Workshop

FAIR research data publication

81. Jahrestagung der Deutschen Geophysikalischen Gesellschaft, 4. März 2021

Andreas Hübner
Deutsches GeoForschungsZentrum GFZ

FID GEO Fachinformationsdienst Geowissenschaften

This work is licensed under a Creative Commons Attribution 4.0 International License.

Interaction

- **Several polls**

- **Mini-group discussion in breakout rooms**

- open the chat
- download links (for the discussion and the full slide set)
- post your questions in the chat
- "!" to indicate an oral contribution

Icons made by Freepik from Flaticon

Workshop

FAIR research data publication

81. Jahrestagung der Deutschen Geophysikalischen Gesellschaft, 4. März 2021

Andreas Hübner
Deutsches GeoForschungsZentrum GFZ

FID GEO Fachinformationsdienst Geowissenschaften

This work is licensed under a Creative Commons Attribution 4.0 International License.

DFG Fachinformationsdienste (FID)

- Initiative to complement existing local information infrastructures in research institutions with national services.
- FID GEO Partners:

SUB | NIEDERSACHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

GFZ Helmholtz-Zentrum Potsdam

- Website: **fidgeo.de**

digitisation
open access
research data
electronic publishing

Why publish research data?

How to publish?

Licences

GFZ Data Services

Wrap up

FID GEO funded by DFG

Political and funders perspectives

G7 Science and Technology Ministers' Meeting

…promoting increasing access to […] scientific data and publications,…

"…open access is the default setting for research data generated in Horizon 2020."

DFG

…research data should be made available as soon as possible.

G7 Science and Technology Ministers 2016: Tsukuba Communiqué
DFG 2015: Leitlinien zum Umgang mit Forschungsdaten
EU 2016: Guidelines on FAIR Data Management in Horizon 2020

Political and funders perspectives

Code of Conduct "Guidelines for Safeguarding Good Research Practice"
German Science Foundation DFG

- making the research data, materials and information on which the results are based, as well as the methods and software used, available and fully explaining the work processes.
- Software is made publicly available along with the source code
- Make available in recognised archives and repositories in accordance with the FAIR principles

DFG 2019: Guidelines for Safeguarding Good Research Practice
https://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/index.html

https://www.fidgeo.de/en/research-data-in-the-dfg-guidelines-for-safeguarding-good-research-practice/

Publishers

Supporting **data must be made available** to editors and peer reviewers at the time of submission for the purposes of evaluating the manuscript. All manuscripts reporting original research published in Nature journals **must include a data availability statement** …

Earth, space and environmental sciences

From January 2019, where community repositories are available, **we will require data sharing** through such repositories […].
Where such repositories are not available, datasets may be hosted in general data repositories such as Figshare, Dryad or Zenodo.

https://www.nature.com/authors/policies/availability.html
https://www.nature.com/nature-research/editorial-policies/reporting-standards

Publishers

All data used in the analysis **must be available** to any researcher for purposes of reproducing or extending the analysis. Data must be available in the paper, deposited in a community special-purpose repository, accessible via a general-purpose repository such as Dryad, or otherwise openly available.

Climate and Earth and Space Sciences data.

Guidelines on data deposition are provided by the **Coalition on Publishing Data in the Earth and Space Sciences (COPDESS)**, together with a searchable online **Repository Finder**.

https://www.sciencemag.org/authors/science-journals-editorial-policies#data-deposition

Publishers

Copernicus Publications **requests depositing data** that correspond to journal articles in reliable (public) data repositories, assigning digital object identifiers, and properly citing data sets as individual contributions.

Data policy

The output of research is not only journal articles but also data sets, model code, samples, etc. Only the entire network of interconnected information can guarantee integrity, transparency, reuse, and reproducibility of scientific findings. Moreover, all of these resources provide great additional value in their own right. Hence, it is particularly important that data and other information underpinning the research findings are "findable, accessible, interoperable, and reusable" (FAIR) not only for humans but also for machines.

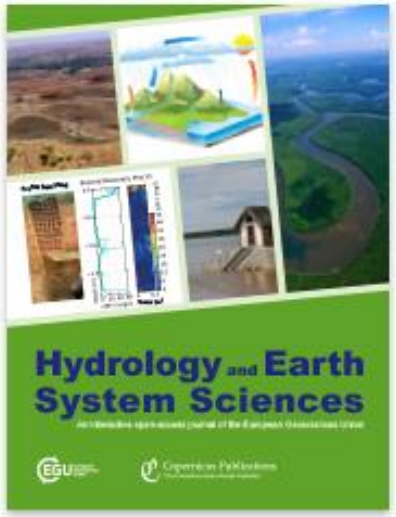
re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Therefore, Copernicus Publications requests depositing data that correspond to journal articles in reliable (public) data repositories, assigning digital object identifiers, and properly citing data sets as individual contributions. Please find your appropriate data repository in the registry for research data repositories: re3data.org. A data citation in a publication resembles a bibliographic citation and needs to be included in the publication's reference list. To foster the accessibility as well as the proper citation of data, Copernicus Publications requires all authors to provide a statement on the availability of underlying data as the last paragraph of each article (see section data availability). In addition, data sets, model code, video supplements, video abstracts, International Geo Sample Numbers, and other digital assets should be linked to the article through DOIs in the assets tab. With Earth System Science Data (ESSD) Copernicus Publications provides a journal dedicated to the publication of data papers, including peer review of data sets. Authors should consider submitting a data paper to ESSD in addition to their research paper in another journal published by Copernicus Publications.

Best practice following the Joint Declaration of Data Citation Principles initiated by FORCE 11:

COPDESS

In addition to promoting these data citation principles, Copernicus Publications is a signatory of the Coalition on Publishing Data in the Earth and Space Sciences (COPDESS) commitment statement and the Enabling FAIR Data Commitment Statement in the Earth, Space, and Environmental Sciences.



https://www.hydrology-and-earth-system-sciences.net/about/data_policy.html

Publishers

Earth science and biodiversity journals can improve support for data publication

"About half (9 out of 20) of the journals from earth sciences in this study don't address data publishing at all."

[...]

"However, in some of the surveyed journals´ texts, ambiguous and inconsistent statements were encountered, making it hard for authors to identify the expectations of the journal on data publishing."

Earth Science and Biodiversity Journals can Improve Support for Data Sharing. *Data Science Journal*, 19(1), p.37. DOI: http://doi.org/10.5334/dsj-2020-037

Scientists and broader perspective

Individual scientist

1. Additional publications
2. Greater citation rate
3. Wider recognition among peers
4. Invitations to meetings, collaborations, consultancy
5. Creators of data are known from citation and so are contactable for more information
6. Citation of data sources adds authority that indicates their quality

Editors, peer reviewers

7. Independent verification and qualification of research findings is possible

Scientific community

8. Data can be reused for similar and new purposes
9. Data can be integrated with other data to create new data resources

Funding agencies

10. Better financial return from research investment as data are used again

Source: Mark J. Costello, Motivating Online Publication of Data https://doi.org/10.1525/bio.2009.59.5.9

Scientists and broader perspective

Individual scientist

1. Additional publications
2. Greater citation rate
3. Wider recognition among peers
4. Invitations to meetings, collaborations, consultancy
5. Creators of data are known from citation and so are contactable for more information
6. Citation of data sources adds authority that indicates their quality

Editors, peer reviewers

7. Independent verification and qualifycation of research findings is possible

Scientific community

8. Data can be reused for similar and new purposes
9. Data can be integrated with other data to create new data resources

Funding agencies

10. Better financial return from research investment as data are used again

Source: Mark J. Costello, Motivating Online Publication of Data https://doi.org/10.1525/bio.2009.59.5.9

Breakout room

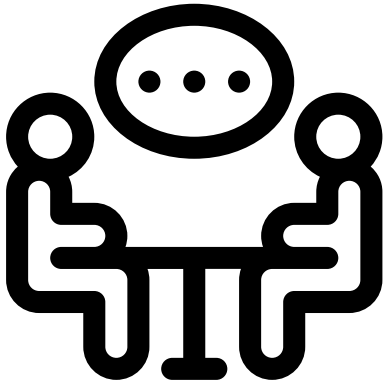
Download pdf-file (link in the chat)

Discuss the advantages

Agree on the 3 most important ones

Rank the 3 most important ones: first, second, third

5 min



Icons made by Freepik from Flaticon

Breakout room

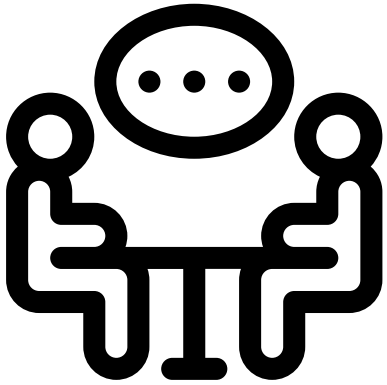
Download pdf-file (link in the chat)

Discuss the advantages

Agree on the 3 most important ones

Rank the 3 most important ones: first, second, third

5 min



Come back to the main room

Participate in 3 polls

Icons made by Freepik from Flaticon

to read....

The citation advantage of linking publications to research data
2020 https://doi.org/10.1371/journal.pone.0230416

A study of the impact of data sharing on article citations using journal policies as a natural experiment
2019 https://doi.org/10.1371/journal.pone.0225883

Sharing Detailed Research Data Is Associated with Increased Citation Rate
2007 https://doi.org/10.1371/journal.pone.0000308

Caution

Rights of other scientist
with Co-authorship, all authors can only jointly decide on the reuse or publication.

Secrecy agreements
In third-party funded projects or by instruction of employer.

Patents
When the research data describe a patentable invention and this invention is to be filed for a patent.

Personalised data
Must be anonymised before publication.

Bild von DavidRockDesign auf Pixabay

Time for questions

Icon made by Freepik

Why publish research data?

How to publish?

