



CRIS 2014

## CASRAI and ORCID: Putting the pieces together to collaboratively support the research community

Laure Haak<sup>a\*</sup>, David Baker, and Thorsten Hoellrigl<sup>c</sup>

<sup>a</sup>*ORCID, Bethesda, United States*

<sup>b</sup>*CASRAI, Ottawa, Canada*

<sup>c</sup>*Thomson Reuters, Karlsruhe, Germany*

---

### Abstract

Researchers and the organizations that support research are stymied by data that are inconsistently specified. Incentives to share data go together with mechanisms to support interoperability. Both are starting to gain traction with the development and implementation of shared standards in research data exchange. The Consortia Advancing Standards in Research Administration Information (CASRAI) provides a peer-reviewed, open dictionary of terminology for the semantics and record-structures of research information. ORCID provides a persistent registry for researchers to obtain a unique identifier and, like CASRAI, works with the community to embed these identifiers in research workflows. Coupled with the CERIF model, which has been adopted as a structural model for research management systems by the European Commission, and CrossRef publication and DataCite dataset identifiers, these underlying exchange standards and services comprise a framework that supports open access and acknowledgement of researcher contributions. In this paper we describe a recent effort to ensure that information exchanged between systems meet the needs of both researchers and data consumers.

© 2014 The Authors. Published by Elsevier B.V.  
Peer-review under responsibility of euroCRIS.

*Keywords:* Persistent identifiers, digital object identifiers, data exchange standards, research administration, CRIS, ORCID, CASRAI, CERIF

---

---

\* Corresponding author. Tel.: +1-301-922-9062  
*E-mail address:* [L.haak@orcid.org](mailto:L.haak@orcid.org), ORCID: <http://orcid.org/0000-0001-5109-3700>

## 1. The problem: Disconnected data

The modern research enterprise is highly collaborative, distributed, and partnership-driven. For this research ecosystem to function, there are myriad research administration processes at every stage of the research lifecycle that must be managed and connected. For the purpose of this article, research administration means any work processes that are necessary to the overall research enterprise but that are not actually involved in ‘doing research’. For researchers, students and research teams, this includes (but is not limited to):

- Finding and establishing collaborations,
- Creating a publicly available presence to show expertise
- Finding potential journals and conferences to participate and submit papers too
- Discovering experts
- Developing and iterating project proposals for funders, often amongst multiple institutions with diverse administrative structures and expectations
- Creating and managing CVs
- Engaging industry and community partners, including submission of the CVs of external partners which may be in various formats
- Discovering well-suited funding opportunities
- Applying and reporting to funders,
- Serving on peer/merit review committees, with the concomitant need to access information on initiatives that have already received funding,
- Managing project-related tasks to meet compliance (conflicts of interest, ethics, biohazards, human and animal care, clinical trials),
- Finding and accessing equipment and infrastructure,
- Interacting with publishers and libraries,
- Reporting to repositories and research offices
- Disseminating work to multiple channels, and
- Managing inventions, licenses and patents
- Translating results into broader societal impacts.

Researchers are integrally involved in each of these administrative processes, enmeshed with research organizations, funders, and publishers in an effort to promote quality, productivity, accountability, and societal returns. The community is challenged with obtaining robust data on the efficiency and effectiveness of research operations and evidence of the difference their interventions make without imposing undue administrative burden. While these data are fundamental to research management, currently available information can rarely be compared, exchanged (in a comparable format), reused, or analyzed.<sup>1</sup> Research teams and administrative personnel, for example, must re-type the same information repeatedly when applying for funding or reporting. Research policymakers, managers and evaluators are consistently frustrated by an inability to draw meaningful conclusions from a growing mountain of ‘disconnected’ data.

## 2. Developing a common infrastructure

It is clear that the research community would benefit from the ability to “connect the dots” between researchers, organizations, and research products. Research administration systems, also known as Current Research Information Systems (CRIS) help to connect the dots by supporting the researchers, students and research teams with administrative tasks throughout the research lifecycle.<sup>2</sup> They help to compile and manage research outputs by reusing existing data from different systems by import, migration and integration, but also enable to manually enter high quality data. Improvement is still needed regarding the interoperability of those systems. Currently a lot of effort is required to map different information models semantically and syntactically, as well as identifying and

resolving duplicate records.<sup>3,4</sup> Interoperability between research management systems would improve both management and discoverability of research, and reduce the reporting burden on researchers. Only a few core components, if adopted across the community, could enable this vision to be realized: unique and persistent identifiers for researchers, their research products and activities, and organizations, and machine-readable means to interpret and exchange these data between data systems.

CASRAI (Consortia Advancing Standards in Research Administration Information - <http://casrai.org>) has been working to develop a common dictionary<sup>5</sup> to support exchange of research administrative data. The CASRAI mission is to advance global standards that enhance the management and flow of information within and between organizations collaborating in research and innovation. To this end, CASRAI provides an open framework comprised of

- a. a network of stakeholder organizations that collectively develop, set, and advance data standards for the research administration domain,
- b. a dictionary of terminology and exchangeable business objects, and
- c. toolkits to guide the community in implementing the CASRAI dictionary and enabling data flow within the research administration ecosystem. International CASRAI working groups have supported the development of data standards for topics including identifiers and related metadata descriptions for researchers, organizations, and data.

