ODIN: ORCID and DataCite Interoperability Network

DASISH workshop on Persistent Identifier Services

December 2014

Rachael Kotarski – British Library

DataCite

- You heard about this from Brigitte yesterday!
- The British Library is the UK allocation agent for DataCite DOIs, working with UK organisations to assign DOIs to their data
- The British Library is the national Library of the UK. We are a legal deposit library and cover all disciplines
- We also have an active role in other persistent identifiers:
  - International Standard Name Identifiers (ISNI)
  - International Standard Serial Numbers (ISSN)

ORCID

- Open Researcher and Contributor ID (ORCID)

"*ORCID is an open, non-profit, community-based effort to provide a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers*"

- Other name identifiers exist:
  - International Standard Name Identifier (ISNI)
  - Country Based ID's

ORCID
Connecting Research and Researchers

isni

The Names Project

What did ODIN aim to achieve and why

- To support development and stimulate adoption of interoperable identifiers for:
  - Researchers
  - Inputs (cited work, data)
  - Outputs (publications and data)
- Facilitate information flow within and between research communities, for greater re-use of data and exploitation of the knowledge created.
- Requirement for a sustainable and participative persistent identifier e-infrastructure in support of data-intensive open science
- ORCID and DataCite are emerging as participative initiatives which, if linked, can play significant role in underpinning such e-Infrastructure
- ODIN proposed to explore the opportunities, highlights gaps and roadmaps, and nurture interoperability solutions, globally, and for specific disciplines and beyond
- Identify and compare workflows for identifier integration between Humanities and Social Science and High Energy Physics

ODIN workstreams

- Communications
- **Proofs of Concept in Humanities and Social Science and High Energy Physics (Discovery)**
- **Interoperability**
- Strategy
- Internationalisation

Proofs of concept in:

Humanities and Social Science (HSS) and High Energy Physics (HEP)

Image credit: smiling_da_vinci on Flickr

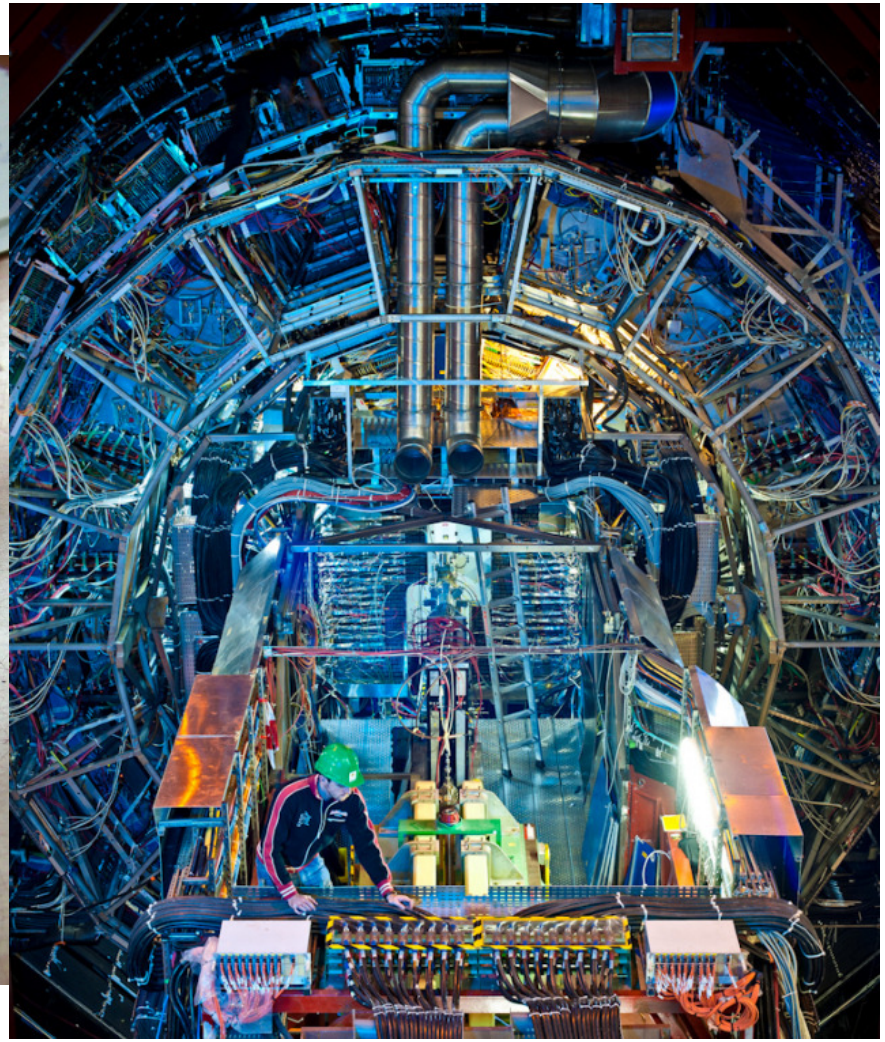
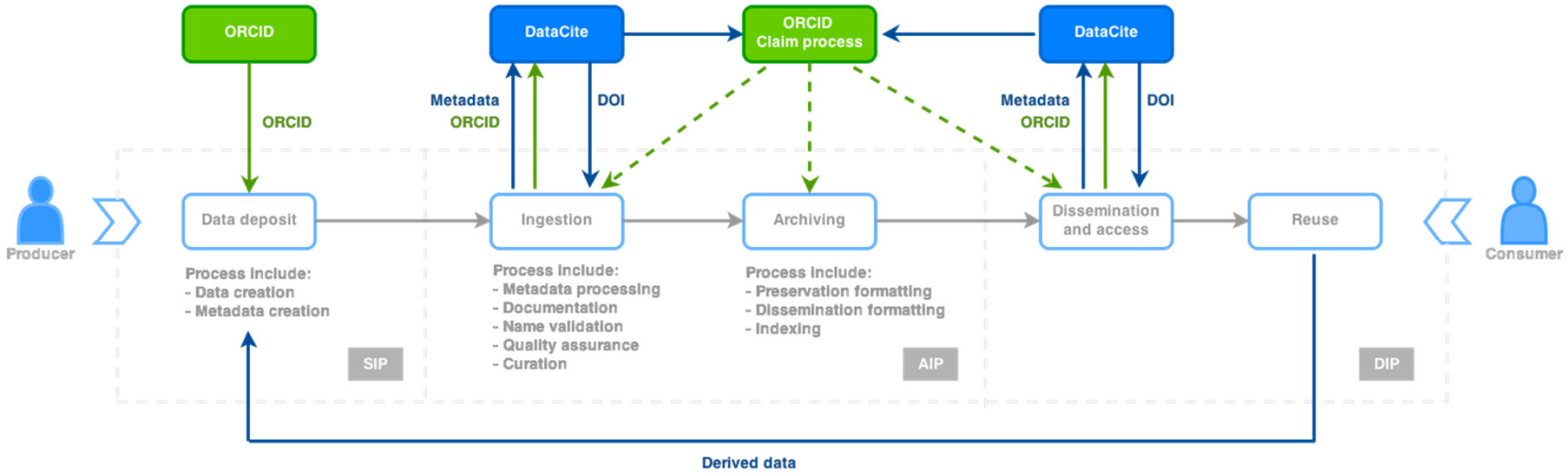


