



FAIR DATA
AUSTRIA

LET'S MAKE OUR DATA FAIR!

19.10.2021

HOSTED BY TU WIEN,
CENTER FOR RESEARCH DATA MANAGEMENT



TECHNISCHE
UNIVERSITÄT
WIEN



EVENT SERIES „RESEARCH DATA MANAGEMENT IN AUSTRIA“

The event series „Research Data management in Austria“ is aimed at researchers and/or research support staff and serves to promote networking and exchange on the topic of research data management, writing a data management plan and similar related topics.

Slides and recordings on the project website:

forschungsdaten.at/en/fair-data-austria/materials/

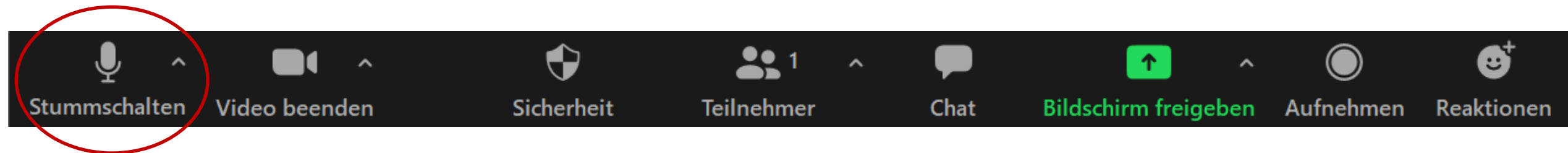


AGENDA

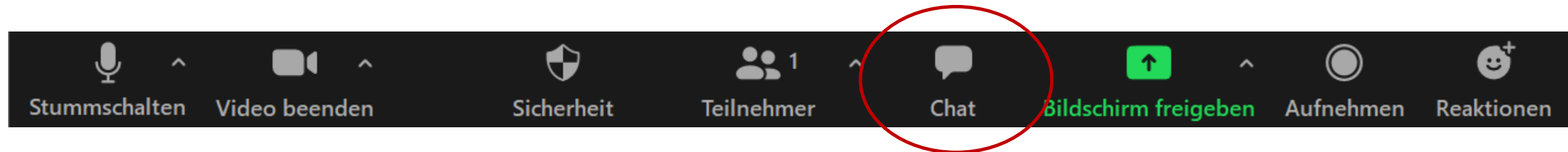
10.00 – 10.10	Welcome and introduction	Barbara Sánchez
10.10 – 11.00	Let's make our data FAIR!	Tomasz Miksa
11.00 – 11.30	Questions & Answers	Barbara Sánchez

FORMALITIES

Please turn off your microphone to avoid background noise



Please post your questions in the chat



LET'S MAKE OUR DATA FAIR!

Tomasz Miksa – TU Wien

Tomasz Miksa is a senior scientist at the faculty of Informatics and a technical coordinator at the Centre for Research Data Management at TU Wien.

He is an expert in data management and design of research data repositories. He also chairs the DMP Common Standards working group at the Research Data Alliance that developed a recommendation on machine-actionable Data Management Plans. Dr. Miksa is a lecturer in the course Data Stewardship which is a specialization within the Data Science curriculum at TU Wien.



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

INTRODUCTION

Hans Rosling and Data Science

- Talk held in 2006
- 15 years later
 - Problems exist
 - But a lot is changing



https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen#t-1144167



Hans Rosling

- “Because the **data is hidden down in the databases**. And the public is there, and the internet is there, but we have still **not used it effectively.**”
- “There are **some web pages** like this,(...), but people **put prices** on them, **stupid passwords** and boring statistics.

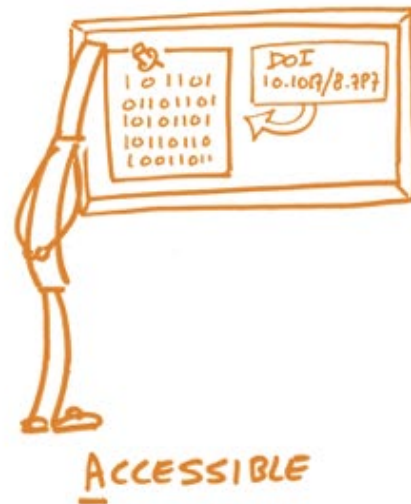
Hans Rosling

- “Some countries accept that their databases can go out on the world. But what we really need is, of course, a search function, a search function where **we can copy the data up to a searchable format** and get it out in the world.”
- “The **publicly funded data** is down here. (...) One of the crucial points is to **make them searchable**, and then people can use the different design tools to animate it there.

FAIR PRINCIPLES

FAIR Principles (very simplified :))

FAIR DATA PRINCIPLES



FAIR Principles (less simplified :))



FAIR principles

Findable
Accessible
Interoperable
Reusable

www.nature.com/scientificdata

SCIENTIFIC DATA

Amended: Addendum

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

SUBJECT CATEGORIES
» Research data
» Publication characteristics

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016

Mark D. Wilkinson *et al.*[#]

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).
[#]A full list of authors and their affiliations appears at the end of the paper.

SCIENTIFIC DATA | 3:160018 | DOI: 10.1038/sdata.2016.18

<https://www.nature.com/articles/sdata201618>

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

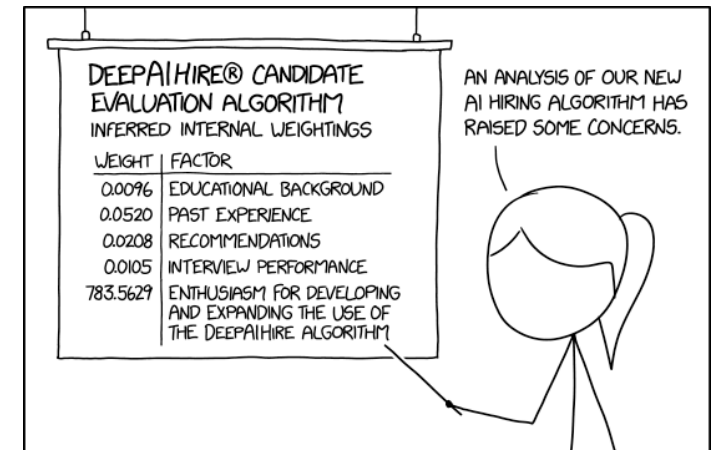
Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

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

FAIR vs fair

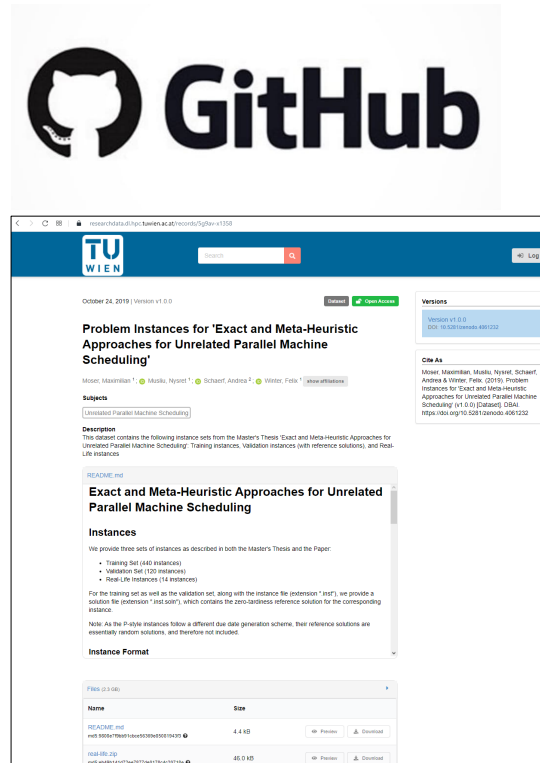
- FAIR principles \neq Algorithmic fairness
- To be FAIR
 - To apply/use FAIR principles
 - Focus on how data is managed, etc.
- To be fair
 - Evade bias
 - Focus on design and implementation



