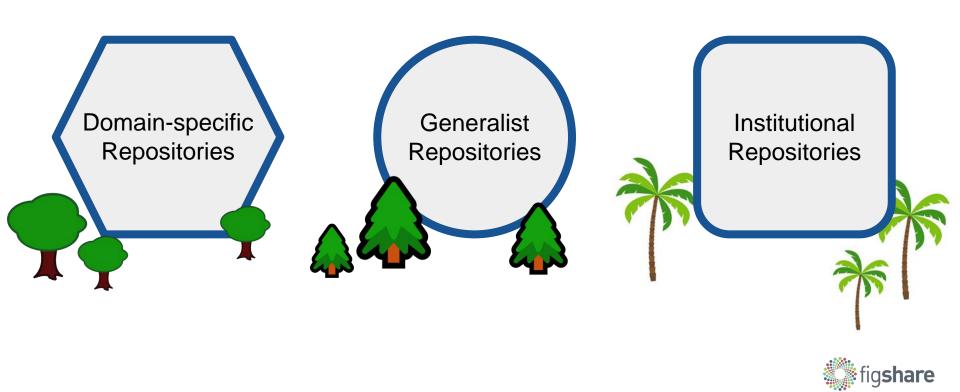
How Figshare and Generalist Repositories support the "R" in FAIR

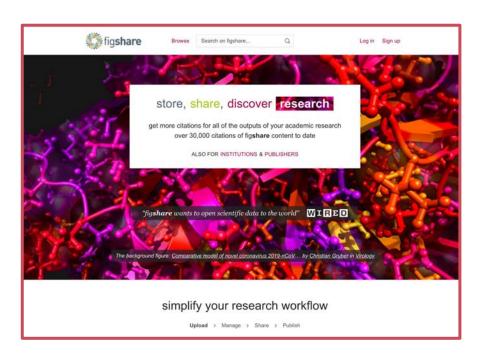
Ana Van Gulick, PhD
Government & Funder Lead, Head of Data Review
Figshare
ana@figshare.com

July 13, 2023

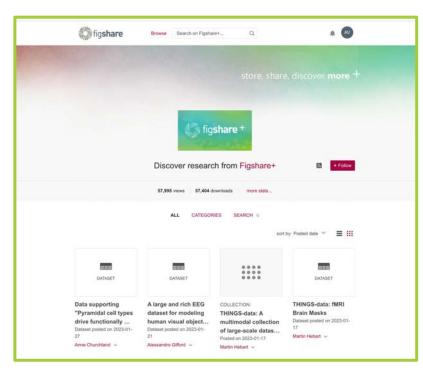


Research Data Repository Ecosystem





Freely available generalist repository figshare.com



Large Datasets plus.figshare.com





a freely available generalist repository for all research disciplines and outputs

Flexibility

- Share any research output or file type
- Files up to 20GB, Support for big datasets
- Preview files in the browser
- Collections

Researcher Workflows

- Open API and FTP
- GitHub, GitLab,
 BitBucket Integrations
- Collaborative spaces
- Restricted Access

Persistent Metadata

- doi Unique DOI for each output, reservable
- ORCID integration
- Link to publications
- Link Funding via

Open Access

- Open Access to all public files and metadata
- CC0 and CC-BY Licenses
- Discoverable across search engines, indexes
- FAIR commitment

Impact

- Public Author Profile
- Views, Downloads, Citations. Altmetrics
- Citations from full text
- TQ Faceted Search



a FAIR repository for big data

- Publish datasets over 20GB+ to 5TB or more
- File sizes up to 5TB

- Expert deposit support
- Dataset review
- One-time data publishing charge

Figshare's role in FAIR data

- Findable DOIs, metadata, discoverable
- Accessible No paywall, open access, programmatic access
- Interoperable metadata, linking, any file type
- Reusable License, metadata gives context





together with





Abstract

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a quideline for those wishing to enhance the reusability of their data holdings. Distinct from

https://doi.org/10.1038/sdata.2016.18



Figshare (and other generalist repositories) as a tool to find and reuse data

Generalist Repositories Offer:

- Flexibility to share any research output, any file type
- Discoverability and open access
- Open APIs programmatic access to files and metadata
- Provide DOIs for all records
- Use DOIs and other Persistent Identifiers (PIDs)
- Structured, standard metadata











