CLUSTER FORSCHUNGSDATEN

FAIR DATA AUSTRIA

LET‘S MAKE OUR DATA FAIR!

19.10.2021

HOSTED BY TU WIEN,

CENTER FOR RESEARCH DATA MANAGEMENT

TU WIEN TECHNISCHE UNIVERSITÄT WIEN

Tomasz Miksa. Let’s make our data FAIR. DOI: 10.25365/phaidra.306. CC-BY4.0

FAIR DATA AUSTRIA

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/

TU Graz
TU WIEN TECHNISCHE UNIVERSITÄT WIEN
universität wien
universität innsbruck
Medizinische Universität Graz
]a[akademie der bildenden künste wien

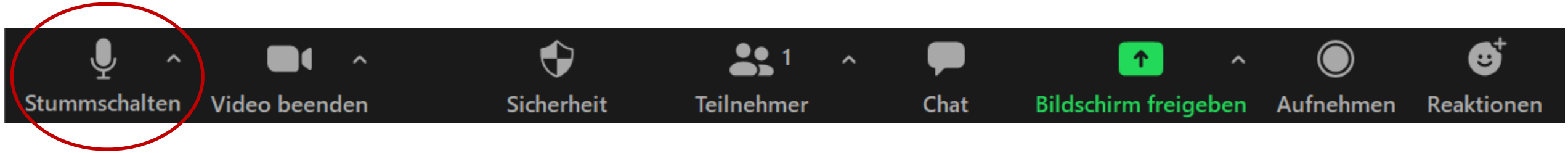
FAIR DATA AUSTRIA

AGENDA

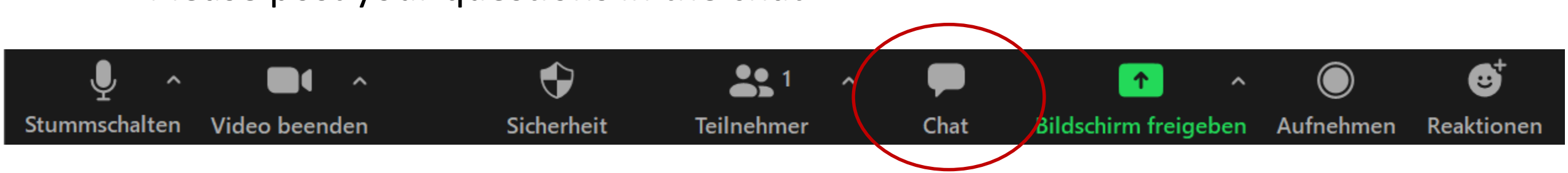
10.00 – 10.10	Welcome and introduction	Barbara Sánchez
10.10 – 11.00	Let‘s make our data FAIR!	Tomasz Miksa
11.00 – 11.30	Questions & Answers	Barbara Sánchez

FORMALITIES

Please turn off your microphone to avoid background noise



Please post your questions in the chat



FAIR DATA AUSTRIA

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

FAIR DATA AUSTRIA

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

TED

FAIR DATA AUSTRIA

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.

FAIR DATA AUSTRIA

Hans Rosling

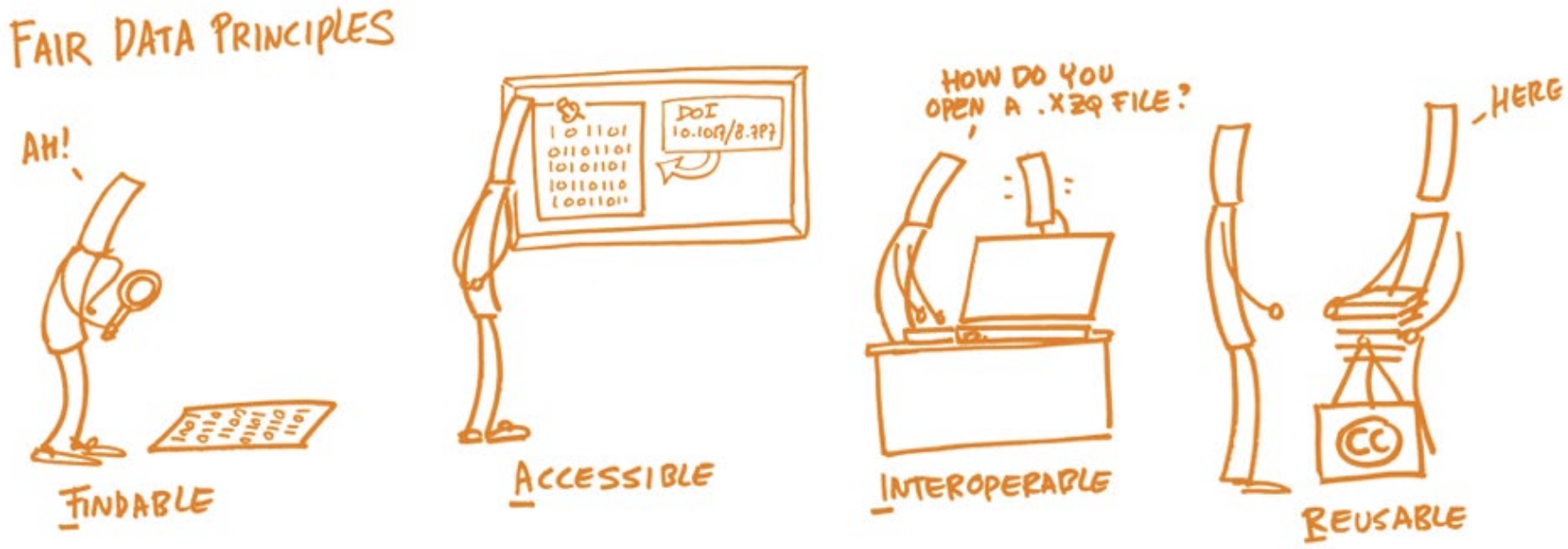
- “Some countries accept that their databases can go out on the world. But what we really need is, of course, a search function, a search function where **we can copy the data up to a searchable format** and get it out in the world.”

- “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 DATA AUSTRIA

FAIR PRINCIPLES

FAIR Principles (very simplified :))



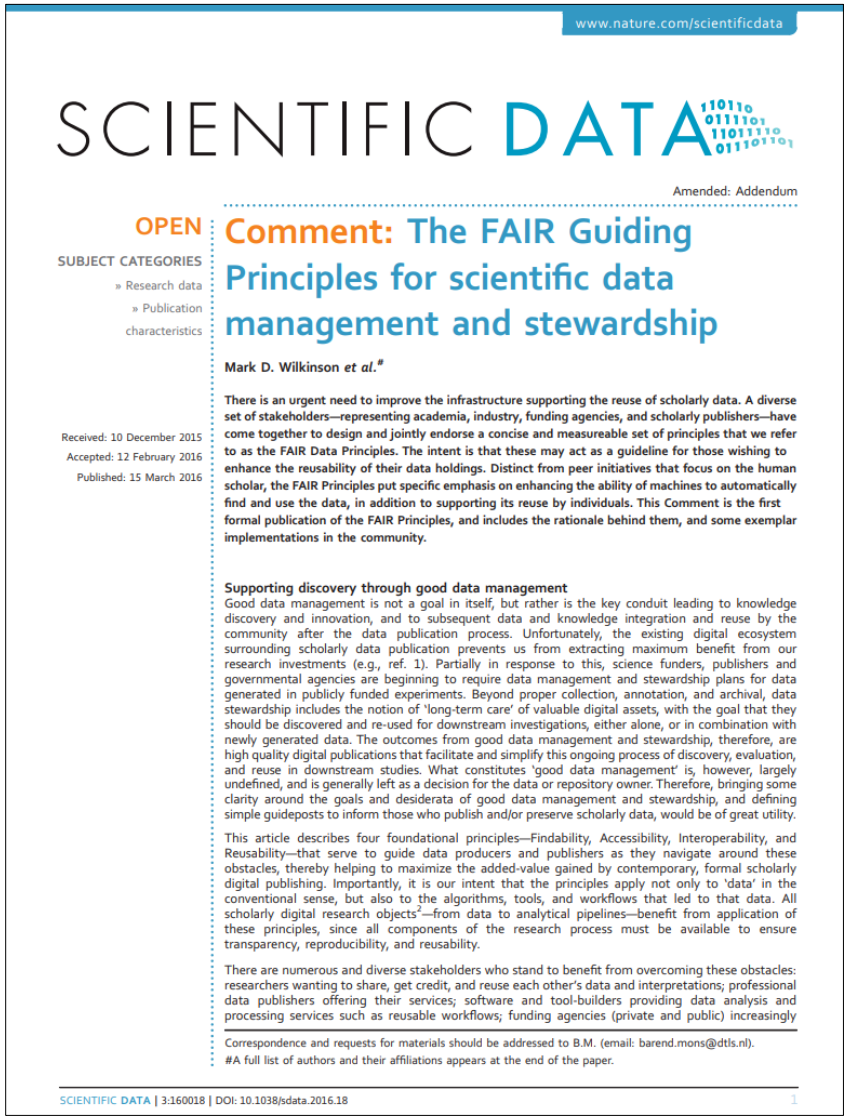
FAIR DATA AUSTRIA

FAIR Principles (less simplified :))

https://www.ands.org.au/working-with-data/fairdata/training

FAIR principles

Findable
Accessible
Interoperable
Reusable



www.nature.com/scientificdata

SCIENTIFIC DATA

Amended: Addendum

OPEN

SUBJECT CATEGORIES
» Research data
» Publication characteristics

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

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

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 measureable 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[2]—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@dtls.nl).
#A full list of authors and their affiliations appears at the end of the paper.

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

GO FAIR

FAIR Principles | Implementation Networks | News | Events | Resources | About GO FAIR

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

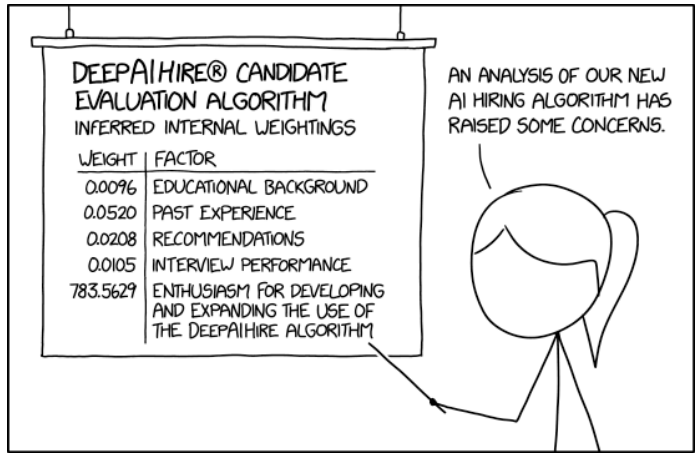
Accessible
Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.

