Access to Public Information in Government Agencies and Archives

New e-services for archives: results from user and expert survey

Final version
23.12.2011

Mari Runardotter, Östen Jonsson (Luleå University of Technology)
Helga Jóna Eiriksdóttir, Julia Pálmadóttir Sighvats, Njóður Sigurðsson (National Archives of Iceland)
Mattias Djupdahl, Magnus Geber (National Archives of Sweden)
Birgit Nurme, Kuldar Aas (National Archives of Estonia)
Raivo Ruusalepp (Estonian Business Archives)
Table of Contents

1. Introduction ................................................................................................................................. 3
   1.1 Problem statement .................................................................................................................... 3
   1.2 Previous work in the project .................................................................................................... 3
   1.3 The public e-services landscape ............................................................................................ 4
   1.4 Structure of the report ............................................................................................................ 4
2. Challenges for e-Service Development in Archives ................................................................. 5
3. User Survey ..................................................................................................................................... 7
   3.1 Summary of responses .............................................................................................................. 7
   3.2 Main survey results .................................................................................................................... 8
   3.3 Expectations for new e-services .............................................................................................. 10
   3.4 Conclusions from survey results ............................................................................................ 10
4. Expert Opinions ............................................................................................................................ 11
5. Enhancing Archives’ e-Services: Three Views .......................................................................... 12
   5.1 Crowd-sourcing for archival description and records’ digitisation ....................................... 12
      5.1.1 Problem statement ............................................................................................................. 12
      5.1.2 Existing technologies and the proposed solution .......................................................... 12
      5.1.3 Discussion ........................................................................................................................ 13
   5.2 Providing seamless access to archived databases ................................................................. 14
      5.2.1 Problem statement ............................................................................................................. 14
      5.2.2 Existing technologies and proposed solution .................................................................. 14
      5.2.3 Discussion ........................................................................................................................ 15
   5.3 Connecting archival content to e-government services ....................................................... 16
      5.3.1 Problem statement ............................................................................................................. 16
      5.3.2 Existing technologies and proposed solution .................................................................. 16
      5.3.3 Discussion ........................................................................................................................ 17
6. Conclusions and Discussion ........................................................................................................ 18
References ........................................................................................................................................... 20
Appendix 1: Survey Questionnaire ............................................................................................... 21
Appendix 2: User Suggestions for New e-Services from Memory Institutions ............................ 25
1. Introduction

The on-going transformation process of e-government and e-services permeates most sectors in society. Initiatives are taken not only by public agencies, but also by the private sector, research communities and individuals. This paper focuses on e-services in the cultural heritage sector and presents the results from a joint networking project of archives and academia from Estonia, Iceland and Sweden - the Access to Public Information in Governmental Agencies and Archives (APIS) project.

1.1 Problem statement

Archives hold vast volumes of information, including primary sources of history, information needed to protect citizens’ rights and public administration records in general. This information is valuable not only to individuals and the state but also to the history and the identity of every region, city or nation. Only a small amount of archival material has been digitized so far and often the only way to access archived information is still to visit the archives and browse through paper records in the archive reading room. Additional limitations are set by the finding aids in archives – a considerable part of finding aids are still in analogue format only and archival material is seldom registered at the individual item level. By developing e-services in archives, archival material can become more accessible, thus facilitating transparency for both democratic processes and the knowledge society.

The work in the APIS project has been guided by quality criteria set for e-services. The project group identified the following as relevant criteria for archives – the e-services should be:

- citizen-centric,
- generic, and
- border-crossing.

The first criterion – citizen-centric – indicates that e-services should be developed in close collaboration with end-users, build on their needs, requirements and visions from the beginning. The second criterion – generic – points to the importance that the e-service is developed through collaboration between different stakeholders, i.e. end-users (citizens), archives, universities, IT developers and service providers. Finally, the third criterion – border-crossing – specifies that the developed e-service must be of interest in all participating countries and internationally.

1.2 Previous work in the project

The APIS project set out to:

1) map and analyse the current access service provision in government agencies and archives, with a focus on the efficiency and effectiveness; and
2) identify the main gaps and challenges in different countries when providing seamless access services to public information in agencies and archives in the Nordic-Baltic region.

The first objective is reported on in three country-specific reports;¹ the identified gaps and challenges are summarised in the first APIS White Paper.² All three countries involved in the project have similar legislation regarding access to public information, records management procedures and archival practices. The countries also have similar restrictions for access to information, that is, protection of personal information³ and protection of information regarding national security.⁴ Because of these

---

1 [https://sites.google.com/site/theapisproject/publications]
2 [https://sites.google.com/site/theapisproject/publications/TheAPISproject_firstwhitepaperver02.pdf]
3 The Personal Data Protection Act (RT I, 2010, 11) in Estonia; the Personal Data Act (1998:204) in Sweden; the Personal Data Act (No. 77/2000) in Iceland
4 The State Secrets and Classified Foreign Information Act (RT I 2007, 16, 77) in Estonia; the Publicity and secrecy Act (2009:400) in Sweden; and the National Archives Act (No. 66/1985) and the Public Information Act (No. 50/1996) in Iceland
similarities between Estonia, Iceland and Sweden, finding a common ground for developing shared e-services for archives is feasible.

