Digitisation based on industrial principles

A new vision for the quality management of mass digitisation projects

Version 1.1-e

MARC HOLTMAN
SEPTEMBER 2017

English translation: Timo Klaassen
ABOUT THE AUTHOR

Marc Holtman has been working at the Amsterdam City Archives since 2001. In his capacity as project leader for digitisation and the development of online access systems, he has been at the forefront of digital developments at the City Archives. He has project-managed the establishment of large-scale digitisation and the development of scanning-on-demand services. He was also responsible for the development of the following City Archives’ applications: Image Bank, Inventories, Indexes and Planning Applications. Key to the integrated approach of his projects, is manageability and simplicity. Marc has given workshops, lectures and advice in the Netherlands and abroad in the field of digitisation and digital services.

ACKNOWLEDGEMENTS

This report draws on practical experience and describes the actual processes we have designed and used. The author could not, and did not, do everything by himself. The process described here is the result of years of incremental developments. Robert Dijkgraaf once said of physicists, that there is no such thing as a “eureka!” moment. The same applies here. The solution is always the result of trying things out, fine-tuning and sharing ideas, which finally lead to further simplification.

For that reason, I would like to thank everyone who has contributed to this report. Firstly, I would like to thank the team for ensuring that, day in day out, everything runs smoothly, and for providing continuous reflection, development and improvement: Nelleke van Zeeland, Shanna Dijkstra and Walter de Korver. Ellen Fleurbaay’s critical scrutiny ensured that we never lost sight of the objective, and she gave me the space and time away from my day-to-day responsibilities to write this report. Conversations with the City Archives’ data manager and digital expert, Jan Huizing, were invaluable. Many of the ideas in this report took shape in cooperation with our suppliers, Picturae and GMS. Both companies were able to translate ideas into concrete solutions. Finally, I would like to thank three people working in the field – Robert Gillesse, Trilce Navarete and Hans van Dormolen – for inspiring conversations from which I greatly benefited. Without the help, effort and enthusiasm of all these people, the ideas presented in this report could not have been developed.

Marc Holtman, September 2017

Note on the illustrations: the photos and prints of factories and laboratories used in this report all come from the Image Bank of the Amsterdam City Archives.
# CONTENTS

| VERSIONS | 1 |
| RATIONALE AND SUMMARY | 1 |
| INTRODUCTION | 3 |

## 1. AIMS AND VISION

| 1.1 AIMS OF THIS REPORT | 5 |
| 1.2 SCOPE | 5 |
| 1.3 OBJECTIVE OF DIGITISATION | 5 |
| 1.4 HISTORY | 5 |
| 1.5 DIGITISATION BASED ON INDUSTRIAL PRINCIPLES | 6 |
| 1.6 VISION FOR DEVELOPMENT | 7 |
| 1.7 VISION FOR WORKING WITH EXTERNAL SUPPLIERS | 8 |
| 1.8 VISION ON TECHNOLOGICAL DEVELOPMENTS | 9 |
| 1.9 VISION FOR QUALITY | 9 |
| 1.10 THE INSTITUTION IN A STEERING ROLE | 10 |
| 1.11 IN-HOUSE DEVELOPMENT | 10 |

## 2. REQUIREMENTS AND TOLERANCES

| 2.1 FOUR PRINCIPLES | 12 |
| 2.2 TOLERANCES | 12 |

## 3. APPLYING THE REQUIREMENTS IN PRACTICE

| 3.1 THERE SHOULD BE NO DAMAGE TO, OR LOSS OF, THE ORIGINALS | 14 |
| 3.2 ALL RELEVANT INFORMATION FROM THE ORIGINALS | 15 |
| 3.3 ALL INFORMATION SHOULD BE DIGITISED | 18 |
| 3.4 THE REPRODUCTIONS SHOULD BE CORRECTLY LINKED TO THE RELEVANT METADATA | 22 |

## 4. ASSESSING QUALITY

<p>| 4.1 METHODS OF ASSESSMENT | 25 |
| 4.2 AUDITING THE PROCESS | 25 |
| 4.3 ASSESSMENT OF REPORTS FROM THE SCANNING PROCESS BY THE SUPPLIER | 26 |
| 4.4 AUTOMATED CHECKS | 26 |
| 4.5 SPOT CHECKS OF SCANS AND ORIGINALS | 27 |</p>
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>THE TWO PARTS OF THE PROCESS</td>
<td>29</td>
</tr>
<tr>
<td>5.2</td>
<td>TENDER</td>
<td>30</td>
</tr>
<tr>
<td>5.3</td>
<td>PREPARATION</td>
<td>31</td>
</tr>
<tr>
<td>5.4</td>
<td>TRANSPORT TO SUPPLIERS</td>
<td>32</td>
</tr>
<tr>
<td>5.5</td>
<td>SCANNING</td>
<td>33</td>
</tr>
<tr>
<td>5.6</td>
<td>RETURN TRANSPORT</td>
<td>35</td>
</tr>
<tr>
<td>5.7</td>
<td>RECEIPT OF SCANS</td>
<td>36</td>
</tr>
<tr>
<td>5.8</td>
<td>MAKING THE SCANS AVAILABLE</td>
<td>37</td>
</tr>
<tr>
<td>5.9</td>
<td>FINALISATION AND EVALUATION</td>
<td>38</td>
</tr>
</tbody>
</table>

**APPENDIX**

| I.      | REQUIREMENTS, TOLERANCES AND CONTROL MEASURES | 39   |
| II.     | KEY TERMS EXPLAINED IN MORE DEPTH             |      |
|         | ARCHIVAL MATERIALS AND IMAGE MATERIAL         | 43   |
|         | ARCHIVES’ DATABASE AND IMAGE BANK             | 44   |
|         | DIGITISATION OF NEGATIVE STORAGES SHEETS      | 49   |
“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 a large-scale digitisation project, which we began in 2006. An important part of our approach is customer-driven: through a scanning-on-demand service, the customer decides what is digitised. We also work on a project basis.

Ten years on, we have met 40,000 client orders and produced 20 million online scans. Meanwhile, the demand for digitised items continues to increase.

It is not just the digitisation project that requires a large-scale approach. This strategy is also driven by the need to manage some 50 kilometres of originals. Our main building in the centre of Amsterdam (De Bazel) is full, so we are developing a new repository in Amsterdam Noord. In conjunction with the construction of this repository, we have received a generous grant for the digitisation of a number of archives that need to be moved. This will take place in 2018 and 2019. The challenge is to digitise one and a half kilometres of archive within two years. This is equal to the production of about ten million scans.

In addition, the large-scale digitisation of the notarial archives started in 2016 and is expected to take several years. This is part of the Golden Agents project and produces 1.5 million scans each year.

Finally, with the implementation of the Law for the Re-use of Government Information, we saw an increase in the number of digitisation requests. We therefore increased the weekly production of scanning-on-demand requests from 15,000 to 20,000 scans in September 2017.

The challenge is to undertake the digitisation of the archive on such a large scale while avoiding making the process too onerous for our organisation, and without compromising quality.

To meet this challenge, the City Archives has developed a new approach which adopts principles from the world of industry. This does not imply smoking chimneys, but a process that remains manageable even on a very large scale, where quality is optimised, and costs remain as low as possible.
Over the last few years, the City Archives has developed a streamlined, scalable digitisation infrastructure. All digitisation takes places within this infrastructure, following a single standardised 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 task.

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 digitisation procedures to focus on individual items. At a rate of 20,000 scans per day, this is no longer possible. For this reason, the 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 City Archives quality manager is therefore 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 quality assessment of the process has been reduced to four essential 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 digitised, 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 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 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 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 City Archives want 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 analysing 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 digitisation. In line with project management principles, a project can only be considered finalised when the scans are online. The cycle time between the production of the scan and its availability online is a few days.

The City Archives want to expand and improve digital services to users using this ‘digitisation 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 have made a huge step forward in this endeavour, and we hope to further expand our digitisation efforts. By the end of 2019 we will still have only a fraction (less than 15 percent) of our archives and collections digitised – we will therefore be pleased to continue our ambitious digitisation drive in the years to come.

INTRODUCTION

Digitisation based on industrial principles is the third report on digitisation published by the Amsterdam City Archives. The first, ‘digitisation unravelled’ discussed the technical aspects of digitisation, such as resolution and file formats. The second report, ‘digitisation simplified’ described the creation of a standard working process for digitisation at the City Archives.

‘Digitisation based on industrial principles’ examines how principles from industry can be applied to the process of digitising historical heritage. This approach does not imply lower quality or that we are looking for ‘quick and dirty’ solutions. Rather the contrary is true: the aim is to set up a streamlined, stable process in which outcomes become predictable. At the same time, it becomes possible to scale up production, lower costs and ultimately produce higher quality results.

