

Digital Author Identifiers summit

13 and 14 March 2012, London

Summary

On 13 and 14 March Knowledge Exchange organised a summit on digital author identifiers. It was well attended by a broad group of international experts, with 31 participants from 10 countries. The first day was directed at identifying the issues and the second day at using the findings and opinions to feed into the development of international initiatives being developed (notably ISNI and ORCID). There are already national identifier systems in place in several countries and there was a strong interest at the summit in connecting these internationally.



Figure 1: Participants at the summit

The summit clearly showed that there is value in aligning and connecting current systems. There are already discussions underway between ORCID and ISNI and VIAF feeds directly into ISNI. These systems are engaging with different groups (ORCID with publishers/researchers, ISNI/VIAF with national libraries) and have different business models but it seems possible that a co-ordinated approach would be feasible. This would definitely be to the benefit of researchers, as well as those working on information infrastructure and research administration.

Key recommendations were:

- All parties should work towards preventing redundancy. It would be great to have one canonical ID bringing together existing systems.
- There is an interest in an open thin layer with clear interfaces so others can build services on this.
- At present there are broadly two approaches to collecting researcher IDs. Solution providers should draw on the relevant strengths of both of these approaches.
- Now is the time for institutions to start doing their homework. They should not make blocking choices but progress and start assigning identifiers and work on linking these with VIAF

Day one: Identifying issues

Setting the stage

Neil Jacobs from JISC welcomed all participants to the summit¹. The value of digital author identifiers (DAI) is now being acknowledged and we live in interesting times with the development of ORCID², ISNI³, other international initiatives⁴, disciplinary systems⁵ and various national identifiers⁶. Bas Cordewener then went on to introduce Knowledge Exchange (KE) and the working group of interoperability of digital repositories.

[Clifford Lynch](#) from CNI⁷ provided his thoughts and perspectives from the US. He considers DAIs to be necessary infrastructure. They are key in providing good quality accurate bibliographic information. At present however there are still substantial quality problems in both established commercial databases and the free citation services (e.g. Google Scholar, Microsoft Academic Research) that are still in early



Figure 2: Clifford Lynch

¹ All presentations, photographs of flipover sheets and links to tweets and blog postings are available on the Knowledge Exchange website at: <http://www.knowledge-exchange.info/Default.aspx?ID=498>

² Open Researcher and Contributor ID: www.orcid.org

³ International Standard Name Identifier, www.isni.org

⁴ For example author claim: <http://authorclaim.org/>

⁵ For example RePec: <http://repec.org/>

⁶ For example [Names](#) in the UK, [TROVE](#) in Australia, [DAI](#) in the Netherlands, [Dissonline](#) in Germany, solutions in Norway and Denmark

⁷ Coalition of Networked Information; www.cni.org

stages of development. One motivating driver can be the wish of institutions and funders to employ metrics such as citation counts and measuring impact factors. This will allow for the rating and ranking of university staff (tenure and promotion) and departments (for funding purposes). A second big driver is the development of social networking in research, for example VIVO⁸. A third is efforts to develop personalized filtering software to identify important publications that a time-constrained researcher should read.

The awards and grants record needs to be thought of as part of a bibliographic record. This does raise questions regarding what should be public and what not. And, at least in the United States, campuses have almost completely failed to connect even the names used for publication to the institutional identity management systems, which include the names used for paychecks, or registration. Cliff presented several issues which should be kept in mind:

- We need a system that spans past, present and future and therefore also connects to historic literature. Ultimately it will extend to, or at least connect to, other historical uses of biography and names as well.
- There are scope issues: does it cover only “scholarly” authors or does it include other authors? There is a strong pressure to be more rather than less inclusive.
- How should we treat anonymity and pseudonymity?
- How should we clean up the existing files in a way that minimizes redundant human intervention?
- How can we reverse course and attach biography for earlier authors?
- Accommodating new forms of scholarship: what about blog posts and tweets, how far shall we go?
- Are links required for authentication and identity management? This might be needed to provide clarity on provenance, transparency and accountability. We started off in a benign world in which this all took place in good faith. If it is presented as the university record, who stands witness? This should be built in from the start.
- There is a challenge in deciding who can determine what is made public. Particularly on scientific misconduct, in which case it would be important that someone other than the author decides.
- The institution will care for a while but might lose interest. Who will then take ownership? A challenge in this respect is the handing down of information from one institution to the next. The current employer is probably the party that cares.
- Don't be limited in what you consider to be an academic, even undergraduates may have scholarly publications. And what about academics that go into business or work at two institutions? A system based solely on higher education institutions won't work.