Data sharing practices for reusable data

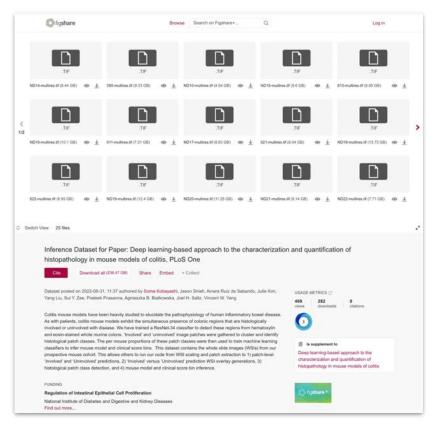
"meta(data) are richly described with a plurality of accurate and relevant attributes" "(meta)data meet domain-relevant community standards"

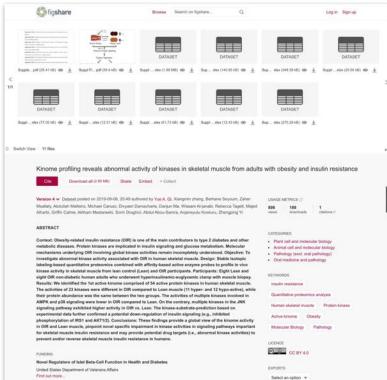
Use Figshare (and other generalist repositories) to:

- Create FAIR records for all outputs of any type
- Use DOIs and PIDs
- Link related materials
- Create descriptive metadata for discovery and reuse incl. discipline-specific standards
- Provide discoverability and access to all files and metadata needed for reuse



Creating FAIR Files and Metadata

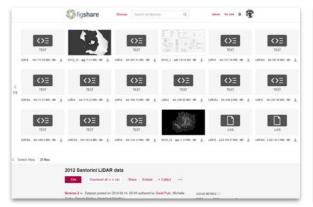








Any File Type, Any Type of Output

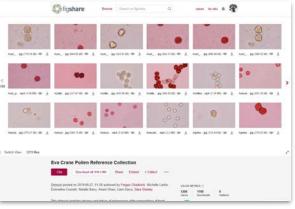




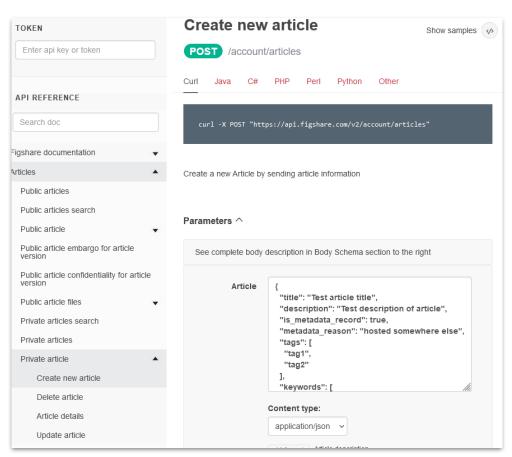




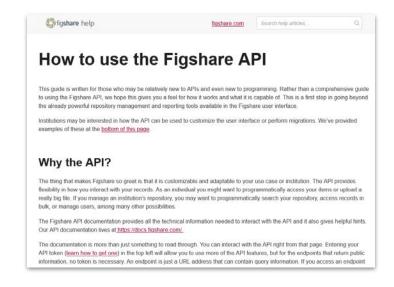




Create Records Programmatically



https://docs.figshare.com



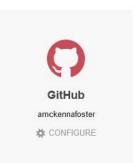


Integrations With Other Tools

Integrations

Connect and sync your figshare account with other apps.











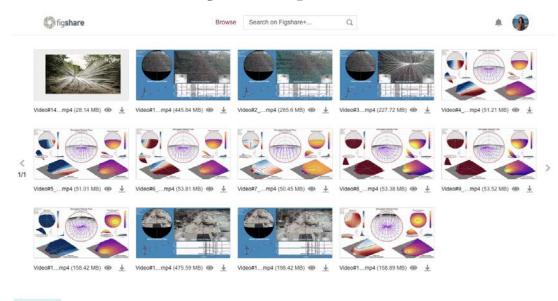
- RSpace
- Overleaf
- Open Science Framework
- ImpactStory
- Lab Folder
- Others (list of apps)

