JISC DEVELOPMENT PROGRAMMES
Project Document Cover Sheet
PROJECT PLAN

Project

Project Acronym		**Project ID**	
Project Title	Metadata Generation for Resource Discovery		
Start Date	12/05/06	**End Date**	18/10/06
Lead Institution	AHDS		
Project Director	Dr. Andrew Wilson, AHDS		
Project Manager & contact details	Dr. Malcolm Polfreman AHDS King's College London 26 - 29 Drury Lane LONDON, WC2B 5RL Tel: 0207 848 1985 Fax: 0207 848 1989 Email: malcolm.polfreman@ahds.ac.uk		
Partner Institutions	Ms. Vanda Broughton, UCL		
Project Web URL	To be advised		
Programme Name (and number)	Metadata Generation for Resource Discovery		
Programme Manager	Balviar Notay		

Document

Document Title	Project Plan		
Reporting Period	n/a		
Author(s) & project role	Dr. Malcolm Polfreman		
Date	26/06/06	**Filename**	
URL	n/a		
Access	☐ ✓ Project and JISC internal	☐ General dissemination	

Document History

Version	Date	Comments
1.0	26/06/06	Final draft for submission to JISC

See Project Management Guidelines for information about assigning version numbers.

Metadata Generation for Resource Discovery project plan.

(Final draft 26 June 2006)

Overview of Project

1. Background

Digital resources present challenges and also opportunities in relation to metadata and resource description. The challenges are well known. Analogue resources usually had long-established formats and a stable, publisher-based, dissemination model. Such resources required few cataloguing standards, most of which were well supported, and cataloguers were usually trained professionals (at least in libraries). The comparative ease with which digital resources can be created and copied has encouraged the emergence of self-publishing and a huge expansion in the number of resources available - for which metadata is required - within UK HE and FE. Metadata creation as an essentially manual process that is carried out by trained professional is expensive, "particularly if it is done well enough to support useful resource discovery services"[1]. Under these circumstances, it may be becoming an unsustainable model.

The extra metadata that is required is not just due to the number of new resources. Resource discovery metadata must record increasingly complicated relationships, such as between digital reproductions and analogue originals, or between pre-and post publication versions. The recycling, combining and nesting that is so typical of digital resources places an onus upon content management metadata and rights metadata. The mutability of digital resources and their vulnerability to technological obsolescence heightens the need for preservation metadata.

Author-created metadata may be a partial solution but concerns remain over the quality of metadata produced by non-professionals. Fortunately, the features of digital resources that caused the problem in the first place – their amenability to automated production, copying, and manipulation – may come to the rescue.

Automated metadata generation is still in its infancy but several approaches are worthy of consideration in relation to the JISC Information Environment.

> Metadata harvesting --- Content creation software, such as Microsoft Word and Macromedia Dreamweaver, may automatically insert useful metadata into file headers or allow manual input of such data during the resource creation process itself. (e.g. Microsoft Word automatically records *file size*, *date created* and *date modified*. It also assigns a *title* based on the first line of a document). Metadata harvesting tools, such as DC-dot[2], can identify and collect such pre-existing tagged metadata, whether it is embedded within the resource or held in a separate database. Such metadata, untouched by human intervention, may present accuracy as well as cost advantages.
>
> Metadata extraction --- Metadata extraction tools, on the other hand, such as Klarity, utilise resource *content* rather than pre-existing metadata. Extraction is feasible to the extent that documents of a given genre tend to towards predicable data structure. E.g. E-prints tend to have a document *title, author, author affiliation, abstract, chapter headings,* etc. Sophisticated algorithms may be used to identify and mine the relevant content and, from it, produce structured resource discovery metadata.
>
> Metadata enhancement --- Online ontologies, thesauri, classificatory systems, and authority files are now widely available within integrated library systems. There may be scope for such tools to support automated, as well as manual, cataloguing processes. (e.g. to match names detected within resource content or metadata to names in the Library of Congress authority file without manual intervention).

[1] http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/resource-discovery-review/jisc-resource-discovery-landscape.pdf, p9.
[2] http://www.ukoln.ac.uk/metadata/dcdot/

Automatic indexing --- New sophisticated automatic indexing or clustering techniques are emerging based, for the most part, either on keyword extraction, whereby key terms are drawn from the text of the document, or classification, whereby keywords are assigned from a controlled vocabulary. The iVia Infomine LCSH software[3], for instance, assigns Library of Congress Subject Headings to resources according to an analysis of the most frequently-occurring LCSH terms in previous, similar, resources.

