National researcher identifier services

Amanda Hill
Names Project
I’d like to talk about some of the work that we’ve done as part of the Names Project recently...

...and how national researcher identifier services fit into today’s researcher identification landscape
Gross generalisation about past approaches to author identifiers

**Libraries**

- Book-level data
- Labour intensive: disambiguation first
- Authors not involved
- Open

**Publishers**

- Article-level data
- Automatically generated: disambiguation later
- Authors can edit
- Proprietary
Current international activity

**ISNI**
- Library-instigated
- Disambiguation first
- Authors not involved
- Broad scope

**ORCID**
- Publisher-instigated
- Disambiguation later
- Authors can submit/edit
- Current researchers
Knowledge Exchange meeting on Digital Author Identifiers in March 2012 encouraged alignment of ISNI and ORCID approaches.

- ISNI has reserved a block of identifiers for use by ORCID: both use 16 digit numbers.
Sources of information

- Both ORCID and ISNI will use existing pools of information to populate their systems
  
  - ISNI: “Leveraging high confidence data from different domains”
  
  - “ORCID will link to other name identifier systems”
National author ID systems

- International systems
  - National system
    - Institution A
    - Institution B
  - National system
    - Institution X
    - Institution Y
Why a national system?

- Reporting relating to researchers often happens at a national or regional level
  - Unique, persistent researcher identifiers within borders of the area are vital for such reporting

- Different countries have different requirements
  - E.g. in UK: Research Excellence Framework (formerly Research Assessment Exercise), Higher Education Statistics Agency
Why a national system?

- National services can interact with local institutions and individual researchers
  - Understand local requirements
  - Can readily respond to requests for customisation/additional services
  - Can define its own scope (e.g. non-active or deceased researchers, historical figures)
Why a national system?

- Disambiguation of authors is key feature of a researcher ID service
  - ...and the hardest part to get right

- Systems like OCLC’s VIAF demonstrate that nationally disambiguated authority files can be successfully matched against each other at an international level
National author ID systems

- 2011: JISC-funded survey and report on national author/researcher identifier systems around the world

- Report published November 2011
  http://ie-repository.jisc.ac.uk/567/
## Maturity of systems (late 2011)

<table>
<thead>
<tr>
<th>System</th>
<th>In development since</th>
<th>Number of identities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lattes (Brazil)</td>
<td>1999</td>
<td>1,600,000</td>
</tr>
<tr>
<td>Frida/Cristin (Norway)</td>
<td>2003</td>
<td>31,000 researchers at 160 institutions</td>
</tr>
<tr>
<td>VIVO</td>
<td>2003</td>
<td>24,400 faculty with profiles 150,000 total IDs including disambiguated co-authors</td>
</tr>
<tr>
<td>Digital Author Identifier (Netherlands)</td>
<td>2005 (1980s for National Thesaurus of Author Names)</td>
<td>40,000 in the NTA 15,000 researchers with Digital Author IDs</td>
</tr>
<tr>
<td>Names Project (UK)</td>
<td>2007</td>
<td>46,000</td>
</tr>
<tr>
<td>New Zealand Electronic Text Centre</td>
<td>2007</td>
<td>2,000</td>
</tr>
<tr>
<td>Trove People and Organisations/NLA Party Infrastructure (Australia)</td>
<td>2007</td>
<td>900,000 people and organisations</td>
</tr>
<tr>
<td>AuthorClaim</td>
<td>2008</td>
<td>200</td>
</tr>
<tr>
<td>Researcher Name Resolver (Japan)</td>
<td>2008</td>
<td>190,000</td>
</tr>
</tbody>
</table>
## Populating identifier systems

<table>
<thead>
<tr>
<th>System</th>
<th>Records created by cataloguers</th>
<th>Records imported from other systems</th>
<th>Records generated by data subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>AuthorClaim</td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Digital Author Identifier (Netherlands)</td>
<td>●</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>Frida/Cristin (Norway)</td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Lattes (Brazil)</td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Names Project (UK)</td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>New Zealand Electronic Text Centre</td>
<td></td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>Researcher Name Resolver (Japan)</td>
<td></td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>Trove People and Organisations/NLA Party Infrastructure (Australia)</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>VIVO</td>
<td></td>
<td>●</td>
<td>●</td>
</tr>
</tbody>
</table>
Features of mature national identifier systems

- With more mature systems:
  - A national organisation generally has oversight: e.g. in Brazil, Norway, Netherlands
  - Integration with research funders, reporting agencies and institutional repositories

- Individual institutions also have defined roles relating to managing information about their own staff
SITUATION IN UK
Work to investigate unique IDs for UK researchers

- Identified in 2006 as part of the call for proposals for the JISC-funded Repositories and Preservation Programme