Image credit CERN

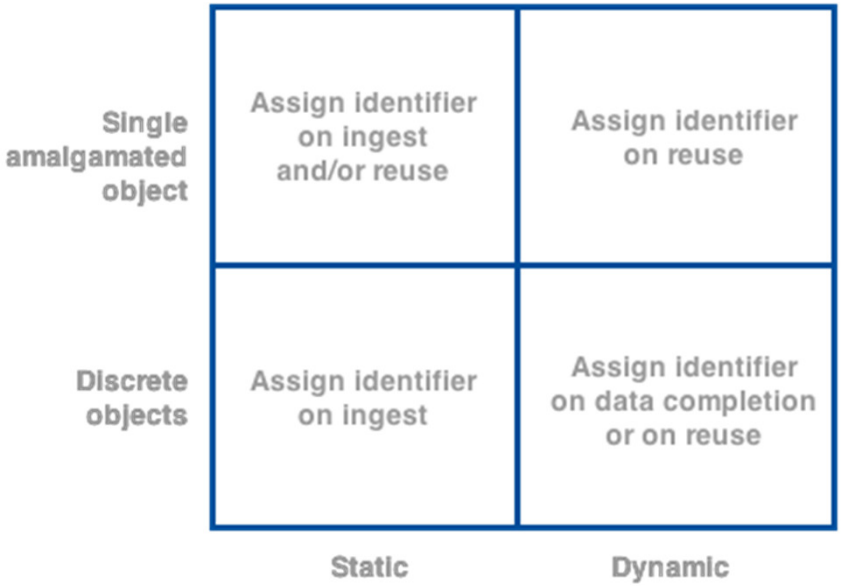
Generic workflow for PIDs

- Proof of concept workflows amalgamated into a single generic workflow
- Generic workflow demonstrated to MRC National Survey of Health and Development (MRC NSHD)
- Feedback from NSHD used to refine generic workflow and outline path for their integration of identifiers
- Feedback on refined workflow solicited from GESIS and UK Archaeology Data Service (DataCite members and clients)



Proof of Concept findings

- Points of integration for workflows across domains are similar, despite major differences in data and documentation
- Name identifiers can be applied at any point, and where there isn't one in wide community use, choice should be kept open
- Options for DOI assignment point can be based on a matrix of two aspects: data storage and fixity



Proof of Concept conclusions

- Remaining challenges for HSS:
  - Trust in content from ORCID
  - Uptake and awareness of name identifier across depositors
  - Availability of year and creator metadata
- Remaining challenges for HEP:
  - Providing adapted services and consolidate data publication
  - Engaging the researchers, obtaining both feedback to design description patterns and better metadata
  - Mapping between different identifiers would ease service compatibility and interoperability
  - Data exchanges and validation with other platforms in the community, i.e. the publishers
- Common challenges:
  - Trust of information from third parties
  - Engagement and input from researchers

Message for interoperability?

- The processes and issues they in integrating PIDs were very similar
- While research and data formats are very different, the ultimate aim of both (data sharing) means management and identification of data and authors are the same

Image credit: Annie Mole on Flickr

Improving interop. between systems

- Building a model of interoperability that is open, discipline-neutral, and inclusive
- Focusing on data citation and attribution as the most tangible, immediate goal
- A relatively lightweight approach to interoperability will deliver the best returns whilst presenting the least risk
- Technical developments to show the way forward in interoperability between PID systems
- Based on open standards and metadata

Gap analysis

Revised gap	Progress
There is only limited access to PID e-Infrastructures for small organisations due to a lack of licenses or memberships on a national level.	ORCID has expanded its public API, adding authentication, to support responsible use of ORCID by projects and small organizations. Some institutions still lack the technical infrastructure or staffing resources to integrate PIDs; overcoming this may require the creation of tools or widgets that make it possible to integrate PIDs with little technical expertise.
Some research communities have little to no experience with interoperable PIDs and no attempts are made to target them explicitly.	There have been more PID initiatives, including direct outreach to universities and to researchers at professional association meetings. Additional work is required to communicate benefits and provide specific training to researchers.

Gap analysis

Revised gap	Progress
Tailored non-operable PID systems are emerging as there is no easy way to integrate interoperable PIDs in local implementations.	First plug-ins to easily integrate ORCID iDs and DataCite DOIs into Eprints and DSpace and other repositories are developed to close this gap.
Documentation for PID integrations is not adequately comprehensive and easily accessible and thus, the development of tools to track research data re-use is constrained.	There is documentation available on e.g. the ORCID and DataCite homepages. Tracking tools are still limited. However, there are collaborations with metrics and altmetrics providers to initiate the development of more citation tools.