<https://xkcd.com/2237/>

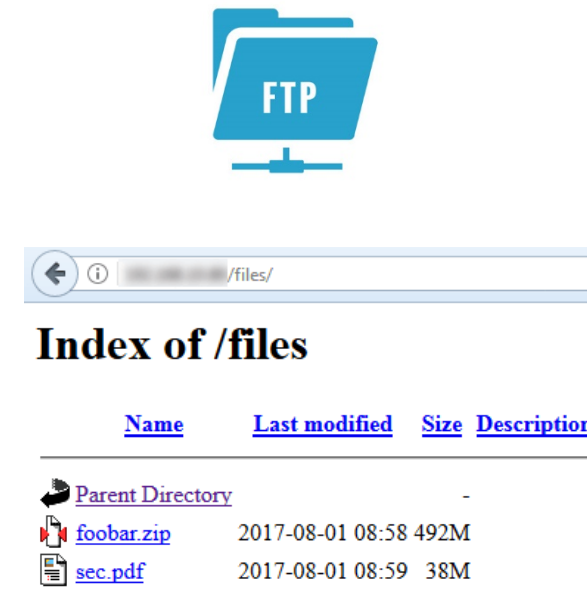
Findable – simplified examples

Yes



Data repository

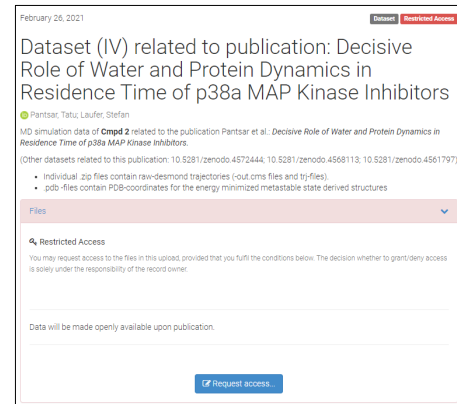
No



Personal website

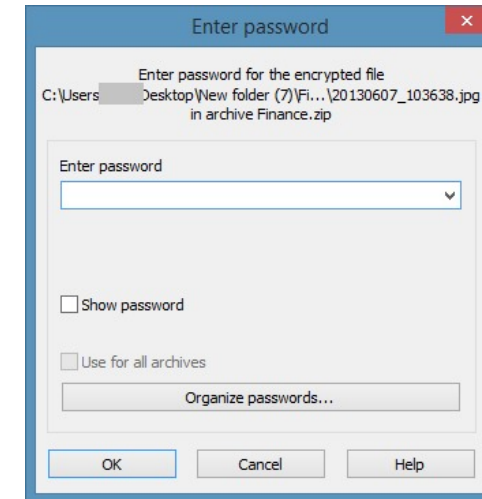
Accessible – simplified examples

Yes



Restricted access,
but a clear way to
request access

No



Interoperable – simplified examples

Yes

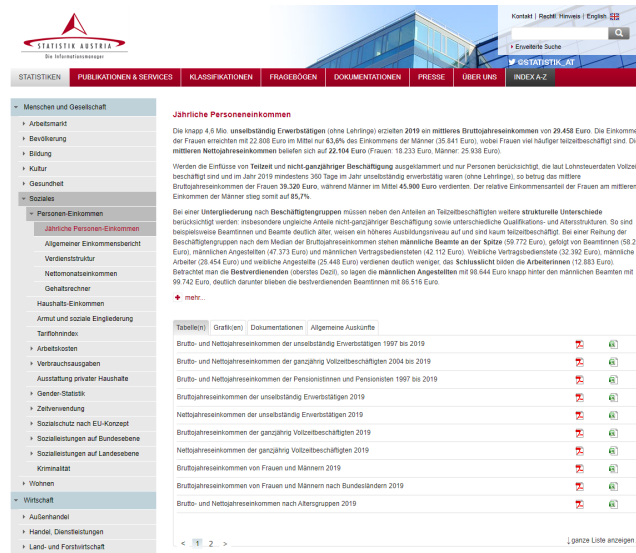
- XML following known XSD Schema
- MP3 for audio recordings

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 **Bruttojahreseinkommen** von 29.458 Euro. Die Einkommen der Frauen erreichten mit 22.808 Euro im Mittel nur 63,9% des Einkommens der Männer (35.841 Euro), wobei Frauen viel häufiger **Teilzeitbeschäftigte** sind. Die mittleren **Nettojahreseinkommen** beliefen sich auf 23.104 Euro (Frauen: 18.233 Euro, Männer: 25.938 Euro).

Wirden die Einkünfte von **Teilzeit- und nicht-ganzjähriger Beschäftigung** ausgeklammert und nur **Personen berücksichtigt**, die laut Lohnsteuerdaten Vollzeit beschäftigt sind und im Jahr 2019 mindestens 300 Tage im Jahr unselbständig erwerbstätig waren (ohne Lehrlinge), so betrug das mittlere **Bruttojahreseinkommen** der Frauen 39.209 Euro, während Männer im Mittel 45.900 Euro verdienten. Der relative Einkommensanteil der Frauen am mittleren Einkommen der Männer stieg somit auf 85,1%.

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 **Teilzeitbeschäftigte**. Bei einer Feinung der Beschäftigungsgruppen nach dem Median der **Bruttojahreseinkommen** stehen **männliche Beamte an der Spitze** (59.772 Euro), gefolgt von **Beamtinnen** (58.213 Euro), **männlichen Angestellten** (47.373 Euro) und **männlichen Vertragsbediensteten** (42.112 Euro). **Weibliche Vertragsbedienstete** (32.392 Euro), **männliche Arbeiter** (28.454 Euro) und **weibliche Jugendliche** (25.468 Euro) verdienen deutlich weniger, der **Schwachsicht** bilden die **Adipösen** (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 stehen die **bestverdienenden Beamtinnen** mit 86.516 Euro.

mehr...

Tabellen	Grafiken	Datensätze	Allgemeine Auskünfte
Brutto- und Nettojahreseinkommen der unselbständig Erwerbstätigen 1997 bis 2019			
Brutto- und Nettojahreseinkommen der ganzjährig Vollzeitbeschäftigten 2004 bis 2019			
Brutto- und Nettojahreseinkommen der Pensionistinnen und Pensionisten 1997 bis 2019			
Bruttojahreseinkommen der unselbständig Erwerbstätigen 2019			
Nettojahreseinkommen der unselbständig Erwerbstätigen 2019			
Bruttojahreseinkommen der ganzjährig Vollzeitbeschäftigten 2019			
Nettojahreseinkommen der ganzjährig Vollzeitbeschäftigten 2019			
Bruttojahreseinkommen von Frauen und Männern 2019			
Bruttojahreseinkommen von Frauen und Männern nach Bundesländern 2019			
Brutto- und Nettojahreseinkommen nach Altersgruppen 2019			

