



# FAIR Principles

Samah Jaber, FAIR Office Austria, 13<sup>th</sup> Sep 2022

with slides from lecture "FAIR Principles" by Tomasz Miksa



# About the Speaker

## Samah Jaber BSc, MSc

### ✓ *Areas of Specialties*

- IT Governance & Operational Management
- Quality Management System
- Process Compliance/Improvement
- Information Architecture.
- FAIRification Process



 MSc. in Data Science at TUWIEN

### ✓ *Research Interests:*

- Data Stewardship
- Data Management Plan
- FAIR Data



# AGENDA

- Why we need FAIR principles?
- FAIR principles (big picture)
- FAIR principles (details)
- How repositories support FAIRness?
- How to FAIRify data?
- FAIR assessment
- GO FAIR
- FAIR Office Austria

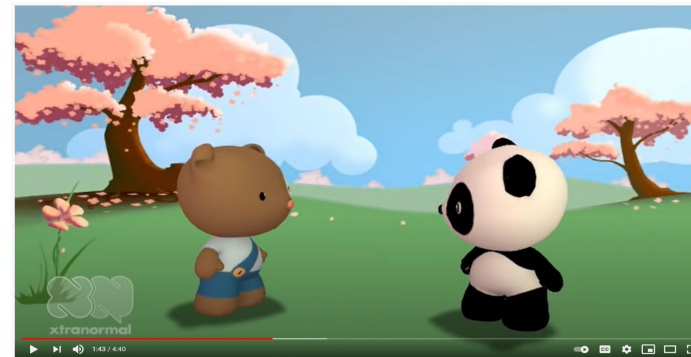


# Why we need FAIR Principles?

# Researchers trying to reuse data...

Conversation of two researchers

- Can I see your data?
- It's on my USB stick
- Can I have it?
- I have in a box and I have moved recently
- Can I have it?
- I forgot to label the boxes...
- (half a year later)
- Thanks, for the USB. However, I cannot read the hexadecimal file on it. How do I open it?
- You need a special program
- What program?
- ...



Hanson, Karen; Surkis, Alisa; Yacobucci, Karen: Data Sharing and Management Snafu in 3 Short Acts.  
<https://doi.org/10.5446/31036>



## Data

- The output of any research processing ( e.g: capturing, pre-processing, transformation, integration, analysis ).

## Search

- To process data we need to make them searchable.
- We're producing more and more data all the time, and if we don't add structure to that data, we only make it useful for ourselves.

## FAIR

- Provide a way to extend the reach of our data.
- Tend to allow more people to make use of it.
- Enhance your recognition as a researcher.



# Research Community

- **Researchers** FAIR data will result in better research output.
- **Funders** FAIR data will add more value to publicly funded research.
- **Publishers** FAIR data will improve the review process.
- **Universities** FAIR data will ensure high research integrity.

# FAIR Principles and Research Lifecycle?



*These four principles should be applied throughout the entire data lifecycle, and they are closely interconnected to increase the reuse of scientific data.*





# SCIENTIFIC DATA

Amended: Addendum

**OPEN** **Comment: The FAIR Guiding Principles for scientific data management and stewardship**

## SUBJECT CATEGORIES

- » Research data
- » Publication characteristics

 Mark D. Wilkinson *et al.*<sup>#</sup>

Received: 10 December 2015

Accepted: 12 February 2016

Published: 15 March 2016

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 measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

### Supporting discovery through good data management

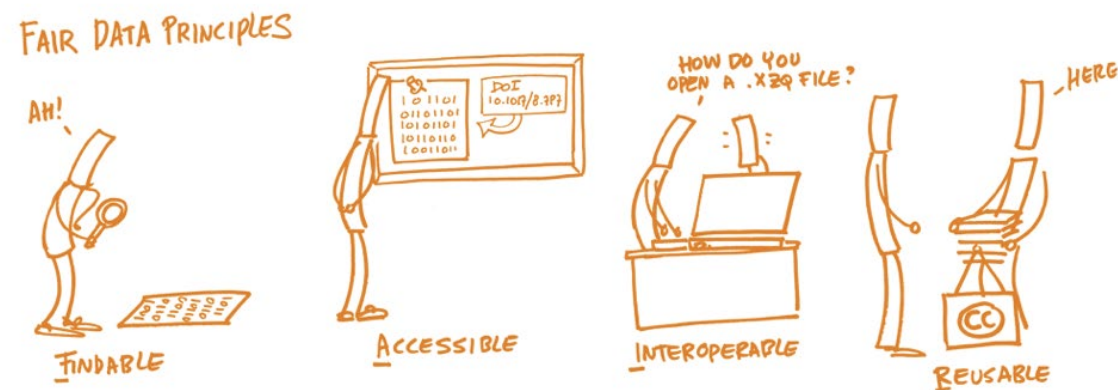
Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process. Unfortunately, the existing digital ecosystem surrounding scholarly data publication prevents us from extracting maximum benefit from our research investments (e.g., ref. 1). Partially in response to this, science funders, publishers and governmental agencies are beginning to require data management and stewardship plans for data generated in publicly funded experiments. Beyond proper collection, annotation, and archival, data stewardship includes the notion of 'long-term care' of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data. The outcomes from good data management and stewardship, therefore, are high quality digital publications that facilitate and simplify this ongoing process of discovery, evaluation, and reuse in downstream studies. What constitutes 'good data management' is, however, largely undefined, and is generally left as a decision for the data or repository owner. Therefore, bringing some clarity around the goals and desiderata of good data management and stewardship, and defining simple guideposts to inform those who publish and/or preserve scholarly data, would be of great utility.

This article describes four foundational principles—Findability, Accessibility, Interoperability, and Reusability—that serve to guide data producers and publishers as they navigate around these obstacles, thereby helping to maximize the added-value gained by contemporary, formal scholarly digital publishing. Importantly, it is our intent that the principles apply not only to 'data' in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All scholarly digital research objects—from data to analytical pipelines—benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability.

There are numerous and diverse stakeholders who stand to benefit from overcoming these obstacles: researchers wanting to share, get credit, and reuse each other's data and interpretations; professional data publishers offering their services; software and tool-builders providing data analysis and processing services such as reusable workflows; funding agencies (private and public) increasingly

Correspondence and requests for materials should be addressed to B.M. (email: [barend.mons@dtis.nl](mailto:barend.mons@dtis.nl)).

<sup>#</sup>A full list of authors and their affiliations appears at the end of the paper.



# FAIR Principles

[Home](#) > [FAIR Principles](#)

## > FAIR Principles

- > **F1: (Meta) data are assigned globally unique and persistent identifiers**
- > **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**
- > **A1: (Meta)data are retrievable by their identifier using a standardised communication protocol**
- > **A1.1: The protocol is open, free and universally implementable**
- > **A1.2: The protocol allows for an authentication and authorisation where necessary**
- > **A2: Metadata should be**

In 2016, the '**FAIR Guiding Principles for scientific data management and stewardship**' were published in *Scientific Data*. The authors intended to provide guidelines to improve the **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse of digital assets. The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.

A practical "how to" guidance to go FAIR can be found in the **Three-point FAIRification Framework**.

### **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 **FAIRification process**.

**F1. (Meta)data are assigned a globally unique and persistent identifier**

**F2. Data are described with rich metadata (defined by R1 below)**

**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**

<https://www.go-fair.org/fair-principles/>



# FAIR vs fair

FAIR principles  $\neq$  Algorithmic fairness

To be FAIR

- To apply/use FAIR principles
- Focus on how data is managed, etc.

To be fair

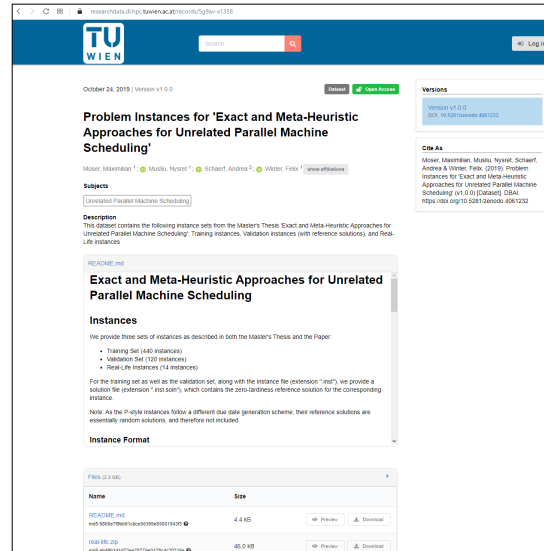
- Evade bias
- Focus on design and implementation



# FAIR Principles (big picture)

# Findable – simplified examples

- Yes

October 24, 2019 / Version v1.0.0

### Problem Instances for Exact and Meta-Heuristic Approaches for Unrelated Parallel Machine Scheduling

Mohr, Maximilian · Muhl, Nysret · Schmitt, Andrea · Weber, Felix · [View statistics](#)

**Subjects**  
 Unrelated Parallel Machine Scheduling

**Description**  
 This dataset contains the following instance sets from the authors' Thesis: Exact and Meta-Heuristic Approaches for Unrelated Parallel Machine Scheduling; Training instances; Validation instances (with reference solutions); and Real-Life instances.

**README.md**  
**Exact and Meta-Heuristic Approaches for Unrelated Parallel Machine Scheduling**

**Instances**  
 We provide three sets of instances as described in both the authors' Thesis and the Paper:

- Training Set (445 instances)
- Validation Set (120 instances)
- Real-Life instances (14 instances)

For the training set as well as the validation set, along with the instance file (extension ".ins"), we provide a solution file (extension ".inst\_soln"), which contains the zero-hardness reference solution for the corresponding instance.

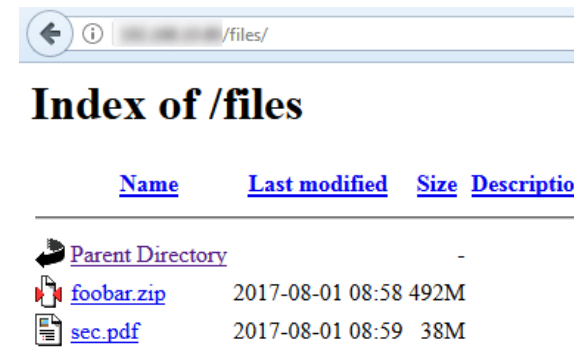
*Note:* As the P-objn instances follow a different data-generation scheme, their reference solutions are essentially random solutions, and therefore not included.

**Instance Format**

Name	Size		
README.md	4.4 KB	<a href="#">Preview</a>	<a href="#">Download</a>
1048-05_20	46.0 KB	<a href="#">Preview</a>	<a href="#">Download</a>




Data repository

- No

← ⓘ [/files/](#)

## Index of /files

	<a href="#">Name</a>	<a href="#">Last modified</a>	<a href="#">Size</a>	<a href="#">Description</a>
	<a href="#">Parent Directory</a>		-	
	<a href="#">foobar.zip</a>	2017-08-01 08:58	492M	
	<a href="#">sec.pdf</a>	2017-08-01 08:59	38M	

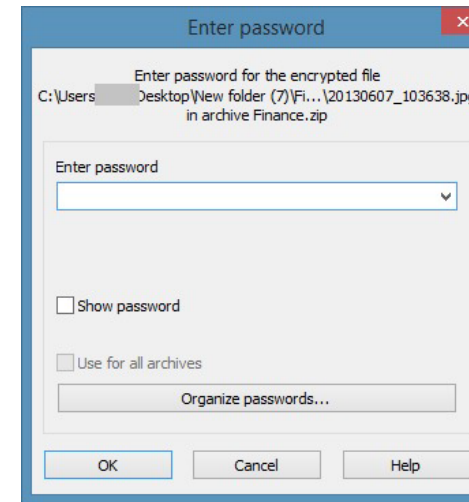
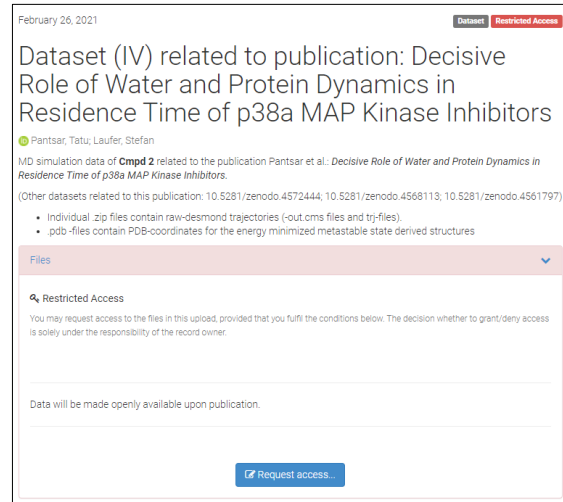
Personal website

# Accessible – simplified examples

- Yes



- No



Restricted access, but a clear way to request access

# Interoperable – simplified examples

- Yes


- XML following known XSD Schema
- MP3 for audio recordings

- No

- Custom XML without any documentation
- M4P (Apple) for audio recordings

# Reusable – simplified examples

• Yes



**Jährliche Personeneinkommen**

Die knapp 4,6 Mio. **unselbständig Erwerbstätigen** (ohne Lehrlinge) erzielten 2019 ein mittleres **Bruttogehalt** von 29.458 Euro. Die Einkommen der Frauen erreichten mit 22.808 Euro im Mittel nur 63,6% des Einkommens der Männer (35.841 Euro), wobei Frauen viel häufiger **teilzeitbeschäftigt** sind. Die mittleren **Nettogehälter** beliefen sich auf 22.104 Euro (Frauen: 18.233 Euro, Männer: 25.938 Euro).

Werden die Einflüsse von **Teilzeit** und **nicht-ganzjähriger Beschäftigung** ausgeklammert und nur Personen berücksichtigt, die laut Lohnsteuerdaten Vollzeit beschäftigt sind und im Jahr 2019 mindestens 350 Tage im Jahr unselbständig erwerbstätig waren (ohne Lehrlinge), so betrug das mittlere Bruttogehalt der Frauen 39.320 Euro, während Männer im Mittel 45.900 Euro verdienten. Der relative Einkommensanteil der Frauen am mittleren Einkommen der Männer stieg somit auf 85,7%.

Bei einer Untergliederung nach **Beschäftigungsgruppen** müssen neben den Anteilen an Teilzeitbeschäftigten weitere **strukturelle Unterschiede** berücksichtigt werden. Insbesondere ungleiche Anteile **nicht-ganzjähriger Beschäftigung** sowie unterschiedliche **Qualifikations- und Altersstrukturen**. So sind beispielsweise **Beamtinnen und Beamte** deutlich älter, weisen ein höheres **Ausbildungsniveau** auf und sind **kaum teilzeitbeschäftigt**. Bei einer Fehlung der Beschäftigungsgruppen nach dem Median der Bruttogehälter stehen **männliche Beamte an der Spitze** (59.772 Euro), gefolgt von **Beamtinnen** (58.233 Euro), **männlichen Angestellten** (47.373 Euro) und **männlichen Vertragsbediensteten** (42.112 Euro). **Weibliche Vertragsbedienstete** (32.392 Euro), **männliche Arbeiter** (28.454 Euro) und **weibliche Angestellte** (25.448 Euro) verdienen deutlich weniger, das **Schlusslicht** bilden die **Arbeiterinnen** (12.883 Euro). Betrachtet man die **Bestverdienenden** (oberstes Dezil), so lagen die **männlichen Angestellten** mit 98.644 Euro knapp hinter den **männlichen Beamten** mit 99.742 Euro, deutlich darunter blieben die **bestverdienenden Beamtinnen** mit 86.516 Euro.

