University of
Strathclyde
Glasgow

Strathprints Institutional Repository

http://strathprints.strath.ac.uk/

Kelly, B. and Williamson, A. and Dawson, A. (2003) Deployment of quality assurance procedures for digital library programmes. Proceedings of the IADIS International Conference WWW/Internet 2003. pp. 849-852.

Deployment of quality assurance procedures for digital library programmes

Brian Kelly, Andrew Williamson and Alan Dawson

Abstract

Many digital library programmes have a development philosophy based on use of open standards. In practice, however, projects may not have procedures in place to ensure that project deliverables make use of appropriate open standards. In addition there will be occasions when open standards are not sufficiently mature for deployment in a service environment or use of open standards will require expertise or resources which are not readily available.

The QA Focus project has been funded to support a digital library development programme by advising on QA procedures which help to ensure that project deliverables are interoperable. Although the methodology developed by QA Focus is aimed primarily at one particular programme, the ideas and approaches are being made freely available and deployment of the approaches by others is being encouraged. This short paper provides an outline of the work of the QA Focus project and an analysis of the relevance and feasibility of QA Focus recommendations from two contrasting digital library projects.

1. Background

The JISC (Joint Information Systems Committee) provides funding for a wide range of digital library development projects. In recent years it has funded development of an ambitious strategy originally known as the DNER (Distributed National Electronic Resource) but now known as the Information Environment (IE) [JISC-1]. Projects funded under the IE programme are expected to comply with a set of documented standards and best practices. The *Standards and Guidelines to Build a National Resource* document [JISC-2] requires use of a range of open standards such as XML, HTML, CSS, etc.

The experience of previous programmes has shown that projects will not necessarily follow recommendations. There are a number of reasons for this: there may be a lack of awareness of the standards document; projects may find it difficult to understand the standards which are relevant to their work; there may be a temptation to make use

of proprietary solutions which appear to provide advantages over open standards, and there may be concerns that use of open standards will require resources or expertise which are not readily available.

2. QA Focus

The QA Focus project was funded under the JISC 5/99 programme [JISC-3] to ensure that projects funded under this programme complied with appropriate standards and best practices in order to maximise interoperability and access to resources. QA Focus is addressing areas such as access, digitisation, metadata, software development and service deployment. A description of the QA Focus work has been published elsewhere [KELLY].

The approach taken by QA Focus is *developmental* which seeks to (a) explain the importance of standards and best practices; (b) review the approaches taken by projects in order to profile the community and obtain examples of best practices and areas where improvements may be made; (c) provide documentation, especially in areas where problems have been observed and (d) encourage projects which have implemented best practices to document their approaches and share their experiences within the community.

QA Focus is also developing a self-assessment toolkit which will provide a checklist for projects to validate their own QA procedures. A self-assessment toolkit designed for use when a project Web site is to be 'mothballed', which will form part of the final toolkit, is currently being tested [QA-FOCUS-1].

Although QA Focus is funded to support JISC's 5/99 programme the QA Focus deliverables are freely available on the QA Focus Web site [QA-FOCUS-2]. Related organisations are encouraged to make use of these methodologies, as this will provide valuable feedback, help refine the work of the project team, validate the methodology and help to ensure that deliverables from other programmes will interoperate with the deliverables of JISC 5/99 projects.

We will now describe the experiences of an organisation which is seeking to deploy the QA Focus methodology across a selection of its own projects. The two case studies have been provided by staff in the Centre for Digital Library Research [CDLR] at the University of Strathclyde. This work was initiated following a presentation on QA Focus work given to CDLR staff [QA-FOCUS-3].

3. Case Study 1: Victorian Times

Victorian Times [VICTORIAN-TIMES] is a large digitisation project funded by the New Opportunities Fund [NOF]. The project is digitising a range of textual and pictorial resources relating to social, political and economic conditions in Victorian Britain (1837-1901). These are supplemented with educational resources written by subject specialists. The project is required to be accessible by a variety of browsers, platforms, automated programs and end users.

