

Figshare, being FAIR and open data everywhere

Stephen Cawley, Head of Institutional Marketing, Digital Science

Figshare is part of the Digital Science Family

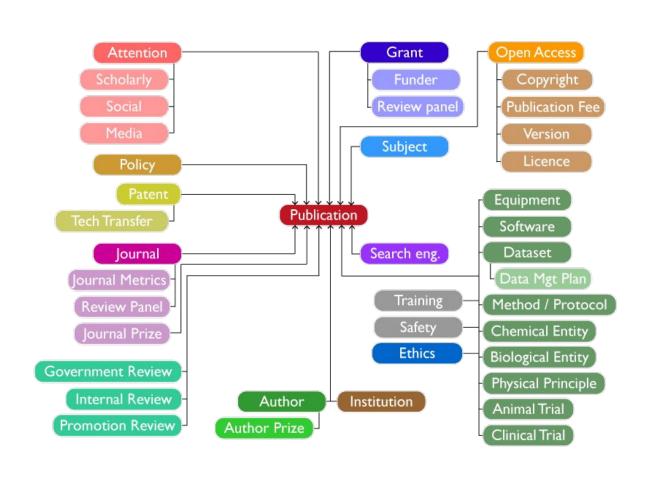
Disclaimer: this co-operative of companies is for profit but also strives for sustainable pricing and business models

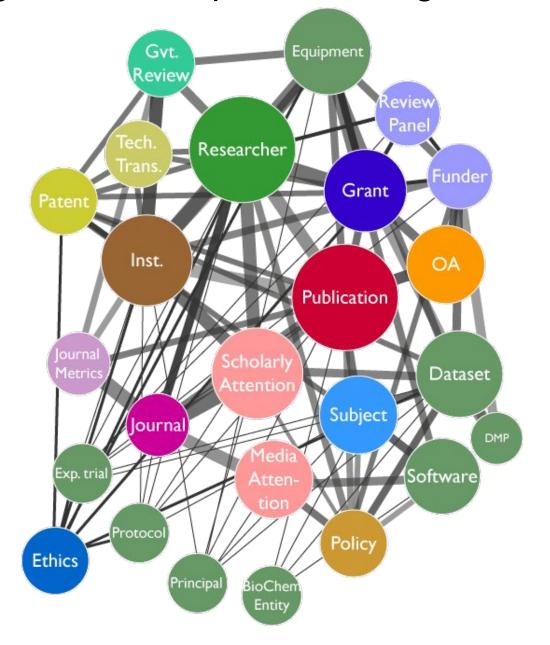
Freemium versions available for several of our software applications





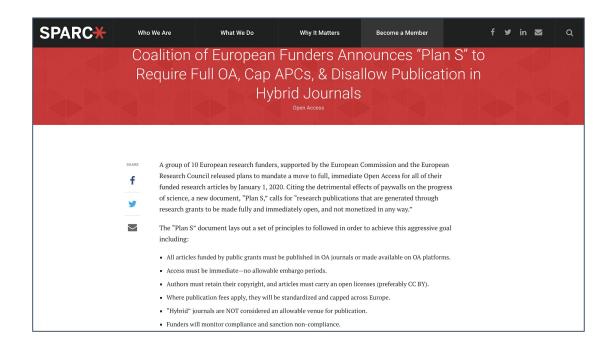
The expanding universe reflected in Digital Science product design





Importance of current and forward-looking data signals for OA compliance strategy

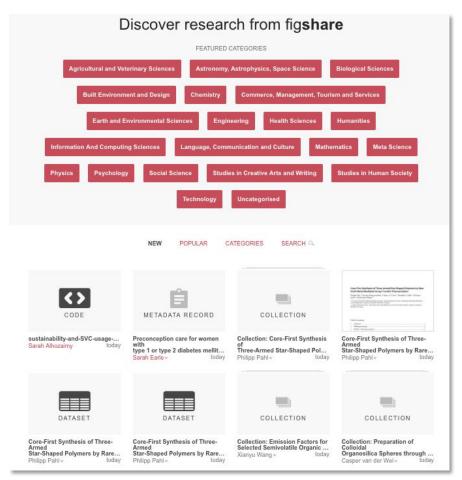
- 89 grants active between 2019 and 2023 with a total value of \$221.3m with an average value of \$4.3m
- 38 of these grants have been awarded by members of the Funder Group Coalition S
- Will this signal of active research projects help enable an OA compliance workflow that can be embedded into the active research workflow?
- We can provide the named investigator, institution, fiscal value in many cases, and the technical summary of the research project, and any resulting publications, patents or associated clinical trials





Search on the research activity of researchers affiliated with University of Luxembourg, Luxembourg Institute of Science and Technology; Luxembourg Institute of Socio-Economic Research, Luxembourg Institute of Health; and the Max Planck Institute Luxemburg for International, European and Regulatory Procedural Law

Figshare: store, share, discover research



Who we are and what we do:

Figshare is an online platform where researchers can preserve and share their digital research content in any file format.

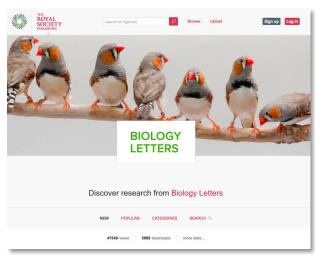
Some features of published content:

- Open licenses for reuse
- DOIs
- Versioning of outputs
- Data availability

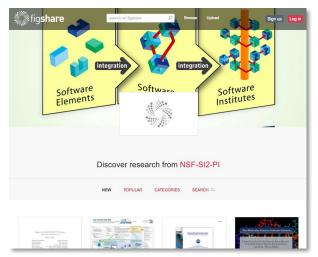




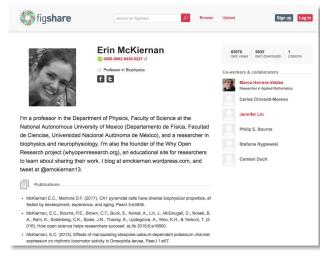
Communities we support and the services we provide



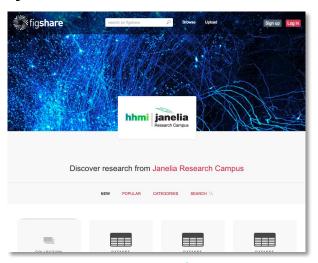
Academic publishers (portal/viewer/preprints!)



Conferences/proceedings



Individual researchers

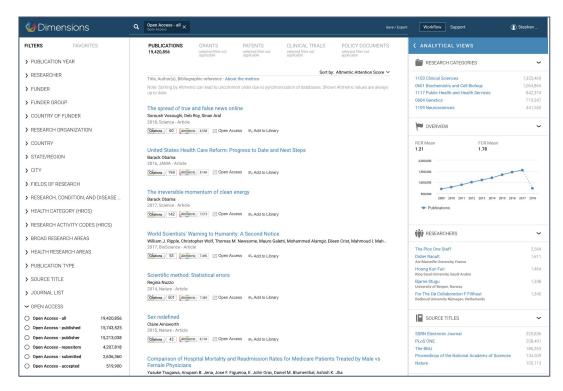


Institutions (research unis/labs)



Foundations and gov'ts (funders)

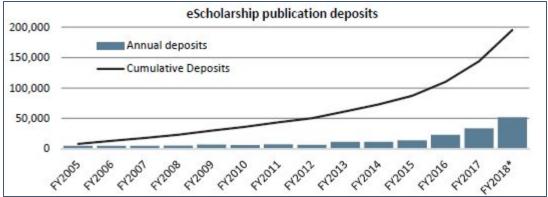
Digital Science is firmly committed to enabling Open Access to Publications and Data