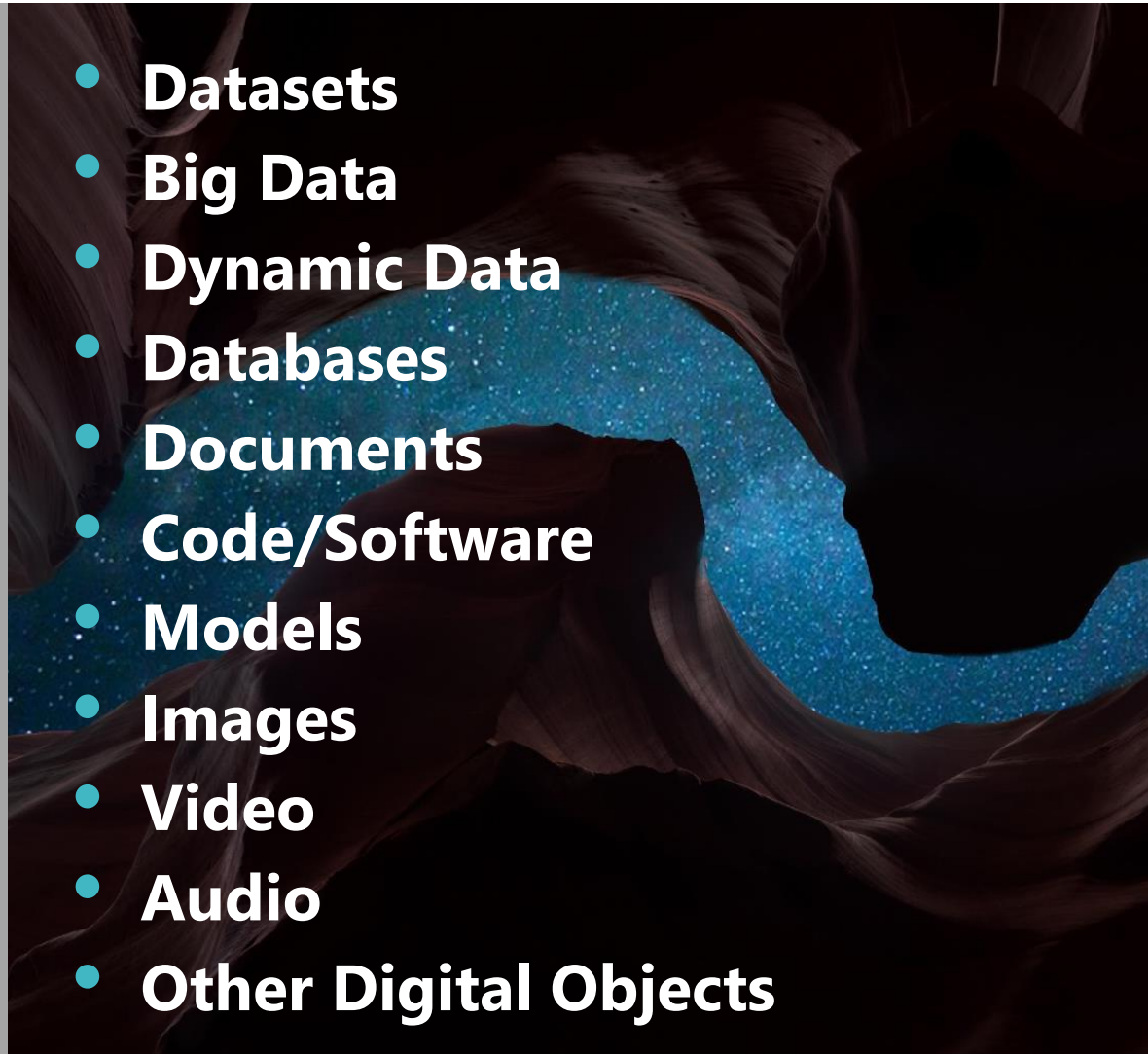
Licences

GFZ Data Services

Wrap up

FID GEO funded by DFG

These Data...



- Datasets
- Big Data
- Dynamic Data
- Databases
- Documents
- Code/Software
- Models
- Images
- Video
- Audio
- Other Digital Objects

Slide from: Stall, Shelley, Townsend, Randy, & Robinson, Erin. (2020, April). The Paper and The Data: Authors, Reviewers, and Editors Webinar on Updated Journal Practices for Data (and Software). Zenodo. http://doi.org/10.5281/zenodo.3744660

FAIR data Guiding Principles

To be Findable:

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

F2. data are described with rich metadata.

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

F4. metadata specify the data identifier.

To be Accessible:

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.

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

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

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

To be 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.

To be Re-usable:

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

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

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

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

https://www.force11.org/fairprinciples https://blogs.tib.eu/wp/tib/wp-content/uploads/sites/3/2017/09/The-FAIR-Data-Principles.pdf https://www.go-fair.org/fair-principles/

Enabling Findable, Accessible, Interoperable and Reusable Data

in the earth, space, and environmental science

Researchers understand and follow expectations related to data management and metadata of the publication

Publishers set standards and follow best practice related to datasets, metadata, accepted repositories and data citation

Scientific repositories are valued for stewardship, data access, improving peer review and digital product quality

https://eos.org/editors-vox/enabling-findable-accessible-interoperable-and-reusable-data

COMMITMENT STATEMENT

IN THE EARTH, SPACE, AND ENVIRONMENTAL SCIENCES

- Repositories
- Publishers
- Societies, communities, and institutions
- Funding agencies and organizations
- Individual researchers

http://www.copdess.org/enabling-fair-data-project/commitment-to-enabling-fair-data-in-the-earth-space-and-environmental-sciences

Background: Gerd Altmann auf Pixabay

COMMITMENT STATEMENT

- Make research outputs FAIR and, whenever possible, open by depositing research outputs (e.g., data, software, physical sample information, etc.) in **trustworthy, community-accepted, FAIR-aligned repositories.**

- **Cite** data, software, physical samples, and other products created or reused for your research in your publications.

- Include a **data availability statement** in your publication to make it clear where the data (and other research outputs as is possible) that supports the paper can be accessed along with any other access information.

http://www.copdess.org/enabling-fair-data-project/commitment-to-enabling-fair-data-in-the-earth-space-and-environmental-sciences

FAIR tools

Top 10 FAIR Data & Software Things Geoscience

https://librarycarpentry.org/Top-10-FAIR//2018/12/01/geoscience/

- for geoscientists
- lots of examples and exercises

FAIR self assessment tool

https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/

FAIR-Aware

https://www.fairsfair.eu/fair-aware

FID GEO funded by DFG

Time for questions

Icon made by Freepik

Data Description

Ensuring that data is „independently understandable“ is crucial.

Bild von silviarita auf Pixabay

Data Description

Impressum

GFZ DATA SERVICES
GEOSCIENCES DATA PUBLISHER

Dataset

3D-URG: 3D gravity constrained structural model of the Upper Rhine Graben

Released

Cite as:

Copy citation to clipboard

Freymark, Jessica; Bott, Judith; Scheck-Wenderoth, Magdalena; Bär, Kristian; Stiller, Manfred; Fritsche, Johann-Gerhard; Kracht, Matthias; Gomez Dacal, Maria Laura (2020): 3D-URG: 3D gravity constrained structural model of the Upper Rhine Graben. GFZ Data Services. https://doi.org/10.5880/GFZ.4.5.2020.004

Description

Files

Download data (zip, 37.3 MB)
Data description

License: CC BY 4.0

Link to journal article

Dataset Description

Supplement to

Freymark, Jessica; Sippel, Judith; Scheck-Wenderoth, Magdalena; Bär, Kristian; Stiller, Manfred; Fritsche, Johann-Gerhard; et al. (2017): The deep thermal field of the Upper Rhine Graben. Tectonophysics. 10.1016/j.tecto.2016.11.013

Related Work

Derived from

Amante, C., & Eakins, B. W. (2009). *ETOPO1 Global Relief Model converted to PanMap layer format* [Data set]. PANGAEA - Data Publisher for Earth & Environmental Science. https://doi.org/10.1594/PANGAEA.769615

Arndt, D., Bär, K., Fritsche, J.-G., Sass, I., & Hoppe, A. (2011). 3D structural model of the Federal State of Hesse (Germany) for geopotential evaluation. *Zeitschrift Der Deutschen Gesellschaft Für Geowissenschaften*, *162*(4), 353–369. https://

Abstract

We provide a set of grid files that collectively allow recreating a 3D geological model which covers the Upper Rhine Graben and its adjacent tectonic domains, such as portions of the Swiss Alps, the Molasse Basin, the Black Forest and Vosges Mountains, the Rhenish Massif and the Lower Rhine Graben. This publication is a complement to the publication of Freymark et al. (2017).

Accordingly, the provided structural model consists of (i) 14 sedimentary and volcanic units; (ii) crystalline crust composed of seven upper crustal units and a lower crustal unit; and (iii) two lithospheric mantle units. The files provided here include information on the regional variation of these geological units in terms of their depth and thickness, both attributes being allocated to regularly spaced grid nodes with horizontal spacing of 1 km.

The model has originally been developed to obtain a basis for numerical simulations of heat transport, to calculate the lithospheric-scale conductive thermal field and assess the related geothermal potentials, in particular for the Upper Rhine Graben (a region especially well-suited for geothermal energy exploitation). Since such simulations require the subsurface variation of physical rock properties to be defined, the 3D model differentiates units of contrasting materials, i.e. rock types. On that account, a large number of geological and geophysical data have been analysed (see Related Work) and we shortly describe here how they have been integrated into a consistent 3D model (Methods). For further information on the data usage and the characteristics of the units (e.g., lithology, density, thermal properties), the reader is referred to the original article (Freymark et al., 2017). The contents and structure of the grid files provided herewith are described in the Technical Info section.

Abstract

Additional Information

We acknowledge Landesamt für Geologie, Rohstoffe und Bergbau (LGRB; Baden-Wuerttemberg) for kindly allocating the digital datasets of the GeORG model and the geological 3D model of Baden-Wuerttemberg.

Methods

The presented 3D structural model is the result of an extensive data integration process. In a first step, we visualized and collectively analysed geological maps, smaller-scale 3D structural models, depth and thickness maps, drilled formation tops and interpreted seismic horizons (See Related Works) using the software Petrel (©Schlumberger). After identifying the main lithological units to be differentiated by the intended 3D model and correcting for inconsistencies between the layers, the scattered information on the top surface elevation of the units was interpolated to obtain regular grids with a horizontal element spacing of 1 km (Convergent Interpolation algorithm of Petrel). More details about the original datasets (e.g., their regional extents, sources etc.) used to model the topology of the structural horizons are listed in the Supplementary Material 1 of Freymark et al. (2017).

