



# Open Access and Data Publishing

#### **Barbara Hirschmann**

Research Data Management Workshop Series, Part 6 5 April 2023

#### Flashback Workshops 1-5 Open Science **WS1** WS2 Data management RDM basics, rules and planning regulations Active research data management Hypothesis / Research question Re-use or collection Reproducible scientific Access and Analysis and interpretation verification calculations and data analysis Open access and data publishing Dealing with (highly) confidential data



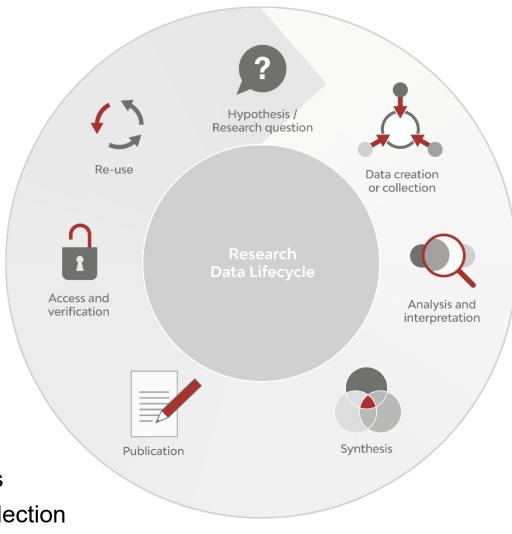
# Agenda

#### Open access publishing

- «Green Road» (self-archiving), incl. copyright questions
- «Gold Road» (OA journals), incl. funding & quality aspects
- ETH & research funders' OA policies

#### Publishing research data

- Why publish research data?
- FAIR Data Principles & trustworthy repositories
- Copyright & Creative Commons Licences
- Preparing your data for publication: Dos and dont's
- How to publish data in ETH Zurich's Research Collection





### Open access publishing

What is Open Access?

accessible

without costs for the reader

 «Open access […] literature is digital, online, free of charge, and free of most copyright and licensing restrictions.»
 (Peter Suber, 2012)

possibility to reuse

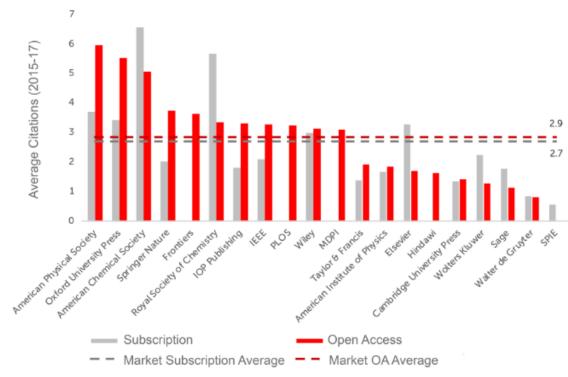
#### © creative commons

- Download
- Copy
- Distribute
- Print
- Textmining
- ...



# Why Open Access?

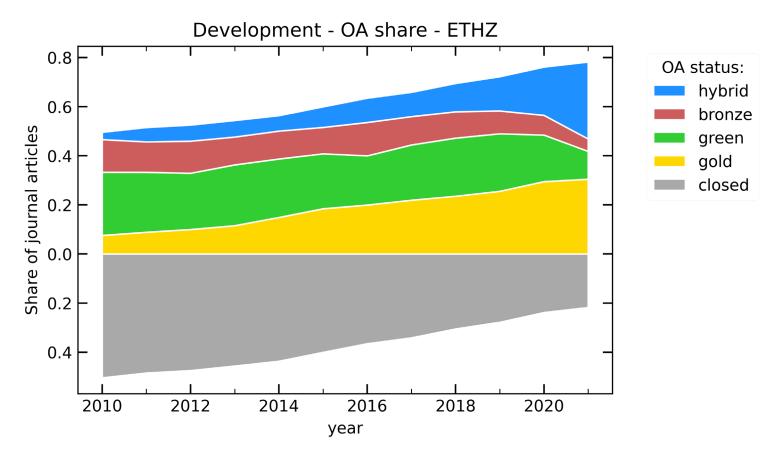
- Visibility of research results, easy access, citations
- Authors retain copyright
- Enables interdisciplinary and international collaboration
- Knowledge produced by public funds should be openly available
- Because your research funder requires it



Source: https://blog.frontiersin.org/2018/07/11/scientific-excellence-at-scale-open-access-journals-have-a-clear-citation-advantage-over-subscription-journals/



# Current developments & state



Source: dimensions.ai, data as of Jan. 2022



#### How to publish open access

#### **Gold Road**



#### Publish directly in

- Open access journals
- Open access books
- Subscription journals with OA option («hybrid journals»)

#### **Green Road**



Self-archive «subscription content» in

- an institutional repository
   (e.g. ETH Zurich's Research Collection)
- a disciplinary repository (e.g. arXiv.org, ...)

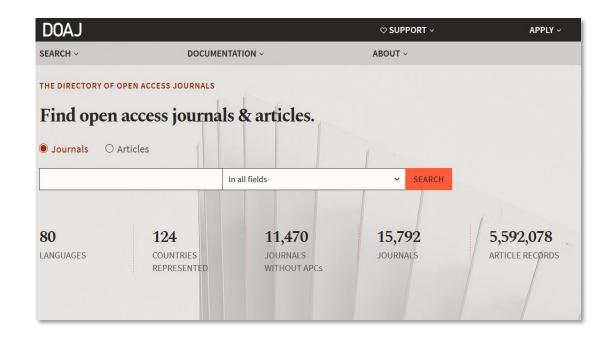


# Gold Open Access



#### Open access journals

- All content free to access and reuse
- Quality Control (Peer Review)
- Author retains copyright (standard licence: CC-BY)
- Different funding mechanisms
  - Institutional, society, library funding
  - Article Processing Charges (APCs): 0-6'000 EUR
- How to find OA journals?
  - Directory of Open Access Journals: <u>www.doaj.org</u>
  - Open access filter in <u>Journal citation reports</u>





### How to assess the quality of OA journals

- Did someone leave a review on QOAM (Quality Open Access Market)?
- Is the publisher a member of **OASPA** (the <u>Open Access Scholarly Publishers Association</u>) or **COPE** (the <u>Committee on Publication Ethics</u>)?
- Have a look at the journal's website:
  - Does the journal conduct a peer review process?
  - Are acknowledged scientists among the authors and editorial board members?
- More tips: <a href="http://thinkchecksubmit.org">http://thinkchecksubmit.org</a>



### Hybrid journals

- Subscription journals with paid open access option
- Mixed content: mainly available to licensing institutions, selected articles available open access
- APCs up to 10'000 EUR (!)
- No comprehensive database of hybrid journals available
  - Most journals not indexed in DOAJ are hybrid journals
  - Check the journal website for author information & open access option



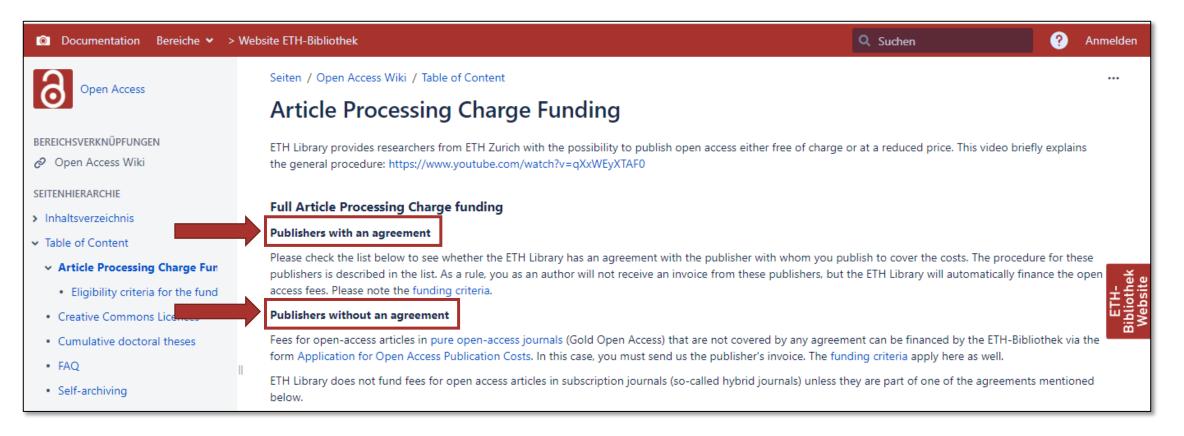
# Open access & hybrid journals: How much does it cost?

