

A review of the current landscape in relation to a proposed Name Authority Service for UK repositories of research outputs

Prepared for the JISC Names Project by

Alan Danskin, Anne Dixon, Michael Docherty,
Amanda Hill, Richard Moore

September 2007

Contents

Introduction	3
International Standards	4
Functional Requirements for Authority Data (FRAD)	4
International Standard for Archival Authority Records for Corporate Bodies, Persons and Families (ISAAR(CPF)).....	6
International Standard Name (Party) Identifier)	8
MARC-related standards	10
MARC 21 Format for Authority Data	10
MARCXML.....	12
MADS: Metadata Authority Description Schema	14
Library and library-derived systems	15
DAREnet.....	15
DAI (Digital Author Identification).....	15
Name Authorities Co-Operative (NACO)	17
National Library of Australia.....	19
People Australia	19
OCLC.....	21
WorldCat Identities.....	21
VIAF (Virtual International Authority File).....	22
Commercial systems	23
Elsevier	23
Scopus Author Identifier.....	23
Thomson-ISI	24
Web of Knowledge – Distinct Author Identification System.....	24
Other related activities	25
EPrints.....	25
The (UK) National Archives.....	26
National Register of Archives (Name Authority Files)	26
Possible sources of data	28
EthOS	28
Je-S.....	29
UK PubMed Central grantees database	31
Zetoc	32

Introduction

This document has been created as a preliminary activity in the process of identifying requirements for a name authority service for use by repositories of research outputs in the United Kingdom.

Existing standards and systems in this area have been analysed and summarised, particularly in the light of their relevance for such a service. These range from long-established library authority files such as those produced by the NACO co-operative to more recent, repository-orientated systems such as the Digital Author Identification aspect of the Netherlands' DAREnet. Some possible sources of data to form the basis of a name authority service have also been identified and described in this report.

The authors are grateful to the managers of these systems who have taken the time to answer questions about their projects.

International Standards

Functional Requirements for Authority Data (FRAD)

Type

A conceptual model encompassing authority data of all types. The final draft is in preparation.

Purpose

The conceptual model has been designed to:

- provide a clearly defined, structured frame of reference for relating the data that are recorded in authority records to the needs of the users of those records;
- assist in an assessment of the potential for international sharing and use of authority data both within the library sector and beyond.

Methodology

FRAD is the first part of Functional Requirements and Numbering of Authority Records. It is an extension and expansion of the *FRBR* (*Functional Requirements for Bibliographic Records*) model. The study was limited to the library sector, but will be useful to other sectors.

FRAD was developed using the same entity analysis technique as *FRBR*. The technique has been employed in order to:

1. identify the key objects (*entities*) that are of interest to users of information in a particular domain. Each of these *entities*, then serves as the focal point for a cluster of data
2. depict relationships between one type of entity and another type of entity
3. identify the primary characteristics, or *attributes* of each entity.
4. depict the relationships that may exist between instances of entities.

The developers of the *FRBR* model saw significant advantages in declaring persons and corporate bodies to be separate entities that could then be related to the other entities laid out in that model. The *FRBR* decision to treat names as entities rather than as attributes has been carried over into this conceptual model of authority data.

Responsibility

The Working Group on Functional Requirements and Numbering of Authority Records (*FRANAR*) was established in April 1999 by the IFLA Division of Bibliographic Control and the IFLA Universal Bibliographic Control and International MARC Programme (UBCIM). Following the end of the UBCIM Programme in 2003, the IFLA-CDNL (International Federation of Library Associations - Conference of Directors of National

Libraries) Alliance for Bibliographic Standards (ICABS) took over joint responsibility for the FRANAR Working Group with the British Library acting as the liaison.

The Working Group is charged:

- To define functional requirements of authority records
- To study the feasibility of an International Standard Authority Data Number
- To serve as the official IFLA liaison to and work with other interested groups concerning authority files.¹

References

Functional Requirements for Authority Data. A Conceptual Model .IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR) Draft 2007-04-01

<http://www.ifla.org/VII/d4/wg-franar.htm>

International Standard for Archival Authority Records for Corporate Bodies, Persons and Families (ISAAR(CPF))

Type

ISAAR-CPF is a standard developed and maintained by the International Council on Archives. It was adopted by the ICA in 2003 and the second edition was published in 2004. The standard was developed because the archival community felt that existing name authority standards did not meet the needs of archivists in describing the creators of archives, their functions and their relationships with other organisations.