There is plenty of literature available on digitisation, but not on how large-scale production can be implemented in daily practice. Innovative projects, such as the digitisation project at Naturalis, have shown that an industrial approach to digitisation is indeed possible in practice. In their project they literally used a conveyor belt to digitise herbaria (dried plants mounted on cardboard). This may be a utopia for scanning, for example, a bound notarial protocol, but the principles behind this process can be implemented on a broader scale.

The starting points and approach described in this report are based on ten years of development and experience in practical digitisation. This report is therefore primarily focused on practice from the perspective of an archival institution. The way in which a scanning system technically works is, therefore, not discussed. The report does, however, describe how large-scale projects with high quality standards can be executed in a streamlined and consistent fashion at a manageable cost.
The main four topics of this report are:

1. The City Archives’ vision of ‘industrial’ principles for digitisation
2. Guidance on defining, establishing and assessing quality
3. Concrete requirements and tolerances
4. A description of the City Archives’ approach
1. AIMS AND VISION

1.1 AIMS OF THIS REPORT

This report has two main aims:

1. Knowledge-sharing and transparency through an explanation and justification of the vision, principles and actual work processes of the City Archives.

2. To discuss standards and the quality requirements for City Archives’ digitisation projects in:
   a. Tender procedures.
   b. The digitisation process itself.

1.2 SCOPE

The processes described are in use in digitisation projects carried out by the City Archives. Two basic conditions apply:

- The items have to be inventoried. This means that, at a minimum, they should have an inventory number and a registered location in the depository.
- The scanning and transport of the originals is outsourced to external, specialised suppliers.

Outside the scope of this report:

- Custom reproductions made at the request of individual clients or for publication purposes. This mostly concerns reproductions with an extremely high resolution or level of detail. Scans of this type are not included in the online access system.

1.3 OBJECTIVE OF DIGITISATION

The primary objective of digitisation is to facilitate archival research: the research should be possible at any time and from any location. The primary objective is to make the archives digitally accessible while we continue to conserve the originals.

All scans produced by the City Archives are made available on the Archives’ Database and in the Image Bank. As of 2017, these contain more than 20 million scans. From 2018 onwards, five million scans will be added to that total each year. A key principle of the process is that a project is only finished when the scans are online.

Digitisation also achieves a conservation objective, because once digitised, the archival documents no longer need to be taken off the shelves. After digitisation, paper originals can no longer be requested for consultation in the study room. This prevents damage caused by using the materials for research and minimises the risk of loss due to theft. In addition, if theft still occurs, a digital reference facilitates investigation and the identification of materials.

1.4 HISTORY

Over a period of about 15 years, digitisation at the City Archives has gone through three phases. The first phase started in around 2001. At this time, digitisation was on a small scale and mostly involved digitising images for the Image Bank. During this phase we gained a lot of knowledge and
experience of various aspects of digitisation in terms of technology and organisation, but also of the provision of access and delivery of scans through web applications.

After putting the archival inventories online in around 2005, it became clear that users expected all items to be available online. A second phase can therefore be seen as starting in 2005 in which large-scale digitisation became the principle. In 2006, a service was developed to enable clients to request the digitisation of any item listed in the publicly available inventory through a web application (the Archives’ Database). The idea behind this was, and remains, that the users prioritise the items for digitisation, which takes place according to demand. In this manner some 15 000 scans (3 metres of archives) are produced on a weekly basis, as users request them. Besides digitisation on demand, each year a number of digitisation projects are selected and carried out by the City Archives.

The third phase of digitisation started in around 2013. The preceding years had shown that large-scale, standardised digitisation was feasible and manageable. The approach was, however, still that of a classical production cycle up to and including checking the end product. Both the efficiency and the stability of the process still left room for further optimisation. The City Archives then, in cooperation with its scanning suppliers, researched and tested the extent to which an approach based on industrial principles was possible. The results have been implemented through revised procedures and a project started in 2016, which will see three kilometres of notarial archives digitised. Due to the construction of a new repository in Amsterdam Noord, one and half kilometres of archives will be digitised during 2018 and 2019 and since September 2017, the weekly on demand scanning rate has been increased to 20 000 scans. From January 2018, the production rate will be scaled up to 20 000 scans daily. On this scale, an industrial approach as described in this report is not an option, but a necessity.

1.5 DIGITISATION BASED ON INDUSTRIAL PRINCIPLES

The industrial principles are:

- Large scale.
- Continuous production.
- Scalable.

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

- The creation of a single standardised process.
- The streamlining of the process.
- Automating everything that can be automated.
- A pragmatic vision for quality management.

The idea behind this industrial approach is that the quality of the product is determined by the quality of the process. Moreover, a more stable process means more predictable outcomes.

To make an industrial approach possible, in 2016 the City Archives revised its principles and procedures for the receipt, assessment and processing of scans. Changes of both a technical and an organisational nature (automation and an integrated approach to the process), have been implemented.
1.6 VISION FOR DEVELOPMENT

The approach of the City Archives is characterised by a number of principles which, through years of practice, have become central to our organisational procedures in the digitisation field. The most important of which are the following:

**Learning by doing**

Most is learned by doing. The City Archives is, therefore, not afraid of experimenting. Even if the outcome is uncertain, there is space to try out ideas in practice. One example is the development of an order tracking application. The first two versions were made in-house in a simple beta set-up using a minimum of resources. This enabled a developer to easily create a third version based on the practical experience already gained by the City Archives. This application release has proved to be stable: it has been running for a few years now and has needed almost no updates.

**Generic solutions**

Solutions should be implementable on all projects, in unchanged form and independent of supplier. One example is the order form for digitisation, which is used for all projects and types of materials. Since its introduction in 2006, the function and content of the order form has remained unchanged. In those ten years, not a single inventory item has been lost during digitisation.

**Continuous development**

The process is never perfect. In consultation with the team and our suppliers we continuously look for improvement opportunities. These are implemented wherever possible, sometimes gradually and sometimes by making major changes. The starting point for every development is that it should contribute to efficiency and/or quality. Each time the tangible benefits are weighed against the investment necessary to implement the change.

**In-house development**

Flexibility is essential for an experimental setting in which continuous development takes place. It should be possible to change procedures quickly and efficiently whenever necessary. The City Archives does not employ any software developers, but it does employ people with skills in the areas of automation and database management. External developers are therefore only needed on an occasional basis.

**(Re)use of resources**

Software is only developed if specific, necessary functionality is not offered in existing applications we already own. One example is the way in which scans are received and checks are carried out. It was tempting to have an application developed to perform spot checks. On closer inspection, however, we concluded that a simple solution with existing Windows software and data access applications did the job just as well.
Automating everything that can be automated

The process contains many steps and actions. In the process used at the City Archives, all basic steps have been automated. However, there is still scope to automate further actions, such as starting a script (which is currently done manually), in the next few years.

Keeping it simple

However a part of the process is automated, human input is still a core component of digitisation. However strict the standard, and however intensive the checks, errors will keep on occurring. Working on a large scale demands an approach that balances scope, quality and manageability. This necessitates a pragmatic approach that aims for a good result, and not a utopian perfection.

1.7 VISION FOR WORKING WITH EXTERNAL SUPPLIERS

The City Archives has chosen to outsource digitisation to external, specialised suppliers and this continues to be the case now we are using industrial principles. Scanning equipment that is available on the general market produces good quality scans, but digitisation is more than making a scan. The processing of the data that is produced requires highly specialised technical infrastructure: scanning systems need to be able to process a wide range of materials, adequate care needs to be taken of the originals, and it should be possible to work outside normal business hours. A simple scanner with a few employees to do the scanning is not sufficient. It is also important not to underestimate the requirements for automation – file conversion and naming, cropping, etc. The process should also ensure that errors are corrected. All of this requires advanced software for workflow management and quality assessment.

The City Archives has generally found working with external suppliers to be a good experience. They have been ready and willing to cooperate and invest in innovation. Compared with suppliers from other countries, Dutch suppliers are front runners in terms of quality management as described in this report.

Outsourcing requires the material to be scanned to be transported to external locations for each project. An alternative solution would be to outsource the work, but to have it carried out on-site. We looked at doing this at the City Archives’ De Bazel site in central Amsterdam. We concluded, however, that this was not a realistic option, because:

- The ICT infrastructure is inadequate for continuous production in a large-scale setting, both in terms of output rate and stability.
- The actual hours when it is possible to work are limited, because of the limited opening hours of the building.
- There is not enough space available.

Substantial investment would have been needed to make digitisation within the City Archives possible. It was therefore decided, for the time being to continue with off-site digitisation.
1.8 VISION ON TECHNOLOGICAL DEVELOPMENTS

In the field of digitisation, developments are rapid. In particular, Google Books has given rise to a new vision for the digitisation of books, developing various digitisation methods in the process. Archival material is generally still digitised in a relatively classical one-shot set-up with a flat or V-shape bookrest and one or multiple cameras above it. This is primarily due to the diversity and fragility of the material. The robot scanners used by Google are not suitable for scanning archival materials.

