One repository solution doesn't fit all

Towards a distributed and collaborative approach to digital preservation services at the Amsterdam City Archives

Ana van Meegen Silva
Stadsarchief Amsterdam
The Netherlands
A.van.Meegen@amsterdam.nl

Abstract – Different digital objects have different preservation requirements involving strategies, policies and practices that go beyond the capacity of a single archival institution. The variety and complexity of digital objects requires specific knowledge for ingesting and the preservation of file formats. Large data files call for a different storage set-up than the approach required for the preservation and distribution of small objects. Building an organizational and technological infrastructure that can cope with the diversity and complexity of digital objects involves an effort that far exceeds what a middle-sized institution can do alone. This paper describes how the Amsterdam City Archives aims to collaborate with other non-profit institutions and partners to improve the quality of the preservation of and access to digital objects while reducing costs.

Keywords – digital preservation services, collaboration, distributed approach, archives repository.

Conference Topics – Collaboration; Exploring New Horizons.

I. Introduction

The Amsterdam City Archives (*Stadsarchief Amsterdam*, SAA) is the historical documentation center of the city of Amsterdam and forms an integrated part of the Amsterdam City Council. The institution is responsible for the City Council's archives and holds records of private institutions and citizens of Amsterdam. The collection is based on two perspectives: information about the city and records formed by the citizens of Amsterdam. The institution currently has approximately 120 terabytes of digital objects, including 28.5 million digitized documents and images, 8 million digital-born objects, and 15,000 moving images and sound files.

In 2020, the SAA will be implementing a whole new digital infrastructure that will renew its archives repository, collection management system and discovery platform. These changes are needed to comply with European regulation and modern technology. SAA uses the Open Archival Information System (OAIS) model as conceptual framework for the information architecture of it's infrastructure. In the period leading up to the roll out of the infrastructure, a comprehensive evaluation of different solutions for the management of digital objects was carried out. The main challenges for the archives repository were the accessibility performance of digital objects versus a low cost storage solution that can provide the preservation functionalities required. During this preparation period, SAA worked with different use cases that reflected particular situations to which the new infrastructure should provide a solution. After investigations, two of the use cases were put aside and new solutions through collaboration are now being explored. In this paper I will explain why these two use cases cannot be integrated in the new infrastructure and how collaboration with other organizations can be a solution for digital assets that do not fit an archival infrastructure based on small-sized digital objects.

II. Exploring collaborations

SAA's digital infrastructure is a standard structure that supports elementary digital object management, from preservation to use. The diversity of digital objects, however, does not always fit in this infrastructure. SAA is working with two cases that go beyond their current capabilities: a cost reduction storage solution for large-sized objects and an integrated solution for preservation and

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

discovery of Council meetings. For both cases, SAA is exploring collaborative solutions, each with their own distinctive features. Different institutions use a shared digital infrastructure, assigning different components of the OAIS model for each institution [1]. In the collaborative cases of SAA however, the collaboration is not a shared digital infrastructure but a service purchase where the providers facilitate parts of the digital infrastructure.

A. Preserving large-sized media files

The Migrant Television Netherlands [2] (MTNL) was a foundation that created television programs with a focus on topics that concerned migrants and migration in the Netherlands. MTNL produced programs from 1984 to 2013, and broadcasted in the four major Dutch cities through local broadcasters. The collection contains a total of 4,000 hours of material. The organization stopped their activities and donated their archives to the City Archives of Amsterdam, Rotterdam, The Hague and Utrecht. During this transition to the archives it was decided to digitize all of the collection's VHS tapes [3]. The digitization was carried out by the Netherlands Institute for Sound and Vision (*Nederland Instituut voor Beeld en Geluid*, NIBG) and for each tape they delivered two files: an HD master file and a low resolution file meant for online purposes. The low resolution files are stored and preserved at the Amsterdam City Archives. It was decided to store the master files temporarily at the NIBG until SAA implements a new repository. The master files can be accessed through the NIBG website [4]. SAA is still working on a solution for the presentation of the low resolution files on its website.

During the preparation period for the renewed digital infrastructure project at the Amsterdam City Archives, the institution investigated possibilities to archive the MTNL master files at the organization. The purpose of the analysis was an integrated approach for all SAA digital objects, knowing that more archives such as those of the MTNL are likely to follow and a solution will need to be in place to accommodate those collections.

