

“The Norwegian approach - automated quality control of digitization on national level in the Digital Archives. A practical approach”

Ottar A.B. Anderson, Head of Photography and Image Services at Intermunicipal archives of Møre og Romsdal, Ålesund, Norway

Abstract

The Digital Archives¹ went in 2019 from being The National Archives of Norway's own digital platform to become Norway's joint national digital platform for receiving, preserving, and publishing digitized/media-converted historical archives. Regardless if you represent state, municipal, or private actors, small or large, the platform is free of charge and use for the Norwegian archive institutions. The digital platform was first published in 1998, marking 25 years in 2023.

The project describes the journey from the pre-project of the Digital Archives, back in 2016, through the research and the development process together with fellow researchers and computer engineering students at NTNU² Gjøvik, to the current on-going stage in the project with the practical implementing of the methods and tools for automated quality control and metadata management. Thoughts on future work on the national dataset are also outlined.

Motivation

The pre-project of the Digital Archives as the joint national digital platform, “National publication platform of digitized archive documents” (NAPSA³, my translation) was established within the “Complete community archives project” (SAMDOK-project⁴, my translation). Selected members from Norwegian archives, museums and libraries started their work in the beginning of 2016 for their main objective and mandate to outline and present a detailed technical plan of a national digital platform. The platform would serve the needs of the archives sector, through a back-end solution with tools and systems for

handling metadata, production and file management, and a front-end solution with digital presentation of the digitized/media-converted historical archives. The pre-project was supervised by the National Archives of Norway and the working group delivered its final report with its recommendations for the main project within the timeframe of the beginning of 2017.

Ottar A.B. Anderson was a member of the working group of NAPSA and carried out the technical part of the work on digitization and quality control. The mandate was to recommend a minimum standard of quality of the digitization and to identify the need in the sector for best practice in the work of digitization of paper-based archive materials.

In February 2017 the Office of the Auditor General of Norway⁵ published the report “Investigation of the digitization on the cultural heritage in Norway”⁶(my translation). The findings in the report outlined the need for a superior plan of the resources and coordination of interactions within the archives, museums, and library sector on the task of digitization of their collections. As a response to the report, The Norwegian Parliament⁷ carried out the necessary exercises over the public administration, in this matter Ministry of Culture and Equality⁸, resulting in a well-balanced coordination between the National Archives of Norway and the National Library of Norway⁹. The community also gained an awareness of the challenges and an open-minded forward-leaning approach to the matter of quality control.

Challenge

Both reports with its recommendations from the pre-project of the Digital Archives, and the report from the Office of the Auditor General of Norway was published before the international standard ISO/TS 19264-1:2017¹⁰ was published in

¹ <https://www.digitalarkivet.no/en/>

² Norwegian University of Science and Technology

³

<https://www.arkivverket.no/arkivutvikling/utviklingsprosjekter/avsluttede-prosjekter/samdok/samdok-delprosjekt-kommunale-arkiver#!#block-body-3>

⁴

<https://www.arkivverket.no/arkivutvikling/utviklingsprosjekter/avsluttede-prosjekter/samdok>

⁵ <https://www.riksrevisjonen.no/en/about-the-oag/about-us/>

⁶ <https://www.riksrevisjonen.no/rapporter-mappe/no-2016-2017/digitalisering-av-kulturarven/>

⁷ <https://www.stortinget.no/en/In-English/About-the-Storting/>

⁸ <https://www.regjeringen.no/en/dep/kud/id545/>

⁹ <https://www.nb.no/en/digitizing-at-the-national-library/#collaboration-projects>

¹⁰ ISO/TS 19264-1:2017 describes a method for analysing imaging systems quality in the area of cultural heritage imaging. The method described analyses multiple imaging systems quality characteristics from a single image of a specified test target.

April of 2017. The result of this was, in lack of published terminology and methodology - the right tools to implement the workflows and procedures for a coordinated national approach on the task. To recommend a minimum standard without the right metrics and parameters, supported by a published international standard was challenging, furthermore, to imagine this standard to be published within a couple of months up front, from both the reports, was impossible. Without been a part of the ISO working group ISO/TC 42 JWG 26¹¹, responsible of the standard, and at that time, no one in the pre-project of NAPSA was on behalf of Standards Norway (SN)¹², the Norwegian representing part of ISO¹³

The main project of the Digital Archives landed on the decision on **not** setting a minimum standard of quality on digitization for publishing on the platform. Despite the recommendations in the final report from the preliminary project, NAPSA clearly stated there should be considerations made of setting a minimum standard. Even with the lack of needed metrics and parameters at the time of post ISO19264-1:2017.

