



JISC Project Plan

Before completing this template please note:

- *The Project Management Guidelines have detailed instructions for preparing project plans.*
- *Please expand tables as appropriate.*
- *Fill in the information for the header, e.g. project acronym, version, and date.*
- *Prepare a cover sheet using the cover sheet template and attach to the project plan.*
- *This template is for completion by JISC funded project managers*
- *Text in italics is explanatory and should be deleted in completed documents.*
- *Please check with your programme manager before completing this form whether they would like to use a specially adapted template specific to your project.*

Overview of Project

1. Background

The project will take more than 50 years of debates from the Upper Chamber of the Northern Ireland Parliament from 1921 to 1972, the Senate Hansard and make it available online, integrating it with the AHRC-funded House of Commons Hansard e-resource, and adding additional functionality (not available in the written volumes) by adding the ability to search and visualise by place across both sets of the Hansards. The debates are wide-ranging in nature concerning the socio-economic and political development of a sometimes troubled part of the United Kingdom over a period of dramatic change. It takes a resource underused by scholars due to the paucity of analogue copies, and the complexity of the printer material, and through innovative search and browse techniques makes it central to the development of Irish Studies in twenty-first century. It meets a clear demand amongst scholars in England and Wales for research material relating to the UK that they simply do not have access to. The combined resource will be far more valuable than its component parts and it will complete the digital availability of historical parliamentary debates for the British Isles. The project offers a strong multifaceted sustainability model and will help to build ICT expertise and infrastructure in the UK.

2. Aims and Objectives

The overall aims of the project are to complete the digitisation and on-line representation of the Historical Hansards for Northern Ireland for both the Commons and the Senate; and to add additional functionality by geo-referencing the place names with coordinates thus enabling visualisation of the debates over time and place through an appropriate web service.

The specific objectives are:

- To digitise the volumes of the Northern Irish Senate Hansard from 1921 to 1972, creating images and OCR text
- To mark up this material by person, role, and topic using the indexes included in the Hansard volumes
- To model and integrate this material with the existing Stormont Papers resource allowing the papers to be searched by name, topic and role
- To add additional functionality by identifying and geo-referencing the place names with coordinates thus allowing users to search by place; and to enable visualisation of the debates over time and place by providing a mash-up with an appropriate web service

Sustainability

Project Acronym: Historical Hansards

Version:1

Contact: Sheila Anderson

Date:17th September 2008

The Centre for e-Research at King's College London has extensive experience of sustaining and preserving web enabled content, largely gained during its stewardship of the Arts and Humanities Data Service. The Centre is now tasked with developing a managed environment within King's for the outputs of its research, including research data. The Historical Hansards material will be developed and managed within that environment.

Separately, CDDA has funding with University College Dublin to establish a dark archive of digital objects held by both institutions with off-site archiving for both CDDA's digital collections of those of UCD. The archive, while funded, has not yet been implemented but it will allow the secure storage of the digital objects. The sustainability of the functionality is equally, if not more, important and CDDA are in discussions with the Digital Humanities Observatory at the Royal Irish Academy in providing long-term functional access to CDDA-developed materials. In the exceptionally unlikely event that CeRch were not able to maintain the resource the DHO would assume this role.

3. Overall Approach

The work is divided between the project partners with CDDA taking responsibility for the conversion of the analogue volumes to digital, including page images of the volumes, and OCR text of the debates. This includes mark-up of the text. The Centre for e-Research is responsible for content modelling and combining the existing Stormont papers with the Historical Hansards, and for their migration and inclusion into a Fedora repository infrastructure in order to ensure sustainability. The Centre will also take responsibility for marking up the indexes included in the volumes; identifying and marking up place names that are in the text to enable enhanced search and browsing functionality to be added to the on-line resource, and adding coordinates to enable spatial visualisation via a mash-up with, for example, Google Earth. The project will also seek to enhance the temporal aspects of the content by adding a timeline function that will allow users to browse debates and follow people over time.