In order to mitigate insufficient coverage of the region with deep seismic profiles revealing the internal

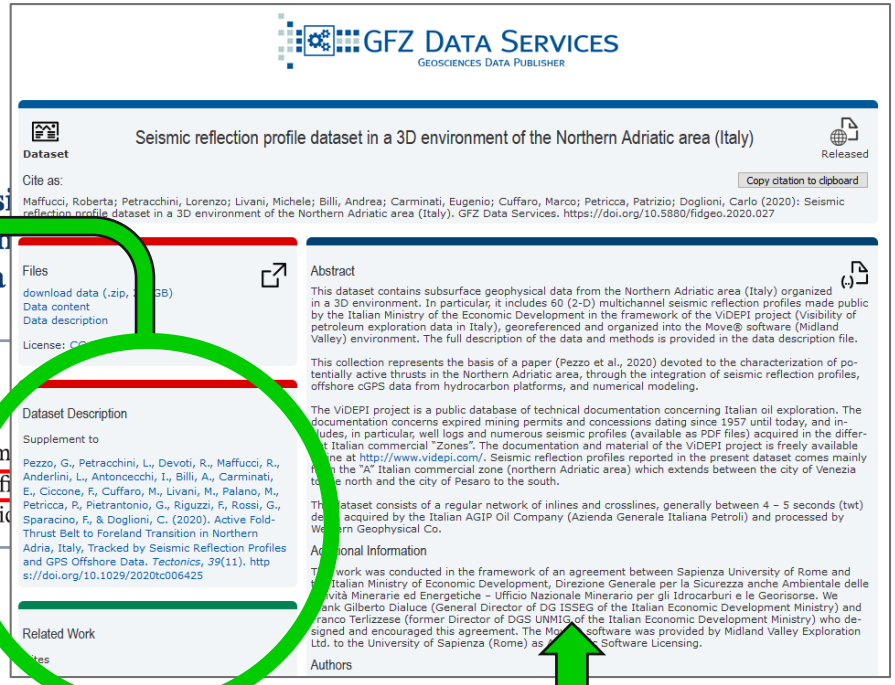
Data Description

Peer-reviewed articles with the description of datasets, data collections, data infrastructures, etc.

No Inter-pretation!

Slide source: Kirsten Elger

Data Description



GFZ DATA SERVICES
GEOSCIENCES DATA PUBLISHER

Dataset

Seismic reflection profile dataset in a 3D environment of the Northern Adriatic area (Italy)

Released

Cite as:

Copy citation to clipboard

Maffucci, Roberta; Petracchini, Lorenzo; Livani, Michele; Billi, Andrea; Carminati, Eugenio; Cuffaro, Marco; Petricca, Patrizio; Doglioni, Carlo (2020): Seismic reflection profile dataset in a 3D environment of the Northern Adriatic area (Italy). GFZ Data Services. https://doi.org/10.5880/fidgeo.2020.027

Files

download data (.zip, [illegible]GB)
Data content
Data description

License: [illegible]

Abstract

This dataset contains subsurface geophysical data from the Northern Adriatic area (Italy) organized in a 3D environment. In particular, it includes 60 (2-D) multichannel seismic reflection profiles made public by the Italian Ministry of the Economic Development in the framework of the ViDEPI project (Visibility of petroleum exploration data in Italy), georeferenced and organized into the Move® software (Midland Valley) environment. The full description of the data and methods is provided in the data description file.

This collection represents the basis of a paper (Pezzo et al., 2020) devoted to the characterization of potentially active thrusts in the Northern Adriatic area, through the integration of seismic reflection profiles, offshore cGPS data from hydrocarbon platforms, and numerical modeling.

The ViDEPI project is a public database of technical documentation concerning Italian oil exploration. The documentation concerns expired mining permits and concessions dating since 1957 until today, and includes, in particular, well logs and numerous seismic profiles (available as PDF files) acquired in the different Italian commercial "Zones". The documentation and material of the ViDEPI project is freely available online at http://www.videpi.com/. Seismic reflection profiles reported in the present dataset comes mainly from the "A" Italian commercial zone (northern Adriatic area) which extends between the city of Venezia to the north and the city of Pesaro to the south.

The dataset consists of a regular network of inlines and crosslines, generally between 4 – 5 seconds (twt) de[illegible] acquired by the Italian AGIP Oil Company (Azienda Generale Italiana Petroli) and processed by We[illegible]rn Geophysical Co.

Additional Information

[illegible] work was conducted in the framework of an agreement between Sapienza University of Rome and [illegible] Italian Ministry of Economic Development, Direzione Generale per la Sicurezza anche Ambientale delle [illegible]ività Minerarie ed Energetiche – Ufficio Nazionale Minerario per gli Idrocarburi e le Georisorse. We thank Gilberto Dialuce (General Director of DG ISSEG of the Italian Economic Development Ministry) and Franco Terlizzese (former Director of DGS UNMIG of the Italian Economic Development Ministry) who designed and encouraged this agreement. The Mo[illegible] software was provided by Midland Valley Exploration Ltd. to the University of Sapienza (Rome) as a[illegible] Software Licensing.

Authors

Dataset Description

Supplement to

Pezzo, G., Petracchini, L., Devoti, R., Maffucci, R., Anderlini, L., Antoncecchi, I., Billi, A., Carminati, E., Ciccone, F., Cuffaro, M., Livani, M., Palano, M., Petricca, P., Pietrantonio, G., Riguzzi, F., Rossi, G., Sparacino, F., & Doglioni, C. (2020). Active Fold-Thrust Belt to Foreland Transition in Northern Adria, Italy, Tracked by Seismic Reflection Profiles and GPS Offshore Data. *Tectonics*, *39*(11). https://doi.org/10.1029/2020tc006425

Related Work

[illegible]tes

Tectonics

RESEARCH ARTICLE
10.1029/2020TC006425

Key Points:

- Seismic reflection profiles in the

Active Fold-Thrust Belt to Foreland Transi[illegible] in Northern Adria, Italy, Tracked by Seism[illegible] Reflection Profiles and GPS Offshore Data

3. Data and Methods

3.1. Seismic Reflection Profiles and Borehole Composite Logs

seismic reflection profiles from the ViDEPI project organized in a 3-D environment b[illegible] m[illegible] software are available in the GFZ Data Services Repository: http://doi.org/10.58[illegible]0/fi[illegible] Maffucci et al., 2020). Furthermore, we collected seismic profiles, structural ge[illegible]logi[illegible]

Data Availability Statement

All data needed to evaluate the conclusions in the paper are present in the paper itself and/or the associated supporting information. All these data are also freely available in external repositories and previous articles. In particular, the geodetic data are available in Palano et al. (2020) (https://doi.pangaea.de/10.1594/PANGAEA.914358). The seismic reflection profiles organized in a 3-D Move® file/project are available in Maffucci et al. (2020) (http://pmd.gfz-potsdam.de/panmetaworks/review/aaf30ce1d97be14e03c64b5a638334ed0c40007bc91f6029b83a149727f47c5f). Supporting figures (Figures S1 and S2) and tables are available online (ftp://ftp.ingv.it/pub/giuseppe.pezzo/TECT_2020TC006425/). In the data repository, we make available the subsurface geophysical data set used to classify the tectonic domains of the studied CGPS stations (i.e., fold-thrust belt, proto-thrust domain, and foreland). The data set is organized into the Move® software (Midland Valley) environment, version 2016.2 and includes 60 public 2-D multichannel seismic reflection profiles deriving from the ViDEPI database (http://www.videpi.com). The dataset and its full description is available on the following link: http://doi.org/10.5880/fidgeo.2020.027 (Maffucci et al., 2020). CGPS data and its full description is available on the following link: https://doi.org/10.1594/PANGAEA.914358 (Palano et al., 2020).

References

Maesano, F. E., Toscani, G., Burrato, P., Mirabella, F., D'Ambrogi, C., & Basili, R. (2[illegible]). Deriving thrust fault slip rates from geological modeling: Examples from the Marche coastal and offshore contraction belt, Nort[illegible] Apennines, Italy. *Marine and Petroleum Geology*, *42*, 122–134. https://doi.org/10.1016/j.marpetgeo.2012.10.008

Maffucci, R., Petracchini, L., Livani, M., Billi, A., Carminati, E., Cuffaro, M., et al. ([illegible]). *Seismic Reflection Profile Dataset in a 3D Environment of the Northern Adriatic Area (Italy)*. GFZ Data Services. https://doi.org/10.5880/fidgeo.2020.027

Malinverno, A., & Ryan, W. B. (1986). Extension in the Tyrrhenian Sea and shortening in the Apennines as result of arc migration driven by sinking of the lithosphere. *Tectonics*, *5*(2), 227–245. https://doi.org/10.1029/TC005i002p00227

Repositories

Repository = (online accessible) database for the recording and publication of research data, texts and other digital objects[1]

Institutional Repository

- members of the institution
- many disciplines

Deposit Once
Repository for Research Data and Publications

Re|fub|ium

Domain-specific Repository

- Researchers worldwide
- discipline-specific

GFZ Data Services

PANGAEA.
Data Publisher for Earth & Environmental Science

- Domain-specific metadata, for example „location“
- Connected to domain-specific data portals
- Better quality-control
- extra services, e.g. integration of IGSN

Generic Repository

- Researchers worldwide
- all disciplines

zenodo

figshare

Mendeley

[1] Einstieg ins Forschungsdatenmanagement in den Geo

Repositories

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

More than
2500 data repositories

https://www.re3data.org/

Peer Reviewers MUST be able to…

- Access the data and/or software used in the research
- Validate that the data supports the science and the visualizations.
- Confirm that the data and/or software citations exist.

FID GEO funded by DFG

Time for questions

Icon made by Freepik

Software

The software, code, workflow, model that is integral to your research.

