ScaDS.AI
DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

Cultivating Open Training

Robert Haase

GEFÖRDERT VOM

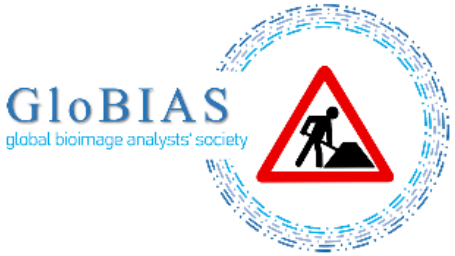
Bundesministerium für Bildung und Forschung

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.

Career update: Cultivating Open Training

- Since 2023: PI of the “Global BioImage Analysts' Society (GloBIAS)”, formerly known as the Society for Knowledge Exchange in BioImage Analysis



- Since 2023: Task area lead “Training and community integration” of the German National Research Data Management Infrastructure for BioImaging NFDI4BioImage

- Since 2024: Lecturer and training coordinator at the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI / Leipzig University)

Closed science

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

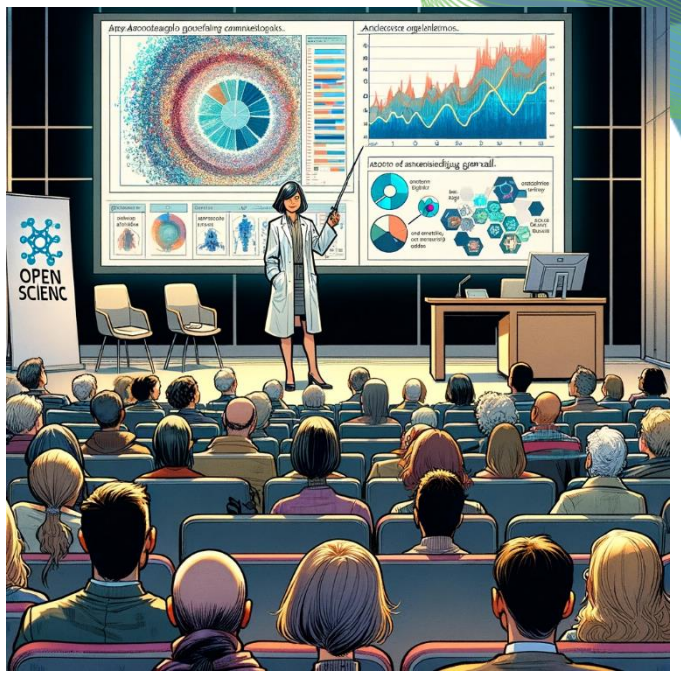
- Reasons
  - Risk of being scooped
  - Fear of blaming oneself (imposter syndrome)
  - Lack of awareness (who is allowed to publish *my work*?)

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



Am I allowed to publish my stuff?

Define responsibilities and procedures early!

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

"Robert will do this by September!"

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.

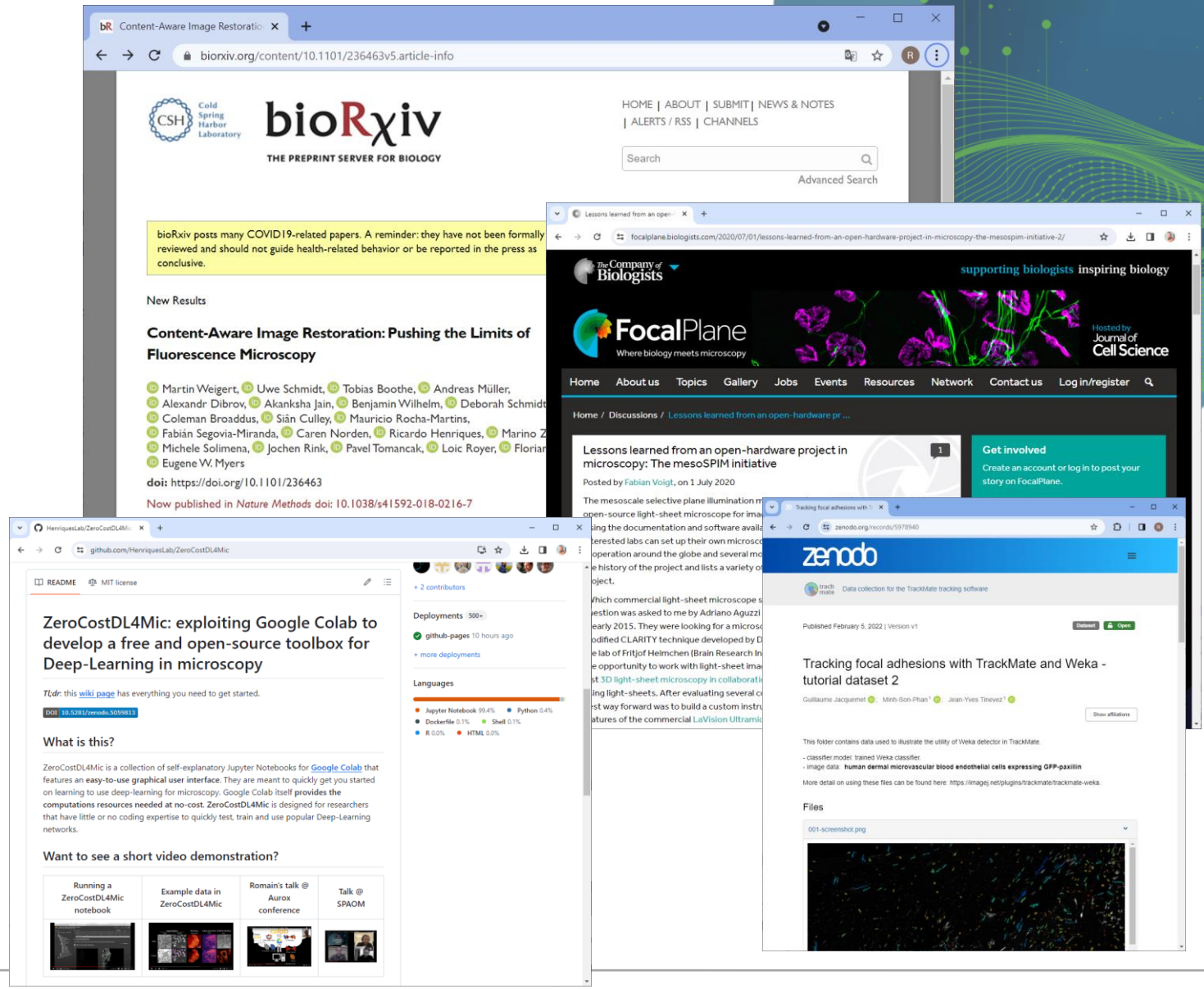
Are we going to publish data / materials / code?

What license can we use?

Deciding by the end of the project is too late!

Where to share?