➔ mehr...

Tabelle(n)	Grafiken	Dokumentationen	Allgemeine Auskünfte
Brutto- und Nettogehälter von unselbständig Erwerbstätigen 1997 bis 2019			
Brutto- und Nettogehälter von ganzjährig Vollzeitbeschäftigten 2004 bis 2019			
Brutto- und Nettogehälter von Pensionistinnen und Pensionisten 1997 bis 2019			
Bruttogehälter von unselbständig Erwerbstätigen 2019			
Nettogehälter von unselbständig Erwerbstätigen 2019			
Bruttogehälter von ganzjährig Vollzeitbeschäftigten 2019			
Nettogehälter von ganzjährig Vollzeitbeschäftigten 2019			
Bruttogehälter von Frauen und Männern 2019			
Bruttogehälter von Frauen und Männern nach Bundesländern 2019			
Brutto- und Nettogehälter nach Altersgruppen 2019			

Trusted source, permission to reuse, well defined meaning of terms used

• No




Provenance and permissions not clear



# Metadata

- What is in the picture?
- Where has it been produced?
- Who has produced it?
- What are the ingredients ?
- Production date?
- Valid till?

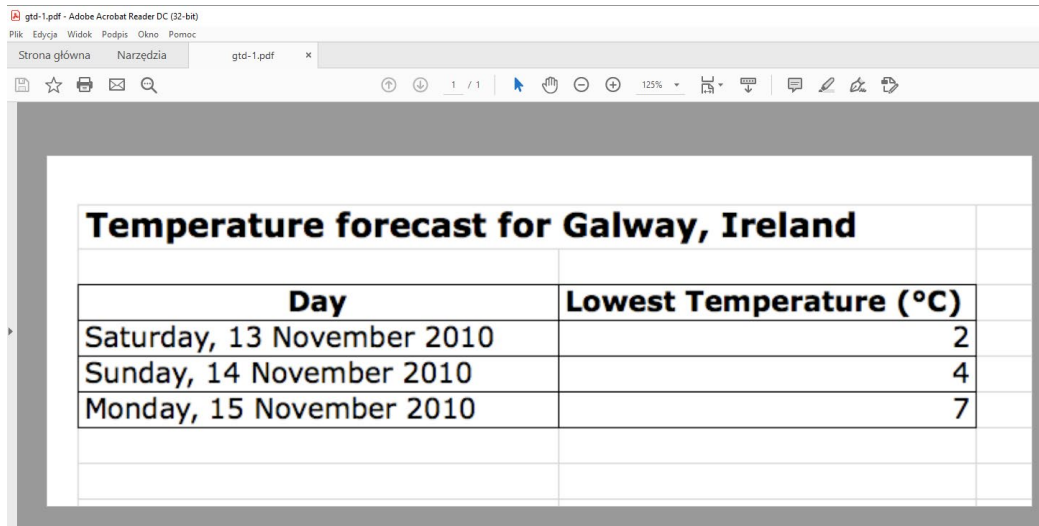


# FAIR vs Open Data

- FAIR data  $\neq$  open data!

Even if the data itself cannot be shared openly, you should create and publish a description of your data so that researchers with a relevant purpose can request permission to reuse the data.

# FAIR and Machine-actionability



gtd-1.pdf - Adobe Acrobat Reader DC (32-bit)

Plik: Edycja Widok Podpis Okno Pomoc

Strona główna Narzędzia gtd-1.pdf x

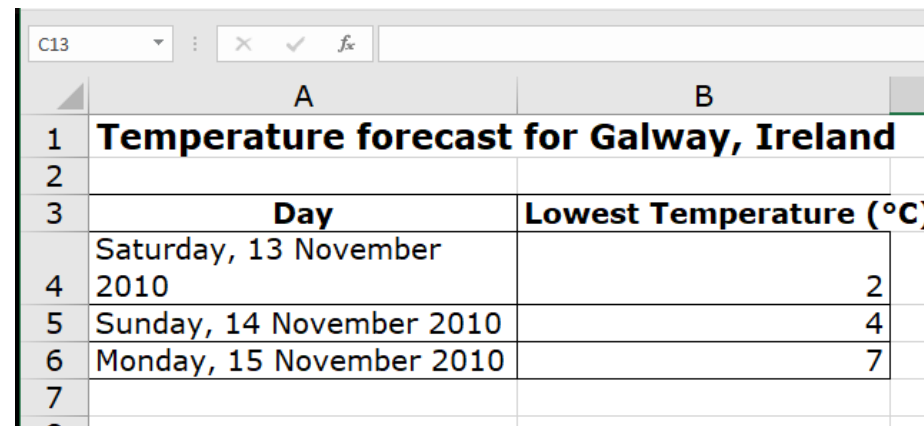
1 / 1 125%

Temperature forecast for Galway, Ireland	
Day	Lowest Temperature (°C)
Saturday, 13 November 2010	2
Sunday, 14 November 2010	4
Monday, 15 November 2010	7

Not machine-actionable



Machine-actionable



C13

	A	B
1	<b>Temperature forecast for Galway, Ireland</b>	
2		
3	Day	Lowest Temperature (°C)
4	Saturday, 13 November 2010	2
5	Sunday, 14 November 2010	4
6	Monday, 15 November 2010	7
7		



# FAIR Principles (details)

# Findable

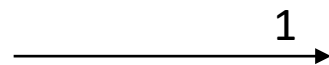
- F1. (Meta)data are assigned a globally unique and **persistent identifier**
- F2. Data are described with rich **metadata**
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a **searchable resource**



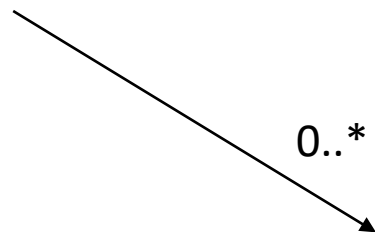
# Persistent identifiers (F1)

- Example

- A car has only one VIN (PID), but can have many number plates over its lifetime (URL)



VIN:	AZUSA1234567892222	
• MODEL:	Awesome Car	•
DATE of MFG:	1970	



# Persistent Identifiers (F1)

- Digital Object Identifier (DOI)
  - Uniquely identify objects
  - DOI assigned once
  - Physical location of data can change
- ORCID ID
  - Unique person ID
  - ORCID assigned once
  - Person can change affiliations (jobs)

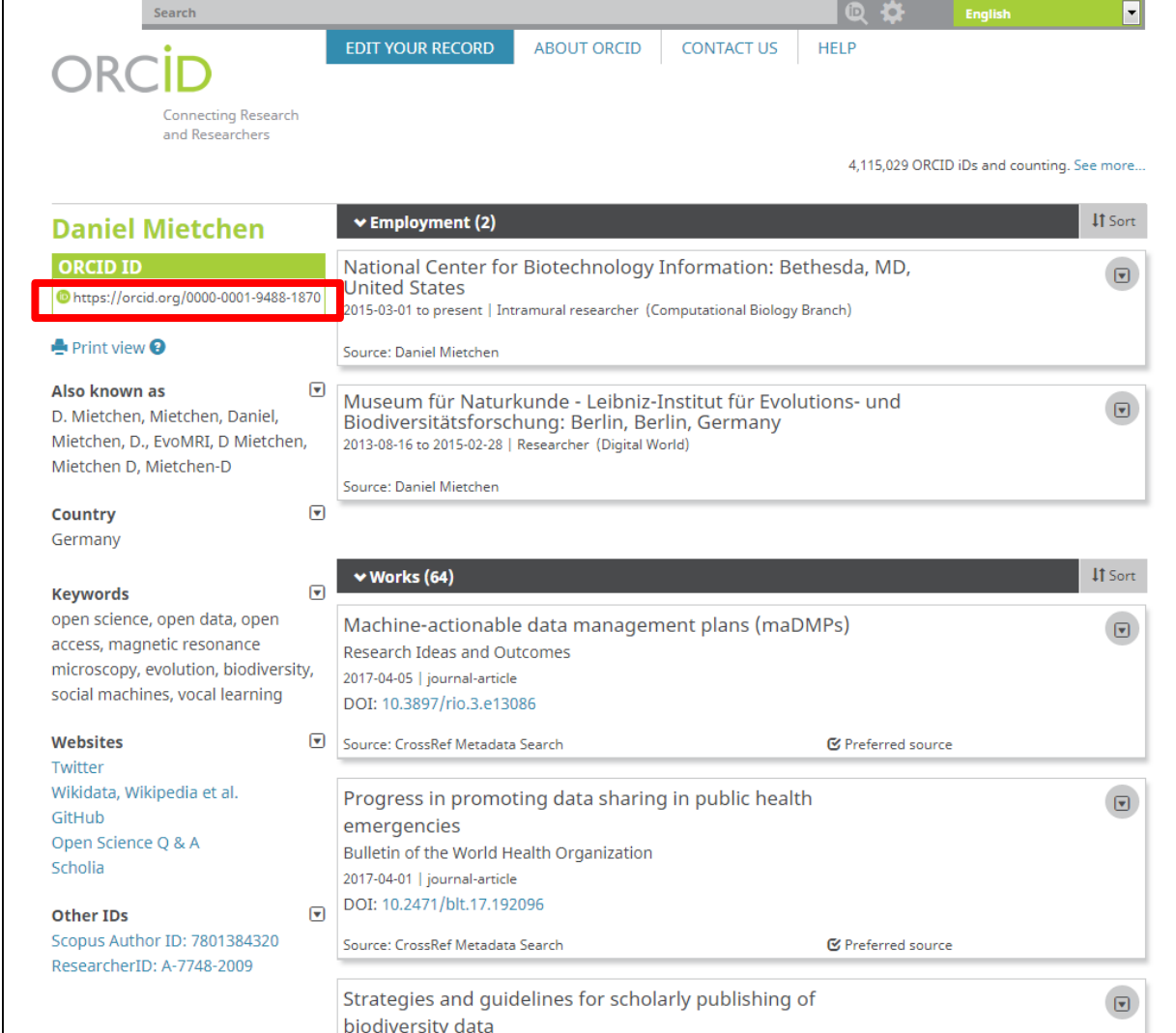


10.5281/zenodo.1068223



0000-0002-4929-7875

# ORCID Example



Search English

ORCID  
Connecting Research and Researchers

4,115,029 ORCID iDs and counting. [See more...](#)

**Daniel Mietchen** Employment (2) Sort

**ORCID ID**  
<https://orcid.org/0000-0001-9488-1870>

**Also known as**  
D. Mietchen, Mietchen, Daniel, Mietchen, D., EvoMRI, D Mietchen, Mietchen D, Mietchen-D

**Country**  
Germany

**Keywords**  
open science, open data, open access, magnetic resonance microscopy, evolution, biodiversity, social machines, vocal learning

**Websites**  
Twitter  
Wikidata, Wikipedia et al.  
GitHub  
Open Science Q & A  
Scholia

**Other IDs**  
Scopus Author ID: 7801384320  
ResearcherID: A-7748-2009

**Employment (2)**

- National Center for Biotechnology Information: Bethesda, MD, United States  
2015-03-01 to present | Intramural researcher (Computational Biology Branch)  
Source: Daniel Mietchen
- Museum für Naturkunde - Leibniz-Institut für Evolutions- und Biodiversitätsforschung: Berlin, Berlin, Germany  
2013-08-16 to 2015-02-28 | Researcher (Digital World)  
Source: Daniel Mietchen

**Works (64)**

- Machine-actionable data management plans (maDMPs)  
Research Ideas and Outcomes  
2017-04-05 | journal-article  
DOI: 10.3897/rio.3.e13086  
Source: CrossRef Metadata Search Preferred source
- Progress in promoting data sharing in public health emergencies  
Bulletin of the World Health Organization  
2017-04-01 | journal-article  
DOI: 10.2471/blt.17.192096  
Source: CrossRef Metadata Search Preferred source
- Strategies and guidelines for scholarly publishing of biodiversity data



# DOI example – assigned to publication

**PLOS** COMPUTATIONAL BIOLOGY

EDUCATION

## Ten principles for machine-actionable data management plans

Tomasz Miksa<sup>1\*</sup>, Stephanie Simms<sup>2\*</sup>, Daniel Mitchen<sup>3\*</sup>, Sarah Jones<sup>4\*</sup>

1 SBA Research & TU Wien, Vienna, Austria, 2 California Digital Library, University of California, Oakland, United States of America, 3 Data Science Institute, University of Virginia, Charlottesville, United States of America, 4 Digital Curation Centre, Glasgow, United Kingdom

\* These authors contributed equally to this work.  
\* [miksa@tuwien.ac.at](mailto:miksa@tuwien.ac.at)

 Check for updates

**OPEN ACCESS**

**Citation:** Miksa T, Simms S, Mitchen D, Jones S (2019) Ten principles for machine-actionable data management plans. *PLoS Comput Biol* 15(3): e1006750. <https://doi.org/10.1371/journal.pcbi.1006750>

**Published:** March 28, 2019

**Copyright:** © 2019 Miksa et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This research was carried out in the context of the Austrian COMET K1 program and publicly funded by the Austrian Research Promotion Agency (FFG) and the Vienna Business Agency (WAW). It was also supported by an NSF EAGER grant awarded to the California Digital Library (Award Number 1745675). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**Abstract**

Data management plans (DMPs) are documents accompanying research proposals and project outputs. DMPs are created as free-form text and describe the data and tools employed in scientific investigations. They are often seen as an administrative exercise and not as an integral part of research practice.

There is now widespread recognition that the DMP can have more thematic, machine-actionable richness with added value for all stakeholders: researchers, funders, repository managers, research administrators, data librarians, and others. The research community is moving toward a shared goal of making DMPs machine-actionable to improve the experience for all involved by exchanging information across research tools and systems and embedding DMPs in existing workflows. This will enable parts of the DMP to be automatically generated and shared, thus reducing administrative burdens and improving the quality of information within a DMP.

This paper presents 10 principles to put machine-actionable DMPs (maDMPs) into practice and realize their benefits. The principles contain specific actions that various stakeholders are already undertaking or should undertake in order to work together across research communities to achieve the larger aims of the principles themselves. We describe existing initiatives to highlight how much progress has already been made toward achieving the goals of maDMPs as well as a call to action for those who wish to get involved.

