DIGITISE MORE, PAY LESS: An industrial evolution in the digitization of cultural heritage

Olaf Slijkhuis; Pictura Imaginis bv; Heiloo, The Netherlands

Abstract

In order to digitize large scale projects like the Dutch 'Images for the Future' project the production will have to become more and more industrialized. By dividing the digitizing process in different steps and by rethinking and innovating each of these steps it is possible to digitize at the highest standard and for a more competitive price.

Introduction

Digitization of cultural heritage is evolving more and more into an industry. With industrialization comes standardization and the development of an efficient workflow. The first step is to figure out a standard for digitization and to get an agreement on such a standard. Such initiatives are for instance the FADGI in the United States, Metamorfoze in the Netherlands and recently the standard set out by the 'Images for the Future' project also in the Netherlands. The next step is a concern for the digitizing companies. What do they have to do to achieve and maintain the set standard on a day-to-day basis? One of the answers is to industrialize the production workflow. In order to use standards you will have to standardize yourself. It means a reevaluation of the entire workflow. In the past few years we have worked on two initiatives, which forced us to reconsider every aspect of the digitization process. The first initiative involves projects that have to be carried out according to the Metamorfoze guidelines, the second initiative is a project for the digitization of over 1 million negatives called 'Images for the Future'. The lessons learned from working with these guidelines have proven to be a valuable asset for the overall production. As a commercial company our goal is not only to produce high volumes but we've also made the choice to always try and achieve the highest possible quality. Our aim is to achieve the Metamorfoze standard on every system that is used for the digitization of reflective material and since the start of the 'Images for the Future' project in October 2009 we've also set up a similar production line for the digitization of transparent material (negatives, slides).

The downside of trying to achieve this goal of the highest possible quality is the time it will take to get there. Digitizing different targets and using the software to measure them is often time consuming. For Metamorfoze projects we use imCheck software [1] to check the Q-62 targets and IE Analyzer from Image Engineering [2] to check the color. Besides the labor some task of checking each target another important aspect is that in order to maintain a constant and significant production flow every system needs to produce constant quality. The way to accomplish this is by innovation and a constant critical view of the production process as a whole.

The schedule shown in figure 1 provides an overview of the typical steps in a digitization process. From this schedule I will elaborate on a few steps to give you an insight in the possibilities for innovation and improvement on efficiency.

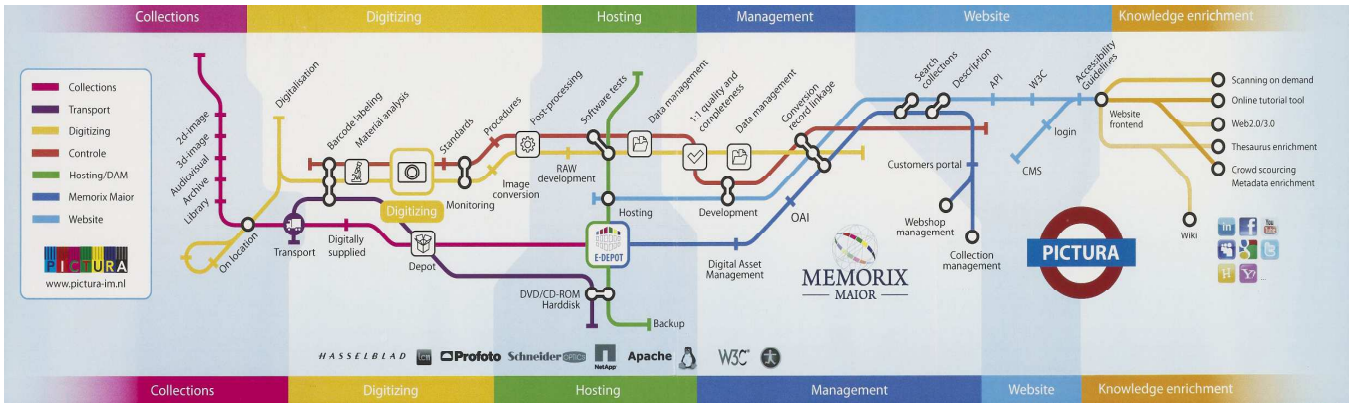

Figure 1

Collection

At the core of each digitization project is the material that needs to be digitized. A typical collection can consist out of 2d-images (reflective and transparent), 3d-objects, audiovisual material, an archive and a library. The first step in preparing a collection for digitization is the identification and labeling of each and every item that is leaving the institution. Every item has to be given a unique ID. This makes it easier to track what has left the institution and what comes back. For the digitizing company it makes it easier to track an item through all stages of the process. In our experience the use of a barcode as a unique ID has proven its merit in all the steps of this process. By scanning a barcode the risk of human error is reduced. Using barcodes makes it easier to automatically check for completeness during scanning. But even

