumers”³. For the purpose of this deliverable, the term “access service” is defined as a setup within an archive that supports the access function. It includes all manual and automatic workflows, processes and tools used to support access to digital material, for example:

- *Finding Aids* for collection overview and identification of *Information Packages* (IPs) of potential interest
- Access restrictions management and processing of access requests including consideration such as confidentiality
- Workflows and tools for retrieving *Archival Information Packages* (AIPs) from storage and creating *Dissemination Information Packages* (DIPs)
- Procedures for making the created *DIPs* accessible for users
- Presentation tools for providing access to *DIPs* and supporting use of content

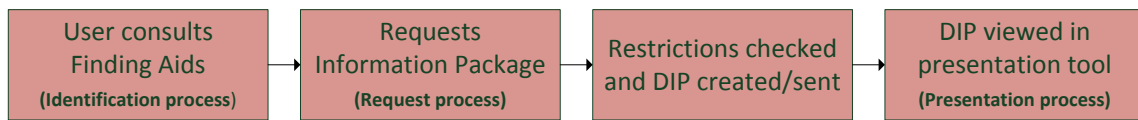
The components of an access service can be illustrated as in Figure 1 and a typical workflow for access is shown in Figure 2:

Figure 1: Conceptual overview of Access services



³ Reference Model for an Open Archival Information System (OAIS) (2012), page 1-8

Figure 2: Typical workflow for access



All access services serve the same purpose of making digital material visible and accessible for users, but the composition and design of access services vary from set-up to set-up. In the following we describe three aspects that are useful to characterise access services for the purpose of the report:

The indexing, accessibility and presentation of metadata: The *Finding Aids*. This concerns which metadata are accessible in *Finding Aids* and to what extent they can be searched for the purpose of identifying *IPs* of potential interest. Is it for example only possible to see and search on descriptive metadata about *IPs*? Or is it possible to search detailed metadata about the content in the *IPs*? This could be a search on specific variables in *IPs*, i.e. a search to identify all *IPs* which ones contain a given variable, e.g., a social security number. It could also be a search on specific values, e.g. to identify all *IPs* containing a particular value e.g., a specific social security number. Similarly, the way metadata are indexed and can be searched is significant. For example, have the detailed metadata been duplicated into the *Finding Aids*? And is it possible to search across metadata from within different *IPs* or is it only possible to search within one *IP* at a time?

The platform for providing access: The presentation tool. This relates to the way content is accessed and presented for users as well as the way it can be used. Is access to content for example provided via a presentation tool whereby *DIPs* can be accessed and used by users? And is that tool online or offline, and what functionalities does the tool include? Some tools may provide read-only access where content only can be viewed, while others may give the option to query the content and others again may give the option to export sub-sets to contemporary formats that can be processed in common software. Or perhaps access is not provided via a presentation tool, but by providing a copy of the *DIP*. And consequently the way content is presented and used is entirely up to the user.

Workflows and technical settings. This concerns the internal processes and workflows in the archive related to access as well as all the technical aspects such as what data structures, formats and asset types that can be handled by the access service. It includes workflows for *DIP* creation, tools for retrieving *AIPs* and creating *DIPs*, but also technical details about the *DIP* format and the digital material that is disseminated. Can it for example handle images, textual documents, databases, three-dimensional data, and which file formats are used?

Providing access to digital material is no trivial task. Legislative and technical restrictions, as well as the characteristics of the digital material itself, pose challenges to access services and define their possibilities and limits.

The characteristics of digital material, e.g. the file formats including the creation formats, preservation formats and access formats, impact the way the material can be used. Furthermore, the preservation strategy and preservation actions employed can influence the possibilities of use. For example a document which was created in MS Word, saved as PDF in an Electronic Document and Records Management System (EDRMS) and migrated to PDF/A-1b or TIFF 6.0 for preservation purposes will have certain limitations of

use, because of the format in which data is accessed. It will obviously not be useable for the end-user in the same way as it was for the producer; likewise for other content types. The quality of metadata is another important factor: If documentation and metadata descriptions about content and context of material are insufficient, meaning could be lost or there could be a risk of misinterpretation / misunderstanding.

Another obstacle for use is the *Knowledge Base* and level of skills the user has. This is especially obvious when users are accessing databases and EDM systems. Searching in databases requires knowledge of what a database is, how to define queries in SQL and how to ensure that the result of the query is correct. In addition, digital material is much easier to manipulate, correlate and query in different ways than paper records. This means that the risk of flaws and misinterpretation occurring is very great. Thus control queries where data are verified are of great importance, but they too demand a high level of knowledge and insight into data and data structures making it difficult for users to handle the data.

Legislation is another major factor which impacts access services. For example Data Protection Acts, Public Information Acts, Copyrights, etc., can be a hindrance for the access service and limit e.g. the amount of metadata that can be made available in the *Finding Aids* or the way digital material can be accessed and manipulated by users. At the same time legislation can place requirements on access services, for example requirements for confidentiality or logging each access that has been made. Although legislation is an important factor for access services, it is not a direct part of the service *per se*, but rather a premise for functioning of the access service.

3.5 Who are the users and what are their requirements for access services?

Users are the consumers of the digital material found in archival collections. Users are typically people, but can also be systems or services. An archive is likely to have several types of users including researchers, scholars, content creators accessing to their own material, legal instances including courts, internal users and the general public. Archives ideally have to know their *Designated Communities*, but in reality this is not, or rarely, the case. As shown by the survey only very few archives have clearly defined user groups and are aware of their *Knowledge Base*, interest, typical use-cases, etc.

The way users interact with access services varies very little in terms of what interactions take place. It consists of three overall steps:

- The **identification process** where a user searches collections in order to identify and investigate potential material of interest
- The **request process** where a user places an access request and the archives process it
- The **presentation process** where content material is made available to the user who can use it for their intended purpose

Figure 2 shows where in the typical access workflow the user interactions occur.

The primary tools relevant for users engaged in these processes are the *Finding Aids* and the presentation tools, and the users primary concerns using these are not surprisingly their usability.

While the way users interact with access services is uniform, the users' requirements for the access services are anything but. Different user types will have different requirements for access services depending on their objective and expected use of the digital material. For example, a researcher studying social-science who wants to do statistical analysis on a whole dataset will have different requirements for an access service than a private person doing genealogy who wants to access a single record holding, who again will have different requirements than content creators who want to reuse their own archived material in their day-to-day work.

Moreover users' expectations and requirements for access services are not expected to be static. They will inevitably evolve as the society and technological possibilities evolve, and users will expect contemporary access services which meet the standards and service levels of contemporary IT services.

3.6 Existing research

There have been several attempts to clarify and compare different aspects of digital archiving over the years. Some of the most recent and significant studies include:

- The study "*Common challenges, different strategies*"⁴ from 2012. This high level study compares strategies and approaches to digital archiving at national archives in Europe. It shows that there are significant differences in the regulative mandate of national archives as well as vast differences in how much experience national archives have in relation to handling and preserving born-digital material. It also shows that the quantity, types, complexities and the age of digital material vary greatly between national archives. The study has played an important role in raising awareness about the differences in strategies and approaches to digital archiving in Europe.
- A study from 2012 entitled "*Database Archiving*"⁵ investigated and compared approaches to database archiving in Europe. The study outlines the common challenges and problem areas related to database archiving and highlights that even though the majority of archives expect to preserve databases in the future, the current experience is limited.
- The study "*Digital Preservation Services: State of the Art Analysis*"⁶ from 2012. It is a high level study that compares and assesses the tools of publically accessible services and tools available to support digital preservation practices. The study shows that the majority of tools are small individual tools adapted for local needs. Furthermore, the study finds that there is a lack of services which orchestrate tools and services into holistic preservation solutions. The study is a central contribution to understanding the differences in digital preservation and illustrates the lack of collaboration as evidenced by the number of different tools available for solving the same tasks.
- A study from 2013⁷ investigated digital preservation practices and how they are implemented at libraries and archives. The main focus was on North America, but the study included respondents

⁴ Kristmar, K. V. (2012): "*Common challenges, different strategies*".

⁵ Velle, K. (2012): "*Database Archiving*".

⁶ Ruusalepp, R. & Dobrova, M. (2012): "*Digital Preservation Services: State of the Art Analysis*".

⁷ Bergin, M. B. (2013): "*Sabbatical Report: Summary of Survey Results on Digital Preservation Practices at 148 Institutions*".

from all over the world. The study found amongst other things that most organisations do digital preservation locally, but that some participate in collaborative efforts especially related to repositories. The study confirms what has been concluded in other studies, i.e. that the approaches taken to digital archiving differ greatly even though the challenges are the same.

Previous studies have contributed to the understanding and awareness of the differences and similarities in approaches taken to digital archiving across institutional and country borders. The current study sets apart from previous work by looking solely at access services and on users and their needs in order to identify gaps where the existing solutions do not meet users' current needs. Furthermore the study sets apart by being more detailed and attempting to go into technical aspects and as such it is not a high-level study like most of the existing work.

3.7 Legislation and access services

Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information⁸ in Article 1(2)(e,f) explicitly excludes documents held by archives from its scope. This has brought a clear distinction between documents in the public sector bodies and archival records held by the archives. Individual EU Member States themselves, therefore, legally regulate access to archival records and their use. In 2004, The International Council on Archives (ICA) published a draft of Principles for Archives and Record Legislation, in which one of the chapters deals with Access⁹. ICA's document defines freedom and liberty of access to records as a right of every citizen. Reasons to restrict access indicated are national security, public interests, protection of privacy, data protection, copyright or respect of individual donor agreements, but emphasising that those restrictions should not last forever.

Review of archival legislation available shows that EU Member States follow the basic principles of the ICA recommendations. On the other hand, legislations governing access to public information often¹⁰ state that they do not regulate domains of archival records and in this fully follow the provisions of the Directive in 2003.

In 2013, the adopted amendment of Directive on the re-use of public sector information¹¹ joined the archives together with other institutions of cultural heritage protection in a group of public sector institutions that are committed to ensure access to public information and their re-use. According to the present practice, which is largely based on the ICA's principles, it is not expected the amended provisions of

⁸ Official Journal L 345, 31/12/2003 P. 0090 – 0096. <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32003L0098>

⁹ DRAFT Principles for Archives and Record Legislation, 2004. <http://icarchives.webbler.co.uk/12188/articles-papers-reports-and-proceedings/draft-principles-for-archives-and-record-legislation-2004.html>

¹⁰ E.g. the Estonian and Slovenian Public Information Acts. <https://www.ip-rs.si/index.php?id=324>; <https://www.riigiteataja.ee/en/eli/ee/Riigikogu/act/514112013001/consolide>

¹¹ Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information. Official Journal L 175, 27/06/2013, p. 1–8. <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32013L0037&qid=1402866122682>

the Directive to be implemented in national legislation by July 2015, will substantially change the conditions of access to archival records.

4. METHODOLOGY

4.1 Approach

The purpose of the deliverable is to identify gaps between requirements for access and existing access solutions. In order to do this it is necessary to have in-depth knowledge about the practices and services currently used for providing access to digital material and detailed information about the requirements for access services from the point of view of users.

To this end we consulted archives and providers of digital archival solutions in order to gather details about their access services and the end-users' needs of archival services. The information is collected in two stages:

- **An online survey** completed by a broad range of stakeholders through which overall information about current digital archiving practices and access services was gathered
- **A series of qualitative interviews** with selected stakeholders through which detailed information about interesting and significant practices and services, workflows as well as end-users and their needs was gathered

The information gathering was done in collaboration with Task 3.2, EARK-SIP Specification, and Task 4.1, EARK-AIP specification. A joint description of the common approach and methodology for the information gathering is located in Appendix A.