1.3 The public e-services landscape

The APIS project has considered different types of e-services, and the project aimed for coordinated, generic and informative/performative e-services (Albinsson, Forsgren, Lind and Salomonsson, 2006). The first e-service type, coordinated, indicates that it requires several involved authorities for its delivery and the second type, general e-service, signifies that the e-service is aimed at the general public (citizens). Regarding the third type, it is clear that e-services built on cultural heritage can be both informative and performative, insofar that the citizen could both study information supplied by the producer, but also act communicatively.

Another way of categorising e-services is to characterise the maturity of a specific e-service, and this is usually done in so called stage models, e.g. the e-service can be divided into one of four stages; 1) information, 2) interaction, 3) transaction and 4) integration (Albinsson et al., 2006). However, both Albinsson et al. (2006) and Lind and Goldkuhl (2008) argue that these models are somewhat tricky since they lead to confusion whether they describe a prescribed trajectory of e-service development or if they indicate varying advanced degrees of the e-service/s.

E-services from archives are, thus far, few and mainly limited to informative services like browsing and searching archive catalogues or viewing limited digitised collections. No evidence of co-ordinated e-services having been implemented in archives was found.

1.4 Structure of the report

The APIS project group discussed at length the types of services that users of archives require and how to conceptualise these. The project group’s initial ideas were expressed as a questionnaire that was distributed to general public through the Botnia LivingLab in Sweden to validate and complement the project group proposals. In parallel with the user survey, consultations with domain experts were held. Three types of e-services were formulated as a synthesis of memory institutions’ view and end-users expectations.

This report is an overview of the work conducted by the APIS project to flesh out new e-services to improve access to public archives. The second and third chapter provide a picture of the problems and challenges archives are facing when trying to deliver e-services. The third chapter reports on a user survey that was conducted for the APIS project by the Botnia LivingLab5 in Sweden and the forth chapter summarises a series of expert interviews and opinions related to archival e-services. Based on the previous chapters three e-service concepts are presented in chapter five. The e-services are set up as scenarios that cover three aspects of the same problem area, however addressing different angles, types of information and target user communities. The report concludes with a discussion of how the proposed scenarios fulfil the chosen criteria for e-services and a comparison of the expected outcomes of the three alternative scenarios. Finally the conclusions summarise the main results of this phase of the APIS project and what are the sustainability plans of the project group.

5 http://testplats.com/doc/start/se/article/2891
2. Challenges for e-Service Development in Archives

In order to develop e-services based on public information in governmental agencies and archives there are several challenges that need to be considered. First and foremost, there is a pressing need for overarching initiatives, such as creating common policies concerning objectives, strategies and methods on a range of issues – from digitisation and digital archives to how to publish public information on archives’ websites in a unified way. It is an important task with direct influence on the ability to synchronise, co-ordinate and create joint search and retrieval tools for end-users of archives. The establishment of these initiatives would constitute a solid foundation and preconditions for the development of common e-services. It would make possible a broadened technological exchange and collaboration, the development of a common infrastructure and standardisation. Moreover, common e-services, digitisation and digital preservation demand engagement from several professional domains and require different competences, such as librarians, archivists, IT personnel, economic and legal expertise, etc. In addition to technical and interoperability issues there are also administrative and organisational challenges in setting up cross-border collaboration for e-services. In many countries there are activities to standardise and improve access to records management systems within the public sector. Another example is web-based solution where institutions like the Swedish e-Delegation provide guidelines on web-based services development for the governmental agencies.

The APIS project has identified practice-based challenges in how access to archival materials is provided today. Some of these are country specific and probably cannot be resolved with an e-service common to the three countries. An example of such challenges is the need for guidelines on how agencies’ public records and records registries should be published on their web sites, and guidelines on what information can be published on the web.

A list of other challenges identified includes

- In recent years, significant funds have been spent on digitization of memory institutions’ analogue collections, yet the majority of collections still remain in analogue format only (especially in smaller institutions). Digitisation of whole collections is costly and only a fraction of existing material can be digitised by archives themselves in the foreseeable future. However, without making content digital, archives will be unable to fully participate in the e-services provision.

- The volume of analogue material in archives with awaiting digitisation is such that it probably cannot be solved with traditional funding by archives. The same is true for traditional archival description that stops at a file or folder level and is rarely available for individual document (item) level. New and low-cost business models for digitising and enhancing the item-level archival description are sought by all memory institutions and many are utilising crowd-sourcing for this (cf. EuropeanaConnect Media Annotation,\(^6\) Waisda\(^7\)).

- Challenges that arise during transfer of records to public archives include securing that the description of transferred records is sufficient (especially when it concerns record-level metadata), as well as ensuring that links are not broken during conversion of record components (electronic computer files) to new (archival) file formats.

- Born-digital records that are beginning to be transferred to public archives are described on the level of individual records, while the traditional archival description stops at a higher (file or folder) level. This leads to a serious limitation in existing finding aids – finding individual items (documents, records) will only be possible for the more recent past. Accessing historic material from archives will require additional description to be created and more efficient search and

---

\(^6\) http://dme.ai.t.ac.at/annotation/

\(^7\) http://woordentikkertje.manbithond.nl/
retrieval tools to be developed. These should be designed so that they could connect archives’ collections to e-government services as well.

- User satisfaction studies show that many of the present access services in memory institutions are more organisation-centric than user-centric, and often too complex to use, especially on hand-held devices. In order to get a better understanding of citizens’ attitudes to e-services within the cultural heritage area, a questionnaire was developed and distributed to the general public (see Chapter 3 below).

