

Demystifying Identifiers for, Global Research Impact

Olatunbosun Obileye

18th of October 2023
IASSIST Africa Regional Workshop 2023



[@datacite](https://twitter.com/datacite)



@datacite@openbiblio.social
@paul4kant@scicomm.xyz

Understanding Identifiers

- ★ A long lasting reference to a digital resource (ORCID)
- ★ A unique long-lasting reference to digital objects of various types, and are a core element of open research. (Westminster)
- ★ PIDs are labels that locate, identify and share information about digital objects
- ★ A PID may be connected to a set of metadata describing an object rather than to the object itself.
- ★ Allow different platforms to exchange information consistently and unambiguously and thus provide a reliable way to track citations and reuse.

PIDs

What is a persistent identifier (PID)?

<https://doi.org/10.34848/GJO6SY>

Unique alphanumeric string referring to a digital resource.



<https://research-data.urosario.edu.co/dataset.xhtml?persistentId=doi:10.34848/GJO6SY>

*Always points to the same resource
(a metadata representation)*

DOIs for research outputs and resources

<https://doi.org/10.5281/zenodo.3630248>



ORCID iDs for researchers

<https://orcid.org/0000-0001-6622-4910>



ROR IDs for research organizations

<https://ror.org/01y2jtd41>



PIDs for people, places and things

**PIDs for people (researchers)
include ISNI and ORCID**



<https://orcid.org/0000-0001-6622-4910>



**PID for places (research
institutions) include ROR**



<https://ror.org/01y2jtd41>



**PIDs for things (research outputs and
resources) include DOIs, handles, IGSN,
ARK and more**



<https://doi.org/10.5061/dryad.708qr>



Making research data FAIR With PIDs



Findable

(Meta)data are assigned a globally unique and persistent identifier



Accessible

(Meta)data are retrievable via an identifier using a standardized protocol
Metadata are accessible, even when the data are no longer available



Interoperable

(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.



Reusable

(Meta)data are richly described with a plurality of accurate & relevant attributes

Why are PIDs important?

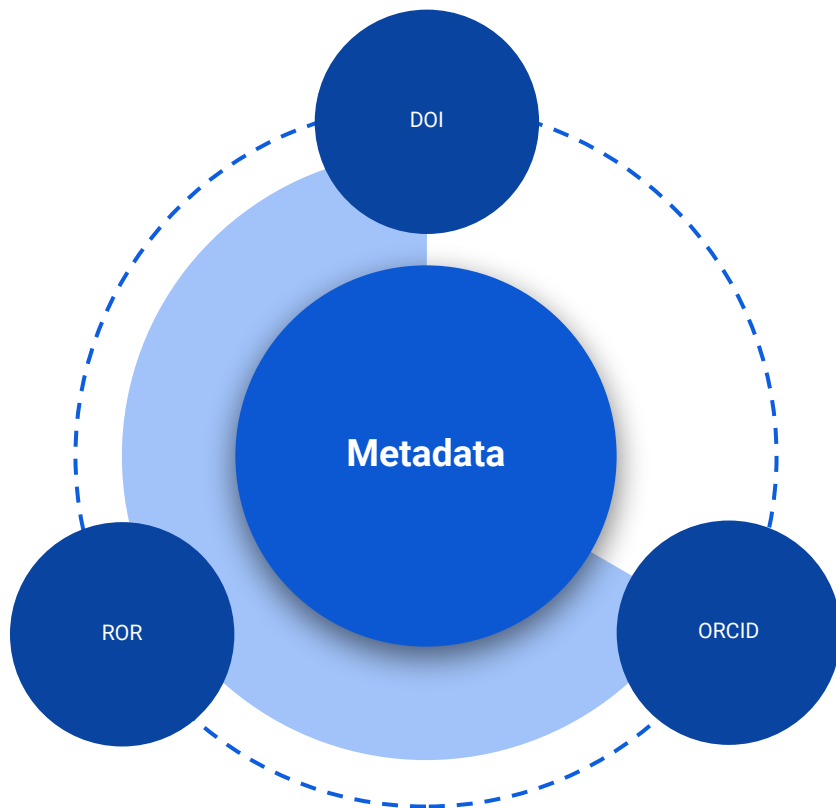
PIDs like DOIs, ORCID iDs y ROR IDs increase the **discoverability, access, citation, reutilization, and recognition** of research outputs and resources



