



Project Document Cover Sheet

Project Information			
Project Acronym			
Project Title	British Newspapers 1620 - 1900		
Start Date	July 2007	End Date	31 March 2009
Lead Institution	The British Library		
Project Director	Patrick Fleming		
Project Manager & contact details	Jane Shaw Colindale Avenue London NW9 5HE Tel: 0207 412 7368		
Partner Institutions	N/A		
Project Web URL	www.bl.uk/collections/newspapers.html		
Programme Name (and number)	JISC Digitisation Programme 2		
Programme Manager	Alastair Dunning		

Document Name			
Document Title	<i>Project Plan</i>		
Reporting Period	N/A		
Author(s) & project role	Jane Shaw, Patrick Fleming.		
Date	29 October 2008	Filename	JISCNewspapers2projectplan6.0final
URL	<i>if document is posted on project web site</i>		
Access	<input type="checkbox"/> Project and JISC internal		<input type="checkbox"/> General dissemination

Document History		
Version	Date	Comments
5.0	26/02/2008	Final
6.0	29/10/08	Updated new draft
6.1	4/11/08	Final draft for approval
6.2	7/11/08	Final approved



JISC Project Plan

Overview of Project

The objective for this project is to digitise some 1 million pages from regional and local newspapers from the 19th century. All of these are intended to be uploaded to the website currently hosted by Gale Cengage in partnership with the British Library by second quarter of 2009. 22 titles covering each region of the UK are in production with a further 70,000 pages of JISC 1 titles that started in the eighteenth century included.

1. Background

The British Library's goals for this project are to deepen and widen the range of digitised content from two earlier projects: the first is British Newspapers 1800-1900 (JISC I) and the second is the Burney collection of British eighteenth-century newspapers, supported by funding from the British Library and the USA National Science Foundation. These total some 2.9 million pages and are broadly representative of British newspapers between 1620 and 1900 and have a primary focus on long runs of regional titles and national newspapers.

By bringing them together to create a single, coherent and enriched resource we plan to digitise a further 1 million pages comprising mostly 19th century regional titles. Titles already included in JISC 1 project, where the publication preceded 1800 and are a good fit for nineteenth century runs already online, are now in production e.g. *Jackson's Oxford Journal*, the *Ipswich Journal* and the *Aberdeen Journal*. Further eighteenth century titles will be selected where the BL has good runs and where a good amount of the run is on microfilm already. Together with the outputs delivered by JISC I and the Burney project, this will create:

- Access to a virtual library containing some 4 million pages of important regional and local newspapers
- The unprecedented ability to search across different newspapers
- Access to a broad range of valuable learning materials; many of them unique and hitherto available only in London-based reading rooms
- A rich resource which opens up new opportunities for learning, teaching and research in HE & FE, from postgraduate researchers to degree students, school pupils to pensioners, lifelong learners and families researching their history.¹

The British Library's approach to JISC II is based upon the experience and expertise we have gained from the JISC I project together with lessons learnt from the Burney project. A solid infrastructure supported by experienced personnel and proven workflows and procedures has been established to deliver JISC II on time and to budget.

Formalised digitisation standards are in place for the project delivery, a contract has been signed with our web hosting partner Cengage Gale JISCI was launched last autumn to the HE sector and the existing website will be upgraded this year in readiness for rollout to a public audience. The upgrade will facilitate content from JISC II in the second quarter of 2009.

¹ BL proposal to the JISC for funding dated 16 August 2006, p2.

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

Whilst building on the experience of JISC I, this project has simplified, streamlined, enhanced and part automated to achieve the best quality source for digitisation with OCR being provided to the greatest accuracy possible.

The greater part of the project will follow the production pattern of JISC1 with scans taken from microfilm. The JISC II production plan was revised to include 100,000 pages scanned from original newspapers at the Library's in-house scanning unit at Boston Spa, all of the one title *The Standard*.

Output in both cases will be sent to a third party supplier, selected after an EU procurement tender process, for digitised content creation: scanning, article segmentation and OCR. The selected supplier, CCS [Content Conversion Specialists] will export the texts to xml files for all of these processes, and align the xml with each digital image.

This updated project plan covers the changes made to content selection, preparation and microfilming, extraction/conversion and production of digital images for hosting on the web and for digital ingest by the Library. It documents improvements and updates to the project scope and approach first presented in the 2006 bid for funding which was successfully achieved in July 2007.

With the agreement of JISC the project completion date has been extended to 31 March 2009. This date will not be subject to extension.

2. Aims and Objectives

The Library regards JISC1 as an exemplar programme that has created a significant and important treasure trove of historic newspaper content in digital format. The overall aim of JISC II is to make available on the web a digitised collection of British newspapers that spans all regions of the British Isles and is representative of all kinds of newspaper between 1620 and 1900.

The specific objective for this project is to digitise some 1 million pages from regional and local newspapers from the 19th century and include eighteenth century runs of JISC 1 titles to complete the full run. These pages will be uploaded to the online database currently established by the Library in partnership with Gale Cengage. The overall target was originally set at 1.1 million pages. This has been reduced to 1 million pages to be in line with the agreed BAFO costs and approved budget. The reason for the overestimation was due to the difficulty in assessing the number of pages in any given volume or title run. The catalogue does not hold this information (it is collected for the first time in our metadata capture.)

The British Library will, at its own cost, provide funding to add further historic pages from the 18th century to this resource to enhance the value of the project.

3. Overall Approach

Strategy

The digitisation of newspapers sits at the very top of the Library's wider newspaper strategy.

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

The Library possesses a newspaper collection that dates back to the 16th century and includes UK and Irish newspapers together with more than 250 international titles at the Newspaper Library in Colindale, North London.

In November 2007 the British Library Board approved a Full Business Case for the long term storage, preservation and access to this collection.

This business case has been approved by the Department for Culture Media and Sport and is currently with HM Treasury for consideration.

Central to this is the move of the content of the Newspaper Library and an associated Legal Deposit building at Colindale to a new newspaper storage building at Boston Spa by end 2012 and the creation of access to newspaper via microfilm and digital surrogates at St Pancras.

To achieve this vision the Library is committed to an intensive programme of development that will lead to increases in both born digital and digitised newspapers.

Core to the newspaper strategy is the relationship with newspaper publishers. The British Library will enter into voluntary arrangements with newspaper publishers to enable the ingest of born digital newspapers. Trials with major newspaper publishers have begun and at present all editions from a major publisher are being ingested into the Library's Digital Library System.

The Library is also working with publishers to assist them in digitising their back catalogues. This will enable them and the Library to create, acquire and provide access to a vast resource of digitised newspaper content.

The experience and knowledge brought about through the JISC and British Library partnership has played a pivotal role in enabling this to happen.

The Library also continues to purchase digital datasets of newspaper content from publishers to enhance our digital content.

Initiatives to provide the best possible long term digital preservation and storage methods are ongoing together with resource discovery projects that will support the current and future needs of researchers accessing the Library's physical and digital collections.

Digitisation will:

- Widen access geographically
- Widen access socially, enabling and stimulating user demand
- Enhance access for all users through improved resource discover
- Enhance the preservation of the nation's historic collections through the creation of digital object surrogates and archive quality microfilms for preservation (as well as the original newspapers being stabilised)

Methodology

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

As with JISC I, newspaper publishers have been kept fully informed regarding the digitisation of potential titles.

The publishers or owners of all titles have been contacted individually by letter.

The Library outsourced a more detailed level of research for all of the newspaper titles within this project and in addition reviewed all of the newspapers that have been digitised as part of the first project, with a view to establishing copyright as clearly as is possible for each newspaper. A sum of money has been ring fenced in the JISC II budget for future work into IPR clearance.

Scope and Exclusions

The Library proposed digitising 1 million pages of 18th and 19th century newspapers in its original proposal and there were three main categories – 75% of regional and local titles mainly from the 19th century, filling gaps in the UK wide coverage from JISC I. As mentioned in previous project reports, JISC 1 titles provide a framework or skeleton of long runs of important nationals and dailies with a few regional titles that act as a virtual index to newspaper history. JISC II titles build on that framework by adding predominately local and regional news to this existing resource.

By their nature these papers are very different in content, layout and pagination to those in JISC 1 and for digitisation purposes, present further challenges. They vary more in periodicity, may start out as a weekly, become a daily and then revert to a weekly and there are very few illustrations. Most of the illustrations seen so far are limited, chess reports, weather reports and graphs or small images contained in advertisements.

