D.3.1 Best Practice: SIP specification, records export requirements, transfer and ingest.

DOI: 10.5281/zenodo.1172506

Grant Agreement Number:	620998
Project Title:	European Archival Records and Knowledge Preservation
Release Date:	13th February 2018

Contributors	
Name	Affiliation
Tarvo Kärberg	National Archives of Estonia
Karin Oolu	National Archives of Estonia
Piret Randmäe	National Archives of Estonia
Kathrine Hougaard Edsen Johansen	Danish National Archive
Alex Thirifays	Danish National Archive
Boris Domajnko	National Archives of Slovenia
Janet Anderson	University of Brighton
David Anderson	University of Brighton
Sharon McMeekin	Digital Preservation Coalition
Jon Garde	DLM Forum
Kuldar Aas	National Archives of Estonia
Hans Fredrik Berg	National Archives of Norway
Björn Skog	ES Solutions
Henrik Ek	ES Solutions
Karin Bredenberg	ES Solutions

EXECUTIVE SUMMARY

The main objective of the deliverable D.3.1 (the best practice report) is to be used internally within the E-ARK project as an input for the E-ARK SIP specification, records export requirements, transfer and ingest recommendations.

The secondary target group is external – the archival institutions which collect digital data and organisations which provide the digital data to archives.

This report provides an overview of the current situation of the digital archiving best practices. Special attention is placed on archival ingest workflows, submission information package formats used for transfer and ingest of digital objects and their metadata. Records export best practices are covered as well.

The report consists of the following parts:

- introduction;
- description of the methods used for the analysis;
- overview of the results with short descriptions of practices, standards and tools;
- recommendations for the E-ARK project;
- appendices (the survey questions, an assessment of the interviewed stakeholders, the questions from the qualitative interview and a terminology list).

The study concentrates on the following topics from the archival workflow:

- Records export (Pre-Ingest workflow steps);
- Steps in Ingest workflow;
- Submission information packages (SIP) used.

Highlighted points of this best practice report for E-ARK work are:

- One high-level (pre-) ingest workflow is proposed in section 4 which consists of 4 phases of the PAIMAS methodology, but several existing workflow parts must be examined more deeply to include the common steps to the E-ARK archiving workflow;
- E-ARK needs to develop detailed and commonly understood requirements for the records export process which include procedures for data selection, extraction, metadata mapping, validation and quality control as these are currently lacking;
- One high-level SIP structure is proposed in section 4. (Recommendation for further work), but several existing SIP physical and logical structures must be examined more deeply to include the common aspects of formats used at archives into the E-ARK SIP specification.

Although, everything described in Chapter 4 is still preliminary and only high-level conceptual models are presented, work will continue and more specific specifications will be available in the coming years/future deliverables.

The authors of this report recognize the fact that the report contains some reservations:

- Many stakeholders mentioned the Open Archival Information System (OAIS) model. Although OAIS is well known by archival organisations and it is widely supported by many digital preservations tools (e.g. DSpace, LOCKSS), the practical implementations can vary a lot.
- Although this report is based on desktop research, online survey and interviews, the main focus was still on online survey and desktop research. The interviews were meant for acquiring complementary information.

This report is prepared on a request for information (RFI) level and therefore it does not provide very detailed modelling requirements for further work in E-ARK project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

The authors of this best practice report would like to thank all those who co-operated with the conduct of the survey and who allowed themselves to be interviewed. In order to preserve their anonymity no names can be made public but, without their participation, this survey could not have achieved its objectives.

1. INTRODUCTION

The comprehensive list of relevant tools and solutions[1] which was produced during the preparation phase of the E-ARK project proposal reflected that differences in digital preservation concerning the whole lifecycle, including how data is prepared and ingested into the archive; how they are stored and preserved in the archives; and how they are disseminated, accessed and used by end-users. Therefore, it was crucial for Work Package 3 to continue the work and look more precisely, especially at the pre-ingest and ingest stages in the scope of this work as seen in Figure 1.

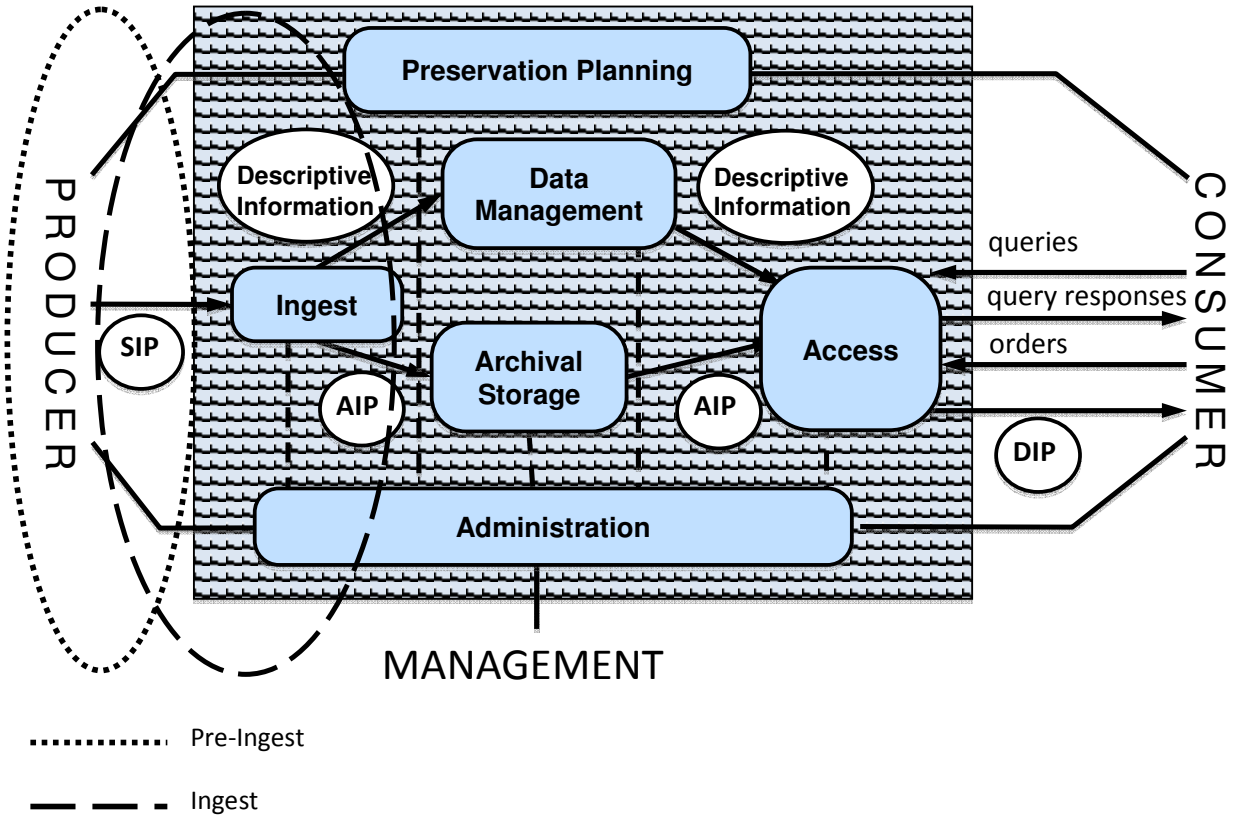


Figure 1: OAIS Functional Entities with Pre-Ingest

Work Packages 4 (Archival Records Preservation) and 5 (Archival Records Access Services) in the E-ARK project have similarly covered the part of existing archival and dissemination formats and services as their first mission. For complete overview it is important to look at reports from Work Packages 4 and 5 as well.

1.1 Structure of the deliverable

The current report is the outcome of the work carried out from February 2014 to July 2014 as part of Work Package 3 (Transfer of Records to Archives) in the E-ARK project.

The report contains:

[1] http://www.fpc.cdpa.org.uk/images/e-ark%20preservation%20tools.pdf

- Introduction;
- Methods (describes the methods used for information gathering for this report);
- Results (presents and describes the information gathered by several methods);
- Recommendations for further work in E-ARK project (concludes the report and gives recommendations to the E-ARK project);
- Appendixes.

1.2 Target group of the deliverable

The report is mainly important for E-ARK partners as it will feed into the onward work of E-ARK. In particular, creating model requirements for records export (T3.1), specifying a common SIP format (T3.2) and specifying the recommendations (T3.3) will benefit from and use the results. These tasks will be documented in further deliverables D3.2 E-ARK SIP draft specification, D3.3 E-ARK SIP pilot specification and D3.4 Records export, transfer and ingest recommendations and SIP Creation Tools. This deliverable is also part of Milestone MS01 "Best practice overview" that will combine information about best practices for *Ingest*, *Archival Storage* and *Access*[2] which are identified by Work Packages 3, 4 and 5 respectively and presented individually in Deliverables D3.1, D4.1 and D5.1.

The secondary goal of this report is to also inform the wider public, especially specialists in the digital preservation field, about best practices in the area.

1.3 Objectives of the deliverable

The purpose of this report is to get an overview of how information is exported from source systems, prepared for transfer and ingested into archival repositories.

The objective of this work is to feed collected information into the E-ARK project to specify common submission information package format(s), pre-ingest and ingest workflow with supporting tools. This means that this report provides valuable input to all three tasks (Records export requirements, EARK-SIP Specification, SIP Creation Tools) in Work Package 3 (Transfer of Records to Archives).

The report gives an overview of the activities performed during the process of gathering best practice about digital archiving on a RFI (request for information) level.

Note: All answers gathered from the online survey and published in this report have been anonymised – as no information provided by the respondents can be publicly attributed to their institution.

Information published in the interviews section has been freely available online or the interviewees have agreed to publish it in this report.

[2] It is assumed that the reader is familiar with the OAIS as terms from that model are used in this report.

2. METHODS

2.1 General approach

Work packages 3 (Transfer of Records to Archives), 4 (Archival Records Preservation) and 5 (Archival Records Access Services) in E-ARK project formed a cross-task collaborating group to analyse current solutions and best practices for *Ingest*, *Archival Storage* and *Access*. 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 of the E-ARK project asking for details about their digital archiving practices.

We conducted our work through

- **Desktop research.** Identifying what relevant information is already available and what can be used in further work in E-ARK project;
- **Online survey** sent to a wide range of stakeholders. Gathering information worldwide across multiple stakeholder groups.
- **Series of qualitative interviews** with selected stakeholders. Gathering more detailed information about relevant solutions from a smaller number of chosen stakeholders.

We gathered information 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 submission tools;
- **Archival Storage.** Available formats and restrictions for storage and different national requirements for authentication for legal purposes (documented in D4.1);
- **Access.** GAPs between requirements for access and current access solutions (documented in D5.1).

2.2 Desktop research

The purpose of the desktop research was to get overall knowledge about current (pre-) ingest practices and solutions.

We began with desktop research as an initial stage of our task. 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.

Results of the desktop research can be seen in section 3.2.

Then the work continued with the online survey.

2.3 Survey

The survey method was chosen as the main 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 addressed five stakeholder groups:

- Archives - As in many countries the records management passive phase and archiving principles are regulated or guided by the archives, the WP3 considered archives as the main target group for gathering information for this report.
- Public Organisations / Government Organisations - creators of digital content (Producers) and regulatory bodies.
- Private Companies / Service Providers - This is the second main target group as service providers may have many clients and it is possible to get information about many clients at once.
- Private Organisations - refer to non-governmental or non-profit organisations. As Private Companies and Private Organisations are not mutually exclusive (by the definition), it was taken into account already at the beginning that there may be only few respondents in one of the groups as respondents may get confused finding the right group. As the work group declared that for themselves at the beginning of the survey, they considered it also in the analysis phase.
- Projects - projects that have developed archiving services (in case if the cross-task group has missed some relevant project or study during desktop research).

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 members of other tasks 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

There were 94 questions all together in the survey. However not all questions were asked every respondent. We created targeted questions depending on which stakeholder group the respondent belongs to. There was also dynamic skip logic[3] 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.

The full set of survey questions for target groups can be found in appendixes on pages 67 - 71. Questions directly relevant for this report are questions:

- 5, 6, 12-18 for Archives (Appendix B: Survey Questions for Archives),

[3] Skip logic is a feature that changes what question or page a respondent sees next based on how they answer the current question. Also known as "conditional branching" or "branch logic," skip logic creates a custom path through the survey that varies based on a respondent's answers. This skip pattern will vary based on rules that you define for the respondent (https://www.surveymonkey.com/mp/tour/skiplogic/).

- 58-67 for Government Bodies (Appendix D: Survey Questions for Government Bodies),
- 44-53 for Private Organisations (Appendix E: Survey Questions for Private Organisations),
- 69-78 for Private Companies / Service Providers (Appendix C: Survey Questions for Private Companies / Service Providers),
- 55-56 for Projects (Appendix F: Survey Questions for Projects).

Construction of survey

- **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

  Survey Monkey's skip logic was used.
- **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.

Quantitative research is good at providing information at a general level, from a larger number of units, but for exploring a topic in depth, quantitative methods can be too shallow. Therefore we continued with the qualitative interviews.

2.4 Interviews

Following the online survey, a series of qualitative interviews were carried out with selected stakeholders to gather detailed information about significant and interesting ingest 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 selected stakeholders' practices and services.

Semi-structured interviewing is more flexible than standardised methods such as the structured survey.[4] 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[5] with set questions to be asked of all respondents. The questions tend to be asked in a similar order and format to allow a form of comparison between answers. 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.

[4] In qualitative interviews the interviewees are given space and time to expand and elaborate their answers and experiences that was not possible to do in the survey.
[5] A joint description of the guidelines for the interviews is located on page 60.

Acknowledging that not all potential relevant stakeholders necessarily participated in the survey, we additionally conducted desktop research to make sure that no significant stakeholders were overlooked just because they did not respond to the survey.

Stakeholders for interview

We used representation and back-tracking for 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). It was considered not to make interviews directly with the Data Providers as they were already represented in the survey and it would have been difficult to get contact with representative amount of Producers and interview them in the scope of this work. It was considered logical to collect information about pre-ingest and ingest only from Archives and Vendors as they are mainly controlling the data preparation principles and archiving processes;
    - 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);
    - Come from different geographical regions (still mainly focused on Europe);
    - 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 selection of stakeholders, and as such, not all interviews will be equally relevant for all tasks.

The detailed schema used for identifying the potential stakeholders is located on page 73. The list of interviewees can be seen on page 39.

Interview questions

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 *Ingest* but also the *Archival Storage* and the *Access* functional entities. The questions directly related to this report cover Pre-Ingest and Ingest workflows and SIP formats used as seen in Figure 2.

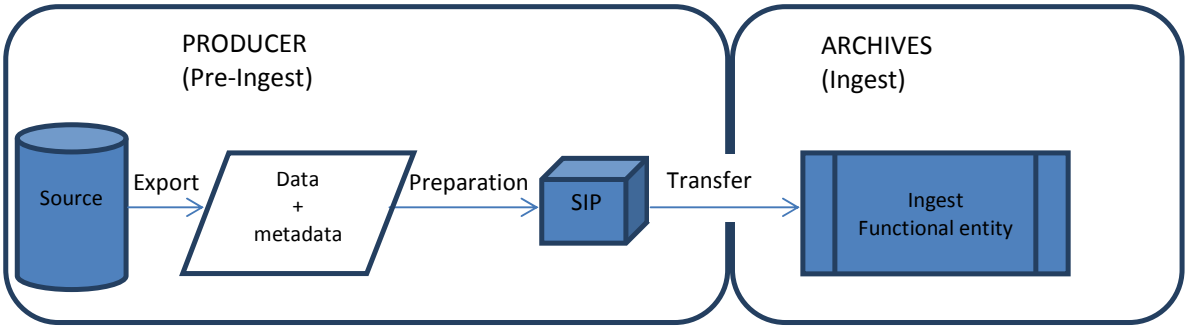


Figure 2: Interview questions coverage

The full set of interview questions are placed in appendixes (Appendix I: Interview questions for Archives; Appendix J: Interview questions for Service Providers).

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 for better results. Each set of questions was reviewed by members of other tasks in our cross-task group and finally all questions were gone through by selected members outside our cross-task group.

We carried out pilot interviews with members of the cross-task group – National Archives of Hungary, The Archives of the Republic of Slovenia, National Archives of Norway and the Danish National Archives prior to other interviews to detect any possible problems that might occur, to see if we fit into the desired one hour time-frame, and to make sure that all questions are well and univocally understood. 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.

Construction of the interviews

Our method used in qualitative interviews comprised elements from semi structured interviews. 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 were carried out in collaboration with other work packages and by making detailed interview guides we ensured that all relevant questions were asked even when persons from that task are not present. In external interview guides, which we sent out to the interviewees in advance, 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.

- **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.

- 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 did not 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 in complicated answers if needed. It was allowed to ask additional questions if the answer was unclear or not detailed enough by the person taking notes.
    - Person who monitored and controlled the process. That person started, observed and closed the interview. He or she 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.

3. RESULTS

3.1 Desktop research

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

1. The study *"Digital Preservation Services: State of the Art Analysis"*[6] from 2012.
   Summary
   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 solutions and illustrates the lack of collaboration among different tools available for solving the same tasks.

[6] Ruusalepp, R. & Dobreva, M. (2012): *"Digital Preservation Services: State of the Art Analysis"* www.dc-net.org/getFile.php?id=467

E-ARK perspective

The study does not cover detailed comparison of features of all these tools and testing them in practice but gives a good overview how the identified tools and services can be grouped into a taxonomy based on stages of the digital archiving workflow. It may be useful when specifying detailed workflow steps in further work in the E-ARK project.

2. "Analysis of Current Digital Preservation Policies: Archives, Libraries and Museums"[7] from 2013.

Summary

The analysis searched for digital preservation policies, strategies or plans published on the Internet by cultural heritage institutions.

E-ARK perspective

The analysis identified a list of policies and made note of the creating body, the document's title, URL; and then grouped the policies into the following categories: archives, libraries, and museums. As the analysis does not go in to detail regarding policies it cannot be used in this report.

3. The study *"Common challenges, different strategies"*[8] from 2012.

Summary

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.

E-ARK perspective

The study gives an overview of what computer file formats are accepted in transfer by various archives, but it does not go in detail describing the SIP formats structure and logic.

4. A study from 2012 entitled "Database Archiving"[9] from 2012.

Summary

This study 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.