4.2 Survey

The purpose of the online survey was to gather overall information about services and approaches to *Access* in order to identify and describe the current landscape for access services in the archives. The survey method was chosen as the first step in the information gathering because it allows for easy distribution to many potential respondents and because the quantitative answers are suitable for comparison and creating an overview.

Because the survey was made in collaboration with the two other above-mentioned E-Ark tasks, it not only includes questions about *Access* but also about the *Ingest* and *Archival Storage* functional entities. Further the survey addresses five stakeholder groups¹² of which only archives and service providers are in the scope of this report. The questions about *Access* in the survey concerns:

- Material types to which access is provided
- Tools used to support the access service
- *Finding Aids*

¹² Archives, Private companies, Government organisations, Private organisations and Projects

- Metadata
- *DIP* format(s)
- Legislation and access restrictions
- End-users of the access service

The full set of survey questions for archives and service providers can be found in Appendix B and Appendix C respectively. Questions directly relevant for Access are questions 19-42 for archives and 79-92 for service providers.

4.3 Interviews

Following the online survey a series of qualitative interviews were carried out with selected stakeholders to gather detailed information about significant and interesting access services and practices. Semi-structured, qualitative interviews were chosen as the method for this part of the information gathering, because the direct interaction and open-ended questions are suitable for getting in-depth insight into a few, selected stakeholders' practices and services.

Based on the Survey responses the stakeholders with the most interesting and relevant access services were identified. To ensure breadth in the interviews, the selected stakeholders cover different acquisition strategies, different collection profiles, different access services, and includes representatives from both stakeholder groups (archives and service providers).

Acknowledging that not all potential relevant stakeholders necessarily participated in the survey we additionally conducted desktop research, drawing upon the expert knowledge of E-ARK partners to make sure that no significant stakeholders were overlooked just because they did not respond to the survey.

The list of stakeholders invited to interview about their access solution is found in Table 1: List of interviewed stakeholders and the detailed schema used for assessing the relevance of potential stakeholders is located in Appendix D. The list of stakeholders invited for interview was circulated and verified within the E-ARK consortium.

Table 1: List of interviewed stakeholders

<i>Stakeholders invited to interview</i>	<i>Stakeholder type</i>	<i>Available for interview</i>
<i>The National Archives UK*</i>	<i>Archive</i>	<i>Yes</i>
<i>Estonian National Archives*</i>	<i>Archive</i>	<i>Yes</i>
<i>National Archives of Hungary</i>	<i>Archive</i>	<i>Yes</i>
<i>Swiss Federal Archives</i>	<i>Archive</i>	<i>No</i>
<i>Danish Data Archive</i>	<i>Archive</i>	<i>Yes</i>
<i>National Archives of Norway</i>	<i>Archive</i>	<i>Yes</i>
<i>The Archives of the Republic of Slovenia</i>	<i>Archive</i>	<i>Yes</i>
<i>Danish National Archives</i>	<i>Archive</i>	<i>Yes</i>
<i>Archivematica</i>	<i>Service provider</i>	<i>Yes</i>
<i>KEEP Solutions*</i>	<i>Service provider</i>	<i>Yes</i>
<i>Preservica</i>	<i>Service provider</i>	<i>Yes</i>
<i>Scope Solutions</i>	<i>Service provider</i>	<i>No</i>

* *These stakeholders answered the interview questions in writing due to difficulties arranging an actual interview.* The qualitative interviews were also conducted in collaboration with the two other above-mentioned E-ARK tasks and therefore the questions asked in the interview not only cover *Access* but also the *Ingest* and *Archival Storage* functional entities. The questions directly related to access cover:

- *Creation of DIPs*
- *Finding Aids* and searching possibilities
- *Access software* and possibilities of use
- *Access restrictions* and confidentiality issues
- *Users* and use-cases
- *User requirements*

The full set of interview questions targeted at archives and service providers are placed in Appendix E and F respectively.

5. RESULTS

5.1 Survey

There were a total of 184 responses to the online survey. 61 of those were from archives and 32 were from service providers. As archives and service providers are the only stakeholder groups in scope for this deliverable, only the answers from these 93 respondents will be considered in the subsequent analyses. Not all respondents completed the whole survey, which means that the number of total respondents to questions varies.

5.2 Interviews

12 stakeholders were invited to participate in the qualitative interviews about their access services. Of these it was only possible to conduct interviews with 10 of them. Table 1: List of interviewed stakeholders

shows the list of interviewed stakeholders. The interviews provided details about users, issues, workflows at the selected stakeholders' organisations – details that were not possible to collect via the survey. Because the interviews were only conducted with selected stakeholders, the information gathered does not necessarily represent the broad landscape of access services. Rather it is an expression of stakeholders and access services that are of particular interest for the purpose of this study.

5.3 The survey results

After analysing the survey results it has become clear that some respondents chose to interpret some questions in slightly different ways to that intended by the authors. This may have arisen because of local interpretation of the English or because of local use of specific terminology. In future surveys, to minimise the risk of this occurring, we will provide definitions of the terms used in the survey questions. We will take account of any suspected alternative interpretations of questions in our final detailed reporting, verifying these interpretations with individual survey participants where appropriate.

6. ANALYSIS OF RESULTS

In this section the results from the information gathering will be analysed and used to create an overview of current solutions and to identify users' needs for access services. The analysis of existing access services will be structured according to the three aspects defined in section 3.4 *What is an access service?* , namely:

- The indexing, accessibility and presentation of metadata: The *Finding Aids*
- The platform for providing access: The presentation tool
- Workflows and technical settings

And the analysis of users' needs will be structured according to the three steps in their interaction with archives' access services:

- The identification process
- The request process
- The presentation process

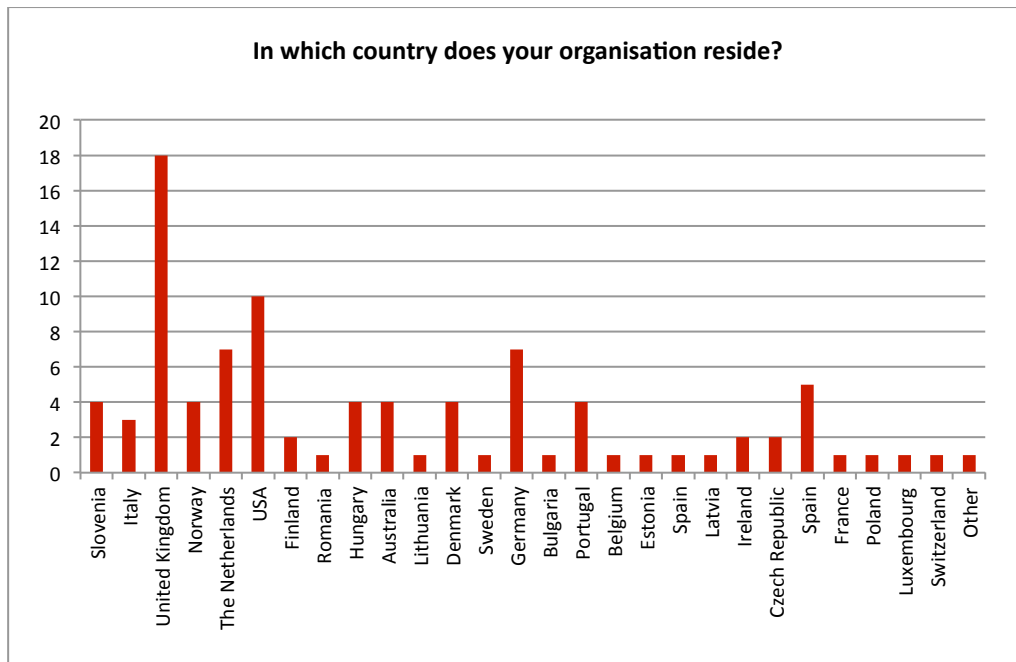
The analysis is made by using the survey results as a starting point and draw upon the interviews and the expert knowledge of E- ARK partners to support and elaborate on the survey findings.

6.1 Respondents

The first part of the analysis concerns the respondents, and outlines the context of respondents which is necessary in order to understand and analyse the survey results.

There were 93 responses to the survey from the stakeholder groups in scope for this report; 61 Archives and 32 Service Providers. The survey was distributed widely, and there is a predominance of respondents from European countries and a remarkably high number of respondents from the United Kingdom as seen from Figure 3.

Figure 3: Distribution of respondents across countries



The size of Organisations in terms of how many people are working in relation to digital curation varies. For archives the most frequent size is 1-10 persons (44%) followed by 10-20 persons (26%) (Q4). For the service providers there is an even distribution on sizes ranging from 1-20 persons to 100+ persons (Q68).

For Archives the size of the organisation also varies in terms of the collection size. The largest category is +20TB (38%), followed by 1-5TB (21%). The other categories are fairly evenly split between 12-15% (Q7). The largest category of digital content types in collections is digitised material (79%) followed by textual data and images (both 62%), databases (53%), then audio-visual data (47%). The other categories are all roughly between 15-20%, with other at 12% (Q9). This also means that most archives have a multitude of digital data with different origins to keep and provide access to. Digital material can be divided into structured and unstructured data. The definition is vague and can seem artificial, but nevertheless it describes a difference. For the purpose of the present work, structured data are defined as databases of any kind with or without files and/or documents. Unstructured data are defined as anything other than databases, such as files and documents in a folder system, hierarchical or not. The survey shows that there is a fairly even split between the two types in archival collections with structured 65% and unstructured 56% (Q10).

For acquisition strategies (Q6) there is a fairly even split between archives taking in single records (36%) and whole systems (44%), with 17% not known (N/K) and specific details from the other 19%:

- System extracts taken
- Files/single records in METS
- Datasets taken from website

For preservation strategies, migration (64%) is the most popular strategy, with the normalisation process used by 55% (Q11). When asked if they provide access to digital material (Q19), only 26 respondents

answered. 88% of those provide access to digital material. The remaining 12%, who do not currently provide access to digital material, all stated that they would do so in the near future.

Service providers were asked if they run any digital curation or access services for archives (Q79). Only 15 respondents answered this question and about half said that they run digital curation or access services for archives. 4 of those adjust their services to clients' individual needs and 7 do not adjust their services. 3 answered that they do not know whether their services are adjusted, and one did not reply to the question at all (Q81).

6.2 Overview of current access services

The level of experience related to providing access to digital material is relatively low. And when archives do provide access, it mainly relates to digitised material, e.g. scanned images of historical records. Providing access to digitised material is generally believed to be less complicated than providing access to born-digital material, because the digitised material often consists of few content types stored in simple, manageable formats, and since the records are historical they often have no or very few access restrictions attached to them. When it comes to born-digital material, the tendency is that more and more archives are beginning to establish practices, but experience is still limited.

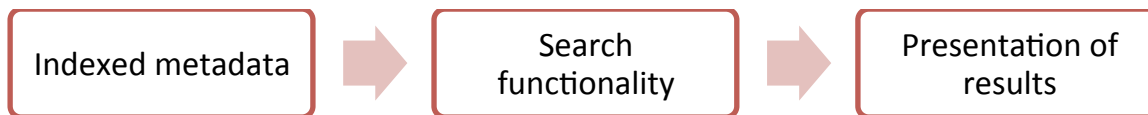
23 archives and 7 service providers out of the 41 respondents who answered the questions about *Access* stated that they provide access to digital material or run access services for archives. The analysis below is based on these 30 respondents and supported with information from the interviews.