When the Digital Archives officially changed from being the National Archives of Norway's own platform, to become the official national publication platform for all archive institutions, in September of 2019¹⁴, a dataset of digitized archives was uploaded and published. This without a minimum standard and the responsibility of the decision of quality and quality control to be determined by the individual archive institutions by themselves.

This was problematic because the back-end solution in the Digital Archives only had a visual quality control module implemented at the time of launch in 2019.



Figure 1.] Fjeldsmarken krets, Vestre Toten skolekommune, 1863–1885. Section of archive published on the Digital Archives where visual quality control was to be conducted, but failed in doing so.

<https://urn.digitalarkivet.no/URN:NBN:no-a1450-el19030000000093.jpg>

The example in figure 1 shows published archives that have been through a visual quality control. Though here a human error failed in the visual quality control. Parts of the archive has been lost in the digitization process and is no longer representing the original primary physical archive source. Even with an automated quality control module based on ISO 19264-1:2021 with the use of reference targets¹⁵ such as an object level target or a device level target, combined with a computer based analyzes, the information would still be lost, if the fingers only were to cover the archive source itself, and not the reference targets. This example substantiates the need for human visual quality control in addition to a computer-based analysis in conformity to ISO 19264-1:2021.

Approach

As the main project of the Digital Archives made the decision of not setting a minimum standard, despite the recommendations in the final report from the preliminary project, NAPSA, a practical approach was needed.

By December of 2018, Ottar A.B. Anderson finalized the training needed provided by Standards Norway in the responsibility of becoming a member of an international working group and contribute as an expert committee member of the ISO/TC 42 JWG 26. This meant that on-going and future work on the matter of imaging system capability qualification for archival recording and approval, as the scope of the working group is, would be within reach of knowledge and of vital importance of success of the up-front tasks and future coming in the project.

The idea of establishing an independent side project run by the Intermunicipal archives of Møre og Romsdal¹⁶, in associations with the Norwegian University of Science and Technology, The Norwegian Colour and Visual Computing Laboratory, Colourlab (NTNU)¹⁷ and the Cultural Heritage Photographers Associations of Norway (IFF)¹⁸ was born. This was made possible by the National Archives of Norway offering research and development founding for a project like this, through an allocation of the application named, "Analyze of the requirement of Quality Assurance of Digitization in the Digital Archives" (AKDD).¹⁹ The R&D project was outlined as a pre-project, with the need for a two-year main project established according to the findings, initiating after completion of the pre-project.

¹¹ <https://www.iso.org/committee/48420.html>

¹² <https://www.standard.no/en/toppvalg/about-us/#.ZA5EDuzMjp8>

¹³ International Organization for Standardization

¹⁴ <https://www.arkivverket.no/nyheter/alle-landets-arkiv-et-tasteklikk-unna?q=digitalarkivet%20grande>

¹⁵ reference target - arrangement of test patterns designed to test particular aspects of an imaging system (ISO 19264-1:2021)

¹⁶ The intermunicipal archive of the county of Møre og Romsdal

¹⁷ <https://www.ntnu.edu/colourlab#/view/about>

¹⁸ <http://institusjonsfotografene.blogspot.com/>

¹⁹ <https://www.arkivverket.no/nyheter/utviklingsmidler-2020-stotte-til-innovasjon-og-nytenking>

Process and results

The pre-project AKDD started in January of 2021 and due to the Covid-19²⁰ pandemic it was extended to March of 2022. Then running simultaneous the 3 first months of 2022 together with the main R&D project “Quality Assurance of Digitization in the Digital Archives”²¹. The pre-project had through established its findings the need of a complexed and automated quality management module, proposed executed in association with a group of three computer science students, at NTNU, Faculty of Information Technology and Electrical Engineering, Department of Computer Science (IDI)²² at Gjøvik. Mikael Falkenberg Krog, Jakob Frantzvåg Karlsmoen and Martin Wighus Holtmoen completed the task in May of 2022 defending their bachelor thesis “Automated Quality Assurance of Digitization in the Digital Archives”²³.

