

Awareness of Quality Assurance Procedures in Digital Preservation

Andrew Williamson

a.williamson@strath.ac.uk

2004

This article appears in Library Review Volume 53, Number 4, 2004

Abstract

Awareness and implementation of appropriate Quality Assurance procedures at each stage in the process of digital preservation is vital to achieving the goals of long-term access and integrity of electronic information, and maximising the return on the high levels of investment being made in digital preservation. This column outlines the four stages of Quality Assurance within the digitisation process suggested in the UK by the JISC QA-Focus, and identifies issues to be considered at each stage.

Introduction: Defining Digital and Preservation Quality Assurance

In recent years, as with many countries worldwide, the UK has witnessed the liberal deployment of public monies to be invested in the digital preservation of a vast range of materials, both through funding by national or international bodies, and direct institutional expenditure. This article considers the Quality Assurance (QA) issues that must be considered at each stage of the digitisation process in order to maximise the return on this investment, and achieve the goals of digital preservation.

A wide range of definitions can be found for the term "Digital Preservation". For example, Chapman (2001) offers the following definition from the Association for Information and Image Management:

The ability to keep digital documents and files available for time periods that can transcend technological advances without concern for alteration or loss of readability.

The NOF-digitise Technical Advisory Service (2003) provide the following definition to grant recipients under the NOF-digitise scheme

The maintenance of digital material over the long-term with a view to ensuring continued accessibility.

and Cornell University Library (2003) offers the following:

The goal of digital preservation is to maintain the ability to display, retrieve, and use digital collections in the face of rapidly changing technological and organizational infrastructures and elements

Fundamental to these definitions are the key goals of long-term access and the integrity of digital objects. Feeney (1999, p.5) argued that an endemic lack of awareness of the need to actively ensure preservation of electronic information created with public money was leading to the loss of this information.

A great deal of money can be wasted if digitisation projects are undertaken without due regard to the long-term preservation of the digital files. It is relatively easy to produce a digital version of a book, manuscript or museum object. Unfortunately it is also easy to do in such a way that either the long-term preservation of the file becomes expensive, because of the way it was created, or with the result that the work will have to be repeated because no plan was in place for archiving the file.

It is important therefore that QA criteria are applied to each stage in the preservation process, in order to ensure that the twin goals of long-term access and integrity are successfully achieved, and to ensure that the return on the investment in digital preservation is maximised.

The remainder of this article considers the 4-stage QA process identified by the JISC-QA Focus project, and provides a brief outline of some of the QA issues that should be considered at each stage of a digital preservation project.

QA in the Digitisation Process

The JISC QA-Focus project educates and advises JISC-funded projects in the UK on the application of Quality Assurance Procedures in their work. QA-Focus (2003a) defines Quality Assurance as being "the process that demonstrates your product is able to satisfy your users" and argues that for quality assurance to be meaningful, appropriate standards must be applied to each stage of the process. The QA Focus Project (2001) Approach proposes 4 stages of Quality Assurance within the Digitisation Process: Strategic QA, Process QA, Sign-Off QA, and Ongoing QA. While it is impossible in this column to provide specific detailed information on the

standards at each stage, it is possible to identify some key principles within the QA-Focus methodology.

1. Use should be made of Open Standards in preference to proprietary standards wherever possible. This will help to maximise the interoperability and platform-independence of resources, facilitate long-term access, and ensure that resources are accessible to users with disabilities.
2. Decisions made and procedures employed at each stage in the process should be fully documented to safeguard against the loss of knowledge, particularly as a consequence of staff turnover. Ensuring the long-term preservation of this knowledge into the long-term future ensures a degree of control over the future of those resources in question.

Stage 1: Strategic Quality Assurance

Strategic decisions underpin the rest of the digitisation process and are therefore the most critical issues to be considered by digitisation projects. The outcomes of this stage are the strategic threads which run through the later stages of the project and mistakes can be difficult, and sometimes impossible, to unpick later. Amongst the issues to be considered at this stage are:

- Identification of materials for digitisation. The nature of the material determines many of the decisions at this stage, including the type of equipment to be used and image capture standards.
- Selection of appropriate file types for digital preservation and delivery.
- Development of guidelines for image capture. This would include issues such as Image Resolution, Bit Depth, and Colour Space.
- Identifying and implementing appropriate metadata standards. These may include:
 1. Cataloguing metadata (e.g. Title, Author, Collection, Subject etc.)
 2. Rights management metadata
 3. Technical capture metadata (e.g. the hardware and software setup used, technical details of the digitised files.)
 4. Process metadata (e.g. details of processing activities carried out on the digitised masters within the workflow process.)
 - Should digitisation be carried out "in-house" or out-sourced to a digitisation agency
 - The equipment for image capture, processing and storage suitable for digitisation of the type of material concerned should be sourced and benchmarked. Storage requirements will be determined by decisions taken on issues such as image capture standards, the

number and type of materials identified for digitisation, and the file types involved.

