Project Management

IMPACT Best Practice Guide

IMPACT project
Niall Anderson, British Library
Jeanna Nikolov-Ramírez Gaviria, Department for
Research & Development at the Austrian National Library
Fedor Bochow, Munich DigitiZation Center
(MDZ) of the Bayerische Staatsbibliothek
Neil Fitzgerald, British Library
Geneviève Cron, BNF - French National Library

Table of Contents

Best Practice Guide to Project Management

Project management is on-going from the inception of the project to its conclusion. The project manager also has a central role in ensuring the project can be sustained by the parent institution, and doesn't become obsolete and out-dated even as it is being run. The best way to minimise such obsolescence is through on-going evaluation of the project delivery methods, and by avoiding unique, proprietary solutions, which may not be available to the institution beyond the life-cycle of the project.

A large number of general guides for managing digitisation projects are available on the web. A substantial list is supplied by Minerva Europe here: http://www.minervaeurope.org/guidelines.htm Retrieved 13.03.2011.

These are some general matters to be considered, particularly at the start of a project:

• Can the project be integrated into the institution's current workflow?

• Can the project be done with existing staff, or must people be hired/trained?

• How can people hours and hardware be combined to provide the most consistent and efficient output?

• Does the institution have enough internal capacity to complete the project, or should some of the work be outsourced?

- What processes can be done automatically, and which need human supervision?

- How to establish quality benchmarks, and embed them into the project workflow

- The requirements of internal and external stakeholders

Costs

Key to deciding how a digital workflow will run is an estimate of the cost of the whole procedure. OCLC provide general guidelines for costing the entire life-cycle of a digitisation project here: http://www.oclc.org/research/activities/past/rlg/digimgtools/RLGWorksheet.pdf Retrieved 13.03.2011.

IMPACT provides a Cost Calculator [https://www.surfgroepen.nl/sites/impactproject/oc/GA%20Annex%201%20Description%20of%20Work/OC2/OC2.2/Cost%20calculator/oc2_cost_calculator_v3.05.xls] for all stages of a digitisation project. Used in addition with the IMPACT Storage Estimator [https://www.surfgroepen.nl/sites/impactproject/oc/GA%20Annex%201%20Description%20of%20Work/OC2/OC2.1/IMPACT_Storage-Estimator_BSB_version3-2.xls], it will provide an institution with a set of benchmark costs for the whole digitisation process over a fixed time, and will aid in deciding which, if any, processes need to be outsourced.

Funding

The scale of a digitisation project is determined substantially by the amount of funding available to it. Here is a list of funding agencies worldwide with an interest in mass digitisation.

Within the UK: http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/potential-sources-of-funding-for-digitisation-projects/ Retrieved 12.02.2010

Within the EU: http://cordis.europa.eu/home_en.html Retrieved 12.02.2010

Workflow Evaluation and Documentation

Project evaluation and documentation starts at the inception of the project and continues throughout its life-cycle. Once a workflow has been proposed for the project, it should be trialled on an agreed batch of source material. This trial will help identify any problems with the proposed workflow, and to refine it for the rest of the project.

Such evaluation, applied throughout a project, naturally has implications for the technical standards set, and the means of quality assuring the outputs. Much of this work can be done automatically by validation software, such as:

- Conformance to metadata schema;

- Presence or absence of auxiliary files;

- Conformance with agreed file formats;

- Conformance with file naming conventions;

- Presence or absence of values and their conformity

Regardless of the scale and purpose of the project, it is always useful to set aside some time for manual sampling of project outputs. Human-eye spot tests should be timetabled into the workflow to assess such matters as:

- Image quality;

- OCR quality;

- Accordance between metadata and the object it describes.

Libraries already engaged in mass digitisation have found it useful to build their own workflow evaluation tools, or to adapt existing software to their own ends. For instance, the Munich DigitiZation Center (MDZ) of the Bayerische Staatsbibliothek uses a program called ZEND[1], while the Digitisation Centre of the University of Goettingen has developed a program called GOOBI[2]. Both of these tools help to monitor the use of resources within a project, to spot bottlenecks in a workflow, and in turn to develop workflows where bottlenecks are less likely to happen.

A number of mass digitisation workflows are described and documented by UKOLN (formerly the United Kingdom Office for Library and Information Networking)[3].

Quality assurance and monitoring

Due to the complexity of both the material being digitised and the production processes being applied it is inevitable that some remedial work will be required in every digitisation project. It is important to regularly analyse and understand why these occur, allowing continuous improvement throughout the project. To ensure that the quality standards set at the outset of the project have been met a Quality Assurance Plan should be produced. The QA Plan should be an integral part of the workflow to provide oversight of the project's outputs. The QA Plan should provide methods by which the outputs of the project can be assessed. This may include research into finding the most appropriate means of digitisation (e.g. from original versus microfilm), parameters for acceptable image quality, and tools for validating metadata.

Some key things to consider:

- What benchmarks are to be used on which outputs?

- What mixture of quantitative and qualitative evaluation is needed?

- Similarly, what level of automatic versus human evaluation is desirable?

- At which points in the workflow will quality assurance be carried out?

- Agree quality benchmarks and milestones with stakeholders and any third-party suppliers to the project;

- What the consequences will be should the established criteria not be met?

Quality assurance for digital images

Benchmarks and standards

At this moment there are no international accepted standards for QA in mass digitisation projects. If digitisation is outsourced, contractors should be asked to provide a QA plan. When production starts, the QA Plan can be adjusted according to real life problems and trends, or updated after running a pilot or formal review point.

