



## Project Document Cover Sheet

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

Document Name			
<b>Document Title</b>	<i>Project Plan</i>		
<b>Reporting Period</b>	N/A		
<b>Author(s) &amp; project role</b>	Jane Shaw, Patrick Fleming.		
<b>Date</b>	28 February 2008	<b>Filename</b>	JISCNewspapers2projectplan5.0final
<b>URL</b>	<i>if document is posted on project web site</i>		
<b>Access</b>	<input type="checkbox"/> Project and JISC internal		<input type="checkbox"/> General dissemination

Document History		
Version	Date	Comments
Final (5.1)	28/02/2008	Final



# JISC Project Plan

## *Overview of Project*

The objective for this project is to digitise some 1.1 million pages from regional and local newspapers from the 19<sup>th</sup> century and all of these are intended to be uploaded to the website currently hosted by Gale Cengage in partnership with the British Library by March 2009.

## 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 1.1 million pages comprising mostly 19<sup>th</sup> century regional titles.

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 <sup>1</sup>

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 on budget.

Formalised digitisation standards are in place for the project delivery, a contract has been signed with our web hosting partner Cengage Gale and a procurement process has been successfully completed to identify a digitisation partner.

JISC1 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 also facilitate content from JISC II.

---

<sup>1</sup> BL proposal to the JISC for funding dated 16 August 2006, p2.

Project Acronym: JISC2 Newspapers  
Version: Final (5.1)  
Contact: Patrick Fleming/ Jane Shaw  
Date: 28.02

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 will also include 200,000 pages scanned from original newspapers at the Library's in-house scanning unit at Boston Spa. The Library is investing in leading-edge technology at Boston Spa to enable and facilitate the increased production of historic newspaper content together with creating a state of readiness for capturing contemporary newspaper production in digital format.

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, Olive will export the texts to xml files for all of these processes, and align the xml with each digital image.

This project plan covers the phases of 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 refreshes 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 JISCII 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 1.1 million pages from regional and local newspapers from the 19th (and some from the 18th) century. These pages will be uploaded to the online database currently established by the Library in partnership with Gale Cengage. By the end of this project in March 2009, four million pages of 18<sup>th</sup> and 19<sup>th</sup> century newspapers will be available online for users. The matter of digitising some small runs of eighteenth century newspapers is being kept under review.

Early discussions are taking place with potential publishing partners to explore the possibility of further additions of historic newspapers to the Library collection.

### **3. Overall Approach**

#### **Strategy**

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

The Library possesses a newspaper collection that dates back to the 16<sup>th</sup> 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.

Central to this is the move of the content of the Newspaper Library and an associated building at Colindale to a new newspaper storage building at Boston Spa by end 2012 and the creation of future access to newspapers 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 digital and digitised newspapers. The Library also continues to purchase digital datasets of newspaper content from publishers to enhance our digital content.

The Library continues to invest in the Digital Object Management programme (DOM) and digital and digitised newspapers are priority content streams for this.

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**

As with JISC1, 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. Legal advice has been taken to ensure that copyright and licensing issues are respected.

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 1.1 million pages of 18<sup>th</sup> and 19<sup>th</sup> century newspapers and in its original proposal there were three main categories – 75% of regional and local titles mainly from the 19<sup>th</sup> century, filling gaps in the UK wide coverage from JISC 1.

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. There is also a question mark around the Irish titles being fit for purpose.<sup>2</sup>

The main workflow is as follows:

- Selection of content
- Preparation and microfilming
- Article zoning
- OCR
- Web hosting

### **Selection of content**

The primary focus of the project is on regional and local titles, predominately from the 19<sup>th</sup> 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.

---

<sup>2</sup> 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.

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

### **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 schema<sup>3</sup> and completed by the prep team when they review and/or repair each page. This new metadata is added to the manifest at issue level to enable rapid ingest into xml by Olive.

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

As in JISC 1 it was decided to refilm up to 75% 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.

Generally it can be said that the higher the resolution in the microfilm camera and the use of a digital image converted from a negative film contribute to an improved OCR accuracy. Exceptions of course can be found and the BL films comply with the highest quality standards.<sup>4</sup>

It is clear that the OCR process is affected by the reduction ratio in the camera, the quality of the original text, the font used and its size. What is important is whether the results are good enough for mass production, for OCR processing of large quantities of text.

### **Article Zoning**

Olive's digitisation process consists of the following steps:

1. Analysis of the source material

---

<sup>3</sup> 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.

