

Report on National Approaches to Researcher Identification Systems

Prepared for the JISC Researcher Identifier Task and Finish Group

November 2011

Amanda Hill

Hillbraith Ltd.

Executive Summary

This report was commissioned to inform the discussions of the JISC-convened Researcher ID Task and Finish Group, which has been charged with formulating recommendations for UK activity in the field of researcher identifiers. The remit of the report was to examine the approaches taken in other countries to the creation and maintenance of researcher identifiers in the light of a growing acceptance around the world that unique identification of creators of research is desirable for recognising, promoting and analysing their work.

It was found that there is considerable variety in the ways in which researcher identifier initiatives have been developed across the world. Many of the systems investigated have only been established in the last five or six years and are still therefore relatively immature in terms of integration with other elements of national research infrastructure. The main exception to this generalisation is the Lattes system in Brazil which has been running for longer and which is now well integrated into national and local systems in that country. Funding for researchers in Brazil is dependent on them having a profile which has been an important factor in the success of Lattes.

Funding for these initiatives has principally been provided by national bodies including ministries, research funding councils and national libraries. The exceptions here are the Digital Author Identifier of the Netherlands which is funded through licencing; the VIVO system which was originally funded by Cornell University and has since been extended elsewhere through a grant; and the (international in scope) AuthorClaim service which is run mainly through the work of volunteers. Centralised funding has usually been used to create centralised systems: sometimes from the outset, as in Brazil, the Netherlands, and Japan, and sometimes after a successful smaller-scale institutional pilot, as in Norway and with VIVO.

There is a general acknowledgement among the initiatives of the importance of making data on researchers openly available, ideally as Linked Open Data, although this has not been possible under the licencing business model used in the Netherlands. There are also data protection restrictions there, due to the nature of the information held about individuals in the system.

Several of the systems polled for this report have been working with international initiatives such as ORCID, VIAF and ISNI to investigate ways of sharing data. These international projects are also relatively immature and there is still uncertainty as to how this integration is best achieved. As research is frequently conducted across international borders, the value of an international mechanism for identifying researcher is recognised. The main benefits of national identifier systems are the ways in which they can be linked into existing national infrastructures for allocation of research funding, assessment of research excellence and collation of research statistics and this is demonstrated by the most mature identifier systems examined here.

Contents

1. Introduction	4
2. Motivation for developing identifier systems.....	5
3. Responsibility for identifier systems.....	5
4. Maturity of systems	5
5. Scope of identifier systems.....	6
6. Population of the identifier system	7
7. Assignment of identifiers	7
8. Form of minted identifiers.....	8
9. Information maintained in systems.....	9
10. Interaction with other systems	9
11. Data licensing	11
12. Funding and development	11
13. Integration with international name identifier initiatives	13
14. Conclusion.....	13
15. References.....	14

1. Introduction

This report is one of several being prepared for the JISC Researcher Identifier Task and Finish Group which was convened in 2011 to develop recommendations for future activity in the field of researcher identifiers in the United Kingdom.¹

The purpose of this report is to examine approaches in national projects around the world which are creating identifiers for individuals beyond the traditional focus of library authority files (i.e. not just authors of printed books) and, in particular, in the area of identifying individuals engaged in academic research. In consultation with JISC, a shortlist of nine such services was drawn up and a survey was sent to each. Seven responses were received. Of these, one was not directly relevant to the investigation, as it covered library authority files only.

The projects and individuals who provided the information which was used to create this report are:

AuthorClaim (international)	Thomas Krichel, Open Library Society
Digital Author Identifier (Netherlands)	Maurice Vanderfeesten, SURF Foundation
New Zealand Electronic Text Centre (NZETC)	Michael Parry, Digital Initiatives Co-ordinator, Victoria University of Wellington Library
People Australia	Basil Dewhurst, Manager, Resource Discovery Services, National Library of Australia
Researcher Name Resolver (Japan)	Kei Kurakawa, Associate Professor, Research and Development Center for Scientific Information Resources, National Institute of Informatics
VIVO (US university consortium)	Jon Corson-Rikert, Head, Information Technology Services, Albert R. Mann Library, Cornell University

Response received but not directly relevant:

National Library of Germany	Uta Ackerman
-----------------------------	--------------

The following services were contacted but no response was received. For the purposes of this report, information on these initiatives has been deduced from websites and related articles, where possible.