Here's what's now possible from ODIN developments…

Supporting users of data

- This research paper is an important new advancement in my research area
- I look at the citations, and see that they cite the datasets they have used. There are 2 of them
- DataCite DOIs have allowed me to access the data for the citation

slavemaking ant with a fragmented distribution. *Conservation Genetics*, **11**, 1701–1710.
Schneirla TC (1971) *Army Ants: A Study in Social Organization* (ed Topoff HR). WH Freeman and Company, San Francisco.
Schöning C, Njagi WM, Franks NR (2005) Temporal and spatial patterns in the emigrations of the army ant *Dorylus* (*Anomma*) *molestus* in the montane forest of Mt Kenya. *Ecological Entomology*, **30**, 532–540.
Seeb LW, Antonovich A, Banks AA *et al.* (2007) Development of a standardized DNA database for Chinook salmon. *Fisheries*, **32**, 540–552.
Segelbacher G, Cushman SA, Epperson BK *et al.* (2010) Applications of landscape genetics in conservation biology: concepts and challenges. *Conservation Genetics*, **11**, 375–385.
Short Bull RA, Cushman SA, Mace R *et al.* (2011) Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Molecular Ecology*, **20**, 1092–1107.
Slatkin M, Maddison WP (1990) Detecting isolation by distance using phylogenies of genes. *Genetics*, **126**, 249–260.
Soare TW, Tully SI, Willson SK, Kronauer DJC, O'Donnell S (2011) Choice of nest site protects army ant colonies from environmental extremes in tropical montane forest. *Insectes Sociaux*, **58**, 299–308.
Soare TW, Kumar A, Naish KA, O'Donnell S (2013) Data from: genetic evidence for landscape effects on dispersal in the army ant *Eciton burchellii*. *Dryad Digital Repository*, doi:10.5061/dryad.j33k3.
Spear SF, Balkenhol N, Fortin M-J, McRae BH, Scribner K (2010) Use of resistance surfaces for landscape genetic studies: considerations for parameterization and analysis. *Molecular Ecology*, **19**, 3576–3591.
Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity*, **98**, 128–142.
Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology*, **19**, 3496–3514.
Sundström L, Keller L, Chapuisat M (2003) Inbreeding and sex-biased gene flow in the ant *Formica exsecta*. *Evolution*, **57**, 1552–1561.
Sundström L, Seppä P, Pamilo P (2005) Genetic population structure and dispersal patterns in *Formica* ants – a review. *Annales Zoologici Fennici*, **42**, 163–177.
Swartz MB (1997) Behavioral and population ecology of the army ant *Eciton burchelli* and ant-following birds. PhD thesis, University of Texas, Austin.
Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends in Ecology & Evolution*, **16**, 330–343.
Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921–935.
Vidal-Riggs JM, Chaves-Campos J (2008) Method review: estimation of colony densities of the army ant *Eciton burchellii* in Costa Rica. *Biotropica*, **40**, 259–262.
Vogt JT, Appel AG, West MS (2000) Flight energetics and dispersal capability of the fire ant, *Solenopsis invicta* Buren. *Journal of Insect Physiology*, **46**, 697–707.
Wang IJ (2010) Recognizing the temporal distinctions between landscape genetics and phylogeography. *Molecular Ecology*, **19**, 2605–2608.
Wang IJ (2011) Choosing appropriate genetic markers and analytical methods for testing landscape genetic hypotheses. *Molecular Ecology*, **20**, 2480–2482.
Watkins JF (1976) *The Identification and Distribution of New World Army Ants* (*Dorylinae*: *Formicidae*). Baylor University Press, Waco.
Willson SK, Sharp R, Ramler IP, Sen A (2011) Spatial movement optimization in Amazonian *Eciton burchellii* army ants. *Insectes Sociaux*, **58**, 325–334.

T.W.S., A.K., K.A.N., and S.O'D. designed the study. T.W.S., A.K., and S.O'D. performed field and laboratory work. T.W.S., A.K., and K.A.N. analysed the data. T.W.S. wrote the first version of this manuscript. All authors contributed to and approved the final manuscript.

Data accessibility

Microsatellite genotypes uploaded as DRYAD entry doi:10.5061/dryad.j33k3.

Supporting information

Additional supporting information may be found in the online version of this article.

Data S1 Methodological details.

Data S2 Analytical details.

Table S1 Date, collection number (coll_no), and number of workers (no_workers) of army ant colonies used in genetic analyses.

Table S2 Annealing temperatures (T_a) for seven primers, number of alleles (N_A), range of allele sizes, and observed (H_O) and expected (H_E) heterozygosities at each locus.

Table S3 Alternative models of landscape resistance predicting pairwise queen relatedness.

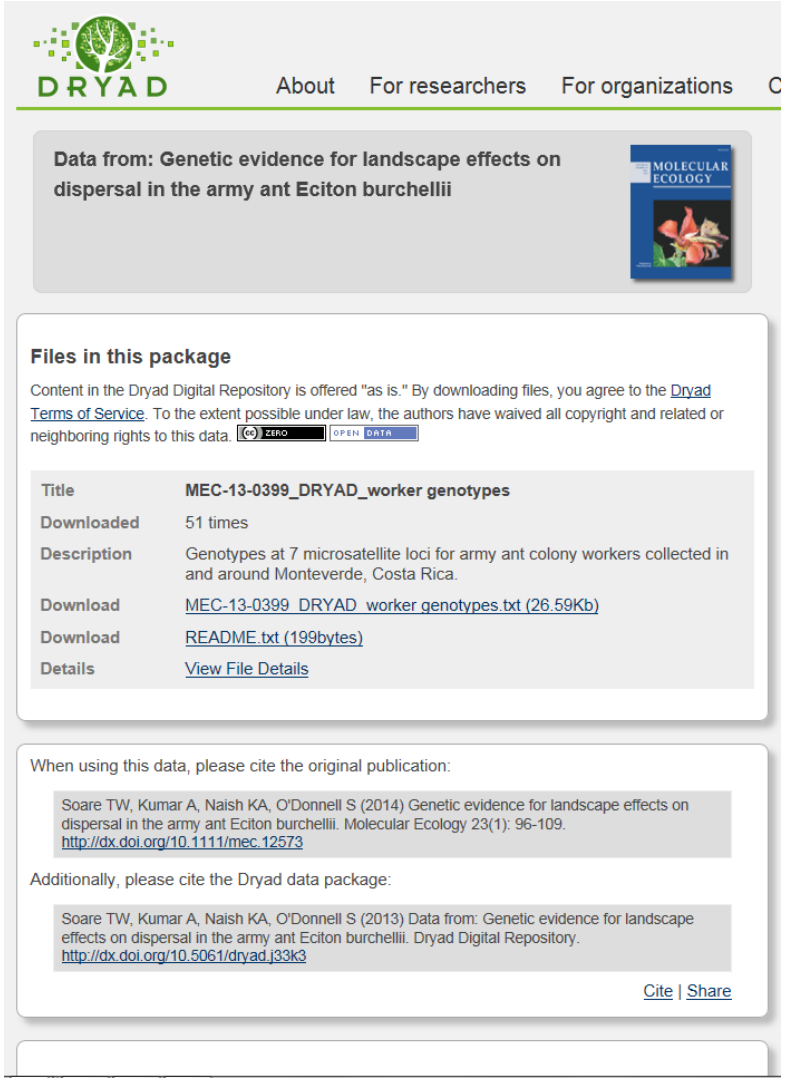
Table S4 Pairwise distance measures among sample locations were correlated.