However, there are developments that may become an alternative to photographic recording methods in the future. In particular, the digitisation of texts through x-ray scanning is interesting, because it means that books will no longer need to be opened. This technology is however still in a very experimental stage. The expectation is that it will not be available for general use in a production setting in the near future.

At the moment, as far as technology and innovation is concerned, the City Archives is primarily looking at the existing workflow to find further opportunities for optimisation. A number of improvements that can be made include:

- Changes to the scanning table to make it possible to work faster.
- Adaptations in software and ICT infrastructure for a more streamlined production flow, data management and delivery process.
- Alternatives to checking each scan against the original – the conventional method – to guarantee completeness.

1.9 VISION FOR QUALITY

Users can access scans through the online City Archives browsing environment. Over the past few years, the digital environment has become the heart of our service provision. Visitors to the Information Centre at the City Archives can still request original items (if they have not yet been digitised), but most visitors are served via terminals through which they can access digitised documents, metadata and browsing systems. Stability, both in terms of the availability and quality of the digital content, is therefore essential.

The challenge we face is to deliver a level of quality that meets users’ needs and that can be guaranteed, at an affordable cost (both in terms of infrastructure and running costs). ‘Quality’ not only refers to the image quality of the scans, but also to the completeness of documents and the order the pages are presented in, and whether everything that has been digitised can really be accessed online.

The City Archives is therefore not really concerned about ‘an error’, but is interested in:

- The cause of the error – analysis and understanding of which can enable the process to be adjusted to prevent it from occurring again.
- The degree of variation from the norm. The outcome of a single measurement is not only compared to the standard, but more importantly, is compared to the outcomes of earlier measurements. What matters is stability, and hence the predictability of outcomes.
1.10 THE INSTITUTION IN A STEERING ROLE

The quality of the product is determined by the quality of the process: a more stable process means more predictable quality. An industrial process is characterised by:

- Clear, standardised and documented procedures.
- The use of objectively measurable standards and tolerances.
- Automating the process wherever possible.
- Automated checks and monitoring.
- Carefully set up, efficient visual/manual checks on aspects that cannot be checked automatically.
- Continuous evaluation and improvement.
- Clear and full administration of information relevant for the project.
- Sound project management, in which the division of tasks and responsibilities is clear.

Once the process is set up properly, the City Archives can take on a steering role, a bit further removed from day-to-day operations. With regard to quality management, the focus shifts from output checks (scans) to process auditing. The City Archives’ controllers/auditors are therefore not image specialists (although they do have the basic knowledge required), but process specialists, who understand the factors and control measures that establish a consistent quality within a project, and who are able to make an impact assessment when an error occurs.

1.11 IN-HOUSE DEVELOPMENT

As far as possible, the City Archives wants to keep the management of the process in-house, using existing facilities. For example, the existing management systems are used at the start and finish of the digitisation process, and the data in these systems prevail. Computer scripts and (if necessary) software is developed around that. To be able to do this we need:

- Management applications which are flexible to set up, and can be fine-tuned independently to meet the requirements of the process (and not the other way around).
- Knowledge and experience of automation and scripting.

In practice, this means that we often make gradual improvements during development, and that we continuously make further developments. For this reason, the ability to automate processes is a key skill. Employees working in the data management team at the City Archives are therefore closely involved in setting up the process. This approach ensures that the City Archives can make quick and effective changes, without needing to hire external developers. Our success in process development over the last few years is to a large extent due to this way of working.
Group of students or employees of the pharmaceutical laboratory during a coffee break, between 1920 and 1925
2. REQUIREMENTS AND TOLERANCES

2.1 FOUR PRINCIPLES

In order to assess whether a process is good, we need to define ‘good’. This means that a quality framework is needed.

The City Archives has kept the framework simple, and defined four basic principles for digitisation. These principles apply to all forms of digitisation, regardless of the aim of the project or the material. They apply to all parts of the process, whether they are the primary responsibility of the City Archives or of an external supplier.

<table>
<thead>
<tr>
<th>Principle</th>
<th>Applies to</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  There should be no (new) damage to, or loss of, the originals</td>
<td>Originals</td>
</tr>
<tr>
<td>2  The reproductions should contain all relevant information from the originals</td>
<td>Scans</td>
</tr>
<tr>
<td>3  All the information should be digitised, and presented in the same order as the original</td>
<td>Scans</td>
</tr>
<tr>
<td>4  The reproductions should be correctly linked to the relevant metadata</td>
<td>Metadata and scans</td>
</tr>
</tbody>
</table>

Each principle is divided into a number of requirements and control measures. For example, one requirement related to the principle that the originals may not be damaged states that documents must be transported in sealable, waterproof containers.

The number of control measures for specific parts of the process for which suppliers are responsible has been kept to a minimum, the idea being that the specialist suppliers need the space to set up and optimise their processes according to their own professional judgement. The supplier is expected to document and demonstrate (on location prior to the start of the execution phase), the measures that have been taken to meet a particular requirement.

2.2 TOLERANCES

Tolerances are linked to the requirements. These are the margins within which the supplier is permitted to operate. The tolerances are, in principle, strict. For example, for completeness of scans, the tolerance is zero percent. This means that not a single error is tolerated.

Strict tolerances have been chosen in order to avoid difficult discussions, such as how a margin of error of three missing scans in 10 000 should be calculated. In this case, would that be calculated on the basis of a consecutive series of 10 000 scans, or per production batch, or for the complete project? And when and what is to be corrected?

This does not alter the fact, however, that digitisation without errors is a utopian dream. When an error is detected, the City Archives is more interested in the causes of the error than in the error itself. For example, if we consider a batch of 40 000 scans containing 12 errors, we can see that when applying a tolerance for completeness of 3 in 10 000, the number of errors falls within the tolerance and the batch has to be approved. But if all these errors are detected in a single inventory item, it is desirable to look at what went wrong. Looking at the causes and estimating the risk of recurrence is what is important. If a procedural change can prevent the risk of recurrence of this particular error, that would be better (and probably more efficient) than dealing with such errors (which may be within the margin) individually time and again. A process oriented approach therefore remains the
starting principle of the City Archives and justifies a zero percent tolerance for a number of requirements.

A zero percent tolerance is also applied for technical parameters such as file names, file formats, bit depth, etc. These aspects are checked automatically. When an error is detected, all scans from the order are destroyed and the whole order needs to be redelivered after correction.

For aspects such as image quality, order and cropping, the tolerances have been set in line with what is reasonable. The reason for being less strict is that the impact of these errors for the user is smaller, and it may be harder to objectively assess these aspects. One example is the order of scans. When addendums and glued-in documents are part of an item, it is not always clear what the logical order of information is. Even in the original this may sometimes be hard to discern. To a certain extent there is a subjective choice to be made. In these cases, the scanning operator is expected to use their judgement, and they may be aided by generally applicable scanning instructions.

\begin{center}
\textit{Bataafse petroleum maatschappij (BP) laboratory}
\end{center}
3. APPLYING THE REQUIREMENTS IN PRACTICE

A set of requirements and tolerances alone are not enough to keep a process manageable. This chapter discusses the most important considerations and choices accompanying the practical application of the requirements.

3.1 THERE SHOULD BE NO DAMAGE TO, OR LOSS OF, THE ORIGINALS

MATERIAL ASPECTS

For each project, a key priority is the prevention of damage to, or loss of originals during the process. This is not only limited to the scanning part of the process. As far as possible, the existing standard procedures and guidelines for document handling and logistics are followed. These are supplemented with points specific to digitisation, such as transport to and from the supplier’s site, storage of the items while at the suppliers’ premises, and movements on the scanning floor.

For the parts of the process for which the supplier is primarily responsible, there are no exhaustive rules. The City Archives consciously chose to work with professionals with proven knowledge of handling fragile historical material. The fact that eating is not permitted on the shop floor is self-evident and does not need to be included in rules set separately by the City Archives.

The requirement for document handling is simple: no originals may get lost or damaged in the process. The City Archives therefore expects suppliers to be transparent about their processes, and to describe and demonstrate the control measures they take to prevent damage or loss. Compliance with national and international standards regarding quality assurance is therefore advisable. Spot checks are done during the process to check whether any originals have sustained damage.

In practice there will always be a certain tension between the physical aspects of digitisation and the preservation of the originals. This is not only the case for the parts of the process carried out by suppliers, but also for what is done by the City Archives itself. An example is the use of glass plates during scanning, which causes pressure on the binding (in particular the spine), which is not ideal from a preservation perspective. Choosing not to use a glass plate however, will lead to blurry and useless scans. The right balance needs to be found in which enough attention is paid to both aspects.

Human input remains a core component of digitisation, and however carefully the work is carried out, some damage will still occur. An important question is to consider when a change in the material condition of an item is such that it is unacceptable, necessitating measures to be taken. When a scanning set-up could lead to pages being ripped out from the binding it is clear and simple. It is more complex when an existing tear in the binding has become larger as a consequence of transportation. Although both examples need to be seen as forms of damage, the impact of each is different and depends on the cause, type, extent, and location of the inflicted damage and (very importantly) the risk of recurrence.