here there is a risk involved. Labeling the objects with a barcode is manual labor and often this is where mistakes occur, it is easy to forget to label an item. In this situation the production line grinds to a hold. The contracting institution needs to be contacted for a solution and while production is going ahead the problem needs to be solved afterwards. This in turn will take more time then it should have.

An important and often overlooked aspect is the packaging of the original material. Does it come neatly packed or is everything assembled in one big box? Is it so tightly packed that removing the original material in order to digitize it will take up more time then necessary? Is the material clean, free from dust and other particles? Eventually all these issues can and often will be solved by the digitizing company but it will come at a cost. Thinking ahead and preparation of the collection is a key factor to keep the cost of digitization down.

In every collection there is a large variety of material. And while every type of material requires an individual and often different approach in digitizing, preparing uniform batches (in size, type, required resolution, etc.) will help the digitizing company to create a more efficient production workflow.

Digitizing

At the start of each project we have to analyze the items that need to be digitized. Is it reflective or transparent? What size, which filename, what are the exceptions, etc.? To create an efficient workflow it's easier to work with batches of uniform material. This is a preparation step that in an ideal situation is already being carried out by the institution. It reduces the time that the supplier has to invest in creating a more efficient production flow. Preparing uniform batches will give the digitizing company the opportunity to create a standard system that will be used for the duration of the project. Thereby contributing to an easier achieved level of quality.

More institutions demand proof of an achieved quality standard (like Metamorfoze or the Dutch 'Images for the Future' project). These demands have put a strain on the workflow of digitization. The amount of time needed to check the targets could reach up to 5 minutes per operating system. Extrapolate that to 4 times a day, 6 days a week and over 20 systems and it added up to a point that we came up with a solution to automate the daily checks of quality targets.

Metamorfoze quality control

The system is based on the Metamorfoze guidelines [3] with our own interpretation derived from putting Metamorfoze into practice for over three years. Our target is made up of a ColorChecker SG chart and 4 Q-62 targets in each corner (figure 2). These are proven and widely accepted standards.
We don't use a Q-62 target in the centre because we assume that sharpness is always less in the corners. The black background is chosen to reduce the flare, which could influence the readout of patches on the outside of the SG chart. Uniformity is checked by comparing the values of the white borders of the Q-62 targets. By doing so we've come up with just one test chart instead of the usual four (Color, Sharpness, Uniformity and distortion).

Figure 2. New Metamorfoze target

We've built a web-based tool that will read out the target in approximately 15 seconds and will show the test results to the operator.

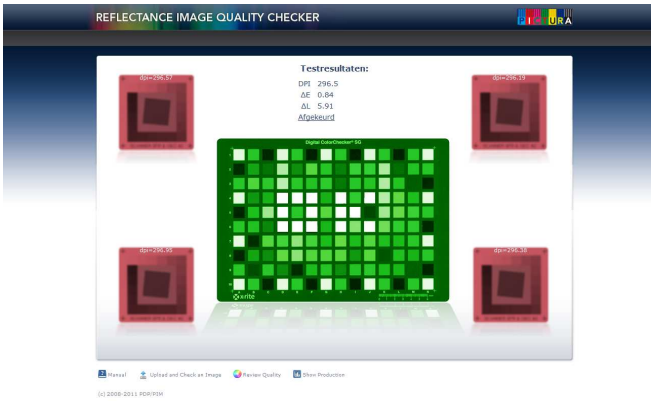

Figure 3. Target Result

When given a green light it is safe to go ahead and if the light is red the operator needs to call an expert to correct the issue (figure 3). The reported test results are indicative for any problem that might occur. Targets are stored with a unique ID based on date, system and project code. It is even possible to incorporate the measurement data in the header of each file. This workflow saves valuable time. From a production point of view it is possible to analyze the performance of each system by checking the statistical output of the data (figure 4), keep track of the performance of dE and resolution on all systems to name a few. More tools for analyzing the production workflow can be added in the future.

