



Project Document Cover Sheet

Project Information			
Project Title	Names Project (phase two)		
Start Date	1 March 2009	End Date	28 February 2011
Lead Institution	The University of Manchester		
Project Manager & contact details	Amanda Hill Hillbraith Ltd. 59 Carnrike Road Consecon Ontario Canada K0K 1T0 Amanda@hillbraith.com , +1 613 394 0259		
Partner Institutions	The British Liibrary		
Project Web URL	http://names.mimas.ac.uk/		
Programme Name (and number)	<i>Information Environment 2009-11 Programme: Shared Infrastructure Services Strand</i>		
Programme Manager	James Farnhill		

Document Name			
Document Title	<i>Project Plan</i>		
Author(s) & project role	Amanda Hill, project manager		
Date	13 August 2009	Filename	Names_project_plan_v4_Aug09.doc
URL			
Access	<input checked="" type="checkbox"/> Project and JISC internal		<input type="checkbox"/> General dissemination

Document History		
Version	Date	Comments
1.0	13 Jul 2009	Initial draft, AH
2.0	23 Jul 2009	Incorporating feedback from Mimas
3.0	28 Jul 2009	BL feedback
4.0	13 Aug 2009	Feedback from JISC

Project: Names
Version: 4.0
Contact: Amanda Hill
Date: 13 August 2009



JISC Project Plan

Overview of Project

1. Background

The Names Project began in July 2007. It was funded to investigate requirements for a name authority service for UK repositories. A prototype name authority system has been developed as part of this work and a number of connections have been made with UK stakeholders and with international projects working in a similar space.

Scoping work

The initial phase of the project involved researching existing services and standards in the name authority area and consulting with the project's stakeholders to determine the requirements for a Name Authority Service for UK repositories. The Landscape Report which summarises this work was published in October 2007 (<http://names.mimas.ac.uk/documents/>) and has been updated several times since, as new services became available. An initial set of usage scenarios for a name authority service was also published in October 2007.

The specific requirements that the prototype and any subsequent service would need to meet were researched with the involvement of the project's stakeholders in the next phase of the project. The Requirements Report was published in February 2008 and is also available from the project's website at the address mentioned above. The British Library team have produced a detailed data analysis, based on the International Federation of Library Association's *Functional Requirements for Authority Data (FRAD)* and existing name authority standards. This data structure has been used as the basis for the prototype.

Prototype development

The prototype¹ has been developed using an iterative approach due to the shifting nature of requirements and exploratory findings. An initial software requirements specification was derived using the outcomes of the requirements gathering phase, followed by design and development work which has been running in parallel, with input from external developers and stakeholders helping shape its course.

Initial prototype work has focused on several main areas.

- 1) A database has been created, required to store name authority records, based on the entities defined in the Data Analysis and *FRAD* mappings outcomes.
- 2) A back end data collection and disambiguation application is under ongoing development, to acquire data from a variety of sources and identify unique entities within them with which to populate the database.
- 3) A web interface is under ongoing development, working with external partners, to provide machine to machine access to the database, with the creation of an API to provide easy, standardised, flexible querying of the system.
- 4) A web based human search interface has been developed to allow human searching of the names records, and also aid in testing of the prototype.
- 5) A client script has been developed in conjunction with Cranfield University, in order to prototype automated methods of externally retrieving data from the Names system for use in other applications.

All of the above work is still ongoing.

Stakeholder engagement

The name authority area is of interest in a number of different domains, some of which are actively working on solutions to the reliable identification of individuals and institutions. During the project's lifetime the project team have been in contact with UK funding councils and name authority service developers from Australia, New Zealand and the USA, as well as the UK repository community (which is the principal target audience for this project).

2. Aims and Objectives

The aim of this continuation to the first phase is to build upon the achievements of the Names Project by further developing the name authority prototype. This will extend the project for a further two years, building the prototype into a form that will be useful for repository services and working with new sources of information to improve the quality of the data.