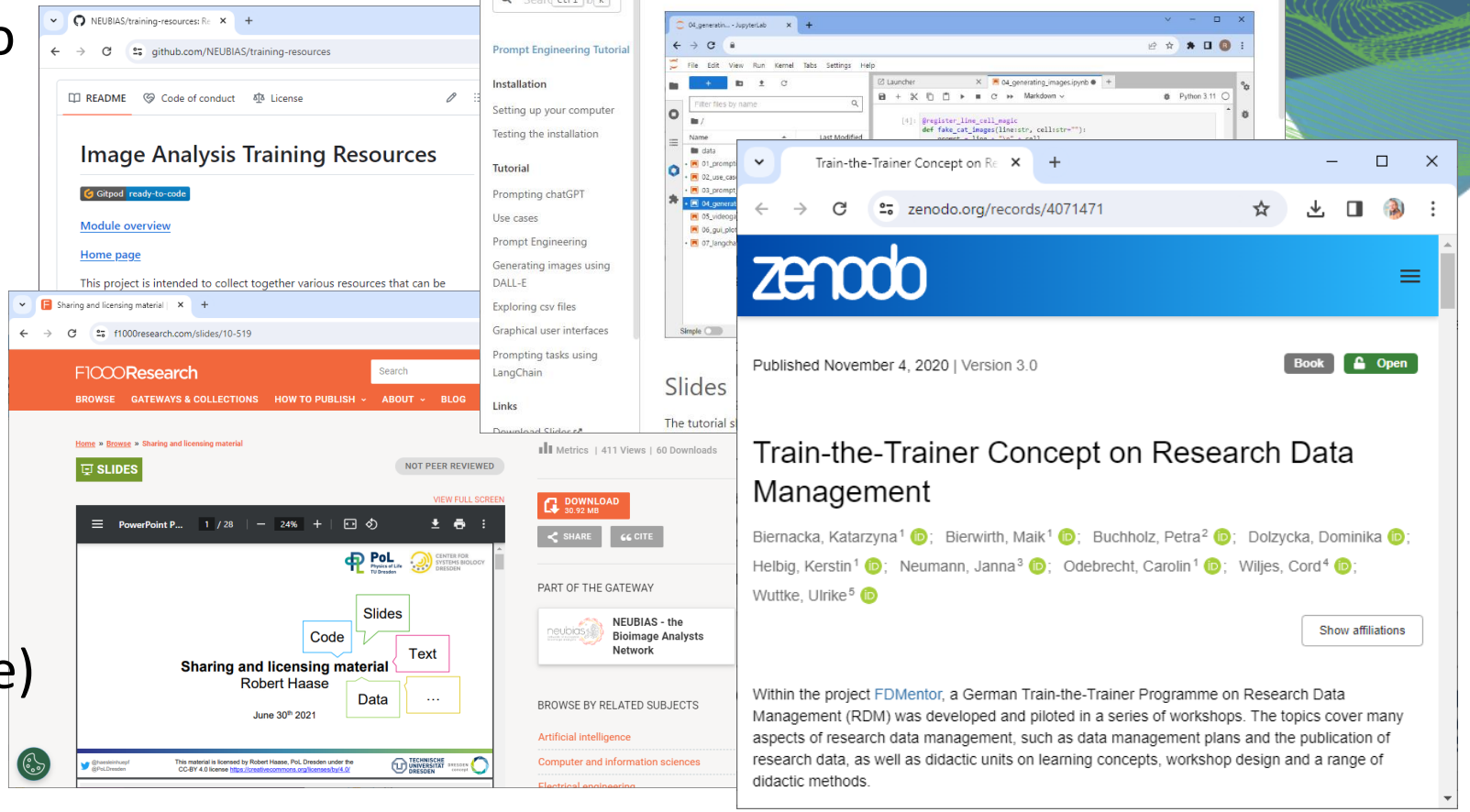
- Open *science* related content
  - bioRxiv (manuscripts, no reviews)
  - Figshare
  - F1000
  - Bioimage Archive (data)
  - Github (code)
  - Zenodo
  - Focalplane
  - Institutional servers (if there is no alternative)



Where to share?

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

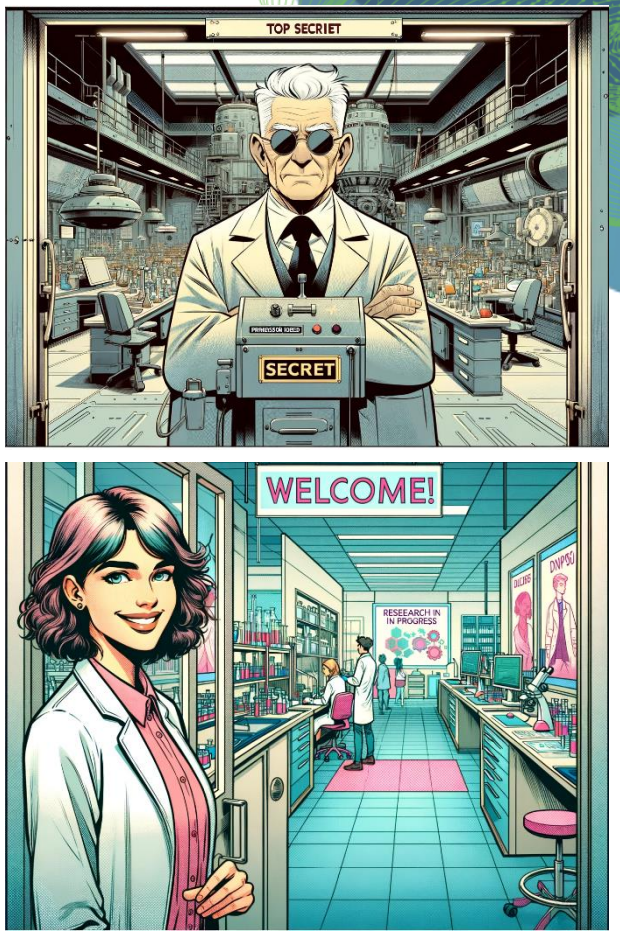
Github pages 🤯



Licensing: Permissive versus restrictive

- Restrictive
  - You can reuse our stuff, but only if you …
    - License your work with the same license we do
    - Make your stuff openly available
    - Make no money with derivatives of our work
  - Examples: GPL, CC-BY-SA, CC-BY-NC, CC-BY-ND

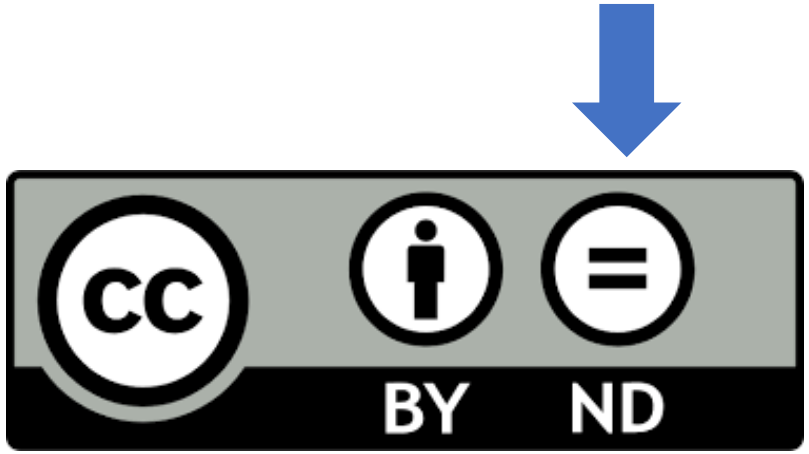
I conclude, these are less *open* in a sense



- Permissive licensing:
  - Do whatever you like with our stuff, just make sure to mention / cite us …
  - Examples: BSD, MIT, Apache, CC-BY

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"

