https://doi.org/10.2352/issn.2168-3204.2018.1.0.24

Quality Assurance - Visual inspection of digitized images

Martina Hoffmann, National Library of the Netherlands, The Hague, The Netherlands

Abstract

In (mass) digitization it is common practice to work according to guidelines such as FADGI or Metamorfoze, to measure and monitor daily targets. Therefore it is common to implement some kind of quality assurance to assess if target values are met. There also is a huge field of digital preservation to ensure data can be stored appropriately. However, what about visual inspection of all those produced digital images? Why would we need to inspect the images? Visual inspection costs time, so how can we make it efficient enough, while maintaining high quality standards? How can we create a workflow for it? In this paper I will try to answer those questions based on experience from past years with a successfully implemented QA-workflow in the Netherlands.

Quality assurance in practice

Metamorfoze [1] is not only the name of preservation imaging guidelines, but it is foremost a project that funds 2D mass digitization projects for cultural heritage on paper. Within the program the objects are divided in two main sections – the books, newspapers and magazines and the archival section. This paper will deal with the latter only. For the archival section of this program there are several features to each project:

- The project has to contain originals that are unique, fragile and from a time between roughly 900 to 1950 AD.
- It concerns mass digitization according to the Preservation Imaging Guidelines Metamorfoze with quality levels Metamorfoze light or Metamorfoze.
- The preservation masters are to be stored in the long term repository at the National Archives of the Netherlands.
- The National Library of the Netherlands is responsible for the technical image quality and the data integrity needs to ensure long term preservation.

In order to ensure the last, the Senior production manager has set up a QA- workflow consisting of three major steps. The first one is called 'data-integrity'. It is the step where a mainly automated workflow ensures that the data derived from digitization is suitable for long term preservation. The second step is to make sure all hardware is up to standards by measuring test targets according to the Preservation Imaging Guidelines Metamorfoze. The last of these three steps is: the Sample. It is a controversial step because in the past years the speed of digitization rapidly went up and therefore the QA needs to speed up as well. Whenever workflows are developed and there is a need to cut time and money, the quick win would be to skip the visual inspections, as they usually take the most time (and therefore money) and capacity from employees. While one can automate quite some steps in QA for mass digitization, visual inspection is not one of them. Given these circumstances it is a challenge to implement a working, fast quality control process which includes visual inspection and sell it to the management. Working according to the Preservation Imaging Guidelines Metamorfoze or in other projects with the Guidelines for photographic material written by the National Archive of the Netherlands, we receive daily target sets which we measure accordingly. This makes a workflow predictable and for a great part easy to plan. It is also a very good wat to ensure objectivity. That is something all projects and approaches for (mass) digitization agree on – it doesn't matter if they are carried out within the Metamorfoze program or are independent projects. Handling huge quantities of objects and data in a high speed workflow with an objective and predictable outcome is great. It is even necessary for large digitization projects. But Guidelines and workflows do not address or are not set up for large visual inspection samples. They also, most of the time, don't answer to the question why one would need to visually inspect the images that closely. After all: We measured the targets so we can predict the outcome.

Visual inspection in a QA-workflow?

Taking a closer look at the Preservation Imaging Guidelines Metamorfoze there is also a section about artefacts. At this point we encounter a problem. Artefacts are, as stated in the guidelines, disturbances in the digital image and have to be checked visually. They tell us that images have to be inspected on 100% (actual pixels) in suitable software e.g. Adobe Photoshop. The guidelines however do not state how many pictures one has to visually inspect. So that is a puzzle we need to solve. Another one is objectivity. Whenever people look at things, they will see different things. When we are talking about halo's, mistaken crops, clipping or alike, one could argue that those are well defined mistakes in a digital image and therefore artefacts/mistakes like these are no issue for objectivity. If we don't take into account that one would need to define what crop would be acceptable beforehand for the project at hand we could indeed argue that things like that can be well defined. So while this is certainly true (to some extend as argued before), the same argument is not valid when talking about color differences. It is by now common knowledge that people see color differently – referring to the internet hype from early 2015 blue/white dress - but also research shows that especially in cultural heritage on paper there are lots of differences [2]. Furthermore, of course we have to keep in mind that there are many forms of color blindness which people may or may not be aware of. Besides that we are actually not very good with defining color as most languages don't have sufficient words to name all the various colors that we can actually see but not describe. The list of obstacles could be much longer but it seems clear enough now: All that are variables which make visual inspection a highly subjective matter far from the needed objectivity in QA-workflows. Besides the issues with or about equipment. Nevertheless it is important to have a visual inspection to spot unusual and unwanted disturbances in high quality imaging for preservation masters. That is one of the beliefs that is core to our understanding of our QA-business. Because, relating to the Metamorfoze program, this digitization effort is often the last rescue for the objects before they deteriorate and vanish.