Text and data mining --- These emerging techniques are proving powerful for unearthing unexpected relationships between resources. They use machine learning, statistical analysis, modelling techniques and database technology, to process large quantities of data to identify recurring patterns that may be undetectable to the human eye within resource content. As such, they perhaps inhabit a grey area between resource discovery and the unearthing of new knowledge. Text and data mining techniques still perhaps more readily reveal the *degree* of similarity or relatedness between resources than the *nature* of the relationship and work is needed to harness the techniques effectively within a metadata-based approach to resource discovery.

Unfortunately, the tools currently available for automatic metadata generation are far from comprehensive. The AMEGA[4] report, conducted by the Library of Congress, the first major survey of automated metadata generation techniques, found that of seven content creation applications studied, only one (Winamp) provided automatic generation for all of its descriptive metadata elements. Cataloguers within the JISC Information Environment will find that there are virtually no tools that can harvest both human-generated and system-generated metadata, and also extract metadata from resource content. Nor do tools always support standard metadata schemas likely to be used within the JISC IE, such as Dublin Core.

Part of the problem is that tools have often been developed to meet the needs of specific institutions or in response to particular commercial opportunities and, consequently, handle a narrow range of formats or generate a restricted element set. For example, Jhove[5], the JSTOR/Harvard tool, can extract preservation metadata from jpeg, jpeg2000, gif and tiff image formats but not png or bmp. The National Library of New Zealand Metadata Extract tool[6] covers some of the other formats but presents a limited element set. Similarly, resource extraction algorithms of the current generation are generally only effective within their rather narrow subject domains or for documents of a certain predictable layout or genre. Nor is there always any guarantee of institutional long-term support for the tools.

A suite of tools for use within the JISC Information Environment is required but can only be developed on the basis of a thorough comparison of, on the one hand, the IE's metadata requirements and, on the other, the practical tools and experimental research carried out thus far. The AMEGA report provides a useful point of departure but it was oriented towards the specific needs of the Library of Congress and covered only digital document-like objects (DLOs) - i.e. "primarily textual resources accessible through a Web browser"[7]. The current study will investigate resource discovery metadata for the most common digital resource types within JISC Information Environment and Portals Programme as appropriate. Indeed, in the long term, it is likely that automatic metadata generation will be particularly useful in relation to non-textual resources (e.g. digital images, sound recordings, or statistical datasets) - which inherently lack machine-searchable text/keywords within their content. Pattern-recognition software may, for example, increasingly be used to translate features, such as shapes, colours and faces within images or timbre and melody within music, into useful MPEG7 descriptors. An important issue to consider will be the relative cost-benefit of channelling research into such difficult, non-textual, scenarios where the need is perhaps greatest or into the perhaps easier task of automating metadata capture from textual resources – where the returns may be more immediate.

[3] http://ivia.ucr.edu/projects/Metadata/LCSH.shtml
[4] Automatic Metadata Generation Application (AMeGa) Project http://www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf
[5] http://hul.harvard.edu/jhove/
[6] http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#extraction
[7] http://www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf, p.viii

In the short term, the generation of most resource discovery metadata is likely to require at least some manual intervention – to initiate the automated procedures and perhaps edit the output. Improved interface and template design and a better understanding of workflows will be necessary if the automated and manual methods of input are to be integrated successfully. Cataloguers, fearful for their employment prospects, may fear automatic metadata generation. If the take-up of such tools is to be maximised, the human resource implications cannot be ignored. Automatic metadata generation is not just a technical issue.

2. Aims and Objectives

The first broad aim is to identify the metadata needs of the JISC Information Environment. It is axiomatic that resource-discovery metadata generation tools should be tailored to the resource profile of that domain and the needs of its stakeholders (particularly end users). Specifically, the project will:

- identify the types of resource found within the JISC IE and ask fundamental questions about the kinds of metadata that are appropriate and the level of granularity and accuracy of description necessary for effective resource discovery and use. The current generation of tools have often been designed in relation to specific resource types and this analysis will reveal the range of needs that metadata generation tools ought to support.
- identify the metadata schemas and element sets used by the JISC portals and which tools should therefore support. It will be necessary to distil the essential recommendations from standards such as Dublin Core, dc:terms, DDI, and MODS that tools will need to incorporate.
- Identify the search strategies favoured by end users and the controlled vocabularies that tools will need to accommodate in order to support them. The controlled subject/keyword vocabularies, classification systems, geo-spatial encoding schemes (e.g. co-ordinates), Learning Object (LOM), audience, and educational-level descriptors that are currently employed or are suitable for future use within the JISC IE will be identified.
- Rank elements in terms of their importance within the JISC IE. This will be important in terms of prioritising future development work. For example, access via author and title may be sufficient for some collections, whereas others, such as a music discography, may require additional descriptors such as recording medium, date of performance, and genre.

The second broad aim is to identify and evaluate the metadata generation and creation processes that are currently used within the JISC IE, and particularly the Portals programmes. Specifically, the project will:

- Identify the broad strategy employed (such as whether metadata is generated by authors, professionals or machines) and the specific methods and tools used by each portal.
- Assess the accuracy, cost and suitability of metadata production by current methods.
- Measure the scale of metadata production and assess whether current practice is sustainable.

The third broad aim is to identify currently available tools that are *not* being used by the JISC portals and discover the reasons for the lack of uptake.

- The objective here is to understand the perceptions that project and portal managers have about metadata tools and their use within the particular circumstances of the JISC IE.

The fourth broad aim, based on the findings of the earlier stages of the project, is to identify the areas most urgently requiring new or improved metadata generation tools. Key questions that will be asked include:

- What is the coverage and functionality of the current tools? Which of the formats and metadata elements identified in the survey of JISC IE portals (above) can/can't they support/generate?
- Which schemas are particularly well or badly served?
- Are the available tools schema-specific or can they be configured, perhaps outputting to multiple schemas or converting metadata between them (as DC-dot can between DC and IEE-LOM when extracting metadata from HTML pages or MS Office files)?

- To what extent do factors such as the textual density or genre of the resource affect the quality of output?
- How effectively do the tools integrate manual and automated processes? (The AMEGA project regarded this issue as out of scope and so did not evaluate the editor function within Klarity and DC-dot).
- How does the quality and cost of the generated metadata compare with that which is manually created?
- How scaleable are the metadata extraction and indexing tools and does it require a prohibitively large number of sample texts to 'train' the software for new subject areas.
- Under what circumstances are automatic metadata generation tools currently used and what factors affect uptake?

The fifth broad aim is to identify the most promising new techniques and approaches emerging from recent *experimental research* into automated metadata generation.

- It is anticipated that the work of Jane Greenberg of the Metadata Research Centre <MRC> will be particularly relevant. Her Metadata Generation Research Project[8] has collaborated with the NIEHS[9] to produce usability studies, comparisons of automated and manual metadata output, and evaluation of tools for author generation of metadata. Some account will also be taken of the OCLC Switch projects[10] and the enrichment of resource description by 'fusion' of metadata from different sources being investigated by Diane I. Hillmann and John Phipps, at the NSDL[11].

    - The sixth, and overall aim, of this study is to identify how and where such recent innovations can be profitably harnessed within the JISC IE in order to maximise the effectiveness of automated metadata generation. The potential uses of such techniques within the JISC portal area and their implications for resource discovery within that environment will be assessed.

3. Overall Approach

Strategy, methodology and how the work will be structured

- A questionnaire and in-depth interviews with portal managers will be the principal methods of data collection for identifying metadata needs and current metadata generation/creation processes within the JISC Information Environment.

- Use will also, of course, be made of any schema and metadata documentation provided on JISC IE portal web pages.

- A full-scale review of the product literature and documentation relating to existing automated generation tools will also be undertaken.

- A gap analysis will be conducted, based upon the output of the research outlined above, to identify the areas most urgently requiring new or improved metadata generation tools.

- An in-depth review of the literature from recent *experimental research* into automated metadata generation will be conducted - with a view to identifying key trends, opportunities and solutions within the areas highlighted by the gap analysis.

- This evaluation of experimental research literature will be augmented by interviews with acknowledged experts in the subject.