**Introduction**

Data management plans (DMPs) are documents accompanying research proposals. They describe the data that are used and produced during the course of research activities, where the data will be archived, which licenses and constraints apply, and to whom credit should be given. DMPs are awareness tools to help researchers manage their data and ensure that it will be of high quality, accessible, and reusable after the project has ended. DMPs are typically created manually, mostly by researchers using checklists and online questionnaires. They are required by funding bodies and institutions all over the world, e.g., the National Science

PLOS Computational Biology | <https://doi.org/10.1371/journal.pcbi.1006750> March 28, 2019 1 / 15

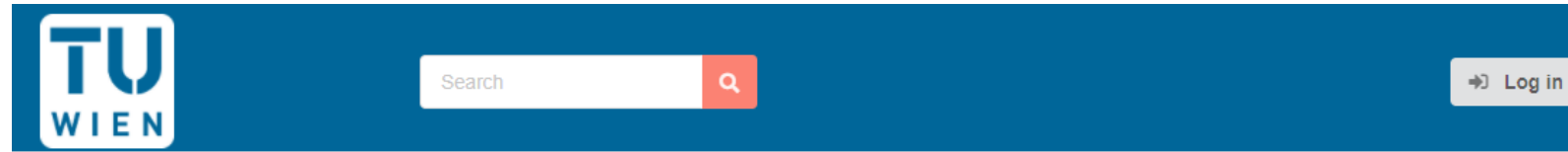


# DOI example – assigned to code

The screenshot shows the GitHub repository page for 'helmuthb / dmp-exercise1'. The repository is at version 1.0.1, has 1 branch, and 3 tags. The commit history shows a 'Corrected DOI link' by 'helmuthb' on April 22, 2019, with 4 commits. The file list includes 'data', 'src', '.gitignore', 'Dockerfile', 'LICENSE', 'README.md', and 'Report.pdf'. The 'README.md' file content is shown below, with a red box highlighting the DOI: `DOI 10.5281/zenodo.2648326`.

The screenshot shows the Zenodo record for 'US Wheat and Salzburg Middle-Aged Marriages - Data Experiment' by 'Helmuth Breitenfeller'. The record has 14 views and 5 downloads. It is available in GitHub and OpenAIRE. The DOI is `10.5281/zenodo.2648326`, which is highlighted with a red box. The record also shows a list of files, including 'LICENSE', 'README.md', 'Report.pdf', 'data', 'src', and 'f\_packages'. The 'Publication date' is April 22, 2019. The 'Versions' section shows three versions: 1.0.1 (DOI: 10.5281/zenodo.2648398), 1.0 (DOI: 10.5281/zenodo.2648396), and 0.1 (DOI: 10.5281/zenodo.2648327).






# DOI example - assigned to data



November 27, 2020 | Version 1.0

Dataset  Embargoed

## The Sentinel-1 Global Backscatter Model (S1GBM) - Mapping Earth's Land Surface with C-Band Microwaves

 Bauer-Marschallinger, Bernhard <sup>1</sup>; Cao, Senmao <sup>1,2</sup>;  Navacchi, Claudio <sup>1</sup>;  Freeman, Vahid <sup>1,3</sup>; Reuß, Felix <sup>1</sup>; Geudtner, Dirk <sup>4</sup>; Rommen, Björn <sup>4</sup>; Vega, Francisco Ceba <sup>4</sup>; Snoeij, Paul <sup>5</sup>; Attema, Evert <sup>4</sup>;  Reimer, Christoph <sup>2</sup>;  Wagner, Wolfgang <sup>1,2</sup> [show affiliations](#)

### Description

This dataset was generated by the Remote Sensing Group of the [TU Wien Department of Geodesy and Geoinformation \(https://mrs.geo.tuwien.ac.at/\)](https://mrs.geo.tuwien.ac.at/), within a dedicated project by the European Space Agency (ESA). Rights are reserved with ESA. Open use is granted under the CC BY-SA 4.0 license.

With this dataset publication, we open up a new perspective on Earth's land surface, providing a normalised microwave backscatter map from spaceborne Synthetic Aperture Radar (SAR) observations. The Sentinel-1 Global Backscatter Model (S1GBM) describes Earth for the period 2016-17 by the mean C-band radar cross section in VV- and VH-polarization at a 10 m sampling, giving a high-quality impression on surface- structures and -patterns.

At TU Wien, we processed 0.5 million Sentinel-1 scenes totaling 1.1 PB and performed semi-automatic quality curation and backscatter harmonisation related to orbit geometry effects. The overall mosaic quality excels (the few existing datasets, with minimised imprinting from orbit discontinuities and successful angle normalisation in large parts of the world. Supporting the design and verification of upcoming radar sensors, the obtained S1GBM data potentially also serve land cover classification and determination of vegetation and soil states, as well as water body mapping.

We invite developers from the broader user community to exploit this novel data resource and to integrate S1GBM parameters in models for various variables of land cover, soil composition, or vegetation structure.

### Versions

Version 1.0  
DOI: 10.48436/n2d1v-gqb91

### Cite As

Bauer-Marschallinger, Bernhard et al. (2020). The Sentinel-1 Global Backscatter Model (S1GBM) - Mapping Earth's Land Surface with C-Band Microwaves (Version 1.0) [Dataset]. TU Data. <https://doi.org/10.48436/n2d1v-gqb91>



# PIDs Examples



Persistent URLs



ORCID

# F2. Data are described with rich metadata

## Resources



### Rockfall Source Areas, Kitzsteinhorn, Austria ...

This dataset contains information on 749 mass movements (out of which 632 are...

Explore ▾

## Dataset Metadata

Export Metadata ▾

Contact Basics Keywords Spatial Time Specifics Quality Conformity

### Basic Information about this dataset

**Dataset Locator - URI** <https://hdl.handle.net/20.500.11756/70ef62e8>

**Abstract** This dataset contains information on 749 mass movements (out of which 632 are rockfalls) that were identified during a six-year (2011-2017) terrestrial laserscanning monitoring at the Kitzsteinhorn, Hohe Tauern Range, Austria. The data documents the significant impact that retreating glaciers have on rockfall occurrence in two deglaciating cirques. The dataset includes: mass movement volume, substrate type, failure depth, height of source area above the glacier surface, slope angle/aspect of source area. An extensive analysis and interpretation of the dataset can be found in two research papers published in the open-access journal "Earth Surface Dynamics" (Hartmeyer et al. 2020). Funding information: Data acquisition was co-funded by the Austrian Academy of Sciences (ÖAW) (Project 'GlacierRocks') and the Austrian Research Promotion Agency (FFG) (Project 'MOREXPART').

**Metadata Language** English

**License** cc-by-sa

**Visibility** public

**Use Limitation** no limitation

# F2. Data are described with rich metadata

## Resources

 **Rockfall Source Areas, Kitzsteinhorn, Austria ...**  
This dataset contains information on 749 mass movements (out of which 632 are...

[↶ Explore](#)

### Dataset Metadata

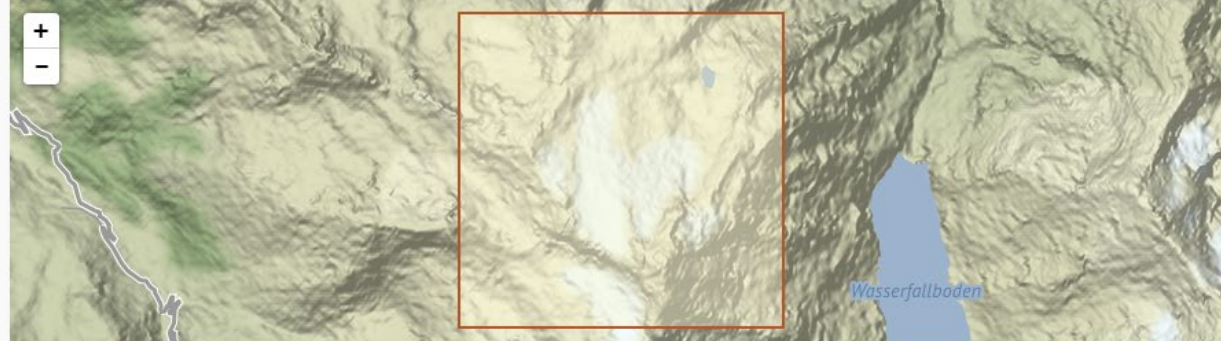
[Export Metadata](#)

Contact Basics Keywords **Spatial** Time Specifics Quality Conformity

#### Geographic Aspects of the Resources

Polygon

 Dataset extent

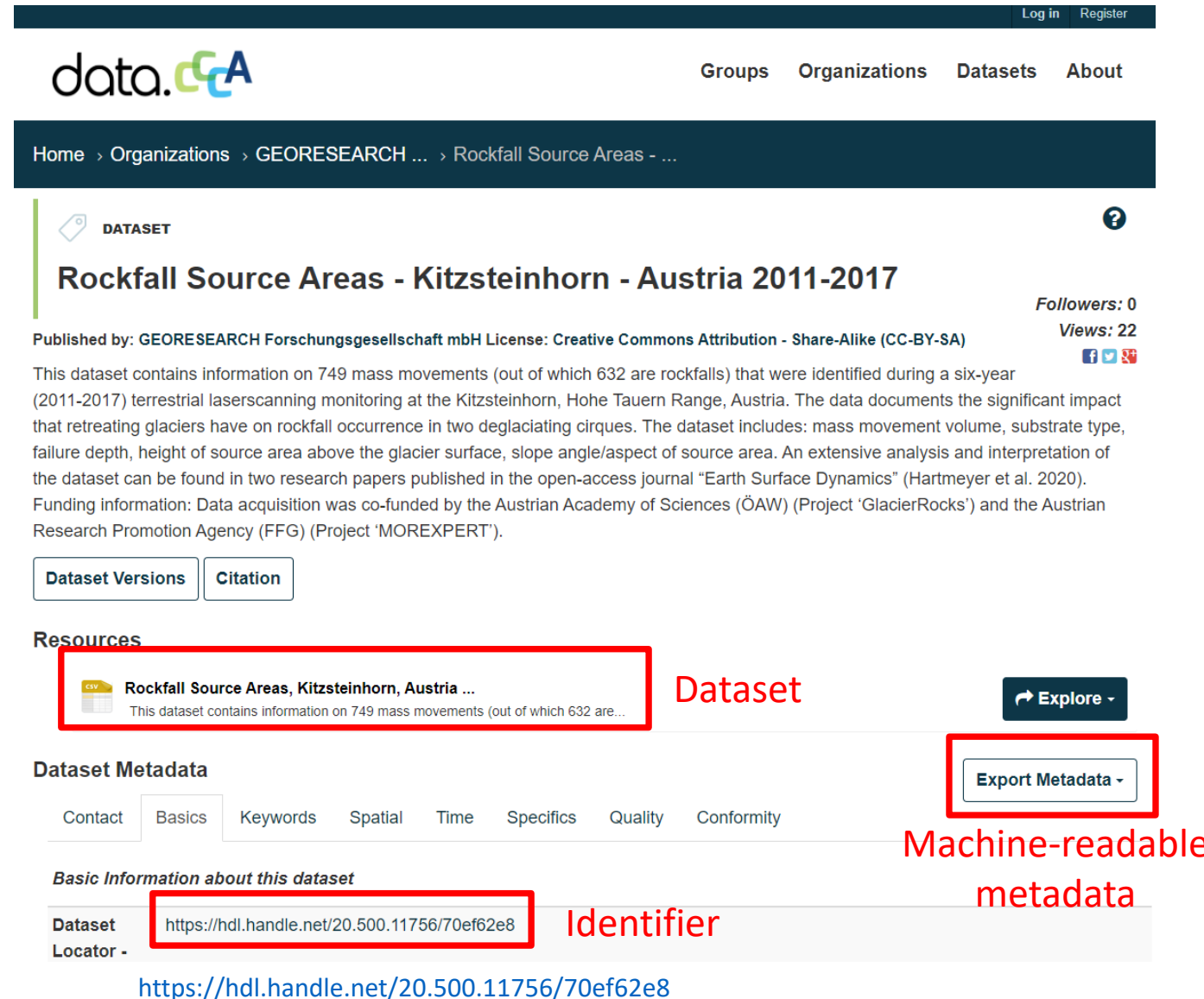


Map data © OpenStreetMap contributors  
Tiles by MapQuest

Coverage

Kitzsteinhorn, Hohe Tauern Range, Austria

# F3. Metadata clearly and explicitly include the identifier of the data they describe



The screenshot shows a dataset page on the data.cca.org website. The page title is "Rockfall Source Areas - Kitzsteinhorn - Austria 2011-2017". The page includes a description of the dataset, published by GEORESEARCH Forschungsgesellschaft mbH, with a Creative Commons Attribution - Share-Alike (CC-BY-SA) license. The dataset contains information on 749 mass movements (out of which 632 are rockfalls) identified during a six-year (2011-2017) terrestrial laserscanning monitoring at the Kitzsteinhorn, Hohe Tauern Range, Austria. The data documents the significant impact that retreating glaciers have on rockfall occurrence in two deglaciating cirques. The dataset includes: mass movement volume, substrate type, failure depth, height of source area above the glacier surface, slope angle/aspect of source area. An extensive analysis and interpretation of the dataset can be found in two research papers published in the open-access journal "Earth Surface Dynamics" (Hartmeyer et al. 2020). Funding information: Data acquisition was co-funded by the Austrian Academy of Sciences (ÖAW) (Project 'GlacierRocks') and the Austrian Research Promotion Agency (FFG) (Project 'MOREXPART').

Key elements highlighted in red boxes:

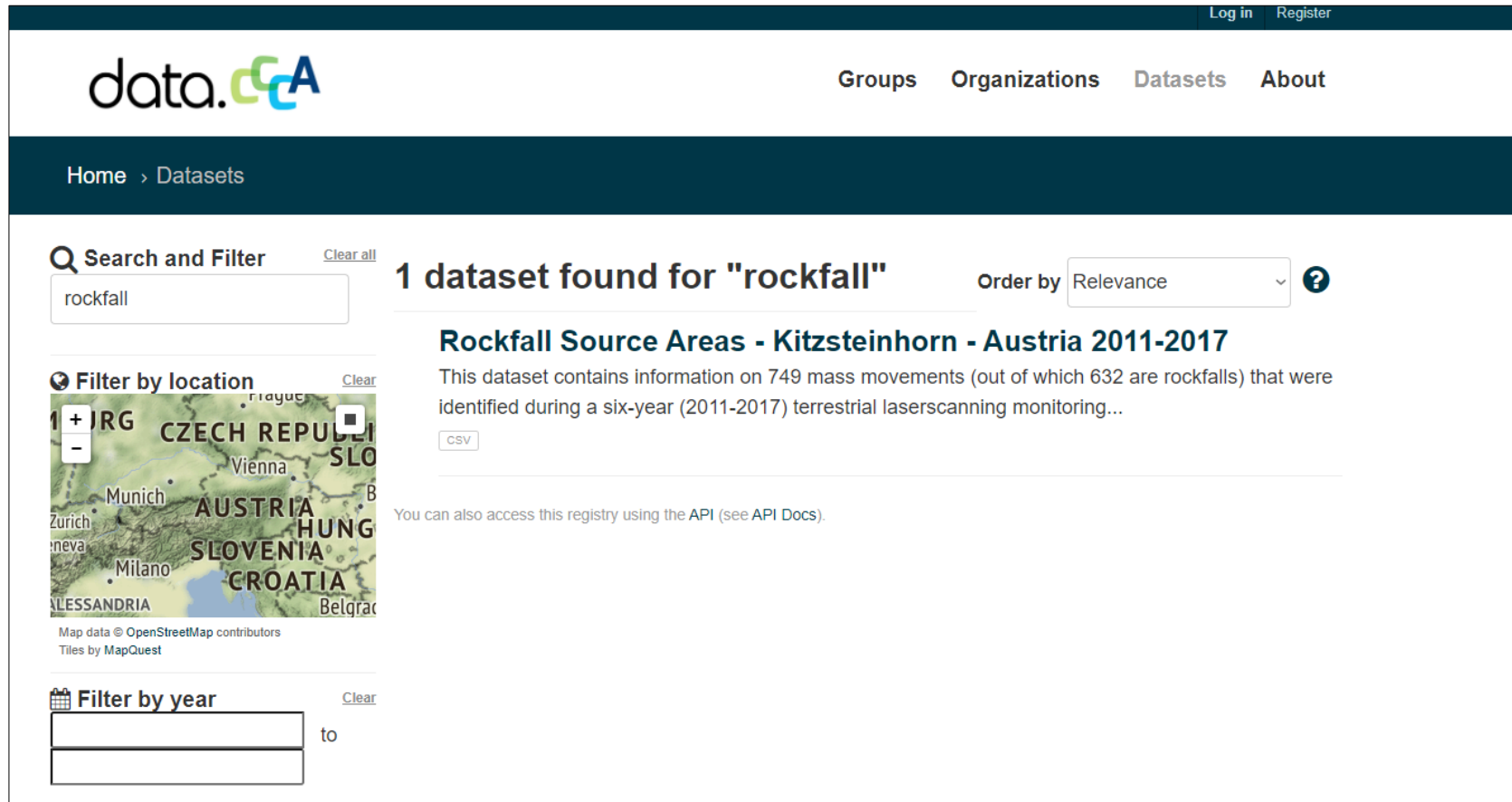
- Dataset Versions** and **Citation** buttons.
- Resources** section with a red box around the dataset entry: "Rockfall Source Areas, Kitzsteinhorn, Austria ...".
- Dataset Metadata** section with a red box around the **Export Metadata** button.
- Basic Information about this dataset** section with a red box around the **Dataset Locator** field containing the URL: <https://hdl.handle.net/20.500.11756/70ef62e8>.