While establishing that it is necessary to visually inspect the digitized products, the next big question arises. How can we ensure that we do this in a QA-workflow? With all the variables and obstacles to take into account? Furthermore, how can we do it in a statistical accountable way, to ensure quality on a large scale? First of all these are totally different questions. The how to and statistically accountable QA can be solved quite easily. Our solution: An AQL process derived from the ISO 2859 part I: 1999 EN "Sampling procedures for inspection by attributes – Part 1: Sampling schemes indexed by acceptance quality limit for lot-by-lot inspection" (see figure 1).

Voor het vaststellen van de grootte van de steekproef en beoordelen van de batch wordt gebruik gemaakt van ISO 2859 deel 1: 1999 EN "Sampling procedures for inspection by attributes - Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection". Uitgegaan wordt van de volgende parameters:
- grootte van deze steekproef is conform het Algemene Testniveau II.
- acceptabel kwaliteitsniveau is 1

Zie voor de steekproefgrootte en de acceptatie van fouten op het Algemene Testniveau II de onderstaande tabel conform ISO 2859 deel 1.

Algemeen testniveau II		Kwaliteitsniveau 1	
batchgrootte	steekproef grootte	acceptatie	afkeuren
		aantal fout	aantal fout
21-150	20	0	≥ 1
151-280	32	≤ 1	≥ 2
281-500	50	≤ 1	≥ 2
500-1200	80	≤ 2	≥ 3
1201-3200	125	≤ 3	≥ 4
3201-10000	200	≤ 5	≥ 6
10001-35000	315	≤ 7	≥ 8
35001-150000	500	≤ 10	≥ 11

Figure 1. *Partial image AQL as used for samples in QA workflow*

An AQL is an 'acceptable quality level' which is usually shown as a table with test levels according to the quantity of a particular element. One should read it as follows: whenever the size of your batch is between X and Y you should take Z as a sample. According to the chosen test level, there is a chosen quality level which indicates at which number of mistakes in a given batch acceptance or rejection is in order. Working with an AQL may seem easy but it actually is rather complex. To fully understand what the results of a test according to an AQL table mean, one has to know that it is never giving any certain results but probable and statistical values. It may not be necessary to understand the statistics or mathematics behind it to be able to use it, but it is definitely necessary to remain aware of the implications of the outcome. There is no certainty with using AQL. It is used by a variety of industries in their QA-workflows. It is a well-known method for sampling and test levels and defines the quality level that is worst tolerable. The mentioned ISO standard gives some further details and there are of course a lot of different national standards and even different acceptable levels of quality in different industries. E.g. Pharmaceutical industries have different 'worst tolerable quality levels' than the house cleaning industry.

In our case we found that this ISO standard gave us sufficient statistical accountability for our purposes. We are not actually claiming that once a batch of digitized images passed through our QA workflow it is completely and 100 % correct. We do allow quite a large margin of error in our samples before rejecting a batch. It has to be stated that while we ask for 100 % correctness of data-integrity this is not the case in the samples. Data-integrity is a fairly automated process which can be strictly defined by parameters accordingly. Visual inspection is far from automated and cannot be defined in total terms. To be able to say something about a batch of digitized images with 100 % certainty we would actually need to check the images individually. As stated before this is not realistically possible when working in large mass digitization programs.

