

**JISC CAPITAL PROGRAMME**  
**Project Document Cover Sheet**  
***PROJECT PLAN***

## Project

<b>Project Acronym</b>	Welsh Journals Online	<b>Project ID</b>	
<b>Project Title</b>	Welsh Journals Online / Cylchgronau Cymru		
<b>Start Date</b>	March 2007	<b>End Date</b>	February 2009
<b>Lead Institution</b>	National Library of Wales		
<b>Project Director</b>	Andrew Green		
<b>Project Manager &amp; contact details</b>	Martin Locock 01970 632885 martin.locock@llgc.org.uk		
<b>Partner Institutions</b>	n/a		
<b>Project Web URL</b>	welshjournals.llgc.org.uk cylchgronau.cymru.llgc.org.uk		
<b>Programme Name (and number)</b>	JISC Digitisation - Phase Two		
<b>Programme Manager</b>	Paola Marchionni		

## Document

<b>Document Title</b>	<i>Project Plan</i>		
<b>Reporting Period</b>			
<b>Author(s) &amp; project role</b>	Martin Locock, Project Manger		
<b>Date</b>	11/6/07	<b>Filename</b>	Project_plan v1.1

<b>URL</b>		
<b>Access</b>	<input checked="" type="checkbox"/> Project and JISC internal	<input type="checkbox"/> General dissemination

## Document History

Version	Date	Comments
1.0	4/5/07	
1.1	25/6/07	Addition of details throughout



## JISC Project Plan Template

### *Overview of Project*

#### 1. Background

The National Library of Wales holds the world's largest collection of periodicals relating to Wales and the Welsh language, including scholarly, literary and cultural, popular, educational and news titles. The material reflects the life of Wales through a wide spectrum of subjects and activities, of potential interest to researchers, students and casual users. NLW is seeking to exploit the potential of mass-digitization of printed material in order to allow free access to all to this material over the Internet; previously Wales and Welsh-language material has been poorly represented in digitisation programmes.

The JISC-funded project will focus on a selection of 90 journals published from 1900 onwards to create a resource of 600,000 pages of searchable text. Approximately 40% of the content will be Welsh language, forming the single largest corpus of Welsh material on the Web. The project will form the cornerstone of the NLW's long-term intention to digitise its entire printed holdings.

#### 2. Aims and Objectives

*Aim:*

To provide remote access to the contents of the main journals relating to Wales

*Specific objectives:*

- To digitise approximately 90 journal titles comprising 600,000 pages of text
- To create page scans and OCR text files for the journals
- To secure clearance from publishers and copyright holders to allow public access to the material and to obscure articles and images for which rights are withheld
- To provide browse, word-search, article title and author search access to the content
- To allow page-scan view and, where agreed, TEI text view
- To provide subject guides and advice to users
- To promote use of the content in learning objects
- To allow users to contribute content

### 3. Overall Approach

Titles will be selected for inclusion based on the recommendations of appropriate NLW staff; the project team will enhance the bibliographic metadata held in the Library's Virtua catalogue prior to its extraction into the project database, where structural metadata will be prepared to inform and control the scanning of each page.

The scanning will be undertaken using two Zeutschel 5000TT Bookscanners; the initial scans will be processed by Omnican v11 software, to create a greyscale TIFF file at 400 ppi. The scans will then be processed to create compressed JPEG/PNG files for presentation.

After checking, the page scans will then be submitted to an external contractor for OCR, and upon its return this data will be incorporated into TEI documents alongside the METS documents derived from the metadata. The main unit of operation will be the journal title; scanning will be managed by physical volume and page, linked to the publication record by printed issue and article.

One of the main concerns of the project will be IPR management, since the majority of the material will be in copyright. Publishers of the journals will be approached for their permission and asked to obtain the consent of their contributors. Where necessary assistance will be provided to publishers in undertaking this work. In addition, umbrella licences will be sought from bodies likely to hold rights in many articles, and particularly images. All copyright material whose holders require a payment for licensing will be excluded.

The web interface will allow users to explore freely the content by searching and browsing. Word-searching of the OCR text will allow the location of phrases and names. In addition, authority-controlled article author search and word search within article title will be enabled. It is planned that fuzzy matching of placenames for searching will be included. (This approach will provide similar functionality to that proposed in the bid without the need for extensive manual editing of the TEI).

User Generated Content will be hosted as comments associated with a specific page of content, so that users will be able to see them in context.