Journal	APC	DOAJ	Impact Factor 2021
Trees, Forests and People (Elsevier)	1'000 USD	<b>✓</b>	-
PLOS ONE	1'805 USD	⊘ DOAJ Seal	3.752
Chemical Science (RSC)	-	<b>✓</b>	9.969
Nature Communications	6'290 USD	<b>✓</b>	17.694
Hybrid Journals	+25-50%	no	



# Institutional APC funding

https://documentation.library.ethz.ch/display/OA/Article+Processing+Charge+Funding





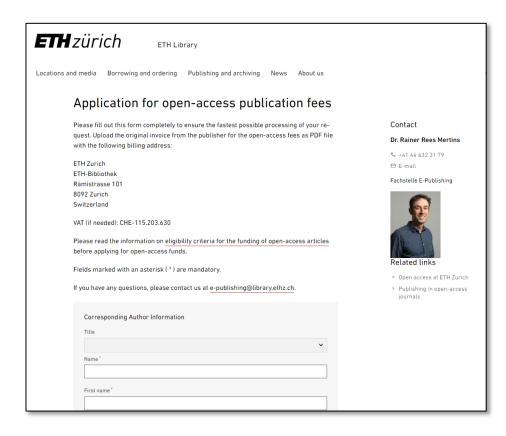
### Institutional funding – publishers with agreement

- > American Institute of Physics
- > Annual Reviews
- > BioMed Central/SpringerOpen..
- > Cambridge University Press
- > Company of Biologists
- > Copernicus Publications
- > Cogitatio
- Elsevier
- > Frontiers
- > IEEE
- > IGI Global
- > Institute of Physics
- Xarger
- > MDPI
- > Public Library of Science (PLOS)
- > Royal Society of Chemistry
- > SAGE
- > SciPost
- > SCOAP3 (Zeitschriften der Hochenergiephysik))
- > Springer Nature
- > Taylor & Francis
- > TransTech Publications
- Wiley

- ✓ Check our list if an agreement with the publisher exists and what conditions apply <a href="https://documentation.library.ethz.ch/display/OA/Article+Processing+Charge+Funding">https://documentation.library.ethz.ch/display/OA/Article+Processing+Charge+Funding</a>
- Submit your manuscript with your ETH email address to the publisher
- ✓ Go through publisher's affiliation check
  - ✓ Email domain
  - ✓ Author chooses
  - ✓ IP range
- ✓ Library approves the article
- Publisher sends the APC invoice directly to the library (or APC is covered up-front by read&publish contract)
- Please check individual conditions and workflow details in the publisher list!



# Institutional funding – publishers without agreement



- ✓ Check via <a href="www.doaj.org">www.doaj.org</a> that your journal is a **Gold OA journal** (hybrid journals not covered!) and that the list price is not higher than USD 5'000
- ✓ Check that your journal is not part of an agreement
- Check that your article meets the other <u>funding criteria</u>
- ✓ Fill in the <u>form</u> (and upload the invoice, if you already have it)
- ✓ Library checks eligibility of your article
- ✓ Library pays invoice to publishers or refunds your paid invoice



### Institutional APC funding: Eligibility

#### **ETH Library covers APCs if:**

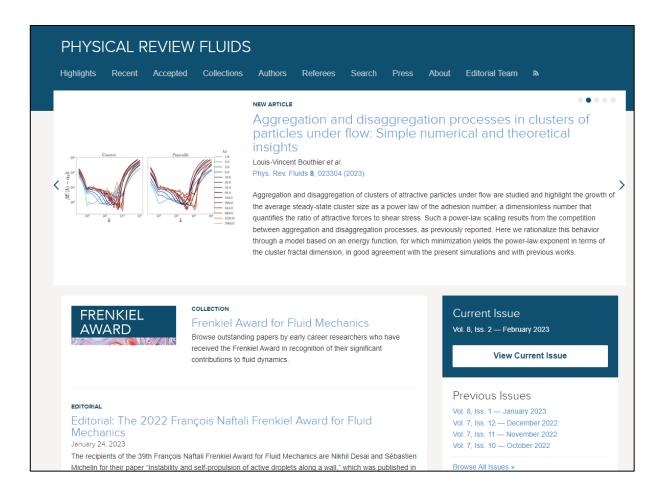
- Submitting or corresponding author must be a member of ETH Zurich
- Article Affiliation details -> must clearly indicate the author's affiliation with ETH.
- An institutional e-mail address (jane.doe@inst.ethz.ch) has to be used.
- ETH publications **submitted after the corresponding author has left**, are generally eligible for funding  $\rightarrow$  Confirmation might be necessary!
- Outside agreements: Maximum fee 5'000\$, only OA fees are covered
- Externally funded publications (SNSF, EU, etc.) have to be financed by the respective funder.
- License Recommendation → Creative Commons (CC or equivalent), as free as possible (e.g. CC BY).
- Open Access Policy: All ETH publications must be self-archived on the Research Collection.



# Green Open Access (self-archiving)



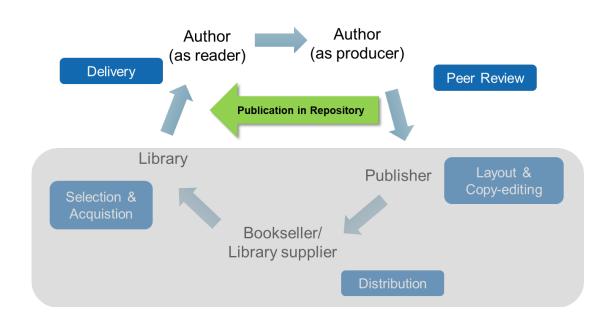
#### Example



- No OA agreement between ETH Libray and APS (yet)
- Hybrid journal (no APC funding through ETH Library)
- What to do?
  - → Green Road to Open Access



# Green Road (Self-archiving)



#### How does it work?

- Publish in a «conventional» subscription journal / conference proceedings /book
- Deposit a peer-reviewed open access version (e.g. author's accepted manuscript) of your paper in a repository

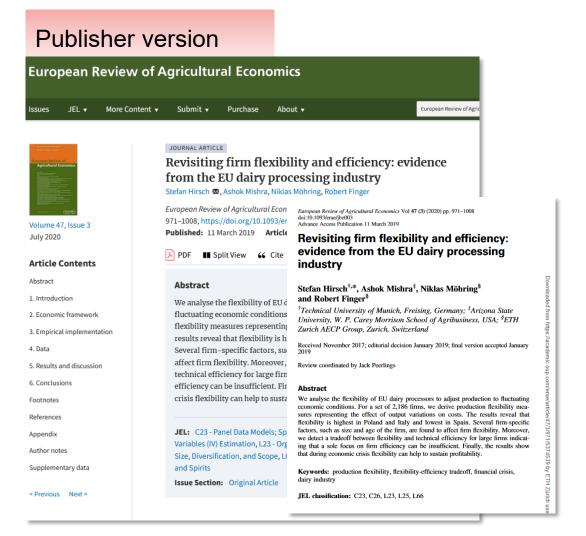


# Self-archiving: Copyright issues

 Subscription publishers usually require authors to transfer copyright or grant them an exclusive license to publish their work

- But: They also usually allow some sort of self-archiving
  - After a certain period of time (embargo)
  - If you do not upload the publisher's PDF but a manuscript version of the paper (no publisher layout & logo)

### Green Road OA publishing via the Research Collection



#### Open access version





# Types of repositories

#### OpenDOAR

(Directory of Open Access Repositories)

Institutional Repositories

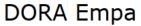
Disciplinary Repositories Multidisciplinary Repositories Academic Networking Site











Digital Object Repository at Empa









- Commercial
- No copyright
   clarification
- Papers might be taken down later
- Not compliant with funder policies

- Not-for-profit
- All uploads checked against publishers' copyright regulations by library staff

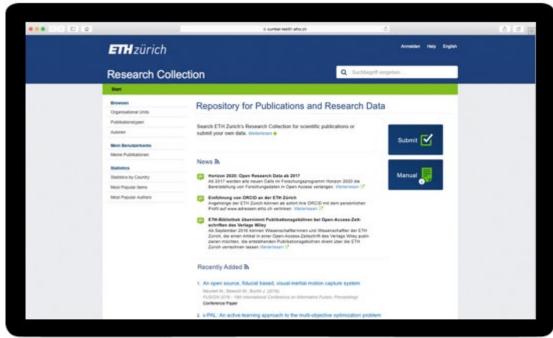




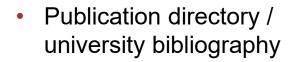
- Usually mainly preprints
- Not all of them publish the author's accepted manuscript



# ETH Zurich's repository: Research Collection













Research data repository



www.research-collection.ethz.ch

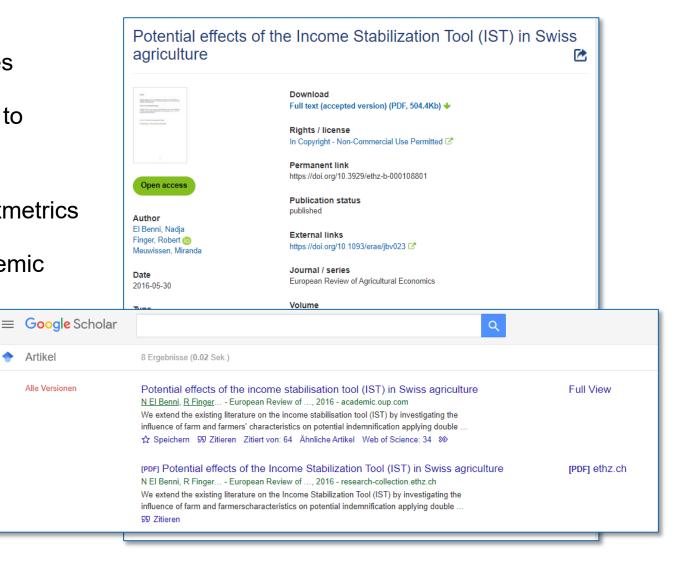


# What does the Research Collection offer for OA publishing?

Artikel

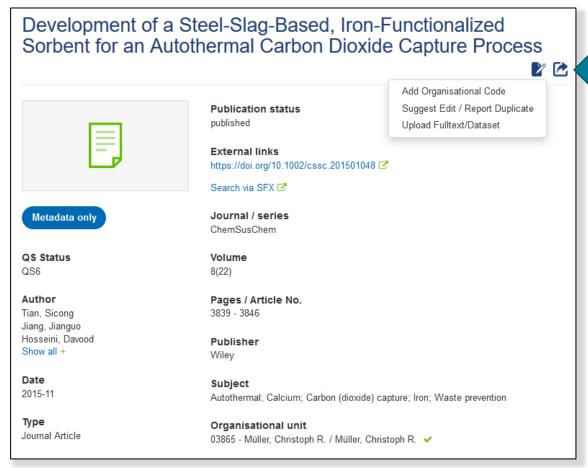
Alle Versionen

- «Automatic» upload of Open Access articles
- Persistent Identifiers (DOIs) and possibility to reserve a DOI
- Usage Statistics, «Cited by» counts and Altmetrics
- Indexed in Google Scholar and other academic search engines
- Long-term archiving in ETH Zurich's IT infrastructure
- Advise on copyright issues
- and more...