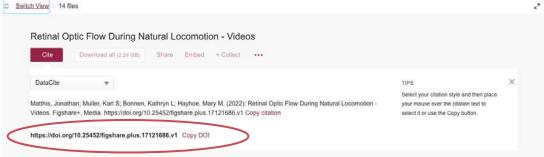
Create items

- snapshot Git-based systems
- integrate with lab notebooks
- pull in metadata from ORCID



Give every output a DOI



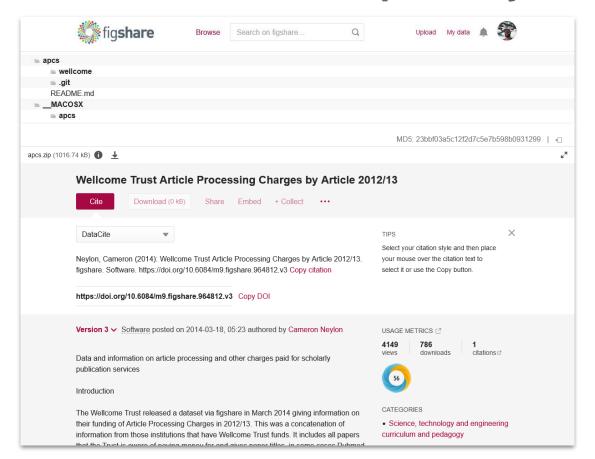


A Persistent Identifier (PID) will

- Provide persistent access
 - Reduce link rot
- Provide clear reference in documentation to outputs
- With DOIs, metadata stored with DOI provider discoverability

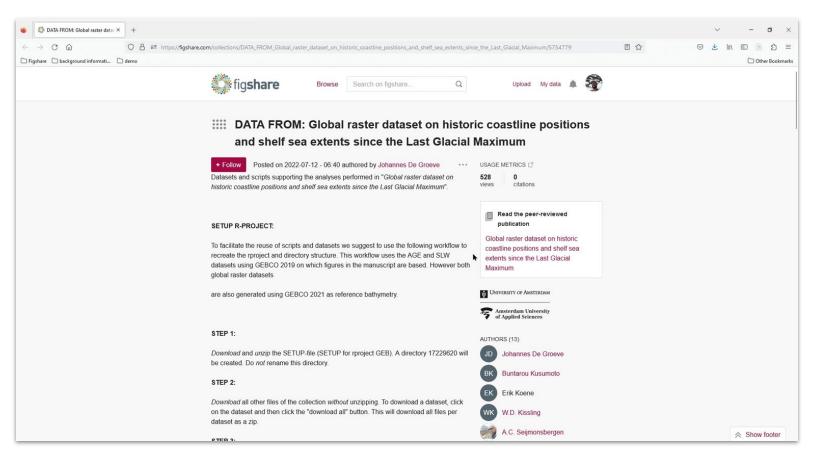


Create a DOI for a GitHub Repository





Create a DOI for a Collection of Items





DOIs are Reservable and Versionable

Item actions

- Add embargo and restricted access
- Share with private link
- DOI Disable DOI Copy DOI 10.0166/FK2.stagefigshare.8403020