[8] Metadata Generation Research project (MRC), School of Information and Library Science, University of North Carolina at Chapel Hill. http://ils.unc.edu/~janeg/mgr
[9] National Institute of Environmental Health Sciences (NIEHS) http://www.niehs.nih.gov
[10] http://www.oclc.org/research/projects/mswitch/
[11] http://metamanagement.comm.nsdl.org/Metadata_Augmentation--DC2004.pdf

Important issues to be addressed

Two areas will be given particular emphasis:

- The quality and coverage of the metadata currently embedded within formats by content-creation software (both automatically and/or by human input) and its suitability for harvesting will be considered in depth. The Automatic Exposure[12] project has suggested that, with encouragement, software developers (e.g. digital camera manufacturers) might be persuaded to automatically embed a fuller range of technical metadata in still image files. It may be that strategies could be devised for further encouraging the insertion and more thorough use of meta tags by resource creators. The study will consider whether the educational sector can learn any lessons in this respect from the efforts of commercial search engines to encourage the creation and embedding of metadata.

- Special attention will be given to subject terms, keywords, automatic classification, and concept or topic mapping. Such features are likely to be particularly significant for resource discovery within a metadata/portal-based environment that, unlike Google, does not expect to index all content. Subject/keyword terms have traditionally been perceived as problematic because of the intellectual input that their creation was believed to require. However, new techniques are emerging. For instance, thesaurus type applications are being developed that map extracted terms to controlled vocabularies and there are high level indexing tools (such as IPSV - the Integrated Public Sector Vocabulary for e-government metadata[13]) and switching languages, which tend to use assigned terms, but in ways that can be manipulated or mapped. Recent work on taxonomies and ontologies (e.g. SKOS[14]) may also be relevant. Sophisticated new automatic indexing or clustering techniques are being investigated by cutting-edge projects, such as Wordsmith and Scorpion at OCLC[15]. The study will assess the potential of such new developments and that of The Automatic Metadata Generation (AMG) system of Eric Duval and the Ariadne Project[16], which uses not only content information but contextual information such as file structure and the community from which it emerged to make educated guesses about the subject of a resource. CiteSeer[17] will be analysed to gauge the applicability of autonomous Citation Indexing (ACI) and its benefits for the JISC IE. Input and insight will also be sought from the National Centre for Text Mining[18]. These initiatives will be emphasised because expectations are likely to grow that resource discovery should be effective *across*, as well as within the various subject communities and institutions that constitute the JISC IE.

  Some attention will be given to new approaches that are sometimes referred to as 'community created metadata' (e.g. social networking/folksonomy etc) and their suitability for use within the JISC IE.

Scope and boundaries of the work.

This study will focus primarily on resource discovery metadata. Where there are areas of overlap between resource discovery and other types of metadata (e.g. preservation and content management metadata), this study will highlight tools and issues that are applicable or useful to them. However, time constraints prevent these other types of metadata being addressed per se. Separate JISC studies on a scale similar to that envisaged here will probably be of benefit in relation to them.[19]

[12] http://www.rlg.org/longterm/ae_whitepaper_2003.pdf
[13] http://www.esd.org.uk/standards/ipsv/
[14] http://www.w3.org/2004/02/skos/
[15] http://www.oclc.org/research/projects/auto_class/default.htm
[16] http://www.cs.kuleuven.ac.be/~hmdb/amg/index.php
[17] http://citeseer.ist.psu.edu/aci-computer/aci-computer99.html
[18] http://www.nactem.ac.uk/
[19] Despite the recent emergence of PREMIS, a profusion of in-house technical and preservation element sets tends to be used. Some rationalisation could be encouraged by studies leading to the development of tools specifically oriented towards the requirements of PREMIS http://www.oclc.org/research/projects/pmwg/premis-

Licensing issues will fall within the scope of the study. Some current tools are proprietary (e.g. Klarity). Others (e.g. DC-dot) are open source and thus easier to redistribute, modify and embed as Web Services within the JISC IE presentation layer.

Critical success factors.

The main critical success factor upon which the success of the project will depend relates to whether the project can engage with the JISC IE portal community and obtain sufficient data from it in the form of completed questionnaires and interviews conducted.

4. Project Outputs

The formal output of the project will be a report submitted to JISC that contains a set of recommendations for future work within the JISC IE to help address the challenges identified. The recommendations will be based on an assessment of the severity of the need, the suitability of the solutions that are emerging, and the likely cost-benefits.

5. Project Outcomes

The project aims to identify and scope future research and development that may facilitate increased automated metadata generation within the JISC IE. Metadata generation underpins the ultimate goal of providing metadata of sufficient quality and quantity within the JISC IE to support improved resource discovery for learning and research within the post-sixteen education sector.