### Upload fulltexts / add organisational code / suggest edit



Instructions: https://documentation.library.ethz.ch/x/ ICM

Full texts have to be uploaded by the copyright/rights holder or an authorized representative (exception CC licenses). https://www.research-

collection.ethz.ch/terms-of-use

 No reason to worry about embargo periods for researchers → we take care

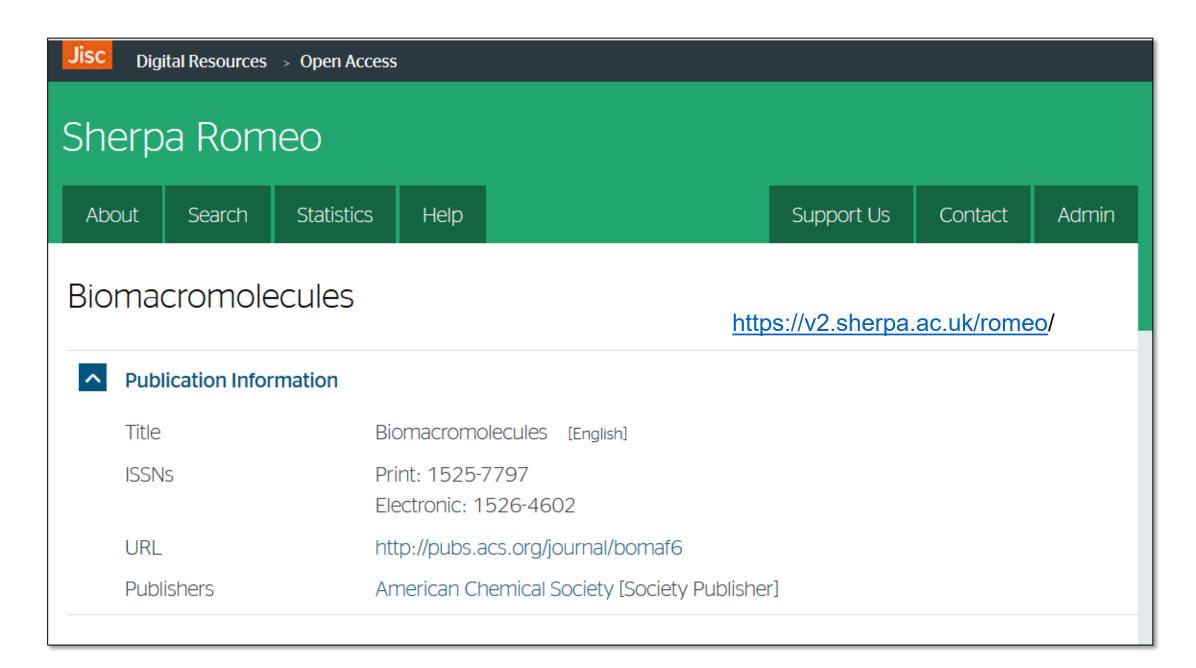
Video tutorial



### Green Road: copyright issues

- Where to look up the regulations?
  - SHERPA/RoMEO Database: <a href="http://www.sherpa.ac.uk/romeo">http://www.sherpa.ac.uk/romeo</a>
  - Publisher/journal website
  - Copyright transfer agreement / publishing contract / license to publish
  - For PhD Theses
     <u>https://documentation.library.ethz.ch/display/RC/Cumulative+doctoral+theses</u>

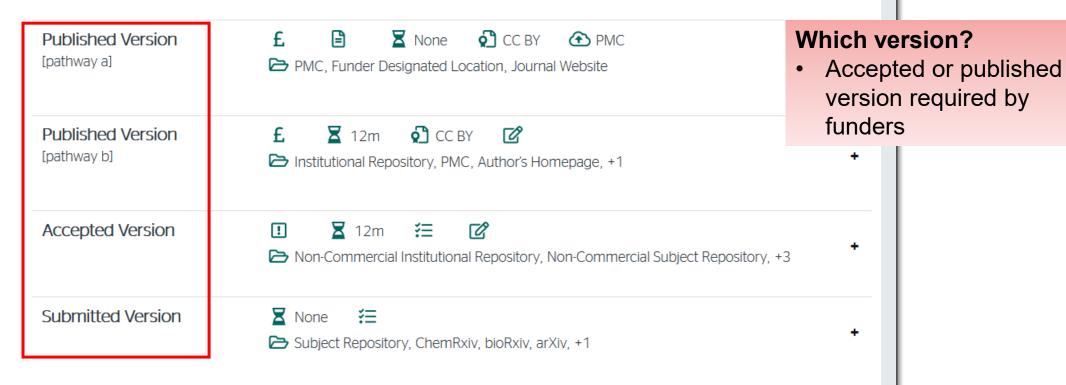






#### Publisher Policy

Open Access pathways permitted by this journal's policy are listed below by article version. Click on a pathway for a more detailed view.



For more information, please see the following links:

- ACS Journal Publishing Agreement
- Copyright & Permissions
- Funder Specific Options U.S. National Institutes of Health
- ACS Journal Publishing Agreement User's Guide



X None CC BY PMC **Published Version** [pathway a] PMC, Funder Designated Location, Journal Website £ OA Fee This pathway has an Open Access fee associated with it OA Publishing This pathway includes Open Access publishing **X** Embargo No Embargo **1** Licence CC BY 4.0 CC BY-NC-ND 4.0 Publisher's Bespoke License © Copyright Owner Publishers Publisher Deposit PubMed Central ▶ Location Funder Designated Location Named Repository (PubMed Central) Journal Website

For most publishers you must pay an Open Access fee for self-archiving the **publisher's version** of the article.

Published Version [pathway b]	£   12m ♠ CC BY   Institutional Repository, PMC, Author's Homepage, +1
£ OA Fee	This pathway has an Open Access fee associated with it
<b>E</b> mbargo	12 Months
<b>♠</b> Licence	CC BY 4.0 CC BY-NC-ND 4.0 Publisher's Bespoke License
© Copyright Owner	Publishers
<b>►</b> Location	Author's Homepage Institutional Repository Named Repository (PubMed Central) Journal Website
✓ Notes	Reduced fee



Accepted Version	<ul><li>I</li></ul>	Commercial Subject Repository, +3	
Prerequisites	If Required by Funder, If Required by Institu	tion	
<b>Embargo</b>	12 Months	When?	
<b>▷</b> Location	Author's Homepage Institutional Website	<ul> <li>Possible embargos between 0 and 48 months</li> </ul>	
	Non-Commercial Institutional Repository	\\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \	
	Non-Commercial Subject Repository Preprint Repository	<ul> <li>Where?</li> <li>Author's website</li> <li>Institutional reposit</li> </ul>	
<b>₹≡</b> Conditions	Must be accompanied by set statement (see policy)  Must link to publisher version		
✓ Notes	If mandated to deposit before 12 months, institution/Funding agency or use AuthorCh		

Submitted Version 

▼ None 

▼ ■

Subject Repository, ChemRxiv, bioRxiv, arXiv, +1

▶ Location Named Repository (ChemRxiv, bioRxiv, arXiv)

Preprint Repository

Subject Repository

E Conditions Must not violate ACS ethical Guidelines

Must note use if preprint server in cover letter and provide link to deposit

Upon publication add a link to published article with DOI

Rules for **preprints** are the most liberal, there is no contract with the publisher yet!