The DOI becomes active when the item is published

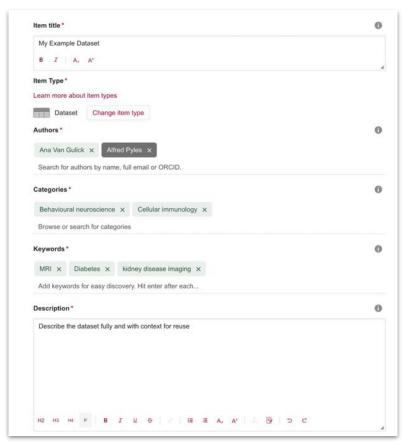
Edit timeline

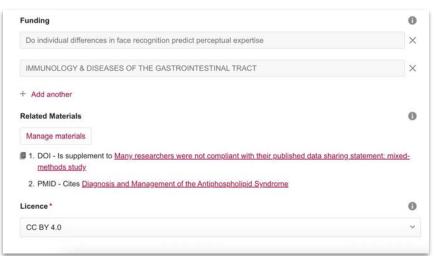
Version 8 ➤ Dataset posted on 2022-01-24, 03:25 authored by Malte Luecken, Maren	
	nterlandi, Michaela Müller, Daniel Strobl, Luke Zappia,
Version 8 2022-01-24, 03:25	atché, Fabian Theis, Kridsadakorn Chaichoompu
Version 7 2020-12-08, 02:45	
Version 6 2020-12-07, 14:55	immune cell, and human and mouse immune cell
Version 0 2020-12-07, 14.33	ks, and all ATAC mouse brain integration tasks from the
Version 5 2020-10-12, 08:16	as-level data integration in single-cell genomics".
Version 4 2020-09-25, 06:14	
	ed from public datasets, cell annotations were
Version 3 2020-09-23, 01:19	d the data was consistently preprocessed using scran
Version 2 2020-09-01, 07:26	on (for RNA tasks). In the immune cell datasets an
Version 4 2020 06 04 00:27	tory was also annotated. Details on dataset the paper and in the accompanying Github at
Version 1 2020-06-04, 09:27	b/scib.
	ID/SCID.

Please cite the paper and the papers the individual datasets were aggregated from wher using this data.



Structured Metadata For all Outputs

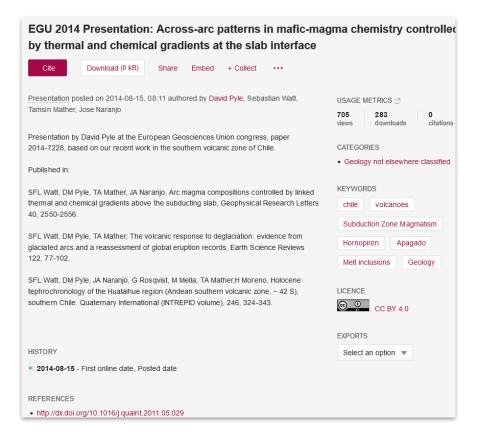




- Provide context for discoverability and reuse
- Add discipline- or method-specific information in the description for reuse
- Use PIDs for authors
- Use DOIs for related outputs
- Link to NIH funding
- Indexed for discoverability



Structured Metadata For all Outputs





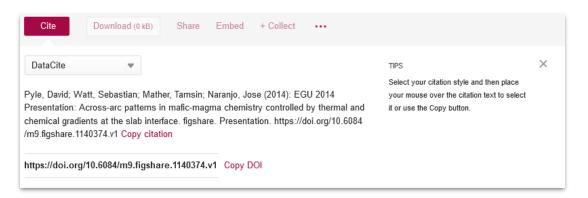
Structured Metadata For all Outputs

```
"authors": [
   "full name": "David Pyle",
   "is active": true,
   "url name": "David Pyle",
   "orcid id": "0000-0002-2663-9940"
   "full name": "Sebastian Watt",
   "is active": false.
   "url name": " ",
   "orcid id": ""
   "full name": "Tamsin Mather",
   "is active": false,
   "url name": " ",
   "orcid id": "
   "full name": "Jose Naranjo",
   "is active": false,
   "url name": " ",
   "orcid id": "
"figshare url": "https://figshare.com/articles/presentation/EGU 2014 Presentation Across arc patter
"funding": null,
"funding list": [].
```

```
H<dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/" 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>EGU 2014 Presentation: Across-arc patterns in mafic-magma chemistry controlled by thermal and chemical
 gradients at the slab interface</dc:title>
 <dc:creator>David Pvle</dc:creator>
 <dc:creator>Sebastian Watt</dc:creator>
 <dc:creator>Tamsin Mather</dc:creator>
 <dc:creator>Jose Naranjo</dc:creator>
 <dc:identifier identifierType="DOI">10.6084/m9.fiqshare.1140374.v1</dc:identifier>
 https://fiqshare.com/articles/presentation/EGU 2014 Presentation Across arc patterns in mafic magma chemistry controlled
 by thermal and chemical gradients at the slab interface/1140374</dc:relation>
dc:description>4lt;p4gt;Presentation by David Pyle at the European Geosciences Union congress, paper 2014-7228, based
 on our recent work in the southern volcanic zone of Chile. </p&gt;
 <p&qt;Published in: &lt;/p&qt;
 6lt;p&qt;SFL Watt, DM Pyle, TA Mather, JA Naranjo, Arc magma compositions controlled by linked thermal and chemical
 gradients above the subducting slab, Geophysical Research Letters 40, 2550-2556.61t;/p>
 6lt;p6gt;SFL Watt, DM Pyle, TA Mather, The volcanic response to deglaciation: evidence from glaciated arcs and a
 reassessment of global eruption records, Earth Science Reviews 122, 77-102. 61t;/p6qt;
 61t;p&qt;SFL Watt, DM Pyle, JA Naranjo, G Rosqvist, M Mella, TA Mather, H Moreno, Holocene tephrochronology of the
 Hualaihue region (Andean southern volcanic zone, ~ 42 S), southern Chile. Quaternary International (INTREPID volume),
 246, 324-343, @lt:/p@gt:
 -<p&qt; &lt;/p&qt;</dc:description>
 <dc:date>2014-08-15 15:11:57</dc:date>
 <dc:subject>chile</dc:subject>
 <dc:subject>volcanoes</dc:subject>
 <dc:subject>Subduction Zone Magmatism</dc:subject>
 <dc:subject>Hornopiren</dc:subject>
 <dc:subject>Apagado</dc:subject>
 <dc:subject>Melt inclusions</dc:subject>
 <dc:subject>Geology</dc:subject>
L</dc:dc>
```