Details

The information elements available within ISAAR-CPF are listed below. The four mandatory elements are highlighted in bold type.

IDENTITY AREA

- **Type of entity**
- **Authorized form(s) of name**
- Parallel forms of name
- Standardized forms of name according to other rules
- Other forms of name
- Identifiers for corporate bodies

DESCRIPTION AREA

- **Dates of existence**
- History
- Places
- Legal status
- Functions, occupations and activities
- Mandates/Sources of authority
- Internal structures/Genealogy
- General context

RELATIONSHIPS AREA

- Names/Identifiers of related corporate bodies, persons or families
- Category of relationship
- Description of relationship
- Dates of the relationship

CONTROL AREA

- **Authority record identifier**
- Institution identifiers
- Rules and/or conventions

- Status
- Level of detail
- Dates of creation, revision or deletion
- Languages and scripts
- Sources
- Maintenance notes

RELATING CORPORATE BODIES, PERSONS AND FAMILIES TO ARCHIVAL MATERIALS AND OTHER RESOURCES

- Identifiers and titles of related resources
- Types of related resources
- Nature of relationships
- Dates of related resources and/or relationships

Note

As a relatively new standard, ISAAR(CPF) is not yet widely used. A corresponding XML exchange format, Encoded Archival Context (EAC), has been developed but this is currently in its beta version and a comprehensive revision of the format is expected.

References

ISAAR(CPF): International Standard Archival Authority Record for Corporate Bodies, Persons, and Families, Second edition <http://www.ica.org/en/node/30230>

EAC web pages <http://www.iath.virginia.edu/eac/>

ISAAR (CPF) and EAC crosswalk

<http://jefferson.village.virginia.edu/eac/documentation/ISAAR2EACbeta.html>

International Standard Name (Party) Identifier

The designation of this project was changed in May 2007 from *ISPI* (International Standard Party Identifier) to *ISNI*: (International Standard Name (Party) Identifier). It is ISO Project 22729.

Type

Draft standard. It is currently at the Working Draft stage, with a view to submitting a Committee Draft in December 2007.

Purpose

The ISNI should enable unambiguous identification of the Public Identity of the parties involved in the creation, production management and distribution chain throughout the media content industries such as music, audiovisual, book, publication.

Methodology

The business model, structure and metadata for ISNI/ISPI are still subject to discussion within the working group. Key principles agreed are as follows below.

- A Public Identity involves the intentional use of a name as a label to represent a specific identity as author, composer, etc. If an author uses two pseudonyms then they have two public identities. If an author's name appears in variant forms on works but there is no intention to represent different public identities then there is only one public identity.
- The ISNI system shall uniquely identify public identities across multiple domains.
- The architecture of the system for achieving this goal is to be defined by the ISNI Registration Authority.
- The ISNI standard will propose a set of required metadata elements to support disambiguation of public identities.

ISNI builds on the work of the EU funded *Interparty* Project.

Scope

To manage the interoperability of natural and corporate names across different domains of creation.

Responsibility

ISO Technical Committee for Information and Documentation (TC46).
Subcommittee for identification and description of information resources (SC9).
Working Group on the International Standard Party Identifier (WG6).

References

ISNI Website

<http://www.collectionscanada.ca/iso/tc46sc9/27729.htm>

InterParty

<http://213.253.134.7/interparty/index.asp>

MARC-related standards

MARC 21 Format for Authority Data

Type

Exchange format/schema.

Purpose

The *MARC 21 Format for Authority Data* is designed to be a carrier for information concerning the authorized forms of names and subjects to be used as access points in MARC records, the forms of these names, subjects and subdivisions to be used as references to the authorized forms, and the interrelationships among these forms.

Although nominally an exchange format, like other MARC formats, the authority format is widely used for input and storage

Methodology

MARC 21 is an implementation of *ISO 2709 Information and documentation – Format for information exchange (ANSI Z39.2) MARC XML* and *MADS* are respectively lossless and lossy expressions of the *MARC 21 Format for Authority Data*.

Character Sets

Nominally *MARC 21* supports 2 character sets: An 8 bit encoding, designated MARC-8, and UNICODE/UCS using the UTF-8 encoding.