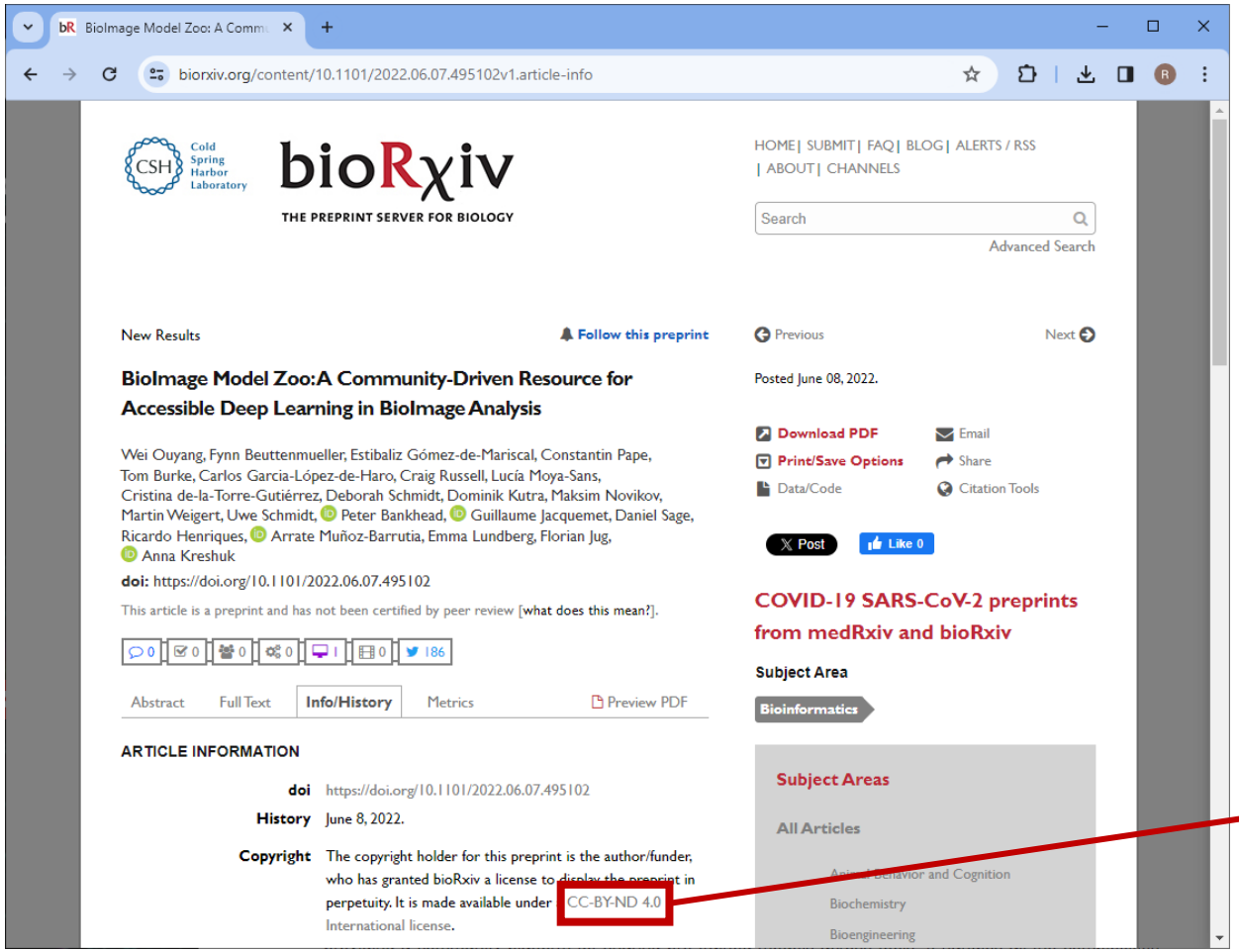
Licensing: Permissive versus restrictive

I hope nobody feels hurt
by the following slides.

I just would like to make a point.

Licensing: Permissive versus restrictive

Example

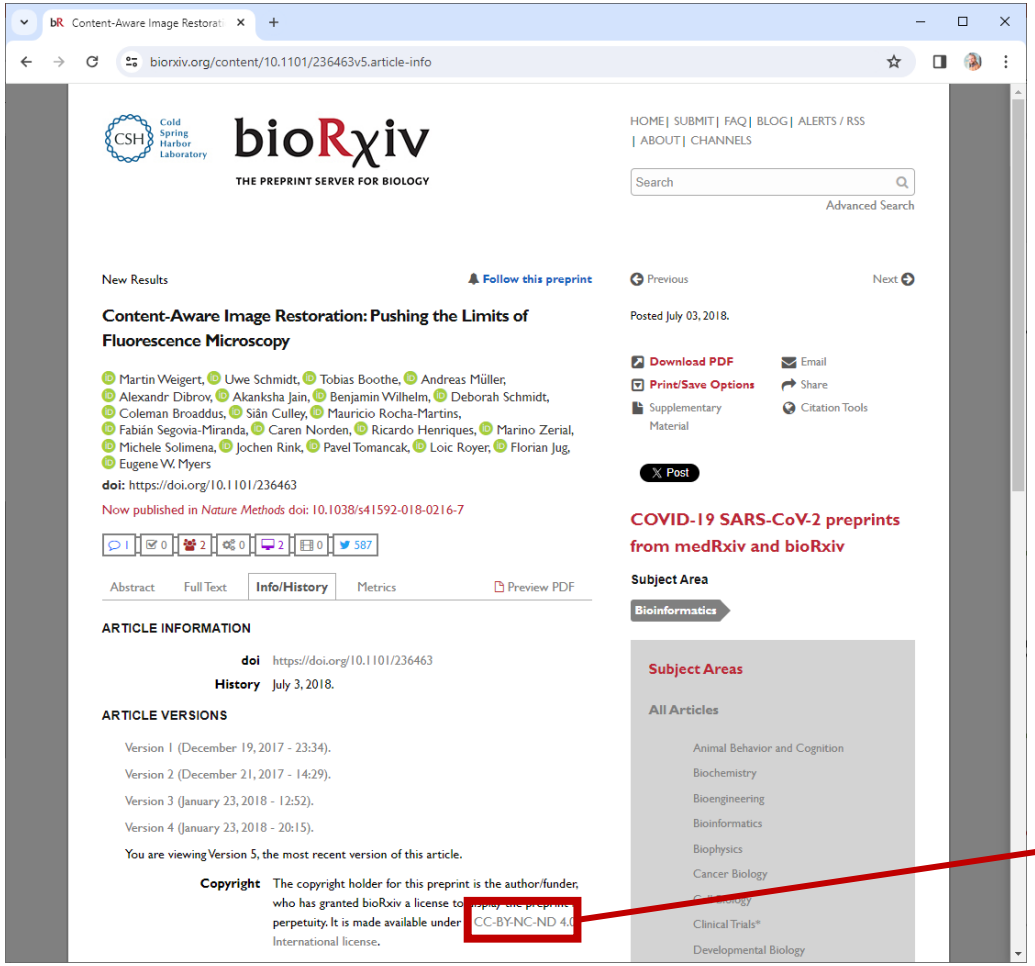


I would love to show you a Figure from this paper!

But I'm not allowed!

