ARCHIVING CONFERENCE
IS&T imaging.org
https://doi.org/10.2352/issn.2168-3204.2020.1.0.43

When perfect is the enemy of good
Quality and sustainability in digitization processes

Millard Wesley Long Schisler
Museum Studies, Johns Hopkins University

Abstract

When we talk about digitization processes, it is necessary to understand that they can be done in countless different ways, qualities and techniques, from cell phones to high resolution captures or more complex systems such as multispectral and three-dimensional digitization. These differences have an impact on the amount and detail of information captured by each technique, the human and material resources necessary for each technique, maintenance, and long-term care for the digital surrogates generated. The desire for digitizing cultural heritage materials must be linked to programs for the preservation of the materials being digitized and the preservation of all the digital files generated by digitization. The institution needs to find the balance between the resulting quality and quantity of the materials that will be digitized and what is possible to sustain in the long term through these processes. When we want the great, this can be the enemy of the good. The good can be done in the best way and be great. It is also possible to work with projects ranging from good to great according to decisions and selections made by the institution on how to deal with digital preservation, digitization, access and preservation in the long term. We want the best for our institutions and collections, and we aim for efficient dissemination programs using the materials generated by digitization. The purpose of this text is to help us think about our wishes for digitization and dissemination within this universe of possibilities.

1. Introduction

1.1 The origins of the expression

"Perfect is the enemy of good" is a phrase found in Orlando Pescetti's book *Proverbi Italiani* (Pescetti, 1603). Voltaire brings this phrase to light, as he quotes this Italian proverb "Il meglio è l'inimico del bene" in his *Dictionnaire philosophique portatif* (Voltaire, 1764). Voltaire starts off his poem *La Bégueule: Conte Morale* (Voltaire, 1772), with "Dans ses écrits un sage Italien Dit que le mieux est l'ennemi du bien" (In his writings, a wise Italian says that the best is the enemy of the good). The expression can also be traced to Aristotle, and other classical philosophers through the principle of the golden mean, where the ideal of the golden middle would be the exact measure between two extremes – one of excess and the other of deficiency.

We can also refer to the Pareto Principle, coined by the American electrical engineer Dr. Joseph M. Juran, in the 1940s, in reference to the Italian economist Vilfredo Pareto. Pareto noted the 80/20 connection while at the University of Lausanne in 1896, in his first work: *Cours d'économie politique*; in it, Pareto showed that approximately 80% of the land in Italy was owned by 20% of the population. This principle has been applied to many fields, and as Juran would say, 20% of the effort yields 80% of the results, and the remaining 20% of the results will require the remaining 80% of effort to be achieved.

Can a digitization project produce 80% of the results with 20% of the budget? Will this 80% be good enough or do we need the other 80% of the budget to complete the additional 20% of effort?

We do see more contemporary discussions of this term in the 20th and 21st centuries presenting ideas like "Do it right the first time" which propose the opposite, where good is the enemy of perfect. Just doing something good means that we might have to do it again to get it better, or perfect. It also might mean that we will never really do it over again and end up living with the good.

Therefore, taking into account these expressions, the goal here is to find the balance between good, perfect and sustainability in digitization projects.

1.2 How this relates to digitization projects

As pointed out in the abstract, there are many techniques used in digitization projects. The higher quality, precision of reproduction and detail of information are important and desirable goals. But to reach this "quasi-perfect" level of reproduction through digitization we will also require human and material resources to produce and preserve this production that are out of reach for most institutions throughout the world. Can we produce 80% of the quality with 20% of the resources? Can we have a project that strives for the golden mean, searching for the balance between a good and a perfect project and a quality that can be produced and sustained for long-term preservation? These are all questions that are relevant and should be asked when planning digitization projects.

2. Digitization as one of the components

2.1 What comes before

When we develop digitization projects, we will be generating digital surrogates of analog materials. These analog materials need to be processed before being digitized. This also means that in almost all cases, we intend to preserve the original in the ideal environment with proper enclosures to protect the object, and environmental controls to provide long term preservation. All the materials will need to be identified and cataloged so that we know what we have, their meaning, and how and where they are housed. Once these procedures have been done, these materials are ready to be digitized.