Scope

MARC was first developed in the 1960s and is widely used by libraries. MARC 21 has superseded the national formats of United States, Canada and the United Kingdom. MARC 21 has been translated into several other languages. It is used throughout the English speaking world. Germany and Austria are currently planning to replace the *Maschinelles Austauschformat für Bibliotheken (MAB2)* with MARC 21. MARC standards have not found favour outside the library community.

The MARC 21 format for authorities supports authority control of names, titles and subjects. Authorised, variant, related, parallel and alternative script forms of headings can be represented. Notes fields support justification and documentation of decisions.

Responsibility

The development of the MARC 21 formats is an international effort. There are three avenues by which users may comment on the format:

- the MARC forum -a list-serve maintained by Library of Congress;
- the appropriate national MARC committee in Canada, United Kingdom or United States;
- directly to the Network Development and MARC Standards Office at the Library of Congress.

The maintenance structure is complex, but the maintenance process is well established and open. Proposals for change are evaluated and approved at semi-annual meetings of the MARC Advisory Committee. Complex or strategic issues may first be surfaced in reports.

References

MARC 21 home page

<http://www.loc.gov/marc>

MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media

<http://www.loc.gov/marc/specifications/>

MARC 21 Format for Authority Data

<http://www.loc.gov/marc/authority/ecadhome.html>

MARXML

Type

MARC XML is a framework and schema

Purpose

The MARXML framework is intended to be flexible and extensible to allow users to work with MARC data in ways specific to their needs. The framework will contain many components such as schemas, stylesheets, and software tools developed and maintained by the Library of Congress.

The core of the MARC XML framework is a simple XML schema which contains MARC data. This base schema output can be used where full MARC records are needed or act as a "bus" to enable MARC data records to go through further transformations such as to Dublin Core and/or processes such as validation. The MARC XML schema will not need to be edited to reflect minor changes to MARC21. The schema retains the semantics of MARC.

Methodology

All of the essential data in a MARC record is converted and expressed in XML. MARC structural elements, such as the length of field and starting position of field data in directory entries are not needed in the XML record. Leader data positions not needed in the XML environment are retained as place holders or carried as blanks.

As a consequence of the lossless conversion from MARC (2709), information in a MARC XML record enables recreation of a MARC (2709) record without loss of data. A MARC (2709) record can also be created without data loss from MARC XML records. Validation with this schema is accomplished via a software tool. This software, external to the schema, will provide three possible levels of validation:

- Basic XML validation according to the MARC XML Schema
- Validation of MARC21 tagging (field and subfield)
- Validation of MARC record content, e.g., coded values, dates, and times.

Scope

MARXML covers all MARC formats, including the authority format. MARC XML could potentially be used as follows:

- for representing a complete MARC record in XML
- as an extension schema to [METS](#) (Metadata Encoding and Transmission Standard)
- to represent metadata for OAI harvesting

- for original resource description in XML syntax
- for metadata in XML that may be packaged with an electronic resource

Responsibility

MARCXML is developed and maintained by Network Development and MARC Standards Office at the Library of Congress.

Reference

<http://www.loc.gov/standards/marcxml>

MADS: Metadata Authority Description Schema

Type

Schema.

Purpose

MADS is an XML schema for an authority element set that may be used to provide metadata about agents (people, organizations), events, and terms (topics, geographics, genres, etc.). MADS was created to serve as a companion to the Metadata Object Description Schema (MODS).

Methodology

MADS is an XML wrapper that enables the use of MARC authority data in environments which do not support MARC 21. Each individual record is contained within `<mads></mads>`, and groups of MADS records can be contained within a MADS Collection (`<madsCollection></madsCollection>`).

A MADS record includes one or more of the following items: an authority, related or variant element. The authority element is mandatory. There may be any number of related or variant elements. A more detailed specification of how MADS works can be found at <http://www.loc.gov/standards/mads/mads-doc.html>

Scope

MADS is a lossy conversion from *MARC 21 Format for Authority Data*. MADS contains some elements which are not present in *MARC 21 Format for Authorities*.

Responsibility

MARCXML is developed and maintained by Network Development and MARC Standards Office at the Library of Congress.

Reference

MADS Home Page

<http://www.loc.gov/standards/mads/>

Library and library-derived systems

DAREnet

DAI (Digital Author Identification)

Type

The DAI project's aim is to assign unique identifying numbers to all active researchers in the Netherlands. This involved building a name authority file which was available to, and updated by, both the library management system in use at the partner institutions and the research information management system (METIS) used by the same organisations.