Fig. S1 Plot of estimates of genetic similarity of simulated dyads of known relationships.

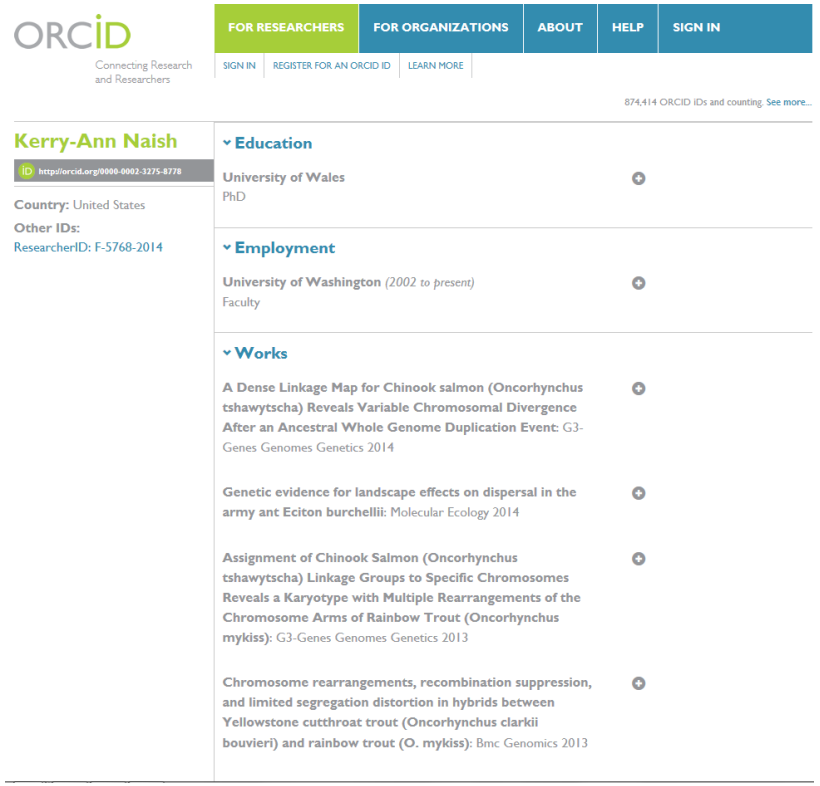
Fig. S2 Ranking of variable importance in random forests (RF) analyses for predicting (a) the pairwise relationship coefficient (r_q) and (b) the pairwise kinship coefficient (F_{ij}) with random variable subsets (m_{try}) of size 4.

Fig. S3 Ranking of variable importance in random forests (RF) analyses for predicting (a) the pairwise relationship coefficient (r_q) and (b) the pairwise kinship coefficient (F_{ij}) with random variable subsets (m_{try}) of size 1.

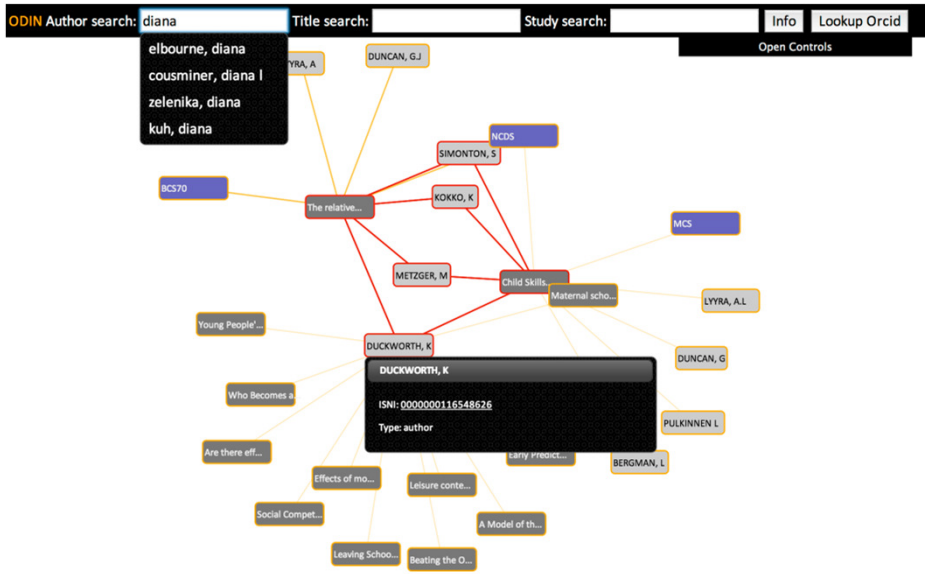
- The data citation has taken me to a page that gives more details about the dataset
- Details about this data link into more papers and further datasets
- How can I find more data?