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

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

FAIR vs fair

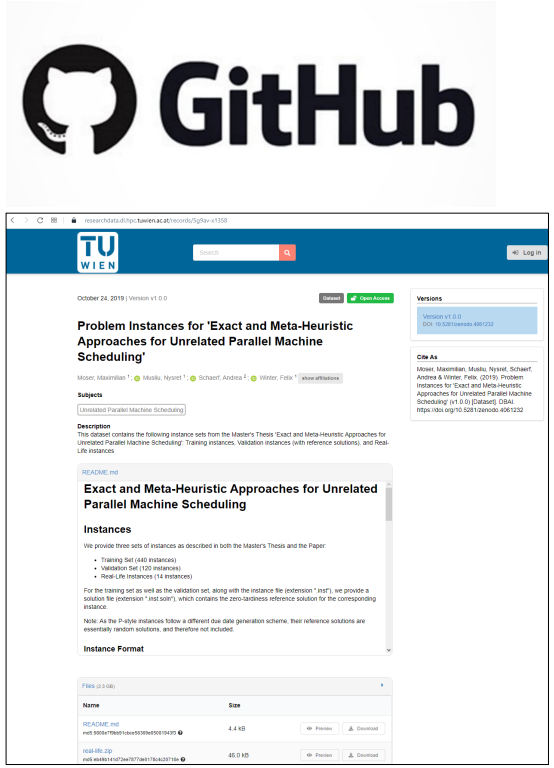
- FAIR principles =/= Algorithmic fairness
- To be FAIR
  - To apply/use FAIR principles
  - Focus on how data is managed, etc.
- To be fair
  - Evade bias
  - Focus on design and implementation



https://xkcd.com/2237/

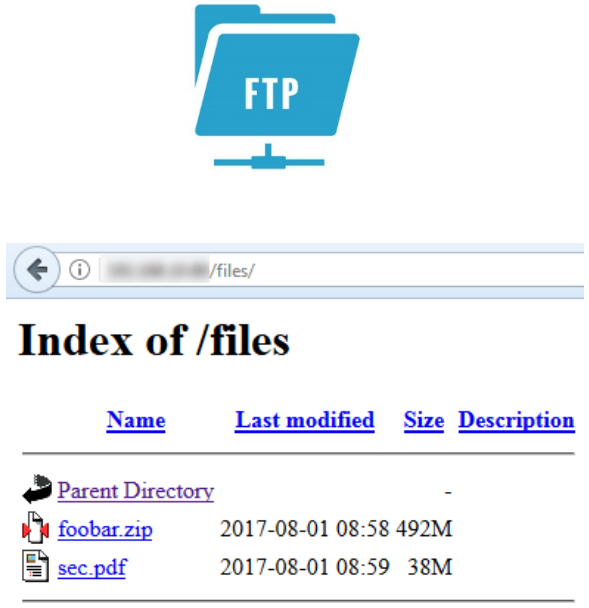
Findable – simplified examples

Yes



Data repository

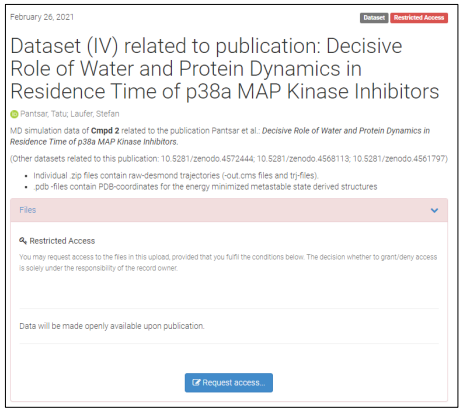
No



Personal website

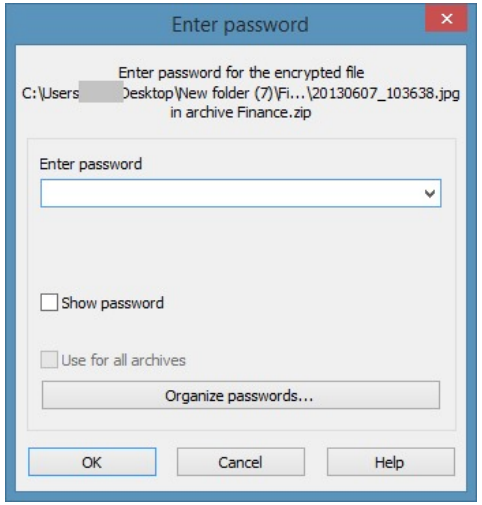
Accessible – simplified examples

Yes



Restricted access, but a clear way to request access

No



FAIR DATA AUSTRIA

Interoperable – simplified examples

Yes

- XML following known XSD Schema
- MP3 for audio recordings

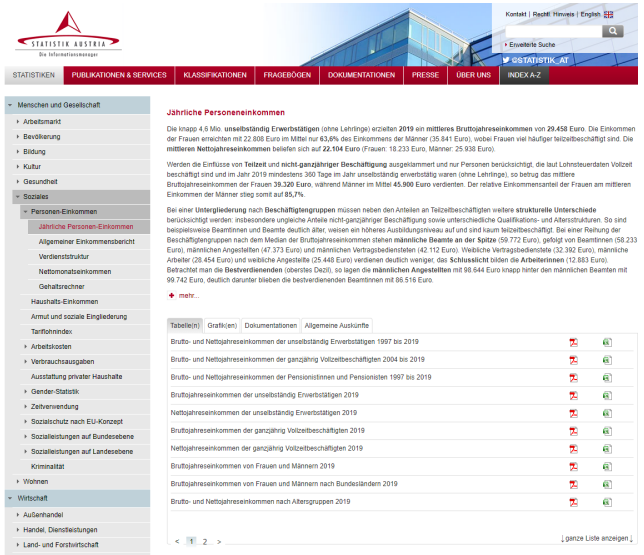
No

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

FAIR DATA AUSTRIA

Reusable – simplified examples

Yes



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

No

Provenance and permissions not clear

FAIR DATA AUSTRIA

MACHINE-ACTIONABILITY

FAIR DATA AUSTRIA

Machine actionability

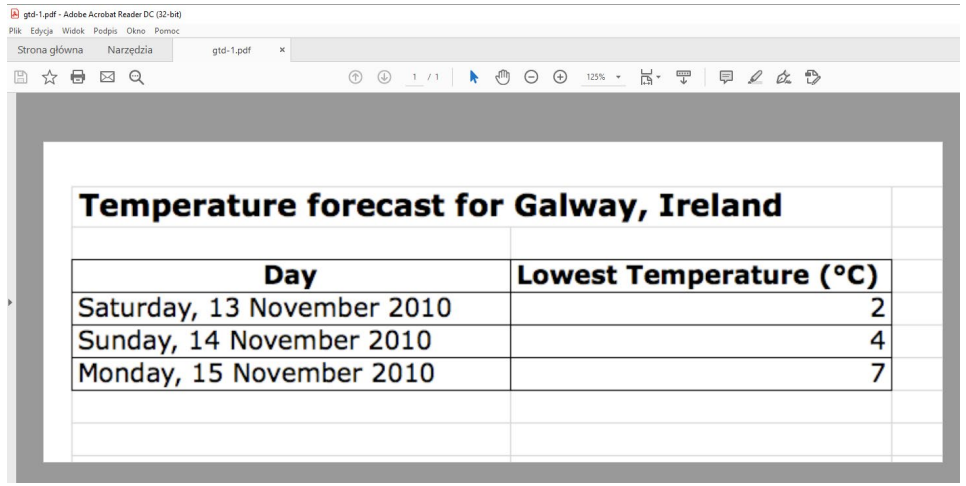
- “the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention” https://www.go-fair.org/fair-principles

- “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

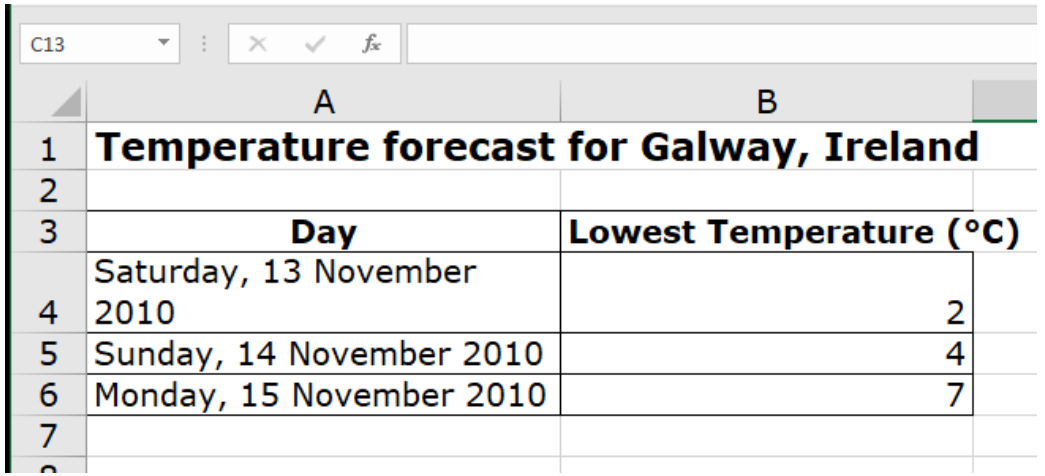
Not machine-actionable



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



	A	B
1	**Temperature forecast for Galway, Ireland**	
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		

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

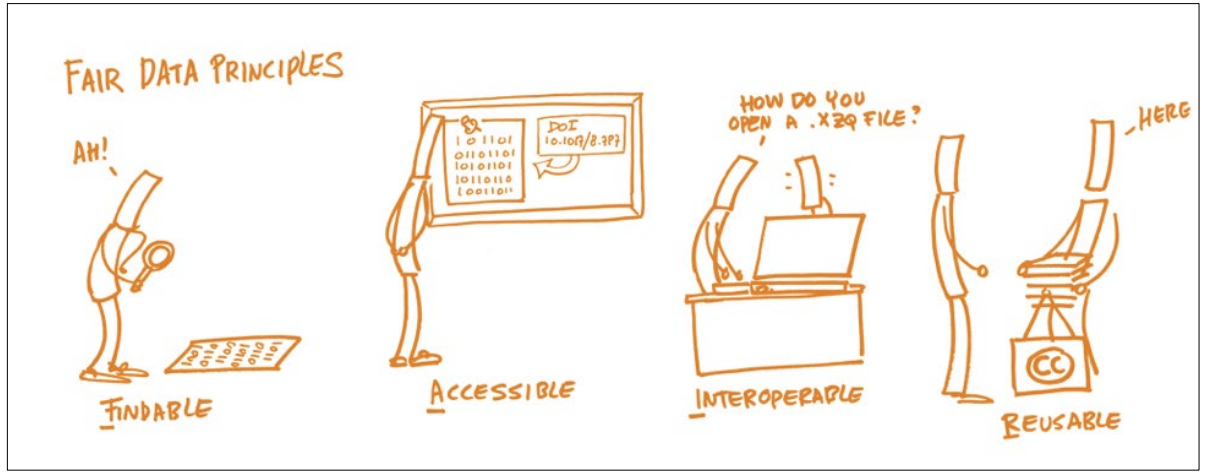
Unit definition
and link to common definition

Values, types, link to forecast

http://graphite.ecs.soton.ac.uk/browser/?uri=http://5stardata.info/en/examples/gtd-5/