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

No



Provenance and permissions not clear

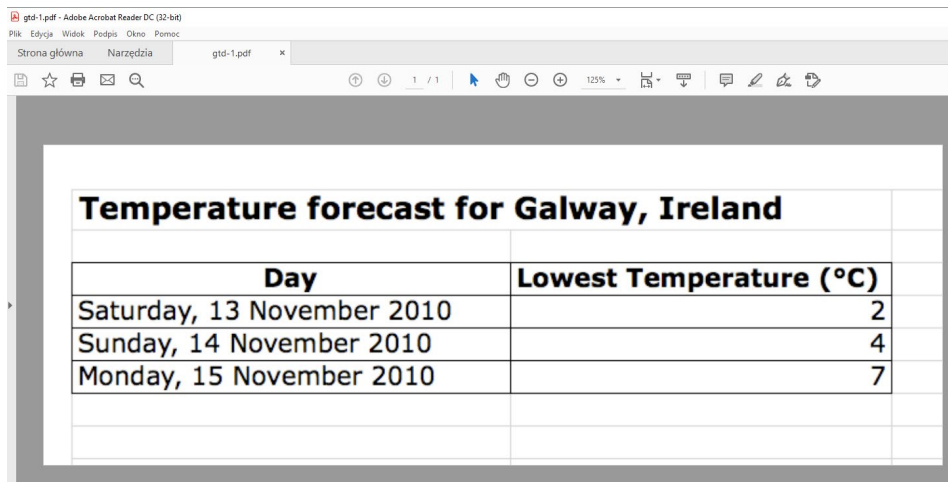
MACHINE-ACTIONABILITY

Machine actionability

- “the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention” <https://www.go-fair.org/fair-principles>
- “information that is structured in a consistent way so that machines, or computers, can be programmed against the structure.” <https://ddialliance.org/taxonomy/term/198>
- **Machine-actionability is core to each of the FAIR principles**

Machine-actionability - example

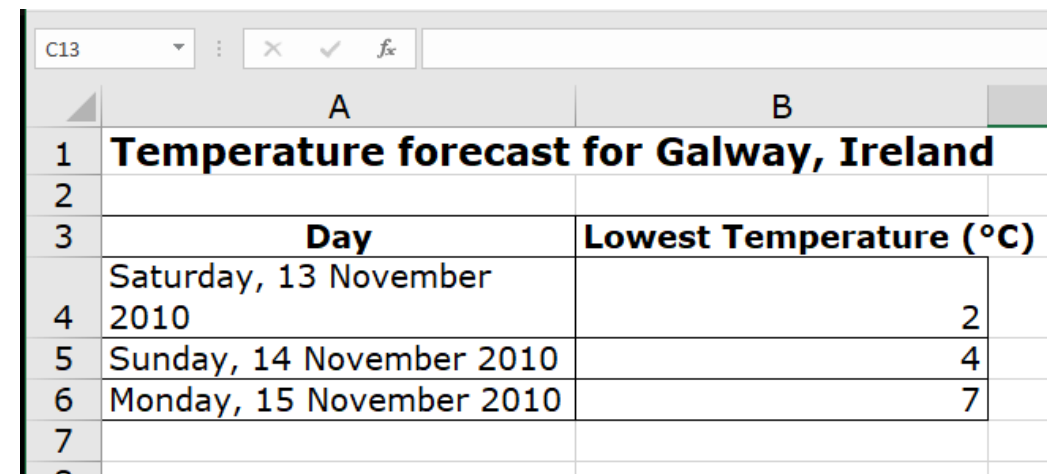
Not machine-actionable



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



Machine-actionable



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



Machine-actionability – example (Linked Open Data)

<https://5stardata.info/en/examples/gtd-5/#temp>

→ `rdfs:seeAlso` → `dbpedia:Temperature`

→ `owl:sameAs` → `dbpedia:Celsius`

<https://5stardata.info/en/examples/gtd-5/#temp20101113>

→ `meteo:celsius` → `"2"^^xsd:decimal`

← is `meteo:temperature` of ← <https://5stardata.info/en/examples/gtd-5/#forecast20101113>

<https://5stardata.info/en/examples/gtd-5/#temp20101114>

→ `meteo:celsius` → `"4"^^xsd:decimal`

← is `meteo:temperature` of ← <https://5stardata.info/en/examples/gtd-5/#forecast20101114>

<https://5stardata.info/en/examples/gtd-5/#temp20101115>

→ `meteo:celsius` → `"7"^^xsd:decimal`

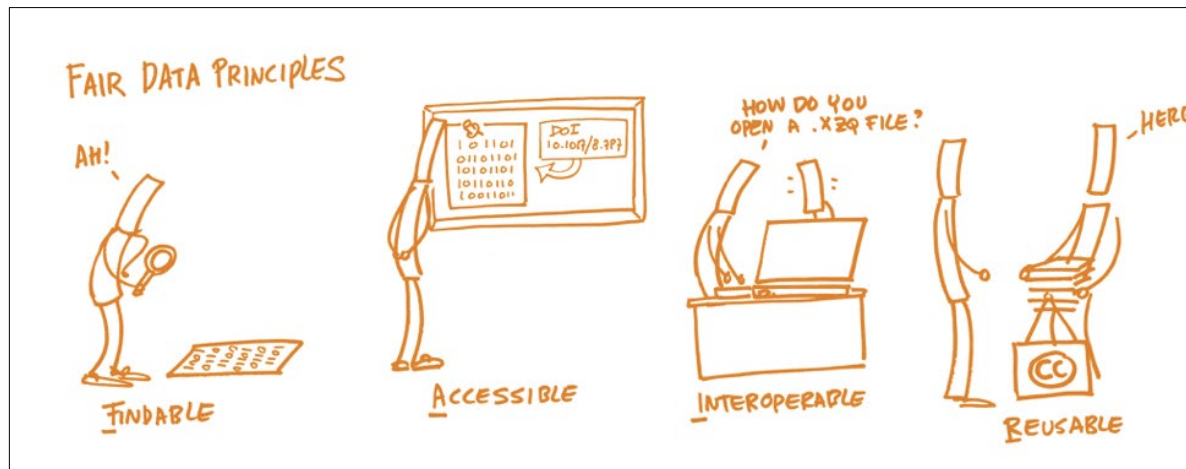
← is `meteo:temperature` of ← <https://5stardata.info/en/examples/gtd-5/#forecast20101115>

Unit definition
and link to common definition

Values, types, link to forecast

Machine-actionability

- The more machine-actionable data is, the better it is!
- FAIR
 - To people!
 - To machines!



Don't get misled by this picture!
Machine is missing in this picture!
That's why it's simplified and thus limited!
(but still useful)

FAIR IN DETAIL

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)




1

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

0..*



 <https://orcid.org/0000-0002-4929-7875>

Website and social links >

SBA Research
Linked in

Countries >

Austria

Is this you? [Sign in to start editing](#) Printable version

Name
Tomasz Miksa

Activities Collapse all

Employment (2) Sort

TU Wien: Vienna, AT

Employment [Show more detail](#)

Source: TU Wien

SBA Research: Vienna, AT

2012-09-01 to present
Employment [Show more detail](#)

Source: Tomasz Miksa

Education and qualifications (2) Sort

Works (19) Sort

The openEO API—Harmonising the Use of Earth Observation Cloud Services Using Virtual Data Cube Functionalities

Remote Sensing
2021-03 | journal-article
DOI: [10.3390/rs13061125](https://doi.org/10.3390/rs13061125) [Show more detail](#)

Source: Multidisciplinary Digital Publishing Institute

ORCID assigned to a person

EDUCATION

Ten principles for machine-actionable data management plans

Tomasz Miksa^{1*}, Stephanie Simms², Daniel Mietchen³, Sarah Jones⁴

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@ifs.tuwien.ac.at

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



OPEN ACCESS

Citation: Miksa T, Simms S, Mietchen 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>

Editor: Patrick Gruber, Genome Canada, CANADA

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.