- The creators names are linked
- I click them and it takes me to their ORCID record
- I can find older datasets by these creators

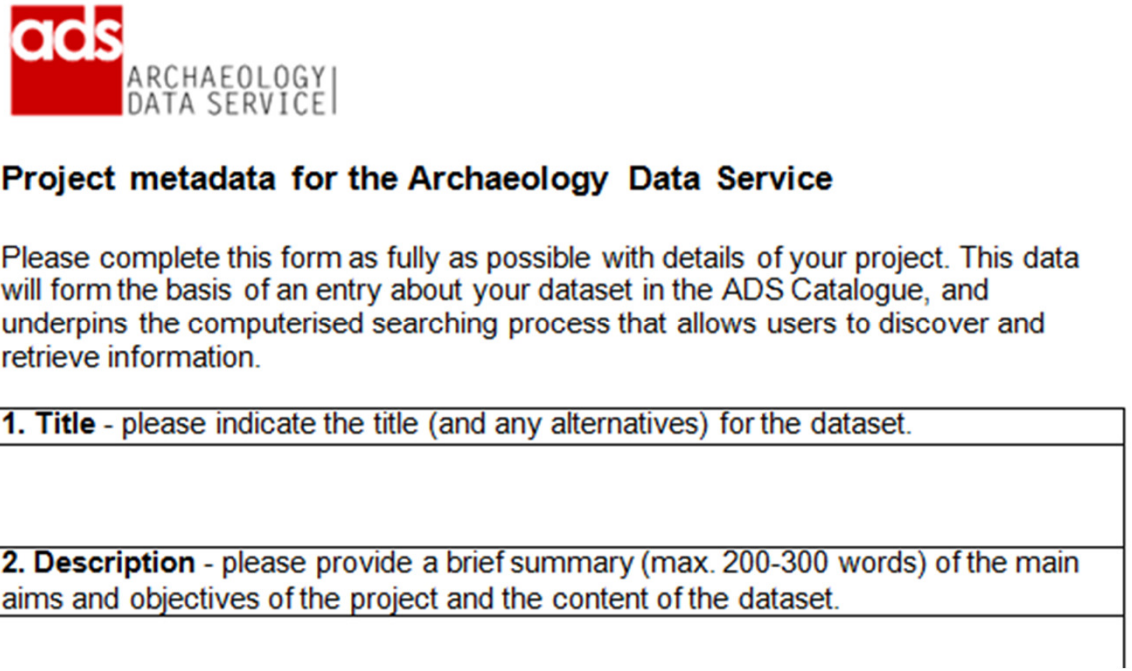


- The DOI takes me to a new data centre
- I see a graph of the authors on this dataset and how they link to various publications and datasets
- On clicking around, I see that the creators of this dataset could be potential collaborators



Supporting creators of data

- I have written my paper and submitted it to a journal
- I have produced data that can be used to reproduce my analysis
- I would like to share my data



