

# DataCite and PIDs at the British Library

*DPC webinar: Choosing & Using Persistent Identifiers*  
*6 September 2023*

**Jez Cope (he/him)** |  [@petrichor@digipres.club](mailto:@petrichor@digipres.club)  
Data Services Lead  
The British Library

*These slides (and others)*  
*at [jezcope.gitlab.io/bl-presentations](https://jezcope.gitlab.io/bl-presentations)*

# Overview

- Open Data, FAIR and CARE
- DataCite for DOIs
- PIDs at the British Library
  - *Persistent Identifiers as IRO Infrastructure* project
  - Persistent Identifier policy

# Research Data is...

- Text, documents, corpora, images, maps, web pages, spreadsheets, interview transcripts, video, audio, 3D models, ...
- Any source of information you use to support a research conclusion or argument

## Research Data is...

- Text, documents, corpora, images, maps, web pages, spreadsheets, interview transcripts, video, audio, 3D models, ...
- Any source of information you use to support a research conclusion or argument

## Research Data **Management** is...

1. Ensuring your data is safe, secure, documented, accurate, findable, useful, preserved, appropriately shared
2. Planning the actions necessary for the above

# Tell the *whole* story

Not just the book/article! Yes, data, but also...

- Methodology
- Analysis code
- Intellectual property
- Primary & secondary sources
- Public engagement activities
- Creative outputs: performances, artworks, poetry...

Wellcome Trust call this "Outputs Management"

## Open Data

Available for others  
to reuse...

...without charge

...without  
unreasonable  
restrictions

## FAIR Data

**F**indable  
**A**ccessible  
**I**nteroperable  
**R**eusable

*Focus more on  
metadata/formats*

## CARE Data

**C**ollective Benefit  
**A**uthority to Control  
**R**esponsibility  
**E**thics

*Focus more on  
people &  
communities*

## Open Data

Available for others  
to reuse...

...without charge

...without  
unreasonable  
restrictions

## FAIR Data

**F**indable  
**A**ccessible  
**I**nteroperable  
**R**eusable

*Focus more on  
metadata/formats*

## CARE Data

**C**ollective Benefit  
**A**uthority to Control  
**R**esponsibility  
**E**thics

*Focus more on  
people &  
communities*

---

**Not all data can be open!**

# What is a DOI?

<https://doi.org/10.5061/DRYAD.6WVPZGMW3>

Prefix: 10.5061 - allocated to repository

Suffix: DRYAD.6WVPZGMW3 - assigned by repository

# What is a DOI?

<https://doi.org/10.5061/DRYAD.6WWPZGMW3>

Prefix: 10.5061 - allocated to repository

Suffix: DRYAD.6WWPZGMW3 - assigned by repository

---

e.g. Borghi, J., & Van Gulick, A. (2020). Data Management and Sharing: Practices and Perceptions of Psychology Researchers (Version 3) [Data set]. Dryad. <https://doi.org/10.5061/DRYAD.6WWPZGMW3>

# DataCite

- DOIs are primarily for objects to be cited
  - Also support persistent access to and linked data about these
- Well-established for "traditional" publication (Crossref)
- Data citation needed an identifier (2000's) — DOI was ideal
  - Also value for other research outputs: software, dissertations, standards, presentations, videos, images, grey literature, ...

# Benefits

- Multi-tiered governance model:  
Handle > DOI Foundation > Registration Agency > Consortium >  
Member > Repository
- Financial sustainability
- Consistent, community-driven metadata schema
- Batteries included: web GUI & API; link checking; resolver(s); platform integrations; ...

# User & developer services

- Machine-readable metadata enables value-added services
- DataCite Commons: [commons.datacite.org](https://commons.datacite.org)
- Event Data: [support.datacite.org/docs/eventdata-guide](https://support.datacite.org/docs/eventdata-guide)
- GraphQL API: [support.datacite.org/docs/datacite-graphql-api-guide](https://support.datacite.org/docs/datacite-graphql-api-guide)
- Wider ecosystem: e.g. reference manager support; altmetrics; ...