DOI assigned to a publication

GitHub repository: [helmuthb / dmp-exercise1](#)

Code Issues Pull requests Actions Projects Security Insights

1.0.1 1 branch 3 tags

Go to file Code

File/Folder	Description	Time
data	First version with full data	2 years ago
src	First version with full data	2 years ago
.gitignore	First version with full data	2 years ago
Dockerfile	First version with full data	2 years ago
LICENSE	Initial commit	2 years ago
README.md	Corrected DOI link	2 years ago
Report.pdf	First version with full data	2 years ago



README.md

DOI [10.5281/zenodo.2648326](https://doi.org/10.5281/zenodo.2648326)

zenodo repository: [US Wheat and Salzburg Middle-Aged Marriages - Data Experiment](#)

Software Open Access

14 views 5 downloads

Available in:  

Publication date: **DOI: [10.5281/zenodo.2648398](https://doi.org/10.5281/zenodo.2648398)**

Related identifiers: <https://github.com/helmuthb/dmp-exercise1/tree/1.0.1>

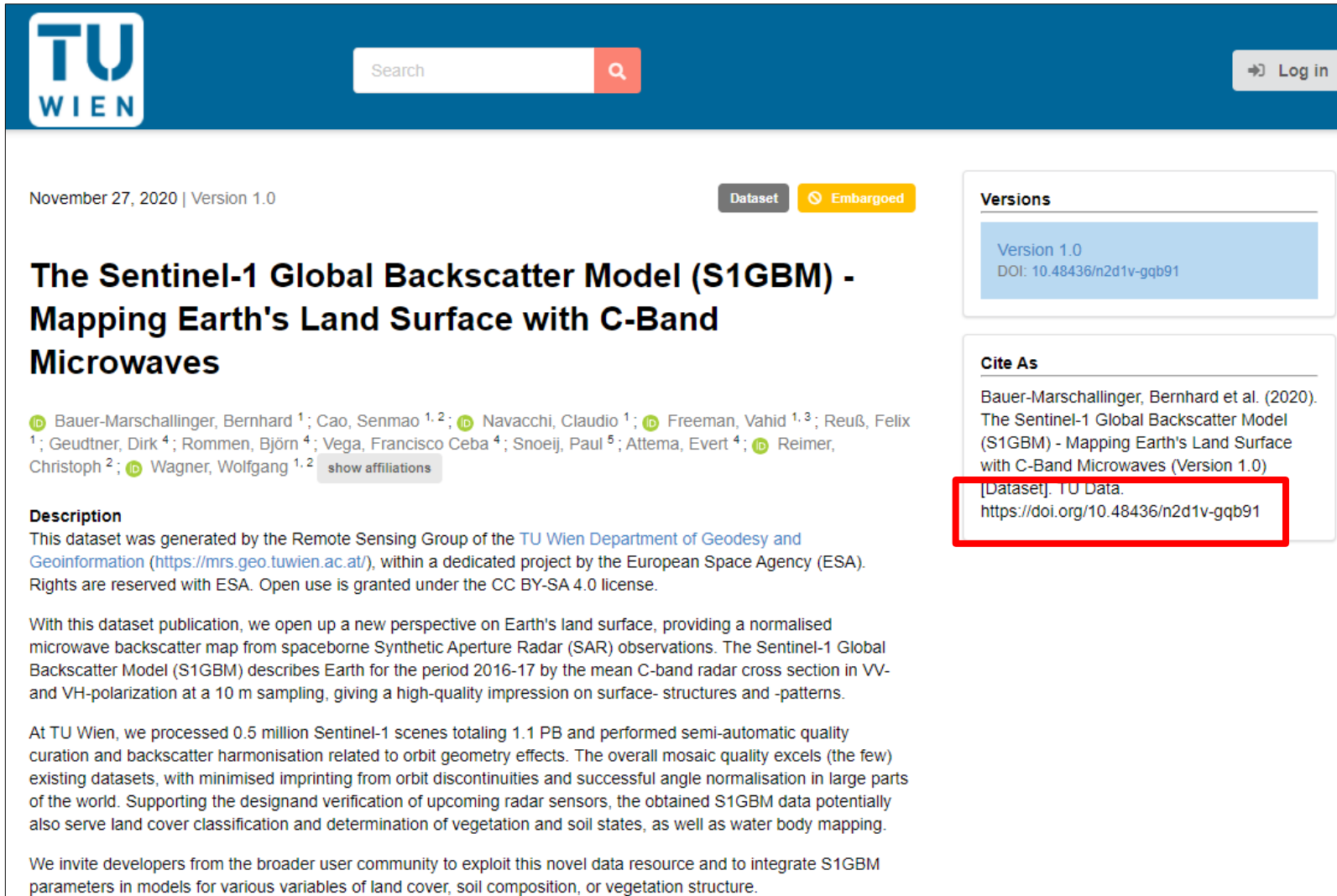
License (for files): [Other \(Open\)](#)


Versions

Version	Date
Version 1.0.1 10.5281/zenodo.2648398	Apr 22, 2019
Version 1.0 10.5281/zenodo.2648396	Apr 22, 2019
Version 0.1 10.5281/zenodo.2648327	Apr 22, 2019

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.2648326](https://doi.org/10.5281/zenodo.2648326). This DOI represents all versions, and will always resolve to the latest one. [Read more](#).






DOI assigned to code





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 ¹; Cao, Senmao ^{1,2};  Navacchi, Claudio ¹;  Freeman, Vahid ^{1,3}; Reuß, Felix ¹; Geudtner, Dirk ⁴; Rommen, Björn ⁴; Vega, Francisco Ceba ⁴; Snoeij, Paul ⁵; Attema, Evert ⁴;  Reimer, Christoph ²;  Wagner, Wolfgang ^{1,2} 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/>), 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 designand 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

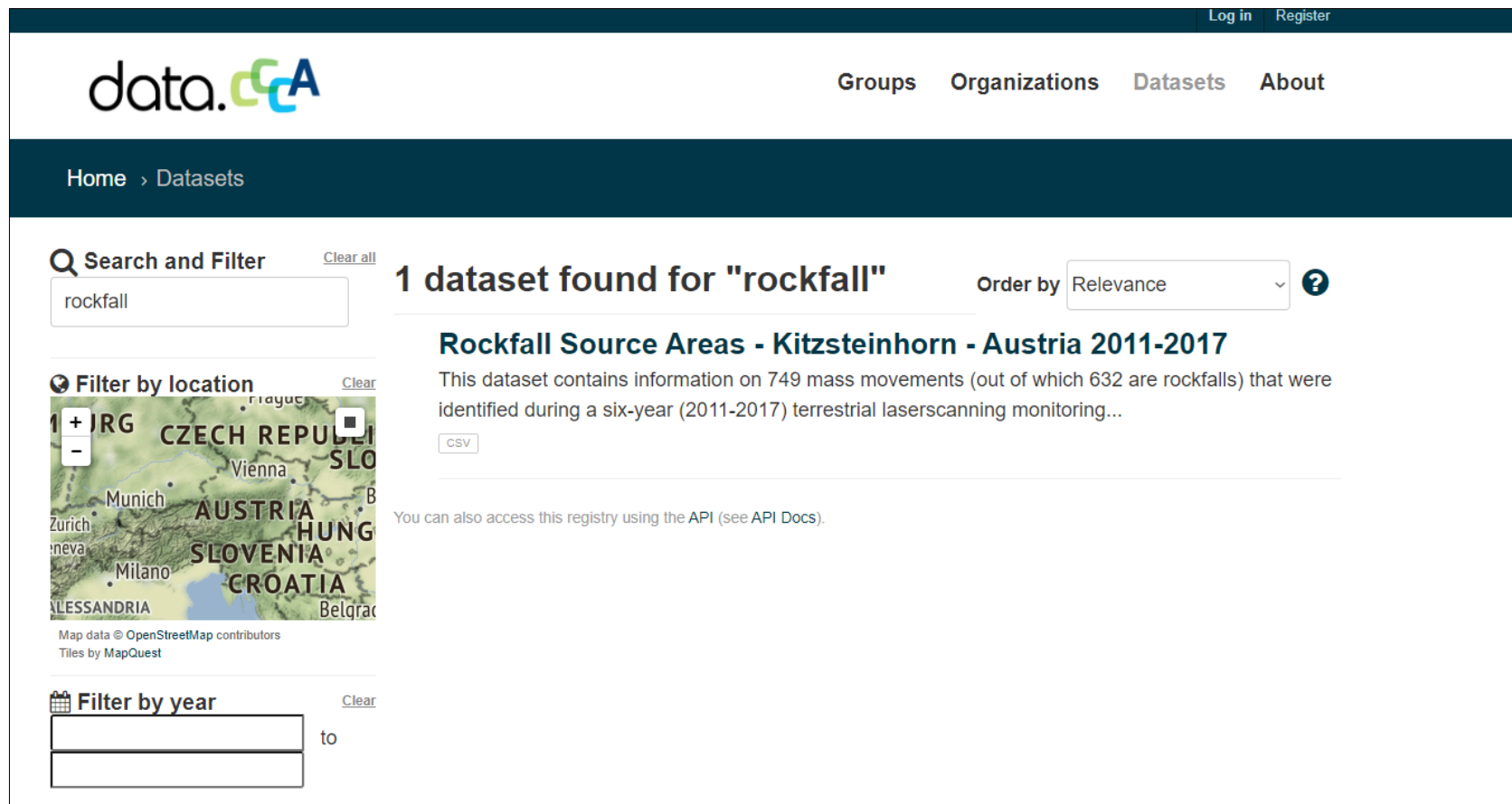
Bauer-Marschallinger, Bernhard et al. (2020). The Sentinel-1 Global Backscatter Model (S1GBM) - Mapping Earth's Land Surface with C-Band Microwaves (Version 1.0) [Dataset]. TU Data. <https://doi.org/10.48436/n2d1v-gqb91>