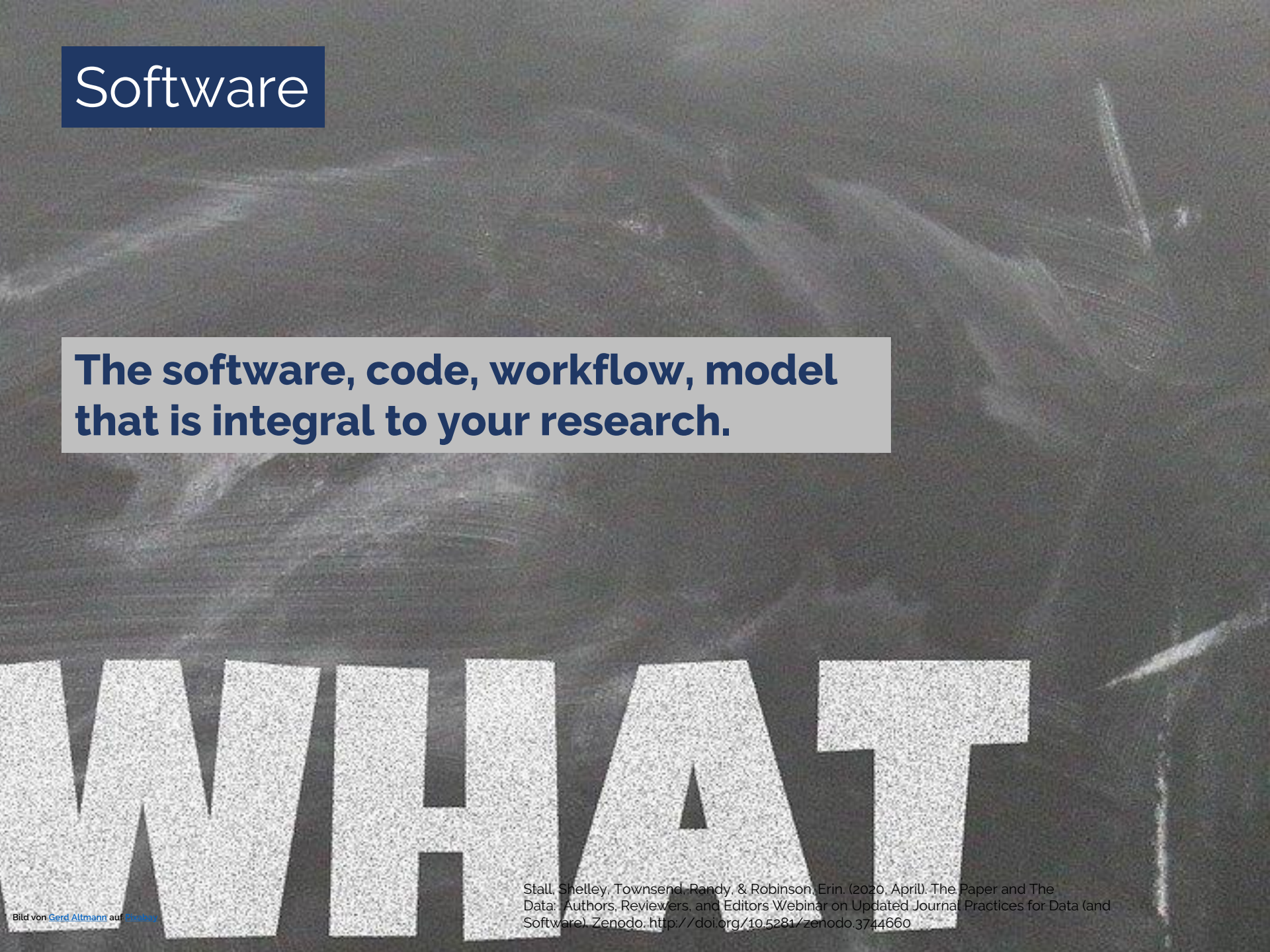


Bild von Gerd Altmann auf Pixabay

Stall, Shelley, Townsend, Randy, & Robinson, Erin. (2020, April). The Paper and The Data: Authors, Reviewers, and Editors Webinar on Updated Journal Practices for Data (and Software). Zenodo. http://doi.org/10.5281/zenodo.3744660

EGU ✓
@EuroGeosciences

"The point of sharing codes is to find mistakes, fix them, and make a software library better. Together."

On the @EGU_Seismo #EGUblogs this week @MTsekhmistrenko ponders the... challenges... of sharing #code in #seismology.

Read more: egu.eu/5JK8RO/

Tweet übersetzen

https://blogs.egu.eu/divisions/sm/2020/08/25/git-or-perish/

Software

what to include in your paper

Availability Statement

1. State **where the version of your software used for your research is preserved**
2. Optionally: Include the GitHub URL or other development platform URL (Note: GitHub is not a preservation resource)

Citation

1. Include in the References Section of your paper the citation to **where your software is preserved**. For GitHub users, there is an integrated connection to Zenodo

Methodology

1. Describe how your software works as it pertains to your research

From: Stall, Shelley, Townsend, Randy, & Robinson, Erin. (2020, April). The Paper and The Data: Authors, Reviewers, and Editors Webinar on Updated Journal Practices for Data (and Software). Zenodo. http://doi.org/10.5281/zenodo.3744660

Software citation methods

- Cite the software itself via **software domain repository** (e.g. Computational Infrastructure for Geodynamics (CGI), Hydroshare),
- Cite the **software journal** (e.g. JOSS) where the software is described **AND the repository where it is preserved**,
- Cite the **general repository supporting software preservation** that provides registration services for a digital object identifier (e.g. Zenodo).

From: Stall, Shelley, Townsend, Randy, & Robinson, Erin. (2020, April). The Paper and The Data: Authors, Reviewers, and Editors Webinar on Updated Journal Practices for Data (and Software). Zenodo. http://doi.org/10.5281/zenodo.3744660

Software

Avoid these incorrect citations resources:

- **GitHub**
- **SourceForge**
- **BitBucket**
- **GitLab**
- **Personal/Institution Page**

Software Citation Checklist for Authors

- Have I identified the software, including my own, which makes a **significant and specialized contribution to my academic work**, and therefore should be cited?
- Have I checked if the software has a **recommended citation from the creators** and used it if available? If this is to a paper, have I also cited the software directly?
- Have I created **as complete a citation as possible** if no recommended citation is given?
- Have I **referenced the software appropriately in my academic work**, in compliance with any citation formatting guidelines?

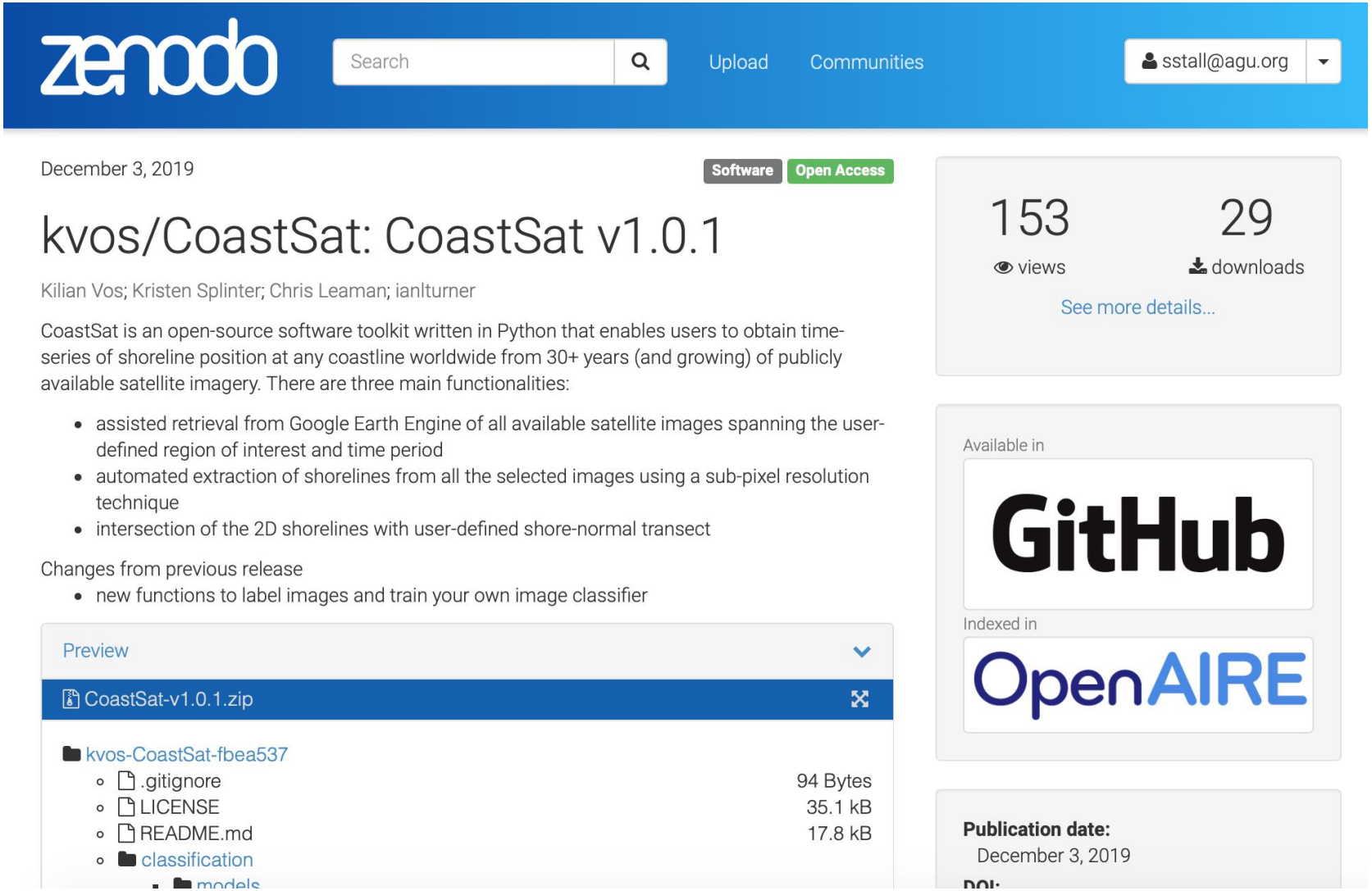
Software Citation Checklist for Authors (Version 0.9.0). Zenodo. https://doi.org/10.5281/zenodo.3479198

Software Citation – Elements

Does the citation include…

1. who created the software, contributed, designed, tested – including usability tests,
2. when it was created,
3. the title of the software (and version if available) and
4. where the software can be accessed (preferably via a persistent identifier to an archival repository)?

FID GEO funded by DFG



Kilian Vos, Kristen Splinter, Chris Leaman, & ianlturner. (2019, December 3). kvos/CoastSat: CoastSat v1.0.1 (Version v1.0.1). Zenodo. http://doi.org/10.5281/zenodo.3560436

Time for questions

Icon made by Freepik

Why publish research data?

How to publish?

Licences

GFZ Data Services

Wrap up

FID GEO funded by DFG

Re-use

- Every user of your published data should exactly know what is allowed to do with your data.

Copyright

- Machine-generated and unprocessed raw data are not protected by copyright.
- For most other data you should assume that data is protected by intellectual property rights.

