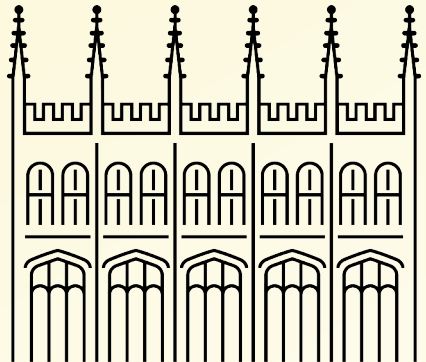
Introduction to Persistent Identifiers

Bodleian iSkills

Neil Jefferies, Head of Innovation
Open Scholarship Support, Bodleian Libraries
University of Oxford

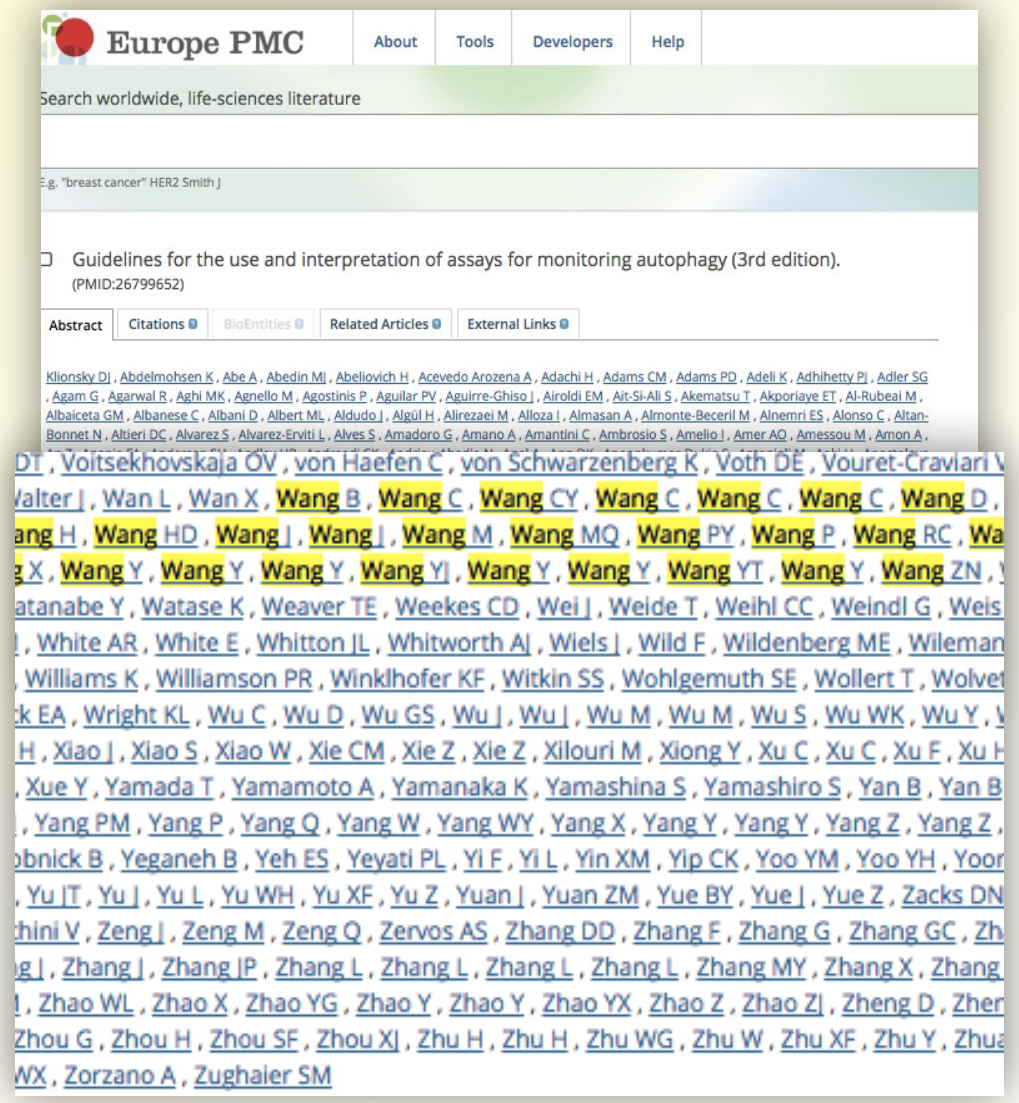
Introduction to Persistent Identifiers

- What is a Persistent Identifier (PID)?
- How do PIDs work?
- Benefits of using PIDs
- PIDs in Practice
- National PID Initiatives