- Another problem area is the access to archived databases. While all three project participant countries have experience in archiving and preserving public sector databases, there is a general lack of user-friendly access methods and tools to archived relational databases. The current prevailing approach is to archive relational databases in standardised formats as periodic snapshots (for example a snapshot can be archived annually or every 5 years) and access is only granted to an individual snapshot by uploading it into a modern database environment. At the same time archived public sector databases hold great potential for serving collaborative e-services which often combine data from both live and archived datasets or over several thematic areas. A more detailed overview of the problems and a possible solution is described in Chapter 5.2 of this white paper.
3. User Survey

In accordance with the citizen-centric approach in the APIS project, we wanted to get a picture of citizens’ overall attitude towards e-services within the cultural heritage area. For that reason we compiled a questionnaire with 20 questions around their general interest in history:

- if and how often they visit or search for information in archives, libraries and museums;
- if they would like to use more e-services in this sector;
- how they would like to access these;
- what suggestions they had for e-services, together with
- a motivation for importance of suggested e-services.

The questionnaire was sent out through Botnia LivingLab, which holds a database of approximately 5,100 persons that have registered voluntarily as participants. These individuals come from all over Sweden but are mostly used for evaluation of technical products and services. In other words, there may be a reason to suppose that the respondents are relatively interested in technology and knowledgeable.

3.1 Summary of responses

A total of 116 answers was received to the survey which is on average considered a good result in Botnia LivingLab experience.

The respondents to the APIS survey do not represent the population of Sweden fully, since they were recruited through Botnia LivingLab which is not a fully random selection from the population of the country. For example nobody under age of 18 answered the questionnaire (see Figure 1). Also persons over 60 are underrepresented compared to their proportion in the national population of Sweden. It can be said that middle-aged people are somewhat over-represented among the respondents. Around 2/3 of respondents were males, which reflects the overall gender balance in the Botnia LivingLab. The figure below shows the age distribution among the respondents.

<table>
<thead>
<tr>
<th>1. Jag år</th>
<th>Response Percent</th>
<th>Response Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>under 16</td>
<td>0.0%</td>
<td>0</td>
</tr>
<tr>
<td>18 – 29</td>
<td>13.8%</td>
<td>16</td>
</tr>
<tr>
<td>30 – 39</td>
<td>22.4%</td>
<td>26</td>
</tr>
<tr>
<td>40 – 49</td>
<td>25.8%</td>
<td>30</td>
</tr>
<tr>
<td>50 – 59</td>
<td>20.7%</td>
<td>24</td>
</tr>
<tr>
<td>60 eller mer</td>
<td>17.2%</td>
<td>20</td>
</tr>
<tr>
<td>answered question</td>
<td></td>
<td>116</td>
</tr>
<tr>
<td>skipped question</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 1. Age distribution of survey respondents

8 See Appendix 1.
The respondents participate in Botnia LivingLab surveys on a voluntary basis and it is plausible that this particular questionnaire on subjects like history and information attracted individuals specifically interested in these topics. This may mean that the answers are somewhat biased and more positive towards memory institutions and services based on their collections than could be expected on average among the population of Sweden. In other aspects there are no specific patterns or circumstances that differ significantly from the averages among population of Sweden.

3.2 Main survey results
The survey was constructed with statements for which the respondent had five available options:

1) agree totally
2) 
3) 
4) disagree totally
5) do not know.

Second and third choices express not a total agreement or disagreement level. We have interpreted the answers so that 1 and 2 correspond to “agree”, while 3 and 4 correspond to “disagree”.

There are no noticeable differences regarding the pattern of answers for using archives, museums or libraries. None of the institutions are frequently visited, but libraries more often than others. The result is the same about websites of memory institutions – these are seldom visited and library websites somewhat more frequently. The figure below shows how frequently the respondents are visiting archives.

<table>
<thead>
<tr>
<th>Question</th>
<th>Response Percent</th>
<th>Response Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varje dag</td>
<td>0.0%</td>
<td>0</td>
</tr>
<tr>
<td>Varje vecka</td>
<td>6.0%</td>
<td>7</td>
</tr>
<tr>
<td>Några gånger i månaden</td>
<td>2.6%</td>
<td>3</td>
</tr>
<tr>
<td>Mer sällan</td>
<td>55.2%</td>
<td>64</td>
</tr>
<tr>
<td>Aldrig</td>
<td>36.2%</td>
<td>42</td>
</tr>
</tbody>
</table>

Figure 2. Frequency of visiting archives

But even if citizens are not visiting memory institutions frequently, nor their websites, a little over 50% of the respondents were interested in history, regardless if it is history in general, national history or local history.

We also found the same proportion that were interest in new e-services based on historical information – around 50% were positive, and once again libraries got the highest score. The figure
below shows the answers to the statement: “I think new information services are needed in the field of archives that are not available on the market today”.

![Table](#)

**Figure 3. The perceived need for new e-services not available in the market today**

Between 30% and 40% of respondents felt that new information services from the cultural heritage sector are needed with the highest score for libraries.