Licensing: Permissive versus restrictive

Example

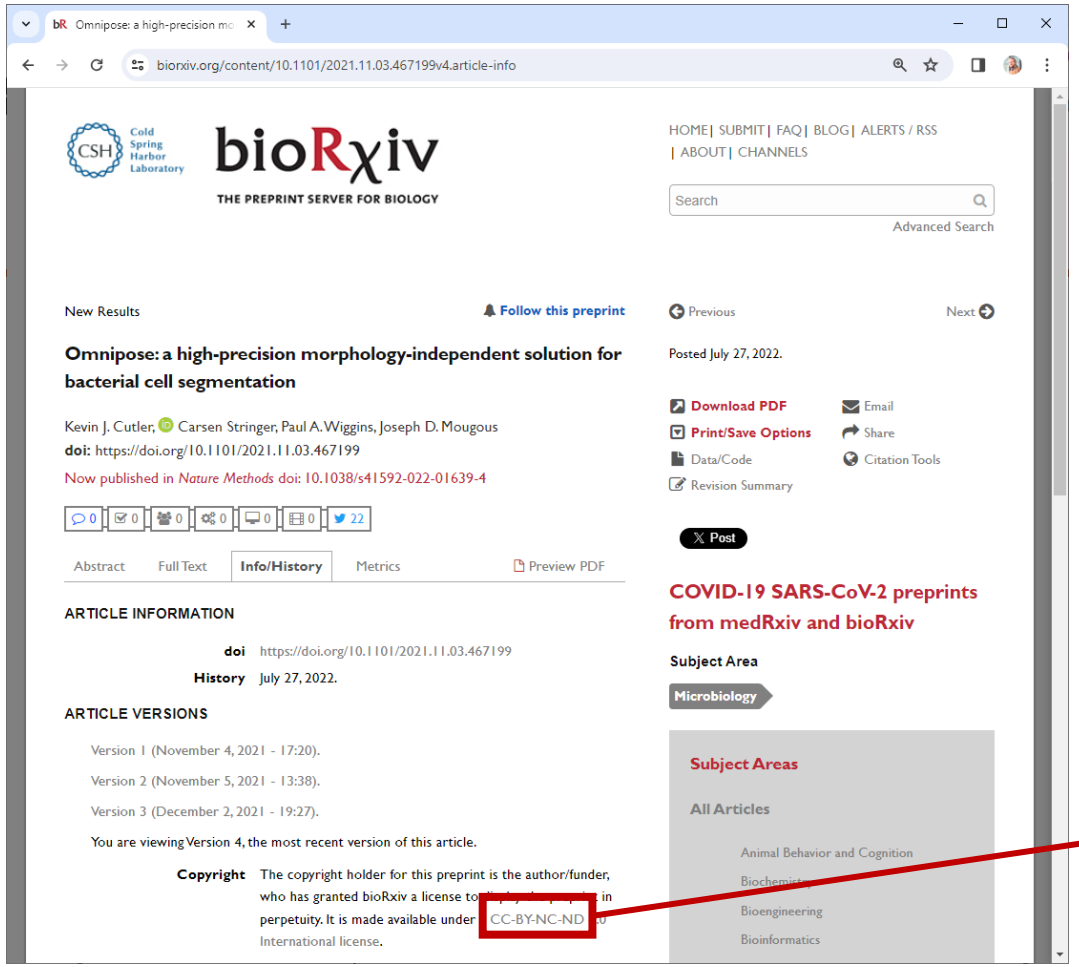


I would love to show you a Figure from this paper!

But I'm not allowed!

Licensing: Permissive versus restrictive

Example

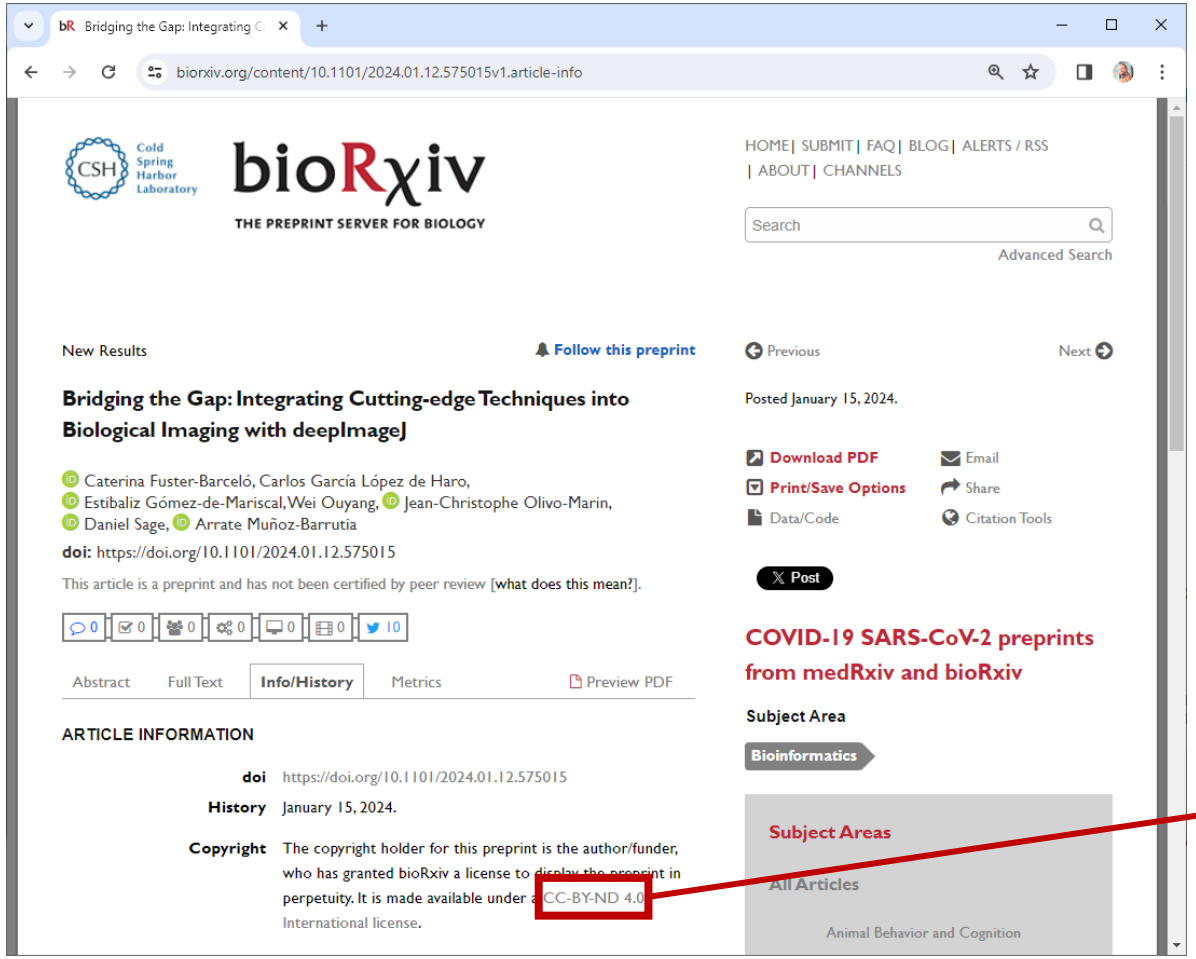


I would love to show you a Figure from this paper!

But I'm not allowed!

Licensing: Permissive versus restrictive

Example

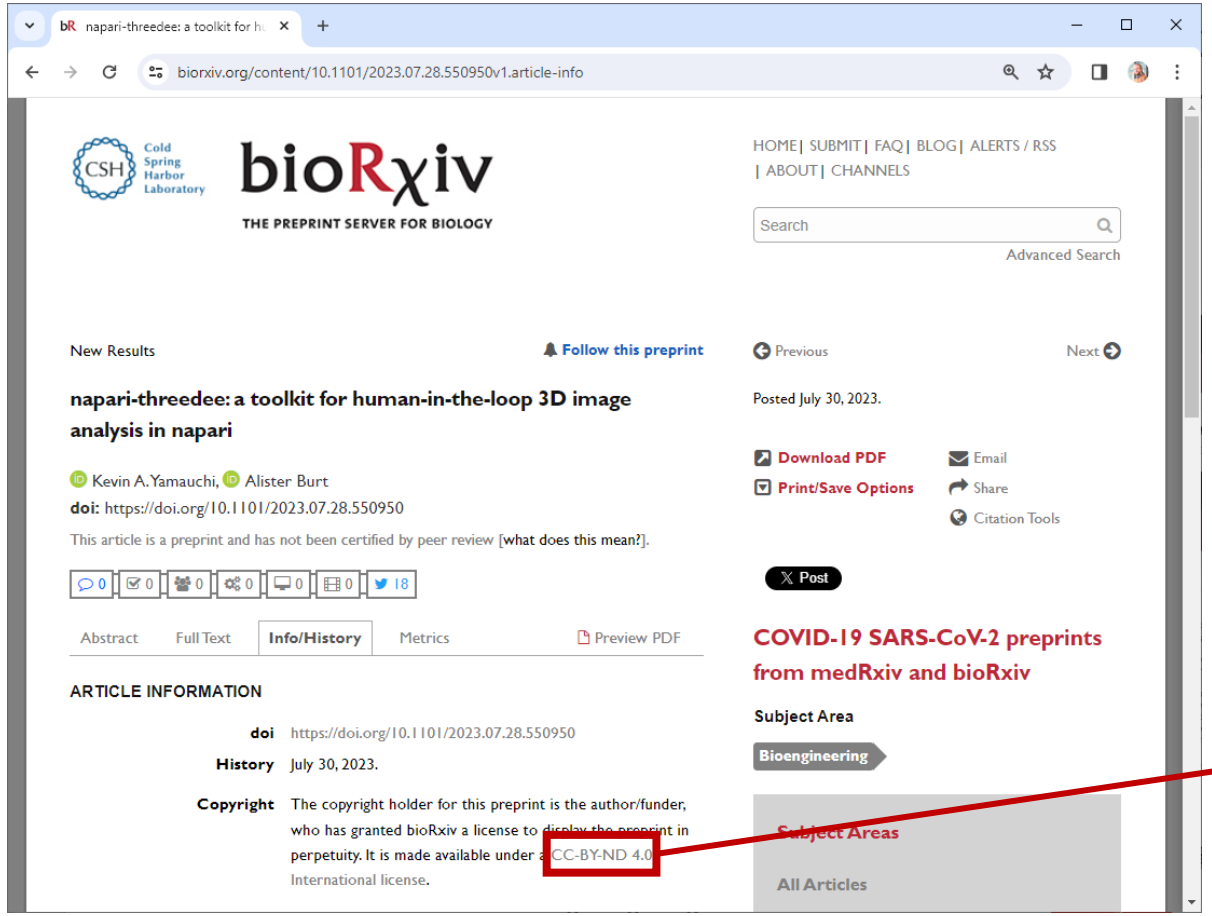


I would love to show you a Figure from this paper!