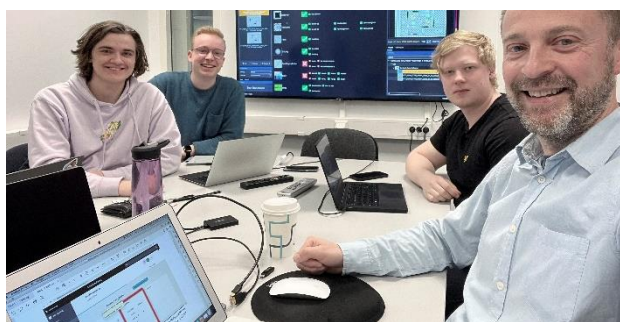


Figure 2.] Photo from one of the meetings at NTNU. From the left: Mikael Falkenberg Krog, Jakob Frantzvåg Karlsmoen, Martin Wighus Holtmoen and Ottar A.B. Anderson

All under the supervision of Professor Rune Hjelsvold and university lecturer Johanna Johansen from NTNU, supported by input from the Colourlab. Gaining the outcome of a quality management application: A backend REST API²⁴ module and a web-based frontend module that serves as a usable example of how the API²⁵ can be integrated into an existing system. The backend module conducts analysis and grades the images in conformity to the ISO 19264-1:2021, validates the images using JHOVE²⁶, and embeds the results as metadata in the files.

The objective of the student projects as a part of the main R&D project was to perform the task of an automated quality control module of digitization in a small-scale project. And due to the fact that the project became a student project, this opened the possibility to even closer corporations with quality analysis software companies. We started out with a wide approach and reached out to several companies, identifying the cons and pros of each specific software and reference target. To get the machine-readable automation needed to perform the huge set of data analyzed, we needed the feature of getting the results in Text file²⁷ or CSV format²⁸. This due to the knowledge of the past events in the Office of the Auditor General of Norway’s report back in February of 2017, we aimed for a future-proof evaluation of the quality analysis. Our strategic choice was to implement the test results of the quality control analysis *into* the metadata of the files itself.

The quality management application is licensed MIT License²⁹, meaning that it potentially can be shared with others and continue to be developed also outside our project.

Conclusions

Acknowledging the content itself, the digitized/media-converted digital files/datasets as the core value is one of our main findings in the project. To make the files/datasets sustainable future systems and potential system migration failures, are also a key admission and may in the end outperform the supreme effort in trying to make the “perfect” system.

Bringing the quality analysis software out to the archive institutions on a national level in the same centralized pipeline as their back-end solutions, where they manage their metadata, production, and file management will hopefully give some benefits. The project’s democratic, equal, and effective way of implementing quality control analysis according to a scientifically recognized methodology, the ISO 19264-1:2021 makes it unique.

We look forward to continuing the work and there are unsolved tasks on identifying and creating an appropriate set of extra metadata schemes for our sector. We need to ensure that the metadata we produce will be useful and that they can be machine readable in the future. Dublin Core³⁰ and the work done by IPTC³¹ together with the ISO system will be the right starting point for standardizing. We believe that integration of metadata on the quality of digitization is the way to go.

And for our project – we will again do the Norwegian approach - the practical approach to make it happen.

²⁰ <https://en.wikipedia.org/wiki/COVID-19>

²¹ <https://www.arkivverket.no/nyheter/her-er-prosjektene-som-far-utviklingsmidler-i-ar>

²² <https://www.ntnu.edu/idi>

²³ <https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3002862>

²⁴

https://en.wikipedia.org/wiki/Representational_state_transfer

²⁵ <https://en.wikipedia.org/wiki/API>

²⁶ <https://openpreservation.org/tools/jhove/>

²⁷ https://en.wikipedia.org/wiki/Text_file

²⁸ https://en.wikipedia.org/wiki/Comma-separated_values

²⁹ https://en.wikipedia.org/wiki/MIT_License

³⁰ <https://www.dublincore.org/specifications/dublin-core/>

³¹ <https://iptc.org/about-iptc/>

References

- [1] https://gitlab.com/bachelor_group9_2022/kvalitetssikring-av-digitisering
- [2] <https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3002862>
- [3] Arkivråd 1/2022 Page 6 – 11
https://www.arkivrad.no/sites/arkivrad/files/arkivrad_1_2022.pdf

Author Biography

Project leader for “Quality Assurance of Digitization in the Digital Archives” supported by the National Archives of Norway.