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 DATA AUSTRIA

FAIR IN DETAIL

FAIR DATA AUSTRIA

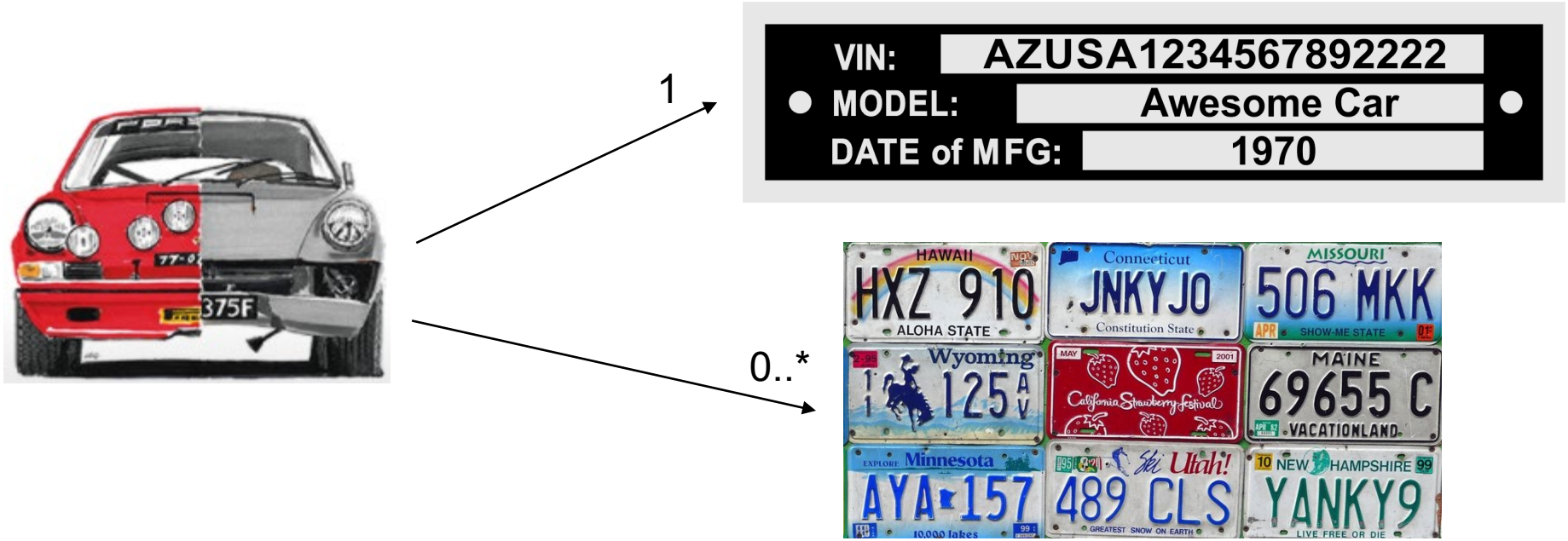
Findable

- 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**

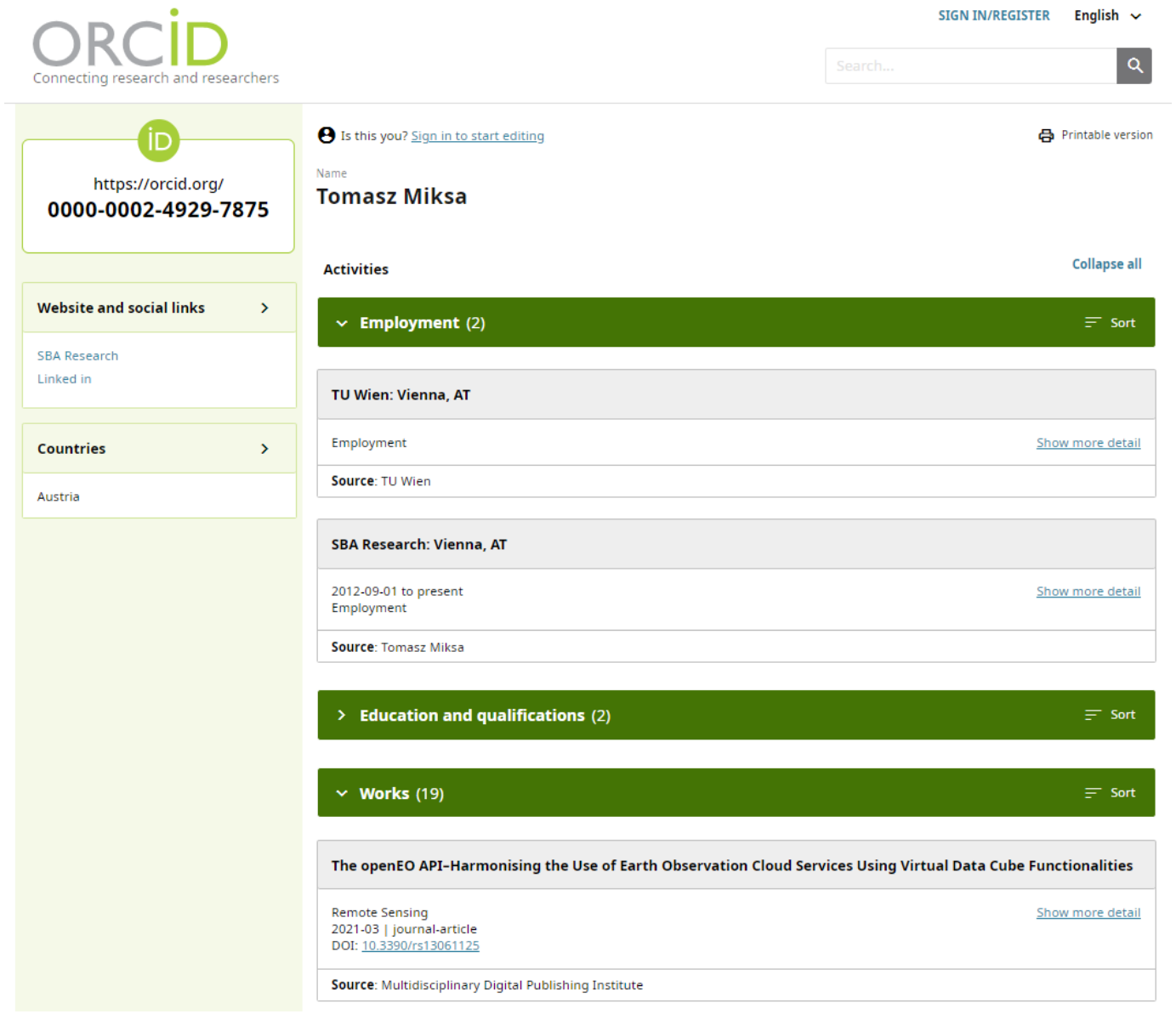
Persistent identifiers (F1)

Example

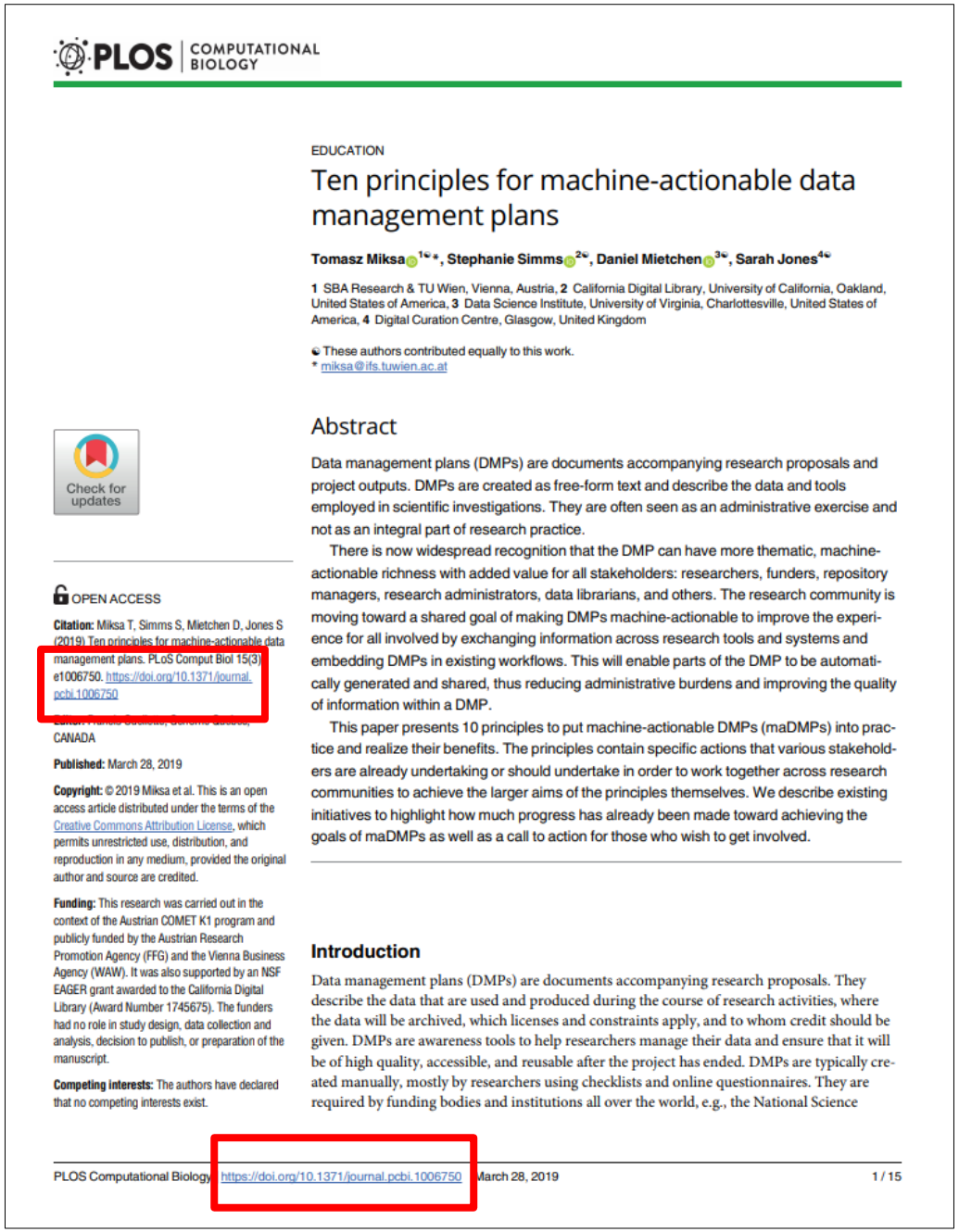
A car has only <u>one</u> **VIN** (PID), but can have <u>many</u> **number plates** over its lifetime (URL)



FAIR DATA AUSTRIA



ORCID assigned to a person



PLOS COMPUTATIONAL BIOLOGY

EDUCATION

Ten principles for machine-actionable data management plans

Tomasz Miksa[1☯*], **Stephanie Simms**[2☯], **Daniel Mietchen**[3☯], **Sarah Jones**[4☯]

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.

Check for updates

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

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, 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.

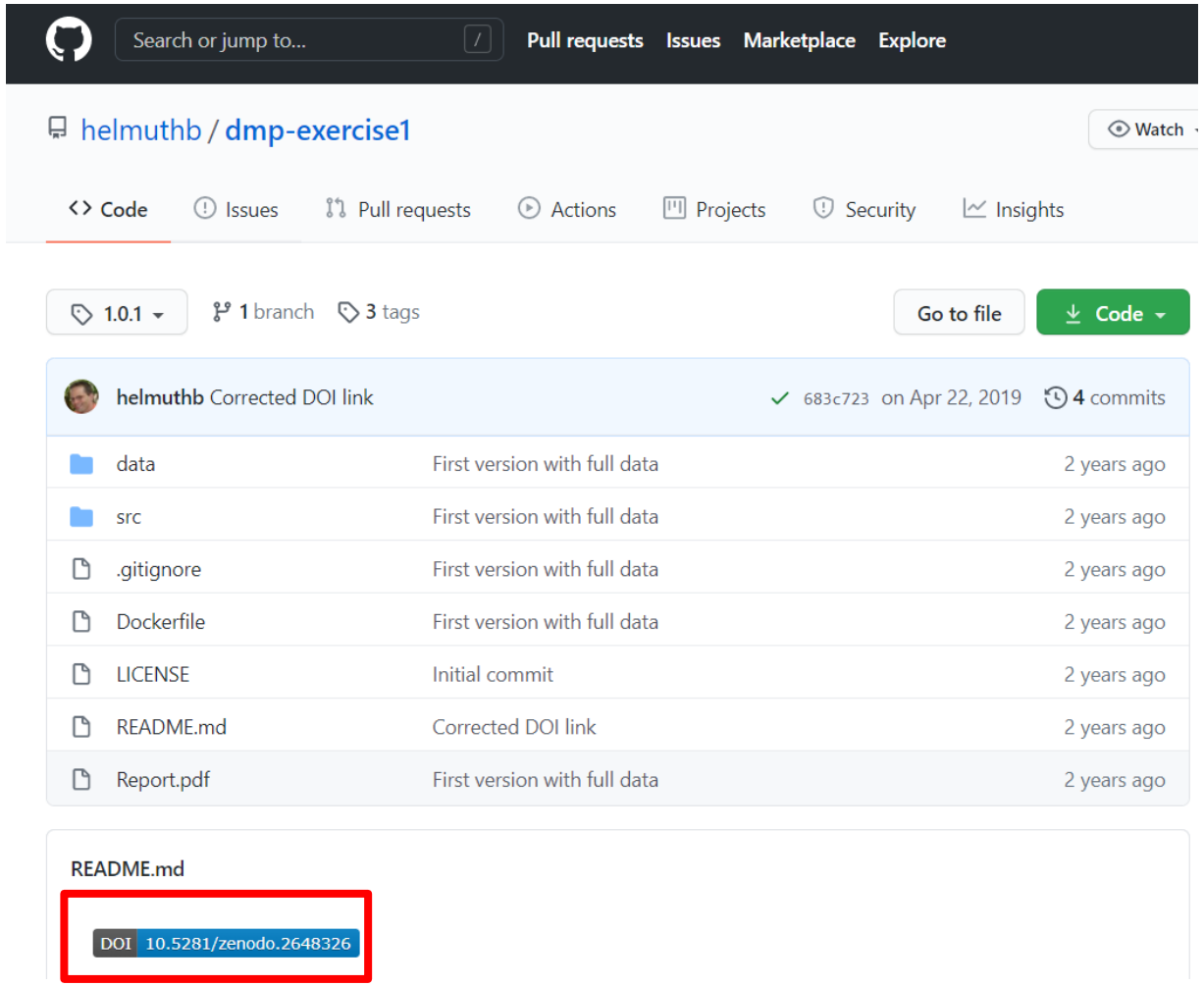
Competing interests: The authors have declared that no competing interests exist.

