Digitization with Use of Principles from the World of Industry

M. Holtman & N. van Zeeland, Public Services, City Archives Amsterdam, Amsterdam, NETHERLANDS

Abstract

"All archival research should be possible 24/7"

In 2005, the Amsterdam City Archives set its digital services department the ambitious target of making it possible to access its entire archive at all times. This meant that we needed to start digitizing our archives and collections on a large scale.

An important part of our approach is customer-driven: through a scanning-on-demand service, the customer decides which archival documents are digitized. Next to scanning-on-demand, we work on a project basis in which we digitize entire archives or collections, of larger amounts of uniform material.

Ten years on, we have met 40 000 client orders and produced 20 million scans, all online for everyone to use. Meanwhile, the customer demand for digitized items continues to increase. At the same time, large digitization projects come our way more often. This requires constant adjustment of our digitization approach and work processes.

In 2018, we meet a new challenge: producing 20 000 scans a day. On this scale, an industrial approach as described in this article is not an option anymore, but a necessity.

Digitization with Use of Principles from the World of Industry

The Amsterdam City Archives

The Amsterdam City Archives holds over 50 kilometers of archival documents, ranging from the municipal administration to church archives and archives of Amsterdam families and organizations. Some famous examples are the archives of Heineken and the Holland Land Company. The Amsterdam City Archives also holds an important collection of historical photographs, drawings and film material.

The archive is located in the heart of the historical center of Amsterdam, housed in an impressive building called *De Bazel*, after it's architect's name. On a yearly basis roughly 100 000 historians, archival researchers, genealogists and people with a general interest in the history of the city of Amsterdam visit the archives, both for archival research and for visiting exhibitions, readings and to participate in educational activities. With these activities the Amsterdam City Archives aims to interest a broad audience in the rich, dynamic history of the city. The importance and relevance of the archives is recently highlighted by adding the Amsterdam notarial archives to the Unesco Memory of the World Register.

Large scale digitization

From around 2005 it became clear that although a beautiful and welcoming building is still important, a building alone is not sufficient anymore. A lot of our visitors wanted to consult the archival documents online. In other words, beside the bricks we needed bits. The Amsterdam City Archives set its digital services department the ambitious target of making it possible to access its entire archive at all times, online.

Even with an optimistic calculation it was clear that digitizing the archives from A to Z would take several hundreds of years to accomplish. To solve this problem we developed a two track approach:

- Customer-driven: through a scanning-on-demand service, the customer decides which archival documents are to be digitized

- Project-based: we digitize entire archives or collections, of larger amounts of uniform material

For making this approach possible we had to start thinking 'large scale'. Both the production costs and structural costs of digitization had to become as low as possible. One of the important decisions we then made was to set the standard for image quality at legibility. In order to reduce structural costs we decided to store all scans as JPEG files, and not as TIFF (which was the custom at that time).

From the beginning an important starting point was that all scans produced, should be visible and thus are put online in the web systems, available for all users. This way slowly, but steadily we create a long tail of digitized archival documents. Regarding scanning-on-demand: as long as we deliver all requested scans within a reasonable time, we state that there is no backlog in digitization.

We invested in a user friendly web environment with different solutions for different target groups, extensive search facilities and easy to use document viewers.

Our approach was a success from the beginning. In ten years we have met 40 000 client orders and produced 25 million scans, which are all online for everyone to use. Are all our customers satisfied now? No. Since doing archival research online becomes more custom and easy, we see an increasing demand for digitized documents. Recently we raised the weekly amount of scans produced in scanning-on-demand to 20 000 scans a week. Are we satisfied now? We were, until it became clear that our scan production rate had be raised to even higher levels.

Two projects that made us rethink the way we manage our digitization projects

Two separate events (with corresponding money flows) followed each other closely and made it possible to multiple our scan production from 20 000 scans a week in 2015 to 20 000 a day in 2018.