For the preservation and metadata management of the files the organization will not require extra functionality; the currently available and renewed systems are also suitable for large files. The organization would, however, need a differentiate workflow for accessibility; these online requests could be easily arranged. The real problem was the storage solution. The analysis showed that a tape storage solution is, at present, the best method to house large-sized object files with lower storage costs. This solution, however, is not suitable for small-sized objects requiring a quick response time and daily use. For an integrated in-house storage approach the organization would need to host two different storage solutions and would therefore not be able to achieve the intended cost reduction.

The best solution remains the storage of the master files at the NIBG. The NIBG is responsible for the archiving of the collections of Dutch broadcasting corporations and is one of the largest media archives in Europe. With more than 17 petabyte of stored collections and more than 3 million online media assets, the NIBG is one of the main experts on media archives in the Netherlands and they have a repository solution that is suitable for the archiving and preservation of large-sized files. The institution has been using tape storage solutions in combination with OAIS-compliant functionalities for preservation and has a Date Seal of Approval certificate [5]. A management storage hierarchy functionality was put in place for the different storage workflows maximizing the data monitoring and the performance capacity [6]. These qualities and storage solutions make the NIBG the best collaborative partner for the preservation of the MTNL master files. There are, however, issues that need to be addressed. Besides the arrangements made for the storage and availability of the files through NIBG's website, there are no arrangements for usability of content, metadata enrichment, statistics on usage frequency and incidents/problems, etc. The roles and responsibilities beyond the preservation of the files have not yet been defined; neither organization knows what they can expect of the other. A better collaboration agreement that defines the mutual objectives with regard to the collection, the future improvements in both organizations and the service level expectations should be discussed and implemented

B. Preservation of and access to council information

Decisions made by the board of Amsterdam City Council are publicly available. Every citizen can follow

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

live streams of the council's meetings and can access meeting agendas and preparatory documents. This rich content information on Amsterdam's decision making is made available through an online information platform [7] hosted by NotuBiz, the vendor that records the video minutes. NotuBiz offers an integrated platform solution that follows the whole political decision-making process. Citizens can view the agenda of upcoming meetings and the submitted documents beforehand. The archive module enables the user to search or browse through different categories such as date, subject, the name of a councilor or document, etc.

In 2017, the council's clerk contacted the City Archives. The eight-year contract period with NotuBiz was coming to an end, a new European tendering process was starting and it was unsure whether NotuBiz would be the new facilitator of the video minutes and online platform. SAA is responsible for the archiving and preservation of all the council's digital assets, including the video minutes. A problem analysis was conducted and the key conclusion was that SAA is very capable of archiving and preserving all the digital objects available on the platform, but that it does not have a platform that can integrate the different digital objects in a user-friendly way. A solution could be to provide a platform that has the same functionalities as the NotuBiz website, but that would mean that Amsterdam City Council would pay twice for the same functionalities. Another solution was to archive the video minutes at SAA and make them available through the new councilor's platform. This solution had the most chance of being implemented until the news was announced that NotuBiz had won the tender and had signed a new eight-year contract with the City of Amsterdam.

During the search for an archival solution for the video minutes, NotuBiz proposed that they facilitate the preservation management of the digital objects in collaboration with SAA. SAA would give advice on and monitor the preservation management implementation at NotuBiz. If the implementation proves successful, the NotuBiz services will be expanded to include the preservation module. This solution is not yet definitive, as both organizations need to think it through and investigate the impact that this model can have on the organization. The organization focusses on creating and sharing content and prior to the discussions with the Archives it had no idea of the problems and concerns of preservation management. The content is stored correctly and the organization takes care of the conversion of objects for accessibility purposes but there are no preservation policies and regulations in place. An external consultant conducted an OAIS analysis on the products of NotuBiz and made recommendations for improvements. The organization is now seriously considering the next steps and how these improvements will fit into its business model. For the Archives it means that they will start a collaboration with an organization that is not OAIS-compliant and that both organizations will work together to achieve the intended preservation goals.

III. Opportunities and challenges

Digital archives are diverse and therefore complex material resulting in a variety of possibilities and choices for storing, describing, preserving, distributing, discovering and accessing the information. Different choices entail many different opportunities and flexible solutions. The Amsterdam City Archives will be switching from an integrated in-house solution to a more flexible distributed approach for the preservation of digital objects, where collaboration is the key to success. But we are not there yet; there are various issues that need to be further analyzed and discussed.

Firstly, it is important to determine what the purpose of the organizations is and what results they wish to achieve for the collection and whether this fits into a collaborative partnership. The choice for collaboration was not a strategic one but more a pragmatic one, that fit the solution required for the arisen problem. As part of the project for renewing the digital infrastructure, SAA investigated whether there were other solutions for these use cases at the technical level and concluded that it was best to move towards a collaborative approach. A storage solution is, however, not the whole picture and a lot of work will still have to be done. In addition, the other organizations will also need to decide which future steps they want to take.