Because clear, objective guidelines are hard to set in this situation, the City Archives has chosen to take a pragmatic approach in line with that of other aspects of digitisation. The following applies:

- Internal and external employees must be adequately instructed in the handling of originals.
- When necessary, employees must participate in training sessions or workshops.
- Work is always organised on a project basis.
- One set of instructions applies for each project.
- Spot checks are made to check for damage.
- There is an ongoing dialogue with the supplier, in which the existing situation is evaluated and potential improvements are discussed.
- If damage is detected, the impact is assessed on a case-by-case basis, to determine whether further examination is necessary. If it is, the following is looked at:
  - What is the cause (internal logistics, transport, handling, or the scanning table?).
  - What measures need to be taken at the spot where the damage occurred to prevent it from recurring.

As with the other aspects of digitisation, open communication and cooperation between the institution and the supplier is key – and this is what is most likely to ensure success.

PROCESS RELATED ASPECTS

However unpleasant damage to the binding of an original is, its impact cannot be compared with the loss of an inventory item. When setting up the process, the City Archives adopted measures to minimise the risk of losing items. The principles are:

- Digitisation can only start if the item has been catalogued. This is checked automatically.
- All inventory items can be identified in a uniform manner.
- At each moment it is known what the actual status of an inventory item is.

To facilitate this, we have developed:

- A standardised process that is applied to all digitisation, independent of type of material or size of the digitisation project at hand.
- Scripts which perform automated checks.
- A standardised order form enabling the supplier to identify the item and receive specific instructions.
- Standardised completeness checks.
- An order tracking application.

3.2 ALL RELEVANT INFORMATION FROM THE ORIGINALS

An important principle of digitisation is that the digital scan reproduces the original as faithfully as possible. The City Archives is not interested in photographic tours de force, but rather in consistent quality. The main question here is how to define that ‘quality’.

TECHNICAL IMAGE QUALITY

It is impossible to make an objective statement about image quality just by looking at it, even if you have the original next to the scan. Noticeable discrepancies can be caused by:
- Incorrect calibration of output devices (monitor or printer) or discrepancies caused by the scanning device itself.
- External influences distorting the viewer’s perception of the image (such as fluorescent lighting).
- Comparing an original object, which can be seen because light is being reflected off it, with an image viewed on a medium that emits light.
- The perception of visual characteristics of an object (especially colour tone) can be strongly influenced by the circumstances under which the object is viewed. For a fair comparison of the quality, the object must be viewed in conditions as similar as possible to those in which the shot was taken (bright, white light!).

The question is, therefore, whether it is possible to make statements about image quality that can reasonably be seen as objective and measurable. Since legibility is a primary aim for textual documents, one might suggest that an image is acceptable if the text can be read. However, there are a number of problems with this proposal:

- The appearance of handwritten documents can differ greatly. One document may be written in various hands, some of which may be dark and use a lot of ink, while others may use very light inks. Documents that have a lot of detail and light text or dark spots can be especially problematic. In this case, a representative sample of this type of scan has to be viewed during the quality check. But it is not clear at the outset of the project which documents are going to be checked.
- To make objective statements about legibility, a relatively large number of documents would need to be checked (and compared with the original in case of doubt).
- It is almost impossible to make objective statements regarding the consistency of, or variation in quality within the scanning process, even though that is an important consideration.
- The equipment used must be properly calibrated and working conditions must be right (a scan that is not legible because the monitor does not have the right settings is not a bad scan).
- Legibility is often subjective (some people like to have a lot of contrast while others prefer less) and is subject to fashion.
- The legibility of ‘relevant information’ in a print or drawing is different from the legibility of a text. With the digitisation of images, in particular, personal preference plays a role when assessing the scan.

The quality of scans can be established objectively by working with calibrated reference originals (mostly named ‘targets’ or ‘technical targets’) which have a number of precise image values that are known. By digitising these elements, and subsequently measuring them with software (such as Photoshop), concrete statements can be made regarding the extent to which these aspects have been transmitted, and with them the quality of the scan. This technique has its origins in analogue photography and microfilming, where measurements were performed on similar kinds of reference originals, but using densitometers instead of software.

In practice, the method developed for the Metamorfoze programme is frequently used. Metamorfoze is a Dutch programme for the preservation of paper heritage, funded by the Dutch
Ministry of Education, Culture and Science. A guideline for the technical quality of the scans was developed as part of this programme, the *Metamorfoze* standard.¹ This standard is applied to the digitisation projects that are financed by and carried out as part of this programme. The following three components constitute the standard:

- A working method: this determines the technical image quality of scans by recording and measuring targets, both during calibration of the scanning system and during production.
- Calculating methods: for establishing quality.
- Standards: for the various aspects with accompanying tolerances.

It is important to recognise it is made up of these three parts, because, in principle, it is possible to use the working and calculating methods with other tolerances. For an original for which colour is less important, the decision could be made to use larger margins for colour reproductions.

The great advantage of using the *Metamorfoze* guidelines is that, in principle, concrete, objective statements can be made regarding quality without having to compare it to the original. Moreover, it is possible simply to refer to the guidelines in a project. The client does not need an extensive theoretical knowledge of the elements that make up the guidelines. However, there are also a few disadvantages:

- To be able to measure targets and assess the results, specialist software and knowledge is required. If the client does not have that knowledge and those tools, they will not be able to verify whether the standard is met. As yet, there are no tools available that enable the non-specialist to measure and assess the targets.
- *Metamorfoze* was developed as part of a programme that focussed on replacing physical paper heritage with digital images, which may not be the main aim of many of our projects. The tolerances are therefore stricter than necessary for many projects. This can have negative consequences in terms of cost. Experience has shown that with many suppliers, rates including or excluding application of the *Metamorfoze* standards no longer differ very much.
- At the moment, working with targets is not really possible when digitising negatives or using rotary scanning, for example.

The City Archives uses *Metamorfoze Light* as a guideline for the digitisation of archival material and negatives of images. For projects in which rotary scanning is used, or in which negatives are digitised, individual agreements are made for each project.

It is important to note that the inclusion of targets in the framework for projects carried out by the City Archives is primarily aimed at monitoring the stability of the process, and not just to determine whether an individual scan is accurate. It is vital that the guidelines can be used on a large scale.

Generally, the primary aim of the City Archives’ digitisation projects is to provide access, and not to replace the original. If there is a quality issue that does not make the text illegible, the scan will not necessarily be rejected. The City Archives expressly sees the *Metamorfoze* guidelines as just that: a

---

¹ https://www.metamorfoze.nl/english/digitization
As it is intended to be. For each project an analysis is made of which aspects are important, and the working method is tailored to that.

Although Metamorfoze Light is the basis, it is not mandatory. Suppliers are, in principle, free to propose an alternative method provided it is possible to measure and determine quality objectively. To achieve this, it is easiest to choose to use Metamorfoze. Another option would be to use the guideline as a starting point, but to deal with the tolerances for some aspects in a slightly different fashion.

Finally, one note regarding tonal range. When making high quality scans, a broad dynamic range is used. This means that as many shades of light and dark should be reproduced from the original as possible. This makes it possible to use the scans for various purposes. However, without editing they may look a bit ‘flat’ to the human eye. Some people prefer to have a more ‘lively’ image for daily use, which can be achieved by editing, such as increasing the contrast. In the early years of digitisation at the City Archives, this was taken as the principle, and the output of digitisation was a (slightly) edited scan, with more contrast. It has since, however, been decided to no longer do that, because:

- It is always possible to make an image with more contrast from a technically correct scan, while the opposite is impossible.
- The quality is no longer measurable.
- It is purely subjective – everyone sees and prefers something different.
- What is considered pleasant is subject to fashion.
- What looks good on a monitor sometimes does not look good in print.
- Contrast enhancement in a production setting is mostly executed in a batch, consequently, one scan may look fine, while another may not, having unclear light and dark sections.

This means that choosing a technically correct, measurable standard is preferable to subjective editing with a single aim (‘it looks good on a monitor’) and an unpredictable outcome. Moreover, functionality such as increasing contrast can also be implemented in document viewers.

3.3 ALL INFORMATION SHOULD BE DIGITISED

Experience with large projects shows that incomplete scans of original documents is the most likely type of error to occur.

These errors include:

- Missing a page in the scanning process.
- The order of the pages in the scan differs from the original.
- Information in the gutter of bound items has been insufficiently reproduced.
- Too much or too little information has been cropped from the scan.
- Incorrect or missing metadata links due to errors in data management.

COMPLETENESS OF SCANS COMPARED TO THE ORIGINALES

However improbable it may seem, the most common error during digitisation is missing out pages during scanning. This is disastrous for the end user, because it is often not clear from the scans alone that something is missing. The scans are the primary media available to users of our service, and the
originals can no longer be requested. Moreover, in case of missing scans, all information on a page is missing, not just a small part that is a bit less legible. For these reasons, completeness is the most important aspect for all projects, and the required standard is high: there should be not a single page missing, independent of the project’s objective.