### Green Road: copyright issues

- Where to look up the regulations?
  - SHERPA/RoMEO Database: <a href="http://www.sherpa.ac.uk/romeo">http://www.sherpa.ac.uk/romeo</a>
  - Publisher/journal website (→ e.g. RSC)
  - Copyright transfer agreement / publishing contract / license to publish
  - For PhD Theses
     <a href="https://documentation.library.ethz.ch/display/RC/Cumulative+doctoral+theses">https://documentation.library.ethz.ch/display/RC/Cumulative+doctoral+theses</a>



# Publisher/journal website

Sharing rights	Accepted manuscript	Version of record
Share with individuals on request, for personal use	✓	✓
Use for teaching or training materials	✓	✓
Use in submissions of grant applications, or academic requirements such as theses or dissertations*	✓	✓
Share with a closed group of research collaborators, for example via an intranet or privately via a scholarly communication network	✓	✓
Share publicly via a scholarly communication network that has signed up to STM sharing principles	Σ	×
Share publicly via a personal website, <u>institutional repository</u> or other not-for-profit repository	互	×
Share publicly via a scholarly communication network that has not signed up to STM sharing principles	×	×



Source: RSC Website

### Green Road: copyright issues

- Where to look up the regulations?
  - SHERPA/RoMEO Database: <a href="http://www.sherpa.ac.uk/romeo">http://www.sherpa.ac.uk/romeo</a>
  - Publisher/journal website (→ e.g. RCS)
  - Copyright transfer agreement / publishing contract / license to publish
  - For PhD Theses
     <a href="https://documentation.library.ethz.ch/display/RC/Cumulative+doctoral+theses">https://documentation.library.ethz.ch/display/RC/Cumulative+doctoral+theses</a>



### Copyright transfer agreement



#### AGREEMENT IN RELATION TO COPYRIGHT IN AN ARTICLE FOR A TAYLOR & FRANCIS/ROUTLEDGE JOURNAL

In order to ensure both the widest dissemination and protection of material published in our Journal, we ask Authors to transfer to the Publisher, Taylor & Francis, the rights of copyright in the Articles they contribute. This enables Taylor & Francis to ensure protection against infringement. The transfer of copyright must be clearly stated in writing.

#### YOUR STATUS

I am the sole author of the Article

Please indicate if any of the statements below also apply to you:

- I am a UK, Canadian or Australian Government employee and claim Crown Copyright
- I am a US Government employee and there is no copyright to transfer
- I am an NIH employee and there is no copyright to transfer. I submit this form together with an NIH addendum.
- I am a contractor of the US Government (includes NIH contractors) under contract number:
  - I am required to sign this form

#### ASSIGNMENT OF PUBLISHING RIGHTS

I hereby assign to Taylor & Francis the copyright in the above specified manuscript (government authors not transferring copyright hereby assign a non-exclusive licence to publish) and any accompanying tables, illustrations, data and any other supplementary information intended for publication in all forms and all media (whether known at this time or developed at any time in the future) throughout the world, in all languages, for the full term of copyright, to take effect if and when the article is accepted for publication. If I am one of several co-authors, I hereby confirm that I am authorized by my co-authors to grant this Licence as their agent on their behalf. For the avoidance of doubt, this assignment includes the rights to supply the article in electronic and online forms and systems.

I confirm that I have read and accept the full terms of the Journal's article publishing agreement attached to this form including my author warranties, and have reviewed the Journal's policies on Author Rights.

Signed:	Name Printed:	
Title and Company (if employer representative):		Date:

Source: T&F Copyright Agreement

Non-exclusive license



Exclusive license



Transfer copyright





### Green Road: copyright issues

- Where to look up the regulations?
  - SHERPA/RoMEO Database: <a href="http://www.sherpa.ac.uk/romeo">http://www.sherpa.ac.uk/romeo</a>
  - Publisher/journal website
  - Copyright transfer agreement / publishing contract / license to publish
  - For PhD Theses
     https://documentation.library.ethz.ch/display/RC/Cumulative+doctoral+theses



### Publishing a Doctoral Thesis in the Research Collection

#### Managing Trade-Offs in Coffee Agroforests





Embargoed until 2019-02-12

Author Nesper, Maike

**Date** 2017

Type

Doctoral Thesis

#### Download

Full text (Adobe PDF, 18.97Mb) ❖
Abstract (Adobe PDF, 171.4Kb) ❖

#### Abstract

Agroforestry systems provide multiple ecosystem services and are refuges for diverse flora and fauna. Coffee is one of the main tropical agroforestry crops and of global economic importance. Over recent decades, coffee agroforestry systems have been increasingly intensified by the replacement of na Show more +

#### Permanent link

https://doi.org/10.3929/ethz-b-000239904

#### Publication status

accepted

#### External links

Search via SFX 3

#### Contributors

Supervisor: Ghazoul, Jaboury Supervisor: Verweij, Pita Supervisor: Küffer, Christoph

#### Publisher

ETH 7urich

#### Chapter 2 |

Shade tree diversity enhances coffee production and quality in agroforestry systems in the Western Ghats

with C. Kueffer, S. Krishnan, C.G. Kushalappa, J. Ghazoul

Published in Agriculture, Ecosystems and Environment (2017), 247, 172 - 181

#### Abstract

Intensification of multispecies coffee agroforests reduces shade tree diversity with implications for tropical biodiversity.

- Mention original publication and coauthors for each chapter
- Publisher's version are forbidden in cumulative thesis @ETH
- Check publishers' requirements in our <u>list</u>



#### Hands-on I

Imagine you want to publish articles in these three journals (see Miro) and you want to publish open access:

- What would be the best open access (funding) option? Drag the journal symbol to the respective frame.
- What OA type is it (Gold, Green, Hybrid)?
- How long is the embargo period for green OA?



# Funder requirements







#### Regulations

OA mandatory	<ul> <li>Peer-reviewed publications (journal articles, books, book chapters, conference proceedings)</li> </ul>
Requirement	Gold or Green
Max. embargo period	<ul> <li>Articles: 6 months (immediate OA → applications 01/23 onwards)</li> <li>Books: 12 months</li> </ul>
Licence	For applications 01/2023 onwards: CC-BY mandatory
Proportion	SNSF funded minimum 50% of the research
Compliance monitoring	Via P3 / mySNF open access platform, <a href="http://www.snsf-oa-check.ch/">http://www.snsf-oa-check.ch/</a>
Funder statement	Mention SNSF funding in the acknowledgements



# Open Access in SNSF-funded projects



#### **Funding**

Publication types	<ul> <li>Journal articles, books, book chapters, conference proceedings (as book chapters)</li> </ul>
OA type	Gold Open Access
Application	• Upload the invoice via ChronosHub → SNSF pays the publisher directly
Timing	up to 6 months after publication, also after project end
Who	Grantees or project team members – mentioned as author
Quality	DOAJ recommended



# Open Access in EU-funded projects



#### Regulation

OA mandatory	Peer-reviewed publications
Requirement	Gold or Green
Max. embargo period	<ul> <li>STM: 6 months (0 Horizon Europe)</li> <li>SSH: 12 months (0 Horizon Europe)</li> </ul>
Licence	Horizon Europe: CC BY mandatory
Funding for OA articles	<ul> <li>APCs for Gold (and Hybrid) journals can be claimed as project expenses → Include APCs in grant application</li> </ul>
Compliance monitoring	Via reporting, institutional repositories and EU portal openAIRE



# ETH Zurich's Open Access Policy



#### Regulation

Mandatory	Research papers, theses and any other research output
Requirement	<ul><li>Gold or Green</li><li>Repository: Research Collection</li></ul>
Max. embargo period	3 years for Doctoral Theses
Funding for OA articles	<ul> <li>ETH Library covers APCs for many OA &amp; subscription journals</li> <li>ETH Grants: APCs for Gold and Hybrid journals can be claimed as project expenses → Include it in application!</li> </ul>



#### Hands-on II

#### Have a look at the solution of Hands-on I

- Would the funder requirements be fulfilled for:
  - A) The SNSF
  - B) The EU under Horizon Europe conditions
- Would there be funding available?



### Challenges re. OA funding and compliance

- Gold OA: Immediate Open Access is sometimes still hard to achieve:
  - Hybrid Journals outside institutional agreements
  - Gold OA journals with APCs >5'000 USD
  - Non-corresponding author articles
- Green OA: Maximum embargo periods of funders often shorter than minimum embargo periods of journals
- New SNSF and Horizon Europe rules (0 months embargo, CC licence) will be challenging
- Books, book chapters, conference proceedings,...



# Publishing research data



### Group survey

### Please share your experiences with data publishing and reuse:

- Have you already published a dataset yourself?
- Have you reused an open dataset published by other researchers?
- Have you generally reused data produced by other researchers?



# Publishing research data

- Why publish research data?
- FAIR Data Principles & trustworthy repositories
- Copyright & Creative Commons Licences
- Preparing your data for publication: Dos and dont's
- How to publish data in ETH Zurich's Research Collection







# Why publish research data?



## Why? – Funder research data policies



Research data should be freely accessible to everyone – for scientists as well as for the general public.