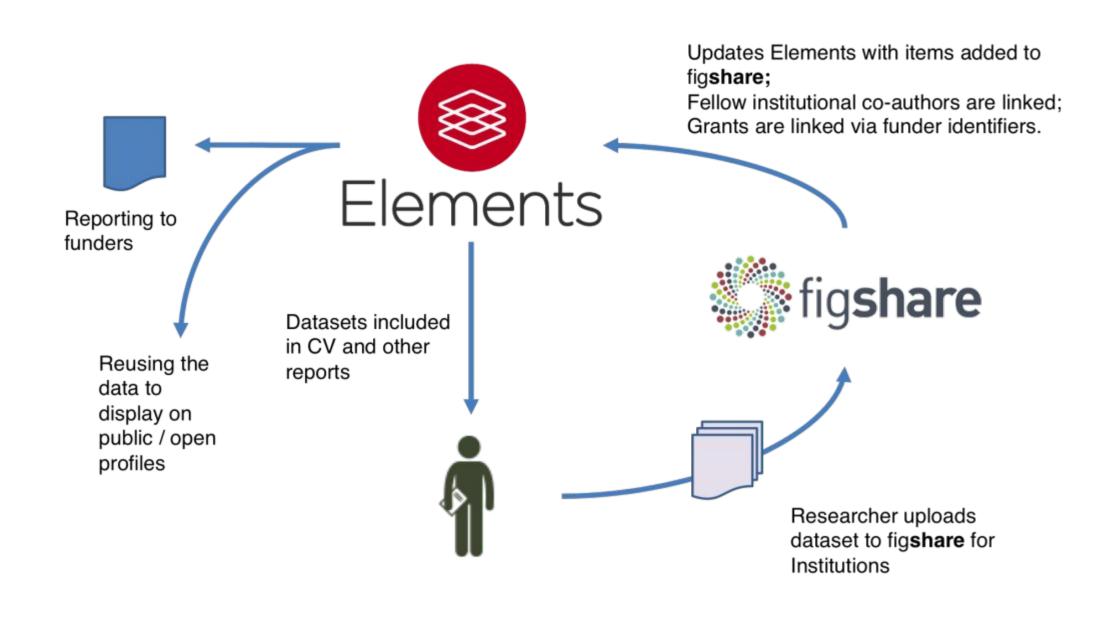


Dimensions OA Filter - 19.4m OA publications discoverable in a free Discovery tool @ Dimensions.ai

Elements OA Monitor Implemented in January 2015
on E-scholarship IR as a
depositing interface >>

















Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the <u>FAIRification process</u>.

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

Accessible

Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- <u>I1.</u> (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- 12. (Meta)data use vocabularies that follow FAIR principles
- 13. (Meta)data include qualified references to other (meta)data

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- R1. Meta(data) are richly described with a plurality of accurate and relevant attributes
- R1.1. (Meta)data are released with a clear and accessible data usage license
- R1.2. (Meta)data are associated with detailed provenance
- R1.3. (Meta)data meet domain-relevant community standards

2007: E-science and data-intensive discovery "The Fourth Paradigm"



The Fourth Paradigm: Data-Intensive Scientific Discovery

By KRISTIN M. TOLLE

D. STEWART W. TANSLEY

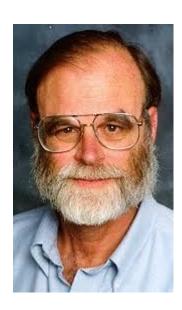
ANTHONY J. G. HEY

External Research, Microsoft Research, Redmond, VA 98052 USA



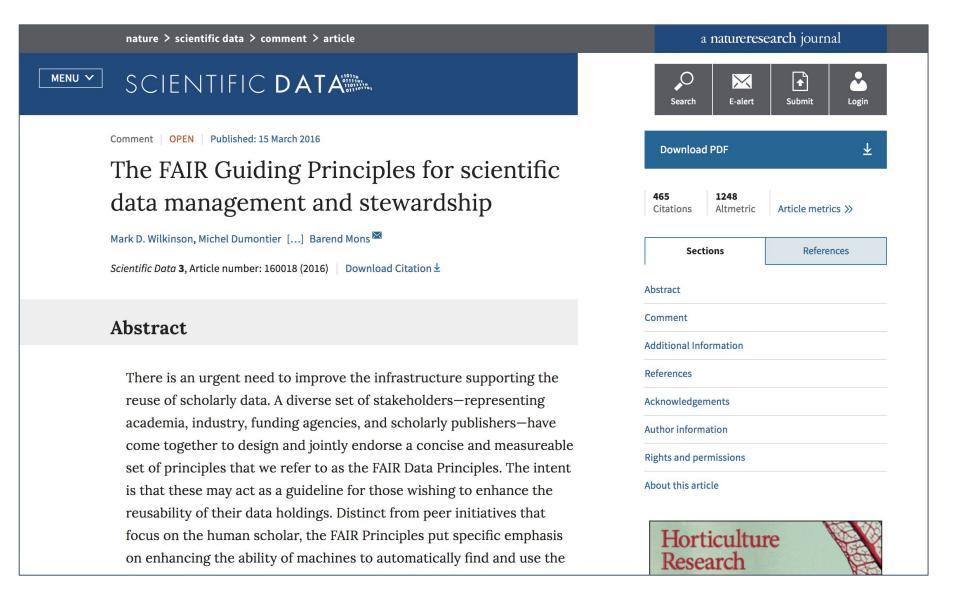
he book The Fourth Paradigm: Data Intensive Scientific Discovery, contains a series of essays by scientists and computer scientists looking forward five years or more to how different scientific fields are being transformed by the exponential increase in scientific

abstracts being deposited in the Medline database corresponds to approximately 1100 papers every day; over 400 000 per year.2 As Faniel and Zimmerman point out, in many fields, scientists have numerous challenges gaining access to the original data to either check the claims of a scientific paper or to combine that data with other data for further analysis.3 Smit goes on to suggest that data and the publications should be "wedded."4 We are now seeing governments and funding agencies looking at ways to increase the value and pace of scientific research through increased or open access to both data and publications. In this point of view article, we wish to look at another aspect of these twin revolutions, namely, how to enable developers, designers and researchers to build intuitive, multimodal, user-centric, scientific applications that can aid and enable scientific research—essentially a

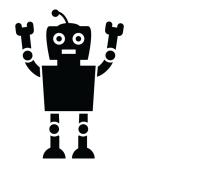


- 1. Empirical Observation and experimentation
- 2. Analytical or theoretical approaches
- 3. Computational science or simulation
-a new method of pushing forward the frontiers of knowledge, enabled by new technologies for gathering, manipulating, analyzing and displaying data

2014 Force 11 task group formed



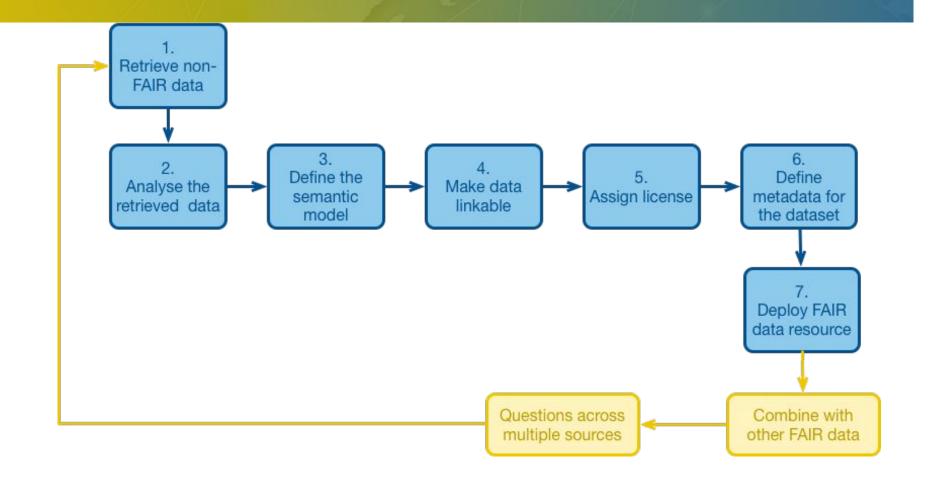




FAIR



F/IR

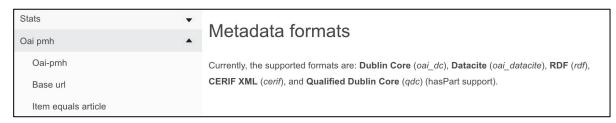


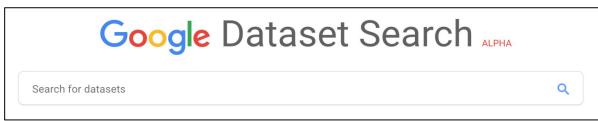
Figshare's beliefs align to FAIR principles