DOI assigned to data

Findable

- 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

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



The screenshot shows the data.cca 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 on the left contains the text 'rockfall' and has a 'Clear all' link. To the right of the search bar, it indicates '1 dataset found for "rockfall"' and an 'Order by' dropdown menu set to 'Relevance'. The search results list a dataset titled 'Rockfall Source Areas - Kitzsteinhorn - Austria 2011-2017'. The description states: '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...'. There is a 'csv' download button. Below the description, a note says 'You can also access this registry using the API (see API Docs)'. On the left side, there are filters: 'Filter by location' with a map of Central Europe showing Austria, Czech Republic, Slovenia, Hungary, and Croatia; and 'Filter by year' with two empty input fields and a 'to' label. The footer of the map shows 'Map data © OpenStreetMap contributors' and 'Tiles by MapQuest'.

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

 DATASET



Rockfall Source Areas - Kitzsteinhorn - Austria 2011-2017

Followers: 0

Views: 22



Published by: GEORESEARCH Forschungsgesellschaft mbH License: Creative Commons Attribution - Share-Alike (CC-BY-SA)


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').

Dataset Versions

Citation

Dataset

Resources

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

Explore

Dataset Metadata

Contact Basics Keywords Spatial Time Specifics Quality Conformity

Export Metadata

Machine-readable metadata


Basic Information about this dataset

Identifier

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

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	<p>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 'MOREXPert').</p>
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...



Dataset Metadata

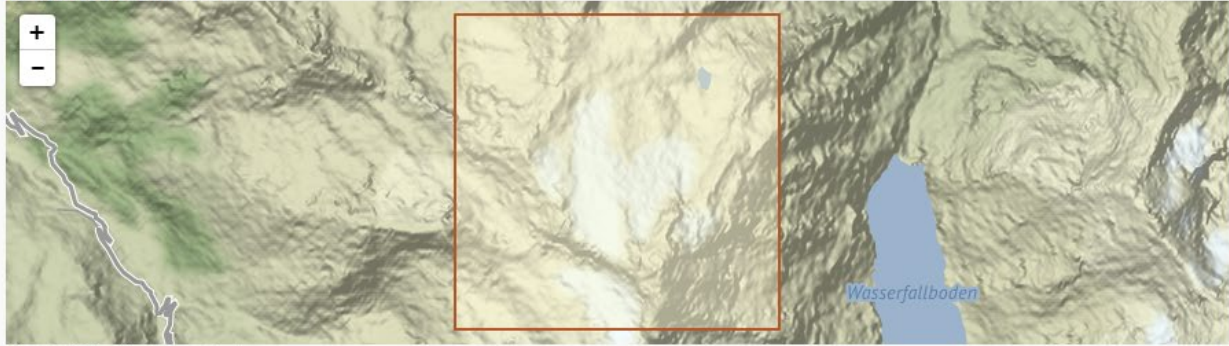
Export Metadata ▾

[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

Metadata - Chemistry


NIH National Library of Medicine
National Center for Biotechnology Information

PubChem About Blog Submit Contact

COMPOUND SUMMARY

Water ethanol


PubChem CID 19096565



Structure 
2D 3D
[Find Similar Structures](#)

Molecular Formula $C_2H_8O_2$

Synonyms ethanol water
ethanol-water
water ethanol
water-ethanol
EtOH water
[More...](#)

Molecular Weight 64.08 g/mol

Parent Compound  CID 702 (Ethanol)

Component Compounds  CID 702 (Ethanol)
 CID 962 (Water)

Dates Modify 2021-02-27 Create 2007-12-04

2 Names and Identifiers

2.1 Computed Descriptors

2.1.1 IUPAC Name

ethanol;hydrate
Computed by LexiChem 2.6.6 (PubChem release 2019.06.18)
[PubChem](#)

2.1.2 InChI

InChI=1S/C2H6O.H2O/c1-2-3;/h3H,2H2,1H3;1H2
Computed by InChI 1.0.5 (PubChem release 2019.06.18)
[PubChem](#)

2.1.3 InChI Key

IDGUHHHCWSQLU-UHFFFAOYSA-N
Computed by InChI 1.0.5 (PubChem release 2019.06.18)
[PubChem](#)

2.1.4 Canonical SMILES

CCO.O
Computed by OEChem 2.1.5 (PubChem release 2019.06.18)
[PubChem](#)

2.2 Molecular Formula

$C_2H_8O_2$
Computed by PubChem 2.1 (PubChem release 2019.06.18)
[PubChem](#)

3 Chemical and Physical Properties

3.1 Computed Properties

Property Name	Property Value
Molecular Weight	64.08 g/mol
Hydrogen Bond Donor Count	2
Hydrogen Bond Acceptor Count	2
Rotatable Bond Count	0
Exact Mass	64.052429 g/mol
Monoisotopic Mass	64.052429 g/mol
Topological Polar Surface Area	21.2 Å ²
Heavy Atom Count	4
Formal Charge	0
Complexity	2.8
Isotope Atom Count	0
Defined Atom Stereocenter Count	0
Undefined Atom Stereocenter Count	0
Defined Bond Stereocenter Count	0
Undefined Bond Stereocenter Count	0
Covalently-Bonded Unit Count	2
Compound Is Canonicalized	Yes

[PubChem](#)

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 authorisation 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