But I'm not allowed!

Licensing: Permissive versus restrictive

Example

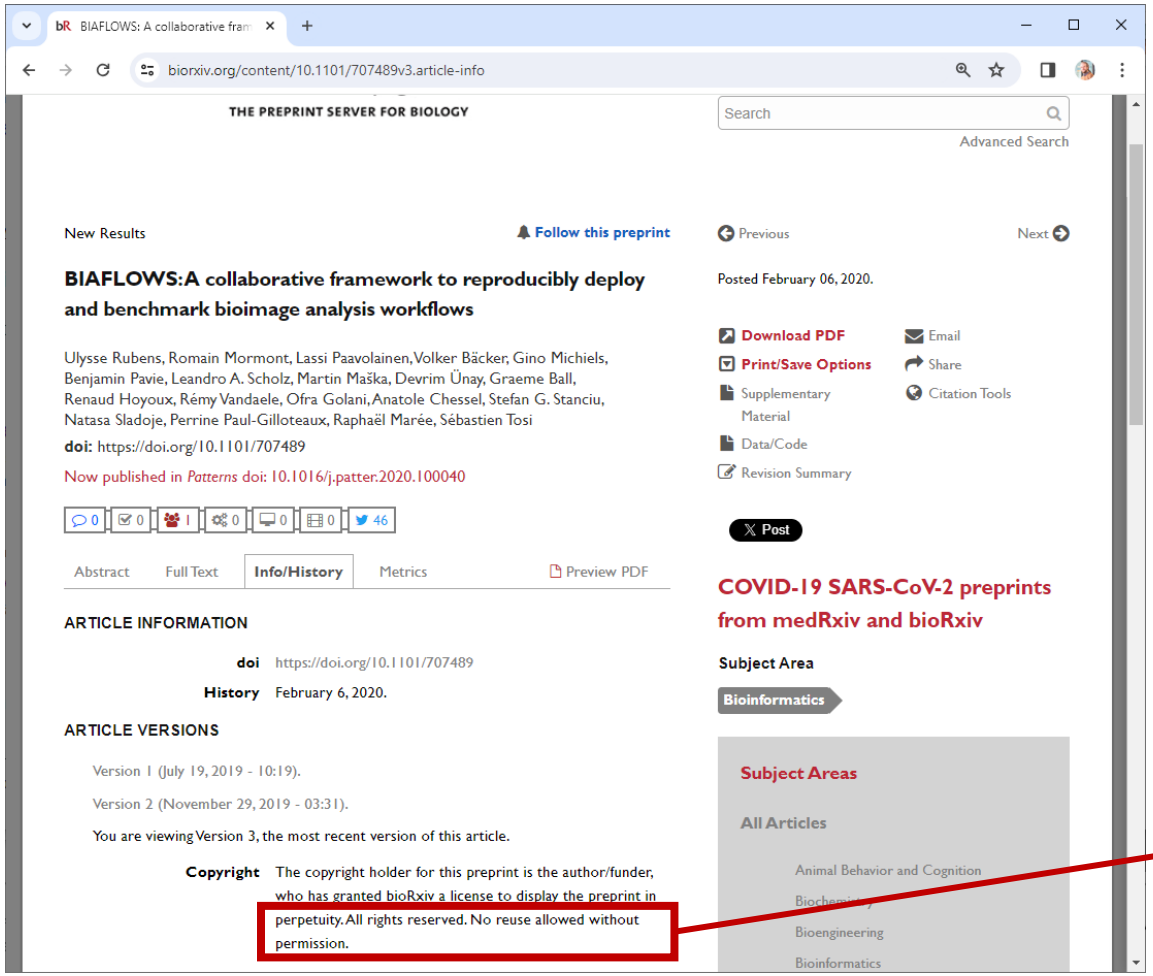


I would love to show you a Figure from this paper!

But I'm not allowed!

Licensing: Permissive versus restrictive

Example

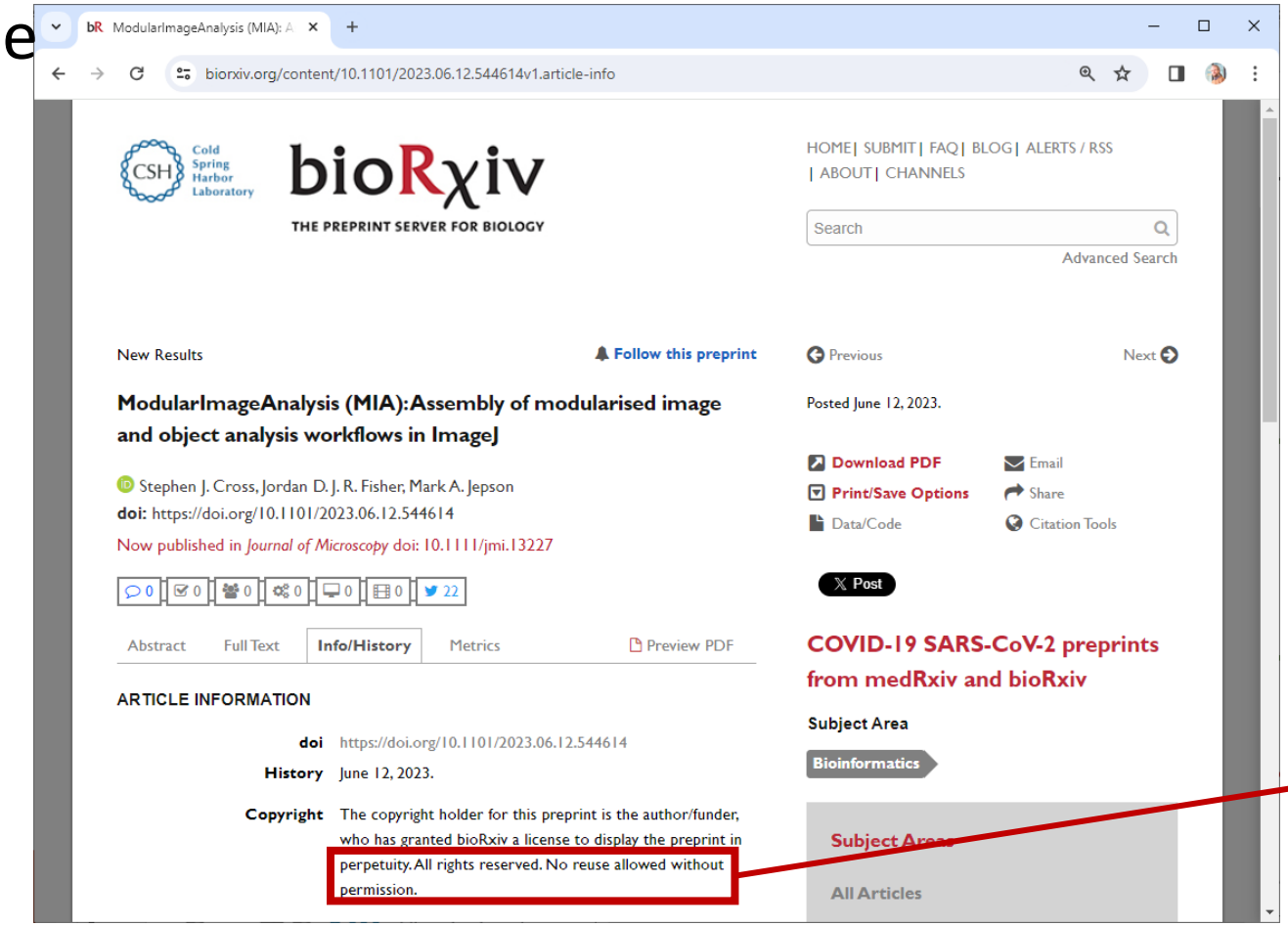


I would love to show you a Figure from this paper!