In addressing some of these challenges perhaps we could work on codes of practice for inter-institutional transfers. It might be worthwhile to consider where to leave the publications which are in the repositories. Current practices are not codified, so this would be a good idea.

Nicky Ferguson presented his work for JISC on the Researcher Identifier Task and Finish Group⁹. This consisted of gathering information on stakeholder use cases and identifier needs. Many of their considerations proved very relevant to the international discussion.

In general everybody was interested having an (non-proprietary) ID, especially to prevent filling in data several times. But to provide a clear case for adoption the benefits or business case do need to be spelled out. The ID should be free of semantic information. The stakeholders saw no reason why you would need a separate researcher ID and an academic ID. On the other hand, the researcher would not have to be member of staff. The ID should be internationally operable and should be able to connect with semantic web services in the future. The adoption can be started without taking final decisions on

⁸ See for more information: <http://vivoweb.org>

⁹ See for example the report at: <http://ie-repository.jisc.ac.uk/568/1/report1-final.pdf>

all questions yet. It should be attractive to allow researchers to self-allocate an ID before having to approach the university. No choice for one system was made, separate systems should be able to connect and exchange.



Figure 3: Andrew Treloar

[Andrew Treloar](#) from ANDS¹⁰ added some further questions to consider. A DAI raises questions about identity and what assertions you want to make. Who is in scope: is it researchers, paid researchers? At a real institution? And what about graduate or undergraduate students or assistants? Also the context does matter, do you want to present the same information in different contexts? What can be used as the source of truth? Will you use a human resource system, an author system, publishing systems, national or discipline systems or large facilities? Which of the spheres do we care about? Finally, when considering digital author identifiers, we need to distinguish between the infrastructure that allows assertions to be

made about who an identifier refers to, the act of making these assertions, and who is making the assertions.

Issues and opportunities

In a panel discussion the perspectives and interest of various stakeholders were presented. The panellists were Andrew McEwan (British Library), Thom Hickey (OCLC), Chris Shillum (Elsevier), Maurice Vanderfeesten (SURF), and Clifford Lynch (CNI). First they were asked to name current issues, which provided several responses. It was considered valuable to support the interchange of biographies. Institutions also want clarity on governance (who controls) and who is going to pay (business model) for the infrastructure. There can be a conflict between openness and business models. There is a need for a piece of infrastructure to connect overlapping systems. This is where ORCID could play a role; not by providing value adding services, but by building as lightweight a system as possible. ISNI is also developing and is welcoming the input of more people. It is working on taking the existing IDs, and also working on legacy and historic material.



Figure 4: Panel discussion underway

The panel members also considered DAIs as infrastructure that could unlock opportunities. For the different stakeholders different opportunities are on offer. DAIs can connect the pieces of the jigsaw by providing the ability to use a record of human achievement from the past and into the future. Through DAIs authors will also have access to more value added services, uncovering new databases. DAIs can play a role by reducing the friction in the system (this was a recurring theme during the summit) and make services easier and lower cost especially on the discovery side. Besides doing old things better it will also enable new possibilities.

¹⁰ Australian National Data Service; <http://ands.org.au>

Magic wand session

All participants were asked to present their three main wishes for author identifiers if they were given a magic wand. The wishes were collected and summarised. In the field of governance there was a repeated wish to align international and national initiatives. In the field of interoperability there was a clear wish for openness. This would allow for value added services which could be generated separate from the ID.

