arbido

Aas Kuldar, Deputy Director, Digital Archives of the National Archives of Estonia

Tags: Archiv, Elektronisch

Developing the « archive it » button

Efficient, seamless, fast, automated – these are some of the words characterising current expectations towards services delivered by governments. Let's have a look at how the National Archives of Estonia is trying to keep up with such expectations in the era of born-digital information.

Background

Have you heard that Estonia is doing everything digitally? If not, then just to let you know that the country has put a lot of effort into digitising most aspects of its public sector. By now, 99% of public services are available through the internet. Most of these have also taken the next step to be "digital-by-default" and are moving towards zero-bureaucracy. For example, we discuss that instead of citizens having to apply for childbirth benefits, these should be assigned automatically once hospital staff registers the birth. [1]

In records and information management it's worth to know that the public sector is creating way more information digitally than on paper. A study carried out in 2014 evaluated that less than 5% of information is created in analogue ways, but also that about three-quarters of digital information is created within structured databases and information systems as opposed to document formats like docx or pdf.

The legal framework of Estonia requires the provision of proactive and efficient access to public information. The Public Information Act states explicitly that all unrestricted public information must be proactively published, and that restricted information about individuals must be available as easily as possible.

In real life, all of the above means that the vast majority of Estonians are very much used to declaring their taxes online within five minutes, establishing companies within half an hour, and having access to relevant information within seconds.

Archiving a digital state

The National Archives of Estonia (NAE) started looking into the development of its digital archiving infrastructure in 2005. At this point it was already clear, that any technological solution to be developed has to follow the same underlying conceptual principles as described above. Most crucially, the transfer of records to archives should not make it worse for citizens to access the information. In other words, a piece of information, which has initially been accessible on the website of an institution, has to be accessible on the website of the archives almost immediately after its transfer.

These requirements led NAE to the understanding that the transfer of born-digital content has to be extensively automated, and that the reuse of available metadata for archival description purposes is one of the most crucial aspects to tackle. In 2007 NAE delivered a software tool called Universal Archiving Module (UAM) which allows agencies to map original metadata created within records management or business systems to archival description standards, validate and amend it if necessary, and deliver very detailed archival descriptions with minimal effort.

By now NAE has more than a decade of experience using UAM in a variety of born-digital transfers from a number of public sector institutions. So what are the practical lessons learned?

Lessons learned

The most important lesson from using UAM was that automation is heavily dependent on the input **data quality**. For example, we have encountered a number of cases where titles of records are misspelled or misleading (leading to the record not being possible to be reasonably found), or even worse – correct information about restrictions is missing in original metadata. Obviously the last case leads to the automated publication of restricted information and ultimately to the violation of individuals' rights.

The options to solve the situation retrospectively are rather limited. Of course, it is theoretically possible to require the archivist to manually check most crucial metadata but taken the amount of information delivered to the archives, it is simply too resource-demanding in practice. There are also technologies in the areas of automated language processing, content analysis and reasoning which might help to detect most errors. However, pilots undertaken by NAE and next to the Archives Portal Europe revealed that for now these technologies are still too immature for most European languages, including Estonian.

Therefore NAE has, for now, concentrated on working on improving data quality proactively by providing consultation, training and practical guidelines for the initial creation of metadata at agencies. While the level of detail which has to be addressed in such consultation has initially been described as overwhelming by archivists, we can already notice some improvement.

The second lesson worth noting is that automation ultimately leads to the **loss of flexibility**. It is rather simple to change little aspects of manual archival description on the go, according to the nature of specific archival records or submitting agency (and at least Estonian archivists are rather used to being a bit "innovative"). Automated metadata reuse, on the other hand, follows strict pre-described routes and any exceptions have to be programmed in advance. As such, even the smallest wish to deviate from the automated route can lead to time-consuming and costly software developments.

Ultimately, the effect of automation is that processes need to be analysed and thought through very carefully and in much detail. At NAE we also try, whenever possible, develop software which allows for manual intrusion in crucial steps of processes which are by default automated.