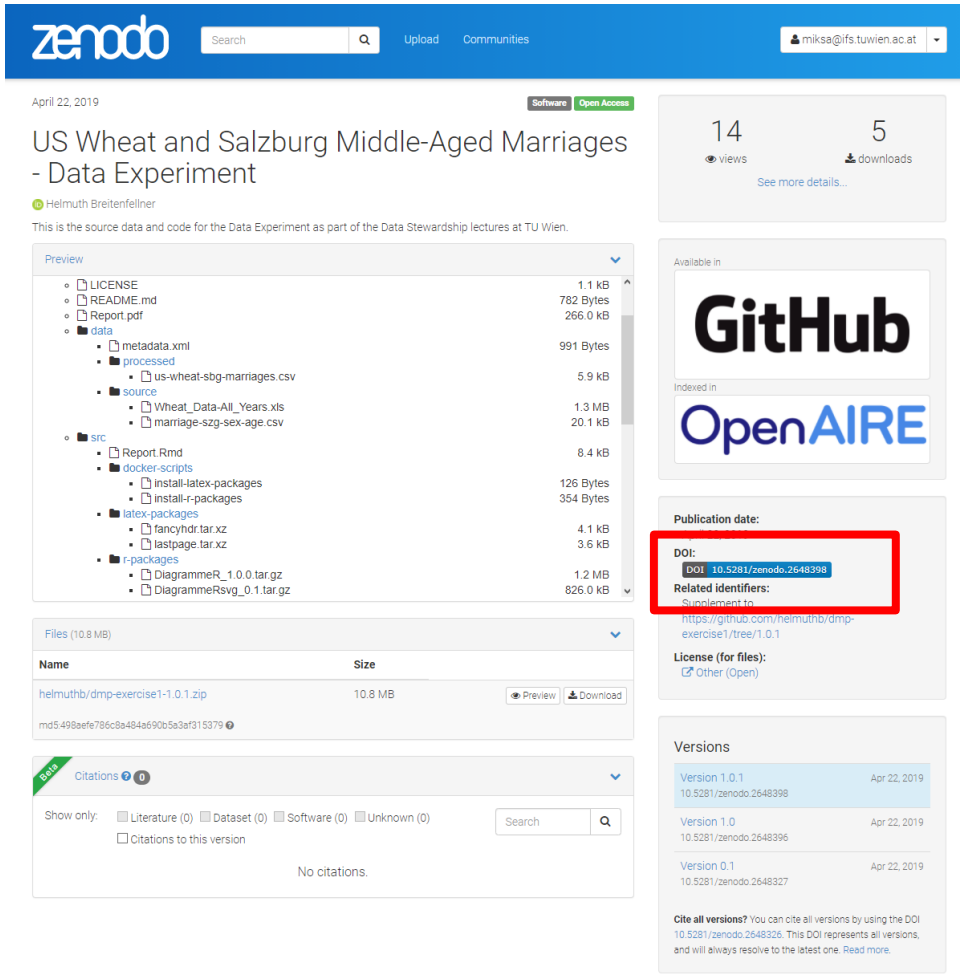
Introduction

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

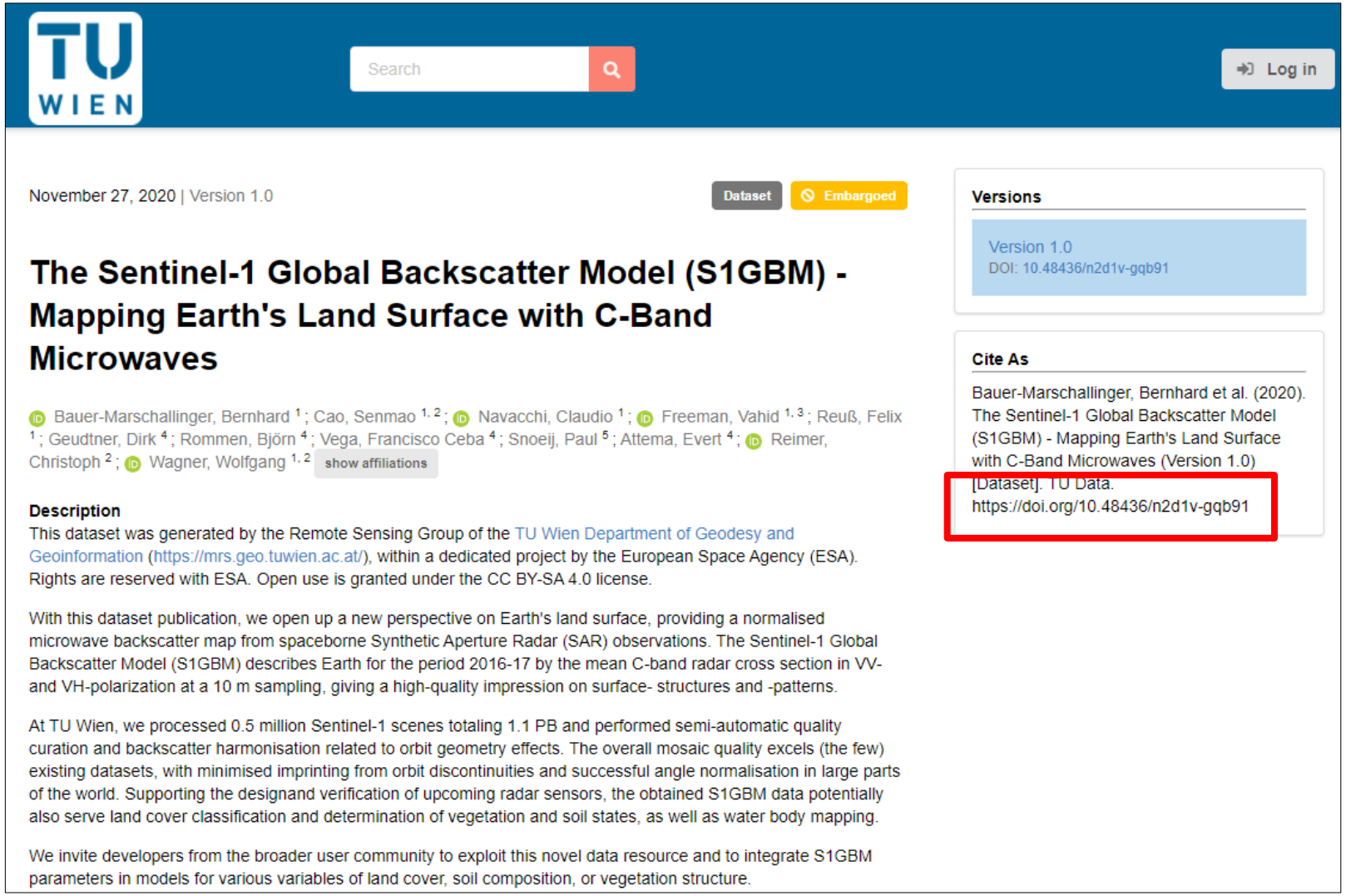
FAIR DATA AUSTRIA

DOI assigned to a publication





DOI assigned to code



DOI assigned to data

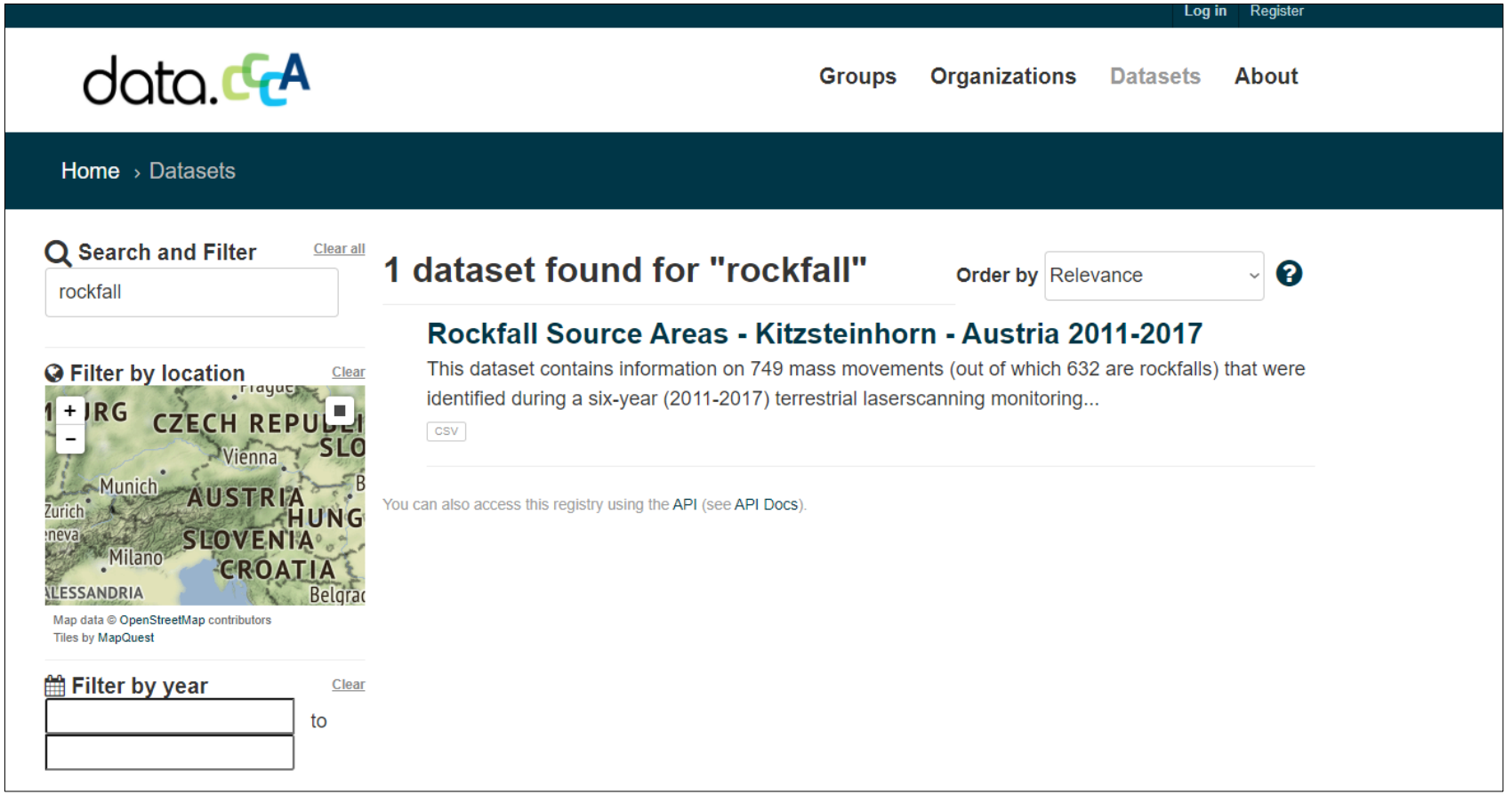
FAIR DATA AUSTRIA

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

FAIR DATA AUSTRIA

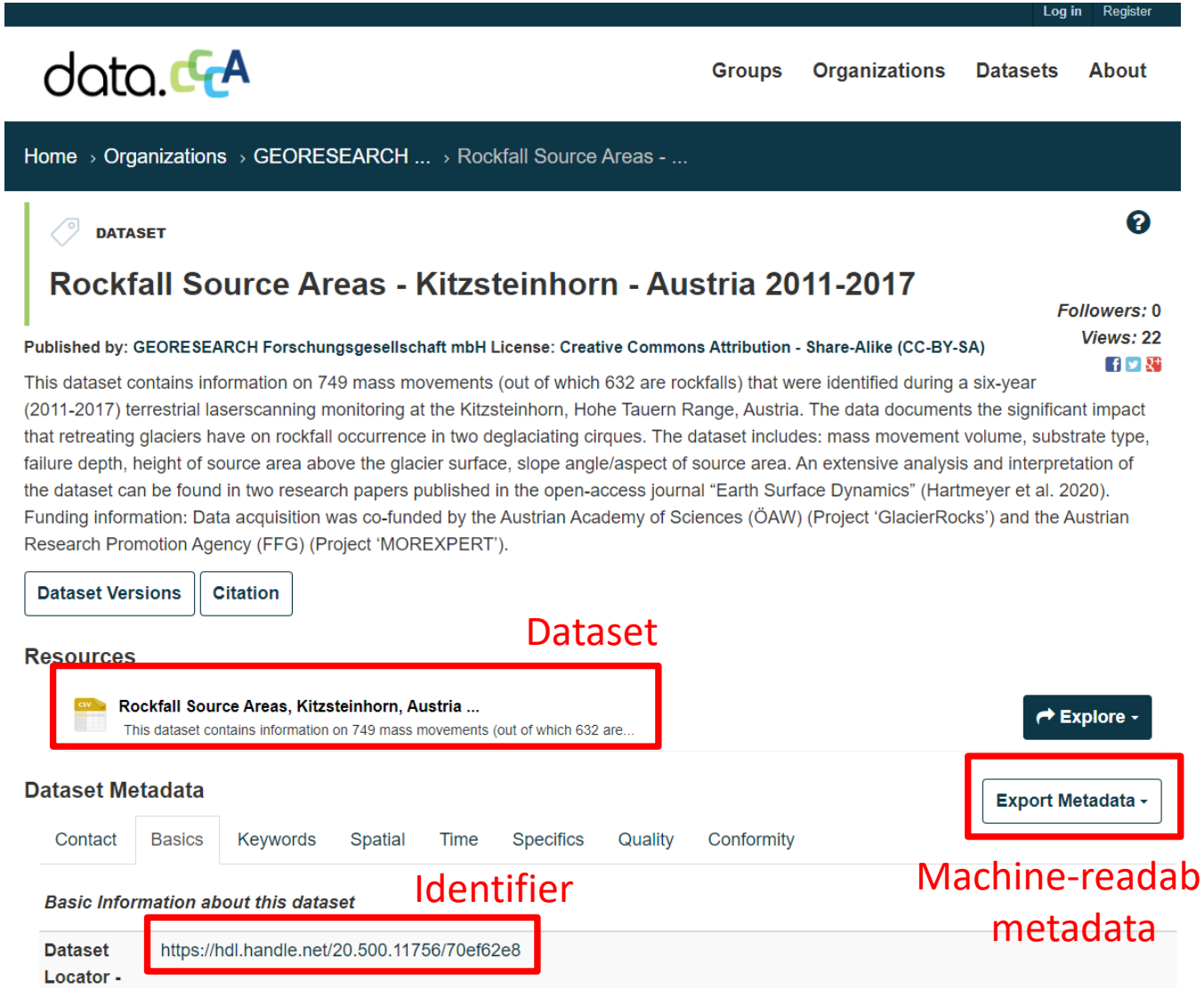
F4. (Meta)data are registered or indexed in a searchable resource



https://data.ccca.ac.at

FAIR DATA AUSTRIA

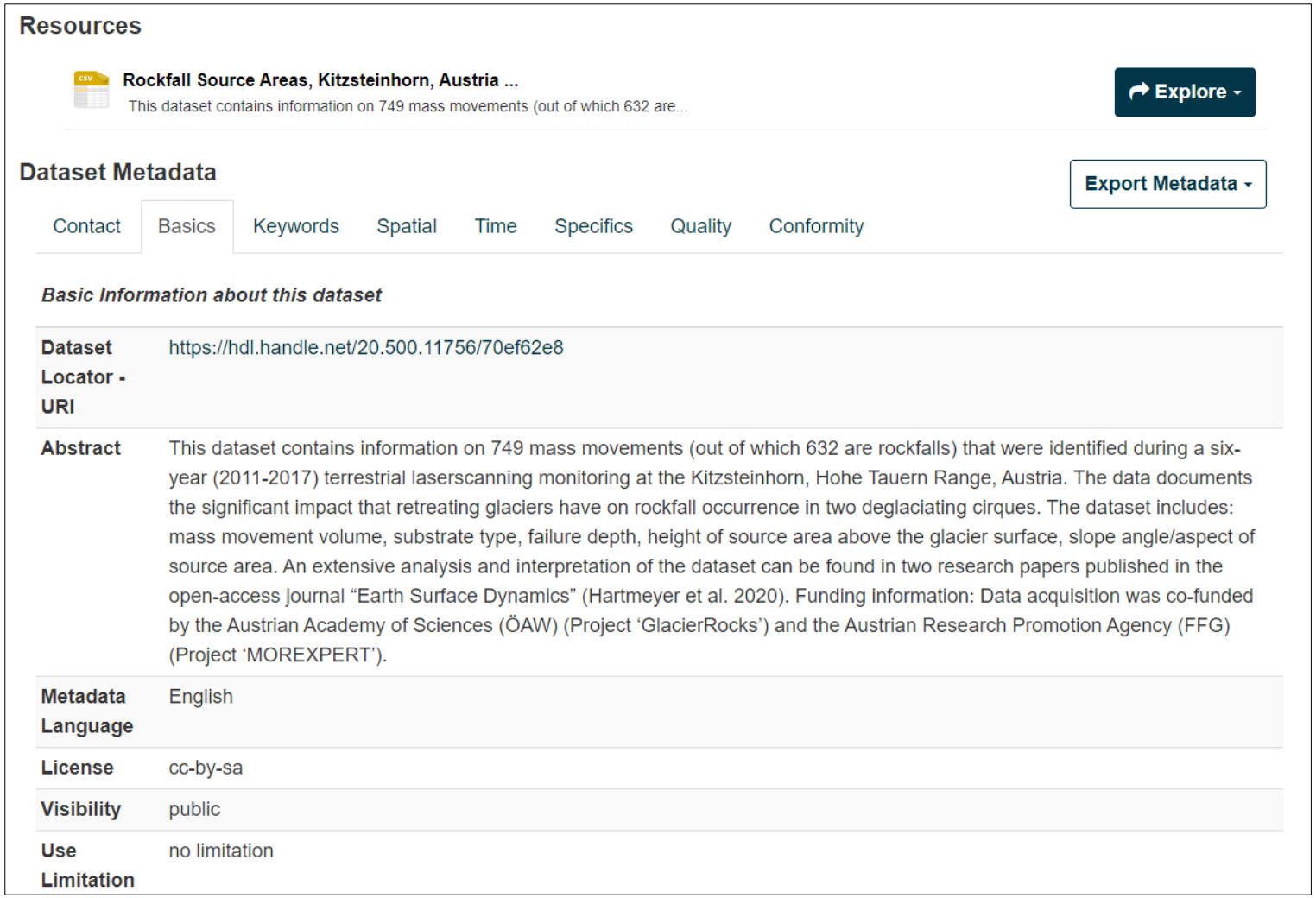
F3. Metadata clearly and explicitly include the identifier of the data they describe