In a summary of the discussions several tensions were identified:

- Should we use a broad versus a narrow definition of an author?
- Should the ID be free of meaning or should it contain a country or institution code?
- Should the system be author controlled or institution controlled?

Breakout sessions

Participants were split into three breakout groups to discuss the business case for author identifiers. This was addressed from the supply side, demand side and the governance perspective. The groups were asked to draw rich pictures to illustrate their ideas.

Supply side

From the perspective of what could be provided, not only author identifiers are important. Also object and organisation identifiers are required to provide valuable services. It is interesting to consider what the critical mass is that is required from the demand side. What value less than 100% is acceptable? This can vary for different use cases, there can be different quality thresholds. Stated differently: how dirty can the data be for specific uses?

Demand side

On the demand side it was considered important to allow for the splitting of identities and merging multiple identities. From the functional perspective it is important to consider all stakeholders that use the record, this includes the funder, the institution, the researcher, but also possible discovery services. For example undertaking a research evaluation is a huge amount of work and a DAI could be instrumental in improving the process. From the economic perspective one needs to consider that actors will have their own systems. A framework might well be required to provide the answer.

Governance

The first observation of this group was that a DAI can be generated or maintained in the cloud. IDs can be obtained by an individual or an institution. A model would be that the individual provides permission for the institution to exchange information about the individual. This is a challenge when looking at an ID that persists past the grave.

A further message was that the word governance might convey a wrong message, the solution could lie in a more federated structure. This might well require trust and to achieve this centrality in rules of engagement, perhaps a central agreed policy.