This method of approaching the sample was used by the National Archives already for their past mass digitization project of photographic material: Images of the future, and has proven to be fast and therefore suitable for a high speed QA-workflow. Using an AQL does not mean one can just define that there is a sample taken according to AQL. It also means that one has to identify test levels and quality levels. Figure 1 shows the choices that were made by the National Archive and the ones we took over in our workflow. We did decide to also have a suitable AQL table for another test level but in practice almost never get to use it. In theory it means that whenever a batch of production gets approved the following production will get inspected at a lower test level. Meaning: taking less samples on the same batch unit with accordingly less acceptable mistakes. Both of the defined test levels can be found in the guidelines written for the delivery of data to the National Library in the archival section of the Metamorfoze program [3]. As it is a Dutch program, the guidelines are written in Dutch but the author can provide translations.

Implementing this step into the QA-workflow for the Archival section of Metamorfoze was a logical step to make in order to maintain high quality standards not only for our vendors but also for the program itself. Unfortunately this was easier said than done. First of all it meant to realize that it was a time consuming part of our quality control while handling high speed digitization of an average of 200 to 300 TB of produced data each year. Secondly we needed to ensure that subjectivity was reduced by establishing proper set ups and the training of people. The most important part was to sell this capacity consuming issue to the management. It was not until we had discovered serious issues in the digital images, which made a valid case for our choice to the management. Luckily we already had some experience with this workflow from the previous project at the National Archives. So we implemented the AQL method for samples in the QA-workflow for the archival section at the Metamorfoze program too. While from the QA point of view this was a natural and logical decision to make it was not easy to explain why we would need such a control at all. One of the big questions here was: What added value would a visual inspection have when we measure targets? Well, fortunately the Guidelines for Preservation Imaging Metamorfoze state that artefacts can't be measured and have to be inspected visually. By implementing an AQL based workflow we at least could say something with a little more reliability and statistically more appropriate than taking random five pictures and inspect those. Keep in mind here that the Guidelines do not state how to implement a visual control but merely state that you have to do this. Once that was cleared the struggle continued because of the time we needed to inspect visually and also how reliable our results were when 'everybody knows that yours eyes can be deceived and are not reliable at all'. At this point it is important to define what we actually trying to see when we visually control an image. What we want to look at on our screen in the archival section of Metamorfoze is a digital image of a physical object with as much as possible near-truth information. We are aware of the fact that we will never get an exact digital copy of the original object. One thing that we all should keep in mind as well is that a digitized object is never a copy of the physical object and there is no such thing as a true picture. Truth is that when the digital file finally arrives at the National Library (or any other client for that matter) it is a highly edited file. In any given digitization workflow (in-house, outsourced, with scanners or digital backs) most of the times even the raw images are edited by internal hardware algorithms. When a client finally gets an image it probably is not a raw image at all. There is a lot of software editing going on along the way to the end product. Therefore we define our objectives

while taking into account that we are in fact looking at nothing more than a resemblance of an original object. What we want is a complete object (refers to proper cropping according to specifications which in our case is always with 10 to 50 pixels background has to be seen around the object in order to prevent loss of information) with as much as possible information. But the most important thing is that we do not want to see information on our digitized images that do not belong to the original material. When keeping in mind that the most important thing in visual controls is that we can spot and identify misinformation in a digital image we need a successful workflow for that besides measuring targets objectively because the things we look for are not measurable in targets.

Objectivity in visual controls

Once we define what the objective is for our visual control, we have to take into account some other valid points about subjectivity of vision or color blindness or some other true arguments that one could argue with against the use of visual inspections.