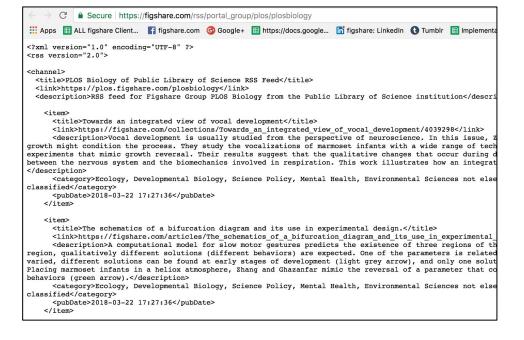
- Academic research outputs should be as open as possible, as closed as necessary
- Academic research outputs should never be behind a paywall
- Academic research outputs should be human and machine readable/query-able
- Academic infrastructure should be interchangeable
- Academic researchers should never have to put the same information into multiple systems at the same institution
- Identifiers for everything
- The impact of research is independent of where it is published and what type of output it is

Support for open standards and de facto standards

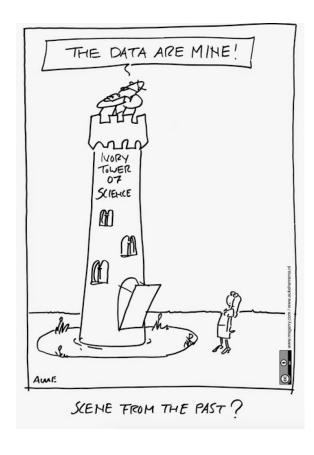
- Dublincore, Datacite, RDF, Cerif XML and Qualified Dublin Core
- All published data items indexed in Google Data Search
- All content marked up with Google Dataset search schema



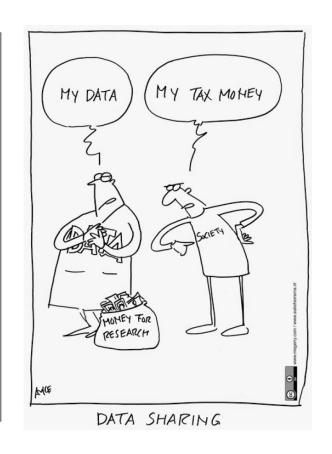






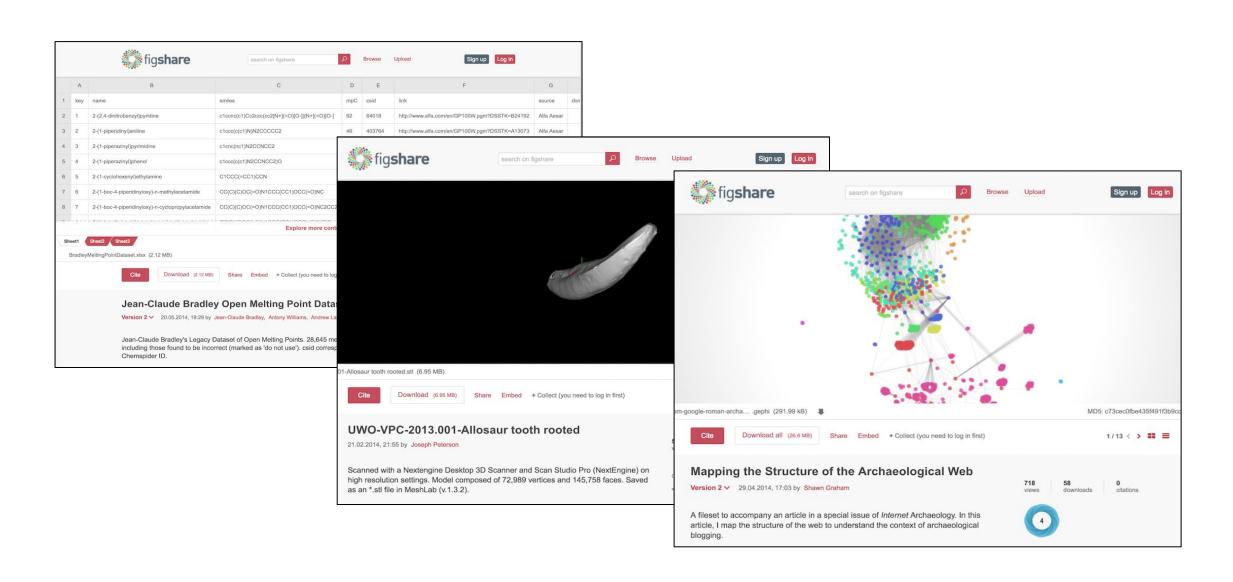






Author: Auke Herrema – Het Bouwteam

Impress the researcher with in-browser visualisers for than 1000 file types



Journal policies around preprints (to accelerate scholarly communication) and publishing data (to support article claims)



cited in the references if they have a digital object identif

made available to qualified researchers for examination

(2003); and Ensuring the Integrity, Accessibility, and Ster

Data in a Digital Age (2009).

[Previous | Top]

References for citation information.) Such deposition ma (i) Articles are considered provided they have not been Published Previously or during the review process and postpublication. In rare ca concurrently submitted for publication elsewhere. What constitutes prior specific repositories are not available, authors may use f publication must take into account many criteria, including the extent of review, or other rare specimens must be deposited in a museum and will be determined on a case-by-case basis. Related manuscripts that are in press or submitted elsewhere must be included with a PNAS submission. about accessibility of data and materials, see the following

Related Data and Materials: Responsibilities of Authorsh Figures, tables, or videos that have been published elsewhere must be identified, and permission of the copyright holder for both the online and print editions of the journal must be provided (see www.pnas.org/site/misc/permissions_letter.pdf).

> (ii) Posting to Preprint Servers, such as arXiv or bioRxiv, is permitted. See the PNAS statement on prior publication for details, and see section vii for media embargo policies.

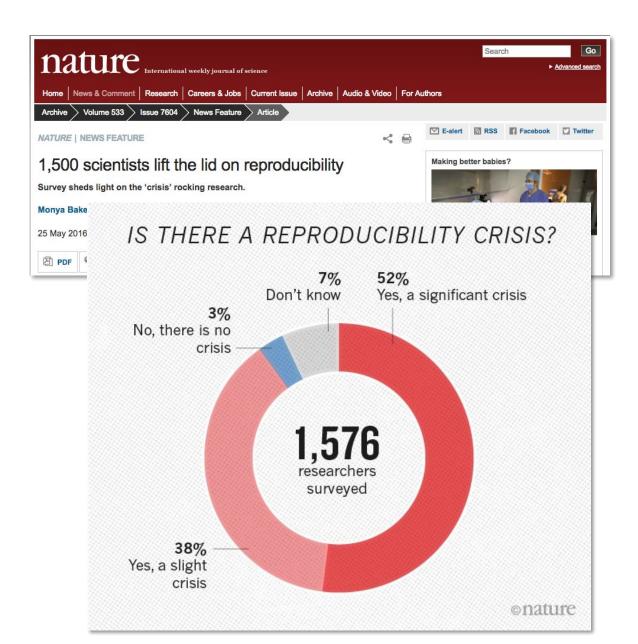
A growing number of journal publishers allow not only the posting of preprints to relevant preprint servers, but also recommend making supporting data and materials available. For example, PNAS states:

To allow others to replicate and build on work published in PNAS, authors must make materials, data, and associated protocols, including code and scripts, available to readers.