The respondents were asked how they want to take advantage of the new e-services. It is interesting to see that the respondents want to use the new services mostly via internet:

![Table](#)

**Figure 4. The preferred means of accessing new e-services**

It is clear that it should be possible to access new services from memory institutions remotely, i.e. via Internet or mobile devices. Without remote access it would be impossible to fulfil citizens’ expectations and increase the use of memory institutions as a source for knowledge.
3.3 Expectations for new e-services

In the end of the questionnaire the participants were asked about new e-services they would like to see developed. The suggestions vary from vague descriptions of quality (e.g., “easy to find”) to more detailed descriptions (e.g. which existing services to build on). Around 20% of the suggestions are clearly geographical location-based services for hand-held devices; a further group of suggestions are linked with mobile devices, even though it is not clearly expressed. Overall, nearly half of the new e-services are expected to run on mobile phone platforms or in other hand-held devices.

The type of suggested services varies from simple search services to experience services, where one could “walk around” in an historic virtual environment. The latter is putting new demands on what information and how memory institutions provide the information to their users. Exactly what requirements this will put on memory institutions is a subject for further studies. All suggestions for new services collected through the user survey are listed in Appendix 2.

3.4 Conclusions from survey results

Our aim with the user survey was to get an understanding of the level of interest in e-services among the general public. Though citizens do not frequently visit memory institutions, they are generally interested in history and would like to gain better access to this information than they can today.

It is notable that 84% of the respondents want to get access to new services from memory institutions via the Internet, and 30% on mobile devices. Only 20% want to go to the institutions to use the service. This means that information in memory institutions has to be made easier to access in order to meet the citizens’ expectation to use the information remotely.
4. Expert Opinions

The APIS project group consists of archivists and IT specialists from the National Archives in the three countries. In order to get a broader view on the subject area of the project it was decided to gather opinions from professionals outside our project group. At the Nordic archivists’ IT meeting in Iceland a focus group was organized to discuss the project initial findings and ideas for e-services. Discussions with archivists in all partner countries were held at several stages of the project. At the beginning of the project, the discussions used more open questions; later on, the meetings were used more for testing our ideas and proposals, and gathering feedback on them. The feedback received from the archives community was very positive, although it was pointed out that the crowd-sourced information would have to pass through a rigorous quality control before it can become part of archival finding aids.

Meetings were organized with representatives from many small archives and museums in Sweden to ascertain their view on e-services and how these could support their functions. In general, the memory institutions and heritage sector does not have enough resources to meet the requirements of a digitized population. This is very clear in small archives and museums that have next to no resources for making the move into digital era. On the other hand, this makes them very interested in new, shared e-services and new models for collaboration with other institutions and their users.

The meetings with representatives of the IT sector demonstrated that there is a general lack of awareness what type of (digital) information archives hold and how it could be utilised for e-services. Developers of e-service software are primarily focused on open data from public agencies, but have rarely looked at historical data (photographs and maps from museums or libraries are practically the only examples of historic content involved in new e-services). The general perception appeared to be that involving content that cannot be always reformatted or repurposed, but needs to be delivered to end-users unchanged and with authenticity proof, will add extra technical complexity to the e-services, which will raise the cost of developing and offering such a service.

Discussions with representatives from the European Commission and EU-funded digital preservation research projects showed that there is a keen interest internationally to introduce new business models around preservation of and access to cultural heritage. Funds at the disposal of memory institutions are not sufficient to match the speed at which the rest of the society is becoming digital and learning to (re-)use information in digital form. There is a real danger that collections of memory institutions will be used less and less if the users, especially the current younger generation of future users, cannot be pro-actively engaged with the content that memory institutions hold. Making digital collections of archives more interoperable with other e-services will expose them to new types of users and is, therefore, well in line with the aims of the Horizon 2020 programme of the European Commission.

Based on the initial gap analysis and project group own ideas that were validated against the end users through a survey and through professional discussions with experts in the field, three e-service descriptions were developed. The e-service use case scenarios describe what they should achieve from the end-users’ (citizens’) perspective and outline the challenges archives are facing when developing such services.
5. Enhancing Archives’ e-Services: Three Views

The project formulated three use case scenarios for e-services that would engage digitised and digital content in archives with citizens. The general topic of the three scenarios is the same – making digital content that is currently difficult to find and retrieve from archives more accessible. The three scenarios use slightly different models and target users to achieve the same end result from the citizens’ point of view:

1) The first e-service is engaging directly with the general public and builds on crowd-sourcing model to create item-level description for archival records.

2) The second e-service is focused on archived databases and methods for disseminating these in a flexible way that meets the requirements of different user groups. This scenario would actively engage government agencies that have created the databases in the first place.

3) The third scenario provides a design for making content in archives interoperable with existing e-services that make use of open public data. The main partner for this scenario would be e-service developers.

5.1 Crowd-sourcing for archival description and records’ digitisation

5.1.1 Problem statement

The scenario tackles two major challenges for archives to be able to meet the citizens’ expectations:

1) Lack of archival description on individual record level
   Born digital records that are beginning to arrive at public archives are described on the level of individual records while the traditional archival description stops at the file or folder level. This results in uneven search results in archival finding aids for the contemporary (digital recordkeeping) and historic (paper-based recordkeeping) periods.

2) Digitisation of analogue collections
   Archival collections can become part of e-services only in digital formats. However, the majority of archival material remains in analogue format, especially in smaller memory institutions. Digitization of whole collection is costly and only a fraction of existing material can be digitized by archives themselves using their own funding.