2.2 What comes afterwards

The digitization process duplicates your collection, creating digital surrogates of the analog objects. This new collection also needs to be taken care of. All the identification and catalog information of the originals needs to be input into the metadata of the digital objects, collection management or similar systems. This new collection of digital surrogates will need to be preserved within a digital preservation program to guarantee long-term access and preservation. These digital assets now will be important tools to enhance access to the collection, integration into new projects, educational use, exhibitions, among other uses. This means that it is imperative that the institution planning for a digitization process considers the resources needed to create and maintain programs that give life to the collection. See *Figure 1*. Digitization and digital preservation can dangerously result in the generation of "dark data", that is carefully stored in collections but not accessed and used by anyone.

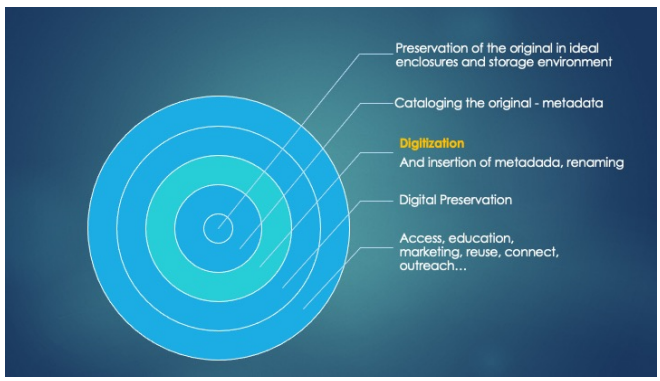

Figure 1 *– Processes before and after digitization*

3. Looking at some of the main variables in digitization projects

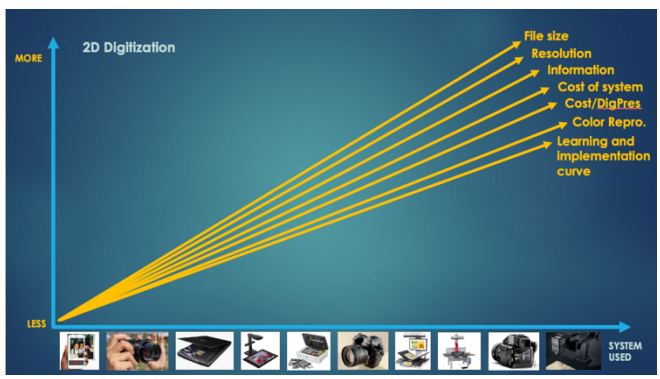

Figure 2 *– Some of the important variables in a digitization project.*

There are many variables in a digitization project, but the purpose of this text is to address a few of the most important ones that are relevant to this discussion. As we look at Figure 2, we have a representation on the X axis of several digitization systems from very simple ones, like a cellphone capture, to more sophisticated and complex systems as we move all the way to the right of the axis. This represents a simplification of the innumerous digitization systems available. On the Y axis we have LESS at the bottom, and MORE at the top. Less resolution or More resolution, for example. The seven items we are discussing on this graph are:

1. File size
2. Resolution
3. Information
4. Cost of system
5. Cost of Long-Term Digital Preservation (LTDP)
6. Color reproduction precision
7. Learning and implementation curve

All these variables share one thing in common. As the systems become more complex, more expensive and with better quality components, we start achieving MORE of each of these items.

If we start with file size, the smaller sensors in cellphones and point and shoot cameras will produce smaller file sizes than a full-frame high resolution DSLR or mirrorless camera, and even a digital back.

The differences in file size can be quite substantial and depending on the system being compared, we can be talking about 20MB for a RAW file on a 20MP pixel camera, and close to 100MB for a RAW file produced by an 80MP digital back. More recent developments of backs with 150MP (Phase One) and 400MP (Multi-shot Hasselblad back) and scanning backs that reach even higher megapixel numbers (Rencay Scanning backs reaching 2808MP) will push these numbers for file sizes even higher.