Review and selection of time and space visualisation software will be undertaken by both partners, and both will also be involved in the design and representation of the content in the on-line resource. The project expects to significantly enhance the value of the collection by this additional functionality, allowing new questions to be explored by the data. The exact scope of the additional functionality will be explored during the digitisation phase and will depend upon the availability and functionality of software and services with which to mash-up and represent different aspects of the data.

The ideal would be to mash-up the data (in particular topic, role and people) over time and space. However, this may prove to be outside the capability and resources available to the project.

Critical success factors include the ease with which additional indexes and text can be marked up and connected over volumes and time to better reflect the roles, topics and people over time; and the ease with which the place names can be parsed and identified, and coordinates added to enable mash-up with spatial visualisation services.

4. Project Outputs

- 45,100 archival TIFF images scanned at 600dpi.
- 45,100 digital surrogates of the above suitable for web-based dissemination.
- 45,100 pages of machine readable text processed through an optical character recognition system and edited to reach 99.9 per cent accuracy
- A website hosted by the Centre for e-Research combined with the existing Stormont Papers website, providing a variety of search techniques not available from the analogue content including free text searching, subject searching based on individual indexes contained in each Hansard volume, searching by office of state, searching by name, and searching and visualising by time and/or location.
- Enhanced functionality of the existing Stormont Papers website matching the range of search techniques described for the Senate project

- The development of a key resource that, once the digitisation of the Westminster Hansards is complete, will provide access to contemporary and historical debates of all parliaments and assemblies in the UK and Ireland
- 6 monthly progress report
- Workshop
- Final Report and Completion Report at the end of the project
- Launch of the new service at the Stormont Parliament

5. Project Outcomes

The debates of the Senate give a clear insight into changes and developments of a constituent part of the United Kingdom over more than a 50-year period. The debates do not, as might be anticipated, concentrate just on sectarian matters and relations with Britain and the Irish Republic. The Northern Ireland Parliament had full powers with the exception of foreign policy. The debates relate to matters such as local government, health provision, industrial development, agriculture, and note changes in society and government over a 50-year period. Thus they mirror the range of socio-economic issues found in Hansard for the Westminster Parliament, as well as following the structure and parameters applied to Hansard for the UK Parliament. Moreover, the analogue material is held by very few libraries and is thus inaccessible. A COPAC search indicates that only the University Libraries at the London School of Economics and Manchester have copies in England and Wales. Even where copies are available the material is difficult to use. Where available it is on a reference only basis and there is no comprehensive index to the Senate Hansard.

We thus envisage the free at the point of use availability of the digitised resource to stimulate and help to answer new research questions. For example, researchers will now be able to track debates across time and place; investigate the differences and similarities in language use across chambers, political parties and office of state; they will be able to compare the debates in Northern Ireland with that of the Westminster Parliament. The resource is also of potential use within teaching and we would hope that teachers and lecturers would use the resource to create learning objects and to stimulate debate among their students.

6. Stakeholder Analysis

Stakeholder	Interest / stake	Importance
JISC	Stake	High
AHRC	Interest	Medium
Other digitisation projects funded by JISC or Research Councils	Interest	Low
Scholars	Stake	High
Politicians concerned with Northern Ireland	Interest	Medium
Wider Public	Interest	Medium
King's College London	Interest	Medium
Queen's University Belfast	Interest	Medium

7. Risk Analysis

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Staffing CeRch	3	3	9	Staff in place and potential substitutes identified

Staff CDDA	2	4	8	All staff in place
Organisational	2	3	6	Both CeRch and CDDA are experienced departments. Possible problems will be detected early and effectively dealt with
Technical	2	4	8	All technical infrastructure is already in place and tested
Legal	1	5	5	All IPR issues have already been addressed, copyright agreements are in place

8. Standards

<List the standards the project will use in the table below. Also indicate:

- Any deviations from the standards that JISC recommends.
- Where choices exist in an area, the reasons for the standards selected.
- Where proprietary standards are selected in an area where open ones are available, the reasons for their use and their scope of deployment.>

Name of standard or specification	Version	Notes
XML		Mark up for text and indexes, defined and available dtd for each.
Javascript;Spring Framework;Dojo Toolkit		Technologies to be used for the front end interface – although these may be subject to change as the project progresses. Any changes will be notified.
Tiff images; JPEG		For preservation; for delivery
Dublin Core		All metadata compliant with Dublin Core