At title level, subject terms and descriptions will be provided, as well as introductions to sources for specific subjects.

The critical success factors will be the speed and efficiency of metadata creation, reliability of scanning and QA, effective rights management, quality of the user interface, and publicity and search engine visibility of content.

## 4. Project Outputs

*List the tangible deliverables (including reports) your project will create, and the less tangible knowledge and experience you hope to build and share.*

1. Website
2. Web content and underlying METS data and TEI files
3. VLE object links resource
4. Evaluation report
5. Final project report
6. Licence agreements with publishers for past content
7. Agreements for future volumes
8. Software applications for data creation and maintenance
9. User-generated content application
10. Guidance to publishers and others on IPR issues
11. Teaching & Learning pack
  
12. Expertise in managing rights
13. Expertise in metadata and scanning for mass digitisation

## 5. Project Outcomes

*List the outcomes you envisage, including their impact on the teaching, learning, or research communities, and what change they will stimulate or enable.*

The creation of the web content will have a transformative effect on access to knowledge relating to Wales:

It will present the single largest corpus of Welsh-language text available on the Internet

It will provide easy access to obscure, forgotten and complex text content through searching

It will provide simple cross-searching across a range of titles of varying dates simplifying historiographical, audience reception and regional studies

It will simplify the study of individual authors and their contributions to knowledge, and the exploration of links between individuals

It will enable a ready appreciation of the bibliographical history of Wales and the social, academic and cultural activities of the publishing societies

It will provide access to a range of journalism, academic writing, and literature dealing with a single topic

It will provide a source of page-image and illustration content for inclusion in VLE objects and other secondary uses

Thus it will in general open up the material to users with a casual interest or engaged in formal research.

Specific outcomes will include:

Use of relevant Welsh language text by users not familiar enough with the language to locate relevant content previously

Use of specialist content by general users  
 Use of wide range of example material in teaching materials and presentations encouraging students to explore further  
 Use as a primary tool for research and coursework by students

## 6. Stakeholder Analysis

Stakeholder	Interest / stake	Importance
NLW	Part of service to public allowing access to its holdings	High
JISC	Delivery of Digitisation Programme and improved resource for educational/public	High
Publishers	Impact on paper/electronic publication and development of web resources	Low
Educational users	Easy access to resources online, including VLE objects	High
General users	Easy access to resources online	Medium
Authors/ creators / copyright holders	Impact on copyright assets' value and use (may be positive or negative)	Medium
HE / FE bodies	Use within VLE; effect on library holdings and e-journal subscriptions	Medium

## 7. Risk Analysis

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Timescale unrealistic	3	5	15	Review/monitor
Lack of suitable staff	1	5	5	Vary job spec if needed
Loss of key staff	3	3	9	Share skills; train deputies
Poor workflow documentation	1	1	1	Improve documentation
Low staff motivation for repetitive tasks	5	3	15	Training, team building

Lack of specialist skills	1	3	3	Training
Equipment/software failure	3	5	15	Maintenance contracts
Financial unsustainability of archive / delivery	1	5	5	Seek funding
Inadequate IT infrastructure and support	1	3	3	Review/monitor
IPR restrictions	3	5	15	Negotiate with publishers
Low OCR accuracy	3	5	15	Pilot, monitor
Theft and misuse of images	5	3	15	Enforce user Terms and Conditions
Archival files unsuitable for preservation	1	3	3	Monitor developments
Poor service delivery	1	1	1	Review customer needs and amend delivery
Departure from metadata and technical standards	1	3	3	QA and monitoring within workflow
Publicity ineffective	3	3	9	Review customer profile and PR strategy; evaluate

## 8. Standards

Name of standard or specification	Version	Notes
TEI level 2	2.1	
METS	1.5	
MIX	1.0	Version 1.0 released January 2007
TextMD	1.0	
Premis	1.1	
WAI A Priority 1	1.0	
MODS	3.0	

## 9. Technical Development

The Project Database and Workflow systems used to hold the metadata will be developed based on open systems with the data structure mapped to MARC21, METS and MODS.

The METS will be held within NLW's Vital system, which has proven effective as a web delivery tool.

The website will conform to best practice in terms of accessibility and cross-browser functionality following Transitional or Strict HTML.

The data will be archived for preservation using Optical Disks, held as part of NLW's Digital Archive.

