

**Repositories and Preservation Proposal Cover Sheet**

<b>Cover Sheet for Proposals</b> (All sections must be completed)		<b>JISC Capital Programme</b>
<b>Name of Capital Programme:</b> Repositories and Preservation Programme		
<b>Bid for Call Area :</b> (Please tick ONE BOX ONLY, as appropriate)		
<b>Tools and Innovation (Strand B)</b>		
<input type="checkbox"/>	<b>Call Area I – Tools and Innovation Projects</b>	Please specify area of proposed project eg 'metadata generation and validation'
<b>Discovery to Delivery (Strand C)</b>		
<b>Call Area II – Discovery to Delivery Projects</b>	<input type="checkbox"/> a) Version identification framework <input type="checkbox"/> b) Persistent identifier interoperability demonstrator <input type="checkbox"/> c) Federated access management and repositories <input type="checkbox"/> d) Semantic interoperability demonstrator	
<b>Repository Start-Up and Enhancement (Strand D)</b>		
<b>Call Area III – Repository Start-Up and Enhancement Projects</b>	<input type="checkbox"/> a) Repository start-up projects <input type="checkbox"/> b) Repository enhancement projects	
<b>Digital Preservation and Records Management (Strand H)</b>		
<b>Call Area IV – Digital Preservation and Records Management Projects</b>	<input type="checkbox"/> a) Digital preservation across the lifecycle <input type="checkbox"/> b) Models and implementation of preservation services <input type="checkbox"/> c) Preservation tools development	
<b>Shared Infrastructure Services (Strand I)</b>		
<b>Call Area V – Shared Infrastructure Services Projects</b>	<input type="checkbox"/> a) Pilot implementation of licence registry <input checked="" type="checkbox"/> b) Pilot national name and factual authority service <input type="checkbox"/> c) Scoping an architecture to support digital policy management <input type="checkbox"/> d) Scoping a terminology registry	
<b>Name of Lead Institution:</b> University of Manchester		
<b>Name of Proposed Project:</b> 'Names'		
<b>Name(s) of Project Partner(s):</b> MIMAS (University of Manchester), British Library		

**Full Contact Details for Primary Contact:****Name: Amanda Hill****Position: Team Leader, MIMAS****Email: amanda.hill@manchester.ac.uk****Address: Manchester Computing, Kilburn Building, Oxford Road, Manchester, M13 9PL****Tel: 0161 275 6055****Fax: 0161 275 6071****Length of Project: 15 months****Project Start Date:****1 July 2007****Project End Date:****30 September 2008****Total Funding Requested from JISC: £176,489****Funding Broken Down over Financial Years (April – March):**

<b>Apr06 – Mar07</b>	<b>Apr07 – Mar08</b>	<b>Apr08 – Mar09</b>
<b>£0</b>	<b>£94,443</b>	<b>£82,046</b>

**Total Institutional Contributions:**

<b>Percentage Contributions over the Life of the Project:</b>	<b>JISC 81%</b>	<b>PARTNERS 19%</b>
---	-----------------	---------------------

**Outline Project Description**

**MIMAS and the British Library will scope and specify a name authority service and proceed to develop a proof-of-concept prototype. The project team will investigate options for the future development and maintenance of the Names service and undertake preliminary testing of the prototype.**

<b>I have looked at the example FOI form at Appendix A and included an FOI form in the attached bid (Tick Box)</b>	<b>YES</b>	<b>NO</b> <b>(nothing in the bid is exempt)</b>  ✓
--	------------	---

<b>I have read the Circular and associated Terms and Conditions of Grant at Appendix B (Tick Box)</b>	<b>YES</b>  ✓	<b>NO</b>
---	---------------------	-----------

## 1. Introduction

1.1. A requirement to be able to reliably and uniquely identify individuals has been articulated across a number of areas of the Information Environment in recent years. For users of repositories of research outputs, it is important to be sure that all the materials submitted by an individual will be retrieved by a search on that person's name, without retrieving items which have been supplied by different people with similar or identical names. For individuals submitting their work, it is similarly important for them to be able to easily assign their own identifier to their materials and to be able to unambiguously identify their institution and department. A shared infrastructure service which will identify individuals and the variants of their names has been recommended by two recent UKOLN reports commissioned by the JISC.<sup>1</sup>