Specific objectives

1. Develop a sustainable business model
 - a. Model the process by which names data is created, maintained, managed and distributed
 - b. Pilot the system using UK stakeholders, for example:
 - o EThOS

¹ Available at <http://130.88.120.172:8080/axis/index.jsp>

- Intute
 - UKPMC
 - HEFCE ; UK Research Councils
 - Institutional Repositories
 - c. Explore opportunities to internationalise collaboration through related projects such as:
 - VIAF (Virtual International Authority File)
 - OCLC's Identities Hub
 - ISNI (International Standard Name Identifier)
 - d. Establish measures for the system to enable illustration of incremental improvement over time
- 2. Develop the names demonstrator into a pilot for an operational system
 - a. Develop the limited test dataset into a comprehensive database using data from:
 - Zetoc (Table of contents)
 - LC/NAF (Library of Congress / NACO Authority File)
 - UKPMC (UK PubMed Central)
 - EThOS
 - Other sources (HEFCE and Research Councils, for example, or individual universities)
 - b. Refine the disambiguation algorithm
 - c. Semi-automate identification and disambiguation of named entities using text mining techniques
 - d. Develop the interface for direct input and edit
 - e. Enhance the search interface
 - f. Implement updating by data subjects
 - g. Work directly with repository developers to embed the pilot in repository workflows
- 3. Review and evaluate the data structure in response to experience, user requirements and external developments

3. Overall Approach

The project will pursue an iterative development path, gradually increasing the quality and functionality of the pilot in response to feedback from the stakeholders. Liaison with related services in other domains and in other countries has been an important part of the first phase of the Names Project and is expected to continue in the pilot phase.

The critical success factors for this next phase will be the creation of a substantial corpus of name authority information uniquely and reliably identifying individuals who are likely to be depositors of materials in UK repositories of research outputs and the institutions to which they are affiliated. There will need to be reliable access to this information through scripts which can be easily implemented by developers of repository services.

4. Project Outputs

Expected outputs are:

1. Project plan
2. Reports produced according to the timetable for meetings of the JISC Infrastructure and Resources committee
3. Business model for future service
4. Demonstrated use of the pilot system in a range of repositories
5. Report on data structure review

5. Project Outcomes

A name authority service which provides unambiguous identification of individuals and institutions is a shared infrastructure service which has been recognised as important to funding bodies and institutions alike.² This continuation of the Names Project's work will enable Mimas and the British Library to build upon the prototype that has been developed by the project team and to populate it with sufficient data to provide a useful source of information for funding bodies and for those depositing or seeking research outputs.

Connections between the Names Project and stakeholders such as the UK Research Councils and HEFCE will help to ensure that data within the pilot system is as comprehensive and up-to-date as possible. This has the potential to be of benefit to administrators throughout the UK Higher Education sector.

An additional technical development officer will be employed at Mimas in the second year of the project. The activities of this member of the team would focus on helping repository developers to embed the pilot name authority data into their services. This would allow the project team to demonstrate the functionality of the pilot and its impact on the work of repository users.

A successful pilot should make it possible to go on and develop a service which would enable contributors of materials to institutional repositories to uniquely identify themselves, their institution(s) and department(s) and their co-authors. For managers of these repositories, such a service would make it possible to provide reliable retrieval of all materials provided by a particular individual or department (and not those of others with similar names). Users of repositories would find that their search results are more complete and comprehensive.

Such a future service would allow repository managers to demonstrate improved functionality for the following resource discovery tasks:

- FIND me everything by X, where X is a person or agency responsible for creating or contributing content.
- Refine an initial search to limit the results to a single identity.
- Enable navigation between identities. For example X collaborates with Y, so retrieve everything by Y.

² See, for example, Recommendation 1 of the JISC-funded 'Report of the Subject and Institutional Repositories Interactions Study', <http://ie-repository.jisc.ac.uk/259/> and Dorothea Salo's article 'Name Authority Control in Institutional Repositories', available at <http://minds.wisconsin.edu/handle/1793/31735>

An investigation of options for the provision for such a future service would form part of the work of this continuation of the project.