Preservation of digital objects is the core business of SAA and needs to be conducted and documented well. NIBG is a non-profit OAIS-compliant

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

organization that shares the same preservation principles as the Archives. This is not the case for NotuBiz, but that should not be an insurmountable problem. In a collaborative setting, the OAIS compliancy is a joint venture between both partners. For both organizations this means that they need only be partly compliant, but together they will form a complete match and will accommodate all the requirements needed for the preservation governance. There will also be overlap in the governance structure, which is needed to guarantee that both organizations see digital preservation as an important activity that ensures the durability and accessibility of digital objects.

Another issue that needs further analysis for preservation is continuity. Continuity of services for the long term is one of the core requisites for an organization working with sustainable archives. In a collaborative setting you can only guarantee continuity for as long as the collaboration lasts. This issue is no different than when an organization chooses for standard third-party services. It is important that expectations and legal requirements are explained and documented. In addition, an exit strategy should be agreed between the partners beforehand. In the governance structure it is important to mention that the archives institute will always be responsible for the continuity of the services if and when the partnership comes to an end.

Ensuring content usability and access is part of the core business of the Amsterdam City Archives. For both use cases the other organization is responsible for the discovery and access of the collection that they manage. In both cases, the Archives have until now had no influence on the presentation and access of information. For the Archives it is important to determine how much influence the organization wants to have in decisions about the description and usability of the collection. It is also important that users who start their search on the Archive's platform are also able to find the collections available on the other platforms. The exchange of metadata and/or content is a topic that should be further explored.

There are many differences in collaborations with a commercial partner or a non-profit institution. Both have their advantages and disadvantages, and for both it is important that the roles, responsibilities and expectations are well documented and that both partners know what to expect of each other. An evaluation should take place from time to time, where users' reports, technological developments and future perspectives are assessed. These topics will be on the agenda over the coming period so as to intensify collaboration and improve the collection's preservation and usability.

IV. Conclusion

The Amsterdam City Archives is convinced that a distributed collaborative approach for parts of the collection is a welcome solution for specific problems entailed by some collections. By making use of the knowledge and infrastructure of others the organization can improve preservation management, reduce costs and improve the collection's usability and user experience. SAA is working with two cases where collaboration seems to be the best outcome. However, there are still some challenges that need to be further explored.

In a collaborative setting there are several topics that should be discussed to avoid disappointments and miscommunication. Firstly, the organization should identify internally which goals and results they want to achieve within the collaboration and how they want to manage gaps and employ strategies to address them. Secondly, and most importantly in a collaborative setting, expectations should be managed and documented. In a partnership with a commercial organization it is not unusual to talk about service level management and agreements but for non-profit organizations this could lead to an uncomfortable situation.

Collaborations can help non-profit organizations professionalize their services for a larger community. For SAA, it is also important that they can make it clear to their partners which expertise or services, as well as financial compensation, they will contribute to the collaboration. In both cases, the collaboration with SAA is not a regular service and there is therefore no standard solution or service in place. SAA and its partners are still working on implementing and further improving the collaboration, with each case having its own challenges and opportunities.

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

References

[1] G. Knight and Mark Hedges "Modeling OAIS compliance for disaggregated preservation services" The International Journal of Digital Curation, volume 2, no. 1, pp. 62-72, 2007. M. Day "Toward distributed infrastructures for digital preservation: the roles of collaboration and trust." The International Journal of Digital Curation, volume 3, no. 1, pp. 15-28, 2008.

[2] See the collection at the website of the Stadsarchief Amsterdam https://archief.amsterdam/inventarissen/inventaris/20002.nl.html

[3] The description of the MTNL Amsterdam Archive Collection is available at https://archief.amsterdam/inventarissen/inventaris/20002.nl.html.

[4] The collection is available at NIBG's website https://zoeken.beeldengeluid.nl/search/program?limit=10&offset=0&q=mtnl%20amsterdam&sort=-nisv.sortdate&x=false

[5] For more information see the "Digital Preservation Policy 2016" http://publications.beeldengeluid.nl/pub/388

[6] D. Steinmeier "Storage binnen OAIS: normatief model en GAP analysis voor Beeld en Geluid", 2013. Online available at: http://publications.beeldengeluid.nl/pub/408. Please note that a major renewing infrastructure project ran in the organization in 2018 and that this publication does not reflect the current situation, but the storage hierarchy functionality is still in place.

[7] See the platform at: https://amsterdam.raadsinformatie.nl.

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.