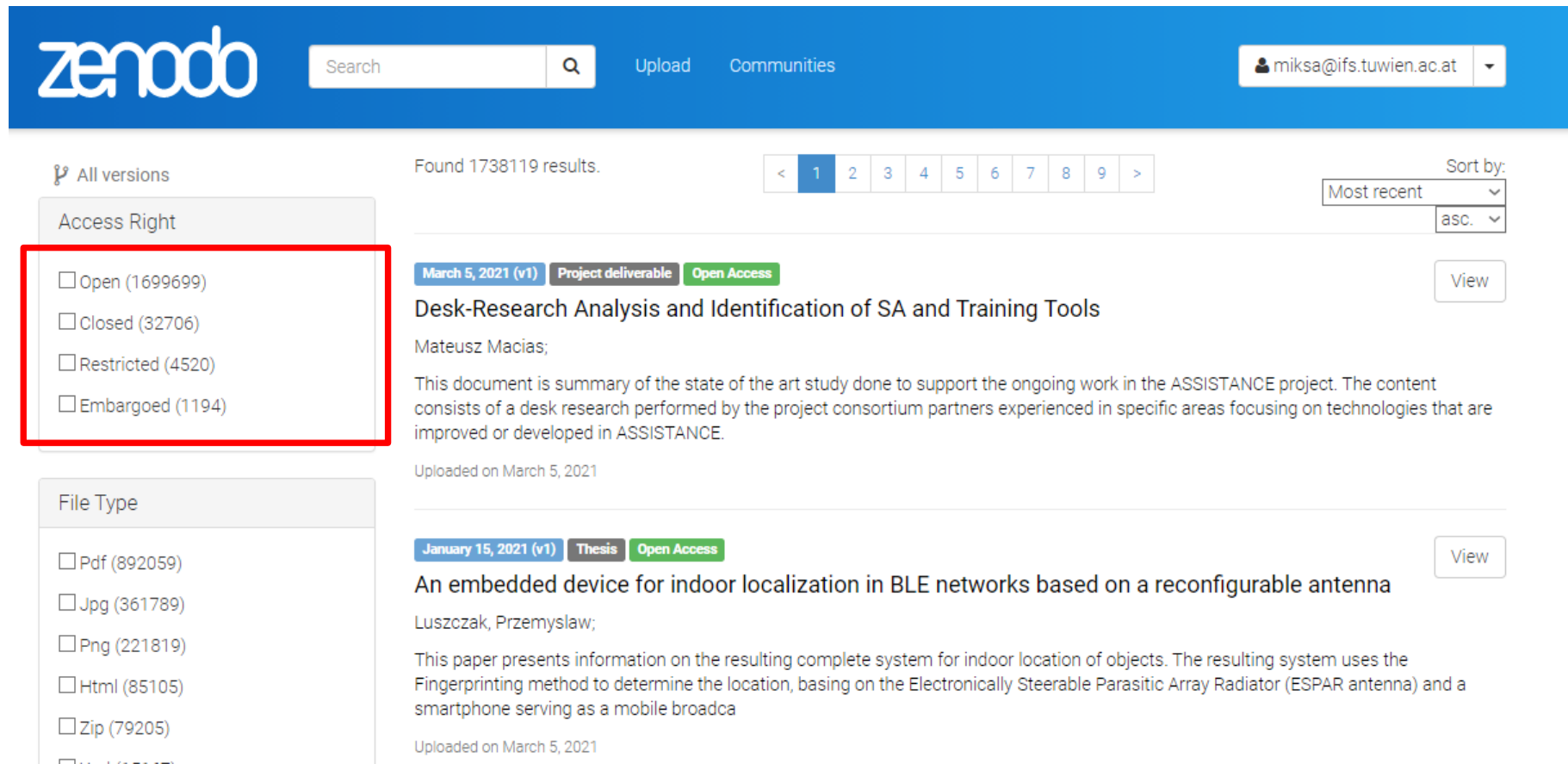
- “Anyone with a computer and an internet connection can access at least the metadata”
- HTTP
 - Open – specification of the protocol is known to everyone
 - Free – no need to pay to “use Internet”
- Proprietary protocols
 - evade

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 authorisation

- 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 interface. The top navigation bar includes the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. A user profile for 'miksa@ifs.tuwien.ac.at' is visible in the top right. The search results section shows 'Found 1738119 results.' with a pagination control on page 1 and a 'Sort by' dropdown set to 'Most recent' and 'asc.'. On the left, 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)
- Embargoed (1194)

The 'File Type' panel contains the following options:

- Pdf (892059)
- Jpg (361789)
- Png (221819)
- Html (85105)
- Zip (79205)

The main content area displays two search results. The first result is titled '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 titled '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.

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

Accessible - example

The screenshot shows the Zenodo interface for a dataset. At the top, there is a blue header with the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. Below the header, the date 'March 5, 2021' is displayed on the left, and two red-bordered buttons labeled 'Dataset' and 'Restricted Access' are on the right. The main title of the dataset is 'Phase-contrast X-ray tomography of free-breathing murine lungs'. Below the title, the authors are listed: Kian Shaker, Ilian Häggmark, Jakob Reichmann, Marie Arsenian-Henriksson, and Hans M. Hertz. A short description follows: 'Full resolution phase-contrast X-ray tomographic dataset acquired of the lungs of a free-breathing mouse (NMRI nude mice, BomTac:NMRI-Foxntm, Taconic Biosciences, DK), weighing roughly 28 g.' Technical details include 'Datasize: 3851x3951 pixels per slice, stack of 1700 slices, 16-bit, .tif' and 'Voxel-size: 8.25x8.25x8.25 micrometer'. A note states 'A sample slice is available for download, prior to downloading the full dataset.' A section titled 'If you use the dataset, please cite:' provides a citation: '"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)'. At the bottom, a red-bordered box highlights the 'Files' section, which contains a 'Restricted Access' notice. The notice explains that users can request access to the files if they fulfill the conditions, and the decision is solely under the responsibility of the record owner. Below the notice is a text input field with the prompt 'Dataset is available upon request. Please state the purpose of your request and the intended usage of the data.' and a blue 'Request access...' button.

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-Foxntm, 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.

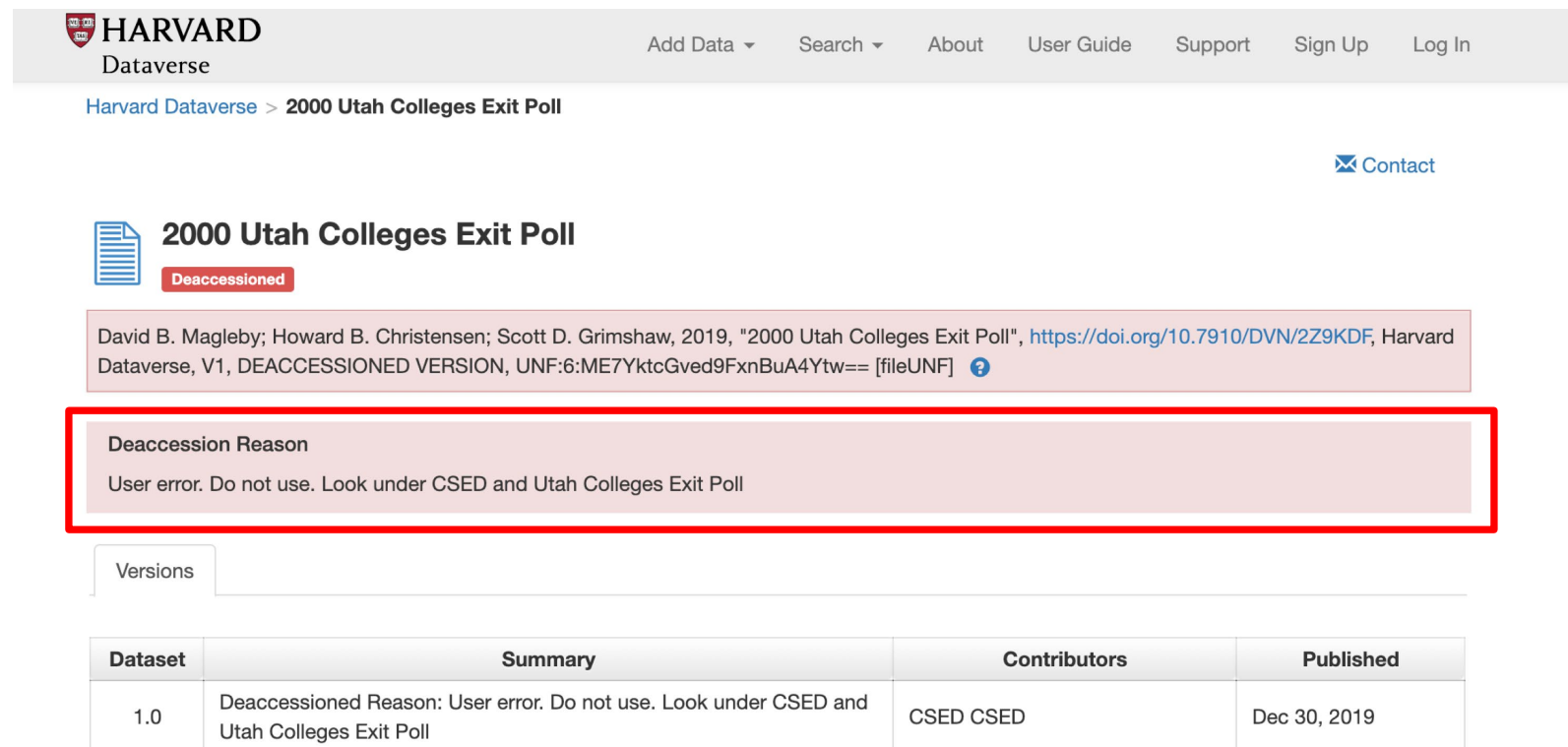
Request access...