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

Restrictive licensing is
a community-wide issue.

I *presume* due to lack of awareness & training

Train the trainers!

Licensing: Permissive versus restrictive

	Download and share for free	Reuse parts, e.g. Figures	Reuse parts, e.g. in paid training
CC-BY	✓	✓	✓
CC-BY-SA	✓ Only under CC-BY-SA	✓ Only under CC-BY-SA	✓ Only under CC-BY-SA
CC-BY-NC	✓	✓ (if free of charge)	✗
CC-BY-ND	✓	✗	✗
CC-BY-NC-ND	✓	✗	✗

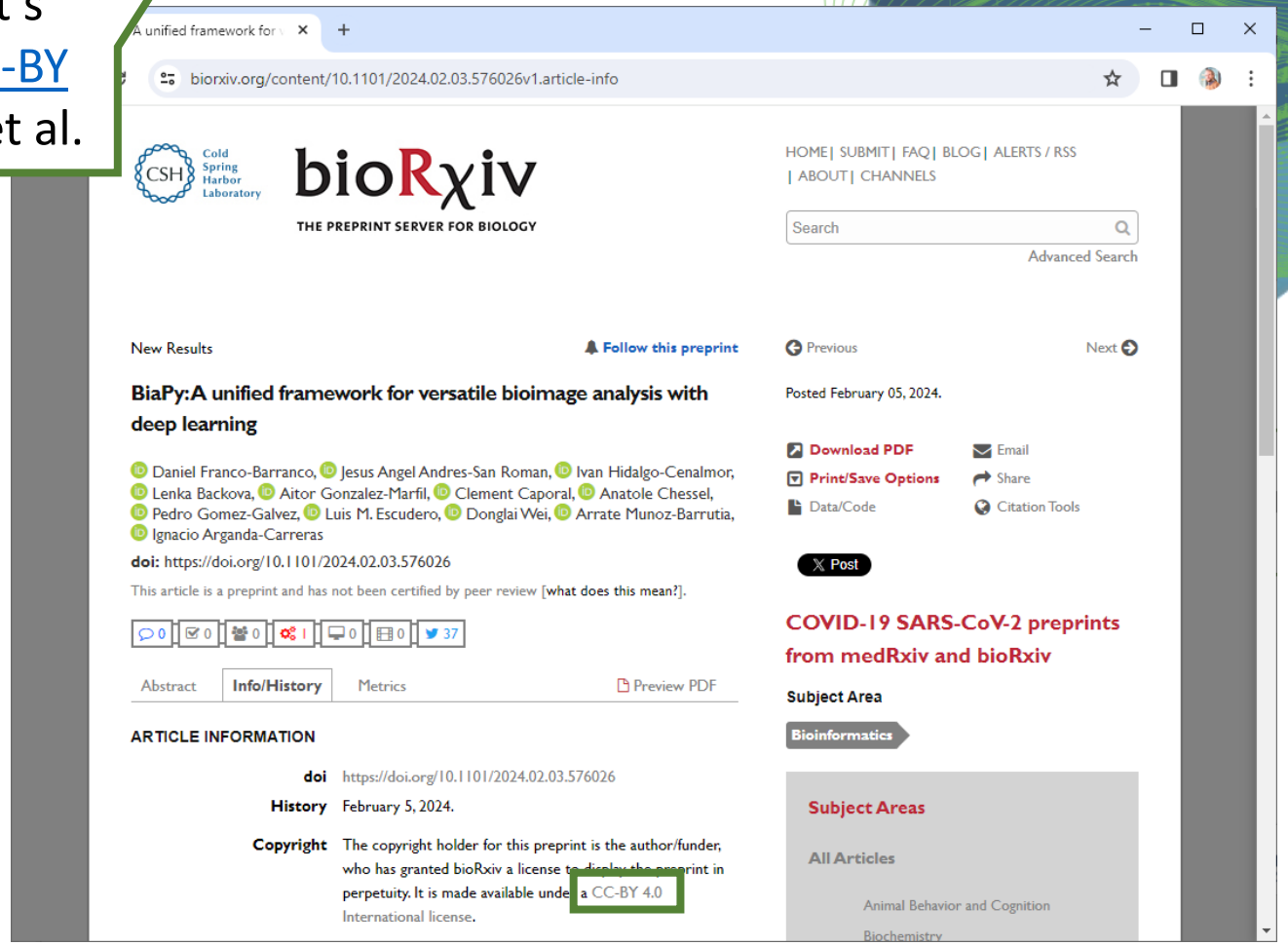
Bad for the progress of science

In particular in the context of training

Licensing: Permissive versus restrictive

Example

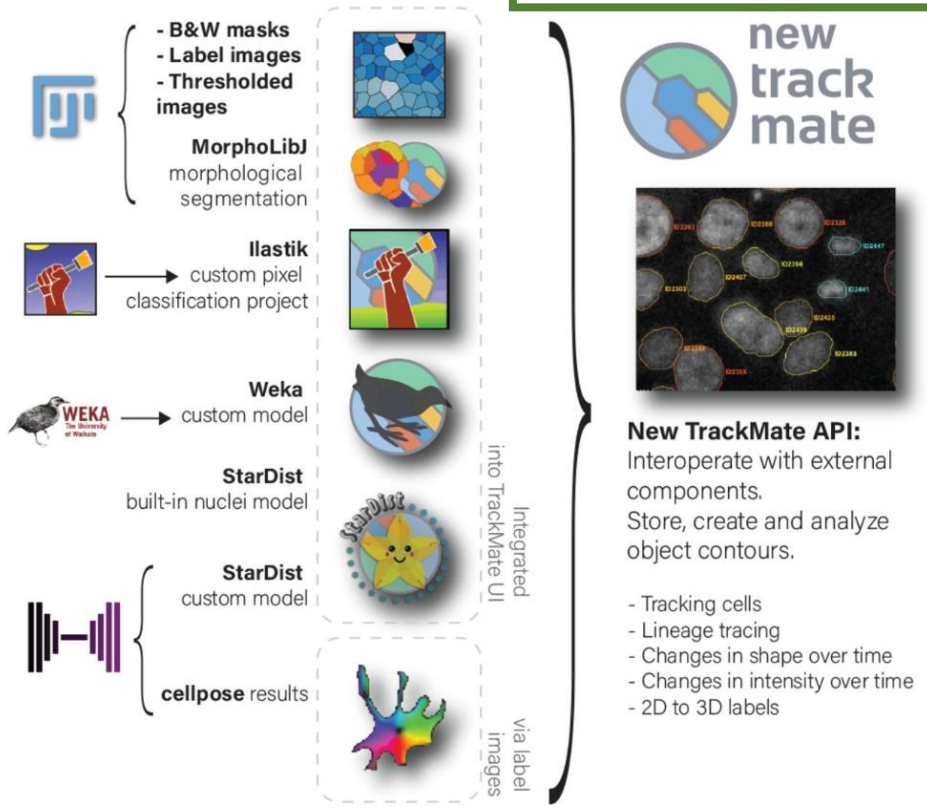
Look at this great figure! It's cropped from / licensed CC-BY 4.0 by D. Franco-Barranco et al.