The traditional and most common method of checking for completeness is to compare each page of a scan against the original. This method gives the best guarantee of completeness, but is very labour intensive and therefore has major cost implications for digitisation.

The City Archives compares scans against originals when carrying out spot checks according to AQL principles, which decrease in frequency if the results are good. For some projects, such as those involving rotary scanning, specific methods are applied, such as digitising twice and comparing the output. Checks tend to use specific properties of the original material. Scans of bound originals can sometimes be checked by using the page numbers (if the originals have page numbers). Only if a discrepancy is detected, is the original used as a reference. In all cases, the supplier should check for completeness, and the City Archives only carries out a final check.

Example of documents showing the complexity surrounding completeness and missing pages

It is up to the suppliers to develop efficient, cost-effective methods that guarantee completeness. Innovation is conceivable in the areas of:

- The development of software to assist with completeness checks.
- Carrying out as many checks as possible directly on the scanning table.
- Incorporating correction into the processes as much as possible, instead of making it a separate task.
- Using well-thought-out spot checking methods, underpinned by statistical principles.
- Technological innovation in the scanning device to prevent missing pages.

Various suppliers are currently working on the implementation of one or more of these solutions. The City Archives sees these as positive developments, and expects further gains to be made over the coming years.

GUTTER

With bound originals, text sometimes runs into the gutter of the binding. This is mostly the case with documents which were originally loose leaf and were bound later. During reproduction there will always be some loss of information compared to the original.
However it seldom leads to the loss of so much information that one needs to see the original, the supplier will still need to show what measures have been taken to avoid this from happening, and show how they are going to record the information in the gutter in the best way possible.

Example of a bound item with gutter issues

Although, at the time of writing, handwriting recognition in a production setting is far from realistic, it is important to consider such a development when aiming for an optimal reproduction quality.

CROPPING

While scans are produced, the recording surface is usually larger than the document that needs to be reproduced. After scanning, the scans are cropped to the document’s border. Most scanning and image editing software comes equipped with an automatic cropping feature. In practice, these standard features do not always work properly in a production setting, especially with historical documents, with more being cropped than should be.

While the City Archives does not impose any particular method to be used, the standard is strict: no information may be lost as a consequence of incorrect cropping.

An example of problems that may arise with automatic cropping:
ORDER OF INFORMATION

With digitisation, the physical properties of the original cannot be reproduced. It will often become difficult to discern properties such as the thickness of a binding. And physical properties such as quires are, after scanning, no more than a number of images. This means that it is sometimes unclear when a unit begins and ends.

Digital browsing, however, also has some important advantages. It makes it possible to browse quickly through a document and jump from one section to the next, and documents can also be quite easily placed in a new context. It is therefore not useful to try to impose the analogue situation on the new digital reality. It is better to invest in a good document viewer.

Whatever purpose the digitised scans are to be used for, it is important that their order follows the logical order of information in the original documents as far as possible. There are many conceivable situations which make this complex in practice—such as glued-in notes on a page which have information on the front and back, or folded-up items. It is impossible to give an instruction for every potential situation. In these cases, the City Archives expects the supplier to make choices which are logical for the user.

There are a number of basic rules that apply:

- The sequence number in the file name follows the logical order of the original documents. In other words, sorting items according to their file name gives the natural order.
- Scanning is only executed in one direction, even if the information in the original reads in multiple directions.
- Bound units are always scanned as a double-page spread (rather than as individual pages).
- Blank pages are not scanned. Blank in this context means that it contains no information, just a page number or a pre-printed table that has not been filled in.
3.4 THE REPRODUCTIONS SHOULD BE CORRECTLY LINKED TO THE RELEVANT METADATA

Scans are useless and cannot be found without metadata. Guaranteeing the accuracy of links between scans and metadata is therefore an important focus in the work process.

The key to this link is the file name which is assigned to each scan. The link is created by registering the filename in the metadata management environment associated with the relevant inventory item.

The City Archives has chosen a method whereby the basis for the link is already made prior to the scanning process itself. However, it is usually unclear at the start how many scans are going to be produced, so each item is given a so-called order number before digitisation is requested. This order number is registered in the metadata management environment of the relevant inventory item. The primary function of the order number is to identify the item to be digitised during the process. But it also has a function when assigning file names to scans. Each scan is given a file name that is made up of the order number (e.g. BRGA00025), together with a sequence number. The first scanned document in the inventory item is assigned sequence number 000001. The subsequent scans within the order number will get the following numbers. Sorting by file name gives the natural order of the items.

The first file name in an order number is assigned using a barcode on the order form. An order form is a sheet of paper containing the most important data related to the digitisation of the inventory item concerned. An order form is manually attached to the inventory item before transport.
Example of an order form

After digitisation, the file names of the scans are collected by means of a digitally generated file list, and are registered to the correct inventory number by an automated script in the metadata management environment. This establishes the link between the inventory number and the individual scan.

This method using order numbers was introduced in 2006 and has been extensively described in the ‘digitisation simplified’ report. Almost no updates have been needed since implementation. There are also, however, a few vulnerabilities:

- Order forms are manually added to the relevant inventory item. Because the work involves large batches (and thus, a large number of order forms), a lot of care and attention is required. If an inventory item gets the wrong order form attached to it, this (and probably also another inventory item) will get linked to the wrong file names.
- The assignment of file names in the scanning process demands a high degree of attention on the part of scanning operators, and the automation of the subsequent file-naming process needs to be well thought out.
- If scans are found to be missing, and items have to be rescanned, all file names from the point where the previously missing scan is inserted have to be changed. This also requires attentiveness and well-thought-out automation. In practice, recovery of this type of error
introduces the significant risk of new errors, for example, due to the incorrect sorting of the files during the renaming process.

- Order forms are placed inside the packages so the scanning supplier will only find out what is on the form when the packages are opened. It would be better if the order number were visible from the outside of the package, but fixing barcodes to the outside of the packages is not possible for this process. This is because bound items are not usually packed separately, and the barcode would then need to be fixed directly to the bindings.
4. ASSESSING QUALITY

A set of requirements and tolerances only becomes meaningful when the quality is assessed and when it is established whether or not the requirements have been met. In the classic set-up for quality assurance, there is a strong focus on the final product – the scans. They are examined to see if they meet the requirements, and if any errors are discovered they are corrected. In itself, this is a good method; but it is also labour intensive and provides little information about the underlying causes. The City Archives has therefore shifted its attention to the quality assessment of the process. The principle is that the scanning supplier delivers a product that meets the set requirements and that does not require another extensive check.

4.1 METHODS OF ASSESSMENT

The City Archives uses four methods to assess whether the quality of internal and external processes meets the required standards:

1. Auditing the process.
2. Assessing production reports.
3. Automated checks.
4. Spot checks of scans and originals.

4.2 AUDITING THE PROCESS

Because each error can be traced to a cause in the production process, an understanding of the process is key in preventing errors, or at least in tracing the cause. The more transparent the process, the smaller the scope for error. Process auditing is therefore an integral part of the City Archives’ quality assessment cycle. The City Archives expects a scanning supplier to be transparent about the way processes have been set up. This is not limited to showing how the originals are stored on location, but also, for example, how the technical process for file naming works, and how error recovery operations are organised. The City Archives does not require to see the scripts in detail, but the principles should at least be clear. The presence of process descriptions is recommended. Assessment takes place during the tender process, prior to granting a contract and the start of work.

Questions that are always central for the City Archives are:

- What measures have been taken to prevent damage to, and loss of, the originals?
- What measures have been taken to guarantee the completeness of the scans compared to the originals?
- What hardware, software and process design solutions have been chosen to guarantee an optimal, consistent image quality of the scans?
- What arrangements have been made for file naming?
- What measures have been taken to prevent or detect errors in automated processing?
- How does error recovery fit into the process?
Depending on the size of a project, interim evaluations may be planned during implementation. Interim audits are intended to be constructive: The City Archives’ auditor and the supplier look at the preceding period together and evaluate what is going well and where improvement is possible. Every evaluation results in a report which also lists all the agreed actions.

Finally, at the end of each project a final report is produced, which at a minimum, includes the project results and lessons learned.

4.3 ASSESSMENT OF REPORTS FROM THE SCANNING PROCESS BY THE SUPPLIER

If the supplier operates a good process, City Archives’ checks can be limited to the most basic aspects. The City Archives can then primarily guide operations on the basis of the supplier’s reporting.

In the best imaginable setting, the City Archives has access to a production environment in which the production reports are available. At a minimum, a report includes:

- City Archives order numbers.
- Completed and pending actions per order.
- Completed and pending checks per order.

For checks to be based on reports, it requires transparency on part of the supplier, a high degree of automation in the scanning system, and trust. This approach is favoured by the City Archives because it is the most efficient way of working – checks are carried out where they should be (in the process) and error recovery actions can also be carried out efficiently. Working with reports is closely linked to process auditing and evaluation. The City Archives will always supplement this with spot checks, the frequency of which will correlate with the level of trust.