<sup>4</sup> A reduction ratio of less than 16 (equals the position of how far the lens is from the original) is good enough for OCR. The project operates a reduction ratio of between 8 and 30 depending on the size of the original. It is also worth noting that you cannot use one reduction ratio for all papers and laser cut resolution targets give better quality targets and better quality resolution variables. Top range microfilm cameras should be able to cope with a wide range of newspapers.

2. Scan tuning and profiling. Olive builds profiles of different samples of the source material and multiple profiles may then be used on reels during scanning as necessary. This is where scan tuning can maximise OCR results
3. Scanning
4. Binding. All pages of the same issue are combined into a single issue and naming conventions are applied. JPEG images are named according to their page number, issue date and publication name
5. Manual QA
6. Image processing i.e. deskewing, cropping etc
7. Structure recognition in which a rough page structure is constructed through recognition of horizontal and vertical lines, text strings and image areas
8. Segmentation or zoning which employs artificial intelligence to automatically recognise components or entities such as articles, images and advertisements on a page
9. OCR run on all the content. Olive uses Advanced Predictive Fuzzy Logic Searching, to compensate for text inaccuracies
10. Manual QA, testing parameters of zoning, searchability and readability
11. Output to XML
12. Indexing. Each issue is indexed for searchability purposes
13. Final QA performed
14. Output of fully indexed XML repository to client

## **OCR Process**

Modern Optical Character Recognition (OCR) software provides a unique opportunity to extract the textual content from printed documents in a low cost automated manner. This computer readable text provides the opportunity for enhanced resource discovery techniques such as full text searching.

Whilst the current state-of-the-art in OCR technologies will provide significant added value to the outputs from the project it is accepted that there are limitations in the performance of OCR software when applied to historic texts.

The Library is committed to supporting research that will advance the state-of-the-art in the OCR of historic texts, particularly the complex issues which surround historic newspapers. As an indication of this commitment the Library is a participant in a European research project in the field of informatics for the humanities. The project, titled IMPACT, seeks to significantly improve the process of mass digitisation of historic texts. Three of the objectives of the project provide primary interest:

- Develop OCR software and technologies which exceed the accurateness of current state-of-the-art software significantly
- Provide a software system which will allow the realisation of new concepts of collaborative correction (in order to lower the costs for full featured full-text) by taking up and integrating the Web 2.0 phenomena
- Develop language tools and lexica in order to provide access to historical textx independently of historical variants of a given language

Whilst many of the outputs of the IMPACT project will be delivered beyond the duration of the JISC II Newspaper project, the Library is committed to sustaining its digital newspaper resource by applying, as appropriate, new technological advances that will improve the end user experience.

### **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 II project. 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 Library's DOM programme.

The BL'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 I. During JISC II the Library will work closely with the external hosting and serving partner to ensure the site keeps pace with current and emerging web tolls 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.<sup>5</sup>

### **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 output
- Computer readable text derived from OCR software provides the opportunity for enhanced resource discovery techniques such as full text searching
- The JISC 11 programme is central to the Library's newspaper strategy and benefits from the ability to draw programme support from all areas of the Library