Additional annotations:

- "Dataset" label in red text next to the resource entry.
- "Machine-readable metadata" label in red text below the "Export Metadata" button.
- "Identifier" label in red text next to the dataset locator URL.



# F4. (Meta)data are registered or indexed in a searchable resource



The screenshot shows the data.ccca.ac.at website interface. At the top right, there are links for 'Log in' and 'Register'. The main navigation bar includes 'Groups', 'Organizations', 'Datasets', and 'About'. Below this, a breadcrumb trail shows 'Home > Datasets'. A search bar contains the text 'rockfall' with a 'Clear all' link. To the right of the search bar, it says '1 dataset found for "rockfall"' and 'Order by Relevance'. The search results section features a map of Central Europe with a red location marker over Austria. The map includes labels for 'CZECH REPUBLIC', 'SLOVAKIA', 'AUSTRIA', 'SLOVENIA', 'HUNGARY', 'CROATIA', 'Munich', 'Vienna', 'Zurich', 'Milano', and 'Belgrade'. Below the map, there is a 'Filter by year' section with two input fields and a 'to' label. The main dataset entry is titled 'Rockfall Source Areas - Kitzsteinhorn - Austria 2011-2017' and includes a description: 'This dataset contains information on 749 mass movements (out of which 632 are rockfalls) that were identified during a six-year (2011-2017) terrestrial laserscanning monitoring...'. A 'CSV' download button is visible below the description. At the bottom of the dataset entry, there is a link: 'You can also access this registry using the API (see API Docs)'.

<https://data.ccca.ac.at>



# Accessible

- 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 authorization procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available



# A1.1 The protocol is open, free, and universally implementable

- Retrieve data by ‘clicking on a link’, without specialized tools or communication methods (http, ftp).
- “Anyone with a computer and an internet connection can access at least the metadata”
- HTTP
  - Open – specification of the protocol is known to everyone
  - Free – no need to pay to “use Internet”
- FAIR data ≠ Open data.

OSI model		
Layer	Name	Example protocols
7	Application Layer	HTTP, FTP, DNS, SNMP, Telnet
6	Presentation Layer	SSL, TLS
5	Session Layer	NetBIOS, PPTP
4	Transport Layer	TCP, UDP
3	Network Layer	IP, ARP, ICMP, IPsec
2	Data Link Layer	PPP, ATM, Ethernet
1	Physical Layer	Ethernet, USB, Bluetooth, IEEE802.11

## A1.2 protocol allows for authentication and authorization

- Protected and private data can be FAIR
- Possible types of access
  - Open – everyone has access
  - Shared or restricted – only a selected/ invited group of people can access
  - Closed – only the owner has access

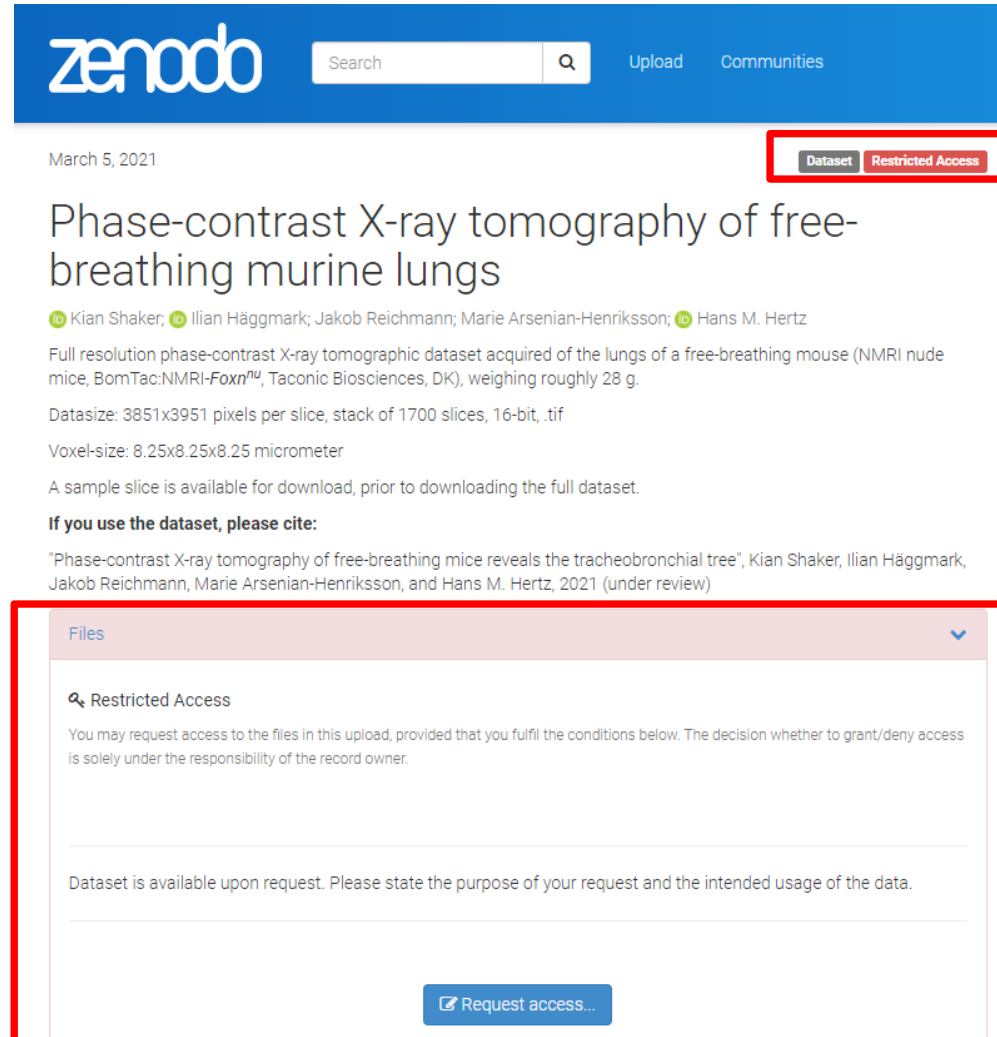


# Accessible - example

The screenshot shows the Zenodo search results page. The top navigation bar includes the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. The user profile 'miksa@ifs.tuwien.ac.at' is visible in the top right. The search results show 1738119 results. The first result is 'Desk-Research Analysis and Identification of SA and Training Tools' by Mateusz Macias, dated March 5, 2021, with 'Project deliverable' and 'Open Access' tags. The second result is 'An embedded device for indoor localization in BLE networks based on a reconfigurable antenna' by Luszczak, Przemyslaw, dated January 15, 2021, with 'Thesis' and 'Open Access' tags. On the left side, there are two filter panels: 'Access Right' and 'File Type'. The 'Access Right' panel is highlighted with a red box and contains the following options: 'Open (1699699)', 'Closed (32706)', 'Restricted (4520)', and 'Embargoed (1194)'. The 'File Type' panel contains options: 'Pdf (892059)', 'Jpg (361789)', 'Png (221819)', 'Html (85105)', and 'Zip (79205)'. The 'Open Access' tags on the search results indicate that the content is freely available.

<https://zenodo.org/search?page=1&size=20&q=>




# Accessible - example



zenodo Search Upload Communities

March 5, 2021 **Dataset** **Restricted Access**

## Phase-contrast X-ray tomography of free-breathing murine lungs

 Kian Shaker;  Ilian Häggmark; Jakob Reichmann; Marie Arsenian-Henriksson;  Hans M. Hertz

Full resolution phase-contrast X-ray tomographic dataset acquired of the lungs of a free-breathing mouse (NMRI nude mice, BomTac:NMRI-Foxn<sup>0/0</sup>, Taconic Biosciences, DK), weighing roughly 28 g.

Datasize: 3851x3951 pixels per slice, stack of 1700 slices, 16-bit, .tif

Voxel-size: 8.25x8.25x8.25 micrometer

A sample slice is available for download, prior to downloading the full dataset.

**If you use the dataset, please cite:**

"Phase-contrast X-ray tomography of free-breathing mice reveals the tracheobronchial tree", Kian Shaker, Ilian Häggmark, Jakob Reichmann, Marie Arsenian-Henriksson, and Hans M. Hertz, 2021 (under review)

Files

**Restricted Access**


You may request access to the files in this upload, provided that you fulfil the conditions below. The decision whether to grant/deny access is solely under the responsibility of the record owner.

Dataset is available upon request. Please state the purpose of your request and the intended usage of the data.

[Request access...](#)

# Tombstone pages (A2)


- Metadata is accessible, even when the data is no longer available

 **HARVARD**  
Dataverse

Add Data ▾ Search ▾ About User Guide Support Sign Up Log In

Harvard Dataverse > **2000 Utah Colleges Exit Poll**

[✉ Contact](#)

 **2000 Utah Colleges Exit Poll**

**Deaccessioned**

David B. Magleby; Howard B. Christensen; Scott D. Grimshaw, 2019, "2000 Utah Colleges Exit Poll", <https://doi.org/10.7910/DVN/2Z9KDF>, Harvard Dataverse, V1, DEACCESSIONED VERSION, UNF:6:ME7YkGved9FxnBuA4Ytw== [fileUNF] [?](#)

**Deaccession Reason**  
User error. Do not use. Look under CSED and Utah Colleges Exit Poll

Versions

Dataset	Summary	Contributors	Published
1.0	Deaccessioned Reason: User error. Do not use. Look under CSED and Utah Colleges Exit Poll	CSED CSED	Dec 30, 2019

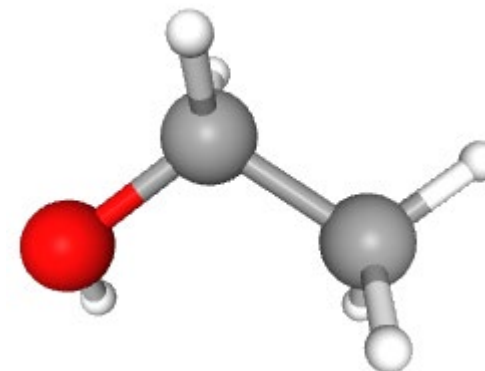
# Interoperable

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable **language for knowledge representation**.
- I2. (Meta)data use **vocabularies** that follow FAIR principles
- I3. (Meta)data include **qualified references** to other (meta)data





- “Data that should be readable for machines without the need for specialised or ad hoc algorithms, translators, or mappings”
- Use:
  - Common formats
    - RDF, JSON (+schema),
    - CSV (+ good README)
  - Well defined/described data models
  - Known representations
    - e.g. InChi Key: IDGUHHHQCWSQLU-UHFFFAOYSA-N
    - Data to have a standard way to describe itself to whomever (machine or human) try to access it without a need for a translator or mappings (common Format: JSON, RDF...)





# Vocabularies

- Vocabulary: Computer-readable file that define meta(data).
- Help evade ambiguities
- “My plane lands in London...” – where exactly?

County	ICAO	IATA	Airport name	Usage
Greater London	EGKB	BQH	<i>London Biggin Hill Airport</i>	Public
Greater London	EGML		<i>Damyns Hall Aerodrome</i>	Private
Greater London	EGLL	LHR	<b><i>Heathrow Airport</i></b>	Public
Greater London	EGWU	NHT	<i>RAF Northolt</i>	Military
Greater London	EGLC	LCY	<b><i>London City Airport</i></b>	Public
Greater London	EGLW		<i>London Heliport</i>	Public

- Controlled vocabularies: IATA and ICAO

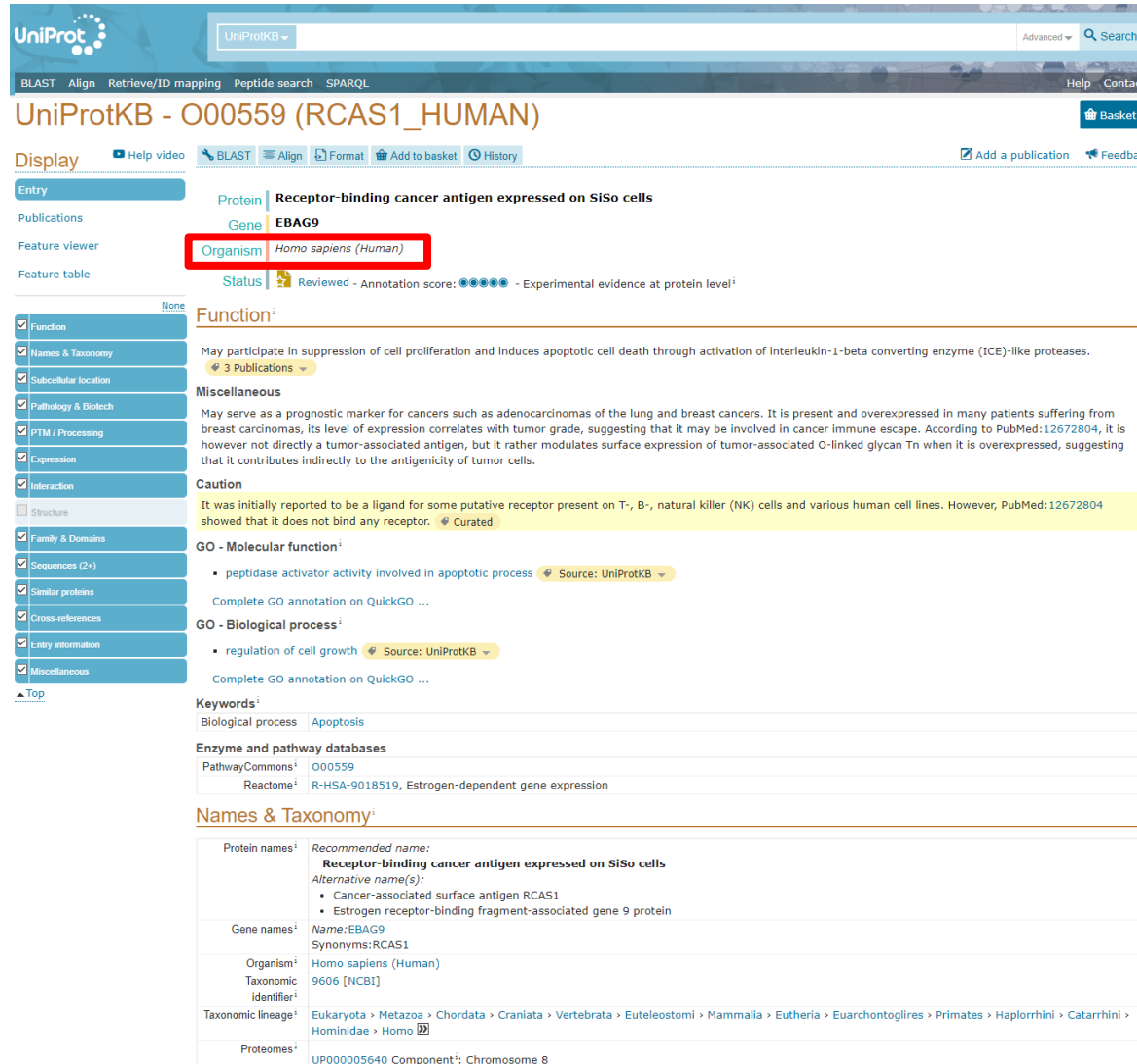
[https://en.wikipedia.org/wiki/List\\_of\\_airports\\_in\\_the\\_United\\_Kingdom\\_and\\_the\\_British\\_Crown\\_Dependencies](https://en.wikipedia.org/wiki/List_of_airports_in_the_United_Kingdom_and_the_British_Crown_Dependencies)