Reproducibility, Replication, and Reuse

Selective reporting is a key culprit in the reproducibility crisis. Publishing all of your results so that researchers can build off of your discoveries and research promotes collaboration, efficiency, and reuse

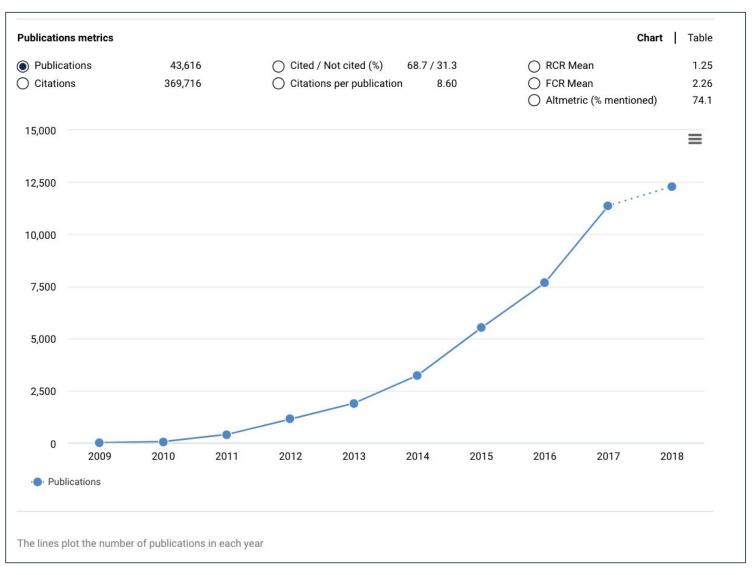


Data citations on the rise providing a new form of credit for researchers











United States Commutes and Megaregions data for GIS

Version 5

✓ Fileset posted on 31.01.2017, 10:01 by Alasdair Rae, Garrett G.D. Nelson

This Figshare dataset contains the files created by and used in a related PLOS ONE paper, entitled 'An economic geography of the United States: from commutes to megaregions', by Garrett Dash Nelson and Alasdair Rae, published 30 November 2016.

Update: 27 January 2017 - see item 7. below

In addition to the files listed below, we have also provided a series of maps here, as high resolution PNGs. The fifth file below can be styled in QGIS using the QML style file provided in number 6.

50068 views 8984 downloads

citations

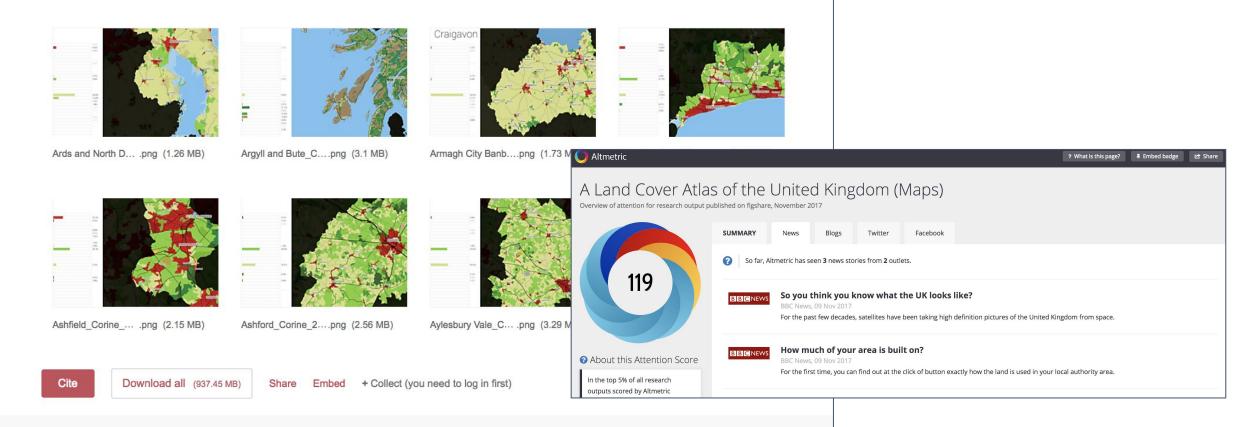




CATEGORIES

- Urban and Regional Economics
- Urban and Regional Planning not elsewhere classified
- Urban and Regional Studies (excl. Planning)





A Land Cover Atlas of the United Kingdom (Maps)

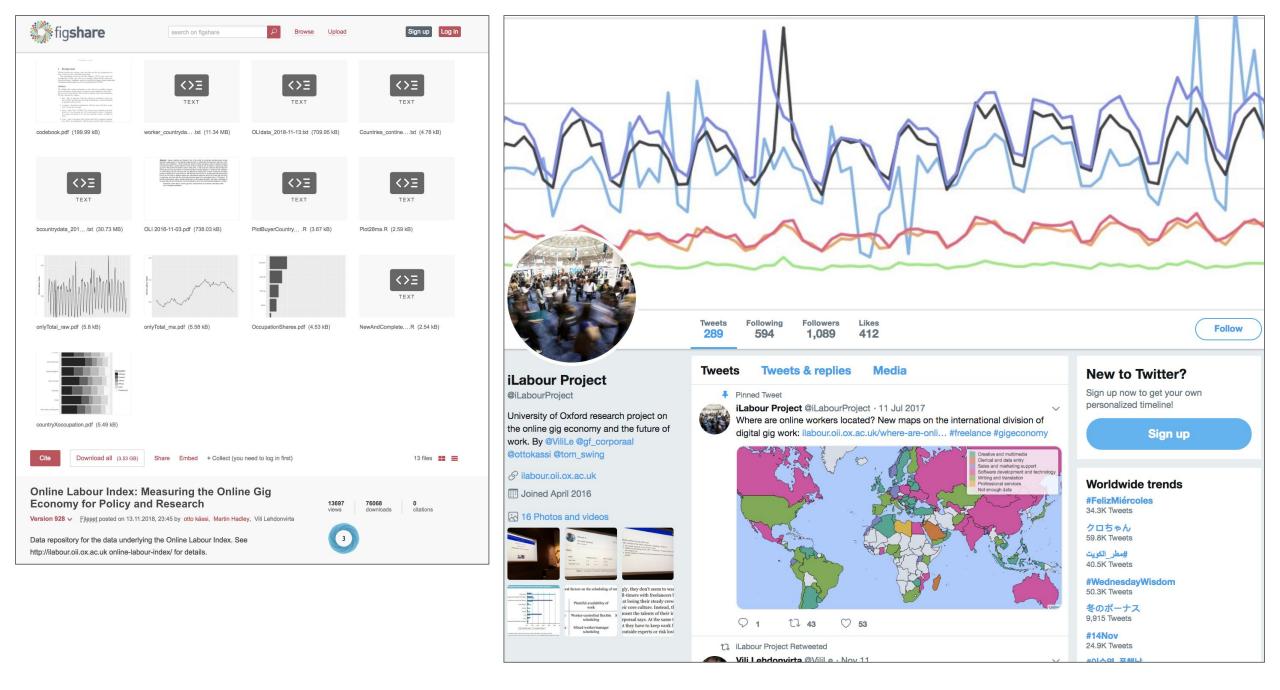
Fileset posted on 03.11.2017, 16:42 by Alasdair Rae

This set of maps accompanies my related publication, entitled 'A Land Cover Atlas of the United Kingdom', https://doi.org/10.15131/shef.data.5266495, which was published at the same time.