Most of the stakeholders offer services that allow for the identification of the *IPs* (*Finding Aids*) and presentation of the *IPs* (presentation tool). In some cases, the tools exist separately, thus introducing a less user friendly and more time consuming access procedure; in other cases, they are integrated, and publically available material (both born digital and digitised) can be identified and accessed immediately using the same tool.

6.2.1 The indexing, accessibility and presentation of metadata: The *Finding Aids*

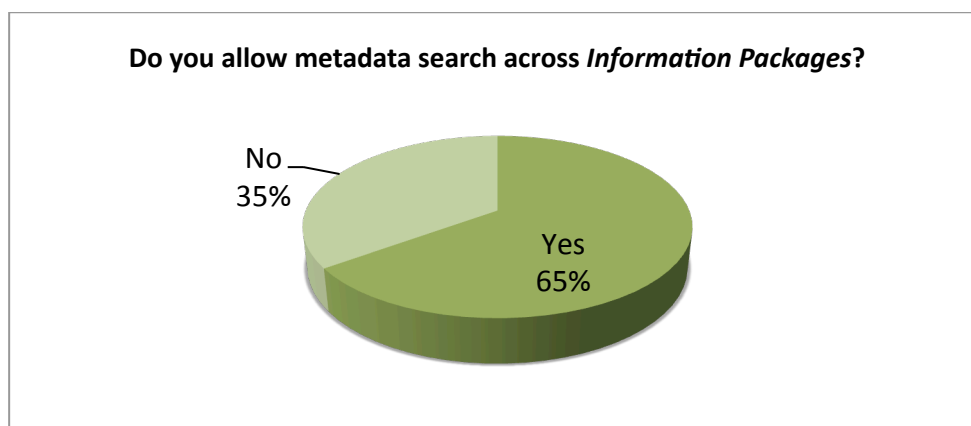
Finding Aids require 3 basic elements to function properly:

1. **Indexed metadata.** The amount, types and pertinence of the metadata are important and the indexing helps in increasing the speed of the search.
2. **Search functionality.** The search functionality will assist in accessing the metadata in a more or less structured fashion: Advanced search functionalities offer, for example, some sort of filtering, and the query results can be accompanied by thumbnails of the identified documents.
3. **Presentation of results.** The display of results can assist in navigating through them, so that they can be consolidated and sorted by relevance, title, size, author, publisher, year, description, etc.

Figure 4: Three necessary elements of the *Finding Aids*

Indexed metadata. The amount, type and pertinence of metadata that are indexed and searchable vary greatly between access services. When archives are asked about what kinds of metadata are visible and searchable in their *Finding Aids*, the picture is motley and no access method to metadata stands out (Q27): 10 out of 18 (56%) cited descriptive metadata/Dublin Core¹³. There are two mentions of technical metadata and one mention of administrative metadata. Other metadata *standards* quoted are DDI-L¹⁴, ISAD (G)¹⁵, EAD¹⁶ compliant metadata, ISAAR (CPF)¹⁷, FGDC CSDGM¹⁸ and DataCite¹⁹, and quoted metadata *elements* are archival reference number, title, time, provenance, geographical coverage, subject, connected persons, institutions, file type, size, dates of important actions, fond, inventory, archival unit, and creator.

65% (13 of 20) of the archives allow metadata search across all descriptions of *Information Packages* when browsing metadata for the purpose of identifying *Information Packages* of potential interest (Q28).

Figure 5: Archives allowing metadata search across *Information Packages*

Search functionality. Good and user friendly *Finding Aids* support multiple entries for search including a broad free-text search on all indexed fields à la “Google search” as well as the possibility for browsing different hierarchical structures, often spanning from fond-level to document-level. Sometimes the basic search function is complemented by an advanced and/or faceted search function allowing for searches

¹³ <http://dublincore.org/>

¹⁴ Data Documentation Initiative: <http://www.ddialliance.org/Specification/>

¹⁵ General International Standard Archival Description: <http://www.ica.org/10207/standards/isadg-general-international-standard-archival-description-second-edition.html>

¹⁶ Encoded Archival Description: <http://www.loc.gov/ead/>

¹⁷ International Standard Archival Authority Record For Corporate Bodies, Persons and Families: <http://www.ica.org/10203/standards/isaar-cpf-international-standard-archival-authority-record-for-corporate-bodies-persons-and-families-2nd-edition.html>

¹⁸ Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata: <http://www.fgdc.gov/metadata/csdgm/>

¹⁹ <http://www.datacite.org/>

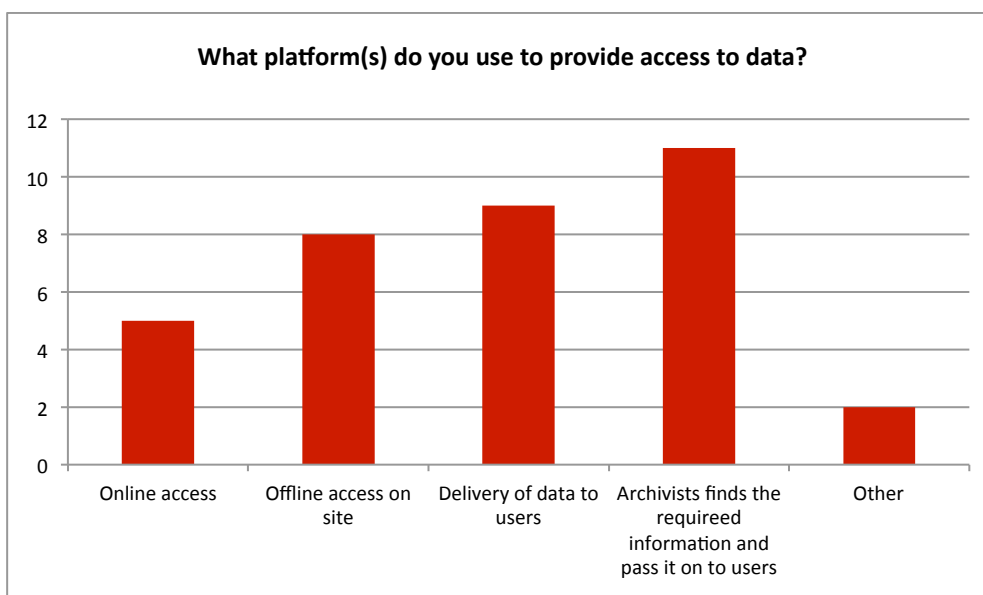
across all types of metadata for all *IPs* visible in the online search catalogue including descriptive metadata, as well as other standardised metadata. This means that search on titles, provenance, time periods, topics, variables, codes and values, etc., can be searched via one field. In a few cases, archives have applied crowdsourcing to allow users to contribute to descriptions and have also made such data searchable. In some cases the *IPs* that can be identified via the *Finding Aids* are the ones that have no access restrictions attached to them; however, some of the tools can be reconfigured to allow searches that target the entire collection, regardless of access restrictions. Currently, there is no common agreement on how to access and search *Finding Aids*.

Presentation of results. In approximately half of the cases the results presented are limited to the *IPs* that have no access restrictions attached to them, because even though one does not get access to the content data via the *Finding Aids*, the display of metadata elements can compromise anonymisation. The presentation of results is quite trivial, and the fonds, series, documents or files are simply displayed accompanied by a number of relevant descriptive metadata elements that users can run through to identify data of potential interest.

6.2.2 The platform for providing access: The presentation tool

According to the survey 65% of archives use tools for data dissemination (Q23). 10 archives use different tools depending on the content types that are being disseminated (Q23), and 5 service providers offer access services that use different tools for access depending on content types (Q84). No tool stands out as widely used from the survey. A wide variety of tools are used for different content types and about half of the tools are developed in-house and half are commercial tools. Most archives use a combination of different platforms to provide access to data. Figure 6 below shows the distribution of platforms used for access by archives (Q26).

Figure 6: Platforms used to provide access to content at archives



21 Archives replied to this question. The majority of archives offer multiple platforms for access. Half of them offer a service where staff find specific information and pass it on to users. About a third of them offer offline access on site and delivery of data to users on a loose medium. And about a fourth of the asked archives provide online access. The category “other” includes partly online access and open data harvesting using OAI-PMH. The most common combination of access platforms is online access, delivery of data and access to specific information via archivists. Service providers were asked a similar question, but because of the flaw in the survey where phrasing of questions in some cases leaves too much room for interpretation resulting in distorting of results, this question (Q85) is not included in the analysis.

Online access. The interviews indicate that most online access services are used to provide access to material which is publically available. Practices for providing access to material with access restrictions, including born-digital material, are still being established and experience is limited. However, some archives do offer online access to material with access restrictions via login where access permissions are managed by user settings.

Some services control access by managing user rights and give users access to general presentation tools, but only show material users have permission to access. Other access services allow users to see the entire collection, but show retention dates/access restrictions along with description of material in the metadata and only allow access to material that the user has permission to access. A third way of handling restricted access to digital material is to store the *DIP* in a separate place on an access platform that only an authorised user can access.

Search within DIPs. The interviews indicate that existing presentation tools (online or offline) generally only allow search on content in one *DIP* at the time, but search on content across *DIPs* is a functionality that many archives say that they would like to offer their users. However, from the interviews no clear conclusions can be made about *how* content can be searched in a *DIP*. Some presentation tools offer multiple entries for search in content and fine-grained indexing while others have more limited possibilities of search and most tools are somewhere in between.

Presentation tools and functionalities. When it comes to the functionalities supported in presentation tools (online or offline) and possibilities of use, the interviews suggest that there are different approaches used. Some services support access and use of several material types in one presentation tool with functionalities optimised for each material type. This could be for instance a tool that can handle access to e.g. text, images and databases, and depending on what material type is being accessed, different functionalities will be available. Other services use individual tools for material type with specialised functionalities. This could be for example one presentation tool for images and another tool for access to databases. For access services where *DIPs* are delivered to users instead of through a presentation tool, the way data can be used only depend on the format in which data are stored and what skills the user has. Based on the interviews there appears to be a slight overweight of services using one tool for all material types and delivering *DIPs* to users without presentation tools but nothing final can be concluded.

Based on the survey and interviews it is not possible to say much about the actual functionalities available in presentation tools as this level of detail was difficult to obtain in the interview form chosen with questions from all tasks in the collaboration. It was found that the design of each service and each tool is different and the specific functionalities and how they are integrated differ.

User-friendliness. Most interviewees assess their access service – or parts of it – to be user-friendly, but they also seem to be aware of where improvements can be made to increase user-friendliness. Some emphasise interfaces as the most use-friendly part while, on the contrary others say that the interfaces are what weaken the user-friendliness and should be improved. Likewise some highlight the indexing of content, possibilities of search, functionalities, etc., as the strengths of their presentation tool while others mention these aspects as weaknesses that should be improved. In conclusion there is no clear picture about which parts of presentation tools are user-friendly and which are not. The only thing on which there is consensus is that online access to digital material (with or without access restrictions) is an important feature that makes access convenient, user-friendly and improves the overall user-experience.

Legislation and access restrictions. The possibilities of use are not only determined by the way access is provided. Legislation and other access restrictions also set boundaries for the use. Access to digital material is limited due to access restrictions at 22 out of 23 archives (Q34), but only 6 out of 23 archives have specific restrictions related to data mining (Q36). Interviews showed that some archives handle access restrictions by anonymisation of sensitive data while others only rely on rights management and limited access.