Moreover all systems can be evaluated according to Metamorfoze or other guidelines. To give you an insight at what this means, we've accomplished to get the value for dE94 from 7.0 just a year a go to a value of dE94 between 1.0 and 2.0 since the beginning of 2011.

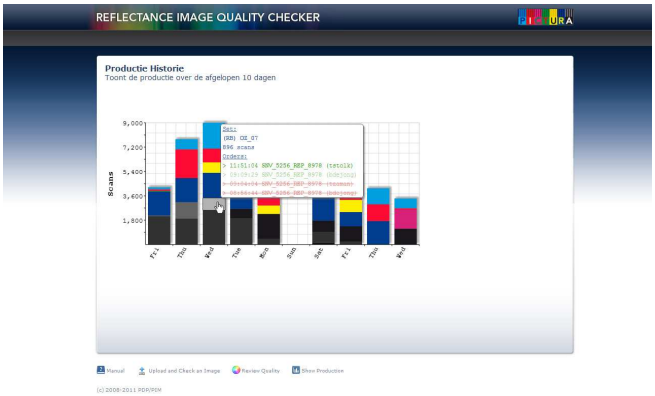

Figure 4. Production data

To achieve such results on an ongoing basis meant that we had to standardize all digitization systems. A steel construction is built around every workstation in which all the equipment has a fixed and stable position. This makes the control of lighting and sharpness easier to maintain. We've chosen for a one-shot system for several reasons. First of all the technological standard is such that the high standards can be easily met (dynamic range, SFR). We use digital camera backs with Dalsa sensors because in our opinion they provide the best range in tonality. The one-shot system is also perfectly suited for a high production rate. A critical factor is the technical skill involved to maintain the daily operation of the systems but by eliminating most variables, the time an expert has to spend to fix a problem is kept to a minimum.

'Images for the Future'

For the digitization of the negatives for the Images for the Future Project (a combination of the collection of the Netherlands Institute of Sound and Vision, the National Archives and the EYE Film Institute) we've set up a production line for scanning film based negatives, glass plate negatives, color slides and recently we've started with the high volume digitization of color negative film.

Quality Control

An elaborate quality control procedure was part of this project [4]. Three different targets needed to scanned at the start of each production day. And each had to be individually checked with the supplied imCheck [6] software.

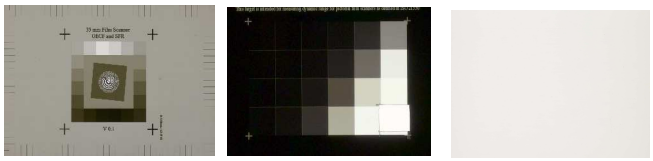

Figure 5. SFR, DNR and Uniformity targets.

A SFR target is used to measure sharpness, resolution, noise and distortion. The second is a target to measure the dynamic range and the third is a target to measure the uniformity. The use of targets and software has proven its value in production. The systems perform at a constant quality and it is easier to evaluate a problem. The experiences with working with quality control targets for all kinds of transparent material have led us to investigate the possibility of making an all-in-one target in the same manner as we did for the Metamorfoze target. Using the same web based software to check the targets would mean a further step in the standardization of the process.

Innovate the process

A large part of the 'Images for the Future' project (over 500.000 6x6 black and white negatives) comes from the collection of the Netherlands Institute for Sound and Vision. To meet the standard that was set out by the Images for the Future partnership and still achieve a high volume of production the idea to shoot 6 images in a single shot was developed into a custom made mask. To take out the individual images from this single shot we've also worked on an automated cropping system but found that the images were too dark (mostly shots from television studios) to make a perfect crop. Checking the images would take as much time as the workflow that we've eventually chosen.

Cropping from a distance

The images are cropped in India. The original high resolution images and the cropping tool to crop the images stay on our servers. A preview image is generated and presented to someone in India in a web browser. He or she will crop the image. The coordinates of the crop will then be sent back and implemented on the original. By dividing the process in different steps and fine tuning each step it made it possible to scan the entire collection in one year on one system. Although our initial start of making a fully automated cropping tool that would work according to our own standard of quality didn't work. It still provided us with the idea of the possibility to use it on other types of images. This is still an ongoing research project.

