

Report: High Volume Digitisation: Issues, Trends & Innovative Robot

Funded by the JISC ITT: Workshops and Seminars - Achievements & Challenges in Digitisation & e-Content Programme.

Workshop Title: High Volume Digitisation: Issues, Trends & Innovative Robot Tech

Venue: NTI Building, Birmingham <http://www.ntibirmingham.co.uk/facilities/>

Date: Wed 23rd September 2009 (all day)

Lead Institution/organisation: **Birmingham City University**

Full contact details for primary contact:

Name: **Beverley Dodd**
Position: **Digital Library Manager**
Email: **beverley.dodd@bcu.ac.uk**
Address: **Library & Learning Resources, City North Campus, Birmingham B42 2TP**
Tel: **0121 - 331 6350**
Fax: **0121 - 331 6510**

1. Executive Summary

There are few practical workshops, events, bodies, single web reference tool or signposting services providing independent up-to-date advice on the latest trends, current issues and innovative solutions specifically on high volume digitisation projects, focusing on bound content e.g. registers, monographs, manuscripts, portfolios, books (including incunabula), journals, magazines and newspapers.

'More of the same...' (Quote from one workshop delegate).

The need for more information is great as the digitisation landscape is constantly changing with many repositories of knowledge and stakeholders looking at mass digitisation whether in house, out sourced or a combination of both. Recent developments in large format (bound and unbound) scanning and fully automated digitisation using innovation robotic arm technology with integrated software solutions now provide greater benefits and more options for both JISC institutions and partners.

The workshop, a key deliverable, contributed to the innovation led JISC Digitisation Strategy (February 2008, sects. 1 & 5) and one of the key JISC 5 year mission statements, *"to provide world class leadership in the innovative use of Information and Communications Technology to support education and research."*

The workshop demystified the term 'high volume, industrial scale and mass digitisation' through a very practical demonstration of the step by step process with digital scanners and integrated software for OCR and metadata creation. , aimed at practitioners and colleagues across the HE and FE community interested in large scale digitisation initiatives (LSDIs),

The workshop was a success, with a tight timescale. Leading speakers and demonstrators from across the UK and Europe contributed to a comprehensive programme of events, open sessions and demonstrations.

Common themes emerged included the importance for preparation and planning, measuring impact and feasibility studies, setting realistic scalable goals and solutions, deciding on in-house or outsourcing, digital access v digital preservation, matching the right bound material format to the right type of scanner, quality assurance methods and devising efficient workflows to ensure maximum economies of scale and cost benefits.

The project web toolkit will be available shortly through www.bcu.ac.uk/dcs. It is intended to act as a roadmap for those considering volume-based digitisation of bound collections and materials in conjunction with existing JISC Services such as JISC Digital Media.

2. Key Issues from the Workshop

No clear definition of what high volume digitisation means. The 1 million + rule (items or pages to digitise) still normally applies. Think about why you want to digitise - for digital preservation or/and access? Preparation and logistical planning is very important. Realistic expectations important from the outset. High Volume projects are not cheap but should be high value. Scanning and automated/robotic technology are advancing rapidly – increasing LSDI viability and allowing greater and more flexible economies of scale.

Match right scanners and technology to the right formats. No one scanner fits all ! Critical in selecting bound materials for digitisation. Throughput rates using robotic and non-robotic arm scanning technology significantly improves with uniform format types batched together. Loss of some quality assurance in large scale projects is inevitable. Need to judge what's acceptable with batch quality assurance standards and protocols Space is important for production/assembly line. Right people are needed, with the right skills and stability to see the LSDI project through.

Metadata – can it really be automated, compromises may be required both for digital preservation and resource discovery. OCR – similar to metadata. Both are key strategic and practical LSDI issues Clarity on copyright is essential prior to start of LSDI. Efficient and speedy workflows are critically important to reduce costs and not overrun schedules. Shift work is employed in larger projects – more cost effective. LSDI projects progress spasmodically according to funding opportunities but are increasingly part of core institutional and cross-group strategic plans. LSDI projects are still largely the domain of larger institutions, national repositories and co-operatives. Measuring the impact of the LSDI project is an important part of future LSDI strategic planning.

Delegates and speakers recognised this was a specialised area of digitisation. Delegates were looking for advice on every aspect of LSDI project management. Delegates and speakers alike recognised there was little impartial and 'free' advice out there on LSDIs, especially on the Internet. Understanding costs and a formula for LSDI projects was difficult to budget for. Generous contingencies in budget planning is also wise.

Some delegates felt they would have liked more in-depth case study information, data and problem solving. More events like this needed as many colleagues, across sectors and professional groupings, felt they did not know where to start.

3. Key Recommendations

- A further series of workshops would provide an opportunity to tailor specific days for specific topics. A brief outline in a single day event was ambitious e.g. covering preparation, planning and costing, robotic arm scanning, OCR and Metadata methods for batch/mass digitisation, Measuring Impact of LSDIs.
- What really came out of the workshop, feedback and discussions is a clear need for further exploration of specific topics and develop the most appropriate itineraries e.g. fewer presenters, more practical demonstrations, more case studies, more problem solving and more open and break out (smaller) group sessions. Specific topics to explore in more depth might include:
 - *What LSDIs are*
 - *Project management, preparation and planning*
 - *Measuring likely impact of digitisation with feasibility studies i.e. why digitise?*
 - *Decisions on digital preservation, archiving and access*
 - *Infrastructure and logistics behind mass digitization*
 - *Importance of setting up efficient scalable workflows and quality assurance methods*
 - *Actual robotic digitisation technologies and real time demonstrations*
 - *Software, metadata, OCR, processing post-production*
- It was also clear from the event there is a need for a co-ordinated, independent and sustainable source of up-to-date information, advice and guidance on bound volume LSDIs that is as far as possible free at the point of access. JISC have started this process for some aspects of bound volume LSDI projects e.g. <http://www.southampton.ac.uk/library/bopcris/documents/Automated-Book-Scanners-Munich-2008.pdf>
- The single 23/9/09 workshop could not achieve this in isolation, even if repeated regularly. The web toolkit, forming part of this JISC funding brief is designed to help address some of these issues in more depth and provide a central news and signposting web service. Pilot is available at www.bcu.ac.uk/lstdi
- There is a commitment by Birmingham City University to up-date and disseminate this resource regularly and explore co-operation with other advisory and training bodies such as JISC Digital Media.

4. Documentation & Multimedia

Workshop presentations, event photographs and follow up information are at http://www.bcu.ac.uk/dcs/workshop_follow-up.htm

Full Report 23/10/09
Beverley Dodd and Beth Delwiche

Birmingham City University