4.4 AUTOMATED CHECKS

Automated checks by the City Archives take place on both the scans and the routines within the work process:

1. Scans – checks on technical parameters such as filenames. These checks are performed on file level for all files. When an error is detected, the whole order is rejected. The outcome of the checks is reported through an automated script.

2. Process – checks on how the process is carried out, and on transfer between locations. Checks that are carried out are:
   o A check that items are properly registered and that the original is available. This takes place when an order is placed for digitisation.
   o The order tracking application checks that all order are processed correctly, making it possible to see whether and/or where items went missing.
   o The metadata management environment gives an automated per project overview of the current status regarding the return of originals, receipt of scans, number of rejections in the order and cancellations.
   o The applications in the data access environment allow the monitoring of missing scans or metadata.
Spot checking methods are used for aspects which cannot be assessed automatically, such as completeness and order. The City Archives uses the principles of the AQL (Acceptable Quality Level) approach. This means that the frequency of checks depends on the phase of the project, in combination with results from earlier checks. The start of a project sees intensive checking. If no errors are detected, the spot check intervals can be decreased, until the minimal frequency is reached. The detection of an error then has two consequences:

- Direct analysis of the cause of the problem and the risk of its recurrence, correction and, if required, making changes to the scanning process.
- An increase in the checking frequency for subsequent deliveries.

Thanks to the use of spot checks and AQL principles, the number of measurements can remain minimal. At the same time, two things should be borne in mind:

- When an error is detected, the quality of all production output since the last approved check is, in principle, open for discussion.
- The time needed for a check within a project is harder to plan.

Spot checks alone cannot give a full picture, and will always need to be used in combination with the methods outlined above.

The frequency and implementation method of spot checks is determined per project, and depends on the practical purpose for digitisation, the characteristics of the originals, and the experience with this particular supplier from previous projects.

Spot checks are used to check the following aspects:

**Scans**

- Checking scans against the originals to ensure nothing is missing.
- Checking that images are complete (gutter and cropping).
- Checking that information is in the right order.
- Checking for technical image quality.

**Originals**

- Checking for damage sustained during the digitisation process.

The number of spot checks carried out is determined on the basis of the order as a unit. A random spot check based on the total number of scans produced would statistically be more in line with AQL, but checks on aspects such as completeness are nearly impossible to carry out in this way. The City Archives therefore chose to do spot checks on the basis of complete orders, and it is mainly the principles behind AQL that are applied.

For completeness checks, the City Archives defines the number of spot checks to be carried out in seven bands. After a spot check has been performed, any errors will be assessed on the basis of type,
impact and risk of recurrence. All errors are communicated to the supplier. If no errors are detected during a spot check, the next spot check will be at rate specified in the next band down.

<table>
<thead>
<tr>
<th>Band</th>
<th>Spot check size (on basis of orders)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1%</td>
</tr>
<tr>
<td>B</td>
<td>1.5%</td>
</tr>
<tr>
<td>C</td>
<td>2.5%</td>
</tr>
<tr>
<td>D</td>
<td>4%</td>
</tr>
<tr>
<td>E</td>
<td>6.5%</td>
</tr>
<tr>
<td>F</td>
<td>10%</td>
</tr>
<tr>
<td>G</td>
<td>16%</td>
</tr>
</tbody>
</table>

This method has been in use since 2017. Experience shows that spot checks can be used to detect any issues that arise in the process. In the first phase of digitisation, judgements made on the basis of a spot check were verified by checking a complete batch of scans against the originals. This established that spot checking is able to provide the necessary information to draw accurate conclusions.

*Steam-powered Factory and Other Tools: Van Rossum en Würtz factory, George Christiaan Schellenberg, 1861*
5. THE PROCESS IN PRACTICE

The previous chapters explained the principles of the City Archives' approach, what requirements are set, how they can be applied in practice, and how quality is assessed. This chapter describes the process in practice.

5.1 THE TWO PARTS OF THE PROCESS

The work process for digitisation consists of two parts:

1. The process the originals go through, from collection from the repository to their return.
2. The process the scans go through, from production to putting them online.

The second part of the process starts at the moment a scan is made. The City Archives' approach does not insist that the workflow between the two parts of the process should be interdependent. This makes it simpler to manage the workflow and means that projects can be set up flexibly. Depending on the type of project, the originals may be returned once the scans are checked, or they may not.

The main stages of the workflow, look like this:

In the following sections, each section of the workflow is briefly explained. The chapter will conclude with an overview of all aspects of the four quality requirements, including all control measures.
Each project begins with a call for tender. The size of the tender depends on the complexity and budget of the project.

For each call for tender, a document is drafted describing the most important features of the project. This includes figures related to its size and an analysis of the materials. Photographs of the materials to be digitised are always added as well.

For a tender process in which multiple bids have to be judged against one another, the working process is an important criterion. The tender participants are asked to give further details of their approach.

Depending on the project, an on-site audit at the supplier’s premises may be part of the tender process. This mostly happens with longer projects, or those with a very innovative character.

When the contract is awarded, a definitive action plan is drawn up. This is shared with all involved and includes planning the ‘kick-off moment’. Depending on the complexity of the project, this can be done by telephone (for a simple project) or at a dedicated meeting to which the main stakeholders are invited.
Each project is set up as a ‘project’ in the metadata management environment and has its own, unique code. This code is used throughout the whole process and is used as the first part of the associated file names. An example of a project code is NOTA – this is used for the notarial archives digitisation project.

After a project is created, a request list is prepared. This is a simple Excel file which lists all items that will be part of the project. This file is imported into the metadata management environment.

The next step is an automated check of the registration of the requested inventory item. If this check is not successful, an order number will not be assigned. The error report will include the inventory number and the reason for the error.

If the check is successful, the inventory item will be allocated an order number.

In practice, an employee can choose to carry out a check before requesting an order number. The procedure is technically the same, however an order number will not be issued as a result. An error report will still be made.

Finally, the order numbers are imported in the order tracking application.
As far as possible, items from the archive are transported by the employees who are also responsible for transport to and from the study room. The standard guidelines for the handling of items are equally applicable.

All transport of items connected to the digitisation process are supported by an order tracking application which tracks the process from start to finish. The application tracks items using the order numbers as issued during step 1 of the process. The application also generates lists that indicate where to find the items in the repository.
If during pick-up from the repository an inventory item appears to be missing, its inventory number is added to the missing items process and the digitisation order is cancelled.

Central to the City Archives approach is the order form, which is physically attached to each inventory item. These forms are also generated in the order tracking application.

A check on the physical state of the originals is a standard step in the procedure. When relevant, potential limitations regarding disclosure or copyright are also checked. If an inventory item does not meet the conditions set for the project, it will be sent for restoration when possible, or its digitisation will be cancelled.

A sample of items is taken to make a damage inventory. The sample size is determined by the type of project in combination with the type of items that are included in the project. For each order, a condition report is drawn up before and after scanning. It contains a description of the original, an overview of any existing damage, and photographs.

Transport is carried out in line with the standard guidelines, one example of which is working with closed trolleys.

### 5.5 SCANNING

<table>
<thead>
<tr>
<th>Storage of originals</th>
<th>Scanning</th>
<th>Data management</th>
<th>Checks and corrections</th>
<th>Delivery of scans</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Production</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This part of the process takes place entirely on the supplier’s site and has the following elements:

- Originals
  - Receipt
  - Storage
  - Handling
  - Return
- Scanning: production, data management and checks.

The City Archives does not impose specific requirements as to how the scanning process and the shop floor should be set up. The supplier is, however, expected to be transparent about its processes and approach, by documenting and demonstrating them. This is a standard element of the tender process. The supplier is also expected to explain what measures have been taken to protect the items from theft, fire and water damage.
The City Archives’ approach starts from the premise that all checks during the production of scans are performed within the supplier’s process. We operate on a basis of trust. It is up to the supplier to persuade the City Archives that the process is right, which enables the City Archives to keep its own checks to a minimum, as well as minimising the need to correct errors after delivery. This trust can be gained in three ways:

1. Providing information about the process in general. This is an important part of the tender process.
2. Providing information about the process during production: what actions have been carried out for each order, and what actions are still to be taken.
3. Positive experience.

For the day-to-day operations, the second point is especially important. A supplier can supply this information through reports. The City Archives is not particularly interested in knowing exactly where each order is in the process (that would be more relevant for project planning purposes), but it is interested in what actions and checks have been carried out. Relevant questions that reveal that information include:

- What has been done to ensure good data management?
- What automated checks have been carried out?
- What spot checks have been done, and what was the sample size?

The outcome of all checks can only be: OK.

This information can be supplied in various ways, but access to the digital environment where a current report can be consulted at any time would be the most straightforward.
Originals are returned to the repository in broadly the same way they were sent to the supplier, but in reverse order.