6.2.3 Workflows and technical settings

The technical settings in access services could be investigated and analysed in great detail, but as this report is focused on users and the gaps between users' needs and existing services, the details about technical settings in existing solutions are kept to a minimum. The report only covers aspects that can be relevant from a user's point of view.

Material types. The survey shows that a variety of different material types are supported by existing access services and almost all services support more than one material type (typically 3-5 types). For archives the major category is images which 95% of all respondents currently provide access to. This is followed by textual data at 70%, databases at 50% and audio-visual material at 35%. 20% of the archives provide access to statistical data, scientific data and complex data, while only 15% provide access to survey data (Q21). For service providers there is no clear tendency of which material types their access service support (Q83), but overall there is a broad coverage of the different types.

DIP formats. About half of the archives (11 of 23) have specific *DIP* formats (Q29). And 7 out of 9 archives use different *DIP* formats depending on the content type (Q30). For service providers 4 of 14 companies use a specific *DIP* format in the access service they offer (Q86) and 3 out of 4 offer different *DIP* formats depending on content types (Q87).

Metadata standards. 14 of 23 archives use metadata standards for dissemination of digital material (Q32) and several archives use more than one standard. Descriptive metadata (Dublin Core) and archival metadata, EAD, are the only ones that emerge as relatively common (Q33). The main dissemination metadata standards used at archives are:

- Dublin Core (36%)
- EAD (36%)
- METS (18%)

- ISAD (G) (18%)
- SIARD (9%)
- MARC21 (9%)
- MODS (9%)
- PREMIS (9%)
- FGDC CSDGM (9%)
- Data Cite (9%)
- iptc-photo metadata (9%)

The interviews only covered very few technical aspects and only at a high level. As such it is neither possible nor the intention to analyse and describe the technical setting in-depth. However, the interviews demonstrated that the technical settings vary greatly from access service to access service. For example the *DIPs* structures vary from XML based formats to single files with their metadata. Another example of differences is the process for *DIP* creation; some services have fully automated *DIP* generation while others have partially automated *DIP* generation, where *DIPs* are created through a combination of automatic and manual processes. Manual processes could for example include quality assurance or metadata enrichment before dissemination. Likewise there are differences in the time at which the *DIPs* are created. One practice is to create *DIPs* upon *Ingest*. The interviews indicate that this is most common for publically available material (and that these *DIPs* are usually kept in a special storage from where dissemination is easy). For material with access restrictions the interviews show that it is common that *DIPs* are created upon request and are typically deleted after use or after a certain time period. But other practices are used and the approach taken depends how the rest of the access service is designed, for instance dependent on what material types are concerned and how restricted access is managed.

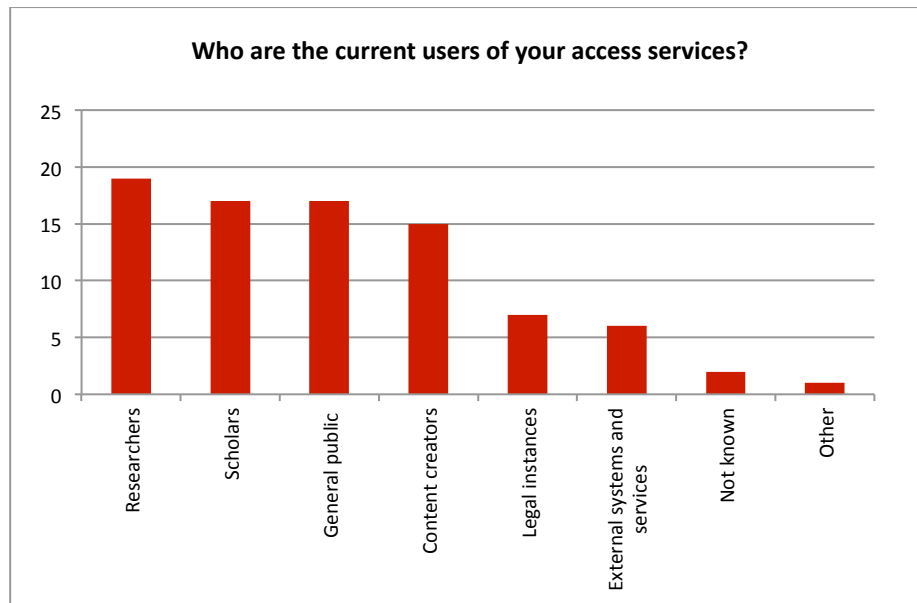
6.3 Users and their needs

Information about users' needs for access services was gathered through archives' and service providers' own studies. Even though this information is not collected directly from the users themselves by the E-ARK project, it is the best available source for users' needs for access services. It is challenging to say anything conclusive about users' needs based on information from this source, since the information gathering has not been tailored for the present study, but it can however be used to deduce a series of the users' needs. Ideally the users themselves should be interviewed about their typical use of access services and what needs they have, but it was not possible within the timeframe or scope of this task.

Who are the users? The survey shows that the most frequent users of archives' access services are researchers, scholars, the general public and content creators (i.e. *Producers*) as seen from Figure 7 below (Q39). This finding is consistent with recent results from the APEX Project.²⁰

²⁰ Aas, K. et al. (2013): "D6.1 First Analysis report: Applying Web 2.0 solutions in archival application", Archives Portal Europe (APEX), page 8-9.

Figure 7: Current users of archives' access services

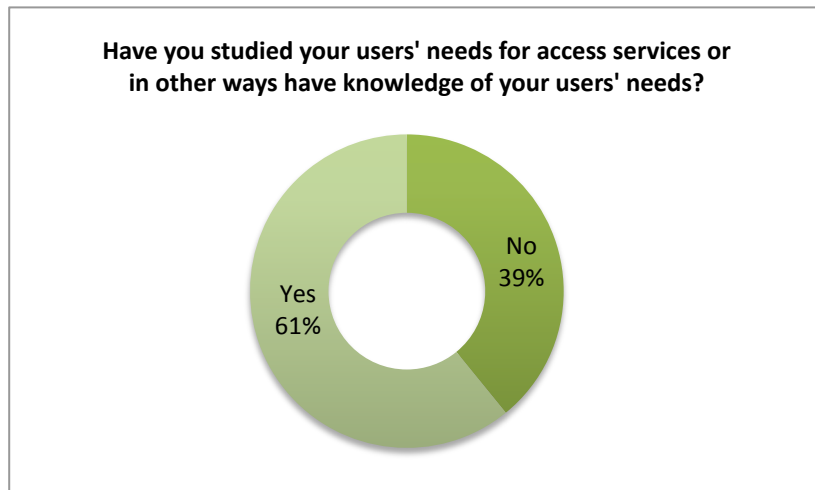


How is material used? The survey and interviews revealed that a large part of access to digital material consists of providing access to digitised material where typical use is history research, genealogy, and quests for specific pieces of information. For born-digital material the picture is varied and typical uses are more vaguely defined. This is partly because archives generally have less experience with providing access to born digital material, and therefore use-cases are not defined yet. But also because there are many more possibilities of use for the different types of born-digital material depending on the type and format. From the interview it is difficult to say much about typical uses for born-digital material, because the interviewed institutions themselves had limited knowledge about the typical use of data.

The only uses that stand out as typical are creators who get access to old records, internal users who access archived material as part of their work or to find a requested piece of information for users, and researchers and scholars who use material for various research purposes including social sciences, humanities, history, etc., with no further specification of how data is used.

What are users' needs? Information about users' needs is also somewhat limited. 14 of 23 (61%) archives said that they have studied their users' needs for access services (Q40) as illustrated in Figure 8. And even those who have studied users' needs do not necessarily have a clear overview.

Figure 8: The percentage of Archives who have studied their users' needs



The inadequate knowledge is probably related to the fact that the experience with providing access to born-digital material is still limited. However, some overall needs for access services have been identified. Generally users are concerned with:

1. Usability and flexibility of services
2. Efficient services and speed of access
3. Integration or interoperability between different parts of an access service e.g. between *Finding Aids* and presentation tools
4. Contemporary solutions that meet the standards of modern IT services
5. Services that are easy to use i.e. that do not require specific technological or human skills
6. Possibilities to search in data and metadata across *Information Packages*
7. Users are concerned with support of their own specific purpose of use. For example a social science researcher could be concerned with support of reuse of data for statistical analysis while a private person who is requesting a specific piece of information only is concerned with support of this function

Through the interviews and consulting of E-ARK partners, more detailed information about users' needs was identified. The needs can be divided into needs for different steps in the access process.

For the **identification process** the following needs have been identified:

1. Need for easy access to *Finding Aids*
2. Need for access to comprehensive metadata about *Information Packages* in *Finding Aids*
3. Need for more than one way of searching in *Finding Aids*
4. Need for a free-text search across metadata elements via one search field i.e. "Google Search" on metadata about *Information Packages*
5. Need for an advanced search
6. Need for a search via browsing a hierarchy
7. Need to search across *Information Packages* using variables, e.g. to identify those containing social security numbers

8. Need for user-friendly and meaningful presentation of search results
9. Need for integration or interoperability between *Finding Aids* and presentation tools

For the **request process** the following needs have been identified:

1. Need for an integrated and easy way of requesting access to material
2. Need for fast processing of access requests and access to content
3. Need for continuous access to old records without having to go through the request process each time records are accessed (this need is specific to *Producers*)

For the **presentation process** the following needs have been identified:

1. Need for online access to disseminated *Information Packages* via presentation tools
2. Need for access to *DIPs* without the use of presentation tools, e.g. delivered on a portable medium
3. Need for support of data mining
4. Need to export and reuse data and/or subsets of data in external tools e.g. MS Word, MS Excel, statistical analysis packages
5. Need to print out metadata, digital documents and records from within *Information Packages*
6. Need to search on the content of digital documents within *Information Packages*
7. Need for free-text search on content in an *Information Package* via one search field i.e. a “Google Search” on the content
8. Need to search on values across different *Information Packages*, e.g. search on a specific social security number
9. Need to search on metadata elements within one *Information Package*
10. Need to relate content to *Representation Information* while viewing and using content
11. Need for the integration of viewing tools in presentation tools, e.g. a document viewer or video viewer

Different user types have different needs for access services, and users may also have different needs depending on content types. Based on the information collected through the survey and the interviews it is not possible to identify specific needs for either user types or content types, because the data basis is too small and the interviewed archives do not have such detailed information about their users. The needs listed above are thus a joint expression of the needs identified and as such they are not necessarily all relevant for all user types.

It can then be assumed that the aim of E-ARK should be to develop atomic components with simple and limited functionality (i.e. viewers), which could be implemented in a variety of scenarios and presentation tools.

7. GAPS BETWEEN EXISTING ACCESS SERVICES AND USERS' NEEDS

The GAP analysis is carried out by taking the identified user needs as a starting point. For each need it is then evaluated how well or poorly that need is met by the existing services. The evaluation is made using the results from the analysis in section 6. A rank will be assigned in order to indicate to what degree each need is fulfilled by existing access services. The ranks are:

- **Good.** This rank is given if the majority of services meet a requirement well and only a few do not meet it or only meet it partially.
- **Average.** This rank is given if around half of the services meet the need well or most of services meet the need partially.
- **Poor.** This rank is given if only a few services meet the requirement either fully or partially and the remaining services do not meet the requirement.
- **Unknown.** This category is given if it is not possible to assess whether or not a need is met based on the analysis of existing solutions because it was not covered by the survey and interview questions. It can also be given if the evaluation gives ambiguous results and it is not possible to conclude if there is a gap or not.