Purpose

To reliably link information about researchers and research groups with their published outputs, using existing management systems.

Methodology

Each researcher in the Netherlands is given a unique identifying number, whether or not they have published a paper. The number forms part of a name authority record for the researcher in the library management system. Authors of theses will already have a name authority file as part of the shared cataloguing process in the Netherlands, so there will already be basic information in the library system for many Dutch researchers. Additional layers of information are added (on top of the basic MARC authority file) which include the name of the researcher in the METIS system and details of their institutional and departmental affiliations.

METIS is an administrative system for maintaining information about researchers, grants, research groups, institutions and research outputs. It has been in existence (and under continual development) since 1993 and is used by all 13 universities in the Netherlands and national organisations such as the Ministry of Education and Culture and the funding institutions.

All 40,000 researchers in the Netherlands will be allocated a unique number by the DAI project, which will improve access to the digital objects in repositories made available through DAREnet. These number more than 120,000.

Name authority records and DAI numbers can be created by library staff through the library management system or by administrative staff using the METIS system. The records are then synchronised between the two systems. Future enhancements include allowing the researchers themselves to update the information in their name authority file.

Responsibility

DAI forms part of the Dutch DARE programme (Digital Academic Repositories). This was a collaborative programme between Dutch research bodies and universities whose aim was to create a shared repository infrastructure for Dutch universities. The programme ran from 2003 to 2006 and the DAREnet site is now the responsibility of the KNAW (Royal Academy of Arts of Sciences)

References

Overview of the system at http://dai-uitrol.ub.rug.nl/logboek/FILES/61/DAI_UKSG_20070417_final.doc

Overview of METIS: <http://metis.hosting.kun.nl/metis/default.cfm?i=aboutmetis>

Name Authorities Co-Operative (NACO)

Type

Data source and service

Purpose

NACO is the name authority program of the PCC (Program for Cooperative Cataloguing). Through this program participants contribute authority records for names and in some cases, name/titles, uniform titles and series to the LC/NACO authority file. This file is maintained by the Library of Congress (LC) and is also known as USNAF or NAF. It has become the de facto international standard for name headings in the English-speaking cataloguing world and contains approximately 7 million authority records and increases at a rate of approximately 250k records per annum.

Methodology

An individual institution may join NACO or a group of libraries may collectively form a “funnel project” and contribute records via a coordinator. There are currently about 510 participating institutions. NACO participants agree to follow a common set of standards and guidelines when creating or amending authority records in order to maintain the integrity of the LC/NACO file. The standards used are AACR2 (Anglo-American Cataloguing Rules, 2nd revised edition), MARC21 and the LC Rule Interpretations (LCRIs). Further information regarding these standards and guidelines can be found at: <http://www.loc.gov/catdir/pcc/naco/nacopara.html>.

Authority records are contributed through OCLC (Online Computer Library Center) and directly by the British Library (BL). Contribution files containing both newly created and changed authority records are retrieved daily by LC via File Transfer Protocol (FTP). OCLC and the BL, retrieve a single distribution file each morning from LC, containing all records that were created, updated, loaded, or deleted in the LC system during the previous day. The LC file is the master file. The British Library and OCLC nodes are dependent files. Z39.50 access enables the nodes to view each other’s versions.

Responsibility

A Policy Committee is responsible for guiding the governance of the PCC as a whole, including NACO. It has four permanent representatives: *British Library, Library of Congress, Library and Archives Canada, and OCLC*. There are also 8 rotating partners elected by PCC members. A Steering Committee composed of 5 permanent members approves PCC membership applications, directs the strategic planning process for the Program, and seeks and manages resources. The entire PCC governance structure can be viewed at: <http://www.loc.gov/catdir/pcc/2001pcc.html>

Schema

MARC 21 Exchange or MARC XML.

Character Set

MARC-8

Exchange of NACO authority data is currently restricted to Roman script and a restricted character repertoire. The NACO nodes are planning to move to UCS/ UNICODE exchange using the UTF-8 character encoding. It is anticipated that some character repertoire restrictions will remain in force pending local system developments.

Availability

Name Authority File is available from Library of Congress Cataloging Distribution Service. As a NACO node the British Library also holds a copy of the file.