The Problem

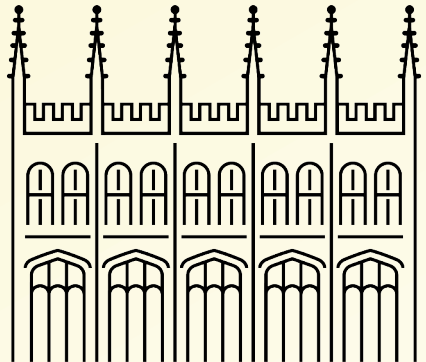
- As outputs become more varied and widely disseminated, name/title becomes an increasingly unreliable indicator
- 38 authors with family name “Wang”, and 6 “Wang, Y”
- Names/titles can be entered many different ways: Zu,Y vs Zue,Y vs Zhu,Y
- Citation and referencing standards do not help much
- At the same time, stakeholders such as funders demand better reporting



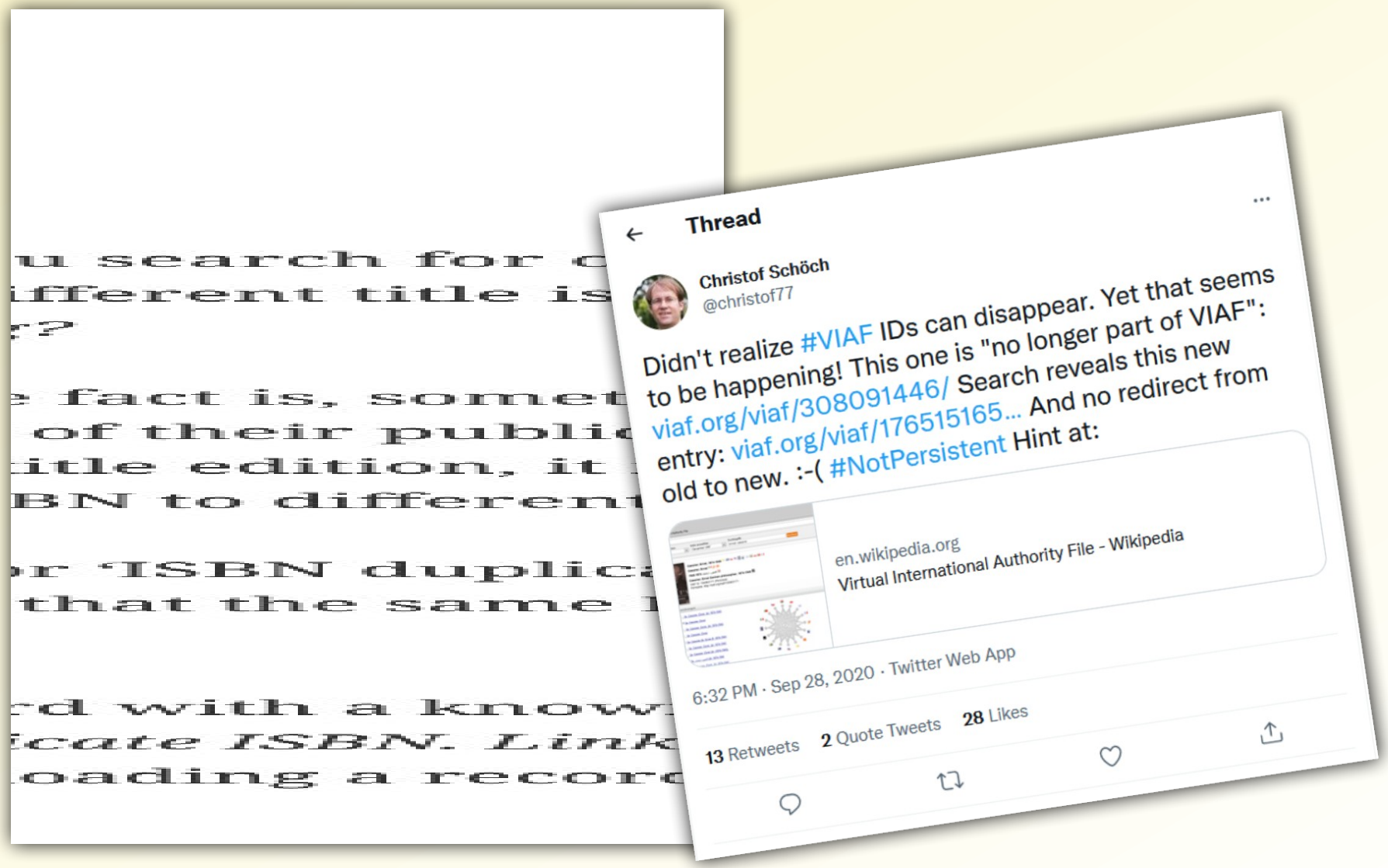
The Solution

- A Persistent Identifier (PID) is a long-lasting (digital) reference to an entity such as a document, file or anything else. They are often expressed as web URL's.

- **Unique** – an identifier should only ever identify one entity. An entity should only have one identifier of any particular type.
- **Persistent** – identifiers should continue to exist as long as necessary. As should the entity to which they refer.
- **Actionable** – identifiers should be usable to retrieve information about the entity it refers to. Ideally, the entity itself can be accessed.
- https://en.wikipedia.org/wiki/Persistent_identifier

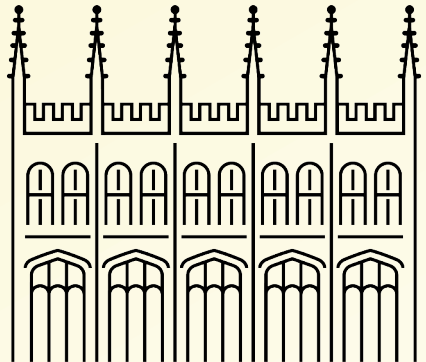


This is harder than you might think!



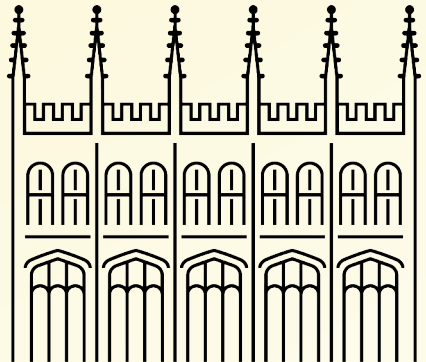
How do PIDs Work?

