ScaDS.AI DRESDEN LEIPZIG

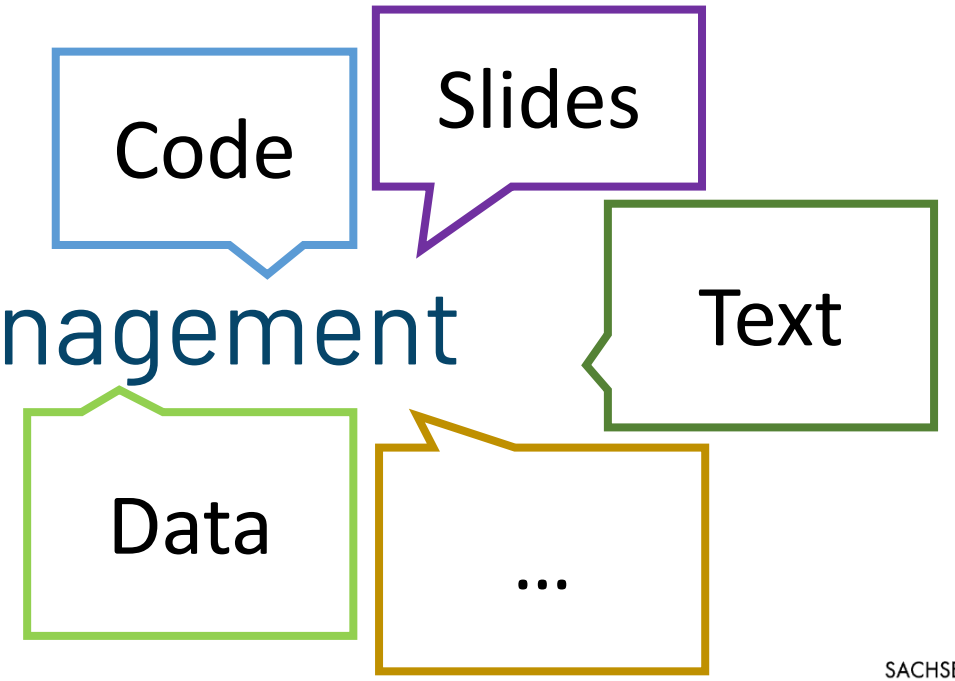
CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

NFDI 4 BIOIMAGE

NATIONAL RESEARCH DATA MANAGEMENT INFRASTRUCTURE FOR MICROSCOPY AND BIOIMAGE ANALYSIS

GloBIAS Global BioImage Analysts' Society

GLOBAL BIOIMAGE ANALYST'S SOCIETY



Research Data Management

Robert Haase

GEFÖRDERT VOM

Bundesministerium für Forschung, Technologie und Raumfahrt

SACHSEN

Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Recap quiz

- Which of the programs outputs „7“?

```
a = 14 / 2
b = 3
c = a - b

print(c)
```

```
n = 2
m = pow(n, 3)
p = m - n

print(p)
```

```
arr = []
for i in range(7):
    arr.append(i)

print(len(arr))
```

```
a = 3
b = 4

print(a + b == 7)
```

Recap quiz

- What's the output of this python code?

```
data = ('A', 'B', 'C', 'D', 'E', 'F', 'G')
print(data[1:3])
```

('A', 'B', 'C') ('B', 'C', 'D') ('A', 'B') ('B', 'C')




Recap quiz

- Which of the following does not raise an error?

```
data = ('A', 'B', 'C', 'D')
print(data[4])
```

```
a = 5
if a > 4
    print("larger than 5")
else
    print("smaller than 5")
```

```
for i in range(0, 10):
    print[i]
```

```
def sum_numbers(a, b):
    result = a + b
    return result

sum_numbers(5,6)
```

ScaDS.AI DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

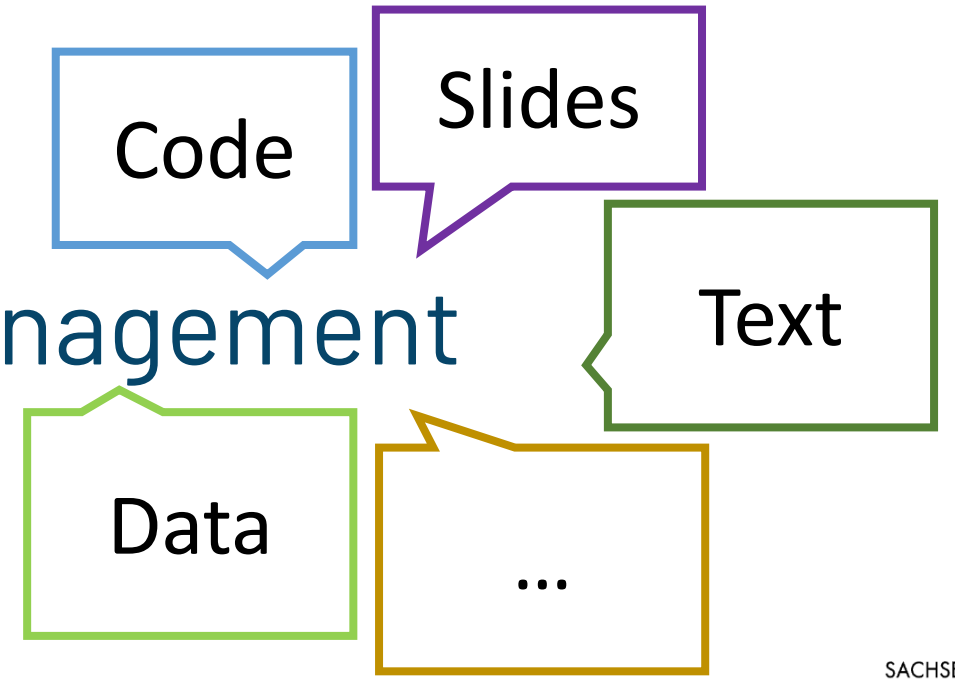
NFDI 4 BIOIMAGE

NATIONAL RESEARCH DATA MANAGEMENT INFRASTRUCTURE FOR MICROSCOPY AND BIOIMAGE ANALYSIS

GloBIAS Global BioImage Analysts' Society

GLOBAL BIOIMAGE ANALYST'S SOCIETY

Research Data Management



Robert Haase

GEFÖRDERT VOM

Bundesministerium für Forschung, Technologie und Raumfahrt

SACHSEN

Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Quiz

- When you published data with *your last publication*, which platform did you use?

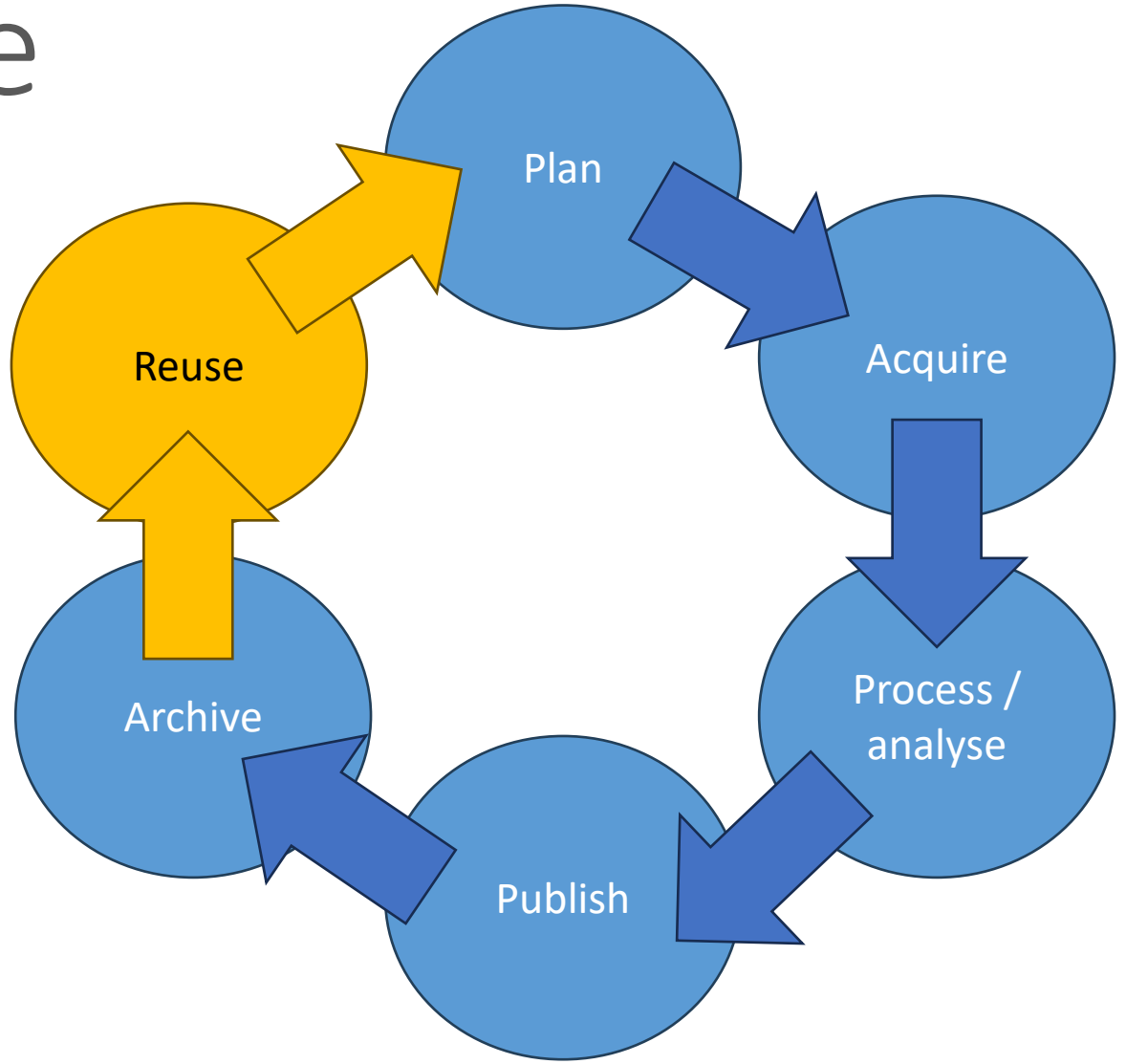
Onedrive/Google cloud/Dropbox/etc.	Zenodo/Figshare/ arxiv/F1000/github	Opara, Institute Website	Other
			

Research Data Management (RDM)

- All activities, processes, terms, persons which have relationships with data
  - Processing
  - Storage
  - Organization
  - Publication
  - …
- In routine: working with data

RDM Life Cycle

- Processes are ideally cyclic
- Closing the loop is a major the challenge



Adaptiert von:
https://forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/

Data Management Plans (DMPs)

- Define responsibilities and procedures early!

"Data / materials we produce will be published under CC-BY 4.0"

"Robert will do this by end of 2025!"

DMP

- Only if procedures are defined early, everyone can follow them.
- Licenses are important when assembling materials (-> Copyright)
- Meta-data might have *higher quality* if the person responsible for publishing the data is aware of their duties.

Are we going to publish data / materials / code?

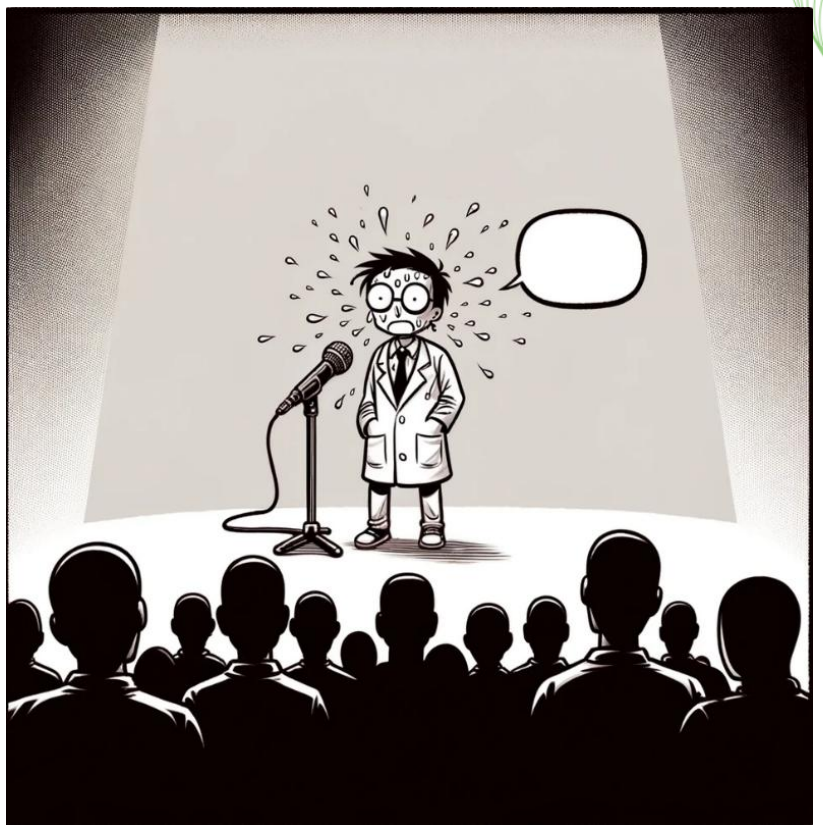
What license can we use?

Deciding by the end of the project is too late!

Closed science

Why are some science-related materials/data/code not shared?

- Risk of being scooped
- Fear of blaming oneself (imposter syndrome)
- Lack of awareness (who is allowed to publish *my work*?)
- Assumption: it’s not worth the effort.

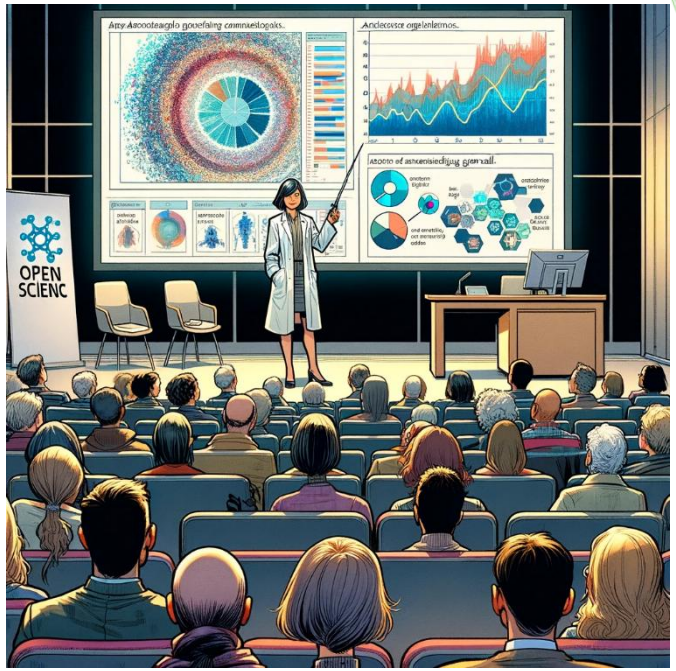


Open Science

- Research related (hot topics)
- Often tailored towards general audience (science communication)
- Earliest at the time a manuscript is published (e.g. as preprint)

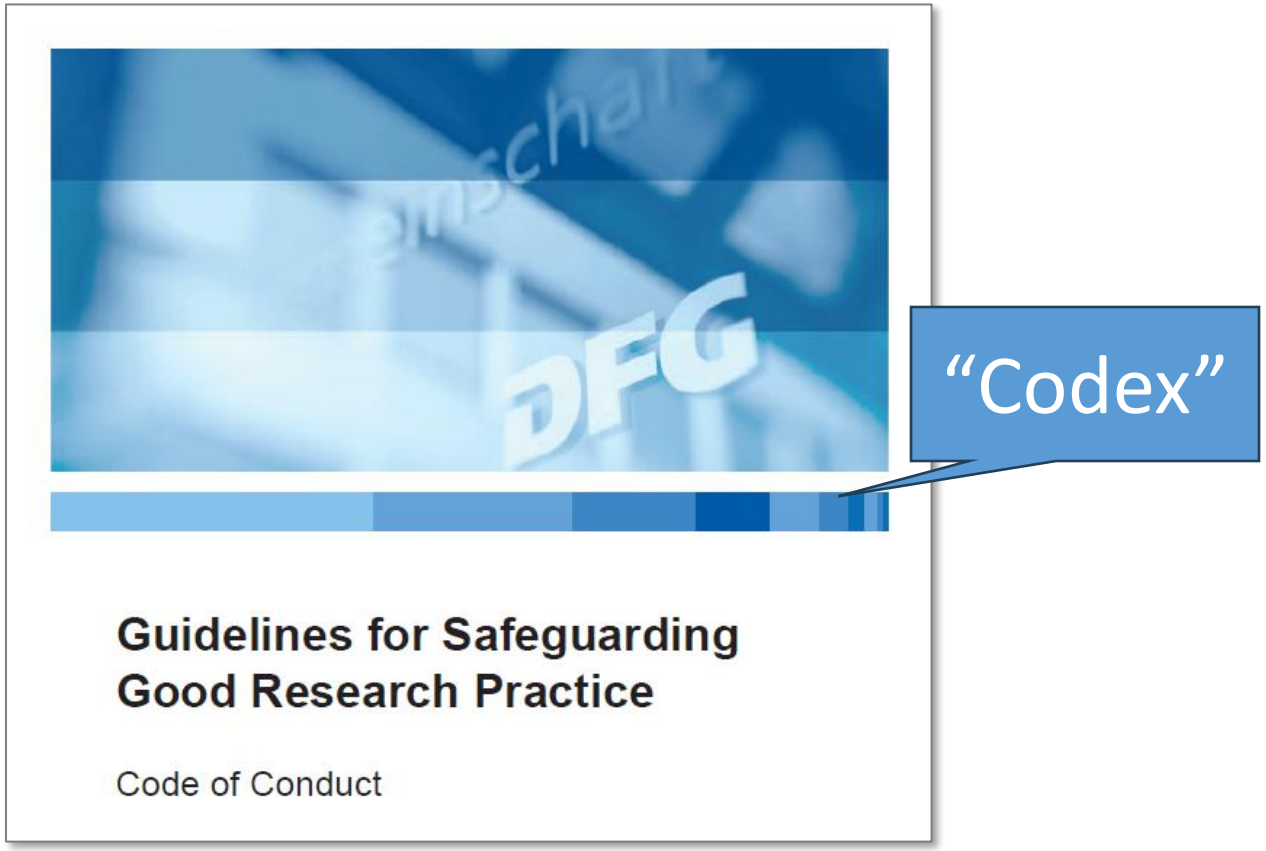
Open Training

- Routine tasks (colder topics)
- Transfer of domain-specific knowledge



Scientific culture

Public access to research results -> Reusability



"Codex"

Guideline 13: Providing public access to research results

- As a rule, researchers make all results available as part of scientific/academic discourse. In specific cases, however, there may be reasons not to make results publicly available (in the narrower sense of publication, but also in a broader sense through other communication channels); this decision must not depend on third parties. Researchers decide autonomously – with due regard for the conventions of the relevant subject area – whether, how and where to disseminate their results. If it has been decided to make results available in the public domain, researchers describe them clearly and in full. Where possible and reasonable, this includes 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 programmed by researchers themselves is made publicly available along with the source code. Researchers provide full and correct information about their own preliminary work and that of others.

Explanations:

In the interest of transparency and to enable research to be referred to and reused by others, whenever possible researchers make the research data and principal materials on which a publication is based available in recognised archives and repositories in accordance with the FAIR principles (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable). Restrictions may apply to public availability in the case of patent applications. If self-developed

https://www.dfg.de/resource/blob/174052/1a235cb138c77e353789263b8730b1df/kodex-gwp-en-data.pdf

Scientific culture

DFG Deutsche Forschungsgemeinschaft

About Us | Funding | Basics and Topics | Funded Projects | News

DFG > News > News and Current Topics > Information for Researchers > Package of Measures to Support a Shift in the Culture of Research Assessment

Information for Researchers, No. 61 | September 1, 2022

Package of Measures to Support a Shift in the Culture of Research Assessment

DFG changes proposal forms and introduces mandatory CV template / The aim is to support a shift in the culture of research assessment / Improvement of equal opportunity practices

DFG changes proposal forms and introduces mandatory CV template / The aim is to support a shift in the culture of research assessment / Improvement of equal opportunity practices

Successful science and research require suitable framework conditions. The Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) ensures these conditions by regularly conducting analyses, providing the relevant information and adapting its procedures accordingly. The DFG set out the challenges and fields of action in a position paper on academic publishing published in May of this year: it sees both the academic community as a whole and itself as a funding organisation as being responsible for initiating a cultural shift towards research assessment that is geared more towards equal opportunity and attaches even greater importance to the substance of research. In the interests of bringing about such a shift, it is up to research funding organisations to broaden the spectrum of accepted publication formats, to attach greater value to content-based evidence of achievement and to strengthen the recipient side of publishing. The DFG has launched a comprehensive and far-reaching package of measures in order to fulfil this mandate.

Binding CV template across all funding programmes

For this reason, the assessment of a researcher's accomplishments must be holistic and based on substantive qualitative criteria. In order to strengthen qualitative evaluation criteria over quantitative indicators, the DFG will be introducing a curriculum vitae template that will be mandatory across all programmes from 1 March 2023 (the template will be adapted shortly for proposals under the Collaborative Research Centre and Research Training Group programmes; information will be provided separately in this regard). The template adopted by the DFG Senate allows applicants to provide both narrative and tabular information, thereby facilitating a holistic view of the applicant's academic career in the review and evaluation process.

In addition to the mandatory information required in order to assess eligibility, applicants may also provide details of special circumstances or additional services to scholarship such as committee activities or the establishment of research infrastructures. As such, the template provides the basis for a qualitatively sound assessment of academic performance that takes greater account of the respective stage of the individual's life and career. Accordingly, reviewers are now instructed to consider applicants' academic performance in the context of their individual curriculum vitae and career stage.

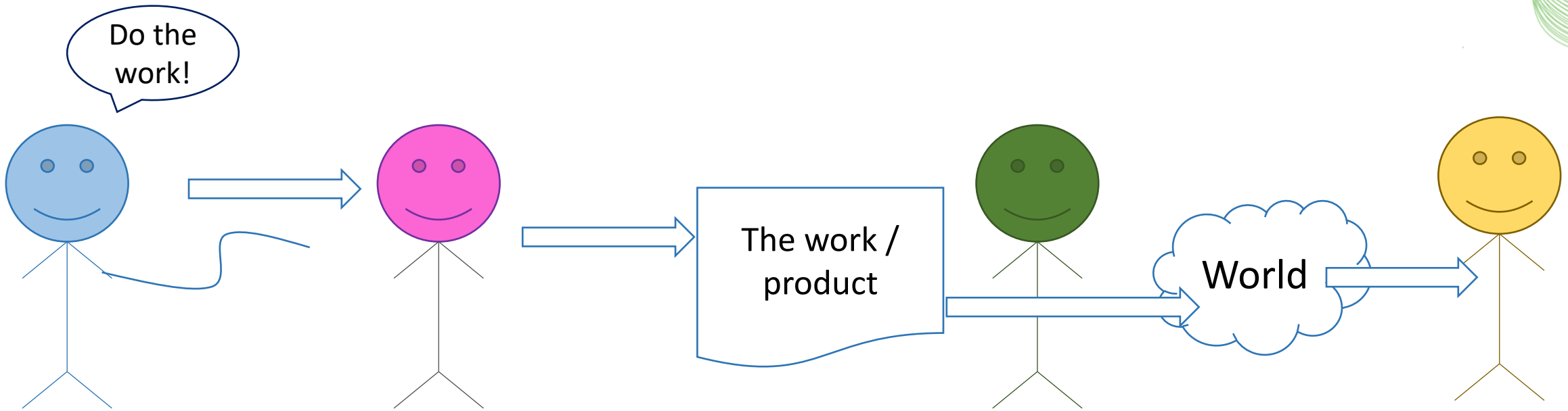
Publication details in proposals and CVs

Performance assessment based on content-related qualitative criteria also explicitly includes ensuring that the entire spectrum of academic publication types are equally displayed and acknowledged in funding proposals and CVs. In addition to a maximum of ten publications in the more common publication formats, the CV can therefore now list up to ten further sets of research outcomes and findings that have been publicised in a variety of other ways, including articles on preprint servers, data sets or software packages, for example. In DFG proposals, the project-specific list of publications will be included in the general bibliography. The intention here is to shift the focus of the review and the evaluation of a proposal away from the list of publications and towards the substance of the applicant's accomplishments. In order to document their own published preliminary work, applicants can typographically highlight (e.g. in bold) a maximum of ten of their own publications in the bibliography that are important for the project. No information on quantitative metrics such as impact factors and h-indices is required in the CV or the proposal, and such information is not to be considered in the review. The relevant details are included in DFG forms and review instructions.