9. Technical Development

CDDA's Digitisation Manager in collaboration with the digitisation staff will select the best analogue content to digitise from. Images will be captured at 600dpi, saved as uncompressed TIFFs and stored on CDDA's filestore automatically backed up to a remote site daily. Attention will be paid to preservation issues when selecting volumes for digitisation, as will the appropriate scanner hardware. Images will be enhanced to remove page curvature, standardise the scanned area of each page, and straightened as required. In terms of image quality assurance the data capture staff will cross-review each other's work and the Digitisation Manager will review all of the images. Enhanced images will be used to develop a machine-readable text using OmniPage. Text accuracy will be enhanced using a variety of macros to address recurrent errors in the OCR process with manual correction and automated spelling and grammar assisting. QA procedures will mirror those detailed for checking image quality.

The content will modelled and ingested into a Fedora Repository Framework to ensure sustainability for the future. The interface will be developed using open source technologies. The text will be marked up using XML with defined dtd. JISC guidelines will be followed at all times. Review and selection of tools and services for the timeline and geo-spatial aspects of the project will take place during months 3-4 of the project once the exact nature of the data and the mark-up possibilities have been thoroughly explored.

Project Acronym: Historical Hansards
Version:1
Contact: Sheila Anderson
Date:17th September 2008

10. Intellectual Property Rights

IPR issues were fully addressed in the earlier AHRC-funded project to make Hansard for the House of Commons available. All copyright material will be freely available through existing agreements with the Northern Ireland Assembly authorities and ultimately HMSO with the only caveat that the material should not be used for significant commercial gain without agreement of both the Assembly and CeRch.

Project Resources

11. Project Partners

Centre for e-Research
King's College London
26-29 Drury Lane
London, WC2B 5RL
Main contact: Sheila Anderson, PI

Centre for Data Digitisation and Analysis
Queen's University Belfast
Geography Building
7 Elmwood Avenue
Belfast, BT7 1 NN
Northern Ireland. UK
Main contact: Dr. Paul Eil

12. Project Management

CeRch

Sheila Anderson, Director of CeRch – PI of the project (0.1 fte)
Sheila.anderson@kcl.ac.uk , 020 7848 1981

Information Officer, CeRch - metadata development, mark-up and content models (0.6 fte)

Repository Developer, CeRch: repository and technical developments

CDDA

Dr. Paul Eil, oversight of data capture work (0.15 fte)
paul.eil@qub.ac.uk , 028 9097 3408

Elaine Yeates, supervision of data capture work (???)
e.yeates@qub.ac.uk , 028 9097 2513

Sheila Anderson as the PI will direct the overall project. The Information Officer and Repository Developer will oversee the content, technical and mash-up work. Dr Paul Eil will oversee the data capture work and liaise with Anderson on overall project management. The data capture work will be supervised by Elaine Yeates, the CDDA Office Manager. Two experienced IT Officers at CDDA will capture the data itself.

A project Management Group will be established headed by Sheila Anderson with Paul Eil as co-director. This group will include the CeRch Head of Information Management and CDDA's Office Digitisation Manager. Meetings will be held both on and off line.

13. Programme Support

None identified to date.

14. Budget

See Appendix A

Detailed Project Planning

15. Workpackages

See Appendix B

16. Evaluation Plan

Timing	Factor to Evaluate	Questions to Address	Method(s)	Measure of Success
Ongoing – June 2008	Data Capture of 45100 pages	Is the quality as high as expected, is selected software and hardware suitable, are standards followed	Ongoing image reviews	Signed off by Project Board
Ongoing – August 09	Machine-readable texts	Has the target of 99.9 % accuracy been achieved?	Ongoing testing	Successful test
- September 09	Search techniques	Allow the chosen tools to search the resource effectively? Is the metadata	Ongoing testing, Advisory group	Signed off by the Workshop participants
Ongoing for as long as the resource is online	Resource website	Is the website easy to use and to understand, are tools well explained, is the underlying infrastructure robust?	Testing before completion, user feedback	Signed off by the workshop participants
- September 09	Geo referencing and visualisation tools	Have the right tools been selected, how could functionality be improved further? Are the tools easy to use, are tool manuals available on the website	testing	Project Board Signed off by the workshop participants
ongoing	Sustainability of resource	Have all measurements been taken to allow preservation of resource data and functionality of website	Following preservation standards and best practice	Sustainability Plan included in King's overall preservation planning