Lattes (Brazil)
Frida/Cristin (Norway)

¹ <http://technicalfoundations.ukoln.info/blog/researcher-id-task-and-finish-group>

Information on the UK JISC-funded Names Project has also been included in the report, where appropriate, for purposes of comparison.

2. Motivation for developing identifier systems

The need for reliable identification of individuals was frequently cited as the original motivation for the creation of an identifier system. In New Zealand and Australia the focus was to provide identifiers for individuals mentioned in texts or who were authors of materials within digital library sites: in other words it was a bibliographic imperative which drove the development.

For the other projects and services the impetus for developing a system has been the need to uniquely identify people engaged in publishing scholarly resources, usually in the context of university research. In Japan, the Netherlands, Norway, Brazil and the UK the initiatives have had a national remit, while the US VIVO service was initially developed for a single institution (Cornell University) and has since been adopted by others, with a cross-institutional search under development.

3. Responsibility for identifier systems

Responsibility for the development and maintenance of the systems covered in this report varies. Some services are run by national bodies, others by individual institutions. The more mature, co-ordinated and comprehensive systems, such as those in the Netherlands, Norway and Brazil, tend to be a combination: a national body has overall control and responsibility, with other institutions having particular roles to play in relation to data about individuals for which they are responsible in some way.

In the Netherlands, for example, SURF is responsible for the governance and strategic developments of the overall identifier system, while the National Thesaurus for Author names is managed by OCLC and each university is responsible for its own Current Research Information System (CRIS) which holds information on members of research staff.

4. Maturity of systems

Development of these identifier projects and services has taken place, for the most part, in the past three to eight years, with the notable exception of the Lattes system of Brazil, which has been under development since 1999. The table below arranges the initiatives on order of the date of establishment and (where known), the number of identities registered.

System	In development since	Number of identities covered
Lattes (Brazil)	1999	1,600,000
Frida/Cristin (Norway)	2003	(researchers at 160 institutions)

System	In development since	Number of identities covered
VIVO	2003	150,000 (in USA institutions)*
Digital Author Identifier (Netherlands)	2005	40,000
Names Project (UK)	2007	46,000
New Zealand ETC	2007	2,000
People Australia	2007	900,000 (people and organisations)
AuthorClaim	2008	200
Researcher Name Resolver (Japan)	2009	180,000

*note that this figure includes undisambiguated records created for co-authors who may not be staff members at those institutions. The figure for profiles of faculty members at those US institutions running VIVO is around 24,400.

As might be expected, it is the longest-running systems which are the most advanced in terms of integration with other services and number of individuals identified within them.

5. Scope of identifier systems

The scope of respondents' systems also varies considerably. Some are broad: for example the People Australia system covers "significant people and organisations... including people and organisations described in the Australian name authority file and VIAF, researchers, people described in various dictionaries of biography etc." The New Zealand system has a similar scope, although it is limited to individuals represented in materials within the New Zealand Electronic Text Centre.

The researcher-centric systems often aim to cover all individuals active in research within institutions. In the Netherlands, the Digital Author Identifier covers researchers in all Dutch universities, with the scope gradually widening to include researchers in other institutions which may not maintain a CRIS. In Brazil, all research, Masters and PhD students are enrolled in the LATTES system. The Norwegian Frida system was based on information exported from universities' human resources databases and therefore covered all individuals working in universities in that country. It has since been expanded (as Cristin) to include researchers in other organisations such as hospitals, colleges and research institutes.

Institutions running the VIVO system make their own decisions about which of their staff will be included. They also have to develop their own policies over what to do with information about staff who move to other institutions or who die while in service.

The international AuthorClaim service is open to any author of published materials. The UK Names Project has also focused chiefly on published authors of journal articles in creating its records for UK researchers, although its scope is not necessarily limited to them.