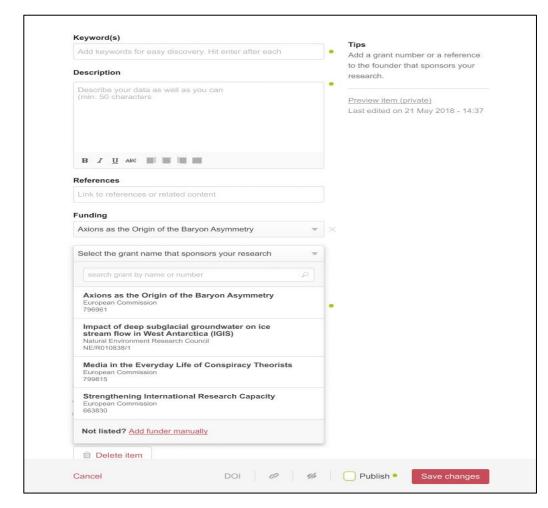
Some of the maps in this set feature in the Atlas, but at a lower resolution. I have deposited them here as high resolution images (300dpi PNG files) so that interested users can access and download them.

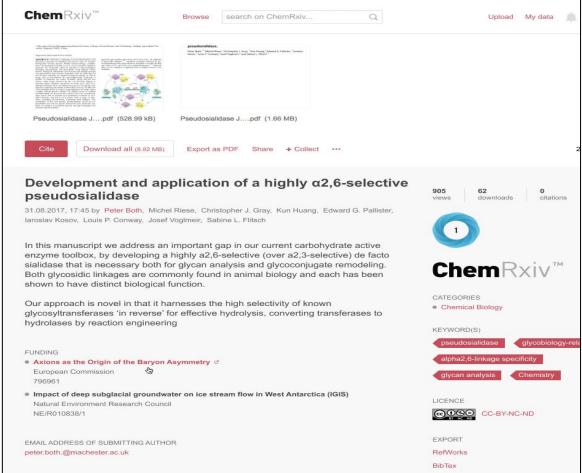






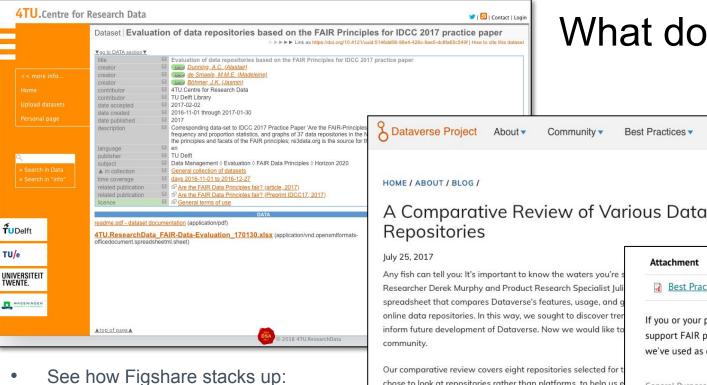
Making compliance easier







FAIR and other relevant reviews



What does a good repo look like?

Contact

f 🏏 🔤 in G+

July 2018 (1)

structured text, software,

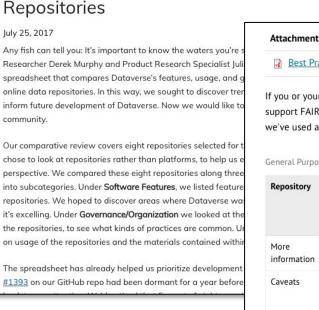
source code, other

will assign a DOI

identifiers

BLOG POSTS BY MONTH

- https://data.4tu.nl/repository/uuid: 5146dd06-98e4-426c-9ae5-dc8fa 65c549f
- https://doi.org/10.7910/DVN/WS9 **OUR**
- https://library.si.edu/research/dat a-repositories



About ▼

Community v

Best Practices v

591.22 KB Best Practices for Choosing a Repository If you or your publisher prefer to deposit with a non-SI repository, there are four general-purpose repositories that support FAIR principles. Their features are compared below. Following the grid is a glossary that clarifies the terms we've used as comparison criteria. General Purpose Data Repositories Compared Repository Figshare Zenodo Dryad Open Science Framework (OSF) More @ Re3 / More @ Re3 / More @ More @ Re3 / More information Re3 / Caveats Dryad's CC-0 license is at odds Figshare has a 5GB per file OSF is best Zenodo is based in Europe, and with SI's general Terms of Use, European laws may apply to data which is closer in spirit to CCactive deposited. There is a 50GB per BY-NC. projects. dataset limit Fees (2018) \$120 per deposit (SI is not a free, premium service for a fee member, and cannot get a discount.) **Formats** office documents, scientific & office documents, images, any (no any (no restrictions on file types) accepted statistical data, plain text, structured graphics, restrictions

audiovisual data, raw data,

supports ORCID, will assign

DOI at time of publication

plain text, archived data

on file

types)

ORCID, will

assign ARK

Size

supports ORCID, will assign DOI or

use provided DOI

A Data Citation Roadmap for Scholarly Data Repositories

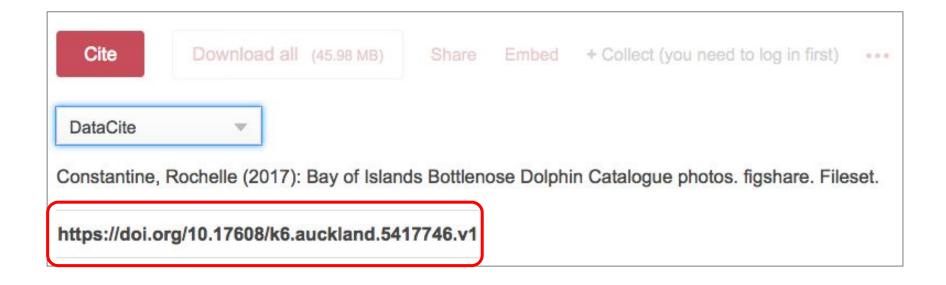
Level	#	Guideline
Required	1	All datasets intended for citation <i>must</i> have a globally unique persistent identifier that can be expressed as unambiguous URL.
	2	Persistent identifiers for datasets <i>must</i> support multiple levels of granularity, where appropriate.
	3	This persistent identifier expressed as URL <i>must</i> resolve to a landing page specific for that dataset.
	4	The persistent identifier <i>must</i> be embedded in the landing page in machine-readable format.
	5	The repository must provide documentation and support for data citation.
Recommended	6	The landing page should include metadata required for citation, and ideally also metadata helping with discovery, in human-readable and machine-readable format.
	7	The machine-readable metadata should use schema.org markup in JSON-LD format.
	8	Metadata should be made available via HTML meta tags to facilitate use by reference managers.
	9	Metadata should be made available for download in Bibtex and/or another standard bibliographic format.