Translated from: Gutachten zu den rechtlichen Rahmenbedingungen des Forschungsdatenmanagements (2018)
https://tu-dresden.de/gsw/jura/igewem/jfbimd13/ressourcen/dateien/publikationen/DataJus_Zusammenfassung_Gutachten_12-07-18.pdf

Copyright

- Machine-generated and unprocessed raw data are not protected by copyright.
- For most other data you should assume that data is protected by intellectual property rights.

Copyright

- Machine-generated and unprocessed raw data are not protected by copyright.
- For most other data you should assume that data is protected by intellectual property rights.

- Public domain
- All rights reserved
- Some rights reserved
- No rights reserved

Open licences

creative commons

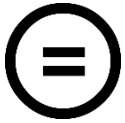
Public Domain (PD): no known copyright restrictions

No rights reserved (0): No Rights Reserved

Attribution (BY): You must give credit to author(s) or licencesor

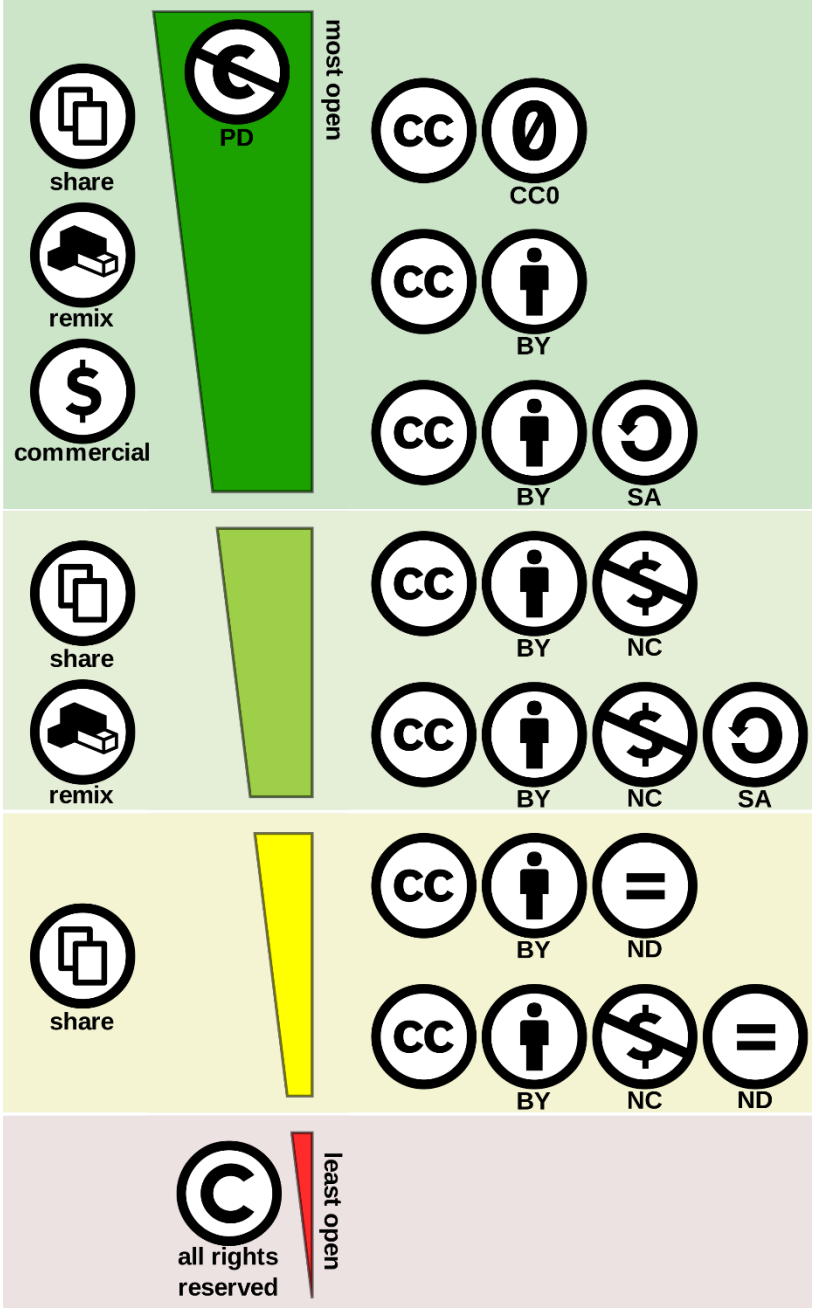
Non-Commercial (NC): reuse not for commercial purposes

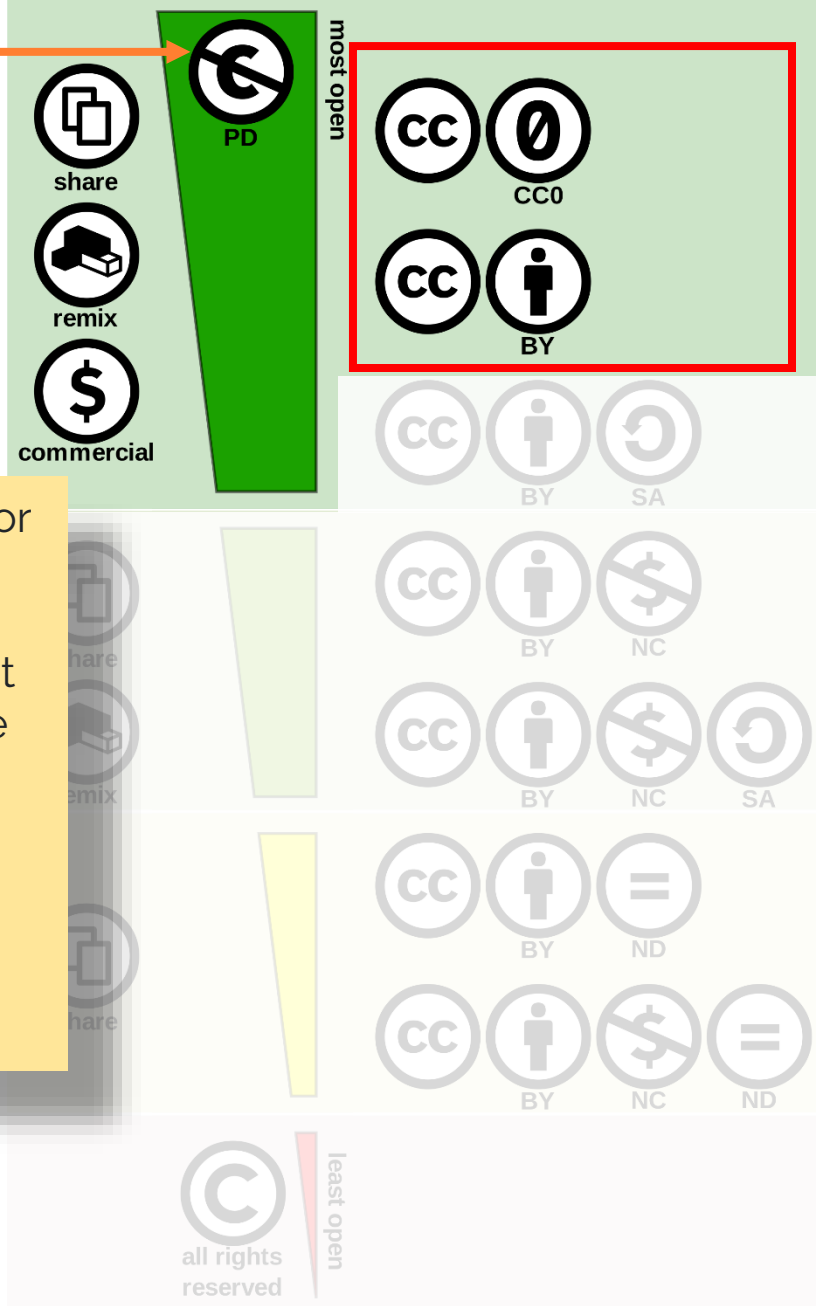
Share Alike (SA): share only with an identical licence

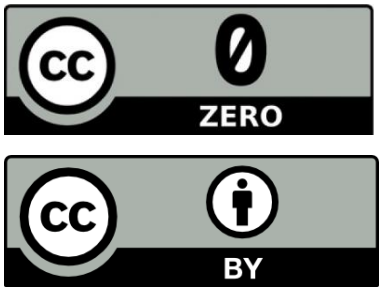


No Derivatives (ND): work must not be changed.

https://creativecommons.org/

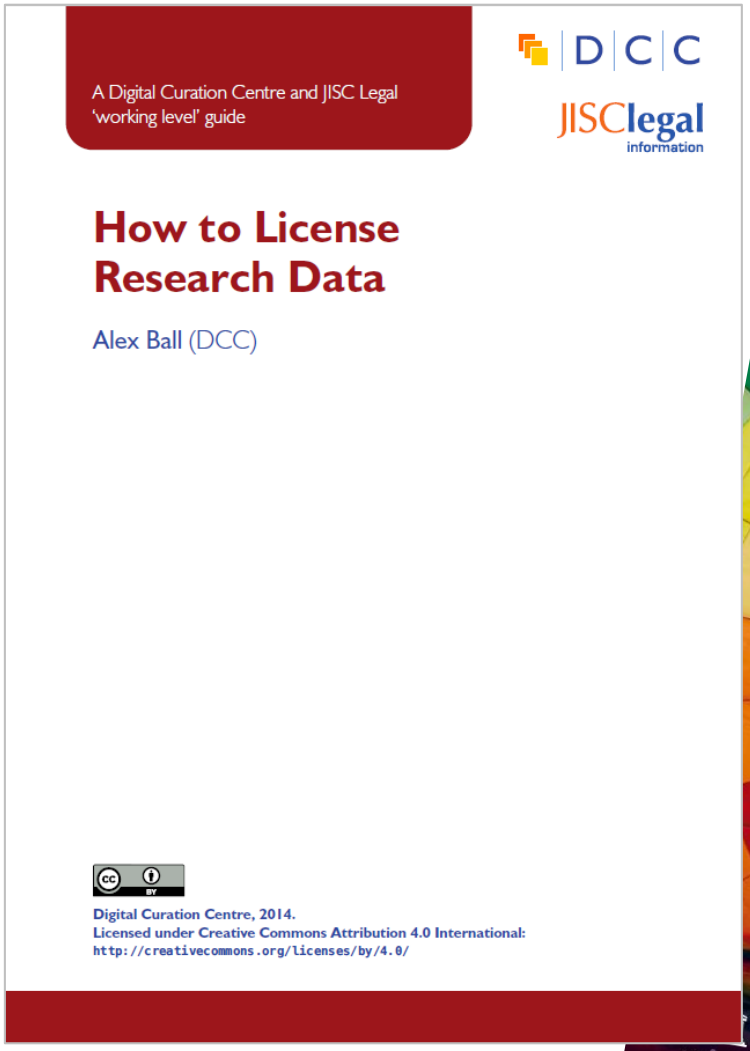

most open
PD
share
remix
commercial
CC0
BY
BY
SA
share
remix
BY
NC
BY
NC
SA
share
BY
ND
BY
NC
ND
all rights
reserved
least open





Use CC0 and simply ask for credit (rather than require attribution), and provide a citation for the dataset that others can copy and paste with ease.