- A PID is typically created/requested by the organisation that holds the entity to which it refers.
- A registration authority can ensure that PIDs are **unique**.
  - They may issue an organisation a range of identifiers or a prefix.
  - Some algorithmic PIDs (such as UUID) can avoid this by being highly random.
- The organisation should ensure that the PID and the underlying entity are **persistent**. Consequently, they are often memory organisations such as libraries or archives.
- The registration authority frequently provides a **persistent resolver service** that links a PID to the entity to which it refers, making it **actionable**.
  - The organisation needs to update the resolver if they move the entity.



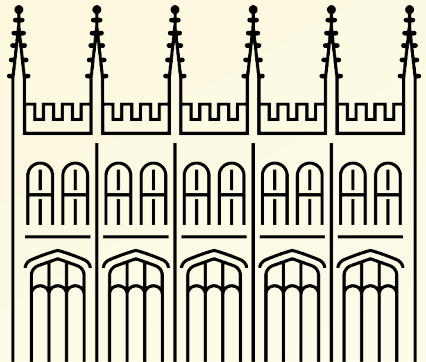
Benefits of Using PIDs

- PIDs do solve the original disambiguation problem
- Assigning a PID should also focus the organisation responsible for the underlying entity on meeting their persistence obligations
- However, a PID is very useful to everyone else when it is re-used...
  - A (good) PID also has associated standardised metadata about the underlying entity
  - Systems can automatically copy and process this metadata. This eliminates re-keying of information, simplifying and speeding up administration for everyone involved