Head of Photography and Image Services at Intermunicipal archive of Møre og Romsdal. Member of the working group ISO/TC 42 JWG 26 - “Imaging system capability qualification for archival recording and approval”. Photography background from the Royal Norwegian Air Force.

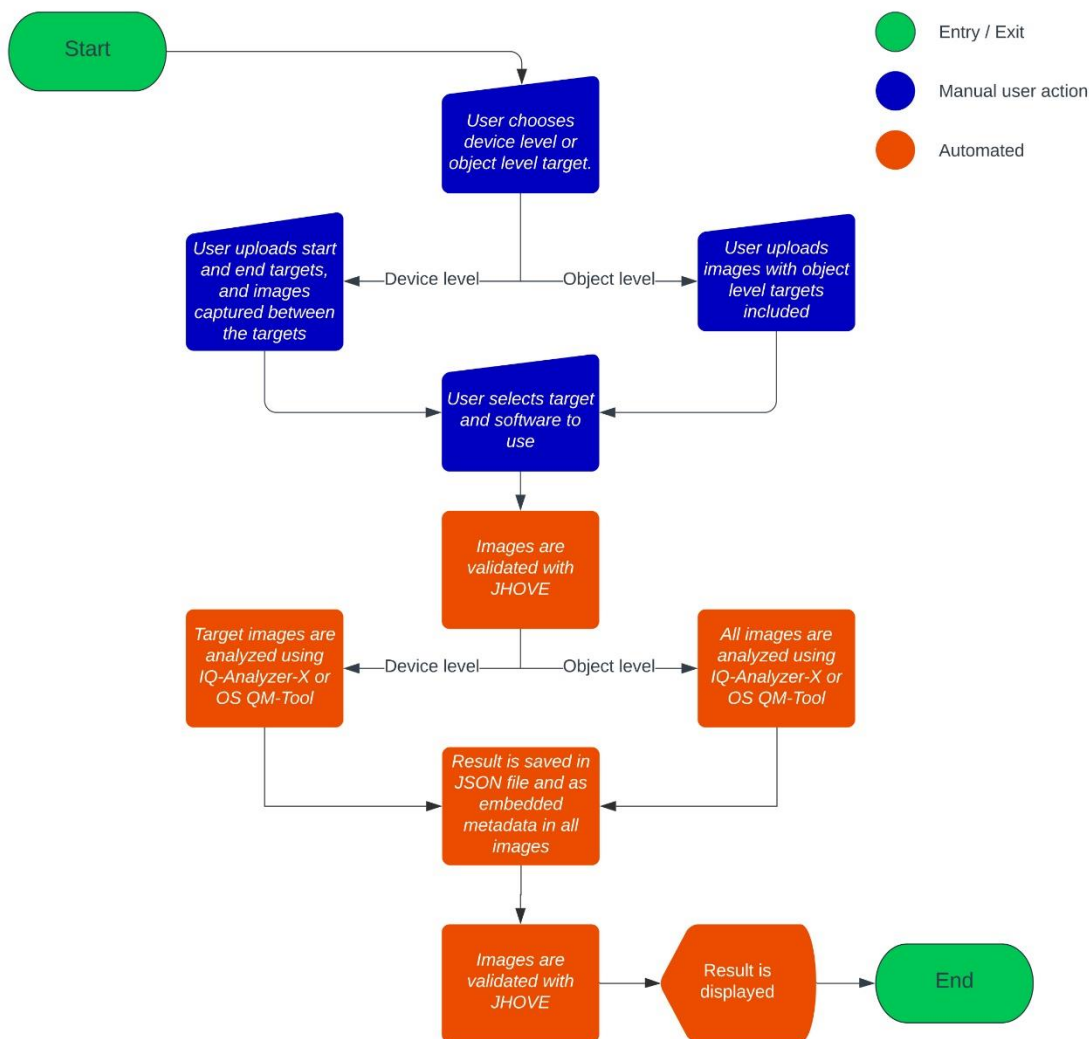


Figure 3.] Flowchart in the quality control module (KAD). Illustration made by Mikael Falkenberg Krog, Jakob Frantzevåg Karlsmoen and Martin Wighus Holtmoen.