Such requests are consistent with scholarly norms for citing source materials.



https://irights.info/wp-content/uploads/2014/11/Open_Content_A_Practical_Guide_to_Using_Open_Content_Licences_web.pdf
http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How_To_License_Research_Data.pdf

Licencing of software

https://www.software.ac.uk/resources/guides/adopting-open-source-licence

https://www.gnu.org/licenses/license-list.en.html

https://choosealicense.com/about/ and references on that page

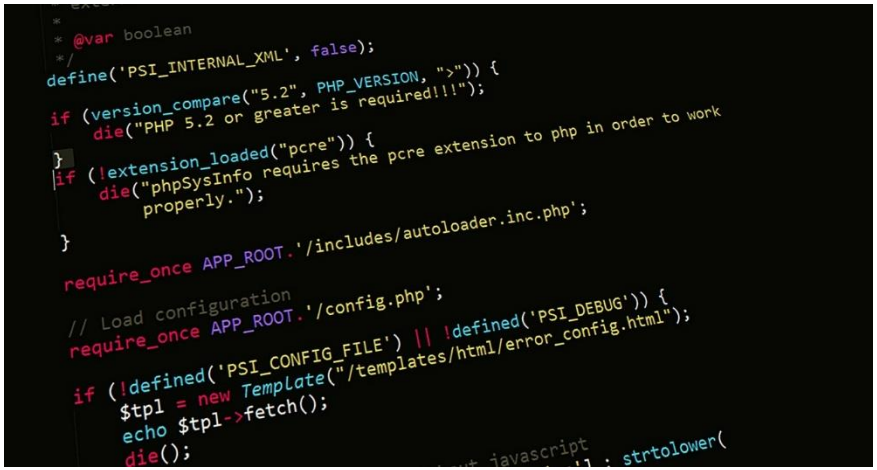


Bild von simplu27 auf Pixabay

Time for questions

Icon made by Freepik

Why publish research data?

How to publish?

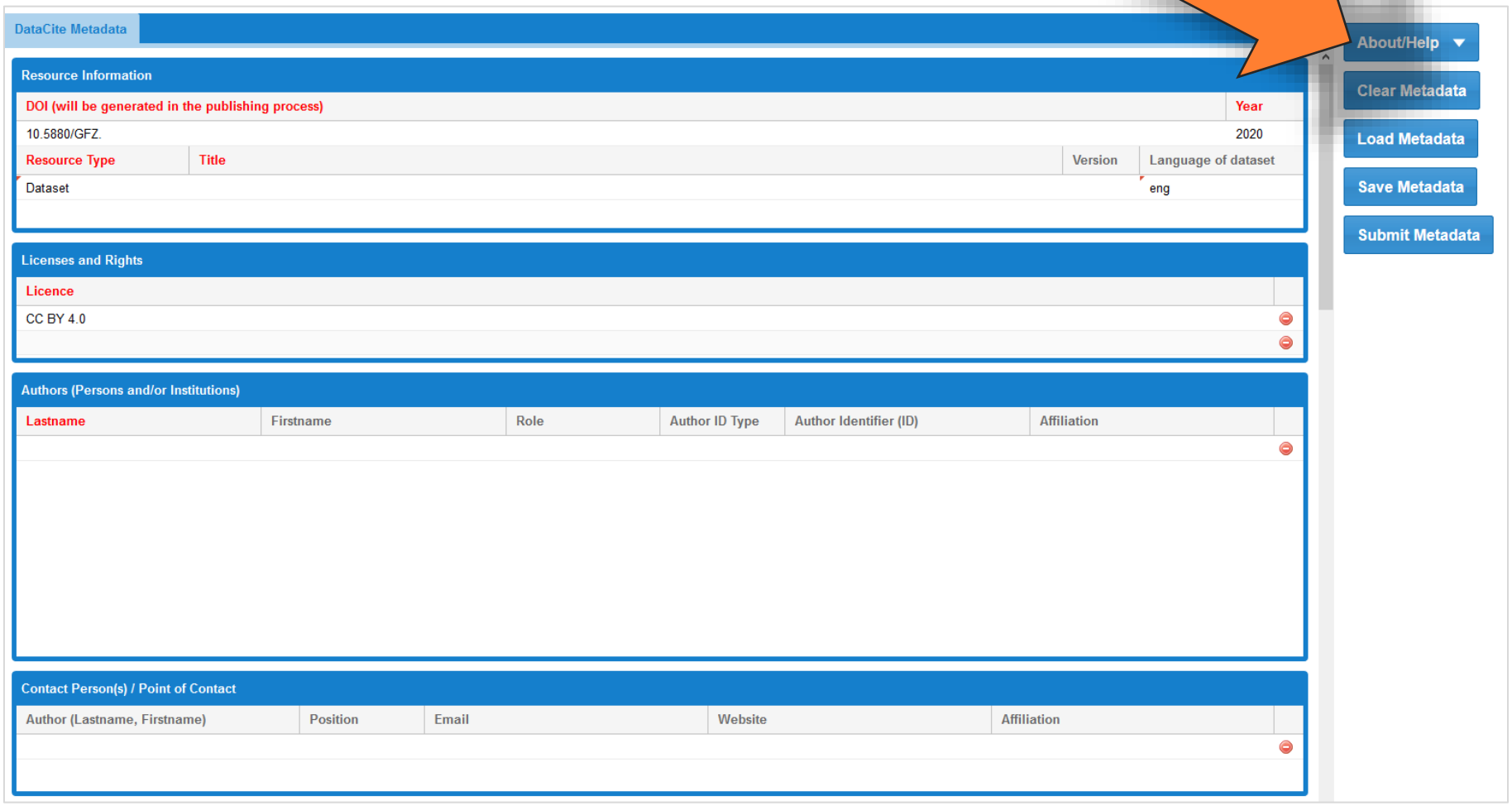
Licences

GFZ Data Services

Wrap up

FID GEO funded by DFG

Metadata-Editor

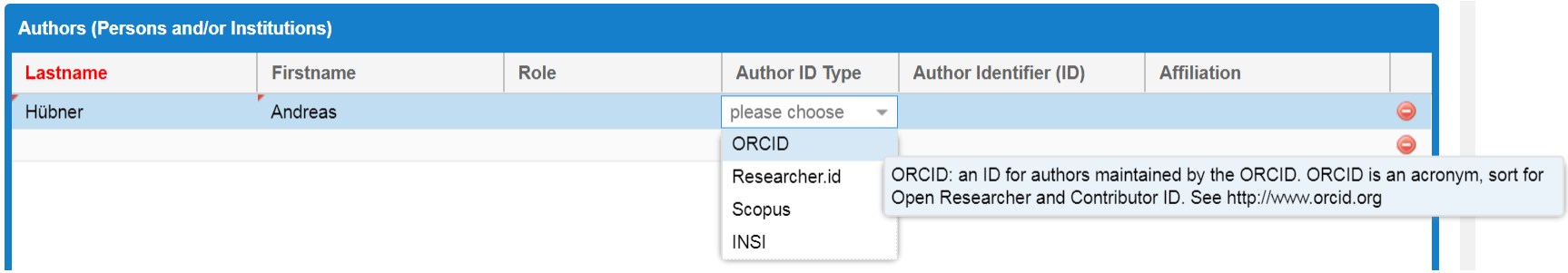


http://pmd.gfz-potsdam.de/panmetaworks/metaedit/

Metadata-Editor

1. ORCID
2. Contributors
3. Related work
4. Embargo

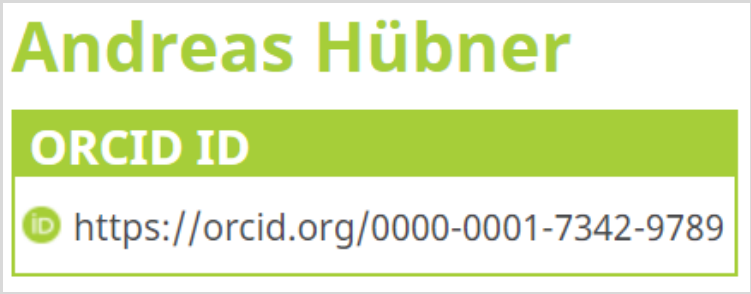
ORCID


Authors (Persons and/or Institutions)
Lastname
Firstname
Role
Author ID Type
Author Identifier (ID)
Affiliation
Hübner
Andreas
please choose
ORCID
Researcher.id
Scopus
INSI
ORCID: an ID for authors maintained by the ORCID. ORCID is an acronym, sort for Open Researcher and Contributor ID. See http://www.orcid.org