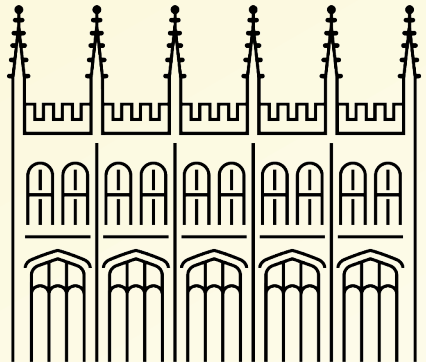
- PIDs feature in the Tickell Reports on Research Bureaucracy and follow-ups

PID's in Practice

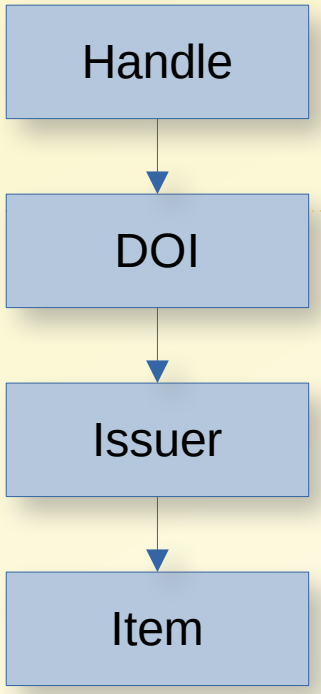


PIDs in Research

- Several PIDs are well-established in the research context.
  - **CrossRef DOIs** (Digital Object Identifiers) for published journal articles
  - **DataCite DOIs** for non-article research outputs are increasingly being adopted and are supported at Oxford by a number of services.
  - **ORCIDs** for researchers
- Demand for improved reporting, transparency and reproducibility has led to other identifier systems emerging.
- UKRI and Jisc in the UK, OpenAIRE in the EU, and others, are looking to standardise on a few…
  - **Grant IDs** (a special type of CrossRef DOI) are self-explanatory
  - **RAIDs** (Research Activity IDentifers) identify research "projects" from a researchers' point-of-view
  - **RORs** (Research ORganisation ID's) for funders, institutions and other corporate bodies



DOIs



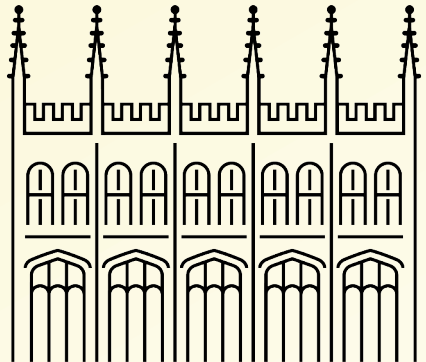
- DOIs are subset of the Handle standard (described in IETF RFC 3650-2) which is operated by the DOI Foundation (DF). They can appear in several forms
  - doi:10.1000/182 (this is the “proper” way to do it)
  - https://dx.doi.org/10.1000/182 (**actionable**, and the CrossRef recommended way to do it)
- 10.1000 is termed the prefix...
  - 10. indicates that this is a DOI, other numbers indicate other subsets of the Handle system
  - 1000 indicates the DOI issuer
- /182 is the suffix, which indicates the individual item which the DOI refers to
- https://dx.doi.org is a DOI resolver provided by DF
- The DF delegates management of DOI issuing to registration authorities.
  - Authorities allocate issuer numbers to approved issuers or they can act as the issuer themselves
  - Set standards for the metadata that accompanies DOIs under their administration
  - Provide services to the users of the DOIs (both entity holders and users)

CrossRef DOIs

- CrossRef is consortium of over 3000 publishers that acts as a DOI registration authority for journal articles. Other article authorities exist, e.g. Airiti for Chinese
- CrossRef allocates prefixes to publishers who are responsible for generating suffixes
- Publishers can allocate DOI's on acceptance rather than publication, which can be useful if you need to include reference in other places
- **ORA** pre-prints or AAM's can link to publisher DOIs
- CrossRef provides a metadata feed which can be used by other services to save rekeying. e.g. **Symplectic** (and thence **ORA**), **ResearchFish**, ORCID profiles