6. Population of the identifier system

Approaches to creating records in the identifier systems investigated here vary. Some have cataloguing staff involved in the creation of authority records, others process data from existing systems to populate records in an automated way and some allow individual data subjects to create their own profiles. More than one of these may be used by a particular service and the table below summarises the methods that the identifier systems examined here are using.

System	Records created by cataloguers	Records imported from other systems	Records generated by data subjects
AuthorClaim		•	•
Digital Author Identifier (Netherlands)	•	•	
Frida/Cristin (Norway)		•	•
Lattes (Brazil)		•	•
Names Project (UK)		•	•
New Zealand ETC	•		
People Australia	•		
Researcher Name Resolver (Japan)		•	
VIVO		•	•

Allocation of responsibility for making changes to information in the systems also varies in line with the source of the data. Where individuals have access to records, as in AuthorClaim, VIVO, Frida and Lattes, authentication systems are used to ensure that they can only edit information relating to themselves. The Frida system has provision for a superuser within a university department who can make changes to all records generated within that department.

7. Assignment of identifiers

Most of the systems examined create identifiers when a new identity is established in the system. The Frida (now Cristin) system in Norway makes use of social security numbers from the human resources databases of contributing universities as its unique identifier.

System	Mint an identifier on creation of identity	Use (or adapt) an existing identifier
AuthorClaim	•	
Digital Author Identifier (Netherlands)	•	
Frida/CRISTIN (Norway)		• (social security number)
Names Project (UK)	•	
New Zealand ETC	•	
People Australia	•	
Researcher Name	• (where other	• (from other national research

System	Mint an identifier on creation of identity	Use (or adapt) an existing identifier
Resolver (Japan)	identifiers do not already exist)	databases)
VIVO	•	

8. Form of minted identifiers

Where identifiers are created from scratch, there are two broad approaches: to algorithmically generate an identifier (which may contain existing information about the data subject), or to use a simple running number.

System	Running number	Algorithmically assigned identifier
AuthorClaim	• (for public identifier)	• (for internal identifier)
Digital Author Identifier (Netherlands)		• (9 or 10 digit MOD 11 ² number)
Names Project (UK)	•	
New Zealand ETC	•	
People Australia	•	
Researcher Name Resolver (Japan)	• (where other identifiers do not already exist)	• (from other national research databases)
VIVO	•	

Most of the systems described here make their identifiers available as URIs. Several expose the records in one or more structured machine-readable formats and some support content negotiation (i.e. delivering different formats of records depending on the request: this allows for provision of data in HTML for web browsers and RDF for Linked Data³ browsers, for example).

System	Example URI	Support content negotiation and/or expose records as structured data
AuthorClaim	http://authorclaim.org/profile/pma2/	• Bulk FTP download
Digital Author Identifier (Netherlands)	info:eu-repo/dai/nl/123456785	
Frida/Cristin (Norway)		• XML export
Lattes (Brazil)	http://lattes.cnpq.br/4727357182510680	• XML export available to

² See http://en.wikipedia.org/wiki/MSI_Barcode#Mod_11_Check_Digit

³ See http://en.wikipedia.org/wiki/Linked_Data for an explanation of the Linked Data principles

System	Example URI	Support content negotiation and/or expose records as structured data
		Brazilian institutions on request ⁴
Names Project (UK)	http://names.mimas.ac.uk/names/individual/345	<ul style="list-style-type: none"> (Various export formats, including RDF)
New Zealand ETC	http://www.nzetc.org/tm/scholarly/name-207379.html	<ul style="list-style-type: none"> (RSS, EAC)
People Australia	http://nla.gov.au/nla.party-503844	<ul style="list-style-type: none"> (OAI, SRU)
Researcher Name Resolver (Japan)	http://rns.nii.ac.jp/nr/1000012345678	<ul style="list-style-type: none"> (RDF)
VIVO	http://vivo.ufl.edu/individual/n25562	<ul style="list-style-type: none"> (RDF)

9. Information maintained in systems

Details maintained on individuals vary according to the purpose of the identifier system. For the library-created New Zealand and Australian services, data can be very rich, including links to external resources, works which mention an individual, genealogical connections and dates of birth and death. The Australian service uses the Encoded Archival Context (EAC) standard as its native record format and this standard is also supported as an output format for records from the New Zealand Electronic Text Centre for its authority records.