5.1.2 Existing technologies and the proposed solution

New and low-cost models for enhancing the item-level description are sought by all memory institutions and many are utilizing crowd-sourcing for this. Crowdsourcing is a way to create a win-win situation for both end-users and archivists – both citizens and archives will gain from archival collections being better documented and supplied with correct metadata. The result of using crowdsourcing to create new archival description will be improved, web-based access to archival material for citizens, while at the same time supplying archives with metadata on individual records to enhance the archives’ catalogues.

Similar services are being experimented with in other sectors (see the EuropeanaConnect and Waisda examples, op cit.). Another form of using crowdsourcing is the European initiative Hack4Europe,\(^9\) where different groups/companies developed and presented different e-service prototypes over a short period of time. The main obstacle found in the Hackathons was finding accessible, well-described and free content from the cultural heritage sector.

Improving metadata quality is an on-going process in all memory institutions as most of their descriptions have, at some stage, been entered manually. Combining user-generated metadata from crowd-sourcing tools with existing finding aids requires rigorous quality control that cannot be done manually. Automation of metadata integration through controlled vocabularies, authority files and

\(^9\) http://www.europeana-libraries.eu/web/api/hackathons
validation routines is becoming a standard practice with an emerging market of software tools tailored specifically for memory institutions.

In order for citizens to become a crowd that helps both themselves and the archives, they need tools and guidance. Existing tools for using crowdsourcing methods do not solely rely on users’ free will to contribute, but often offer them something in return. Competitions, club-like organisations or acknowledgment of effort are frequently used to maintain interest in contributing. Cultural heritage institutions may engage in providing access to their digital collections for specific technical solutions.

A future scenario of how crowd-sourced information can be re-used in e-services is presented below.

You are in an unfamiliar town. What is there to do? Fancy a coffee? Where is the nearest café? It is easy to find via your mobile phone. A café nearby seems to have special offers. You also get directions on map and how far it is.

It was easy to get there because the directions were easy to follow. Once there, you find the café charming and would like to know more about its history. Check! And yes, there is something available – history of the café, its building and the events associated with the site. The house was built in 1882 by a farmer. Previously, there was a popular hotel on the site where the locals used to gather in the pub on Saturday nights. But the house burned down in 1879, and the new building was built to house a grocery store. The shop moved in 1956 to a new building a block away, and the building was used for various purposes, among other things, as tailors, a shoe shop, and a pizzeria. After a renovation and refurbishment in 1992 the café was re-founded, sharing the premises with a hairdressing salon. If you want to know more about past activities, just click on the appropriate link. There are also pictures from different eras that can be looked at.

Interesting, but now it is time to go to the railway station. On the way there you leave the information service in your mobile phone on. Continually it brings up information about the places you pass. There seems to be a story for each building and location, and it is amazing that there is so much interesting history. It certainly was interesting, you almost forgot the time where you were walking and checked into the historic backdrop of the road. Next time you go to this place, you will turn on this service at the arrival.

5.1.3 Discussion

The scenario above is essentially based on spatial information about objects, supplemented by links to the corresponding information. The sources are mixed from updated sources (as supply in the café) to extracts from old archives. And it is the historical information, which needs to be in the focus of the crowdsourcing exercise.

The historical information is gathered and brought together mostly from paper-based information in archives, museums and libraries, and made available by voluntary action. The compilation, which was also supplemented with different metadata, is a result of the voluntary work of historically-minded people, school classes work on native history, research associations, village associations, etc.

The collected and processed information is openly accessible and usable for the service providers that wish to use it. This means the users can access the same information from multiple applications. Since the information is tagged using international standards, the same application can be used in several countries.

A follow-up proposal to the APIS project – YEAH10 – will begin the development of tools and guidelines to make this vision a reality.

10 http://www.vinnova.se/sv/Resultat/Projekt/Effekta/YEAH/
5.2 Providing seamless access to archived databases

5.2.1 Problem statement

During the last decades a vast amount of public and private sector data has been collected, managed and used with the help of electronic databases. Part of this data is also of long-term value (retention period longer than 10 years) or even of archival value which means that it needs to be transferred to suitable long-term repositories and curated according to long-term preservation principles.

The current approach to database preservation in the national archives of Sweden, Iceland and Estonia is quite basic: the institutions rely on open, text-based formats to transfer and preserve the data. More technically speaking – all countries use a combination of XML, flat file and SQL technologies with minor differences in details. This approach allows preserving the most relevant components of the original database (database description, data itself, structure of the database, relevant queries and services) in a format which is secure against format and database platform obsolescence. In short – the approach is well grounded for long-term preservation purposes.

However, it also means that accessing archived databases is a rather complicated process and requires extensive technical knowledge, in particular the user needs to get into technical details of the database structure and content in order to access it. In addition, the tools currently available for accessing archived databases usually only allow uploading a single archived database into a modern database environment and executing pre-defined queries.

At the same time it is easy to understand that users of archived databases would also gain from the possibility of regrouping and restructuring the databases for different (research) purposes as well as in the possibility to deliver e-services that seamlessly combine data from both archived and active datasets. Current solutions make it hard to unlock this potential. More specifically the following obstacles can be detailed:

- descriptive metadata of the database and data are insufficiently described which makes it hard to recover all relevant datasets;
- technical description of data, data structures and relationships are currently insufficiently standardised and therefore do not allow for automated reuse of the data;
- database preservation techniques assume that the complex data structures applied in operational database systems are also kept in place during archiving. At the same time data structure complexity is one of the main reasons that makes the use of archived data difficult and highly technical;
- data must be used only according to access restrictions – publicly available data must be “cleaned” of sensitive data while authorised users shall be able to access also sensitive data relevant for their purposes. Currently the process of negotiating and giving access to databases that contain sensitive data is done only manually and therefore no automated and effective public access is possible.

