

Large-scale digitization: the £22-million JISC programme and the role of libraries

Based on a paper presented at the 31st UKSG annual conference, Torquay, April 2008

The article explains JISC's ground-breaking programme of digitizing large collections across a variety of formats and subjects. The first phase is described, together with the challenges faced and significant lessons learned. The second phase is also outlined, and the methods for selecting the successful projects explained. These two digitization phases are providing UK higher and further education researchers and students with a very large amount of fascinating online content taken from the collections of some of the UK's major national and university libraries. The article concludes with some thoughts about a further large phase of digitization and a discussion of the issue of sustainability into the future.



JEAN SYKES

Librarian and Director of IT Services
London School of Economics
and Political Science

This article is about the Joint Information Systems Committee (JISC) large-scale digitization programme, which started in 2004 when the Higher Education Funding Council for England (HEFCE) gave JISC a one-off grant of £10 million. JISC handed the responsibility for this fund to its Content Services committee (JCS), of which I was a member at the time, and from there the Digitization Advisory Group (DAG) was set up to manage the programme. I was asked to be chair and I have continued in that role despite having completed my term of office on the committee a couple of years ago.

Back in 2004, suggesting a small number of very large digitization projects was a bold and ambitious thing to do. But it has proved to be a seminal turning point in the UK where mass digitization is concerned. In particular, such a major programme has permitted JISC and the community to discover a host of issues and problems associated with the digitization of a variety of formats and for the most part overcome them. The result is that the state of knowledge and skills associated with large-scale digitization within the higher education community, especially the library community, has grown from a very low base to an enviable level in the UK. The learning curve with regard to digitization issues was initially very

steep, but with the successful bid from JISC in 2007 for funds for a second phase of the programme, worth £12 million, the skills base in the UK is continuing to improve all the time.

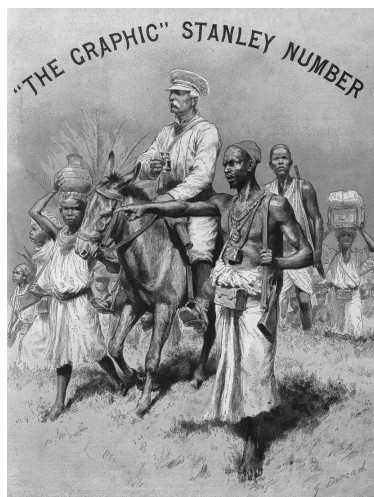
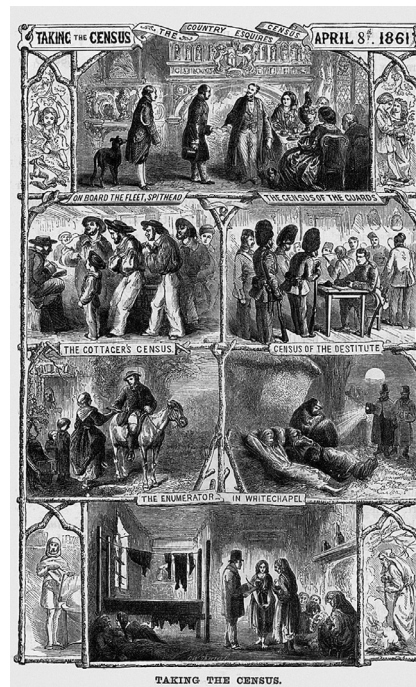
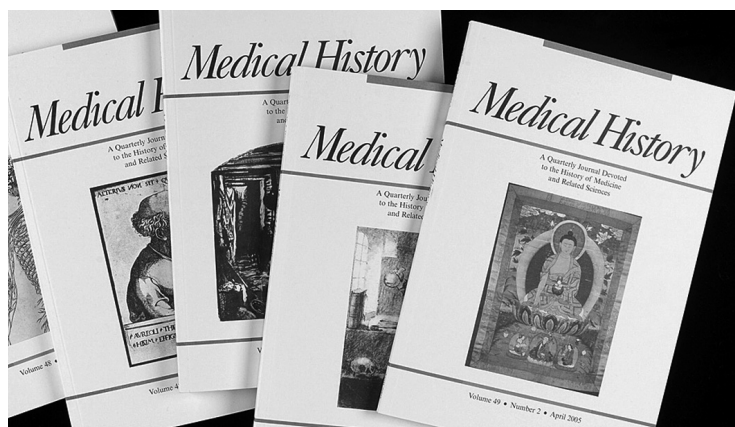
The HEFCE funding for Phase One came with a tight timescale – just three years – given the mountain to be climbed. Nevertheless, the DAG, consisting of a number of members of the JCS plus some other experts, wished to take time to consult with the community before deciding on the collections to be digitized. JISC had been presented with some extant proposals from a number of bodies, and the DAG asked librarians and others in the sector, through an online consultation process, to rank the existing proposals and to add any major collections they thought should also be considered for digitization. After an iterative process the six most highly ranked projects, which included one which had not been in the initial project list, were funded. They are:

18th-century British parliamentary papers (University of Southampton) – 1 million pages

Medical journals back-files 1809 to the present day (Wellcome Institute and the National Library of Medicine in the US, with matching Wellcome funding) – 2 million pages



Phase One: the material to be digitized included a variety of formats



Online historical population reports from 1801 to 1936 (University of Essex) – 200,000 pages

British Library archival sound recordings – 3,000 hours of listening to a wide range of music and spoken word recordings

British newspapers 1800-1900 (British Library) – 2 million pages

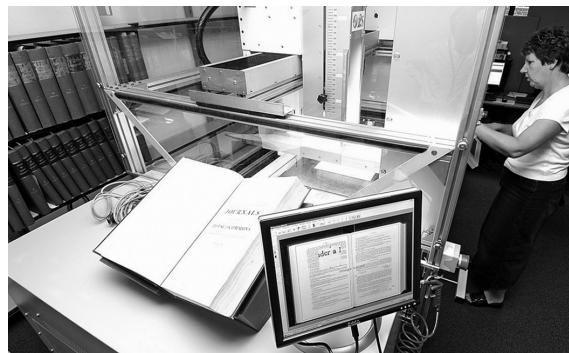
Newsfilm Online (British Universities Film and Video Council) – 60,000 clips and 3,000 hours of broadcasting.

There was clearly a significant variety of formats in the materials to be digitized and this in itself presented major challenges as well as excellent opportunities to learn how best to address the issues and build a body of knowledge to help with future digitization projects. Among the interesting

features in this varied set of projects are the first robotic scanner in the UK, funded by JISC, for the 18th-century British parliamentary papers project and installed in the University of Southampton library; the fact that the British Library set up a panel of users to help them choose the recordings to be digitized from an enormous sound archive; also a panel to choose the British Library newspapers to be digitized; and some very tricky technical and intellectual property rights (IPR) issues to be addressed in a number of projects, notably Newsfilm Online. All of the projects except Newsfilm have already been launched and are fully available via JISC collections, and Newsfilm is due to launch on 2 October 2008. See <<http://www.jisc.ac.uk/digitization>> for a description of the collections.

The Phase One programme was probably the single biggest public sector digitization activity in Europe back in 2004, and with the addition of Phase Two it is quite possible that the UK still leads the field in Europe for mass digitization funded by the public sector.

JISC appointed Evidence Base to evaluate the Phase One programme as it neared completion, and their report can be found on the JISC digitization web page mentioned above by clicking onto *Evaluation* in the left-hand column. The following were the main points identified in the report. Issues to do with standards (technical, metadata, and presentational) had to be grappled with by most of the projects early on in the process, and because of the variety of formats, the standards also varied considerably, meaning that one project could not necessarily learn much from another. Yet on the plus side some ground was covered once and for all, allowing for re-use and adaptation by future digitization projects across a wide range of materials. Some unique contractual agreements and contracts had to be reached with rights holders, sometimes with commercial partners. The most obvious one was Newsfilm Online, whose major partner is ITN. What resulted from the programme as a whole was an improved framework for IPR across a number of formats, providing templates for the future. Exemplars for increased interoperability also emerged, including open source applications. Most of the project work was straying into completely unknown territory and yet the evaluation report shows that considerable achievements were realized and most of the problems, particularly acute in the multimedia projects, were overcome. Out of Phase One also came an



The robotic scanner

embryonic overarching JISC digitization strategy, which is still being refined as Phase Two reaches its midway point.