VALUE

Connecting Research PIDs and Metadata

Find and connect Metadata



Metadata schema - DataCite DOI Example



Add your DOI metadata following DataCite Metadata Schema

DataCite DOIs are suitable for a wide range of research outputs

Mandatory	Recommended	Optional
Identifier	Subject	Language
Creator	Contributor	AlternateIdentifier
Title	Date	Size
Publisher	RelatedIdentifier	Format
PublicationYear	Description	Version
ResourceType	GeoLocation	Rights
		FundingReference
		RelatedItem

Resource types

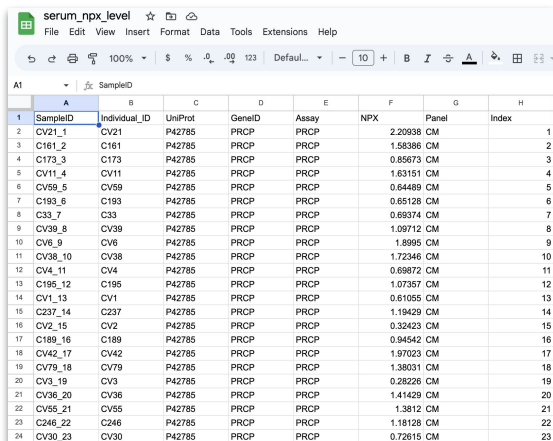
Audiovisual	Model
Book	OutputManagementPlan
BookChapter	PeerReview
Collection	PhysicalObject
ComputationalNotebook	Preprint
ConferencePaper	Report
ConferenceProceeding	Service
DataPaper	Software
Dataset	Sound
Dissertation	Standard
Event	Text
Image	Workflow
InteractiveResource	Other
Journal	
JournalArticle	

Advancing Knowledge Citation and Reuse

Citations

Citation means reference to another object the researchers have used as part of their work, this may be articles, data, software or other objects.

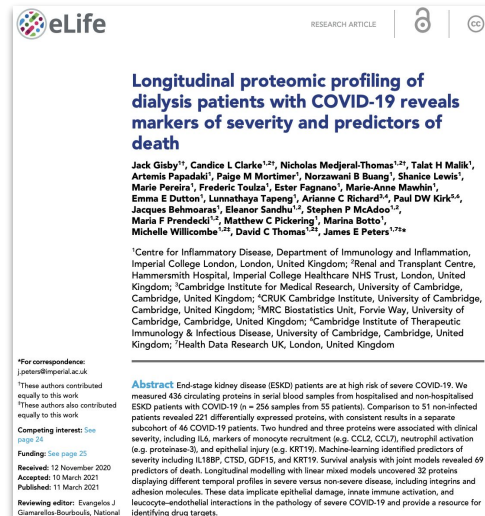
Citations are useful as they clearly point to the object being used or reused in research - they provide a link between the research objects