E-ARK perspective

[7] Sheldon, M. (2013): " Analysis of Current Digital Preservation Policies: Archives, Libraries and Museums" http://www.digitalpreservation.gov/documents/Analysis%20of%20Current%20Digital%20Preservation%20Policies.pdf?loclr=blogsig

[8] Kristmar, K. V. (2012): *"Common challenges, different strategies"*.

[9] Velle, K. (2012): *"Database Archiving", https://www.sa.dk/media(4588,1033)/EBNA-Minutes,_CPH_29-30_May_2012.pdf*

The study highlights Plain Text, CSV in combination with XML as submission formats and EAD, SIARD as metadata formats used by respondents in database archiving. This information could be used in E-ARK SIP defining process for structured data.

5. Analysis of digital documents in other national archives[10] from 2013

   Summary

   The survey focused on the following issues: current approaches to the analysis of digital documents, cooperation and projects in connection with the use of digital documents, trends and future challenges in the use of digital documents."

   E-ARK perspective

   The study has a strong focus on *Access* and the results are based primarily on the findings of the Internet research as the archive questionnaire generated few responses. There is no additional information for the current report.

6. DCH-RP Project / DCH-RP-Survey[11] from 2013

   Summary

   The survey presents the standards, best practices, and identifiers that are of interest for the Digital Cultural Heritage (DCH) sector

   E-ARK perspective

   The survey provides short descriptions and references to various types of important standards and discusses issues and challenges regarding these standards. It also states that practical tests made within DCH-RP project have shown that already developed e-infrastructures must be modified and/or improved in order to provide a “pan-European” solution for the DCH community. The survey confirms the need of the current report.

7. Survey on Digital Preservation, 2013[12]

   Summary

   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 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.

   E-ARK perspective

   The study is not detailed enough for E-ARK work regarding submission information packages, ingest workflows or records export.

[10] Swiss Federal Archives SFA, Historical Analysis Services (2013): ”Analysis of digital documents in other national archives”

[11] Justrell B., Toller E. (2013): ”Standards and interoperability best practice report” http://www.dch-rp.eu/getFile.php?id=165

[12] Bergin, M. B. (2013): “Sabbatical Report: Summary of Survey Results on Digital Preservation Practices at 148 Institutions” http://works.bepress.com/cgi/viewcontent.cgi?article=1012&context=meghan_banach

8. SCAPE survey on preservation monitoring[13] from 2014.
   Summary
   Its purpose is to understand digital preservation incidents, threats and opportunities which are relevant to organisations, and the ways they would like to detect them.
   E-ARK perspective
   The survey focus is on preservation watch systems, thus not providing detailed information about (pre-) ingest and submission information packages.

As these previous studies have shown, they are too high-level or have a different focus, so it is still a need to deepen the research to get a detailed overview of how the information is exported from the source systems, prepared for transfer, transferred and ingested into archival repositories. Therefore we continued with the online survey.

3.2 Survey

There were a total of 184 responses to the online survey. Not all respondents completed the whole survey, which means that the number of total respondents to questions varies. It is also important to note that survey did not reveal any other relevant project from WP3 point of view what has not been covered by the desktop research already in the previous stage.

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.

3.2.1 Respondents profiles

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 60 responses to the survey from the Archives, 31 from the Private Companies (Service Providers), 9 from the Private Organisations and 43 from the Public (Government) Organisations as seen in Figure 3.

[13] Faria L, Duretec K., Kulmukhametov A., Moldrup-Dalum P., Medjkoune L., Pop R., Barton S., Akbik A. (2014): "SCAPE survey on preservation monitoring" http://www.scape-project.eu/wp-content/uploads/2014/05/SCAPE_D12.2_KEEPS_V1.0.pdf

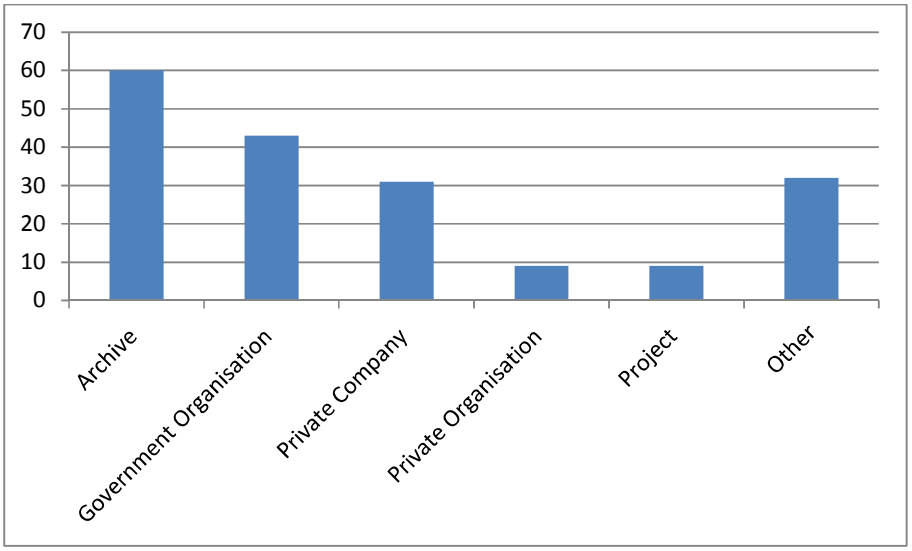


Figure 3: What type of organization do you represent?

After looking more closely the stakeholder group “Other” (as it contained many respondents) we identified that group “Other” includes 5 Government Organisations, 4 Private Companies, 11 Libraries, 11 Universities and one organisation which could be placed in to Archives group. Taking that information into account we made small correction to the profile and came up with the updated distribution as seen in Figure 4.

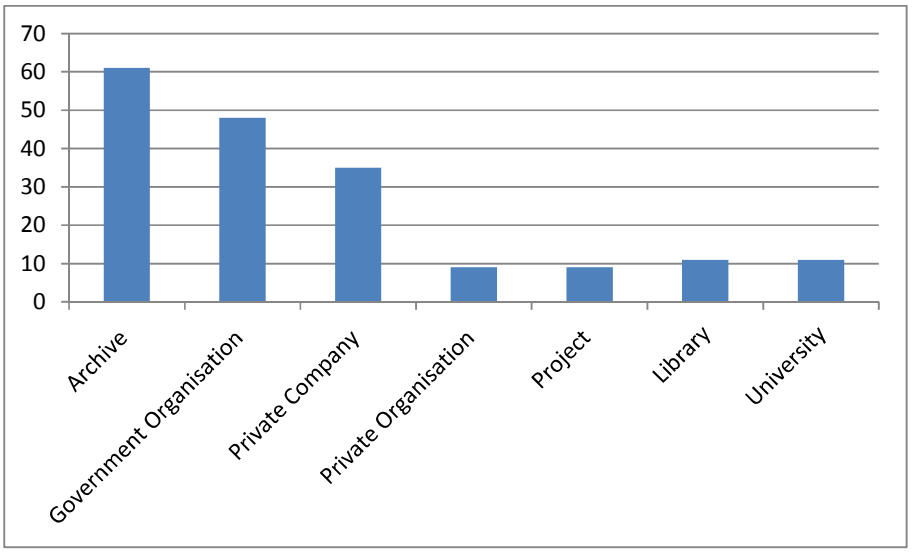


Figure 4: Updated question "What type of organization do you represent?”

Since the survey was constructed with individual sets of questions targeted at each stakeholder group, the consequence was that libraries were given a set of questions which was meant for group "Other". As traditional[14] libraries were not the target group for WP3, no relevant information got lost.

The survey was distributed widely and got responses from 32 countries. Most respondents came from the United Kingdom; 3 respondents did not reveal their country, but as they chose option "Other" then we can assume that they did not find suitable value from the list of countries as seen below:

Figure 5.

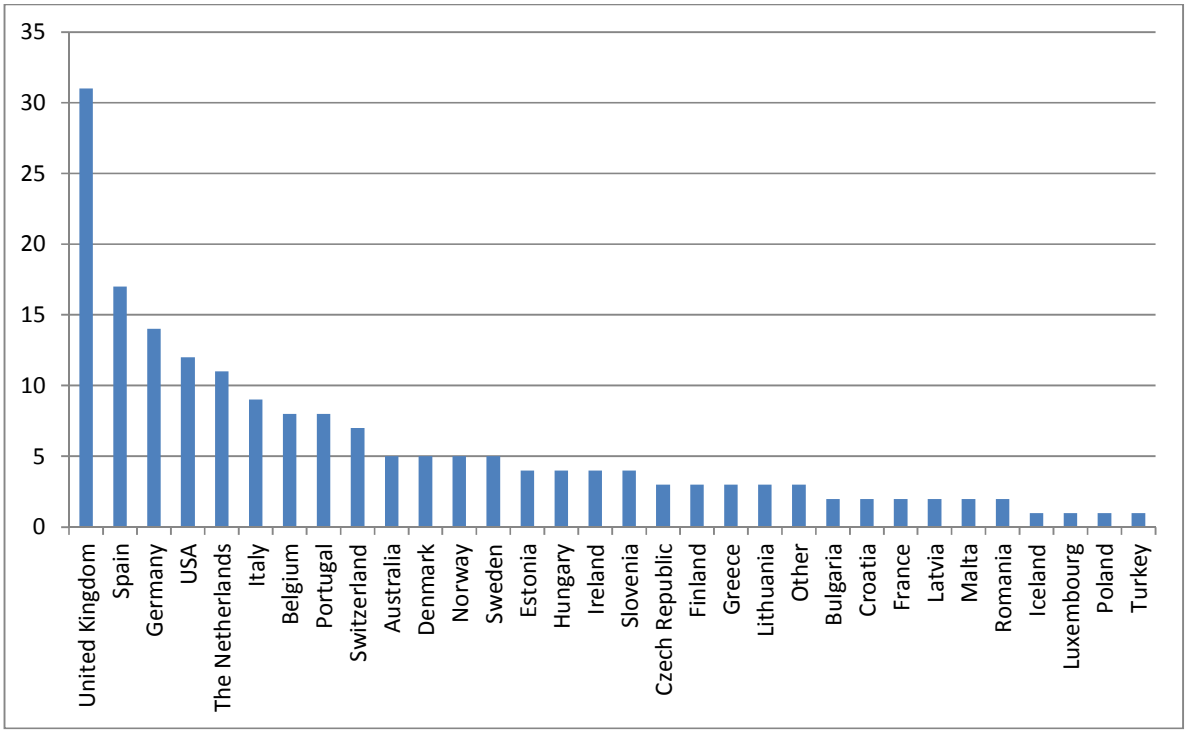


Figure 5: Distribution of respondents across countries

3.2.2 Archives

Questions directly relevant for archives in the context of this report were questions 5, 6, 12-18 (Appendix B: Survey Questions for Archives).

The answers to the question (Q5) about national legislation gave several links to the legislation specifically covering pre-ingest and ingest. Also OAIS (ISO 14721:2003) was mentioned.

Analysing the information gathered in this question in detail is meant to be a part of the legal analysis carried out in the E-ARK project, and therefore the further analysing of this question does not belong to the scope of this report.

[14] libraries that dont act as archives

The answers to the question (Q6) about acquisition strategy showed that both archiving as single records and whole systems are remarkably equally represented (see Figure 6).

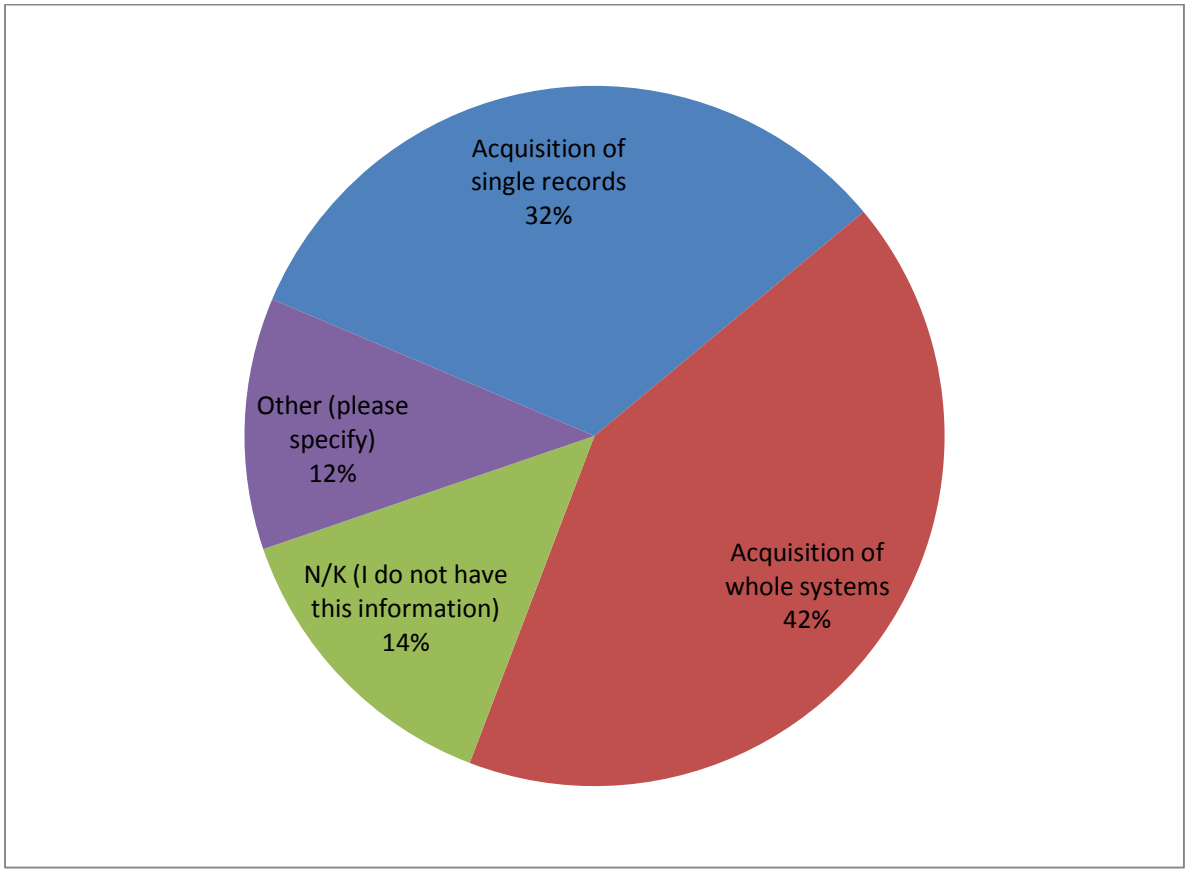


Figure 6: What acquisition strategy does your organisation employ for data from databases and Records Management Systems?[15]

The answers reflect also that some organisations employ both strategies for data archiving. "Other" means that the organisation does not currently ingest digital data or they have not explicitly answered what strategy they employ.[16] This supports the E-ARK approach that the archiving of electronic records is two-fold. While in some cases agencies and archives prefer to only archive single records along with their metadata, in other scenarios full systems (e.g. the bulk content of relational databases) are archived.

[15] The definition of single records and whole systems can be vague, but for the purpose of this work, acquisition of single records means that information is extracted from the source systems as records (with metadata) and acquisition of whole systems means that information is extracted as whole databases.
[16] As the results depend on the interpretation of the choice "Other" then the percent's of acquisition of single records and acquisition of whole systems may vary to a small extent (+/- 5%).

The answers to the question (Q12) about following any general rules or guidelines for pre-ingest, ingest or digital preservation, gave the result that 82% of respondents are following general rules or guidelines for pre-ingest, ingest or digital preservation.

Questions (Q13, Q14) about current ingest workflow gave 21 responses including a web link along with the description. According to the answers three are using / aim to use Tessella's SDB (now Preservica EE), two respondents claim to have OAIS compliant workflows, one is using Fedora, one is using ESSArch Tools.

Question (Q15) "What tools and services are currently used for (pre)ingest and active digital preservation?" showed that many different tools to support the workflow are used (see Table 1).

Table 1: List of tools

Workflow step	Name of the tool
Transfer to SIP	Elev SIP Creator; Preservica, MetsCreation; UAM; DRI; rsync*.
Ingest	DRI; Archaeology Data Service (ADS) Data Seal of Approval (after ADS DSOA); kleio; SFTP*, maior memorix.
Identification	DROID; maior memorix.
Normalisation	Preservica; METS, DRI, ADS DSOA; maior memorix; AdobePhotoshop; AdobePremiereCS5.5.2; Matrox MAX H264 Capture.
Characterisation	JHOVE (via kleio); Preservica; MODS; ADS DSOA; PRONOM; DROID.
Additional metadata	Preservica, DRI, UAM, MARC to MODS; maior.memorix; PIT+AIS; Adobe Bridge; EZID, gencat (Catalonian metadata); FTK Imager.

* file transfer protocol not specific to archiving

Answers to the question (Q16) "Are there any details of information packages (SIP, AIP) formats used in your organisation or supported by your solution(s) available online" show that 58% of the SIP or AIP descriptions are available online (including 5 respondents who answered "Yes", but who did not share the URL in response to the next question).

Questions (Q17) "Please, briefly describe the submission and archival information package formats used in your organisation or supported by your solution(s) and provide a URL link" and (Q18) "Please, briefly describe the submission and archival information packages formats used in your organisation or supported by your solution(s)" resulted in various SIP formats:

- METS;
- SDB XIP/Preservica;
- PREMIS;

- EAD;
- EAC;
- SIARD;
- XML
- PDF/A
- Windows folder
- Bagit.

When looking those answers and SIP formats more closely, we can see that most popular are METS (different variations) and Preservica/Tessella XIP.

3.2.3 Government Organisations

There were a total of 48 responses from Government Organisations. More than three responses per country came from Belgium, United Kingdom, Switzerland and Spain.