17. Quality Plan

Output	45,100 archival TIFF images scanned at 600dpi				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)

03-10-08	Well trained digitisation staff	Staff are familiar with chosen hard- and software, supervised by EY		EY	
30-06-09	All TIFF images are at 600 dpi, best possible readability, Completeness (no missing pages)	Constant check-ups of hard- and software Review of images by Digitisation Manager, supervision by EY	Data complete in Repository	PE	

Output	45,100 digital surrogates of the above suitable for web-based dissemination				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
03-10-08	Well trained digitisation staff	Staff are familiar with chosen hard- and software, supervised by EY		EY	
30-6-09	45100 JPEG files for web dissemination	Review by EY	Data complete	PE	

Output	45,100 pages of machine readable text processed through an optical character recognition system and edited to reach 99.9 per cent accuracy				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
31-7-09	99.9% accuracy	Extensive testing	Test statistics	PE/EY	

Output	XML marked up text				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
30-9-09	Compliant with standards; dtd available	Tested by project	Signed off by Project Board	Information Officer; Technical Officer	

Output	Search techniques				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
30-9-09	Geo-referencing works with Google Map or MS world	Tested by project and reviewed by Advisory Committee	Signed off by Project Board and Workshop participants Functionality available on new website	Information Officer; Technical Officer	

Output	Website hosted by the Centre for e-Research combined with the existing Stormont Papers website offering enhanced functionality				
---------------	--	--	--	--	--

Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
30-9-09	New combined / integrated website	Review by Project Board and Workshop participants	Signed off by Project Board and Workshop participants; successful launch of new website	SA/PE	

18. Dissemination Plan

Timing	Dissemination Activity	Audience	Purpose	Key Message
ongoing	Conference papers and info material	Interested scholars	Promotion of resource	Value of the resource for research
01-10-09	Project website	Interested scholars and public	Dissemination + information	Promote best practice; promote future use of the website; encourage feedback
15-10-09	Final Reports available on King's College London and Queen's University Belfast websites	Interested scholars, other digitisation projects	Information, possibility for review	Contribute to best practice; promote use
Sept 09	Workshop	Participants include scholars, librarians, humanities computing specialists; lecturers	Review, sign off, encourage use for research and teaching	Testing and evaluation; Value of resource for research and teaching
Sept/Oct 09	Launch of resource at Stormont	Public, press	Promotion of resource	Value of resource for research; value of JISC funding for research and widening access
After launch	Mailing lists, press releases	Press, interested scholars	Promotion of resource	Promote use and value; value of JISC funding for research

19. Exit and Sustainability Plans

Project Outputs	Action for Take-up & Embedding	Action for Exit
On-line resource	Available through KCL and Queen's; Metadata harvested by Intute; Available via major search engines	Secure and stable server environment; security measures in place; OAI compliant
Content	Ingested into KCL and Queen's repository and preservation systems	High quality documentation produced and content normalised for ingest into