With the increase in file size we obtain more resolution, capturing more details from the original object. This translates also as more information and precision in the reproduction. There is also less risk of not being able to capture all the information in the original when we have more detail and resolution.

If we want the best results then, we must strive for systems that can produce the highest resolution. These systems have a higher price tag, so it is imperative to be looking at their cost. Digital equipment has a life span. They will eventually need maintenance, become obsolete, or we might want to purchase better equipment after the existing equipment has been used for half a decade or so. Therefore, the investments we make in digitization systems are not long-term investments and need to be understood this way because they will eventually need to be renewed.

The overall cost of systems as we move to the right on the X-axis will increase substantially. Most of the systems towards the end of the axis are completely out of reach for the majority of cultural heritage institutions around the world.

If we want or need more resolution or more information, this translates into larger file sizes and higher cost digitization systems. These large file sizes will also impact the cost of the infrastructure to process and store these for the long-term. The implementation and maintenance costs of long-term digital preservation programs that are caring for 20TB of data are very different from those caring for 20PB of data. Even before starting a digitization program it is necessary to have a clear vision of the digital assets that will be generated and a digital preservation program in place to care for these. This has to be a program that the institution can manage and sustain for the long haul. According to a well-known idiomatic expression: "one's eyes cannot be bigger than one's stomach".

Color reproduction precision is another topic that is very important in digitization. As we create a digital surrogate, it should be as close as possible to the original. And reproducing the tonal and color scale to match the original requires a series of targets, tools, knowledge and software that are all part of a color managed workflow. Better systems are more capable of producing higher

color fidelity in reproduction. This will be addressed further in the section 5, on standards for digitization.

The amount of learning and the timeline to implementing a digitization program will also depend on the complexities of the system being used. Teaching someone to use a simple flatbed scanner can be a one-day task and this user would already be capable of performing well with this equipment and processing the digitized files. The same learning process with higher end equipment, digital backs and specialized software, for example, will already require an operator with previous skills and further specialized training to understand and know what they are doing. For learning and implementation, we need less time, and less qualified people to start work on simpler systems and more time and more qualified people to start work on more complex systems.

3.1 Where is the sweet spot

Having the best of everything on this list can give you excellent results. It may not be something attainable, and if attainable, not sustainable for most institutions. Let us go back to the initial discussion and try to find the golden middle, or the sweet spot. In finding the golden middle, we strive to find solutions with reasonable amounts of resources, human and material, that can make digitization projects more widespread and accessible around the globe.

4. The expansion of digitization into many other areas

In this text we are discussing mainly two-dimensional digitization programs. On the graph exhibited in *Figure 2*, where we talk about all the different systems, we have left out other forms of digitization that produce a lot more information, even beyond what the eyes can see. We need to mention this here, as these systems are important for institutions because of the results they produce and how these results can be used. They also represent MORE of everything that is already on our scale.

Even though these methods of gathering data are very interesting, they are outside of our initial discussion. We can name a few of these, such as multispectral imaging or imaging with all the many different wavelengths, photogrammetry, RTI–Reflectance Transformation Imaging and many diverse methods of 3D digitization as exemplified below in *Figure 3*.

Figure 3 – Going beyond 2D digitization and capturing much more information about objects.

5. Quality standards in digitization

We have developed many standards for reproduction before the digital era that dealt with resolution, color and tonal reproduction, lighting, among others factors. With the development of digitization technology, from hardware to software, we now have published standards that define very clearly how to measure different aspects that are captured through a digitization system and rate them accordingly. The FADGI (Federal Agencies Digital Guidelines Initiative) document entitled: *Technical Guidelines for Digitizing Cultural Heritage Materials, 2016*[1], reflects the most recent advances in imaging science and cultural heritage best practices for digital imaging. These guidelines are intended to be used in conjunction with DICE, Golden Thread, OpenDICE and AutoSFR (Digital Image Conformance Environment) targets and software. Together, these guidelines and the DICE testing and monitoring system provide the foundation for a FADGI-compliant digitization program. According to the compliance, a digitization system will receive a star rating of FADGI 1★, 2★, 3★ and 4★. A FADGI 4★ rating means that the system used for digitization excels in all the areas of image reproduction. Simpler digitization systems are not capable of achieving this level of compliance.