A Data Citation Roadmap for Scholarly Data Repositories

Martin Fenner, Mercè Crosas, Jeffrey Grethe, David Kennedy, Henning Hermjakob, Philippe Rocca-Serra, Gustavo Durand, Robin Berjon, Sebastian Karcher, Maryann Martone, Timothy Clark bioRxiv 097196; doi: https://doi.org/10.1101/097196



F1 (meta)data are assigned a globally unique and persistent identifier



F3 (meta)data are registered or indexed in a searchable resource

Top referrals			
Source	Total views		
1. figshare.com	261		
2. auckland.figshare	40		
3. www.informas.org	32		
4. toolbox.google.com	20		
5. www.google.com	15		
6. presto.auckland.ac	10		
7. search.datacite.org	10		
8. www.google.co.nz	9		
9. xueshu.baidu.com	4		
10. unidirectory.auckl	4		

F4 metadata specify the data identifier



Transforming the Monash University Research Ecosystem

Monash University

Andrew Harrison (Aggregated by) Beth Pearson (Aggregated by) David Groenewegen (Aggregated by) David Groenewegen (Aggregated by) Neil Dickson (Aggregated by)

http://doi.org/10.4225/03/5975999c33419

Access data via landing page http://doi.org/10.4225...

Pearson, Beth; Groenewegen, David; Dickson, Neil; Harrison, Andrew; Splawa-Neyman, (2017): Transforming the Monash University Research Ecosystem. figshare. Figure.

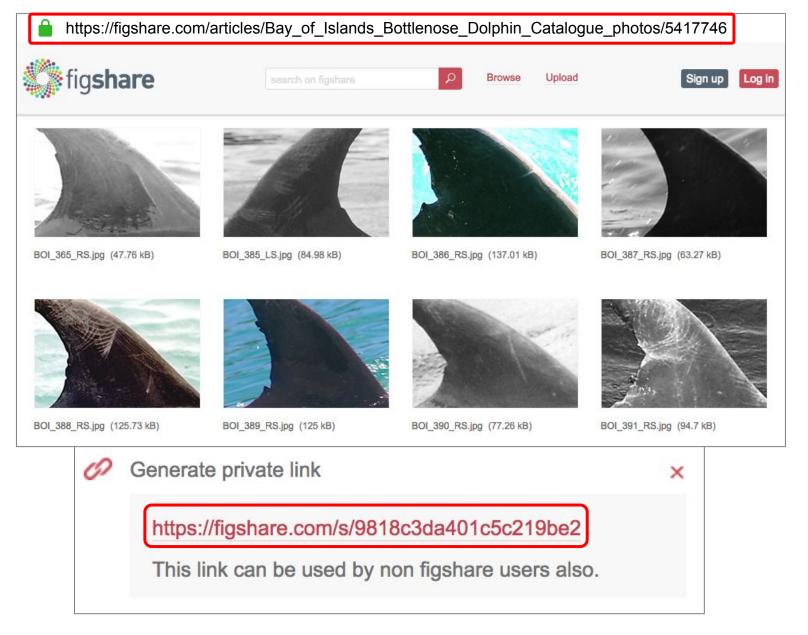
https://doi.org/10.4225/03/5975999c33419

Select your citation style and then place your mouse over the citation text to select it.

Transforming the Monash University Research Ecosystem

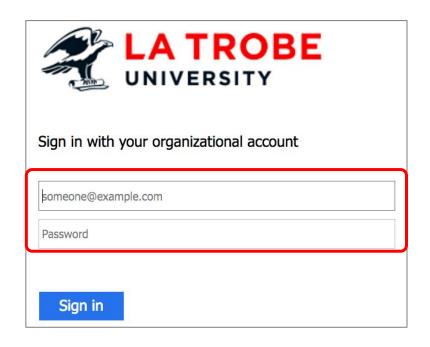
Figure posted on 24.07.2017, 16:54 by Beth Pearson, David Groenewegen, Neil Dickson, Andrew Harrison, Patrick Splawa-Neyman

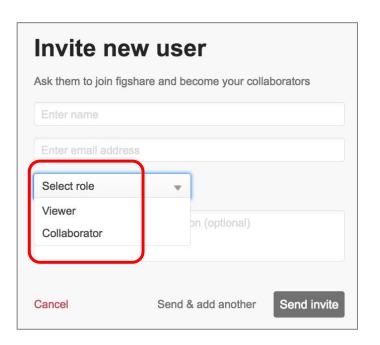
A1 (meta)data are retrievable using a standardized communications protocol

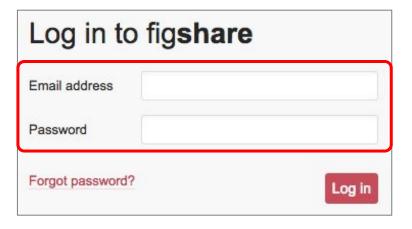


A1.1 the protocol is open, free, and universally implementable

A1.2 the protocol allows for an authentication and authorization procedure







A2 metadata are accessible, even when the data are no longer available













I1 (meta)data use a broadly applicable language for knowledge representation

```
<dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/"</pre>
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:title>Bay of Islands Bottlenose Dolphin Catalogue photos</dc:title>
<dc:creator>Rochelle Constantine</dc:creator>
<dc:identifier>10.17608/k6.auckland.5417746.v1</dc:identifier>
<dc:relation>https://figshare.com/articles/Bay_of_Islands_Bottlenose_Dolp
hin_Catalogue_photos/5417746</dc:relation>
<dc:description>This catalogue is a collection of uniquely identified
bottlenose dolphin dorsal fin photos from the Bay of Islands, New Zealand
1993-2013. Where possible there is a right and left side image of each
dorsal fin.<div&gt;&lt;br&gt;&lt;/div&gt;&lt;div&gt;The photograph
codes are as follows: < br&gt; &lt; div&gt; BOI = Bay of
Islands</div&gt;&lt;div&gt;XXX = unique number for each
dolphin&lt:/div&gt:&lt:div&gt:RS/LS = right side/ left
side<br&gt;&lt;div&gt;&lt;br&gt;&lt;/div&gt;&lt;div&gt;All work was
conducted by researchers at the University of Auckland and the catalogue
is curated by Dr Rochelle Constantine, University of Auckland
(r.constantine@auckland.ac.nz). Please contact Rochelle if you have any
queries.</div&gt;&lt;/div&gt;&lt;/div&gt;</dc:description>
<dc:date>2017-11-12 22:04:08</dc:date>
<dc:subject>bottlenose dolphin</dc:subject>
```

I2 (meta)data use vocabularies that follow FAIR principles

```
010101 Algebra and Number Theory
```

010102 Algebraic and Differential Geometry

010103 Category Theory, K Theory, Homological Algebra

010104 Combinatorics and Discrete Mathematics (excl. Physical Combinatorics)

010105 Group Theory and Generalisations

010106 Lie Groups, Harmonic and Fourier Analysis

010107 Mathematical Logic, Set Theory, Lattices and Universal Algebra

010108 Operator Algebras and Functional Analysis

010109 Ordinary Differential Equations, Difference Equations and Dynamical Systems

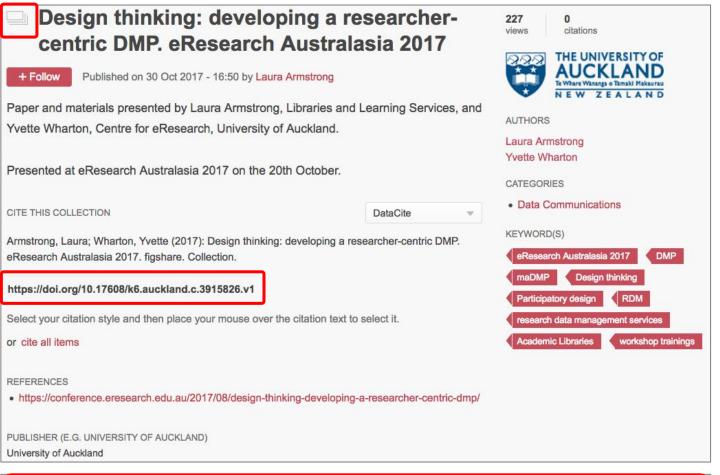
010110 Partial Differential Equations

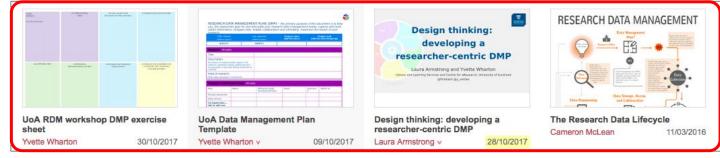
010111 Real and Complex Functions (incl. Several Variables)

