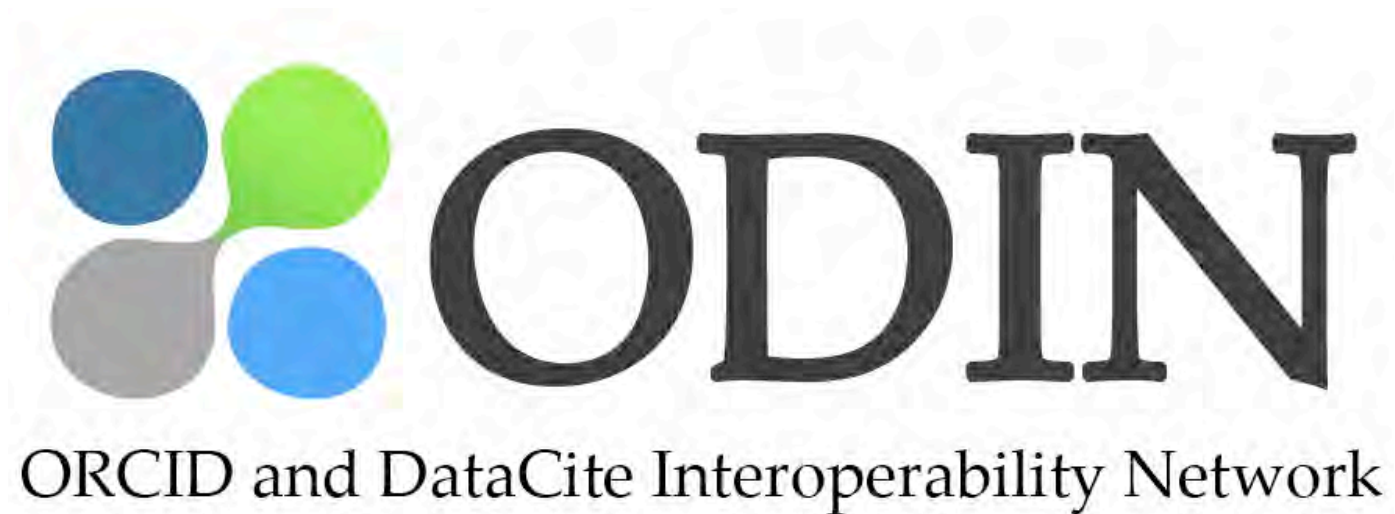


Interoperable persistent identifiers



Agenda

- ▣ ORCID: Laura Paglione
- ▣ Datacite: Sebastian Peters
- ▣ INSPIREHEP – Implementation: Sünje Dallmeier-Tiessen
- ▣ Dryad – Implementation: Ryan Scherle
- ▣ Overarching challenges
- ▣ Q&A



ODIN

ORCID and DataCite Interoperability Network





Connecting Research
and Researchers

The ORCID Registry

Open Repositories 2014

June 12, 2014

Laura Paglione

Technical Director, ORCID

L.Paglione@ORCID.org

<http://orcid.org/0000-0003-3188-6273>

What is ORCID ?

The ORCID Organization

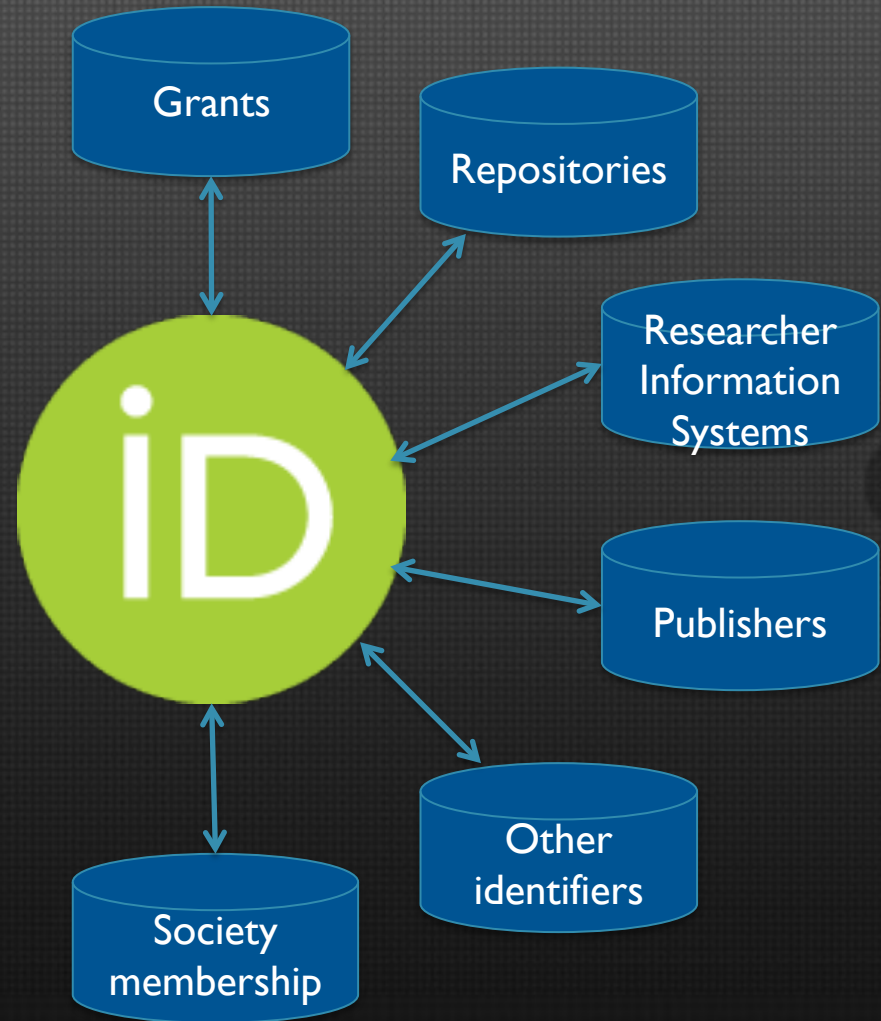
- Non-profit, non-proprietary, open, and community-driven
- Global, interdisciplinary
- Independent
- Supported by the membership of organizations using the ORCID API
 - Funding organizations
 - Professional societies
 - Universities & research institutes
 - Publishers

The ORCID

- Unique, persistent identifier for researchers, scholars & analysts
- Free to researchers
- Used throughout one's career
- Used across professional activities, affiliations, disciplines, nations & languages
- Embedded into workflows & metadata
- API enables interoperability between siloed systems