The National Library of the Netherlands published their guidelines for digitization programs, the Metamorfoze Preservation Imaging Guidelines[2]. On the Metamorfoze website they state that "*The images produced for Metamorfoze must adhere to quality standards and retain a verifiable relation to the original in such a way that they can serve as a replacement of the original object, as the originals are withdrawn from use after preservation. However, there will be different requirements for different types of material* (2010)."

These different requirements are similar to the star system defined in the FADGI compliance. The Metamorfoze guidelines use Metamorfoze Extra Light, Metamorfoze Light and Metamorfoze (Full). In measuring 13 different technical image criteria, a Metamorfoze project must adhere to very strict standards to achieve the necessary goals specified in these 13 criteria. An Extra Light project is less strict, or less demanding in terms of reaching all the goals of a Metamorfoze project; an example can be the digitization of newspapers versus the digitization of historical paintings.

For both FADGI and Metamorfoze systems to be used we also need specialized targets and software so that we can measure the data generated through the digitization of these targets as we can see in *Figure 4*.

[1] http://www.digitizationguidelines.gov/guidelines/digitize-technical.html

[2] https://www.metamorfoze.nl/english/digitization

Figure 4 – FADGI, Metamorfoze, targets and software for evaluating technical image criteria.

There is also here a direct relationship between the cost of the entire system (hardware, software, targets) and the quality standards that they can achieve. Entry level digitization systems are not capable of reaching the higher levels of image reproduction laid out by the FADGI and Metamorfoze specifications.

If we want to build a digitization project, we need to understand the possibilities of each system and the output according to these standards. The graphs in *Figures 5* and *6* are for illustration purposes to show the different standards, FADGI and Metamorfoze, and their approximate equivalence when mapped to a range of digitization systems.

Figure 5 – Simple systems for digitization achieve FADGI 1★ and 2★. Only more expensive systems, with the use of targets and software are capable of reaching FADGI 4★.

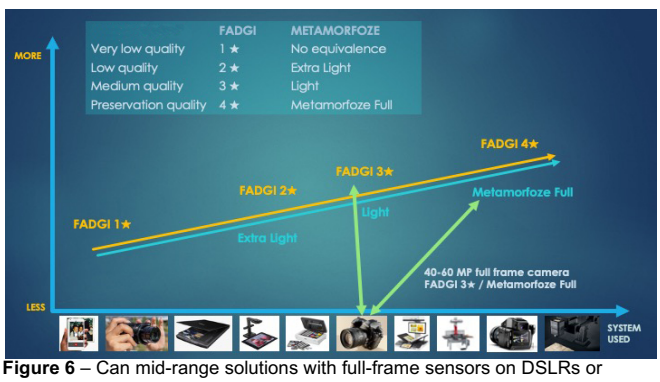

Figure 6 – Can mid-range solutions with full-frame sensors on DSLRs or Mirrorless cameras with resolutions from 40-60MP, aim to achieve FADGI 3★ and Metamorfoze Light or Metamorfoze Full?

As we can see in *Figure 6*, we can have mid-range solutions that fall into the Pareto Principle where 20% of the effort/cost yield 80% of the results, and the remaining 20% of the results will require the remaining 80% of the effort/cost to be achieved. I added the word "cost" to this principle because cost has to be an important consideration for most of the institutions throughout the world. It is not possible for all institutions to aim for the best systems capable of achieving the best results.