---

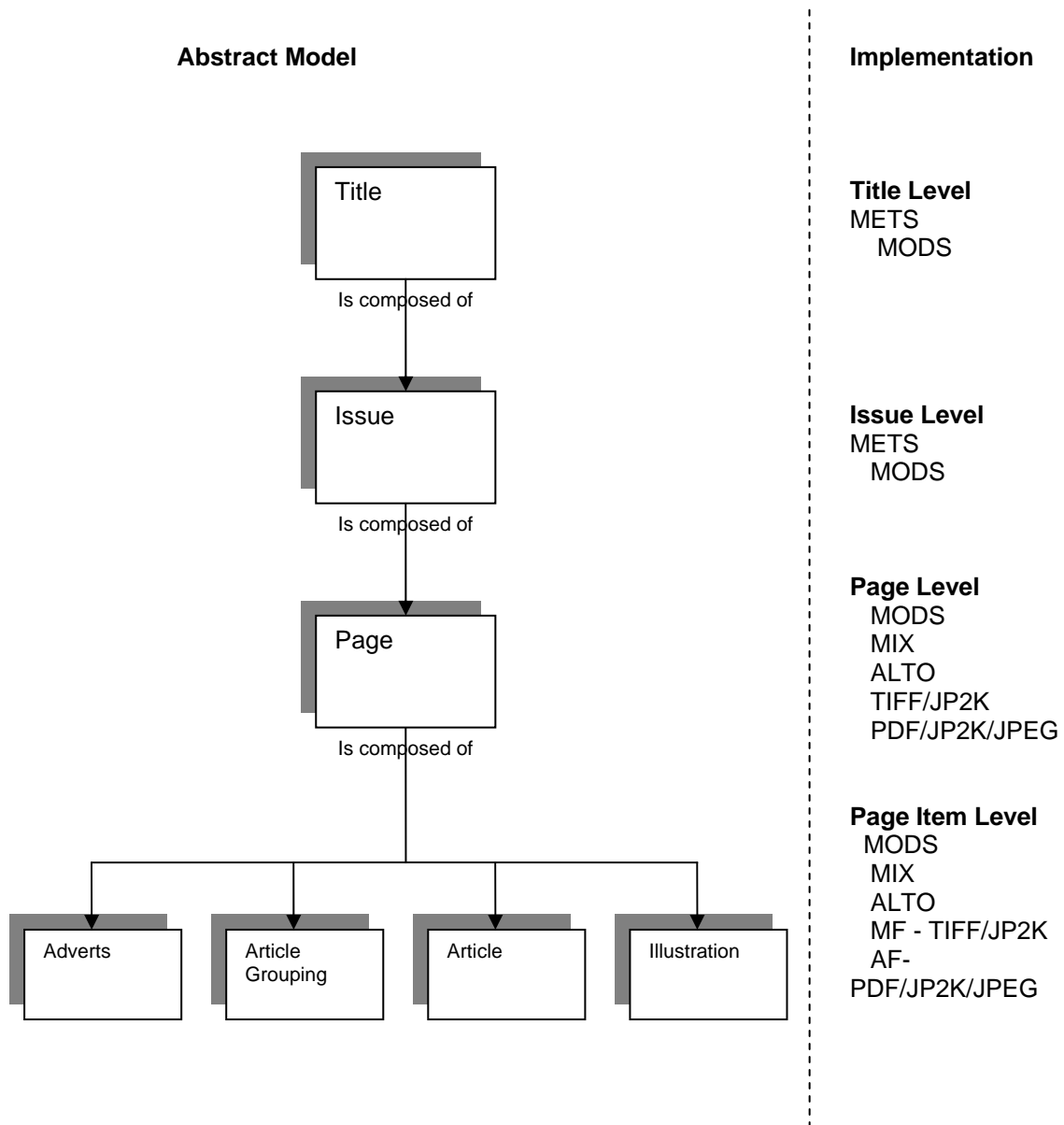
<sup>5</sup> BL proposal to the JISC for funding dated 16 August 2006, p5.

## **4. Project Outputs**

### **Engagement with the Material/Selection of Content**

31 newspapers have been identified to date and have been divided into seven 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.

## Metadata



METS<sup>6</sup> – XML Schema for encoding descriptive, administrative and structural metadata for digital objects.

MODS<sup>7</sup> – XML Schema for bibliographic elements.

<sup>6</sup> <http://www.loc.gov/standards/mets/>

<sup>7</sup> <http://www.loc.gov/standards/mods/>

MIX<sup>8</sup> – XML Schema for technical metadata related to digital still images.

ALTO<sup>9</sup> – Stores layout information and OCR recognized text

At the start of the project the Library will agree with the Supplier the detailed metadata elements to be used.

See Appendix for the project DTD schema agreed for this project.

## OCR<sup>10</sup>

One of the challenges of this project is to gain an understanding of how useful any produced OCR is to the end user.

Calculating OCR character and word accuracy<sup>11</sup> up until recently could not be automated and is extremely time consuming to do.<sup>12</sup> The main purpose of providing any text from these articles is to give the end user search and find capabilities and it is this, rather than character and word accuracies, that should be the focus of any study and future projects.<sup>13</sup>

We already know that early newspapers were not bound by orthographic rules which make 100% accuracy an unreasonable request. This does however; make advanced searchability of the platform a necessity.

There was always an option to utilise metadata to describe the source in more detail and include keywords which might be used for searching and this is the option that we have

---

<sup>8</sup> <http://www.loc.gov/standards/mix/>

<sup>9</sup> <http://www.ccs-gmbh.com/alto/>

<sup>10</sup> OCR: Optical Character Recognition: the process by which a type written source is converted into machine readable text.

<sup>11</sup> Word accuracy is generally a better approach when considering a text's suitability for searching as it only takes a single character error to render the entire word unrecognisable. It is possible to have 80% character accuracy and yet only 40 – 50% word accuracy.