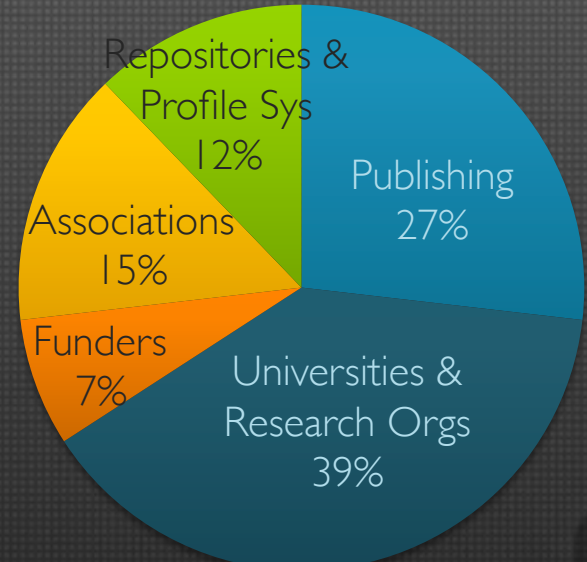
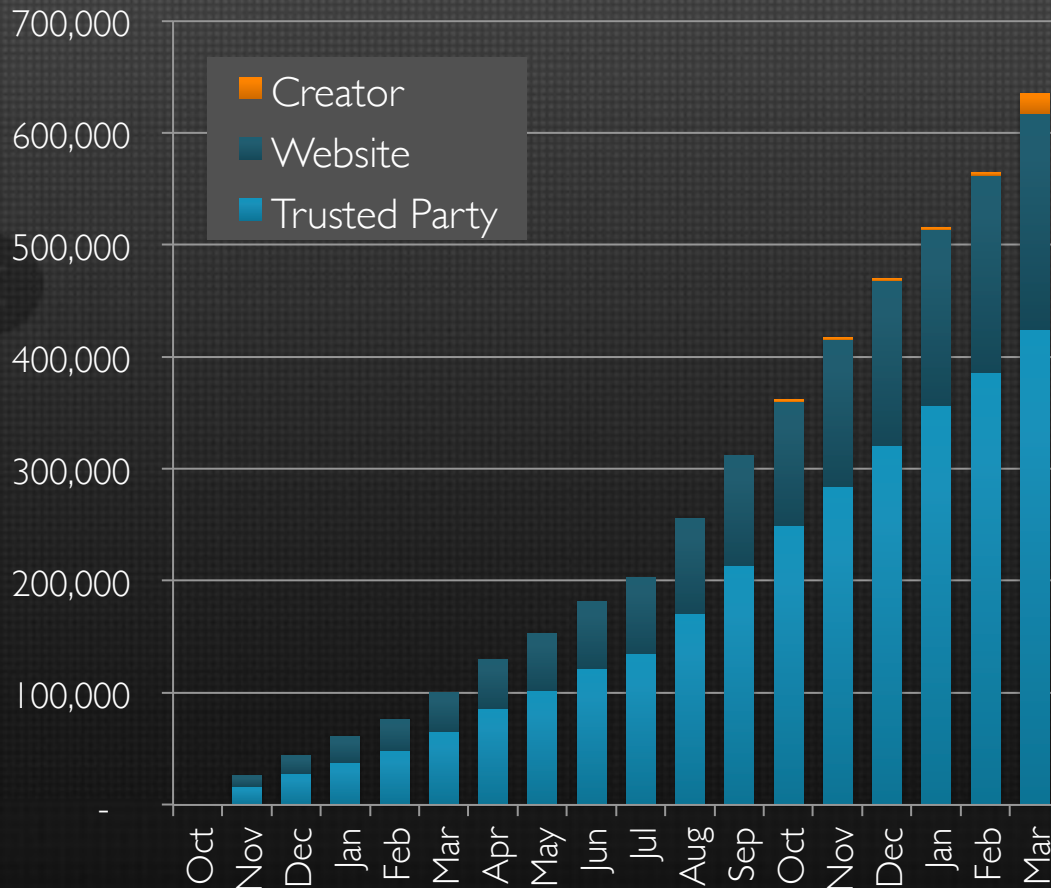
The ORCID API enables information exchange between systems:

- Less time re-keying
- Improved data
- Easier maintenance
- Better sharing across systems

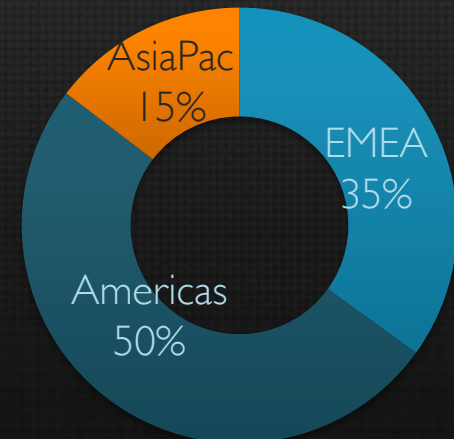


Adoption and Integration

ORCID has issued over 740,000 iDs since our launch in October 2012. Integration and use is international.



Over 145 members, from every sector of the international research community



Laura AD Paglione 16 Works 3 Education 3 Employment 0 Grants

<http://orcid.org/0000-0003-3188-6273>

Personal Information [Update](#)

Biography

Laura Paglione is an entrepreneurial, technically versatile, resourceful leader who thrives at the intersection of creative, technical and business environments. She currently serves as Technical Director of ORCID, where she is directing the technical efforts in ORCID's mission to address name ambiguity for researchers, and serve as a gateway to connect their research activities from disparate sources.

Previously as Director, Advancing Innovation at the Kauffman Foundation, Laura directed the efforts of the iBridge Network, an innovation catalyst for university collaboration and technology commercialization. In prior positions at Ford Motor Company and Avid Technology, as well as several start-up/gazelle companies, Laura has turned around, launched and led 4 other high-profile initiatives, the most visible of which was for Ford Motor Company's Board of Directors.

Education [Add Education Manually](#)

- Massachusetts Institute of Technology Sloan School of Management (1999) MBA
- Brown University ScM
- Stevens Institute of Technology BEng

Employment [Add Employment Manually](#)

- ORCID (06-2012 to present) Technical Director
- Management Leadership for Tomorrow (04-2010 to 06-2012) VP, Operations
- Wildcat (2009 to 2010) Chief Operating Officer

Works [Import Works](#) [Add Work Manually](#)

- Work Title: Work Subtitle | 2-2013
- ORCID: A system to uniquely identify researchers: Learned Publishing 2012

Funding [Add Funding Manually](#)

- One more thing - present: Wellcome Trust Sanger Institute (2013 to present)
- Another fake grant: National Institutes of Health Office of Technology Transfer (2011-03 to 2024-04)

Contains:

← Biographical Data

← Education Data

← Employment Data

← Works Data

← Funding Data

ADD EDUCATION

Who can see this?

Institution: *

- Massachusetts Institute of Technology
Cambridge, MA, academic
- Massachusetts Group Insurance Commission
Boston, MA, public
- Massachusetts Interlocal Insurance Association
Boston, MA, other/assoc
- Associated Industries of Massachusetts
Boston, MA, other/advocates
- Massachusetts Eye and Ear Infirmary
Boston, MA, academic/hospital

Department:

Degree/title:

ADD FUNDING

Who can see this?

Funding type: *

Title: *

Funding amount: *

Start date: Day Month Year

FUNDING AGENCY

Funding agency name: *

Funding agency city: *

Funding agency region:

Funding agency country:

Funding agency
list consistent
with FundRef

↑ Organization
list from Ringgold
(an ISNI Registrar)

Search & link Wizards ↓

External IDs
(DOIs, ISBNs, ISNIs, PubMed IDs, etc)
Other Person IDs
(Scopus ID, ResearcherID, ISNIs,
Institution IDs, etc)

LINK FUNDING

ÜberWizard for ORCID

IMPORT WORKS

ORCID has been working with many member organizations to make it easy to connect your ORCID iD, and import information from their records. Use any of the imports wizards to get started.

Australian National Data Service (ANDS) Registry
Import your research datasets into ORCID from Australian Research Data Australia (RDA). ANDS is producing agencies to improve discovery and reusability of research data across many research domains from earth science to technology and engineering.