1.2. Between July 2007 and September 2008 MIMAS and the British Library propose to scope the requirements of such a service and specify and develop a pilot which would test the processes concerned. This would include determining the data format, setting up an appropriate database, populating it with records from a variety of sources and testing the use of the data. This will provide important information about the future usefulness of a name authority service for institutional and subject-based repositories, and other applications beyond the repository sector.

## 2. Project Description

2.1. As recommended in the call for proposals, the project will be divided into two principal phases: scoping and pilot development.

### ***Phase I : Scoping (July – December 2007)***

2.2. An expert panel for Names will be established in the early months of the project. This will include representation from the JISC and other organisations with an interest in the development of a name authority service. This group will act as an advisory committee for the project, permitting the sharing of existing expertise and ensuring that the project develops a prototype that is useful to as broad a constituency as possible.

2.3. During the initial phase of the project, the project manager will investigate the requirements of a name authority service. The needs of the UK's repositories will be established as a matter of priority, but it will also be necessary to liaise with related initiatives beyond the repository environment. Initial work on this area in preparation for this proposal has already identified a number of organisations in the UK who would be keen to participate in the development of a name authority service, either by providing test data or by sharing existing expertise.

2.4. The project manager will be based within MIMAS, but will be working closely with colleagues at the British Library, who are already creating name authority data for bibliographic records. The British Library has maintained an authority list since 1981. In 1993 the British Library and Library of Congress initiated the alignment of their authority files through the Anglo-American Authority File project. The British Library joined the

---

<sup>1</sup> *JISC Shared Infrastructure Services Synthesis Study* by Ann Chapman and Rosemary Russell and *Terminology Services and Technology* by Douglas Tudhope, Traugott Koch and Rachel Heery

Name Authority Cooperative (NACO)<sup>2</sup> in 1996 and in the following year the BL Name Authority List was formally superseded by the LC/NACO Authority File.

2.5. The LC/NACO file contains approximately 6.3 million records for names and titles, increasing at approximately 250,000 records per-annum. The LC/NACO Authority File has in excess of 200 institutional contributors from around the world, including several in the UK. The British Library is the largest contributor to LC/NACO, contributing approximately 26,000 records in 2006. Records are contributed in the MARC21 Authorities format. For bibliographic records the LC/NACO Authority File is the standard source of information, but this does not include many of the individuals whose works may be deposited in the UK's institutional and subject-based repositories. As the UK's principal contributor of authority data to the Name Authority Cooperative, and as a potential longer-term maintenance agency for a UK Name Authority Service, it is important that the British Library staff are closely involved in the scoping phase.<sup>3</sup>

2.6. Initial discussions have also been held with colleagues at The National Archives (TNA), who are currently developing a name authority service from the 290,000 personal, corporate and family names which are held in the National Register of Archives.<sup>4</sup> Records in this system are being created in accordance with the International Standard Archival Authority Record for Corporate Bodies, Persons and Families: ISAAR (CPF). It is expected that this service will be launched in January 2007. Archivists have long had similar problems to those of repositories with unambiguously identifying individuals and institutions. The overlap between the names in TNA's system and those that will be needed for the UK's institutional and subject-based repositories may be relatively small, but the methodology may well be similar. Nicholas Kingsley, head of the National Advisory Services department of TNA has agreed to represent the organisation on the expert panel for the Names project, should this proposal be successful.