A1	SampleID	Individual_ID	UniProt	GeneID	Assay	NPX	Panel	Index
1	CV21_1	CV21	P42785	PRCP	PRCP	2.20938	CM	1
2	C161_2	C161	P42785	PRCP	PRCP	1.58386	CM	2
3	C173_3	C173	P42785	PRCP	PRCP	0.85673	CM	3
4	CV11_4	CV11	P42785	PRCP	PRCP	1.63151	CM	4
5	CV99_5	CV99	P42785	PRCP	PRCP	0.64489	CM	5
6	C193_6	C193	P42785	PRCP	PRCP	0.65128	CM	6
7	C33_7	C33	P42785	PRCP	PRCP	0.69374	CM	7
8	CV99_8	CV99	P42785	PRCP	PRCP	1.09712	CM	8
9	CV6_9	CV6	P42785	PRCP	PRCP	1.8895	CM	9
10	CV38_10	CV38	P42785	PRCP	PRCP	1.72346	CM	10
11	CV4_11	CV4	P42785	PRCP	PRCP	0.69872	CM	11
12	C195_12	C195	P42785	PRCP	PRCP	1.07357	CM	12
13	CV1_13	CV1	P42785	PRCP	PRCP	0.61055	CM	13
14	C237_14	C237	P42785	PRCP	PRCP	1.19429	CM	14
15	CV2_15	CV2	P42785	PRCP	PRCP	0.32423	CM	15
16	C189_16	C189	P42785	PRCP	PRCP	0.94542	CM	16
17	CV42_17	CV42	P42785	PRCP	PRCP	1.97023	CM	17
18	CV79_18	CV79	P42785	PRCP	PRCP	1.38031	CM	18
19	CV3_19	CV3	P42785	PRCP	PRCP	0.28226	CM	19
20	CV36_20	CV36	P42785	PRCP	PRCP	1.14129	CM	20
21	CV55_21	CV55	P42785	PRCP	PRCP	1.36112	CM	21
22	CV46_22	CV46	P42785	PRCP	PRCP	1.18126	CM	22
23	CV30_23	CV30	P42785	PRCP	PRCP	0.72615	CM	23

Gisby J, Clarke CL, Medjeral-Thomas N, Malik TH, Papadaki A, Mortimer PM, Buang NB, Lewis S, Pereira M, Toulza F, Fagnano E, Mawhin M, Dutton EE, Tapeng L, Kirk P, Behmoaras J, Sandhu E, McAdoo SP, Prendecki MF, Pickering MC, Botto M, Willicombe W, Thomas DC, Peters JE (2020) **Dryad Digital Repository** Longitudinal proteomic profiling of high-risk patients with COVID-19 reveals markers of severity and predictors of fatal disease.

<https://doi.org/10.5061/dryad.6t1g1jwxj>



eLife RESEARCH ARTICLE

Longitudinal proteomic profiling of dialysis patients with COVID-19 reveals markers of severity and predictors of death

Jack Gisby^{1*}, Candice L Clarke^{1,2†}, Nicholas Medjeral-Thomas^{1,2†}, Talat H Malik¹, Artemis Papadaki¹, Paige M Mortimer¹, Norzawani B Buang¹, Shanice Lewis¹, Marie Pereira¹, Frederic Toulza¹, Ester Fagnano¹, Marie-Anne Mawhin¹, Emma E Dutton¹, Lunnathaya Tapeng¹, Arienne C Richard^{1,3}, Paul DW Kirk^{4,5}, Jacques Behmoaras¹, Eleanor Sandhu^{1,4}, Stephen P McAdoo^{1,5}, Maria P Prendecki^{1,5}, Matthew C Pickering¹, Marina Botto¹, Michelle Willicombe^{1,2†}, David C Thomas^{1,2†}, James E Peters^{1,2†}

¹Centre for Inflammatory Disease, Department of Immunology and Inflammation, Imperial College London, London, United Kingdom; ²Renal and Transplant Centre, Hammersmith Hospital, Imperial College Healthcare NHS Trust, London, United Kingdom; ³Cambridge Institute for Medical Research, University of Cambridge, Cambridge, United Kingdom; ⁴CRUK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom; ⁵MRC Biostatistics Unit, Forvie Way, University of Cambridge, Cambridge, United Kingdom; ⁶Cambridge Institute of Therapeutic Immunology & Infectious Disease, University of Cambridge, Cambridge, United Kingdom; ⁷Health Data Research UK, London, United Kingdom

Abstract End-stage kidney disease (ESKD) patients are at high risk of severe COVID-19. We measured 434 circulating proteins in serial blood samples from hospitalised non-hospitalised ESKD patients with COVID-19 (n = 236 samples from 55 patients). Comparison to 51 non-infected patients revealed 221 differentially expressed proteins, with consistent results in a separate cohort of 40 COVID-19 patients. Two hundred and three proteins were associated with clinical severity, including IL6, markers of monocyte recruitment (e.g. CCL2, CCL7), neutrophil activation (e.g. proteinase-3), and epithelial injury (e.g. KRT19). Machine-learning identified predictors of severity including IL1BP, CTSD, GOLT3, and KRT19. Survival analysis with joint models revealed 69 predictors of death. Longitudinal modelling with linear mixed models uncovered 32 proteins displaying different temporal profiles in severe versus non-severe disease, including integrins and adhesion molecules. These data implicate epithelial damage, innate immune activation, and leucocyte-endothelial interactions in the pathology of severe COVID-19 and provide a resource for identifying drug targets.