010112 Topology

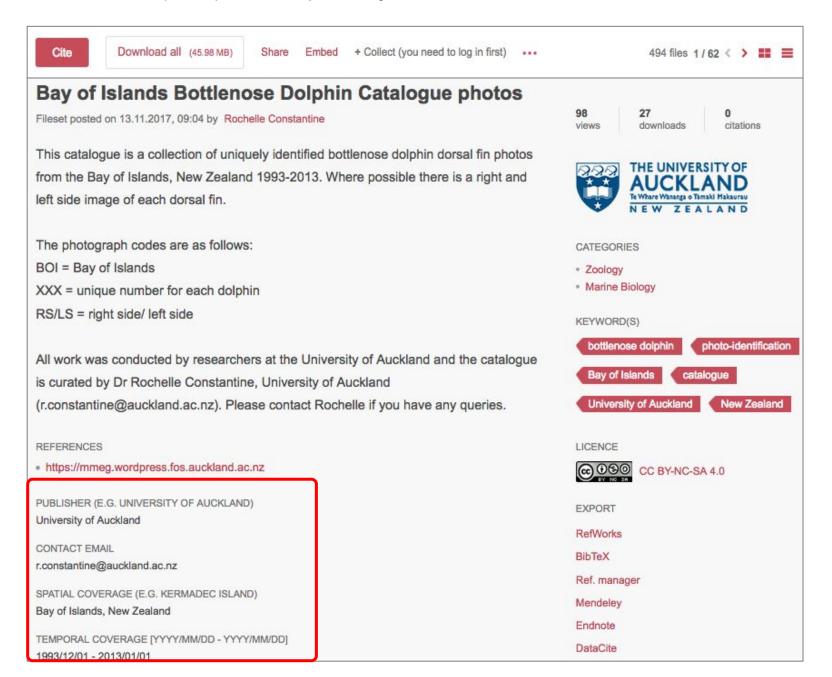
010199 Pure Mathematics not elsewhere classified

13 (meta)data include qualified references to other (meta)data

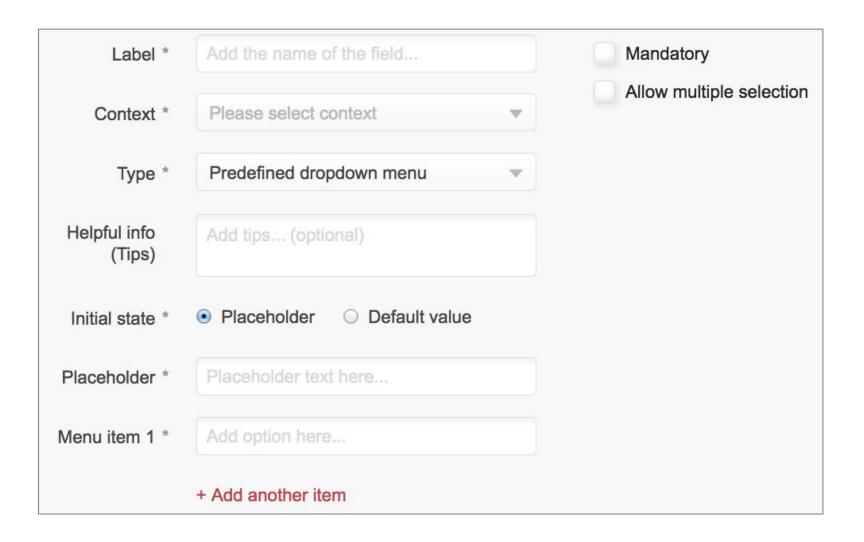




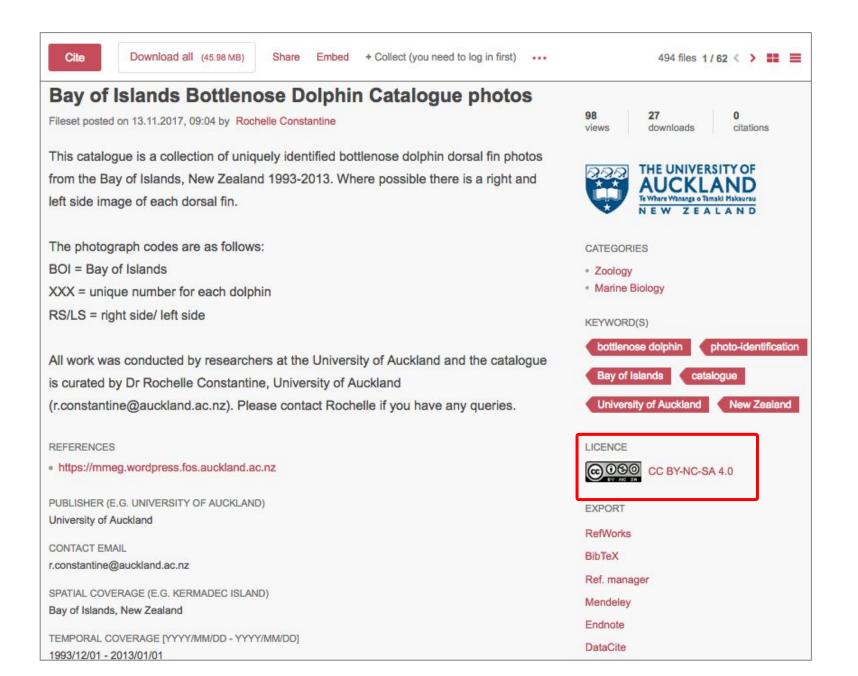
R1 meta(data) have a plurality of accurate and relevant attributes



R1 meta(data) have a plurality of accurate and relevant attributes



R1.1 (meta)data are released with a clear and accessible data usage license



R1.1 (meta)data are released with a clear and accessible data usage license

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:



Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



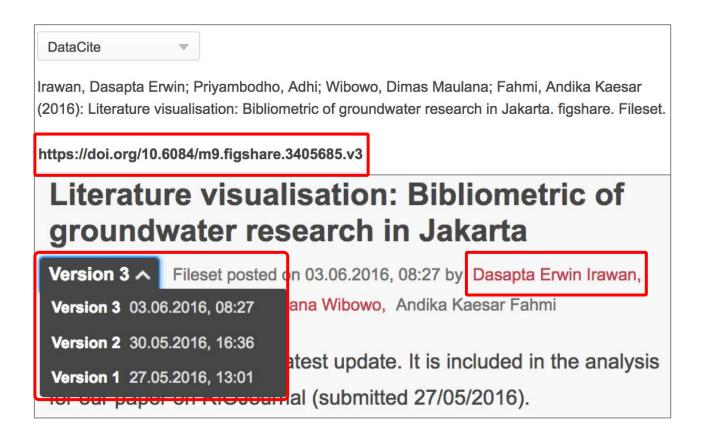
NonCommercial — You may not use the material for commercial purposes.