<sup>12</sup> The JISC commissioned Simon Tanner, of Kings College Digital Services in September 2007 to undertake an independent evaluation of the OCR produced in JISC I and the Burney projects. His report is due out in April 2008, and he has developed an OCR analysis tool as part of this work which will be made available as a service from Kings College.

<sup>13</sup> Most OCR software can produce an OCR confidence level, for each character they read, usually from 0 to 9. 100% confidence score of 9 does not necessarily mean it has read the character correctly – it could read an 'e' as 'a' and be completely confident. There are two different ways of calculating an average confidence per word which lead to different results. Both methods are valid, one is easier than the other which is more real. One is to add all the individual scores and divide by the number of characters and the second which is more real is to just add the character passes and fails and average them. Depending on where you set the acceptability target level, one way could be a pass and the other a fail and the same applies at overall article or page level.

Project Acronym: JISC2 Newspapers  
Version: Final (5.1)  
Contact: Patrick Fleming/ Jane Shaw  
Date: 28.02

The metadata sheet that the preparation team uses to capture data follows the logic and structure of the xml in the project DTD allowing a more streamlined approach.

JISC II will benefit from Simons Tanners evaluation into OCR accuracy from JISC I, both in understanding ways to analyse OCR at word level but also, by utilising any software tool that he has developed as part of his evaluation.

### **Quality Assurance**

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.

These three are heavily reliant on human intervention. We discovered on both the JISC I and Burney projects that defining an article is not straightforward when dealing with historic newspapers that change frequency, layout and format over the length of their run.

Questions raised include: Does an article start and end with different fonts and capitalisation, or begin when font and layout change? What rules should be followed to allow a computer to consistently apply them when font size, punctuation and layout are constantly changing throughout the length of a run? How will a computer decide what subject category to apply to an article when many of them contain more than one subject? And are they really useful anyway for searching by an end-user? Someone has to read or scan the article and this is a huge task in time and cost.

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 1% sampling of Olives 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 the Library will use forensic QA on the Pilot and for the first three months of production and then tail this off.

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
- Development of tools for access and repurposing

Project Acronym: JISC2 Newspapers  
Version: Final (5.1)  
Contact: Patrick Fleming/ Jane Shaw  
Date: 28.02

- 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

The headline figures of Website usage stats for just October and November last year are as follows:

	October	November
Users	679	1,194
Page hits	54,107	170,177

Multiple users from the same institution show up as a single entry in the list.

## 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.

## 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

This project plan has one Gantt Chart the project outline updated since November 2007 and described as a Tracking Gantt. It includes Olive'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.

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	2	BL seeking to sign off selected titles with newspaper publisher organisations. Current search on copyright for JISC 2 titles	Open	16/11/07
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	3	BL, Gale & Olive work collaboratively on an agreed quality plan. Supplier has to pass the pilot before final contract is awarded.	Open	21/01/08

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	2	Consider using overtime or increasing resource allocation as a contingency.	Open	21/01-008
7	26-Apr-04	Loss of key staff and length of time to recruit is a minimum of 3 months.	Project Director	Medium	High	6	Ensure adequate handovers and swift, efficient recruitment.	Open	16/11-007
8	03-Aug-04	Loss or damage of material in transit, within BL or at suppliers	Project Manager and Team	Medium	Low	2	Insurance included in shipment costs. Redupe if lost or damaged by customs. Ongoing audits by project team. Supplier agrees to pay invoice for damaged films so they can be replaced.	Open	16/11-007
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	16/11-007
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 minimises us of BL acetate master microfilms. Metadata still has to be collected either from dupe or from originals.	Open	16/11-007
11	01-Jul-07	Preparation team failing to deliver work to Microfilm Studio, impacts 140K pages target for delivery to Olive by end January 2008.	Project Manager	Medium	Medium	4	140k pages completed and QA-ed. Ready to ship to Olive after successful completion of pilot.	Open	21/01-008
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	16/11-007
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	16/11-007

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	21/01-008
19	07-Jan-08	New requirements could delay schedule and increase project costs.	Project Board/SRO	High	High	9	BL determines scope of work and all parties agree and sign off; if Olive fail the pilot, we can move to the runner up CCS.	Open	21/01-008
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	21/01-008

## 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. <sup>14</sup> 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

<sup>14</sup> 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.

		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.<sup>15</sup>

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.

