

“Zeutschel prospective on how AI will change our digitization workflow”

Christian Hohendorf, Andrea de Polo Saibanti

Zeutschel GmbH, Germany

Christian.Hohendorf@zeutschel.de, Andrea.Depolosaibanti@zeutschel.de

English summary: Artificial intelligence (AI) is a technology that allows us to generate, classify, and perform tasks such as video and image analysis and speech recognition, discover meaning and learn from past work. Machine learning and AI can indeed speed-up the entire workflow for example of a digitization system, including cataloguing, content retrieval, auto-tagging, digital restoration, etc. This paper analyze practical case studies in the cultural heritage sector, providing some ultimate recommendations for best practice digitization.

Keywords English: workflow, Kitodo, ISO 19264-1, image quality, standard, Metamorfoze, FADGI, Universal Test Target, Permanent Image Quality Control, guidelines, CMS, open source, accessibility

Българско резюме: Изкуственият интелект (AI) е технология, която ни позволява да генерираме, класифицираме и изпълняваме задачи като анализ на видео и изображения и разпознаване на реч, да откриваме смисъл и да се учим от предишна работа. Машинното обучение и AI наистина могат да ускорят целия работен процес, например на система за дигитализация, включително каталогизиране, извличане на съдържание, автоматично маркиране, дигитално възстановяване и т.н. Този документ анализира практически казуси в сектора на културното наследство, предоставяйки някои крайни препоръки за най-добри практики в цифровизацията.

Ключови думи български: работен процес, Kitodo, ISO 19264-1, качество на изображението, стандарт, Metamorfoze, FADGI, универсална тестова цел, постоянен контрол на качеството на изображението, насоки, CMS, отворен код, достъпност

1. INTRODUCTION

The participation in and engagement with culture creates value in our sector and society at large will build a highly accessible service for cultural heritage digitization and preservation which meet the licensing rights and IPR from the various content owners. This paper is related to the best practices in digitization in a world where new technologies like AI and machine learning are changing our digital environment.

We want everyone to have the opportunity to participate in the digital cultural heritage - to access it, to share it, to enjoy it, contribute to it and to be enriched by it. We want to target in this proposal large community of users, including Europeana community, spanning from humanities researchers to the creativity industry, archivists, art historians, graphic designers, bloggers, teachers, students, advertising and marketing, policy makers.

In this respect Zeutschel (<https://www.zeutschel.de/en/>) is providing concrete solutions for digitization (capturing solutions), content management systems (Kitodo open source solution (<https://www.kitodo.org>)) and quality control including cataloging and preservation.

Today, preservationists worldwide are increasingly looking at digitization as one viable solution for protecting heritage under threat. Both tangible and intangible heritage artifacts—from physical relics to oral histories—are now being captured and preserved via diverse digital tools and techniques. In addition to supporting the resilience of cultural and historical heritage, digitization can help make heritage sites and artifacts more accessible to a wider audience.

2. DIGITIZATION FOR THE FUTURE

Zeutschel have been delivering the highest quality, technologically advanced products to our customers worldwide for over 60 years and we are the driving force in the digitization of cultural assets. Due to these many years of experience and our worldwide history with libraries, archives and museums, we are familiar with the requirements for high-quality digitization . This understanding of industry requirements shapes our product strategy and development. The internationally recognized UTT test charts were developed and have now found their way into the ISO 19264-1:2021 standard "Photography - Archiving systems - Imaging systems quality analysis - Part 1: Reflective originals". A working group of international experts (ISO TC42 JWG26) developed this standard. In addition to Zeutschel GmbH, well-known institutions from almost all nations (eg Metamorfoze , Danish National Library, Swedish National Archives, Cambridge University, Library of Congress, Metropolitan Museum, National Archives, etc.) are all represented in this working group, as well as the development of general standards, innovative methods of digitization are created as part of this cooperation and exchange.

All scanning and processing technologies should be designed to process the originals gently and in the highest quality and to treat them with the greatest care, to rule out any damage.

The fundamental aim must be to digitize each original only once, if possible, so that the content can be researched worldwide. Therefore, when using any scan system, in addition to the image content, an optimal image quality in terms of detail (resolution, scale, geometry) and tonal reproduction (brightness, dynamic range, color, noise) play the greatest role. The image quality is not only assessed on the actual image, but also on test charts which are digitized on the system under identical conditions. A software-supported measure of the image quality can then be determined in these scans.