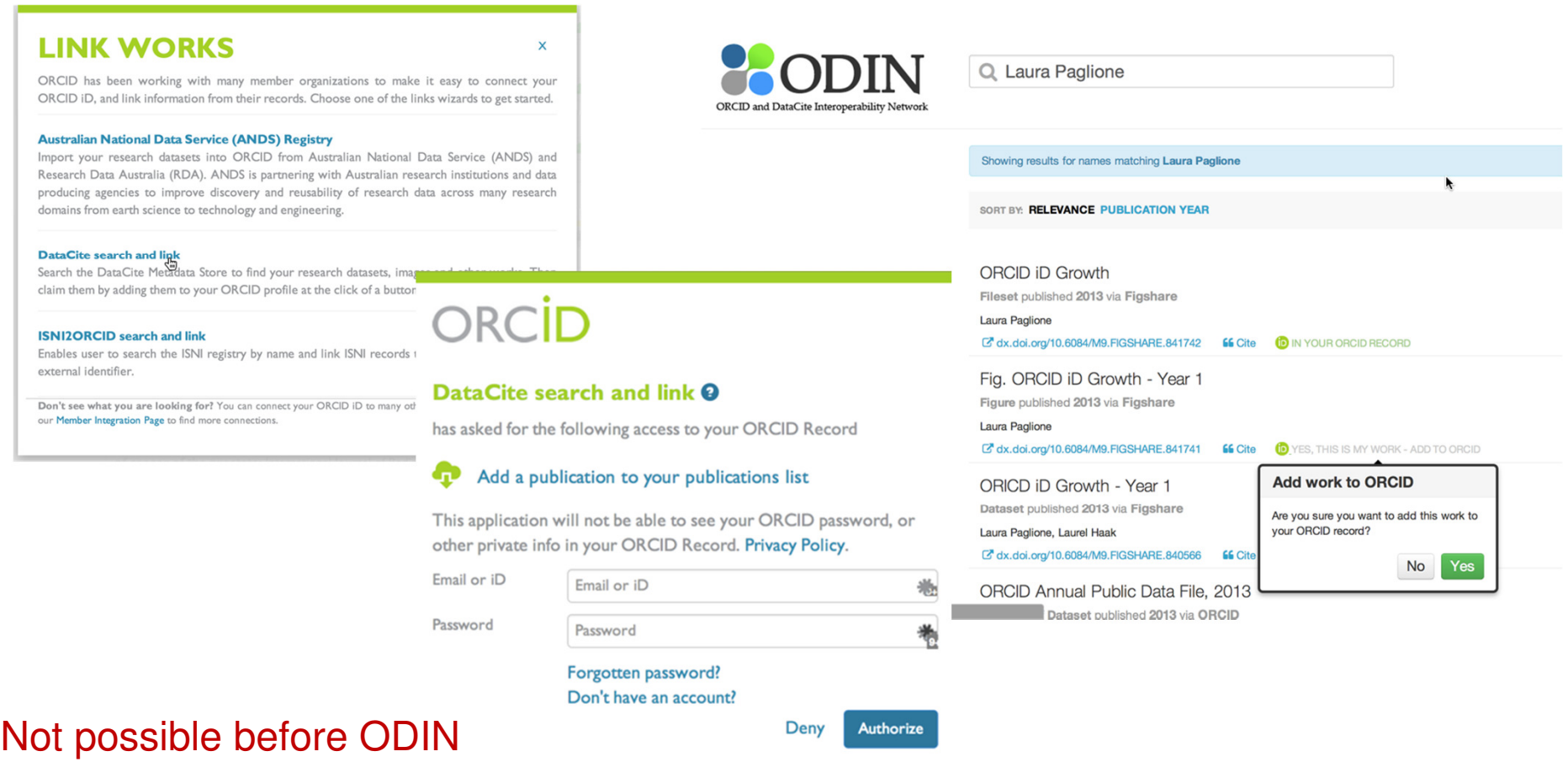
ads ARCHAEOLOGY DATA SERVICE

Project metadata for the Archaeology Data Service

Please complete this form as fully as possible with details of your project. This data will form the basis of an entry about your dataset in the ADS Catalogue, and underpins the computerised searching process that allows users to discover and retrieve information.

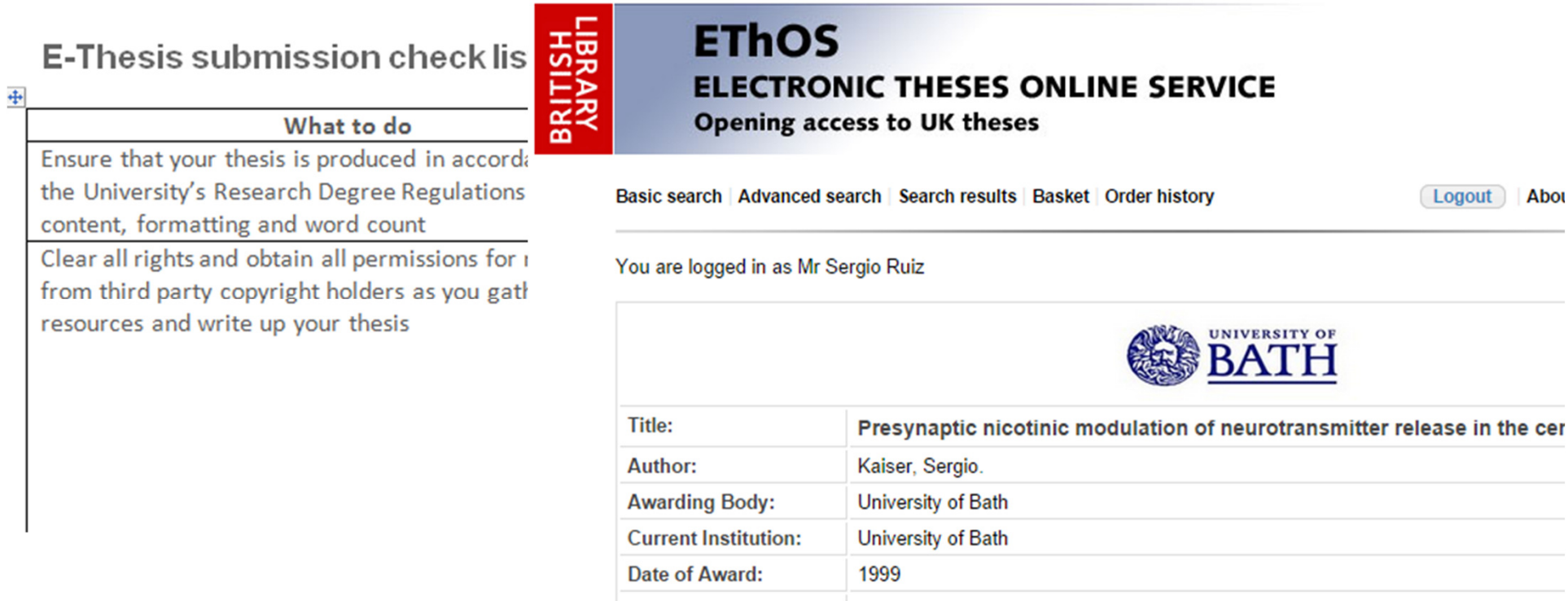
1. Title - please indicate the title (and any alternatives) for the dataset.
2. Description - please provide a brief summary (max. 200-300 words) of the main aims and objectives of the project and the content of the dataset.

- I can link my paper AND my data to my ORCID profile
- There may be thousands of authors named like me



Not possible before ODIN

- I sign up for a PhD programme
- I write my doctoral thesis and deposit it, along with my data
- I see the thesis in EThOS



E-Thesis submission check lis

What to do
Ensure that your thesis is produced in accorda the University's Research Degree Regulations content, formatting and word count
Clear all rights and obtain all permissions for r from third party copyright holders as you gatl resources and write up your thesis