Not all needs can be evaluated using this method. This is partly because of the methodology design where information about users' needs was identified at the same time as details about the existing services. This means that it was not possible to design the study around the users' needs and target questions to identify information about these particular aspects of access services. Instead the approach chosen was to make a comprehensive study of the current access services to get an overall insight into each service. This was done presuming that this would be adequate for the GAP analysis. In most cases this holds true, but for a few of the identified needs the information gathered about access services is not detailed enough to allow evaluation and assessment of potential gaps.

7.1 Evaluation of gaps for identification process

Table 2 gives an overview of how well existing services meet users' needs for the **identification process**. Detailed evaluation of each need is found below.

Table 2: Overview of evaluation of GAPS in the identification process

User need	Rank
→ Need for easy access to <i>Finding Aids</i>	Good
→ Need for access to comprehensive metadata about <i>Information Packages</i> in <i>Finding Aids</i>	Poor
→ Need for more than one way of searching in <i>Finding Aids</i>	Good
→ Need for free-text search across metadata elements via one search field i.e. "Google Search" on metadata	Average

↪ Need for an advanced search	Average
↪ Need for a search via browsing a hierarchy	Poor
↪ Need to search across <i>Information Packages</i> for variables e.g. a search to identify those containing social security numbers	Unknown
↪ Need for user-friendly and meaningful presentation of search results	Poor
↪ Need for integration or interoperability between <i>Finding Aid(s)</i> and presentation tools	Poor

Detailed evaluation of needs:

- Need for easy access to *Finding Aids*
Good. Most services have online access to *Finding Aids* where collections can be searched.
- Need for access to comprehensive metadata about *Information Packages* in *Finding Aids*
Poor. The analysis shows that there are vast differences in the amount, type and pertinence of metadata in *Finding Aids*.
- Need for more than one way of searching in *Finding Aids*
Good. Most services offer multiple entries for search. However, if the accessibility and indexing of metadata which can be searched is poor, this will be the limiting factor in the identification process.
- Need for free-text search across metadata elements via one search field, i.e. "Google Search" on metadata about *Information Packages*
Average. Around half of the studied services offer this kind of search. However, the quality and usability of the search depends on the accessibility and indexing of metadata.
- Need for advanced search
Average. Around half of the studied services offer this kind of search, but the quality and usability of the search depends - again - on the accessibility and indexing of metadata.
- Need for search via browsing a hierarchy
Poor. Only some services offer this kind of search.
- Need to search across *Information Packages* using variables, e.g. a search to identify those containing social security numbers
Unknown. The analysis does not cover this aspect of *Finding Aids* in sufficient detail to evaluate whether this need is met or not.
- Need for user-friendly and meaningful presentation of search results
Poor. The analysis showed that the ways search results are presented in *Finding Aids* generally are poor. They serve the identification purpose, but without being remarkably user-friendly.
- Need for integration or interoperability between *Finding Aids* and presentation tools

Poor. Only some of the studied access services have integrated *Finding Aids* and presentation tools. The analysis further indicates that integration between services is most common for digital content with no access restrictions.

In summary all in all the existing *Finding Aids* meet the needs moderately. The biggest gap is that the *Finding Aids* do not always allow users to find what they are looking for: The lack of qualitative and comprehensive metadata compromises the *Finding Aids*' performance and efficiency, which directly impacts the user experience and the user's access to the archival holdings in their entirety^{21, 22}. To circumvent this to a certain extent, the possibility of browsing the hierarchical structure of a collection is available in some archives. Another impeding factor when conducting searches in archival holdings, is national legislation that limit the possibility of identifying the *IPs*. Because metadata risk revealing restricted information and breach anonymisation, some access services only allow for searches in publically available *IPs*, even though they can be configured otherwise. It should be stressed that these restrictions related to metadata search are not imposed everywhere.

7.2 Evaluation of gaps for request process

Table 3 gives an overview of how well existing services meet users' needs for the **request process**. Detailed evaluation of each need is found below.

Table 3: Overview of evaluation of GAPs in the request process

User need	Rank
– Need for an integrated and easy way of requesting access to material	Unknown
– Need for fast processing of access requests and access to content	Unknown
– Need for continuous access to old records without having to go through the request process each time records are accessed	Unknown

Detailed evaluation of needs:

- Need for an integrated and easy way of requesting access to material
Unknown. The way access is requested is not covered directly by the question in the survey or in the interviews. Furthermore, the access request process is not necessarily handled exclusively within the archive: For example national data protection agencies can also be involved in the process.
- Need for fast processing of access requests and access to content

²¹ Note that metadata are not necessarily created and indexed as part of *Access*, so unless there is continuity in the way archives conceive their *Information Packages* (from *SIP* over *AIP* to *DIP*), inadequate metadata represent a condition that the *Access* service is subject to.

²² Whether *Information Packages* are found easily or not also depends on the performance of the algorithms that constitute the search engine. This study has not evaluated these.

Unknown. The time frame from a user requests access to the *DIP* until it can be accessed depends on the access restrictions, the transportation of the *DIP* and on the access request procedures that archives follow. For example it depends on whether or not the access requests are processed within the archive or if third parties e.g. data protection agencies, are involved.

- Need for continuous access to old records without having to go through the request process each time records are accessed

Unknown. This aspect is not covered by the study and cannot be evaluated.

In summary the elements which could be tweaked to allow for an optimum user experience include the integration in the *Finding Aids* of a functionality allowing for the access request to be initiated; of documentation explaining the conditions of providing access to any identified *IP*; and access conditions in general privileging the speed of granting access and providing access to the *IP* (tiered storage, online access, rapid case management). However, because the request process was not covered by the study, the extent to which the user needs are met remains unknown.

7.3 Evaluation of gaps for presentation process

Table 4 gives an overview of how well existing services meet users' needs for the **presentation process**. Detailed evaluation of each need is found below.

Table 4: Overview of evaluation of GAPS in the presentation process

User need	Rank
↪ Need for online access to disseminated <i>Information Packages</i> via presentation tools	Average
↪ Need for access to <i>DIPs</i> without the use of presentation tools e.g. delivered on a portable medium	Good
↪ Need for support of data mining	Unknown
↪ Need to export and reuse data and/or subsets of data in external tools e.g. MS Word, MS Excel, programs for statistical analysis	Unknown
↪ Need to print out metadata, digital documents and records from within <i>Information Packages</i>	Unknown
↪ Need to search on the content of digital documents within <i>Information Packages</i>	Poor
↪ Need for free-text search on content in an <i>Information Package</i> via one search field i.e. a "Google Search" on the content	Unknown
↪ Need to search on values across different <i>Information Packages</i> e.g. search on a specific social security number	Unknown
↪ Need to search on metadata elements within one <i>Information Package</i>	Unknown
↪ Need to combine content to <i>Representation Information</i> while viewing and using content	Average
↪ Need for integration of viewing tools in presentation tools e.g. a document viewer or video viewer	Average

Detailed evaluation of needs:

- Need for online access to disseminated *Information Packages* via presentation tools
Average. The analysis shows that this need is met by some access services, but not all.
- Need for access to *DIPs* without the use of presentation tools, e.g. delivered on a portable medium
Good. Most services include the possibility of giving access to *IPs* without going through presentation tool.
- Need for support of data mining
Unknown. Only the legislative aspects of this is covered by the study and shows that use of digital material for data mining purposes generally is not restricted by special legislation. Only a few archives stated in the survey that use of digital material for data mining purposes is restricted, but it is likely that the general access restrictions will be the biggest obstacle.
- Need to export and reuse data and/or subsets of data in external tools e.g. MS Word, MS Excel, programs for statistical analysis
Unknown. The respondents did not supply enough detail to allow us to evaluate this need.
- Need to print out metadata, digital documents and records from within *Information Packages*
Unknown. The respondents did not supply enough detail to allow us to evaluate this need.
- Need to search on the content of digital documents within *Information Packages*
Average. Several services include this functionality both for born-digital and digitised material, but the way the searches can be done and the quality of the search vary.
- Need for free-text search on content in an *Information Package* via one search field i.e. a “Google Search” on the content
Unknown. The respondents did not supply enough detail to allow us to evaluate this need.
- Need to search on values across different *Information Packages*, e.g. search on a specific social security number
Unknown. The respondents did not supply enough detail to allow us to evaluate this need.
- Need to search on metadata elements within one *Information Package*
Unknown. The respondents did not supply enough detail to allow us to evaluate this need.
- Need to combine content to *Representation Information* while viewing and using content
Average. Most services include this functionality to some extent.

- Need for integration of viewing tools in presentation tools, e.g. a document viewer or video viewer Average. Most presentation tools include this functionality to some extent, but most presentation tools only offer viewing tools for some content types.

In summary there are many gaps in relation to the specific needs for the presentation process. This is partly because only a few presentation tools that deal with born-digital material exist²³ and partly because some of the archives simply have not yet ingested any born digital material; additionally many of those who have not prepared it for Access. When archives do offer access to born digital material, the needs for platforms to make it available (online, portable medium) are quite well covered. However, the functionalities including options for export, analysis and searching within *IPs* appear to be poorly covered even though it was not possible to evaluate these needs thoroughly. The interviews included direct questions about functionalities in presentation tools and possibilities of use of data, and if the interviewed stakeholder had had sophisticated solutions with advanced functionalities, they would most likely have highlighted them during the interview. Nevertheless, this was not the case and there is thus reason to assume that most presentation tools do not have such advanced functionalities. Overall it is safe to say that as few archives have born digital material to grant access to, this area is somewhat underdeveloped.

7.4 Overall assessment of how well needs are met and recommendation for further E-ARK work

Overall the analysis shows that there are some considerable gaps between the users' needs and the existing access services. Although most needs are met in one or more of the services and all services meet some of the needs, there is no one service which meets all the needs. As seen from the above it was not possible to evaluate all needs and the extent to which some of the needs are met remains unknown.

Generally the existing access services have limited usability, due to lack of integration of services (*Finding Aids*, presentation and analytical/viewing tools), but also due to external factors that are hard to amend, such as unevenly well indexed metadata, legislation (restricted access) and, of course, resources. The best coverage of specific needs is in relation to *Finding Aids* (identification process) and the biggest gaps are found in relation to the specific needs for the presentation process. The request process was not covered by the interviews and the coverage of the needs for this process remains unknown.

Little practice exists in regard to access to born-digital material. This means that E-ARK has the opportunity to lead the way and propose components especially for the technical and semantic interoperability of access services. For example, there is the possibility to propose a core set of DIP descriptions and provide an example on how/whether these should be made available for searching in the finding aids. E-ARK should focus on complex born-digital material types as they pose a bigger challenge and have more gaps because archives have less experience regarding these than with simple material types such as digitised copies of historic records. Furthermore, E-ARK should focus on the metadata level in *Finding Aids*, e.g. by providing guidelines for which metadata should be searchable, rather than on the *Finding Aids* which, in general, are performing adequately.

²³ Access to simple content types, e.g. digitised images of historical records, is trivial.

8. CONCLUSIONS

This report has studied access services and the requirements they must live up to, and identified the gaps between the two.

Firstly the existing services were studied and analysed to gain an overview of the landscape of existing access services and the set-ups used to provide access to digital material. The analysis was made based on three overall characteristics of access services: The way collections can be searched in *Finding Aids*, the platform for providing access to digital material, and the internal workflows and technical settings used to support the access services.

Next the users' needs for access services were identified according to the different stages in their interaction with access service; namely the identification process, the access request process and the presentation process.

Following this, gaps between existing services and users' needs were evaluated. Taking as a basis the user needs and evaluating how well they are met by the existing access services. The analysis showed that there are considerable gaps between the users' needs and the access services. Generally the existing access services have limited usability, and do not meet the users' need for modern solutions that meet the standard of contemporary IT solutions.