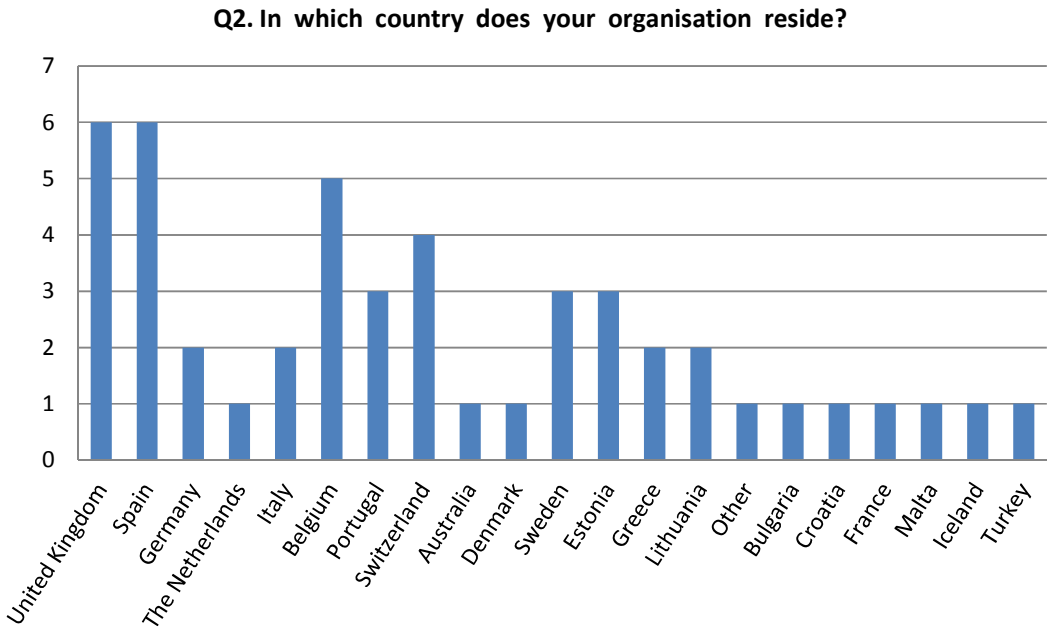


Figure 7: Distribution of respondents across countries

Not all respondents completed the whole survey, which means that the number of total respondents to questions varies.

The size of Government Organisations in terms of how many people are working in relation to information management varies. Please see the details from Figure 8.

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

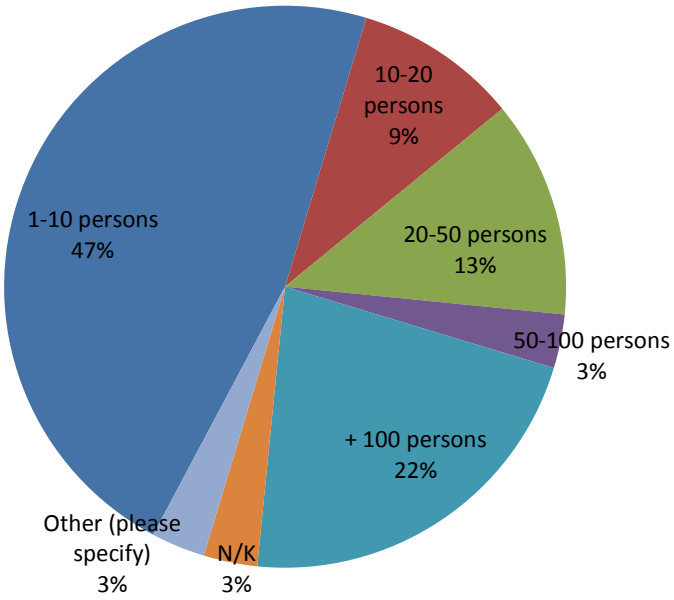


Figure 8: Size of Government Organisations

Questions directly relevant for Government Organisations in the context of this report were questions 58-67 (Appendix D: Survey Questions for Government Bodies).

Information of (Q58) **National legislation that regulates Pre-ingest and ingest** was impossible to select from answers which included also legislation of Archival storage/preservation and Access service and Access restriction (see Table 2). 4 respondents claimed there is no national general legislation or they were not sure. 18 responses provided many different acts or links; most of the legislation is in local languages:

Table 2: Legislation

Description	Country	URL
National Library Act 1968	Australia	
Copyright Act	Australia	
Archives Act	Estonia	https://www.riigiteataja.ee/en/eli/530102013053/consolide
Public Information Act	Estonia	https://www.riigiteataja.ee/en/eli/514112013001/consolide
Personal Data Protection Act	Estonia	https://www.riigiteataja.ee/en/eli/512112013011/consolide
Government regulation "Archival Rules"	Estonia	

(available in Estonian)		
AFNOR NF Z 42-013 for general electronic archival system	France	
AFNOR NF Z 42-020 for electronic safe deposit for archive	France	
RM AFNOR NF Z 44-022	France	
SEDA (Standard d'Echanges de Données pour les Archives) for exchange rules both in ingest and access parts "Livre II du code du patrimoine" : rules for archive in all public agencies	France	
Various others	France	http://www.archivesdefrance.culture.gouv.fr/archives-publiques/lois/
Legal deposit law for some of the material for all 3 functions	Denmark	
Commission Decision 2002/47/EC, ECSC, Euratom of 23 January 2002 amending its Rules of Procedure, annexing the provisions on document management (OJ L 21, 24.1.2002, p. 23)	Belgium	
Commission Decision 2004/563/EC, Euratom of 7 July 2004 amending its Rules of Procedure, annexing the Commission's provisions on electronic and digitised documents (OJ L 251, 27.7.2004, p. 9);	Belgium	
Implementing rules for Decision 2002/47/EC, ECSC, Euratom on document management and for Decision 2004/563/EC, Euratom on electronic and digitised documents (SEC(2009)1643, 30.11.2009), adopted by the Secretary-General, in agreement with the Directors-General of Personnel and Administration and of Informatics.	Belgium	
UK Public Records Act 1958	United Kingdom	http://www.nationalarchives.gov.uk/information-management/legislation/public-records-act.htm
Freedom of Information Act 2000	United Kingdom	http://www.legislation.gov.uk/ukpga/2000/36
Data Protection Act 1998	United Kingdom	http://www.legislation.gov.uk/ukpga/1998/29/contents
Environmental Information Regulations 2004	United Kingdom	http://www.legislation.gov.uk/uksi/2004/3391/contents/made
The Re-use of Public Sector Information Regulations 2005	United Kingdom	http://www.legislation.gov.uk/uksi/2005/1515/contents/made
	Croatia	www.kultura.hr
Arkivlagen	Sweden	
Arkivförordningen	Sweden	
Offentlighets- och sekretesslagen	Sweden	
Personuppgiftslagen	Sweden	
Skattedatabaslagen	Sweden	
Skattedatabasförordningen	Sweden	

The Freedom of the Press Act, which states the basic rights of the public to have access to public records (official documents) and also defines the term public record	Sweden	http://www.riksdagen.se/sv/Document-Lagar/Lagar/Svenskforfattningssamling/Tryckfrihetsforordning-19491_sfs-1949-105/?bet=1949:105
The Archives Act which defines the scope of activities that the SNA and the municipal archives are responsible for. As well as defining the goals of these "archival" activities.	Sweden	http://www.riksdagen.se/sv/Document-Lagar/Lagar/Svenskforfattningssamling/Arkivlag-1990782_sfs-1990-782/?bet=1990:782
The Archives Ordinance which mandates the SNAs right to regulate records management and archival activities at state public agencies. From procurement of Writing materials to storage facilities. Including all facets of Electronic public records. It also extends the definition of public record in the Freedom of the Press Act to specifically include any single data in a database.	Sweden	http://www.riksdagen.se/sv/Document-Lagar/Lagar/Svenskforfattningssamling/Arkivforordning-1991446_sfs-1991-446/
Regulations concerning access and secrecy, documentation of paper as well as electronic public records can be found in the Public Access to Information and Secrecy Act	Sweden	http://www.riksdagen.se/sv/Document-Lagar/Lagar/Svenskforfattningssamling/Offentlighets--och-sekretessla_sfs-2009-400/?bet=2009:400
The Personal Data Act is the Swedish implantation of the EU directive	Sweden	http://www.government.se/content/1/c6/01/55/42/b451922d.pdf
General regulations issued by the SNA include rules governing everything from creation of records to disposal of them or transfer to the SNA. They also cover such things as storage facilities, description of records and archives etc. All on a very general level that does not include any specifics regarding Electronic public records, but are applicable to them as well as paper records, sound recordings etc.	Sweden	http://www3.ra.se/ra-fs/ra-fs_1997-04.pdf
An addition concerning and especially applicable to the description of (electronic) public records	Sweden	http://www3.ra.se/ra-fs/ra-fs_1997-04.pdf
Specific regulations issued by the SNA concerning electronic public records	Sweden	http://www3.ra.se/ra-fs/ra-fs_2009-01.pdf and http://www3.ra.se/ra-fs/ra-fs_2009-01.pdf
General regulations concerning storage facilities	Sweden	http://www3.ra.se/ra-fs/ra-fs_2013-04.pdf
"Codice dei beni culturali e del paesaggio" legislative decree 42/2004, modified 2008 ;	Italy	
"Codice dell'amministrazione digitale" legislative decree 82/2005 modified 2010	Italy	
Justid manages a edepot.	The Netherlands	www.justid.nl
Legal deposit including ingest, preservation and access	Germany	http://www.gesetze-im-internet.de/dnbg/index.html

Most widely used (Q59) **standards for electronic document and records management** which are being used by Government Organisations are ISO15489-1, ISO23081-1 and Moreq2. There were 24 responses in total: please find details from Figure 9: EDRMS standards. The vertical axis shows, how many times the standard was marked, as there was possibility to mark several choices.

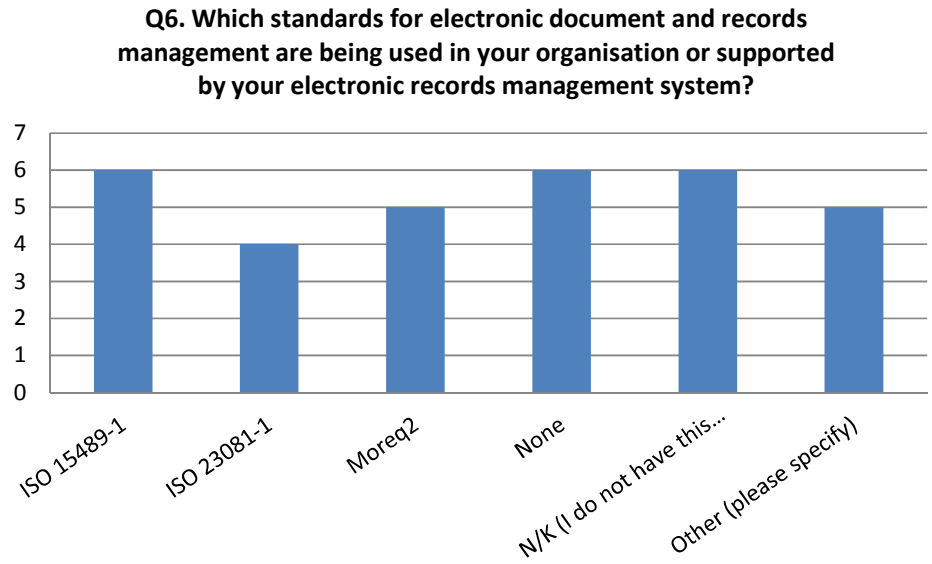


Figure 9: EDRMS standards

Other mentioned standards (by 5 respondents):[17]

- EAD/EAC
- SEDA/NF Z44-022
- MOREQ2010
- ICA-Req
- ISO 27001*
- ISO 24721
- OAIS
- MARC
- DIN 31644:2012-04

* an information security management system (ISMS) standard

69% of the (Q60, Q61) **details of the export functions of the records management system(s) used by the respondent's organisation** are not made available online (total of 29 responses). One organisation has

[17] In addition we know also about NOARK 4 and 5 which includes both records management and archival guidelines.

marked this information is confidential and one, that the export functions are not online yet. Seven respondents provided the next URL links for available online export functions:

- http://www.archivesdefrance.culture.gouv.fr/gerer/classement/normes-outils/
- https://www.aoc.cat/Inici/SERVEIS/Gestio-interna/iArxiu
- www.kultura.hr
- http://www.cuevapintada.org/imagenes
- www.carare.eu
- www.3dicons-project.eu
- http://www.africamuseum.be/collections
- http://sci-gems.math.bas.bg/jspui/

Previously listed links reveal the possible misunderstanding of the question. All provided links are describing different projects or portals for digital cultural heritage, with no connection to records management systems. The question was meant for describing the export workflows.

63% of Government Organisations (total answers 27) are currently following (Q62, Q63) **general rules or guidelines for pre-ingest, ingest or digital preservation.** 12 of the respondents provided either the link or the title of used guidelines, please see Table 3:

Table 3: General rules and guidelines

Description	Country	URL
The guidelines of the National Archives	Estonia	http://rahvusarhiiv.ra.ee/en/principles-standards-guidelines/
Evaluation of Electronic Archival System	France	http://www.archivesdefrance.culture.gouv.fr/static/7109
Standard d'Echange de Données pour l'Archivage (SEDA and recently NF Z 44-022)	France	http://www.boutique.afnor.org/norme/nf-z44-022/medona-modelisation-des-echanges-de-donnees-pour-l-archivage/article/814057/fa179927
Some directives for email archiving	France	http://www.archivesdefrance.culture.gouv.fr/static/2822 http://www.archivesdefrance.culture.gouv.fr/static/2823
Study "proof of concept" from VITAM project on email archiving	France	http://www.archivesdefrance.culture.gouv.fr/static/7140
References and "good practice" from Head IT for French government	France	http://references.modernisation.gouv.fr/archivage-numerique
The VITAM project aims to produce also some experiments and tools to enhance and facilitate both pre-ingest, ingest and access, while producing also the electronic archival core system. This project is at his beginning.	France	

	Spain	http://suport.aoc.cat/Portal/Tots-els-serveis/Integracio-serveis-Consorci-AOC
Condicions específiques de prestació del servei iARXIU	Spain	https://www.aoc.cat/content/download/13501/32409/file/Cond_espec%C3%ADfiques_iARXIU_amb_annexos.pdf
iArxiu: Estructura i creació de Paquets d'Informació de Transferència (PIT) utilitzant el model METS	Spain	https://www.aoc.cat/content/download/6657/24722/file/estructuraPitMets.pdf
Metadata guidelines, format guidelines	Croatia	http://www.kultura.hr/Sudjelujte/Preuzimanja-i-dokumenti
Guidelines and regulations issued by Parliament, Government and the National Archives ourselves	Sweden	
3D Icons	Spain	http://www.3dicons-project.eu/
EUROPEANA, Biodiversity Heritage Library, Global Biodiversity Information Facility , Biodiversity Information standards (TDWG)	Belgium	
	Bulgaria	http://sci-gems.math.bas.bg/jspui/handle/10525/2104/browse?type=dateissued&sort_by=2&order=DESC&rpp=20&etal=0&submit_browse=Update
Specific metadata profile special designed for permanent archival for governmental use	The Netherlands	http://www.nationaalarchief.nl/sites/default/files/docs/Toepassingsprofiel_metagegevens_rijksoverheid.pdf
PDF 1.4; PDF/A 1b	The Netherlands	
DIN 31645 ("Information und Dokumentation - Leitfaden zur Informationsübernahme in digitale Langzeitarchive"): A guidance for ingests in digital archival systems	Germany	http://www.dnb.de/EN/Netzpublikationen/Ablieferung/ablieferung_node.html

*Title added by author of this report

The question about (Q64) **tools and services, which are currently used for (pre-) ingest and active digital preservation** by Government Organisations showed how different approaches respondents might have. Answers are shown in the next list (Table 4):

Table 4: Tools and services

Workflow step*	URL
Destruction of data with no archival value or after the retention period is no longer valid	www.3dicons-project.eu/
Disposal of data with an archival value from the source system	www.3dicons-project.eu/
Transfer to SIP creating tool	www.3dicons-project.eu/
Transfer to archives' ingest module	Waarp or other transfer tools with secured and managed transfer system (http://waarp.github.io/Waarp/);

	www.3dicons-project.eu/; Electronic Messages Services; own developed tools
Identification	Droid (Pronom) for file format, www.3dicons-project.eu/; Adlib Express conversion, FITS
Normalisation	www.3dicons-project.eu/, PDF A, Format and metadata normalisation
Characterisation	FITS and its subordonates (JHOVE, EXIFTool, Droid), www.3dicons-project.eu/, Metadata, FITS (including JHOVE, DROID and other tools)
Additional metadata description	in the future: semantic analysis probably based on Apache Mahout, www.3dicons-project.eu/, Dublin Core plus some other info, Technical functional and own developed tools
Validation	www.3dicons-project.eu/, Adlibserver, FITS and own developed tools
Storage	www.3dicons-project.eu/, IBM DIAS (including Content Manager, TSM)

* categories what were defined as part of the survey question

79% of answered Government Organisations said there are NO details of (Q65) **information packages (SIP, AIP) formats** available online. Some organisations has provided the description (Q66, Q67) or URL of used submission and archival information packages formats as follows:

- Universal Object Format: http://kopal.langzeitarchivierung.de/downloads/kopal_Universal_Object_Format.pdf
- They are described in the reports from the project BHL-Europe, OPen up! Etc ...
- The submission is made from archive service or directly from IT service, depending on the ingest contract.
  The transfer protocol might vary according to the context (Waarp, FTP, USB, CDROM, ...).

  The SIP will be defined in the 2014 year by the VITAM project. Mainly it will be based first on a global ZIP or TAR to package all information. Then inside the SIP will be organized as follow:

  a) archive files themselves (binary format)
  b) transport XML file (close to SEDA/NF Z44-022) to list all files and their preliminary technical identification (uri, digest, size mainly) and some other general information (sender, contract id, submission id...)
  c) technical description metadata for each file in XML format (schema to finalize)
  d) business/archival description metadata and management metadata (life cycle, archive rights and rules) for each file according to a "DAG" (Directed Acyclic Graph or Multiple Trees representation, close to an extension of Moreq2010 model) (schema to finalize)
- Mainly PDF

Two respondents have claimed the submission and archival information packages formats are under development, and one it is not yet implemented.

If we compare the answers with the same question which was asked from Archives, we see that there are some differences (58% of Archives answered that the SIP or AIP descriptions are available online) which may mean that Archives must share the information more broadly to raise the awareness among Government Organisations.

The previous section described the results of the survey from Government Organisations concerning the process of pre-ingest and ingest. The results revealed there is a lot of national regulation, mostly in local languages, which regulates the fields of Pre-ingest and ingest, but also Archival storage/preservation and Access service and Access restrictions. Government Organisations use mostly ISO15489-1 standard for electronic document and records management, also different kinds of general guidelines. There is a need to emphasize the development of online access to details of information packages (SIP, AIP) formats.