There are two checks that are made on receipt of the originals:

- All orders in the transport batch are checked against the transport list.
- There is a spot check for damage made on the basis of condition reports that have been drawn up earlier in the process.

If an item is found to be missing the supplier is contacted immediately. If there is damage, the follow-up is decided on a case-by-case basis.

The originals are returned to the shelves on the basis of a list generated by the order tracking application.
Scans are delivered by the supplier in one of two ways:

- Using a portable USB drive, or
- By SFTP

The City Archives prefers SFTP.

The receipt process is automated. To prevent conflicts with transfers to the City Archives that are already under way, a delivery consists of two parts:

- A folder containing all scans for a given order, with the order number as its title.
- A .txt file with the name ‘ready [order number]’. This file is added after the transfer of the file with the scans has been completed.

An order is only sent for further processing when there is a ‘ready’ file.

During receipt in the storage depository a number of automated checks are carried out on the files. If an error is detected the whole order is erased and the error and the order number are reported in an error report. This report is sent to the supplier with the request to correct the error and to redeliver the order.

Approved orders are included in the storage depository and the file names are automatically registered in the metadata management environment.

Finally, again automatically, the scans and metadata are made available for distribution to the relevant browsing application.
After importing the scans and metadata a completeness check is done. Are there scans without metadata? Is there metadata without scans? There are standard reports for these checks available within the applications. Any discrepancies are solved straight away.

The scans are then made available online and the processing procedures are complete.

After putting the scans online, there will be spot checks for completeness and order. This takes place within the access environment. This has a number of advantages:

- All functionality to access scans quickly can be used.
- It is easy to broaden the sample when necessary.
- All orders are processed quickly and do not need to be checked before they are put online.

The third advantage is wholly in line with the philosophy underpinning the City Archives’ approach. These spot checks should be seen as a check on the checks that have already been performed.

There is also a disadvantage, however. If a serious error requiring recovery is detected, a relatively complex recovery process follows. In principle, the whole production process is followed again: the scans are removed, the registration of the file names is erased, and the inventory item will be included in the process again with a new order number.
Each project is formally finalised, and evaluation is part of the project. Finalisation is done in a number of steps:

- Final checks are based on a standard report from the metadata management environment. The main questions are:
  - Have all inventory items listed been digitised?
  - Have all originals been returned?
  - Have all scans been delivered?
  - Or has the digitisation of this inventory item been cancelled?

After this check, the project is formally closed in the management environment.

The final step is the evaluation of the project. The form this takes depends on the size and complexity of the project. For each project, a final report is drawn up which, at a minimum, includes the extent to which the project objectives have been met and what lessons have been learned.
APPENDIX

I. REQUIREMENTS, TOLERANCES AND CONTROL MEASURES

The Amsterdam Superphosphate factory in Pernis, 1927
<table>
<thead>
<tr>
<th>Requirement</th>
<th>Applies to</th>
<th>Parts of the process</th>
</tr>
</thead>
<tbody>
<tr>
<td>There should be no (new) damage to, or loss of, the originals</td>
<td>Originals</td>
<td>Logistics and document handling,</td>
</tr>
<tr>
<td><strong>Aspect</strong></td>
<td>Control measure</td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>Each unit to be digitised has a unique identifier in the process.</td>
<td>Each unit to be digitised is given a unique order number via an order form that is attached to the original.</td>
</tr>
<tr>
<td></td>
<td>All inventory items that are eligible can be digitised.</td>
<td>Automated check on a number of control measures when the order number is requested.</td>
</tr>
<tr>
<td></td>
<td>At any moment it is clear which inventory items are part of a given project.</td>
<td>Automated registration of order numbers and change of the borrowing status into ‘digitisation’ in the metadata management environment after assigning an order number.</td>
</tr>
<tr>
<td>Internal transport</td>
<td>All originals belonging to a given project are sent for digitisation.</td>
<td>The order tracking application generates a list of order numbers to be collected.</td>
</tr>
<tr>
<td></td>
<td>In the repository, it is clear when an original has been removed from the shelf for digitisation.</td>
<td>A marker is placed in the shelf space in the repository normally occupied by the original.</td>
</tr>
<tr>
<td></td>
<td>All originals are returned to their places in the repository.</td>
<td>The tracking system generates a list to facilitate the return of originals, which is used in conjunction with scanning the barcode on order forms. The tracking system is also used to verify if all orders have been returned.</td>
</tr>
<tr>
<td></td>
<td>The borrowing status in the metadata management environment returns to normal after digitisation.</td>
<td>The metadata management environment automatically updates after the tracking system reports the return of an item.</td>
</tr>
<tr>
<td></td>
<td>All missing items are reported.</td>
<td>In case of a missing item, the order is cancelled, and the inventory number is included in the standard process for missing items.</td>
</tr>
<tr>
<td>Order tracking</td>
<td>It is always clear where every order is in the process.</td>
<td>Status changes are entered in the originals tracking system during the process.</td>
</tr>
<tr>
<td>Readying originals for digitisation</td>
<td>If the digitisation method chosen in the project leads to damage to the original, digitisation is cancelled.</td>
<td>There is a visual check before the originals are placed in the transport trolley. If damage is found, the order is repaired where possible and otherwise cancelled.</td>
</tr>
<tr>
<td></td>
<td>There should be no paperclips or staples in the originals.</td>
<td>All staples and paperclips are removed.</td>
</tr>
<tr>
<td></td>
<td>For digitisation on demand, items covered by copyright or other limitations are not included.</td>
<td>Check with regards to right of disclosure and copyright.</td>
</tr>
<tr>
<td>Transport to and from the supplier (general)</td>
<td>Transport of the originals has been outsourced.</td>
<td>Transport is carried out by a supplier.</td>
</tr>
<tr>
<td></td>
<td>Day and time of transport are known to all who are involved in the project.</td>
<td>A transport schedule is drawn up before a project is started.</td>
</tr>
<tr>
<td></td>
<td>Orders are packed for transport in a secure and safe way.</td>
<td>The scanning supplier provides sealable, waterproof containers in which originals can be tightly packed.</td>
</tr>
<tr>
<td></td>
<td>Orders are transported securely.</td>
<td>The vehicle and method used by the supplier is tailored to the transport of fragile historical items.</td>
</tr>
<tr>
<td>Transport to the supplier</td>
<td>The content of each consignment is known to ACA and supplier.</td>
<td>A transport list is compiled for each consignment.</td>
</tr>
<tr>
<td></td>
<td>The responsibility for the care of the originals is formally transferred at the moment of transport.</td>
<td>For each consignment, the supplier and the City Archives sign a transport form. A consignment is never combined with that of another client.</td>
</tr>
<tr>
<td>Storage</td>
<td>Orders are stored securely on the supplier’s premises.</td>
<td>Measures are taken at the supplier’s storage and scanning facility to prevent loss or damage to items caused by theft or damage from water or use.</td>
</tr>
<tr>
<td>Handling</td>
<td>Originals are handled carefully.</td>
<td>The supplier applies guidelines for the handling of originals which is strictly observed by all employees and which are enforced.</td>
</tr>
<tr>
<td>Scanning set-up</td>
<td>The scanning set-up is suitable for the digitisation of fragile historical items.</td>
<td>The supplier uses scanning systems in which damage to the originals is prevented.</td>
</tr>
<tr>
<td>Return transport</td>
<td>The content of a consignment is known to both ACA and the supplier.</td>
<td>A transport list is compiled for each consignment.</td>
</tr>
<tr>
<td></td>
<td>The responsibility for the items is formally transferred.</td>
<td>For each consignment, the supplier and the City Archives sign a transport form.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Applies to</th>
<th>Parts of the process</th>
</tr>
</thead>
<tbody>
<tr>
<td>The reproductions contain all relevant information present in the originals</td>
<td>Final product</td>
<td>Scanning</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Control measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tolerance</td>
</tr>
<tr>
<td></td>
<td>Instructions</td>
</tr>
</tbody>
</table>
### Image quality

| All relevant information from the originals is present in the reproduction. For text documents this is based on legibility. For images, all details that are visible in the original should be visible. |
| The use of reference originals (targets) for calibration and production linked to a standard which suits the intended use of the scans. |
| Dependent on item type and the objectives of the project. |

### Colour space

| ECI-RGBv2 |
| Process Automation |
| 100% correct |

### Technical parameters

| Image format JPEG, 1st or 2nd generation. |
| Compression 1:10 |
| Resolution is determined per project. |

### Requirement

| All the information should be reproduced, and in the right order |
| Final product |
| Scanning |

### Aspects

| Complete ness |
| All pages with relevant information are digitised. |
| Visual check against the original. |
| 100% completeness |

| All information on a page is digitised. |
| Visual or software check (or a combination of both). |
| 100% completeness |

| Order |
| All inventory items are scanned in the same order as the original. |
| Visual check against the original. |
| Principle of 'reasonableness' |

### Requirement

| All reproductions are linked to the correct metadata |
| Final product |
| Scanning, Data management |