An initiative to provide an international registry of unique and persistent identifiers for researchers has been launched by ORCID (<http://orcid.org>).<sup>6</sup> This initiative is built upon the principle that researchers control what information is connected to their identifier and how that information is shared. At the same time, for an identifier to be useful, it must become part of regular research workflows and systems. ORCID identifiers connect researchers and scholars with their contributions by incorporation into the processes for making public articles, datasets, theses, grants, and other research works. One example of this is the ODIN project collaboration between ORCID and DataCite,<sup>7</sup> which has led to the integration of person identifiers into the DataCite metadata schema 3.0, a list of core metadata properties chosen for the accurate and consistent identification of data for citation and retrieval purposes, along with recommended use instructions. This means, all new research datasets can be associated with both a DOI and ORCID for the contributors. The collaboration has also produced tools for researchers to link extant datasets with their ORCID identifier, allowing researchers the opportunity to bring together their research datasets, past and present. In addition to works, ORCID supports connections with organizational identifiers and other person identifiers.<sup>8</sup> Brought together through interactions in standard workflows, these *ex ante* connections between people, places, and things can be validated at the source and are made available generally through application programming interfaces (APIs).

### 3. Use cases and test beds for research data exchange

While a digital object identifier such as ORCID can be used to resolve an entity, it does not in and of itself define an entity. Systems need to collect, store, and share the identifier, the identifier source, and a set of data fields to describe the entity being identified. ORCID has focused on establishing the minimal set of metadata to accurately describe and resolve a research contribution. For example, to link with research works, ORCID started with a set of data fields that included title, author, author role, document identifier, work type, publication date, and other information that was coded by work type. We found quickly that data consumers wanted more and different fields. We formed a Works Metadata technical working group, chaired by Richard Rodgers of MIT, a few months after launch of the ORCID Registry, and charged it with examining the works metadata with respect to fitness for purpose in the ORCID registry and utility to producers and consumers of such metadata.<sup>9</sup> During these discussions, a few broad themes emerged. The first was that machine-to-machine, ‘M2M’, use cases (registry interoperability with supplying or consuming automated information systems) furnished important drivers to metadata representations. The Working Group strongly argued for actionable (resolvable) identifiers, rather than opaque strings. For example, a parsable BibTeX citation would possess much higher value in this context than an unstructured text representation, as would a DOI over an article title. A second, related point was that works metadata interoperability rose in

proportion to its alignment with standard, recognized, published and maintained vocabularies (ontologies). Several specific efforts were cited, including CASRAI, the euroCRIS (<http://www.eurocris.org>) family of standards and the SPAR (<http://sempublishing.sourceforge.net/>) ontologies.

The central insight was that where they were defined and relevant (for example, in work type taxonomies, or contributor roles), use of standard terms could greatly enhance the semantic reach and reuse of ORCID registry metadata. In the near term, ORCID has re-aligned its metadata specifications for works to align with CASRAI and euroCRIS standards where possible, such as the Common European Research Information Format (CERIF).<sup>10</sup> CERIF is a data model under the custodianship of euroCRIS that allows for a representation of research entities, their activities and their output whilst maintaining high flexibility with formal relationships, enabling quality maintenance, archiving, access and interchange of research information and supporting knowledge transfer to decision makers for research evaluation, research managers, strategists, researchers, editors and the general public. CERIF is a standard that is already widely used in different federated scenarios, in which research information is exchanged. An example for this is SweCRIS a joint initiative between 10 Swedish universities that uses the CERIF standard data model to exchange grant funded research data from Swedish universities and research organizations.<sup>11</sup> The national platform is intended for end users to search for research projects across the country in different fields, to track the funding from different bodies or analyze that received by different researchers and to look at collaboration between researchers and institutions, all of which is made possible only by using the CERIF standard data model.

The ORCID Works Metadata Working Group recommendations are helping inform the ongoing process of improving ORCID registry services, and they provide a roadmap for future analysis and directions in ORCID works metadata, one of which is determining how to acknowledge peer review activities.<sup>12</sup> After being contacted by several organizations about linking peer review documents to ORCID identifiers, ORCID decided to engage with CASRAI and leverage its community Working Group process. CASRAI has established a group, co-chaired by F1000 Research and ORCID to review existing standards for applicability, and ultimately recommend data fields for citation and exchange of data about peer review activities, as well as propose workflows for associating the review and reviewer with persistent identifiers. The group's recommendations will be submitted to an external review circle prior to finalization, and then CASRAI will translate into fully defined record-types, fields, and classifications to be added to the CASRAI dictionary for open use by the community. ORCID will be using the group's recommendations to develop methods for linking review activities with ORCID identifiers and for posting review metadata to the ORCID Registry. This is just one example of how the CASRAI dictionary and associated standards working groups can assess common approaches to research management challenges. Other CASRAI projects include standards for Data Management Plans, Research Classification codes, Research Impacts, Organizational IDs, Authoritative Lists, Funding Results Announcements, Research Services Office KPI, and Contributor Roles