3.2.4 Service Providers

There were 32 responses from Service Providers (Private Companies). There is a preponderance of respondents from Spain, USA, Germany and United Kingdom as seen from Figure 10.

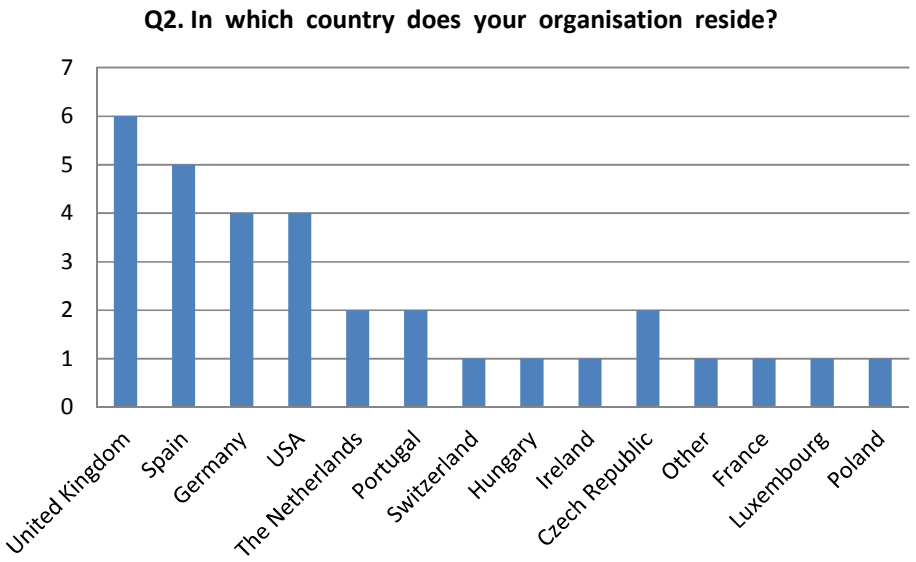


Figure 10: Distribution of respondents across countries

Not all respondents completed the whole survey, which means that the number of total respondents to questions varies.

The size of Service Providers in terms of how many people are working in relation to information management varies. There is an even distribution on sizes ranging from 1-20 persons to 100+ persons.

As (Q69) (**Please specify national legislation that regulates: Pre-ingest and ingest, Archival storage/preservation, Access service and Access restriction**) included also information about the archival storage and access part, and most of the answers were in local languages, it was not possible to deduce (without further analysis) what legislation regulates exactly the pre-ingest or ingest part. (see

Table 5). 16 responses provided different links and comments (open-ended responses);

Table 5: Legislation

Description	URL
US Government laws	

Only tax office regulations, usually handled through paper	
We provide a data archiving service to our customers. Regulations that they need to comply with include the Data Protection Act, ISO27001 information security, IL levels for government information e.g. IL2 or IL3, and in the case of healthcare/pharma there's FDA in the US, Eduralex in the EC, and UK regulations from MHRA. For example. MHRA guidelines on GCP and FDA 21 CFR part 11. The list is quite long. We'd be happy to provide more information and links if needed.	
Zákon 499/2004 Sb., archival and records management Vyhláška 259/2012 Sb., the details of Record Management Národní standard (VMV 64/2012), National standard for ERMS, including the definition of SIP and communication (XML) between ERMS and Archives Zákon 300/2008 Sb., electronic acts and authorized conversion of documents	
din tr-esor e-goc gestz	
Technical guidelines on long-term preservation of legal value of signed documents	https://www.bsi.bund.de/DE/Publikationen/TechnischeRichtlinien/tr03125/index_htm.html
PDF/A (ISO 19005) is an international standard that has been adopted by many members of the EU, as well as most countries in Latin America and Asia.	
National Archives Act 1986 applies Government records.	http://www.irishstatutebook.ie/1986/en/act/pub/0011/index.html
UK Data Protection Act 1998 Dutch Data Protection Act 2000	
Depends on sector (e.g., Public Records Act related only to public records). Other sectors might be regulated which might ultimately be backed up by legislation but the legislation won't specify details.	
Personal data protection law.	
Articles 16, 109, et 189 du Code de Commerce	http://www.legilux.public.lu/leg/textescoordonnes/codes/code_commerce/L1_du_commerce.pdf
Loi du 14 août 2000 sur le commerce électronique	http://eli.legilux.public.lu/eli/etat/leg/loi/2000/08/14/n8
Articles 1322-2, 1334, 1341, 1348 du code civil	
Règlement grand-ducal du 22 décembre 1986	http://www.legilux.public.lu/rgl/1986/A/2748/1.pdf
Loi du 5 avril 2003 sur le secteur financier.	http://eli.legilux.public.lu/eli/etat/leg/loi/1993/04/05/n1
National and various Cantonal archiving and records management laws	

One comment gave information about future plans: “A new legal framework for digital archiving is on the way (draft in French here: http://www.legilux.public.lu/ldp/2013/20130021_I.pdf, some inputs in English here: http://www.linklaters.com/Publications/Publication1403Newsletter/TMT-News-18-July-2013/Pages/Luxembourg-Draft-laws-encourage-paperless-offices.aspx), with technical requirements for

digitisation and electronic archiving provided by ILNAS (standardisation body of Luxembourg): http://www.ilnas.public.lu/fr/confiance-numerique/archivage-electronique/documents-obtention-statut-psdc/ilnas-technical-regulation-psdc-en-v1-3.pdf"

3 respondents claimed there is no national general legislation with a couple of longer comments: "Not aware of any national legislation for private companies with regards to archives other than general legislation such as Data Protection Act"; "There is not a national legislation as such; each archive is following international recommendations and internal procedures."

The most used (Q70) **standard for electronic document and records management** among Service Providers is ISO15489-1. Please see detailed info from Figure 11. Vertical axes shows how many times the standard was marked as there was possibility to mark several choices.

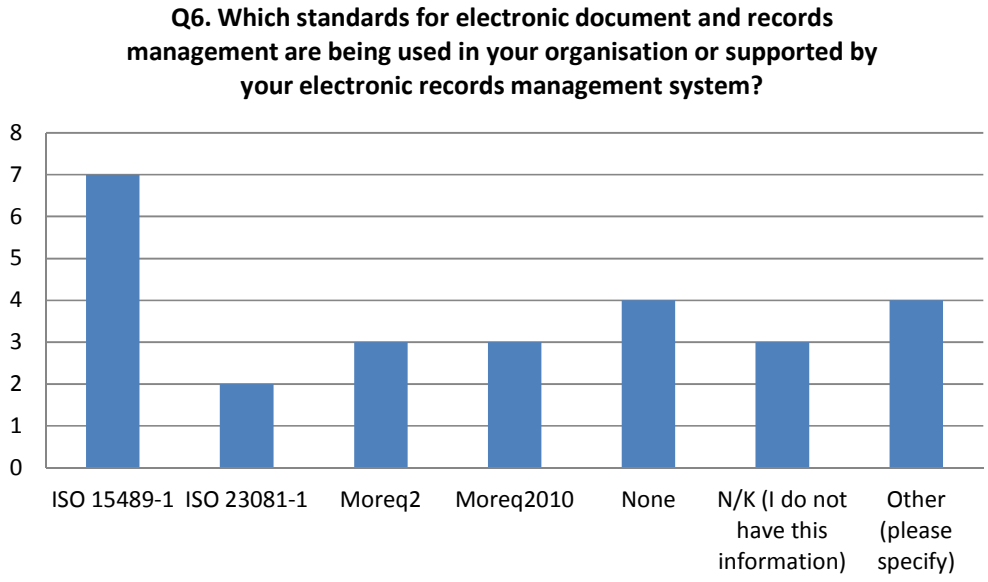


Figure 11: EDRMS standards

Other standards that were pointed out were:

- PDF/A - ISO 19005
- NF Z42013
- ISO 27001
- OAI-PMH
- NSESSS (Czech national derivate of MoReq2)

One respondent commented: "These are record management standards. We are more concerned with long term preservation (as in ISO 14721)."

65% of the (Q71, Q72) **details of the export functions of the records management system(s) used or provided by the respondent's company** are not made available online or not online yet. Only two companies have provided links of online export functions or details:

- http://www.scope.ch
- http://docs.oracle.com/cd/E28280_01/doc.1111/e26693/part4_record_mgmt.htm#CIHCDBGI

Other comments were:

- "Data escrow is part of our service"
- "Metadata that describes the files we store is included as part of escrow using XML data structures and BagIt from the Library of Congress"

Approximately half (58%) of Service Providers are following (Q73) **general rules or guidelines for pre-ingest, ingest or digital preservation**. Used guidelines (Q74) were commented on by 7 companies as following:

- We operate at the file storage/bit preservation level and make extensive use of checksums for data integrity validation. We follow the OAIS model where appropriate (e.g. we provide archive storage for AIPs) and we follow the applicable parts of ISO16363. General best practice includes multiple copies of data in multiple locations with active integrity management and regular technology/media migration to address obsolescence.
- As per customer guidelines and rules
- PREMIS
- Go through a number of steps to ensure quality assurance. Can include: Virus checking. Verification that metadata documents are compliant with stated schemas
- Documents must follow defined internal templates and define a set of mandatory metadata. Documents are confidential.
- BS10008 - Evidential weight and legal admissibility of electronic information
- OAIS

(Q75) **Tools and services, which are currently used for (pre-)ingest and active digital preservation** by Service Providers are shown in the next list (Table 6):

Table 6: Tools and services

Workflow step*	URL
Destruction of data with no archival value or after the retention period is no longer valid	In-house bespoke function with approval workflow; SharePoint; scopeOAIS
Disposal of data with an archival value from the source system	Part of ingest workflow (is possible). In fact, rare that this is possible; scopeOAIS
Transfer to SIP creating tool	Bespoke workflow or 'SIP Creator' tool, scopeOAIS
Transfer to archives' ingest module	Bespoke workflow or 'SIP Creator' ; scopeOAIS
Identification	File Investigator, http://fid3.com/products/fi-api; DROID; scopeOAIS
Normalisation	archivematica; Depends on format and target format. Lots of tools used; scopeOAIS
Characterisation	File Investigator, http://fid3.com/products/fi-api; Depends on format; scopeOAIS
Additional metadata description	XML metadata; Embed a schema (no tool used within system

	but might be outside); scopeOAIS; Adobe Bridge, Filework Pro; File Investigator, http://fid3.com/products/fi-api
Validation	File Investigator, http://fid3.com/products/fi-api; bagit and checksums; Depends on format; scopeOAIS
Storage	data tape and hard drives; A series of adaptors available to link to different storage systems with different storage structures; Windows fileserver...; scopeOAIS;
Other relevant	Bespoke workflow; scopeTKS

* categories what were defined as part of the survey question

62% of Service Providers said there are no details of (Q76) **information packages (SIP, AIP) formats** available online.

Submission and archival information packages formats (Q77, Q78) were described only by 5 (out of 9) Service Providers as follows:

- EAD, eCH-0160, METS, PREMIS, XBARCH, EDIAKT, XISADG
- We use XIP. We are likely to publish this more widely once we have new web site up.
- There are no formalized packages, submission is made by web form, archival package is specific of the used project and document management system used.

One organisation has provided the URL of used submission and archival information packages formats:

- Národní standard (VMV 64/2012) - National standard for ERMS incl. definition of SIP http://www.mvcr.cz/soubor/priklad-xml.aspx

The results of the Service Providers group are similar to previously described Government Organisations. In conclusion we can say there is a lot of National regulation, mostly in local languages, which regulates the fields of Pre-ingest and ingest, but also Archival storage/preservation and Access service and Access restrictions. The most used standard for electronic document and records management also ISO15489-1 and over than half of Service Providers are following different kinds of general guidelines. A lot of work is still ahead concerning online access, both with export functions of the records management system(s) or information packages (SIP, AIP) formats.

Private Organisations

Answers from Private Organisations (9) will be analysed in this chapter. According to the stakeholders definition on page 13 the Private Organisations will be analysed separately.

Questions directly relevant for Private Organisations in the context of this report were questions 44-53 (Appendix E: Survey Questions for Private Organisations).

There were only 2 sources of (Q44) **National Legislation that regulates Pre-ingest and ingest, Archival storage/preservation, Access service and Access restriction**. It is not clear whether respondents do not know or do not have any. A couple of comments were as follows:

- The German signature and eGoverment laws. For preservation we have a product according to the technical guideline from BSI TR-ESOR aka TR-03125

- DIN 31644, 31645,31646, 31647 (draft) - E-Government-act - Signaturgesetz (digital signatures) - BSI TR-03125 - BSI TR-RESISCAN - RFC 4998, RFC 6283

The most used (Q45) **standard for electronic document and records management** is ISO15489-1 (3 responses) and MoReq2 (2 responses) or its national derivate. Other mentioned standards are:

- OAI-PMH
- ISO 23081-1
- DOMEA new,
- ISO 303xx family,
- BSI-TR-03125, BSI-TR 03138,
- DIN 31644, DIN 3647 (draft)

There are 3 organisations who mentioned the **details of the export functions of the records management system(s) made available online** (Q46, Q47). Two of respondents have provided the next comments:

- Yes BSI - TR-03125 E+F XOEV-standard SAGA 5.0 https://www.bsi.bund.de/EN/Publications/TechnicalGuidelines/TR03125/BSITR03125.html
- http://fid3.com/products/fi-api

Currently followed general rules or guidelines (e.g. data preparation guidelines, transfer recommendations, data validation rules) for pre-ingest, ingest or digital preservation (Q48, Q49) are the following:

- https://www.bsi.bund.de/DE/Publikationen/TechnischeRichtlinien/tr03125/index_htm.html
- https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/TechGuidelines/TG03125/TG-03125AnnexTR-ESOR-F.pdf?__blob=publicationFile
- http://www.nabd.din.de/projekte/DIN+31647/de/117989686.html
- http://www.nabd.din.de/cmd?level=tpl-art-detailansicht&committeeid=54738855&artid=145158117&bcrumblevel=3&languageid=de
- ATHENA
- Dublin-core metadata guidelines http://dublincore.org/

Currently used tools and services for (pre) ingest and active digital preservation (Q50) were named as follows:

- https://www.governikus.com/de/governikus_lza/5952804, http://www.fujitsu.com/de/products/computing/storage/software/data-protection/backup-archiving/secdocs/
- Microsoft Access
- Adobe Bridge, Filework Pro
- File Investigator, http://fid3.com/products/fi-api

There are only 2 responses out of 8 who **marked the details of information packages (SIP, AIP) formats are available online** (Q51). The respondents commented on this as follows (Q52, Q53):

- XAIP see also https://www.bsi.bund.de/DE/Publikationen/TechnischeRichtlinien/tr03125/index_htm.html
- https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/TechGuidelines/TG03125/TG-03125AnnexTR-ESOR-F.pdf?__blob=publicationFile
- Digital files are edited, catalogued by nr/topic, archived at own server. Physical copies are all preserved.
- as defined by NSESSS

As the target group of Private Organisations was quite small (only 9 responses) it does not make sense to generalize the results widely. In spite of that, numerous links provided by respondents are valuable data and sources for further investigation. Most of results are similar to previously described stakeholders groups (Government Organisations and Service Providers).

Please look for full table of standards, guidelines and legislation used by stakeholders on page 73.

3.3 Interviews

13 stakeholders were invited to participate in the qualitative interviews. With 11 of these it was possible to conduct interviews. Table below (Table 7) shows the list of interviewed stakeholders. The detailed schema used for identifying the potential stakeholders is located in Appendix H: Assessment of stakeholders for interview from point of view of D3.1.

Table 7: List of interviewed stakeholders

Stakeholders invited to interview	Stakeholder type
The National Archives UK *	Archive
Estonian National Archives *	Archive
National Archives of Hungary	Archive
Swiss Federal Archives **	Archive
Danish Data Archive ******	Archive
National Archives of Norway	Archive
The Archives of the Republic of Slovenia	Archive
Danish National Archives	Archive
Archivematica	Service provider
KEEP Solutions *	Service provider
Preservica	Service provider
Scope Solutions **	Service provider
ESSArch Tools ***	Service Provider
Arkivum ******	Service Provider

** These stakeholders answered the interview questions in writing due to difficulties arranging an actual interview.*

*** These stakeholders were invited for an interview, but they are not included in the results due to difficulties getting contact or finding suitable time for the interview or writing answers.*

**** These stakeholders shared their product specifications, no additional information was needed – the need for an interview was cancelled.*

***** Not important for this work, but is relevant for the cross-task group.*

The interviews provided details about ingest 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 during interviews does not necessarily represent the broad landscape of ingest, but it complements the information gathered in previous steps. If some additional information was needed during the analysing phase, it was collected from the Web.

To achieve better regional coverage, some countries which were not in the respondents' list of the survey and which were not interviewed, were included into additional online research. Their information about SIP formats and ingest workflows is considered also in this report.

3.3.1 Archives

Hungarian National Archive

The SIP format at Hungarian National Archives is based on Tesella SDB (now named as Preservica) software.

The Hungarian National Archives uses OAIS compliant workflow which is assisted by Preservica.

According to the current regulations Hungarian National Archives makes regular inspection of creators and collect information about their records. If records are considered valuable the archives starts a negotiation process. During the negotiation process the archives defines the material to be submitted to the archive and determines the format and structure of and the day of transfer. The archives has an internal regulation about what can be transferred to the archive, but sometimes the producers have difficulties to meet these regulations, so the process is flexible.

When receiving material the first step is hash-sum checking and virus checking. Then content is put to quarantine for one month. After one month the content is virus checked again and the delivered metadata is validated. Subsequently a SIP is created from the content and metadata. After SIP creation archival metadata about the content is added and all metadata is validated. Then a characterization of the content is made using DROID. A manual, intellectual check of the content is also made to ensure that the content is the same as what was agreed in the negotiation process. Then an AIP is created and ingested to archival storage. The Hungarian National Archives also store a copy of the original data received from the producer.

Slovenian National Archive

The Slovenian National Archives has divided SIP formats into 3 categories according to data type:

- Computer files (metadata for each computer file can be optionally prepared in a separate XML file, XML schema is based on international standards and is extensible);

- ERMS data (custom built XML Schema, core set of PREMIS);
- Databases (in the future SIARD will be used).