6. Stakeholder Analysis

Stakeholder	Interest / stake	Importance
JISC	Funding body	High
Repository managers	Possible future users of service	High
Funding Councils	Possible sources of data, possible future users of service	High
Managers of cross-repository services	Possible future users of service	High
Providers of related name authority services (in UK and internationally)	Source of information/possible collaboration	High
Depositors of materials	Possible users of service	Medium
Project partners		High
End users	Improved resource discovery	High (in longer term)

7. Risk Analysis

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Problems recruiting and/or retaining staff	2	5	10	Early advertisement, with possibility of secondment. Embedding staff and project within Mimas/British Library
Breakdown of partnership	1	3	3	Maintain good levels of communication. Hold bi-monthly conference calls.
Inability to meet expectations of project stakeholders	2	4	8	Manage expectations – be realistic about what the project is aiming to achieve.
Expert panel members do not engage with project	2	4	8	Maintain communication with panel members.
Services supplying data cease to operate	2	3	6	Obtain data from a variety of sources. Persistence and sustainability should be criteria for using data.
Data Protection issues limit possible sources for service	2	3	6	Publish only data in public domain and seek to avoid reliance on privileged information. Inform contributors on how any information contributed by them will be used.
Changes in technological environment that renders project unnecessary	1	5	5	Close liaison with JISC and continued monitoring of developments in the area of name authorities

8. Standards

Name of standard or specification	Version	Notes
MARC standard for authority data		Output format for Names records
EAC (Encoded Archival Context)	New edition forthcoming	XML exchange format for Names records
JSON		Exchange standard used for sharing Names records
FOAF		RDF output format for Names records
Names format		Local XML format devised by the Names project for output of Names records

9. Technical Development

The project will take an iterative development approach which will include the involvement of external users and developers for the purpose of testing and refining the pilot system. As part of this the team will work specifically on the data interfaces with their intended users to make sure they are flexible and usable for their purposes. The team will also work iteratively on the back end of the system, working closely with data providers to allow easy import and manipulation of external data sources for use within the disambiguation process.

Therefore the pilot system will be publicly available for use and testing, and continually updated following feedback from its intended audiences. Information about changes to the pilot will be disseminated through the website, the project blog and through Twitter to keep people updated. It will be made clear that data is subject to change, and currently only transient. Though there is currently only one technical developer working on the project, the software will be maintained in a version control system, to allow easy manipulation of the iterative development process in use.

10. Intellectual Property Rights

Ownership of intellectual property rights is as determined by the consortium agreement. There may be rights associated with data supplied by third parties and this will need to be negotiated as part of the process of obtaining the data.

Any outputs from the project will be made available, free at the point of use and under Open Access or Open Source principles where possible, to the UK and HE community in perpetuity. JISC, on behalf of HEFCE, will receive an irrevocable, non-exclusive royalty-free licence in perpetuity to exploit the outputs in any way it sees fit, including enabling the JISC to use, archive, preserve and disseminate the outputs. This may include, where appropriate, the delivery of project outputs to the community under a suitable Creative Commons and/or Open Source licence. In all cases, JISC will also retain the right to modify or adapt the project outputs.

Project Resources

11. Project Partners

Mimas, The University of Manchester
Project management (subcontracted to Amanda Hill of Hillbraith Ltd.)
Development
Data checking and editing

The British Library (Authority Control)

Project: Names
Version: 4.0
Contact: Amanda Hill
Date: 13 August 2009

Expertise in bibliographic authority control
Liaison with international developments

12. Project Management

Members of the project team are listed below. The project will be managed through regular contact between members of the team.