## 10. Intellectual Property Rights

All IPR in the new digital data (scanned images of pages and OCR text) will be owned by NLW.

Third-party content will comprise copyright over the physical layout of the published (paper) edition and copyright of the creative content contained in the journals scanned. The use of third-party content will be governed by a licence agreement between each journal publisher and NLW, underlying which are individual agreements (consent letters) between the publisher and copyright holders. Where no agreement can be reached with the publisher, the journal title will not be included in the project. Where the publisher is unable to secure the consent of copyright holders for specific parts of the published work, these will appear on the website blanked out.

Publishers are required to provide a warranty to NLW that they have taken reasonable steps to contact copyright holders and to identify any for which permission was refused; these reasonable steps will include publishing information about the project's proposal in the journal, newsletter etc, and the direct mailing of contributors where address lists exist. Where the publisher is unable to undertake this mailing themselves, the project team will assist by doing so on their behalf.

The warranty to NLW constitutes the authorisation to digitise, and will specify the rights users will be granted. Should copyright holders subsequently withdraw their consent, or come forward without having previously been consulted, NLW will take down the material as soon as is practicable.

The licence agreements and where provided the consent letters will form part of the project records held at NLW to be used in the event of future disputes.

Blanket agreements will be sought from rights holders likely to be represented in numerous titles or where (in the case of literary estates) permission may be hard to obtain.

## **Project Resources**

### **11. Project Partners**

The project is being undertaken by NLW.

The project will be assisted by members of WHELF (Welsh Higher Education Libraries Forum) through the Project Advisory Board; the assistance is to comprise advice and support, and the loan of material to fill gaps in NLW holdings.

Subcontractors will be appointed to deliver:

- Software development
- Website design
- Creation of OCR output from images

### **12. Project Management**

#### **Project management framework**

The project team will be managed by the Project Manager, reporting to the Head of Digital Developments. The project is overseen by a Project Management Board comprising representatives of Library Sections affected by the project, chaired by Director of Collection Services.

A Project Advisory Board, chaired by the Librarian, is attended by representatives of WHELF bodies (universities in Wales), JISC, FE, and the Welsh Books Council. Advisory Board members provide a link with the main user communities.

A Technical Panel reports to the PMB and considers issues relating to systems development, interoperability and metadata standards.

#### **Project team**

Martin Locock	Project Manager	x885	(100% Project Management)	mlc@llgc.org.uk
Gwenno Dafydd	Liaison Officer	x969		gda@llgc.org.uk
Branwen Rhys	Metadata Officer	x967		bmi@llgc.org.uk
Iris Thompson	Metadata Assistant	x326		elt@llgc.org.uk
(tba)	Metadata Assistant			
Dan Field	System developer	x582		dof@llgc.org.uk

Tba	Scanning technicians (x 2)	
Tba	Scanning QA	
Sebastian Clark	IT support officer	x422
Tba	IT support assistant	

scl@llgc.org.uk

**Training needs:**

NLW cataloguing, METS, system usage	NLW staff
IPR and Copyright	External courses

### 13. Programme Support

Legal advice on the forms of Model Licences, including 3<sup>rd</sup> party licences, consent forms etc.

Shared best practice on technical specifications for OCR, and generally for QA.

Guidance on evaluation and usability surveys.

Guidance on VLEs.

Guidance on User Generated Content applications and implementation issues.

Guidance on making Open Source applications available to others.

### 14. Budget

The total grant awarded to this project by JISC is £588,465.

## Detailed Project Planning

### 15. Workpackages

W P	[Months>]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	Proj. Mgt																								
2	Identification																								
3	Digitisation																								
4	OCR																								
5	QA																								
6	Preservation																								
7	Web Delivery																								
8	Evaluation																								
9	Marketing																								
10	T&L Pack																								
11	IPR																								

Month 1= March 2007

Month 24 = February 2009

**Note:** Activities on the project will run Jan 2007-March 2009; those outside the JISC funding window will be funded by NLW as part of its contribution to the project cost.

### 16. Evaluation Plan

Indicate how you will evaluate the quality of the project outputs and the success of the project. List the factors you plan to evaluate, questions the evaluation will answer, methods you will use, and how success will be measured. Expand as appropriate on how you will conduct the evaluation.