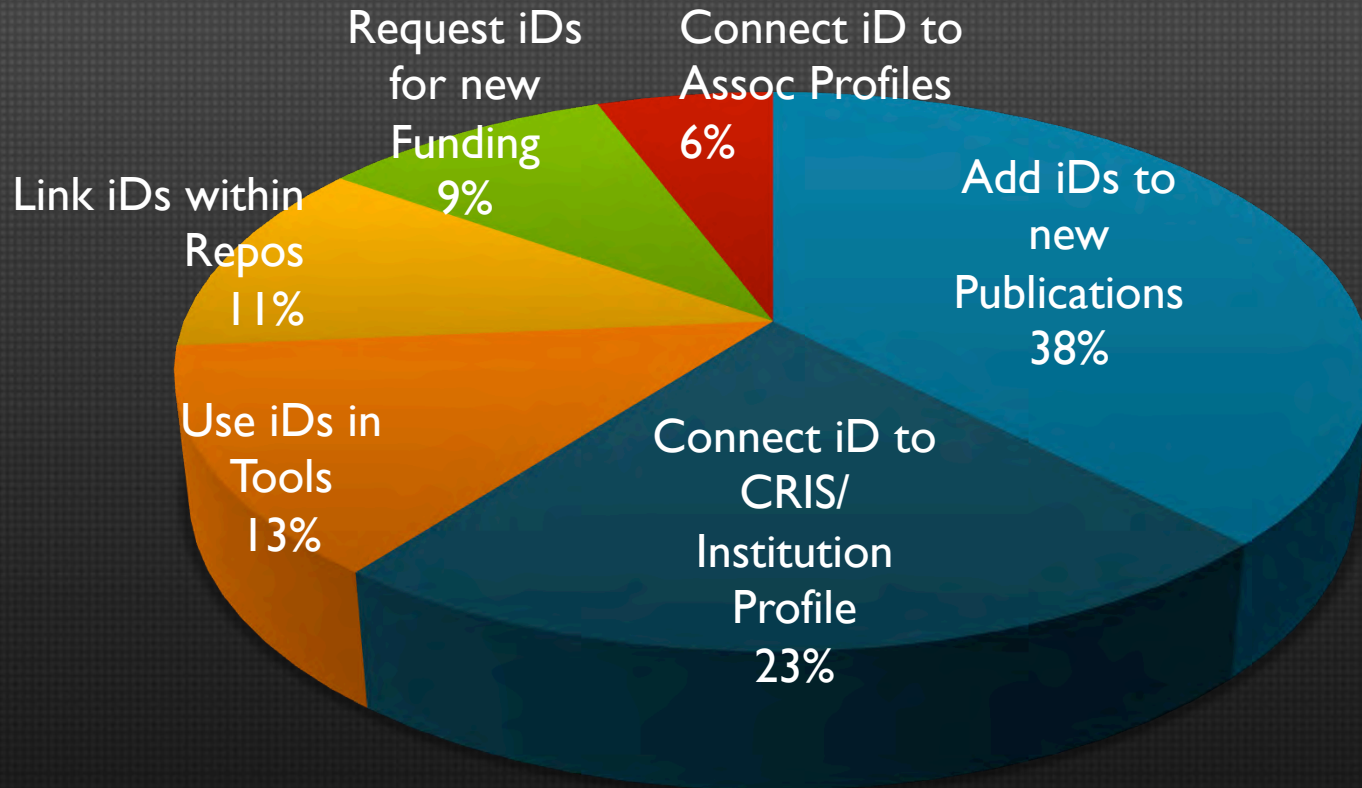
- ANDS Registry
- CrossRef Metadata Search
- DataCite Metadata Store
- Europe PubMed Central
- ResearcherID
- Scopus
- ISNI number & ISBNs
- ÜberWizard (funding)

Who is Integrating?

- **Research Funders**
- **Professional Associations**
- **Publishers**
- **Repositories, CRIS Vendors, Metrics Providers**
- **Universities and Research Orgs**

For a list of organizations and integrations see
<http://orcid.org/organizations/integrators>

Over 55 integrations by members and others using the public API



<http://orcid.org/organizations/integrators/current>

❖ ORCID works collaboratively with the research community to ensure use and adoption of research information exchange standards

ORCID to begin using Ringgold as Registration Agency

CASRAI and ORCID Collaborate to Advance Integrated Research Data

A CERIF-XML Person Record + Vocab

crossref & orcid

ODIN
ORCID and DataCite Interoperability Network

- Remembering all of those identifiers!
- Identifiers without other info are hard; they don't provide enough info (resolvable iDs help)
- Type the ORCID iD
 - Prone to errors
 - Requires researcher to remember the iD
- Search for someone to associate an iD
 - Disambiguation problems – which Smith / Wang / Johansson are you?
- iD display – long

Thank you!

- Find out more at <http://orcid.org>
- More on membership at <http://orcid.org/about/membership>
- Learn about tools to embed ORCID iDs at <http://support.orcid.org/knowledgebase/>
- Subscribe to our blog at <http://orcid.org/about/news> and follow [@ORCID_Org](https://twitter.com/ORCID_Org) on Twitter
- Contact the ORCID Executive Director at l.haak@orcid.org

DataCite

or

DOI is more than just CrossRef

Sebastian Peters

DataCite Tech Team Lead

German National Library of
Science and Technology

tech@datacite.org / sebastian.peters@tib.uni-hannover.de

TIB | GERMAN NATIONAL LIBRARY OF
SCIENCE AND TECHNOLOGY



DOI Registration Agencies

- 1) Airiti, Inc.
- 2) China National Knowledge Infrastructure (CNKI)
- 3) **CrossRef**
- 4) **DataCite**
- 5) Entertainment Identifier Registry (EIDR)
- 6) The Institute of Scientific and Technical Information of China (ISTIC)
- 7) Japan Link Center (JaLC)
- 8) Multilingual European DOI Registration Agency (mEDRA)
- 9) Publications Office of the European Union (OP)

What is DataCite?

- Global consortium carried by local institutions
- focused on improving the scholarly infrastructure around datasets and other non-textual information
- focused on working with datacentres and organizations that hold data
- Providing standards, workflows and best-practice
- **Make research data citable** (data publication)
- Initially, but not exclusively based on the DOI system

<http://www.datacite.org>

DataCite History

- Founded December 1st 2009 in London
- Forerunner (since 2004), founder member and managing agent: German National Library of Science and Technology (TIB)
- 1st DOI for a dataset registered by TIB in 2004
- Happy 10th Anniversary:
<http://www.datacite.org/node/121>

Facts & Figures

- 22 full members, 9 associated; from 18 countries
- ~ 3.4 Mio. DOIs
- ~ 3.2 Mio. metadata objects
- > 300 DOI prefixes
- > 300 datacentres (publishing agents)

DOI – Digital Object Identifier

- *Digital Identifier of an Object*
- Maintained by the International DOI Foundation (IDF)
- Based on Handle System
- Actionable, interoperable, persistent
- ISO standard 26324:2012
- ~ 100 Mio. DOIs registered
- Example: doi:10.5524/100001, <http://doi.org/10.5524/100001>
- Resolve to Landing Pages