Project Team
Mimas

Project manager: Amanda Hill (amanda@hillbraith.com)
Technical officer: Daniel Needham (daniel.needham@manchester.ac.uk)

British Library (name authority expertise)
Alan Danskin (alan.danskin@bl.uk)
Richard Moore (richard.moore@bl.uk)
Data editing role (to be appointed)

13. Programme Support

The project will look for continued support from the JISC programme management in helping to identify stakeholders, possible data suppliers and in liaison with funding bodies and other organisations which might be able to make use of the Names data.

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	Effectiveness of project	Is the pilot developing into a useful resource for repositories and other stakeholders?	Iterative development and communication with repository developers, communication with expert panel	Feedback from testing sites and expert panel

	Effectiveness of Names pilot	Benchmark Conduct searches now using Intute and other stakeholder interfaces. Analyse results to identify duplication; false drops; omissions; Repeat search as system is developed.	We may have to set up a database somewhere on which to run the potted search, but the closer it is to real data the more convincing it will be.	
	Effectiveness of disambiguation	Benchmark Prepare a test file of records from different sources. Run the file against the algorithm whenever it is upgraded. Measure results: number of records input; number of records matched; number of correct matches; number of mismatches; levels of match		
	Coverage of database	Number of names Number of sources Number and level of links		
	Ease of use	Users test interface to input names. Measure how long it takes and how many errors are introduced.	Would really need fresh users each time.	

17. Quality Plan

Output Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
Project plan	Adherence to project plan guidelines, comprehensiveness	Peer review, review by JISC	Acceptance by JISC	Hillbraith	
Project progress reports	Adherence to report guidelines	Review by JISC	Acceptance by JISC	Hillbraith	
Business model	Comprehensiveness	Review by JISC	Acceptance by JISC	Project team	
Demonstrated	Usefulness of pilot	Peer review	Use by	Project team	

Project: Names
Version: 4.0
Contact: Amanda Hill
Date: 13 August 2009

use			community		
Report on data structure	Comprehensiveness	Review by JISC	Acceptance by JISC	Project team	

18. Dissemination Plan

Timing	Dissemination Activity	Audience	Purpose	Key Message
Ongoing	Presentations, articles, reports, meetings, blog entries, website updates	Stakeholders, expert panel, JISC repository programme	To keep interested parties up to date with project progress	What we're doing and why

19. Exit and Sustainability Plans

Project Outputs	Action for Take-up & Embedding	Action for Exit
Project plan		Preservation as part of Names website by UK Web Archiving Consortium
Project progress reports		
Business model		
Report on data structure		
Pilot system	May be suitable for further development (see below)	

Project Outputs	Why Sustainable	Scenarios for Taking Forward	Issues to Address
Pilot name authority system	May have potential to become part of the repository landscape	Could become a JISC-funded service, or might fall under the purview of the British Library, to complement other name authority activity undertaken there	Responsibility for the service, future funding, maintenance of data contributions

Project: Names
Version: 4.0
Contact: Amanda Hill
Date: 13 August 2009

Appendixes

Appendix A. Project Budget

Project: Names
 Version: 4.0
 Contact: Amanda Hill
 Date: 13 August 2009

Appendix B. Workpackages

Workpackages

WORKPACKAGES	Mon th	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: Project Management																									
2: Standards/activity watch																									
3: Stakeholder liaison																									
4: Data structure review																									
5: Pilot development																									
6: Data editing																									
7: Dissemination																									
8: Evaluation																									

Project start date: 1 March 2009

Project completion date: 28 February 2011

Duration: 24 months

Project: Names
Version: 4.0
Contact: Amanda Hill
Date: 13 August 2009

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Milestone	Responsibility
				Milestone	Responsibility
WORKPACKAGE 1: Project Management <u>Objective:</u> <i>To ensure timely performance of the project activities</i>					Hillbraith Ltd. for Mimas
1. Write project plan	1 Jun 2009	31 Jul 2009	Project plan		
2. Organise project meetings and monitor project progress against plan, internal reporting	Ongoing				
3. Produce progress reports for JISC, in time for meetings of the Infrastructure and Resources Committee.	Ongoing		Reports produced according to the timetable for meetings of the JIR committee		
4. Develop Business Model <ul style="list-style-type: none"> • Work with repositories and other stakeholders to validate use cases and requirements for a service • Estimate usage levels for service • Estimate service growth • Estimate resources needed to maintain and develop service 	1 Mar 2010	28 Feb 2011	Business model for future service		