2.7. The project manager will spend approximately 6 months reviewing the existing provision for name authorities and in drawing up requirements for the pilot Names service. This will include determining the most appropriate format for the authority data and identifying the functions that the service will need to support. The project officer will need to be aware of the provisions of the Data Protection Act and its implications for sharing of personal information. This phase of the Names project will build on the work previously undertaken by the IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR).<sup>5</sup> It will also take into account the recent work in this area that has been done on the JISC's behalf by UKOLN. Rachel Heery has agreed that the Names project should have a representative from UKOLN on its expert panel. The British Library's Authority Control staff will be well placed to undertake initial comparative analysis of name heading standards in anticipation of collaboration. The team will additionally examine options for the creation of crosswalks between formats and related issues including those of character sets and institutional or personal identifiers. A key objective of the work will be to compare the differing formats and related data models to FRANAR in order to identify any potential gaps in coverage prior to the formulation of a pilot.

2.8. An assessment will be made regarding the possible data sources for the pilot, including rights to incorporate and coverage. Recommendations will be made regarding which data sources can be taken forward into the pilot. In addition to the potential sources specified in 2.9 below the project will look at other sources and seek to discuss the project with the Je-S team (responsible for the research councils' Joint Electronic Submissions grant application system), to get access to the active name data in their Common Data

---

<sup>2</sup> <http://www.loc.gov/catdir/pcc/naco/naco.html>

<sup>3</sup> <http://www.bl.uk/services/bibliographic/authority.html>

<sup>4</sup> <http://www.nationalarchives.gov.uk/nra/>

<sup>5</sup> *Functional Requirements for Authority Records: A Conceptual Model*, June 2005, <http://www.ifla.org/VII/d4/FRANAR-Conceptual-M-Draft-e.pdf> and *Mandatory Data Elements for Internationally Shared Resource Authority Records*, February 1999, <http://www.ifla.org/VI/3/p1996-2/mlar.htm>

Repository (CDR). The active data in the CDR has the virtue of being self-managed by the data subject, hence should be robust in quality.

## ***Phase II : The Pilot - Specification and development (September 2007 – June 2008)***

2.9. When the scoping phase is complete, the project will enter its pilot development phase. This will involve the employment of an additional team member with technical skills who will be responsible for working closely with the project manager to specify, design and produce a prototype service.

2.10. The project manager and technical officer will identify and install appropriate software for the prototype Names service, using the existing MIMAS hardware platform. This bid assumes the use of open source software, but a contingency sum has been included in case this should not be possible. The team will then populate the Service with test data from a variety of sources. Some potential sources that have already been identified for this purpose include:

- A subset of author names from the British Library's Zetoc service<sup>6</sup>. The Zetoc database contains the names of journal article authors for around 20,000 journals published from 1993 to the present day.
- A subset of author and organisational names from the Intute database<sup>7</sup>. Intute cataloguers record the names of authors of web resources of interest to the JISC community. Some 30,000 of these have been standardised according to the Anglo-American Cataloguing Rules.
- Names of Principal Investigators for grants from the Wellcome Trust, Research Councils and charities involved in UK PubMed Central. Recent mandates from these funding bodies regarding the deposition of research outputs to open access repositories suggest that UK PubMed Central is likely to become a comprehensive source of names of UK biomedical researchers in the future.
- Names from ESRC's Society Today lists of grant-holders
- Institutional names from MLA's Institutional Server. MLA (the Museums Libraries and Archives Council) have offered to supply the project with the c.6,000 names from this application
- Agent names from the Information Environment Service Registry<sup>8</sup>
- Institutional and departmental names from the Archives Hub<sup>9</sup>. The Archives Hub contains information about archives held in UK universities and colleges and, as a result, holds some departmental names, as well as institutional ones. There are also a large number of personal names, some of which may be useful for the Names service.

2.11. It is expected that analysing the existing data and devising procedures to convert them into a standardised format will be a considerable part of the work of this phase. Where possible, automated routines will be developed to expand the minimal information that will be provided from some of the data sources. This approach has been shown to be successful in a commercial environment by the Scopus Author Identifier service,<sup>10</sup> as well as in the VIAF project undertaken by OCLC and Die Deutsche Bibliothek.<sup>11</sup> If time and resources allow, it may be possible to extend the list of data suppliers. The British Library's Authority Control staff would contribute to this part of the project by building upon their work during the scoping phase and implementing the data crosswalks necessary to