Quality is not all about targets

There is more to achieving the highest quality than checking targets. In the case of scanning transparent material it is also a case of removing dust and other artifacts. In our goal to industrialize the process we've built a laboratory like clean room which is pressurized to keep dust from coming in. For 35mm and 6x6cm negatives we've made a machine which uses ionized air to remove dust without the use of brushes. Even the operators wear antistatic clothing. Much can be done in advance. In preparing a collection to be digitized one has to consider the packaging material and the preparation of the collection before packaging. If it's possible to get rid off as much dust and other artifacts beforehand the quality of a digital image would gain considerably without affecting the production process too much.

Scanning on location

By critically reviewing the different parts of the production process and by taking that knowledge into account when creating new systems it was possible for us to set up two similar systems on location at the National Archives in The Hague for the digitization of over 50.000 glass plate negatives, another part of the 'Images for the Future' project. We used the same workflow with the exception of the overnight transfer of the digital files to our own storages for the post processing and data management part. A script was developed to automatically copy the files and when successfully

transferred the files are deleted from the scan system. The project was finished within three months after which we digitized other collections on location. The amount of files in the overall collection of the National Archives is around 500.000 images.

We've started with the digitization of the different collections within the Images for the 'Images for the Future' project in November 2009 and in march 2011 we have nearly completed the digitization of over 1 million negatives ranging from film based negatives to glass plate negatives, color slides and lantern slides.

Post processing

The next step in digitization is the processing of the images. A customer can opt for a raw scan as a digital substitute and derivatives that are used for presentation purposes. Because of the variety in material this used to be a process entirely done by hand. Recently we've figured out a way to automate the processing of both black and white and color (positive and negative) images. The post processing has shifted from adjusting every individual image by hand to a quick check of automatically processed images and the correction of just a few instead of all the images. At first this doesn't sound like an innovation. The innovative part is in the fact that we can automate this process for all types of images.

Management

In the schedule in figure 1 we can distinct different types of file management. One type has to do with collection management and another has to do with data management. This last type of management is another important step in the industrialization of digitization.

Data management

This is a step in the process where all files undergo a 1:1 check by which the original object is compared to the digital file. The barcode system proves its usefulness in this step of the process. The search for missing barcodes can be automated as can a check on the correctness of the filename. It is possible, if so desired, to generate an XML File which contains information regarding collection data, file size, resolution, etc. At the final stage of the delivery of files it is important to look back at the preparation that an institution has to do before starting a digitization project. To ensure an efficient workflow enough time has to be taken to prepare this stage of the project. If not all information is clear at this point the delivery of the images will be halted yet again.

In our experience by far the easiest and safest way to deliver digital files is by transfer by wire. This can be done overnight and it can be scripted so it checks if the files were sent correctly. Delivery of hard disks and occasionally DVD is still an option but it takes a large amount of time to copy the files to and from a hard disk. Delivery from server to server works more efficient for both parties concerned. In this regard the contracting institution will have to be aware of the amount of data that will be transferred. Do they have the storage capacity to store all the files that are being delivered?

Conclusion

Making digitization into an industry requires the commitment of both parties concerned, the institution for cultural heritage and the digitizing company. To ensure an efficient and affordable workflow the institution will have to pay more attention to the preparation of a project. This will often take place in collaboration with the digitizing company but this initial stage of a digitization project is for a large part the concern of the institution. A well planned and prepared project will give the digitizing company the opportunity to make a more efficient planning of their resources and thereby providing a more competitive price.

References

[1] imCheck software developed by Peter Burns, Image Science Associates.

[2] IE Analyzer developed by Dietmar Wüller, Image Engineering.

[3] http://www.metamorfoze.nl/publicaties/Richtlijnen/guidelinespijune07.pdf

[4] http://beeldenvoordetoekomst.nl/en/backstage/pictura-imaginis-images-images

[5] http://en.archief.nl/sites/default/files/docs/guidelines_digitisation_photographic_materials_0.pdf

[6] imCheck software for 'Images for the Future' developed by Peter Burns, Image Science Associates.

Author Biography

Olaf Slijkhuis's background is in Communication Science, Art History and Photography (practitioner). Since 2007 he works for Pictura Imaginis (the Netherlands). First as datamanager responsible for High-End scanning projects like Metamorfoze and since 1,5 years as production manager transparencies and as project manager for the 'Images for the Future' projects