In 2016 the City Archives Amsterdam received a large grant with which three kilometers of the notarial archives is being digitized. This digitization project is part of a scholarly project called Golden Agents, that researches the creative industry in the seventeenth century (Golden Age) in Amsterdam.

To facilitate this project, we will produce 1.5 million scans from the notarial archives each year. In a way, this is also scanning-on-demand, but on a completely different scale than before. To support the scholarly research, the selection of this project is pinpointed on certain periods and notaries and serves as a broad breeding ground for several socio-historical research questions.

Secondly, the upscaling of our digitization capacity got an enormous boost when we got green light to build a second repository. Our main building in the center of Amsterdam reaches its maximum capacity, so we will build a new repository in the north of Amsterdam. In conjunction with the construction of this repository, we received a generous grant for the digitization of a number of archives that need to be moved. This will take place in 2018 and 2019. The challenge is to digitize 1.5 kilometers of archives within two years. This is equal to the production of about ten million scans.

With the scanning on demand and digitization of both the notary archives and the archives that have to be moved to the new repository our production rate increases to an amount of 20 000 scans per day. On such a large scale we have to avoid making the process too onerous for our organization (and rather small team), without compromising quality. It was clear that, although we already had a standardized workflow, some things have to be done differently.

To meet this challenge, we adopted principles from a world in which large scale and efficient workflow is daily business: the world of industry. Industry in this regard does not imply smoking chimneys, but a process that remains manageable even on a very large scale, where quality is optimized, and costs remain as low as possible.

Basic digitization principles

To understand the implication of this industrial approach, we first have to take a step back. Over a period of almost 10 years, the Amsterdam City Archives had developed a streamlined, scalable digitization infrastructure. All digitization takes places within this infrastructure, following a single standardized approach.

This work process has been kept as simple as possible and uses existing tools wherever possible. Steps such as file name checks have been automated using scripts which have been developed in-house. Other procedures include monitoring and reporting at relevant intervals and spots, which provide information required to perform a check or carry out another.

Our general principles for digitization projects include:

- Digitization is always project-based

- All digitization, no matter what kind of project, follows the same process

- Before starting a project all details and instructions are covered in a project plan

- Metadata is always added before the digitization project, not during digitization

- Pre-processing only on a minimal level

- Digitization of the original document as is, so no reordering, adding page numbers and so forth

- Digitization only of complete inventory numbers

The main advantages of this approach are the scalability and consistency. Because all digitization projects follow the same principles and workflow, we can work with different teams and suppliers, without huge changes. Do we need to execute an extra project? We hire a temporary worker and contract another digitization company. Do the requests for scanning-on-demand suddenly gain more popularity? Only a new agreement with the digitization company is needed. Further changes are unnecessary.

By approaching digitization as a more or less isolated process, we can create a clear workflow and consistency in the end result. We do not bother with metadata, further indexing of the documents, restoration, reordering et cetera. These are all activities that are executed before or after the digitization project, never during. This makes clear what we can and cannot do and what the end user can expect from the scans they encounter on one of our websites.

Integrating principles form the world of industry in our working process

The most important change accompanying the increase to 20 000 scans per day, is a shift in the quality management focus from individual scans to the scanning process itself. In the heritage sector, it is common for digitization procedures to focus on individual items. At a rate of 20 000 scans per day, this is no longer possible.

For this reason, the Amsterdam City Archives has chosen an industrial approach with the motto: a more stable process means more predictable outcomes. In other words, the quality of the product is determined by the quality of the process. The Amsterdam City Archives quality manager is therefor not an image specialist, but a process expert who understands how parts of the process are interdependent, has knowledge of automation, and a clear, analytical vision of the process as a whole: from an item in the depository to an online scan.

The industrial principles are:

- Large scale

- Continuous production

- Scalable

The implication of these principles can be summarized in the following four measures:

- Creating a single standardized process

- Streamlining the process

- Automating everything that can be automated

- A pragmatic vision for quality management

The idea behind this industrial approach is that once you get grip on the production process, a one on one check of the end product is no longer necessary. A more stable process means more predictable outcomes.