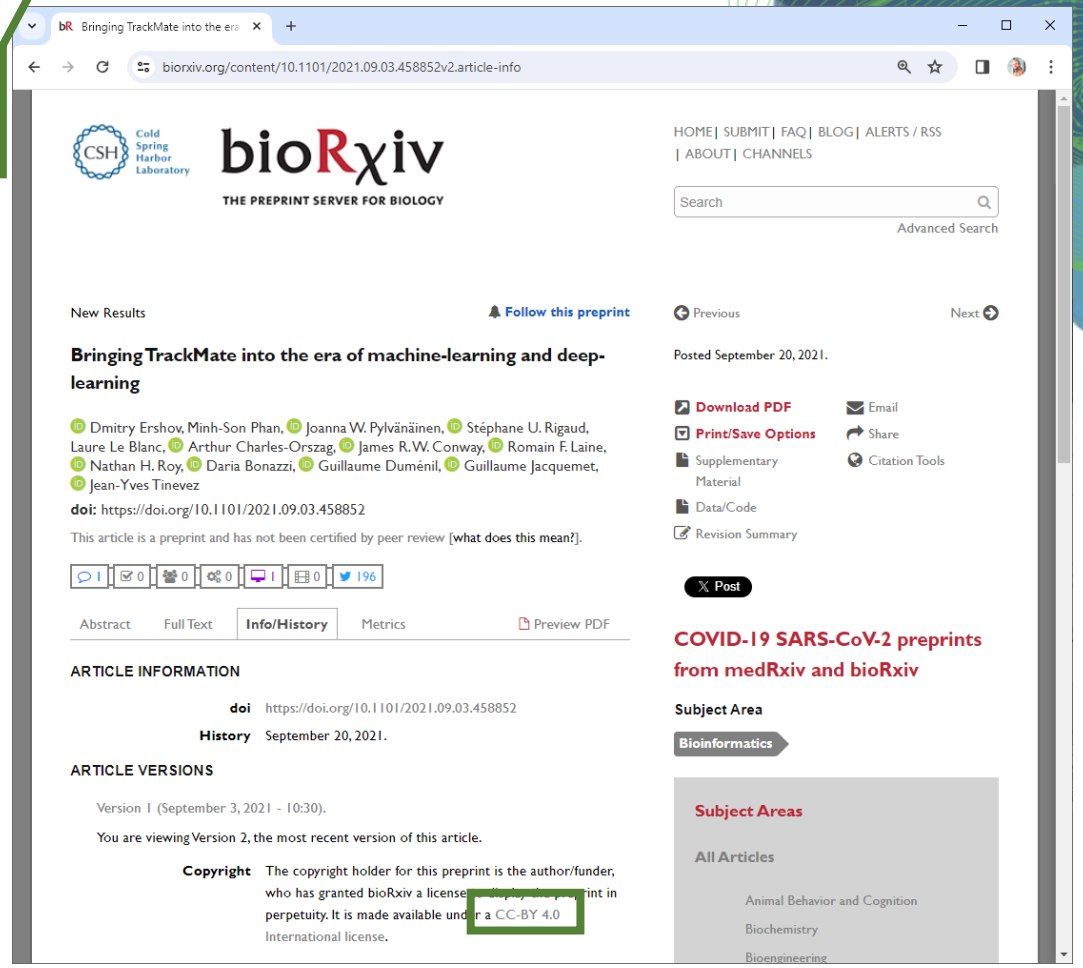


Licensing: Permissive versus restrictive

Example

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

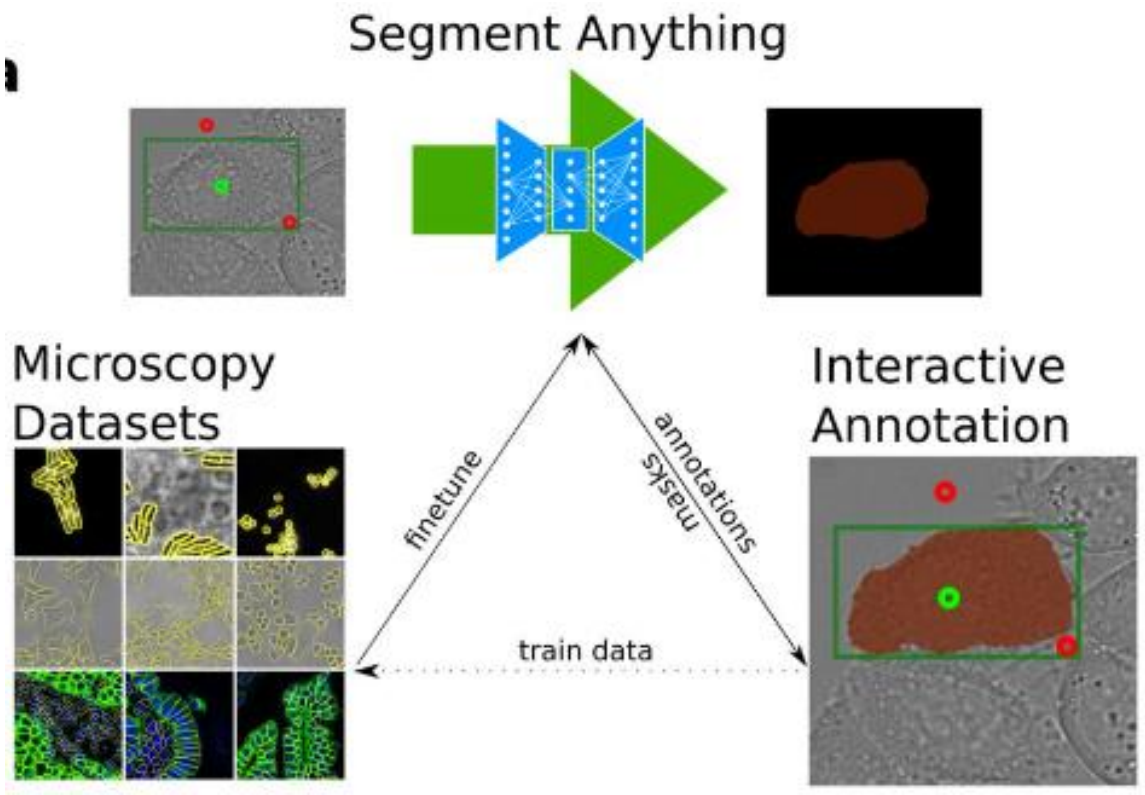


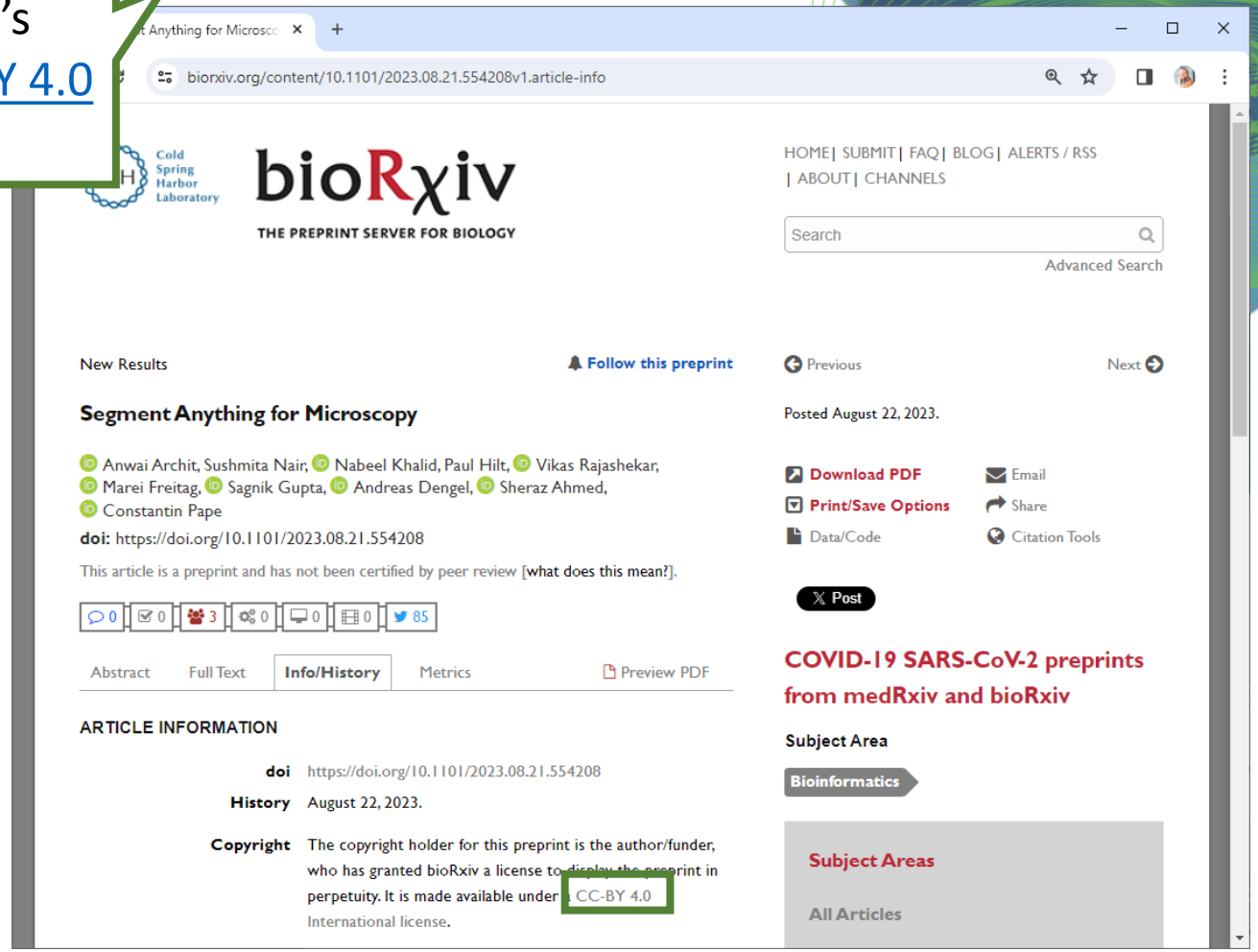


Licensing: Permissive versus restrictive