<http://www.doi.org>

Challenge: Getting DOI Metadata

Problem

Every RA has a custom metadata format and API

Solution

Content Negotiation on doi.org level

Content Negotiation – CrossRef & DataCite

- HTTP Content Negotiation with [dx.]doi.org proxy
- You don't need to know the actual DOI RA
- Common Formats of CrossRef & DataCite:
 - RDF, RIS, BibTeX, Citeproc, ...
- <http://crosscite.org/cn/>
- Example:
 - `curl -LH "Accept:application/rdf+xml" http://doi.org/10.5524/100001`

Citation Formatter

- Using Content Negotiation
- More than 500 styles (<http://citationstyles.org>)
- <http://crosscite.org/citeproc/>
- <https://github.com/crosscite/citeproc-doi-server>



DOI:

Grobe, Hannes (2005) Data policy of the information system PANGAEA. doi:
10.1594/PANGAEA.327791



DataCite Metadata Generator and ORCID

- Metadata formular with ORCID lookup
- Result from the last code sprint at CERN
- <http://www.datacite.org/node/102>
- <https://github.com/koelnconcert/datacite-metadata-generator>

DataCite Metadata Generator - Kernel 3.0

Mandatory Elements	Metadata
DOI: [DOI]	<pre><?xml version='1.0' encoding='UTF-8'?> <resource xmlns='http://datacite.org/schema/k xsi:schemaLocation='http://datacite.org/schem <identifier identifierType='DOI'></identifier> <creators> <creator> <creatorName>jan brase</creat <nameIdentifier nameIdentifie </creator> </creators> <titles> </titles> <publisher></publisher> <publicationYear></publicationYear> </resource></pre>
Title(s): [TITLE]	
[titleType]	
Creator(s): jan brase	
Search ORCID ORCID 0000-0002-8250-6253	
Lookup ORCID	
ORCID Identifier 0000-0002-8250-6253	
Biography	

Combining Identifiers for Citing Fragments

- Media Fragment Identifier (MFID), W3C
- Prototype for TIB's AV-Portal
- <http://doi.org/10.5446/12780#t=00:20,00:27>
- Out of the box solution
- DOI resolver agnostic
- Content negotiation still works

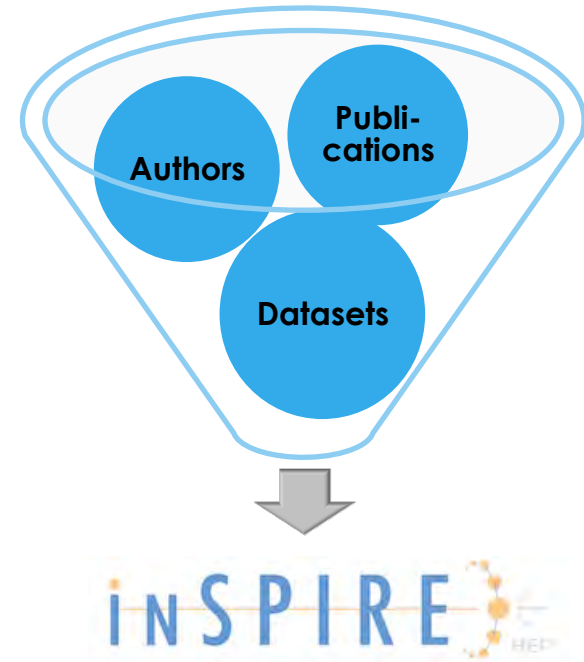


Shared identifiers in a large-scale digital library: INSPIRE-HEP

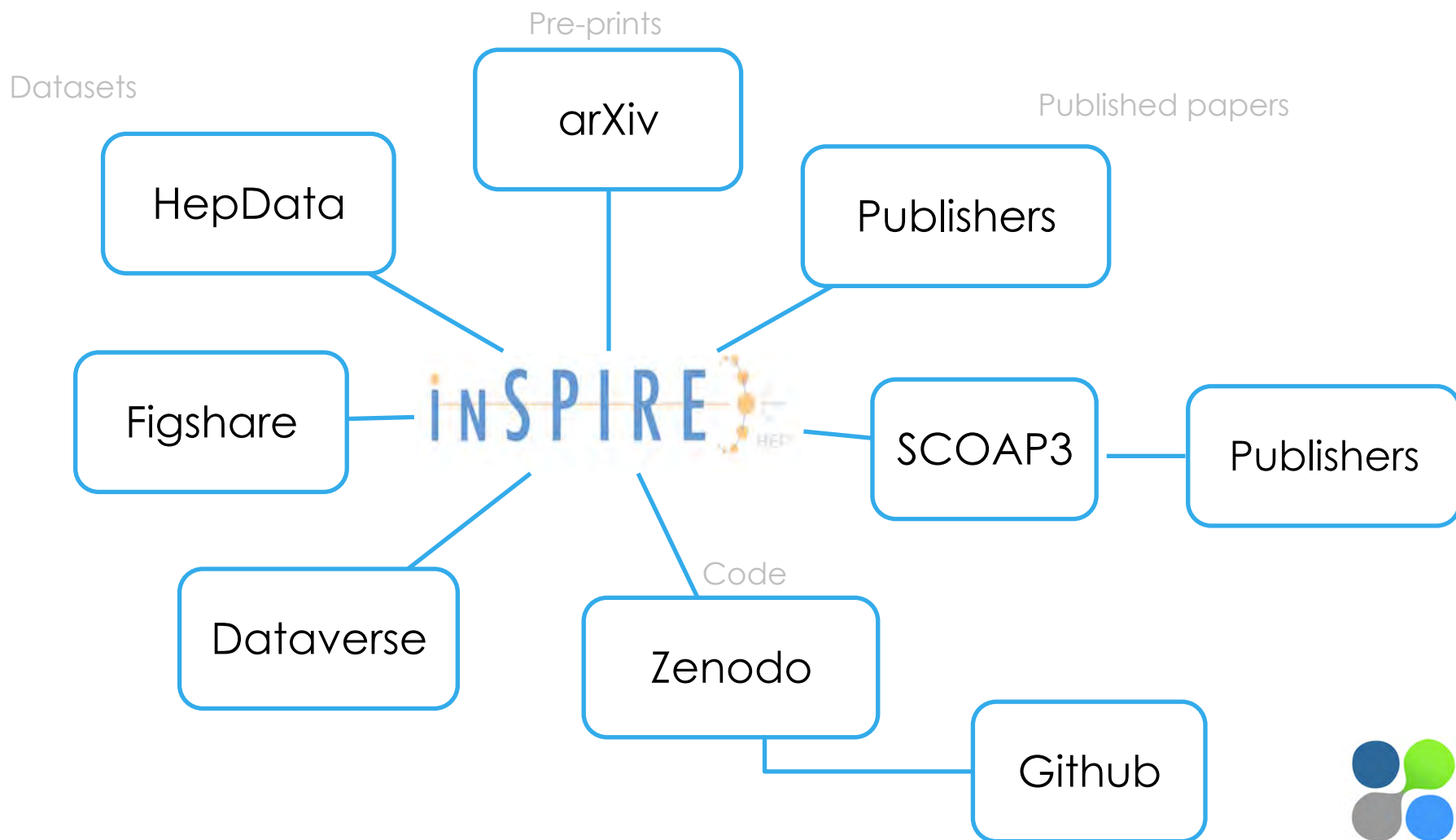
sunje.dallmeier-tiessen@cern.ch

What is INSPIRE?