---

<sup>6</sup> <http://zetoc.mimas.ac.uk/>

<sup>7</sup> <http://www.intute.ac.uk/>

<sup>8</sup> <http://iesr.ac.uk/>

<sup>9</sup> <http://www.archiveshub.ac.uk/>

<sup>10</sup> [http://info.scopus.com/etc/authoridentifier/br\\_author\\_identifier.pdf](http://info.scopus.com/etc/authoridentifier/br_author_identifier.pdf)

<sup>11</sup> <http://www.oclc.org/research/projects/viaf/>

support the pilot. A further strand of the work will be the formulation of proposals for the supply of Names data files to third party name authority systems in order to improve the efficient sharing and re-use of information.

2.12. Ensuring alignment between the development of the Names service with other related systems will be essential to provide maximum interoperability for users of name authority files. Likewise, it will be vital for any system to be able to use interoperability standards to share its files so that they can be re-used in the other national and international systems that are already in existence or are under development. Implementation of such standards will be planned for in this phase, though their introduction may be constrained by the project's short timescale. Co-operation with DAREnet<sup>12</sup> in the Netherlands and with the recently-announced 'People Australia' service,<sup>13</sup> based on the Australian Name Authority File will be important elements of the Names pilot. British Library staff would also examine the issues relating to the supply of new or changed headings from a national database to long established international systems used by libraries such as NACO.

2.13. During this phase it is planned that the project will test the use of the prototype in exemplar external systems to ensure the service is fit for purpose and to refine the system accordingly. Services already identified for this purpose include:

- the Intute Repository Search service
- the UK PubMed Central ingest system

### ***Dissemination***

2.16. Sharing information about the project's progress will be undertaken by the appropriate members of the project team. A project website will be established and maintained providing access to information about the project and encouraging feedback and contact from those with an interest in the development of this initiative.

2.17. Participation in JISC Programme Meetings will be important, but it is expected that this will be supplemented by articles, conference papers and presentations in other settings.

### ***Sustainability***

2.19. A successful pilot project is likely to prove to be of long-term value to the JISC community and beyond. Part of the project manager's remit will be to investigate options for continued support of the prototype beyond its initial project phase. As both a partner in UK PubMed Central and a key player in the national and international library world, The British Library is particularly interested in the long term potential for a national name authority system. It would therefore wish to participate fully in examining options for the creation of a sustainable Names service which could be utilised in its various activities as a core resource.

---

<sup>12</sup> <http://www.darenet.nl/en/page/language.view/home>

<sup>13</sup> <http://www.nla.gov.au/nla/staffpaper/2006/jpearce1.html>

## **Intellectual Property Rights**

2.20. It is planned that any software used and developed by this project would be open-source in nature, to enable the project outputs to be shared beyond the partner institutions. The data within the Names prototype would also be freely shared, probably under a Creative Commons licence, with anyone wishing to make use of it. The work during the scoping phase will incorporate investigation of data access rights.

## **Risks**

2.21.

<b>Risk</b>	<b>Probability (1-5)</b>	<b>Severity (1-5)</b>	<b>Score (P x S)</b>	<b>Action to Prevent/Manage Risk</b>
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 communication levels
Technical problems with software/data conversion	2	4	8	Thorough analysis of requirements, options and capabilities in scoping phase
Inability to meet expectations of project stakeholders	2	4	8	Manage expectations – be realistic about what pilot can achieve in relatively short timeframe
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
Data Protection issues limit possible sources for service	2	3	6	
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.

## **Value to the community**

2.22. The Names project will establish the feasibility of a name authority service for the UK repositories community and investigate options for its future development and support. An easily-accessible name authority service would simplify the data input process for individuals submitting their work to a repository, allowing them to select their name from a list. The future introduction of automated procedures for updating the service by repositories will keep the Names service current and help to ensure its usefulness to the community. It is probable that the data within the service will be exported to other authorities, including the Library of Congress Name Authority File. This would bring benefits to the wider community. The British Library would undertake a key liaison role with the Library of Congress and NACO concerning interoperability issues with Names.