Linking and PID use



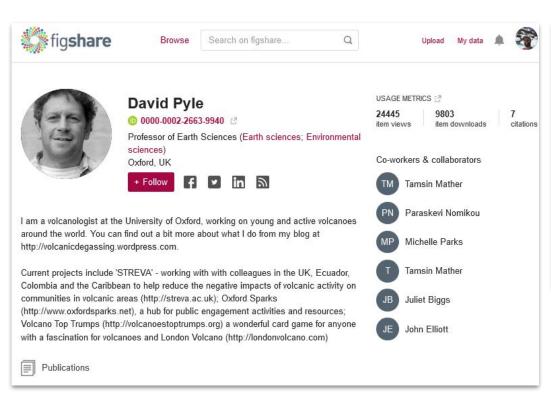


RELATED MATERIALS

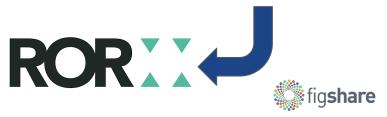
- 1. URL References https://doi.org/10.25452/figshare.plus.c.5983795
- 2. URL References https://github.com/Rajaram-Lab/cancres-2022-intratumoral-heterogeneity-dl-paper
- 3. URL References https://github.com/Rajaram-Lab/cancres-2022-intratumoral-heterogeneity-dl-paper/tree/master/Data_riles
- 4. DOI Is supplement to Intratumoral resolution of driver gene mutation heterogeneity in renal cancer using deep learning view PDF



Linking and PID use



ROR ids coming soon!



Link Funding

FUNDING

Regulation of Intestinal Epithelial Cell Proliferation

National Institute of Diabetes and Digestive and Kidney Diseases

Find out more...





DOCUMENTS - e.g. plastic AND instrument

ave / Expo

. .

upport

Grant

Regulation of Intestinal Epithelial Cell Proliferation

Funder: National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)
Grant number: R01DK052230 - Original description A

Investigators

VINCENT W YANG - Stony Brook University

Research organization

Stony Brook University, United States

Abstract

PROJECT SUMMARY/ABSTRACT The mammalian intestinal epithelium is a continuously renewing and highly regenerative tissue in which numerous biological processes such as proliferation, differentiation, migration, and apoptosis are carefully choreographed to achieve homeostasis. Studies have indicated that multiple signaling pathways including the WNT, NOTCH, BMP and HH form important components of the regulatory network and converge upon the intestinal crypts where intestinal stem cells (ISCs) reside. Recent studies have identified distinct populations of ISCs based on the markers that they express. At the present time ISCs are divided into two relatively broad functional groups: active ISCs (aISCs), a population of crypt base columnar (CBC) cells expressing LGR5, function as the multipotent stem cells during homeostasis, and quiescent or reserve ISCs