- ▣ Based on Invenio
 - ▣ Digital library in High-Energy Physics
 - ▣ And an information hub:
 - ▣ Pre-prints
 - ▣ Published papers
 - ▣ Authors
 - ▣ Data
- } **Linked!**
- ▣ Some numbers:
 - ▣ 1.000.000 publications
 - ▣ 100.000 authors
 - ▣ 50.000 datasets
 - ▣ And institutions, conferences, journals, job offers...



What's inside?



Research data (citation)



Welcome to [INSPIRE](#), the High Energy Physics information system. Please direct questions, comments or concerns to feedback@inspirehep.net.

HEP :: HEPNAMES :: INSTITUTIONS :: CONFERENCES :: JOBS :: EXPERIMENTS :: JOURNALS :: HELP

Four

Information Citations (4) Files

Data from Figure 7 from: Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC

ATLAS Collaboration ([Aad, Georges \(Freiburg U.\)](#) **One** [Show all 2923 authors](#))

Cite as: ATLAS Collaboration (2013) HepData, <http://doi.org/10.7484/INSPIREHEP.DATA.26B4.TY5F> **Two**

Description: -2 log Likelihood for the $H \rightarrow WW^* \rightarrow l\nu l\nu$ channel in the $(\mu_{ggF+ttH} * B/BSM, \mu_{VBF+VH} * B/BSM)$ plane for a Higgs boson mass $m_H = 125.5$ GeV.

Preview not available

Note: * Temporary entry *

This dataset complements the following publication: [Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC](#) **Three**

Record created 2013-09-11, last modified 2013-12-16



Authors on INSPIRE



Welcome to [INSPIRE](#), the High Energy Physics information system. Please direct questions, comments or concerns to feedback@inspirehep.net.

HEP :: HEPNames :: INSTITUTIONS :: CONFERENCES :: JOBS :: EXPERIMENTS :: JOURNALS :: AYUDA

Cranmer, Kyle S.

[View Profile](#) [Manage Profile](#) [Manage Publications](#) [Ayuda](#)

Profile Name

2014-05-20 14:11:33

PERSONAL INFORMATION

PUBLICATIONS AND OUTPUT

STATS

Personal Details (HepNames)

Name Kyle S. Cranmer
Current Institution New York U.
E-mail cranmer@cern.ch
Links <http://physics.as.nyu.edu/obje...>
<http://twitter.com/KyleCranmer...>
<http://theoryandpractice.org/>

Fields HEP-EX
HEP-PH
PHYSICS

Experiments FNAL-E-0830
CERN-LHC-ATLAS
CERN-LEP-ALEPH
FNAL-TEV-CDF

Identifiers INSPIRE: INSPIRE-00074922
ARXIV: ARXIV-CRANMER-K-1
CERN: CERN-538804
ORCID: <http://orcid.org/0000-0002-3789-7054>
BAI: K.S.Cranmer.1

Period	Rank	Institution
1995 - 1999	UG	Rice U.
1999 - 2005	PHD	Wisconsin U., Madison
2005 - 2007	PD	Brookhaven

Publications

[Datasets](#) [External](#)

Two

- Supplementary Material for "A Novel Approach to Higgs Coupling Measurements"
- decouple software associated to arXiv:1401.0080
- Code associated to "Kernel Estimation in High Energy Physics"
- Data from figure 6 from: Measurement of charged-particle event shape variables in $\sqrt{s} = 7$ TeV proton-proton interactions with the ATLAS detector
- Data from figure 6 from: Measurement of charged-particle event shape variables in $\sqrt{s} = 7$ TeV proton-proton interactions with the ATLAS detector
- Data from figure 5 from: Measurement of charged-particle event shape variables in $\sqrt{s} = 7$ TeV proton-proton interactions with the ATLAS detector
- [irecord/1269077](#)
- Additional data from: Measurements of $W\gamma$ and $Z\gamma$ production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector at the LHC
- Additional data from: Measurements of $W\gamma$ and $Z\gamma$ production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector at the LHC
- Additional data from: Measurements of $W\gamma$ and $Z\gamma$ production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector at the LHC

Co-Authors

B.Mellado.1 (13)
W.Quayle.1 (11)
C.T.Potter.1 (9)
I.Aracena.1 (9)
S.L.Wu.1 (9)
B.Vachon.1 (8)
D.Damazio.1 (8)
R.Goncalo.2 (8)
S.H.Robertson.1 (8)
S.J.Hillier.1 (8)
[more](#)

Papers

	All papers	Single authored
All papers	446	10
Book	0	0
ConferencePaper	28	8
Introductory	0	0
Lectures	0	0
Published	366	2
Review	5	0
Thesis	1	1
Proceedings	1	0

Three

Citations Summary

446 papers found, 436 of them citeable (published or arXiv)

	Citeable papers	Published only
Number of papers analyzed:	436	366
Number of citations:	31380	29280
Citations per paper (average):	71	80
h_{HEP} index [?]	77	76

Breakdown of papers by citations:

	Citeable papers	Published only
Renowned papers (500+)	7	6
Famous papers (250-499)	10	10
Very well-known papers (100-249)	39	37
Well-known papers (50-99)	82	82
Known papers (10-49)	200	190
Less known papers (1-9)	68	38
Unknown papers (0)	30	3

Warning: The citations count should be interpreted with great care.
[Read the fine print](#)

One

Kyle Cranmer

ID <http://orcid.org/0000-0003-5769-7094>

Keywords: physics

Websites:
theoryandpractice.org

Personal information

Biography

Kyle Cranmer is an Associate Professor of Physics at New York University's Center for Data Science. He is an experimental particle physicist at the Large Hadron Collider, based in Geneva, Switzerland. Professor Cranmer joined the University of Wisconsin-Madison in 2005 and his B.A. in Mathematics from Rice University in 2007, he was awarded the Presidential Early Career Award for Scientists and Engineers via the Department of Energy's Office of Science and the Sloan Foundation's Career Award. Professor Cranmer develops statistical modeling, which was used extensively for the discovery of the Higgs boson. Associate professor of physics at NYU.

Education

University of Wisconsin Madison (2000 to 2005)
PhD

Rice University (1995-09 to 1999-05)
B.A.

Employment

New York University (2007-09 to present)
Associate Professor

Works

One

Data from Figure 7 from: Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC 2013-09

Very soon... !

ORCID

This profile is already connected to the following ORCID: **0000-0003-0762-2235**

Push my INSPIRE claimed publications to ORCID

?

LINK WORKS X

Australian National Data Service (ANDS) Registry

Import your research datasets into ORCID from Australian National Data Service (ANDS) and Research Data Australia (RDA). ANDS is partnering with Australian research institutions and data producing agencies to improve discovery and reusability of research data across many research domains from earth science to technology and engineering.

CrossRef Metadata Search

Search CrossRef's comprehensive metadata on journal articles, conference proceedings and monographs. Easily add search results to your ORCID profile.

INSPIRE-HEP author profiles

Link your ORCID profile with the INSPIRE-HEP portal to import your claimed publications and datasets in High-Energy Physics.



And a bit more complex...

The screenshot shows the GitHub interface for the repository 'svenkreiss / decouple'. It includes the repository name, commit count (44), branches (4), releases (9), and contributors (1). A file browser on the left lists files like 'Decouple', 'ModelGenerators', 'Plot', 'output', 'plots', 'plotsForPaper', 'scripts', '.gitignore', 'LICENSE', 'Makefile', 'README.md', 'requirements.txt', and 'requirements_dev.txt'. A Zenodo logo is overlaid on the repository content.