For reasons of condition/preservation, the group of London titles proposed in the bid to be digitised from the Burney collection has been excluded from the scope and for the moment no specialist newspapers on particular themes have been selected, nor is there an intention to build on the Chartists sub cluster from JISC 1. The main workflow is as follows:

- Selection of content
- Preparation and microfilming
- In-house scanning
- Digitisation by third party supplier
- Article zoning
- OCR
- Web hosting

Selection of content

The primary focus of the project is on regional and local titles, predominately from the 19th century. This selection will balance the majority of titles from the two previous collections which were mostly London based.

The preliminary list of titles attached under Appendix 1 of the 2006 proposal was drawn up after wide consultation with the HE & FE communities, with first choices underlined. Those underlined titles were endorsed as the final selection for digitisation in discussion with the Chair of the user panel Dr Peter Mandler of Cambridge University.

The criteria for selection did not include a detailed survey of the condition and quality of the existing microfilm and/or original newspapers but on:

- The intrinsic research value of each individual newspaper title
- Avoiding any risk, through consultation via NEWSPLAN, of duplicating the digitisation efforts of other organisations
- The completeness of the runs of issues

Omissions

Missing issues or pages

The project is based on British Library holdings only. We do not intend on filling gaps in runs from other collections during the main production phase, but have considered filling gaps post production as some pages were not on the shelf or misplaced during the preparation stage. We will produce a list of missing material as part of project closure.

Irish titles

The question mark around the Irish titles being fit for purpose was resolved.² We have dropped a work package The Irish Titles / Work Package Four and Four A, as in January 2008 we added an extra 2 titles to meet the earlier page number target. These were *The Standard* and *The Morning Post*. There were several issues surrounding the retrieval of the volumes in work package four, the Irish titles, which were a cause for concern. Although we were provided with an alternate selection of titles the nature of the collection and its storage quirks mean that the Irish Titles are generally quite uniform in the problems (for JISC) which might effect our ability to retrieve, film and replace these volumes.

The volumes currently stored at 120 are in the process of being boxed and re-shelved by the ASP preparation team in expectation of the buildings closure and its holdings eventual move to the ASP facility in Boston Spa. This makes locating volumes within 120 difficult.

The volumes are stacked horizontally one on top of the other and by size rather than in yearly sequences, a decision made due to space constraints, this slows retrieval significantly because every volume above that required needs to be removed from the shelf and replaced on retrieval and replacement.

Most of the storage is above head height and so would require a trained member of staff for retrieval as specialist equipment is required for accessing these shelves. The ASP preparation team is not available to retrieve volumes for JISC. The Legal Deposit Newspaper Team is available but time constraints and a heavy work load (in preparation for the buildings closure) mean that all retrieval work for JISC would have to be carried out after standard working hours. This would involve additional (overtime) payment for two or three members of staff.

The possibility of scanning directly from existing microfilm duplicate negatives has been explored and while we feel it would be possible to achieve a good quality image from existing negatives this would mean no metadata would be collected. We asked CCS for an estimate for the scanning costs, unfortunately they were prohibitively high and we have made a decision to drop the Irish titles.

Titles culled due to an over estimation of pagination and budget constraints in September 2008 were:

Cambridge Chronicle and Journal
Northampton Mercury

² If it is decided that these titles are too fragile to handle or would be too slow to prepare, we may substitute a sub cluster of titles on a specialist theme to make up our numbers.

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

Norfolk Chronicle or The Norwich Gazette
Chronique (Chroniques) de Jersey
Inverness Courier and General Advertiser For the Counties of Inverness
Cambrian [Swansea]

Each region of the UK is represented by at least one complete run of a title still.

Preparation and microfilming

The preparation stage involves engaging with the source material and starts with pre-generation of metadata from the newspaper catalogue. Elements such as newspaper title, type, the date ranges and availability of each title are structured into a metadata capture sheet/manifest like the project DTD schema³ and completed by the prep team when they review and/or repair each page. We decided not to capture volume number as this is so intermittently printed on newspapers that it could increase QA effort if it was only occasionally present. This new metadata is added to the manifest at issue level to enable rapid ingest into xml by CCS.

The manifest sheets for each title are audited for accuracy and consistency before shipment with the corresponding microfilm reels to CCS.

As in JISC I it was decided to refile up to 90-% of the originals as the quality of the existing microfilms would not be good enough to be converted via the OCR process to enable free text searching.

Article Zoning

CAS's digitisation process closely mirrors our own and consists of the following steps:

- Selection: Books, newspapers, Journals etc. From original, microfilm and digital image formats.
- Capture: scans from originals and from microfilms; import from digital image files, quality control.
- Convert: image pre-processing, layout analysis, OCR (text), ISR (Structure recognition), metadata, automated quality control.
- Preserve: non-proprietary format, OS independent, Library standard, METS/ALTO compliant.
- Access: Improved search and retrieval through structural information, compatible with multiple content discovery systems; optional, e-commerce systems, information push services.

OCR Process

Simon Tanner⁴ from King's College London presented his findings from the evaluation he performed on the OCR produced by the JISC 1 Nineteenth Century British Newspaper

³ DTD or document type definition defines the valid syntax of a class of xml documents. A schema is used to list and describe all of the possible data content or elements of a document such as a newspaper in a very formalised way.

⁴ With acknowledgement to Simon Tanner.

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

and Burney Projects. This work was commissioned by the Library in July 2007 as part of the final evaluation of JISC 1 and was paid for by the JISC.

Simon has 20 years experience of working in over 450 digitisation projects. The OCR accuracy was assessed against the actual xml produced from the projects in relation the newspaper images not against some ideal.

In order to start the evaluation he had to decide what could be measured and was fairly directive in that he was not assessing the performance of the OCR engines or the vendor's processes (Apex) just presenting ground truths and guidance for Library to infer from the results.

Punctuations were excluded from the analysis as these are discarded by the search engines. The principle of the evaluation was to find the highest OCR accuracy on every page. This provides a solid benchmark as the accuracy would not get any better than this and because it is very easy to come up with methods to show OCR is rubbish. The evaluation covered 20,000 pages, which was sent to Digital Divide Data in Cambodia for rekeying to enable the comparison. They were told to avoid text that would be hard to rekey e.g. contained tables.

Simon suggested that in future we should use the OCR results to tell us about the corpus of our material and focus on improving significant word accuracy as this would have the highest impact on searchability. Eighteenth century text presents specific challenges such as the long s, use of apostrophes or abbreviations and external factors that will affect the OCR e.g. quality of the paper and typesetting. The lower the significant word accuracy the more the search results will be negatively affected and this is why he recommends further work being done on unique significant words which could be handled by booking a slot for grid computing at his College. We could also enhance significant word recognition on CCS processes by using some of this analysis and BL could further populate its own dictionary through the analysis.

Simon plans to publish his findings with BL acknowledgment and recommends making the data set available under the website and building the OCR analysis as part of the JISC 2 workflow.

Interoperability

An important task for the future will be a final assessment of the DTD implemented in this project and ways to improve it for harvesting and sharing with others. This is already taking place in the JISC 2 project. A copy of the newspaper dtd is included in this report. By examining the inclusion of more extensive DC elements 'The project DTD' being implemented will work to deliver the data in a form for hosting on the website and for ingest into BL DLP programme.⁵

The Library's digital newspaper resource can be further enhanced by potential interoperation with important historic digital newspaper content managed by Gale Cengage, with their nineteenth century serials programme, with which we have avoided overlap, and with their resource for electronic versions of eighteenth century books.

The website went live in October 2007 with material from JISC 1. During JISC 2 "the Library will work closely with the external hosting and serving partner to ensure the site

⁵ Digital Library Management Information Project replaced DOMS

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

keeps pace with current and emerging web tools and technologies and trends in the provision of web-based learning materials. The Library will ensure that the platform meets the JISC's technical requirements for the Common Information Environment."⁶

Keys to Success

- Pre-generated metadata sourced from the newspaper catalogue checked against each individual page and anomalies resolved with new additional metadata harvested
- Title changes, periodicity, pagination, condition and missing material will have been identified allowing for improved collections management
- The consistent delivery of microfilmed and digital content to the supplier with ongoing QA to ensure the best possible outputs
- Computer readable text derived from OCR software provides the opportunity for enhanced resource discovery techniques such as full text searching

The JISC II programme is central to the Library's newspaper strategy and benefits from the ability to draw programme support from all areas of the Library.