(https://www.snf.ch/en/dMILj9t4LNk8NwyR/topic/open-research-data)



"as open as possible, as closed as necessary"

(https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination en.htm)

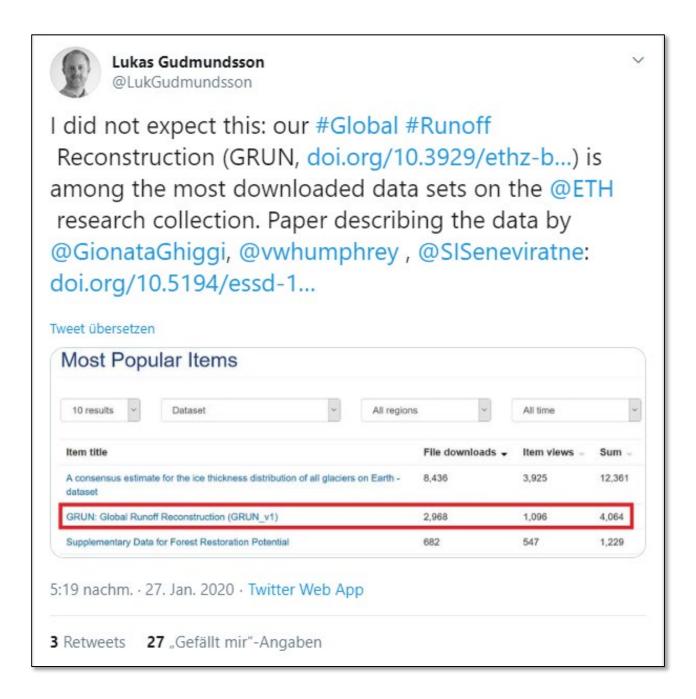


### Why? – Journal research data policies

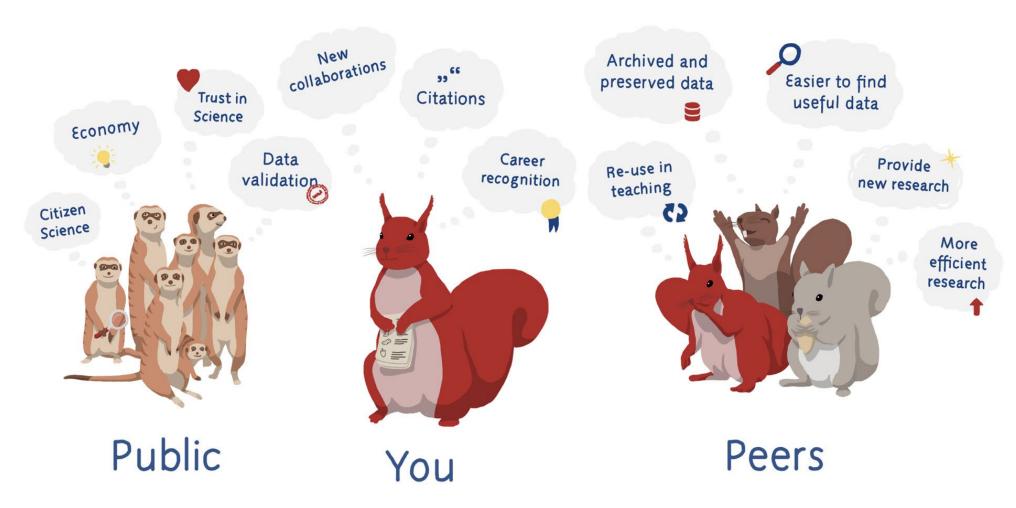
- SpringerNature Research Data Policies
   https://www.springernature.com/gp/authors/research-data-policy
- PLOS Data Availability Policy <a href="https://journals.plos.org/plosone/s/data-availability">https://journals.plos.org/plosone/s/data-availability</a>
- Copernicus Data Policy
   https://www.publications.copernicus.org/services/data\_policy.html







# Benefits of Data Sharing

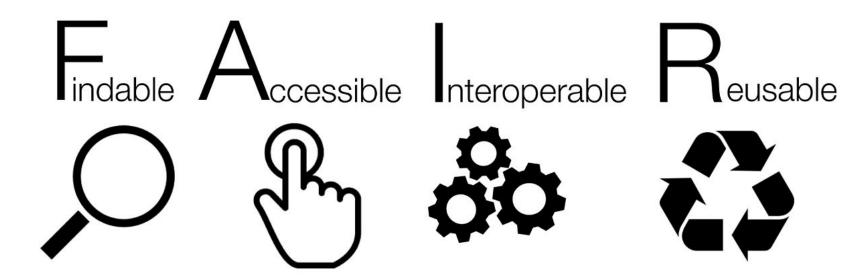




# FAIR Data Principles & trustworthy repositories



## How to publish data according to the FAIR Data Principles



CC-BY-SA 4.0 Sangya Pundir https://upload.wikimedia.org/wikipedia/commons/a/aa/FAIR\_data\_principles.jpg

The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, Issue 3, 2016. 10.1038/sdata.2016.18.



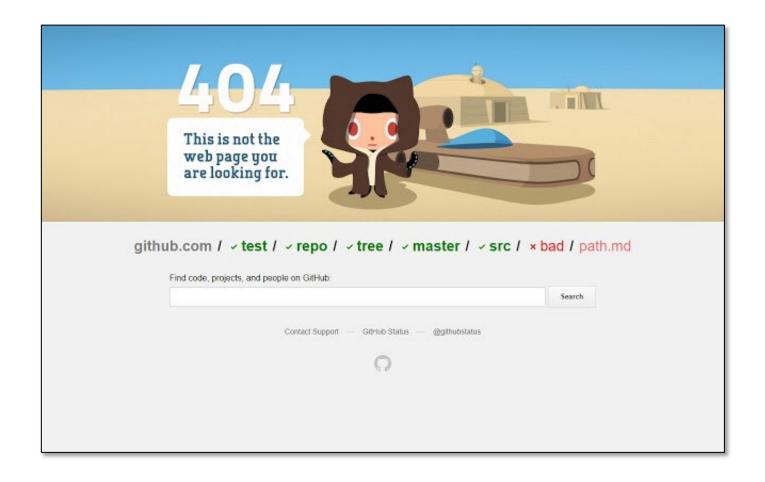
#### Hands-on III

### How to publish data in line with the FAIR Data Principles?

- Have a look at the features and requirements written on the sticky notes.
- Which of the four aspects of FAIR does it help you to implement in practice?
   Move the note to the respective letter on the right side.
- You can also copy a stick a note if it enables you to implement more than one aspect of FAIR.



# DOIs – Solving link rot with Persistent Identifiers





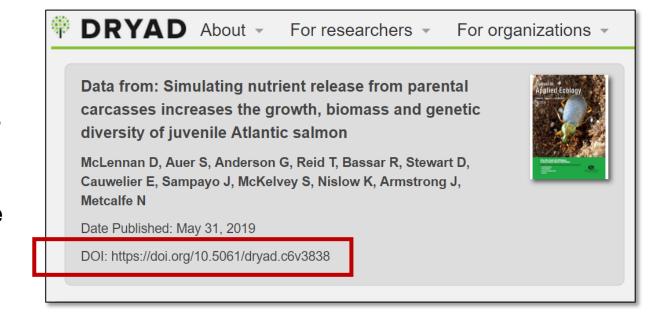
### DOIs – Solving link rot with Persistent Identifiers

- A Persistent Identifier is a long-lasting reference to a document, file, web page, or other object
  - It usually comes with a resolving mechanism where each Persistent Identifier resolves to a URL
  - While the URL where a publication / dataset is located may change over time, the PI does not change and always leads to the most current location of a resource



### DOIs – Solving link rot with Persistent Identifiers

- A DOI is a Persistent Identifier that is «citable» because
  - ✓ Publishers / repositories update the URLs associated with DOIs
  - ✓ Publishers / repositories don't change the content associated with a DOI after its publication





#### Metadata

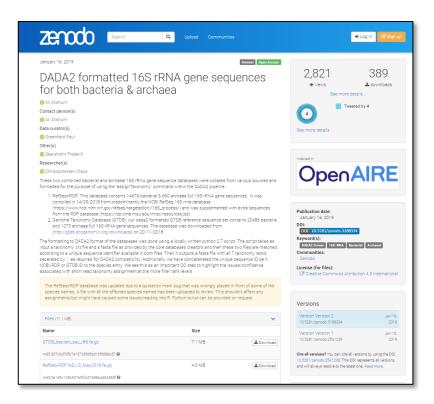
- Remember?
  - Metadata is the data about your data
  - Metadata can be supplied in the file header (e.g. TIFF, netCDF), as sidecar files (e.g. XML, JSON), as README text file or as columns in a database
- For publishing data in compliance with the FAIR Data Principles, you must provide this metadata to the repository
  - via a manual submission form
  - via XML upload, an API, ...