In order to eliminate subjectivity as much as possible, we took several steps into a controlled environment. We searched for a place without interference of daylight and painted the walls in a gray tone. We took the lights off and replaced them with non-direct light away from our monitor screens. The screens we bought were calibrated monitors which we do now re-calibrate on a regular basis. The room where visual inspections are carried out is adapted to minimize distraction for the QA employee in terms of color or reflections. We take care to use high-end equipment which we maintain on a regular basis in order to monitor not only the incoming quality but also the quality of our inspection.

With this set up we can eliminate subjectivity of hardware output, daylight interference and even artificial light fluctuations which all contribute to objectivity. What we cannot eliminate is human vision and the interpretation of a digital image in relation to a physical object.

The most challenging part of visual inspection is the human eye and for that matter each human eye vision is a challenge. But on the other hand it is not. Our solution to that issue is extensive testing and training. Every QA-employee working in the archival section of Metamorfoze has had the same training and testing before they start visual inspection themselves. To ensure that everybody is equipped with the same information, the trainer always remains the same person and is the one most experienced in visual inspection – in the practical case it is the Senior Production Manager. The goal being that everybody can spot the same things in an image and we translate the visual abnormality understandably for others. In order to train people we need people with several different backgrounds. We preferably train people who understand digitization and know what happens when a physical object is reproduced digitally - which makes photographers preferred trainees one would think - and who have extensive knowledge about original materials they would probably encounter in our projects - which would make archivists preferred trainees. Truth is: We can teach people what happens when someone digitizes an object as there are lots of publications some more scientific than others about that and we can even show them what happens on hand. What we cannot teach people is what an original would look like if they have never encountered an object live. But that is what we at the end expect them to do: We do visual inspection solely based on digital images and decide which information on the image is probable to be in the original and which is not without seeing the actual original material.

Figure 2. *Partial image original with mold (tones of red color)*

What we expect people to be able to detect is that in fact there is something off with the color of the image shown in figure 2 and also make the connection that this is probably due to the original being damaged – in this particular case by mold. It means one has to know that some coloration - red is especially difficult to interpret - can have more than one cause and that it might appear worse in a digital image than in an original. When training people for visual inspection it is also very important to train the accompanying language. Visual inspection is relying on good communications as we cannot tell each other that we see 'something odd in red on the bottom of the right page'. We actually have to train people to identify - if they are not trained in conservation matters already - the specific language of damages for instance. That means we need both specialties in one person and besides that really good eye vision. For example: While color blindness is definitely a problem for visual inspection it is on the other hand easy to detect. In figure 2 we should at least agree that the color is something red rather than green. But there are many other things that are much harder to identify as shown in figure 3. The differences in tones are far less obvious here.

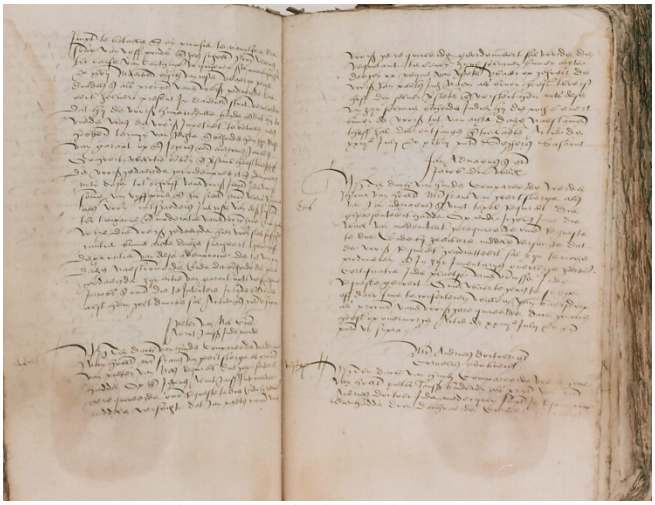

Figure 3. *Partial image original with water damage*

While these examples may seem obvious to most people working with handwritten archival material it is not that obvious when we talk about spotting dust on lenses or glass plates covering an original purely on digitized images.