DataCite DOIs

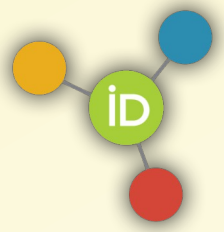
- The DataCite consortium acts as a DOI registration authority for research outputs that are not journal articles
  - Originally focussed on data, but now includes preprints, software, methologies etc.
  - DataCite has its own metadata schema to reflects this
- DataCite allocates prefixes to repositories who are responsible for generating suffixes
- **Oxford is a member of DataCite** (through the BL)
  - ORA Data deposits can get an Oxford DataCite DOI
  - The Bodleian provides a DataCite DOI service so that Oxford outputs stored elsewhere can get an Oxford DataCite DOI (provided they can give some assurance of persistence!)
- Material in the **Sustainable Digital Scholarship** service can get DOIs from FigShare
- Material managed using the **Open Science Framework (OSF) at Oxford** can get DOIs from the Center for Open Scholarship
  - For software, GitHub can feed into OSF (or Zenodo)
- **Symplectic** and ORCID profiles can pick up a DataCite metadata feed

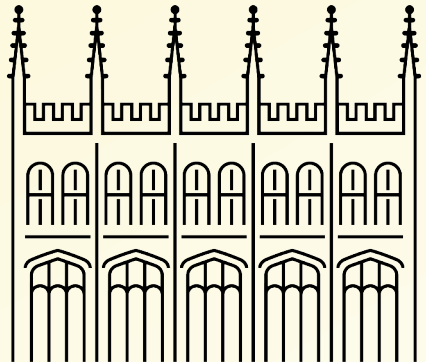


Researcher IDs

- Many different researcher identifiers already exist
- Most are not controlled by researchers (who are, after all, technically responsible for their own persistence)
- Most are incomplete, so none are authoritative sources
- Some, like academia.edu, don't even check you are who you claim you are
- Some aren't very persistent (VIAF, as mentioned previously)
- Metadata standards and services vary between identifiers

ORCID





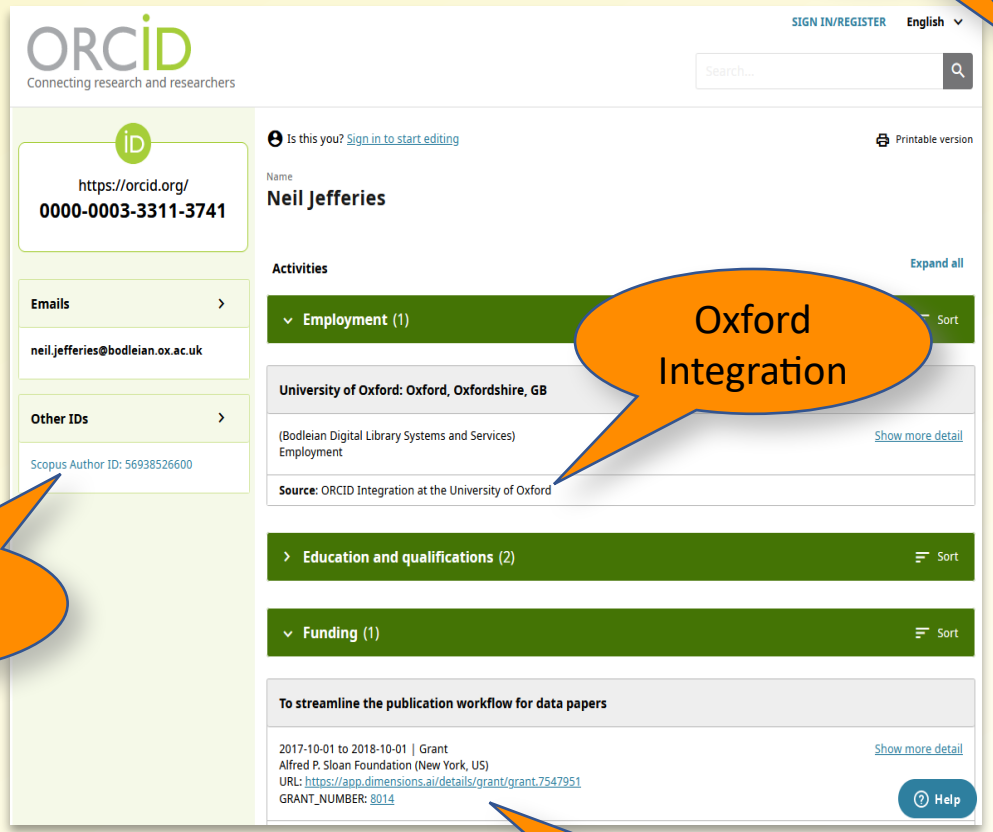
~~Open Researcher and Contributor IDentifier~~ ORCID