# Vocabularies

- Less time/money spent on data cleaning
  - Different languages
  - Spelling mistakes
  - Abbreviations
  - Capital letters

Vienna	<b>Beč</b> (Croatian, Serbian, older Bulgarian), <b>Beç</b> (older Turkish)*, <b>Bech</b> or <b>Vidnya</b> (Romani), <b>Bécs</b> (Hungarian)*, <b>Bin / Pin</b> - 빈 (Korean), <b>Dunaj</b> (Slovene)*, <b>Fienna</b> (Welsh), <b>Vedunia</b> (Celtic), <b>Vena</b> - Вена (Russian), <b>Videň</b> (Czech)*, <b>Viden' / Videň</b> (Ukrainian)*, <b>Viedeň</b> (Slovak), <b>Viên</b> (Vietnamese), <b>Viena / Vijena/ Виена</b> (Belarusian, Bulgarian, Macedonian), <b>Viena</b> (Catalan*, Lithuanian, Portuguese*, Romanian*, Spanish*, Tagalog*), <b>Vienna</b> (Italian)*, <b>Vienne</b> (French)*, <b>Viénni - Βιέννη</b> (Greek), <b>Vieno</b> (Esperanto), <b>Viin</b> (Estonian), <b>Vin</b> - ווין (Yiddish), <b>Vín</b> (Irish, Icelandic), <b>Vina</b> - וינה (Hebrew), <b>Vínarborg</b> (Icelandic variant), <b>Vindobona</b> (Latin), <b>Vīne</b> (Latvian)*, <b>Viyana</b> (Turkish)*, <b>Vjenë</b> (Albanian), <b>Vjenna</b> (Maltese), <b>Vyana</b> (Azeri), <b>Wean</b> (local Viennese, Austrian and Bavarian dialects)*, <b>Weiyena</b> - 維也納 (Chinese)*, <b>Wene</b> (Afrikaans), <b>Wenen</b> (Dutch)*, <b>Wiedeń</b> (Polish)*, <b>Wien</b> (Danish*, Finnish*, German*, Norwegian*, Swedish*), <b>Wīn</b> - ウィーン (Japanese)*, <b>Wina</b> (Indonesian), <b>فيينا</b> (Arabic), <b>وين</b> (Persian)
--------	--

# Vocabularies (I2)



UniProtKB - O00559 (RCAS1\_HUMAN)

Display [Help video](#) [BLAST](#) [Align](#) [Format](#) [Add to basket](#) [History](#) [Add a publication](#) [Feedback](#)

Entry **Protein** Receptor-binding cancer antigen expressed on SISO cells  
Gene **EBAG9**  
**Organism** *Homo sapiens (Human)*  
Status [Reviewed](#) - Annotation score: ●●●●● - Experimental evidence at protein level<sup>1</sup>

Function<sup>1</sup>  
May participate in suppression of cell proliferation and induces apoptotic cell death through activation of interleukin-1-beta converting enzyme (ICE)-like proteases. [3 Publications](#)

Miscellaneous  
May serve as a prognostic marker for cancers such as adenocarcinomas of the lung and breast cancers. It is present and overexpressed in many patients suffering from breast carcinomas, its level of expression correlates with tumor grade, suggesting that it may be involved in cancer immune escape. According to PubMed:12672804, it is however not directly a tumor-associated antigen, but it rather modulates surface expression of tumor-associated O-linked glycan Tn when it is overexpressed, suggesting that it contributes indirectly to the antigenicity of tumor cells.

Caution  
It was initially reported to be a ligand for some putative receptor present on T-, B-, natural killer (NK) cells and various human cell lines. However, PubMed:12672804 showed that it does not bind any receptor. [Curated](#)

GO - Molecular function<sup>1</sup>  
• peptidase activator activity involved in apoptotic process [Source: UniProtKB](#)

Complete GO annotation on QuickGO ...

GO - Biological process<sup>1</sup>  
• regulation of cell growth [Source: UniProtKB](#)

Complete GO annotation on QuickGO ...

Keywords<sup>1</sup>  
Biological process [Apoptosis](#)

Enzyme and pathway databases  
PathwayCommons<sup>1</sup> O00559  
Reactome<sup>1</sup> R-HSA-9018519, Estrogen-dependent gene expression

Names & Taxonomy<sup>1</sup>

Protein names <sup>1</sup>	<b>Recommended name:</b> <b>Receptor-binding cancer antigen expressed on SISO cells</b> <b>Alternative name(s):</b> <ul style="list-style-type: none"><li>Cancer-associated surface antigen RCAS1</li><li>Estrogen receptor-binding fragment-associated gene 9 protein</li></ul>
Gene names <sup>1</sup>	<b>Name:</b> EBAG9 Synonyms:RCAS1
Organism <sup>1</sup>	<i>Homo sapiens (Human)</i>
Taxonomic identifier <sup>1</sup>	9606 [NCBI]
Taxonomic lineage <sup>1</sup>	Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorhini > Catarrhini > Hominoidea > Hominidae > Homo <a href="#">[?]</a>
Proteomes <sup>1</sup>	UP000005640 Component <sup>1</sup> : Chromosome 8

# Vocabularies (I2)

- Each metadata field has its definition

## Organism

**Last modified** April 10, 2018

This subsection of the [Names and taxonomy](#) section provides information on the name(s) of the organism that is the source of the protein sequence.

The organism designation consists of the Latin scientific name, usually composed of the genus and species names (the binomial system developed by Linnaeus), followed optionally by the English common name and a synonym.

Examples: *Bacillus subtilis*, *Homo sapiens* (Human), *Cardamine pratensis* (Cuckoo flower) (Alpine bitter cross)

The synonym can be a common name in English (or in Latin in the case of some historical legacy names).

Example: *Radianthus magnifica* (Magnificent sea anemone) (*Heteractis magnifica*)

In the case of viruses, the designation does not follow the binomial system. The English common name is used as the scientific name, sometimes followed by an acronym. When possible, viruses are named according to the nomenclature of the International Committee on Taxonomy of Viruses (ICTV).

Examples: Human immunodeficiency virus type 1 (isolate BRU/LAI group M subtype B) (HIV-1) , Influenza A virus (strain A/Aichi/2/1968 H3N2)

The organism name can differ from that given by the international nucleotide sequence databases for the same taxon. This is mainly due to our efforts in providing the most descriptive common names and synonyms to our users.

Note that the proteome for a given organism, when available, can be accessed through the [proteomes](#) page of our website.

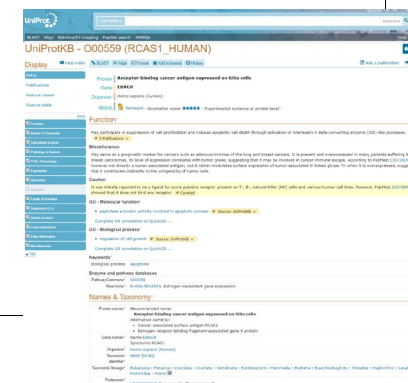
## Related documents

[Taxonomy](#)

[Controlled vocabulary of species](#)

[What are proteomes?](#)

[What are reference proteomes?](#)



# Vocabularies (I2)

**Taxonomy - Homo sapiens (Human)** (SPECIES)

[Map to](#) [Format](#)

UniProtKB (194,609)  
 Reviewed (20,397)  
 Swiss-Prot  
 Unreviewed (174,212)  
 TrEMBL  
 Proteomes (3)

Mnemonic	HUMAN
Taxon identifier	9606
Scientific name	Homo sapiens
Taxonomy navigation	<ul style="list-style-type: none"> <li>&gt; Homo</li> <li>&gt; [Choose one] <span>All lower taxonomy nodes (2)</span></li> </ul>
Common name	Human
Synonym	-
Other names	<ul style="list-style-type: none"> <li>&gt;Home sapiens</li> <li>&gt;Homo sampiens</li> <li>&gt;Homo sapeins</li> <li>&gt;Homo sapieins</li> <li>&gt;Homo sapian</li> <li>&gt;Homo sapians</li> <li>More »</li> </ul>
Rank	SPECIES
Lineage	<ul style="list-style-type: none"> <li>&gt; cellular organisms</li> <li>&gt; Eukaryota</li> <li>&gt; Opisthokonta</li> <li>&gt; Metazoa</li> <li>&gt; Eumetazoa</li> <li>&gt; Bilateria</li> <li>&gt; Deuterostomia</li> <li>&gt; Chordata</li> <li>&gt; Vertebrata</li> <li>&gt; Gnathostomata</li> <li>&gt; Teleostomi</li> <li>&gt; Euteleostomi</li> <li>&gt; Sarcopterygii</li> <li>&gt; Dipnotetrapodomorpha</li> <li>&gt; Tetrapoda</li> <li>&gt; Amniota</li> <li>&gt; Mammalia</li> <li>&gt; Theria</li> <li>&gt; Eutheria</li> <li>&gt; Boreoeutheria</li> <li>&gt; Euarctontoglires</li> <li>&gt; Primates</li> <li>&gt; Haplorrhini</li> <li>&gt; Simiiformes</li> <li>&gt; Catarrhini</li> <li>&gt; Hominoidea</li> <li>&gt; Hominidae</li> <li>&gt; Homininae</li> <li>&gt; Homo</li> </ul>

Each metadata **value** comes from a controlled vocabulary – no free form answers.

UniProtKB - O00559 (RCAS1\_HUMAN)

**Display** [View](#) [Print](#) [Download](#) [Share](#)

**Name** Receptor binding cancer antigen expressed on tille cells

**Publications** [View](#) **EBAC3**

**History** [View](#) Home sapiens (human)

**History** [View](#) **Reviewed** annotation score: **★★★★** Experimental evidence at protein level

**Function** [View](#) [Download](#)  
 May participate in suppression of cell proliferation and induces apoptotic cell death through activation of caspase-3-like proteases.

**Keywords** [View](#)  
 May serve as a prognostic marker for cancers such as adenocarcinoma of the lung and breast cancer. It is present and overexpressed in many patients suffering from breast carcinoma. Its level of expression correlates with tumor grade, suggesting that it may be involved in cancer metastasis. According to PubMed (217279), it is however not directly a tumor associated antigen, but it rather modulates surface expression of tumor associated G-protein. Its when it is overexpressed, suggesting that it contributes directly to the prognosis of cancer cells.

**Caution** [View](#)  
 It has been reported to be a ligand for some positive regulator present on T-, B-, natural killer (NK) cells and various human cell lines. However, PubMed (137264) showed that it does not bind any receptor. **Caution**

**Biological process** [View](#)  
 • positive regulator activity involved in apoptotic process **KB** [UniProtKB](#)

**Cellular component** [View](#)  
 • integral cell membrane protein **KB** [UniProtKB](#)

**Biological process** [View](#)  
 • regulation of cell growth **KB** [UniProtKB](#)

**Cellular component** [View](#)  
 • integral cell membrane protein **KB** [UniProtKB](#)

**Keywords**  
 Biological process: apoptosis

**Enzyme and pathway databases**  
 UniProtKB: [UniProtKB](#)

**Names & Taxonomy**  
 UniProtKB: [UniProtKB](#)

**Public names** [View](#)  
 UniProtKB: [UniProtKB](#)

**Other names** [View](#)  
 UniProtKB: [UniProtKB](#)

**Gene names** [View](#)  
 UniProtKB: [UniProtKB](#)

**Organism** [View](#)  
 UniProtKB: [UniProtKB](#)

**Species** [View](#)  
 UniProtKB: [UniProtKB](#)

**Protein** [View](#)  
 UniProtKB: [UniProtKB](#)

# Qualified References (I3)

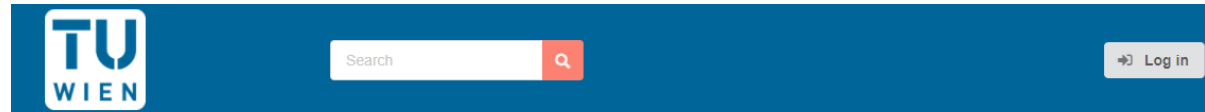
- Meaningful links to describe connections
  - Dataset X was *derived from* dataset Y
  - Dataset Y was *produced* using code Z
- Standard relations define by Data Cite
  - <https://schema.datacite.org/meta/kernel-4.3/>
- Use persistent identifiers

```
<language>en</language>
<resourceType resourceTypeGeneral="Workflow">Software</resourceType>
<relatedIdentifiers>
  <relatedIdentifier relatedIdentifierType="DOI"
    relationType="IsReferencedBy">10.5072/2047-217X-1-1</relatedIdentifier>
  <relatedIdentifier relatedIdentifierType="DOI"
    relationType="Compiles">10.5072/100038</relatedIdentifier>
</relatedIdentifiers>
<sizes>
  <size>31 MB</size>
</sizes>
```

<https://schema.datacite.org/meta/kernel-4.3/example/datacite-example-workflow-v4.xml>

IsContinuedBy  
Continues  
IsDescribedBy  
Describes  
HasMetadata  
IsMetadataFor  
HasVersion  
IsVersionOf  
IsNewVersionOf  
IsPreviousVersionOf  
IsPartOf  
HasPart  
IsReferencedBy  
References  
IsDocumentedBy  
Documents  
IsCompiledBy  
Compiles  
IsVariantFormOf  
IsOriginalFormOf  
IsIdenticalTo  
IsReviewedBy  
Reviews  
IsDerivedFrom  
IsSourceOf  
IsRequiredBy  
Requires  
IsObsoletedBy  
Obsoletes

# Qualified References (I3)



January 19, 2021 | Version 1.0

Dataset  Open Access

Versions

Version 1.0  
DOI: 10.48436/tkkfs-11b75

## European Sentinel-1 Forest Type and Tree Cover Density Maps

 Dostalova, Alena <sup>1</sup>; Cao, Senmao <sup>1,2</sup>;  Wagner, Wolfgang <sup>1,2</sup> 

### Description

This dataset was generated by the [TU Wien Department of Geodesy and](#)

European Sentinel-1 forest type and tree cover density maps represent first order Sentinel-1 C-Band Synthetic Aperture Radar (SAR) backscatter data. For the European continent with 10 m and 100 m sampling for forest type and tree cover density maps derived using the method described in <https://www.tandfonline.com/doi/full/10.1080/01431161.2018.1479788>

The forest type map shows the dominant forest type class (coniferous, broadleaf, or mixed) and the percentage of forest canopy cover within the 100 m pixel.