Timing	Factor to Evaluate	Questions to Address	Method(s)	Measure of Success
March 2008	Website usability	Ease of use; documentation; reliability	Usability test by sample users; opt-in feedback	If less than 80% successful/satisfied, redesign
February	Website and UGC	Coherent; complete;	User panel	All areas of concern lie

2009	content	understandable	and advisory board feedback	outside core deliverables
January 2009	Marketing/ awareness	Degree of awareness in target groups	Commissioned research	HE users: 10% aware General: 2% aware
February 2009	Project success as a whole	Compare to aims and objectives	Commissioned research: stakeholders	All stakeholders 50% satisfied or very satisfied Key stakeholders 75% satisfied or very satisfied
January 2009	Data quality	Integrity of METS and TEI files	QA check by Technical Panel	95% files ok; 100% files structure correct
January 2009	VLE and T&L pack resources	Useful to HE/FE community	Advisory Board feedback	Board approves content

## 17. Quality Plan

Output	Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools
Website	Feb 2008	Usability; WAI A compliance	Usability test; feedback; automatic validation	Panel results; - ; validation results	Project manager; web designer	
Data: scans	(throughout)	Sequence check; quality check	File check; sample check	(in Workflow program)	QA assistant	Workflow system
Data: metadata	(throughout)	Complete; accurate	Sample check at multiple stages	(in Workflow program)	Metadata Officer	Workflow system

Data: OCR	(throughout)	Accuracy within reliability limit (98% character)	Part of OCR process	Report	QA assistant	
-----------	--------------	---	---------------------	--------	--------------	--

## 18. Dissemination Plan

Timing	Dissemination Activity	Audience	Purpose	Key Message
March 2008	Leaflet circulation	General users	Raise profile of live data	Website live
Summer 2008 (?) <sup>1</sup>	Academic conference	HE/FE staff, Information Science specialists	Raise profile within HE//FE community	Vital research tool, large corpus of Welsh language content
Autumn 2008	Regional roadshow events	Students/public	Raise profile with users	Available now, free, for all
Spring 2009	Launch	Publishers and other project partners	Maintain links and set out future	Website will continue to grow
(throughout project)	Press releases	Media > general public	Encourage new users	Project and website are for all

## 19. Exit and Sustainability Plans

Project Outputs	Action for Take-up & Embedding	Action for Exit
Project archive	Transfer to NLW Project archive	Arrange and document

<sup>1</sup> Following discussion with Advisory Board, the project now proposes to send speakers to other conferences to reach the academic audience rather than holding a single conference for the project. Further details to be provided.

Project Outputs	Why Sustainable	Scenarios for Taking Forward	Issues to Address
Website	Forms part of NLW web presence	NLW core task	Documentation/handover, inclusion in forward plan
Rights management	NLW act as caretaker for future changes	NLW core task	Information path and ability to amend and regenerate rights controls
Future issues of publications	NLW committed to future inclusion	NLW core task	Establish workflow for scanning and release

## Appendixes

### Appendix B. Workpackages



W P	[Months>]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	Proj. Mgt																								
2	Identification																								
3	Digitisation																								
4	OCR																								
5	QA																								
6	Preservation																								
7	Web Delivery																								
8	Evaluation																								
9	Marketing																								
10	T&L Pack																								
11	IPR																								

Project start date: 24 March 2007

Project completion date: 27 February 2009

Duration: 24 months

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Milestone	Responsibility
<b>YEAR 1</b>					
<b>WORKPACKAGE 1: Project Management</b>					
<b><u>Objective:</u></b>					
1. Coordinate team activities	24/3/07	27/2/09			ML
2. Report to Project Boards and JISC	24/3/07	27/2/09	<b>Progress report</b> to Project Management Board (monthly) <b>Progress report</b> to Project Advisory Board (quarterly) <b>Progress report</b> to JISC (six-monthly)	<b>Sept 2007</b> <b>March 2008</b> <b>Sept 2008</b> <b>March 2009</b>	ML
3. Procurement	24/3/07	27/2/09	System development; OCR contract; Website; Evaluation		ML
4. Budgeting	24/3/07	27/2/09	Predicted and actual expenditure accounts	<b>Mar 2007</b> <b>Sept 2007</b> <b>Mar 2008</b> <b>Sept 2008</b> <b>Feb 2009</b>	ML/ FD
5. Training and staff development	24/3/07	27/2/09	Probationary Year assessments Annual Performance Appraisals		ML