5.2.2 Existing technologies and proposed solution

ICT technologies already available and/or under research provide some tools and principles that will help overcome the above issues:

- Data warehousing techniques: The problems related to accessing data from multiple and often highly complex database systems are also relevant in the management of active databases. For that purpose the concept of data warehouses and business information warehouses have been developed and significantly improved during the last years. In principle a data warehouse is a relational database that is specially designed for query and analysis instead of transaction processing. During the data warehousing process data structures are usually
simplified and presented in a subject oriented way. This includes often also de-normalisation of data. Data warehousing techniques could be applied both during ingest as well as access processes which would make the resulting data preserved at the archives or accessed by the users easier to understand and therefore also easier to preserve and use.

- Semantic data description: Using current tools in the area of semantic description would allow automating the combination of data originating from different (topical, spatial or temporal) datasets into a single query. According to current state-of-the-art semantic description requires the underlying creation or uptake of topic-based ontologies (as an example in OWL format) as well as creating crosswalks between those (as an example in SWRL or SKOS). According to suitable ontologies current technical and descriptive metadata could be enriched with semantic annotations (as an example as a combination of RDF and SA-WSDL).

- Semantic data access tools: If the data is annotated in a machine-readable way it would be possible to apply semantic query languages like SPARQL or SeRQL to allow for data access which would be independent of the underlying data structures or even repositories. As an example the technologies would allow to easily create and execute a query over properties in a certain region even if the underlying data is preserved as multiple separate snapshots with different preservation data structures.

- Access constraints definition: In order to automate the process of negotiating access to data there needs to be a generalised and formalised possibility for describing access restrictions. While principles on how to achieve this are available in archives already now there is a need for a more centralised approach and added specification of data description regulations. If such a (technical) regulation is available it would also be possible to build common automated access tools which could automate the process of “cleaning” the original data or negotiating access for authorised users.

5.2.3 Discussion
When applying the methods and tools described in the previous section it would be possible to widen the use of archived databases on multiple levels:

- Less need for technical knowledge: databases that are archived according to warehousing techniques have a highly simplified structure that is easier to understand to both archivists who take care of their preservation as well as for end-users;

- Unified approach to preservation: databases will not be preserved as individual snapshots but rather as interconnected semantic “units”;

- Simple to find and use: finding relevant information (databases) and creating queries is possible by the means of ontologies, users do not have to get familiar with the terminology used in each separate database and their technical structures. Additionally the methods allow for over-arching queries that include different databases or a wide time-span;

- Simple to re-use: as semantic description methods are getting more widespread also in active databases it will be easy to reuse archived databases to form common queries over both active and archived information.
5.3 Connecting archival content to e-government services

5.3.1 Problem statement

Archives’ digital collections are at present poorly being re-used in e-government services. The reasons for this are manifold, but primary hurdle is difficulty of accessing the collections, both technically (e.g., bespoke databases that manage digitised collections) and intellectually (varying description schemas for different collections that are managed separately). A typical digital database or digitised collection in an archive has its own platform (e.g., MS Access or Progress database) and has a specific metadata schema that holds the description of the collection. The description present in the collection’s database is often an enhanced subset from a more extensive description in the main archival catalogue or finding aid. For example:11

- A collection of digitised historical maps that are stored as image files on a file server with their metadata in a database that can be accessed only in the reading room of the archives. The maps may come from different archival fonds and the metadata database refers back to the main archives catalogue for description of the originating fonds.
- A collection of digitised photographs that has its own metadata schema describing both the photographer and the subject matter of photos.
- A selective collection of digitised records that were picked for a specific exhibition, are linked with a person, building, location, etc. The digitised collection is typically not exhaustive on the subject it covers but includes only highlights or most interesting records.
- A database with information copied from records (e.g. transcribed from original records) for a specific period, e.g. a database of property sales between 1926 and 1936.

Because many such databases use different metadata from the standard archival description, they are not accessible through a single common search interface provided by the archives. Therefore, locating and retrieving an item from these small collections is nearly always time consuming and cumbersome since their search tools are different.

For an e-service that is run from outside the archives, e.g. from a government agency or a private company, such archival collections are practically inaccessible. Technically they are invisible to web search engines, they can only be used through local search interface in the web browser and material can often be retrieved only from within the local area network (LAN) of the archive. In other words, it is difficult to link to these databases to harness their content with search and retrieval tools other than their own native platform and interface. Developing a separate interface for each such collection or database would be too costly and offer low ROI. Intellectually the main hurdles are non-standard and different metadata schemas that make it difficult to access the collections with a single query command; and separation of the context from items that can be accessed (i.e. part of the collection’s description is available only in the main archival catalogue). From the users’ point of view the different file formats, large file sizes and poor readability of images can be seen as problems when using digitised objects from archives. Users who access e-services on their mobile phones or other handheld devices may not appreciate large files that download slowly and are difficult to read on small screens.