These modifications and innovations reflect the fact that the DFG is continuing to promote the cultural shift in research assessment that was advocated in May with the publication of the position paper on academic publishing. The DFG hopes that this refocus – away from quantitative indicators and towards the substance of scholarship – will lead to improved equality of opportunity and a higher-quality basis for review overall.

Am I allowed to publish my stuff?

- … it depends… on who is responsible



Copyright holder, e.g. employer / PI of the author

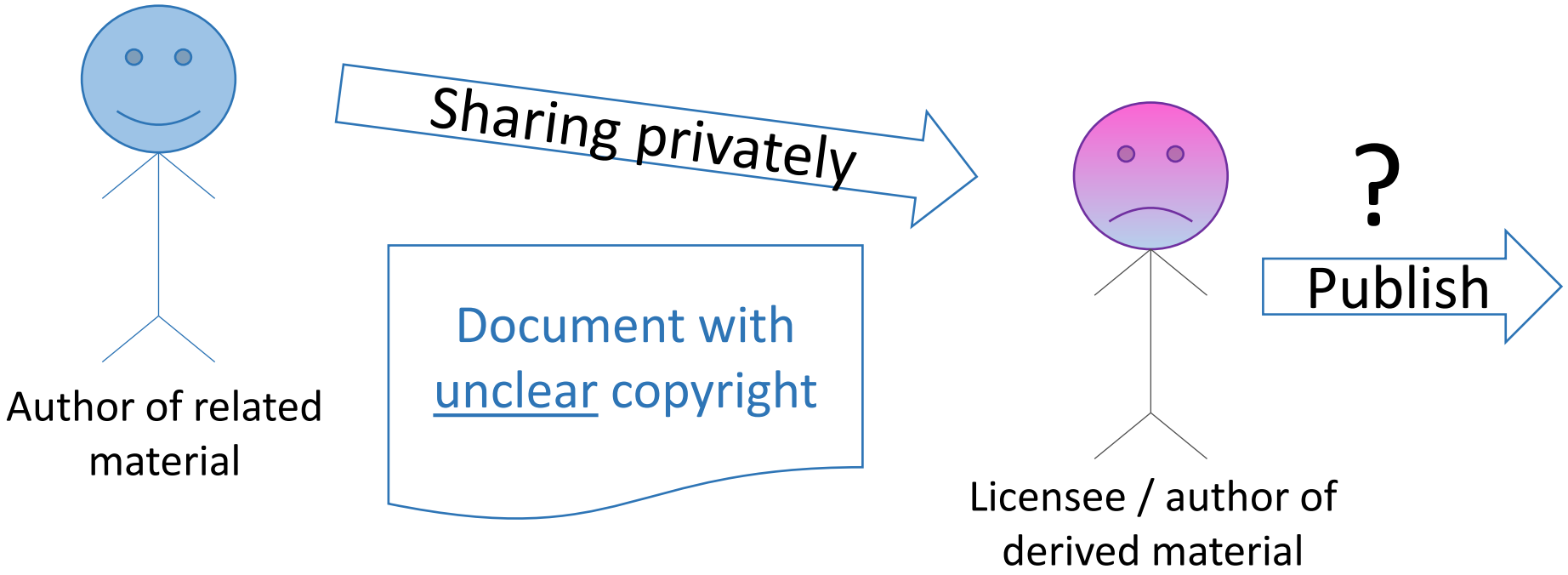
Author
Creator of the work

Publisher
The person who made the work available

Licensee
The person who uses / reuses the material

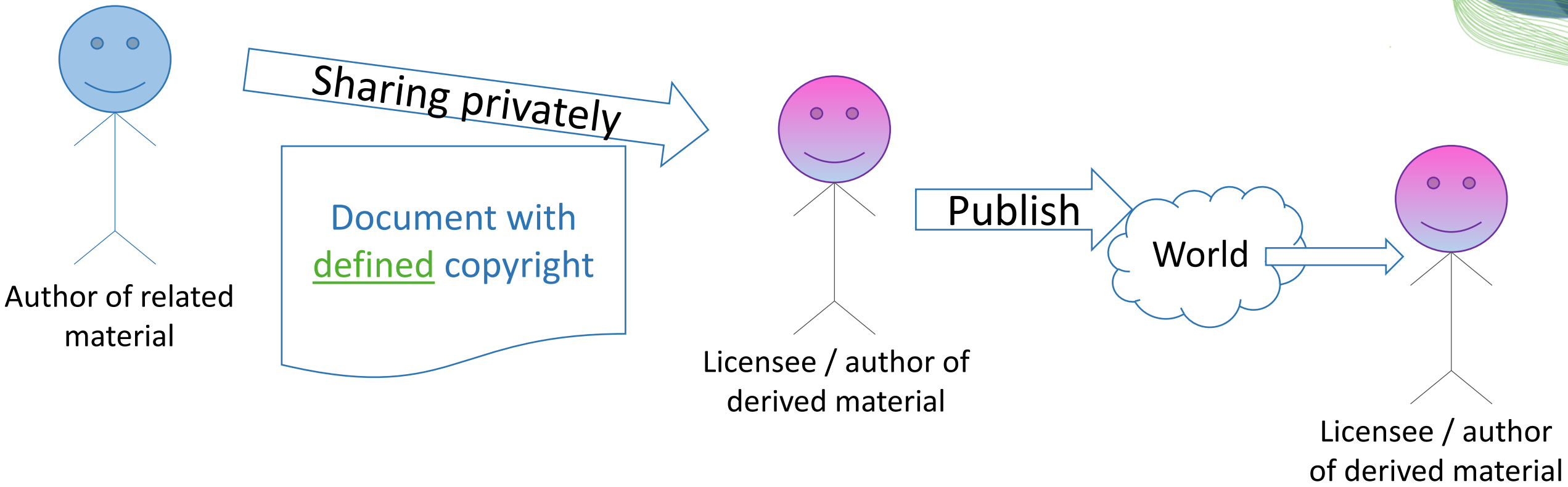
Am I allowed to publish my stuff?

- … it depends… on what materials served as basis



Am I allowed to publish my stuff?

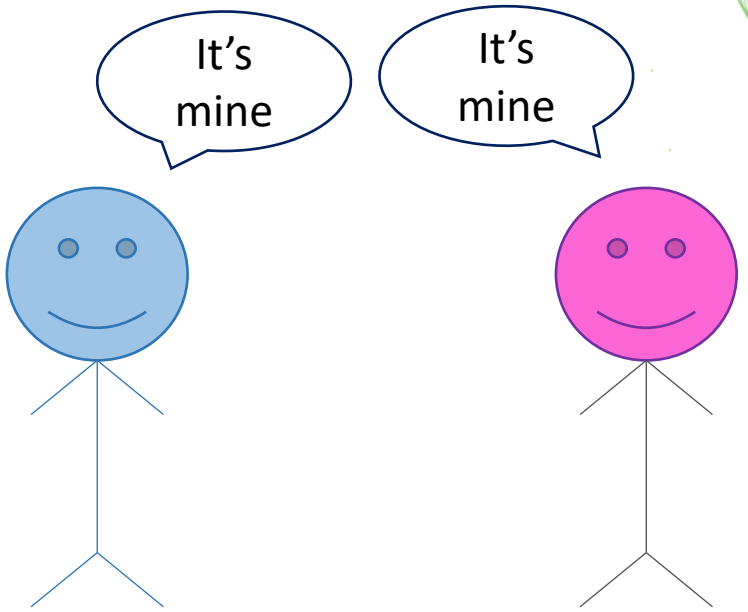
- … it depends… on what materials served as basis



Public versus institutional repositories

Conflict of interest between employers and scientists

- Universities seek to keep things secret and potentially exploit them commercially (licenses, startups, …)
- Scientists need to publish to advance their career.
- Hints:
  - Decide early during the project what will be published and by who (-> DMP)
  - Check your job description! (Is “Programming” or “model training” part of it?)



Copyright holder, e.g. employer / PI of the author

Author
Creator of the work

Quiz

- When you publish data on Zenodo.org, which role has Zenodo?

Copyright holder

Author

Publisher

Licensee

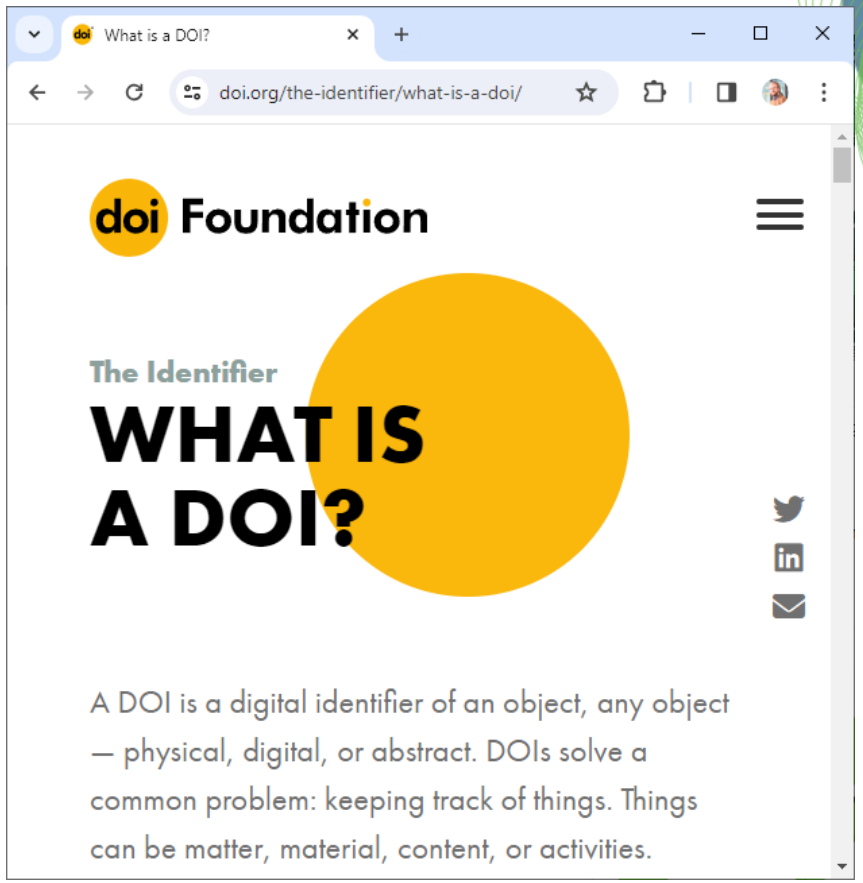
Standard for sharing: The FAIR-principles

- Findable
- Accessible
- Interoperable
- Reusable

The FAIR-principles

Findable

- F1. (Meta)data are assigned a globally unique and persistent identifier
  - Universal Resource Identifier (URI)
  - Digital Object Identifier (DOI)
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource



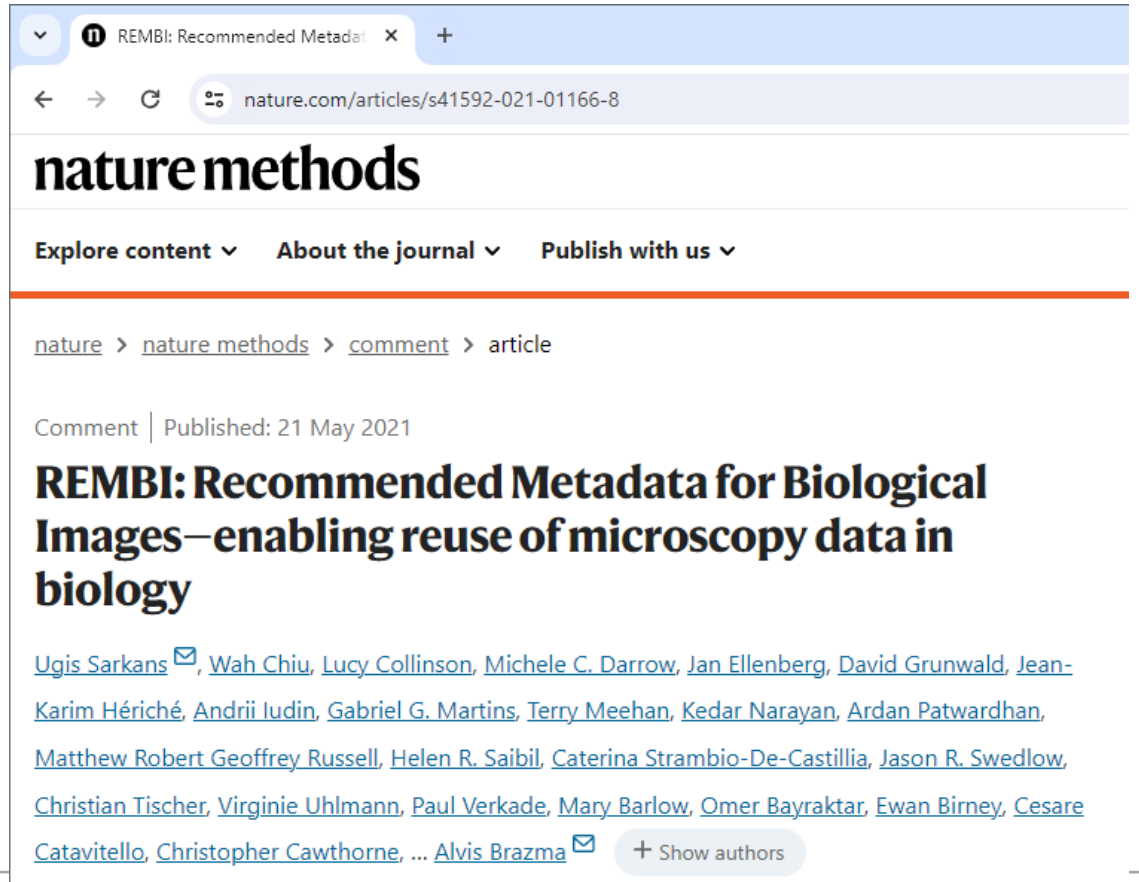
Screenshot source: https://www.doi.org/the-identifier/what-is-a-doi/ licensed CC-BY 4.0
https://www.go-fair.org/fair-principles/
https://biii.eu/introduction-3d-analysis-3d-imagej-suite

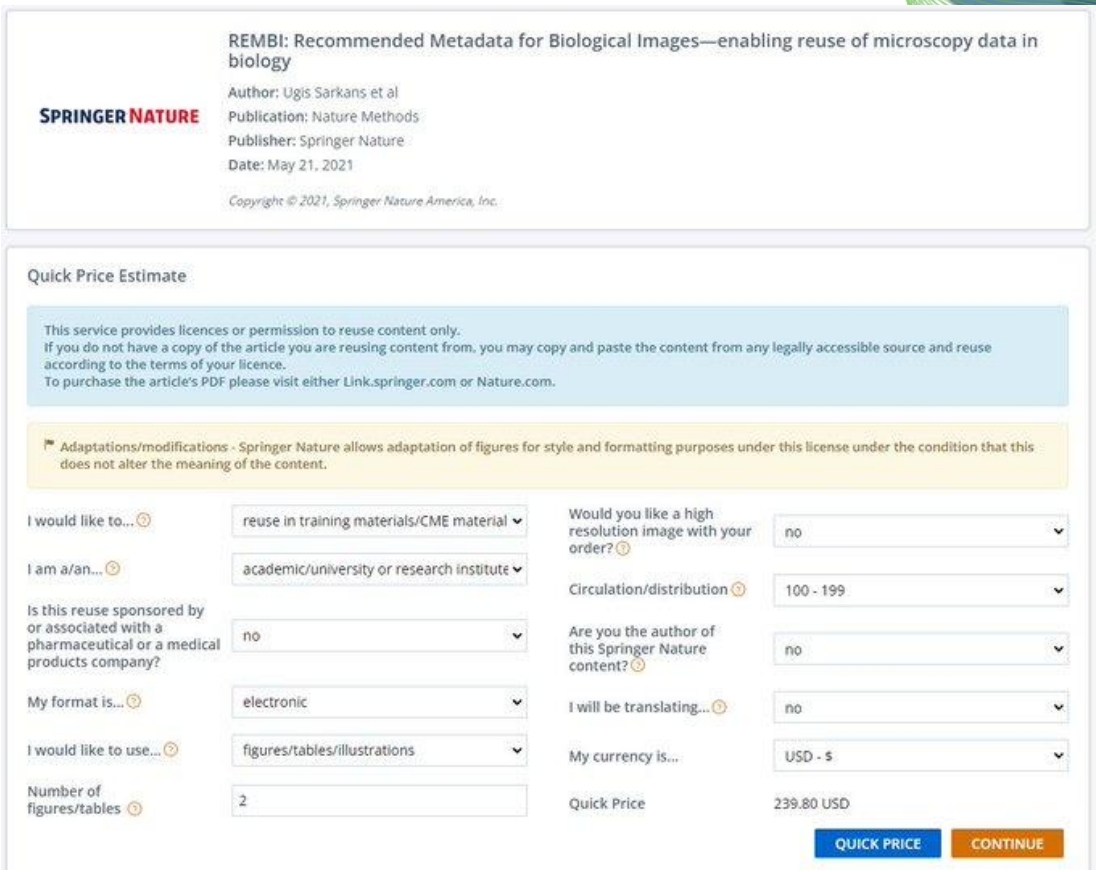
Meta data

- Generic
  - Author
  - Usage license
  - Creation date
- Field-specific (microscopy)
  - Exposure time
  - Wavelength (colour)
  - Microscope type/vendor
- Field-specific (software)
  - Dependencies
  - Requirements
  - Purpose of the code
  - User documentation

REMBI: Recommended Metadata for Biological Images—enabling reuse of microscopy data in biology

- Read more:





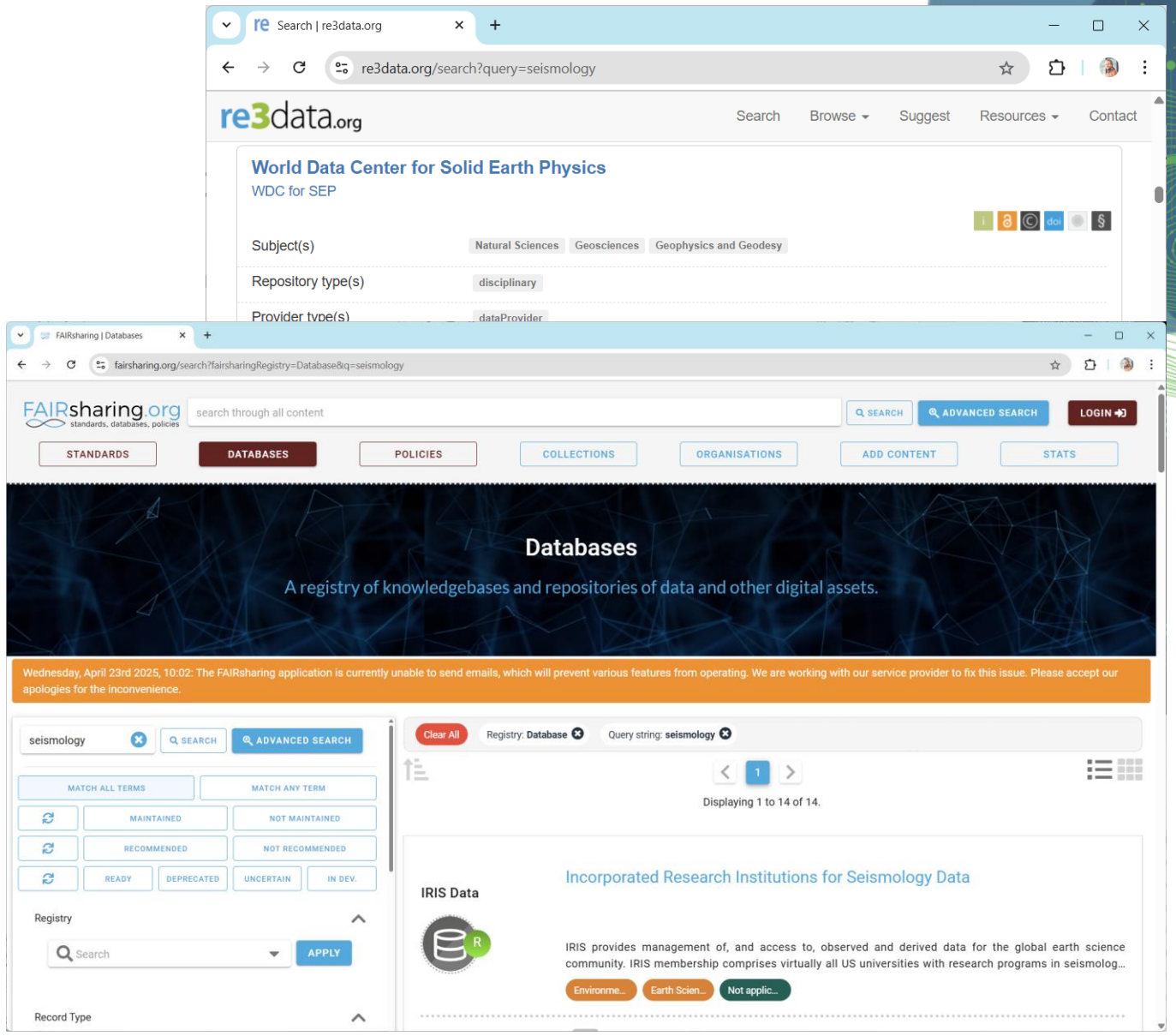
Findability

Domain-specific

- Search repository registries for your field!