Evidence Base also identified lessons learned during Phase One, which were extremely important for the second phase programme as they allowed a streamlining of the JISC bidding guidelines with a much tighter set of requirements for bidders to fulfil. This would ensure that the selected Phase Two projects could hit the ground running with some of the lessons learned during Phase One already resolved, or at least better understood, before the bids were submitted. The lessons learned during Phase One were well documented by the JISC programme manager as the work progressed, and then used in the preparations for Phase Two. They encompassed all aspects of the digitization process, including:

- the need for user consultation before and throughout the project
- procurement lessons
- the importance of standards of all kinds, particularly metadata
- the need for extremely tight management both at the individual project level and the JISC programme level
- the absolute requirement to address IPR issues meticulously and to persist in negotiations
- inbuilt quality assurance procedures
- the need to maintain a risk register and to index risks and lessons learned
- the importance of user-friendly interfaces to match the type of material as well as the target audience
- attention to accessibility issues for the visually impaired
- last but not least, the requirement to plan the promotion and dissemination of the collection when launched.

A second level of lessons learned, which for the most part were not able to be successfully addressed in Phase One projects because in hindsight we can see that they ought to have been built in from the start, included: the need to address metadata *from the outset*; the fact that evaluation needs to be inbuilt and take place *throughout the project*; the desirability of impact assessment, requiring projects to install *from the very beginning* the necessary licences and arrangements, for example for web log analysis; and the fact that an expert user panel, to be in place from the start, is best for selection of content from a large collection which cannot be covered in its entirety.

Despite some delays and missed deadlines and some changes in emphasis along the way due to unforeseen circumstances, Phase One was already proving its worth before the projects were completed, and JISC was receiving credit for having had the vision to restrict the programme to a small number of very large projects, in order to have maximum impact. This approach clearly paid off when JISC was awarded a further £12 million for Phase Two in 2007. This time JISC launched a conventional programme call for bids. But unusually for JISC programmes, the bids were both marked by experts (normal JISC procedure) and ranked by the community via another online consultation exercise. This combined method of selection may be unique, and out of 48 bids submitted, a long list of 24 was drawn up, with the first 16 of them being chosen for funding. The quality of the bids was extremely high, reflecting the fast progress made in the awareness of, and expertise in, digitization issues which had taken place in only a few years since 2004. The 16 Phase Two projects, which are due to finish in 2009, are as follows:

Radio news archive 1973–mid 1990s (Bournemouth University)

British cartoon archive (University of Kent) – 15,000 Giles cartoons cross-searchable with 123,000 other cartoons already digitized

Pre-Raphaelite resource site (Birmingham Museums and Art Gallery) – 3,000 digital images

Electronic ephemera from the John Johnson Collection (University of Oxford) – 65,000 items and 150,000 images

East London theatre archive (University of East London) – 15,000 digital objects

Modern Welsh journals 20th and 21st century (National Library of Wales)

British Library archival sound recording: phase two

British governance in the 20th century (The National Archives) – 500,000 pages

Core e-resources on Ireland (Queen's University Belfast)

19th-century pamphlets online (University of Southampton and six other university libraries) – 30,000 pamphlets

British Newspapers 1620–1900 (British Library) – 1.1 million pages

UK theses (British Library) – 5,000 theses

First World War poetry – five poets (University of Oxford)

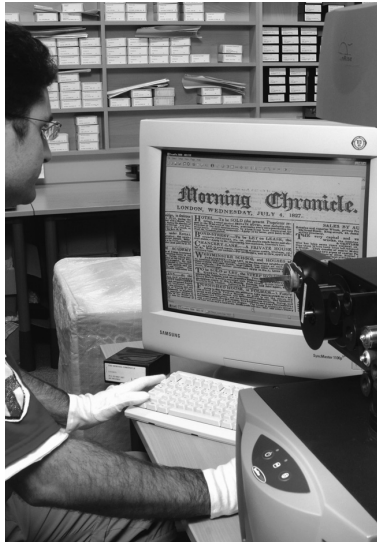
Voices: moving images in the public sphere (British Film Institute)

Historic boundaries of Britain (University of Portsmouth)

Historic polar images (Scott-Polar Research Institute, Cambridge).