Competing interest: See page 25

Funding: See page 25

Received: 12 November 2020

Accepted: 10 March 2021

Published: 11 March 2021

Reviewing editor: Evangelos J Giannakoulis-Bourboulis, National

Citations

Most commonly citations happen when researchers cite scholarly objects in articles.

Something to bear in mind in the context of journal articles is that citations can appear in different locations in the article.



Journal Article
Preprint
Report



Data



Software

Data availability

All data generated during this study are included in the manuscript and supporting files. Underlying source data for all analyses (individual-level proteomic and clinical phenotyping data) are available without restriction as Source Data Files 1-4. In addition, these data have been deposited in the Dryad Digital Repository (<https://doi.org/10.5061/dryad.6t1g1jwxj>). Code is available in the following GitHub repository: https://github.com/jackgisby/longitudinal_olink_proteomics copy archived at <https://archive.softwareheritage.org/swh:1:rev:32f08137859d44707ec4f086eed9af9b9ee91a87/>.

Materials and Methods

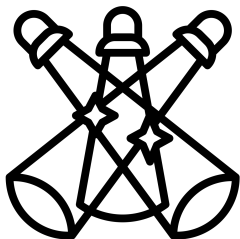
Library preparation by reverse complement PCR

The protocol described in this peer-reviewed article for SARS-CoV-2 amplicon library preparation and sequencing in wastewater RNA samples is published on protocols.io ([dx.doi.org/10.17504/protocols.io.81wgb7bx3vpk/v3](https://doi.org/10.17504/protocols.io.81wgb7bx3vpk/v3)) and is included for printing as [S1 File](#) with this article. In brief, wastewater nucleic acid samples are purified by 1.8x magnetic bead clean-up using Mag-Bind® TotalPure NGS beads (Omega Bio-tek) before cDNA is synthesised using the LunaScript® RT SuperMix Kit (New England Biolabs, UK). This protocol then utilises the EasySeq™ RC-PCR SARS CoV-2 (novel coronavirus) Whole Genome Sequencing Kit (NimaGen, The Netherlands) for library preparation, which generates SARS-CoV-2 amplicons

References

Abdill, R. J. and Blekhman, R. (2019). Meta-research: tracking the popularity and outcomes of all bioRxiv preprints. *ELife* 8, e45133. <https://doi.org/10.7554/eLife.45133>

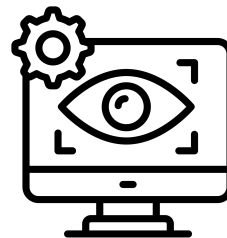
The benefits of Citations



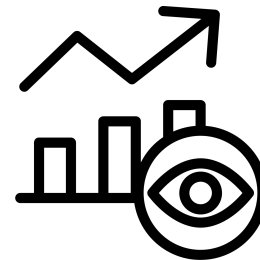
Acknowledgement for
the researchers who
produce the object



Transparency &
reproducibility for
those using the object



Visibility for the
repositories that
host research
objects



Evaluation of
open outputs

Data Citations: a role for everyone



There is a role for all actors in the research process in supporting best practices for data citation

Researchers

- Deposit dataset at repositories that assign DOIs
- Include citations to the datasets they have used

Funders

- Encourage researchers to cite datasets they use - their own and others' - in their research outputs
- Consider information on datasets shared and their reuse as part of evaluation processes

Data repositories

- Collect citation information for datasets
- Include those citations in the metadata deposited with DataCite