ORCID

ORCID
Connecting Research and Researchers

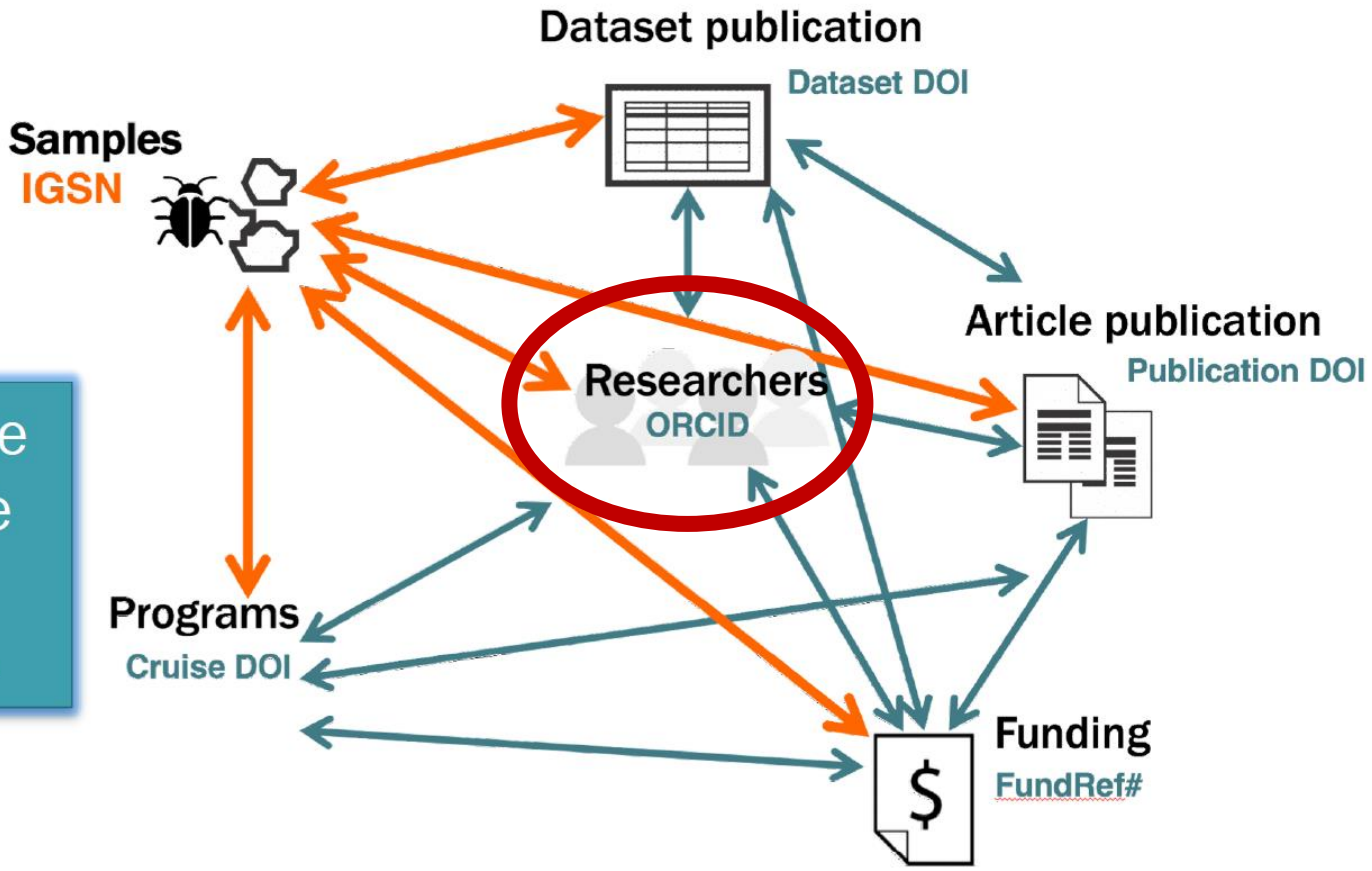
https://orcid.org
http://www.orcid-de.org/



- The **Open Researcher and Contributor ID** distinguishes you from every other researcher.
- ORCID connects all your research outputs to your person.
- Use "Auto-update" (for Crossref, Datacite, etc.) http://bit.ly/ORCID_Trust

ORCID

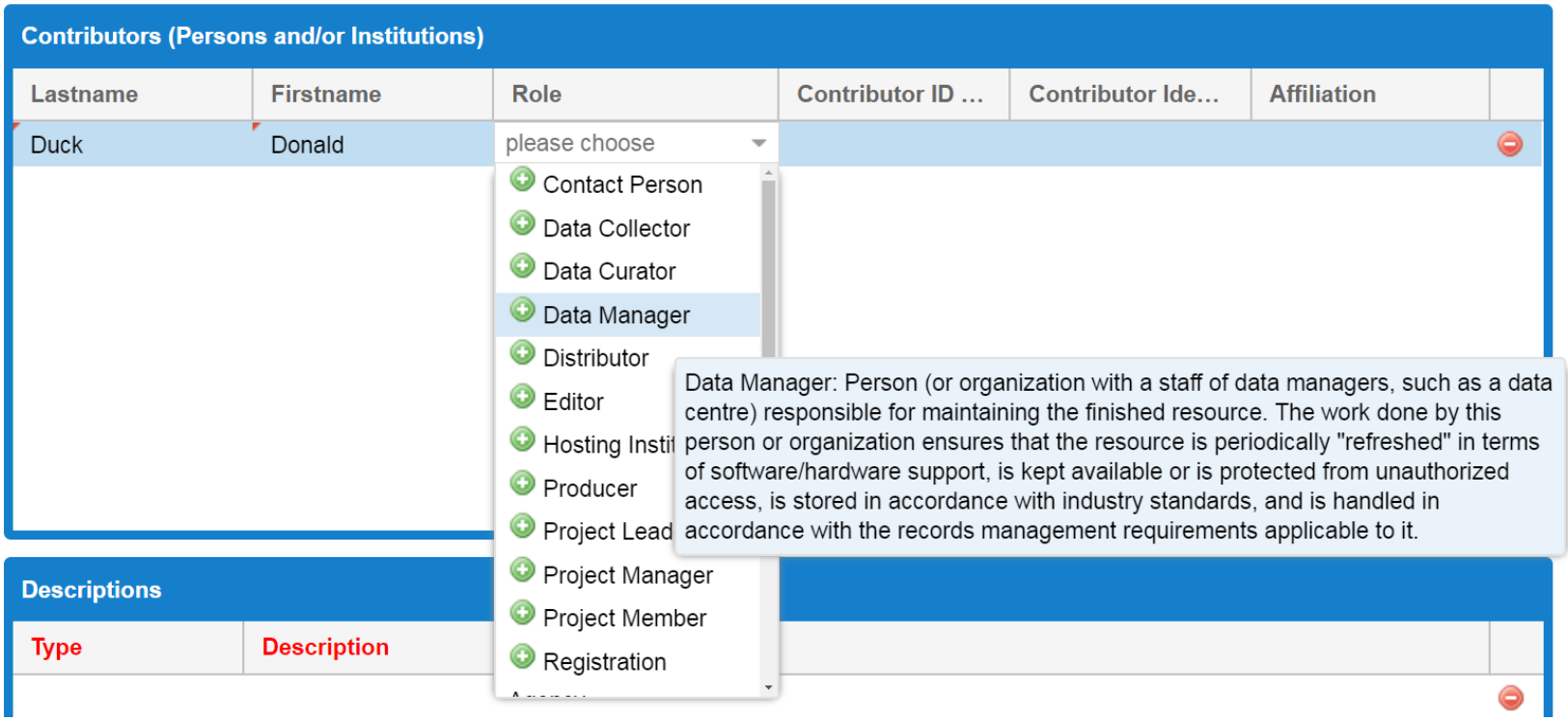
Identifiers are the foundation of the FAIR Research Data Ecosystem.



Lehnert et al., 2020, Digital Data Infrastructure for Physical Samples: The IGSN and the Internet of Samples, GeoUtrecht2020 virtual conference

Contributors

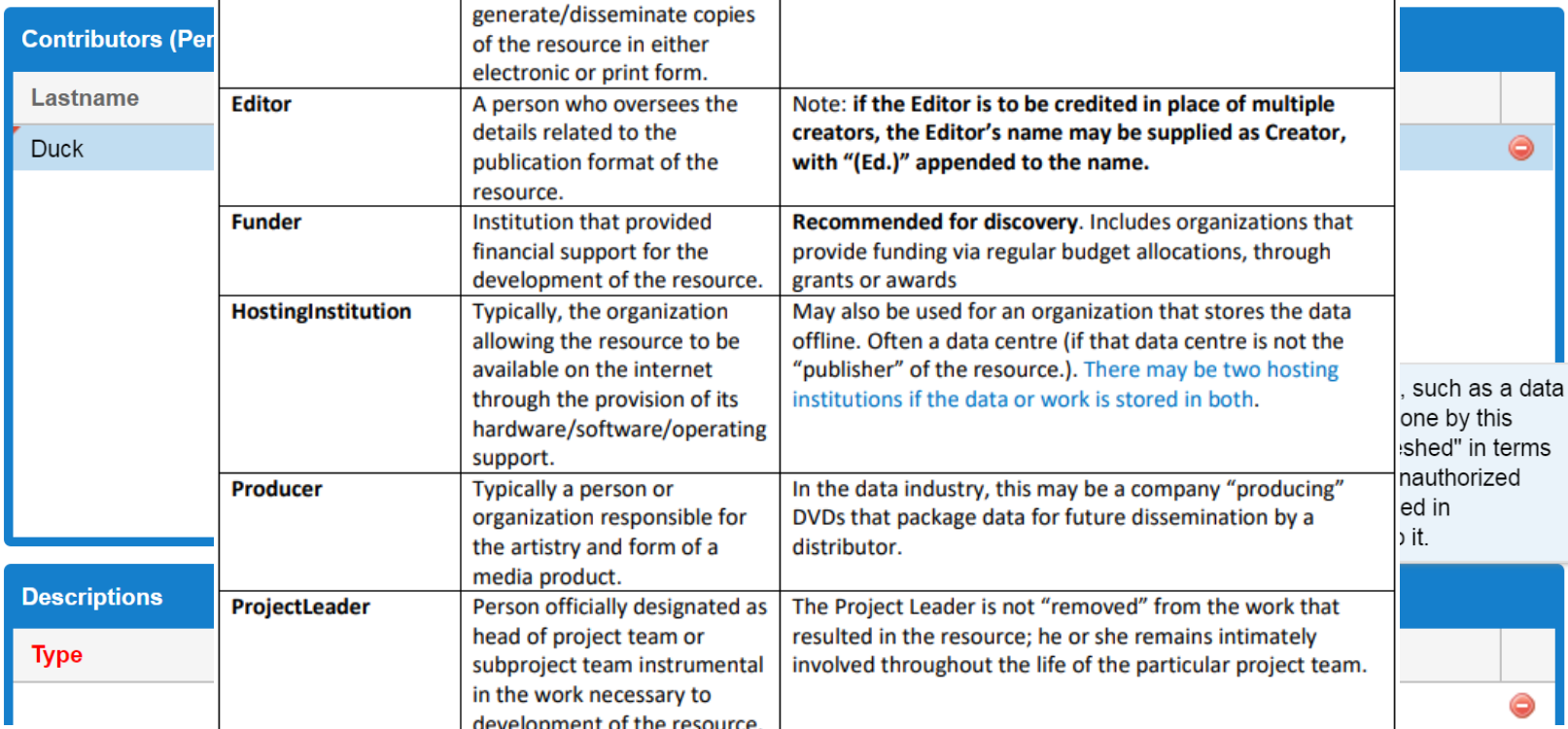
With “contributor” you have the possibility to acknowledge additional persons or institutions related to the dataset but which you would normally not mention as authors. These are not named in the citation, but always related with the dataset and searchable as all the other metadata fields.