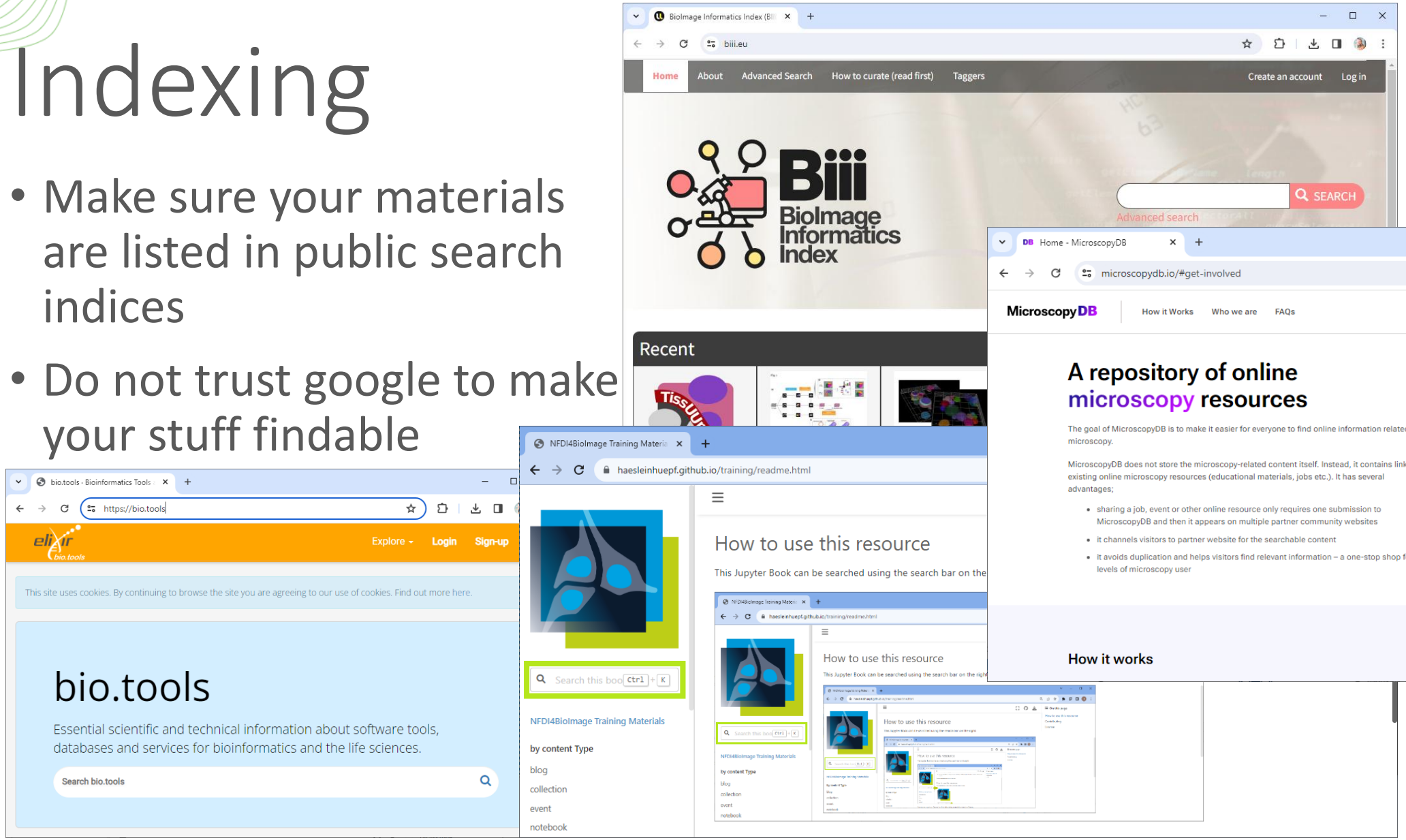
Guidelines

- Publish where your community publishes
- Publish where everyone publishes (beyond your community)
- Publish in your local institute's infrastructure



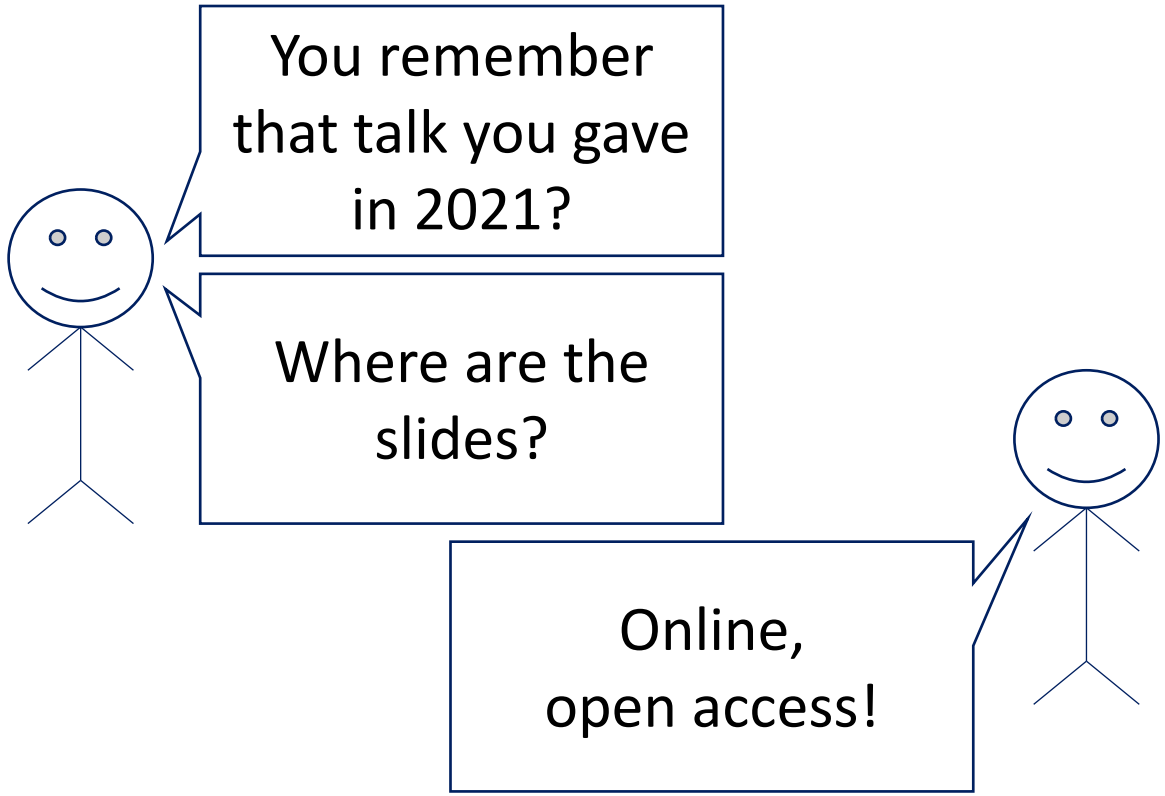
Indexing

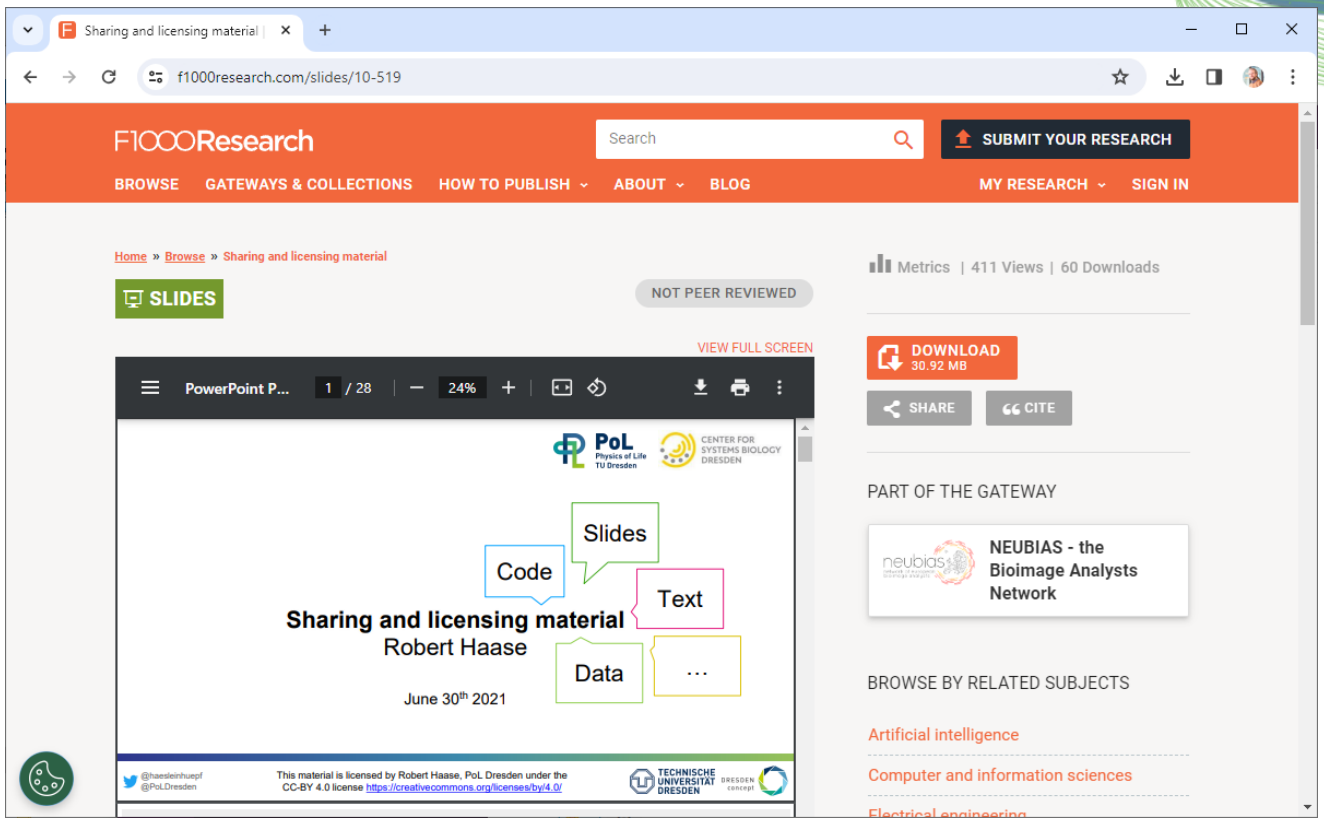
- Make sure your materials are listed in public search indices
- Do not trust google to make your stuff findable



Incentives: Findability

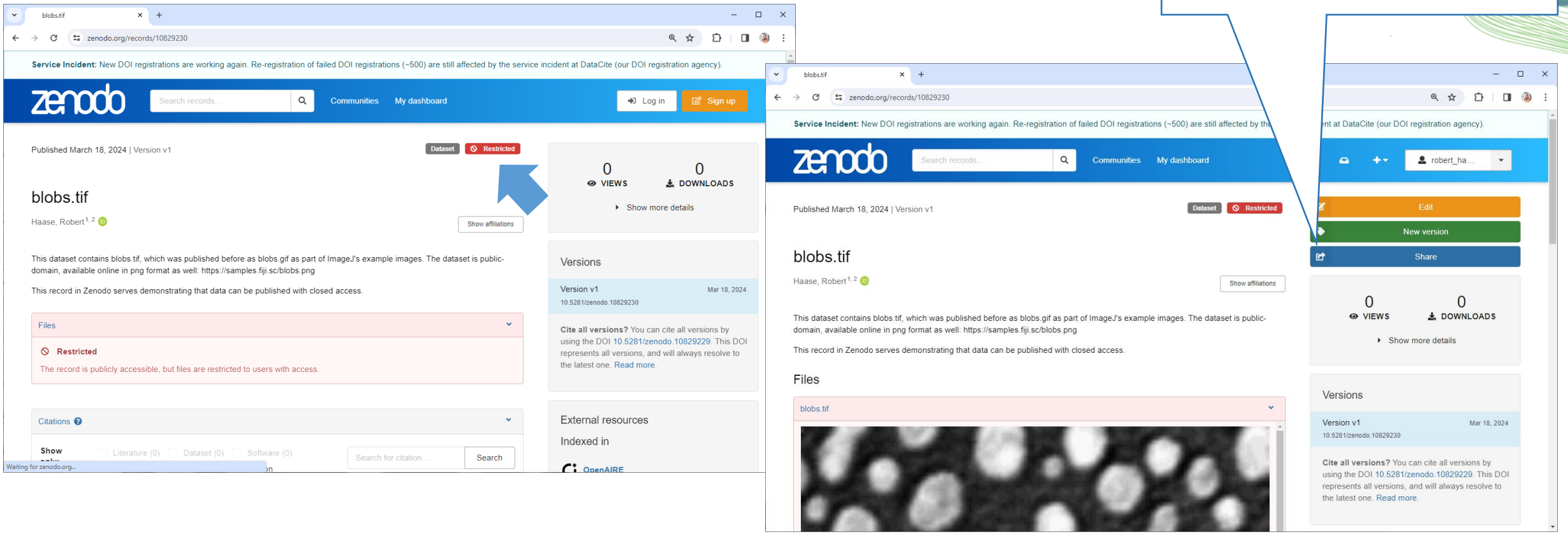
- Your *future-self* will thank you, because they will find your work





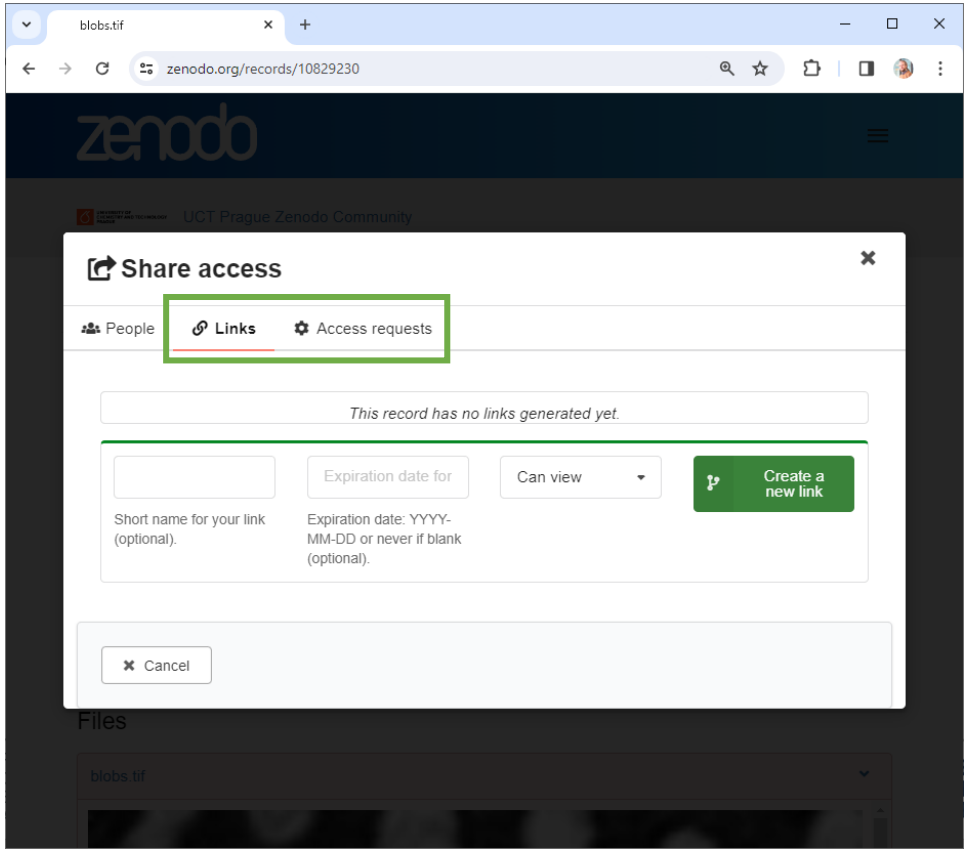
Accessibility

- The A in FAIR does not necessarily stand for Open Access



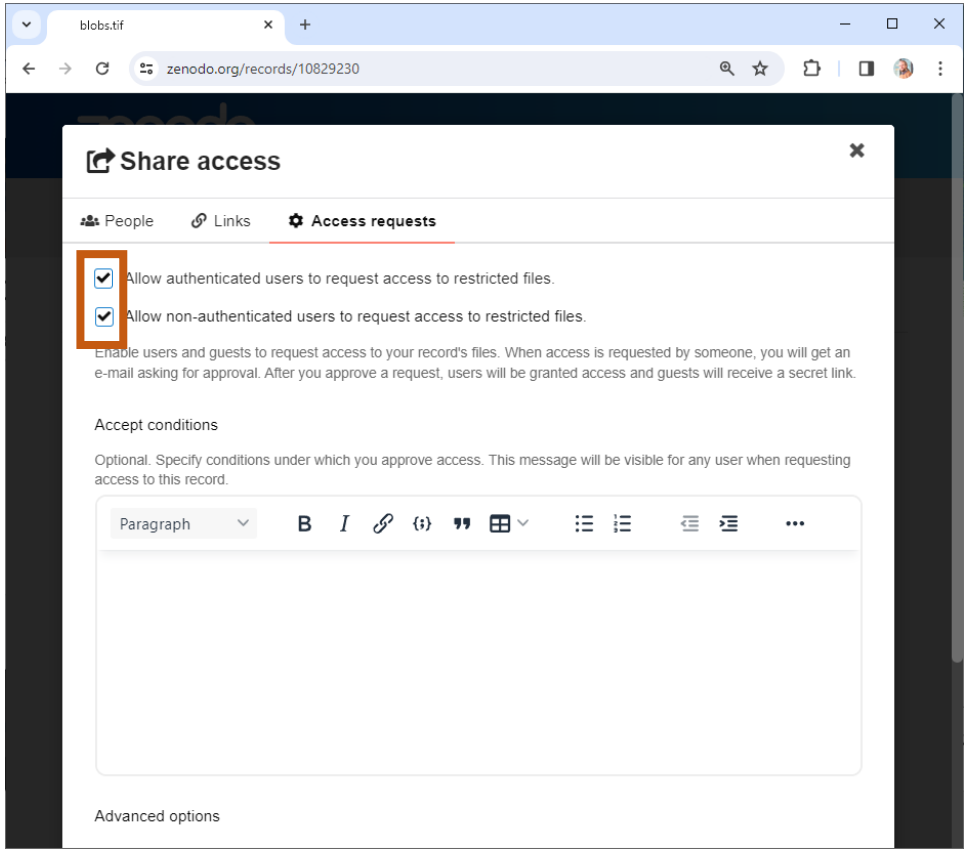
Accessibility

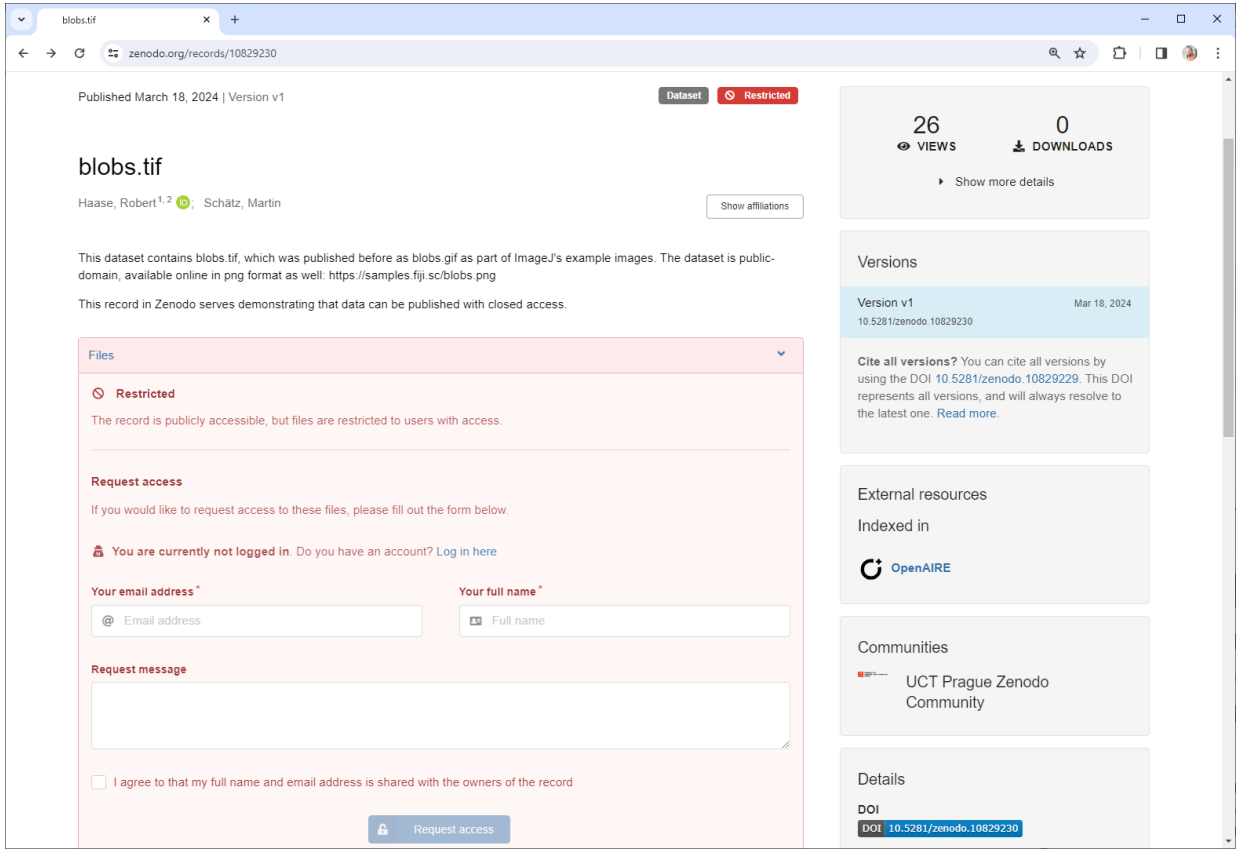
- The A in FAIR does not necessarily stand for Open Access



Accessibility

- The A in FAIR does not necessarily stand for Open Access





Interoperability

- 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

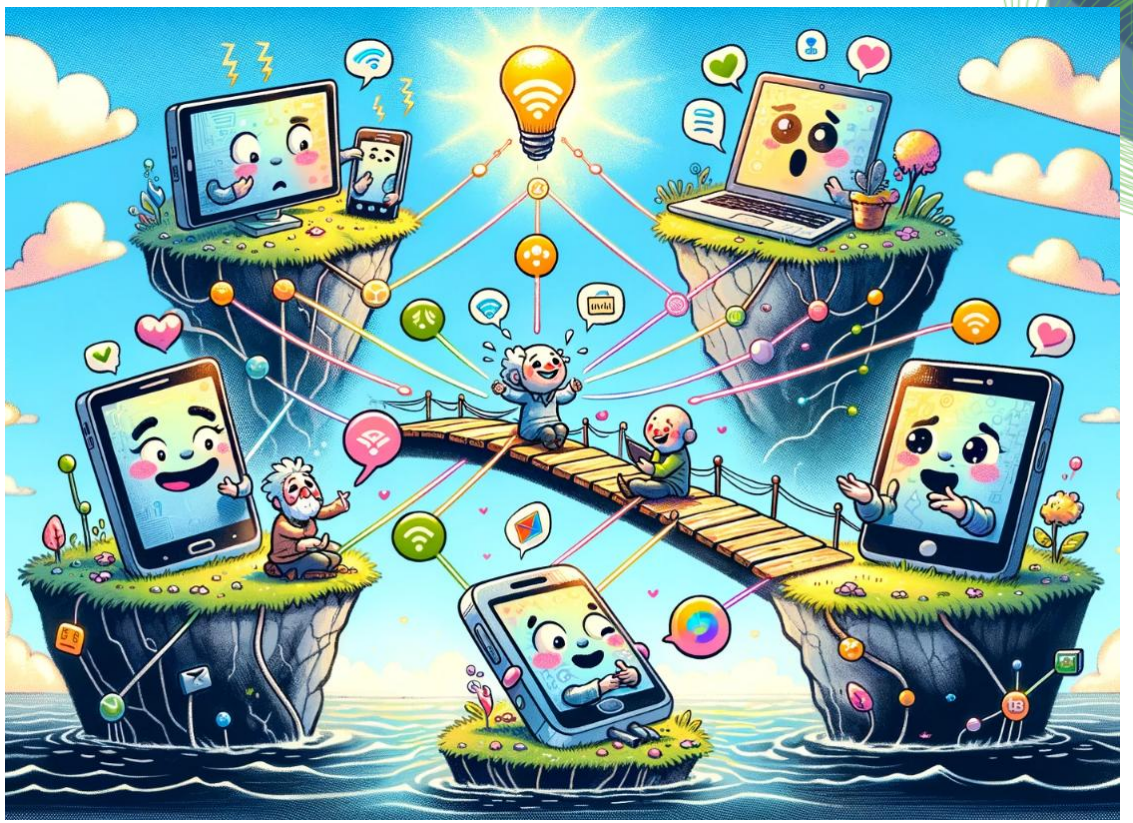
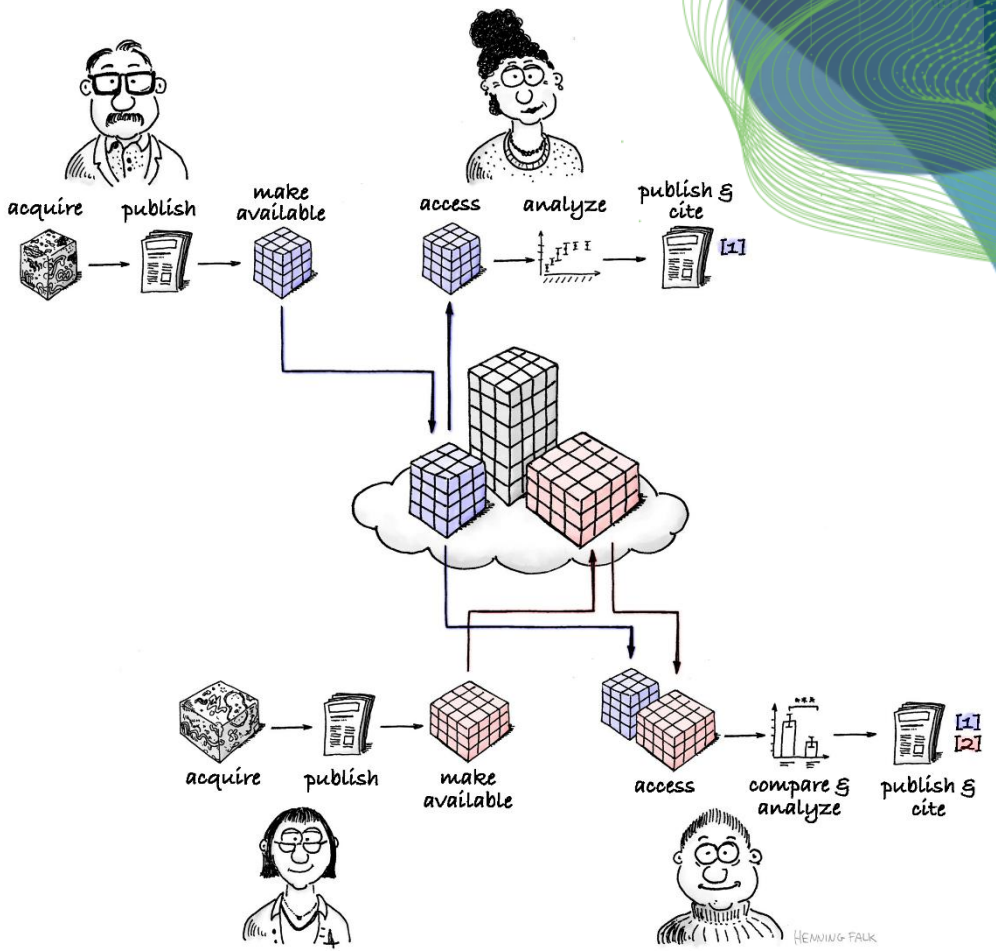


Image was created using Generative AI (Dall-E)
https://www.go-fair.org/fair-principles