The biggest gap for the identification process, where users browse/search collections to identify material of potential interest is the lack of comprehensive metadata available and indexed in *Finding Aids*. This compromises the performance and efficiency of *Finding Aids*, which directly impacts the user experience and the user's access to the archival holdings in their entirety.

For the request process, where users request access to digital material and await that request to be processed, the biggest gap is the lack of speed. The timeframe from the access is requested to the material can be accessed should be as short as possible to accommodate users and increase the user experience. There is no evidence of this from the examinations of user needs, but it is a well-known fact, and especially legislation delays the request process.

The biggest gap for the presentation process, where users get access to content, is the general lack of established experience of offering such services. For simple content types like digitised images of historical records, access is simpler and there is plenty of experience. But for born-digital material it is a different story. Many archives still do not provide access to born-digital material and only some of those who do have actual presentation tools to provide access to it. Furthermore the tools that do exist are simplistic and lack functionalities to support advanced use and analysis. All in all it is still an underdeveloped area.

It should be noted that some of the gaps that we have identified exist because of external factors that the E-ARK project of course have no means to remedy. This is particularly true when it comes to the national legislations that regulate the access conditions. It is also true, however, when it comes to the state and quality level of metadata in the archives. And lastly, it is true that many archives are in the development phase as regards to digital curation in general, and to Access in particular.

Future work should attempt to bridge the gaps identified in this report to better meet the user needs in future access services. Especially the E-ARK tasks on the development of common *DIP* formats and on the development of *Access* tools in general should address those gaps. Regarding the gaps that exist because of the external reasons recapitulated above, it is not possible to do anything about them within the scope of the E-ARK project, but to raise awareness about them.

9. REFERENCES

Aas, K. et al. (2013): "D6.1 First Analysis report: Applying Web 2.0 solutions in archival application", Archives Portal Europe (APEX):

http://www.apex-project.eu/images/docs/D61_Web20_In_Archival_Applications.pdf

Bergin, M. B. (2013): "*Sabbatical Report: Summary of Survey Results on Digital Preservation Practices at 148 Institutions*", University of Massachusetts, Amherest:

http://works.bepress.com/cgi/viewcontent.cgi?article=1012&context=meghan_banach

Data Documentation Initiative: <http://www.ddalliance.org/Specification/>

Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32003L0098>

Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32013L0037&qid=1402866122682>

DRAFT Principles for Archives and Record Legislation, 2004: <http://icarchives.webbler.co.uk/12188/articles-papers-reports-and-proceedings/draft-principles-for-archives-and-record-legislation-2004.html>.

Dublin Core Metadata Initiative: <http://dublincore.org/>

Encoded Archival Description (EAD): <http://www.loc.gov/ead/>

Estonian Public Information Act:

<https://www.riigiteataja.ee/en/eli/ee/Riigikogu/act/514112013001/consolide>

Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata:

<http://www.fgdc.gov/metadata/csdgm/>

Data Cite: <http://www.datacite.org/>

General International Standard Archival Description: <http://www.ica.org/10207/standards/isadg-general-international-standard-archival-description-second-edition.html>

International Standard Archival Authority Record For Corporate Bodies, Persons and Families:

<http://www.ica.org/10203/standards/isaar-cpf-international-standard-archival-authority-record-for-corporate-bodies-persons-and-families-2nd-edition.html>

Kristmar, K. V. (2012): *“Common challenges, different strategies”*, EBNA, 29 May 2012, Copenhagen
http://www.sa.dk/content/us/about_us/danish_national_archives/25th_european_board_of_national_archivists_conference/presentations

Reference Model for an Open Archival Information System (OAIS) (2012), The Consultative Committee for Space Data Systems:
<http://public.ccsds.org/publications/archive/650x0m2.pdf>

Ruusalepp, R. & Dobрева, M. (2012): *“Digital Preservation Services: State of the Art Analysis”*, Digital Cultural Heritage Network, DC-NET:
<http://www.dc-net.org/index.php?en/201/publications>

Slovenian Public Information Act: <https://www.ip-rs.si/index.php?id=324>

Velle, K. (2012): *“Database Archiving”*, EBNA, 29 May 2012, Copenhagen:
http://www.sa.dk/content/us/about_us/danish_national_archives/25th_european_board_of_national_archivists_conference/presentations

Appendix A: Joint description of the cross-task methodology for information gathering

Tasks T3.1, T4.1 and T5.1 formed a Cross-task collaborating to analyse current solutions and best practices for Ingest, Archival Storage and Access respectively. This was done to align work, be effective and avoid redundancy but also to ensure that stakeholders were not approached several times by different tasks leaders from the E-ARK project asking for details about their digital archiving practices.

The objective of this cross-task work was to build a knowledge base about best practices, tools, requirements and restrictions relevant to archiving solutions. The collected information will feed into the onward work within E-ARK to specify common formats for OAIS Information Packages (SIP, AIP and DIP) and in the work of developing common tools for archival services. We conducted our work through desktop research, an online survey sent to a wide range of stakeholders and a series of qualitative interviews with selected stakeholders. We included the following groups into this work:

- Archives (national, municipal and private archives. This group was also intended to include libraries),
- Private companies that have developed archiving services
- Private organisations that have developed archiving services
- Projects that have developed archiving services
- Public organisations (creators of digital content (Producers) and regulatory bodies)

We contacted Organisations throughout Europe, as well as in North America, Australia and New Zealand. Our findings gave a unified view of three areas of research, each specified to support work in one of our reports:

- **Ingest.** Best practices for pre-ingest, ingest and ingest tools
- **Archival Storage.** Available formats and restrictions for storage and different national requirements for authentication for legal purposes.
- **Access.** Gaps between requirements for access and current access solutions.

Desktop research

The purpose of the desktop research was to get overall knowledge of current practices and access solutions.

Method

We began with desktop research as an initial stage of our task followed by primary research – a quantitative online survey that was ensued by qualitative interviews. Our desktop research comprised of data collation, gathering overall knowledge from available published resources. That information, reports and publications on similar matters, were then analysed and cross-referenced.

Online survey

The purpose of the survey was to gather overall information about the practices for digital archiving from a broad range of stakeholders.

Method

- **Survey type.** Quantitative survey via an online questionnaire with a mix of question types
 - Yes/No questions
 - Multiple choice and comment
 - Choose from list (drop-down)
 - Essay box questions
- **Media.** Online survey using SurveyMonkey. Survey invitation sent out to numerous stakeholders via e-mail.
- **Period.** The initial survey period was from 02-20 April 2014, which was later extended to the beginning of May.
- **Documentation.** Survey question sets and survey results can be found online <http://www.cdpa.co.uk/EARK/showquestions.php?Group=All>

Qualitative research is good at providing a breadth of information from a larger number of units, but if wanted to explore a topic in depth, quantitative methods can be too shallow. Hence we decided to use quantitative data collection on our survey in order to collect data from as many respondents as possible so we could achieve broad-based results. To distinguish best practice used worldwide we needed to go for in-depth qualitative techniques, in our case qualitative interviews. From collected answers we then chose eleven Organisations based on their given answers that interested us especially.

Stakeholders

The survey was sent out to broad range of stakeholders from all five stakeholder groups.

Questions for the survey

The questions for the survey were created considering the needs of each task. We used two level internal quality assurance to ensure that the questions were appropriate, understandable and covered all relevant topics for better end results. Each set of questions was reviewed by other task members in the cross-task group and finally all questions went through quality assurance by E-ARK partners outside our cross-task group.

The questions from the survey can be divided into four categories

1. General questions about background, legislation and contact information
2. Questions concerning pre-ingest, ingest and ingest tools
3. Questions about preserving archival information packages and file formats
4. Questions about requirements for access and current access solutions

Construction of the survey

There were 94 questions all together in the survey. However not all questions were put to every respondent. We created targeted questions depending on which stakeholder group the respondent belongs to. There was also a dynamic logic on given answers. For example if (Q.19) *Does your Organisation*

provide access to digital material? was answered “Yes” then the survey logic skipped (Q.20) *Why do you not provide access to assets?* and went straight to (Q.21) *Which specific content types do you currently provide access to?*. This was done to ensure that respondents only were asked relevant questions.

Note about libraries in the survey

The intention was to include libraries in the stakeholder group “Archives”. However, libraries responding to the survey generally identified themselves as belonging to the category “Other”. Since the survey was constructed with individual sets of questions targeted at each stakeholder group, the consequence was that libraries were not asked the right set of questions, and they were for example not asked questions about Archival Storage and Access

Qualitative interviews

The purpose of the qualitative in-depth interviews was to gather details about selected (interesting/significant) solutions used worldwide. The answers collected are used for description of best practices and as input for the onward work of WP3, WP4 and accordingly WP5 to create a common SIP and DIP specifications and ingest/access tool(s).

Method

Our method used in conducting qualitative interviews comprised elements from structured as well as semi structured interviews.

- **Interview type.** structured/semi-structured interview
- **Platform.** Media used for conducting the interviews was Skype
 - and face-to-face in the very few cases when it was possible
 - 4 persons (institutions) answered in writing to our qualitative interview questions
- **Interview period.** Interviews were held throughout May 2014. Interviews lasted on average one hour; the shortest interview was 45 minutes while longest was about 1h 15 minutes.
- Questions were sent to interviewees beforehand to enable them to familiarise themselves with and think about the questions before the interviews.
- Interviews held on Skype were recorded using MP3 Skype Recorder. A summary of the interview was written and sent to interviewees for verification afterwards. There were 3 interviewers’ roles in our interviews:
 - Person who asked questions. Interviewer's mission was to have a conversation with the respondent by asking key questions and other related questions. The exact set of questions depended on the responses of the respondent. The interviewer played a neutral role and didn't give his or her opinion in the interview process.
 - Person who took notes. The notes in written form were the primary source for the later analysis. The voice recordings were used for making sense of complicated answers if needed. It was allowed to ask additional questions if the answer was unclear or not detailed enough for the person taking notes.
 - Person who monitored and controlled the process. That person started, observed and closed the interview. S/he was encouraged to interrupt the interview whenever needed to gain and maintain the control over process. This person could also ask follow-up questions if something was left unclear or of particular interest, but the interrupting should not be consistent.

After a few interviews conducted with the three interviewer's roles it was discovered, that the same work can be done just as efficiently by two interviewers. So the tasks of a person monitoring the overall process of an interview were then divided by person taking notes and person asking most of questions.

In the qualitative interviews, the interviewees are given space and time to expand and elaborate their answers and experiences in a way that wasn't possible to do in the survey. Moreover, their answers are not pre-categorised in the interview.

Semi-structured interviewing is more flexible than standardised methods such as the structured survey. Although the interviewer in this technique will have some established topics for investigation, this method allows for the exploration of emergent themes and ideas rather than relying only on concepts and questions defined in advance of the interview. The interviewer would use a standardised interview guide with set questions which will be asked of all respondents. The questions tend to be asked in a similar order and format to make a form of comparison between answers possible. However, there is also scope for pursuing and probing for novel, relevant information, through additional questions often noted as prompts on the schedule. The interviewer frequently has to formulate impromptu questions in order to follow up leads that emerge during the interview.

We created internal and external interview guides to ensure that all relevant topics would be covered and to allow clarification and discussion about interesting aspects. We chose to make detailed internal interview guides with comprehensive questions. Because interviews are carried out in collaboration with T5.1, T3.1 and T4.1 and by making detailed interview guides we ensured that all relevant questions are asked even when persons from that task are not present. In the external interview guides we explained shortly the process of the interview and added also questions asked in the interview so that the interviewee can think about the answers and be prepared if needed.