5.3.2 Existing technologies and proposed solution

The problem of discreet digital collections and their bespoke database platforms is already being tackled in the museum domain where the concept of a smart museum has been developed to link up

---

11 For some examples of such databases see: http://rahvusarhiiv.ria.ee/ehaest/andmebaasid/ or http://www.ra.ee/vau/ (switch to English in the top right corner of the page); http://www.riksarkivet.se/default.aspx?id=6489&refid=1189; http://www.arkisto.fi/se/aineistot/; http://www.sa.dk/content/dk/temaer_til_inspiration/webudstillinger/kobstadskort/kobstadskort
disparate digital collections. In the records management and archives domain the concept of a metadata broker has been elaborated by a project at the Monash University in Australia. Other variations utilising established methods like metadata crosswalks or ontology-based search are found in other domains.

Similar to these examples, a software layer can be conceptualised to act as a metadata broker between various e-services and archival collections, databases and catalogues. The purpose of the software tool is to help e-services with finding content from disparate archival collections by “translating” the incoming queries to metadata elements typically used in archives. The results of the query would be mediated back to the e-service with standard query language and standard keywords attached. The software tool should act as a middleware between archive databases and citizen-oriented e-services, and does not need to be visible to the end-users. It should offer a standard query interface for incoming connections that various e-services could plug into. Ideally, it should also be open for any type of database with digitised items to be connected to the broker, whether it resides in an archive, museum, library or other type of memory institution.

5.3.3 Discussion
The described metadata broker would facilitate:

- overcoming the apparent isolation of archival content into discreet silos dependent on proprietary platforms and bespoke software;
- further standardisation of archival metadata and its interoperability;
- significantly improved access to archival content via other e-services that are already known to end-users;
- obtaining hitherto unavailable historical and time-series data.

The metadata broker would require:

- development of interfaces that allow external e-services to access archival collections;
- development of multilingual crosswalks between metadata used for archival collections and standard e-government metadata;
- existing e-government services that would benefit the end-users from being enhanced with archived data sources;
- very likely application of semantic software tools for metadata translation, especially in a multilingual setting.

---

12 See for example: [http://www.smartmuseum.eu/deliverables.html](http://www.smartmuseum.eu/deliverables.html)
13 See: [http://infotech.monash.edu/research/groups/rcrg/crkm/outcomes.html](http://infotech.monash.edu/research/groups/rcrg/crkm/outcomes.html)
6. Conclusions and Discussion

The first APIS white paper (Access to public information in governmental agencies and archives - main challenges) discussed what e-services should be developed from a citizen-centric perspective. We concluded that there are three main questions:

- What public information is of interest to citizens?
- How would they like to access this information?
- What kind/s of e-service/s would they prefer?

These questions were addressed in the user survey conducted by the project. From the answers it can be seen that there is sufficient interest in information in historical subjects among the general public (more than half of the respondents demonstrated active interest). There also exists a considerable group interested in gaining access to information in the memory institutions. The clearly preferred access method for this information is remotely, via internet or mobile phones. The suggestions for new services varied across the board and no clear message from the user survey can be identified.

The project also had professional discussions in the archival field, the broader ALM-sector and with the IT-sector. There was also contact with the European Commission and EU-funded projects in the field. This gave the project further input in forming the three scenarios for enhanced archival e-services of interest for the citizens.

A general problem that is addressed in all the developed three use case scenarios is the lack of usable metadata as a way to facilitate access to archival material. On the one hand there is a gap in the existing description of archives collections on the individual object level that sets limits to using the individual items. On the other hand there are digital records and records systems transferred to the archives with quality metadata also on the object level, like record management systems and databases. But even if this metadata exist there is a lack of tools and standards to reuse it and make it accessible. One reason for this is that metadata is archived in different schemas and platforms. In our scenarios these problems are addressed either by involving users into voluntarily supplying additional metadata and descriptions or proposing more advanced metadata reuse and mapping tools.

This goes well in hand with some of the practice oriented challenges which were identified in the first APIS white paper. Those were:

- securing that the description of transferred records is sufficient (especially record-level metadata);
- further develop methods for searching and retrieving single archival records.

The first APIS white paper identified e-services criteria suitable for archives: the solutions should be citizen-centric, generic and border-crossing. All three scenarios that were presented in this report can be said to fulfil these criteria. They are citizen-centric in the way that they would be built on the needs of end-users and would be developed in co-operation with them. They are generic because they are expected to be developed in co-operation of different stakeholders like universities, companies, archives and other cultural heritage institutions. And they are border-crossing because they will be suitable for different countries, like the ones in the APIS project.

The three use case scenarios share a common outcome – citizens will gain improved and wider access to information in archives. This is partly achieved by the use of new technical tools, partly by linking archival collections with existing and new e-services, and partly by engaging the general public more pro-actively in generating digital content in archives.

However there are some visible differences between the scenarios. The second and third scenario focus on making existing collections more accessible by technical and semantic development and
standardisation. The first scenario will increase access by adding metadata to the collections and also by having more documents scanned into digital form. Scenario two and three are executed with a focused strategic development by software developers and archival experts. The first scenario to some extent also requires development by IT experts but mainly relies on the work of volunteers.

The first APIIS white paper was also emphasising challenges arising from different legal context in different countries. This area is covered in all three scenarios but scenario two will have to resolve more issues than the others since public data is often more tightly regulated than textual or photographic sources.