The screenshot shows the INSPIRE record page for 'decouple software associated to arXiv:1401.0080'. It features the INSPIRE logo, a navigation bar with links like 'HEP', 'HEPNames', 'Institutions', 'Conferences', 'Jobs', 'Experiments', 'Journals', and 'Help'. The record title is 'decouple software associated to arXiv:1401.0080' by Cranmer, Kyle; Kreiss, Sven (New York University). It includes a description, a list of files, and a table of files.

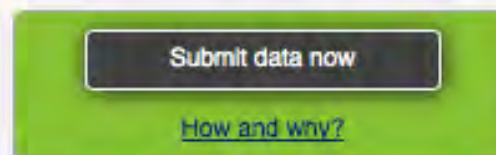
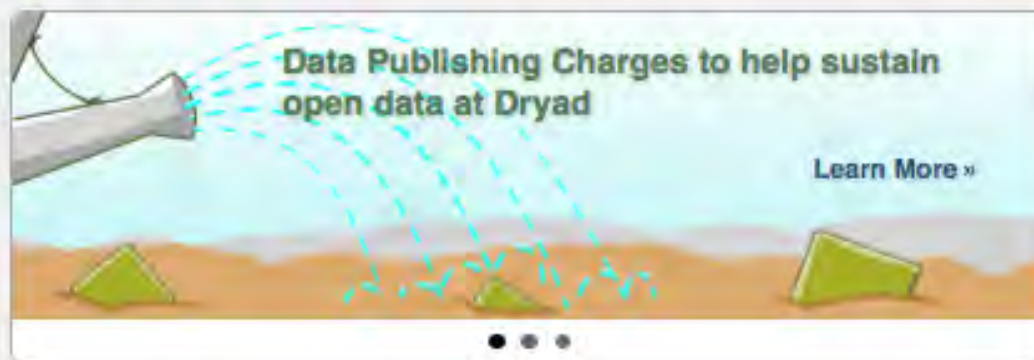
Name	Date	Size	Download
decouple-v1.2.5.zip	08 Mar 2014	266.6 kB	Download





Identifiers and Interoperability: Dryad Perspective

Ryan Scherle
Dryad Digital Repository



Search for data

Enter keyword, author, title, DOI, etc.

[Advanced search](#)

Browse for data

Recently published

Popular

By Author

By Journal

Recently Published Data

Phadke S, Zufall R, Paixao T, Pham S, Pham T (2013) Data from: Genetic background alters dominance relationships between mat alleles in the ciliate *Tetrahymena thermophila*. *Journal of Heredity* doi:10.5061/dryad.qv731

Chiu C, Jost L, Chao A (2013) Data from: Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. *Ecological Monographs* doi:10.5061/dryad.vr4k2

Salgado-Salazar C, Rossmann AY, Chaverri P (2013) Data from: Not as ubiquitous as we thought: taxonomic cryptsis, hidden diversity and cryptic speciation in the cosmopolitan fungus *Theloneotria discophora* (Nectriaceae, Hypocreales, Ascomycota). *PLoS ONE* doi:10.5061/dryad.q3s66

Morgan EMJ, Green BS, Murphy NP, Strugnelli JM (2013) Data from: Investigation of genetic structure between deep and shallow populations of the southern rock lobster, *Jasus edwardsii* in Tasmania, Australia. *PLoS ONE* doi:10.5061/dryad.656gf

Robinson BW (2013) Data from: Evolution of growth by genetic accommodation in Icelandic freshwater stickleback. *Proceedings of the Royal Society B*

Be part of Dryad



Publishers, societies, universities, libraries, funders, and other stakeholder organizations are invited to become **members**. Tap into an active knowledge-sharing

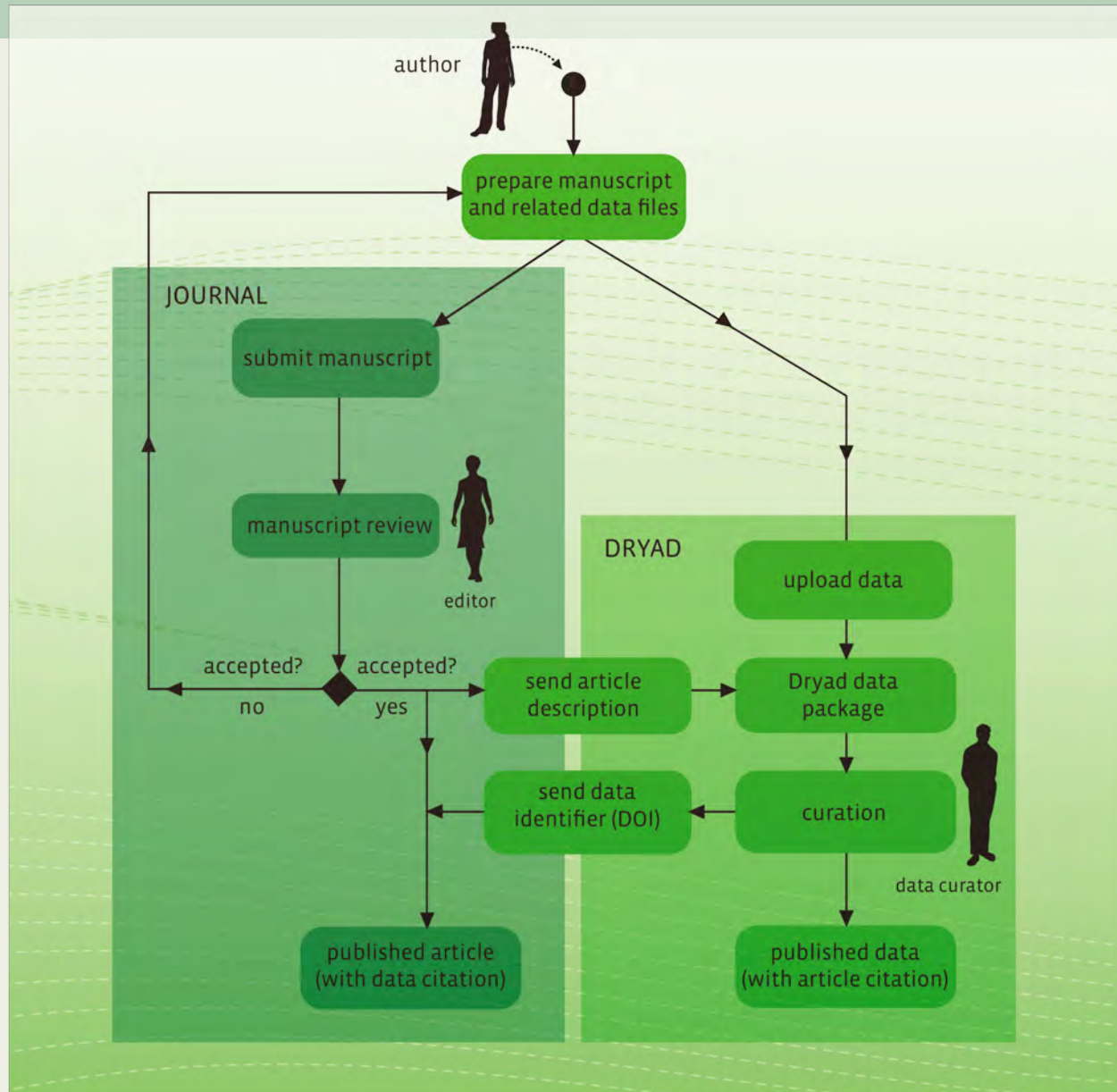
network, receive discounts on submission fees, and help shape Dryad's future.