## **Evaluation**

2.23. An evaluation of the project will be commissioned in the July to September of 2008.



#### **4. Key Personnel**

##### ***Project partners***

4.1. The partners in this proposal are MIMAS and the British Library.

4.2. MIMAS provides a wide range of national services to the UK academic community, from the provision of data sets and bibliographic resources to repository-related services such as JORUM (in partnership with Edina) and the forthcoming Intute Repository Search service. MIMAS is also the lead partner in the Information Environment Service Registry (IESR), a shared infrastructure project which is now running as a service-in-development.

4.3. The British Library is the national library of the United Kingdom and one of the world's greatest libraries. It is one of the major contributors of authority data to the Name Authorities Co-Operative (NACO), which supplies bibliographic name authorities to the Library of Congress Name Authority File. There is therefore existing expertise within the Library in the area of name authority data, which will be of direct relevance to this project.

4.4. MIMAS and the British Library have successfully provided the Zetoc service to the JISC community since September 2000 and, more recently, have extended their collaboration into the repositories area with their successful joint bid to provide the UK PubMed Central service.

##### ***Expert panel***

4.5. The project partners plan to establish an expert panel to act as an advisory group for the Names project. Representation on this group has already been confirmed by The National Archives and UKOLN. It is expected that the full membership of the expert panel would be established in the first few months of the project, in consultation with the JISC programme manager.

##### ***Project manager***

4.6. Amanda Hill is an archivist by profession and is the service manager for the Archives Hub and the IESR at MIMAS, where she has worked since 2001. In these roles she has had considerable experience of developing standards-based services for the JISC community and beyond, including experience of an early implementation of a test name authority service using the Encoded Archival Context (EAC) standard, as part of the Archives Hub development. IESR is, like the proposed Names service, a shared infrastructure service, and the process of bringing IESR from pilot to service-in-development will be directly relevant to the development of the Names service.

##### ***British Library staff***

4.7. Alan Danskin is a librarian by profession and is Data Quality & Authority Control Manager at the British Library. In this role he has overall responsibility for the fitness for purpose of the British Library Integrated Catalogue. Alan has worked at the British Library since 1987 and has been involved in authority control activities since 1990. He has worked on numerous internal and international authority control projects, including Project AUTHOR and Interparty, and has spoken and written widely on the subject. Alan was responsible for the development and implementation of the Anglo-American Authority File project to integrate the British Library Name Authority List with LC/NACO Authority File. He was a member of the IFLA Working Group which specified the Minimal Level Authority Record and

editor of the UNIMARC Authorities Format, Second Edition and is a member of IFLA's FRBR Review Group.

4.8. Richard Moore has specialised in name and subject authority control at the British Library since 1991. Since 1998, he has managed its Authority Control team, which provides controlled name access to the Library's catalogues, and contributes extensively researched data to the LC/NACO authority file. He is the Library's NACO and SACO coordinator. He has been closely involved in the Anglo-American Authority File project, to reconcile policies and data content between the LC/NACO authority file and the Library's former name authority list. He is responsible for Authority Control policy documentation and training within the British Library.

### ***Technical officer***

4.10 The technical officer will need to start work approximately six months into the project. This may involve a recruitment process, which would need to begin in October 2007.



THE BRITISH LIBRARY

96 Euston Road  
London  
NW1 2DB

T +44 (0)20 7412 7000  
www.bl.uk

THE WORLD'S KNOWLEDGE

Repositories and Preservation Programme  
Joint Information Systems Committee

Dear Colleague

I am writing to confirm the support of The British Library for the attached bid from MIMAS and The British Library to scope and pilot a national name authority service. We consider this work to be a natural extension of our current Authority Control activity and are keen to explore the issues associated with extending this into the repository domain.

MIMAS and The British Library have successfully collaborated on a number of projects in the past and I am sure that our combined expertise will yet again prove to be of value in realising a effective demonstrator for a name authority service.

Yours faithfully,

A handwritten signature in cursive script that reads "Caroline Brazier".

Caroline Brazier  
Head of Collection Acquisition & Description