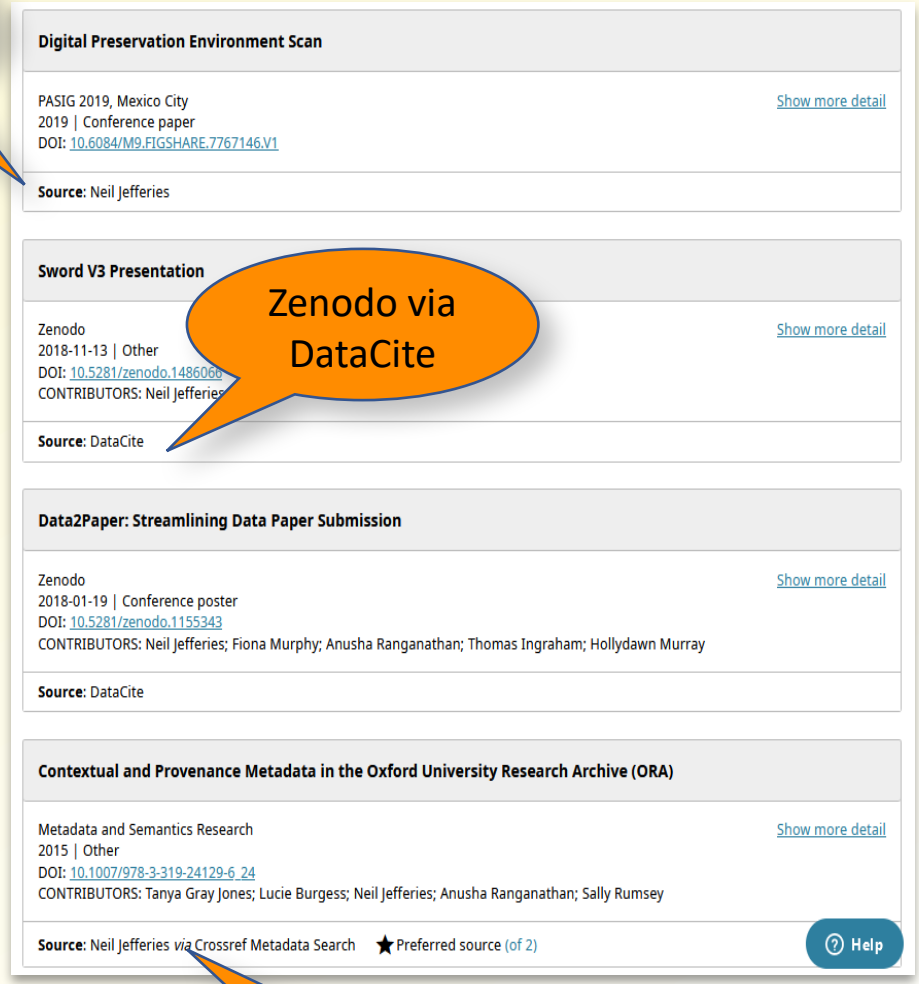
- **The ORCID iD**: a unique, persistent identifier free of charge to researchers
- An **ORCID record** connected to the ORCID iD
- A set of **Application Programming Interfaces** (APIs) to integrate with other services
- Supported in the UK by the UK ORCID Consortium operated by Jisc

ORCID Principles (simplified)