https://hdl.handle.net/20.500.11756/70ef62e8

FAIR DATA AUSTRIA

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

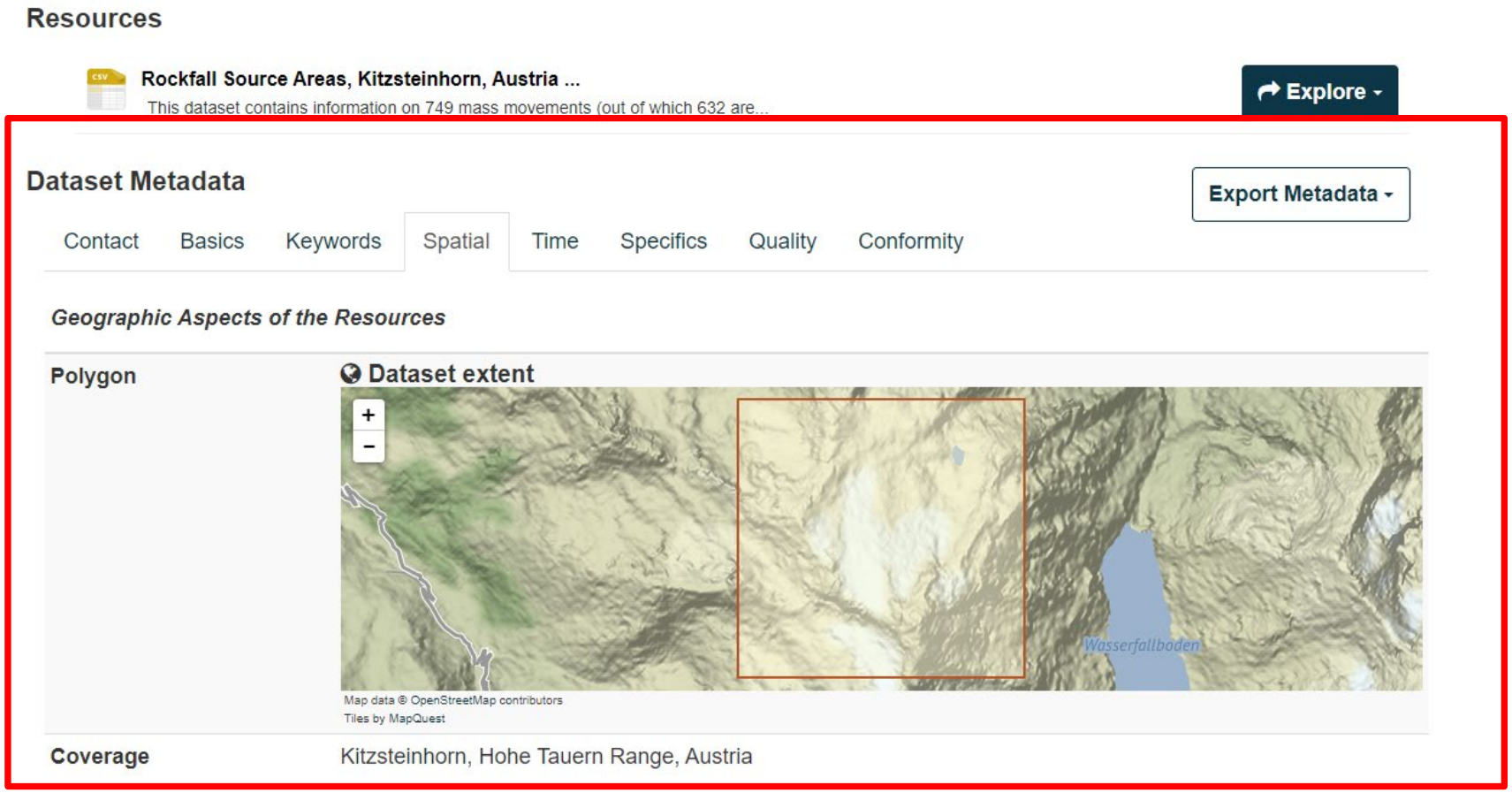
Contact | Basics | Keywords | Spatial | Time | Specifics | Quality | Conformity

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 ‘MOREXPERT’).
Metadata Language	English
License	cc-by-sa
Visibility	public
Use Limitation	no limitation

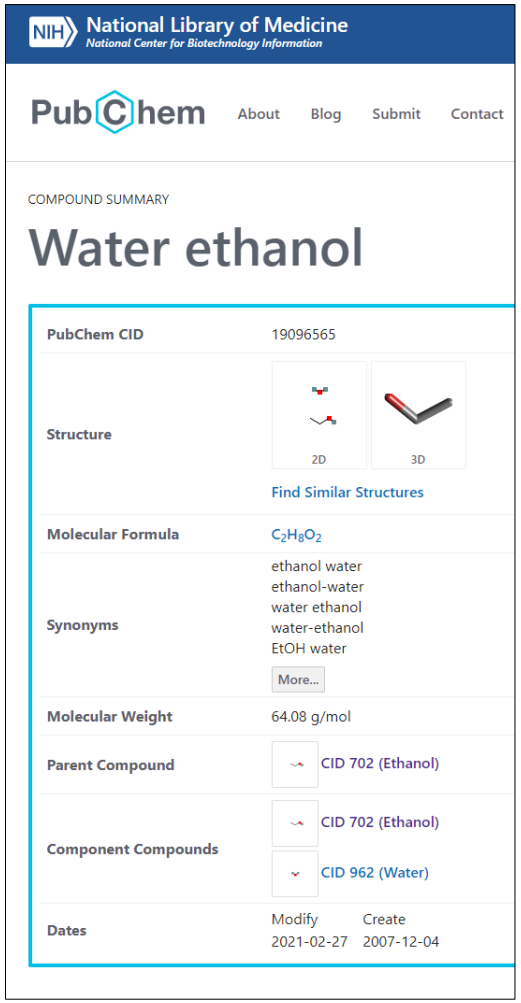
FAIR DATA AUSTRIA

F2. Data are described with rich metadata



Metadata - Chemistry

FAIR DATA AUSTRIA



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

IDGUHHHQCWSQLU-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)