Closely related to the previous lesson is the third one - **long-term cost of technology**. It does not matter how well you have designed and built your automation software, as everyone close to digital preservation knows, sooner or later it becomes obsolete either technologically (i.e. underlying technical components like databases, programming languages, security mechanisms etc. become unusable) or conceptually (i.e. the underlying concepts and processes have changed). This means that once you automate a process and create a piece of software to support it, you also take the responsibility to constantly update and maintain it. Even worse, a typical archive has to deal with a large variety of different data types ranging from usual office file formats to architectural drawings, social media and interactive content. All of these might require different pieces of software to manage, meaning that the institution has to tackle the long-term maintenance of tens of different tools, leading to serious long-term IT costs.

The most straightforward way to overcome this issue is collaboration. Estonia started to centralise its digital archiving activities in around 2010 and by now there is only one digital transfer, preservation and access infrastructure shared by all archival institutions. At around the same time NAE also started to push for international collaboration which led to the execution of the international [E-ARK project](#) in 2014 – 2017. The sole purpose of E-ARK was to standardise some of the most crucial aspects in archival transfer, preservation and access, and to develop internationally reusable software components for these processes.

Summary

Automation is, with no doubt, relevant and inevitable in the current situation where more and more information has to be processed at increasing speeds. However, the experiences of the National Archives of Estonia have shown that there are multiple shortcomings associated with automation. In our opinion the most crucial ones are related to the quality of automated processes and data, the loss of flexibility and the long-term cost of associated technology. I would also like to note, that it has been rather difficult for Estonian archivists and information managers to "reset their minds" – think about archiving as a set of abstract but detailed processes, as opposed to a broad activity being executed slightly differently for any given set of archival records.

The description above should also be seen as the start of a long road. A few years ago we defined our ultimate long-term goal to be an "archive it" button which allows agencies to send information of archival value to public archives within minutes. But indeed – there is still a lot of work to be done and more lessons to be learned!

1 for more information on e-Estonia visit www.e-Estonia.com.

Kuldar Aas

Deputy Director, Digital Archives of the National Archives of Estonia

Abstract

English
Automation is, with no doubt, relevant and inevitable in the current situation where more and more information has to be processed at increasing speeds. However, the experiences of the National Archives of Estonia have shown that there are multiple shortcomings associated with automation. In our opinion the most crucial ones are related to the quality of automated processes and data, the loss of flexibility and the long-term cost of associated technology. I would also like to note, that it has been rather difficult for Estonian archivists and information managers to "reset their minds" – think about archiving as a set of abstract but detailed processes, as opposed to a broad activity being executed slightly differently for any given set of archival records.

The description above should also be seen as the start of a long road. A few years ago we defined our ultimate long-term goal to be an "archive it" button which allows agencies to send information of archival value to public archives within minutes. But indeed – there is still a lot of work to be done and more lessons to be learned!

Deutsch
In der heutigen Zeit, wo mehr und mehr Informationen immer schneller verarbeitet werden müssen, ist die Automatisierung zweifellos ein wichtiges und unumgängliches Thema. Die Erfahrungen der Nationalen Archive von Estland zeigen jedoch, dass es in diesem Bereich noch viele Defizite gibt, vor allem was die Qualität der automatisierten Prozesse und Daten, der Verlust der Flexibilität und die langfristigen Kosten der entsprechenden Technologien betreffen. Eine weitere Schwierigkeit für die Estnischen Archivare war es, den Vorgang des Archivierens nicht mehr als ein dem jeweiligen Archivgut angepasster Vorgang zu sehen, sonder als Teil eines abstrakten, detaillierten Prozesses.

Dieser Erfahrungsbericht ist zugleich auch der Start eines noch langen Weges. Vor ein paar Jahren wurde ein "archive it" button, welcher es den Behörden erlaubt, archivwürdige Informationen innnert Minuten an die öffentlichen Archive zu senden, als langfristiges Ziel definiert. Die Umsetzung benötigt jedoch noch viel Arbeit und es gibt noch viel zu Lernen.