Reusability

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

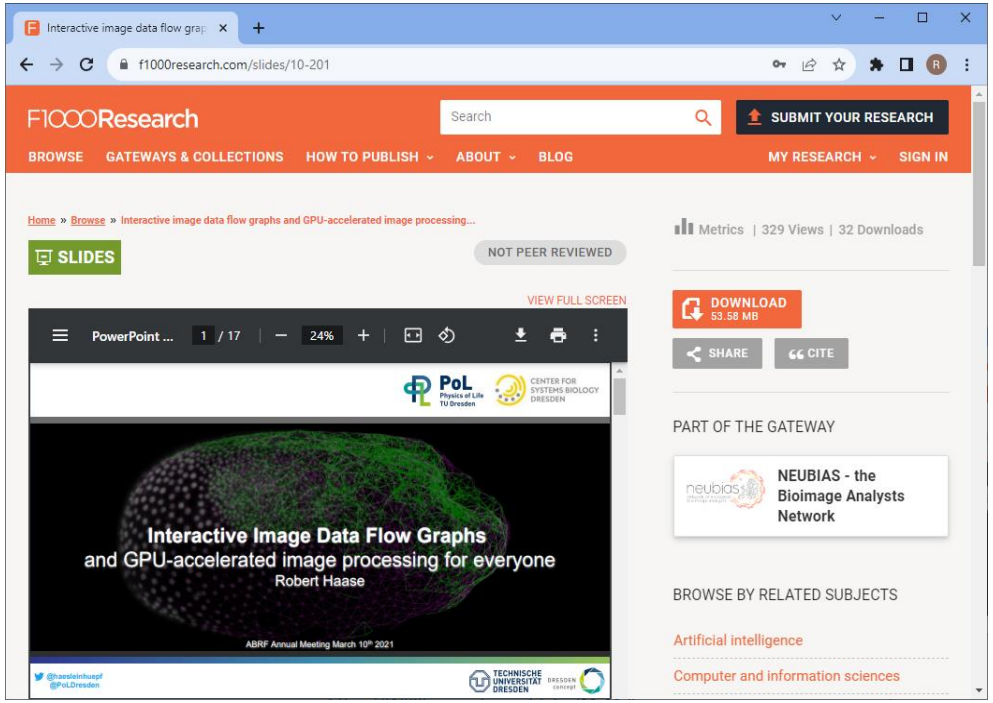


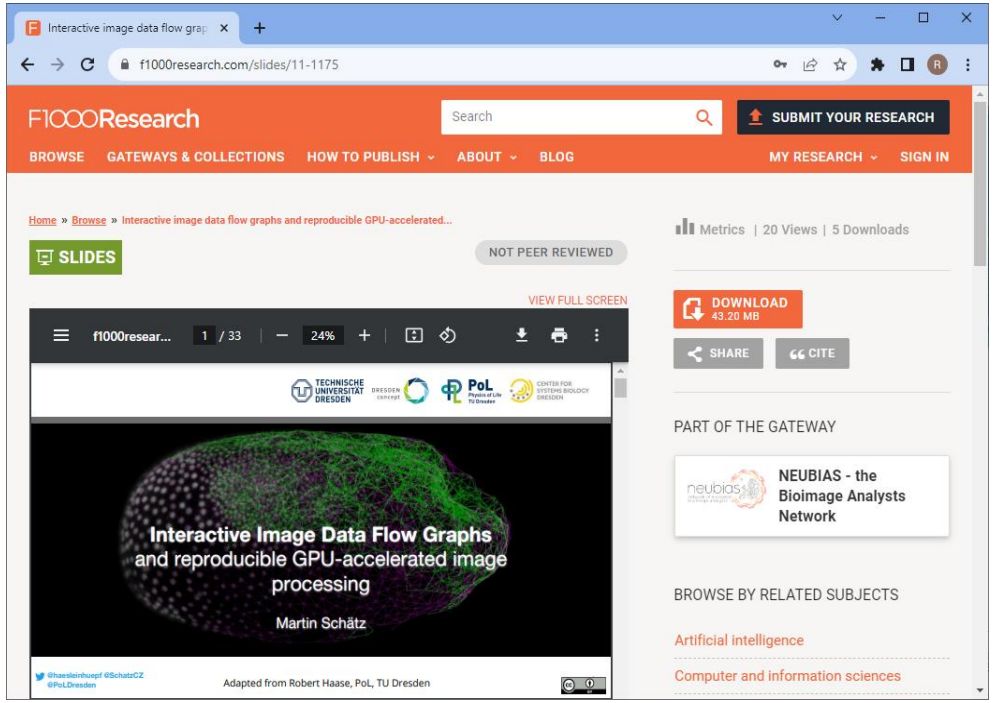
https://www.go-fair.org/fair-principles/

Drawing made by Henning Falk, ©2022 NumFOCUS, reused under a CC BY 4.0 license

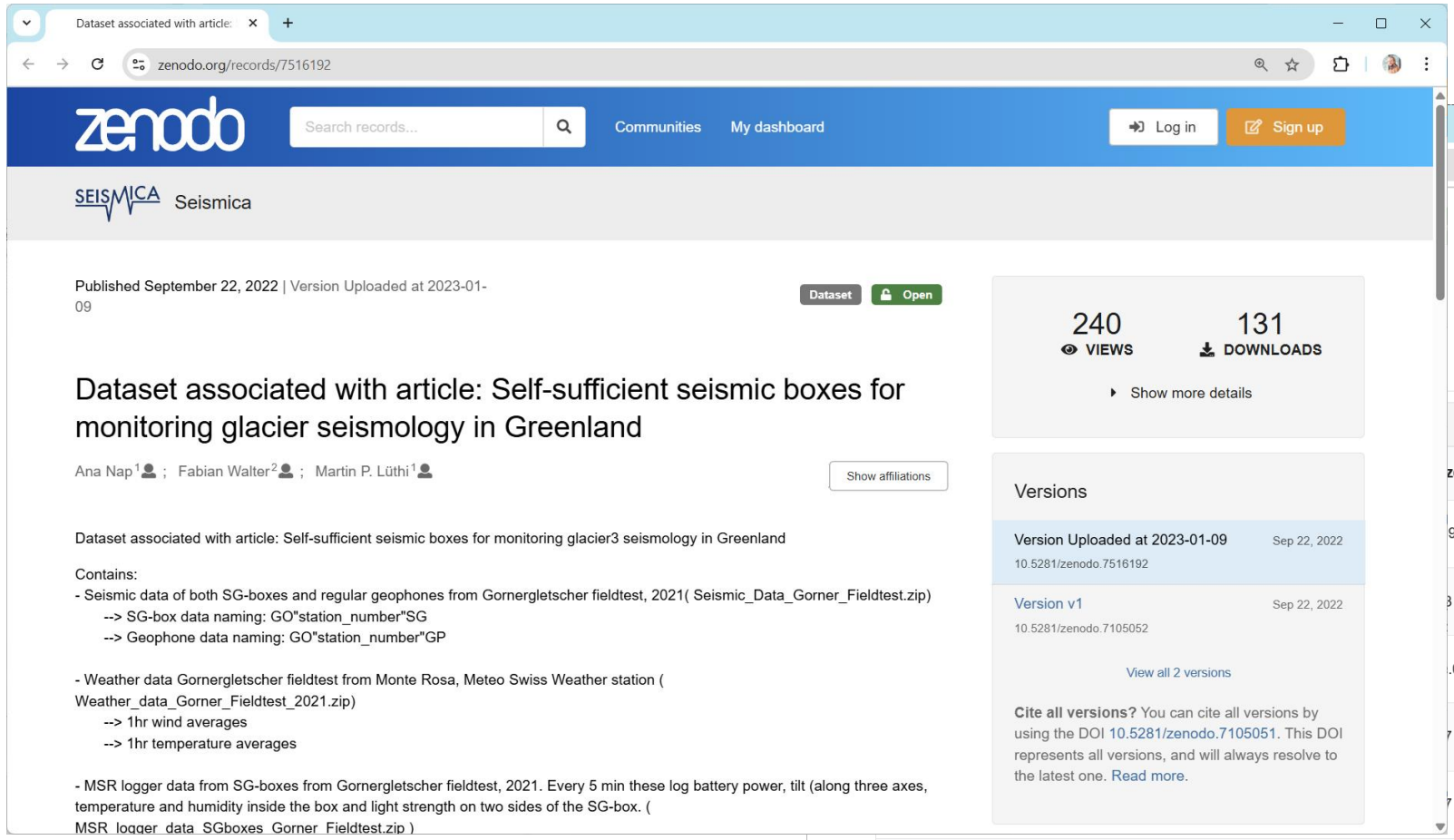
Incentives: Reusability

- Open Educational Resources
  -> Others teach how to use your *tools & methods*



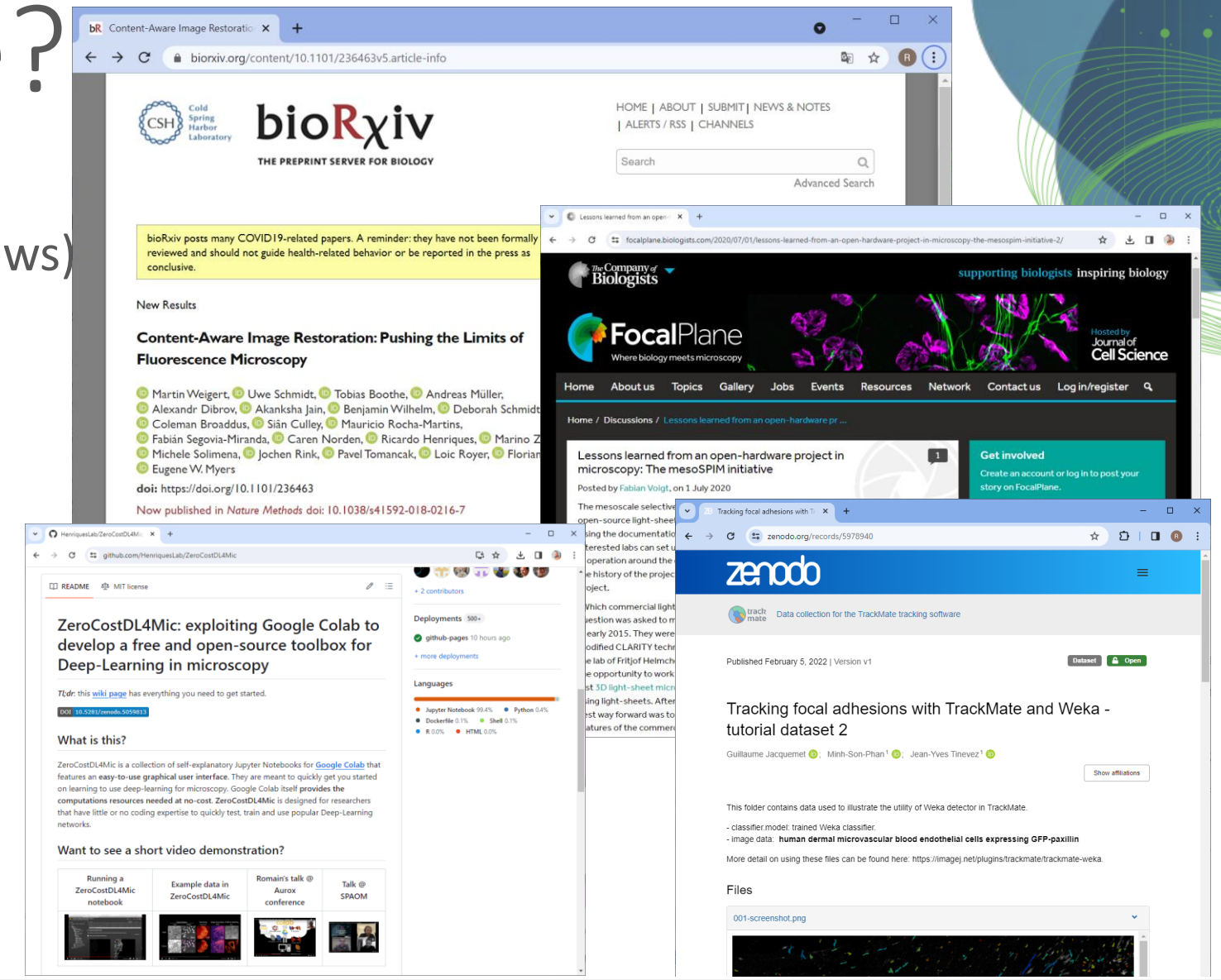


Incentives: Citability



Where to share?

- Open *science* related content
  - bioRxiv (manuscripts, no reviews)
  - Figshare
  - F1000
  - Bioimage Archive (data)
  - Github (code)
  - Zenodo
  - Focalplane
  - Institutional servers
    (if there is no alternative)



https://www.biorxiv.org/content/10.1101/236463v5.article-info
https://github.com/HenriquesLab/ZeroCostDL4Mic https://focalplane.biologists.com/2020/07/01/lessons-learned-from-an-open-hardware-project-in-microscopy-the-mesospim-initiative-2/
https://zenodo.org/records/5978940

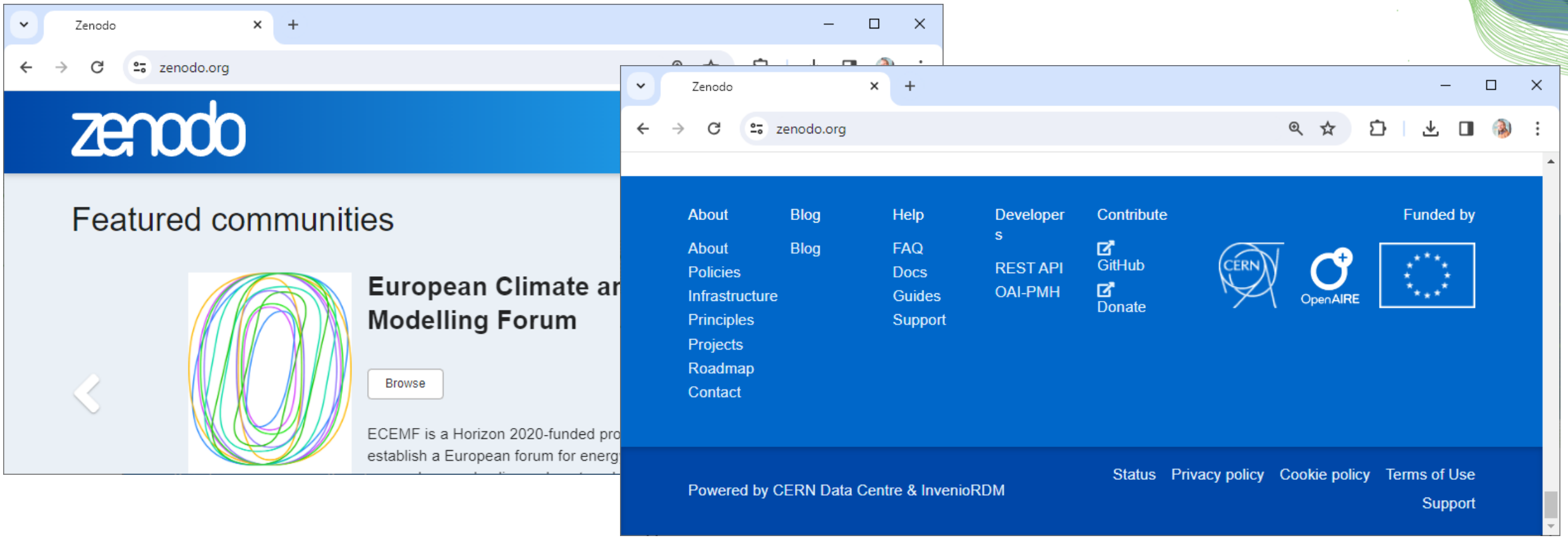
TECHNISCHE UNIVERSITÄT DRESDEN
UNIVERSITÄT LEIPZIG

Quiz

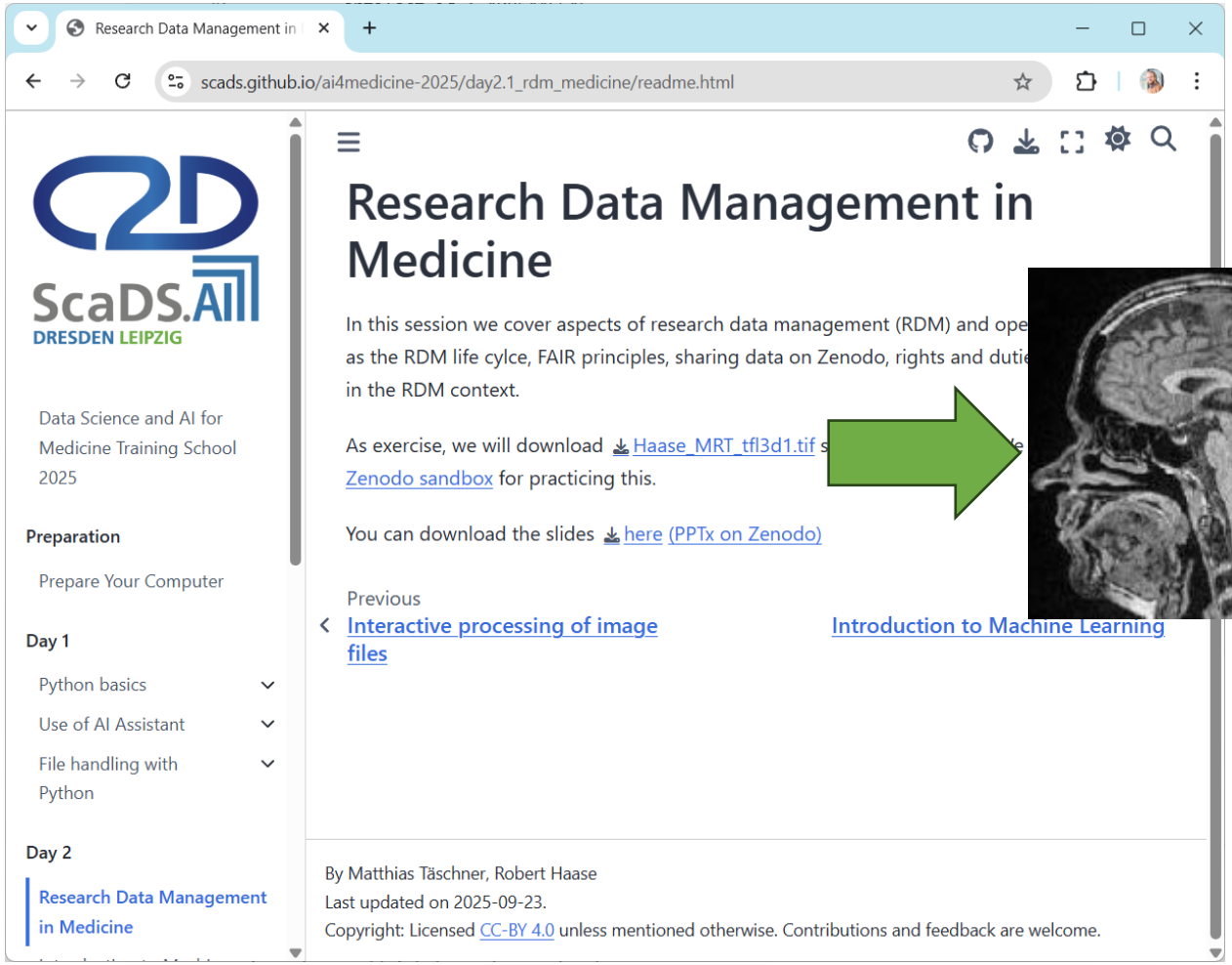
- I'm sharing data in a public repository on Github.com in a file format we developed using an open source license.
- Which FAIR principle am I violating?

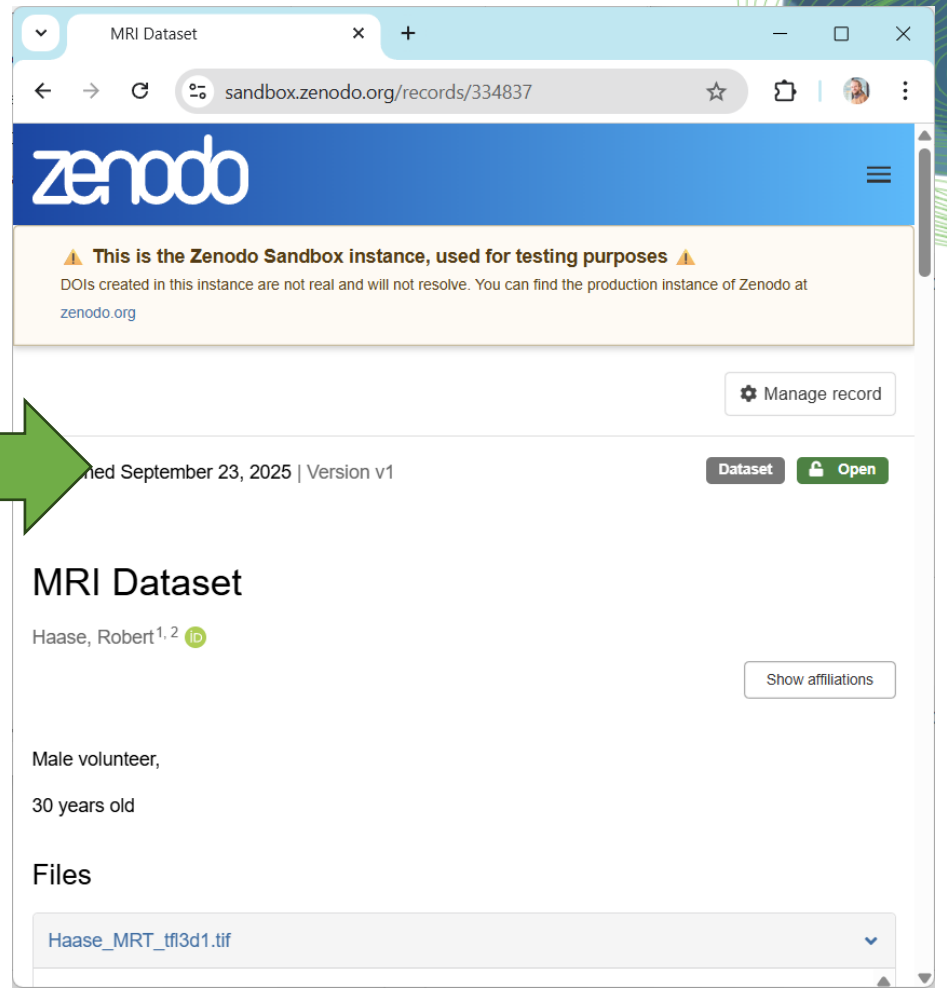
Zenodo

- Publicly funded infrastructure @ CERN / Switzerland



Exercise: Sharing files on Zenodo



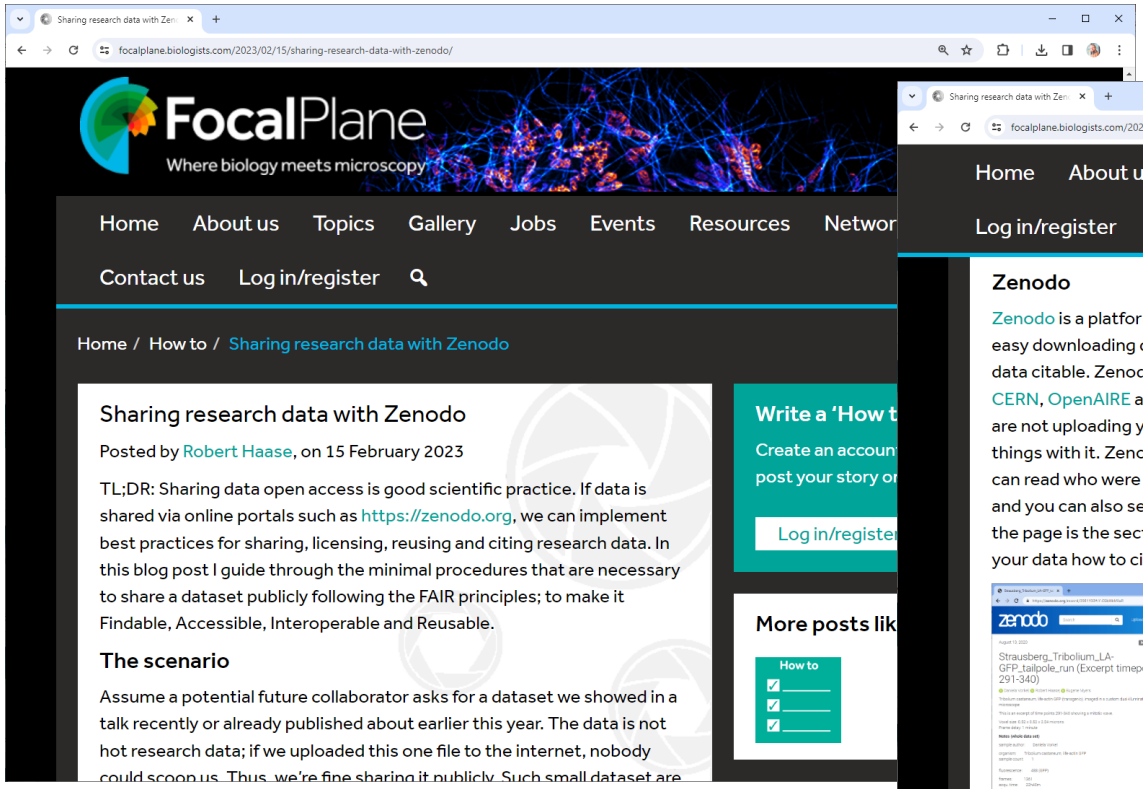


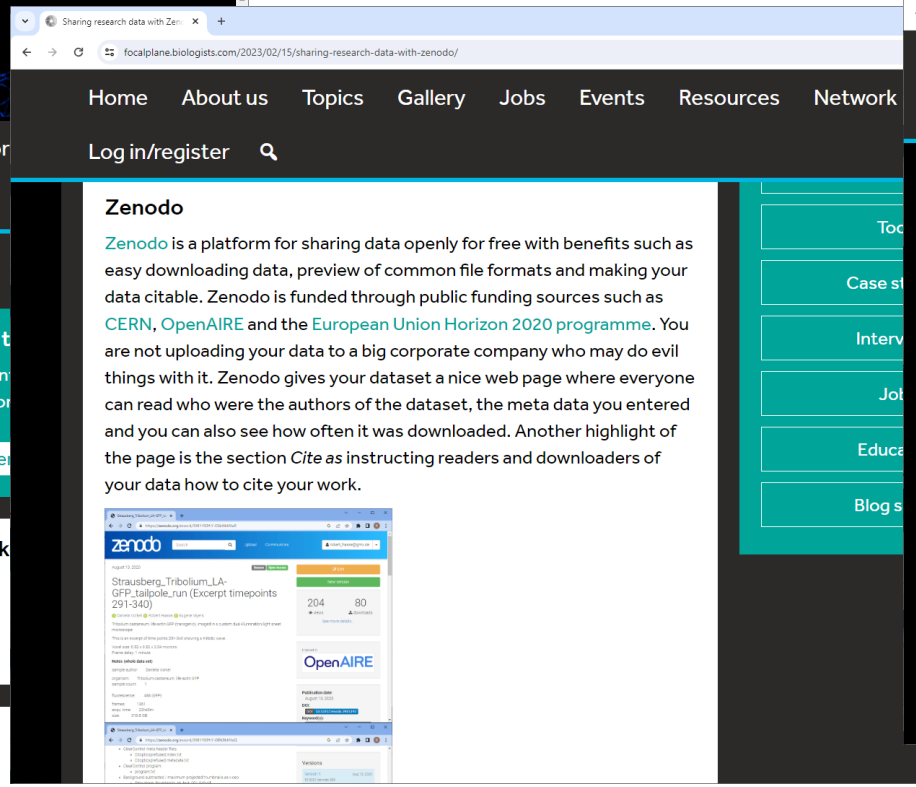
Exercise!