- A standard scheme for file naming should be devised. This will depend on a number of factors, including the size of the collection. Resources should be consistently named, and the scheme devised should be sufficiently scalable.

Stage 2: Process QA

Quality control methods must be developed within the image production process to facilitate the highest possible quality of capture. Some of these issues are dependent on whether the digitisation is being carried out "in-house" or by an external agency. In the latter case, it is important to establish during the tendering process that potential suppliers satisfactorily address these issues.

- The security and integrity of (potentially valuable) materials during the digitisation process must be safeguarded. If the digitisation is being out-sourced, then procedures for transit and storage off-site must be considered.
- Development of a methodology for capturing and processing images
- Identification of image delivery/ transfer methods (e.g. CD-R, DAT tape, File Transfer Protocol, Local Area Network).
- Ensuring that the equipment used in the scanning and processing of the digital images is properly maintained, serviced, and regularly calibrated.
- Staff training should emphasise awareness of quality issues at all stages of the workflow process.

Stage 3: Sign-off QA

This stage occurs after the capture and initial processing of the digitised object, and requires the operator performing the digitisation to inspect the quality of each image created and "sign-off" each image as meeting the quality standards defined in the earlier stages, before proceeding to the next stage of the process. Similarly, the operator should check metadata created. These checks should be recorded as part of the process metadata.

Stage 4: Ongoing QA

Systems should be established which ensure the accessibility and reliability of the digitised objects in the future. These issues may be especially problematic in situations where project funding has come to an end, and project staff have either left or taken on other responsibilities. For this reason it is important that standards and procedures are fully documented. Some of the issues to be considered are:

Responsibility for ongoing maintenance of the digital collection should be clearly defined.

Storage and backup procedures for digitised objects and associated metadata should be developed and implemented. Similarly, all necessary steps should be taken to ensure the security of the computer systems used for storage and delivery of the materials.

A user interface to the digitised materials - which meets accessibility criteria - should be created and maintained.

As "human error" type mistakes are inevitable during the sign-off stage, systems should be put in place to allow users and project staff to report faults with the digitised materials and ensure that these are rectified.

Conclusion

The notions of accessibility and integrity of digital information are central to the process of digital preservation. While it is now commonplace for large-scale funding bodies such as the New Opportunities Fund (NOF) to put in place a support infrastructure to advise and inform projects on Quality Assurance issues, and for compliance with standards to be a condition of grant payment, there is a need to ensure that the issues are publicised and understood wherever digital preservation activities are undertaken. Although this article considers QA-Focus initiatives which are specific to the UK, the issues raised are universal.

Obviously this column merely provides a stepping stone for those engaged in digital preservation or digitisation initiatives to ensure all corners are covered. I've provided references below which provide further information on some of the QA issues discussed in the article.

References

Chapman, S (2001), "What is digital preservation?", OCLC, Ohio, available at <http://www.oclc.org/education/conferences/presentations/2001/preservation/chapman.htm> (accessed 27 November 2003)

Cornell University Library (2003), "Moving Theory into Practice Digital Imaging Tutorial", Cornell University, Ithaca, NY, available at <http://www.library.cornell.edu/preservation/tutorial/preservation/preservation-01.html> (accessed 27 November 2003)

Feeney, M (1999), "Digital Culture: maximising the nations' investment: a synthesis of JISC/ NPO studies on the preservation of electronic materials", National Preservation Office, London

NOF-digitise Technical Advisory Service (2003), "Programme Manual Section 2 Digital Preservation", UKOLN, Bath, available at <http://www.ukoln.ac.uk/nof/support/manual/digital-preservation/> (accessed 27 November 2003)

QA Focus Project (2001), "Image QA in the Digitisation Workflow", UKOLN, Bath, available at <http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-09/briefing-09-A5.doc> (accessed 27 November 2003)

QA Focus Project (2003a), "Introduction to Quality and its Importance", UKOLN, Bath, available at <http://www.ukoln.ac.uk/qa-focus/presentations/iwm-2003/pres3.ppt> (accessed 27 November 2003)

QA Focus Project (2003b), available at <http://www.ukoln.ac.uk/qa-focus/> (accessed 27 November 2003)

TASI: Technical Advisory Service for Images (2003), available at <http://www.tasi.ac.uk/> (accessed 27 November 2003)