NOF provide extensive guidance on the use of open standards, supported by online discussion forums, and access to a technical advisory team. They also require quarterly reports from projects documenting their implementation of standards and, when decisions have been taken to set aside standards, to provide strategies for migrating to suitable standards in the future.

This case study provides a brief account of some of the QA issues faced by the project, highlighting instances where it was necessary to compromise on adoption of standards for financial, technical, or service quality reasons.

Based on NOF guidance for **creation standards** the project decided that high-quality digital master images should be created in uncompressed TIFF format at 400 dpi resolution. This would meet preservation requirements and maximise options for creating digital surrogates as new open standards emerged. Later consideration of QA Focus guidelines shows this decision to be in full accord with recommendations.

Three surrogate formats were identified for **delivery formats** of the digitised resources; JPEG image files, plain-text OCR output, and PDF text files. It was also decided that Web content would be delivered in HTML 4, utilising cascading style sheets, and meeting W3C WAI accessibility criteria where possible.

Although PDF is a proprietary format it was judged to be acceptable since free viewers were available and materials would also be offered in other open formats.

It soon became apparent that the quality of OCR output from the digitisation was variable and highly dependent on the quality of the source materials. The option of manual correction of the text proved prohibitively expensive. As the OCR output was being used to support free-text searching, the variation in quality was accepted as inevitable in the short-term. The high-quality digital masters allowed the possibility of repeating the process if there were significant improvements in the technology.

Comparison of project delivery formats with QA Focus recommendations showed mixed results, with areas of full compliance, partial compliance and non-compliance. However, the realistic and flexible nature of the guidelines meant that it was possible to comply with the QA framework even where recommended standards were not being followed, provided suitable procedures were followed and documented.

NOF guidelines on **resource identification** specify that digitised resources should be 'unambiguously identified and uniquely addressable'. This posed difficulties for the Victorian Times project, since its content is delivered by a bespoke content management system (CMS), with pages being dynamically generated based on the profiles of individual users. URIs are therefore lengthy strings of characters which are unique - if largely meaningless - to the user. It is, however, possible for users to uniquely address individual images from the collection, though this removes the images from the context of the Web service.

In this area it was necessary to balance the benefits to users of strict adherence to the identification standards against the richer service quality which implementation of the CMS would deliver. As the project would also be implementing an Open Archives Initiative [OAI] gateway to its digitised resources, it was decided that it would be

appropriate in this case to set aside the standard. Again, the decisions made did not follow QA Focus recommendations, but could still be regarded as following best practice as a cost-benefit analysis had been carried out and informed decisions were made after considering the alternatives available. In addition consideration of the QA Focus recommendations helped to raise awareness of the issues.

4. Case Study 2: Glasgow Digital Library

The Glasgow Digital Library [GDL] has a long-term aim to create a wholly digital resource to support teaching, learning, research and public information at all levels in the city of Glasgow, bringing together material separated by ownership and physical location. Funding was obtained for two years to research the feasibility of a co-operative and distributed approach to developing a regional digital library, but not to provide an ongoing service.

By early 2003 the library had a collection of around 5,000 publicly available digital objects, and is being supplemented by further collections as small amounts of funding are obtained for specific digitisation projects. However, unlike the Victorian Times project, no funding is available for technical support, content management or ongoing maintenance and development.

There is a need to consider the extent to which it is feasible to apply the recommendations and procedures of the QA Focus project to an existing digital library with little time or money available. Many aspects of QA Focus guidance for Web sites and digital libraries have been considered, but particular attention is given to the use of open standards, the migration from HTML to XHTML, compliance with accessibility guidelines, and implementation of the `<LINK>` element to assist navigation, as recommended by QA Focus [QA-FOCUS-4].