## Metadata / Landing Pages

 The repository will then provide the metadata to end-users in human as well as machine-readable form, via the dataset's "Landing Page"



```
html lang="en" dir="ltr">
meta http-equiv="X-UA-Compatible" content="IE=edge":
meta_name="viewport" content="width=device-width, initial-scale=1">
   <meta name="google-site-verification" content="5fPGCLllnWrvFxH9QWI011TadV7byeEvfPcyK2VkS s"/>
   <meta name="google-site-verification" content="Rp5zp04IKW-s1IbpT0GB7Z6XY60o1oZD5C3kTM-AiY4"/</pre>
 meta name="norton-safeweb-site-verification" content="umenay8zh4kswbi568zqp19bqb-jvngusibublygib0x3jne9rig0fnmtofm8abb7lkzgltqp5yhm68s5qz4iqqkm39xl2o-p5foixd-1xfq4yig07ugcd1sp5kmyvpe" />
   <title>DADA2 formatted 16S rRNA gene sequences for both bacteria &amp; archaea | Zenodo</title>
   <link rel="shortcut icon" href="/static/favicon.ico"/>
   k rel="apple-touch-icon-precomposed" sizes="114x114" href="/static/apple-touch-icon-114-precomposed.png"/>
   k rel="apple-touch-icon-precomposed" sizes="72x72" href="/static/apple-touch-icon-72-precomposed.png"
   k rel="apple-touch-icon-precomposed" sizes="57x57" href="/static/apple-touch-icon-57-precomposed.png"/
 ink href="https://fonts.googleapis.com/css?family=Roboto:300,400,500,700,100,italic" rel="stylesheet">
link href="/static/gen/zenodo.6d737863.css" rel="stylesheet">
 -- HTML5 shim and Respond.js for IE8 support of HTML5 elements and media queries -->
 <script src="https://oss.maxcdn.com/html5shiv/3.7.2/html5shiv.min.js"\times/script>
 <script src="https://oss.maxcdn.com/respond/1.4.2/respond.min.js"></script>
    <meta name="description" content="These two combined bacterial and archaeal 165 rRNA gene sequence databases were collated from various sources and formatted for the purpose of using the %#34;assignTaxonomy%#34; command</p>
ithin the DADA2 pipeline. RefSeq+RDP: This database contains 14676 bacterial & 660 archaea full 165 rRNA gene sequences. It was compiled in 14/05/2018 from predominantly the NCBI RefSeq 165 rrna database
https://www.ncbi.nlm.nih.gov/refseq/targetedloci/165 process/) and was supplemented with extra sequences from the RDP database (https://rdp.cme.msu.edu/misc/resources.jsp). Genome Taxonomy Database (GTDB): our dada2 formatte
TDB reference sequence set contains 20486 bacteria and 1073 archaea full 165 rRNA gene sequences. The database was downloaded from (http://gtdb.ecogenomic.org/downloads) on 20/11/2018. The formatting to DADA2 format of the
atabases was done using a locally written python 2.7 script. The script takes as input a taxonomy .txt file and a fasta file as provided by the core databases creators and then these two files are matched according to a
indige sequence identifier available in both files. Then it outputs a fasts file uith all 7 sources as a compared by $833;8834, as required for DADA2 compatibility. Additionally, we have concatenated the unique sequence is in NESI/KDP on GIDB ID to the species entry. We see this as an important QC step to highlight the issues/confidence associated with short read taxonomy assignment at the more finer rank levels." />

(meta name—"citation, title" content—"DADA2 formatted 15s FRNL agence sequences for both bacteria & Bang, anchase" />
    <meta name="citation_author" content="Ali Alishum" />
   <meta name="citation_publication_date" content="2019/01/16" />
<meta name="citation doi" content="10.5281/zenodo.3188334" />
    <meta name="citation_keywords" content="DADA2 format" /:</pre>
   <meta name="citation_keywords" content="165 rRNA" />
<meta name="citation_keywords" content="Bacterial" />
   <meta name="citation_keywords" content="Archaeal" />
    <meta name="citation_abstract_html_url" content="https://www.zenodo.org/record/3188334" />
   <meta property="og:title" content="DADA2 formatted 16S rRNA gene sequences for both bacteria & amp; archaea" />
   <meta property="og:description" content="These two combined bacterial and archaeal 165 rRNA gene sequence databases were collated from various sources and formatted for the purpose of using the %#34;assignTaxonomy%#34;</p>
 mmmand within the DADA2 pipeline. RefSeq+RDP: This database contains 14676 bacterial & 660 archaea full 165 rRNA gene sequences. It was compiled in 14/05/2018 from predominantly the NCBI RefSeq 165 rrna database
https://www.ncbi.nlm.nih.gov/refseq/targetedloci/165 process/) and was supplemented with extra sequences from the RDP database (https://rdp.cme.msu.edu/misc/resources.jsp). Genome Taxxonomy Database (GTDB) our dada2 formatter
TDB reference sequence set contains 20486 bacteria and 1073 archaes full 165 rRNA gene sequences. The database was downloaded from (http://gtdb.ecogenomic.org/downloads) on 2011/2018. The format GTDB of the contains 20486 bacteria and 1073 archaes full 165 rRNA gene sequences.
atabases was done using a locally written python 2.7 script. The script takes as input a taxonomy .txt file and a fasta file as provided by the core databases creators and then these two files are matched according to a
unique sequence identifier available in both files. Then it outputs a fasta file with all 7 taxonomy ranks separated by 8#34;8#34; as required for DADA2 compatibility. Additionally, we have concatenated the unique sequence be it NCBI/ADP or GTDB ID to the species entry. We see this as an important QC step to highlight the issues/confidence associated with short read taxonomy assignment at the more finer rank levels." />
   <meta property="og:url" content="https://www.zenodo.org/record/3188334" />
   <meta property="og:site_name" content="Zenodo" />
   <meta name="twitter:card" content="summary" /:</pre>
```



# What's a trustworthy repository?

- A repository
  - that provides all or most of the features described in the previous slides
- PLUS:
  - that is run by a long-lived institution
  - and that has sustainable funding sources





## Criteria for choosing a data repository

- Compliance with FAIR Data Principles
- Sustainability and trustworthyness
- Location of the service and its servers
- Legal regulations on data protection
- Immediate/long-term costs
- How can you get your data back?





### Data repositories – examples









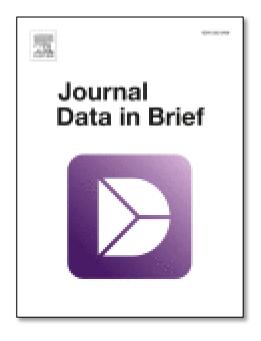


Data Publisher for Earth & Environmental Science

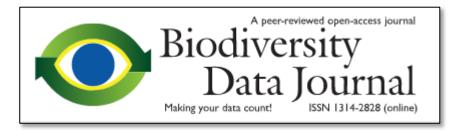


#### **Data Journals**



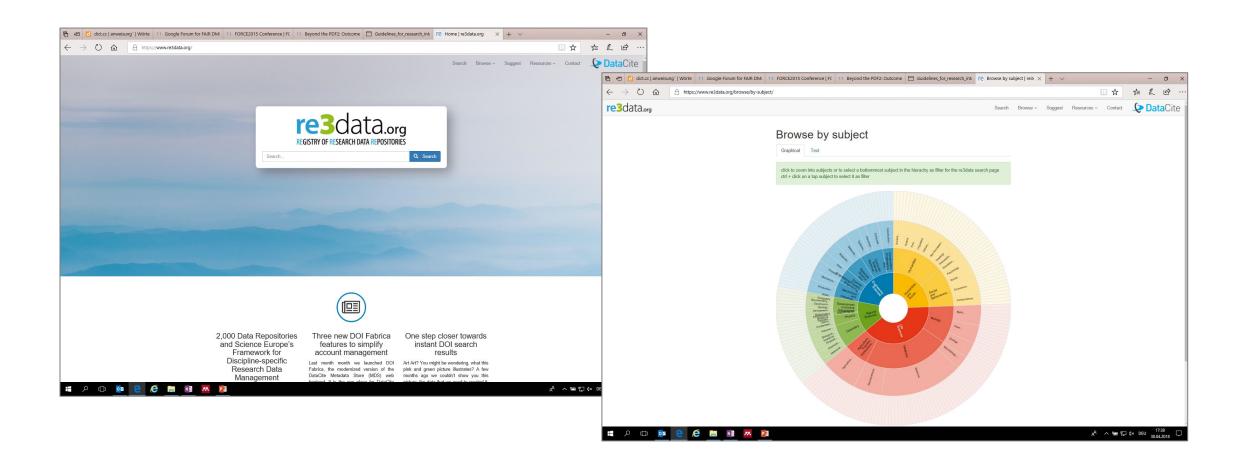








# Finding trustworthy data repositories – www.re3data.org





#### Hands-on IV

# How many repositories can you find in www.re3data.org that

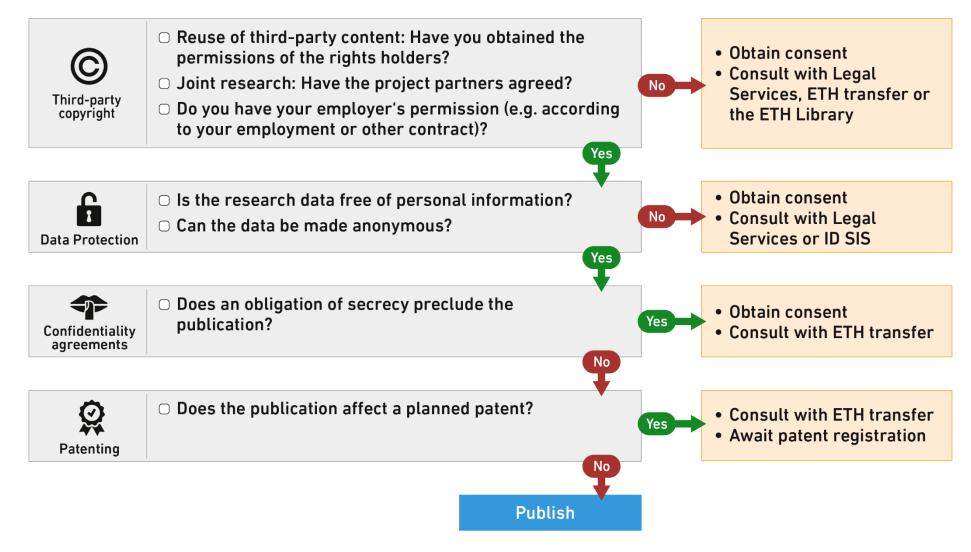
- are based in Switzerland
- and provide DOIs?



# Copyright & Creative Commons Licences



## Legal aspects: what to clarify before publishing your data





#### **Creative Commons Licences**



#### **What are Creative Commons Licences?**

 CC licences allow authors / data producers to define what types of reuse are permitted for their works



# Who can grant a Creative Commons licence?



- CC licences can only be granted by authors or copyright holders themselves
- All co-authors / data producers / rights holders must agree to the licensing



### Why should I apply a Creative Commons licence to my work?

- Because works published without CC licence can only be reused
  - with explicit **permission** from the copyright holder
  - or based on an exemption defined in copyright law
     (e.g., in Switzerland, for «personal use» or for «quotation in scientific works»)
     → c.f. The right of quotation



### Why should I apply a Creative Commons licence to my work?

- Whereas works published with CC licence can be reused as specified in the licence, for example
  - for illustrating a presentation
  - for distribution in classrooms
  - for republication in Wikipedia
  - for commercial purposes
  - ..