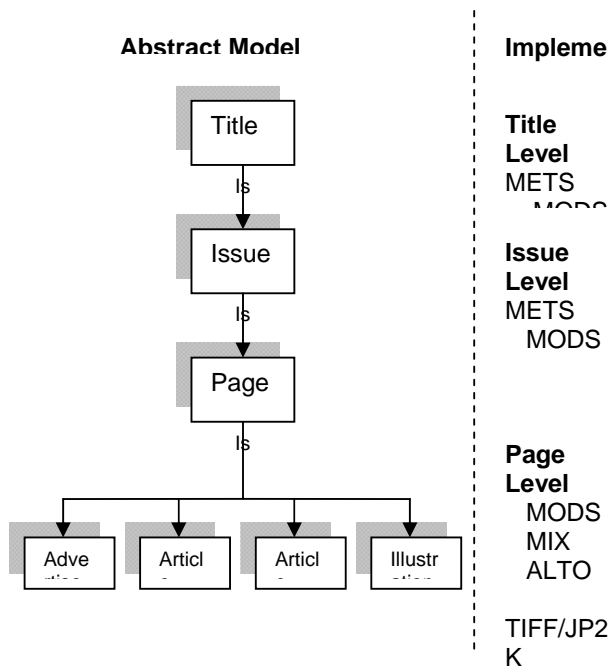
4. Project Outputs

Engagement with the Material/Selection of Content

22 newspapers have been identified to date and have been divided into six work packages. On completion of the project and its availability alongside JISC I and Burney, it is expected that HE & FE will have access to a planned 1349 titles from the period 1620 – 1900 and an estimated 20 million articles. It is also expected that, further to discussions taking place in parallel with this project, that this resource will be available to other audiences i.e. all organisations – worldwide, and all individuals, from their desktop, on a pay per view basis.

⁶ BL proposal to the JISC for funding dated 16 August 2006, p5.
Page 9 of 38
Document title: JISC Project Plan

Metadata



METS⁷ – XML Schema for encoding descriptive, administrative and structural metadata for digital objects.

MODS⁸ – XML Schema for bibliographic elements.

MIX⁹ – XML Schema for technical metadata related to digital still images.

⁷ <http://www.loc.gov/standards/mets/>

⁸ <http://www.loc.gov/standards/mods/>

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

ALTO¹⁰ – Stores layout information and OCR recognized text

See next page for the project DTD schema agreed for this project.

⁹ <http://www.loc.gov/standards/mix/>

¹⁰ <http://www.ccs-gmbh.com/alto/>

Description of the Empire NCBP DTD Fields (DTD Version 1.0)

Elements	Link between the elements and values captured in final XML file			DTD Description
	Captured from	Equivalent field	Format	
BLJISCI Note: Changed from UKP				Root element for the project
Periodical	Data Captured			Element specific to periodical content
Newspaper	Data Captured			Element specific to newspaper content
Issue	Data Captured	--		This is applicable for both Newspaper and Periodical. The terminology Issue remains the same for both Newspaper and Periodicals
metadataainfo		--		Element that holds metadata information
newspaperID	Metadata	NCBP#### corresponding to the title on the source image	NCBP####	newspaperID is a unique identifier in the format NCBP####. Where first # denotes the source location and remaining 3 #s denotes the unique identifier for a Title
JournalTitle*	Metadata	Title	--	Journal Title will not have the output
sourceLibrary	Metadata	SourceLibrary	--	Source library denotes the location where the microfilm has been sourced.
id	Data captured	--	newspaperID-iiii	Identifier for the Newspaper Issue
ed*	Data captured	--	Alpha numeric field	Edition of the Issue
is?	Data captured	--	Alpha numeric field	Source Issue Number of the Issue
volNum?	Data captured	--	Alpha numeric field	Volume Number of the Issue
da	Data captured	--	Month DD, YYYY	Date of the Publication
pf	Data captured	--	YYYYMMDD	Fixed Publication Date
dw*	Data captured	--	Alpha field	Day of the week
ip	Data captured	--	Numeric	Total Page Count for an issue
page+	Auto Generated	--		Page level elements hold the page level data
pageid	Auto Generated	--	newspaperID-iiii-pppp	Element that holds the image file name.
pa?	Data Captured	--		Page number captured from the source image
article*		--	--	Article elements
id	Auto Generated	--	newspaperID-iiii-pppp-aaa	Identifier for Newspaper Article
ocr?	Data captured	--	--	Article level OCR Confidence Level
sc	Data captured	--	Alpha	Starting column: a string to label the column where article starts

pi+	Auto Generated	--	newspaperID-iiii-pppp	Page Image file name
ci+	Auto Generated	--	newspaperID-iiii-pppp- aaa-ccc	Clipped Image file name
pc	Auto Generated	--	Numeric	Page Count: total number of pages containing sections of the article
ti	Data captured	--	--	Title of the Article
ta*	Data captured	--	--	Subtitle of the Article
detailed_au*	Data Captured	--	--	Element that holds the break up author name as prefix, first, middle, last, suffix.
prefix?	Data Captured	--	--	Prefix of the author
first?	Data Captured	--	--	First name of the author
middle?	Data Captured	--	--	Middle name of the author
last	Data Captured	--	--	Last name of the author
suffix?	Data captured	--	--	Suffix of the author
altSource*	Data captured	--	--	Alternate Source Paper of the Article
ct+	Data captured	--	--	Category: uses predefined list of article categories as per the client's specification
ll*	Data captured	--	--	Illustration: uses predefined list of illustration types like map, chart, drawing-painting, etc.
text	Auto Generated	--	--	Element that holds the OCR output
text.preamble?	Auto Generated	--	--	Element which holds the word coordinates of the Article Preamble (any text preceding article title)
pg	Auto Generated	--	--	Position guide for Article Preamble
p	Auto Generated	--	--	Paragraph
wd*	Auto Generated	--	--	Word
text.title?	Auto Generated	--	--	Element which holds the word level coordinates of the Article Title
pg	Auto Generated	--	--	Position guide for Article Title
p	Auto Generated	--	--	Paragraph
wd*	Auto Generated	--	--	Word
text.cr+	Auto Generated	--	--	Element which holds the word coordinates of the Article Content Clip Rectangle
pg	Auto Generated	--	--	Position guide for Article Content Clip
p*	Auto Generated	--	--	Paragraph
wd*	Auto Generated	--	--	Word

Quality Assurance

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

The original bid/proposal based the QA process on continued use of the online QA tool developed by Apex for the JISC 1 project. This tool enabled simultaneous QA by all of the team on a randomly generated sample of articles, images and xml with the facility to generate QA reports and track progress. As Apex was replaced first by Olive and then this year by CCS, we lost access to this tool.

Instead, we will be accessing data from hard drives and online for our QA and still use the same method which is to look for the following:

- Good quality images with no loss of text, image skew or truncation
- Metadata integrity
- Zoning of each page
- OCR
- Correct page sequence
- File linking

During the procurement process, Olive was asked to undertake a small pilot and process 19,916 pages from 3 titles. A report documenting the results of 1% QA on the pilot was sent to JISC May 2008. In summary, a paper copy print produced in-house from the microfilm reel was generally more readable than the Olive scan, even where density varied within a page. Further QA would need to be done to provide confidence in image quality if Olive rescanned every reel and this remedy would cause further risk to the schedule. It was decided to offer the work to the runner up CCS who in turn had to pass a small pilot.

One key lesson learnt from both the JISC 1 and Burney projects was how time consuming and subjective QA can be. It can become one of the challenges of mass newspaper digitisation along with over zealous subject categorisation and the effort involved in achieving 100% OCR accuracy, even if that were a realistic goal.

The JISC II project is about using more automation and software tools to screen out inconsistently applied quality acceptance criteria via human intervention. It also recognises there is a limited return on 100% QA that in itself can introduce further errors and anomalies.

For JISC II the bulk of the QA is front loaded during the prep and microfilming stages followed by up to 2% sampling of CCS's outputs. With the extra QA that Gale Cengage are planning it makes no sense for the BL to spend time in a tight schedule on introducing labour intensive and inconsistent QA procedures. It is planned that BL will use forensic QA on the Pilot and for the first three months of production and then tail this off. The QA team will be two members of staff, the project manager and the team leader, supplemented when necessary by BL imaging resources.