6. Risk Analysis

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Staffing	1	2	2	Document the project sufficiently so that other staff can take over if/when project partners are unable to continue with their work.
Organisational	1	2	2	Explicitly define and allocate the tasks, and set deadlines during the initial planning stage in order to prevent disagreement between project partners. Hold regular meetings between project members. They are based in close proximity to one another.
Technical	1	1	1	The technical risks are minimal because the project is primarily desk research and the output is a report rather than an application or prototype.
External suppliers	2	3	6	The biggest potential risk would be a lack of response to the survey. Effectively communicate the advantages that the project and

final.pdf and the preservation element sets recommended by the JISC-funded AHDS studies on the archiving of E-prints, Still Images, and Moving Images and Sound.

				metadata generation may bring to portals. Design the questionnaire with care and, if possible, with the input of a consultant experienced in questionnaire design. Rely on multiple survey method (e.g. interviews as well as questionnaires) Carefully scheduled action to chase up late replies.
Legal	1	1	1	The legal risks are low because the project is desk research and the output is a report rather than the creation of a resource that uses other organisations' 'content'.

Project Resources

7. Project Partners

The lead organisation is the Arts and Humanities Data Service (AHDS). The study will be carried out under the direction of Dr. Andrew Wilson, AHDS Preservation Manager. The day-to-day research for the study will be carried out by Dr. Malcolm Polfreman, AHDS Information Officer, and Vanda Broughton, Lecturer, School of Library, Archive and Information Studies (University College London). Dr. Malcolm Polfreman will be responsible for writing the final report.

Main Contact:

Andrew Wilson, Preservation Manager
Arts & Humanities Data Service (AHDS)
King's College London
26 - 29 Drury Lane
LONDON, WC2B 5RL
Tel: 0207 848 1982
Fax: 0207 848 1989
Email: andrew.c.wilson@ahds.ac.uk

Other contacts (the other project partners):

Malcolm Polfreman
Information Officer
Arts & Humanities Data Service (AHDS)
King's College London
26 - 29 Drury Lane
LONDON, WC2B 5RL
Tel: 0207 848 1985
Fax: 0207 848 1989
Email: malcolm.polfreman@ahds.ac.uk

Vanda Broughton
Lecturer in Library & Information Studies & Departmental Tutor
School of Library Archive & Information Studies
University College London
Gower Street
LONDON, WC1E 6BT

Tel: 020 7679 2291
Email: v.broughton@ucl.ac.uk

Detailed Project Planning

Appendix A. Project Budget

	JISC Contribution Requested			**Institution Contribution**			**Total**
	YR1 (£)	**YR2**	**YR3**	**YR1**	**YR2**	**YR3**	(£)
Staff							
Project director, Andrew Wilson (9 days @ £300)	2,700						2,700
Project partner, Malcolm Polfreman (48 days @ £300)	14,400						14,400
Project partner, Vanda Broughton (38 days @ *£300*)	11,400						11,400
Consultant (12 days @*£500*)	6,000						6,000
Travel & Subsistence	3,500						3,500
Equipment *(office costs)*	2,000						2,000
Dissemination activities							
Evaluation activities							
Other							
Total	40,000						40,000
Total requested from JISC	40,000						40,000

NB. The budget is the same as in the agreed project proposal.

Appendix B. Workpackages

Metadata Generation Project

Schedule	*Month*	**1**	**2**	**3**	**4**	**5**
		May - Jun	Jun - Jul	Jul - Aug	Aug - Sep	Sep - Oct
1: Planning and Reporting		**XXXX**	**XXXX**	**XXXX**	**XXXX**	**XXXX**
2: Current Research and Tools		**XXXX**	**XXXX**			
3: Review of Current Practice			**XXXX**	**XXXX**		
4: Automated Metadata Generation			**XXXX**	**XXXX**	**XXXX**	
5: Metadata and Tool Deployment					**XXXX**	**XXXX**

Project start date: *12 May 2006*
Project completion date: *18 October 2006*
Duration: *6* months

Work Packages

Workpackage and Activity	**Earliest start date**	**Latest completion date**	**Outputs (clearly indicate deliverables & reports in bold)**	**Milestone**	**Responsibility**
WORKPACKAGE 1: Planning And Reporting **<u>Objective</u>: Coordination of activities of project team, report as required, assessment as project progresses.**					
1 Extended work plan and timetable	12 May	26 Jun	This document.		MP/AW
2 Draw up an outline of all deliverables.	12 May	26 Jun	This document.		MP/AW