Stakeholders for interview

We used representation and back-tracking for the identifying of stakeholders with best/good practices for the interviews

- Representation: we chose a representative cross section of stakeholders that
 - Come from different Organisation types (i.e. Archives, Vendors)
 - Hold different data types (both format types and structured/unstructured data)
 - Are subject to different legal requirements (e.g. retention periods, dispensations, confidentiality)
 - Use different strategies/methods (e.g. normalization of data on Ingest, on demand access, offline/online storage, emulation/migration)
 - Use different systems
- Back-tracking: We identified the stakeholders who provided us the most interesting answers in the quantitative survey and then chose them as interviewees for the qualitative interview. Each task has different interest and criteria for the selection of stakeholders, and as such not all interviews will be equally relevant for all tasks.

Interview questions

The questions for the interview were created also considering the needs of each task. We used two level internal quality assurance just like we did on creating survey questions in order to obtain better results.

Each set of questions was reviewed by other task members in our cross-task group, and finally all questions were gone through by members outside our cross-task group.

We carried on pilot interviews with the National Archives of Hungary, The Archives of the Republic of Slovenia, the National Archives of Norway and the Danish National Archives prior to other interviews to detect any possible problems that might occur. This was also to check to see if we fit in desired one-hour time-frame and make sure that all questions are well and unequivocally understood. Also the questions were amended based on feedback from the pilot interviews, and they were further refined iteratively throughout the whole interview process based on feedback from interviewees.

Interview guidelines

The following guidelines were developed to give the best possible conditions for interviews and ensure consistency.

General principles

- All potential respondents should be contacted prior to the interviews.
- All terms and rules should be introduced during the contact making process.
- All key questions should be sent beforehand.
- All privacy concerns should be regulated with the legal agreement.
- All prior information about the respondents and their current situation should be clear to all interviewers beforehand.

Questions

- The questions will be created prior to the interview.
- Open-ended questions will be allowed. However, when open-ended questions are used it is a good idea to have a list of topics that should be covered in the question to ensure that the necessary information is obtained.
- Questions will be grouped by respondent's type.
- The interviewer will ask each respondent's group the same set of key* questions.
- Ordering and phrasing of the key* questions will be kept consistent from interview to interview.

**All key questions should be easily identified in the questions list.*

Establishing the connection and recording the interviews

- Interviewers use Skype even if the respondents use telephone because of the agreed recording functionality and constant quality.
- All conversations will be recorded with the MP3 Skype Recorder tool. If the respondent rejects the recording agreement then the recording should not take a place.
- Recordings will not be shared with third parties.
- All recordings will be deleted at the latest by the end of 2014.
- Interviewers are aware of possible technical issues with the sound quality, microphone malfunctions, and a lag in the Internet connection speed and have a backup plan prepared in advance.

Things which should be avoided (based on QDATRAINING guidelines)

- Talking over the participant
- Interrupting the participant (not allowing the participant time to finish talking before asking the next question)
- Finishing sentences for the participant (putting words in their mouths)

- Asking more than one question at a time (very often, you will only get a response to the last one the participant heard)
- Asking narrow questions (framing the question too narrowly)
- Asking leading questions
- Filling up silences (not giving the participant time to think or expand) which is very common amongst less experienced (and also some very experienced) qualitative interviewers
- Not following the topic guide (not to be confused with not allowing emergent topics) or being consistent across and between interviews in relation to key topics from the topic guide which should have been drawn from the research question itself
- Not allowing interesting and emergent topics to be developed because of a rush to get to the next question or prompt
- Not being courteous enough
- Not having due cognisance where a power relationship exists between the interviewer and participant.
- Arguing with the participant
- Being judgemental
- Not signalling when the end of the interview is approaching allowing the participant to say anything they may have on their mind
- Fumbling with equipment and being unfamiliar with the equipment being used
- Failing to record the interview altogether
- Recording in a noisy and distracting environment (only limited control available to the researcher on this one but cognisance is important nevertheless where choices do exist)

Things do before the interview starts

- The leader will state “With the permission of the interviewee, this interview is being recorded for accuracy purposes only”.
- State that that interviewee will receive the written summary from the interview for reference and to correct any mistakes before it is used in the reports
- The leader will introduce the participants.

Appendix B: Survey Questions for Archives

Survey Questions for Archives

(Q.1) What type of Organisation do you represent?

(Q.2) In which country does your organisation reside?

(Q.3) What is your role/position within the Organisation?

(Q.4) How many persons in your organisation undertake work related to digital curation?

(Q.5) Please specify national legislation that regulates: Pre-ingest and ingest, Archival storage/preservation, Access service and Access restriction.

(Q.6) What acquisition strategy does your organisation employ for data from databases and Records Management Systems?

(Q.7) What is the size of your Organisations digital collection? (In TB)

(Q.8) What is the size of your Organisations digital collection? (number of assets)

(Q.9) What are the primary content types in your collection?

(Q.10) In what technical structure is your assets primarily stored?

(Q.11) What preservation strategy does your Organisation employ?

(Q.12) Do you currently follow any general rules or guidelines (e.g. data preparation guidelines, transfer recommendations, data validation rules) for pre-ingest, ingest or digital preservation?

(Q.13) Please, briefly describe the current workflow and provide a URL link.

(Q.14) Please, briefly describe the current workflow for pre-ingest, ingest or digital preservation.

(Q.15) What tools and services are currently used for (pre)ingest and active digital preservation?

(Q.16) Are there any details of information packages (SIP, AIP) formats used in your organisation or supported by your solution(s) available online?

(Q.17) Please, briefly describe the submission and archival information packages formats used in your organisation or supported by your solution(s) and provide a URL link.

(Q.18) Please, briefly describe the submission and archival information packages formats used in your organisation or supported by your solution(s).

(Q.19) Does your Organisation provide access to digital material?

(Q.20) Why do you not provide access to assets?

(Q.21) Which specific content types do you currently provide access to?

(Q.22) What other content types do you expect to provide access to in the next 10 years?

(Q.23) Do you use any software tools for data dissemination? This could be e.g. an access system, a DIP creation tool or other tools.

(Q.24) Do you use different software tools according to different technical and/or content types?

(Q.25) For each tool please describe the name, purpose, kind (proprietary, commercial, open source) and any other key features you wish to highlight.

(Q.26) What platform(s) do you use to provide access to data?

(Q.27) What kinds of metadata about your assets are accessible and searchable?

(Q.28) Do you allow metadata search across information packages?

(Q.29) Do you have specific format(s) for Dissemination Information Packets (DIP's)?

(Q.30) Do you have different dissemination formats depending on the type of content (e.g. formatted text, geodata, statistical data, etc.) and/or the technical structure (i.e. databases/not databases)?

(Q.31) Is there any publicly available information about your DIP format(s) e.g. descriptions, specifications, articles etc.

(Q.32) Do you use metadata standards for dissemination?

(Q.33) Which metadata standards do you use for dissemination?

(Q.34) Is access to your assets limited by any restrictions caused by e.g. copyright, Data protection acts, archival acts, etc.

(Q.35) What are the restrictions and how are they regulated?

(Q.36) Do you have any restrictions related to data mining?

(Q.37) What are the restrictions and how are they regulated?

(Q.38) How many requests do you serve on a yearly basis?

(Q.39) Who are the current users of your access services?

(Q.40) Have you studied your users' needs for access services or in other ways have knowledge of your users' needs?

(Q.41) Would you be willing to share this information with the E-ARK project?

(Q.42) If you wish to provide any further details about your access system or have references to publicly available material that can help the EARK project to understand your access system, please do so here.

(Q.93) Would you allow us to contact you at a later point in the project for an interview or other engaging activities?

(Q.94) Please provide contact information

Appendix C: Survey Questions for Service Providers

Survey Questions for Service Providers

(Q.1) What type of Organisation do you represent?

(Q.2) In which country does your organisation reside?

(Q.3) What is your role/position within the Organisation?

(Q.68) How many persons in your organisation undertake work related to information management?

(Q.69) Please specify national legislation that regulates: Pre-ingest and ingest, Archival storage/preservation, Access service and Access restriction

(Q.70) Which standards for electronic document and records management are being used in your organisation or supported by your electronic records management system?

(Q.71) Are any details of the export functions of the records management system(s) used in your organisation or provided by your company made available online?

(Q.72) Please, provide a URL link to the details of the export functions of the records management system(s) used or provided by your organisation.

(Q.73) Do you currently follow any general rules or guidelines (e.g. data preparation guidelines, transfer recommendations, data validation rules) for pre-ingest, ingest or digital preservation?

(Q.74) Please, briefly describe the guidelines, and provide a URL link if the document is available online.

(Q.75) What tools and services are currently used for (pre)ingest and active digital preservation?

(Q.76) Are there any details of information packages (SIP, AIP) formats used in your organisation or supported by your solution(s) available online?

(Q.77) Please, briefly describe the submission and archival information packages formats used in your organisation or supported by your solution(s) and provide a URL link if the document is available online.

(Q.78) Please, briefly describe the submission and archival information packages formats used in your organisation or supported by your solution(s).

(Q.79) Does your company run any digital curation or access services for archives or public sector agencies?

(Q.80) How many public sector clients (worldwide)?

(Q.81) Are your access services adjusted to individual clients?

(Q.82) What technical structure of data does your access service support?

(Q.83) Which specific content types does your access service support?

(Q.84) Does your access service use different software tools according to different technical and/or content types?

(Q.85) What platform(s) does your access service use to provide access to data?

(Q.86) Do you have a specific format for Dissemination Information Packets (DIP's)?

(Q.87) Do you have different dissemination formats depending on the type of content (e.g. Formatted text, geodata, video, etc.) and/or the technical structure (i.e. databases/not databases)?

(Q.88) Which metadata standards do you use for dissemination?

(Q.89) Is there any publicly available information about your DIP format(s) e.g. descriptions, specifications, articles etc.

(Q.90) Where can it be found?

(Q.91) Have you studied your users' needs for access services or in other ways have knowledge of your users' needs?

(Q.92) Would you be willing to share this information with the EARK project?

(Q.93) Would you allow us to contact you at a later point in the project for an interview or other engaging activities?