In view of the importance of XHTML and the potential of languages such as XSLT for repurposing XML resources, it was felt desirable to migrate the Glasgow Digital Library from HTML to XHTML format. The GDL Web site consists of a large number of static but *automatically generated* web pages and a small number of *manually created* pages. In order to migrate the automatically generated pages, one page was converted manually, so that all the changes were understood. Once this had been validated, the programs and templates used to generate multiple pages were modified to produce the desired results, after which a small random sample was tested to validate the migration. In contrast, for the manually created pages, a batch conversion tool was used to carry out the migration from HTML to XHTML. Both approaches were feasible, but the exercise emphasised the value of automatic generation over manual creation for quality assurance as well as content maintenance.

5. Conclusions

The experiences in addressing QA in the context of real-world issues have helped QA Focus to refine its methodologies. It is clear that the approaches taken by projects to the use of open standards and best practices will be strongly influenced by issues such as resource implications, time scales and technical expertise.

Two digital library projects (one under development, with a CMS being implemented, one largely complete with a large collection of static pages) have attempted to follow QA Focus guidelines retrospectively and to implement appropriate recommendations. This exercise showed that the extent of compliance with guidelines could be categorised into four areas: (1) Areas of full compliance, where the project had already made decisions in accordance with QA guidelines; (2) Areas where compliance could be achieved with relatively little extra work or with minor changes to workflow procedures; (3) Areas where QA guidelines were considered desirable but impracticable or too expensive and (4) Areas where QA guidelines were not considered appropriate for the project.

The conclusion from the project managers involved was that consideration of the QA guidelines improved the value, flexibility and accessibility of the digital library deliverables, provided they were interpreted as guidelines and not rules. Rather than the QA process imposing additional constraints, the exercise validated decisions that had been made to vary from recommended standards, provided the issues had been considered and the decisions documented. What had been seen as a potentially burdensome exercise was regarded in retrospect as beneficial for the user service, for accessibility, interoperability, future flexibility and even for content management. It was felt that there are a number of areas in which simple developments to scripts or use of tools can provide a significant development to interoperability.

The developmental approach taken by QA Focus appears to have been largely validated in recommending that any compromises taken are *documented* and agreed with funding bodies, steering committees, etc. rather than *mandating* strict compliance with open standards. The feedback on real-world deployment issues is being addressed by QA Focus through a number of internal QA Focus documents which will provide examples of documentation which describe compromises which may be necessary [QA -FOCUS-5].

References

[CDLR] **Centre For Digital Library Research**, University of Strathclyde, <http://cdlr.strath.ac.uk/>

[GDL] **Glasgow Digital Library**, University of Strathclyde, <http://gdl.cdlr.strath.ac.uk/>

[JISC-1] **Information Environment Home**, JISC, <http://www.jisc.ac.uk/index.cfm?name=ie_home>

[JISC-2] **Standards and Guidelines to Build a National Resource**, JISC, <http://www.jisc.ac.uk/index.cfm?name=projman_standards>

[JISC-3] **Learning and Teaching (5/99) Programme**, JISC, <http://www.jisc.ac.uk/index.cfm?name=programme_learning_teaching>

[KELLY] **Developing A Quality Culture For Digital Library Programmes**, Brian Kelly, EUNIS 2003, <http://www.ukoln.ac.uk/qa-focus/documents/papers/eunis-2003/>

[NOF] **New Opportunities Fund**, <http://www.nof.org.uk/>

[OAI] **Open Archives Initiative**, <http://www.openarchives.org/>

[QA-FOCUS-1] **Mothballing Self-Assessment QA**, QA Focus, <http://www.ukoln.ac.uk/qa-focus/toolkit/mothballing/>

[QA-FOCUS-2] **QA Focus**, UKOLN, <http://www.ukoln.ac.uk/qa-focus/>

[QA-FOCUS-3] **QA For Digital Library Projects**, QA Focus, <http://www.ukoln.ac.uk/qa-focus/presentations/strathclyde-04-2003/>

[QA-FOCUS-4] **Enhancing Web Site Navigation Using The LINK Element**, QA Focus, <http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-10/>

[QA-FOCUS-5] **QA Focus In-House QA**, QA Focus, <http://www.ukoln.ac.uk/qa-focus/qa/>

[VICTORIAN-TIMES] **Victorian Times**, <http://www.victoriantimes.org/>