Please be referred to our peer-reviewed article at <https://doi.org/10.3390/rs13030337> for an assessment across Europe.

### Dataset Record

The forest type and tree cover density maps are sampled at 10 m and 100 m and georeferenced to the Equi7Grid and divided into square tiles of 100km x 100km. The maps consist of 728 tiles over the European continent, with data volumes of 100MB per tile.

The tiles' file-format is a LZW-compressed GeoTIFF holding 16-bit integer values and georeference. Compatibility with common geographic information system libraries as GDAL is given.

In this repository, we provide each forest map as tiles, whereas two zip files are available for download below.

### Code Availability

For the usage of the [Equi7Grid](#) we provide data and tools via the python package available on GitHub at <https://github.com/TUW-GEO/Equi7Grid>. More details on the grid reference can be found in <https://www.sciencedirect.com/science/article/pii/S0098300414001629>.

### Acknowledgements

The computational results presented have been achieved using the Vienna Scientific Cluster (VSC).

### Details

#### Licenses

#### Resource type

Dataset

#### Formats

application/x-geotiff

#### Related identifiers

##### isreferencedby

10.3390/rs13030337 ( doi )

Paper citing this dataset

##### issupplementto

10.5281/zenodo.3515933 ( doi )

<https://github.com/TUW-GEO/Equi7Grid> ( url )

Code

##### references

10.1080/01431161.2018.1479788 ( doi )

Paper describing the method to produce this dataset

10.1016/j.cageo.2014.07.005 ( doi )

<https://researchdata.dl.hpc.tuwien.ac.at/records/tkkfs-11b75>

# Qualified References (13)

README.rst

## [Equi7Grid](#)

build passing coverage 32% pypi package 0.0.12 docs passing

A python class for working with Equi7Grid - how to convert to - how to use the tiling system - etc.

It's a python package that handles the geometric and geographic operations of a gridded and tiled projection system. It was designed for data cubes ingesting satellite imagery and builds the basis for the Equi7Grid (see <https://github.com/TUW-GEO/Equi7Grid>).

A detailed documentation on the Equi7Grid definition is at:

`~/docs/doc_files/`

Overlays for visualisation in Google Earth can be found here:

`~/docs/doc_files/google_earth_overlays/`

## Citation

DOI [10.5281/zenodo.1048530](https://doi.org/10.5281/zenodo.1048530)

If you use the software in a publication then please cite it using the Zenodo DOI. Be aware that this badge links to the latest package version.

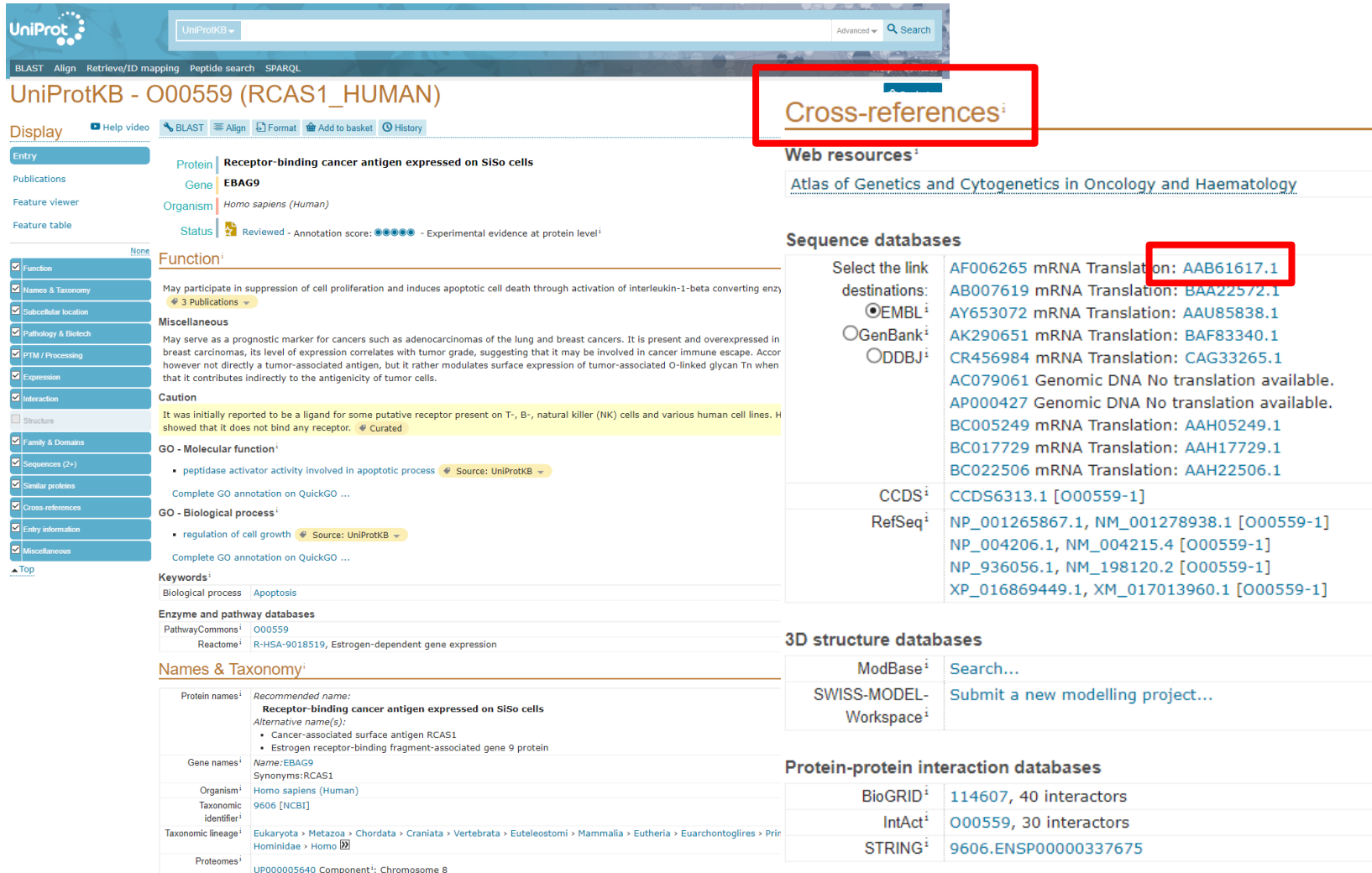
Please select your specific version at <https://doi.org/10.5281/zenodo.1048530> to get the DOI of that version. You should normally always use the DOI for the specific version of your record in citations. This is to ensure that other researchers can access the exact research artefact you used for reproducibility.

You can find additional information regarding DOI versioning at <http://help.zenodo.org/#versioning>

<https://github.com/TUW-GEO/Equi7Grid>



# Qualified References (I3)



**UniProtKB - O00559 (RCAS1\_HUMAN)**

**Cross-references<sup>i</sup>**

**Web resources<sup>i</sup>**

[Atlas of Genetics and Cytogenetics in Oncology and Haematology](#)

**Sequence databases**

Select the link destinations:	AF006265 mRNA Translation: <b>AAB61617.1</b>
<input checked="" type="radio"/> EMBL <sup>i</sup>	AB007619 mRNA Translation: BAA22572.1
<input type="radio"/> GenBank <sup>i</sup>	AY653072 mRNA Translation: AAU85838.1
<input type="radio"/> DDBJ <sup>i</sup>	AK290651 mRNA Translation: BAF83340.1
	CR456984 mRNA Translation: CAG33265.1
	AC079061 Genomic DNA No translation available.
	AP000427 Genomic DNA No translation available.
	BC005249 mRNA Translation: AAH05249.1
	BC017729 mRNA Translation: AAH17729.1
	BC022506 mRNA Translation: AAH22506.1
CCDS <sup>i</sup>	CCDS6313.1 [O00559-1]
RefSeq <sup>i</sup>	NP_001265867.1, NM_001278938.1 [O00559-1] NP_004206.1, NM_004215.4 [O00559-1] NP_936056.1, NM_198120.2 [O00559-1] XP_016869449.1, XM_017013960.1 [O00559-1]

**3D structure databases**

ModBase <sup>i</sup>	<a href="#">Search...</a>
SWISS-MODEL-Workspace <sup>i</sup>	<a href="#">Submit a new modelling project...</a>

**Protein-protein interaction databases**

BioGRID <sup>i</sup>	114607, 40 interactors
IntAct <sup>i</sup>	O00559, 30 interactors
STRING <sup>i</sup>	9606.ENSP00000337675

**Function<sup>i</sup>**

May participate in suppression of cell proliferation and induces apoptotic cell death through activation of interleukin-1-beta converting enzyme. [3 Publications](#)

**Miscellaneous**

May serve as a prognostic marker for cancers such as adenocarcinomas of the lung and breast cancers. It is present and overexpressed in breast carcinomas, its level of expression correlates with tumor grade, suggesting that it may be involved in cancer immune escape. Accor however not directly a tumor-associated antigen, but it rather modulates surface expression of tumor-associated O-linked glycan Tn when that it contributes indirectly to the antigenicity of tumor cells.

**Caution**

It was initially reported to be a ligand for some putative receptor present on T-, B-, natural killer (NK) cells and various human cell lines. H showed that it does not bind any receptor. [Curated](#)

**GO - Molecular function<sup>i</sup>**

- peptidase activator activity involved in apoptotic process [Source: UniProtKB](#)

Complete GO annotation on QuickGO ...

**GO - Biological process<sup>i</sup>**

- regulation of cell growth [Source: UniProtKB](#)

Complete GO annotation on QuickGO ...

**Keywords<sup>i</sup>**

Biological process | Apoptosis

**Enzyme and pathway databases**

PathwayCommons <sup>i</sup>	O00559
Reactome <sup>i</sup>	R-HSA-9018519, Estrogen-dependent gene expression

**Names & Taxonomy<sup>i</sup>**

Protein names <sup>i</sup>	<b>Recommended name:</b> Receptor-binding cancer antigen expressed on SiSo cells <b>Alternative name(s):</b> <ul style="list-style-type: none"><li>Cancer-associated surface antigen RCAS1</li><li>Estrogen receptor-binding fragment-associated gene 9 protein</li></ul>
Gene names <sup>i</sup>	Name: EBAG9 Synonyms: RCAS1
Organism <sup>i</sup>	Homo sapiens (Human)
Taxonomic identifier <sup>i</sup>	9606 [NCBI]
Taxonomic lineage <sup>i</sup>	Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Hominidae > Homo
Proteomes <sup>i</sup>	UP000005640 Component: Chromosome 8

# Reusable

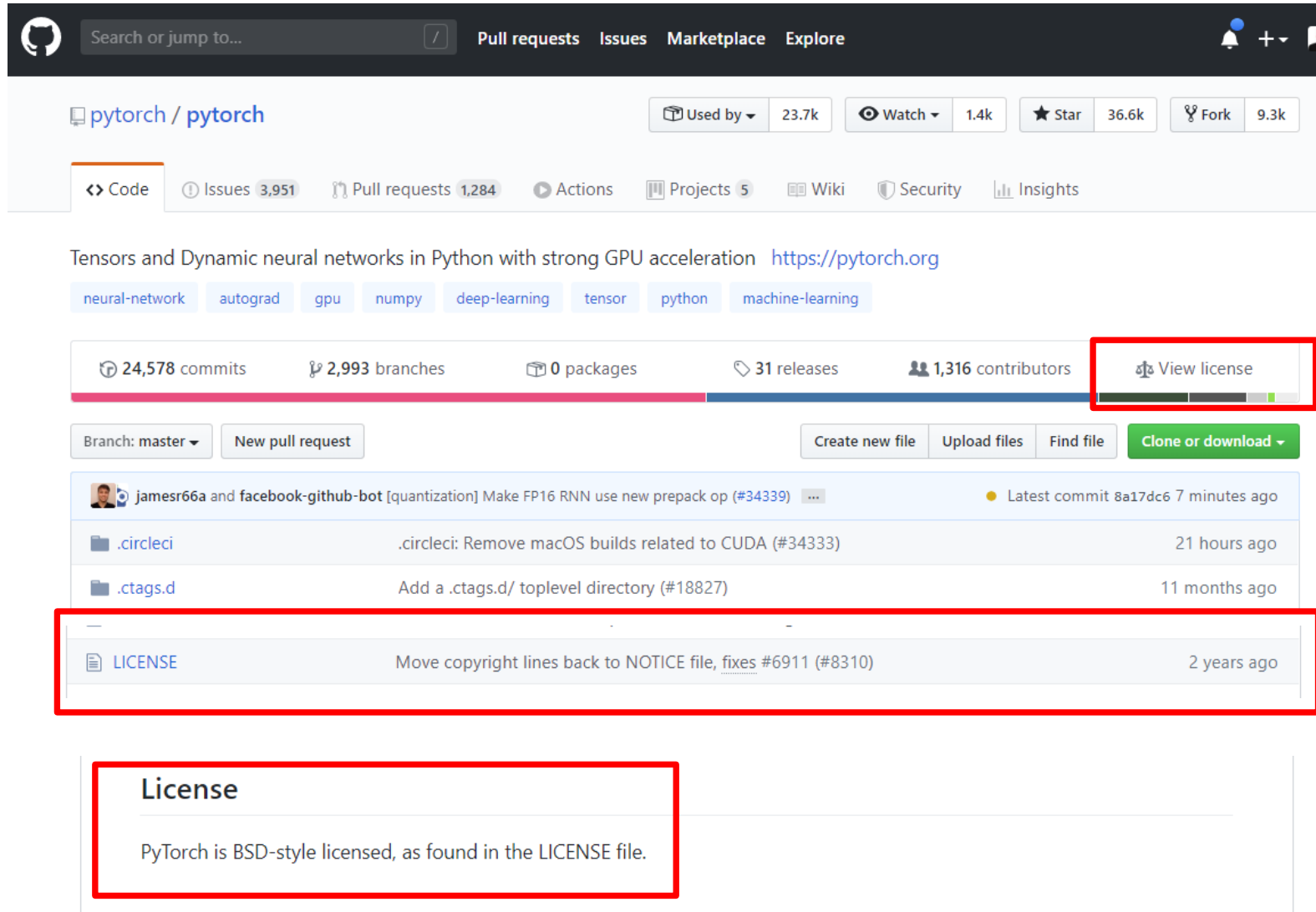
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
  - R1.1. (Meta)data are released with a clear and accessible data usage **license**
  - R1.2. (Meta)data are associated with detailed **provenance**
  - R1.3. (Meta)data meet domain-relevant **community standards**



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

- Public repository on GitHub
  - May suggest that authors are willing to share code
- No license
  - no possibility for reuse
  - can only be viewed (only because terms of use enforce that)
- Code without a license is like an object in a museum
  - You can watch and admire it, but you cannot touch it!