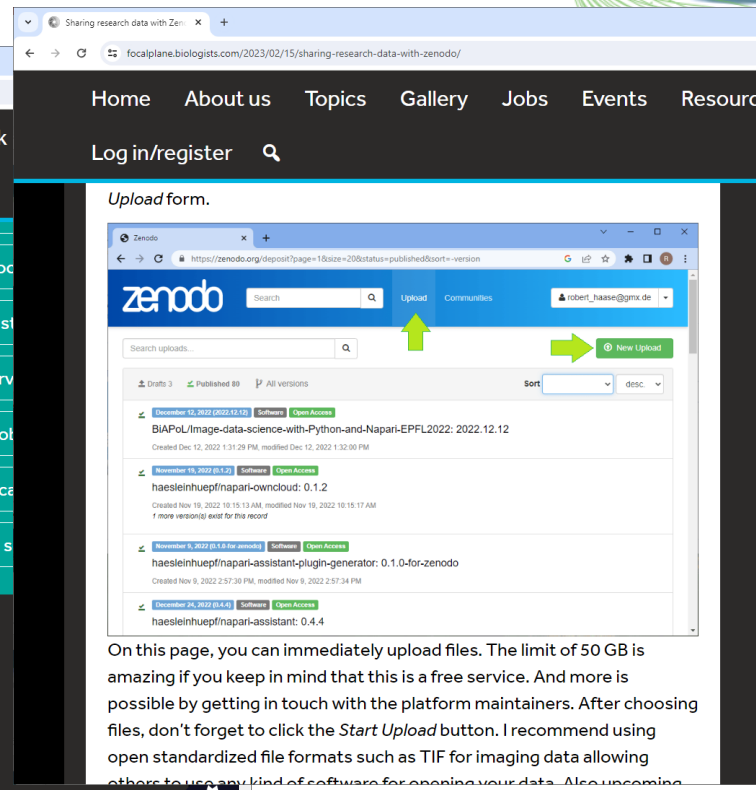
https://sandbox.zenodo.org/

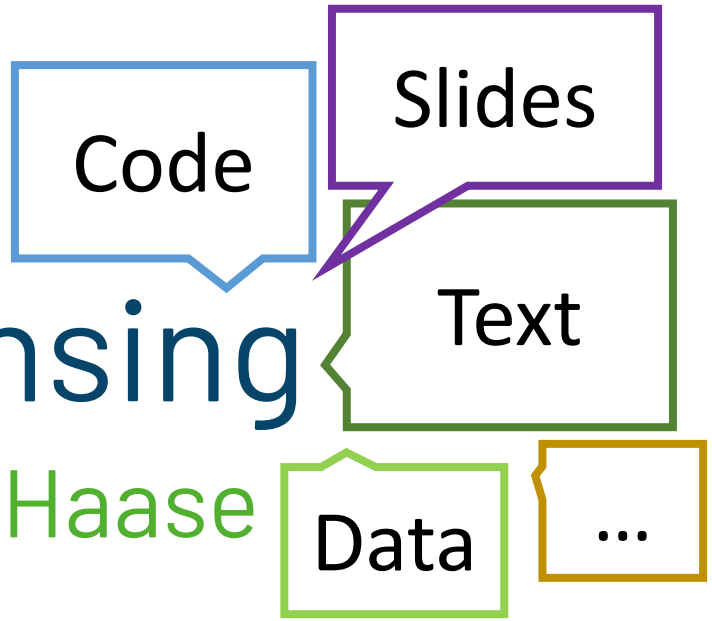
Sharing files on Zenodo

- … is easier than you think









Licensing

Robert Haase

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

SACHSEN
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

What are the consequences of this sentence?

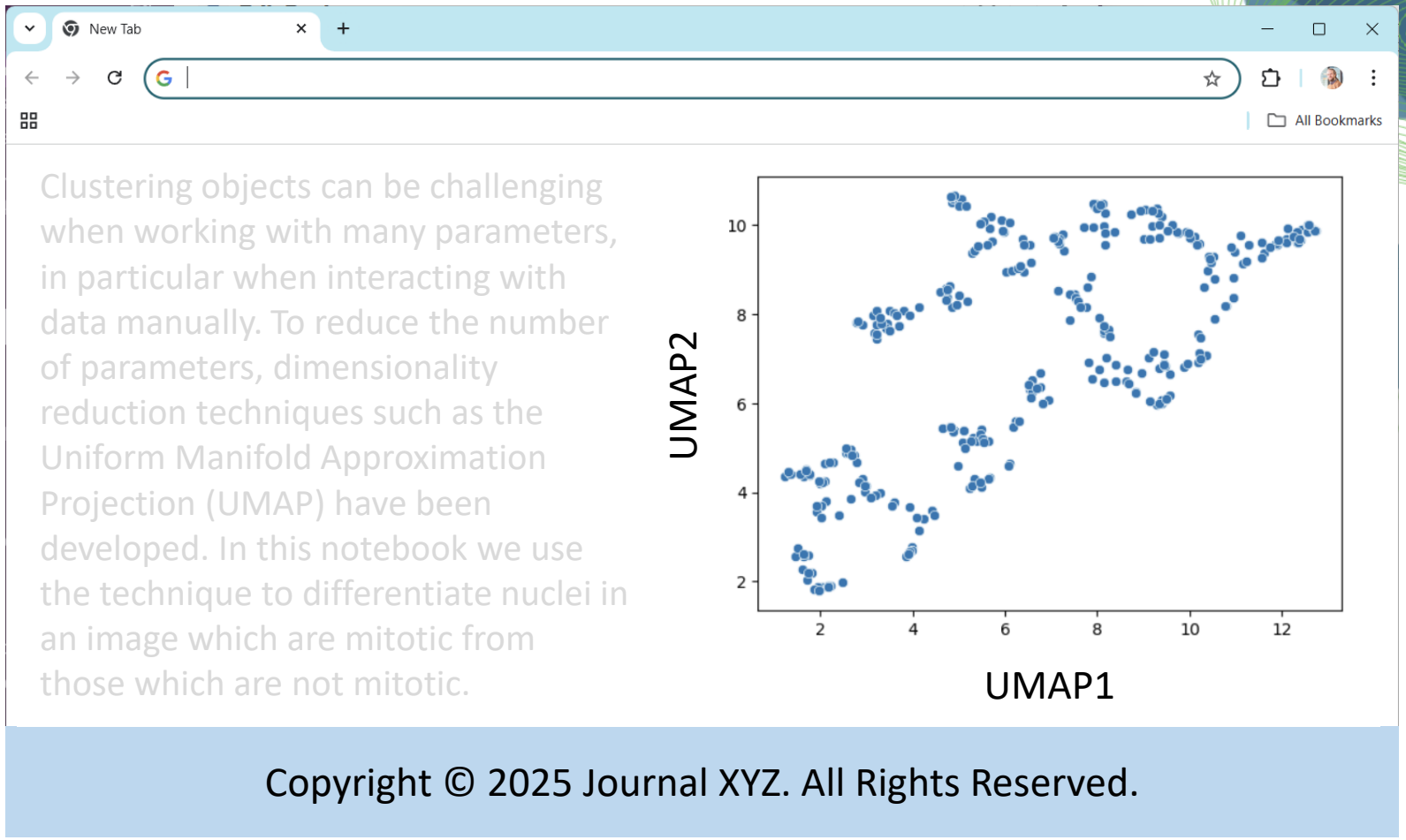
Quiz

- When you shared materials publicly on the internet, which *license* did you use?

None	Public Domain	Creative Commons	BSD/GPL/ MIT/…
			

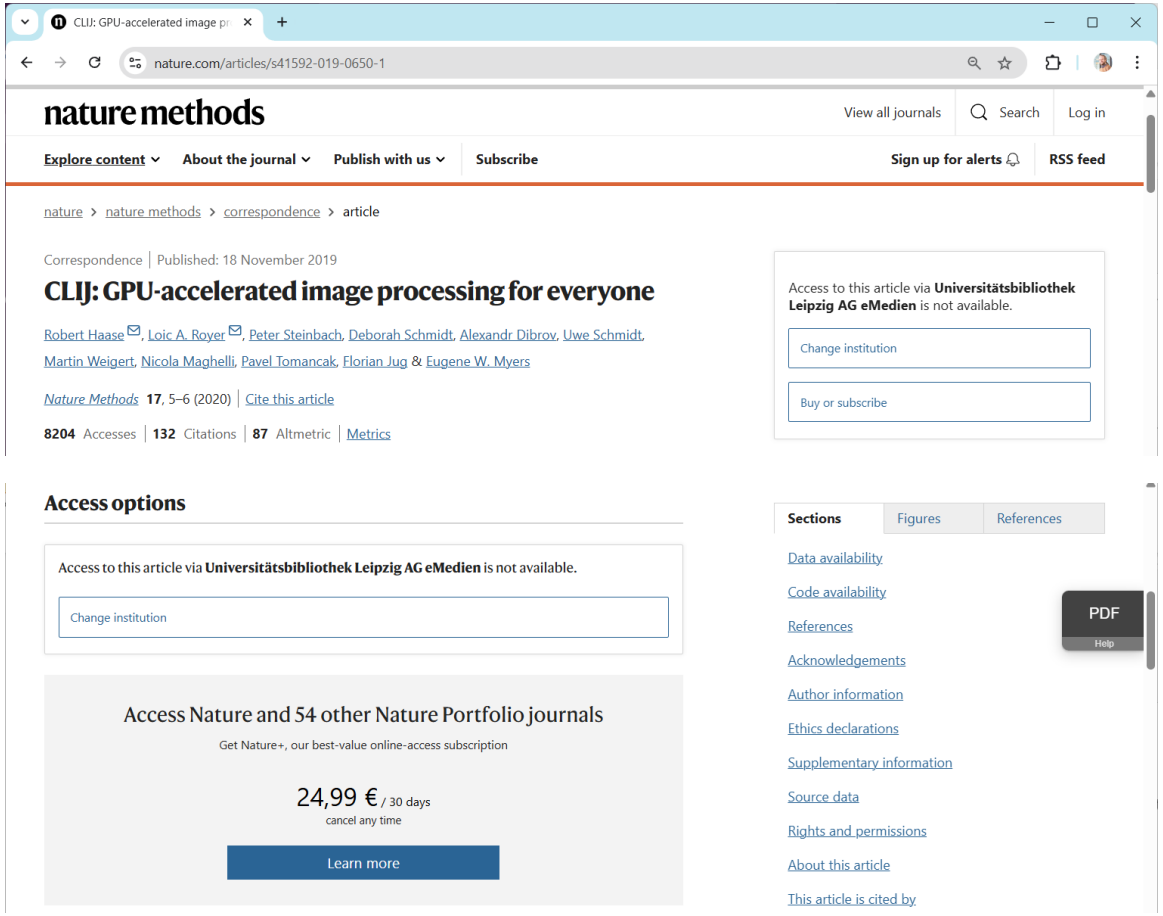
Quiz

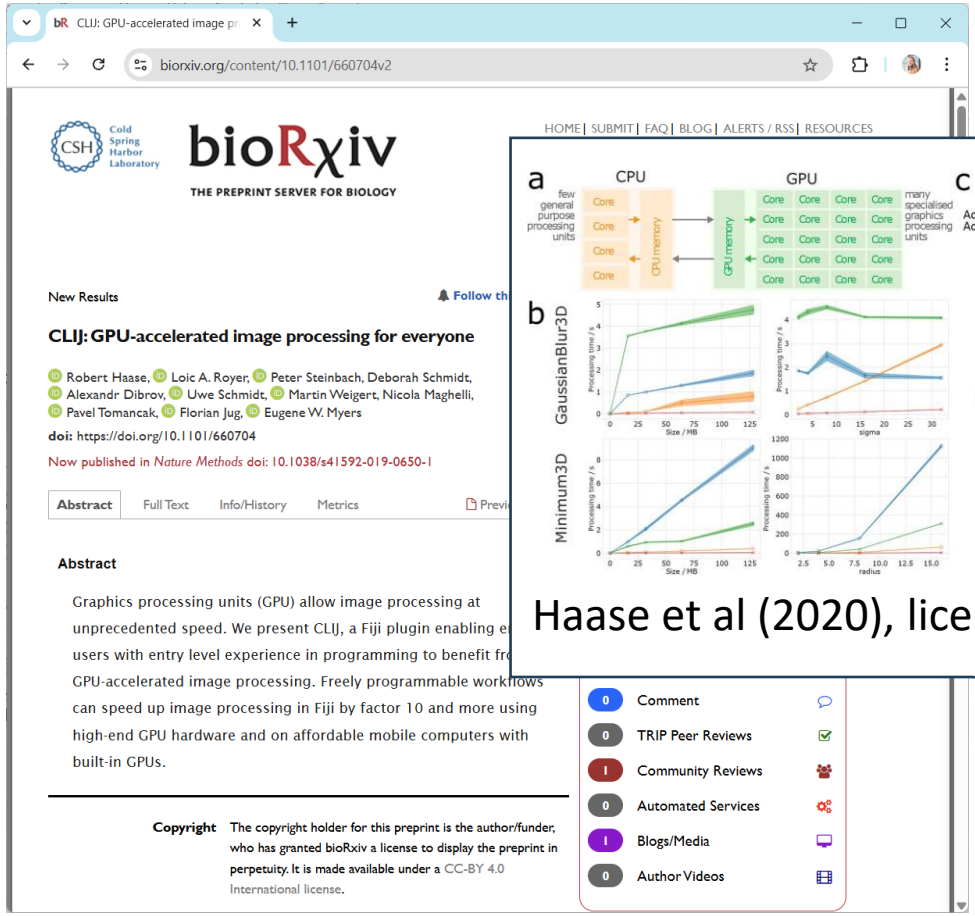
How can you reuse this plot?
What is allowed?



Hint: Search for pre-prints

In case a journal doesn't allow reusing figures from a paper, search for the corresponding preprint!

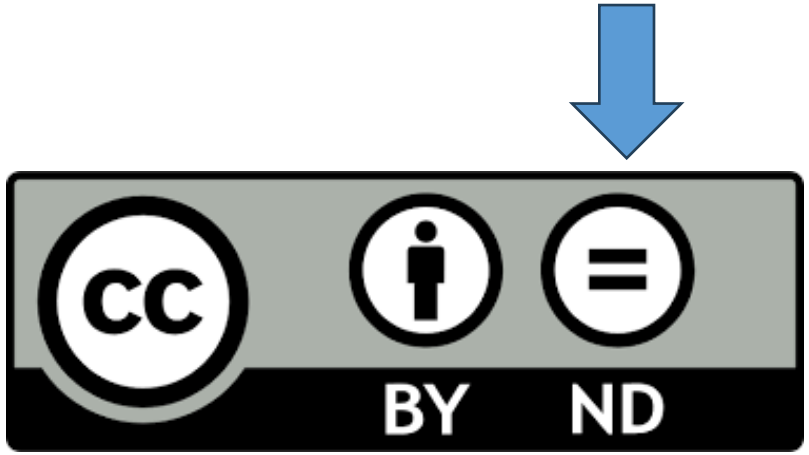




Haase et al (2020), licensed CC-BY 4.0

Licensing: Permissive versus restrictive

- Who knows what the ND stands for?



You are free to:

Share — copy and redistribute the material in any medium or format for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

"permissive"

Under the following terms:

Attribution — You must give appropriate credit , provide a link to the license, and indicate if changes were made . You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NoDerivatives — If you remix, transform, or build upon the material, you may not distribute the modified material.

"restrictive"

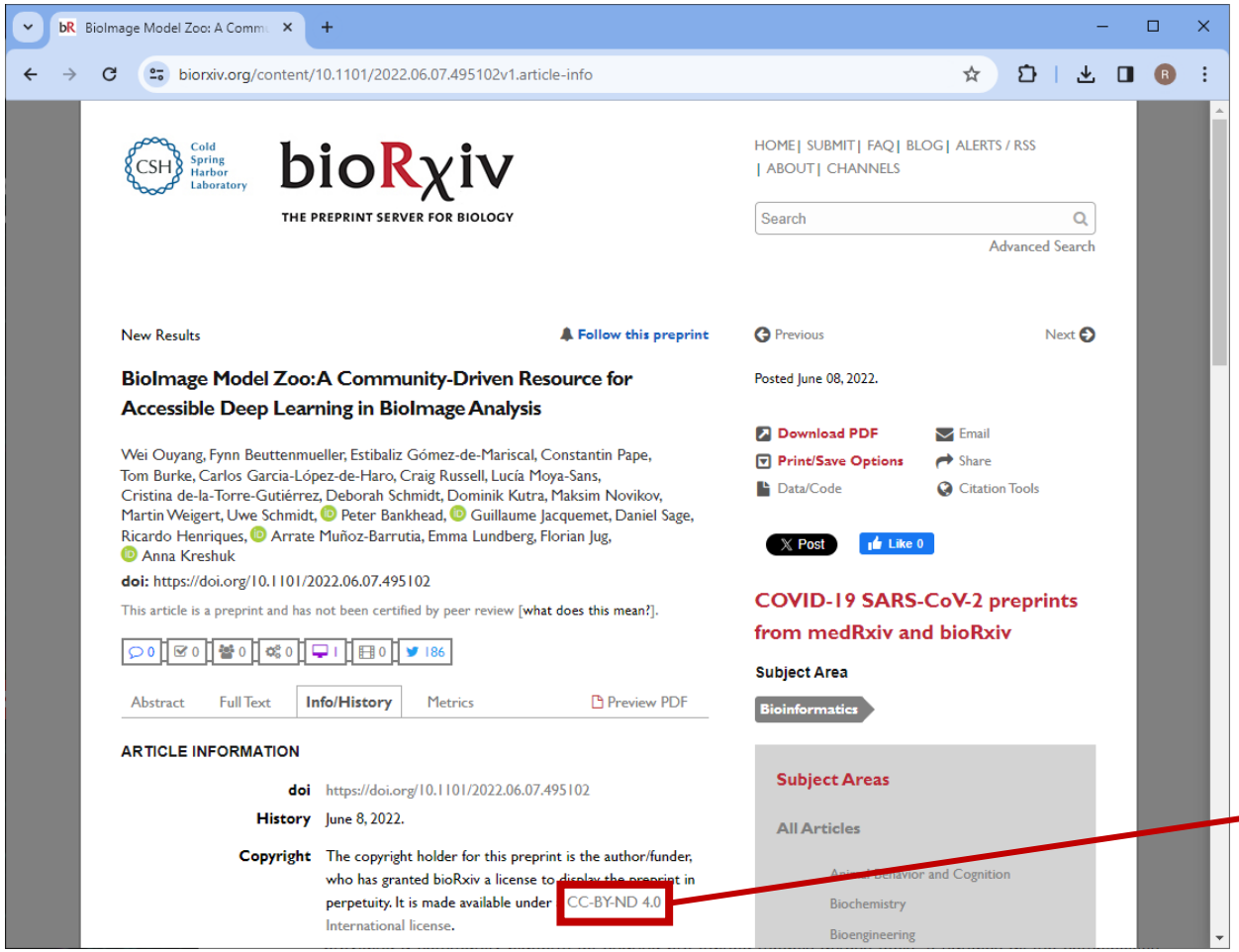
Licensing: Permissive versus restrictive

I hope nobody feels hurt
by the following slides.

I just would like to make a point.

Licensing: Permissive versus restrictive

Example

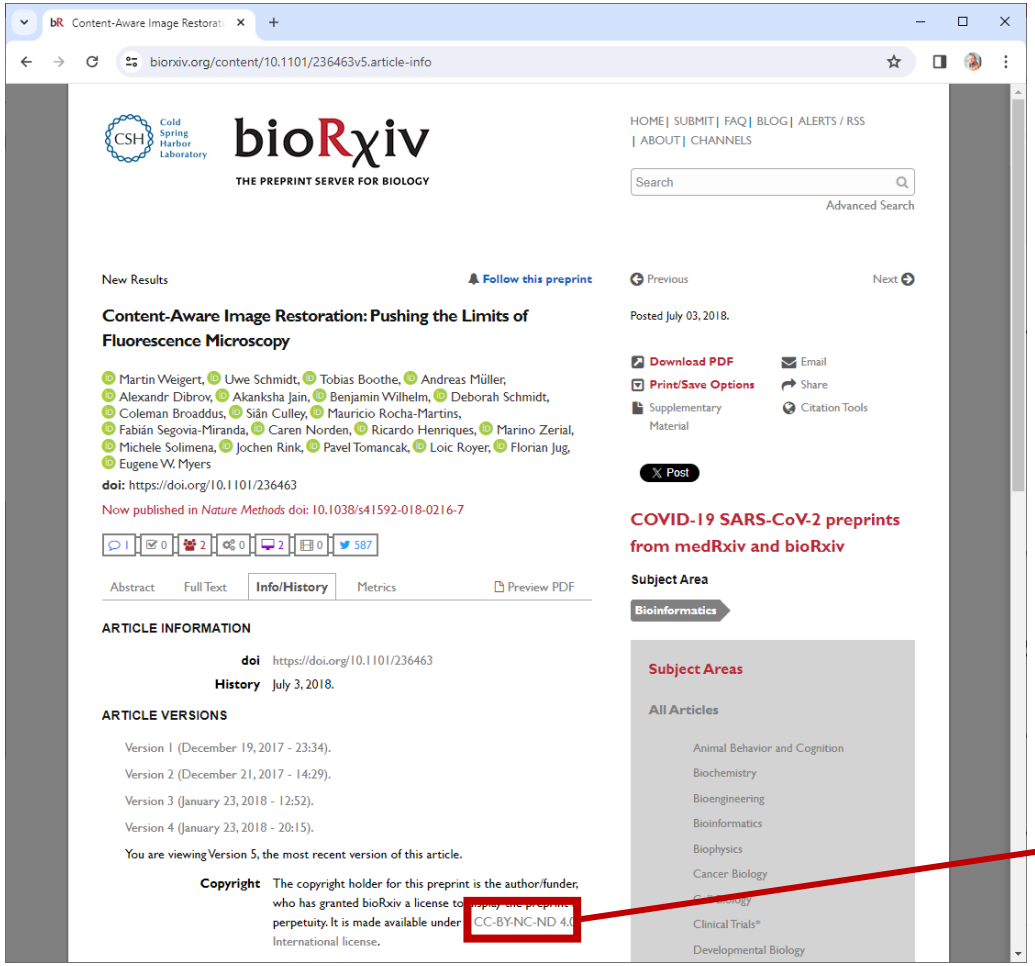


I would love to show you a Figure from this paper!