The other, more researcher-focused, systems contain information which is more narrowly defined for that particular class of person, including funded research projects, publications, collaborators and institutional affiliations. There is not an internationally-agreed standard format for this type of data, although the Japanese system uses the fields supported by ThomsonReuters' ResearcherID service, with extensions for local needs.

10. Interaction with other systems

There are considerable differences of approach between the various initiatives examined for this report when it comes to interaction and data-sharing with other systems and services. It is difficult to make generalisations about these, so they are presented here in sequence.

AuthorClaim: Journal article information is provided to AuthorClaim by bibliographic data providers. Data within the AuthorClaim system is used by other services developed by the Open Library Society.

⁴ See <http://lattes.cnpq.br/english/conteudo/acordos.htm> for details

Digital Author Identifier (Netherlands): The National Thesaurus for Author names (NTA) interacts with the CRIS systems of universities which in turn interact with institutional repositories. These are harvested by the NARCIS metadata aggregator which provides a gateway to research and researchers in the Netherlands.⁵ The NTA information is subject to data protection laws and is not made directly available to third party sites.

Frida/Cristin (Norway): Cristin obtains information on researchers from the human resources systems of Norwegian universities, research institutes and regional health authorities (using social security numbers as unique identifiers internally) to create researcher profiles. Cristin is designed to be a national research information system, with the information within it available for use by funders, institutions and staff for the production of reports and statistics.

Lattes (Brazil): Lattes is integrated with a number of other systems, including SciELO⁶, LILACS⁷, SCOPUS, Crossref and universities' databases. Funding for researchers is dependent on having a profile in the Lattes system, which is administered by the federal bureau responsible for funding science and technology research, CNPq, in collaboration with other funding agencies, universities and research institutes. Data in the system is used by these different bodies to determine the allocation of funding, develop policies and make decisions on tenure and promotion of individuals.⁸

Names Project (UK): Data has been extracted from universities' returns to the 2008 Research Assessment Exercise and from UK institutional repositories. Sample data has been shared with the ISNI (International Standard Name Identifier) initiative for cross-matching. A plug-in for the EPrints repository software has been developed which provides an auto-complete facility using the Names API.

New Zealand Electronic Text Centre: Cross-matching of individuals (using EAC records) has been undertaken with Archives New Zealand and the Auckland War Memorial Museum. Within the local NZETC system, interaction is via XML topic maps, which include links out to external resources of relevance to the individuals and entities identified. Records on digital items in NZETC are made available to the Matapihi⁹ and DigitalNZ¹⁰ digital collection aggregation sites.

People Australia: The people and organisations data from People Australia is used by the National Library of Australia's Trove cross-searching service¹¹, the Australian National Data Service (ANDS)¹², the University of Melbourne eScholarship Research Centre¹³ and Party Infrastructure contributors. Interaction with the VIAF authority data is currently being tested.

⁵ <http://www.narcis.nl/>

⁶ Scientific Electronic Library Online: <http://www.scielo.org/>

⁷ Latin American and Caribbean health library: <http://lilacs.bvsalud.org/en/>

⁸ Lane, J. (2010) Let's make science metrics more scientific *Nature* Vol. 464 pp. 488-489

<http://www.nature.com/nature/journal/v464/n7288/pdf/464488a.pdf>

⁹ <http://www.matapihi.org.nz>

¹⁰ <http://www.digitalnz.org/>

¹¹ <http://trove.nla.gov.au>

¹² <http://www.ands.org.au>

¹³ <http://www.esrc.unimelb.edu.au/>

Researcher Name Resolver (Japan): The Researcher Name Resolver experimentally interacts with JAIRO, the federated search portal for Japanese institutional repositories, for disambiguating names of authors of items among institutional repositories in Japan.

VIVO: A correctly-configured installation of VIVO provides Linked Open Data which can be re-used in web pages and applications both within the host institution and externally. There is a prototype multi-institutional search using information exposed from the institutions using VIVO at <http://beta.vivosearch.org/>.

11. Data licensing

Most of the systems analysed here have open licences on their data, but this is not universally the case. The table below summarises the situation for the survey respondents.