The (pre-) ingest workflow used in the Slovenian National Archive includes:

- The process starts with deciding which types of data have an archival value. Deciding is usually based on a classification scheme.
- Then the form (digital or not) of the data will be identified.
- Then starts the evaluation process (what should be the SIP content, what procedures are going to be applied, etc). The list of steps depends very much on the data types (records /computer files/, ERDMS, databases). The outcome is a draft of the Submission agreement.
- Submission agreement - after signing the submission agreement, the preparation process can start.
- Preparing SIP – the producer can check the SIP.
- Transfer – the producer can send the SIP to archives.
- Validation – the archive makes the validation of the content (including technical). Ingest steps are specific (i.e. can include computer file migrations) and depend on the plan created before submission.
- Preliminary DIP creation - creating preliminary DIP (enables testing the use of the data and their validation).
- AIP generation.

Norwegian National Archive

The SIP structure used in Norwegian National Archive (NAN) is shown in Figure 12.

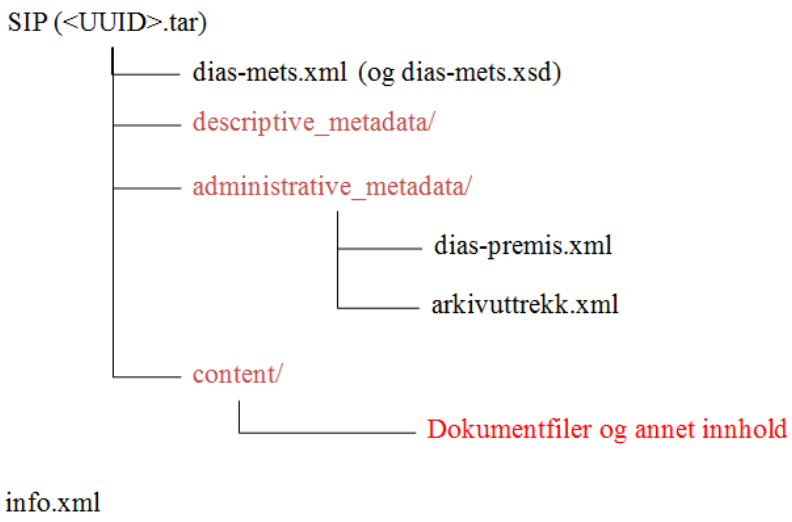


Figure 12: SIP used in National Archives of Norway

Arkivuttrekk.xml is an ADDML (Archival Data Description Markup Language) file containing information about extract.

Info.xml is a METS file and contains the checksum of SIP.

The digital delivery (pre-ingest and ingest) workflows include following.

Pre-Ingest

- Extraction – The specified data from the specified archival period is extracted in the specified extraction format (the current standard for records is Noark 5, but few extracts are following the current standard).
- SIP creation – The extract and information about the extract (such as period, creator, description, etc.) is packaged in NAN's SIP format, producing an SIP (DIAS-METS, DIAS-PREMIS, EAD, EAC-CPF, ADDML) as a TAR-file. The SIP is created using ESSArch Tools.
- Transfer – Transfer of the SIP. An e-mail from the Producer is sent to NAN, containing basic information about the extract and a hash value calculated on the SIP file. The e-mail is recorded in NAN's records management system. The SIP itself is transferred via another channel (DVD, portable disk, FTP, etc.)
- Submission – The received SIP is registered in the information system and goes through a virus check. A hash value is computed and compared to the value received on e-mail. The overall SIP content is compared to the agreement with the archives creator. The SIP is placed in a three week quarantine before a new virus check is performing with updated virus signatures.

Ingest

- Test – The SIP content is validated against the format specifications of both metadata and document files using various in-house tools for different types of content (Noark-3, Noark-4, Noark 5 and non-Noark content). If there are significant deviations, the SIP is rejected, and the archives creator is requested to deliver a corrected SIP. Additional preservation metadata are attached to the SIP.
- AIP creation – The validated SIP, with metadata describing any repository operations is packaged into an AIP using EPP.
- Archival storage – The AIP is registered in NAN's catalogue and is stored in secure digital repository along with an AIC (Archival Information Collection). The AIC keep track of the generations of AIPs after format conversions etc.

National Archives (UK)

The National Archives (UK) has defined a set of rules to the SIP construction and delivery for producers:[18]

- Hard drives must contain a single, NTFS formatted, file-system. The file-system volume label(s) provide an identifier for the physical media and the producers will also need to record these on the accompanying Delivery and Transfer forms.
- If the producers are sending closed records, they will need to save a copy of their application for closure form to the root of the file system on every drive that contains closed records referenced

[18] Packaging and delivery
http://www.nationalarchives.gov.uk/information-management/manage-information/selection-and-transfer/digital-records-transfer/digital-transfer-steps/

on that form. There is no need to sub-divide the closure form if a transfer spans several series or hard drives.

- At the root of the file-system, the producers will need to create a folder representing the series to which the transferring records belong. The series code must be used as the folder name, but with an underscore character in place of the space.
- The metadata.csv file must be put directly into this series folder. The checksum for this metadata file must be sent to the archives – it should be created exactly as the checksums for the records, and saved in the same folder as the metadata.csv file. The checksum should be in a simple text file called 'metadata.sha256'.
- The producers should create a further folder called 'content' inside the series folder. This will act as a container for the records themselves. The structure within the content folder has been not defined as this will depend on the records that are transferred. It is assumed that the contents of the folder will be described by the metadata files supplied by the producers.
- If it is a need to transfer records from multiple series, these steps should be repeated to create additional series folders at the root of the file system.
- The National Archives hard drives currently have a 2TB capacity. If a series will not fit on a single drive the producers should divide the records logically between two or more drives. They will also need to divide the metadata file so that each set of records remains with its associated metadata (and the checksum for the metadata file). It may be easier to generate the metadata for the two parts of the series separately rather than dividing the file. There is no need to sub-divide the closure file in this way.

Before digital records can be transferred to The National Archives, they must be appraised and selected for permanent preservation and reviewed for sensitivity through the following steps:[19]

- Appraisal;
- Selection;
- Sensitivity review (applying for closure on transfer);
- Preparation for transfer (test transfer of records and metadata, technical evaluation, metadata);
- Packaging and delivery.

Once these stages have been completed, the records may be delivered to The National Archives. The digital archiving workflow at National Archives after delivery is constructed as follows:

[19] Digital transfer steps
http://www.nationalarchives.gov.uk/information-management/manage-information/selection-and-transfer/digital-records-transfer/digital-transfer-steps/

1. Preparation:

Virus scans, file format characterisation, check if the file formats are on a "white list" using DROID, validate the metadata against relevant schema, checks the metadata is UTF-8 valid, create a checksum of each file and compare it with the checksum supplied

2. Pre-Ingest:

- updates DRI (Digital Records Infrastructure) catalogue with status that the pre-ingest has started, and then continues to update the status;
- uses 2 different antiviruses to check the data;
- validates the checksums of all files;
- checks if the file format is on a "white list";
- validates the content against the metadata;
- checks the metadata is utf-8 valid;
- generates a SIP package.

3. Ingest (for born digital):

- updates and checks the status information in the DRI catalogue;
- copies the SIP package in a processing area. This makes the metadata and content files available to all subsequent steps in the workflow for processing;
- fixity checks. Verifies that the file and the metadata exists and the content was not changed;
- validates the csv files against the schema to check that it conforms to it;
- file characterization. This involves identifying the formats of the content files, validating those formats and extracting the key properties associated with each file;
- CSV to XML transformation, incorporating the closure information in to the DRI catalogue
- adds TNA catalogue references;
- stores the files and the metadata in the archive as an AIP;
- updates the search index in SDB (The "Update Search Index" workflow step).

Danish National Archives

The SIP format is a Danish version of SIARD (known as SIARDDK). Please refer to the Executive Order on Submission Information Packages for further details.[20]

The usual workflow for digital archiving (including pre-ingest steps) is constructed as follows:

[20] The Executive Order on Submission Information Packages
http://www.sa.dk/media%283367,1033%29/Executive_Order_on_Submission_Information_Packages.pdf

- Notification and approval - Authorities are obliged to notify the National Archives when commissioning an IT-system used for the collection and storage of information that is created or obtained in conjunction with an authority's activities. The Danish State Archives will then evaluate whether the system should be preserved and, if so, determine a date on which the data in the system is to be transferred for the first time. This will normally take place after a period of approximately 5 years. All IT systems that are to be preserved must be approved upon commissioning.
- Submission agreement meeting - At the determined time an agreement about submission of data and the National Archives issue a submission provision that describes in detail what must be included in the SIP.
- SIP creation - The authority migrates the digital content from the original IT-system to the SIP format specified by law in the "Executive Order on Submission Information Packages".
- Submission and quality assurance - When a SIP is received it is virus checked and the integrity of the content is checked via comparison of checksums. Then the SIP goes through thorough quality assurance consisting of both automated and manual steps where the content and structure of the submission is verified. One important step is to ensure that the submitted data are meaningful and useful and that the meaning is preserved. SIPs must comply with the Executive Order on Submission Information Packages, which describe the structure of SIPs in detail. If the SIP does not comply with the executive order it is returned to the provider who will amend the SIP and re-submit it. This process is iterative and continues until the SIP fully complies with the requirements.
- Ingest to repository - After quality assurance the SIP is repacked and ingested into the repository as an AIP.

National Archives of Estonia

The SIP used at the National Archives of Estonia (NAE) consists of:

- Archival structure XML container (contains descriptions of the agency, fund, functions, series and case-files);[21]
- Record container with computer files (encoded to BASE64 format). Each record (metadata + computer files) is in a separate XML container;[22]
- Index files (one HTML file for human browsing and one XML file for automated processing).

The record containers are grouped together in a computer folder as seen in Figure 13. "sisukord.xml" and "sisukord.html" are index files which contain the information about SIP (including checksums). "EHA.3_[2014-06-25_14-40-10].arh" is a XML file what contains information about classification scheme and archival descriptions. ARH filename extension is used only for determining that this particular file's format is used by pre-ingest software universal archiving module (UAM). [23]

[21] XML Schema http://rahvusarhiiv.ra.ee/public/Digiarhiiv/UAM/UAM_Eksport_arhiiviskeem_v2.0.xsd
[22] XML Schema http://rahvusarhiiv.ra.ee/public/Digiarhiiv/UAM/UAM_Eksport_arhivaal_v2.0.xsd
[23] Universal archiving module
rahvusarhiiv.ra.ee/en/universal-archiving-module/

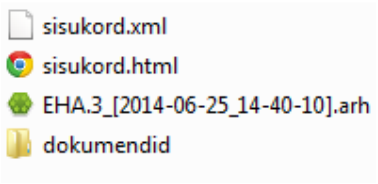


Figure 13: SIP structure used at the National Archives of Estonia

The usual workflow for digital archiving (including pre-ingest steps) is constructed as follows (A – archivist of NAE, P – archivist of the producer):

- The producer informs NAE (via e-mail or telephone) that s/he is willing to submit a collection of digital objects to the archives for long-term preservation. The head of appraisal at NAE and his/her colleagues inspect the collection and look which of the documents are fixed in Appraisal Act. Only documents that are included in this list will be exported to NAE.
- A guides P through the export process. S/he introduces P to the general rules and procedures of submitting digital objects as well as relevant software for pre-ingest preparation and transfer of digital documents. The general rule is that, before export, the collection has to be described and arranged on behalf of the producer.
- The software that P uses for arrangement and export is called the Universal Archiving Module, UAM (http://rahvusarhiiv.ra.ee/en/universal-archiving-module/).
- UAM holds the archival descriptions in archival schema of the producer (an XML file). In case P has ever given any objects (either digital or analogue) over to NAE, then the current state of the archival schema of this producer can be found in the archival information system of NAE (AIS, http:/ais.ra.ee), A exports it from there and sends to P (by e-mail). P uses this pre-filled schema in UAM for describing the new collection and creating the SIP. In case the archival schema cannot be found in archival information system (e.g the producer has never submitted anything to the archives, neither analogue nor digital material) then P opens a blank schema in UAM.
- The tool A uses for obtaining the archival schema from AIS is the ingest module of NAE.
- P sends the corrected schema (only schema, not the documents) back to A for supervision. For this, A uses ingest module which has functionality to compare and validate archival schemas. Should A find information missing, s/he informs P about the need for correction, P corrects mistakes and this process goes in cycles until the archival schema is accepted by A.
- P now adds digital content to the schema and writes its descriptions (using UAM). When this is finished P compiles the archival schema and digital objects into a SIP (using UAM).
- The producer sends the SIP to NAE via the national document exchange centre DEC (https://www.ria.ee/dec/). The SIP ends up in a folder that is readable by ingest module. The archivist starts a new submission project in ingest module, opens the SIP and carries out several checks. If the SIP is not acceptable (faults or missing data found), then A contacts P and asks P to do correct the SIP and perform export once again.

3.3.2 Service Providers

Interviews with Service Providers (4) allow to analyse the process of pre-ingest and ingest in more breadth at archives using commercial archive services.

3.3.2.1 RODA (KEEPS)

The RODA SIP is basically a compressed ZIP file containing a METS envelope, the set of files that compose the representations and a series of metadata records as shown in Figure 14. Within the SIP there should be at least one descriptive metadata record in EAD-Component format.[24]

One may also find preservation and technical metadata inside a submission package, although this last set of metadata is not mandatory as is seldom created by producers.

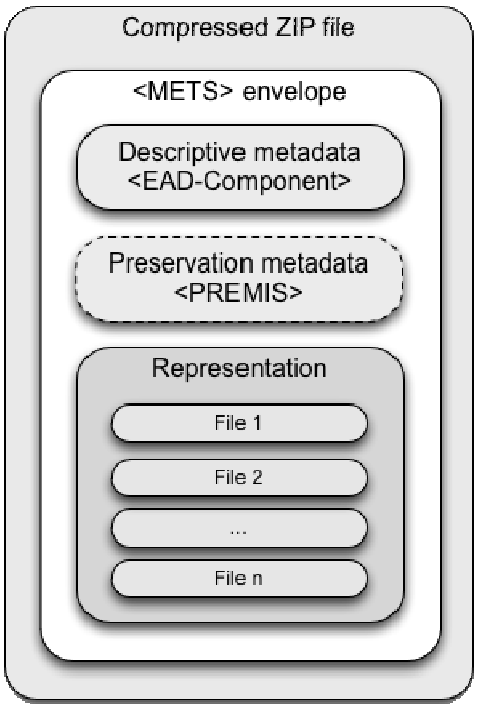


Figure 14: Structure of a Submission Information Package in RODA

Pre-ingest procedures contain institutional agreement between producer and archive, definition of a classification plan, user authorization and SIP creation with a tool RODA-in (by Producers).

The ingest workflow contains:[25]

[24] An EAD description is used to describe an entire collection of representations, but RODA SIP needs only a segment of EAD which is sufficient to describe one representation.
[25] RODA Community http://www.roda-community.org/features/

- Decompression of the SIP – ZIP file is decompressed.
- Virus check – SIPs are checked for viruses. Clam anti-virus is being used under the hood to perform this task.
- Envelope syntax check – Verify that the METS envelope is well formed.
- SIP completeness check - Check if all files referred in the METS envelope exist within the SIP.
- File integrity check – Files included in the SIP are accompanied by a checksum string. This information is used to check if any of the files have suffered corruption of some sort.
- Descriptive metadata check – Verify that an EAD-component is included in the SIP and that its syntax is correct.
- Preservation metadata check – Check if a PREMIS record has been included in the SIP and that its syntax is correct.
- Representation check – Verify that at least one representation exists within the SIP.
- Representation check – Depending on the type of the representation in the SIP, a series of more specific tests are conducted to verify if the representation is complete and format-wise compliant with the ingest policy in place.
- Specific representation check – Depending on the type of the representation in the SIP, a series of more specific tests are conducted to verify if the representation is complete and format-wise compliant with the ingest policy in place.
- Normalization – Representations whose format does not conform to the preservation formats defined by the preservation policy are automatically converted to the correct format. The original representation is maintained by the repository for diplomatic reasons.

3.3.2.2 Preservica

Preservica uses a workflow system and, as such, does not require any specific SIP structure. Hence, it can receive SIPs that are ingested in a variety of structures (e.g., national standards like ARELDA in Switzerland, EDIAKT in Austria or SAHKE2 in Finland).

However, it is most efficient to convert whatever SIP is received into a standard Preservica-defined SIP structure. This allows existing ingest workflows to be reused.

The standard Preservica-defined SIP structure has exactly one root or top-level directory. The name of the root directory is the string representation of a randomly generated (version 4) UUID as seen in Figure 15. The root directory contains a subdirectory named content, and the associated metadata in a file called metadata.xml. The content subdirectory contains all the physical files that make up the SIP; any arbitrary directory / file hierarchy is allowed within the content subdirectory.

```
'/UUID'
        /content
                '/file hierarchy'
                ...
        metadata.xml
'UUID'.protocol
```

Figure 15: Structure of a Submission Information Package in Preservica

The metadata must conform to the XIP schema and it must include a protocol file.

Preservica provides also a locally installable (optional) "SIP Creator" to

- Build submission packages from locally held files;
- Assign descriptive metadata from fragments created elsewhere or by using a GUI;
- Select where in the hierarchy to place the submission;
- Upload content to Preservica.

The minimum set of steps which are required to transfer digital records into the archive are to copy the records into Preservica's working area, ensure that the metadata etc. is correctly formed and then to store the files in Preservica's storage area(s) and the metadata in Preservica's metadata store database. However, this does not carry out all of the quality checks on the records being submitted that should be expected in a long-term repository, nor does it characterise the content files. Characterisation needs to be carried out because the information it provides is needed in order to be able to preserve the ingested information objects; however, such information objects can be characterised post-ingest, so it does not have to be included in an ingest workflow. Preservica includes workflow steps which implement both these additional processes, and standard workflows that incorporate them. An example of such a workflow is shown in Figure 16.