[Submission integration](#) is a free service that allows publishers to coordinate manuscript and data submissions. It makes submitting data easy for researchers; makes linking articles and data easy for journals; and enables confidential review of data prior to publication.



Submission fees support the cost of keeping Dryad's content free to use. Flexible [pricing plans](#) provide volume discounts.

Integrated article/data submission



Encouraging citations



When using this data, please cite the original article:

Levy J, Foulsham T, Kingstone A (2012) Monsters are people too. *Biology Letters* 9(1): 20120850. [doi:10.1098/rsbl.2012.0850](https://doi.org/10.1098/rsbl.2012.0850)

Additionally, please cite the Dryad data package:

Levy J, Foulsham T, Kingstone A (2012) Data from: Monsters are people too. Dryad Digital Repository. [doi:10.5061/dryad.4rk06](https://doi.org/10.5061/dryad.4rk06)

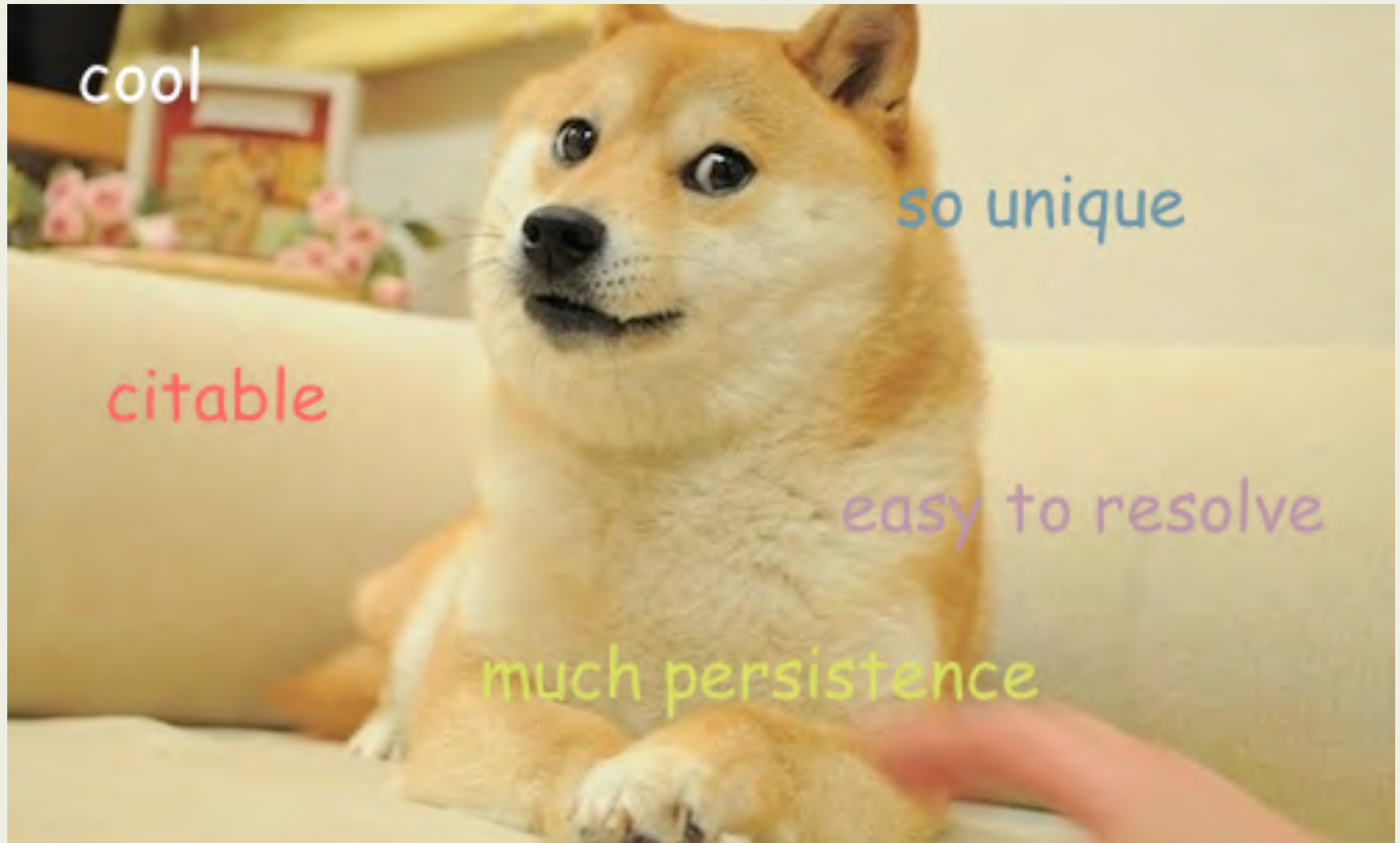
[Cite](#) | [Share](#)

Principles of Citable Identifiers



- Use DOIs
- Keep identifiers simple
- Use syntax to illustrate relationships
- Create new identifiers when (and only when) the meaning of the content changes

From: Creating Citable Data Identifiers. Ryan Scherle and Mark Diggory. Open Repositories 2012.