Accessible (A1) - comments

- FAIR data \neq open data!
 - Common misconception stemming from A1.1 (open protocol)
 - FAIR data *can* be open data, but it has nothing to do with A1
- Access can/should be realized over APIs
 - SPARQL endpoints
 - HTTP APIs
 - Client libraries
- Access is not only “click to download”

Tombstone pages (A2)

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



The screenshot shows the Harvard Dataverse interface for a dataset titled "2000 Utah Colleges Exit Poll". The page includes a navigation bar with options like "Add Data", "Search", "About", "User Guide", "Support", "Sign Up", and "Log In". A breadcrumb trail shows "Harvard Dataverse > 2000 Utah Colleges Exit Poll". A "Contact" button is visible in the top right. The dataset title is accompanied by a document icon and a red "Deaccessioned" badge. Below the title, a text box provides the citation: "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] ?". A red-bordered box highlights the "Deaccession Reason" section, which states: "User error. Do not use. Look under CSED and Utah Colleges Exit Poll". Below this, a "Versions" tab is active, and a table lists the dataset's history.

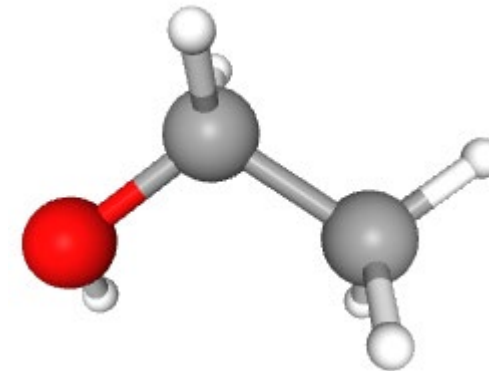
Dataset	Summary	Contributors	Published
1.0	Deaccessioned Reason: User error. Do not use. Look under CSED and Utah Colleges Exit Poll	CSED CSED	Dec 30, 2019

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

Language for knowledge representation (I1)

- “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



Vocabularies

- 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	<i>Heathrow Airport</i>	Public
Greater London	EGWU	NHT	<i>RAF Northolt</i>	Military
Greater London	EGLC	LCY	<i>London City Airport</i>	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

Vocabularies

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

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

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
Publications **Gene** EBAG9
Feature viewer **Organism** *Homo sapiens (Human)*
Feature table **Status** Reviewed - Annotation score: ●●●●● - Experimental evidence at protein level¹

Functionⁱ
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 functionⁱ

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

[Complete GO annotation on QuickGO ...](#)

GO - Biological processⁱ

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

[Complete GO annotation on QuickGO ...](#)

Keywordsⁱ
Biological process: Apoptosis

Enzyme and pathway databases
PathwayCommonsⁱ: O00559
Reactomeⁱ: R-HSA-9018519, Estrogen-dependent gene expression

Names & Taxonomyⁱ

Protein names ⁱ	Recommended name: Receptor-binding cancer antigen expressed on SiSo cells Alternative name(s): <ul style="list-style-type: none">Cancer-associated surface antigen RCAS1Estrogen receptor-binding fragment-associated gene 9 protein
Gene names ⁱ	Name: EBAG9 Synonyms: RCAS1
Organism ⁱ	Homo sapiens (Human)
Taxonomic identifier ⁱ	9606 [NCBI]
Taxonomic lineage ⁱ	Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Haplorrhini > Catarrhini > Hominidae > Homo [?]
Proteomes ⁱ	UP000005640 Component ¹ : 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.

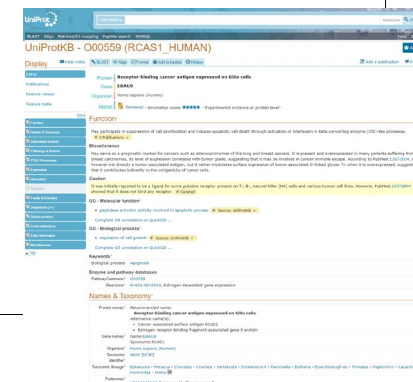
Related documents

[Taxonomy](#)

[Controlled vocabulary of species](#)

[What are proteomes?](#)

[What are reference proteomes?](#)



Qualified References (I3)

The screenshot shows the UniProtKB entry for O00559 (RCAS1_HUMAN). The 'Cross-references' section is highlighted with a red box. Below it, the 'Sequence databases' section is also highlighted with a red box, showing a list of mRNA translations with 'AF006265 mRNA Translation: AAB61617.1' highlighted.

UniProtKB - O00559 (RCAS1_HUMAN)

Cross-referencesⁱ

Web resourcesⁱ

- Atlas of Genetics and Cytogenetics in Oncology and Haematology

Sequence databases

Select the link destinations:	AF006265 mRNA Translation: AAB61617.1
<input checked="" type="radio"/> EMBL ⁱ	AB007619 mRNA Translation: BAA22572.1
<input type="radio"/> GenBank ⁱ	AY653072 mRNA Translation: AAU85838.1
<input type="radio"/> DDBJ ⁱ	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 ⁱ	CCDS6313.1 [O00559-1]
RefSeq ⁱ	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 ⁱ	Search...
SWISS-MODEL-Workspace ⁱ	Submit a new modelling project...

Protein-protein interaction databases

BioGRID ⁱ	114607, 40 interactors
IntAct ⁱ	O00559, 30 interactors
STRING ⁱ	9606.ENSP00000337675

Functionⁱ

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 functionⁱ

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

Complete GO annotation on QuickGO ...

GO - Biological processⁱ

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

Complete GO annotation on QuickGO ...