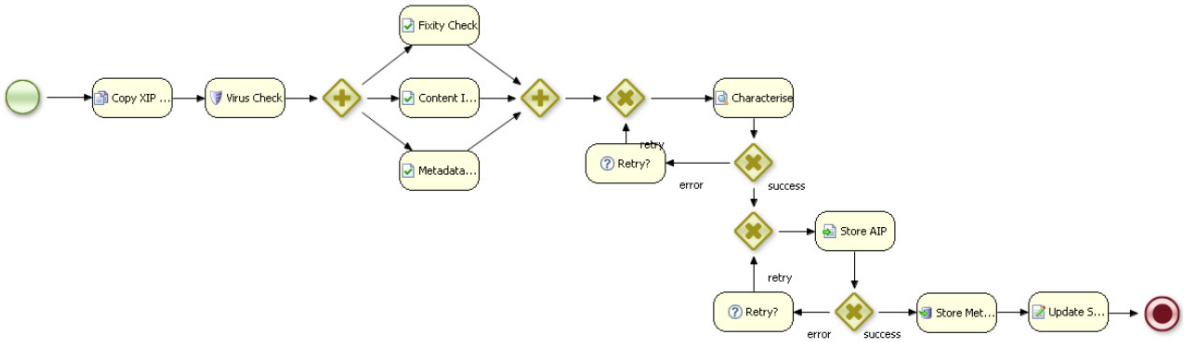


Figure 16: Ingest workflow in Preservica

- Pre-Ingest - SIPs must be created in or transformed into defined format.

- Receive Submission - Once a suitable SIP has been created, the ingest process needs to be initiated (manually, by monitoring a shared network, scheduled time).
- Quality Assurance Steps - One set of checks is intended to detect problems with the files, such as file corruption or viruses. Another set detects any mismatches between the metadata and the contents, such as content files which are not described in the metadata or missing content files that are described in the metadata. Finally, the metadata file can be validated against the XIP schema.
- Characterisation – There are two aspects to this. The first is to be able to determine whether a record is in need of attention and this requires the technology-dependent, technical properties of the record's content files to be measured. This involves identifying the formats of the content files, validating those formats and extracting the key properties associated with each file. The second aspect is to be able to determine the essential characteristics of each record that should be preserved in any preservation action and this requires the technology-independent, significant properties of the manifestations of the record to be measured. This involves detecting the presence of technology-independent "components", for example a document, and recording their properties regardless of the technology the component is manifested in, for example a PDF file with embedded images or a web page consisting of multiple HTML, CSS and image files.
- Store Files - The creation of an AIP from a SIP is a gradual process in Preservica, with each ingest step potentially adding extra metadata.
- Store Metadata – This step stores in the metadata store database the metadata contained in the XIP file in the current workflow instance's working area.
- Update Search Index – This step updates the index based on the descriptive metadata, if any, held in the XIP metadata xml file and the text of those content files in formats supported by the indexer.

3.3.2.3 ESSArch

Generally about ESSArch

ESSArch[26] consist of ESSArch Tools (ET) and ESSArch Preservation Platform (EPP). Together they support the whole process when information are structured and packaged as SIPs, delivered to a preservation platform, stored as AIPs and made accessible as DIPs. Together they bring cost effective functionality for creating and managing archived information. ESSArch is a multi-platform licensed as Open Source.

Short description of ESSArch Tools (ET)

ET is briefly a SIP package tool with logging (eq. notes/events) capabilities. It provides mechanisms for preparing, creating, transferring and receiving SIPs and along the way creates manually notes about the steps taken. It uses METS to describe SIP content and SIP packages (TAR-files) as well as PREMIS for content preservation. Notes and events are stored as PREMIS events. The SIP metadata (content/package/notes) are described as xml-files and based on the specifications for SIP packages used in Sweden and Norway. Both xml structure and the physical content represented in the xml structure are validated during the create process as well as when receiving SIPs. ET can be installed as windows 32/64 binaries and on Linux. A basic installation of ET is profiled as a producer (OAIS terminology) but can easily be switched to a receiver of SIPs as an archival institution (PreIngest/Ingest, OAIS). ET can also be profiled as being used within highly

[26] ESSArch – http://www.essarch.org

secured environments with only logging capabilities. All these three profiles can easily be switched within the application. ET is a stand-alone application and can be used as a complement to EPP.

ET can be used by those who produce information to be archived as well as by organizations which receive and preserve information. ET is configurable and adaptable to the processes and procedures that exist within a producer organization as well as within a preservation organization (public/restricted/secured zones).

SIP format and structure

The format and structure of a SIP is based on the conceptual idea of being able to describe and package any kind of content. In order to do so we need to use common specifications (CS) for different specific type of deliveries, the package itself and the delivery description of the package (SIP). CS is used to facilitate searching for and retrieving information for all sectors and with this including both the public and private sector. A CS is a structured description of the functional and technical requirements that meet the needs of all or part of the organization administration. A specification provides guidance when developing regulations, specifications for system procurement and when writing contracts.

ET does not create the specification (CS) for delivery types since it will be a part of the export from the producers system. ET will however create references to it in the package description.

The physical structure is normally a hierarchical map structure which basically contains a map for the content and a map for metadata (eq. context for content).

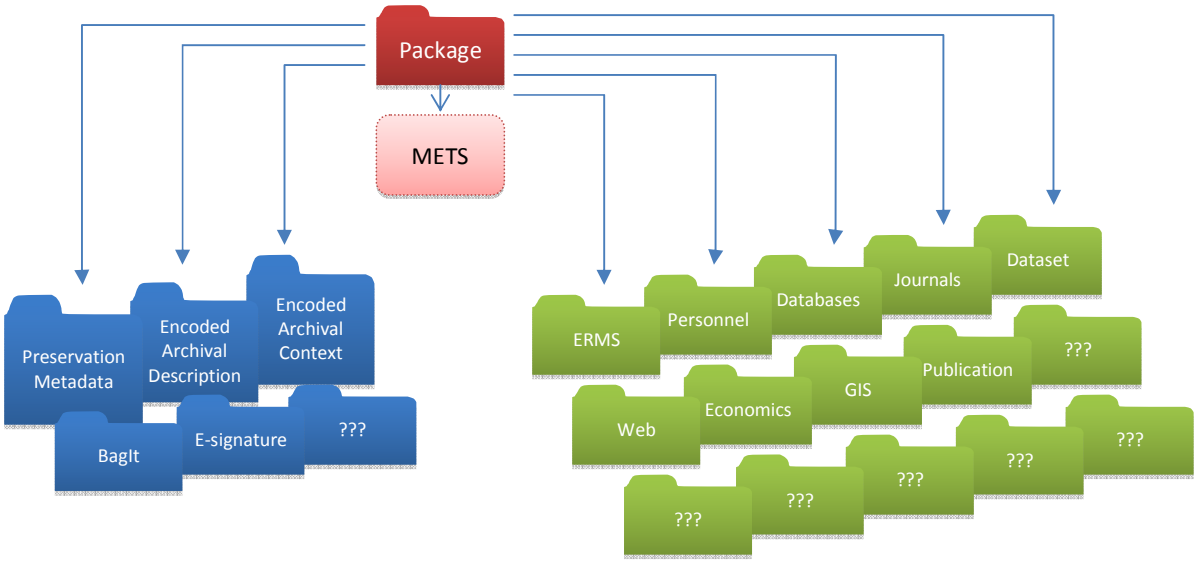


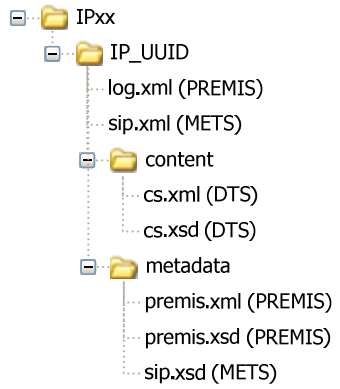
Figure 17: Common specifications for different types of delivery

The table below (Table 8) describes examples of delivery types. Each of the examples needs to be specified and described with its own common specification (CS).

Table 8: Examples of delivery types

Abbreviation	Description
ERMS	Information exported from any kind of case- or document management systems
Personnel	Information from any kind of personnel systems
Databases	Information from any kind of database or register systems
Journals	Medical journals from any kind of healthcare systems
Dataset	Any kind of information eq. physical files
Web	Information from web sites and intranets
Economics	Information from any kind of economic or business systems
GIS	Information from any kind of geographical information systems
Publication	An electronic publication

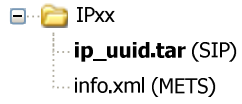
An information package (SIP) can be created in a directory structure like the one described in Figure 18. If the delivery consists of only one SIP, as an open directory structure, the package description will be represented by the package description *sip.xml*. If the deliver consist of one packed SIP or several packed SIPs a package delivery description *info.xml* will be created. Events related to a delivery, SIP, will be manually registered in ET and saved in enclosed log file *log.xml*.
Content in SIPs are also described by *premis.xml*, as preservation metadata. Associated schemas (xsd-files) for used metadata description files will be stored in the SIP.



IPxx – numbered directory for information package (SIP) to be prepared
IP_UUID – unique directory for information package (IP)
log.xml – notes / log file
sip.xml – package description for SIP
content – exported content from any kind of system
cs.xml – common specification for content delivery type
cs.xsd – schema for common specification for content delivery type
metadata – producers metadata for ”content” and archival descriptions
premis.xml – preservation metadata description file
premis.xsd – schema for PREMIS
sip.xsd – schema for METS package description

Figure 18: Structure of a Submission Information Package in ESSArch Tools

If an information package (SIP) will be delivered in a container format it will be packaged as a TAR-file with a unique identifier and described in the package delivery description *info.xml*.



IPxx – numbered directory for information package (SIP) to be transferred
ip_uuid.tar – information package (SIP) with unique identifier
info.xml – package delivery description for the information package (SIP)

Figure 19: A SIP in container format in ESSArch Tools

Rules for a SIP:

- Always contain an XML-file by an agreed name for example “sip.xml” which should contain general metadata describing the SIP and be based on the metadata standard METS and using the METS profile developed by the project E-ARK. This XML-file shall have the same format and structure regardless of the delivery type.
- The file “sip.xml” shall be placed at the root (top) level in the map structure used in the SIP.
- A SIP shall belong to one and only one delivery type.
- A SIP shall belong to one and only one Submission Agreement, SA.
- A SIP can contain one or many data files referenced in “sip.xml”.
- A SIP shall be size and volume independent. This means that there should be no limitations in the size and volumes of data files in the SIP and that the delivery type specifications shall be general enough to allow this.

Pre-Ingest and Ingest workflow (ET)

ET and EPP can be used in one organization where different roles and responsibilities are consolidated. They can also be installed and cooperate in separated organizations where different roles and responsibilities exist, preferably addressed as different zones. The basic delivery workflow between the producer and the preservation organization can be divided into four zones, one at the producer and three within the preservation organization:

- A producer zone where the information to be preserved is produced and packaged, and where the producer (creator) is responsible for the information. They could also be the consumer.

- A public zone at the preservation organization where the access and influence of information is regulated.
- A restricted zone, not public and where the access and ability to influence the information is highly restricted and controlled.
- A secured zone, not public and where access to the zone is strictly limited. The ability to influence the information is regulated to only a few functions.

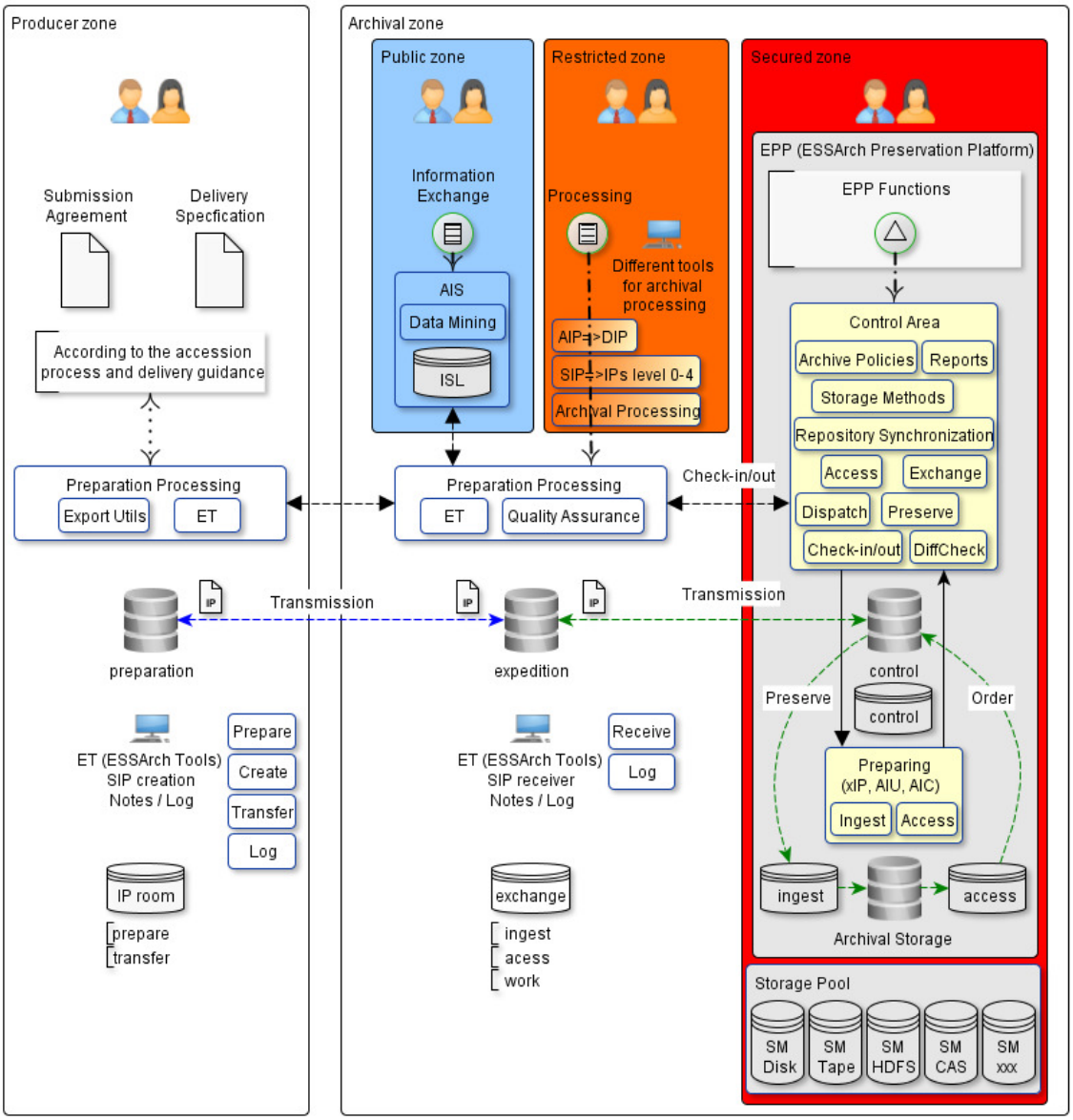


Figure 20: Zones where ET and EPP can exist

ET provides functionality to create an information package, SIP, with a fixed content exported from any kind of system. The SIP is described by the content delivery type description (e.g. *erms.xml*) which is a part of the export from the producers system, a package description (e.g. *sip.xml*) and the package delivery description (e.g. *info.xml*). ET will also create a log file, *log.xml* for events related to a delivery.

Information packages can be created by ET as TAR-files with associated checksums. The package is saved and transferred either on appropriate medium (carrier) by courier/mail or by any other transmission technique like ftp / scp etc.

The package delivery description (e.g. *info.xml*) is preferably sent to the preservation organization by e-mail. The preservation organization will after receiving the SIP and its package delivery description perform quality assurance controls.

ET has functionality to receive transferred SIPs if installed as such at the receiver organisation, equally the preservation organisation. When a SIP is about to be received and interpreted it also will, at the same time, be validated of its structure and content. If the SIP does not pass validation a notification must be sent to the producer, requesting a retransmit of the SIP.

If the SIP passes validation while it is received by ET it will be transferred to an expedition area and ET will prepare for generation and version management (AIC, AIU etc) in EPP. ET will also prepare for further processing, eq. check-in to EPP.

Pre-Ingest workflow (ET) Producer

Prepare SIP

- enter into form name of archivist organization and archive label of prepared SIP
- create map structure, unique id for IP and log file
- manually add export from system into content map
- add events manually

Create SIP

- check for any locked files
- get schemas from internet locations or locally
- create preservation metadata file (*premis.xml*) and package description file (*sip.xml*)
- create checksums for all files in SIP, store them in package description file
- run schema validation for PREMIS and METS files (*premis.xml* / *sip.xml*)
- check physical content against logical representation in package description file and vice versa
- create tar-file
- create package delivery description file (*info.xml*)
- run schema validation for METS file (*info.xml*)
- check physical content against logical representation in package description file and vice versa
- move tar file and *info.xml* to transfer file area

Deliver SIP

- fill in e-mail form e.g. to receiver of package delivery description file *info.xml*
- transfer SIP and package delivery description file *info.xml*

Ingest workflow (ET) Preservation organization

Before SIPs are received they are stored in quarantine for virus checks etc. which is not performed by ET.

Receive SIP

- select SIP to receive, interpret and search for physical tar-file and package description file *info.xml*
- perform validation of package delivery description file *info.xml*
- perform validation of physical content in package (SIP) and vice versa
- create unique AIC map structure and new log file in expedition area
- copy SIP to AIC map structure and media area in the expedition area
- add events manually

When the SIP is received it can be checked-in to EPP and further archival processing can be performed. EPP workflows are not explained since they are not a part of the pre-ingest and ingest workflow.

3.3.2.4 Archivematica

The physical structure of SIP used in Archivematica is shown in Table 9.

Table 9: Physical structure of SIP in Archivematica[27]

Path	Object Type	Function
/SIP_Folder	folder	top level container for the the SIP
		can take on any name
/SIP_Folder/logs	folder	
/SIP_Folder/logs/filemeta	fodler	
/SIP_Folder/metadata	folder	
/SIP_Folder/metadata/checksums.md5	text file	contains md5 hash values for objects in /SIP_Folder/objects
/SIP_Folder/objects	folder	

Archivematica SIP uses METS, PREMIS, Dublin Core and other metadata standards.

Pre-ingest and Ingest workflow consist mainly of two parts (transfer, ingest) which can contain many micro-services.[28]

Transfer services are shown in Table 10.