Size

Retrospective: 1977-2006. 6,650,000 records. File size: 3,132 MB. Avg. record length: 471 bytes

2007 Approximately 450,000 records; including 250,000 new records

References

General information about the PCC can be found at:

<http://www.loc.gov/catdir/pcc/2001pcc.html>

Specific information relating to NACO can be found at:

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

Information on the NACO FTP process can be found at:

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

Information on Library of Congress Cataloging Distribution Service's MARC products can be found at:

<http://www.loc.gov/cds/mds.html#tocmds>

National Library of Australia

People Australia

Type

Re-development of the existing Australian Name Authority File into a web resource for information on significant Australians

Purpose

The aim of People Australia is to make the contents of the Australian Name Authority File available on the Web and to link its contents with information found in other resources as well as in the Australian National Bibliographic Database.

Methodology

A separate repository will be created from the data in the Australian Name Authority File. This will then have interfaces to services of partner organisations and, in turn, provide access to its own data through a variety of protocols. The People Australia project will enhance the data in the Australian Name Authority File, which contains around 120,000 personal names and 90,000 corporate names

For users of the system, a search on the People Australia service will provide a directory page for each individual or organisation, with links to related material in other services.

People Australia is using a new data format for names that had been developed for the related Music Australia service: MAPS (Metadata for Australian People Schema). This was used because existing standards were considered either too basic or too complex. A task to investigate the use of MADS or EAC instead was planned in the feasibility study for the People Australia service.

Note

Basil Dewhurst, the project manager for People Australia, also manages the ARROW Discovery Service which provides a search interface to materials in a variety of Australian repositories. He expects the People Australia work to be used by ARROW and other services to improve the accuracy of retrieval of work by individual researchers.

References

Overview of People Australia:

<http://www.nla.gov.au/initiatives/peopleaustralia/index.html>

Feasibility study:

<http://www.nla.gov.au/initiatives/peopleaustralia/PeopleAustraliaFeasibilityStudyWebsite.doc>

Gatenby, P., 'Reaching new audiences: the People Australia and Picture Australia projects at the National Library of Australia', (Paper presented at IFLA conference in Durban, August 2007): <http://www.ifla.org/IV/ifla73/papers/147-Gatenby-en.pdf>

OCLC

WorldCat Identities

Type

WorldCat Identities is an overview of authors in OCLC's WorldCat union catalogue of library materials from contributing institutions. The authors include both individuals and corporate bodies, showing works by and about the entity, a publication timeline, a link to the entry in the Library of Congress Name Authority File, if one exists, alternative forms of the name, book cover illustrations and links to Wikipedia.

Purpose

The original aim was to provide an overview of authors in WorldCat and a route into the catalogue, but it is recognised by OCLC that the service has the potential for wider use than this.

Methodology

Data is mined from the WorldCat union catalogue which contains 15-21 million personal names. WorldCat includes theses and dissertations, whose authors' names are not established as authorities until they publish monographs. These names would not therefore be found in LCNAF. Likewise, the creators of archival material and internet resources appear in WorldCat Identities, but would not necessarily appear in LCNAF.

WorldCat Identities supports non-Latin scripts, which LCNAF does not. Chinese, Japanese and Korean names may lose their uniqueness when transliterated into Latin script, so supporting the original form reduces ambiguity of these names.. There are around 500,000 non-Latin names in WorldCat Identities.

WorldCat Identities contains information on corporate bodies, at the level of institution (e.g. University of Oxford), but not on lower levels of hierarchy such as departments or schools. WorldCat contains 30,000,000 article-level records, but the names of authors of these are not included in the Identities resource.

User feedback

Users can email OCLC if they notice anomalies in the data (e.g. works of more than one individual grouped under a single name), but at present it is not possible to make suggestions of this nature within the interface.

Responsibility

WorldCat Identities is an internal OCLC project.

References

The beta service at <http://orlabs.oclc.org/Identities/> - will be moved to a more permanent location later this year

VIAF (Virtual International Authority File)

Type

VIAF is a project which makes connections between existing national name authority files – currently LCNAF (Library of Congress Name Authority File) and the Deutsche National Bibliothek's PND (Personennormdatei).

Purpose

The project is working towards creating a virtual international name authority file by matching data from different national name authority files. This allows for local variations in standard forms of names to be linked, and would make it possible to retrieve relevant records from systems that are using different standards.