- Mimas and the British Library proposed a two-year project to:
  - Investigate requirements for a UK name authority service
  - Build a pilot system to demonstrate potential
Original plan

- Use data from British Library’s Zetoc service to create author IDs
  - Journal article information from 1993->
  - Last names, initials, paper titles, subject classifications

- But...
  - International in scope
  - Lack of information on affiliations and first names to help with making matches
  - Huge dataset -> processing issues
Revised plan

- Used 2008 Research Assessment Exercise data (as cleaned up by JISC Merit project) to pre-populate the Names system
  - Names matching algorithm identified unique individuals and assigned identifiers
  - Quality assurance process checked results
Results

- Data quality good, included institutional information and article titles

- Disambiguation process -> high accuracy, despite only having initials, not full first names

- Except for...
Professor Derek Siveter

Professor Emeritus

Email: dereks@earth.essex.ac.uk
TEL: +44 (01655) 272953
FAX: +44 (01655) 272970

Research Profile

My research is broadly concerned with the palaeontology and stratigraphy of the Lower Palaeozoic, and it is in particular focused on the exceptionally preserved faunas of the Herefordshire (Silurian; UK) and the Chengjiang (Cambrian; Yunnan Province, China) Konservat-Lagerstätten, and also on fossil arthropods.

Selected Publications

Search results for...

Names = Siveter, David
Assigned =
ResultPublicationIdentifier =
Family Name = Siveter
Fields of interest =
Given Name =
Affiliation =
names
Order by Id

Siveter, David
Assigned identifiers on other systems
36249 - MERIT
Result Publications
An oostegite crustacean with soft parts from the Lower Silurian.
A Silurian sea spider.
Brood care in a Silurian ostracod
A Silurian sea spider
An ostracode crustacean with soft parts from the Lower Silurian
More...
Fields of Interest
Earth Systems and Environmental Sciences
Collaborations
Sutton, M.D.
Siveter, Derek
Institutional Affiliations
University of Leicester
http://names.mimas.ac.uk/individual/14839 Last modified 01:34 PM May 05, 2011

Siveter, Derek
Assigned identifiers on other systems
141541 - MERIT
Result Publications
An ostracode crustacean with soft parts from the Lower Silurian.
A Silurian sea spider.
A new probable stem lineage crustacean with three-dimensionally preserved soft-parts from the Herefordshire (Silurian) Lagerstätte, UK.
A Silurian sea spider
An ostracode crustacean with soft parts from the Lower Silurian
More...
Fields of interest
Earth Systems and Environmental Sciences
Collaborations
Sutton, M.D.
Siveter, David
Institutional Affiliations
University of Oxford
http://names.mimas.ac.uk/individual/14840 Last modified 01:34 PM May 05, 2011
Building on Merit...

- Merit data covers around 20% of active UK researchers

- Working to enhance records and create new ones with information from other sources
  - Institutional repositories
  - British Library data sets (Zetoc)
  - Direct input from researchers
## Submission form

**The Names Project**

### Submit

The Names Project is working to uniquely identify UK researchers, to make it easier to find works by an individual.

If you are not already in our system, you can create a record for yourself by completing this form to create a Names identifier. You will receive an email with your own unique identifier when your submission has been processed, usually within a week.

Find out more about Names.

Items marked with an * are required.

Press the help button (?) for more information about a particular section.

Press the plus sign (+) to show a section of the form that contains additional fields.

If you are having trouble viewing some elements of the form then you can view a simplified version here.

### My details

#### My Name Information

#### General Name Information

<table>
<thead>
<tr>
<th>Field</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family Name *</td>
<td>e.g. Smith or Johnson-Smith</td>
</tr>
<tr>
<td>Given Names / Initials*</td>
<td>e.g. Jane or J</td>
</tr>
<tr>
<td>Preferred Name Form *</td>
<td>e.g. Smith, J. or J Johnson-Smith</td>
</tr>
</tbody>
</table>

#### Previously Known As

<table>
<thead>
<tr>
<th>Field</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family Name</td>
<td>e.g. Johnson</td>
</tr>
<tr>
<td>Given Name / Initials</td>
<td>e.g. Jane or J</td>
</tr>
</tbody>
</table>

Add another previous name

#### Published Name Forms

<table>
<thead>
<tr>
<th>Field</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name Form</td>
<td>e.g. Smith, J. or J Smith</td>
</tr>
</tbody>
</table>
Quality matters

- Automatic matching can only achieve so much
  - Dependent on data source
  - Very important to get this right, from researcher’s point of view

- British Library team perform manual check of results of matching new data sources
  - Allows for separation/merging of records

- Plan to allow people to update their own information in future
Access to Names

- API allows for flexible searching of Names data

- EPrints plugin released in 2011: allows repository users to choose from a list of Names identities
  - ...and to create a Names record if none exists
### Creators

Type three letters of the Family Name to begin searching the Names identity service. You can narrow the search further by typing an initial into the Given Name / Initials box. Move the cursor down the drop-down list to show information about the person to help in identification. Otherwise, enter the name in the format given on the first page or title page. If there are more than four authors, click on the [More input rows] button. To have your own details added to the Names identity service, visit [http://names.mimas.ac.uk/submit](http://names.mimas.ac.uk/submit).