Another use case that would benefit from shared standards is the ability to express linkages between identifiers in a CV format. Each discipline and research organization has its own format and data needs, so creating one CV interface for the entire community is neither practical nor desirable. Rather, we need to think about using a “container” to carry information used in a CV to a variety of interfaces. This is precisely what the CASRAI CV standard supports.<sup>13</sup> The CASRAI standards recognize that data about people comes in many forms beyond a one-size-fits-all CV; for example, abridged versions, student versions, and internal reporting versions. In CASRAI, the concept of a CV is a collection of data profiles suited to the purpose. If both source and destination databases use this standard, it is then possible to query across multiple systems to gather data and then transform and express in any number of relevant formats. This is the goal of platforms such as SciENCV<sup>14</sup>, a cross-agency initiative in the US to gather information to support grant applications. It is also the process by which the Portuguese Fundação para a Ciência e a Tecnologia (FCT) are building a national research information management infrastructure by coupling a CERIF-compliant CRIS with ORCID identifiers to ensure interoperability among the systems involved.<sup>15</sup> Local CRIS systems are also looking to data standards and persistent person-place-contribution identifiers to streamline data exchange between university systems, such as thesis repositories, library collections, disciplinary data repositories, and multiple instances of faculty profile systems launched in different departments and schools.

Together, persistent identifiers and data exchange standards will move the research community toward an “enter once use many times” model, reducing the current burden for reporting and improving data quality and discoverability, desirable outcomes for researchers and the community broadly.

## References

1. Haak LL, Baker D, Ginther DK, Gordon GJ, Probus MA, Kannankutty N, Weinberg, BA (2012) Standards and Infrastructure for Innovation Data Exchange. *Science* **338** (6104): 196-197.
2. Joerg B, Ruiz-Rube I, Sicilia M-A, Dvorak J, Jeffrey K, Hoellrigl T, Rasmussen HS, Engfer A, Vestdam T, Barriocanal EG (2012) Connecting close world research information systems through the linked open data web. *Int. J. Software Engineering and Knowledge Engineering* **22**: 345-364.
3. Lai R, D'Amour A, Doolin DM, Li G-C, Sun Y, Torvik VI, Yu AZ, and Fleming L (2013) Disambiguation and Co-authorship Networks of the US Patent Inventor Database (1975–2010). Working Paper. [http://funginstitute.berkeley.edu/sites/default/files/Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database \(1975-2010\)\\_0.pdf](http://funginstitute.berkeley.edu/sites/default/files/Disambiguation%20and%20Co-authorship%20Networks%20of%20the%20U.S.%20Patent%20Inventor%20Database%20(1975-2010)_0.pdf).
4. Weinberg BA, Owen-Smith J, Rosen RF, Schwarz L, McFadden Allen B, Weiss RE, and Lane J (2014) Science Funding and Short-term Economic Analysis. *Science* **344**(6179): 41-43.
5. CASRAI Dictionary, <http://dictionary.casrai.org>.
6. Haak L, Fenner M, Paglione L, Pentz E, Ratner, H (2012). ORCID: a system to uniquely identify researchers. *Learned Publishing* **25** (4): 259-264.
7. ODIN Project Website, <http://odin-project.eu>.
8. Paglione L (2013) Organizational affiliations are not part of ORCID record. ORCID Blog, 9 December, 2013. <http://orcid.org/blog/2013/12/09/organizational-affiliations-now-part-orcid-record>.
9. Paglione L (2013) Works Metadata: Recommendations from our Working Group. ORCID Blog, 9 September 2013. <http://orcid.org/blog/2013/09/09/works-metadata-recommendations-our-working-group>.
10. CERIF Releases, <http://www.eurocris.org/Index.php?page=CERIFreleases&t=1>.
11. Sweden ScienceNet webpage. <http://www.sciencenet.se/converis/static/about>.
12. Haak LL (2014) ORCID and CASRAI: Acknowledging peer review activities. ORCID Blog 8 April 2014. <http://orcid.org/blog/2014/04/08/orcid-and-casrai-acknowledging-peer-review-activities>. SweCRIS
13. CASRAI CV modeller. <http://dictionary.casrai.org/documents/academic-funding-cv/1.1>.
14. Hutcherson KL (2013) My NCBI Curriculum Vitae web application: ScienCV. *NLM Technical Bulletin* **394**:e3. [http://www.nlm.nih.gov/pubs/techbull/so13/so13\\_sciencv.html](http://www.nlm.nih.gov/pubs/techbull/so13/so13_sciencv.html).
15. deCastro P (2014) Building pioneering functionality around ORCID integration: FCT and Portugal. GrandIR blog, 18 February 2014; <http://grandirblog.blogspot.com/2014/02/building-pioneering-functionality.html>.