The methods used are described in international ISO standards and relevant technical guidelines. These include the classification of the measurement results in defined quality levels. Subsequent manual correction using image processing should only be necessary in exceptional cases. Therefore, in addition to the selection of the right scanning system, the competence of the scanning employees is key for the optimal results.

The best “tool” is worth nothing if the employee working with it does not have the necessary qualifications, both in operating the systems used and in dealing with archive materials. Due to our work with highly sensitive, historical templates and valuable art collections, knowledge of the principles of conservation when dealing with media of all kinds is a matter of course for all of our employees. One of the most important components for working with cultural assets is therefore the continuous training of employees with regard to dealing with and handling fragile and valuable documents, as well as behavior in the vicinity of archival media. We therefore not only offer these training courses for our employees, but also offer them to all interested parties and create video tutorials on the relevant topics. The final content of these tutorials comes from active participation and regular exchanges with the relevant specialist standard and quality committees.

In addition to well-trained employees the scanning hardware technology used is also key for the "perfect" digital copy - whenever necessary, contact-free processing must be used, which can still deliver the highest possible quality as a result. The photographic process of reprography has proven particularly useful for this and not only offers the advantage of increased quality and depth of field, but with the appropriate choice of perspective, a three-dimensional recording and compensation of the curvature of the original can also take place. Furthermore, the light exposure plays an important role in the processing of light-sensitive originals: depending on the imaging principle, a low exposure is achieved by short sequential illumination or a low illuminance.

In this respect the Kitodo is a perfect solution to match and being integrated into the Zeutschel digitization ecosystem. Why Kitodo than is the next question? The answer is very simple: Kitodo is an open source software suite intended to support the digitisation of cultural assets for libraries, archives, museums, and documentation centers of all sizes. A range of modules with open interfaces support the production, presentation, and archiving of digital assets. The software can be flexibly used for a multitude of digitisation strategies and scalable business models – for in-house projects, purely corporate services, or hybrid endeavors. Kitodo is backed and continually updated by a dynamic user and developer community and the non-profit association Kitodo e. V.

Our proposal goes toward the participatory practices, standards, platforms, tools and guidelines best practice recommendations to encourage, stimulate and help communities (private and public sectors) to actively being engaged, participate and collaborate in a digitization and preservation practice.

Building partnership, encouraging stakeholders, private and public institutions to get involved in this initiative also through crowdfunding and crowdsourcing activities, creating networks of users, involving and stimulating various class of citizens to join our proposal is the earth of our project to make it financially sustainable and attractive to the industry too.

3. DIGITIZATION BEST PRACTICE

In today's digital business ecosystem, the efficient management of physical images and documents has become essential. Photos and document digitization is not simply converting files into digital format through a scanner or digital camera; it is a complex strategy to optimize information technology and the digitization workflow. Below, we explore what this process consists of, its benefits and the best practices for a successful implementation in the cultural heritage and library sectors.

This paper will facilitate and boost further the aggregation and harmonization of old and new content and will stimulate, facilitate and encourage citizens, private and public institutions to digitize their content in a more efficient way.

The aim it to make more visible and accessible digital content laying in the national repositories, facing obstacles of different catalogue and descriptive schemas, different languages, different organisation of collections and themes for searching. Our approach will work on the digital formats producing new contents and new description and will facilitate the use of these historical valuable contents by the new social networks approach and mechanisms enlarging extremely the potential user target arena.

All of these points inspire and influence us as Zeuschel in our daily work and we do our part to preserve unique cultural assets and to secure art and knowledge for future generations.

Document and imaging digitization is not just converting files into digital format through a scanner or digital camera; it is a strategy to optimize information management. Below, we explore in depth what this process consists of, its benefits and the best practices for a successful implementation.