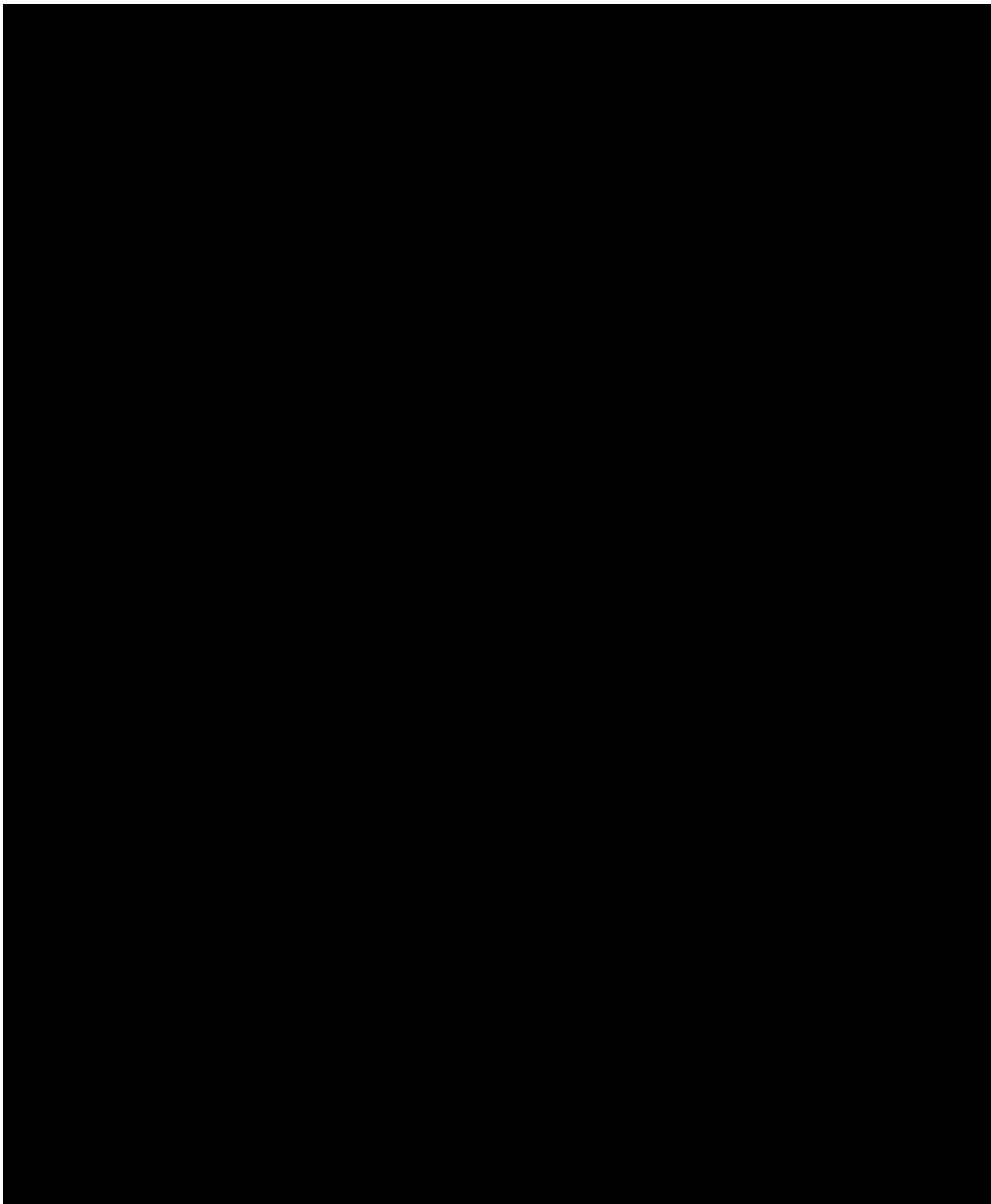
Project Acronym: Historical Hansards
Version:1
Contact: Sheila Anderson
Date:17th September 2008