### File naming

| All scans are assigned the correct file name. |
| File name assignment takes place on the basis of the barcode on the order form of the relevant order. |
| 100% correct |

### Registration

| All scans are registered with the correct inventory number. |
| File names are registered in the metadata management environment during receipt of the scans. |
| 100% completeness |
Visit of Queen Wilhelmina, Prince Hendrik and Princess Juliana to the Kromhout Motoren Fabriek, the royal guests arrive by car at the factory site at around half past two in the afternoon, Jan van Veen, 20 May 1927
ARCHIVAL MATERIALS AND IMAGE MATERIAL

PHOTOGRAPHIC REPRODUCTION

In photography, the light that is reflected or emitted by the original object is projected using a convergent lens on to a carrier with a photosensitive emulsion (the photographic negative). Thus 3D reality is reproduced on a 2D surface. In digital reproduction, the emulsion and carrier are usually replaced by three light sensitive image sensors, which translate the image to a grid, in which every square grid element is called a pixel. Each pixel has a specific colour value. The transferred image can be divided into three elements which together form the image:

- Detail
- Colour
- Tone (nuances between light and dark)

The extent to which a photographic system – whether it is analogue or digital – has the potential to be a good solution will be limited to a certain level by the properties and limitations of the lens, photographic emulsion, or image sensors used. The original document will contain information in all three areas which cannot be reproduced, even if the most advanced photographic devices are used. A photographic reproduction for that reason is in principle a compromise in which information from the original document, will be lost.

INTENDED USE

For the digitisation of archival materials, the intended use of the document is decisive for the requirements that need to be set and the necessary degree of transfer of the above-mentioned three elements. The aim is to be able to do substantive research. To be able to read a text, you need to be able to discern the dot on the i, and no more than that. A reproduction in which the paper fibres are visible up to the micro-level may be beautiful, but for the aim of ‘reading’ that does not have a lot of added value. Higher requirements are generally set for image material that will be included in a publication.

The fact that physical features are lost during reproduction is generally not an issue as far as research is concerned, except if it concerns the original order of complex multi-part inventory items, for example, folded up pieces of small letters in envelopes which, in turn are also folded. The size of the original is often also hard to imagine when viewing a reproduction. These disadvantages can be eliminated by using a document viewer which is tailored to doing research using a monitor. A good viewer offers substantial advantages compared to doing research on the original. Two examples are smart navigation functionality using thumbnails, and a zooming feature. It is not useful to try to reproduce the analogue reality in a digital reproduction. It is more interesting to take the new digital reality as the starting point, and develop software based on that. Digital working starts with digital thinking.

TYPES OF MATERIAL

Archives and collections contain various different types of material. Loose pages, charters, bound documents (which are sometimes extremely thick), hybrid items, card indexes, photos, architectural
drawings, glass negatives, all these and more are found in the archives and it is all eligible for digitisation.

In our projects a distinction is made between two main categories: archival items and images. Images can be subdivided into 2D materials and negatives. The distinction is made mostly because of the difference in their intended use.

**ARCHIVAL ITEMS**

By archival items, we mean paper textual documents in bound or unbound form.

Inventory items that include only unbound or only bound documents are rather the exception. You often find additions to bound documents, and unbound inventory items often contain sections such as quires or booklets. In practice, this means that in some cases multiple scans have to be made of a single page to be able to reproduce all the information.

The concept of ‘relevant information’ is therefore closely intertwined with the intended use – carrying out substantive research on the archive – hence the ‘reading’ of the information in the documents. In general, this concerns textual information in written or printed form, together with some images, such as graphs, grids and illustrations. The principle here is also ‘readability’. This means in practice that the transmission of details from the original has to be good. An exact colour rendering is a lot less important.

**IMAGE MATERIAL – 2D IMAGES INTENDED FOR INCLUSION IN THE IMAGE BANK.**

What we mean by image material are photos, prints, drawings and maps in loose or fixed form.

For reproductions of image material, the principle is that all details visible in the original should be visible in the digital reproduction. In addition to the normal viewing of the image, reproductions of image material are frequently used in paper or digital publication, for which a faithful rendering of the original is important. This poses high demands for the reproduction of detail, colour and tone. One note should be added regarding what is meant by faithful – this is meant in relation to the original document in its current state. The current state will in many cases not be the same as the document in its original form (due to discoloration of the paper, etc.).

**IMAGE MATERIAL – NEGATIVES**

Negatives are, in comparison to the other image material described above, generally scanned at a much higher resolution. In practice, this sometimes leads to confusion because for many projects ‘300 ppi at true size of original’ is the standard. Negatives are generally quite small and have a much higher information density than prints. For this reason, a higher resolution is used for scanning negatives which is calculated on the basis of a specific printing format. A 35-mm negative that needs to have a resolution of 300 ppi at a A3 format is digitised at 3600 ppi.

After scanning a negative, a positive reproduction is made. This often results in a need to make some adjustments, comparable to making a traditional print. The reproduction of a negative is therefore quite different from the appearance of the original document. Specific working method and image quality requirements are therefore set for each project involving the digitisation of negatives. Hence image quality is to some extent subjective.
The City Archives sometimes treats sheet film collections as archival material which is scanned through a fully transparent negative sheet. The result is excellent for most research purposes, but not good enough for publication.
ARCHIVES’ DATABASE AND IMAGE BANK

ARCHIVES’ DATABASE – COMPLETE DOCUMENT, IMAGE BANK – SELECTIONS

When archival materials are digitised, all material listed under a given inventory number is always digitised. The scans are listed in the inventory, and when it is indexed also in the Indexes.

When image material is digitised for inclusion in the Image Bank, it is sometimes decided to digitise only a selection of the items in a given inventory number. For example, three photos from an album of 100 photos.

The principle is that an inventory number is always digitised in full and included in the Archives’ Database. Hence, the user finds the full context here. A selection is then included in the Image Bank. The Image Bank only contains individual photographs, and does not include the reverse side, for example.

Example:

An inventory in the City Archives’ Database showing photographs with a front and back

30602: Collectie Bart de Kok en Jozef van Poppel

<table>
<thead>
<tr>
<th>Documentaire foto’s door Bart de Kok</th>
<th>566 bestanden</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Foto’s van de joodse zondagsmarkt in de Eerste Van Swindenstraat 1941</td>
<td>4 bestanden</td>
</tr>
<tr>
<td>2 Foto’s van mensen op straat in de Eerste Van Swindenstraat 1941</td>
<td>8 bestanden</td>
</tr>
<tr>
<td>3 Foto’s van straathandelaars en handelswaar op het Waterlooplein 1941</td>
<td>14 bestanden</td>
</tr>
</tbody>
</table>
**ARCHIVES’ DATABASE – WHOLE DOCUMENT, IMAGE BANK – CROPPED TO SHOW THE PICTURE ONLY**

The principle for inclusion in the Archives’ Database is that the document is always scanned in full, including its cover. Images are cropped to the border of the document. Scans of image material for the Image Bank are cropped to the image border.

If a photograph album is digitised, for example, the best working method is to start by digitising the whole album for use in the Archives’ database. This enables the user to see the full document in context (the number of photos per page, descriptions with the photos, order). Additionally, individual photographs can be cropped to the image border, and then can be included as separate images in the Image Bank. In all cases care needs to be taken to link scans with the relevant inventory number, allowing the user to navigate between scans of the album and scans of the individual pictures.
Viewing the photo album as an archival document

Viewing photos from the same photo album as images with a description for each picture
DIGITISATION OF NEGATIVE STORAGE SHEETS

The conventional problem with digitising negatives is that these are generally large collections (especially of 35 mm negatives) which sometimes contain hundreds of thousands of individual pictures. Digitisation of individual negatives is, because of the relatively large number of actions that need to be taken to make a scan, labour intensive and hence relatively expensive. It is therefore mostly impossible to digitise a complete collection at a high quality within the budgets available (and taking into account the high costs it is perhaps also not desirable). Making selections of negatives is however, from a practical standpoint, very hard. Moreover, making selections also leads to collections never becoming optimally accessible.

In 2013 the City Archives began using an alternative method for the digitisation of negative storage sheets. Instead of scanning individual negatives, the complete sheet is digitised, right through the transparent sleeve. The result is a picture of the sheet of a quality that is good enough for most research purposes. The degree of detail and tonality in the reproduction is limited, however.

The advantage of this method of digitisation is that the whole collection can be made digitally accessible at a relatively low cost.

After scanning:

- Individual negatives can be selected using a selection tool. The City Archives calls this type of scans ‘preview’.
- The preview images can then be described individually, for which crowdsourcing could be used. and the images are included in the Image Bank.
- The user can select and order a high-quality scan, on the basis of a preview.

The disadvantage is that setting objective standards for image quality of preview images is nearly impossible, and not very useful.
Würtbg.-Metaalwaren-Fabriek

MIELE & Co

AMSTERDAM

Kalverstraat 35-37.

PRIJSCOURANT
1896/97.