System	Licence on data	Nature of licence
AuthorClaim	•	Creative Commons 0
Digital Author Identifier (Netherlands)	•	Data is licensed to OCLC (restricted by Data Protection Act). Access to the name thesaurus beyond users of the shared cataloguing system (GGC) is chargeable.
Names Project (UK)		None yet (intention is for data to be open)
New Zealand ETC	•	Metadata released without restriction. CC licence applied to public content where possible.
People Australia	•	Creative Commons Attribution-NonCommercial-ShareAlike 2.1 Australia
Researcher Name Resolver (Japan)		None yet
VIVO	•	Copyrightable information: Creative Commons Attribution 3.0 Non-copyrightable information: Creative Commons 0

12. Funding and development

The answers to the question 'How is the system funded?' produced another wide range of responses from the initiatives investigated. The findings are summarised in the table below.

System	Funding source
AuthorClaim	Volunteer labour
Digital Author Identifier (Netherlands)	Licence fees for use of the shared cataloguing system and CRIS

System	Funding source
Frida/Cristin (Norway)	Ministries of health and education
Lattes (Brazil)	National council for scientific research
Names Project (UK)	Project funding from JISC (through UK higher education funding councils)
New Zealand ETC	Developed as an internal project
People Australia	National Library of Australia operational funding
Researcher Name Resolver (Japan)	National Institute of Informatics
VIVO	Initial funding from Cornell University. Two year grant of \$12million from the National Institutes of Health to fund development by VIVO consortium in 2009-2011 (extended into 2012). Further grant proposals pending.

Of the systems examined, it appears that only the Netherlands' Digital Author Identifier currently approaches self-sustainability in funding. The majority receive funding from national bodies of one sort or another for maintenance and development activities. Not all of the systems are under active development at present. The table below captures the current situation in this respect.

System	Actively being developed	Comment
AuthorClaim	•	
Digital Author Identifier (Netherlands)		Currently in 'maintenance mode'. Looking to create APIs to open up the data in the future. Also interested in integration with federated authentication system.
Frida/Cristin (Norway)	•	
Lattes (Brazil)		Not clear from website whether Lattes is currently being developed or just maintained
Names Project (UK)	•	
New Zealand ETC		Former staff member continuing to develop system independently of the NZETC
People Australia		Actively maintained as part of the Trove infrastructure
Researcher Name Resolver (Japan)	•	
VIVO	•	

13. Integration with international name identifier initiatives

The library-based systems in Australia and the Netherlands are working or planning to work with OCLC's VIAF (Virtual International Authority File) service to match up their authors. The ISNI (International Standard Name Identifier) was mentioned as another possible service to align with by the Dutch respondent and test data has also been matched with ISNI by the UK Names Project. The main service of interest to the more researcher-based systems is the evolving ORCID initiative, which is being monitored by all concerned, with many being represented on the ORCID technical or stakeholder groups.

Of the services surveyed, VIVO is the only one to have funded an active development with ORCID, having used some of its National Institutes of Health funding to fund a mini-project to "explore the interaction of VIVO and ORCID in the scholarly identity ecosystem".¹⁴ Jon Corson-Rikert reports that "the project has developed a connector from VIVO to CrossRef for searching and pulling in publications and has explored the potential for VIVO to provide batch submissions of authoritative institutional information on researchers. In the future, we hope it will be possible for individuals to edit or supplement information in their ORCID profiles from VIVO and vice-versa using open identity exchange mechanisms".

All the systems which are making data available through open APIs with permissive licences are doing so with the intention that the information within them will be readily integrated with other systems, be they local, national, or international.

14. Conclusion

The systems examined in this report vary considerably in their levels of maturity, funding support and capacity for future development. Common to them all is an acknowledgement that unique identifications of individuals on the web is a necessary step to enable due recognition of the work of researchers and other creators of materials. Several of the more mature researcher-focused systems are integrated with other services such as national grant funders' databases, university research information systems and institutional repositories, performing an important role in promoting, assessing and enabling research. In the most mature of these systems, Brazil's Lattes, funding for researchers is only available to individuals who have registered a profile. As a consequence, Lattes covers a large number of individuals (over 1.6 million people) and is an integral part of Brazil's research infrastructure.