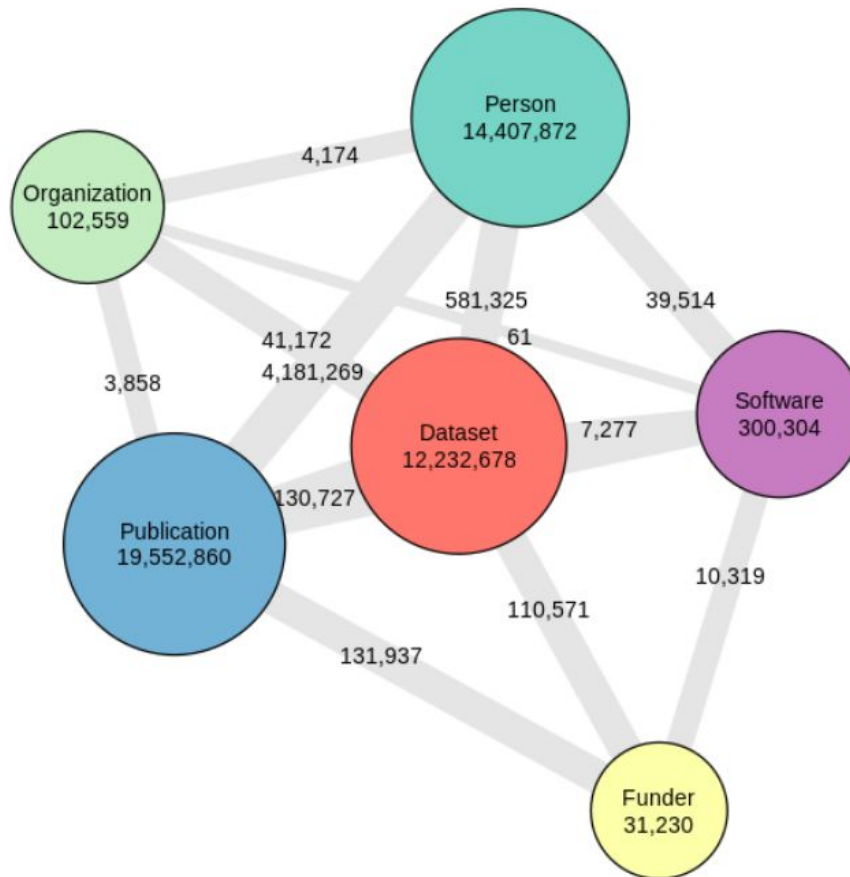
Publishers

- Ensure data citations are included in the article, in a machine-readable format
- Index citations with Crossref

Research Impact

The PID Graph

Number of nodes and connections
(August 2022)



DataCite Commons: Work



Add to ORCID Record

Download Metadata

Cite as

DataCite Metadata Working Group. (2021). *DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs v4.4*. <https://doi.org/10.14454/3W3Z-SA82>

APA

Share

Email

Twitter

Facebook

Filter Works

Type to search... X

Creators & Contributors ?

☐ Burger, Felix

4

☐ White, Andrew

1

☐ Barluzzi, Luciano

1

☐ Mies, Thomas

1

☐ Layfield, Richard

1

☐ Barrett, Anthony

1

☐ Pzena, Henry

1

DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs v4.4

<https://doi.org/10.14454/3w3z-sa82>

7 Citations

Description

Creators

Contributors

Registration

1 Introduction 1.1 The DataCite Consortium 1.2 DataCite Community Participation 1.3 The Metadata Schema 1.4 Version 4.4 Update 2 DataCite Metadata Properties 2.1 Overview 2.2 Citation 2.3 DataCite Properties 3 XML Example 4 XML Schema 5 Other DataCite Services Appendices Appendix 1: Controlled List Definitions Appendix 2: Earlier Version Update Notes Appendix 3: Standard values for unknown information Appendix 4: Version 4.1 Changes in support of software citation Appendix 5: FORCE11 Software Citation Principles Mapping