(Q.94) Please provide contact information

Appendix D: Assessment of stakeholders for interview from point of view of D5.1

Colour codes used in the schema:

Relevant for interview	Could be relevant for interview but deselected	Out of scope or not relevant for interview
------------------------	--	--

Schema for assessment of access services and identification of stakeholders for interview:

<i>Stakeholder</i>	<i>Organisation type</i>	<i>Acquisition strategy</i>	<i>Preservation strategy</i>	<i>Content types to which access is provided</i>	<i>Details about access service and users that make the stakeholder interesting</i>	<i>E-ARK partner</i>	<i>References</i>
STAKEHOLDERS IDENTIFIED BASED ON THE ONLINE SURVEY							
Estonian National Archives	Archive	Single records	Normalisation on ingest and migration	Textual data, images, Audio-visual data	Many different user groups including courts. The archives is subject to an interesting legal requirement where all material in principle must be immediately accessible online - See http://wiki.ra.ee/doku.php?id=estonia	x	Survey
National Archives of Hungary	Archive	Single records and whole systems	Normalisation on ingest and migration	Textual data, images, Audio-visual data, databases	Several platforms for access, uses many metadata standards for access and has a broad range of users including external services and courts. The Archives use services from Preservica and ScopeArchive. They have limited amount of born-digital material but have a newly developed access system that will be used to provide access to born-digital material. The National Archives of Hungary was used to test the interview methodology.	x	Survey
The National Archives UK	Archive	NA	Migration	Textual data, images, audio-visual data	Uses an access system developed in-house. They provide online access and broad user group. The online service "Discovery" allows users to search collections and submit requests for data that is not publicly available; buy downloads of data or view online (this includes snapshots of databases which is exported to excel spreadsheets, webpages and digitised material)		Survey
Bulgarian Archives State Agency	Archive	NA	NA	Textual data, images, databases	Have a SharePoint based system for acquisition, management and access to material. And an online service for searching collections and accessing material but it does not handle born-digital material. The services do not seem interesting enough to be elaborated on in an interview		Survey
Bergen city Archives	Archive	Whole systems	Migration	Textual data, images, databases	Have three services for access including a system developed in-house for databases. Uses services from fotoware.com and geomatikk-ikt.no.		Survey

Stakeholder	Organisation type	Acquisition strategy	Preservation strategy	Content types to which access is provided	Details about access service and users that make the stakeholder interesting	E-ARK partner	References
Consorti Administració Oberta de Catalunya	Archive	Single records	Migration	Textual data, images, audio-visual data	Interesting system where users can access and export data in different formats in the online access system. Primary users of the access system are the content creators. The access system itself seems interesting, but the system only handles simple data types.		Survey
Archaeology Data Service (York University)	Archive	Whole systems	Normalisation on ingest and migration	NA	Access is provided to archaeological data, geo-tagged and related to pictures. Assessed to be out of scope for this work.		Survey
Stanford Digital Repository	Archive	NA	Normalisation on ingest	NA	Libraries are out of scope for this work.		Survey
Danish National Archives	Archive	Whole systems	Normalisation on ingest and migration	Digitised material, databases with preservation formats for text, sound, video and geodata	In-house developed access system, no metadata standards used for dissemination, broad user group ranging from general public to records creators. Users are Researchers, general public, content creators and internal users. Have well established practice for providing access to born-digital records, tool for access to born-digital material online and in reading rooms/on site	x	Survey
KEEPS	Service provider	NA	NA	The service supports many different content types	Run services for several archives, the services are adjusted to individual client	x	Survey
Arkivum	Service provider	NA	NA	Clients determine content and AIP format	Do not provide access to end users but only access to AIPs (to data owners). The content owners themselves are responsible for the access services.		Survey
Preservica	Service provider	NA	NA	NA	Runs service for many archives, the service is not adjusted to clients' needs. The service handles many data types, but not databases. Use different DIP formats depending on content type. Service widely used at National Archives which is seen from the survey.		Survey
Scope Archive	Service provider	NA	NA	Supports a wide range of content types including complex data, survey data, scientific and statistical data	Run services for many archives. The services are adjusted to clients' needs. Their services are widely used at archives, which is also seen from the survey.		Survey
ADDITIONAL STAKEHOLDERS IDENTIFIED BASED ON E-ARK KNOWLEDGE AND DESKTOP RESEARCH							
Danish Data Archive	Archive	NA	NA	Research data, survey data,	Uses the DDI-L standard which is widely used in Data archives and participates in CESSDA collaboration. The archive is considered to be representative for data archives that uses DDI-L for preservation and access		http://samfund.dda.dk/da/default-en.asp and http://samfund.dda.dk/da/default-en.asp
National Archives of Norway	Archive	NA	NA	NA	Provides access to born-digital material in reading rooms. National Archives of Norway was used to test the interview methodology.	x	http://www.arkivverket.no/eng/Using-the-Archives/Acc

<i>Stakeholder</i>	<i>Organisation type</i>	<i>Acquisition strategy</i>	<i>Preservation strategy</i>	<i>Content types to which access is provided</i>	<i>Details about access service and users that make the stakeholder interesting</i>	<i>E-ARK partner</i>	<i>References</i>
							ess-to-archival-material/How-to-gain-access
National Archives of Sweden	Archive	NA	NA	NA	Interesting legal requirements for access. Have a relatively large collection of born-digital material which originates in the 1960s.	x	http://riksarkivet.se/handla-bestall
National Archives Slovenia	Archive	NA	NA	NA	Interesting workflow where a test DIP is created under SIP creation to allow assessing if the data will be meaningful and usable for access purposes and the SIP is amended accordingly to improve usability. The National Archives of Slovenia was used to test the interview methodology.	x	http://www.arhiv.gov.si/en/use_of_archival_records/
German Federal Archives	Archive	Single records	Migration	Images, databases (born-digital)	No tools for access and it seems as though they do not provide access to born-digital material, have online finding aids to search metadata about collections. They do not seem to have significant solutions from the access point of view.		http://www.bundesarchiv.de/index.html
Swiss National Archives	Archive	NA	NA	NA	Have a practice for providing access to digital material. Have three different ways of providing access to data: Standardised DIPs, non-standardised DIPs, and databases. Each data-type uses different tools. For standardised DIPs they have developed a tool which allows end-users to consult data and see content, metadata in a DIP.		http://www.bar.admin.ch/archivgut/01675/index.html?lang=en http://www.bar.admin.ch/dienstleistungen/00823/01559/index.html?lang=en
ES Solutions	Service provider	NA	NA	NA	ESSArch access is an interface to consumers, workflow in 9 steps in described. ES Solutions are widely used in Scandinavian countries. Although the service seems interesting the available material did not give reason enough to prioritize interview.	x	http://www.essolutions.se/ESSArch
Archivematica	Service provider	NA	NA	Supports many different content types including vector, email, audio, video, images, text	Open source software that supports the entire digital preservation process. Archivematica is integrated with the access system Atom.		https://www.archivematica.org/wiki/Main_Page

Appendix E: Interview questions for Archives

The (pre-)ingest of digital objects

1. Steps in pre-ingest process

- Please describe the usual negotiation process between producer and archive.
- Please describe the usual records export process and procedures at agencies of what your archive is aware of.

2. Steps in ingest process

- Could you briefly describe your usual workflow for digital archiving (including pre-ingest steps)?
- Could you briefly describe any other more complicated workflows you use in your institution?

The processing and storage of digital objects

1. Maintenance of AIP

- Please explain how your AIPs are stored: what kind of logical and physical containers do you use?
- How are your AIPs preserved over time, which strategies do you apply?
- How do you ensure authenticity (in a legal context) for your stored data?

2. Access to AIP

- Do you keep track of every access that has been made to a specific AIP while it is in storage (e.g. who accessed it, when etc.)?
- How do you handle restricted access to certain data (and thus to AIPs)?

The accessing of digital objects

1. Data and creation of DIPs

- What are the typical steps in your workflow when providing access to data?
- What happens to the DIPs after use?
- Could you briefly describe the information packages you use in your institution?

2. Dissemination and access

- Which tools do you use for providing access to your collections?
- How can users search your collections and find out what data he/she needs? (In other words: how can users find the correct DIP(s))
- How can the content of one or more DIPs be searched?
- How can disseminated data be used by users?
- What access restrictions and requirements must your access service comply with?
- How does your system handle confidentiality, retention dates, dispensations, user identification/authorization etc.?

3. Users

- What are the most typical use-cases for your access services?
- What do you know about your end-users' needs?
- How user friendly is your access system in your opinion?

4. General

- What would you say are the biggest advantages/weaknesses of your access service?
- What kind of access would you like to offer but are not capable of offering currently?

Appendix F: Interview questions for Service Providers

Ingest process

- How does your solution support negotiation process between producer and archives?
- Could you briefly describe your customers usual workflow for digital archiving (including supported pre-ingest steps)?
- Could you briefly describe any other more complicated workflows what are supported by your solution?

The processing and storage of digital objects / maintenance of AIPs

- Please explain how your AIPs are stored. What kind of physical containers do you recommend?
- Please explain the logical structure of data stored by your system.
- How are your AIPs preserved over time, which strategies can be applied by your solution?
- How do you ensure authenticity in your system?
- Please explain how and on what circumstances your system creates DIPs from AIPs?
- Does your solution keep track of every access that has been made to a specific AIP while it is in storage (e.g. who accessed it, when etc.)?
- How does your solution handle restricted access to certain data (and thus to AIPs)?

Access to stored data / access service details

- What are the typical steps in the workflow when providing access to data using your system?
- What typically happens to DIPs after use?
- Are your access service adjusted to your clients' local conditions?
- What functionalities does your access system have? (if possible you are very welcome to support your answer with snapshots of the interfaces in your access system?)
- How users (e.g. a researcher) search collections for the purpose of identifying which IPs contain the specific information he/she wants?
- How can content in one or more DIPs be searched?
- How does your system handle confidentiality, retention dates, dispensations, user identification/authorization etc.?
- Do you have any knowledge of how end-users typically use your access services?
- What do you know about the needs of the end-users of the access service?
- How user friendly is your access system to end-users in your opinion?

General

- What would you say are the biggest advantages/weaknesses of your access system?

Appendix G: Terminology

Access Functional Entity	The <i>OAIS</i> 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 <i>Archival records</i> .
Access Services or Solutions	Services and solutions developed in order to underpin the processes that give <i>Consumers</i> access to <i>Archival records</i> .
Access Software or tools	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	<i>OAIS</i> : 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 <i>OAIS</i> .
Archival records	A document whose long term value justifies its permanent retention.
Archive	An Organisation that intends to preserve information for Access and use by a <i>Designated Community</i> .
Asset types	This refers to different content or data types, e.g. geo-data, spreadsheets, image files.
Consumer	The role played by those persons or client systems, which interact with <i>OAIS</i> services to find preserved information of interest and to access that information in detail. This can include other <i>OAISs</i> , as well as internal <i>OAIS</i> persons or systems.
Descriptive metadata	Metadata that describes the data content.
Digital material	The term used to describe the digital assets of an archive, contained in <i>Information Packages</i> .
Digital Object	An object composed of a set of bit sequences.
Dissemination Information Package (DIP)	<i>Dissemination Information Package</i> , an <i>Information Package</i> , derived from one or more <i>AIPs</i> , and sent by <i>Archives</i> to the <i>Consumer</i> in response to a request to the <i>OAIS</i> .
Electronic Documents and Records Management System (EDRMS)	Is a type of content management system and refers to the combined technologies of document management and records management systems as an integrated system.
Finding Aid	A type of Access Aid that allows a user to search for and identify <i>Information Packages</i> of interest.
Information Package	A logical container composed of optional <i>Content Information</i> and optional associated <i>Preservation Description Information</i> . Associated with this <i>Information Package</i> is <i>Packaging Information</i> used to delimit and identify the <i>Content Information</i> and <i>Package Description</i> information used to facilitate searches for the <i>Content Information</i>
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 <i>Designated Community</i> .
Presentation tool	The tool and functionalities that provide access to <i>Archival records</i> .
Producer	The role played by those persons or client systems that provide the information to be preserved. This can include other <i>OAISs</i> or internal <i>OAIS</i> persons or systems.

Representation Information	The information that maps a <i>Data Object</i> into more meaningful concepts. An example is JPEG software which is used to render a JPEG file; rendering the JPEG file as bits is not very meaningful to humans but the software, which embodies an understanding of the JPEG standard, maps the bits into pixels which can then be rendered as an image for human viewing.
Service providers	Companies providing services to archives ranging from developing software to performing services
Submission Information Package (SIP)	An <i>Information Package</i> that is delivered by the <i>Producer</i> to the <i>OAIS</i> for use in the construction or update of one or more <i>AIPs</i> and/or the associated <i>Descriptive Information</i> .