The first scenario is, perhaps, the closest to citizens as it engages them actively in content creation through crowdsourcing, offers improved access to content and services (e.g. content-enriched storytelling on the web) to them in return for their effort. Citizens who have never been in contact with original archival documents may discover their value and find a new interest. This scenario will also offer a way to engage members of the public in digitisation of archival records. This is a need which has been strongly emphasised by small archives and museums that the APIIS project surveyed.

Based on an overall evaluation of the three options the project group decided to propose the first scenario as a follow-up project to the APIIS project. The scenario will be implemented within the new project YEAH (You! Enhance Access to History). Other scenarios will be developed into concrete project proposals for other funding streams.
References


Appendix 1: Survey Questionnaire

Enkät om e-tjänster som bygger på kulturarvet

Denna enkät har som syfte att ta reda på medborgarnas intresse för och syn på den historiska information som finns i våra arkiv, museer och bibliotek.


Kontaktperson:
Mari Runardotter, projektledare API

DEL 1. Personfrågor

Ålder
☐ Under 18
☐ 18 – 29
☐ 30 – 39
☐ 40 – 49
☐ 50 – 59
☐ 60 – och över

Kön
☐ Kvinna
☐ Man

Jag bor i
☐ Storstad
☐ Medelstor stad
☐ Liten stad
☐ Landsbygd

Jag är
☐ Arbetssökande
☐ Student
☐ Anställd
☐ Sjukskriven
☐ Pensionär
☐ Egen företagare
☐ Annat, nämligen.................
**DEL 2. Historia**

**Historia generellt**
Hur väl stämmer detta påstående på dig? ”Jag är väldigt intresserad av historia överhuvudtaget”.

- [ ] 1 helt rätt
- [ ] 2
- [ ] 3
- [ ] 4 helt fel

**Svensk historia**
Hur väl stämmer detta påstående på dig? ”Jag är väldigt intresserad av svensk historia”.

- [ ] 1 helt rätt
- [ ] 2
- [ ] 3
- [ ] 4 helt fel

**Bygdens historia**
Hur väl stämmer detta påstående på dig? ”Jag är väldigt intresserad av min hembygds historia”.

- [ ] 1 helt rätt
- [ ] 2
- [ ] 3
- [ ] 4 helt fel

**DEL 3. Arkiv, museer och bibliotek**
Hur väl stämmer följande påståenden på dig?

**Jag besöker arkiv**

- [ ] Varje dag
- [ ] Varje vecka
- [ ] Några gånger i månaden
- [ ] Mer sällan
- [ ] Aldrig

**Jag besöker museer**

- [ ] Varje dag
- [ ] Varje vecka
- [ ] Några gånger i månaden
- [ ] Mer sällan
- [ ] Aldrig

**Jag besöker bibliotek**

- [ ] Varje dag
- [ ] Varje vecka
- [ ] Några gånger i månaden
- [ ] Mer sällan
- [ ] Aldrig

**Jag besöker ett arkivs webbsida**

- [ ] Varje dag
- [ ] Varje vecka
- [ ] Några gånger i månaden
Jag besöker ett museums webbsida
☐ Varje dag
☐ Varje vecka
☐ Några gånger i månaden
☐ Mer sällan
☐ Aldrig

☐ Om ja, vilken information letade du efter?---------------------------

Jag besöker ett biblioteks webbsida
☐ Varje dag
☐ Varje vecka
☐ Några gånger i månaden
☐ Mer sällan
☐ Aldrig

☐ Om ja, vilken information letade du efter?---------------------------

DEL 4. E-tjänster

E-tjänster ger dig ökade möjligheter att söka information i arkiven direkt via Internet. Det kan exempelvis vara att via din mobil söka historisk information om platsen där du för tillfället befinner dig.

Inställning till nya e-tjänster som bygger på arkivmaterial
Jag skulle beskriva mig själv som
☐ 1 oerhört intresserad
☐ 2
☐ 3
☐ 4 inte alls intresserad av e-tjänster inom detta område

Inställning till nya e-tjänster som bygger på museimaterial
Jag skulle beskriva mig själv som
☐ 1 oerhört intresserad
☐ 2
☐ 3
☐ 4 inte alls intresserad av e-tjänster inom detta område

Inställning till nya e-tjänster som bygger på biblioteksmaterial
Jag skulle beskriva mig själv som
☐ 1 oerhört intresserad
☐ 2
☐ 3
☐ 4 inte alls intresserad av e-tjänster inom detta område
DEL 5. Nya arkiv, museum och bibliotekstjänster

Nya tjänster och produkter
Jag känner ett behov av nya informationstjänster som inte finns på marknaden idag inom arkivområdet

☐ 1 helt rätt
☐ 2
☐ 3
☐ 4 helt fel

Jag känner ett behov av nya informationstjänster som inte finns på marknaden idag inom museiområdet

☐ 1 helt rätt
☐ 2
☐ 3
☐ 4 helt fel

Jag känner ett behov av nya informationstjänster som inte på marknaden idag inom biblioteksområdet

☐ 1 helt rätt
☐ 2
☐ 3
☐ 4 helt fel

Hur vill du nå de e-tjänster du är intresserad av?

☐ 1 Vill ta del av dem på arkivet/museet/biblioteket
☐ 2 Vill ta del av dem via Internet
☐ 3 Vill ta del av dem via min mobil
☐ 4 Annat, nämligen...........................................................................................................

TACK! för din medverkan!
Appendix 2: User Suggestions for New e-Services from Memory Institutions