cool

so unique

citable

easy to resolve

much persistence

Opportunities



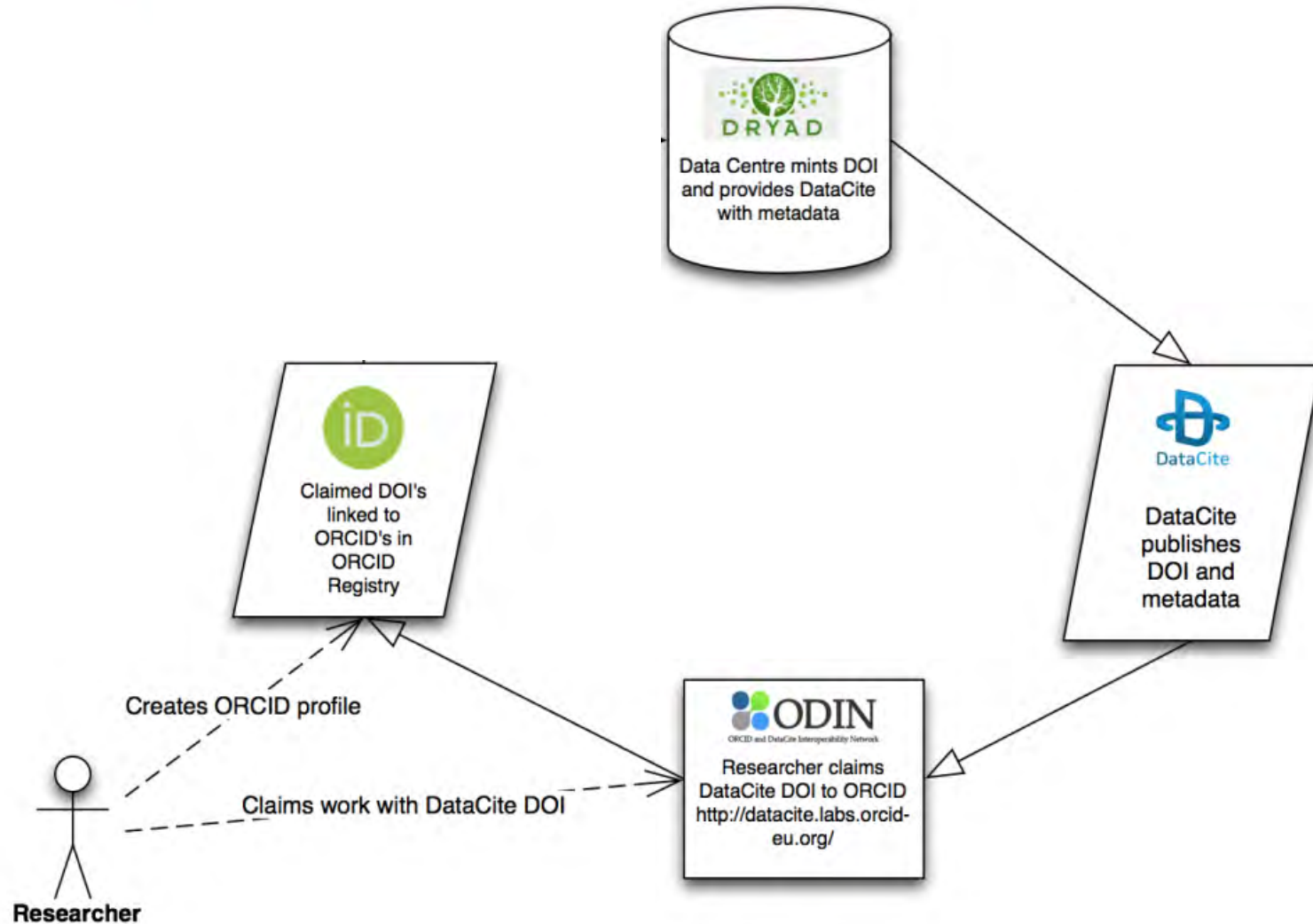
DataCite DOIs

- Leverage central infrastructure to distribute and transform metadata

ORCIDs

- Reduce ambiguity
- Permissions
- Add “missing” metadata

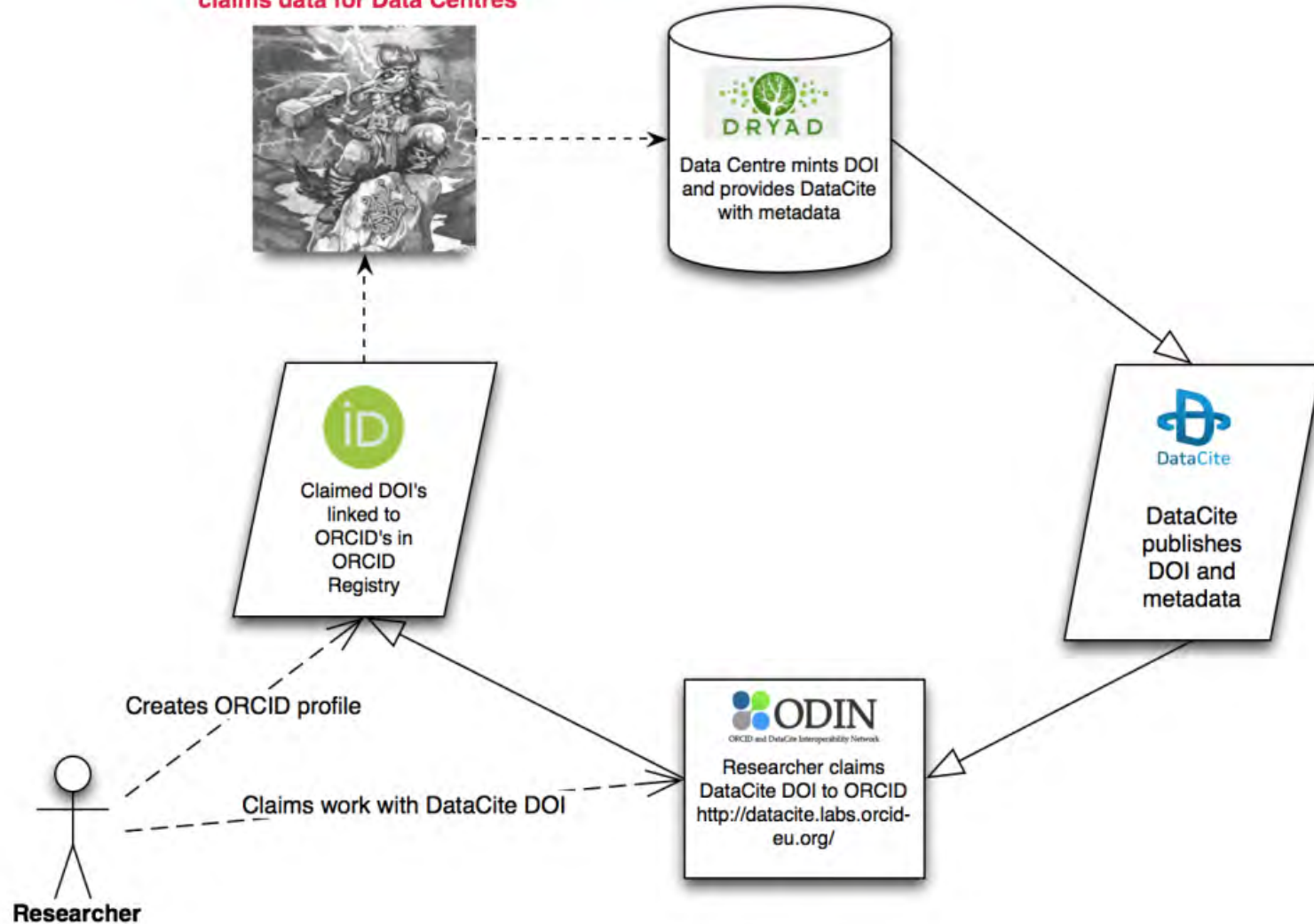
ODIN's HAMR



ODIN's HAMR



ODIN's HAMR provides ORCID claims data for Data Centres



Challenges



DOI metadata requirements vs privacy

Challenges



DOI metadata requirements vs privacy

Identifier: doi:10.5061/DRYAD.PK85D [Get link](#)

About the identified object

Location (URL): <http://datadryad.org/publicationBlackout>

Creators: (:tba)

Title: Dryad Item 10.5061/DRYAD.PK85D

Publisher: Dryad Digital Repository

Publication year: 9999

Subjects: (:tba)

Resource type: Dataset/DataPackage

Related identifiers: 10.5061/DRYAD.PK85D/1 [DOI] [HasPart]

Challenges



Prioritizing tool development for identifiers

Challenges



Prioritizing tool development for identifiers



Challenges



Levels of assurance – when using ORCID^s for permissions

Challenges



Levels of assurance – when using ORCIDiDs for permissions

ADD WORK

Who can see this?



Work category

Publication *

Role

Author

Work type

Book *

Credited

First

Title

Harry Potter and the Deathly Hallows *

OTHER WORK IDENTIFIERS

Identifier type

Challenges



Identifiers and content mutability

Challenges



Identifiers and content mutability



Challenges



Identifiers and content mutability



Challenges & Opportunities

- ▣ Confusion – between identifiers & services
- ▣ Tension between data needs and privacy concerns
- ▣ Identifier not enough – prioritizing tool development
- ▣ Levels of assurance of identifier & relationships
- ▣ Resolvability – of identifiers; consistency in metadata & APIs