[27] SIP Structure
https://www.archivematica.org/wiki/SIP_Structure

[28] Archivematica implements a micro-service approach to digital preservation. The Archivematica micro-services are granular system tasks which operate on a conceptual entity that is equivalent to an OAIS information package: Submission Information Package (SIP), Archival Information Package (AIP), Dissemination Information Package (DIP).
https://www.archivematica.org/wiki/Micro-services

Table 10: Transfer micro-services in Archivematica[29]

Micro-service	Description
Approve Transfer	This is the approval step that moves the transfer into the Archivematica processing pipeline.
Verify transfer compliance	Moves the transfer to a processing directory based on selected transfer type (standard, zipped bag, unzipped bag, DSPace export or maildir). Verifies that the transfer conforms to the folder structure required for processing in Archivematica and restructures if required. The structure is as follows: /logs/, /metadata/, /metadata/submissionDocumentation/, /objects/.
Rename with transfer UUID	Directly associates the transfer with its metadata by appending the transfer UUID to the transfer directory name.
Include default Transfer processingMCP.xml	Adds a file named processingMCP.xml to the root of the transfer. This is a configurable xml file to pre-configure processing decisions. It can configure workflow options such as creating transfer backups, quarantining the transfer and selecting a SIP creation option.
Assign file UUIDs and checksums	Assigns a unique universal identifier and sha-256 checksum to each file in the /objects/ directory and sets file permission to allow for continued processing.
Verify transfer checksums	Checks any checksum files that were placed in the /metadata/ folder of the transfer prior to moving the transfer into Archivematica.
Generate METS.xml document	Generates a basic METS file with a fileSec and structMap to record the presence of all objects in the /objects/ directory and their locations in any subdirectories. Designed to capture the original order of the transfer in the event the user chooses subsequently to delete, rename or move files or break the transfer into multiple SIPs. A copy of the METS file is automatically added to any SIP generated from the transfer.
Quarantine	Quarantine's the transfer for a set duration, to allow virus definitions to update, before virus scan.
Scan for viruses	Uses ClamAV to scan for viruses and other malware. If a virus is found, the transfer is automatically placed in /sharedDirectoryStructure/failed/ and all processing on the transfer is stopped.
Clean up names	Some file systems do not support unicode or other special characters in filenames. This micro-service removes prohibited characters and replaces them with dashes. Original filenames are preserved in the PREMIS metadata.
Identify file format	Identifies formats of the objects in the transfer using either FIDO or file extension based on user choice. Format types are managed in the Format Policy Registry. This micro-service can be skipped and done in Ingest instead.
Extract packages	Extracts objects from any zipped files or other packages. Extracts attachments from maildir transfers.
Characterize and extract metadata	Characterizes and validates formats and extracts object metadata using the File Information Tool Set (FITS).
Complete transfer	Indexes transfer contents, then marks the transfer as complete.
Create SIP from Transfer	This is the approval step that moves the transfer to the SIP packaging micro-services (Ingest) if user chooses to Create single SIP and continue processing. User can also choose to Send transfer to backlog at this time.

Ingest services are shown in Table 11.

[29] Archivematica 1.0 Micro-services
https://www.archivematica.org/wiki/Archivematica_1.0_Micro-services

Table 11: Ingest micro-services in Archivematica[30]

Micro-service	Description
Verify SIP compliance	Verifies that the SIP conforms to the folder structure required for processing in Archivematica. The structure is as follows: /logs/, /metadata/, /metadata/submissionDocumentation/,/objects/.
Verify transfer compliance	Verifies the METS from the transfer.
Rename SIP directory with SIP UUID	Directly associates the SIP with its metadata by appending the SIP UUID to the SIP directory name and checks if SIP is from Maildir transfer type to determine workflow.
Include default SIP processingMCP.xml	Copies the processing configuration file added to the transfer in Include default Transfer processingMCP.xml, above, to the SIP.
Remove cache files	Removes any thumbs.db files.
Clean up names	Some file systems do not support unicode or other special characters in filenames. This micro-service removes prohibited characters and replaces them with dashes. Original filenames are preserved in the PREMIS metadata.
Normalize	Determines which normalization options are available for the SIP and presents them to the user as choices. Normalizes (i.e. generates preservation and/or access copies) based on selection. Thumbnail files are also generated during this micro-service.
Process submission documentation	Processes any submission documentation included in the SIP and adds it to the /objects/ directory.
Process metadata directory	Processes metadata.
Prepare DIP	Creates a DIP containing access copies of the objects, thumbnails and a copy of the METS file.
Upload DIP	Allows the user to choose to upload the DIP to either ICA-AtoM or CONTENTdm.
Upload DIP to ICA-AtoM	The user uploads the DIP to a selected description in ICA-AtoM.
Upload DIP to CONTENTdm	The user uploads the DIP to a selected description in CONTENTdm.
Prepare AIP	Creates an AIP in Bagit format. Creates the AIP pointer file. Indexes the AIP, then losslessly compresses it.
Store AIP	Moves the AIP to /sharedDirectoryStructure/www/AIPsStore/ or another specified directory. Once the AIP has been stored, a copy of it is extracted from storage to a local temp directory, where it is subjected to standard BagIt checks: verifyvalid, checkpayloadoxum, verifycomplete, verifypayloadmanifests, verifytagmanifests.

[30] Archivematica 1.0 Micro-services
https://www.archivematica.org/wiki/Archivematica_1.0_Micro-services

4. RECOMMENDATIONS FOR FURTHER WORK

This report has studied available practices of archival ingest of digital objects and their metadata including records export, preparation of the submission information packages and existing workflows what support that all in practice.

As the concluding recommendations derive from multiple sources (desktop research, online survey, interviews) it is not possible to link them directly to one source or answer. The recommendations are based on previously described results and are describing common principles among researched archival stakeholders.

Workflows

The gathered information reflects that currently the workflow for ingesting digital data is common only at a very high level. We can distinguish the data export phase, preparation, transfer and quality control as shown in Figure 21.

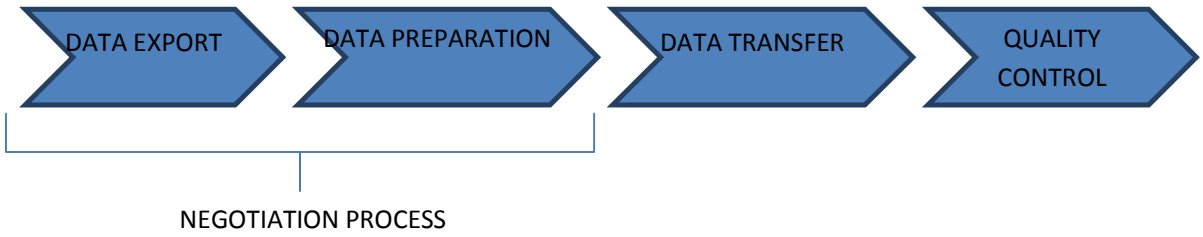


Figure 21: Common workflow

If some errors or issues (i.e. missing metadata) are encountered during the Quality Control then a new transfer may be initiated.

Many respondents also claim that they are using or plan to use the OAIS (Reference Model for an Open Archival Information System) compliant tools and standards.

In fact the workflows used for ingest are very similar to PAIMAS (Producer Archive Interface Methodology Abstract Standard, CCSDS 651.0-M-1, ISO 20652:2006) methodological standard which is tightly related to OAIS. PAIMAS consist of 4 phases:[31]

[31] PAIMAS (Producer Archive Interface Methodology Abstract Standard http://public.ccsds.org/publications/archive/651x0m1.pdf

- **Preliminary** (includes the initial contacts between the Producer and the Archive and any resulting, feasibility studies, preliminary definition of the scope of the project, a draft of the SIP definition and finally a draft Submission Agreement);
- **Formal definition** (includes completing the SIP design with precise definitions of the digital objects to be delivered, completing the Submission Agreement with precise contractual transfer conditions such as restrictions on access and establishing the delivery schedule);
- **Transfer** (performs the actual transfer of the SIP from the Producer to the Archive and the preliminary processing of the SIP by the Archive, as it is defined in the agreement);
- **Validation** (includes the actual validation processing of the SIP by the Archive and any required follow-up action with the Producer).

The processes and outcomes of PAIMAS can be seen in Figure 22.

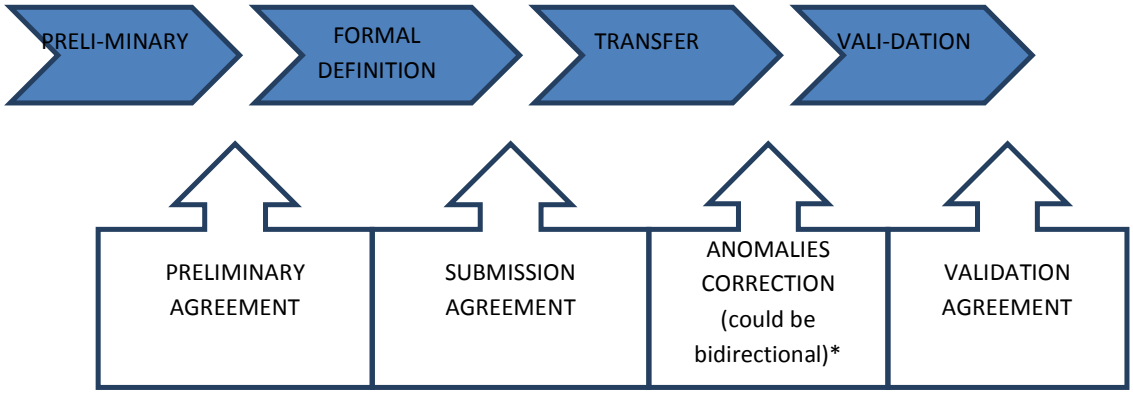


Figure 22: PAIMAS phases

* might cause a revision to the submission agreement and initiate a second Transfer

The authors of the report encourage the E-ARK project to consider taking PAIMAS standard for basis when designing ingest workflows for E-ARK further work.

As the process of creating archival information packages (AIPs) from the SIPs after validation can be complex (one SIP can produce one or multiple AIP, one AIP can be produced from multiple SIPs, produced at different times etc.) it is assumed that those rules are analysed and agreed later in the work of E-ARK project.

Records export

As respondent's activity was quite low regarding description of data export, we may assume there is no widespread or common practice of this process. As survey results also revealed, there is no certain standard for export process, the most used one is ISO15489-1, followed by MoReq and others. Both standards are not detailed enough for technical development. We should note that also OAIS does not describe technical implementation details.

As

- many national regulations state that records creators are responsible for the extraction of data and SIP creation;
- the regulations are at a high level and do not include technical details for records and metadata export;

the authors of this report recommend developing detailed and commonly understood requirements for records export process which include procedures for data selection, extraction, metadata mapping, validation and quality control. Apparently, they should include the clear roles for both sides – records creator and archives.

Submission information packet format (SIP)

The information gathered during desktop research, online survey and interviews reflect that understanding of the SIP format can be very different. Some respondents count simple computer folders as SIP, other ones see metadata standards as SIP etc, but still we can notice some general principles that SIP formats tend to have among our report target groups. According to the results we can look at a SIP in a two ways:

1. Physical view (structure how the physical (i.e. computer) files are located and naming conventions). The physical structure can be also very different, but the results show that mainly one manifest XML file is used which describes the physical structure of SIP, one another XML file which contains descriptive metadata and one folder with the content. Most SIPs have also some unique identifier (UID).

   Therefore, the authors of this report propose to use this (Figure 23) top-level structure for physical construction of the E-ARK SIP as these elements were most frequently represented in explored SIP structures.*

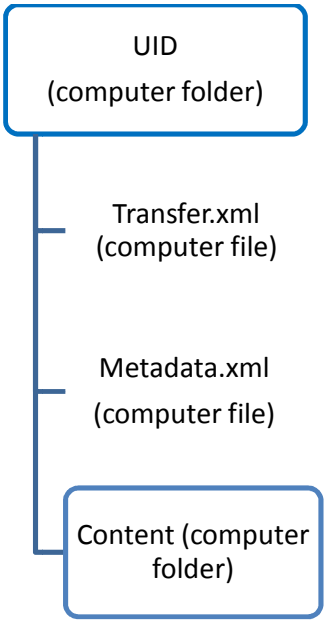


Figure 23: SIP physical view

*The metadata blocks for physical transfer or compressing options (TAR, ZIP etc.) are not looked at and discussed in the scope of this report.

2. Logical view answers to questions "What information is inside the physical folders and files?", "What metadata standards are used?"
   This report cannot recommend exact structure for logical view of SIP as the results do not reflect any common logical structure.
   Still, the logical structure seems to be influenced by METS standard[32] in many cases according to the gathered information. Also PREMIS[33] and EAD[34] are well represented. The suggested set of metadata could include blocks:
   - for automated transfer validation (integrity, checksums, technical metadata);
   - for describing the SIP structure;
   - for describing whole collection and Producer (e.g. EAC-CPF, EAD; catalogue metadata);
   - for describing single objects (e.g. MoReq, EAD);
   - for describing any relevant actions performed before or during the ingest;
   - for describing the access restrictions.

[32] METS provides a means of associating the metadata related to an object and describes the relationships with other objects.
http://www.loc.gov/standards/mets/
[33] PREservation Metadata: Implementation Strategies
http://www.loc.gov/standards/premis/
[34] Encoded Archival Description, is a non-proprietary de facto standard for the encoding of finding aids for use in a networked (online) environment.
http://www.loc.gov/ead/

Both physical and logical views can be harmonized as part of project results. We cannot expect to create one universal technical specification as a result of this project, but the close cooperation between all partners will enable synergies, minimization of financial input because of shared input, and further harmonization in next steps if the partners will be able to follow the recommendations.

As the paragraph about workflows pointed out the PAIMAS standard for describing pre-ingest and ingest workflows, then the authors of this report suggest looking also at PAIS (Producer Archive Interface Specification)[35] as one of the SIP possible candidates.

To conclude, the E-ARK needs to be broader than any previous approach in practice today, specifically:

- The E-ARK process(es) should manage ingest with a preliminary/pre-ingest phase, as well as ingest without a preliminary/pre-ingest phase;
- The E-ARK process(es) should manage ingest where there is a “contract” between the provider and the archive, and where there is no contract;
- The E-ARK process(es) should manage a SIP that is part of a series of such SIPs that are regularly transferred from a particular provider under a standing agreement, as well as a SIP that is unique or "standalone".

To address these points, as a general rule, the more pre-ingest/preliminary and formal definition work that is done up front, the less detail needs to be included in any individual SIP. The less work that is done during the preliminary phase, the richer the descriptive information in the SIP needs to be. Therefore the SIP structure for E-ARK needs to be flexible and handle situations where detailed information is provided within the SIP, as well as when detailed information is referenced (e.g. by URL) from documents hosted outside the SIP.

[35] PAIS (Producer Archive Interface Specification)
http://public.ccsds.org/publications/archive/651x1b1.pdf

5. REFERENCES

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

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

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

Faria L, Duretec K., Kulmukhametov A., Moldrup-Dalum P., Medjkoune L., Pop R.,
Barton S., Akbik A. (2014): "SCAPE survey on preservation monitoring"
http://www.scape-project.eu/wp-content/uploads/2014/05/SCAPE_D12.2_KEEPS_V1.0.pdf

Justrell B., Toller E. (2013): "Standards and interoperability best practice report"
http://www.dch-rp.eu/getFile.php?id=165

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. & Dobreva, M. (2012): *"Digital Preservation Services: State of the Art Analysis",* Digital Cultural Heritage Network, DC-NET:
www.dc-net.org/getFile.php?id=467

Swiss Federal Archives SFA, Historical Analysis Services (2013): "Analysis of digital documents in other national archives"

Velle, K. (2012): *"Database Archiving"*, EBNA, 29 May 2012, Copenhagen:
https://www.sa.dk/media(4588,1033)/EBNA-Minutes,_CPH_29-30_May_2012.pdf

6. APPENDIXIES

Appendix A: Guidelines for conducting interviews

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 interviews.
- All terms and rules should be introduced during the contact taking 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. But 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 needed 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 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 participant
- Interrupting participant (not allowing participant time to finish talking before asking the next question)
- Finishing sentences for 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 (yes we are serious and have an excellent example in the workshop)
- Being judgemental (we have a wonderful example in the workshop)
- 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 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 Private Companies / 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: Survey Questions for Government Bodies

Survey Questions for Government Bodies
(Q.1) What type of organization do you represent?
(Q.2) In which country does your organisation reside?
(Q.3) What is your role/position within the organization?
(Q.57) How many persons in your organisation undertake work related to digital curation?
(Q.58) Please specify national legislation that regulates: Pre-ingest and ingest, Archival storage/preservation, Access service and Access restriction.
(Q.59) Which standards for electronic document and records management are being used in your organisation or supported by your electronic records management system?
(Q.60) 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.61) Please, provide a URL link.
(Q.62) 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.63) Please, briefly describe the guidelines, and provide a URL link if the document is available online.
(Q.64) What tools and services are currently used for (pre)ingest and active digital preservation?
(Q.65) Are there any details of information packages (SIP, AIP) formats used in your organisation or supported by your solution(s) available online?
(Q.66) 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.67) Please, briefly describe the submission and archival information packages formats used in your organisation or supported by your solution(s).
(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 E: Survey Questions for Private Organisations

Survey Questions for Private Organisations
(Q.1) What type of organization do you represent?
(Q.2) In which country does your organisation reside?
(Q.3) What is your role/position within the organization?
(Q.43) How many persons in your organisation undertake work related to digital curation?
(Q.44) Please specify national legislation that regulates: Pre-ingest and ingest, Archival storage/preservation, Access service and Access restriction.
(Q.45) Which standards for electronic document and records management are being used in your organisation or supported by your electronic records management system?
(Q.46) 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.47) Please, provide a URL link.
(Q.48) 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.49) Please, briefly describe the guidelines, and provide a URL link if the document is available online.
(Q.50) What tools and services are currently used for (pre) ingest and active digital preservation?
(Q.51) Are there any details of information packages (SIP, AIP) formats used in your organisation or supported by your solution(s) available online?
(Q.52) 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.53) Please, briefly describe the submission and archival information packages formats used in your organisation or supported by your solution(s).
(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 F: Survey Questions for Projects

Survey Questions for Projects
(Q.1) What type of organization do you represent?
(Q.2) In which country does your organisation reside?
(Q.3) What is your role/position within the organization?
(Q.54) How many persons in your organisation undertake work related to digital curation?
(Q.55) Please specify national legislation that regulates: Preingest and ingest, Archival storage/preservation, Access service and Access restriction
(Q.56) What tools and services are currently used for (pre)ingest and active digital preservation?
(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 G: Standards, guidelines and legislation used by stakeholders.