Contri

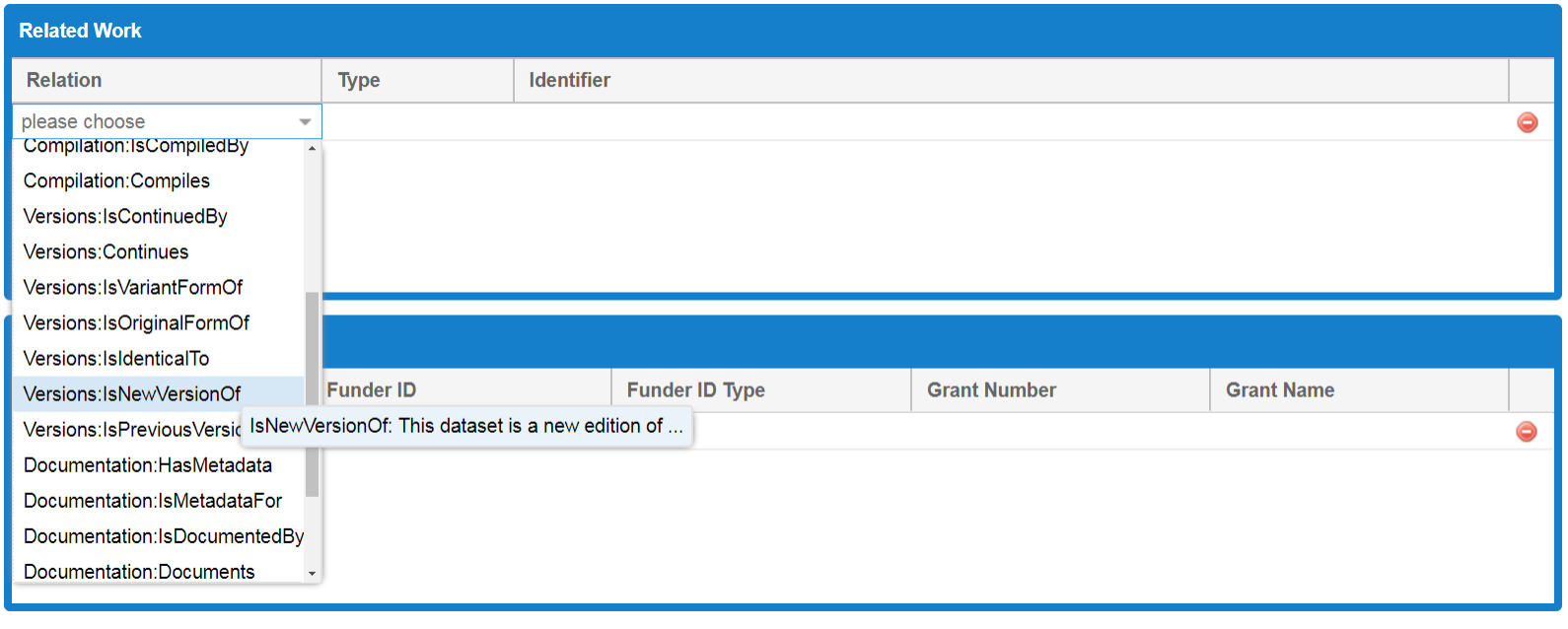
With "cor
persons c
not menti
related w

litional
uld normally
t always
ata fields.



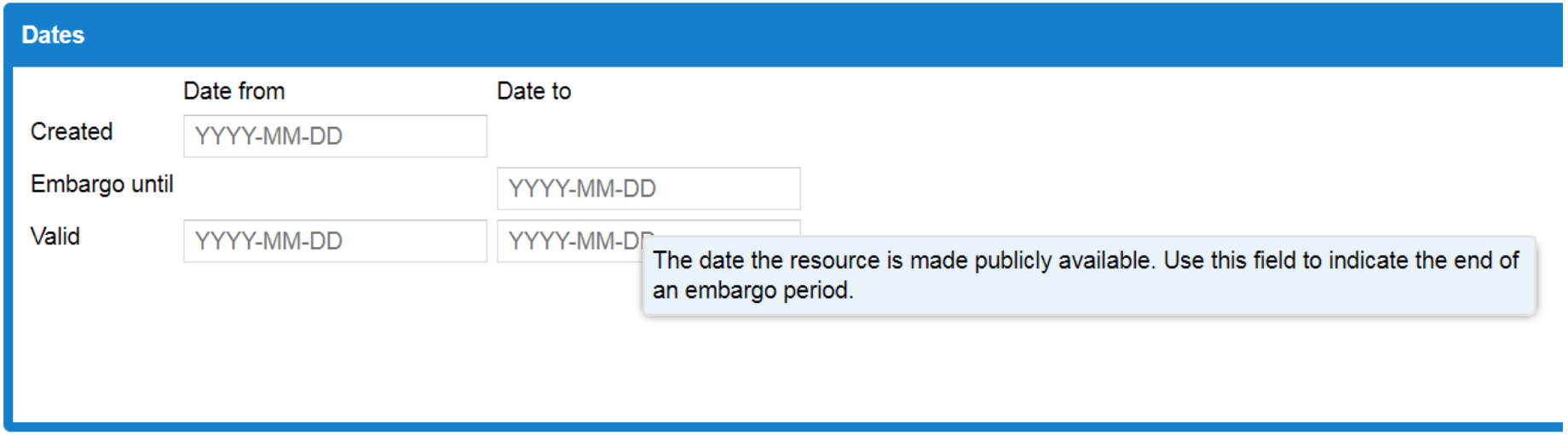
DataCurator	Person tasked with reviewing, enhancing, cleaning, or standardizing metadata and the associated data submitted for storage, use, and maintenance within a data center or repository	While the "DataManager" is concerned with digital maintenance, the DataCurators' role encompasses quality assurance focused on content and metadata. This includes checking whether the submitted dataset is complete, with all files and components as described by submitter, whether the metadata is standardized to appropriate systems and schema, whether specialized metadata is needed to add value and ensure access across disciplines, and determining how the metadata might map to search engines, database products, and automated feeds.
DataManager	Person (or organization with a staff of data managers, such as a data centre) responsible for maintaining the finished resource.	The work done by this person or organization ensures that the resource is periodically "refreshed" in terms of software/hardware support, is kept available or is protected from unauthorized access, is stored in accordance with industry standards, and is handled in accordance with the records management requirements applicable to it.
Distributor	Institution tasked with responsibility to generate/disseminate copies of the resource in either electronic or print form.	Works stored in more than one archive/repository may credit each as a distributor.
Editor	A person who oversees the details related to the publication format of the resource.	Note: **if the Editor is to be credited in place of multiple creators, the Editor's name may be supplied as Creator, with "(Ed.)" appended to the name.**
Funder	Institution that provided financial support for the development of the resource.	**Recommended for discovery**. Includes organizations that provide funding via regular budget allocations, through grants or awards
HostingInstitution	Typically, the organization allowing the resource to be available on the internet through the provision of its hardware/software/operating support.	May also be used for an organization that stores the data offline. Often a data centre (if that data centre is not the "publisher" of the resource.). There may be two hosting institutions if the data or work is stored in both.
Producer	Typically a person or organization responsible for the artistry and form of a media product.	In the data industry, this may be a company "producing" DVDs that package data for future dissemination by a distributor.
ProjectLeader	Person officially designated as head of project team or subproject team instrumental in the work necessary to development of the resource.	The Project Leader is not "removed" from the work that resulted in the resource; he or she remains intimately involved throughout the life of the particular project team.
ProjectManager	Person officially designated as manager of a project. Project may consist of one or many project teams and sub-teams.	The manager of a project normally has more administrative responsibility than actual work involvement.

Related Work



Related Work		
Relation	**IsCitedBy**	indicates that B includes A in a citation (recommended for discovery).
Relation	**Cites**	indicates that A includes B in a citation (recommended for discovery).
Relation	**IsSupplementTo**	indicates that A is a supplement to B (recommended for discovery).
Relation	**IsSupplementedBy**	indicates that B is a supplement to A (recommended for discovery).
Relation	**IsContinuedBy**	indicates A is continued by the work B
Relation	**Continues**	indicates A is a continuation of the work B
Relation	**HasMetadata**	indicates resource A has additional metadata B
Relation	**IsMetadataFor**	indicates additional metadata A for a resource B
Relation	**IsNewVersionOf**	indicates A is a new edition of B, where the new edition has been modified or updated
Relation	**IsPreviousVersionOf**	indicates A is a previous edition of B
Relation	**IsPartOf**	indicates A is a portion of B; may be used for elements of a series (recommended for discovery).
Relation	**HasPart**	indicates A includes the part B (recommended for discovery).
Relation	**IsReferencedBy**	indicates A is used as a source of information by B
Relation	**References**	indicates B is used as a source of information for A
Relation	**IsDocumentedBy**	indicates B is documentation about/ explaining A
Relation	**Documents**	indicates A is documentation about/B
Relation	**IsCompiledBy**	indicates B is used to compile or create A
Relation	**Compiles**	indicates B is the result of a compile or creation event using A
Relation	**IsVariantFormOf**	indicates A is a variant or different form of B, e.g. calculated or calibrated form or different packaging
Relation	**IsOriginalFormOf**	indicates A is the original form of B
Relation	**IsIdenticalTo**	indicates that A is identical to B, for use when there is a need to register two separate instances of the same resource
Relation	**IsReviewedBy**	indicates that A is reviewed by B
Relation	**Reviews**	indicates that A is a review of B
Relation	**IsDerivedFrom**	indicates B is a source upon which A is based
Relation	**IsSourceOf**	indicates A is a source upon which B is based

Embargo


Dates
Date from
Date to
Created
YYYY-MM-DD
Embargo until
YYYY-MM-DD
Valid
YYYY-MM-DD
YYYY-MM-DD
The date the resource is made publicly available. Use this field to indicate the end of an embargo period.

Time for questions

Icon made by Freepik

Why publish research data?

How to publish?

Licences

GFZ Data Services

Wrap up

FID GEO funded by DFG

Key messages

- Publish your data
- Use a domain repository
- Provide rich metadata
- Use an „open as possible“ licence

huebner@gfz-potsdam.de

FID GEO funded by DFG