As both the FADGI and Metamorfoze standards are very clear in specifying, not all materials have to or need to be digitized with the highest standards. Institutions should be able to create priorities of what types of materials receive what kind of level of digitization. Not all institutions are capable of the necessary human and material resources to be able to perform digitization projects in all these different levels of quality.

Therefore, it would make a lot of sense to have a very high-quality digitization setup capable of achieving FADGI 4★ and Metamorfoze (Full), and also other more complex digitization systems (hyperspectral, 3D, etc.) as a central lab that can produce the "quasi-perfect" digitization. This lab can serve as the hub of a larger complex of digitization labs spread around the region or country that are producing "good" quality digitizations, that can fall into the category of FADGI 3★ and Metamorfoze Light or Metamorfoze (Full).

We can have the best possible *Good* digitization, and the *Perfect* digitization when possible. This can allow us to create more good quality digitization programs spread across all our institutions, regions or country. The high-quality labs help establish the best practices and image quality control to be used in the more sustainable, less expensive "good" digitization labs.

6. Digitization projects to consider
Example 1 – Southern Historical Collection

The Southern Historical Collection from the University of North Carolina was awarded a grant from the Andrew W. Mellon Foundation. Through this grant, SHC archivists promote community-driven archiving by aiming for grassroots efforts to record and preserve local history. The archivists at SHC provide training, technical know-how, supplies and equipment to perform the archiving. The backpacks contain items to assist a community-based historian to record oral history interviews and also have the capacity to digitize photographs, letters, documents and meaningful artifacts.

This is an interesting example of digitization that is gathering important information from the underrepresented minorities in the North Carolina region. The equipment used has not been tested under the FADGI or Metamorfoze guidelines, but most likely the results would put this digitization project into a FADGI 1★ or 2★ and Metamorfoze Extra Light.

As we mentioned in the beginning of this paper, digitization is a component of a much bigger ecosystem. In many, if not most, situations, our goal is to make the information gathered from these documents accessible to researchers, educators and the public. We want these documents to become embedded in our societal knowledge. Decision makers will have to evaluate long-term costs and benefits of the different digitization systems. Therefore, as was stated in the beginning of this paper, the goal here is to find the balance between good, perfect and sustainability in digitization projects.

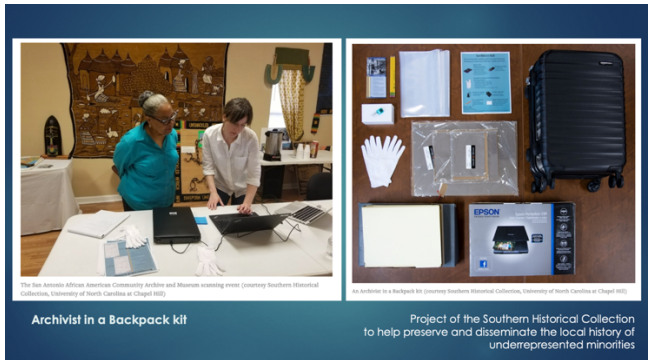

Figure 7 – The Archivist in a Backpack Kit from the Southern Historical Society[3].

7. Digitization projects to consider Example 2 – the *Rede Memorial* digitization kit

Another example of a digitization project was developed in 2016 in Brazil to train teams in nine cultural heritage institutions spread throughout the country to be able to digitize their collections. Each institution received an intensive one-week training on all the digitization equipment and then online and email monitoring during the duration of the project. This digitization kit was performing camera scanning setup using a 24MP full-frame digital camera with a macro lens to capture the images.

At the time of this project we did not use any targets or specialized software to define the image criteria for the digitization. None of the institutions involved in this project had any experience with camera scanning and these kits were the first tools they had for digitization. Several of these institutions have now started developing their digital preservation plans for not only the digitized materials but also the born digital ones. An entry level kit like this is can be an excellent way to introduce the production, care and management of digital files and get the institution involved in working with digital files, building expertise, and moving into a digital preservation program.