# License (R1.1)



The screenshot shows the GitHub repository for PyTorch. The repository name is `pytorch / pytorch`. It has 23.7k users, 1.4k watches, 36.6k stars, and 9.3k forks. The repository description is "Tensors and Dynamic neural networks in Python with strong GPU acceleration" with the URL <https://pytorch.org>. The repository has 24,578 commits, 2,993 branches, 0 packages, 31 releases, and 1,316 contributors. A red box highlights the "View license" button. Below the repository information, there is a list of recent commits. A red box highlights a commit by `JamesR66a` and `facebook-github-bot` titled "Move copyright lines back to NOTICE file, fixes #6911 (#8310)" from 2 years ago. Below the commit list, there is a section titled "License" with a red box around it. The text in this section reads: "PyTorch is BSD-style licensed, as found in the LICENSE file."

Search or jump to... Pull requests Issues Marketplace Explore

Used by 23.7k Watch 1.4k Star 36.6k Fork 9.3k

Code Issues 3,951 Pull requests 1,284 Actions Projects 5 Wiki Security Insights

Tensors and Dynamic neural networks in Python with strong GPU acceleration <https://pytorch.org>

neural-network autograd gpu numpy deep-learning tensor python machine-learning

24,578 commits 2,993 branches 0 packages 31 releases 1,316 contributors View license

Branch: master New pull request Create new file Upload files Find file Clone or download

JamesR66a and facebook-github-bot [quantization] Make FP16 RNN use new prepack op (#34339) Latest commit 8a17dc6 7 minutes ago

.circleci .circleci: Remove macOS builds related to CUDA (#34333) 21 hours ago

.ctags.d Add a .ctags.d/ toplevel directory (#18827) 11 months ago

LICENSE Move copyright lines back to NOTICE file, fixes #6911 (#8310) 2 years ago

**License**

PyTorch is BSD-style licensed, as found in the LICENSE file.

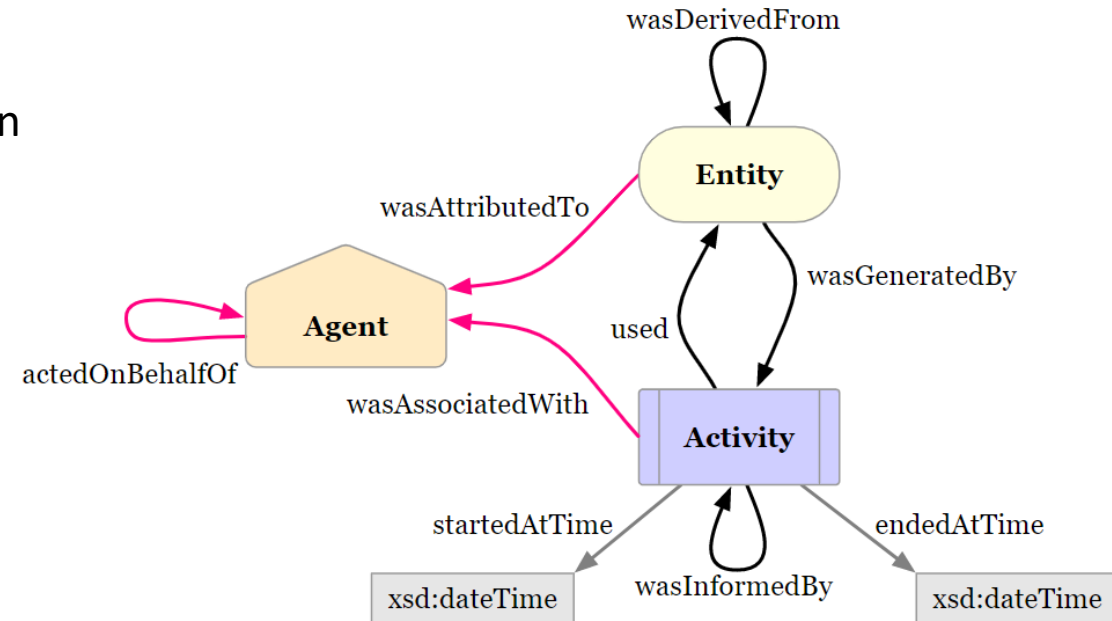
# R1.2 (Meta)data are associated with detailed provenance

- Provenance
  - Describes origin of data
  - Who? What? When? How?
- Supports evaluation and can build trust in data
  - ‘Officially, North Korea claims to have identified zero cases of COVID-19 inside its territory’ <https://www.npr.org/sections/goatsandsoda/2020/02/20/807027901/north-korea-claims-zero-cases-of-coronavirus-infection-but-experts-are-skeptical?t=1615196582563>



# R1.2 (Meta)data are associated with detailed provenance

- PROV-O: The PROV Ontology
  - Machine-actionable way to express provenance
- Licenses (level of compliance for repository preservation /policy/strategy/action plan)
- To what extent the data is useful for others.
- Rich Metadata that describe the context under which the data was generated.
- What usage rights do you attach to your data?
- Provenance means to describe origin of data (Who? What? When? How? ) → Ontology



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



- Who is the “community” ?
- What is the “standard” ?
  - English vs other languages
- Metadata
  - Domain independent
    - e.g. Dublin Core
  - Domain specific
    - e.g. EXIF for images
- Sometimes no common standard exist
  - Good documentation and README
- There is no universal guideline – it always depends!

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

- Does not have to be a standard for everyone!

## COVID-19 Gs

**3G-Regel:** 3G steht für geimpft, genesen oder getestet. Beim Test gelten SOWOHL Antigentests ALS AUCH PCR-Tests. Sind also 3G für einen Zutritt nötig, muss man ENTWEDER geimpft, genesen ODER getestet sein.

**2G-Regel:** 2G steht für geimpft oder genesen. Werden für einen Zutritt 2G gefordert, dann muss man ENTWEDER geimpft ODER genesen sein.

**2G+ oder 2Gplus** steht für geimpft oder genesen PLUS getestet, und zwar mit einem PCR-Test. Wird für einen Zutritt die 2G+-Regel verlangt, muss man also ENTWEDER geimpft ODER genesen UND ZUSÄTZLICH getestet sein, und zwar mit einem PCR-Test.

**2,5G-Regel:** 2,5G steht für geimpft oder genesen oder PCR-getestet. Wird für einen Zutritt also 2,5G verlangt, dann muss man ENTWEDER geimpft, genesen ODER PCR-getestet sein.

## NATO Phonetic Alphabet

A	Alpha	N	November
B	Bravo	O	Oscar
C	Charlie	P	Papa
D	Delta	Q	Quebec
E	Echo	R	Romeo
F	Foxtrot	S	Sierra
G	Golf	T	Tango
H	Hotel	U	Uniform
I	India	V	Victor
J	Juliet	W	Whiskey
K	Kilo	X	X-ray
L	Lima	Y	Yankee
M	Mike	Z	Zulu



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

- Good documentation supports reuse
  - Removes ambiguities (especially where there are no common controlled vocabularies or others standards)
- Example
  - Confirmed cases of COVID-19: testing date vs reporting date

Indicators	Definition
Tests	Cumulative number of tests carried out for SARS-CoV-2, from 27 February 2020 up to and including the reporting date. Responsible for data consolidation: Office of the respective federal state government (Land), data status: morning of the reporting day.
Laboratory-confirmed cases	Cumulative number of laboratory-confirmed cases of SARS-CoV-2 infection (sum of "Active cases", "Recovered cases" and "Deceased cases") with laboratory diagnosis date since 27.02.2020 up to and including the <b>reporting date</b> .
Active cases	Cumulative number of laboratory-confirmed cases of SARS-CoV-2 infection with laboratory diagnosis date from 27.02.2020 up to and including the reporting date, which have not been classified as "recovered" or "deceased" on the reporting date.
Recovered cases	Cumulative number of laboratory-confirmed cases of SARS-CoV-2 infection with laboratory diagnosis date from 27.02.2020 up to and including the report date, which are classified as "recovered" on the report date. Definition of "recovered" (since 9 July): in the case of home care, 10-day home isolation after the onset of symptoms or laboratory diagnosis; in case of severe disease progression, the earliest 10 days after onset of symptoms, at least 48 hours without symptoms AND the following result by RT-PCR according to the Charité protocol: no nucleic acid detection of beta-coronavirus SARS-CoV-2 or nucleic acid detection of beta-coronavirus SARS-CoV-2 at a Ct value of more than 30. Further details can be found in the recommendation for the release of COVID-19 cases, recommendation for the release of COVID-19 cases from isolation.
Deceased cases	Cumulative number of laboratory-confirmed cases of SARS-CoV-2 infection with a laboratory diagnosis date from 27.02.2020 up to and including the report date, which are classified as "deceased" on the report date. Definition of "deceased": COVID-19 death is defined, for surveillance purposes, as one laboratory-confirmed case of COVID-19 resulting

[https://covid19-dashboard.ages.at/basisinfo\\_en.html?!=en](https://covid19-dashboard.ages.at/basisinfo_en.html?!=en)

# To summarize FAIR Principles...

- Data can be **Findable** by adding metadata and a persistent identifier.
- Data can be **Accessible** by defining who can access data and how, but keep in mind that, if you can't publish data openly, you should provide access to the metadata, for example through a data repository.
- Data can be **Interoperable** by using common standards and open data formats.
- Data can be **Reusable** by adding documentation that help others understand data and an appropriate data license that determines how data can be reused.



# How repositories support FAIRness

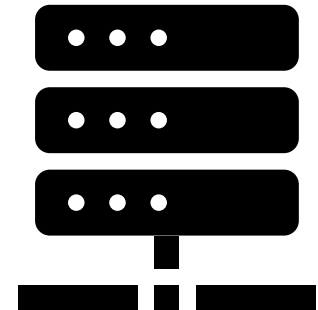


# Repositories and FAIRness

**Data repositories** are key in putting the FAIR principles into practice.

Not only do they enable findability and access, but they also provide persistent identifiers, documentation and metadata, thus fostering reusability for humans and machines.

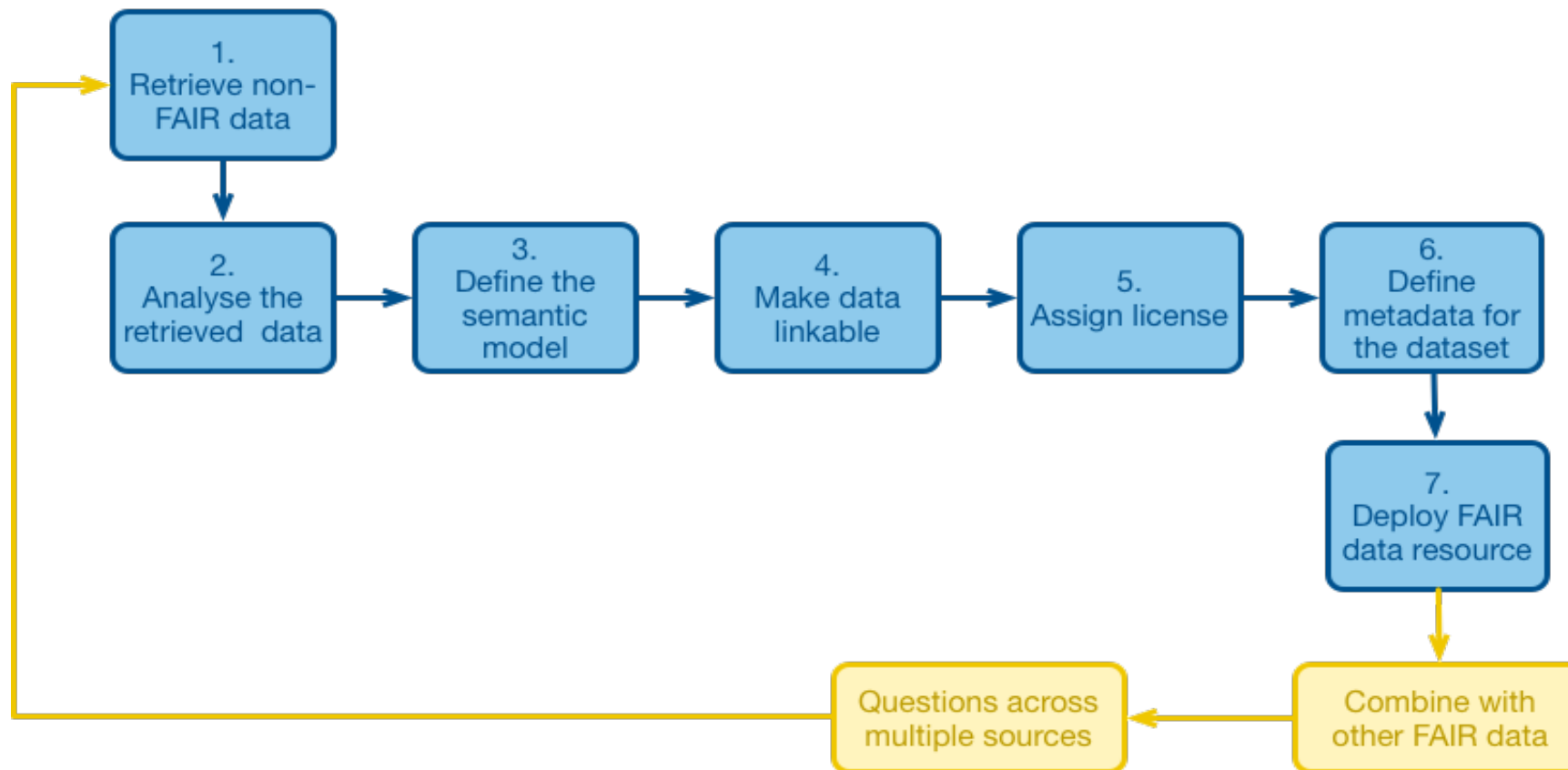
- A “form” needs to be filled –metadata by default.
- The form complies with a specific metadata standard.
- Metadata will then become machine-actionable and searchable in an online resource.
- A persistent identifier for the data is automatically generated.
- References to other data or metadata can be included.
- Authentication and authorization procedures are in place.
- Access can be regulated from closed to open.
- The provision of machine-readable licenses enhances the reusability of the data.
- The use of standards and controlled vocabularies is enforced.
- Interfaces for external services like OAI-PMH allow harvesting of metadata for stored records.