Country	Stakeholder	Form	Description	URL
Australia	government org.	legislation	National Library Act 1968	
Australia	government org.	legislation	Copyright Act	
Belgium	government org.	guidelines	EUROPEANA, Biodiversity Heritage Library, Global Biodiversity Information Facility , Biodiversity Information standards (TDWG)	
Belgium	government org.	legislation	Commission Decision 2004/563/EC, Euratom of 7 July 2004 amending its Rules of Procedure, annexing the Commission's provisions on electronic and digitised documents (OJ L 251, 27.7.2004, p. 9);	
Belgium	government org.	legislation	Implementing rules for Decision 2002/47/EC, ECSC, Euratom on document management and for Decision 2004/563/EC, Euratom on electronic and digitised documents (SEC(2009)1643, 30.11.2009), adopted by the Secretary-General, in agreement with the Directors-General of Personnel and Administration and of Informatics.	
Belgium	government org.	standards	Moreq2, MARC	
Bulgaria	government org.	guidelines		http://sci-gems.math.bas.bg/jspui/handle/10525/2104/browse?type=dateissued&sort_by=2&order=DESC&rpp=20&etal=0&submit_browse=Update
Croatia	government org.	guidelines	Metadata guidelines, format guidelines	http://www.kultura.hr/Sudjelujte/Preuzimanja-i-dokumenti
Croatia	government org.	legislation		www.kultura.hr
Czech Republic	service provider	legislation	Zákon 499/2004 Sb., archival and records management Vyhláška 259/2012 Sb., the details of Record Management Národní standard (VMV 64/2012), National standard for ERMS, including the definition of SIP and communication (XML) between ERMS and Archives Zákon 300/2008 Sb., electronic acts and authorized conversion of documents	
Czech Republic	service provider	standards	ISO 15489-1, Moreq2	
Denmark	government	standards	iso 27001	

	org.			
Estonia	government org.	guidelines	The guidelines of the National Archives	http://rahvusarhiiv.ra.ee/en/principles-standards-guidelines/
Estonia	government org.	legislation	Archives Act	https://www.riigiteataja.ee/en/eli/530102013053/consolide
Estonia	government org.	legislation	Public Information Act	https://www.riigiteataja.ee/en/eli/514112013001/consolide
Estonia	government org.	legislation	Personal Data Protection Act	https://www.riigiteataja.ee/en/eli/512112013011/consolide
Estonia	government org.	legislation	Government regulation "Archival Rules" (available in Estonian)	
Estonia	government org.	standards	ISO 15489-1, ISO 23081-1, Moreq2	
France	government org.	guidelines	Evaluation of Electronic Archival System	http://www.archivesdefrance.culture.gouv.fr/static/7109
France	government org.	guidelines	Standard d'Echange de Données pour l'Archivage (SEDA and recently NF Z 44-022)	http://www.boutique.afnor.org/norme/nf-z44-022/medona-modelisation-des-echanges-de-donnees-pour-l-archivage/article/814057/fa179927
France	government org.	guidelines	Some directives for email archiving	http://www.archivesdefrance.culture.gouv.fr/static/2822 http://www.archivesdefrance.culture.gouv.fr/static/2823
France	government org.	guidelines	Study "proof of concept" from VITAM project on email archiving	http://www.archivesdefrance.culture.gouv.fr/static/7140
France	government org.	guidelines	References and "good practice" from Head IT for French government	http://references.modernisation.gouv.fr/archivage-numerique
France	government org.	guidelines	The VITAM project aims to produce also some experiments and tools to enhance and facilitate both pre-ingest, ingest and access, while producing also the electronic archival core system. This project is at his beginning.	
France	government org.	legislation	AFNOR NF Z 42-013 for general electronic archival system	
France	government org.	legislation	AFNOR NF Z 42-020 for electronic safe deposit for archive	
France	government org.	legislation	RM AFNOR NF Z 44-022	

France	government org.	legislation	SEDA (Standard d'Echanges de Données pour les Archives) for exchange rules both in ingest and access parts "Livre II du code du patrimoine" : rules for archive in all public agencies	
France	government org.	legislation	Various others	http://www.archivesdefrance.culture.gouv.fr/archives-publiques/lois/
France	government org.	standards	EAD/EAC, SEDA/NF Z44-022, MOREQ2010, ICA-Req	
France	service provider	standards	NF Z42013	
Germany	government org.	guidelines	DIN 31645 ("Information und Dokumentation - Leitfaden zur Informationsübernahme in digitale Langzeitarchive"): A guidance for ingests in digital archival systems	http://www.dnb.de/EN/Netzpublikationen/Ablieferung/ablieferung_node.html
Germany	government org.	legislation	Legal deposit including ingest, preservation and access	http://www.gesetze-im-internet.de/dnbg/index.html
Germany	government org.	standards	ISO 15489-1, DIN 31644:2012-04	
Germany	service provider	legislation	din tr-esor e-goc gestz	
Germany	service provider	legislation	Technical guidelines on long-term preservation of legal value of signed documents	https://www.bsi.bund.de/DE/Publikationen/TechnischeRichtlinien/tr03125/index_htm.html
Germany	service provider	standards	ISO 15489-1	
Ireland	service provider	legislation	National Archives Act 1986 applies Government records.	http://www.irishstatutebook.ie/1986/en/act/pub/0011/index.html
Italy	government org.	legislation	Guidelines and regulations issued by Parliament, Government and the National Archives ourselves	
Italy	government org.	legislation	"Codice dell'amministrazione digitale" legislative decree 82/2005 modified 2010	
Italy	government org.	standards	Moreq2	
Luxembourg	service provider	legislation	Articles 16, 109, et 189 du Code de Commerce	http://www.legilux.public.lu/leg/textescoordonnes/codes/code_commerce/L1_du_commerce.pdf
Luxembourg	service provider	legislation	Loi du 14 août 2000 sur le commerce électronique Articles 1322-2, 1334, 1341, 1348 du code civil	http://eli.legilux.public.lu/eli/etat/leg/loi/2000/08/14/n8
Luxembourg	service provider	legislation	Règlement grand-ducal du 22 décembre 1986	http://www.legilux.public.lu/rgl/1986/A/2748/1.pdf
Luxembourg	service provider	legislation	Loi du 5 avril 2003 sur le secteur financier.	http://eli.legilux.public.lu/eli/etat/leg/loi/1993/04/05/n1
Portugal	service provider	legislation	Personal data protection law.	
Portugal	service provider	standards	ISO 27001	

Spain	government org.	guidelines		http://suport.aoc.cat/Portal/Tots-els-serveis/Integracio-serveis-Consorci-AOC
Spain	government org.	guidelines	Condicions específiques de prestació del servei iARXIU	https://www.aoc.cat/content/download/13501/32409/file/Cond_espec%C3%ADfiques_iARXIU_amb_annexos.pdf
Spain	government org.	guidelines	iArxiu: Estructura i creació de Paquets d'Informació de Transferència (PIT) utilitzant el model METS	https://www.aoc.cat/content/download/6657/24722/file/estructuraPitMets.pdf
Spain	government org.	guidelines	3D Icons	http://www.3dicons-project.eu/
Spain	government org.	standards	ISO 15489-1, ISO 23081-1, ISO 24721, OAIS	
Spain	service provider	guidelines	PREMIS	
Spain	service provider	standards	ISO 15489-1, Moreq2, Moreq2010	
Sweden	government org.	guidelines	Guidelines and regulations issued by Parliament, Government and the National Archives ourselves	
Sweden	government org.	legislation	Offentlighets- och sekretesslagen	
Sweden	government org.	legislation	Personuppgiftslagen	
Sweden	government org.	legislation	Skattedatabaslagen	
Sweden	government org.	legislation	Skattedatabasförordningen	
Sweden	government org.	legislation	The Freedom of the Press Act, which states the basic rights of the public to have access to public records (official documents) and also defines the term public record	http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Tryckfrihetsforordning-19491_sfs-1949-105/?bet=1949:105
Sweden	government org.	legislation	The Archives Act which defines the scope of activities that the SNA and the municipal archives are responsible for. As well as defining the goals of these "archival" activities.	http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Arkivlag-1990782_sfs-1990-782/?bet=1990:782
Sweden	government org.	legislation	The Archives Ordinance which mandates the SNAs right to regulate records management and archival activities at state public agencies. From procurement of Writing materials to storage facilities. Including all facets of Electronic public records. It also extends the definition of public record in the Freedom of the Press Act to specifically include any single data in a database.	http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Arkivforordning-1991446_sfs-1991-446/

Sweden	government org.	legislation	Regulations concerning access and secrecy, documentation of paper as well as electronic public records can be found in the Public Access to Information and Secrecy Act	http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Offentlighets--och-sekretessla_sfs-2009-400/?bet=2009:400
Sweden	government org.	legislation	The Personal Data Act is the Swedish implantation of the EU directive	http://www.government.se/content/1/c6/01/55/42/b451922d.pdf
Sweden	government org.	legislation	General regulations issued by the SNA include rules governing everything from creation of records to disposal of them or transfer to the SNA. They also cover such things as storage facilities, description och records and archives etc. All on a very general level that does not include any specfics regarding Electronic public records, but are applicable to them as well as paper records, sound recordings etc.	http://www3.ra.se/ra-fs/ra-fs_1997-04.pdf
Sweden	government org.	legislation	An addition concerning and especially applicable to the description of (electronic) public records	http://www3.ra.se/ra-fs/ra-fs_1997-04.pdf
Sweden	government org.	legislation	Specific regulations issued by the SNA concerning electronic public records	http://www3.ra.se/ra-fs/ra-fs_2009-01.pdf
Sweden	government org.	legislation		and http://www3.ra.se/ra-fs/ra-fs_2009-01.pdf
Sweden	government org.	legislation	General regulations concerning storage facilities	http://www3.ra.se/ra-fs/ra-fs_2013-04.pdf
Sweden	government org.	standards	ISO 15489-1, Moreq2	
Switzerland	service provider	guidelines	OAIS	
Switzerland	service provider	legislation	National and various Cantonal archiving and records management laws	
Switzerland	service provider	standards	ISO 15489-1, ISO 23081-1, Moreq2, Moreq2010	
The Netherlands	government org.	guidelines	Specific metadata profile special designed for permanent archival for governmental use	http://www.nationaalarchief.nl/sites/default/files/docs/Toepassingsprofiel_metagegevens_rijksoverheid.pdf
The Netherlands	government org.	guidelines	PDF 1.4; PDF/A 1b	
The Netherlands	government org.	legislation	Justid manages a edepot.	www.justid.nl
The Netherlan	government org.	standards	ISO 23081-1	

ds				
The Netherlands	service provider	legislation	Only tax office regulations, usually handled through paper	
The Netherlands	service provider	legislation	UK Data Protection Act 1998 Dutch Data Protection Act 2000	
United Kingdom	government org.	legislation	UK Public Records Act 1958	http://www.nationalarchives.gov.uk/information-management/legislation/public-records-act.htm
United Kingdom	government org.	legislation	Freedom of Information Act 2000	http://www.legislation.gov.uk/ukpga/2000/36
United Kingdom	government org.	legislation	Data Protection Act 1998	http://www.legislation.gov.uk/ukpga/1998/29/contents
United Kingdom	government org.	legislation	Environmental Information Regulations 2004	http://www.legislation.gov.uk/uksi/2004/3391/contents/made
United Kingdom	government org.	legislation	The Re-use of Public Sector Information Regulations 2005	http://www.legislation.gov.uk/uksi/2005/1515/contents/made
United Kingdom	government org.	standards	ISO 15489-1	
United Kingdom	service provider	guidelines	BS10008 - Evidential weight and legal admissibility of electronic information	
United Kingdom	service provider	guidelines	We operate at the file storage/bit preservation level and make extensive use of checksums for data integrity validation. We follow the OAIS model where appropriate (e.g. we provide archive storage for AIPs) and we follow the applicable parts of ISO16363. General best practice includes multiple copies of data in multiple locations with active integrity management and regular technology/media migration to address obsolescence.	
United Kingdom	service provider	legislation	We provide a data archiving service to our customers. Regulations that they need to comply with include the Data Protection Act, ISO27001 information security, IL levels for government information e.g. IL2 or IL3, and in the case of healthcare/pharma there's FDA in the US, Eduralex in the EC, and UK regulations from MHRA. For example. MHRA guidelines on GCP and FDA 21 CFR part 11. The list is quite long. We'd be happy to provide more information and links if	

			needed.
United Kingdom	service provider	legislation	Depends on sector (e.g., Public Records Act related only to public records). Other sectors might be regulated which might ultimately be backed up by legislation but the legislation won't specify details.
United Kingdom	service provider	standards	ISO 15489-1, ISO 23081-1, Moreq2010
USA	service provider	legislation	US Government laws
USA	service provider	legislation	PDF/A (ISO 19005) is an international standard that has been adopted by many members of the EU, as well as most countries in Latin America and Asia.
USA	service provider	standards	PDF/A - ISO 19005

Appendix H: Assessment of stakeholders for interview from point of view of D3.1

Colour codes used in the schema:

Relevant for interview	Could be relevant for interview but deselected

Schema for identification of stakeholders for interview:

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*
STAKEHOLDERS IDENTIFIED BASED ON THE ONLINE SURVEY							
Arkivum	Service provider	NA	NA	Clients determine content and AIP format	Not relevant for this work as it deals mainly with bit preservation.		Survey
Bulgarian Archives State Agency	Archive	NA	NA	Textual data, images, databases	Although the stakeholder uses SharePoint based system, the solution does not seem interesting enough to be elaborated on in an interview		Survey
Bundesarchiv	Archive	Acquisition of single records	Migration	Textual data	Interesting PreIngest-Toolset PIT. Information about PIT and SIP received directly, no separate interview needed.		Survey
Consorci Administració Oberta de Catalunya	Archive	Single records	Migration	Textual data, images, audio-visual data	The answers contain good links to sufficient online material, no separate interview needed.		Survey
Danish National Archives	Archive	Whole systems	Normalisation on ingest and migration	Digitised material, databases with preservation formats for text, sound, video and geodata	Interesting ingest procedures.	x	Survey
Estonian National Archives	Archive	Single records	Normalisation on ingest and migration	Textual data, images, Audio-visual data	Custom-built SIP and interesting ingest workflow.	x	Survey
Italy	Archive	Whole systems	Normalisation on ingest	Digitised material, images,	No contact information provided.		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*
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
National Archives of Hungary	Archive	Single records and whole systems	Normalisation on ingest and migration	Textual data, images, Audio-visual data, databases	The Archives use services from Preservica, ScopeArchive and a tool Elev SIP Creator. The National Archives of Hungary was used to test the interview methodology.	x	Survey
Portugal	Archive	NA	Normalisation on ingest, Migration	Textual data, images	No contact information provided.		Survey
Preservica	Service provider	NA	NA	NA	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
Stanford Digital Reposity	Archive	NA	Normalisation on ingest	NA	Libraries are out of scope for interviews.		Survey
The National Archives UK	Archive	NA	Migration	Textual data, images, audio-visual data	Interesting pre-ingest and ingest procedures.		Survey

ADDITIONAL STAKEHOLDERS IDENTIFIED BASES ON E-ARK KNOWLEDGE AND DESKTOP RESEARCH

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
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.		http://samfund.dda.dk/dda/default-

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*
					The archive is considered to be representative for data archives that uses DDI-L for preservation and access. It is not in the scope of this work.		en.asp and http://samfund.dda.dk/dda/default-en.asp
ESSArch Tools	Service provider	NA	NA	NA	Open source software that supports the entire digital preservation process.. It is widely used in Scandinavian countries. As the information about workflows and SIP format used received directly, no separate interview would be needed.	x	http://www.essarch.org
ExLibris	Service provider	NA	NA	NA	Widely used software at libraries, but libraries are out of scope for interviews.		http://www.exlibrisgroup.com/category/RosettaOverview
Library and Archives Canada (LAC)	Archive	NA	NA	NA	LAC is building Trusted Digital Repository (TDR). As the work is currently in process it is reasonable to not conduct the interview yet.		http://www.bac-lac.gc.ca/eng/Pages/home.aspx
National Archives of Norway	Archive	NA	NA	NA	National Archives of Norway was used to test the interview methodology.	x	http://www.arkivverket.no/eng/
National Archives of Sweden	Archive	NA	NA	NA	Have a relatively large collection of born-digital material which origins back to the 1960ies. Has enough information available online.	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/
Swiss National Archives	Archive	NA	NA	NA	Tool Package Handler for creating, examining and validating digital packages.		http://www.bar.admin.ch/dienstleistungen/00823/01559/index.html?lang=en

Appendix I: 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 J: 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 K: Terminology

AIP	OAIS: An *Archival Information Package*, consisting of the *Content Information* and the associated *Preservation Description Information (PDI)*, which is preserved within an OAIS.
Archive	An Organisation that intends to preserve information for Access and use by a *Designated Community*.
Descriptive metadata	Metadata that describes the data content.
Digital material	The term used to describe the digital assets of an archive, contained in *Information Packages*.
Digital Object	An object composed of a set of bit sequences.
Dissemination Information Package (DIP)	*Dissemination Information Package*, an *Information Package*, derived from one or more *AIPs*, and sent by *Archives* to the *Consumer* in response to a request to the *OAIS*.
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.
Information Package	A logical container composed of optional *Content Information* and optional associated *Preservation Description Information*. Associated with this *Information Package* is *Packaging Information* used to delimit and identify the *Content Information* and *Package Description* information used to facilitate searches for the *Content Information*
Ingest	PAIMAS: The OAIS entity that contains the services and functions that accept Submission Information Packages from Producers, prepares Archival Information Packages for storage, and ensures that Archival Information Packages and their supporting Descriptive Information become established within the OAIS.
OAIS	The Open Archival Information System is an archive (and a standard: ISO 14721:2003), consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a *Designated Community*.
Producer	The role played by those persons or client systems that provide the information to be preserved. This can include other *OAIS'es* or internal *OAIS* persons or systems.
Service providers	Companies providing services to archives ranging from developing software to performing services
Submission	An *Information Package* that is delivered by the *Producer* to the *OAIS* for use in the

Information Package (SIP)	construction or update of one or more *AIPs* and/or the associated *Descriptive Information*.