The project website will also have been developed significantly and will carry reports of lessons learnt. Conference papers and presentations will have been given. The sharing of knowledge and dissemination will have added to the community of practise in this important area. Case studies will become available.

Less tangible aspects include:

- Continuous development of expertise within the Library with regard to newspaper digitisation
- Engagement with the JISC community; understanding of different users needs
- Engagement with other key players e.g. NLA, Loc

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

- Development of tools for access and repurposing
- Improved understanding of OCR accuracy and development and use of OCR analysis tools

The Library will deepen its knowledge of the historic collections currently based at Colindale.

It will be able to publicise further:

- The average numbers of pages for different types of newspapers
- The length of time it takes to prepare newspapers by gathering metadata essential to input to the DTD schema
- The length of time it takes to repair torn pages in volumes of newspapers, or to scan directly from the original

5. Project Outcomes

Relevance to UK learning, teaching & research.

The British Newspapers 1800-1900 project was launched on the 22 October 2007. Since this time, the Gale Cengage has signed licences for 159 UK HE/FE institutions. The database, entitled for marketing purposes "C19 British Library Newspapers" has also been available in the entire Library's reading rooms at St Pancras, and also at the reading room for newspapers in Colindale. The Library has received many dozens of messages from users all over the UK in support of the online database. Some allude to how much it will transform their research.

It is anticipated that the additional one million pages that are being digitised as a result of this project will significantly add to the enrichment of the user experience of the database that Gale Cengage operates. The texts of the newspaper will also complement the digitised text being made available as a result of the JISC funded projects in particular for C19 Pamphlets.

It is envisaged that the Library will continue to engage with the UK HE and FE sectors, to explain the background to these projects, and what content is available. E-Learning and Webinars are two instances of how the Library can spread awareness of the content to new audiences.

Newspapers Website

Background:

Since building the HE/FE site last year, Cengage have moved to an enhanced platform for all their products, it allows more flexibility for bespoke design and functionality; this

[REDACTED]

[REDACTED]

Business Model

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Website Functionality

Key areas of additional functionality, beyond the existing HE/FE interface will include:

- Browse by Location another approach to get to specific titles
- Research Tools section has contextual essays explaining historical importance of newspapers as well as topical articles, providing insight into the 19th Century for novice searchers
- An interactive chronology ('History through Documents) with live links back to results related to the timeline entry
- Users will be able to save and download article, as well as mark them as favourites.

User Testing

On 6 August 2008 WUP (Web Usability Partnership) under guidance from the British Library, undertook user testing of an early prototype of the new commercial site. In particular the British Library wished to gain insight into:

- Usability – ease and intuitiveness of the prototype site
- Users' attitudes to the prototypes.

The feedback focused on:

Project Acronym: JISC2 Newspapers
 Version: 0.2
 Contact: Patrick Fleming/ Jane Shaw
 Date: 30.1.2008

- Users and goals – can users understand what the site is, what goals they can achieve on the site and who the site is aimed at?
- Getting to goals – how well does the navigation aid users in achieving their goals? How effective is the search?
- Content – users’ reactions to the brief and full articles.
- The pricing model – what are users’ expectations and understanding of the pricing model.

Three testers individually explored the site during observed research. The individuals included a teacher, a local historian and a family historian.

Overall, testers thought that the site could be extremely useful to them and they managed to use it without major difficulty. All testers said they would use the site again (two specifically mentioned they would pay to use it), and all testers said they would recommend the site.

Homepage screenshot (10/10/08 iteration):



6. Stakeholder Analysis

Stakeholder	Interest / stake	Importance
JISC Executive	The funding body	High
BL Executive Team	BL ownership	High

All UK HE/FE institutions	Research, tailored to individual projects, and also to research by students attending courses	High
---------------------------	---	------

7. Risk Analysis

ID	Date Raised	Risk Description	Owner	Current Probability	Impact	Ranking	Mitigation Actions / Response	Status	Last Review
1	01-Oct-03	Newspaper Publishers continue challenge over the right to digitise/provide access to the selected materials	Ronald Milne and ET	Medium	Low	1	BL seeking to sign off selected titles with newspaper publisher organizations. Current search on copyright for JISC 2 titles	Open	15/10/08
3	Oct 2003	Failure of selected supplier to consistently deliver the required quality standard, causing unacceptable rework or delays to schedule.	Project Board, Risk Group	Low	High	6	Olive failed the pilot. Runner up (CCS) conducted a successful parallel pilot during April/May 2008 and is now the preferred supplier.	Closed	12/08/2008
4	01-Oct-03	Insufficient BL resources to undertake the microfilming to schedule and complete all filming and duping work by end 2008.	Andrew Kenny	Low	Medium	4	Resources at Colindale have been made available to ensure project success	Open	29/10/08
7	26-Apr-04	Loss of key staff.	Project Director	Medium	Low	2	QA activity matches staff resource and is reduced in 2008 as staff are redeployed.	Open	29/10/08
8	03-Aug-04	Loss or damage of material in transit, within BL or at suppliers	Project Manager and Team	Low	Low	1	Insurance included in shipment costs. Redupe if lost or damaged by customs. Ongoing audits by project team.	Open	7/07/2008

							Supplier agrees to pay invoice for damaged films so they can be replaced.		
9	29-Apr-04	Insufficient or irregular feedback from JISC.	Project Manager	Low	Medium	2	PM and SRO keeps in regular contact with JISC Programme Manager. JISC attends monthly project board meetings.	Open	07/07/2008
10	29-Apr-04	BL Acetate duping machine is slow and likely to break down, impacts three Irish titles	MFU	Low	Medium	2	Agreed an approach that minimizes us of BL acetate master microfilms. Metadata still has to be collected either from dupe or from originals.	Closed	12/-8/2008
11	01-Jul-07	Preparation team failing to deliver work to Microfilm Studio, impacts on schedule. Material is in worse condition than JISC 1 and contains many more duplicates and variants which slows down production.	Project Manager	Low	Medium	2	New recruits joining team and consistent effort by prep team has resulted in all prep team achieving their 1 million targets by end of September 2008.	Closed	29/10/08
12	29-Sep-04	Unforeseen increases in costs may reduce the number of pages able to be digitised under preset funding	Project Analyst, BL Finance	Low	Medium	2	JISC will not agree to a Standing Contingency fund. Careful budget management required.	Open	16/11/2007
17	01-Oct-07	Possible delays to EU procurement of scanners at Boston Spa.	Martin Lunn	Low	Medium	2	Continue to monitor EU procurement of scanners at Boston Spa and any possible impact on project productivity to contractor.	Open	16/11/2007

18	16-Nov-07	£75,000 accrued interest from JISC 1 needs to be added to JISC 2 to bring total budget up to £1,143,000.	Project Manager	Low	Low	1	Secure agreement with Phil Michel to ring fence this money and call on it as needed. JISC to agree how shortfall is funded.	Open	21/01/2008
19	07-Jan-08	New requirements could delay schedule and increase project costs.	Project Board/SRO	High	High	9	BL determines scope of work; all parties agree and sign off. Olive producing price proposal for scope change by week ending 25/01/2008	Closed	21/01/2008
20	09-Jan-08	Paying for post 1865 content which we may not be able to load immediately, due to copyright permissions being needed.	SRO	Low	Medium	2	Publisher and printer names being captured during preparation stage. SRO leads on copyright research.	Open	21/01/2008

8. Standards

Name of standard or specification	Version	Notes
<i>Web Publishing</i>		BL is aware of the MINERVA technical guidelines and standards for digitisation, including Dublin Core metadata standards, the use of DTD/XML formats and open access file types such as TIFF and JPEG2000. We already comply and are actively addressing a few remaining issues.
<i>Interoperability and XML interface requirements –</i>		Metadata planning and compilation creates XML data which is consistent with the Dublin Core framework. BL requires that all new major systems developed by or on behalf of the

		organisation should include a Z39.50-compliant interface. ¹¹ This should be implemented using SRW/U, the Search and Retrieve Web Service (SRW)/Search and Retrieve URL Service (SRU) protocol, which uses easily available technologies (XML, SOAP, HTTP, and URI) to perform tasks traditionally done using proprietary solutions. It can be carried either via SOAP (SRW) or as a URL (SRU) and all search/retrieve requests use the intuitive 'Common Query Language'.
--	--	--

9. Technical Development

As part of the EU procurement process undertaken in 2007, the Library required from bidding companies:

1. The Ability to deliver XML metadata and service images of the required specification within the project schedule
2. Overall quality of the solution proposed for (in descending order of priority):
 - a. OCR scanning, with any enhancements for achieving best accuracy.
 - b. Content analysis of the structure of each issue (accurate zoning).
 - c. Type classification, title capture and further analysis of newspaper articles.
 - d. Accurate reproduction of graphical content and illustrations in the service images.