Methodology

Data fields from bibliographic records are extracted and combined with name authority files for individuals to create enhanced name authority files. These enhanced files from both the Library of Congress and the Deutsche Bibliothek are then compared and linked where there are clear matches in the data.

The VIAF service will be launched later this year.

Responsibility

VIAF is currently a project between OCLC, Library of Congress and the Deutsche Bibliothek. Bibliothèque Nationale de France will be joining the project and the consortium can be expanded to other national libraries in the future.

References

Overview of the project: <http://www.oclc.org/research/projects/viaf/>

Commercial systems

Elsevier

Scopus Author Identifier

Type

Part of a commercial product – Scopus from Elsevier – which provides information about articles, web resources, conference proceedings

Purpose

To disambiguate author names and retrieve all items from a particular author.

Methodology

The approach is similar to that of the Distinct Author Identifier in Web of Knowledge (see below): author names are grouped based on their affiliation, address, subject area, source title, dates of publication citations, and co-authors. It uses full names of authors, where available, rather than just surname and initials.

The Author Search interface permits users to enter surname and forename (or initials), affiliation and broad subject area (life sciences, physical sciences, health sciences and social sciences). If records for the same author have not been grouped by the algorithm the interface allows users to suggest groupings (providing they are registered with Scopus or have an Athens ID). Over 2,000 items of feedback were received to the end of 2006, mainly grouping articles under a single author.

References

Overview:

http://help.scopus.com/Robo/BIN/Robo.dll?mgr=agm&tpc=%2Frobo%2Fprojects%2Fsc_help%2Fh_authrch_intro.htm&project=schelp&wnd=schelp%7CScopusHelp&agt=wsm&ctxid=scopus%2Fresults%2FauthorNamesList.url

Thomson-ISI

Web of Knowledge – Distinct Author Identification System

Type

Author Finder is part of ISI Web of Knowledge – a commercial service providing access to information about articles, conference proceedings and web resources. It is found in the Web of Science product.

Purpose

Author Finder allows users to more precisely identify an author before searching for their articles, or once an article has been found, the other works by an author can be accessed by following the link on the author's name.

Methodology

Data within Web of Science is used to cluster articles under authors' names, using their institutional affiliations, subject of research and articles cited by them.

The Author Finder search is based on surname and one or more initials, then further refined using broad subject areas (arts & humanities, life sciences & biomedicine, multidisciplinary science & technology, physical sciences and social sciences) and institutional affiliation. Not all authors have yet been covered by the clustering algorithm.

There is an opportunity to provide feedback on the Author Finder, by means of an online questionnaire, but this is less sophisticated than the option to group distinct authors that is found in the Scopus equivalent of this service. It also has 12 questions, mostly mandatory, one per page of the questionnaire, which will be offputting to many users. In order to give feedback on the Author Sets page, users have to register with the Web of Science: another barrier.

References

Overview: <http://scientific.thomson.com/support/faq/wok3new/dais/>

Other related activities

EPrints

Type

Repository software author disambiguation.

Purpose

To provide an auto-complete option for authors depositing their materials and to unambiguously identify authors within the system to enable analysis of citation impact of their articles.

Methodology

The new version of the EPrints software (EPrints 3) has a name authority component. This can be formed either from existing metadata within the repository, a separate local name authority file or an external file.

References

Overview of features of EPrints 3: <http://www.eprints.org/software/v3/>

The (UK) National Archives

National Register of Archives (Name Authority Files)

Type

Name authority file for the creators of archives: both individuals and organisations.

Purpose

The National Register of Archives (NRA) provides access to summary information about the location of archival collections in the UK and, in some cases, overseas. It was established in 1945. Access to the NRA has always been available via a name index and this has been available online since 1995. This year the information in the name data has been restructured to be compliant with the International Standard for Archival Authority Records (ISAAR). The name entries can now carry links to related resources and detailed historical notes, providing access to a broader range of information than just the NRA.

Methodology

The NRA name authority files are held in a SQL Server 2000 database. A data creation interface is available for NRA staff, with drop-down lists wherever possible. Data is structured according to the International Standard for Archival Authority Records (ISAAR). The process is quite hand-crafted and time-consuming and at present there is no easy way for external organisations to automatically add or upload information or files for individuals. The National Archives is looking into adding data provided in XML form by other archives: the CALM system used by many UK archives provides an XML output based on ISAAR (but not using the Encoded Archival Context format).