If anything, the Phase Two projects offer an even wider variety of formats and materials to be digitized, adding a rich set of new subjects while also awarding significant second tranches of funding to the British Library for sound archives and British newspapers. This will ensure the availability online of major collections from the nation's richest treasure house, providing for example a total of 4 million pages from newspapers. Between them, the two phases of digitization, covering 22 projects and costing £22 million, will have achieved a combination of a few very large collections, together with other collections still of considerable size and offering a critical mass of content across an imaginative range of subjects and formats, much of it unique, representing our UK heritage.

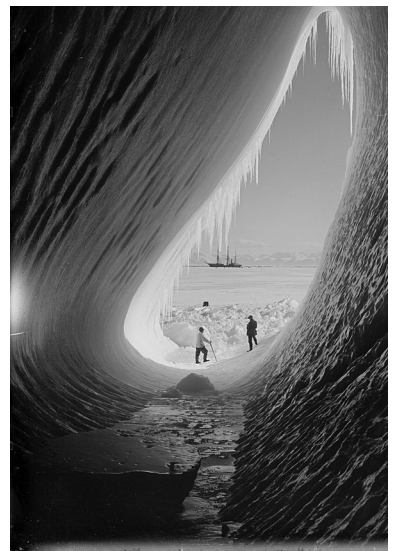
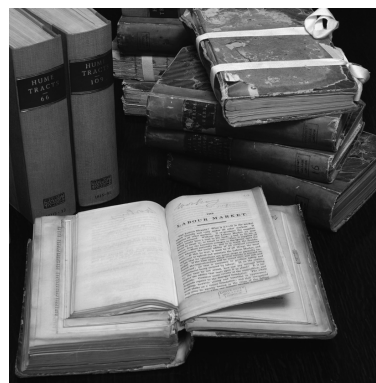
Throughout Phase Two so far, the JISC and the DAG have been discussing key issues for the future and have been trying to keep half an eye on the possibility of a further phase of major digitization funding. Certainly, the group wants to be ready with a plan should such an opportunity arise, and an underspend from Phase One will



If anything, the Phase Two projects offer an even wider variety



Anthem to Dead Youth -
 ...ing bells for you who die
 ...ly the monstrous anger of
 ...stuttering rifles' rattled
 ...an patter out your has
 ...nts for you, nor balms,
 ...for any voice of mourning
 ...will demented choirs of
 ...And bugles calling for y



**VINOLIA
 WAR FUND.**

EVERY TABLET OF
Vinolia Soap
 you buy from November 6th to December 5th
 means a **HALFPENNY** for the
"TRANSVAAL WAR FUND."

Every Chemist, Store, and Grocer is an
 Agent for Vinolia Soap.



**CYLCHGRAWN-CYMDDEITHAS
 ALAWON-GWERIN-CYMRV**

"Moes erddygian a chanu
 Dwg i'n gerdd dig awengau,
 Traw'r dolydd taro'r delyn,
 Oni bo'r idd yn y bryn."

**JOURNAL OF THE
 WELSH FOLK SONG SOCIETY
 VOL. I • 1909 • PART I**

allow a number of useful studies and tools to be undertaken. [*Post-conference note: unexpectedly, the sum of £2 million was made available in early summer 2008 and JISC put out a call for bids for much smaller digitization projects, to be completed by September 2009, and therefore to be more or less shelf-ready for a standing start. Marking of the bids is taking place during the summer and the results are likely to be announced in September. This is not the major Phase Three which the DAG would have liked to see, and indeed hoped would happen in 2009/10. It is impossible at this time to tell whether a further round of large-scale digitization projects will become a reality in the next few years. The remaining part of the presentation made at the UKSG conference in April addresses some of the fundamental issues the DAG has identified for a large-scale Phase Three. Not all of them will apply to the small project programme now announced but they are still important in the wider context of any further major digitization funding phase.*]

Among the ideas which the DAG has approved for funding from the underspend are:

- an updating of JISC's standards catalogue to incorporate digitization projects
- attempting a post-hoc usage/impact assessment of Phase One projects (even though statistics of use were not built in from the start)
- a broader impact study focusing on some quantitative and qualitative indicators
- a gap analysis to find out what the community needs
- an analysis of the availability of relevant large or significant collections already in the community that could be digitized
- the production of guidelines for securing IPR rights for moving images and sound
- the creation of a UK Forum for Digitization
- investigating thematic portals to make resources more comparable and usable, possibly extended to cover JISC Collections.