### How can a work be published under a CC license?

- Referring to the licence in the work itself (e.g. title page, legal notice section, readme.txt)
- Additionally: Adding the license to the landing page and metadata on a repository or publisher website
- The reference to the license should contain the name and stable URL of the licence as well as, optionally, the licence logo.
- Example:
  - This work is licensed under a <u>Creative Commons Attribution 4.0 International license</u>



CC licenses can be generated by combining the following elements:



Attribution (BY) – mandatory element in all licenses



Share-Alike (SA)



NonCommerical (NC)



NoDerivatives (ND) – recommended not to use in OA / ORD context



#### CC-BY

License	Terms	Pros	Cons
Creative Commons Attribution 4.0 International <a href="http://creativecommons.org/licenses/by/4.0/">http://creativecommons.org/licenses/by/4.0/</a>	The name of the author must be stated.	Allows unrestricted copying, redistribution, publication, adaptation	The author completely gives up control of re-use of his work.
© €Y		and use of a work.  Many authors and publishers consider it best practice to license open access articles under a CC-BY license.	Attribution can be complicated to achieve when automated methods are used to assemble a dataset from various CC-BY-licensed sources.



#### CC-BY-SA

License	Terms	Pros	Cons
Creative Commons Attribution-ShareAlike 4.0 International <a href="http://creativecommons.org/licenses/by-sa/4.0/">http://creativecommons.org/licenses/by-sa/4.0/</a>	The name of the author must be stated.  The work must be distributed under the same license after any changes.	This license has the character of a "viral license" which means that new works resulting from changes must be distributed under the same license.	It may lead to undesired limitations in re-use due to "license incompatibilities": For example, the material used within a CC-BY-SA article can only be combined and republished with other CC-BY-SA content.



#### **CC-BY-NC**

License	Terms	Pros	Cons
Creative Commons Attribution-NonCommercial 4.0 International <a href="http://creativecommons.org/licenses/by-nc/4.0/">http://creativecommons.org/licenses/by-nc/4.0/</a>	The name of the author must be stated.  No commercial use is permitted.	This license is suitable if there is a realistic prospect of commercial users paying for use of the material.	May prevent use in education and research (e.g. use in fee-paying courses of study or public-private partnerships).
			May also prevent re-use in non-commercial services if they require a more liberal license (e.g. Wikipedia).

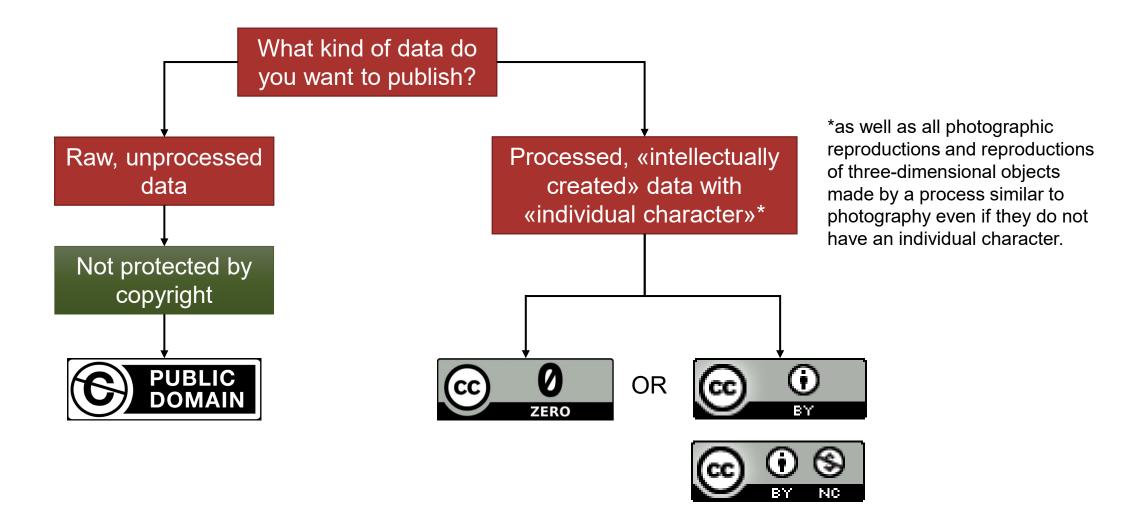


- ✓ The main factor in deciding in favor of a specific license should be your individual intention.
- ✓ Consider that each additional license limitation always also leads to increased legal uncertainty.

This means that more restrictive licenses are generally linked to disadvantages for licensors and users alike.



#### How to mark / licence Open Research Data for reuse





# Preparing your data for publication: Dos and dont's



## Document your data

- Include a README
- Use informative file and folder names
- Include metadata
- Link it to an article or other publication



### Check your data for legal compliance

- Do you and your group members have the rights to all the data?
- If not, do you have permission to publish the data?
- Are you complying with the licence terms of any third party data?
- Are you complying with the regulations of your University (e.g. regarding software licences)?
- Under wich licence will you publish your dataset?
- Have you included the licence terms in your README file?



### Edit your data for publication

- Remove temporary and backup files
- Rename files and folders where helpful
- Remove third party files and software for which you don't have permission
- Check for hardcoded file paths, symbolic links, references
- Don't include your manuscript
- Spell check your text files

 Good folder structure and regular cleanup should ideally already be part of the data management process



### Choose appropriate file formats

- Domain specific
- Open standards
- Avoid proprietary, undocumented filetypes
- For long term preservation: use recommended formats
   (https://documentation.library.ethz.ch/display/RC/File+formats+for+archiving+research+data)
- Use standard containers
  - Zip or TAR-Files
  - No encryption/passwords



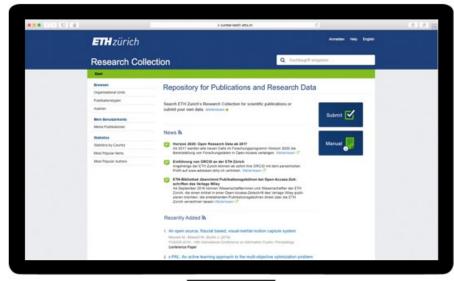
# Uploading your data into the Research Collection



#### ETH Research Collection – www.research-collection.ethz.ch

- One-stop-shop for depositing research output
  - Publications, research data
  - Web upload, DOI-reservation and registration,
     ORCID, export to OpenAire...
  - Long-term preservation in ETH Data Archive

 Aligned with FAIR principles according to SNSF guidelines







#### Research Collection for research data

- Publish, share or archive research data
- As supplementary material or stand-alone publication
- All file formats allowed, recommendations for suitable formats provided
- Max. file size: 10 GB (50 GB per entry)
  - New service «libdrive» for larger datasets
- Free service for ETH members





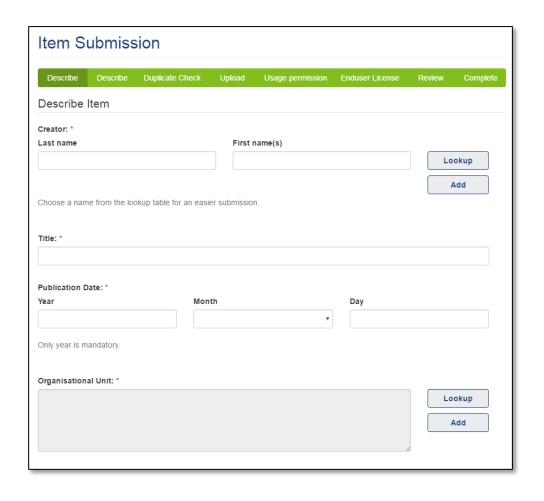
### Step 1: Choose a research data type

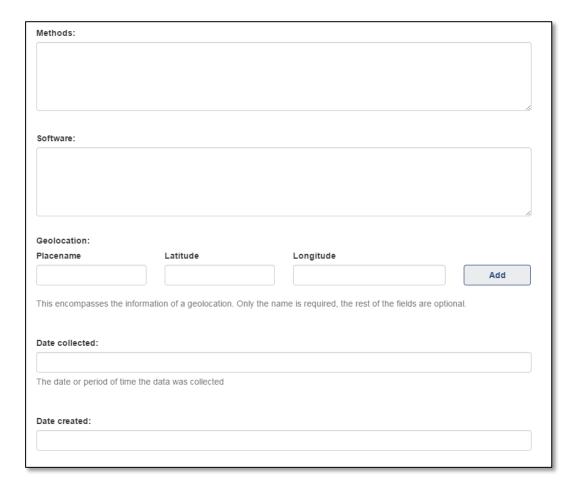






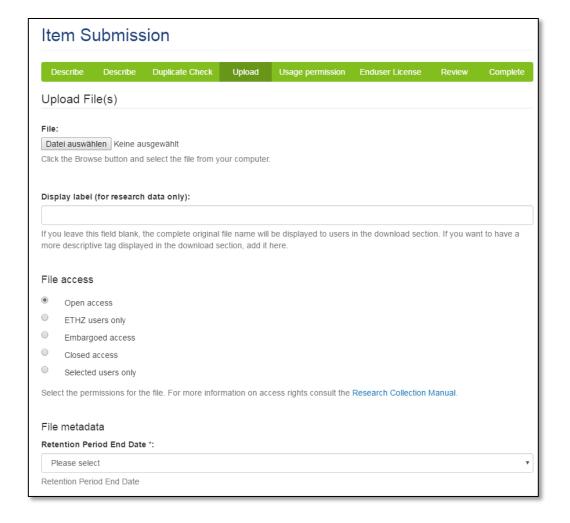
## Step 2: Describe your dataset







### Step 3: Upload file(s)





### Access rights options

	Open access	Embargoed	ETHZ users	Selected users	Closed access
Publications	✓	✓			
Research data	$\checkmark$	$\checkmark$	✓	<b>√</b>	$\checkmark$