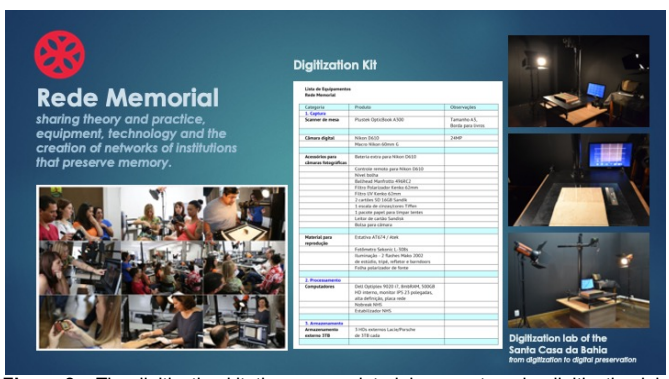

Figure 8 – The digitization kit, the one-week training event, and a digitization lab in one of the institutions.

8. History bias in favor of certain cultures

It is very important to consider the ideas of Margaret Hedstrom, an associate professor at the University of Michigan's School of Information. In an interview for the Digital Preservation Pioneers spearheaded by the Digital Preservation initiative at the Library of Congress, Margaret talks about a digitization and digital preservation project where she and her team of students helped assemble the archives of the liberation movements, including the records of the African National Congress at South Africa's University at Fort Hare (UFH). The article[4] is really worth reading.

It states that "*Margaret's experience gives her a unique perspective. Though UFH is not a Third World institution, it faces enormous resource obstacles. As Western institutions blaze forward, preserving terabytes and petabytes of their cultural content, institutions like UFH lag behind. "We need to appreciate the privileged position developed countries are in regarding technology," Margaret says. "The most threatened content in the world is the unique content in developing countries."*

She continues talking about the skewed versions of history: *"…we need to be cognizant of the fact that because almost all the (digital preservation) emphasis is coming from a small number of developed countries, we're going to have a skewed historic perspective." Even if Western countries help digitize the collections of developing countries, it is crucial to also help them create a solid infrastructure and advance their capacity to care for their own digital heritage."*

We could substitute digital preservation for digitization in the last paragraph, and even when she says that the initiatives are coming from a "small number of developed countries", we could narrow that down to a "select group of institutions in a small number of developed countries". The biases and skewed visions will also exist when we consider the disparities amongst regions and institutions within the United States, Brazil, as well as other countries.

Considering these ideas, we should strengthen the ideal of having many different levels of digitization programs going on at the same time. The higher-level labs could be strategically placed and be beacons for countless less expensive labs that are aiming for the sweet spot or golden mean. This will also help spread and democratize the use of technology, digitization, use of digital data and digital preservation programs to all regions and enrich the data that we will give access to and preserve.

[3] https://library.unc.edu/wilson/shc/community-driven-archives/archivist-in-a-backpack/

[4] http://www.digitalpreservation.gov/series/pioneers/hedstrom.html

All images from this text are from this author's presentation at the ApoyOnline 30th Anniversary conference, Rio de Janeiro, Brazil, September 23-27, 2019.

Author Biography

Millard Schisler is a photograph and media conservator, with a background in teaching. He holds an MFA from the Visual Studies Workshop in Rochester, New York. He also completed the two-year Certificate Program in Photographic Preservation held at the George Eastman House. He later taught historical photographic processes for three years in the same program. He was a professor at the School of Photographic Arts and Sciences and School of Print Media at the Rochester Institute of Technology from 1996 to 2006, where he was actively involved in teaching during the transition of traditional to digital photography. Mr. Schisler moved back to Brazil in 2006 and became the Director of Preservation of the Cinemateca Brasileira (Brazilian Film Institute), where he was heavily involved in the preservation of all types of media in analog and digital formats, from 2007 to 2012.

In the last ten years, he has worked as a lecturer and consultant on many different projects in the field of organization, preservation and digitization of photographic collections and moving images, and becoming largely concerned and active with digital preservation.