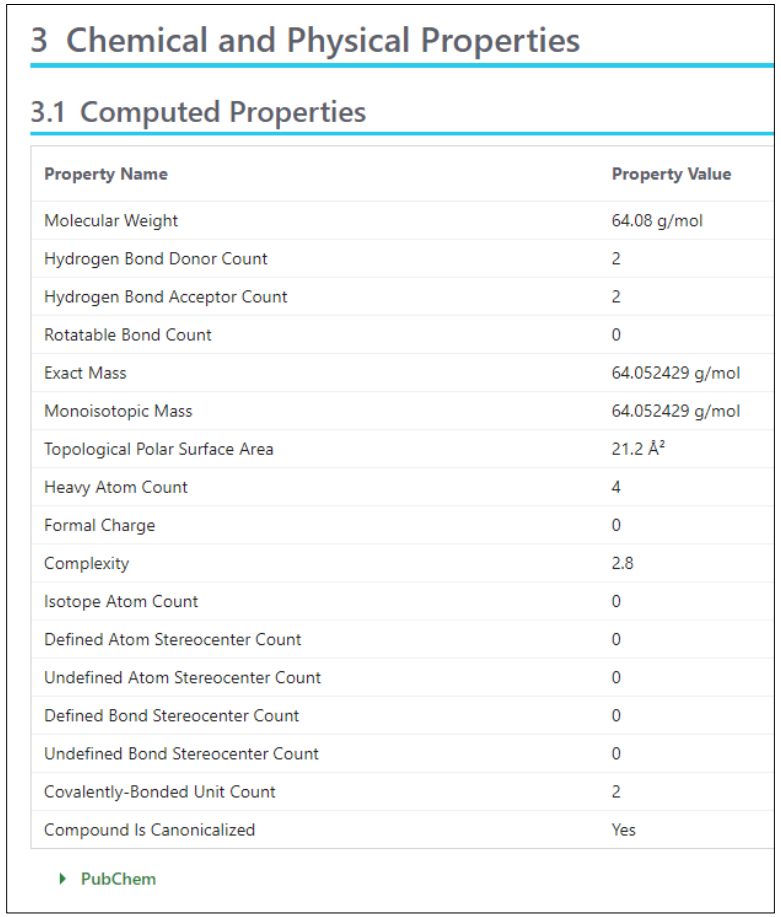
PubChem

2.2 Molecular Formula

C2H8O2

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 Å²
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

https://pubchem.ncbi.nlm.nih.gov/compound/19096565

FAIR DATA AUSTRIA

Accesible

- 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

FAIR DATA AUSTRIA

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

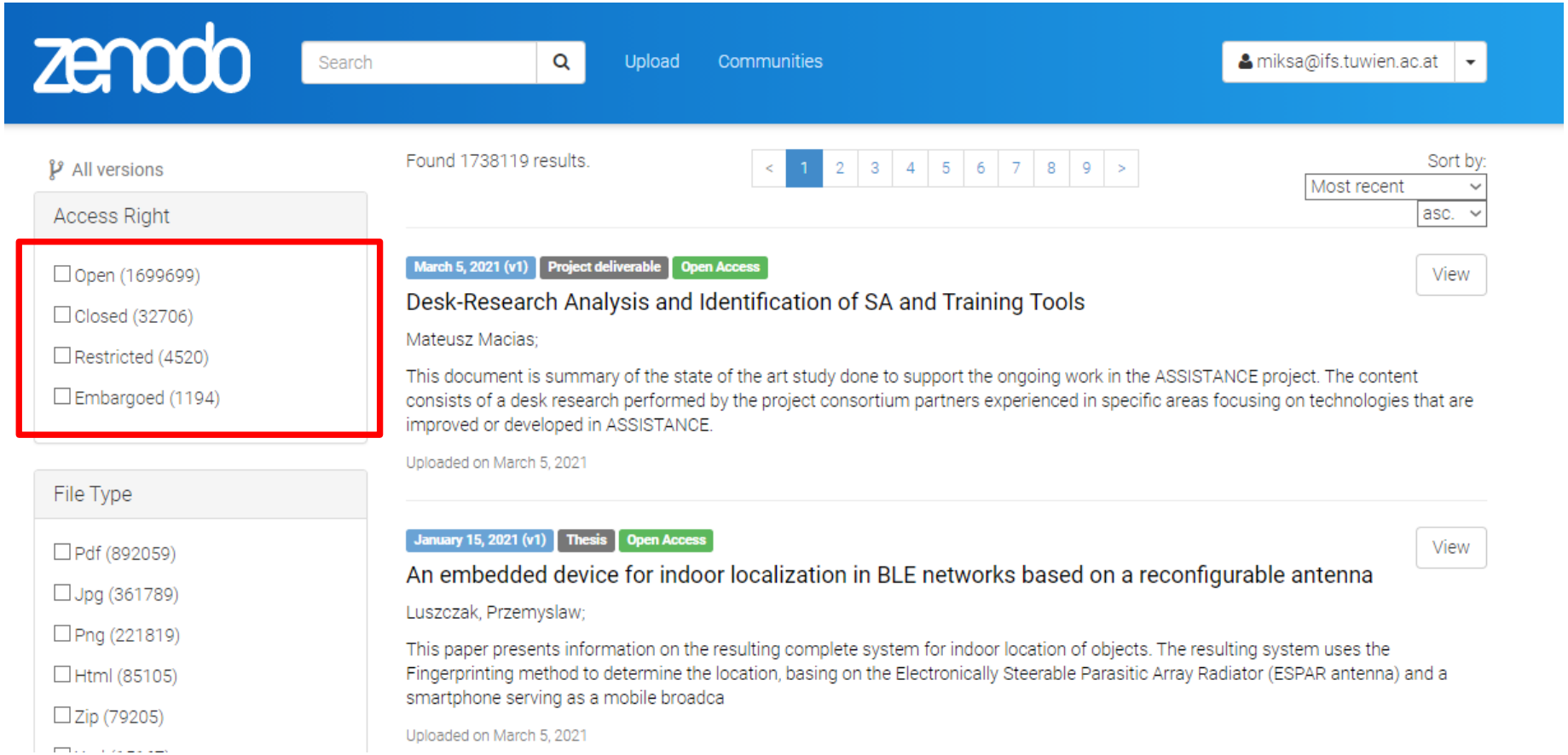
Layer	Name	Example protocols
7	Application Layer	HTTP, FTP, DNS, SNMP, Telnet
6	Presentation Layer	SSL, TLS
5	Session Layer	NetBIOS, PPTP
4	Transport Layer	TCP, UDP
3	Network Layer	IP, ARP, ICMP, IPSec
2	Data Link Layer	PPP, ATM, Ethernet
1	Physical Layer	Ethernet, USB, Bluetooth, IEEE802.11

FAIR DATA AUSTRIA

A1.2 protocol allows for authentication and authorisation

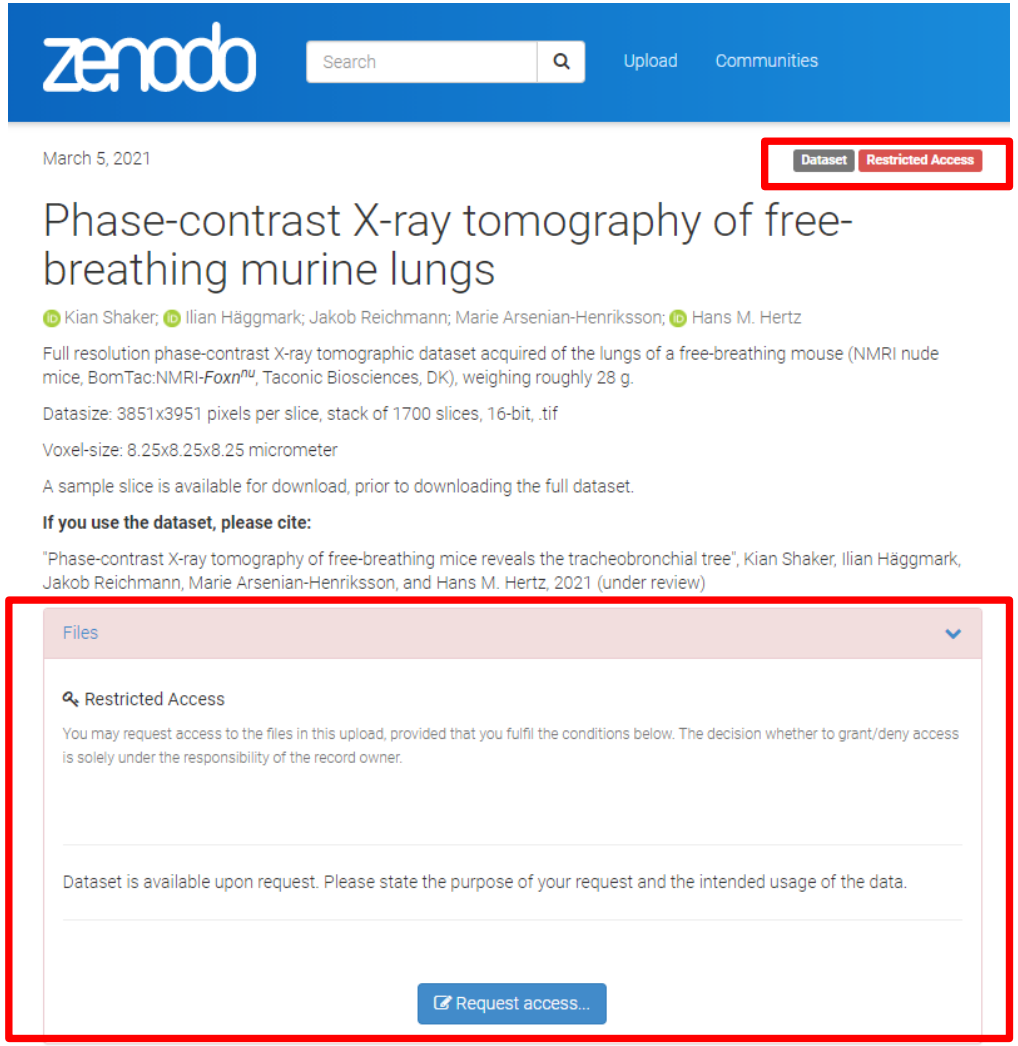
- Protected and private data can be FAIR
- Possible types of access
  - Open – everyone has access
  - Shared or restricted – only a selected/ invited group of people can access
  - Closed – only the owner has access