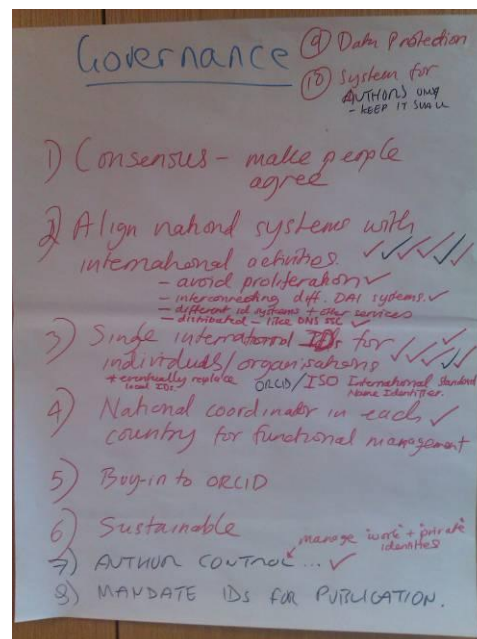


Figure 5: Summary of wishes in the field of Governance

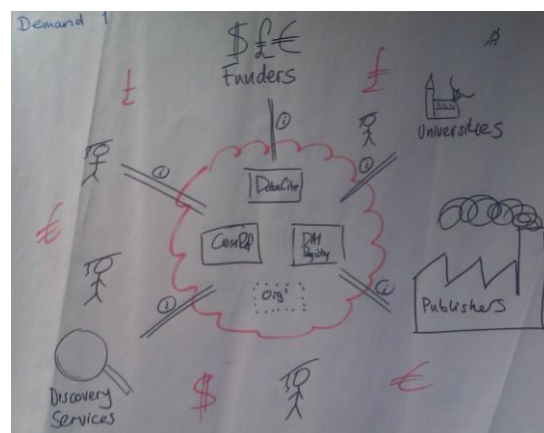


Figure 6: Rich picture from the demand breakout group

Summarising day one

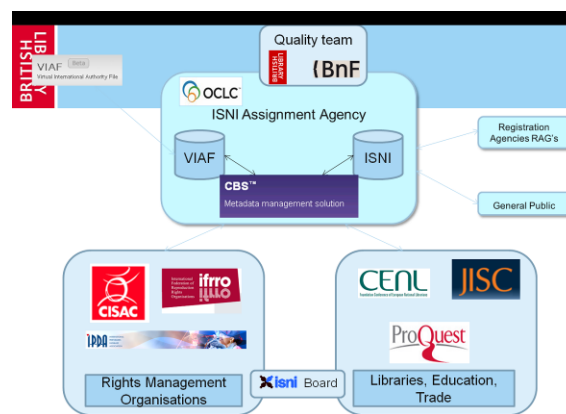
Collecting the issues and challenges had provided the impression that there were still a lot of issues to be resolved and questions to be answered. Questions had been raised on whether a centralised or a federated solution was required. There was a clear interest in DAIs and recognition of the added value they could offer to the various stakeholders. A digital author identifier would not only reduce friction and allow to do things better but also allow for new things. Questions had also been raised regarding interoperability, is there going to a single DAI? Perhaps the infrastructure could be constructed like the infrastructure required for the web (e.g. DNS). In the case of the web it was hard to make the argument for such a thing before it was in place, but once this had occurred a range of services became possible. The same should be the case for DAIs. Finally the business case should be directed to different parties as different stakeholders will have different requests.

Day two: Taking action

Day two was set up to identify actions which could be taken. Though several challenges had been identified on day one there were initiatives already working on connecting author identifiers. To offer background for the discussions the various solutions present briefly presented their current status.

ISNI

Andrew McEwan provided some information on the development of [ISNI](#) which is an ISO standard for author names. It is also an assignment agency that is connected with an infrastructure. The initial database has been created and the first million ISNIs have been assigned. It is based on VIAF (see below) and amongst other things is connected with the JISC Names project. It will register anybody that might be listed as an author in a library and also has connections with rights management organisations. Various views on the database are possible including a highly filtered view for the public. It is a variable registration agency with varied pricing. ISNI is at present discussing interoperability with ORCID.



Figur 7: Slide showing ISNI connections

ORCID

Chris Shillum from Elsevier, also a board member for the Open Researcher and Contributor ID (ORCID) initiative, provided some information on the development of the initiative. [ORCID](#) has been set up to act as glue between systems in this space. It is intended as an open and transparent system, built using Open Source software. The data will be made available under a CC-0 licence at least once a year. It is a non profit organisation based in the US. At present there are 300 participants in the initiative. Once the membership arrangements are clear a selection process will take place. The underlying principle used is that the author is in charge of their own ORCID. By the end of the second quarter of 2012 the technical development of the first system should be complete. This will be based on author self-claim. A use case for the first version will be that publishers will connect ORCID to DOIs¹¹ in Crossref.

The phase 2 system is still in concept, and they are looking into ensuring that assertions can be established on authors (e.g. which author works at which institution and the provenance of this assertion.) The initiative may well be funded through a mix of sponsorships and loans. Membership fees would be starting in 2013.

VIAF

Thom Hickey from [OCLC](#) presented the Virtual Internet Authority File (VIAF)¹². This initiative was started 10 years ago and is run by libraries and OCLC. It will become an OCLC service in the future. VIAF covers author names and organisations. Data is made available according to linked data

¹¹ Digital Object Identifiers, see <http://www.doi.org/>

¹² See <http://viaf.org>

principles and can be used for auto completion of fields when entering information. Information from VIAF feeds into ISNI. 90% of access is through APIs and two thirds of use is European. The web service is also actively used.

VIVO

Clifford Lynch from CNI provided an overview of [VIVO](#). VIVO was started by Cornell to enable the discovery of researchers and encourage collaboration within the institution. More recently, funding was made available from the National Institutes of Health to make it into a cross institutional network, and it is being deployed in number of institutions, with some emphasis on medical faculty. VIVO is open source software that can connect information made available according to semantic web standards. It is also being used outside of the US (e.g. in Australia as a solution supported by ANDS), and interest has also been shown from other countries.

Another development in the US is also relevant. [Kuali](#)¹³ is a community of universities, colleges and commercial firms partnering to develop open-source software for higher education; much of their early work has been in core administrative systems. They are now building [Kuali Coeus](#) which provides extensive modules for research management.