Example: [Chance-Collins] [Chris] [http://names.mimas.ac.uk/individual/42131]  
Example: [Smith] [J.P.] [j.smith@soton.ac.uk]  
Example: [Fu] [Yan-Lih] []  
Example: [von Hüpner] [Ludwig G.] []

<table>
<thead>
<tr>
<th>Family Name</th>
<th>Given Name / Initials</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[More input rows]

### Corporate Creators

<table>
<thead>
<tr>
<th>1.</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[More input rows]

### Editors

<table>
<thead>
<tr>
<th>Family Name</th>
<th>Given Name / Initials</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[More input rows]
<table>
<thead>
<tr>
<th>Creators</th>
<th>Family Name</th>
<th>Given Name / Initials</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Roberts</td>
<td>A</td>
<td><a href="http://names.mimas.ac.uk/ind">http://names.mimas.ac.uk/ind</a>...</td>
</tr>
<tr>
<td>2.</td>
<td>Cox</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Cox, A</td>
<td></td>
<td><a href="http://names.mimas.ac.uk/ind">http://names.mimas.ac.uk/ind</a>...</td>
</tr>
<tr>
<td>4.</td>
<td>Cox, A G</td>
<td></td>
<td><a href="http://names.mimas.ac.uk/ind">http://names.mimas.ac.uk/ind</a>...</td>
</tr>
<tr>
<td>5.</td>
<td>Cox, A L</td>
<td></td>
<td><a href="http://names.mimas.ac.uk/ind">http://names.mimas.ac.uk/ind</a>...</td>
</tr>
<tr>
<td>6.</td>
<td>Cox, A M</td>
<td></td>
<td><a href="http://names.mimas.ac.uk/ind">http://names.mimas.ac.uk/ind</a>...</td>
</tr>
<tr>
<td>7.</td>
<td>Cox, A W</td>
<td></td>
<td><a href="http://names.mimas.ac.uk/ind">http://names.mimas.ac.uk/ind</a>...</td>
</tr>
<tr>
<td>8.</td>
<td>Cox, B E</td>
<td></td>
<td><a href="http://names.mimas.ac.uk/ind">http://names.mimas.ac.uk/ind</a>...</td>
</tr>
<tr>
<td>9.</td>
<td>Cox, B T</td>
<td></td>
<td><a href="http://names.mimas.ac.uk/ind">http://names.mimas.ac.uk/ind</a>...</td>
</tr>
<tr>
<td>10.</td>
<td>Cox, C L</td>
<td></td>
<td><a href="http://names.mimas.ac.uk/ind">http://names.mimas.ac.uk/ind</a>...</td>
</tr>
<tr>
<td>11.</td>
<td>Cox, D</td>
<td></td>
<td><a href="http://names.mimas.ac.uk/ind">http://names.mimas.ac.uk/ind</a>...</td>
</tr>
<tr>
<td>12.</td>
<td>Cox, D R</td>
<td></td>
<td><a href="http://names.mimas.ac.uk/ind">http://names.mimas.ac.uk/ind</a>...</td>
</tr>
</tbody>
</table>

**Names author information**

**Affiliations:**
University of Manchester

**Fields of interest:**
Physics

**Selected publications:**
- "Energy flow and rapidity gaps between jets in photoproduction at HERA"
- "Double diffractive higgs and di-photon production at the Tevatron and LHC."
- "Observing a light CP-violating Higgs boson in diffraction."
Ultimate aim

- High-quality set of unique identifiers for UK researchers and research institutions

- Available to other systems (national, international and institutional)
  - Names has already obtained ISNIs for over 40,000 unique identities

- Possible additional services (to generate income and aid sustainability)
  - Disambiguation of existing data sets
  - Identification of external researchers in institutions’ data
Why a national system?

- Reliably disambiguated unique identifiers at national level can be high-quality resource that can be used for
  - International systems
  - National reporting
  - Improving institutional repositories/researcher management systems

- ...in a scalable, flexible, way
Project updates

- Names: http://names.mimas.ac.uk
- Blog: http://namesproject.wordpress.com
- Twitter: @NamesProject