## 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

---

15

Project Acronym: JISC2 Newspapers  
Version: Final (5.1)  
Contact: Patrick Fleming/ Jane Shaw  
Date: 28.02

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 Gale Cengage in March 2007. The BL is in partnership with Gale to:

- Make the JISCI 19<sup>th</sup> 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

Contact at British Library for Gale contract: Ben White, Licensing & Copyright  
Compliance Manager  
The British Library  
96 Euston Road  
London NW1 2DB  
Tel. 020 7412 7307

Project Acronym: JISC2 Newspapers  
Version: Final (5.1)  
Contact: Patrick Fleming/ Jane Shaw  
Date: 28.02

## **Agreement between The British Library Board and Cengage Gale for The British Library Newspaper Platform**

Non-exclusive, non-transferable license

License made: 25/03/07

Duration: 7 years, to 24/03/14

Automatic-renew for one year terms, 12 months notice needed to from either party not to renew.

Gale has no copyright or IP rights except for the platform and supporting materials created or funded by them.

**Company details:** Cengage Learning, formerly Thomson Learning, delivers customised learning solutions for the HE and FE marketplace, as well as reference centers, government agencies, corporations and professionals around the world. They also have a growing programme of B2C activity. Gale, part of Cengage Learning, serves the world's information and education needs through its vast and dynamic content pools, which are used by students and consumers in their libraries, schools and on the Internet. It is best known for the accuracy, breadth and convenience of its data, addressing all types of information needs in a variety of formats, books and eBooks, databases and microfilm.

Gale is based in Farmington Hills, Michigan, with additional offices elsewhere in the U.S. and in the U.K., Australia, Germany, Switzerland and Malaysia.

The Library have a dedicated Relationship Manager. VP and Publisher / VP Brandon Nordin, we are also in weekly (and often more frequent) contact with Ray Bankovski, Gale's Vice President of Electronic Asset Management. Marketing and PR activities are run through Gale's London and Reading offices.

The Library signed a preliminary contract with Olive software in September 2007 to realise:

1. Scanning of newspaper pages from microfilm copies - includes greyscale scanning at 300 dpi with images produced as JPEG2000 files
2. Extraction of the content by OCR and content analysis
3. Production of XML metadata and service images for presentation on the web
4. Deliver completed, prescribed, metadata and images to the Library's web hosting organisation

Contact at British Library for Olive software contract: Alasdair Fraser, Head of BL Procurement  
The British Library  
Boston Spa  
Wetherby  
LS23 7BQ  
Tel. 01937 546460

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 Olive.

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]

**British Library  
 Newspaper and Digitisation Strategy**

**JISC senior-level  
 EXTERNAL**  
 Patrick Fleming  
 BL JISC Account Manager / HE  
 representation  
 John Tuck  
 JISC: Catherine Grout  
 JISC Alasdair Dunning

**Programme Board: JISC Newspapers and Gale Platform  
 INTERNAL**  
 SRO: Patrick  
 Newspaper Programme Implementation Manager: TBC  
 Head of Imaging Services: Andrew Austin  
 Project Manager: Jane Shaw  
 Head of British Collections: John Tuck  
 Product Development inc.Gale: Hannah Porter  
 Head of HE: Ian Snowley  
 Finance: Lydia Devenny  
 Head of Web Services: Adrian Arthur

**Vendor / Partner Meetings  
 EXTERNAL**  
 Gale Relationship  
 Patrick Fleming (as needed)  
 Hannah Porter  
 Andrew Austin  
 Gale: Brandon Nordin and Ray Bankovski (etc)  
 Olive Relationship  
 Patrick Fleming (as needed)  
 Andrew Austin  
 Jane Shaw  
 UK Marketing  
 Hannah Porter  
 Joanne Newman / Ginevra House  
 Other markets inc education: TBC  
 Lawrence Christensen  
 Gale UK marketing and PR: Rossella Proscia and Sophie  
 Villiers

**1. Risk Meeting**  
 Patrick Fleming, John Tuck, Jane Shaw, Andrew Austin, Hannah Porter and  
 Adrian Arthur

**2. JISC Newspapers Production Project Board  
 EXTERNAL**  
 SRO: Patrick Fleming  
 Operational standards: Andrew Austin  
 Production project manager: Jane Shaw  
 Digitisation standards: Aly Conteh  
 Internal digitisation: Martyn Lunn  
 Stakeholder (Gale) representative: Hannah Porter  
 Stakeholder (JISC): Alasdair Dunning

**3. Content Management Board  
 INTERNAL**  
 SRO: Hannah Porter  
 Operational standards: Andrew Austin  
 Production project manager: Jane Shaw  
 Editorial: Moira Goff and Ed King  
 Website design and branding: Adrian Arthur  
 Marketing HE: Ginevra House  
 PR: Lawrence Christensen (as needed)