But I'm not allowed!

Licensing: Permissive versus restrictive

Example

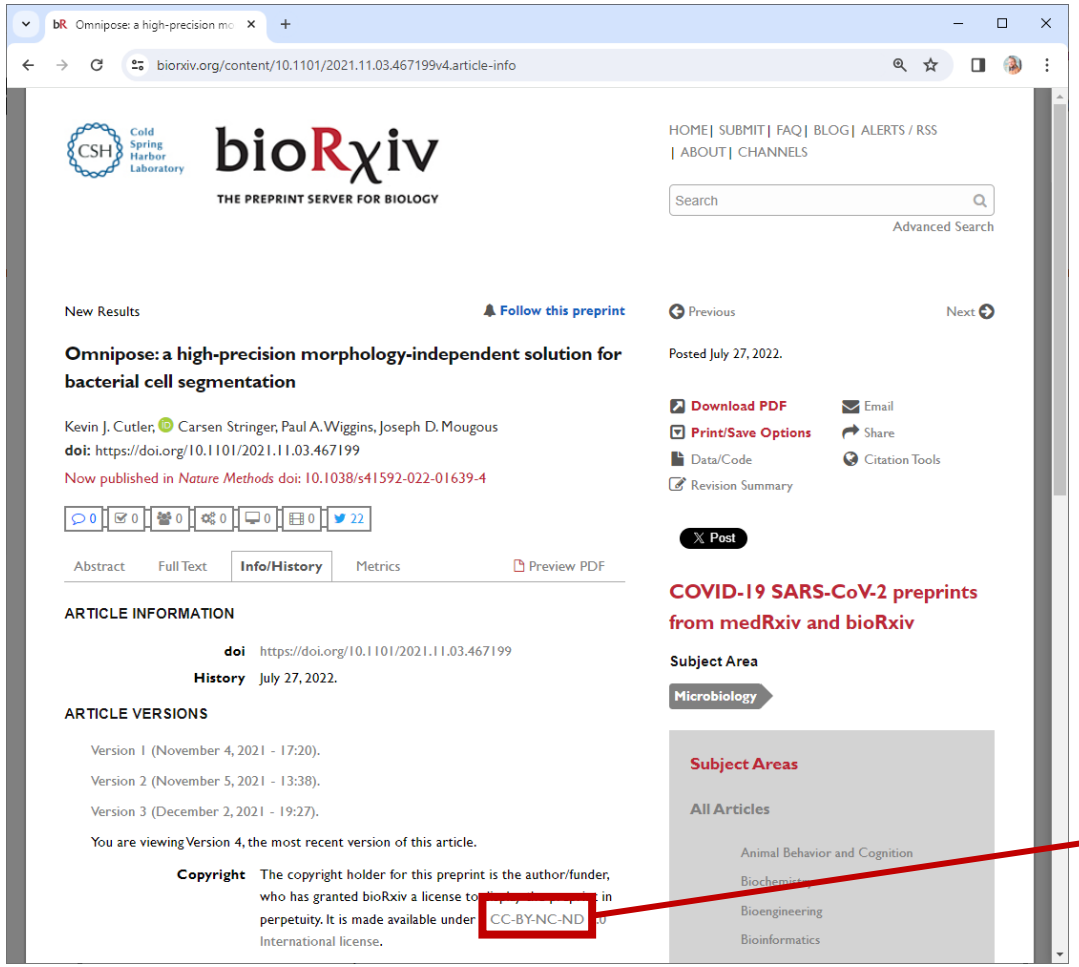


I would love to show you a Figure from this paper!

But I'm not allowed!

Licensing: Permissive versus restrictive

Example

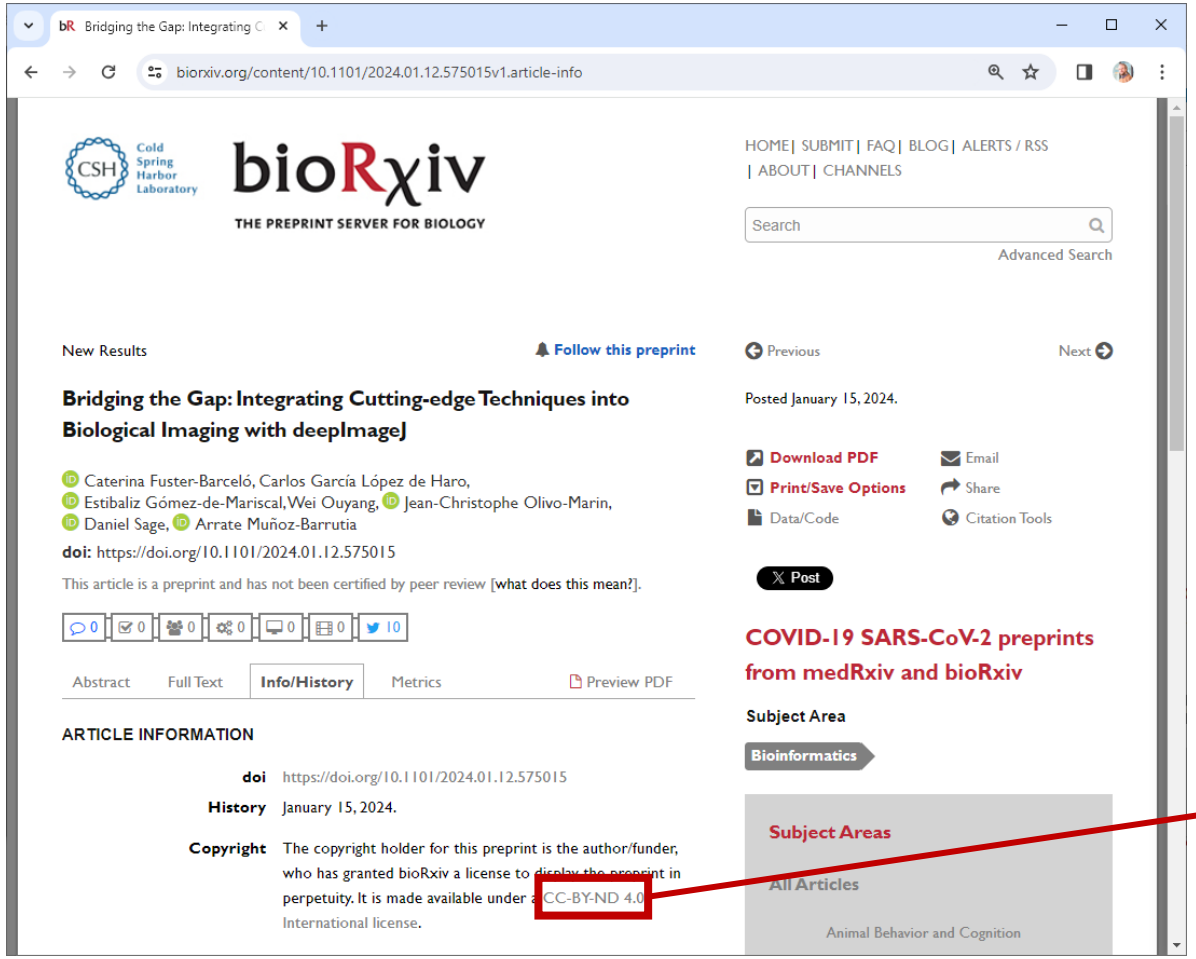


I would love to show you a Figure from this paper!

But I'm not allowed!

Licensing: Permissive versus restrictive

Example

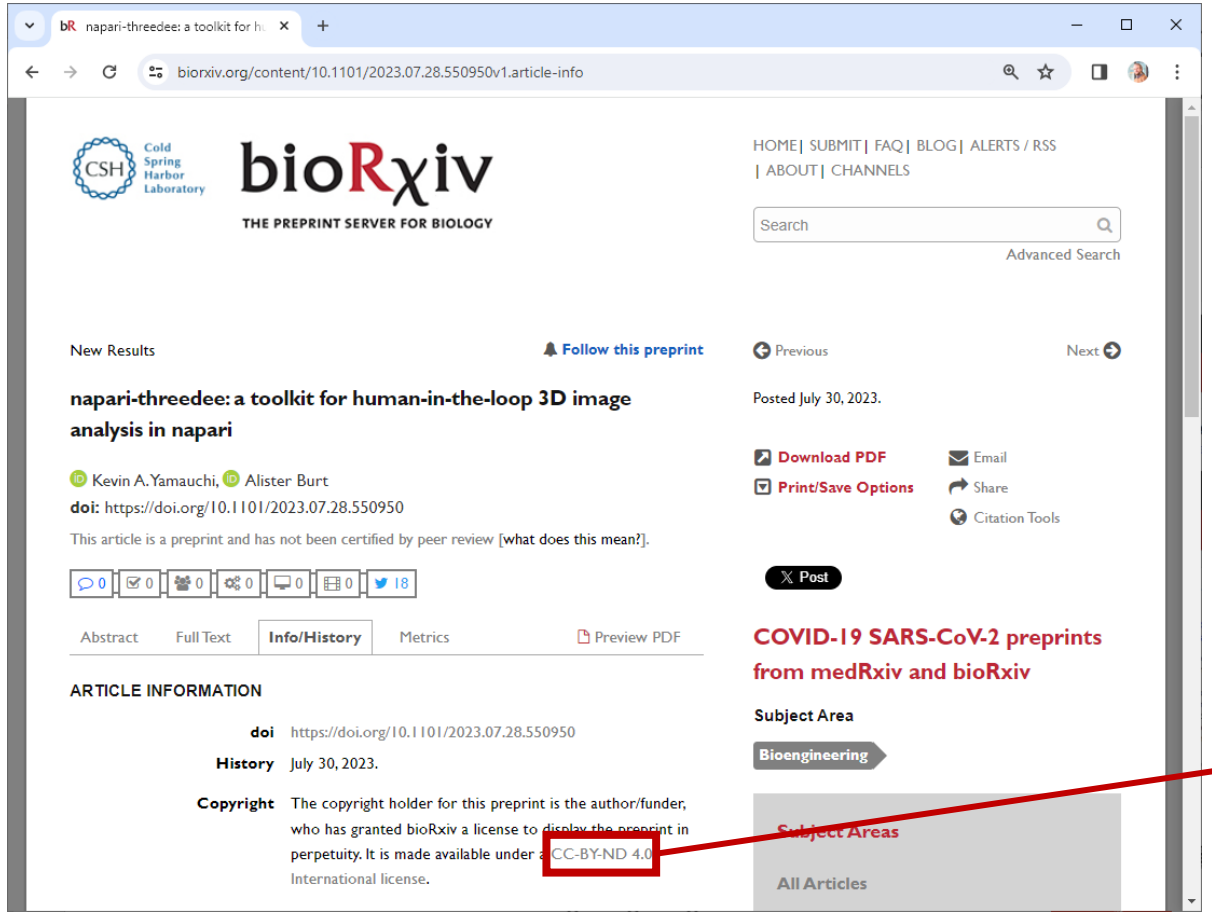


I would love to show you a Figure from this paper!

But I'm not allowed!

Licensing: Permissive versus restrictive

Example

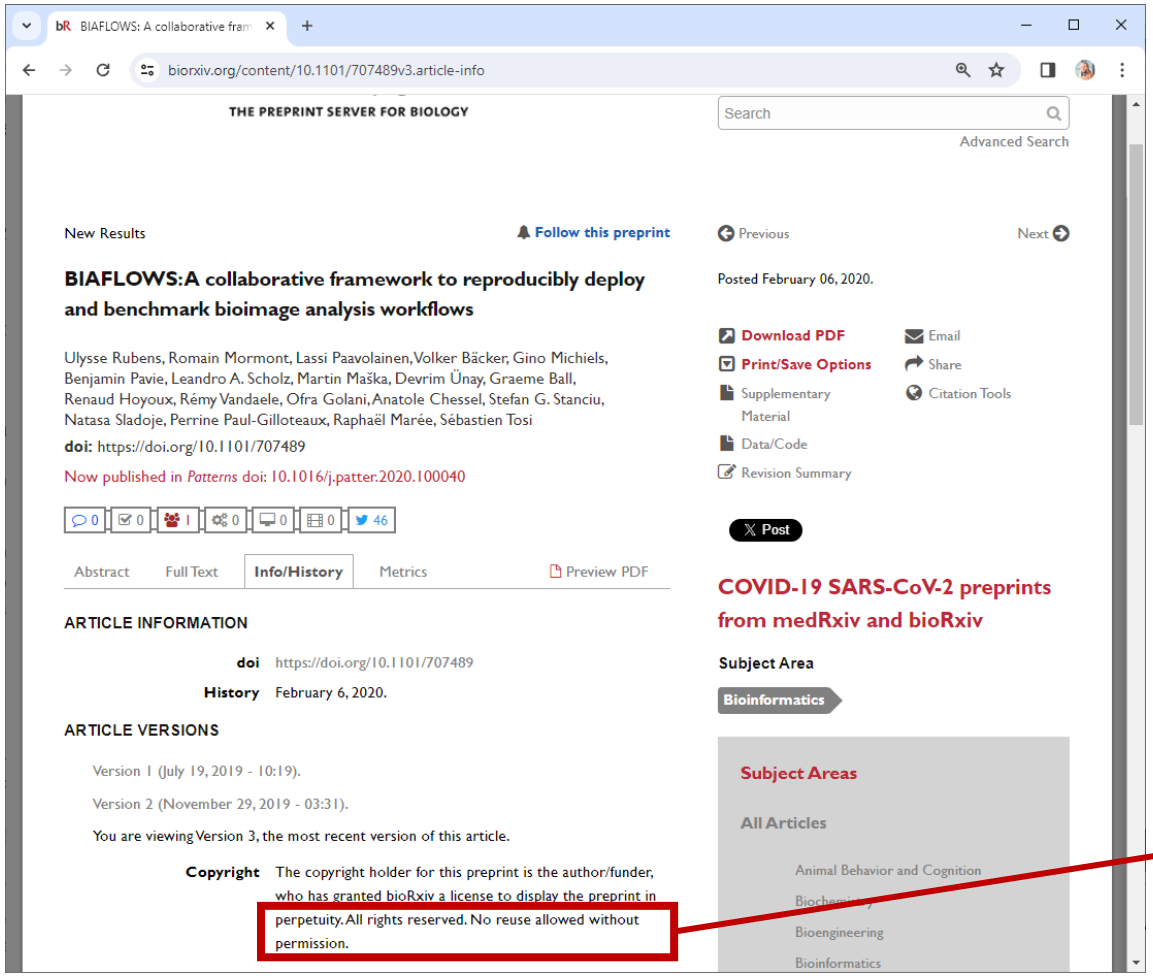


I would love to show you a Figure from this paper!

But I'm not allowed!

Licensing: Permissive versus restrictive

Example

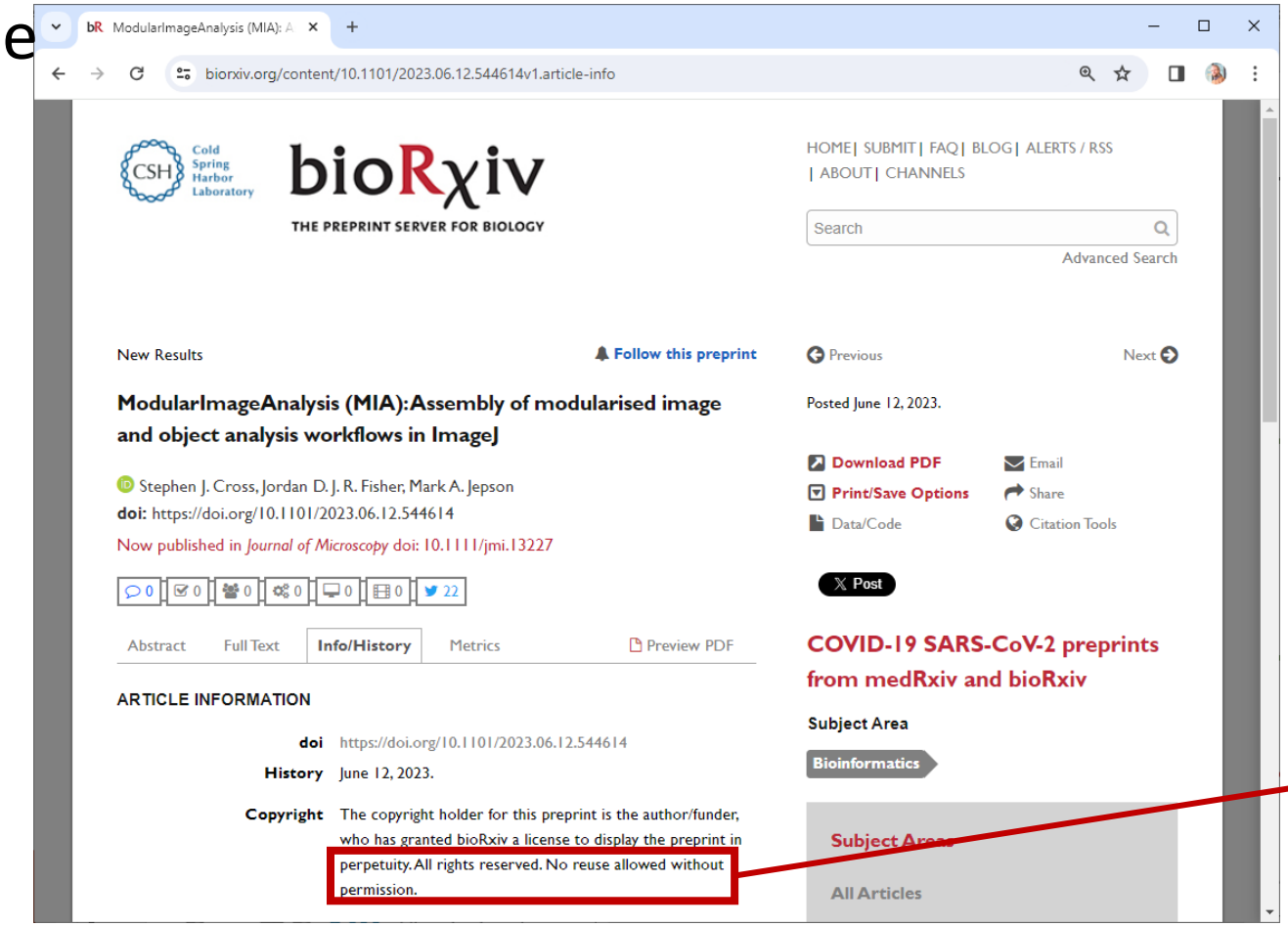


I would love to show you a Figure from this paper!

But I'm not allowed!

Licensing: Permissive versus restrictive

Example

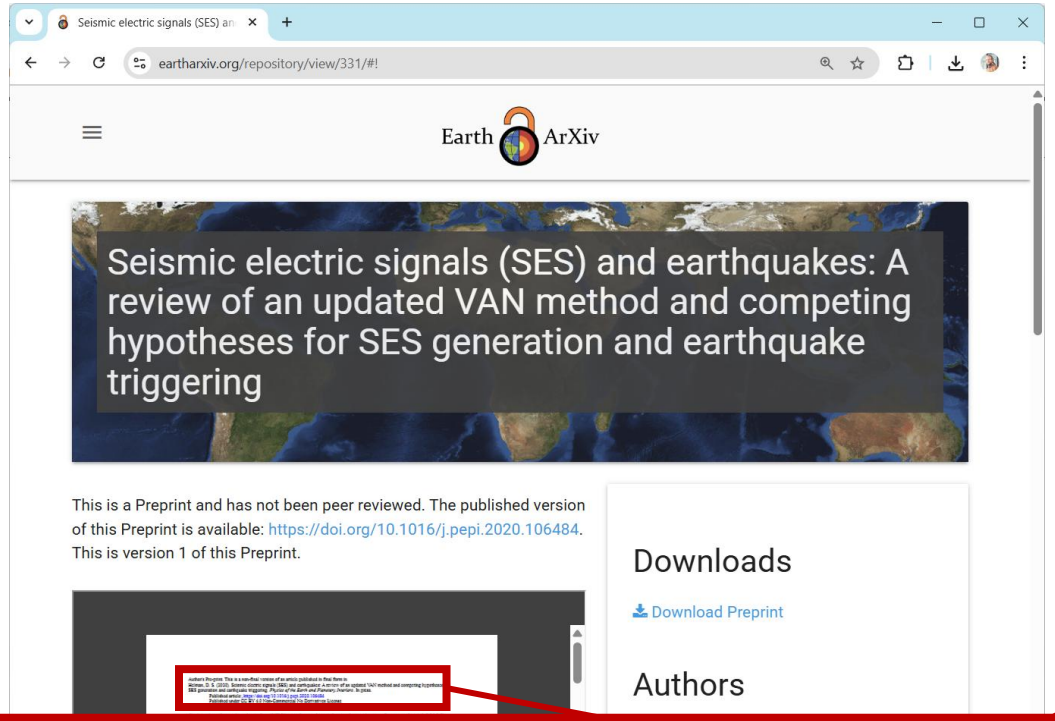


I would love to show you a Figure from this paper!

But I'm not allowed!

Licensing: Permissive versus restrictive

Example



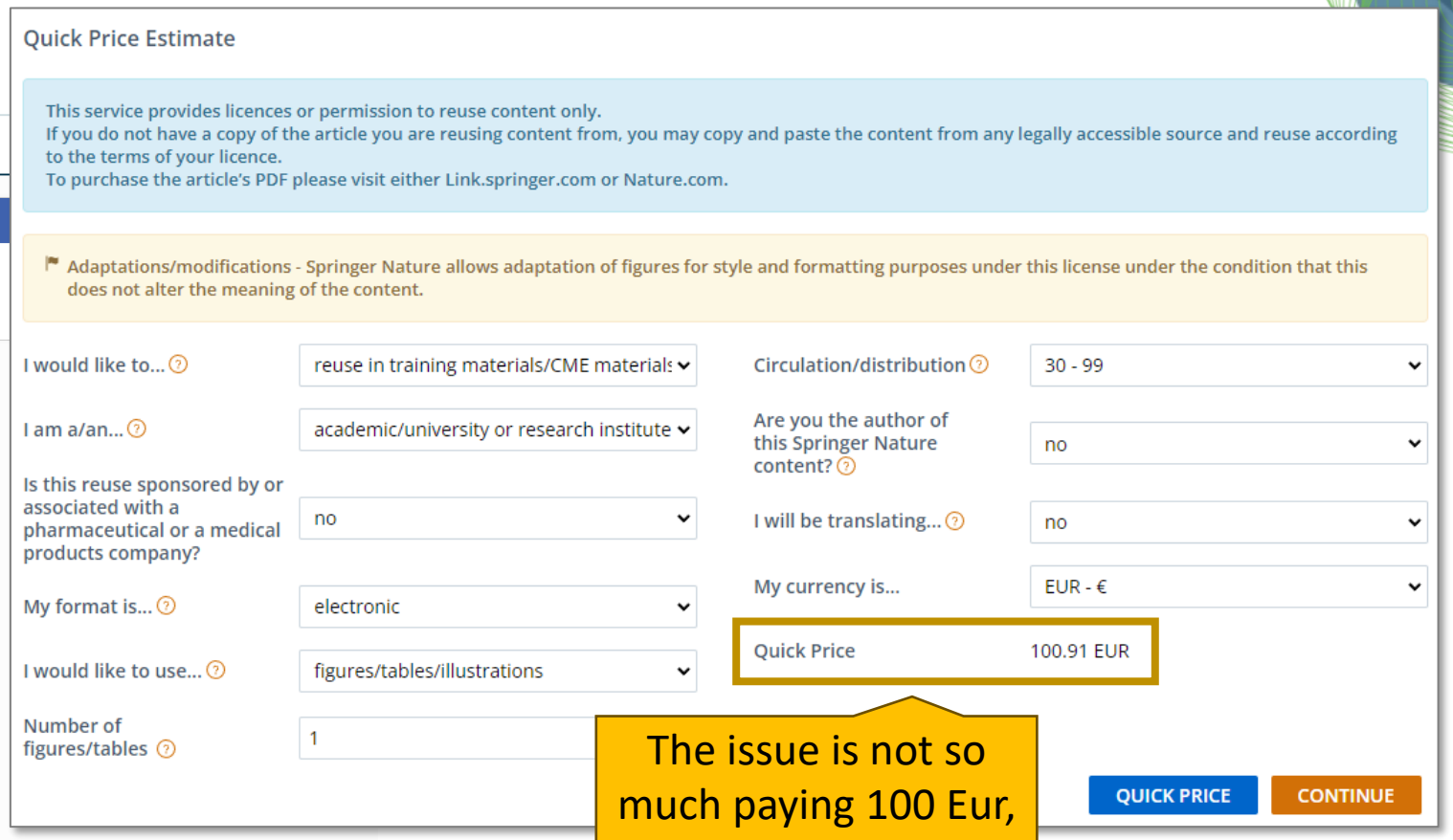
I would love to show you a Figure from this paper!