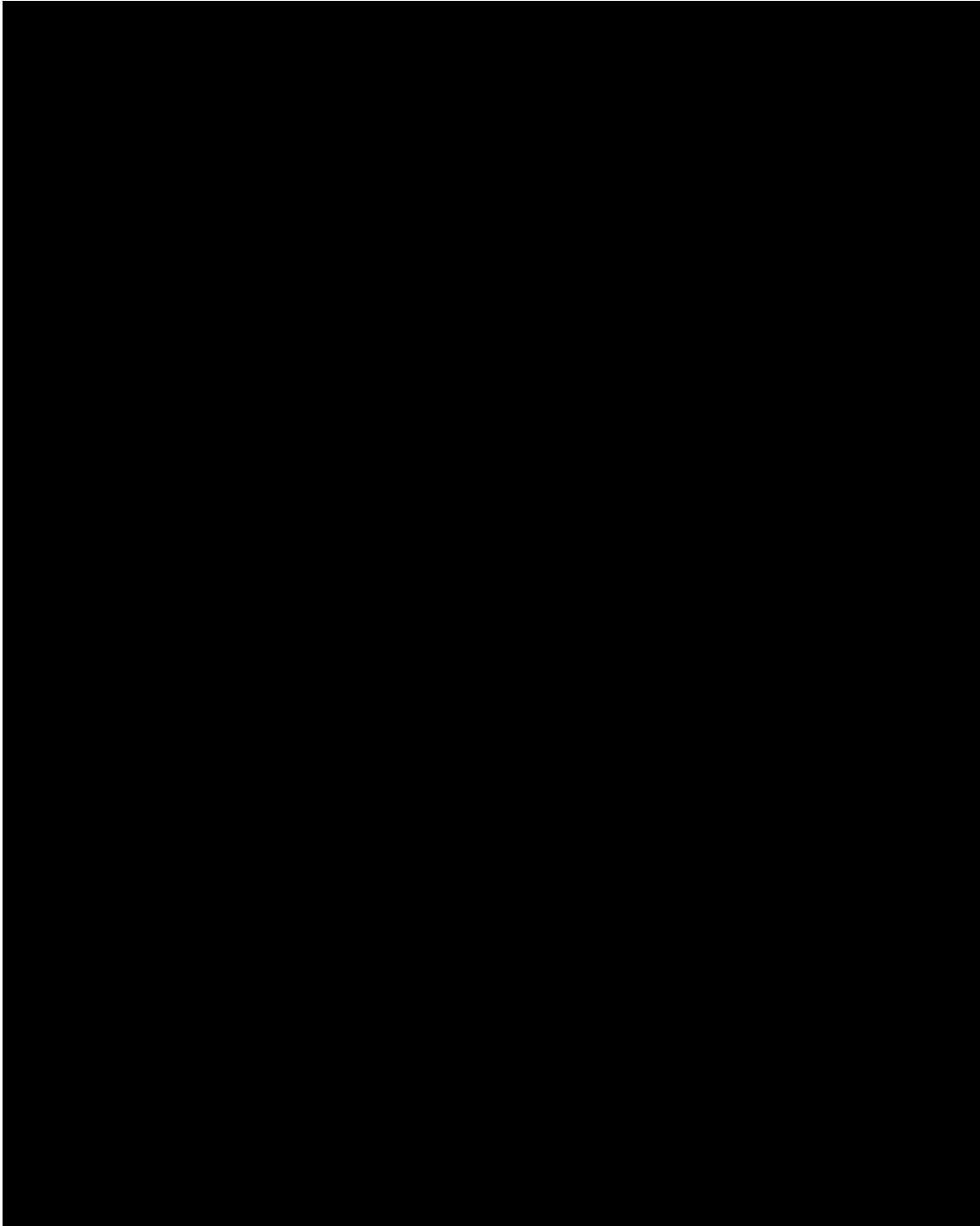
		preservation systems

Project Outputs	Why Sustainable	Scenarios for Taking Forward	Issues to Address
As above			

Appendixes

Appendix A. Project Budget





Appendix B. Workpackages

JISC

JISC WORK PACKAGE

Before completing this template please note:

- *The Project Management Guidelines have detailed instructions for preparing project plans and work packages..*
- *Please expand tables as appropriate.*
- *Fill in the information for the header, e.g. project acronym, version, and date.*
- *Text in italics is explanatory and should be deleted in completed documents.*