3	Identify the portals and projects that fall within the scope of the study.	12 May	26 Jun	Working document		MP
4	Draft outline for interim and final project report.	12 May	26 Jun	Table of contents		AW
5	Create website.	26 Jun	3 Jul	Project website		MP/AW
6	Update website with project deliverables.	26 Jun	18 Oct	Project reports		MP
7	Write and submit interim report.	3 Jul	18 Aug	Interim final report.		MP
8	Write and submit draft final report.	21 Aug	29 Sep	Draft final report.		MP
9	Revise draft report and submit final report.	2 Oct	18 Oct	Final report.		MP + AW
WORKPACKAGE 2: Current Research and Tools **Objective: Review of current metadata/resource discovery research, and of existing metadata generation tools**				**Report and gap analysis**		
1	Desk research to identify research and development work into metadata generation strategies (e.g. whether metadata is generated by authors, professionals or machines)	12 May	10 Jul	Working document		VB
2	Desk research to identify current tools available for metadata generating and editing (including analysis of licensing requirements)	19 Jun	10 Jul	Working document		MP
3	Create report listing and assessing current activity and tools	11 Jul	21 Jul	Report on project website		VB + MP
4	Undertake gap analysis (with results from WP 3) and write report	7 Aug	21 Aug	Gap analysis report		MP
5	Gap Analysis Report	22 Aug	23 Aug	Gap analysis report on website		MP

WORKPACKAGE 3: Review of current practice **<u>Objective</u>: Review of Metadata Generation/Creation Tools and Processes Currently Employed in UK HE and FE portals.**			**Report**		
1 Scope interview questions and determine information required	21 Jun	30 Jun	List of questions		MP
2 Devise and design questionnaire, design its layout and.	3 Jul	10 Jul	Appropriately constructed questionnaire (paper and/or electronic)		Consultants
3 Send questionnaire to portal managers and conduct follow-up interviews by email or visits	11 Jul	4 Aug			MP
4 Analyse questionnaire results and interviews to determine: • metadata schemas and element sets currently used within the JISC portals • importance for discovery of individual descriptive elements • specific methods and processes used to create metadata by individual portals • scale and sustainability of current metadata creation processes • tools used by portals to assist with metadata creation • level of granularity needed to improve resource discovery across the JISC portals	24 Jul	11 Aug	Report analysing questionnaire and interview results		MP + VB
5 Write chapter for final report.	7 Aug	18 Aug			MP

WORKPACKAGE 4: Automated metadata generation research **<u>Objective</u>: Identify New Opportunities for Automatic Metadata Generation**			**Report with recommendations.**		
1 Desk research to review the findings of recent experimental research in the field of metadata generation and the documentation relating to emerging tools that are not yet employed within the JISC IE.	21 Jun	21 Jul	Working document setting out research results		VB
2 Identify key trends and new tools, opportunities, and solutions within the areas highlighted by the gap analysis.	21 Aug	1 Sep	Report discussing new trends and emerging tools, and suggesting opportunities for further work and tool development		MP + VB
3 Write chapter with recommendations for final report.	1 Sep	15 Sep			MP + VB
WORKPACKAGE 5: Metadata and Tool Deployment **<u>Objective:</u> Desk research, analysis, and consultation with experts to identify issues and challenges relevant to metadata and tools deployment and processes**			**Report with recommendations.**		
1 Review and analysis of outputs from WP2 - 4.	14 Aug	25 Aug			AW
2 Identify areas appropriate for automated and for manual metadata practices	18 Aug	1 Sep			VB
3 Analysis to develop possible workflow scenarios for integrated automated/manual metadata practices.	4 Sep	15 Sep			AW/MP

4	Identify quality control and organisational structures necessary for integrating automated and manual metadata practices.	18 Sep	29 Sep			MP
5	Identify practical and perceptual obstacles (including an assessment of costs) that prevent maximal uptake of automated metadata tools.	2 Oct	6 Oct			VB
6	Write chapter with recommendations for final report.	2 Oct	13 Oct			MP

Members of Project Team:
AW=Andrew Wilson
MP=Malcolm Polfreman
VB=Vanda Broughton