Share

Details

Funding amount USD 7,755,285

Funding period 1997 - 2023 1 Jul 31 Mar

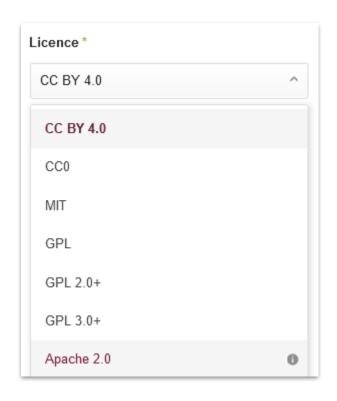
Program

Non-SBIR/STTR (Funding Mechanism)

Resulting publications 124



Clear, Machine-readable licenses for reuse



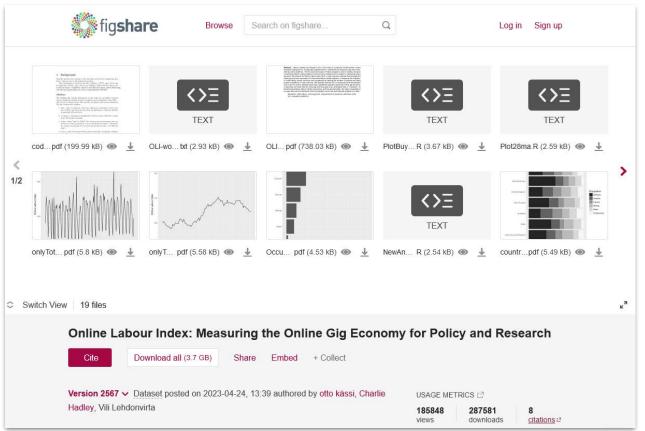
Machine readable licenses

Open licenses are encouraged





Accessing Files and Metadata

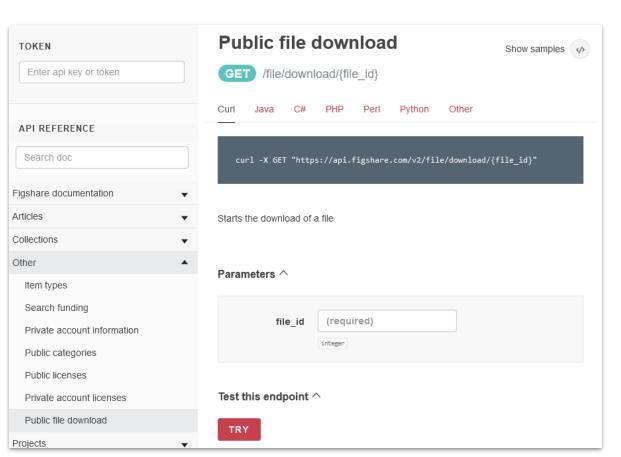


https://doi.org/10.6084/m9.figshare.3761562.v2567

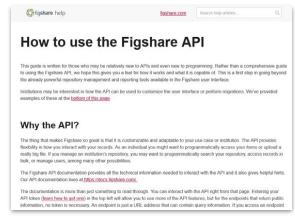
https://doi.org/10.6084/m9.figshare.3761562