- Access rights refer to files. Metadata are always visible and indexed in search
- Closed access and selected user items do not receive a DOI
- Access to restricted data can be requested via a form, decision is taken by submitter / rights holder



### Retention periods

Submitters choose between one of the following retention periods:

10 years «Good scientific practice»

15 years «Good scientific practice» + safety margin

Indefinite
 Use of <u>recommended file formats</u> strongly encouraged

> Gives ETH Library the possibility to take necessary archiving / deletion decisions



### Step 4: Grant usage permissions and end user license

#### Licenses for research data

Creative Commons Attribution 4.0 International

Creative Commons Attribution-ShareAlike 4.0 International

Creative Commons Attribution-NonCommercial 4.0 International

Creative Commons Attribution-NoDerivatives 4.0 International

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International

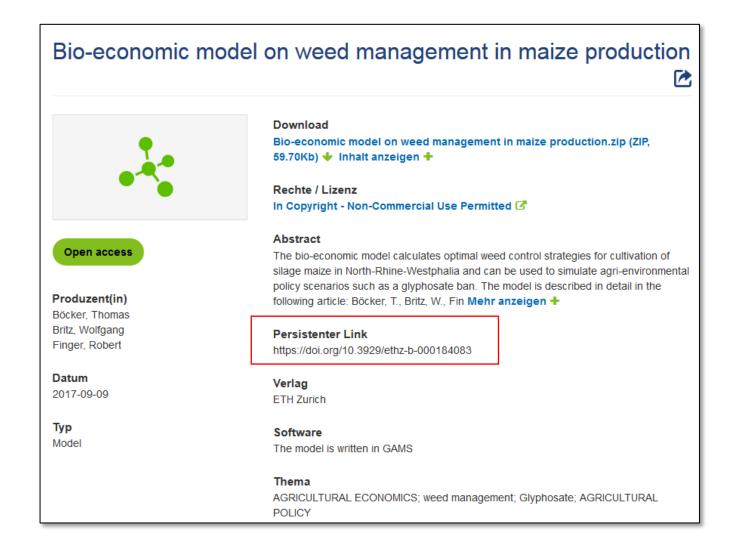
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

Public Domain Dedication (CC0 1.0 Universal)



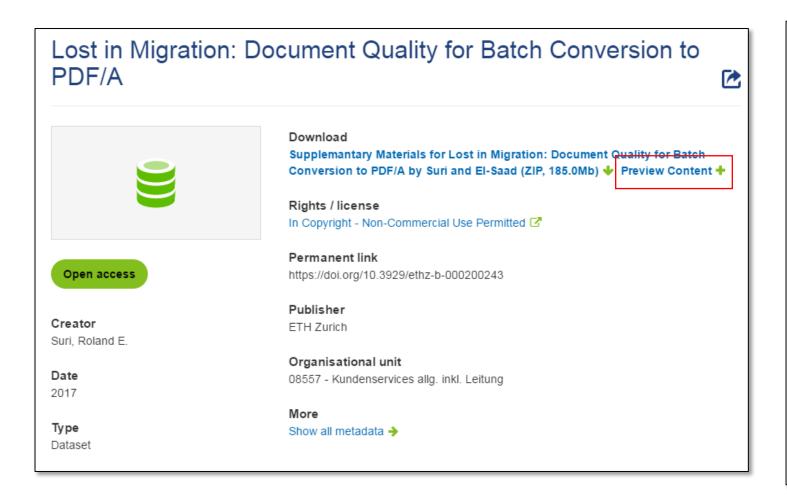


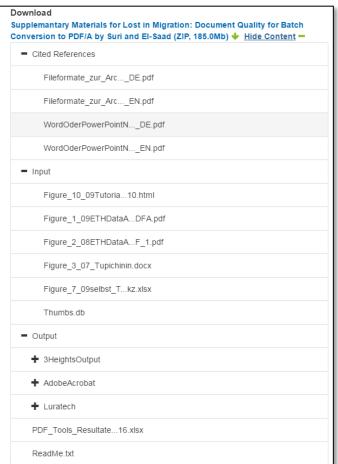
### Step 5: Cite your dataset with a DOI





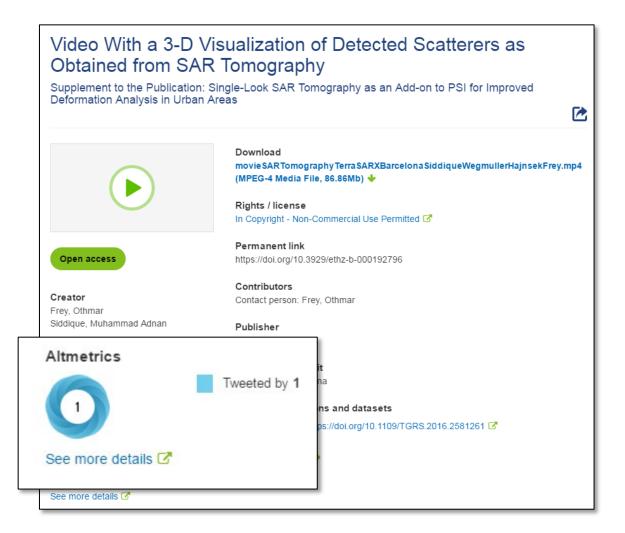
### Content preview for ZIP- and tar containers

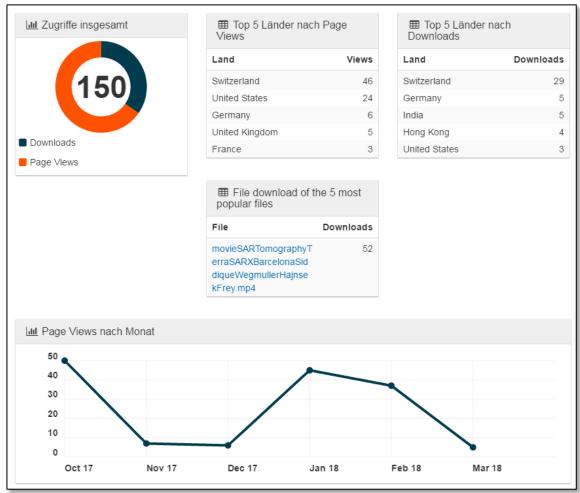






#### Altmetrics & Download statistics







### Dataset versioning

Option 1: Link new and previous version

#### Related publications and datasets

Is new version of: https://doi.org/10.3929/ethz-b-000130802 📝

Option 2 - Add an additional file

#### Download

Full text (updated version) (Adobe PDF, 1.580Mb) ❖
Full text (published version) (Adobe PDF, 1.710Mb) ❖



### Linking between publications and datasets





### What happens after the upload?

- While still in review: Metadata visible but files not accessible
- Within few days: review by our team
  - Checking and supplementing the entries metadata
  - Identification of formats → Informing submitters of issues with long term preservation (depending on retention period)
  - Superficial check for obvious legal issues (copyrighted material, software licenses)
    - > Submitters are responsible for the contents of their upload!
  - No content review
- Informing submitter of finished review
- DOI registered overnight after item finalization



#### Hands-on V

#### How can a data publication be improved?

- Look at an example dataset in the Research Collection
- Discuss the data publication in your breakout group
- Collect the findings by posting them in your room's section on the Miro board



#### Best practice examples?

- https://doi.org/10.3929/ethz-b-000449966
  - + Contains README
  - + Recommended licence: CC0
  - + Software separately licenced and declared
  - + Linked to publication
  - + Well supported formats (except binary RDS-file, which can be recreated)
- https://doi.org/10.3929/ethz-b-000455732
  - + Well supported formats (CSV, PDF)
  - + Good licence: CC-BY
  - + Linked to publication
  - + Well documented: Readme + description of variables



#### Take away: Recommendations for publishing research data

- Deposit in trustworthy repositories (→ check <u>www.re3data.org</u>)
  - Provides landing pages for individual datasets
  - Provides human and machine-readable metadata for datasets
  - Provides **DOIs** or other persistent identifiers for data
  - Provides the option to use open content licences
  - Has sustainable funding sources
- Before submitting your dataset...
  - Think about which data should be published
  - Check for legal and ethical compliance
  - Edit your data: Select well supported formats, remove unnecessary files
- Provide enough descriptive metadata so others can find and make sense of your data



#### And then...



- Make your research data a first class citizen in your scholarly record!
  - Cite and link your data publications in your formal reference list
  - Mention your data publications on your CV
  - Share your data publications on social media and track their attention (e.g. via <u>Altmetric</u>)



Barbara Hirschmann @library.ethz.ch

ETH Library Rämistrasse 101 8092 Zurich

www.library.ethz.ch



Feedback survey:

https://www.umfrageonline.ch/s/a13b937