The Library has defined a technical development approach that will be implemented by the project. The approach has been driven by three factors:

- the approach taken and lesson learnt on previous newspaper digitisation projects undertaken by the Library
- to meet standards for the long term preservation of the digital files
- review of approach taken on similar projects in Europe, US and Australia.

¹¹ Z39.50 – Gale has this capability but it is not available in the current BL products. It will be added during a future migration scheduled to be completed by the end of this year.

10. Intellectual Property Rights

Because all newspaper titles in British Newspapers 1620-1900 (JISC II) will be more than 100 years old, any Intellectual Property Rights in the titles or their contents will normally have lapsed. The final title selection will remove any titles which continued into the 20th century. This approach is consistent with British Newspapers 1800-1900 (JISC I), and experience from that project confirms that newspaper publishers now accept that such material is out of copyright.

Nevertheless, as a matter of courtesy, the Library will again inform publishers of the intention to digitise this out-of-copyright content. We will also continue an ongoing dialogue with publishers of newspapers and The Newspaper Society. Should any small amount of material still be in copyright, we have the in-house expertise to identify this and would naturally seek to clear rights. Since JISC II started, and since the launch of the online database for the UK HE/FE sectors, we have stepped up our efforts to be specific regarding copyright for all newspapers in JISC I and JISCII. The work to date shows some newspapers are now out of copyright; however, for other titles, it can be very difficult to establish the owners in 1900, and show that the title is out of copyright.

One way that might assist us is capturing additional metadata such as printer and publisher names and we decided to adapt the project DTD schema and manifest sheet to allow the preparation team to gather this level of detail. Both of these are new elements to the DTD schema prepared during tendering. We believe this could greatly assist with IPR clearance.

The Library has already agreed to grant to HEFC in perpetuity a non-exclusive, royalty-free licence to use the intellectual property rights identified as owned by the Library at the start of the project, or such additional rights as will be generated for BL during the course of the project, and which are necessary to provide access and allow use of the on-line British Newspapers 1800-1900 by FE and HE institutions. Making a separate licence to effect this will be put in hand.

This will ensure that users from HEIs and FEIs will be able to use freely not only the digital newspapers themselves but also the digital objects created as part of the development of learning or teaching packages and that these can be made freely available through JORUM and other such repositories.

Project Resources

11. Project Partners

The BL signed a contract with Cengage Gale Learning in March 2007. The BL is in partnership with Gale to:

- Make the JISC I 19th century newspapers content available free of charge to all institutions in the UK HE/FE sectors
- Ensure that the database is sustained and publicly available for a number of years
- Ensure that the database is marketed to users (both organisations and individuals) worldwide during the course of 2008
- Upload the JISC II newspaper content to its web platform

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

Owing to several parts of these contracts being confidential between the Library and each company, the Library is the first point of contact for enquiries.

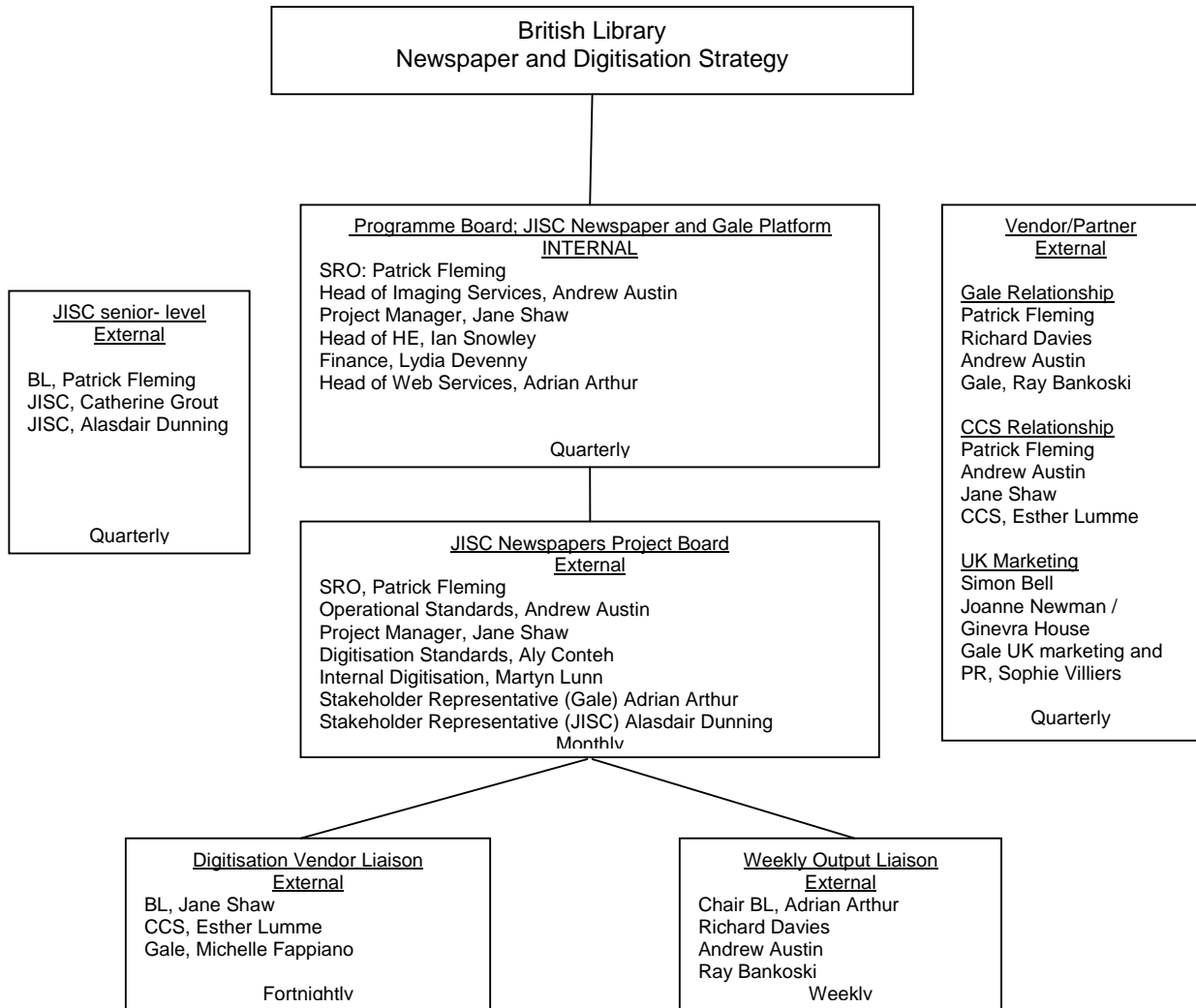
12. Project Management

The project is managed by the same project manager as JISC I and she provides continuity with the funder, the board and with the lessons learnt from JISC I.

She reports to a monthly operational Project Board consisting of key stakeholders in the Library, the funder JISC and the Chair of the User Panel. The project manager works closely with the SRO to realise all of the deliverables.

The decision making lies with the SRO according to PRINCE 2 principles, although he is supported and informed by both the project manager and board members. The project manager manages all production on the project, including reporting to the JISC programme manager on progress, risks and issues and acting as the contract manager for the operational running of the digitisation contract with CCS.

The JISC II project sits within a newspaper programme/framework and is one strand of work that reports into the Newspaper Programme Board. To avoid an over large and cumbersome project board on JISC II, it was decided to introduce a different project management framework this time, namely, there will be a leaner Production Project Board with its own SRO, another Risk Panel and Content Development Project Board focusing on all matters relating to Gale delivery of content and marketing of content. The JISC II project manager will attend as many of these satellite meetings as necessary. [See framework chart]



The project manager runs monthly Checkpoint meetings with the internal suppliers such as the Microfilm Unit to discuss issues and progress and holds conference calls with CCS and the Boston Spa scanning unit.