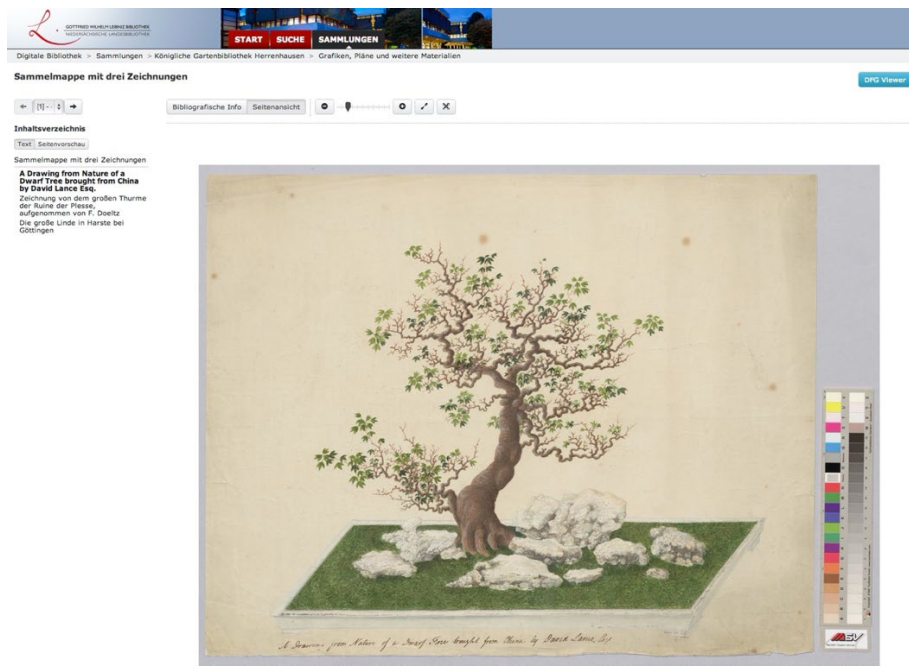
Below the most important and essential steps in the digitization best practice workflow, taking into consideration also the latest advances in machine learning and AI.

1. Budgeting, project time frame, defining usage of the digital content
1. Preliminary analysis: content selection and tasks prioritization
2. Preparation of physical documents (remove unwanted items like staples, paper clips, etc.)
3. Scanning/camera digitization preparation and Quality Preservation
4. OCR and cataloguing for easy searching
5. Backup, storage and long term preservation

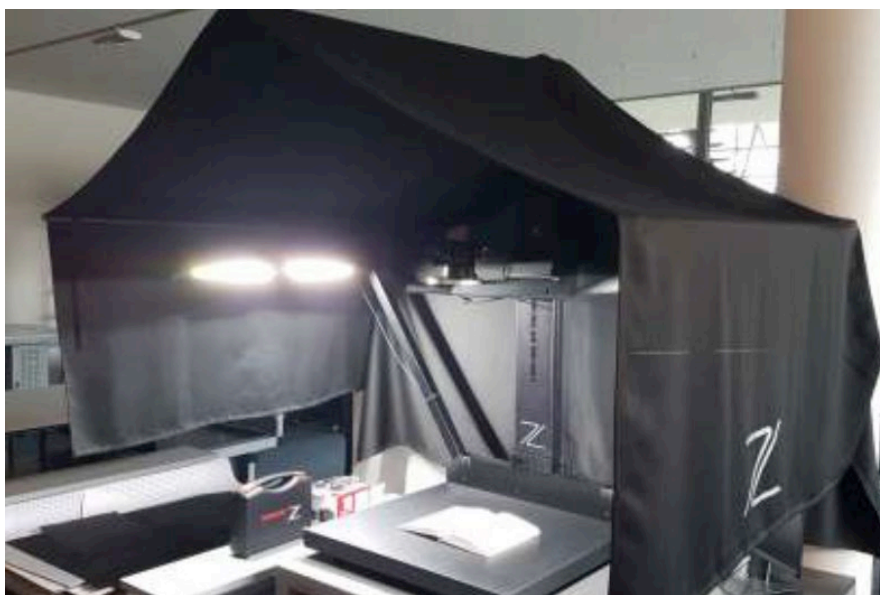
6. Dissemination, communication, marketing and sustainability activities

7. Maintenance, future updates and upgrades (hardware and software)

Below some examples of the art work captured through Zeutschel's technology:



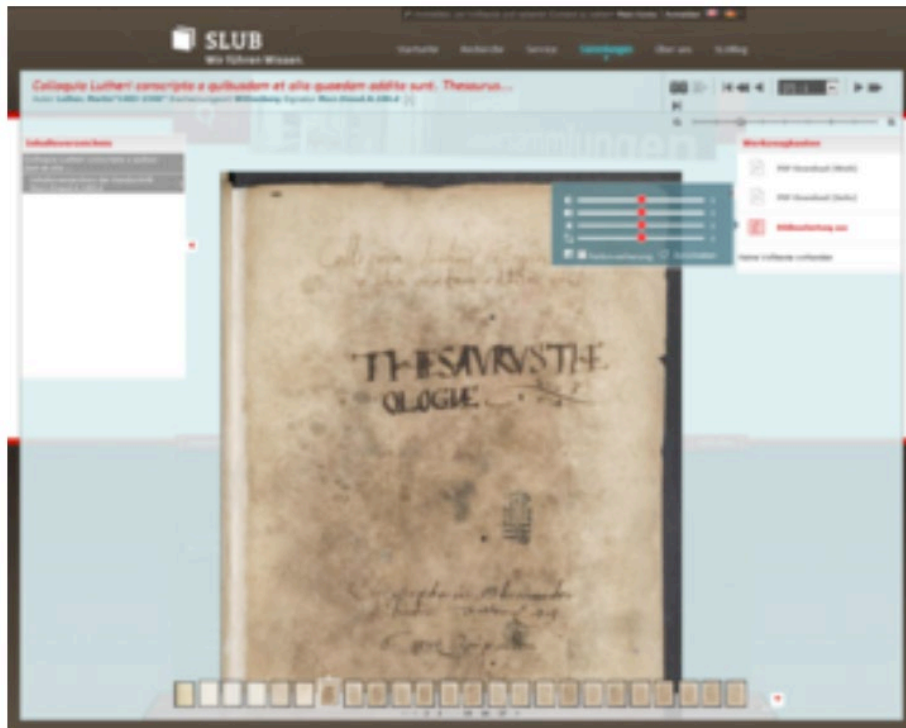
Example of digitization result and color calibration table



Example of digitization set-up (Zeutschel OS A1)



Example of digitization of a fragile historical book



Example of Kitodo CMS interface

4. CONCLUSION

Artificial intelligence, machine learning and ChatGPT automation are indeed very attractive solutions to reduce the timing in digitization, optimize and resources and speed-up the entire eco-system. However, AI poses certain

risks in terms of ethical and social consequences. Although AI may create new jobs, these may require more technical skills than the jobs AI has replaced and this can be a problem.

Privacy is another matter that concern many people. As AI often involves collecting and processing large amounts of personal information, there is the risk that this data will be used by the wrong people. With generative AI, it is even possible to manipulate images and create fake content, profiles and news. AI can also be used to survey populations and track individuals in public spaces.

In the digitization and cultural heritage sector, AI and machine learning can indeed help archivists to speed-up the cataloguing activity (ie. auto-tagging, metadata enrichment, etc.), help photographers and digitization post processing person ell during the digital photo manipulation, digital restoration and enhancement. However, the visual interpretation of a digital defect on an image, or the understanding of the historical value of the provided content can play an important challenge to automatic operations , including AI and machine learning processing. As consequence, time for human verification and quality check could still be expensive and complex to operate, especially on a large scale of data.

In this respect, Zeutschel is still play with confidence a strong role to focus and promote standards, quality control over the entire digital workflow not only with their Germany made top notch digitization portfolio solution but also with its great free Kitodo CMS software, which provide 3 specific components very handy for the librarians and archivists and that can be translated in many languages including Bulgarian:

- **Commons** is a group of classes that extend the TYPO3 API with Kitodo-specific functions – e.g. processing METS structures and library metadata in various formats or integrating external resources such as OAI interfaces or Solr search engines. The API extensions are available to all extensions throughout TYPO3, i.e. they can be used by both Kitodo.Presentation and other extensions. Consequently, commercial or highly specific functions that are only required at single locations need not become part of the official Kitodo.Presentation release. They are implemented in independent extensions.
- **Modules** is a group of components that is integrated in the TYPO3 backend and used to configure the extension and manage digital objects, clients, and collections. At present, there is a module for client management and a module for manual import of new or updated indexes for existing digital assets. In addition, structure data and metadata handling allows for highly granular configuration and database management. A command line interface that can be used, for example, to automate the import of new digital objects is also available.
- **Plug-ins** are components that generate frontend output and provide a wide range of functions. All of the plug-ins can be configured, in part in great detail, via a GUI and freely placed within websites managed using TYPO3. All frontend outputs use design templates and are thus freely customisable. At the present time, Kitodo.Presentation includes 12 plug-ins.