**4. Digitisation Vendor Liaison Teleconference  
 EXTERNAL**  
 Aly Conteh  
 Andrew Austin  
 Jane Shaw

**5. Weekly Output Liaison Teleconference  
 EXTERNAL**  
 Chair: Adrian Arthur  
 Aly Conteh  
 Hannah Porter  
 Andrew Austin  
 Digitisation vendor rep (As needed)

Project Acronym: JISC 2 Newspapers  
Version: final (5.1)  
Contact: Jane Shaw/Patrick Fleming  
Date: 28.02.2008

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

The project manager has an informal link to the Gale Cengage relationship 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  
Mohammed Jaffer, Project Analyst 020 7412  
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  
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 March 2008, the team will be increased by 2 extra recruits for 9 months only to assist on preparation and QA.

All of the project team including the project manager have training needs. The Quality Assurance Officers would benefit from more involvement in O & S directorate generally, plus developing new skills, such as minute and note taking and improving their computer skills. Many have an interest in continuous improvement in areas such as website design and conservation/preservation techniques.

Due to the matrix reporting structure around the project, the project manager would benefit from attending courses on facilitation and programme management.

The costs for these courses will be met by the Library training budget.

### **13. Programme Support**

None at present.

### **14. Budget**

[See Appendix A](#)

Project Acronym: JISC2 Newspapers  
Version: Final (5.1)  
Contact: Patrick Fleming/ Jane Shaw  
Date: 28.02

## ***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.

Project Acronym: JISC2 Newspapers  
 Version: Final (5.1)  
 Contact: Patrick Fleming/ Jane Shaw  
 Date: 28.02

We have developed a Production Forecast Formula which models changes in staff numbers, operating capacity and average production each day. By altering any of these three variables we can plan better for the future.

This chart provides an average hourly production rate per person.  
The benefit here is that it helps management in planning for  
future requirement at productivity side.

### **Production Forecast**

Average Hourly Production	122
No of productive hours (excluding idle time)	5
Average production per person per day	610
No of employees	7.5
Average Production per day	4575
Operating Capacity (Considering holidays)	75%
Average Expected Production per day	3424.845

### **Expected Production**

<u>Month</u>	<u>No Of Working Days</u>	<u>Expected Output</u>
February	21	71922
March	19	65072
April	20	68497
May	22	75347
Output expected in 4 months with 5.5 staff		280837

## **16. Evaluation Plan**

We would consider the use of independent evaluators for image quality of the scans and for OCR accuracy as done on the first project. We plan to use Focus Groups for further evaluation of the website.

<b>Timing</b>	<b>Factor to Evaluate</b>	<b>Questions to Address</b>	<b>Method(s)</b>	<b>Measure of Success</b>
03-04/2008	QA Pilot production	Quality conforms to contract?	Online QA	tbc
- 03/09	QA	Ditto	Online Qa	tbc

## 17. Quality Plan

QA will be carried out at key points in the process – at preparation and microfilming, prior to shipment to Olive, during processing at Olive, 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.

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, Olive and Gale, with two months in hand will allow for easier rework if needed, smoother billing and capacity plans.

The digitisation supplier Olive software 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 NPI 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</i>		

<i>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 &amp; 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 Olive 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 sound 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 1

- 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

The Quality Improvement Agency QIA event 10-11th December at The University of Hull was attended by ASR2 engagement officer on our behalf:

Purpose of meeting: to work with Further Education teachers to begin developing learning packages utilising JISC-funded online resources.

This was the second event organised by the QIA. In the first event, 6 online resources were presented to participants, who began thinking about the type of learning package they wanted to create and how they might use the available resources. The project manager will attend the next workshop taking place on 17 March 2008, as she has a particular interest in FE take up of the JISC 1 resource.

## 19. Exit and Sustainability Plans

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 for these files to be ingested into its digital object management (DOM) programme, to ensure preservation	Arrange with BL DOM

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

Project Acronym: JISC2 Newspapers  
Version: Final (5.1)  
Contact: Patrick Fleming/ Jane Shaw  
Date: 28.02





## ***Appendixes***

### **Appendix A**

The total grant awarded to the project by JISC is £1,143,000.