6. Documentation	24/3/07	27/2/09	JISC Project Plan Boards Terms of reference Reports to Boards	12/05/07 12/05/07	ML
7. Tracking of progress	24/3/07	27/2/09	Revised programme		ML
<b>WORKPACKAGE 2: Identification</b>					
<u>Objective:</u>					
8. Draft initial list of titles for inclusion	24/3/07	1/4/07	Initial list of titles	1/4/07	ML/GD
9. Research contact addresses and approach	1/4/07	8/4/07		8/4/07	GD
10. Prioritise reserve titles	8/4/07	1/5/07	Reserve list of titles	1/5/07	Advisory Board/ GD
11. Research contact addresses and approach	1/5/07	25/6/07		25/6/07	GD
<b>WORKPACKAGE 3: Digitisation</b>					
<u>Objective:</u>					
12. Prepare bibliographic metadata	25/5/07	30/12/08	Completed metadata by title	Completion of each title	BR
13. Prepare structural metadata	25/6/07	30/12/08	Completed metadata by title	Completion of each title	BR
14. Scan titles	10/7/07	30/12/08	Completed scans by title	Completion of each title	SW
15. Generate METS	10/10/07	25/02/09	Completed METS by title	Completion of each title	BR

<b>WORKPACKAGE 4: OCR</b>					
<u>Objective:</u>					
16. Send scan files to contractor	10/10/07	30/12/08	Batch of files dispatched	Batches sent	SW
17. Ingest OCR files	10/12/07	1/2/09	Batch of files ingested	Batches ingested	BR
<b>WORKPACKAGE 5: QA</b>					
<u>Objective:</u>					
18. Technical QA on images	10/7/07	10/1/09	Checked images	[Through-out]	SW
19. QA on metadata	25/6/07	20/1/09	Checked metadata	[Through-out]	BR
<b>WORKPACKAGE 6: Preservation</b>					
<u>Objective:</u>					
20. High resolution scans to archive	10/7/07	10/1/09	Digital archive	Complete 10/1/09	SW
21. Web scan files to archive	10/7/07	10/1/09	Digital archive	Complete 10/1/09	SW
22. Metadata files to archive	10/10/07	20/2/09	Digital archive	Complete 20/2/09	BR
23. System files to archive	1/1/08	20/2/09	Digital archive	Complete 20/2/09	ML
<b>YEAR 2</b>					

<b>WORKPACKAGE 7: Web delivery</b>					
<u>Objective:</u>					
24. Specify web design	1/11/07	1/12/07	Web specification	1/12/07	ML
25. Procure web design	10/12/07	1/2/08	Website	1/2/08	ML
26. Test query and delivery	1/2/08	28/2/08	Test report	28/2/08	ML
27. Usability testing	28/2/08	1/7/08	Reports	1/7/08	ML
<b>WORKPACKAGE 8: Evaluation</b>					
<u>Objective:</u>					
28. Specify evaluation	1/10/08	1/11/08	Specification	<b>1/11/08</b>	ML
29. Procure evaluation	1/11/08	1/12/08			ML
30. Undertake evaluation	1/12/08	30/1/09	Report	<b>30/1/09</b>	ML/contractor
31. Review report	30/1/09	14/2/09			PMB
<b>WORKPACKAGE 9: Marketing</b>					
<u>Objective:</u>					
32. Project start publicity	1/3/07	1/5/07	Press release	1/5/07	ML/GD
33. Live launch	1/3/08	1/7/08	Launch event	1/7/08	ML/GD/ Advisory Board
<b>WORKPACKAGE 10: T&amp; L pack</b>					
<u>Objective:</u>					
34. Define target audience and content	1/3/08	1/5/08	Specification	1/5/08	ML
35. Prepare pack	1/5/08	1/7/08	T&L Pack	1/7/08	ML
36. Test pack and amend	1/7/08	1/10/08	Revised pack	1/10/08	ML

<b>WORKPACKAGE 11: IPR</b>					
<b>Objective:</b>					
37. Identify rights holders	24/3/07	1/7/08	Contact list	1/7/08	GD
38. Contact rights holders	24/3/07	1/12/08	Response letters	1/12/08	GD
39. Incorporate rights data in METS	1/7/07	1/1/09	METS statements	1/1/09	GD/BR

Members of Project Team:

*ML Martin Locock*  
*GD Gwenno Dafydd*  
*BR Branwen Rhys*  
*SW Scott Waby, Digitisation Unit*  
*FD Fei Du, Finance Section*

*PMB Project Management Board*