National initiatives

Participants are almost all working on author identifiers, some at the national level. These solutions are at different states of maturity. In some cases the IDs have been connected to grant information and are quite complete. This for example is the case in Denmark; however the bibliographic information is not open at present.

Breakout session: working towards recommendations

The participants were split up into breakout groups focusing on three perspectives: governance, interoperability and added value. After analysing the opportunities they were asked to agree on strategic approaches and activities which could be carried forward by the different players in this field. They came up with the following recommendations:

Recommendations

	Governance	Interoperability	Added Value
To solution providers:	<ul style="list-style-type: none"> • Important that business model allows for open and re-usable data • Business model needs to ensure sustainability of service/data • Alignment needed between business model and possible funding sources • Provide clear and transparent locus and scope of control (Need to clearly communicate their plans) • Aim for wide community consultation (KE and 	<ul style="list-style-type: none"> • ORCID, ISNI, etc. should seek to cooperate and to coordinate their activities: one should become the 'canonical' ID. (Running both will come with costs but it may make sense to leave market forces to make the choice.) • Undertake the mappings, disambiguation to the canonical ID – canonical IDs linked to local IDs (with provenance). • Define the APIs and interfaces to make all the other services work properly 	<ul style="list-style-type: none"> • Develop concept for IDs in local setting • Provide standard way of adding author IDs to various (reference management) systems. • Exposing the data. Letting others know what you can do with the data (providing APIs) • Everyone who produces research objects needs to have a concept for author information • Support linking to object IDs (could be to any contributor)

¹³ See <http://kuali.org>

	<p>partner organisations, CNI, ANDS) could assist with this)</p> <ul style="list-style-type: none"> • Encourage existing solutions providers to continue discussions around collaborative / complementary activity • Explicitly discourage redundant and competitive behaviour (e.g. where individuals/institutions /national solutions connect in) 		
To KE partners:	<ul style="list-style-type: none"> • Identify importance of relationship between KE partners and relevant national libraries (and in this domain they need to work together) • Investigate possible role for research administrators /managers at institutions • Draw on the relevant strengths of the library-focussed and researcher-focussed approaches • Identify roles and responsibilities and use to structure relationships with the solution providers 	<ul style="list-style-type: none"> • Undertake the mappings, disambiguation to the canonical ID (canonical IDs linked to local IDs (with provenance)) • The KE partners can contribute to the bootstrapping using their existing IDs.(This needs a big dataset to start it running - whoever is the canonical ID will have a set of constituents) • Define the APIs and interfaces to make all the other services work properly 	<ul style="list-style-type: none"> • Explore linking to object IDs (could be to any contributor)
Those with existing systems	<ul style="list-style-type: none"> • Link existing authority files with VIAF • Do your homework on what is happening • Assign identifiers now • Don't make blocking choices 		<ul style="list-style-type: none"> • Everyone who produces research objects needs to have a concept for author information • Explore linking to object IDs (could be to any contributor)

The recommendations were supported by the participants at this meeting. It was agreed that it would be worthwhile to send these to the current solution providers, the partners and others with existing systems. These recommendation were also considered of interest to funders of research and research information infrastructure, including the EC.

In the discussion some challenges were addressed. For example, the choice for one system or several systems does have repercussions for governance. It was agreed that agencies or services should talk

and find a common denominator approach. They should talk about terms of connecting. The KE partners would be happy to offer help if this was required.

Closing remarks

Clifford Lynch closed the summit with some reflections. He considered the recommendations to be very helpful, also for the situation in the US. They send messages in several directions. It can be a starting point to bring groups together to talk to each other (e.g. IT staff, library staff, research funders, research administrators, research groups). ISNI and ORCID are both young organisations, figuring out their way. One issue which should not be forgotten is the challenge of assigning and managing author IDs of deceased authors; ultimately this will bring not just librarians but a wide range of humanistic scholars into the discussion. To this end some scale scoping could be done identifying how big the retrospective problem is and how hard it will grow. He closed by stating that new, network-based and recombinant bibliographies and biographies are obvious and high-payoff opportunities for quick wins for author IDs.