BRITISH LIBRARY

EThOS

ELECTRONIC THESES ONLINE SERVICE

Opening access to UK theses

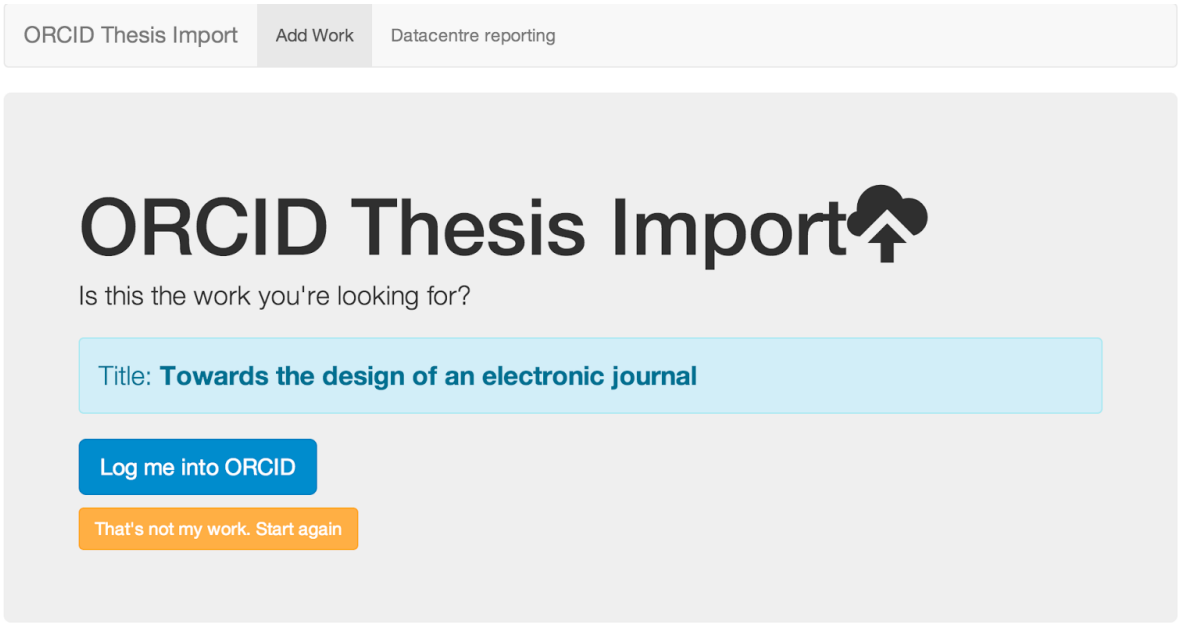
Basic search | Advanced search | Search results | Basket | Order history | Logout | Abou

You are logged in as Mr Sergio Ruiz

UNIVERSITY OF BATH

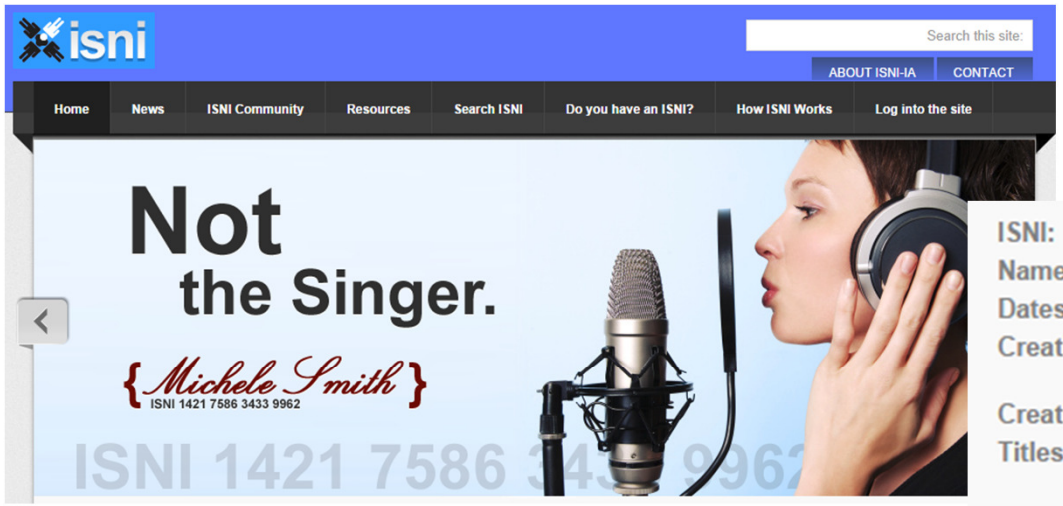
Title:	Presynaptic nicotinic modulation of neurotransmitter release in the cer
Author:	Kaiser, Sergio.
Awarding Body:	University of Bath
Current Institution:	University of Bath
Date of Award:	1999

- I can link my thesis AND the data behind to my ORCID profile



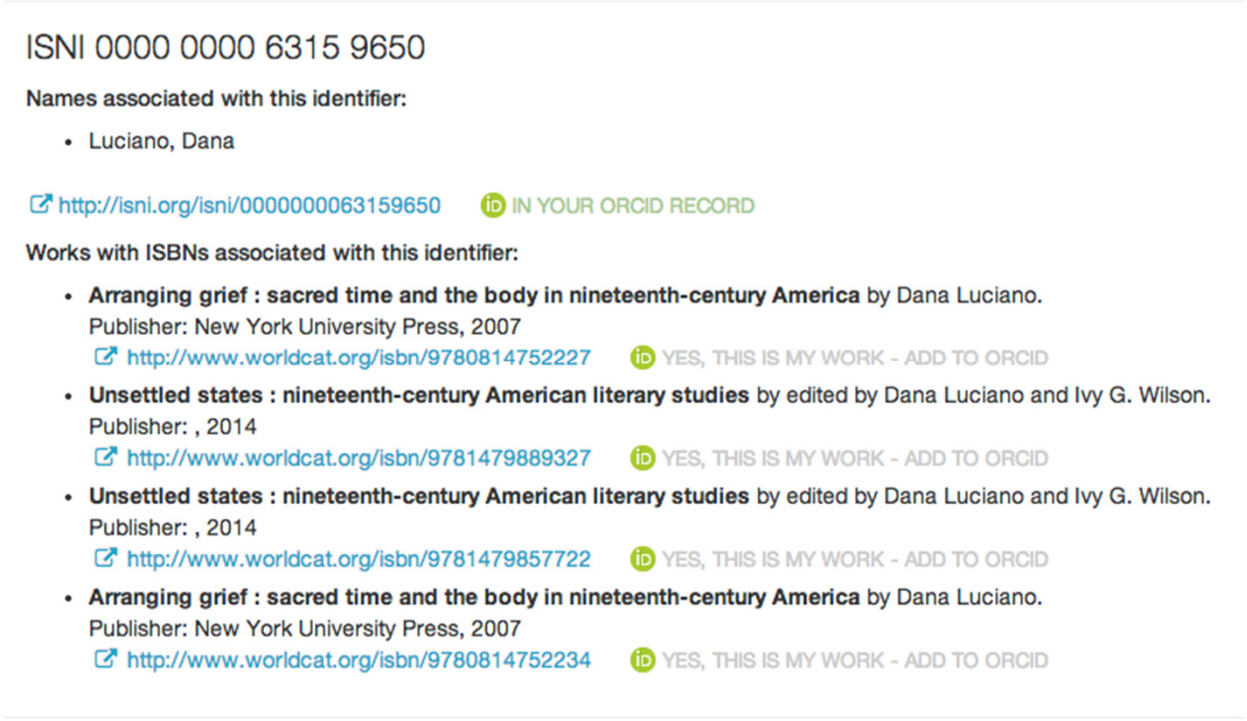
Not possible before ODIN

- A publisher reads my thesis and asks me to do further work
- I publish a book with the results which leads to my having a record in ISNI