But I'm not allowed!

Author's Pre-print. This is a non-final version of an article published in final form in
Helman, D. S. (2020). Seismic electric signals (SES) and earthquakes: A review of an updated VAN method and competing hypotheses for SES generation and earthquake triggering. *Physics of the Earth and Planetary Interiors*. In press.
Published article: https://doi.org/10.1016/j.pepi.2020.106484
Published under CC BY 4.0 Non-Commercial No Derivatives License

Licensing: Permissive versus restrictive

Example



The issue is not so much paying 100 Eur, but the related administrative effort.

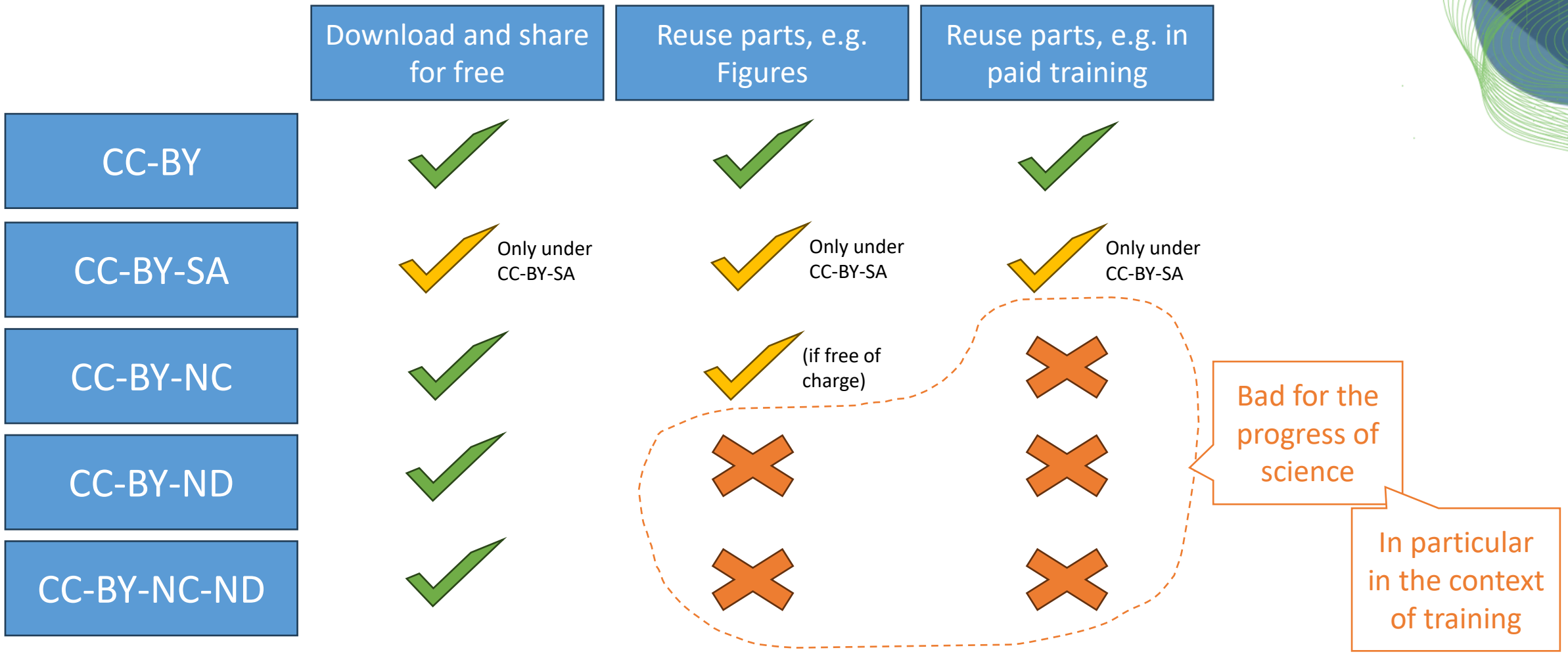
Licensing: Permissive versus restrictive

Restrictive licensing is
a community-wide issue.

I *presume* due to lack of awareness & training

Train the trainers!

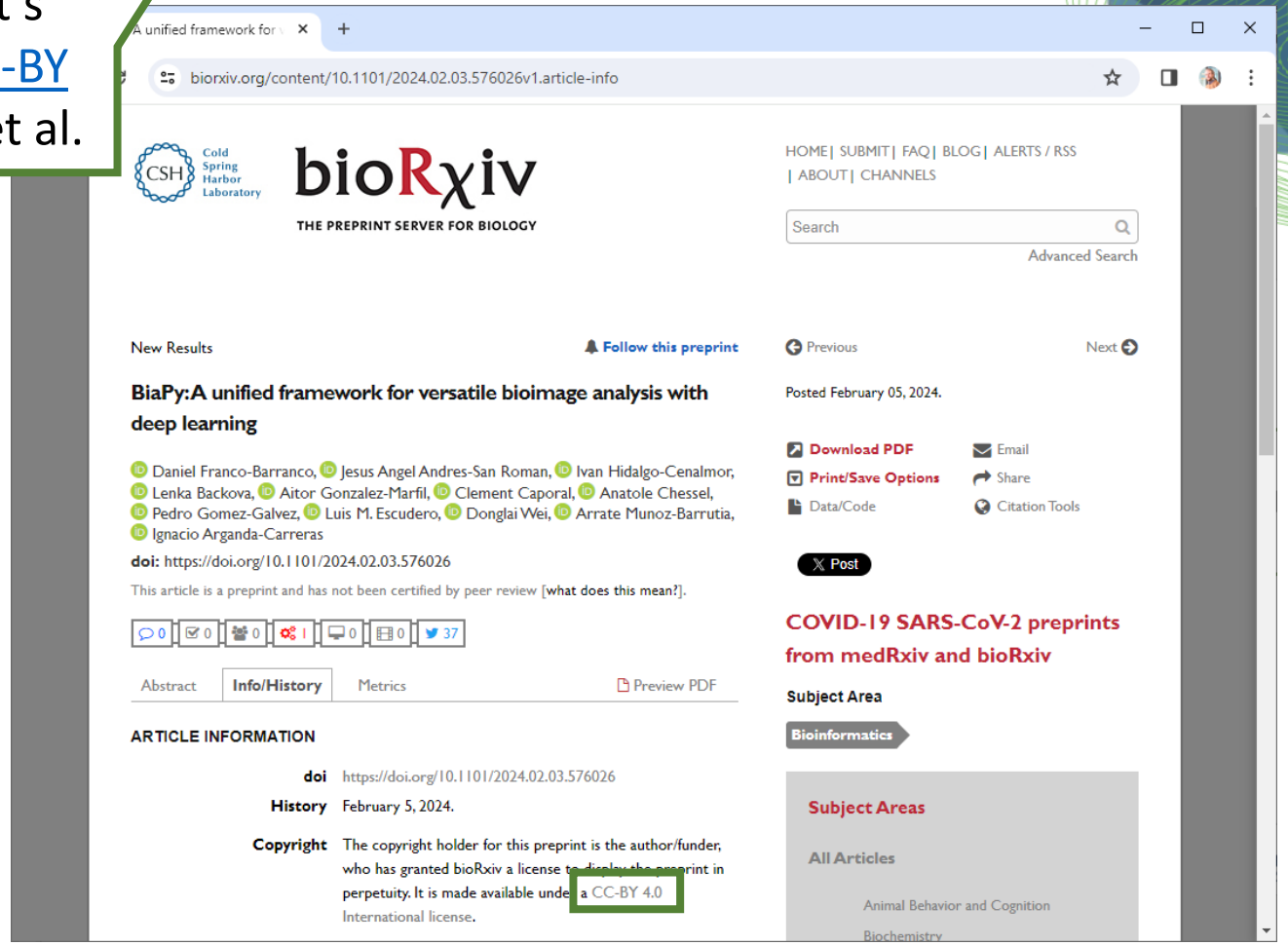
Licensing: Permissive versus restrictive



Licensing: Permissive versus restrictive

Example

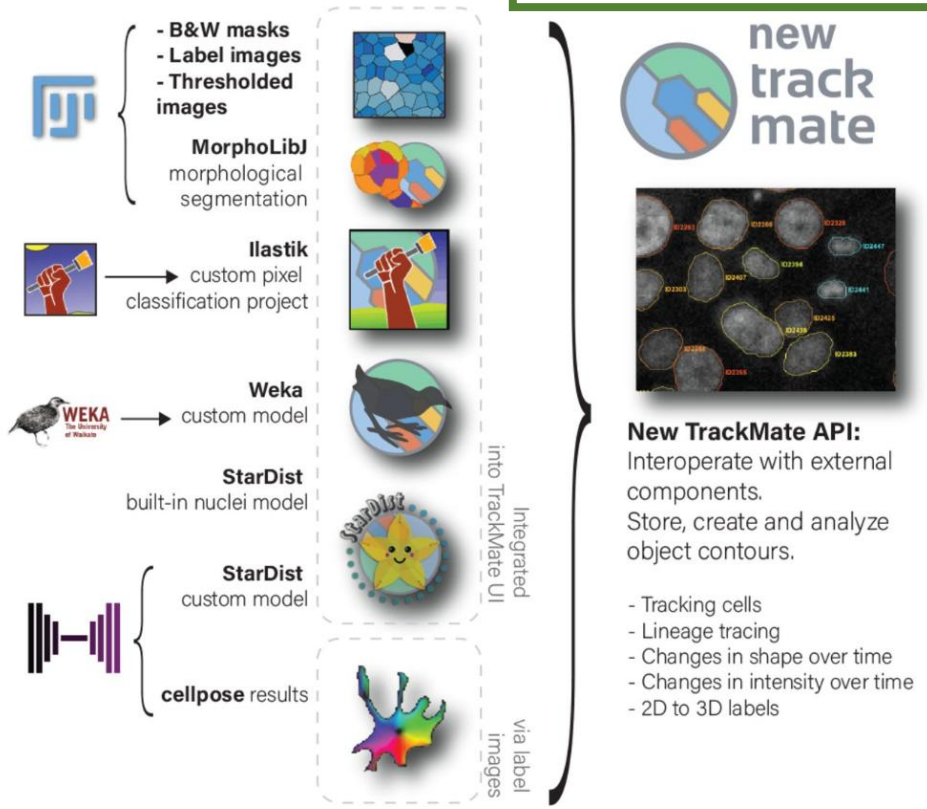
Look at this great figure! It's cropped from / licensed CC-BY 4.0 by D. Franco-Barranco et al.

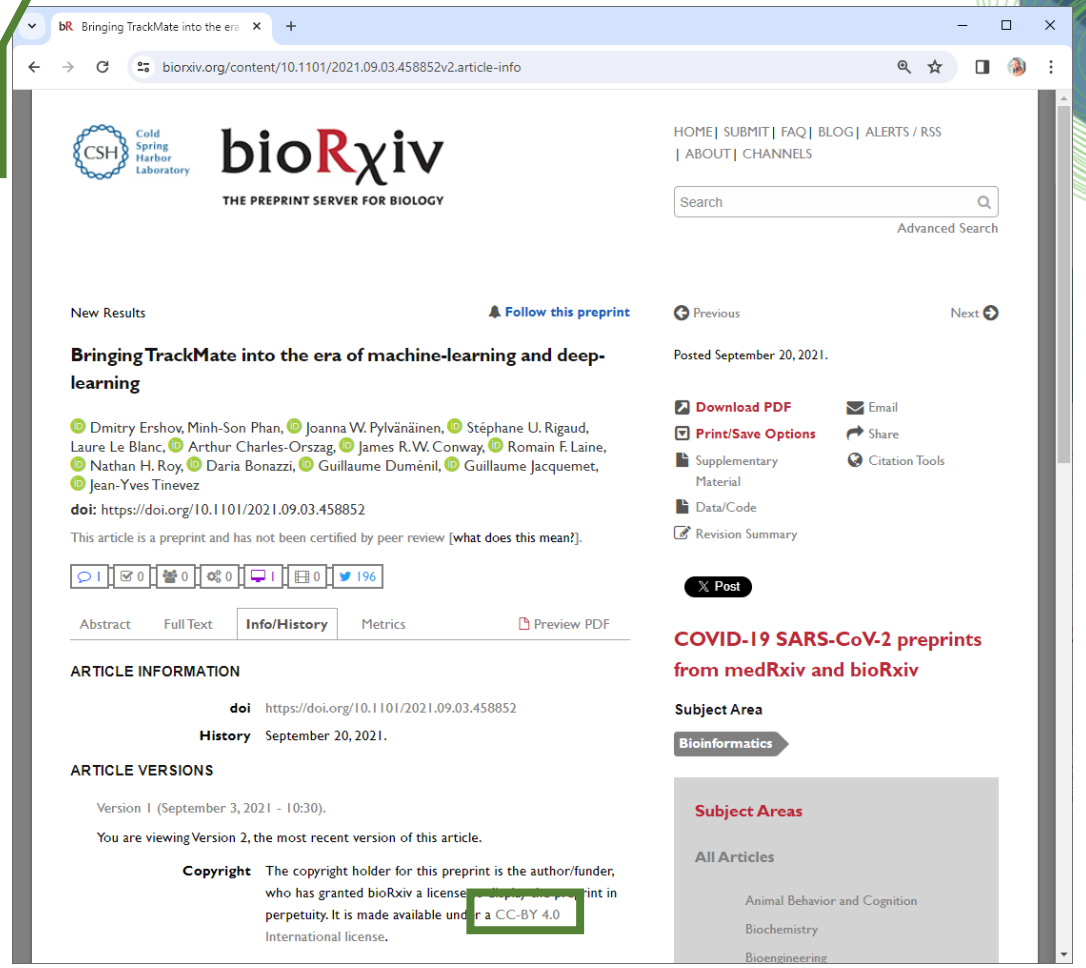


Licensing: Permissive versus restrictive

Example

Look at this great figure!
It's take from / licensed CC-BY 4.0 by D. Ershov et al.

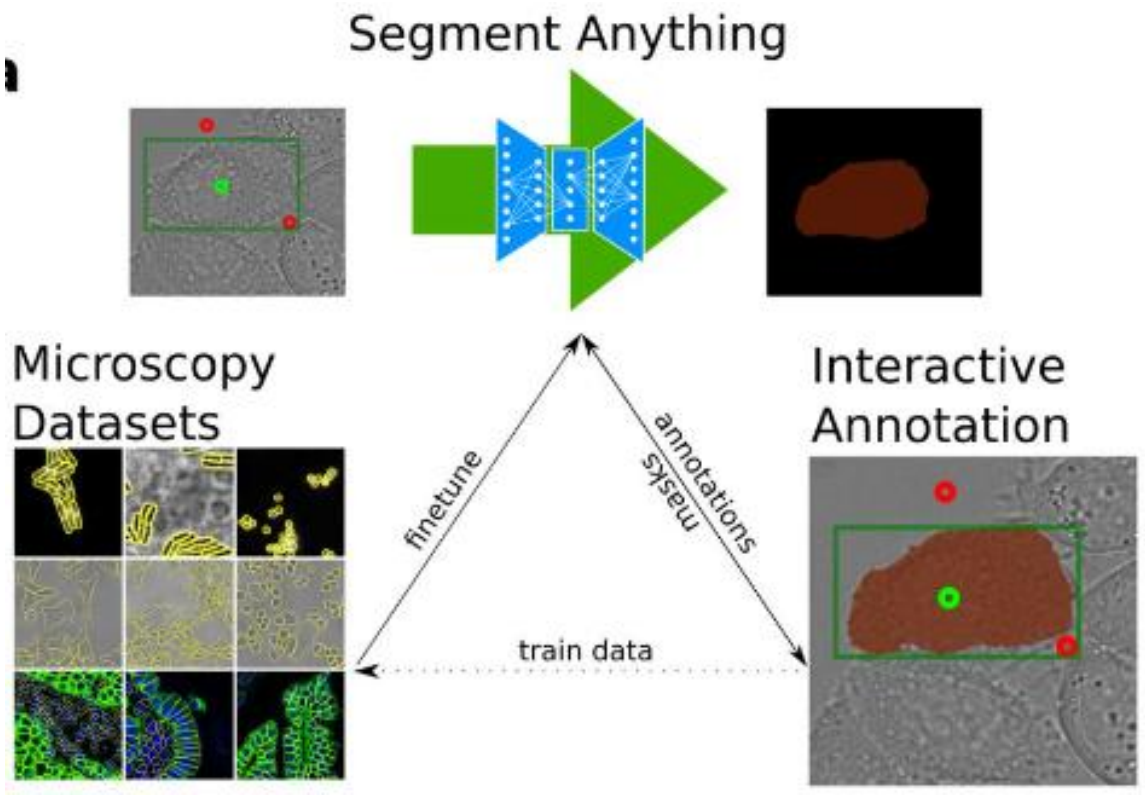


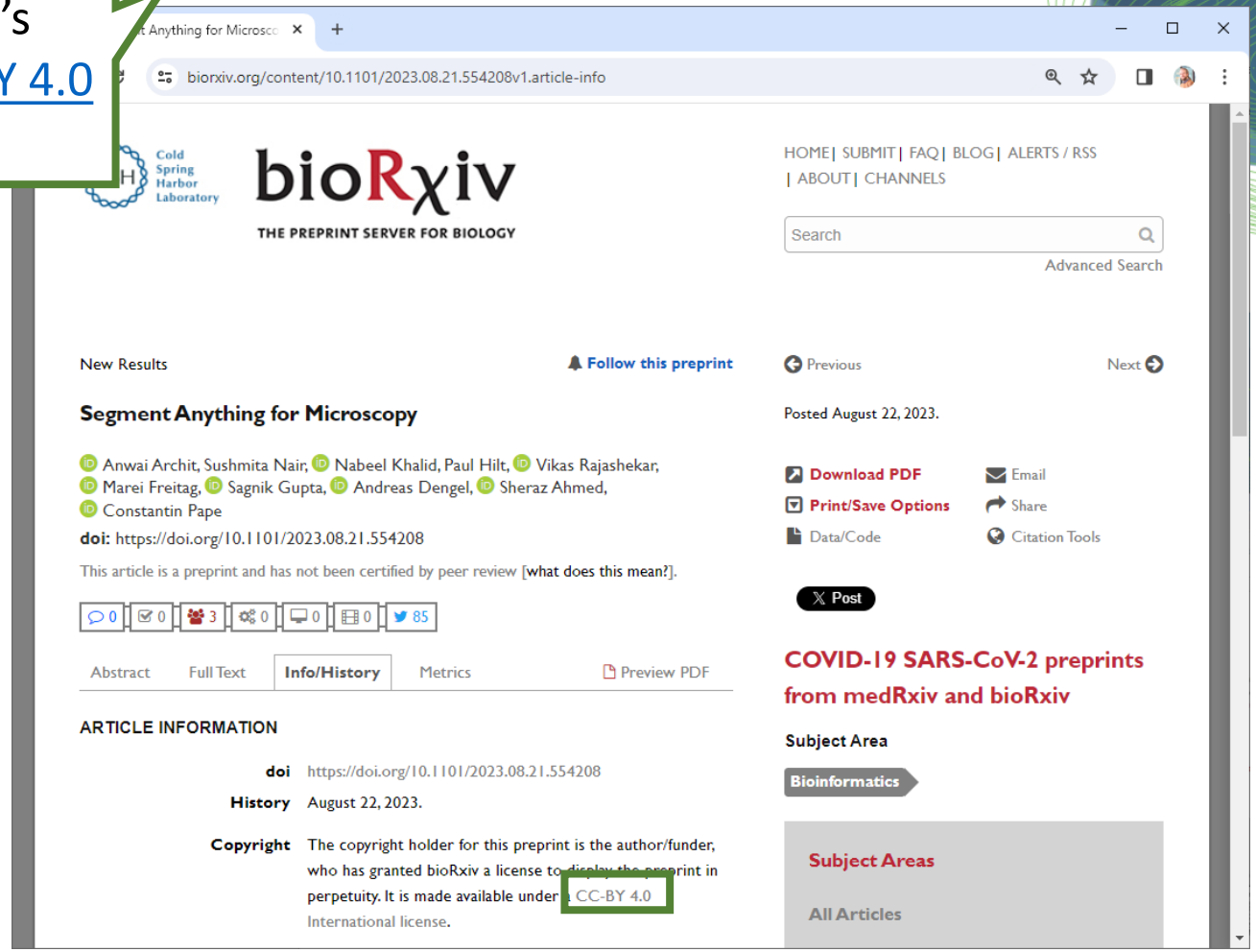


Licensing: Permissive versus restrictive

Example

Look at this great figure! It's cropped from / licensed CC-BY 4.0 by A. Archit et al.

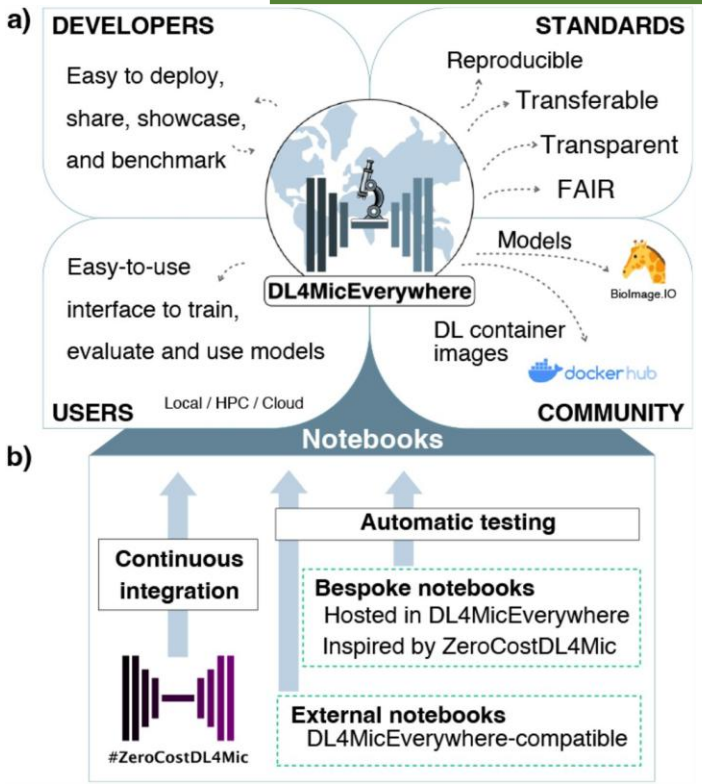


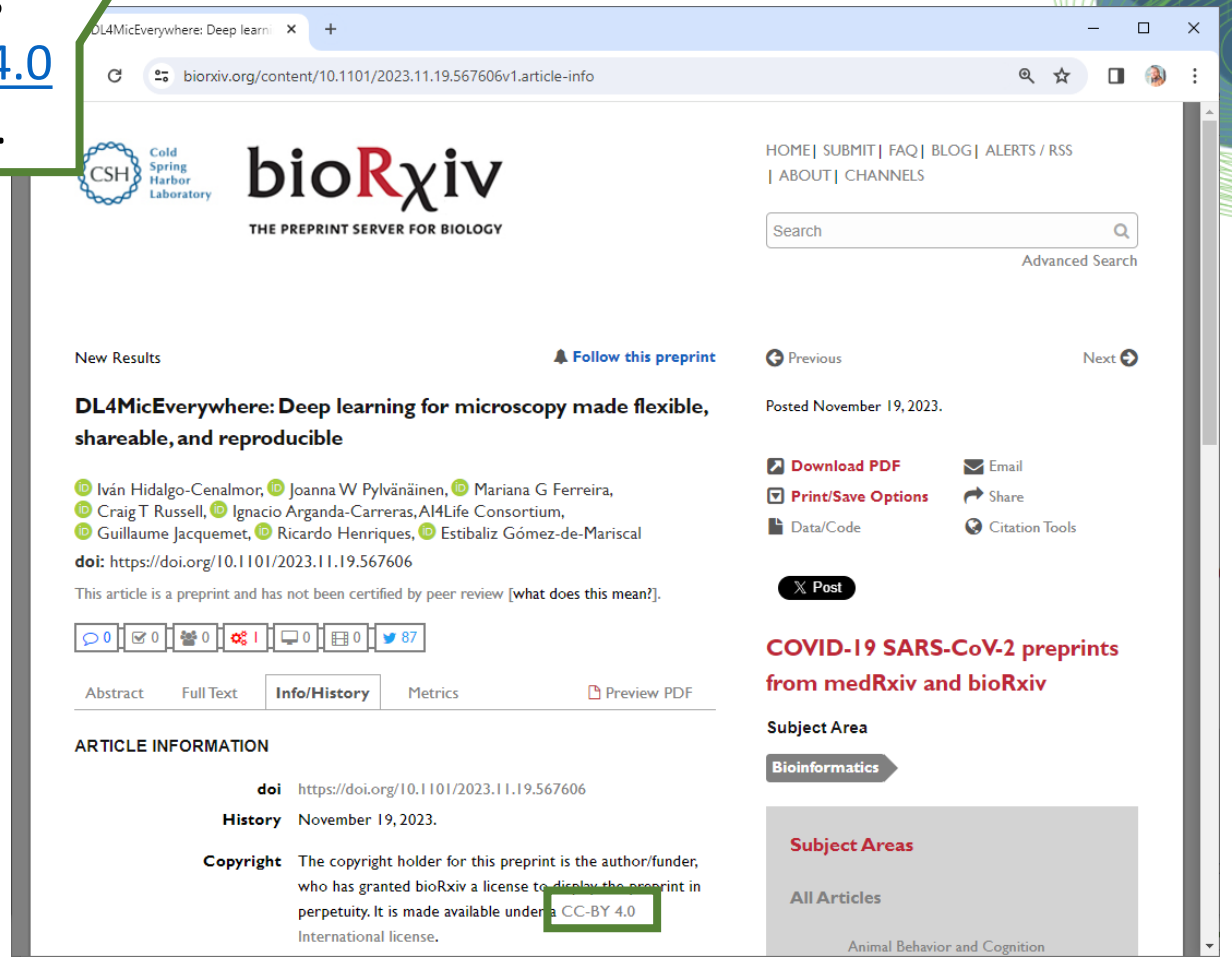


Licensing: Permissive versus restrictive

Example

Look at this great figure! It's cropped from licensed CC-BY 4.0 by I. Hidalgo-Cenalmor et al.