Contractors might apply ISO certifications[4] like:

---

[1] *Digitisation Workflow*; 2004; BSB: http://www.digitale-sammlungen.de/index.html?c=digitalisierung-zend&l=en Retrieved 13.03.2011

[2] *Bausteine für die Digitale Bibliothek*; 2008; UGOE: http://gdz.sub.uni-goettingen.de/entwicklung/workflow-goobi/ Retrieved 13.03.2011

[3] *Documenting Digitisation Workflow*; 2004; UKOLN: http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-20/briefing-20-A5.doc Retrieved 13.03.2010

[4] International Organization for Standardization homepage: http://www.iso.org/iso/home.htm Retrieved 15.02.2010

- ISO 90001 V.2000;

- ISO 2895-1: Sampling Procedures for Inspection by Attributes – Part 1: Sampling Schemes Indexed by Acceptance Quality Limit (AQL) for Lot-by-Lot Inspection;

Types of issues

The following may occur:

- Images don't meet the required specifications;

- Incorrect file structure, naming or relationship;

- File corruption or missing derivatives;

- Inconsistencies in the metadata schemes;

- OCR does not meet the required specifications;

- Missing or duplicate pages;

- Pagination errors in the original

Types of checks

The following automatic and/or manual checks can be applied:

- Human eye control of images;

- Batch sampling;

- Feedback by the end user once the files are made available on the internet. For images this might not be an effective method because of the high effort for re-digitisation. For the OCR or the metadata, correction by the end user can however be an option to consider;

- Automatic script-based expected file parameter threshold checking of outputs[5]

- Validation of the well-formedness of files using an on-board tool like JHOVE[6]

Quality Assurance for Software

Many institutions involved in mass digitisation have developed dedicated software that performs automatic checks on expected file parameters thresholds. Automatic QA can be done for the validation of file names and XML schemes. Unfortunately most software is dedicated to specific circumstances within the separate institutions and is not available for other organisations.

Automatic QA can also be done for images, using targets scanned with the same device used for the images. For this end there are a lot of test targets available. The Koninklijke Bibliotheek and the German company Image Engineering have developed the Universal Test Target (UTT). The UTT is a single test target which is designed to provide insight into the complete image quality of all types of high end cameras and scanners following current ISO-standards.[7]

---

[5]JISC has recently established a new project in the UK to address automated quality assurance of this sort. AQuA Project summary; the Joint Intelligence Systems Committee; 2011: http://www.jisc.ac.uk/whatwedo/programmes/inf11/digpres/aqua.aspx Retrieved 13.03.2011

[6] Introduction to JHOVE; President and Fellows of Harvard University Library; 2003-2009: http://hul.harvard.edu/jhove/ Retrieved 13.03.2011

[7] Universal Test Target homepage: http://www.universal-test-target.com/ Retrieved 23.03.2010

Commercial software is becoming available that automates QA by measuring image quality trough the use of test targets. By using these tools an institution will always know if the digitisation was done according to the requirements. This software can help making the digitisation process more efficient and can also improve communication with suppliers. Tools on the market include:

Certifi Pedigree from Certifi Media[8]: http://www.certifi-media.com/

GoldenThread from Image Science Associates[9]: http://www.imagescienceassociates.com/

The Koninklijke Bibliotheek and Image Engineering are also developing software that will be able to analyse the UTT.

For managing logistics commercial software is available. An example is MyBib that can be used for QA issues and as a management tool for logistics.[10]

The Munich DigitiZation Center (MDZ) has developed ZEND publishing software, with a workflow tool that also supports QA activities. ZEND is developed in-house and will be made available as ZENDPRO for others. [11]

QA in mass digitisation projects

Because of the absence of generic software there still tends to be a lot of manual work involved in performing QA in mass digitisation. Unless robust processes are implemented at the start of the project, remedial work may absorb a large amount of the available budget.

For mass digitisation projects a scalable approach, which can deal with very large volumes of data is required. The type of material digitised will also have an impact on the method used, because mass digitisation requires an approach, which adds value to the underlying data in the long term. Trend analysis rather than identification of individual file issues (although this is an important secondary element) should be the main driver.

In many institutions individual digitisation projects have established their own standards and methodology, determined to a large extent by available budget, the type of material being digitised, the historical approach of that institution to digitisation, and its available resources. In general, the level of QA should also have a relationship to the scientific and material value of the sources: partnership working with third parties may provide an opportunity to share the QA workload. A balance between quality and quantity is required.

Serious consideration should be given to the provision of centralised generic automated QA services, which can be used by all projects in a given organisation, and for the identification of remedial work to reduce costs and increase standardisation where feasible.

QA and workflow

While the individual aspects may differ from institution to institution, QA should be integrated into the total workflow of each digitisation project. One of the main differences can be the choice made between outsourcing or digitising in house. When digitisation is outsourced the supplier should perform an extensive quality check on the digital files.

---

[8] Retrieved 12.03.2011
[9] Retrieved 12.03.2011
[10] MyBib eDoc; ImageWare Components GmbH: http://www.imageware.de/de/software/loesungen/MyBib_eDoc/ Retrieved 12.03.2011
[11]Digitisation Workflow; MDZ: http://www.digital-collections.de/index.html?c=digitalisierung-zend&l=en Retrieved 12.02.2011