Project: Names
 Version: 4.0
 Contact: Amanda Hill
 Date: 13 August 2009

<p>WORKPACKAGE 2: Standards Watch</p> <p>Objective: <i>To maintain awareness of evolving standards that are of relevance to the project</i></p>				British Library/Mimas
<p>5. Update Landscape Report with new information</p>	Ongoing			
<p>6. Update metadata/data structure if required</p>	Ongoing			
<p>WORKPACKAGE 3: Stakeholder Liaison</p> <p>Objective: <i>To ensure that the Names prototype system meets the needs of its users and contributors</i></p>				Mimas/British Library
<p>7. Maintain contact with Expert Panel</p>	Ongoing			
<p>8. Establish contact and explore potential for collaboration with UK funding councils</p>	1 Mar 2009	28 Feb 2010		
<p>9. Maintain contact with UK repository developers and managers</p>	Ongoing			
<p>10. Work directly with repository developers on embedding the pilot in their services</p>	1 Mar 2010	28 Feb 2011	Demonstrated use of the pilot system in a range of repositories	

Project: Names
Version: 4.0
Contact: Amanda Hill
Date: 13 August 2009

WORKPACKAGE 4: Data structure review					Mimas/British Library
Objective: To ensure that the data structure adopted for the Names prototype continues to be fit-for-purpose					
11. Review data structure	1 Aug 2010	31 Oct 2010	Report on data structure review		
WORKPACKAGE 5: Pilot Development					BL, Mimas and key stakeholders
Objective:					
12. Model data flows and document system	Ongoing		Initial design and user documentation will be available in Spring 2010.		
13. Continue to work on API in conjunction with repository managers: start getting the system embedded in other services	Ongoing		Release of official API: early 2010		
14. Expand data included in system	Ongoing				
15. Test and improve disambiguation algorithm	Ongoing		Initial data sets (Zetoc, UKPMC, HESA, Open access list of institutions) by early October 2009: will increase data considerably LC/NACO records disambiguation by early November 2009		
16. Testing the use of text-mining techniques to improve data in system	Ongoing				
17. Develop user management tool to allow people to update their own information/merge/split records	Ongoing		Merging and splitting functionality: by December 2009. User management tool: February 2010.		

Project: Names
 Version: 4.0
 Contact: Amanda Hill
 Date: 13 August 2009

18. Evaluate web interface and further develop it in light of user feedback	1 Mar 2010	28 Feb 2011			
WORKPACKAGE 6: Quality Assurance <u>Objective:</u> To ensure that updates to the data in the system are appropriate and accurate and to manage relationships between Names records					Mimas/British Library
19. Sampling user-generated changes to the data and checking them to ensure that they are not inappropriate	1 Mar 2010	28 Feb 2011			
WORKPACKAGE 7: Dissemination <u>Objective:</u> To raise awareness of the Names project and to promote it to potential data contributors and consuming services					All
20. Maintain web site	Ongoing		Current web pages		
21. Publish articles and present conference papers	Ongoing		Articles and papers		
22. Present project to potential contributors and services that might make use of it	Ongoing				

Project: Names
 Version: 4.0
 Contact: Amanda Hill
 Date: 13 August 2009

<p>WORKPACKAGE 8: Evaluation</p> <p>Objective: To assess the effectiveness of the Names project</p>				<p>Hillbraith Ltd./Mimas/BL</p>
<p>23. Establish measures for the system to enable illustration of incremental improvement over time.</p>	<p>Ongoing</p>			
<p>24. Liaise with Expert Panel and testing groups to gather feedback on project outcomes and achievements to include in final report</p>	<p>1 Jan 2011</p>	<p>28 Feb 2011</p>		