To make an industrial approach possible, from 2016 onwards, the Amsterdam City Archives revised its principles and procedures for the receipt, assessment and processing of scans. Changes of both a technical and an organizational nature (automation and an integrated approach to the process), have been implemented. While implementing these changes we kept close collaboration with the scan suppliers.

Quality assessment

In the industrial approach of digitization, the quality assessment of the process has been reduced to four main aspects:

1. There should be no damage to, or loss of, the originals.

2. The scans should contain all relevant information from the originals.

3. All the information should be digitized, and in the right order.

4. The scans should be correctly linked to the relevant metadata.

Quality assessment of these four aspects is based on process auditing and takes place in various ways. The Amsterdam City Archives wants to have a clear insight into the work processes of its suppliers. In this context, transparency, consultation and cooperation are key.

Suppliers' work processes are assessed during the tender process and during regular checks on the shop floor by the Amsterdam City Archives quality manager. During the tender process, price is not the main criterion; it is the quality of the project execution plan that carries more weight in the assessment.

At the same time, we set relatively few conditions on the way in which suppliers meet a requirement. Completeness checks, for example, can be carried out by the tried and tested method of checking each scan against the original, but it does not have to be done this way.

At the Amsterdam City Archives we want to leave room for innovation. This has, for example, led to a development by a supplier in which laser technology is used at the scanning table to detect whether a page of a bound book has been missed. This makes a final completeness check of each scan against the original unnecessary.

Any checks that can be automated have been fully automated. Completeness, damage and image quality are checked on a sample basis. Acceptable Quality Level (AQL) principles are applied while performing these checks. This means that at the start of a project the sampling takes place at short intervals, which can be reduced to a minimal sampling interval if good results are achieved. This approach presumes that the process is reliable and that fewer checks are necessary if good results are achieved.

The quality standards are high, however. For completeness and damage to the originals, a margin for error of zero percent is applied. This may seem excessively strict – human error can always occur – but the Amsterdam City Archives wants to avoid difficult discussions about the numbers of errors allowed and how to calculate them.

Our approach is therefore simple: an error is an error. This does not necessarily mean that each error is immediately corrected. We are more interested in analyzing why the error occurred. This is because we believe that solving structural errors in a process is more useful than treating the symptoms.

The final step is to add the scans to the online system. This is carried out by a permanent team of employees, who also have other roles apart from digitization. In line with project management principles, a project can only be considered finalized when the scans are online. The cycle time between the production of the scan and its availability online is a few days.

The future

The Amsterdam City Archives wants to expand and improve digital services to users through digitization based on industrial principles. The modern user of archives expects to be able to consult their items directly from home. We see our digital, online study room as equal to our physical study room in our main building, *De Bazel*. Thanks to a production rate of ten million scans in two years, we can make a huge step forward in this endeavor, and we hope to further expand our digitization efforts.

By the end of 2019 we will still have only a fraction (less than 15 percent) of our archives and collections digitized – we will therefor be pleased to continue our ambitious digitization drive in the years to come. And yes, we're already dreaming of a fully automated assembly line…

Authors

Marc Holtman (1973) is project leader digitization and digital services at the Amsterdam City Archives since 2001. He has been concerned with the development of digitization workflow and digital webservices. This includes a scanning on demand service, image database (Beeldbank) and archives database (Archiefbank). His motivation is to create durable digital cultural heritage services for a broad audience at manageable costs.

Nelleke van Zeeland (1983) works at the Amsterdam City Archives as project leader digitization and crowdsourcing since 2010. She stood at the base of Dutch crowdsourcing platform VeleHanden.nl (literally: Many Hands) and managed a wide range of digitization projects. The planning and implementation of large scale digitization is now one of her primary tasks.