Figure 4. *Partial image white cardboard (edited for print)*

The question we want to answer, without having to inspect all the images that we received, is whether the abnormality actually belongs to the original or not. In figure 4, we see a small bow shaped figure with in the original file, but also a wider, lighter smudge that follows the shape of the bow. There are two options here: either there is a smudge on the cardboard or it is a scratch in the glass plate. We know that these different issues can look very similar. The simple option would be to check all the following images for the same smudge, but unfortunately once the images are cropped they do not have to have the exact same size or frame. If the actual image could be shown in this paper, one would be able to see that there are differences in height that can give clues about where in the actual setting the smudge appeared. In this case, we are rather confident that the smudge is on the cardboard. If it is in fact a scratch, we would recommend the vendor to look at their machines closely to prevent damage to the set, originals and images.

Of every batch of digitized images which is on average between 10.001 to 25.000 images, we will randomly select 315 images to visually inspect according to our chosen test level in the AQL table from figure 1. A QA-employee has only a few hours per batch to measure the targets and to fully and completely inspect those images on actual pixel size and decide whether the image is correct in terms of cropping, straight position or artefacts. We have automated the taking of the sample with some simple scripts to prevent subjective interference here too.

Does visual inspection pay off?

When we implemented this workflow, due to really good vendor performance, we didn't have a lot of things that caught our attention or was reason for rejection of the batches based on visual inspection. Digitization grew and to pick up speed it was suggested to skip visual inspection because there were so little actual results. While it was still a necessary step from a QA-point of view, it can be considered a time/financial issue, especially if the management does not see the benefit in this inspection. We got lucky.

Not that long after we started carrying out visual inspections in the archival section of Metamorfoze we encountered a strange pattern in images. There were random blocks of slightly different color than expected. They were not visible in the targets and the measurements were normal. They were also not consistent in each image or the same size and shape in each image. We got our first big riddle to solve.

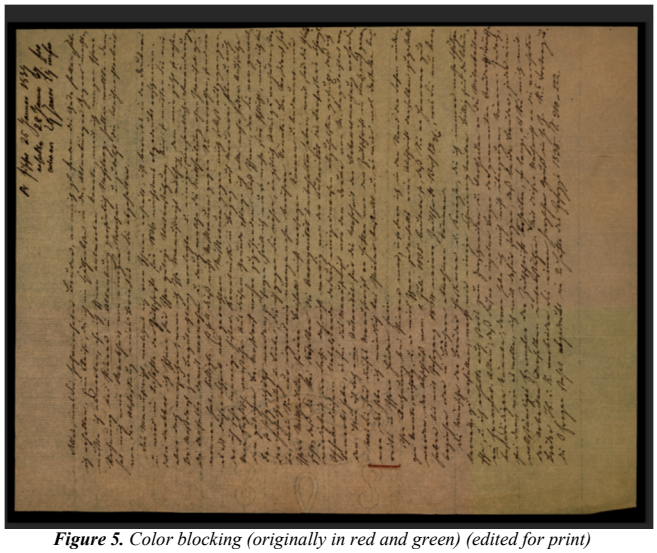

Figure 5. *Color blocking (originally in red and green) (edited for print)*

Lots of people, time and discussions further, with the combined effort with our vendor, the solution to our riddle was a hardware problem in a specific type of digital back. The hardware supplier took responsibility and fixed the problem. This was one of our show cases that proved to the management that implementing visual inspection like we did was worthwhile. When we want to aim for such a high level quality of digitization with mass and speed we also need a QA-workflow to ensure high quality. We don't believe we can afford to skip visual inspections because it is cheaper or easier. Our management was convinced. Since then we encountered numerous abnormalities in images and trained several people to catch those abnormalities. We encounter sharpened images as well as images with shadows or reflection, scratches that (could) lead to loss of (textual) information, dust problems or even problems with dirty glass. The above mentioned color-blocking returned a few years after the first encounter due to a software update in the digital backs. Visual inspection is now a highly valued step in our QA-workflow.