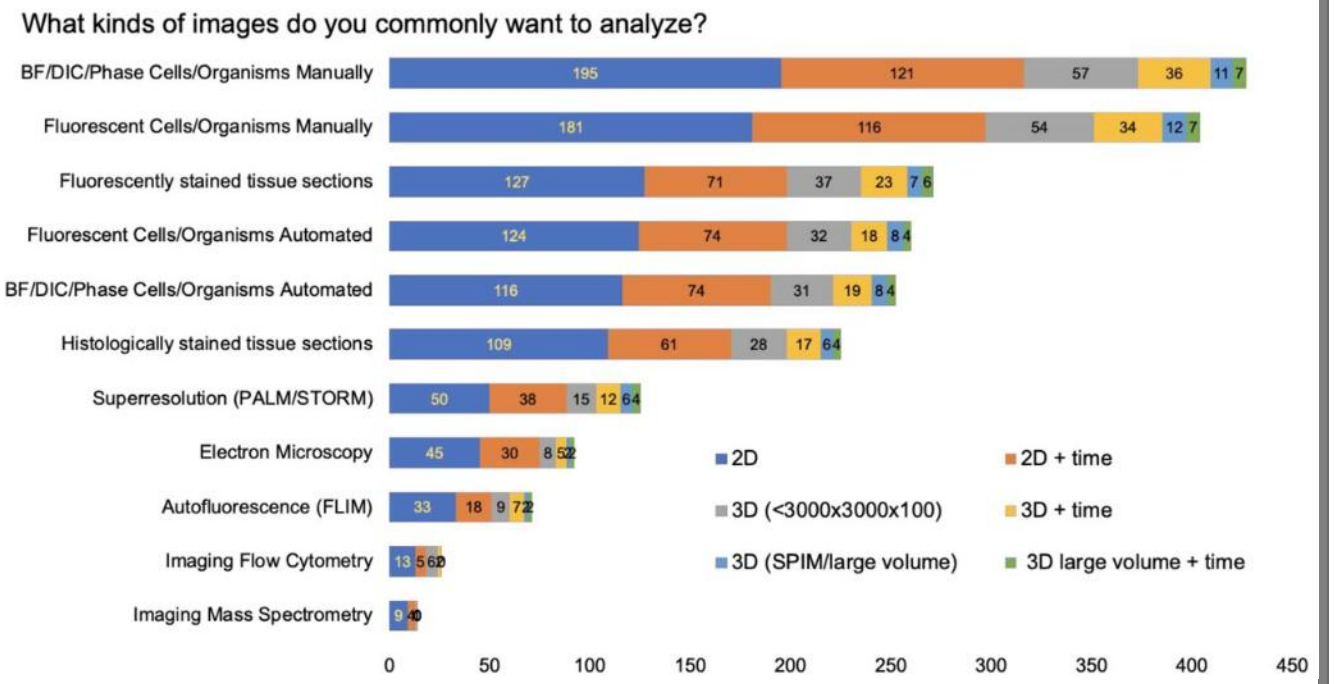


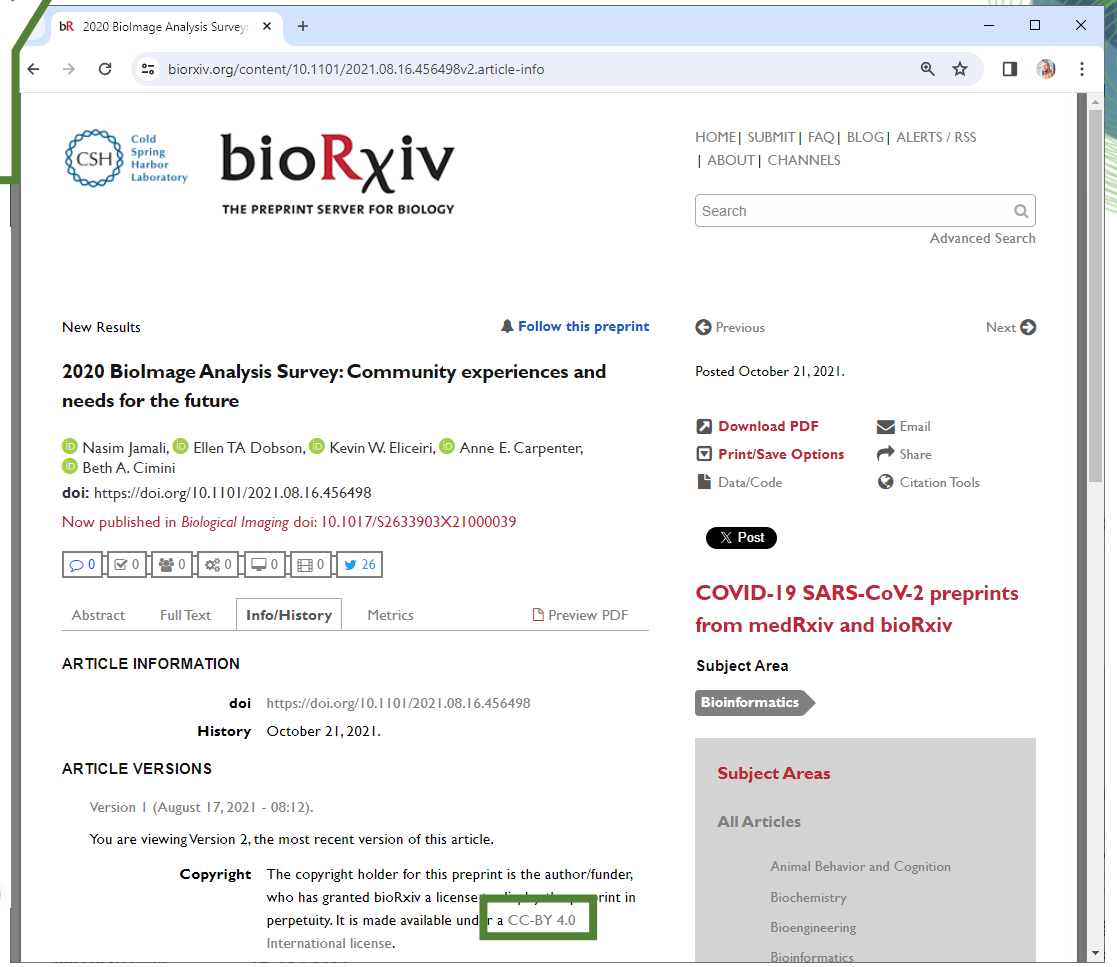


Licensing: Permissive versus restrictive

Example

Look at this great figure! It's cropped from / licensed CC-BY 4.0 by N. Jamali et al.

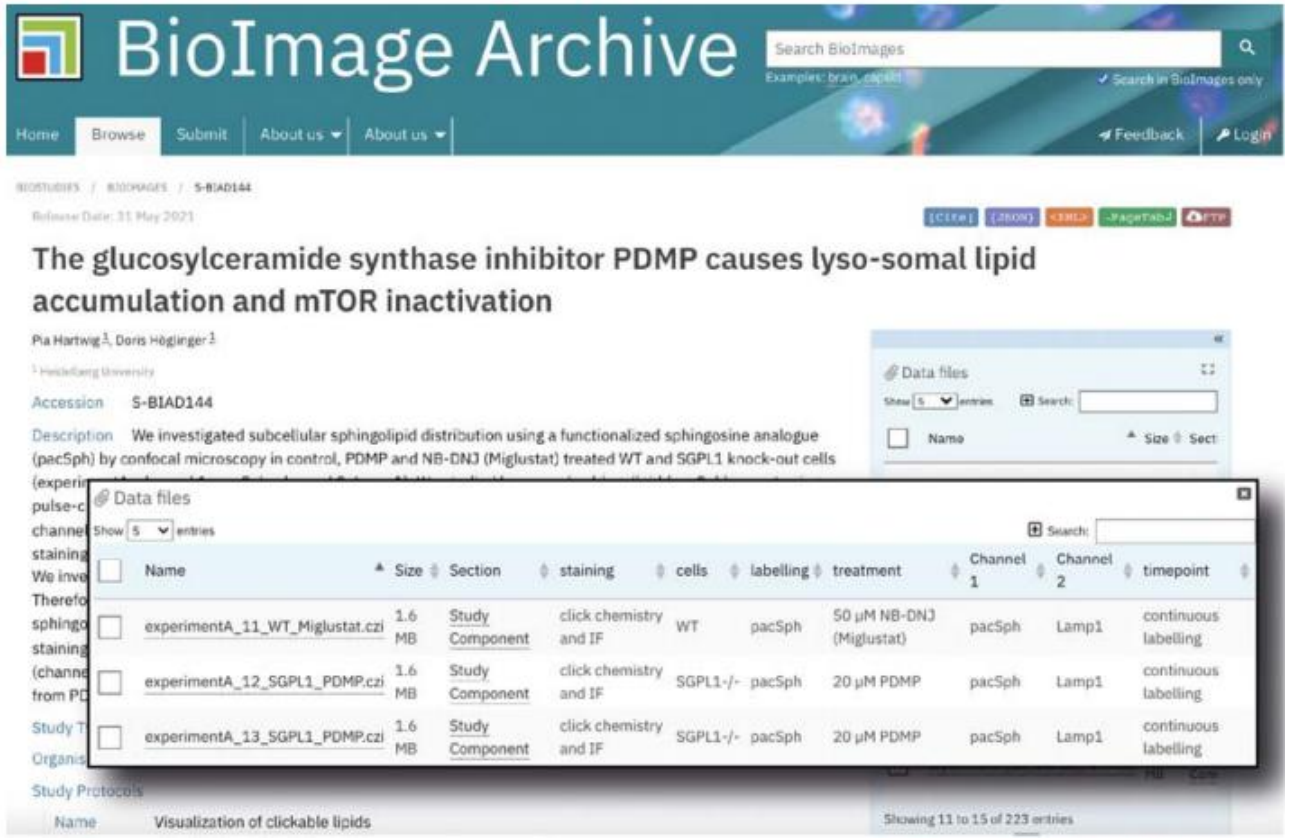




Licensing: Permissive versus restrictive

Example

Look at this great figure! It's taken from M. Hartley et al.



Incentives

The system is changing currently towards more openness (thankfully)

Career goal:	Open Science	Open Training
Research PI / Professor	✓	✓
Lecturer*	✓	✓
Academic staff scientist	✓	✓
Industry engineer	✗	✗

In industry, secrecy plays a key role because of $$

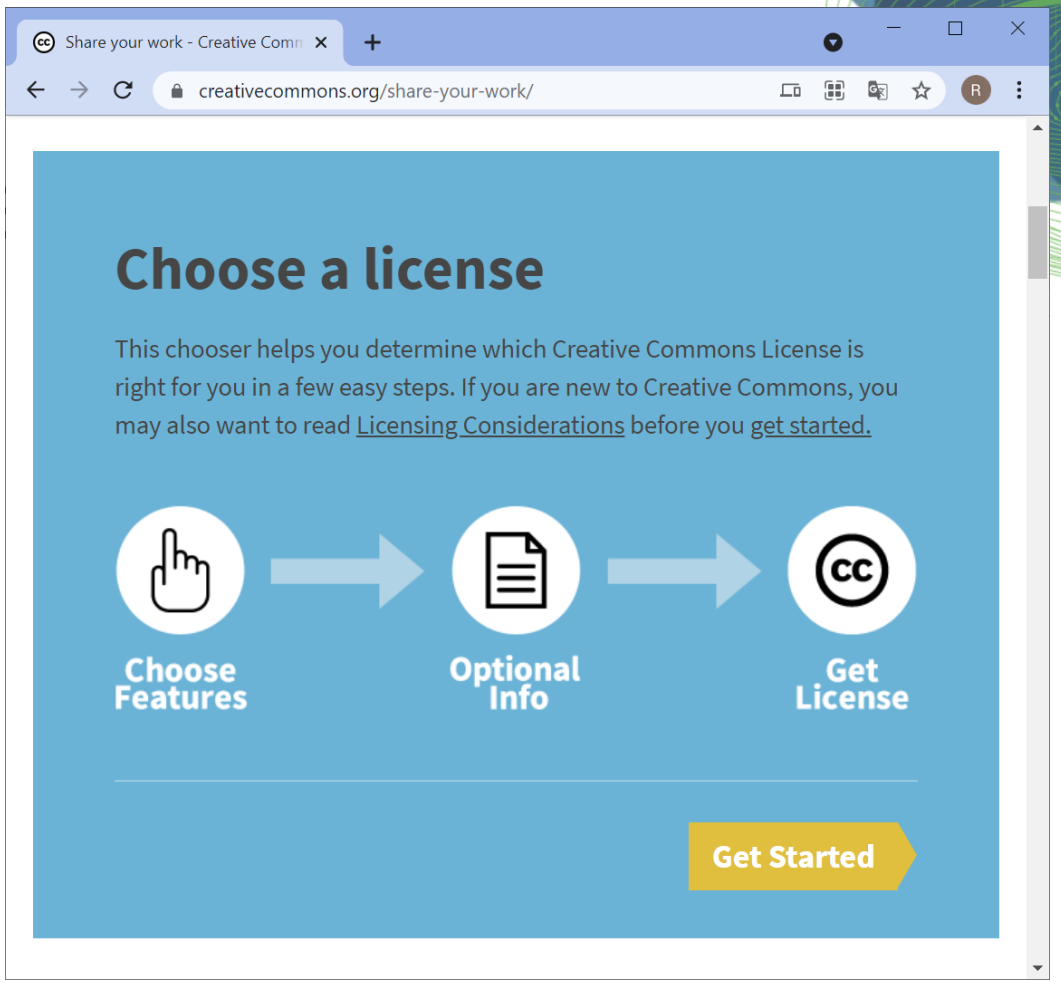
Also this seems to be changing thanks too new business models…

* Note: this may differ depending on the country. In the US, lecturer is a career path, in Germany not really.

Licensing: Creative Commons (CC)

- Public domain (CC0)
- Attribution International (CC-BY)
- Attribution ShareAlike Int. (CC-BY-SA)
- Attribution Non-Commercial Int. (CC-BY-NC)
- Attribution NoDerivatives Int. (CC-BY-ND)

+ Combinations, e.g. CC-BY-NC-ND



Licensing: Creative Commons (CC)

- **Public domain (CC0)**
- Everyone can reuse without mentioning the source or author of the shared resource.
- Public domain licenses cannot be revoked.
- The author must own the right to copy (copyright) the resource.
  - If you authored work as part of your job, you may not be the copyright holder. (check your employers' guidelines)

Employers don't like this one because you give away the rights to *exploit* your work.

Licensing: Creative Commons (CC)

"By attribution"

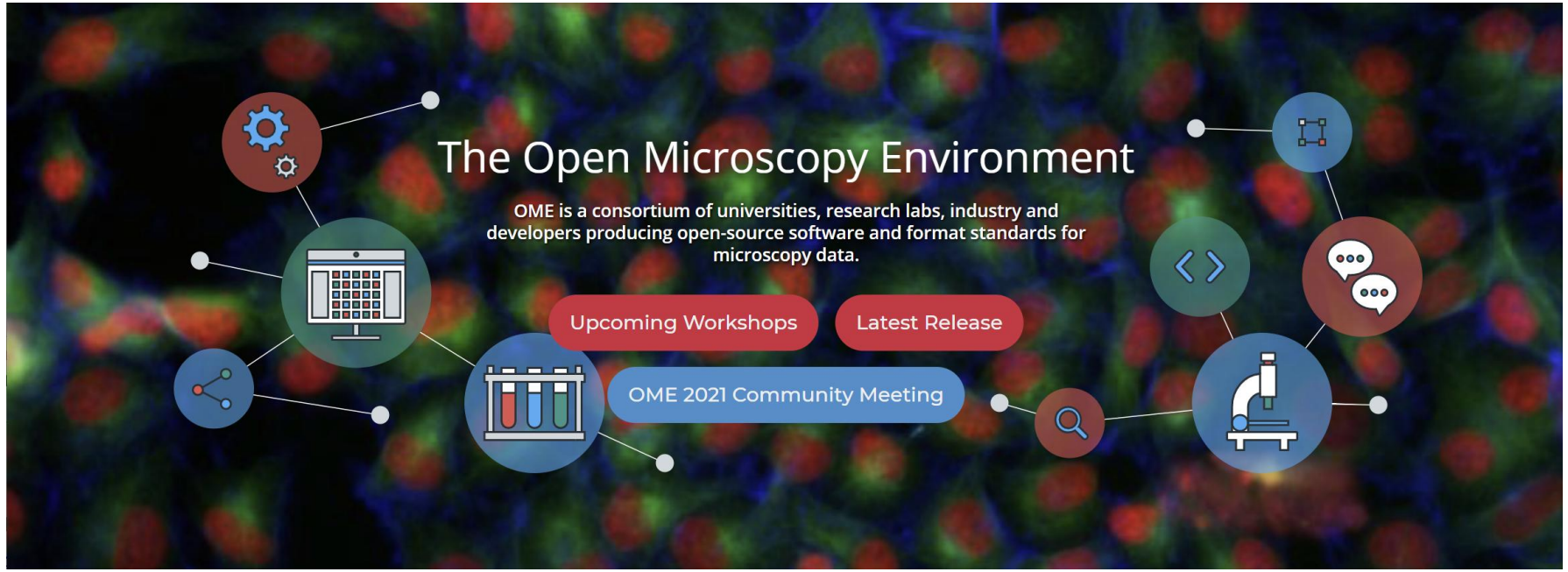
CC BY: This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use.

CC BY includes the following elements:

BY – Credit must be given to the creator

Licensing: Creative Commons (CC)

Example



You must put such a sentence and keep the link to CC-BY

Figure adapted from https://www.openmicroscopy.org/ licensed by University of Dundee & Open Microscopy Environment under Creative Commons Attribution 4.0 International License

Licensing: Creative Commons (CC)

"Share alike"

CC BY-SA: This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use. If you remix, adapt, or build upon the material, you must license the modified material under identical terms.

CC BY-SA includes the following elements:

BY – Credit must be given to the creator

SA – Adaptations must be shared under the same terms

"Restrictive" licensing

Licensing: Creative Commons (CC)

CC BY-NC: This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator.

It includes the following elements:

BY – Credit must be given to the creator

NC – Only noncommercial uses of the work are permitted

"Restrictive" licensing

Licensing: Creative Commons (CC)

CC BY-ND: This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, and only so long as attribution is given to the creator. The license allows for commercial use.

CC BY-ND includes the following elements:

BY – Credit must be given to the creator

ND – No derivatives or adaptations of the work are permitted

"Restrictive" licensing

Licensing: Creative Commons (CC)

CC BY-NC-ND: This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator.

CC BY-NC-ND includes the following elements:

BY – Credit must be given to the creator

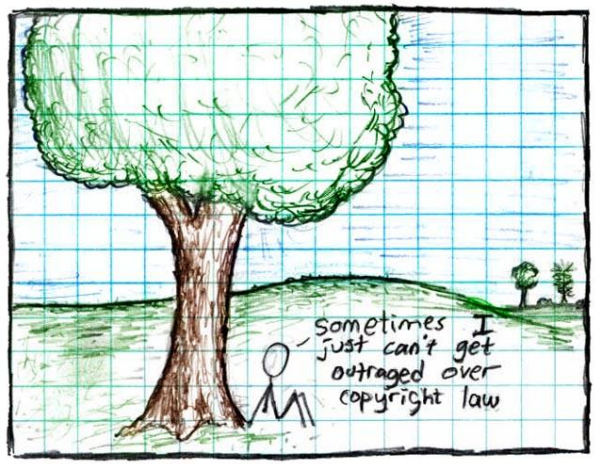
NC – Only noncommercial uses of the work are permitted

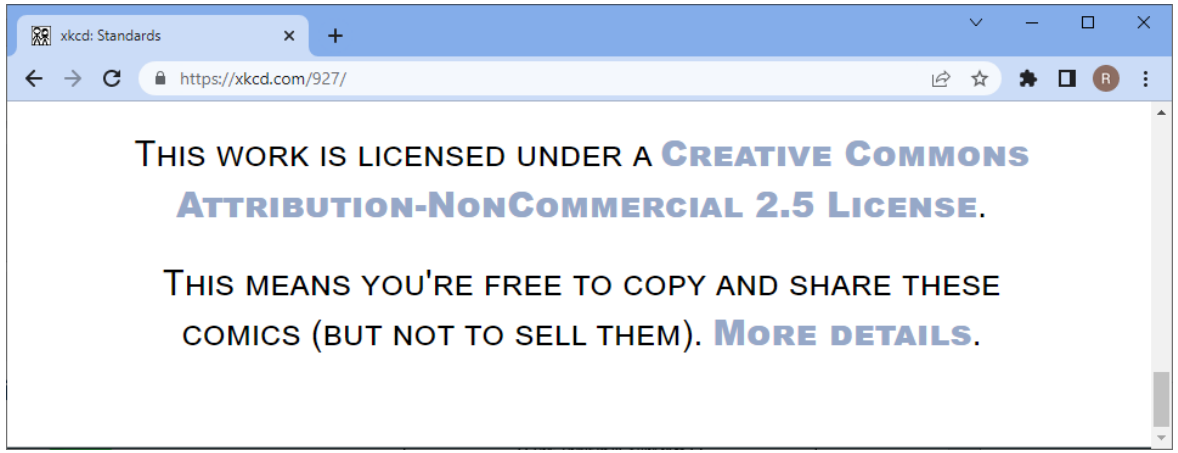
ND – No derivatives or adaptations of the work are permitted

"Restrictive" licensing

Quiz

- It’s ok to reuse Figures from XKCD.com if…





Mention author

Ask the authors

Link to the license

Copy the copyright statement

Source: https://xkcd.com/14/

Take home message

If you share material (openly or not)

license it

and it'll be harder to steal it

Summary

- If you want to make your stuff reusable:
  - Use permissive licenses
  - Share it on community-wide platforms (preferably not institutional servers)
  - Register your data and software in search-indices
- Read more:
  - Sharing on Zenodo
    https://focalplane.biologists.com/2023/02/15/sharing-research-data-with-zenodo/
  - Sharing on Figshare
    https://focalplane.biologists.com/2023/07/26/sharing-your-poster-on-figshare/
  - Collaborative work on github
    https://focalplane.biologists.com/2021/09/04/collaborative-bio-image-analysis-script-editing-with-git/
  - Licensing
    https://focalplane.biologists.com/2023/05/06/if-you-license-it-itll-be-harder-to-steal-it-why-we-should-license-our-work/

Acknowledgements

Communities & platforms

image.sc

NFDI4 BIOIMAGE

GloBIAS Global BioImage Analysts' Society

neubias network of european bioimage analysts

FocalPlane Where biology meets microscopy

SaxFDM

BIDS team UL

Lea Gihlein, Lea Kabjesz, Mara Lampert

Funding

Bundesministerium für Bildung und Forschung

SACHSEN

Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Chan Zuckerberg Initiative

DFG Deutsche Forschungsgemeinschaft

Some Figures were generated using DALL·E