## Appendix B. Workpackage

MASTER PROGRESS SHEET  
British Newspapers 1620 - 1900  
As At- 15 - February - 2008

Progress Bar	Condition Of Title
Complete 	Good   A
Incomplete 	Average   B
Set-aside 	Poor   C

<u>ENGLAND CENTRAL Work Package One</u>										
Title	Normalised Title	Abbreviated Title	Language	British Library 1800 - 1900 Years to be completed	Est'd No. Of Pages	Volumes	Actual No. Of Pages Found	Condition	Volumes Unavailable	Y
Cheshire Observer And General Advertiser For Cheshire and North Wales	Cheshire Observer	CHOR	ENGLISH	13 May 1854 to 1900	19136	44	18505	B	1880, 1898	
Leicester Chronicle, or Commercial and Agricultural Advertiser	Leicester Chronicle	LECH	ENGLISH	23 March 1811 to 1900	30057	70	27672	B	1811-1826, 1848, 1850, 1873,	
Nottinghamshire Guardian and Midland Counties Advertiser	Nottinghamshire Guardian	NOGN	ENGLISH	1 May 1846 to 1900	26172	51	25829	A	1848, 1872, 1888, 1897	18 18

Berrows Worcester Journal	Berrows Worcester Journal	WOJL	ENGLISH	6 Jan 1808 to 1900					1815, 1817, 1864, 1871, 1888, 1893, 1897, 1898, 1899	18
					27586	84	27129	B		
					102951	249	99135			
					Avg Pgs Per Vol		398			
					Ave Vols Per Title		62			
					Avg Pgs Per Title		24784			

<u>ENGLAND NORTH Work Package Two</u>										
Title	Normalised Title	Abbreviated Title	Language	British Library 1800 - 1900 Years to be completed	Est'd No. Of Pages	Volumes	Actual No. Of Pages Found	Condition	Volumes Unavailable	Y
Bradford Observer	Bradford Observer	BROR	ENGLISH	6 Feb 1834 to 1900	73840	94	18322	C	1835, 1836, 1851, 1864, 1867, 1871, 1872	18
Blackburn Standard	Blackburn Standard	BLSO	ENGLISH	21 Jan 1835 to 1900	19060	63	16688	A	1841-1844, 1870, 1871, 1873, 1874, 1883, 1885, 1896, 1897	18 18 18 18
Isle Of Man Times And General Advertiser	Isle Of Man Times	IMTS	ENGLISH	17 Jul 1869 to 1900	12480	26	12411	A	1870, 1871, 1873, 1882, 1885, 1896	18 18 18
Lancaster Gazette, And General Advertiser For Lancashire West	Lancaster Gazette	LAGE R	ENGLISH	20 June 1801 to 1900	32526	65	32727	A	1837, 1839, 1840, 1848, 1849, 1886, 1895 - 1900	18
Daily Gazette For Middlesbrough, Stockton and District	Daily Gazette For Middlesbrough	DGMH	ENGLISH	15 Nov 1869 to 1900	38192	44	0	B		

Sheffield Independent and Yorkshire and Derbyshire Advertiser	Sheffield Independent	SHIN	ENGLISH	11 Dec 1819 to 1900	88844	155	72128	A	1819-1830, 1875 (jul-dec), 1881, 1882 (oct-dec) 1885 (jan-apr), 1887, 1894, 1897, 1898
York Herald	York Herald	YOHD	ENGLISH	3 Jan 1801 to 1900	97852	148	81868	B	1802, 1821, 1822, 1827, 1897, 1898
Huddersfield Chronicle and West Yorkshire Advertiser	Huddersfield Chronicle	HUCE	ENGLISH	6 April 1850 to 1900	22932	73	0	A	
					385726	668	234144		
					Avg Pgs Per Vol		351		
					Ave Vols Per Title		84		
					Avg Pgs Per Title		58536		

<a href="#">ENGLAND SOUTH Work Package Three</a>										
Title	Normalised Title	Abbreviated Title	Language	British Library 1800 - 1900 Years to be completed	Est'd No. Of Pages	Volumes	Actual No. Of Pages Found	Condition	Volumes Unavailable	Y
Bury and Norwich Post	Bury and Norwich Post	BNPT	ENGLISH	1800 to 1900	28741	78	0	A		
Cambridge Chronicle and Journal	Cambridge Chronicle	CACE	ENGLISH	3 May 1811 to 1900	29120	77	0	B		
Essex Standard	Essex Standard	ESSD	ENGLISH	7 Jan 1831 to 1900	19344	67	0	B		
Isle Of Wight Observer	Isle Of Wight Observer	IWOR	ENGLISH	4 Sept 1852 to 1900	11520	45	0	A		
Northampton Mercury	Northampton Mercury	NOMY	ENGLISH	1800 to 1900	20375	81	0	A		
Norfolk Chronicle Or The Norwich Gazette	Norfolk Chronicle	NOCE	ENGLISH	5 June 1802 to 1900	24856	85	0	A		