The project manager has an informal link to the Gale Cengage production manager and the e-is Digitisation Programme Manager to ensure they are both kept up to date with the details of progress, although they are board members under the newspaper framework.

Jane Shaw, Project Manager 020 7412 7368
 Lisa Kenny, Quality Assurance Officer 020 7412 4800
 Fiona Mc Henry, Quality Assurance Officer 020 7412 7353
 Stephen Leary, Quality Assurance Officer 020 7412 4834
 Rita Huggins, Quality Assurance Officer 020 7412 7356

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

Kawaljit Singh, Quality Assurance Officer 020 7412 7357
Lee Campbell, Quality Assurance Officer 020 7412 7378

Jane Shaw works full-time on the JISC2 project and is the relationship manager for JISC for the project.

As of November 2008, the above team will be decreased as the preparation work has finished. The team will return to their previous posts at the end of November, four months earlier than planned allowing for an orderly closed down. There will be two staff left to complete the QA and manage the closedown, the project manager and team leader.

13. Programme Support

None at present.

14. Budget

The budget has been under review and we have made some cost savings to remove the over spend and deliver the project to its original target. This has been achieved mainly from staff salaries due to an earlier close down. The current budget is attached.

Detailed Project Planning

15. Workpackages

See Appendix B

The Work Package List presented here is our preferred format for tracking progress. There is also a Gantt Chart.

We have developed and implemented a system for tracking the movement of collection items once they are removed from shelf by the JISC team. Once a volume is removed from the shelf a marker is placed in its stead to indicate to any other potential user the date of removal and that the item is part of the JISC Newspaper Digitisation Project. Each volume is recorded individually, its date of removal from shelf, the date it was sent to MFU for filming and the date of return. Those volumes sent to Boston Spa for direct scanning from the volumes are recorded on a folder in a shared drive. The system was put into practice some time after the beginning of the project so the records are not complete but do ensure that a volume can be located at any point in the work flow. We track the movement of the microfilm negatives in the same way. Each reel is identifiable by both its unique ID number allocated when we receive the reels from MFU and also the title and date range as it appears on the reel. We must periodically send microfilm to be scanned by Hermann & Kraemer in Germany. On these occasions comprehensive lists are compiled which we keep as a movement

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

control record, our partners CCS ingest these lists into a tracking sheet which they share with us to inform us of the location and progress of each title. When the reels are returned and we have finished our QA the reels will be accessioned into the collection.

Title List Linked to Metadata

A document is currently being compiled to display the bibliographic information captured in our metadata. Based on the JISCI version the document will track all title changes, publication days and significant differences in format. Until the QA is completed the list will remain a work in progress to ensure accuracy of information

16. Evaluation Plan

Background

The 19th Century Newspapers project was launched in October 2007. It provides online access to over 2 million pages of regional, local and national newspapers from 1800 to 1900. The project is served by Gale Cengage Learning, who provides the web interface and search functionality. Funded by the JISC, the service is free to UK higher and further education, and available by subscription to other institutions. A further 1 million pages are currently being digitised under the JISC 2 project, and should come online in 2009.

Users can search through all pages using advanced Optical Character Recognition (OCR) technology and download pages for teaching, learning and research purposes, e.g. for inclusion in institutional repositories, Virtual Learning Environments (VLEs) and course material packs.

Gale Cengage have been engaged in marketing activities aimed at increasing sign-up, and so far (as of July 2008) 244 institutions have signed up. This currently includes 120 schools/ FE institutions and 124 HE institutions

A further project to digitise 1 million pages of the Charles Burney Collection of newspapers and related ephemera, dating from 1600 to 1800, has also recently come online. If funding is obtained to make Burney available for free to UK HE and FE institutions, then the project should also be incorporated into this marketing plan. There may also be scope to include the library's Nineteenth Century Serials Edition (ncse) in these marketing activities, once this project is fully launched.

A new Business to Consumer model is planned. This will allow individual users to access the resource by paying for a 24 hour subscription. Anticipated new markets will be the genealogy market, local historians, lifelong learners and private or company researchers. This marketing plan deals primarily with promoting the service within HE and FE and will not, at this stage, include marketing to the above groups.

More information about the project can be found at <http://newspapers.bl.uk>

Marketing Objectives

- To continue to gain sign-up to the service, in collaboration with Gale Cengage
- To increase usage of the service by students and academic staff from universities and colleges in the UK, measured through web statistics.
- To gain an understanding (quantitatively through web stats, and qualitatively through case-studies and consultation) of who is using the service, how and why
- To raise the profile of the service within the academic community and the British Library
- To raise the profile of the role of the BL and the JISC in underpinning the research infrastructure of the UK
- To deliver key messages for the project (see below)

Key Messages

- The British Library Newspapers website has digitised two million pages from forty-six 19th Century newspapers, thanks to funding from the Joint Information Systems Committee (JISC)
- Further funding from JISC will allow the addition of another 1 million pages, offering further regional and London papers, covering a greater expanse of the British Isles
- The collection includes 19th century national and regional titles, chosen by leading experts in the field
- The content is uniquely rich and fully searchable online using Optical Character Recognition (OCR) software
- The resource will greatly enhance research into the society, culture and history of the UK between 1800-1900 (or from 1600-1900 if Burney included. In this case, the resources would form a single, cross-searchable resource)
- Digitisation is central to the Library's strategy in the 21st Century: this project is just the latest in a range of initiatives (e.g. Archival Sound Recordings, Turning the Pages, Collect Britain, EThOS) we are pursuing in order to make a 'critical mass' of digitised material available online to users

Target audience

Target individuals

- Teachers, lecturers and academics in FE and HE
- HE and FE librarians, information professionals, learning resource managers
- HE and FE students, taught and research courses (end users)
- BL readers
- BL stakeholders
- Named academics (who might also endorse our activities)
- Key associations (who can help promote the service amongst academic communities)

Target disciplines

- History
- English/ Literature
- Social Studies: economics, business, politics, sociology, cultural studies, anthropology, human and social geography
- Mass Communication: journalism, media studies, publishing
- History
- Creative Arts and Design
- Family historians

Tactics

- **HE e-newsletter**
To feature 2 items in HE Newsletter between now and March 2009.
 - Content highlights - topical subjects
 - Research highlights
- **Lis-servs**
 - To offer promotional material (leaflets) and support materials (Power Point guide) on the service to academic librarians through lis-SCONUL, RLUK, RSCs etc
 - To promote service to academic communities through subject-based lis-serves such as JISC-lists and Higher Education Academy Subject Centres
- **Direct Mail**
 - Letter to signed institutions offering promotional and support materials (see above) for circulation and materials for staff/ student training sessions (see Design Work)
 - Letter to librarians of unsigned HE and FE institutions with history departments, encouraging sign-up
- **Design Work**
 - To create generic leaflet incorporating 19th C Newspapers and Burney without naming them specifically, e.g. Historic Newspapers Online.
 - To create a Power Point guide to using the resource which can be distributed to institutions for use in staff/ student training sessions?
- **Events**
 - To hold at least 1 workshop for the academic community on the use of the resource.
 - To be held in 2009
 - For academics and post-graduates
 - Aimed at finding potential case studies and increasing use of service by encouraging academics to use with students
 - May incorporate other 19thC online resources

- To hold a launch event for JISC 2 in 2009
- **Key conferences**
 - To ensure appropriate presence at targeted conferences with engagement tactics including
 - Workshop or paper on Newspapers
 - Conference stand
 - Leaflets in delegate packs
 - At least 4 conferences over coming year
 - Outcomes: increased sign-up, raise profile (and usage) of resource, make contacts for case studies, distribute promotional material
 - To continue to update conference list throughout the project.
 - Attended so far:
 - SCONUL conference – 30/01/08 – HE stand
 - UKCGE – 12/2/08 – HE stand
 - JISC 2008 15/4/08 – HE stand
 - Thomason Tracts conference – 30/6/08 leaflets
 - IHR Conference – 2/7/08 - leaflets sent
 - nsce launch – 13/5/08 - stand
 - Negotiating the Mediascape – 2/5/08 – stand
 - British Library Training Days for Postgraduates
- **Press coverage**
 - To gain coverage of 19th Century Newspapers along with Burney and ncse as they become fully accessible for HE.
 - To encourage at least 1 review of services by academic community. Possible journals include:
 - Nineteenth Century Gender Studies <http://www.ncgsjournal.com/>
 - Cultural and Social History Journal <http://www.socialhistory.org.uk/journal.htm>
 - History Workshop Journal http://www.oxfordjournals.org/our_journals/hiwork/about.html
 - Reviews in History <http://www.history.ac.uk/reviews/index.html>
 - Historical Research <http://www.history.ac.uk/historical>
 - Rethinking History <http://www.tandf.co.uk/journals/titles/13642529.asp>
 - Intellectual History Review <http://www.tandf.co.uk/journals/titles/17496977.asp>
- **Feedback**
 - To gain monthly statistics on sign-up and usage of the service, to be obtained through Gale Cengage/ GlobalTech
 - To develop 5 case-studies of users to help understand how people are using the resource and to help encourage usage
 - To identify a mixture of academics and students through contacts with the old user panel, and others made at conferences