- **Researcher control** of the content of your record, and who can access it
- **Inclusive community governance**, mostly not-for-profit stakeholders
- **Openness** of software, data and governance
- **Persistence**, institutional fees are enough to ensure sustainability
- **Free** for researchers

ORCID Integrations





Manual Entry

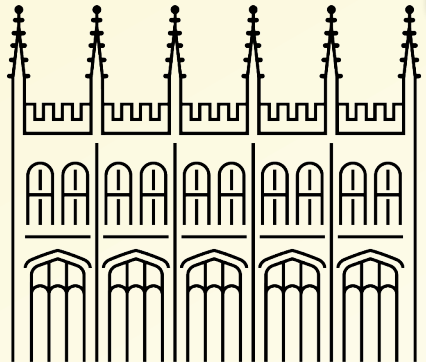
Zenodo via DataCite

Oxford Integration

Scopus Integration

Sloan Foundation

Chapter via CrossRef



ORCID at Oxford

In Oxford

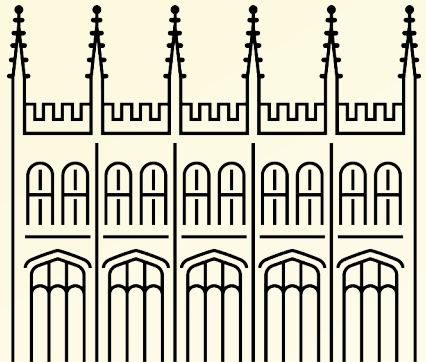
- Integration developed as part of the Jisc/ARMA ORCID pilot in 2014-15
- Your ORCID can be linked to your Single Sign-On ID
  - Doing this updates your ORCID record with a validated University affiliation
- Systems that use SSO can look up your ORCID, and fill it in automatically (e.g. ORA)
- Symplectic is a third-party service so you need to give it permission to access your ORCID too
- REF 2021 recommended researchers use ORCID

Outside Oxford

- Sign in to ORCID with your SSO
- This also works for systems that accept ORCID logins such as the Open Science Framework and Zenodo
  - If you have used your SSO to access Outlook, then you should be automatically logged in to them
  - OSF can get your affiliation from your ORCID record
- Some funders accept ORCIDs in lieu of a resume. NSF fast-tracks such proposals.
- If you leave Oxford, your original ORCID password will still work

UK PID History

- 2000 CrossRef founded and pilots (protoype) DOIs and services
- 2012 ISO 26324 published, describes the DOI standard for reliable online referencing of objects
  - OpenURL was not really successful, though still used to link to paywalled material
- 2012 ORCID created to disambiguate authors
- 2012 DataCite UK launch, led by the British Library
- 2015 UK ORCID Consortium launch, led by Jisc
- Other services and benefits emerged later (e.g. SCHOLIX)
- 2018 Tickell report on research bureaucracy recommends the use of PIDs and that Jisc should lead the work on identifying suitable candidates
- 2019 Jisc, supported by UKRI starts work on developing a national PID Strategy, publishing outcomes in 2020
  - Identifying the “Priority PIDs” that are needed to achieve (most of) the Tickell efficiency goals
  - Developing a cost-benefit analysis that demonstrated the value of widespead PID usage
  - Recommend mechanisms for enabling the UK research community to realise these benefits



Towards a National PID Strategy

- Priority PIDs
  - ✓ ORCID
  - ✓ DOI (article and DataCite)
  - Grants can be registered with CrossRef and get a Grant ID
    - New DOI type, limited adoption
    - Supports ORCID and ROR
    - GrantID is something completely different
  - ROR IDs identify organisations
    - RingGold data etc, needs cleanup
    - Supported by OSF
  - RAID identifies research “activities”
    - Unfunded or multiple grants
    - Run by Australian Research Data Commons

- Funding and research are not limited by national boundaries
- PID initiatives (e.g. OpenAIRE in the EU) need to be cordinated worldwide
- RDA (Research Data Alliance) National PID Strategy Working Group
- Governance of systems an issue – RAID needs resolving
- UK Research Identifier National Coordinating Committee (RINCC)
  - Working with UKRI and Jisc on next steps
  - National PID Support structures?

Questions