Lessons learned – so far

The biggest lessons we learned from visual inspections are that targets are definitely a mean to an end but are far from sufficient in quality control for mass digitization of preservation masters as they do not account for all the (automated) steps that a digital file undergoes after actually producing the image itself. There are lots of things that can go wrong even with the right hardware setup, for example automated crops. But there is more to it. Measuring targets gives some important information about the equipment. It also does give a hint of what to expect when looking at an image. That unfortunately is not the whole process: When working with digital files anything can go wrong in any step of the line. Bits and bytes are not always saved correctly and the impact on a digital image cannot be predicted when the network has a power peak or there is a lack of power. The hardware and software that is used to process the image has an influence on the final image. The type of editing can vary within a batch of data that comes from different hardware sets. The other big influence comes from the actual hardware that is used to take the image or make the scan itself. Of course there is a lot of software used to produce an actual image hardwired into the equipment, but also after taking the image there is a lot of things going on that might influence the

actual preservation master. For example while halo's most likely occur already in the first step of production, corrupting pixels can also occur later on in the process and not necessarily in all images alike. The lesson we also learned is that one can actually produce batches of images and we cannot see abnormalities in all of the images within the same production. Often we encounter random abnormalities within a production that even occur on the exact same equipment with exact same software and editing process. Some of those are still riddles while other can be explained after thorough investigation. Therefore we need to have an established statistical accountable quality workflow for visual inspection. Another big lesson was that visual inspection costs relatively much time in a high speed digitization and thus costs a lot of money. When you are not lucky enough to have a show case like we did, it is definitely hard to sell why a QA-workflow needs visual inspection. Also we learned along the way that while things were obvious to us working directly and consistently with digitized images it was absolutely not obvious to others. We really had to work on our definitions and explanations. One of the big misunderstandings for example was that we would actually be telling that the color of paper was off in the digital image. But that is not what visual inspection is all about. Another big misunderstanding was – and still is – that software could do visual inspections. Maybe in the future we have neural networks of AI that is able to catch abnormalities like we define them in digitized images but at this moment in time the human eye is still the fastest and best tool for it. Which doesn't mean that we do not try to work towards better software with the results we got so far. Actually one of our previous riddles let to a piece of software that we now run within our data-integrity software. It does not mean we stopped looking at those pictures but we can detect a certain abnormality faster and pick extra images to determine the impact on the batch of data.

Conclusion

Visual inspection can be implemented with low level costs for equipment but it takes time to train peoples vision and it costs time in the process. While hopefully we do not see any abnormalities in the digitized images, practice shows that visual inspection is a very valuable step in a QA-workflow that, at this moment in time, can increase the overall quality of the images as well as inspire innovations in hard- and software or adjustments in production processes. It is important to clearly define what the outcome will be and what aspects visual inspection does not cover. It is also not the only QA-step. Our workflow is based on controlled settings in location, hardware but also production setting – all images are produced according to given values from the guidelines. Visual inspection as a standalone QA would have to be set up differently and would also produce different outcomes with probably different definitions and objectives. In our case: We do want to find the riddles and solve them so everybody in the field can profit from the time and effort we put into it together with our vendors We do not aim for perfection and we do make mistakes too but we have a workflow up and running that gives pretty good probable predictions about the quality of the images.

References

[1] Metamorfoze, program: https://www.metamorfoze.nl/english

[2] Susan P. Fernand: Evaluating perceived capture quality for digitization of cultural heritage objects (IS&T, Riga, Latvia 2017) pg. 88-92

[3] Guidelines delivery of data for the archival section (Dutch only): https://www.metamorfoze.nl/sites/metamorfoze.nl/files/publicatie_documenten/Specificatiedocument%20KB%2010-12-2014_def.pdf

Author Biography

Martina Hoffmann is Senior Production Manager digitization at the National Library in the Netherlands for the archival section of Metamorfoze. She was operational manager quality control of digitized products in the National Archives in the Netherlands. She co-designed several quality assurance workflows for different mass digitization projects in the Netherlands. Starting with only image quality QA processes her main focus now are QA processes including several fields of expertise from metadata to long term preservation.