- To obtain 'sound bite' quotes for publicity materials and press releases
 - To create more detailed case studies of the research the way the resource was used, with the aim of encouraging greater usage in teaching, learning and research
 - To determine appropriate placement of case studies – whether on BL website, 19C newspaper website, or simply to be used in presentations and workshops
-
- ***Raising awareness of the service with existing BL Readers***
 - Leaflets in Reading Rooms (StP and BSpa)
 - Two articles in the Reader Bulletin (e-mail and printed) during the two year period.
 - 2 staff briefings for reading room staff to ensure they are aware and fully briefed on the service by March 2009.
 - ***Stakeholders and Partnerships***

To continue to develop our relationships with:

- JISC
- SCOUNL
- CURL
- UKADIA
- The Higher Education Academy and their Subject Centres
- Former User Panel members

To continue to identify key organisations that can help encourage the use of the resource, e.g.

- Subject Centre for History (Higher Education Academy)
- Institute for Historical Research
- Centre for 19th Century Studies at Birkbeck University
- Centre for 19th Century Studies at Sheffield University
- MECCSA – Media, Communications and Cultural Studies Association
- Association of Media Practice Educators
- Others to be identified.

What will a productive relationship look like?

- Distribution channel for promotional material (printed, e.g. leaflets, or via the web/ email)
- Partners in developing workshops
- Speaking opportunities at key meetings, conferences or workshops
- Source of feedback/ case-studies for the service.

Milestones

July 2008

Item in HE Newsletter on ncse

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

August 2008

Budget to be submitted

September 2008

Leaflet draft to be submitted
Old User Panel members to be contacted for case studies/ journal reviews
1 new case study identified
Digital Resources in the Humanities and Arts Conference
ALT-C – Rethinking the Digital Divide Conference

October 2008

New leaflet printed
1 Case Study identified
Power Point presentation for institutions complete
Briefing for reading room staff

December 2008

Contact journals to discuss possible review or feature
SCONUL Conference
1 new case study identified

January 2009

Item in Reader Bulletin

Mailout to signed institutions offering publicity material
Item in HE Newsletter
1 Case Study identified

February 2009

Briefing for reading room staff

March 2009

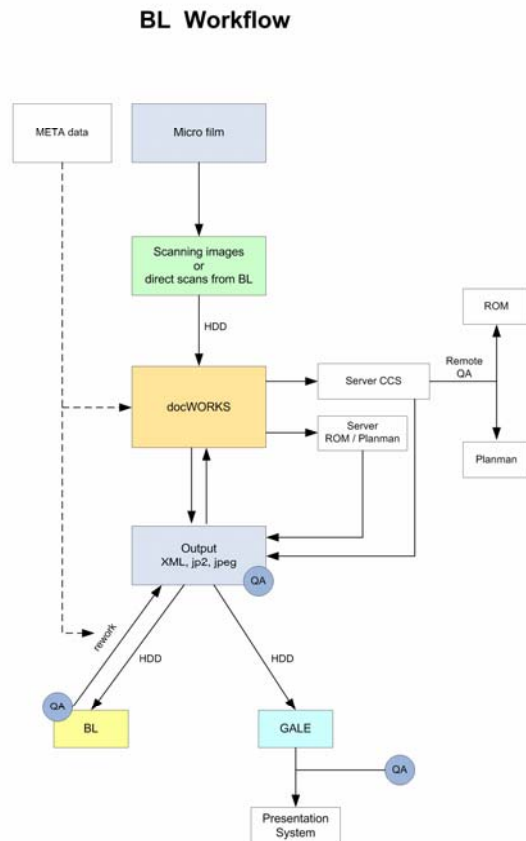
Mailout and lis-serve postings to encourage sign-up to service

April 2009

Social History Society Annual Conference

17. Quality Plan

The following diagram¹² shows there are 6 QA points in this project. At preparation and microfilming, prior to shipment to CCS, during processing at CCS, during online checking at page, article and image level by the Library project team, during checking of file naming conventions and structures of final outputs by BL preservation team and finally intensive QA by Gale Cengage.



With such a high volume of newspaper texts being digitised in a shorter time, and continuous QA done on metadata accuracy, scan quality using scan logs, OCR quality, the segmentation quality and sequencing using chck sum files only 1% of the total volume of pages will be checked in detail by the project team. The earlier project was able to QA 2% of 2 million pages so this would be in the right percentage.

¹² CCS has not given permission for this diagram to be included on a public domain website as it is commercially in confidence.

The quality plan dictates the workflow and size of the milestones. It has been agreed that a regular turn around of small batches of data between the Library, CCS and Gale, with two months in hand will allow for easier rework if needed, smoother billing and capacity plans.

The digitisation supplier CCS will follow their own internal QA procedures as will Gale Cengage and the BL preservation team who are currently working with the Library project team on a joint quality plan.

Tables below indicate how sampling will be tested against the acceptance criteria and current tolerances, i.e. the maximum designated value of "percent nonconforming".

1. Newspaper preparation and repair. Where filming from original newspapers is required, each page is checked for condition and repaired as appropriate (100% sampling). Metadata elements are captured during this process.

Attribute	Tolerance	Unit of Measure
Condition of pages and binding (minimum handling and avoiding over-repair)	Within BL guidelines	Newspaper volume
Metadata completeness for key elements (title, issue, page, pagination)	1%	Pages/records per volume

2. Microfilming and duplication films: Since 1990 all BL microfilming and duplication has been done to British standard BSI-ISO 4087 (205) and NPO guidelines (100% sampling).

Attribute	Tolerance	Unit of Measure
Density (highest possible)	0.9 – 1.5	Reel
Reduction (least possible)	< 17 times	Reel
Resolution (highest possible)	Min.100	Lines per millimetre.

3. For digital scanning and creation of final output, sampling of image and XML files output will be random and conform to ISO-2859-1.

Attribute	Tolerance	Unit of Measure
<i>Scanning, OCR and creation of master digital images</i>		
Automated data testing	0%	Various file formats
Image Link	1.0%	XML link element
Image Skew, Sequence & Inventory	1.0%	Image File
Image Quality	1.5%	Image File
<i>Article Recognition, categorisation & XML output</i>		
XML File Inventory	2.5%	XML file
XML Element Content	4.0%	XML element
XML Metadata Character Error	n/a	OCR accuracy

Article recognition will conform to the CCS standard of profiling and subject categorisation will conform to the Gale standard. OCR accuracy will not be QA-ed as a level will not be set; we require samples to review only and OCR confidence levels on articles.