# DataCite history

- Established 2009 after a pilot project by German National Library of Science & Technology (TIB)
- Initial focus on datasets; soon expanded to "research outputs & resources" broadly
- Now has 1,375 members internationally, independently or via one of 33 consortia

# UK DataCite Consortium in numbers

Open to UK universities and research(-supporting) organisations

- **111** Consortium Members  
with **164** Repositories
- **715,926** Findable DOIs  
of **49,566,046** total for DataCite
  - **62,199** in 2022
  - **51,615** so far in 2023...

# *PIDs as IRO Infrastructure* project

- Funded by Arts & Humanities Research Council [Towards a National Collection \(TaNC\) programme](#)
- Goal: make recommendations on Persistent Identifier use in AHRC Independent Research Organisations
- Investigators:
  - The British Library (lead)
  - Royal Botanic Garden Edinburgh
  - The National Gallery
  - University of Glasgow
- Partners: Science Museum Group; V&A; Natural History Museum

# *PIDs as IRO Infrastructure* key outputs

Practical guidance: [tanc-ahrc.github.io/PIDResources](https://tanc-ahrc.github.io/PIDResources)

Case studies: Persistent Identifiers at...

- ... British Library — [doi:10.23636/1242](https://doi.org/10.23636/1242)
- ... National Gallery — [doi:10.23636/1243](https://doi.org/10.23636/1243)
- ... Natural History Museum — [doi:10.22020/k99s-we61](https://doi.org/10.22020/k99s-we61)
- ... RBG Edinburgh — [doi:10.23636/pvfs-n308](https://doi.org/10.23636/pvfs-n308)

More at [tanc-ahrc.github.io/HeritagePIDs/resources.html](https://tanc-ahrc.github.io/HeritagePIDs/resources.html)

# PIDs at the British Library

## In our own collection

- Collection items: ARK; ISBN/EAN; ISSN; DOI; ...
- People/organisations: ISNI; ORCID; ROR; ISIL; ...
- Bibliography: FAST; LCCN; Shelfmarks; ...

## External PID services

- DataCite DOIs
- ISNI (International Standard Name Identifier)

# British Library Persistent Identifier policy

*"More IROs, higher education institutions and heritage organisations should implement policies on use of PIDs to support linking of items and their metadata across institutional boundaries, and identify a minimum technical passive provision for PIDs that future-proofs new tools and systems for their use"  
— PIDs as IRO Infrastructure recommendation*

# British Library Persistent Identifier policy

- **Scope:** identifiers used
- **Excludes:** namespaces; temporary identifiers; external identifier services

# British Library Persistent Identifier policy

- **Scope:** identifiers used
- **Excludes:** namespaces; temporary identifiers; external identifier services
- **Uses:**
  - Create: we create new identifiers to apply to entities in the collection
  - Contribute: we use these identifiers and have a role in their maintenance
  - Consume: we use externally-allocated and -maintained identifiers

# British Library Persistent Identifier policy

Also sets out:

- **Principles:** high-level intentions for the policy (e.g. persistence, interoperability)
- **Requirements:** expectations for PIDs we create/contribute/consume
- **Responsibilities:** who is the primary internal owner of each PID scheme

# British Library Persistent Identifier policy

Also sets out:

- **Principles:** high-level intentions for the policy (e.g. persistence, interoperability)
- **Requirements:** expectations for PIDs we create/contribute/consume
- **Responsibilities:** who is the primary internal owner of each PID scheme

Full policy available at [doi:10.23636/kwgh-pc35](https://doi.org/10.23636/kwgh-pc35)

# Thanks!

- Rachael Kotarski
- Frances Madden
- Many other colleagues at the British Library, DataCite and elsewhere!

---

Contact: [datasets@bl.uk](mailto:datasets@bl.uk)



**Thank  
you**