Research Identifiers

Andrew Johnson
Research Data Librarian
CU Boulder Libraries
andrew.m.johnson@colorado.edu

Outline

- What are “research identifiers” and why are they important?
- Identifiers and profiles for researchers
- Identifiers for research outputs (e.g., journal articles, data sets, software/code)
- Identifiers in action
- Q&A

What are “research identifiers” and why are they important?

Research identifier

- A unique string, typically assigned by an organization, that is used to identify some object or entity involved in the research ecosystem
- Usually resolvable to or associated with a URL
- Intended to be persistent
- Intended to be interoperable

Research ecosystem

- Researchers interact with many organizations (e.g., universities, funders, publishers)
- Researchers interact with many systems (e.g., annual reporting, manuscript submission, repositories, social media and other third party services)
- Researchers produce many outputs (e.g., journal articles, data sets, software/code)

Disambiguation

- How do we know that Article A in Journal B cited by Article C was written by Researcher D at Institution E based on Code F using Data Set G housed in Repository H and created with funding from Agency I?
- Need unique and persistent identifiers for every object/entity involved
- Names are not unique
- URLs might not be persistent
- Local identifiers are unknown to outside systems

Metrics

- Trend toward using (relying on?) metrics for promotion/tenure, annual performance, departmental/institutional reputation, research impact
- Traditional metrics examples: article citations, h-index, Journal Impact Factor
- Altmetrics examples: article views/downloads, mentions on social media, citations in policy documents, saves in Mendeley
- Identifiers allow for more accurate tracking and measuring of outputs (not just articles)
- Identifiers allow outputs to be more easily associated with researchers

Reproducibility

- Increasingly important for computational science
- “Replication crises” in some disciplines
- Emphasized by funders and publishers
- Requires access to software/code and data underlying research findings published in articles
- Must be able to identify which software/code was used to create/analyze which data
- Many funders and publishers requiring identifiers as “proof” of access

Identifiers and profiles for researchers

Researcher identifiers or just profiles?

- ISNI: http://www.isni.org/
- ORCID: http://orcid.org
- ResearcherID: https://www.researcherid.com/
- Scopus Author ID: https://help.elsevier.com/app/answers/detail/a_id/2845/p/8150/
- Google Scholar Citations: https://scholar.google.com/citations
- CU Experts (VIVO): https://experts.colorado.edu/

ORCID iDs

- One of the most widely used identifiers for researchers (over 2.7 million registered to date)
- Example: 0000-0001-5109-3700
- Associated URL and profile: http://orcid.org/0000-0001-5109-3700
- Anyone can register here: https://orcid.org/register
- Additional instructions for faculty at CU Boulder: http://www.colorado.edu/fis/orcid

ORCID at CU Boulder

- Over 1400 known claimed ORCID iDs for faculty at CU
- Appear in FRPA (annual reporting system) and CU Experts: https://experts.colorado.edu/
- Example: https://experts.colorado.edu/display/fisid_151402
- Plans for additional uses and implementation in other systems

Identifiers for research outputs

Too many to name here...

- ...so we will focus on the Digital Object Identifier (DOI)
- Most prominent identifier for journal articles
- Also frontrunner for identifier of choice for data sets
- Possible to register DOIs for other outputs like software/code
- Interested in other identifiers? See this article by Duerr, Downs, Tilmes, et al. (2011): http://dx.doi.org/10.1007/s12145-011-0083-6

DOI governance

- Governed and maintained by the International DOI Foundation (IDF)
- DOIs are registered by Registration Agencies (or their members), including:
  - Crossref (primarily journal articles)
  - DataCite (primarily data sets)
- Registration Agencies also maintain and provide requirements for metadata associated with DOIs (especially important for data citation)
- Membership with a Registration Agency typically requires fees

DOI structure

- Made up of prefix and suffix separated by “/”
- Prefix always begins with “10” followed by a period and four digits (representing the registrant)
- Suffix is the id for the individual object/entity (can include alphanumeric characters subdivided by periods)
- Example: doi:10.6084/m9.figshare.2066037.v16
- Made actionable by using the DOI resolver
- Example: https://dx.doi.org/10.6084/m9.figshare.2066037.v16

How to get DOIs for research outputs

- Register DOIs yourself with EZID subscription (annual membership fee): http://ezid.cdlib.org/
- Upload outputs to Figshare: https://figshare.com/
- Create a public project on the Open Science Framework: https://osf.io/
- Connect GitHub to Zenodo or Figshare to automatically create DOIs whenever software/code releases are made: http://ivory.idyll.org/blog/2016-using-zenodo-to-archive-github.html

A word of caution

- To a certain degree, a DOI is a good indicator of persistence
- Sometimes a DOI is an indicator of access
- A DOI is NOT an indicator of quality
- Funders, publishers, and others sometimes seem to assume that quality is inherent to things with DOIs (especially for data)

Identifiers in action

Impactstory

- Register for free using Twitter or ORCID: https://impactstory.org/
- Another profile system
- Pulls info from ORCID (which itself pulls info from a number of sources, including Crossref and DataCite)
- Creates a profile based on ORCID info and tracks impact of works on Twitter, Mendeley, etc.
- Also looks for open access versions of paywalled content
- Tracking primarily relies on ORCID iDs and DOIs

CU Boulder Elements

- CU Boulder faculty can log in with IdentiKey at https://elements.colorado.edu
- More information here: http://www.colorado.edu/fis/CUBE
- Effort to make it easier to track faculty accomplishments
- Also helps populate CU Experts profiles and FRPA annual reporting system
- Searches Crossref and other data sources (using DOIs and other identifiers) to find publications and other research outputs from CU Boulder faculty

Questions?