Accessible - example



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

Accessible - example



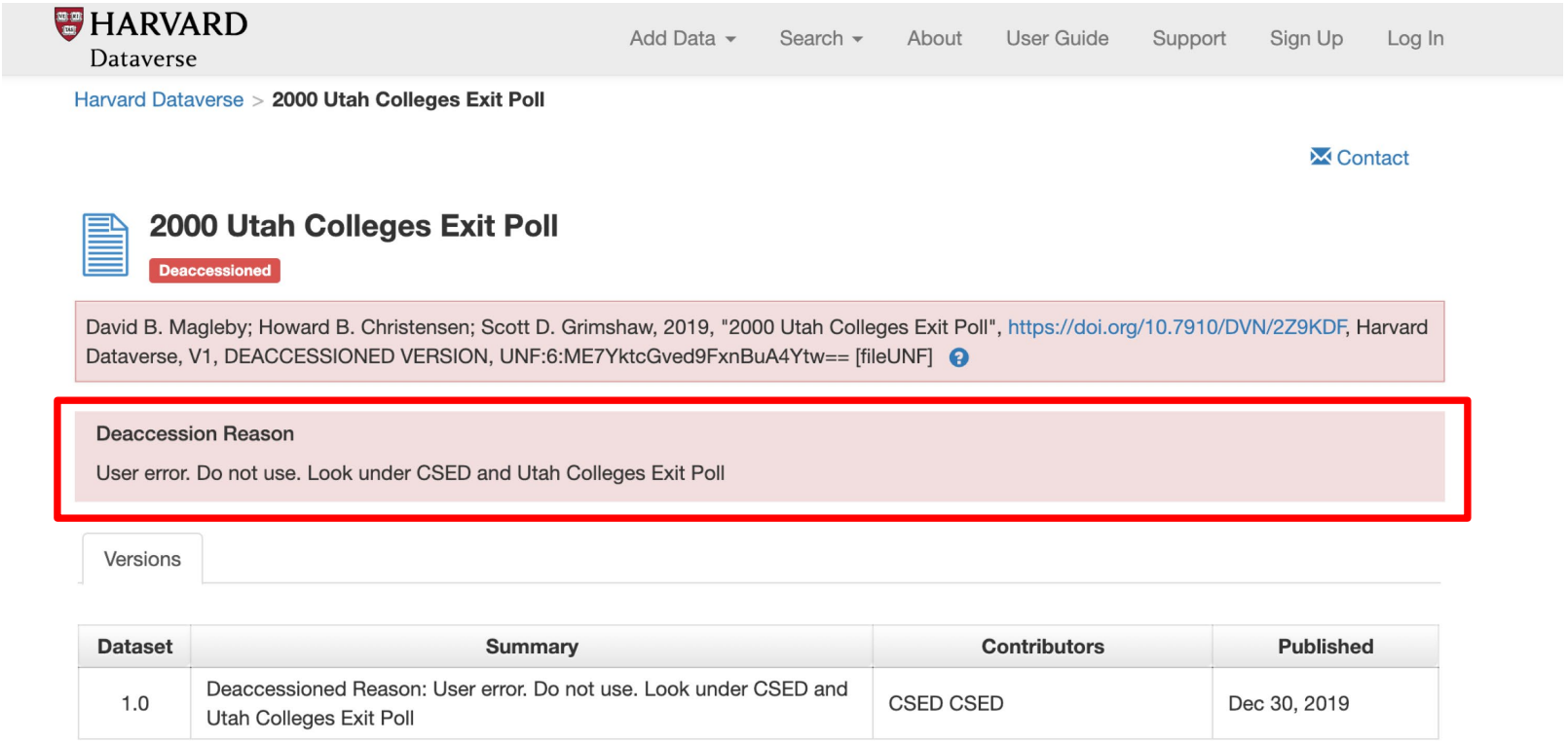
Accessible (A1) - comments

- FAIR data =/= 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”

FAIR DATA AUSTRIA

Tombstone pages (A2)

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



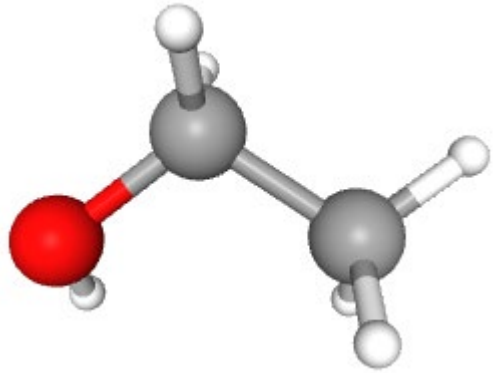
FAIR DATA AUSTRIA

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	*London Biggin Hill Airport*	Public
Greater London	EGML		Damyns Hall Aerodrome	Private
Greater London	EGLL	LHR	***Heathrow Airport***	Public
Greater London	EGWU	NHT	*RAF Northolt*	Military
Greater London	EGLC	LCY	***London City Airport***	Public
Greater London	EGLW		*London Heliport*	Public

- Controlled vocabularies: IATA and ICAO

https://en.wikipedia.org/wiki/List_of_airports_in_the_United_Kingdom_and_the_British_Crown_Dependencies

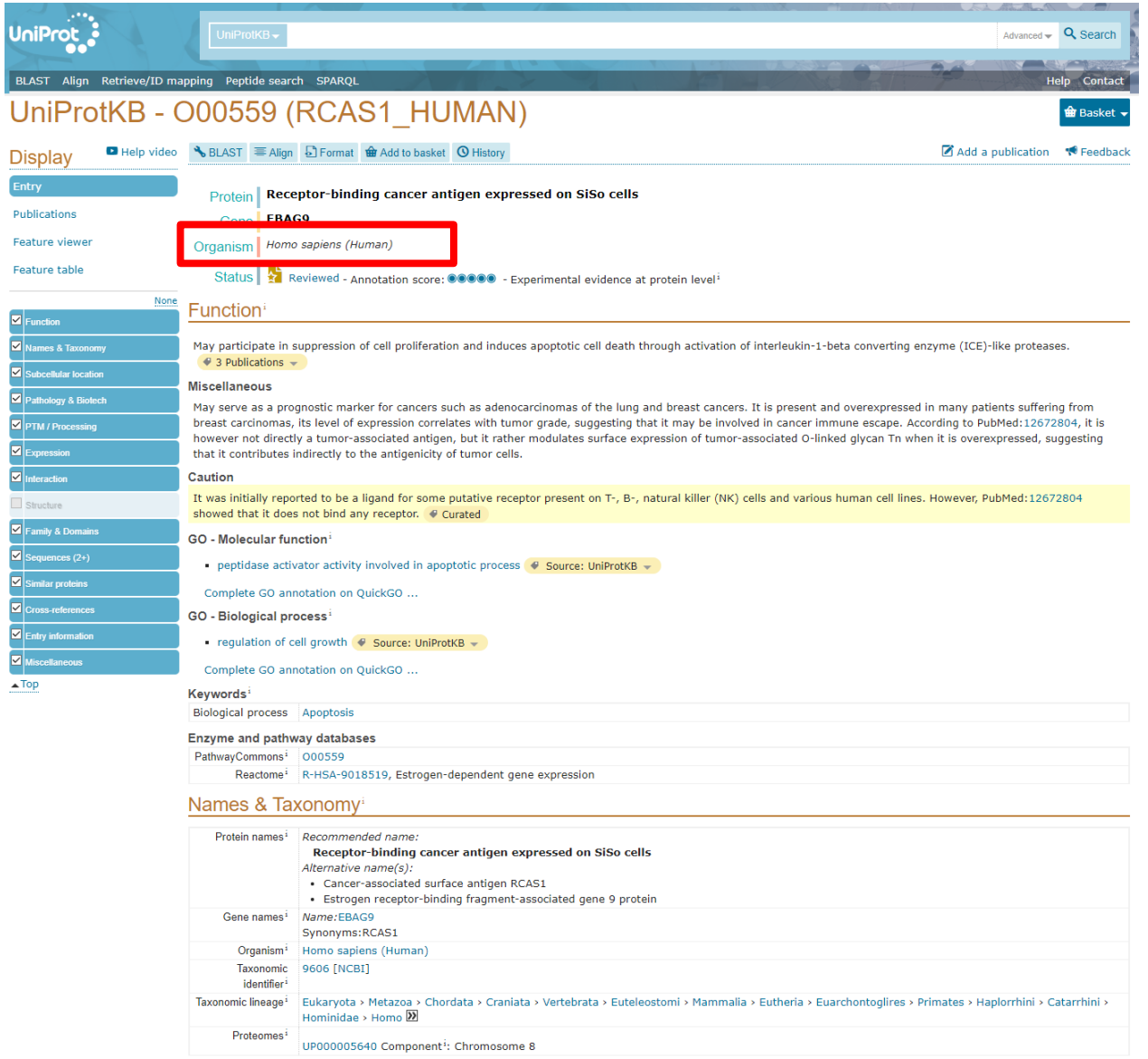
FAIR DATA AUSTRIA

Vocabularies

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

Vienna	**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), وین (Persian)

Vocabularies (I2)



https://www.uniprot.org/uniprot/O00559

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 cress)

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.

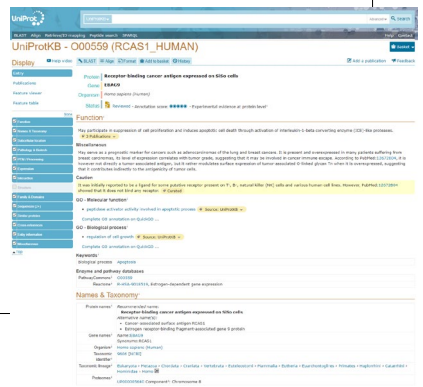
Related documents

Taxonomy

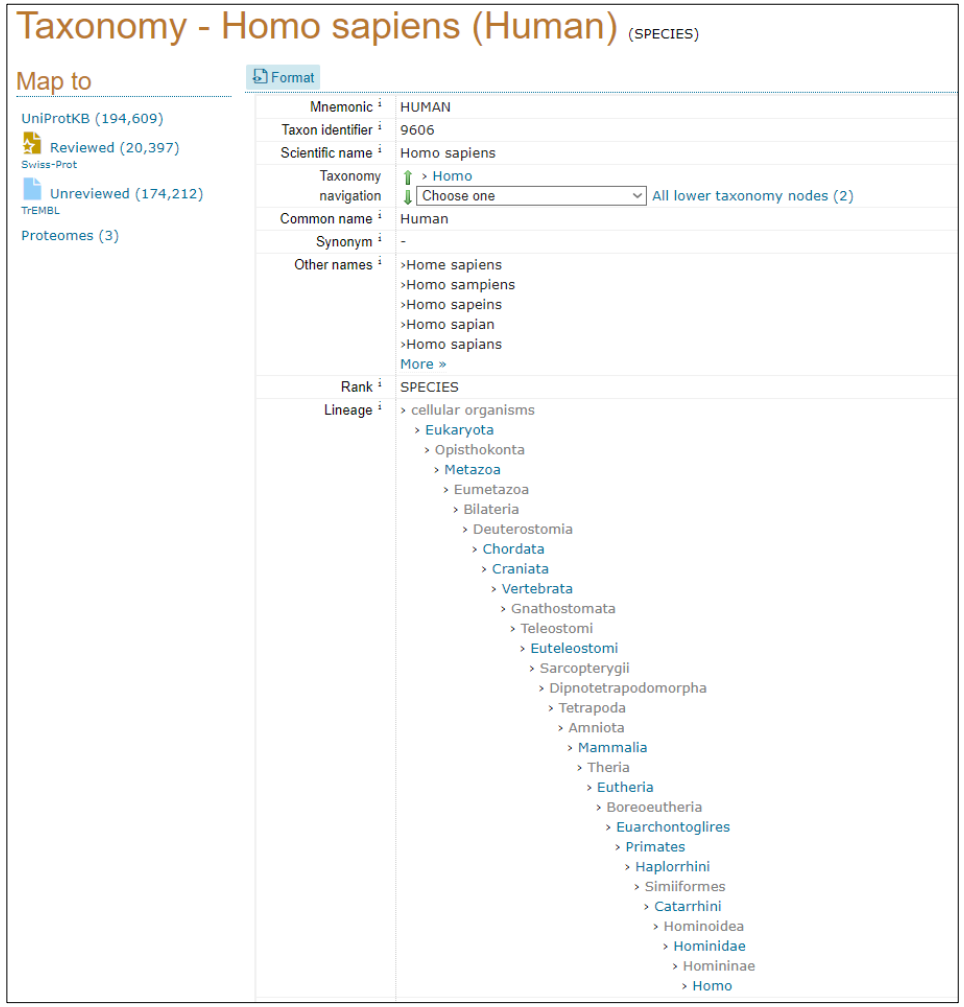
Controlled vocabulary of species

What are proteomes?

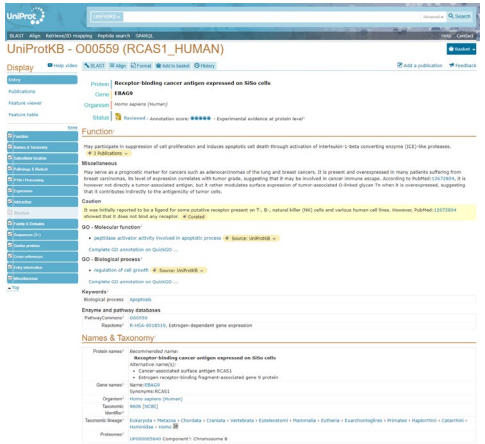
What are reference proteomes?