Keywordsⁱ

Biological process | Apoptosis

Enzyme and pathway databases

PathwayCommons ⁱ	O00559
Reactome ⁱ	R-HSA-9018519, Estrogen-dependent gene expression

Names & Taxonomyⁱ

Protein names ⁱ	Recommended name: Receptor-binding cancer antigen expressed on SiSo cells Alternative name(s): <ul style="list-style-type: none">Cancer-associated surface antigen RCAS1Estrogen receptor-binding fragment-associated gene 9 protein
Gene names ⁱ	Name: EBAG9 Synonyms: RCAS1
Organism ⁱ	Homo sapiens (Human)
Taxonomic identifier ⁱ	9606 [NCBI]
Taxonomic lineage ⁱ	Eukaryota > Metazoa > Chordata > Craniata > Vertebrata > Euteleostomi > Mammalia > Eutheria > Euarchontoglires > Primates > Hominidae > Homo
Proteomes ⁱ	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 page for PyTorch. The repository name is 'pytorch / pytorch'. The page includes navigation links for Code, Issues (3,951), Pull requests (1,284), Actions, Projects (5), Wiki, Security, and Insights. The repository description is 'Tensors and Dynamic neural networks in Python with strong GPU acceleration' with a link to 'https://pytorch.org'. The repository statistics show 24,578 commits, 2,993 branches, 0 packages, 31 releases, and 1,316 contributors. A 'View license' button is highlighted with a red box. Below the repository statistics, there is a list of recent commits. The commit titled 'LICENSE' is highlighted with a red box. The commit message is 'Move copyright lines back to NOTICE file, fixes #6911 (#8310)' and it was made 2 years ago. Below the commit list, there is a 'License' section with a red box around the text: 'PyTorch is BSD-style licensed, as found in the LICENSE file.'

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

pytorch / pytorch 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>



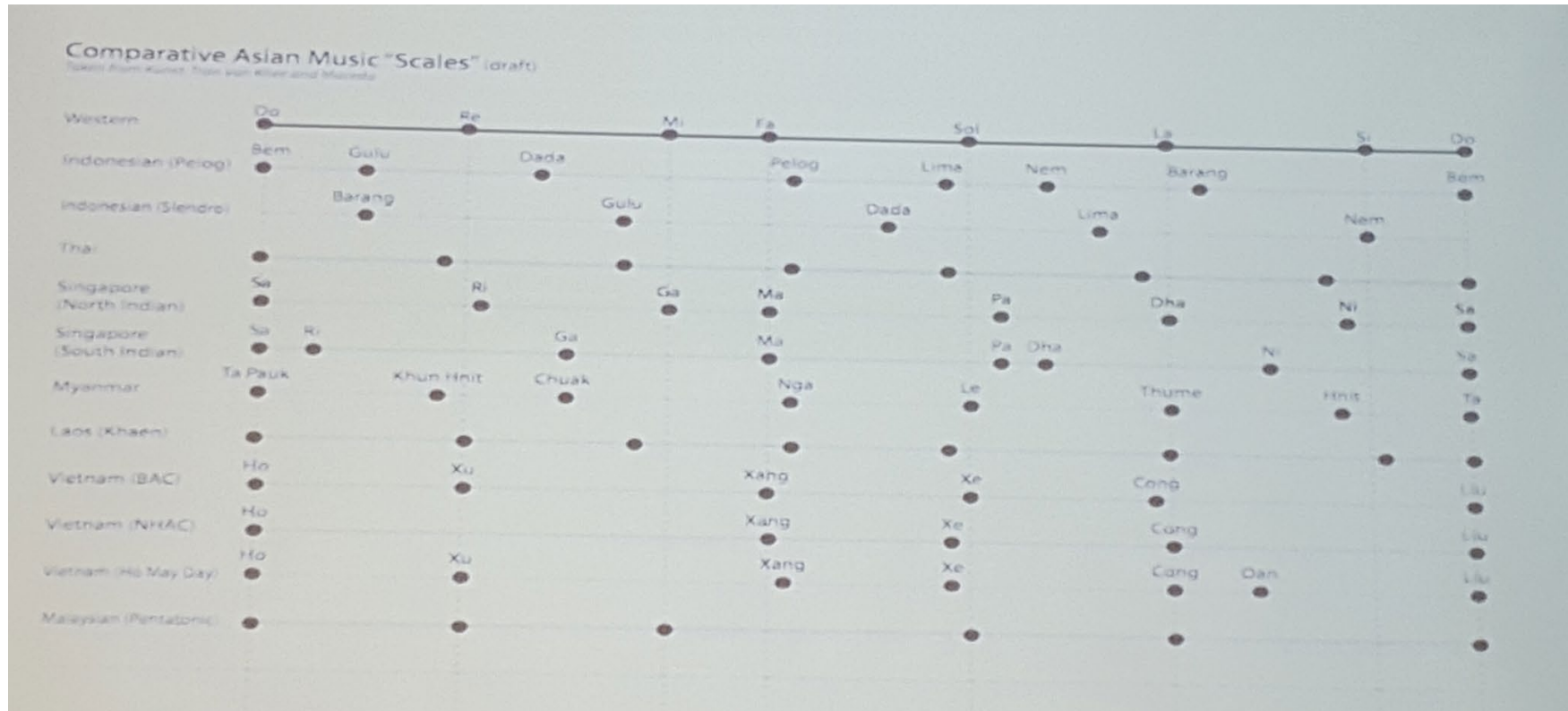
R1.3. (Meta)data meet domain-relevant community standards

- 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!

Camera manufacturer	Canon
Camera model	Canon EOS 1200D
Author	Praveen. P
Exposure time	1/60 sec (0.016666666666667)
F-number	f/11
ISO speed rating	200
Date and time of data generation	22:29, 22 November 2018
Lens focal length	41 mm
Show extended details	

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

- Do Re Mi Fa Sol La Si Do
 - Does not have to be a standard for everyone!



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

- Follow standards and domain specific conventions
- Examples
 - Sharing COBOL code
 - with Data Science students?
 - with mainframe operators?

```

000024
000025 PROCEDURE DIVISION.
000026 0001-MAIN.
000027     INSPECT FUNCTION REVERSE(STR-1)
000028         TALLYING WS-LEN1 FOR LEADING SPACES.
000029     COMPUTE WS-LEN = LENGTH OF STR-1 - WS-LEN1.
000030     DISPLAY WS-LEN.
000031     MOVE 1 TO I.
000032     MOVE WS-LEN TO J.
000033     PERFORM REV-PARA WS-LEN TIMES.
000034     DISPLAY STR-1.
000035     DISPLAY STR-2.
000036     GOBACK.
000037 REV-PARA.
000038     MOVE STR-1(J:1) TO STR-2(I:1).
000039     SUBTRACT 1 FROM J.
000040     ADD 1 TO I.
000041     EXIT.
***** Bottom of Data *****

```



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 reporting date .
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

SUMMARY

Summary

- There is no silver bullet that makes data FAIR immediately!
- Making data FAIR is a shared process and a joint responsibility
 - Infrastructure operators and management
 - Provide infrastructure, e.g. PIDs, repositories, etc.
 - Research communities
 - Define common vocabularies, standards, models, etc.
 - Individual researchers
 - Use systems and standards, prepare documentation, manage the data
- Machine-actionability of information is essential



SAVE THE DATE

REPOSITORIES FOR RESEARCH DATA (INVENIO RDM)

<https://forschungsdaten.at/en/fair-data-austria/materials/research-data-repositories/>

18th November 2021, 10:00-11:30, online

DATA MANAGEMENT PLANS AT TECHNICAL UNIVERSITIES

<https://forschungsdaten.at/en/fair-data-austria/materials/data-management-plans-dmps-at-technical-universities/>

25th November 2021, 10:00-11:30, online

INTRODUCTION TO GIT

<https://forschungsdaten.at/en/fair-data-austria/materials/introduction-to-git/>

7th December 2021, 10:00-11:30, online



**THANK YOU FOR YOUR
PARTICIPATION & ATTENTION!**