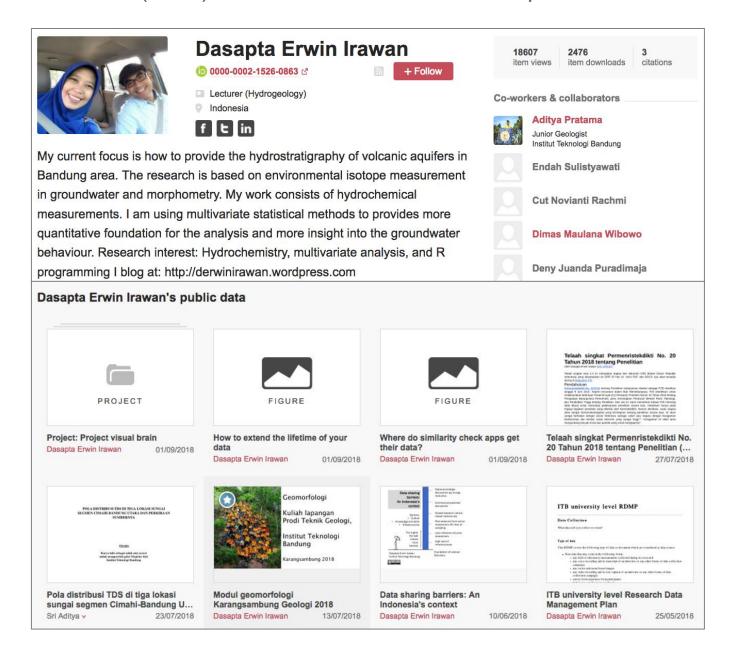
ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the <u>same license</u> as the original.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

R1.2 (meta)data are associated with their provenance



R1.2 (meta)data are associated with their provenance



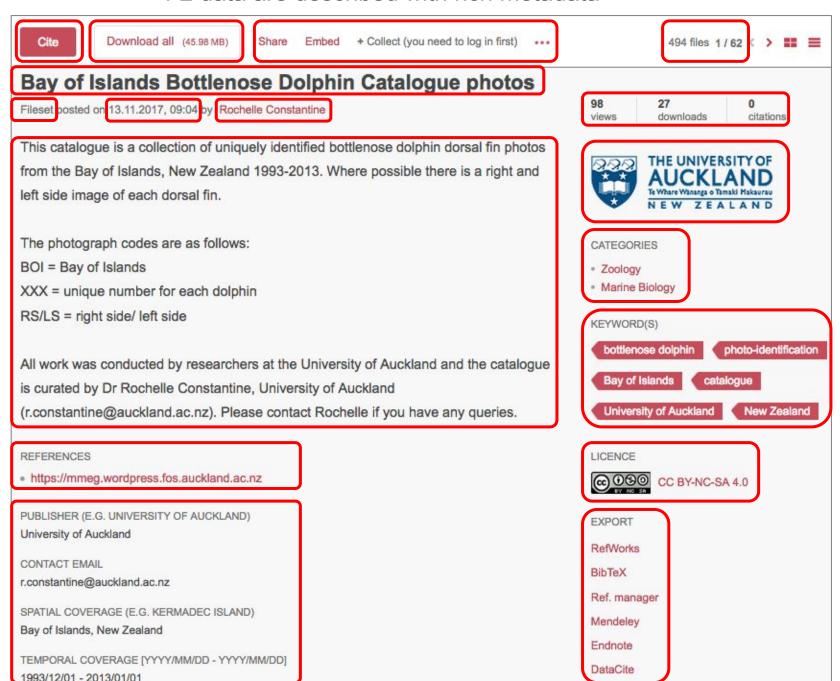


Ongoing Improvements aligned with FAIR

Facted and full text search will be added

astronomical instruments papers	
Authors Max Maxeltowner (123) J. H. J. Dzubiella (121) Vahid Etminan Farooji (76) George Adam (56) P. D. Smith (32)	Tuning the Collapse Transition of Weakly Charged Polymers by Ion-Specific Screening and Adsorption of elements Preprint submitted on 10.08.2018, 14:53 and posted on 10.08.2018, 17:46 by Richard Chudoba, Jan Heyda Joachim Dzubiella, Max Maxeltowner
more authors v	Collection: Predicting bed shear stresses in vegetated channels Published on 12 Aug 2018 - 14:31 by Vahid Etminan Farooji, P. D. Smith, G. Adam
Date published to Journal or Publication PLOS One (123) Nature (121) Cosmic Journal (76) The Journal for Astronomical Science and Data (56)	Understanding the Role of AgNO3 Concentration and Seed Morphology to Achieve Tunable Shape Control in Gold Nanostarsem ipsum dolor sit amet, consectetur adipiscing elit. Nullam malesuada magna vitae tortor lacinia pulvinar. Proin quam justo, mollis ut interdum Preprint submitted on 09.08.2018, 18:25 and posted on 10.08.2018, 15:06 by Supriya Atta Collection: Object personification in autism: This paper is very sad Published on 11 Aug 2018 - 15:00 Vahid Etminan Farooji, P. D. Smith, G. Adam
A Z (32)	Project: Systems thinking and physical education: Crossing borders Published on 11 Aug 2018 - 08:27 by Supriya Atta, Laura Fabris
Code Search codes	Tuning the Collapse Transition of Weakly Charged Polymers by Ion-Specific Screening and Adsorption of elements Preprint submitted on 10.08.2018, 14:53 and posted on 10.08.2018, 17:46 by Richard Chudoba, Jan Heyda Joachim Dzubiella, Max Maxeltowner

F2 data are described with rich metadata



R1.3 (meta)data meet domain-relevant community standards

ID	Property	Obligation
1	Identifier (with mandatory type sub-property)	М
2	Creator (with optional given name, family name, name identifier and affiliation sub-properties)	М
3	Title (with optional type sub-properties)	М
4	Publisher	М
5	PublicationYear	М
10	ResourceType (with mandatory general type description sub- property)	М



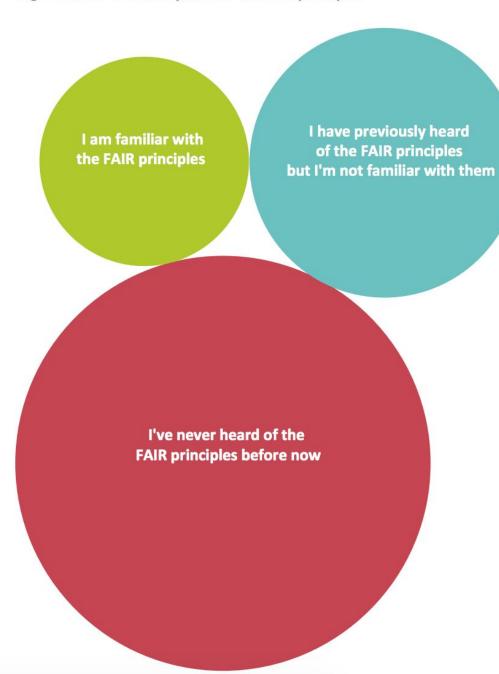


https://doi.org/10.6084/m9.figshare.7195058.v1

Key results:

- 64% of respondents revealed they made their data openly available in 2018, a 7% rise on 2016.
- Data citations are motivating more respondents to make data openly available, increasing 7% from 2017 to 46%.
- The percentage of respondents in support of national mandates for open data is higher at 63% than in 2017 (55%).
- Respondents who revealed that they had reused open data in their research continues to shrink. In 2018, 48% said they had done this, whereas in 2017, 50% had done so, with 57% in 2016.
- Most researchers felt that that they did not get sufficient credit for sharing data (58%), compared to 9% who felt they do.
- Respondents having lost research data has decreased from 2017 (36% versus 30% in 2018).

Fig. 4 How familiar are you with the FAIR principles



60% of respondents had never heard of FAIR principles (Findability, Accessibility, Interoperability and Reusability), which provide a guideline for data producers and publishers to enhance the reusability of academic data.



Stephen Cawley

Function: Head of Institutional Marketing

Email: s.cawley@digital-science.com