Chronique (Chroniques) de Jersey	Chronique (Chroniques) de Jersey	CHJY	ENGLISH	1 Jan 1814 to 1900	17566	50	0	B	
Star [Guernsey]	Star [Guernsey]	STGY	ENGLISH	17 Jul 1869 to 1900	18096	29	0	B	
Southampton Herald and Isle Of Wight Gazette	Southampton Herald	SOHD	ENGLISH	25 Jul 1823 to 1900	34976	69	32283	B	1823-1829,1833,1850(jul-Dec), 1853, 1864 (jul-Dec), 1866(jul-Dec)1896,1898,
Royal Cornwall Gazette, Falmouth Packet and Plymouth Journal	Royal Cornwall Gazette	COGE	ENGLISH	2 Jul 1803 to 1900	31132	74	25734	B	1803-1810,1838,1866,1889,1897
					235726	655	58017		
					Avg Pgs Per Vol		89		
					Ave Vols Per Title		66		
					Avg Pgs Per Title		14504		

<b>IRELAND Work Package Four</b>										
Title	Normalised Title	Abbreviated Titles	Language	British Library 1800 - 1900 Years to be completed	Est'd No. Of Pages	Volumes	Actual No. Of Pages Found	Condition	Volumes Unavailable	Year
Constitution or Cork Morning Post	Constitution or Cork Advertiser	CNCK	ENGLISH	18 Aug 1823 to 1900	52840	113	0	C		
Derry Journal	Derry Journal	DEJL	ENGLISH	15 Feb 1825 to 1900	39200	71	0	C		
Kerry Evening Post	Kerry Evening Post	KEPT	ENGLISH	24 May 1813 to 1900	30160	74	0	C		
					122200	258	0			
					Avg Pgs Per Vol		0			
					Ave Vols Per Title		86			

	Avg Pgs Per Title		0
--	----------------------	--	---

**SCOTLAND Work Package Five**

Title	Normalised Title	Abbreviated Titles	Language	British Library 1800 - 1900 Years to be completed	Est'd No. Of Pages	Volumes	Actual No. Of Pages Found	Condition	Volumes Unavailable	Y
Dundee Courier	Dundee Courier	DUCR	ENGLISH	29 Oct 1839 - 1900	62364	90	0	B		
Inverness Courier and General Advertiser For the Counties of Inverness	Inverness Courier	INCR	ENGLISH	4 Dec 1817 to 1900	29120	58	0	B		
					91484	148	0			
					Avg Pgs Per Vol		0			
					Ave Vols Per Title		74			
					Avg Pgs Per Title		0			

**WALES Work Package Six**

Title	Normalised Title	Abbreviated Titles	Language	British Library 1800 - 1900 Years to be completed	Est'd No. Of Pages	Volumes	Actual No. Of Pages Found	Condition	Volumes Unavailable	Y
Cambrian [Swansea]	The Cambrian [Swansea]	CASA	WELSH & ENGLISH	10 Apr 1813 to 1900	18512	59	0	B		
Wrexham Weekly Advertiser	Wrexham Weekly Advertiser	WRWA	WELSH & ENGLISH	March 1854 to 1900	16432	45	0	A		
					34944	104	0			
					Avg Pgs Per Vol		0			
					Ave Vols Per Title		52			
					Avg Pgs Per Title		0			

**LONDON PAPERS Package Seven**

Project Acronym: JISC2 Newspapers  
 Version: Final (5.1)  
 Contact: Patrick Fleming/ Jane Shaw  
 Date: 28.02

<u>Title</u>	<u>Normalised Title</u>	<u>Abbreviated Titles</u>	<u>Language</u>	<u>British Library 1800 - 1900 Years to be completed</u>	<u>Est'd No. Of Pages</u>	<u>Volumes</u>	<u>Actual No. Of Pages Found</u>	<u>Condition</u>	<u>Volumes Unavailable</u>	<u>Y</u>
The Standard	The Standard	SDLN	ENGLISH	21 May 1829 to 1900	95450	137				
Morning Post	Morning Post	MOPT	ENGLISH	8 Jan 1803 to 1900	97840	141				
					193290	278	0			
					Avg Pgs Per Vol		0			
					Ave Vols Per Title		139			
					Avg Pgs Per Title		0			

Project Acronym: JISC 2 Newspapers  
Version: final (5.1)  
Contact: Jane Shaw/Patrick Fleming  
Date: 28.02.2008

## **Appendix C Project Plan**

**Sent under separate cover**