WORKPACKAGES	Mon th	1 10	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Project Management																	
1. Project plan		X															
2. PM group meeting				X													
3. Consortium Agreement		X	X	X													
4. Risk assessment		X	X	X													
5. project group meetings	ongoing	X	X	X	X	X	X	X	X	X	X	X	X				
6. Progress report								X									
7. Draft final report												X					
8. Final report													X				
9. Completion report														X			
2. Data Capture																	
10. Volume selection		X															
11. Scanning	ongoing	X	X	X	X	X	X	X	X								
12. Quality review	ongoing	X	X	X	X	X	X	X	X								
13. OCR image to text	ongoing	X	X	X	X	X	X	X	X								
14. Correction of text	ongoing	X	X	X	X	X	X	X	X	X							
15. Quality control checks	ongoing	X	X	X	X	X	X	X	X	X							
16. Transfer to KCL	ongoing	X	X	X	X	X	X	X	X	X							
3. Text Enhancement																	
17. Review and content analysis		X	X	X													
18. content model development			X	X	X												
19. XML tagging			X	X	X	X	X	X	X	X	X	X					
20. place names and coordinates / mark-up			X	X	X	X	X	X	X	X	X	X					
4. Repository and Interface development																	
21. Content model review			X	X	X	X											
22. Migration of content to Fedora				X	X	X	X										
23. Ingest new content into Fedora							X	X	X	X	X	X	X				
24. Timeline functionality						X	X	X	X	X	X	X	X				
25. Geo mash-up						X	X	X	X	X	X	X	X				

26. Interface redesign								X	X	X	X	X						
27. Testing												X	X	X				
5. Dissemination																		
28. promotional material / website			X	X	X	X	X	X	X	X	X	X	X	X				
29. Promotion	ongoing				X	X	X	X	X	X	X	X	X	X				
30. Workshop													X					
31. Launch at Stormont															X			
6. Sustainability																		
32. Sustainability plan		X	X															
33. Ingest into preservation repository															X			

Project start date: *1st October 2008*

Project completion date: *30th September 2009*

Duration: <12> months

YEAR 1			
<i>WORKPACKAGE 1:</i>			
<u><i>Objective:</i></u> Project Management			
1. JISC Project Plan		31st October 08	Project Plan, Workp Project Budget
2. Project Management Group meeting		10 th Dec 08	
3. Consortium Agreement		31st December 08	Consortium Agreem
4. Develop risk assessment and plan		31 st Dec 08	
5. Organisation of Project group meetings		ongoing	Start up meeting; mid meeting; final meeting
6. Progress report		15th April 09	Progress report
7. draft final report		30th August 09	draft final report
8. final report		30th September 09	final report
9. Completion report		31st October 09	Completion report
<i>WORKPACKAGE 2:</i>			
<u><i>Objective:</i></u> Data Capture			

10. Selection of volumes for digitisation		31 st October 08	
11. Scanning of 45100 pages		Ongoing	
12. Review of image quality		Ongoing	Any adjustments to s process
13. OCR image to text		Ongoing	
14. Manual and semi-automatic correction of machine readable text		Ongoing	
15. Quality control checks		Ongoing	Review and adjustme necessary
16. Transfer of images and corrected QA text to KCL		Ongoing	
WORKPACKAGE 3:			
<u>Objective:</u> Text Enhancement			
17. Review and analysis of Stormont content and HH content for similarities and differences		20 th December 08	Short internal docume
18. Develop content model for new content			Content model (intern
19. XML tagging / mark-up of new text and indexes by topic, person name, and role		30th August 09	
20. Identify place names and add coordinates to both sets of content; add mark-up		30 th August	Review and develop methodology
WORKPACKAGE 4:			
Objective: Develop enhanced repository and interface to collections			
21. Review and revise content model to incorporate new content		Feb 09	
22. Develop Fedora instance for collection and migrate existing Stormont content		Mar 09	
23. Ingest new content into Fedora		Mar onwards	
24. Develop timeline functionality		April onwards	

25. Develop geo mash-up and search and display functionality		Feb onwards	
26. Redesign of interface to accommodate additional content, tools and functionalities		August 09	
Testing		September 09	
WORKPACKAGE 5: Dissemination			
Objective: to promote use of the collections and raise awareness			
28. Creation of promotional material and project website		30th September 09	
29. Promotion of project and resource		ongoing	
30. Workshop		Sept 09	
31. Launch at Stormont		Oct 09	
<i>WORKPACKAGE 5:</i>			
<u>Objective: Sustainability</u>			
32. Develop sustainability plan			
33. Ingest content into preservation repository at King's College London			

Members of Project Team:

SA Sheila Anderson
MP Malcolm Polfreman
PE Paul Ell
EY Elaine Yeates