It is possible for other people to expand on the information held in the NRA by using the link 'Your Archives' wiki in each record and put information in the wiki instead. Institutions with name authority information in an ISAAR-compliant XML form can submit them to The National Archives for inclusion in the service. The name authority file data is not available via an Application Programming Interface or in XML format for use by other services – it is a web interface only.

There are around 50,000 personal names in the NRA (most with fairly minimal authority file data) and over 96,000 organisations, of which 239 are universities.

Note

The overlap with names of interest to the Names project is likely to be small, but data could be made available for use by Names if required.

References

Warner, A., 'The National Register of Archives, developing into a Name Authority File database', *Recordkeeping Magazine* (Summer 2007)
<http://www.nationalarchives.gov.uk/documents/recordkeeping-aug-07.pdf>

Possible sources of data

EthOS

Type

E-theses union catalogue project at the British Library.

Details

Information about theses in institutional repositories is harvested and presented with information about digitised theses held in a central hub at the British Library. The metadata for the theses includes author names, institutions and departments (and subject headings). The service will be launched in early 2008.

Note

Some overlap in data with Names is highly probable.

References

EthOS web site: <http://www.ethos.ac.uk/>

Final report on Ethos project by Key Perspectives Ltd:

<http://www.keyperspectives.co.uk/openaccessarchive/reports/EThOS%20Report%20-%20final%20published%20version.pdf> (2006)

Russell, J., 'EthOSnet: building a UK e-thesis community' *Ariadne* 52 (July 2007)

<http://www.ariadne.ac.uk/issue52/russell/>

Je-S

Type

A service used by the UK Research Councils for grant application submission (and a range of subsequent actions). Je-S holds data on research organisations, departments and individual researchers.

Details

The Research Councils link by web services to the Common Data Repository (CDR) which is an integral part of Je-S. Applications for funding need to be assigned to Research Organisations that are registered for funding - so, the CDR also holds data on these institutions and their related departments, with details of which individuals are affiliated to them. Individuals are expected to keep their details in the system up to date, while the research organisations maintain the departmental and staff details.

Types of Data held in the Je-S/CDR:

1. Research Organisations (including Unique ID and Name)
 - 1b. - Departments (Unique ID, Name, Type)
2. Individuals (Unique ID, Title, Initials, Forenames, Surname, (Preferred form of address), (Preferred Salutation), DoB, Gender, LinkedRO, LinkedDept, RegistrationType, Email, Tel, WebURL, Fax, PostalAddress, EthnicOrigin, CountryOfNationality, DisabilityFlag/Type, CurrentPost and StartDate, Sector, Function, ExpertiseAreas, Qualifications, (and soon also any periods of Unavailability)).

There are over 400 Research Organisations in the system and more than 60,000 active individual user accounts.

Note

The CDR data would cover a large proportion of the individuals likely to be of interest to the Names service. Preliminary discussions around the use of this data for the prototype service have taken place. Not all the data would be needed for the Names service, but the unique identifiers and names and the links between the different entities would be of significant interest.

The Terms and Conditions of the service state, under the Privacy/Disclosure section (3.3) that:

The following information contained in funded research proposals may routinely be made publicly available:

- name of the host research organisation;

- details of applicants (title, forenames, initials, surname, research organisation and department);

References

Je-s Terms and Conditions: <https://je-s.rcuk.ac.uk/JESHANDBOOK/jesHelp.aspx?m=s&s=711>

Researcher details in Je-S: <https://je-s.rcuk.ac.uk/JESHANDBOOK/jesHelp.aspx?m=s&s=54>

UK PubMed Central grantees database

Type

Details of grant-holders associated with biomedical research grants.

Details

The database associated with UK PubMed Central holds information on the funding body, the grant identifier, the principal investigator, the name of the project and the principal investigator's email address.

Note

Permission will need to be obtained from the funding bodies to re-use this data in the Names prototype.

Zetoc

Type

Table of contents data for journal articles.

Details

The data includes author names in the format: surname, initial(s) and article and journal metadata. It does not include institutional or departmental affiliation for authors. The data also includes information on the Dewey classification of journals (although this is not searchable through the web interface to Zetoc).

Note

As the data belongs to the British Library, there should be few barriers to its use by the project.

References

Zetoc data types: <http://zetoc.mimas.ac.uk/technical/etocdatamap.html>