Programmatic Access to Files and Metadata

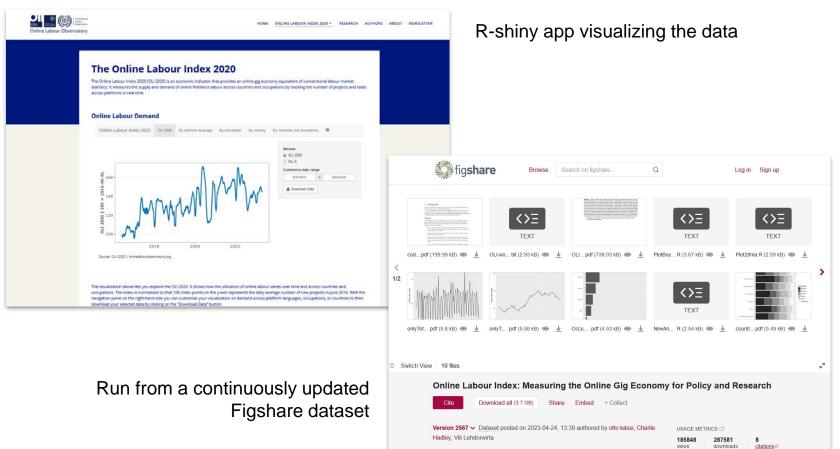


https://docs.figshare.com



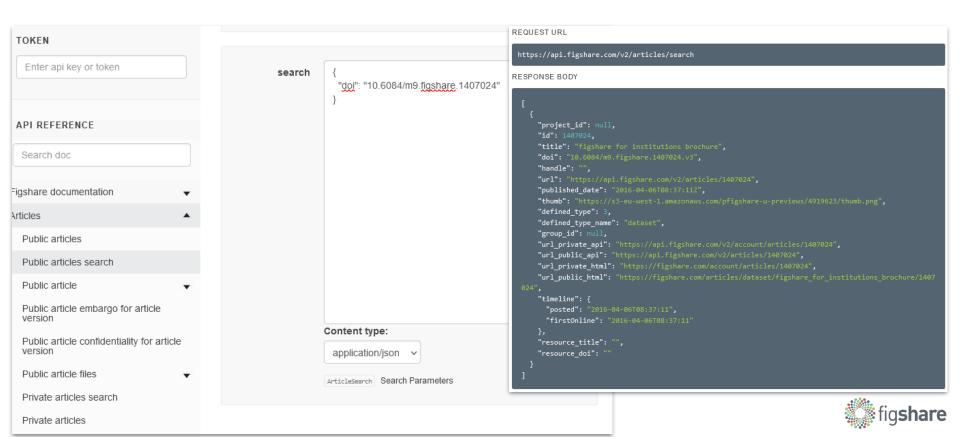


Programmatic Access to Files and Metadata





Programmatic Access to Files and Metadata



Discoverability

- Research that can be found and built upon
- Discoverable FAIR outputs offer doors opportunities for innovation and collaboration
- Well described research outputs can be reused
- Someone searching by topic can find many related outputs needed for replication or reuse

Google, Google Scholar, and Google Dataset Search

We mark up all content uploaded to Figshare to enable identification by Google. This involves configuring meta-tags in line with Google indexing guidelines so that their 'parser' software can identify your data when crawling the Figshare site. Data from DataCite suggests that Figshare content gets 10x more views than any of the other DataCite DOI minting repositories.

Please note: it can take a few weeks for Google to index your research.

Data Citation Index

Data Citation Index is a citation database powered by Clarivate. If your organization has access to the Data Citation Index through a subscription, you'll have access to Figshare citation data.

For more information, please visit https://clarivate.com/webofsciencegroup/solutions/webofscience-data-citation-index/.

DataCite

Not only is Figshare a DataCite consortium member, but items published on Figshare with a DataCite DOI are harvested into DataCite's vast database of scholarly records.

For more information, please visit https://datacite.org/.

DataMed

DataMed is a biomedical data search engine that includes thousands of biomedical research items published on Figshare.

For more information, please visit https://datamed.org/.

Dimensions

Dimensions is an Abstracting & Indexing (A&I) database that indexes grants, publications, datasets, clinical trials, patents, and policy documents. Datasets published in Figshare are harvested into Dimensions automatically.

For more information, please visit https://www.dimensions.ai/.

PubMed LinkOut

PubMed LinkOut is a service that allows you to link directly from PubMed and other NCBI databases to a wide range of information and services beyond the NCBI systems. LinkOut aims to facilitate access to relevant online resources in order to extend, clarify, and supplement information found in NCBI databases. Examples of LinkOut Resources include full-text publications, biological databases, consumer health information, research tools, and more.







Figshare Guides for NIH Data

Guide to Sharing NIH Research on Figshare https://help.figshare.com/article/guide-to-sharing-nih-funded-research-on-figshare-com

Guide to Including Figshare in a DMP
https://help.figshare.com/article/how-to-write-a-data-management-plan-dmp-and-include-figshare-in-your-data-sharing-plans

Guide to sharing Data on Figshare+ https://help.figshare.com/article/guide-to-sharing-dataon-figshare-plus

Visit help.figshare.com





Getting started!

Sign up for figshare.com
https://figshare.com/account/register

Publish a dataset on Figshare+ https://knowledge.figshare.com/plus#purchase-figshare-plus

Figshare API https://docs.figshare.com/

Visit help.figshare.com

Thank you!

Ana Van Gulick, PhD

Government & Funder Lead Head of Data Review ana@figshare.com

@anavangulick@figshare

figshare.com



open data
open access
striving for FAIR-ness
empowering researchers
...for the goodness of humankind



Thank you