Version 4.4 of Documentation published 2021 in DataCite

Text

English

doi

<https://doi.org/10.14454/3w3z-sa82>

7 Citations

Publication Year

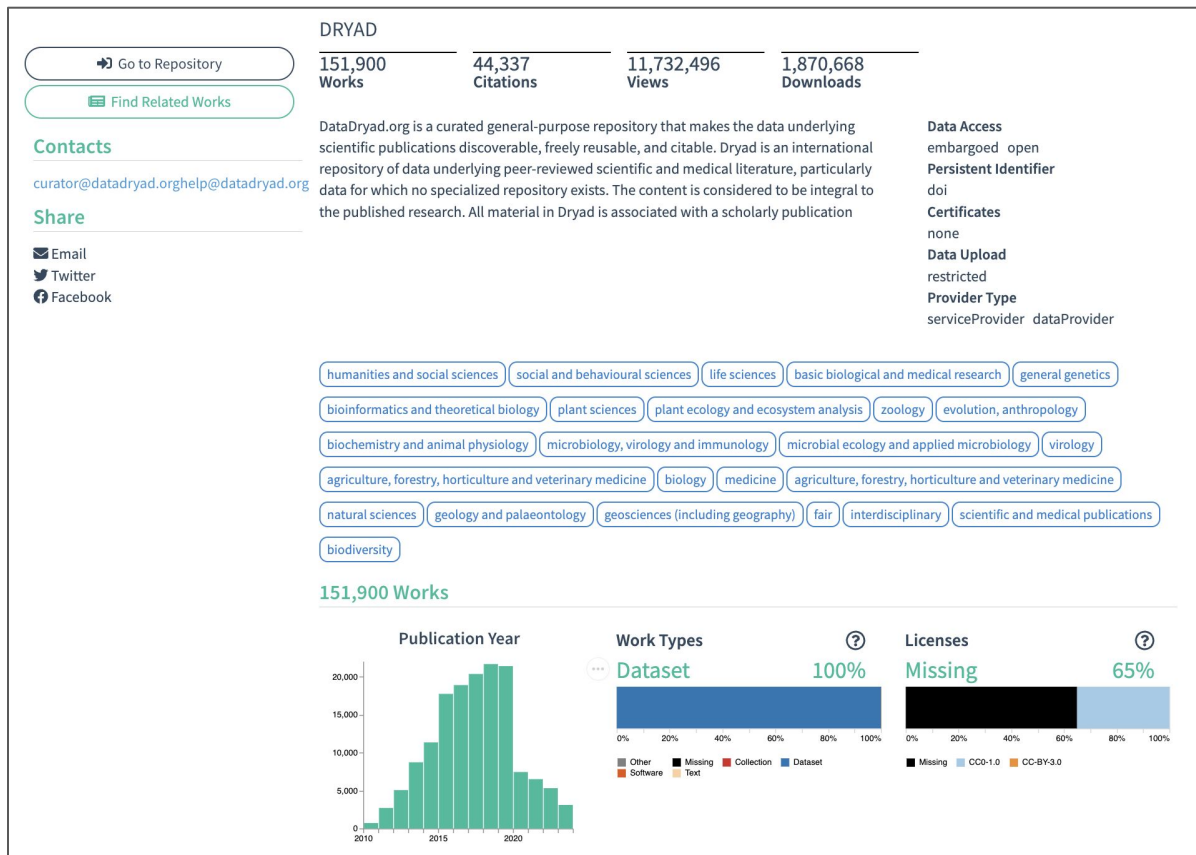
Work Types

Licenses

Best Practices mit dem DataCite-MetadatenSchema 4.4

<https://commons.datacite.org/doi.org/10.14454/3w3z-sa82>

DataCite Commons: Repository



<https://commons.datacite.org/repositories/nxrc8v>



CONNECTING RESEARCH,
ADVANCING KNOWLEDGE



info@datacite.org



pidforum.org



datacite.org
blog.datacite.org



support.datacite.org
support@datacite.org



[@datacite](https://twitter.com/datacite)



[DataCite](https://www.youtube.com/DataCite)



[@datacite](https://www.linkedin.com/company/datacite)