18. Dissemination Plan

Marketing Objectives

- To significantly increase usage of the service by students and academic staff from universities and colleges in the UK
- To raise the profile of the service, the British Library and the JISC Programme
- To raise the profile of the use (and issues around) of historic text in teaching, learning and research
- To gain an understanding (quantitatively through web stats, and qualitatively through case-studies and consultation) of who is using the service, how and why
- To continue to gain sign-up for the service from organisations who did not do so during JISC 1

Target audience

In the first Newspapers Digitisation project we focused on engaging with academic librarians to increase sign-up to the service. There will be a shift in focus in JISC II towards the end users; academic staff, students and existing BL readers. However, the target audiences can be defined as:

Those responsible for signing the license agreements:

- HE and FE librarians

Key stakeholders who will be able to promote the service to others, act as advocates:

- HE and FE librarians, information professionals, learning resource managers
- Teachers, lecturers and academics in FE and HE
- HE and FE students, taught and research courses (to peers)
- BL stakeholders

End-users:

- Teachers, lecturers and academics in FE and HE
- HE and FE students, taught and research courses
- BL Readers (on site)

Target disciplines:

The target audiences are similar to those outlined in JISC I

- Art and design - architecture, painting, design, audio visual media, fine art, photography, postmodernism, sculpture
- Cultural studies - cultural and ethnic identity, cultural history
- Media studies - broadcast media, journalism
- Social sciences
- History and classics

May- Jun- Jul-08 Aug- Sep-

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

	08	08		08	08
No. Institutions	234	243	249	252	258

Sign-up campaign:

Actions – June 2008

- Direct mail campaign to HE and FE librarians
- Posting to SCONUL lis-serve
- Emails sent by RSCs to member institutions

Results:

- New license requests: 24
- Requests for publicity materials: 13

Conferences

28/5/08 – ncse launch at BL - stand

24/6/08 – SEARCH! Conference – ASR publicity materials in delegate packs

2/7/08 – Institute for Historical Research Conference – ASR publicity materials sent

24/9/08 – ARLIS Conference – Newspapers included in British Library paper

9/9/08 – ALT-C conference – exhibition stand

16/9/08 – Digital Resources in Humanities and Arts Conference – Newspapers mentioned in paper

22/10/08 – Oxford History Graduate Fair – stand and presentation incl. Newspapers

Testimonials

Action: Testimonials and case studies sought from academic users through previous user panel contacts, librarians and contacts from conferences.

Results:

19th Century British Library Newspapers gives us an invaluable new perspective on the way Victorian Britain came to understand itself as a community of citizens, consumers and commentators. It is hard to imagine any researcher with nineteenth-century interests to whom it does not offer important insights on their subject: and it is dangerously addictive!

Prof. Arthur Burns, Kings College London

The well chosen geographical range of provincial newspapers and the sophisticated search facility could well change the face of British historiography over the coming decades, as genuinely national treatments of topics are now much more manageable.

Andrew J. Hobbs, PhD student, University of Central Lancashire

Work that would have taken painful weeks on a microfilm reader can now be done in an afternoon. One undergraduate, for example, tracked the early weeks of the newspaper reception of Darwin's Origin of Species, the first time this had been attempted for half a century. I used to spend much of my time showing students how to find things; now there is more time for them to think.

Prof. James A. Secord, History and Philosophy of Science, University of Cambridge

Design Work

Leaflet for British Library Newspapers Online currently under development

Page 36 of 38

Document title: JISC Project Plan

19. Exit and Sustainability Plans

The prep team are returning to their previous posts at the end of November, four months earlier than initially planned. The last of the set aside volumes are being prepped by the team and all volumes will be in the MFU or ready to move to Boston Spa by the end of November. The QA stage will now be conducted in St Pancras.

Project Outputs	Action for Take-up & Embedding	Action for Exit
All images, plus xml for each image on tape	BL will arrange with Gale Cengage to upload the tapes to its web platform	Plan and Arrange with Gale Cengage
All images plus xml	BL will arrange to these files to be ingested into it digital object management (DLP) programme, to ensure preservation	Arrange with BL DLP

Project Outputs	Why Sustainable	Scenarios for Taking Forward	Issues to Address
Further development of the historic newspaper collection	Support from the BL newspaper strategy and potential investment from publishers	Ongoing discussions with publishers	Licensing, copyright and terms of reference for future partnerships

Appendixes

Appendix A. Project Budgets see attachments for audit trail.

removed from this version

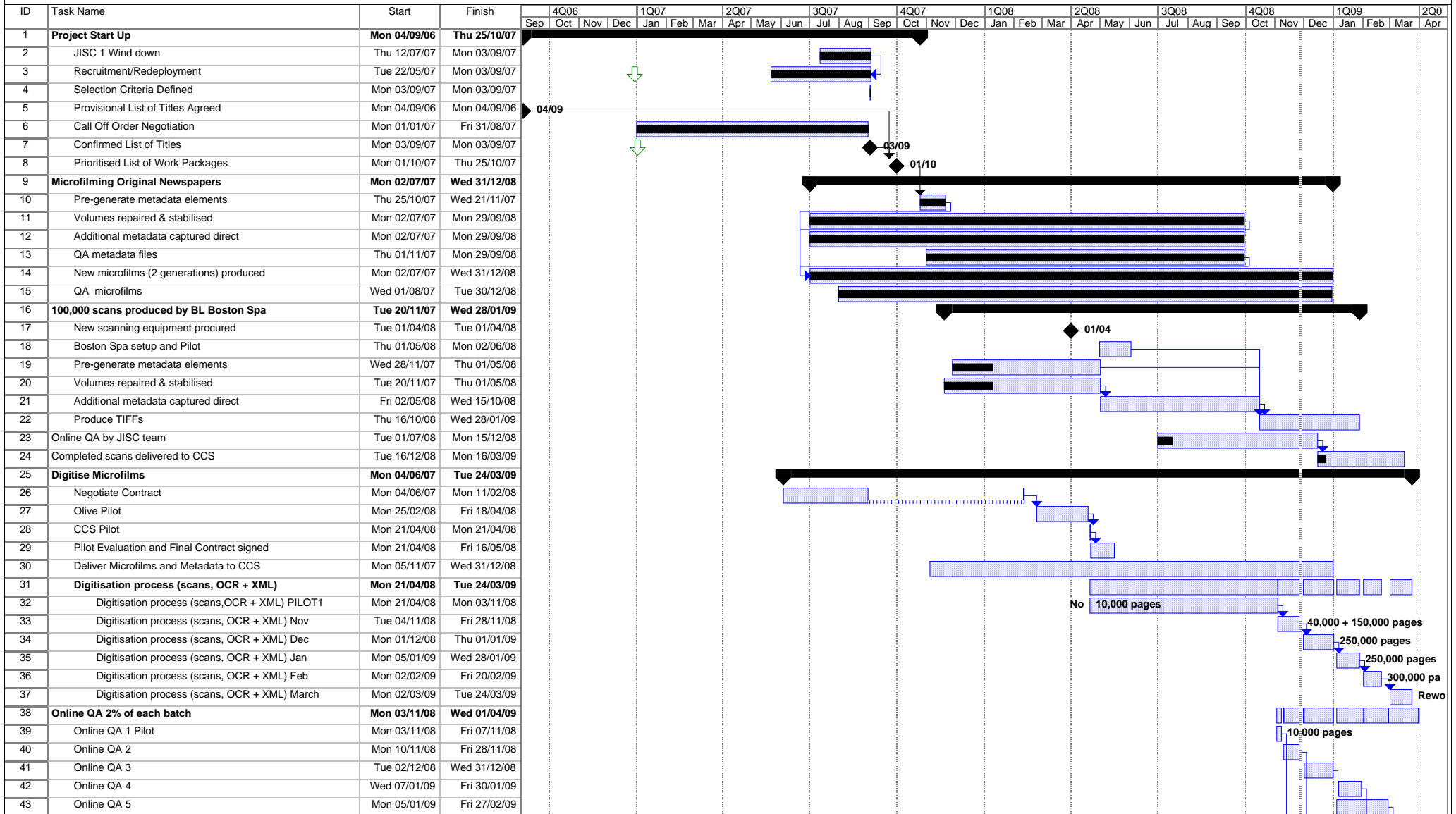
Appendix C Project Plan

This project plan has one Gantt Chart the project outline updated since November 2007 and described as a Tracking Gantt. It includes CCS's latest schedule and milestones

The main overriding risk now is the schedule. A late start, an early scope change and delays on the digitisation pilot have necessitated the creation of a Tracking Gantt chart.

Project Acronym: JISC2 Newspapers
Version: 0.2
Contact: Patrick Fleming/ Jane Shaw
Date: 30.1.2008

BRITISH NEWSPAPERS 1620 - 1900
PROJECT OUTLINE
1 JUNE 2007 - 30 MARCH 2009
updated Nov 2008



Task: [Blue hatched bar] Progress [Black bar] Summary [Black arrow] External Tasks [Grey bar] Deadline [Green arrow]

Split: [Dotted line] Milestone [Black diamond] Project Summary [Grey arrow] External Milestone [Grey diamond]

7.7 hours per day
37.5 hours per week
20 working days per month