The importance of having publicly-available URIs for individuals and of making information available in standard, re-usable ways is understood by all of the individuals contacted for this report. There is also a general enthusiasm for open licensing, although in the Netherlands this has not been possible due to the model there of raising funds for the identifier service through licensing fees. In most of the other national researcher identifier systems, ongoing funding has been provided by a central national body of some kind, providing a degree of permanence and authority to the systems which is important for the adoption of identifiers and integration of them by other services.

¹⁴ <http://orcid.org/node/229>

Centralisation of data is another key theme of the approaches here. With the notable exception of the VIVO system developed by Cornell University, all of the researcher identifier systems examined here are managed by a central organisation. In Norway, Frida began as separate research information systems held at different universities but these have more recently been merged into a single CRIS for the whole country, including researchers outside the university system in health authorities and research institutes. The Lattes system in Brazil, on the other hand, was designed as a centralised national resource from the beginning. In the Netherlands, there is a central authority file for individuals which is then used by the universities' repository and CRIS services (it is worth noting that all of the universities in the Netherlands are using the same CRIS). With the VIVO system, it is up to universities to maintain their set of identifiers and to develop ways of dealing with researchers who move on to other institutions and who therefore may be represented in more than one instance of VIVO, causing potential duplication of individuals in a combined search across VIVO installations in the future.

The administrators of the systems focusing on the identification of researchers are all aware of the evolving ORCID initiative and are intending to interoperate with this when it becomes available, although there is still a degree of uncertainty in most cases as to how this will be achieved. With the international nature of research, ORCID will perform an important role in identifying individuals beyond national boundaries. The advantages of a national identifier system include the ability to tailor it for interaction with other national systems such as those of funding organisations and statistics-gathering bodies and with institutional systems of universities, research institute and health authorities, as has been the case in the more mature systems of Brazil, Norway and (to some extent) the Netherlands. ORCID and other international initiatives such as the International Standard Name Identifier (ISNI) and OCLC's Virtual International Authority File (VIAF) offer further opportunities for integrating national systems into a global research context.

15. References

Websites for initiatives mentioned in this report

AuthorClaim	http://authorclaim.org/
Digital Author Identifier (Netherlands)	http://www.surffoundation.nl/en/themas/openonderzoek/infrastructuur/Pages/digitalauthoridentifiertai.aspx
Frida/Cristin (Norway)	http://www.cristin.no/english/
ISNI (International Standard Name Identifier)	http://www.isni.org/
Lattes (Brazil)	http://lattes.cnpq.br/
Names Project (UK)	http://names.mimas.ac.uk/
New Zealand Electronic Text Centre	http://www.nzetc.org/
ORCID	http://orcid.org/
People Australia	https://wiki.nla.gov.au/display/peau/About+People+Australia

Researcher Name Resolver (Japan)	http://rns.nii.ac.jp/
VIAF (Virtual International Authority File)	http://viaf.org/
VIVO	http://vivoweb.org/

Articles

Kurakawa, K., Takeda, H., Takaku, M., and Aizawa, A. (2009). Researcher Name Resolver: A framework for researcher identification in Japan. *The 5th International Conference on Open Repositories OR2010*
<http://www.or10.es/Resources/documentos/GSabstracts/or2010-kurakawa-v1.pdf>

Lane, J. (2010) Let's make science metrics more scientific *Nature* Vol. 464, pp. 488-489
<http://www.nature.com/nature/journal/v464/n7288/pdf/464488a.pdf>

Lingjærde, G. C. & Sjøgren, A. M. F. (2010) Remodelling Frida – From institutional registration to common registration and responsibility across member institutions *EuroCRIS Conference*
http://www.eurocris.org/Uploads/Web%20pages/cris2010_papers/Papers/cris2010_Lingjaerde.pdf

Norrish, J. & Stevenson, A. (2008) Topic Maps and Entity Authority Records: an Effective Cyber Infrastructure for Digital Humanities, (paper presented at Digital Humanities conference, 2008)
<http://researcharchive.vuw.ac.nz/bitstream/handle/10063/336/paper.pdf?sequence=2>