# How to FAIRify data?

# FAIRification Process

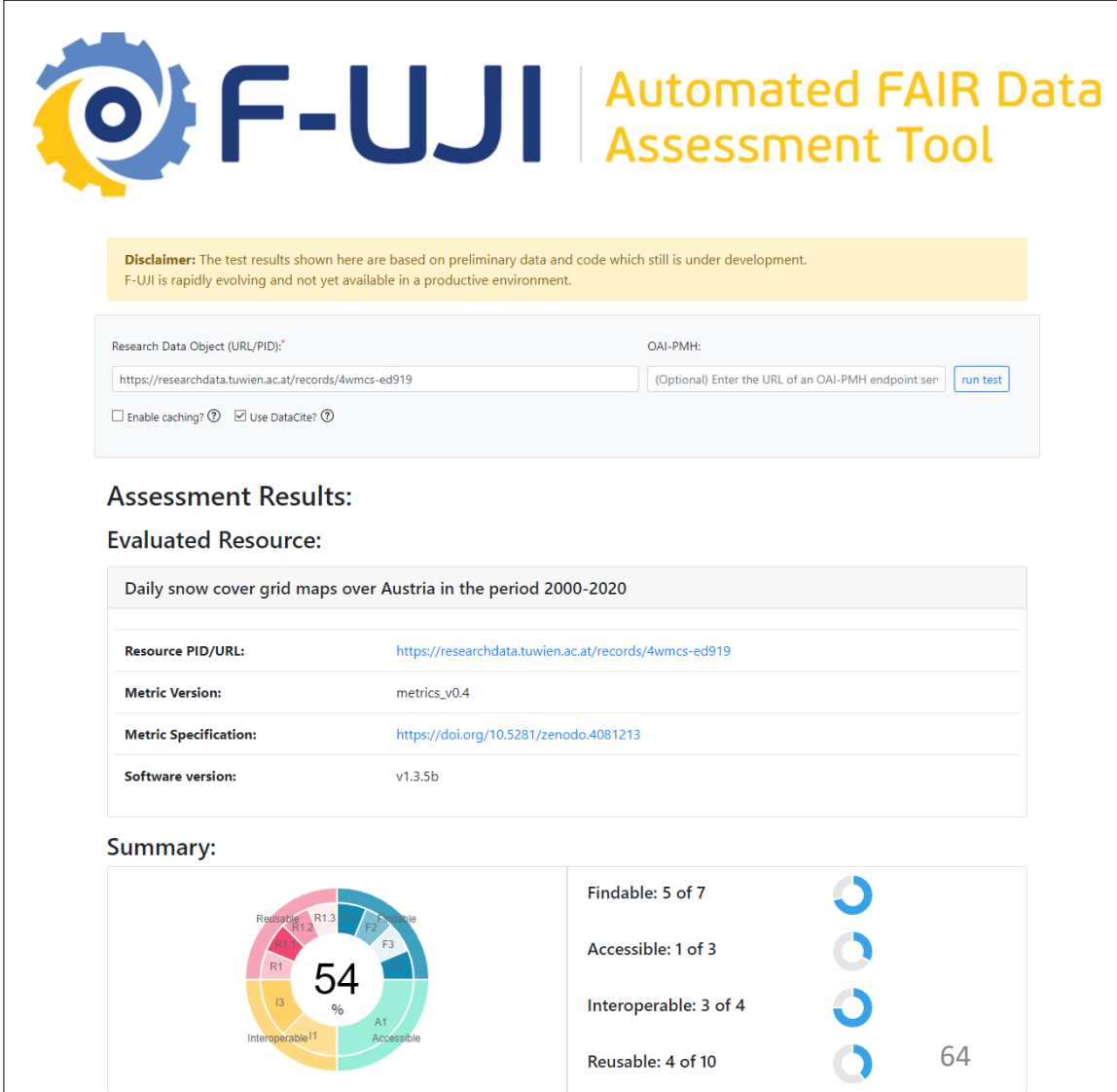


[FAIRification Process - GO FAIR \(go-fair.org\)](https://go-fair.org)

# FAIR Assessment - F-UJI Tool

- F-UJI is a web service to programmatically assess FAIRness of research data objects based on metrics developed by the FAIRsFAIRproject.

<https://f-uji.net>



**F-UJI** | Automated FAIR Data Assessment Tool

**Disclaimer:** The test results shown here are based on preliminary data and code which still is under development. F-UJI is rapidly evolving and not yet available in a productive environment.

Research Data Object (URL/PID):  OAI-PMH:

Enable caching?  Use DataCite?

**Assessment Results:**

**Evaluated Resource:**

Daily snow cover grid maps over Austria in the period 2000-2020

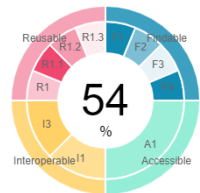
**Resource PID/URL:** <https://researchdata.tuwien.ac.at/records/4wmcs-ed919>

**Metric Version:** metrics\_v0.4

**Metric Specification:** <https://doi.org/10.5281/zenodo.4081213>

**Software version:** v1.3.5b

**Summary:**



Findable: 5 of 7	
Accessible: 1 of 3	
Interoperable: 3 of 4	
Reusable: 4 of 10	

64

**Report:**
**Findable**

 FsF-F1-01D - Data is assigned a globally unique identifier. ✔

 FsF-F1-02D - Data is assigned a persistent identifier. ✔

 FsF-F2-01M - Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability. ✔

 FsF-F3-01M - Metadata includes the identifier of the data it describes. ✘

 FsF-F4-01M - Metadata is offered in such a way that it can be retrieved programmatically. ✔
**Accessible**

 FsF-A1-01M - Metadata contains access level and access conditions of the data. ✘

 FsF-A1-03D - Data is accessible through a standardized communication protocol. ✘

 FsF-A1-02M - Metadata is accessible through a standardized communication protocol. ✔
**Interoperable**

 FsF-I1-01M - Metadata is represented using a formal knowledge representation language. ✔

 FsF-I1-02M - Metadata uses semantic resources ✘

 FsF-I3-01M - Metadata includes links between the data and its related entities. ✔
**Reusable**

 FsF-R1-01MD - Metadata specifies the content of the data. ✘

 FsF-R1.1-01M - Metadata includes license information under which data can be reused. ✔

 FsF-R1.2-01M - Metadata includes provenance information about data creation or generation. ✔

 FsF-R1.3-01M - Metadata follows a standard recommended by the target research community of the data. ✔

 FsF-R1.3-02D - Data is available in a file format recommended by the target research community. ✘
**Metric tests:**

Test:	Test name:	Result:
<b>FsF-R1-01MD-1</b>	Minimal information about available data content is given in metadata	✔
<b>FsF-R1-01MD-1a</b>	Resource type (e.g. dataset) is given in metadata	✔
<b>FsF-R1-01MD-1b</b>	Information about data content (e.g. links) is given in metadata	✘
<b>FsF-R1-01MD-2</b>	Verifiable data descriptors (file info, measured variables or observation types) are specified in metadata	✘
<b>FsF-R1-01MD-2a</b>	File size and type information are specified in metadata	✘
<b>FsF-R1-01MD-2b</b>	Measured variables or observation types are specified in metadata	✘
<b>FsF-R1-01MD-3</b>	Data content matches file type and size specified in metadata	✘
<b>FsF-R1-01MD-4</b>	Data content matches measured variables or observation types specified in metadata	✘

**Debug:**

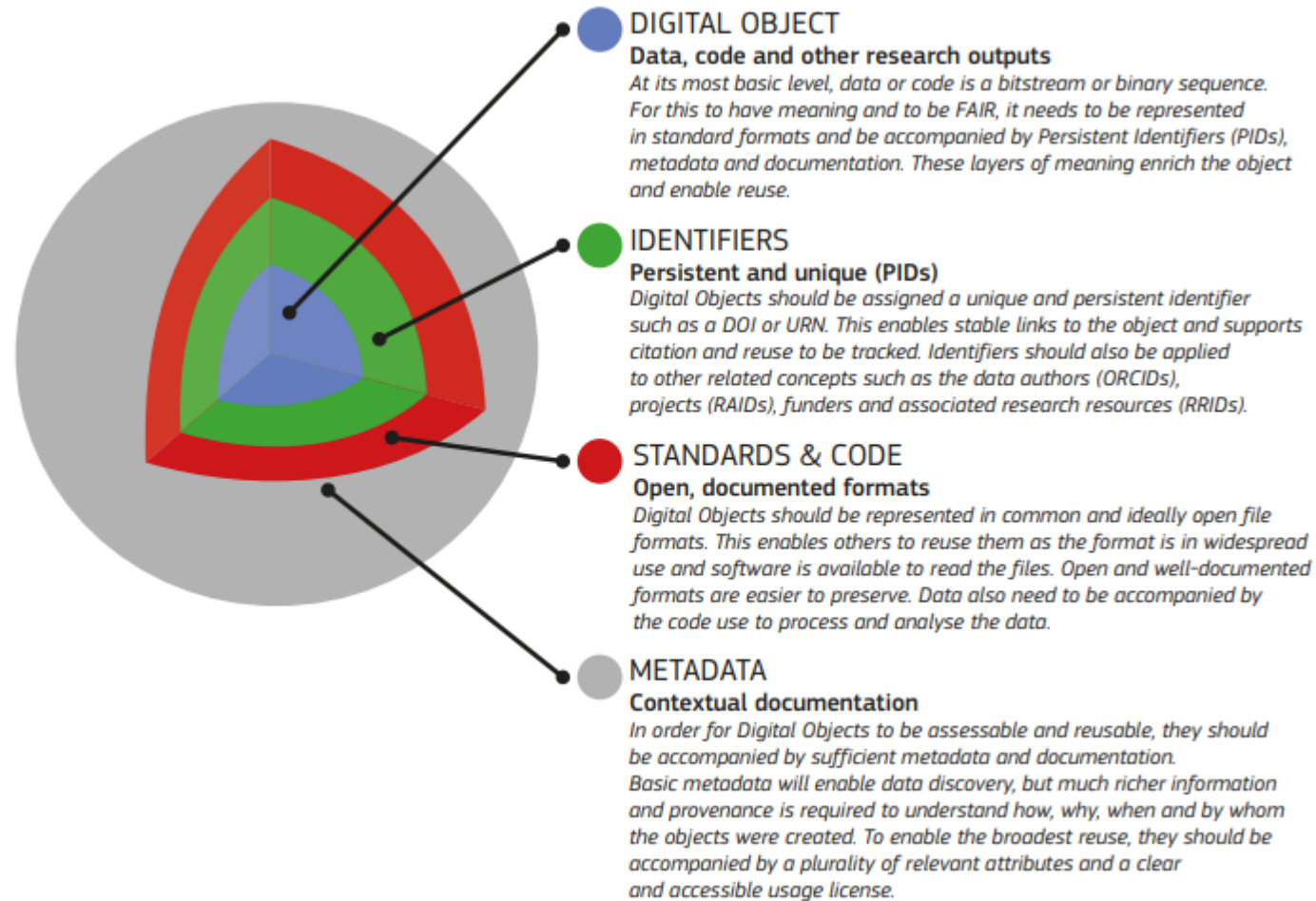
Level:	Message:
INFO	Object landing page accessible status -: True
SUCCESS	Resource type specified -: dataset
WARNING	NO data object content available/accessible to perform file descriptors (type and size) tests
WARNING	NO measured variables found in metadata, skip 'measured_variable' test.
WARNING	Measured variables given in metadata do not match data object content

Code still under development!

<https://www.f-uji.net/index.php>



# FAIR Digital Object ( FAIR ecosystem)



Turning FAIR into reality <https://op.europa.eu/s/oM5N>



GO FAIR



## GO FAIR?

- Community that provides Practical Guidance to support research stockholders to implement the FAIR data Principles.

## GO FAIR Framework

- Three-point FAIRification framework provides practical guidance to stockholders seeking to go FAIR.
  - Metadata for Machines workshops (M4M)
  - FAIR Implementation Profiles (FIP)
  - FAIR Data Point (FDP)

## GO FAIR Implementation Network

- Work groups to establish specific materials and tools as elements of the Internet of FAIR data and Services (IFDS).

# GO FAIR : How to join?

- You can do so either by joining an [existing Implementation Network](#) or by forming a new one.
- The INs are active in three activity pillars: [GO CHANGE](#), [GO TRAIN](#) and [GO BUILD](#)
- IN Example: [Food Systems](#)



<https://www.go-fair.org/>



# FAIR Office Austria

# FAIR Office Austria



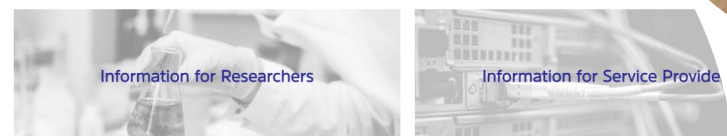
- Founding team = Consortium TU Wien, Graz University of Technology & University of Vienna
- Mission: **We connect stakeholders from research communities and service providers. Together, we help to advance the FAIR principles.**
- Support of researchers at partner institutions
  - FAIR in project proposals and DMPs
  - FAIR in research processes
  - FAIR through optimal use of tools and services
  - Support from data stewards
- Support of service providers
  - FAIRisation of repositories
  - Enhancing automation and machine actionability
- Partner with EOSC Mandated Organization and RDA Austria
  - Exchange with national and international initiatives
- Monitoring of FAIR processes in Austria
- Since June 2021: FAIR Office Austria is GO FAIR National Support and Coordination Office
- More information see: <https://fair-office.at/?lang=en>



 **FAIR**  
OFFICE AUSTRIA

We connect stakeholders from research communities and service providers. Together, we help to advance the FAIR principles.

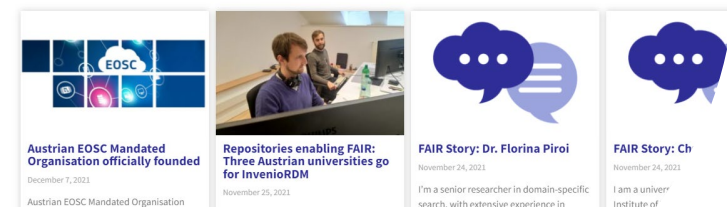
Deutsch English



Information for Researchers

Information for Service Providers

## News & Events



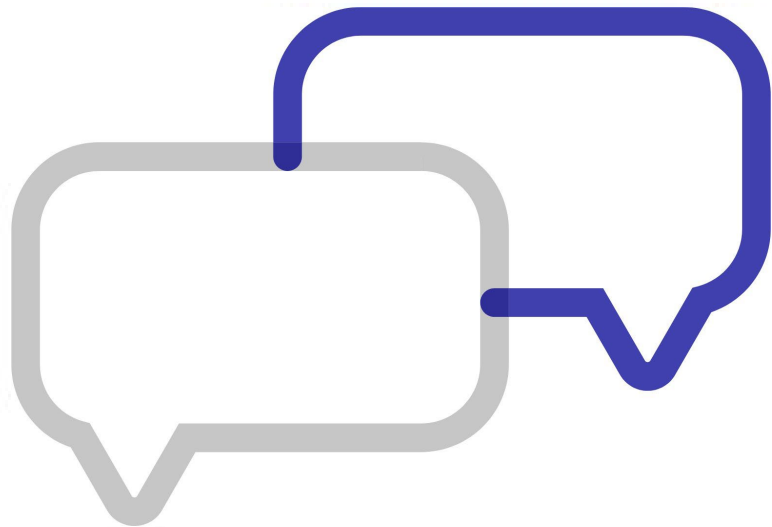
**Austrian EOSC Mandated Organisation officially founded**  
December 7, 2021  
Austrian EOSC Mandated Organisation

**Repositories enabling FAIR: Three Austrian universities go for InvenioRDM**  
November 25, 2021

**FAIR Story: Dr. Florina Piroi**  
November 24, 2021  
I'm a senior researcher in domain-specific search, with extensive experience in

**FAIR Story: Ch**  
November 24, 2021  
I am a univer  
Institute of

# Q&A





# Contact

[contact@fair-office.at](mailto:contact@fair-office.at)  
[samah.jaber@tuwien.ac.at](mailto:samah.jaber@tuwien.ac.at)







Thank you!