Vocabularies (I2)

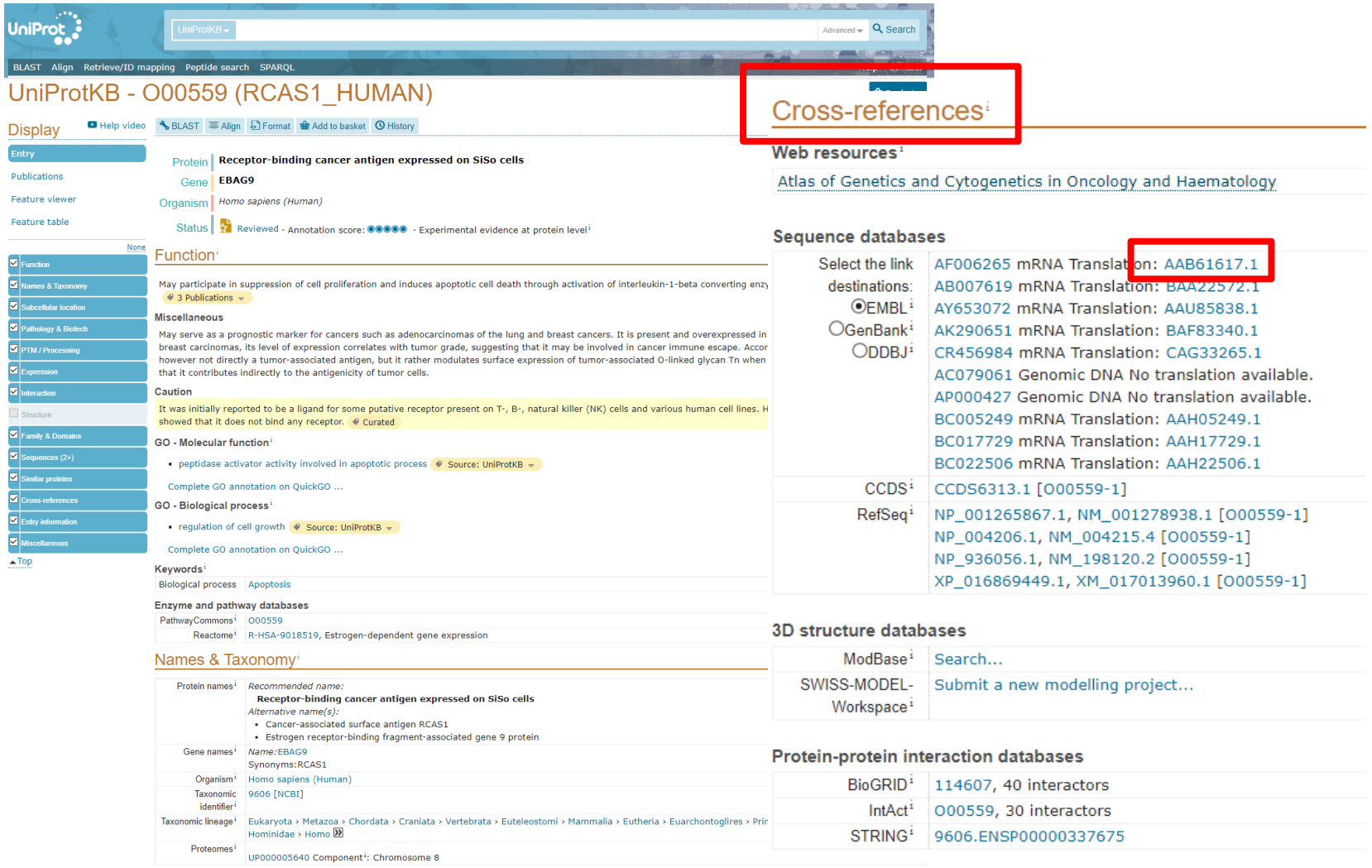


Each metadata **value** comes from a controlled vocabulary – no free form answers.



Qualified References (I3)

FAIR DATA AUSTRIA



FAIR DATA AUSTRIA

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**

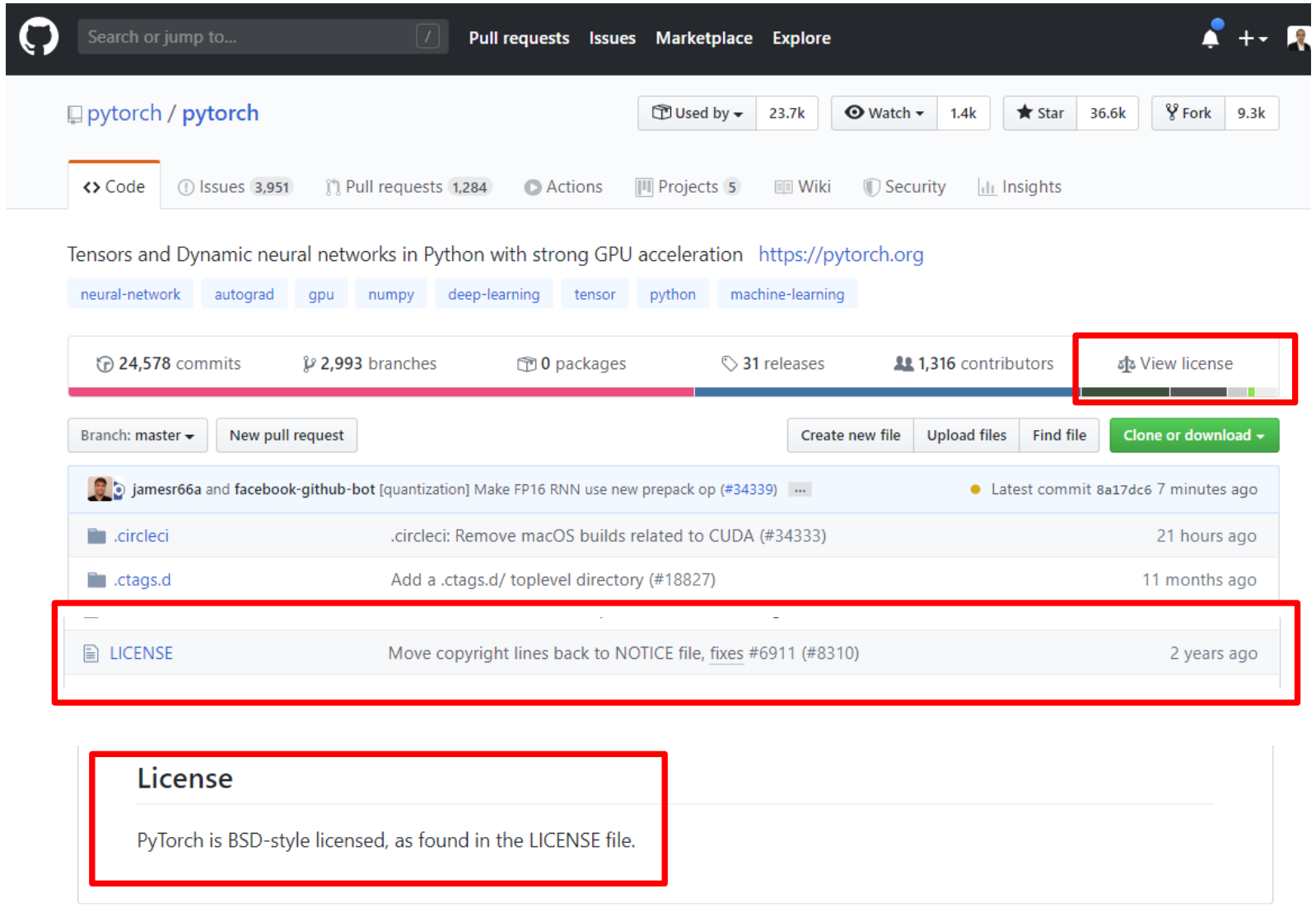
FAIR DATA AUSTRIA

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 by 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)

FAIR DATA AUSTRIA



FAIR DATA AUSTRIA

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

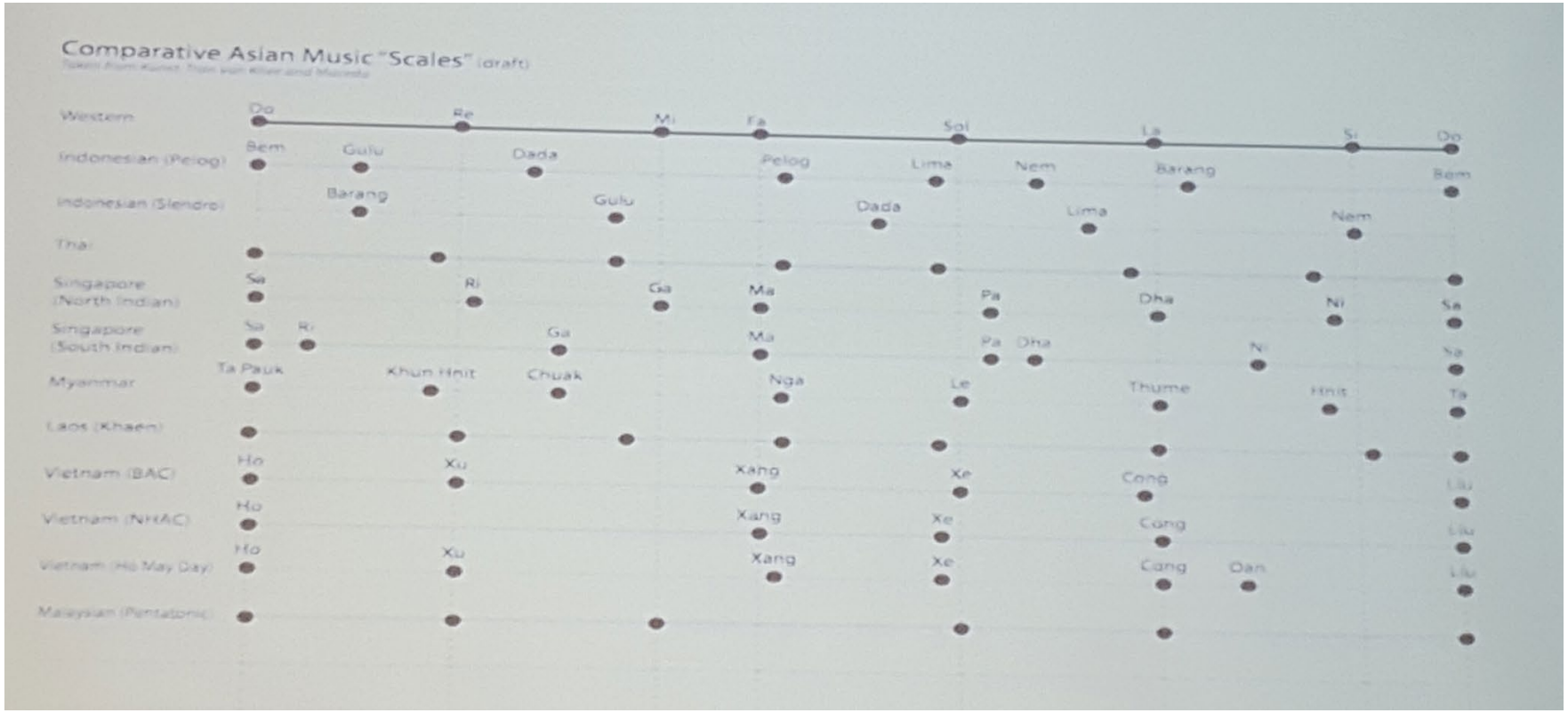
- Who is the “community” ?
- What is the “standard” ?
  - English vs other languages
- Metadata
  - Domain independent
    - e.g. Dublin Core
  - Domain specific
    - e.g. EXIF for images
- Sometimes no common standard exist
  - Good documentation and README
- There is no universal guideline – it always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

FAIR DATA AUSTRIA

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.

https://covid19-dashboard.ages.at/basisinfo_en.html?l=en

FAIR DATA AUSTRIA

SUMMARY

FAIR DATA AUSTRIA

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


Q
A

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

FAIR DATA AUSTRIA

THANK YOU FOR YOUR PARTICIPATION & ATTENTION!