ISNI:	0000 0003 8522 5964
Name:	**Mele, Salvatore**
Dates:	1965-
Creation class:	Language material Text
Creation role:	author
Titles:	Apuleio e L'asino d'oro : saggio critico Dorgali ellenismo nei dialetti della Calabria media, L' Gallura felix, c2009: Gallura felix : il sud del Giudicato di Gallura e il territorio del Ca : Dorgali
Notes:	
Sources:	VIAF BAV LC BOWKER

- I can import the ISNI record on my book to my ORCID profile



ISNI 0000 0000 6315 9650

Names associated with this identifier:

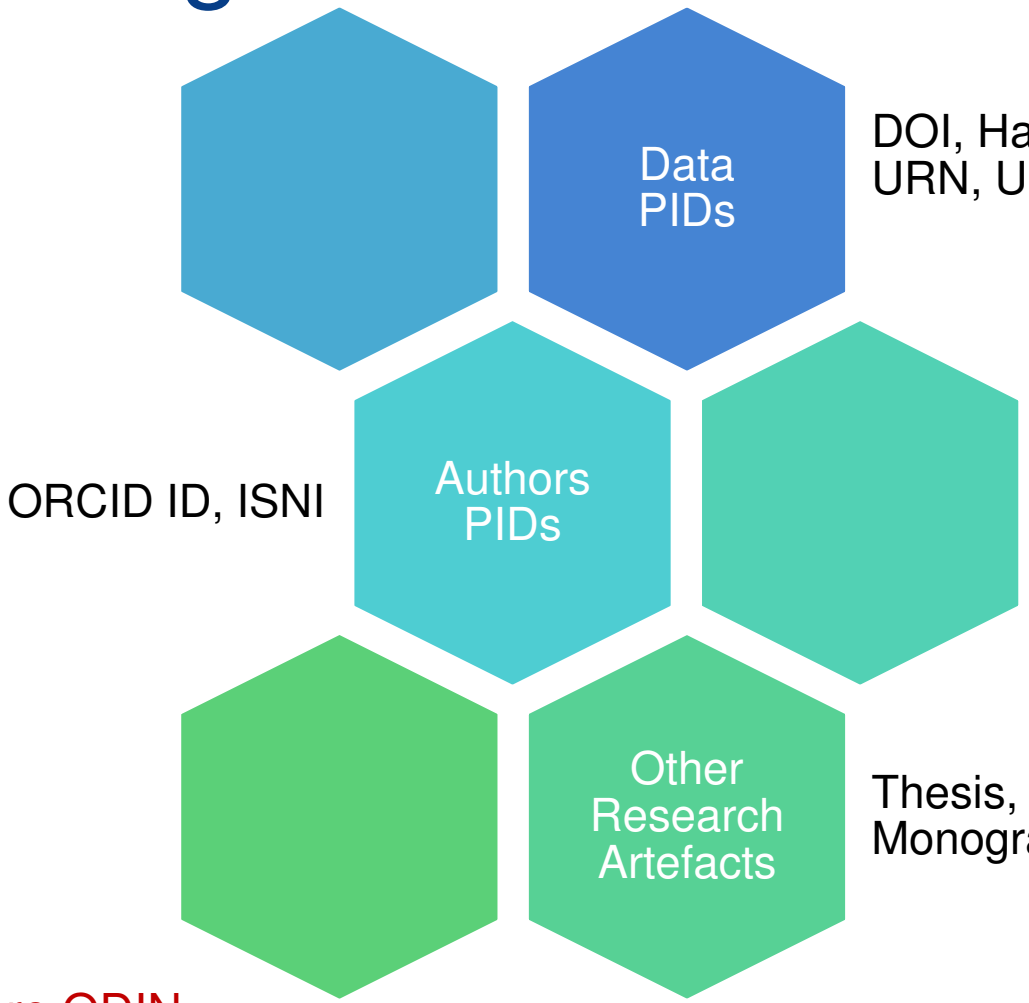
- Luciano, Dana

http://isni.org/isni/0000000063159650 IN YOUR ORCID RECORD

Works with ISBNs associated with this identifier:

- **Arranging grief : sacred time and the body in nineteenth-century America** by Dana Luciano.
  Publisher: New York University Press, 2007
  http://www.worldcat.org/isbn/9780814752227 YES, THIS IS MY WORK - ADD TO ORCID
- **Unsettled states : nineteenth-century American literary studies** by edited by Dana Luciano and Ivy G. Wilson.
  Publisher: , 2014
  http://www.worldcat.org/isbn/9781479889327 YES, THIS IS MY WORK - ADD TO ORCID
- **Unsettled states : nineteenth-century American literary studies** by edited by Dana Luciano and Ivy G. Wilson.
  Publisher: , 2014
  http://www.worldcat.org/isbn/9781479857722 YES, THIS IS MY WORK - ADD TO ORCID
- **Arranging grief : sacred time and the body in nineteenth-century America** by Dana Luciano.
  Publisher: New York University Press, 2007
  http://www.worldcat.org/isbn/9780814752234 YES, THIS IS MY WORK - ADD TO ORCID

Connecting PIDs



Not possible before ODIN

ODIN Partners

ORCID
Connecting Research and Researchers

DataCite
Helping you to find, access, and reuse data

BRITISH LIBRARY HSIEH
THE BRITISH LIBRARY
Explore the world's knowledge

CERN
European Organization for Nuclear Research

ands
AUSTRALIAN NATIONAL DATA SERVICE

DRYAD

arXiv.org