The last idea is worth pausing over. We could imagine pooling access to digital resources thematically from both the JISC large-scale digitization collections and the growing portfolio of commercial materials in JISC Collections, to which libraries subscribe at tightly negotiated prices. The result would be portals leading to a rich mix of content from the public sector community, especially higher education library collections, and

commercial content. To take the 18th century as one possible theme, a fantastic resource could open up including the wonderful ECCO (Eighteenth Century Content Online) from JISC Collections, plus the 18th-century parliamentary papers from Phase One and the British Library's 18th-century newspapers and parts of the John Johnson ephemera collection from Phase Two of the JISC digitization programme. Even better would be the ability to cross-search these resources, if feasible, offering a uniquely stimulating service for researchers.

Finally, a look at the wider landscape and the implications of the JISC digitization programme for libraries. One objective is that Phase One and Two collections should be free at the point of use to UK higher and further education users and some will also be free at the point of use to schools and public libraries. Access will continue to be password-protected for the eligible communities, using Athens and/or the new JISC Federated Access Management scheme which is to replace Athens. A few of the Phase Two collections aim to be on completely open access. The thematic portals will be a good step forward for enhancing the user experience though they may cause access challenges, especially if commercial content from JISC Collections is included and cross-searchable, because that will be a mix of open access, free at the point of use access, and commercial content available only to libraries which subscribe via the various JISC licences. Looking further ahead, a future stage could see linking up with some of the European and US heritage digitization projects, and this would present even more challenges, especially for IPR.

Perhaps the main issue for libraries and the community at large is going to be sustainability into the future. The free at the point of use and open access objectives of the Phase One and Two projects are currently time-limited. No one yet knows how costly it is going to be to keep the collections going, to develop them technically, to improve the interfaces, and in some cases to add more content. Even those collections which are finite, that is they have been digitized in their entirety, will want to keep up with future technological and presentational developments. Some collections, and possibly the whole programme taken as an entity, may be able to raise income from overseas and commercial subscribers which might offset at least some of the sustainability costs. But it is my view that if we want to continue to attract funding from JISC and other bodies to

digitize more materials, we cannot also expect them to carry all the sustainability costs as well. It is my prediction that libraries are going to have to be prepared to pay for licensed access in due course, albeit at subsidized rates.

This programme has placed JISC among the very top players in the global digitization league. It is already well known and admired across the world, certainly in Europe and also in the US and Australia. Higher and further education communities are going to benefit from fantastic online resources across a wide range of subjects, gaining access to some of the richest content held in the UK's great national and university libraries. The collections chosen for digitization are for the most part those which would have found difficulty attracting commercial sponsorship, and the JISC funding has been able to unlock them and offer them widely online. Libraries must be ready to promote these valuable resources vigorously as they become available. Five collections from Phase One are already available and librarians should be making them accessible now to their users via JISC Collections and alerting their staff and students to their existence. The final Phase One project,

Newsfilm Online, which ranked number one in the first online user consultation exercise back in 2004 and will be of significant interest to the widest possible audience covering as it does all kinds of disciplines, was due to come on stream in October 2008. The sixteen Phase Two collections will become available in the second half of 2009 and librarians must look out for them, publicize them and arrange access to them as they appear. All the parties involved in this major programme can be very proud of its achievements so far, and can look forward to more rich content becoming available in 2009 and beyond.

Article © Jean Sykes

■ Jean Sykes

Librarian and Director of IT Services

London School of Economics and Political Science

10 Portugal Street

London WC2A 2HD, UK

Tel: +44 (0)20 7955 7218

E-mail j.sykes@lse.ac.uk

To view the original copy of this article, published in *Serials*, the journal of the UKSG, click here:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=21&issue=3&spage=167>

The DOI for this article is 10.1629/21167. Click here to access via DOI:

<http://dx.doi.org/10.1629/21167>

For a link to the table of contents for the issue of *Serials* in which this article first appeared, click here:

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=21&issue=3>