Example

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

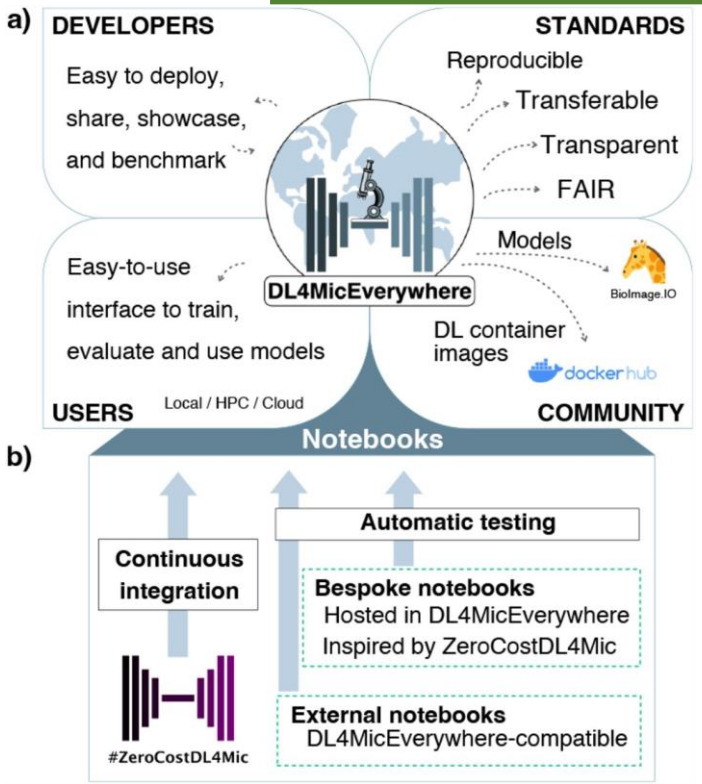


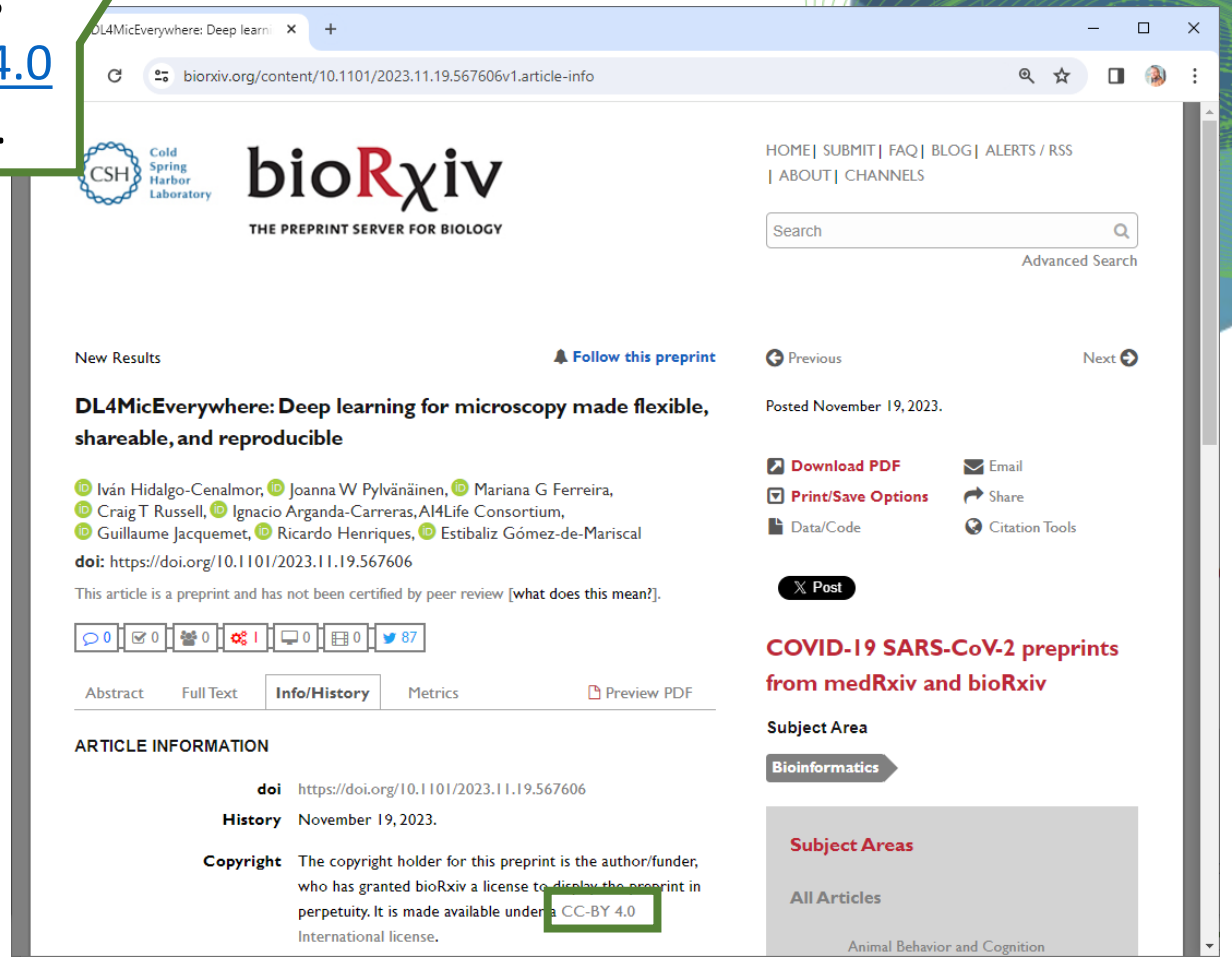


Licensing: Permissive versus restrictive

Example

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

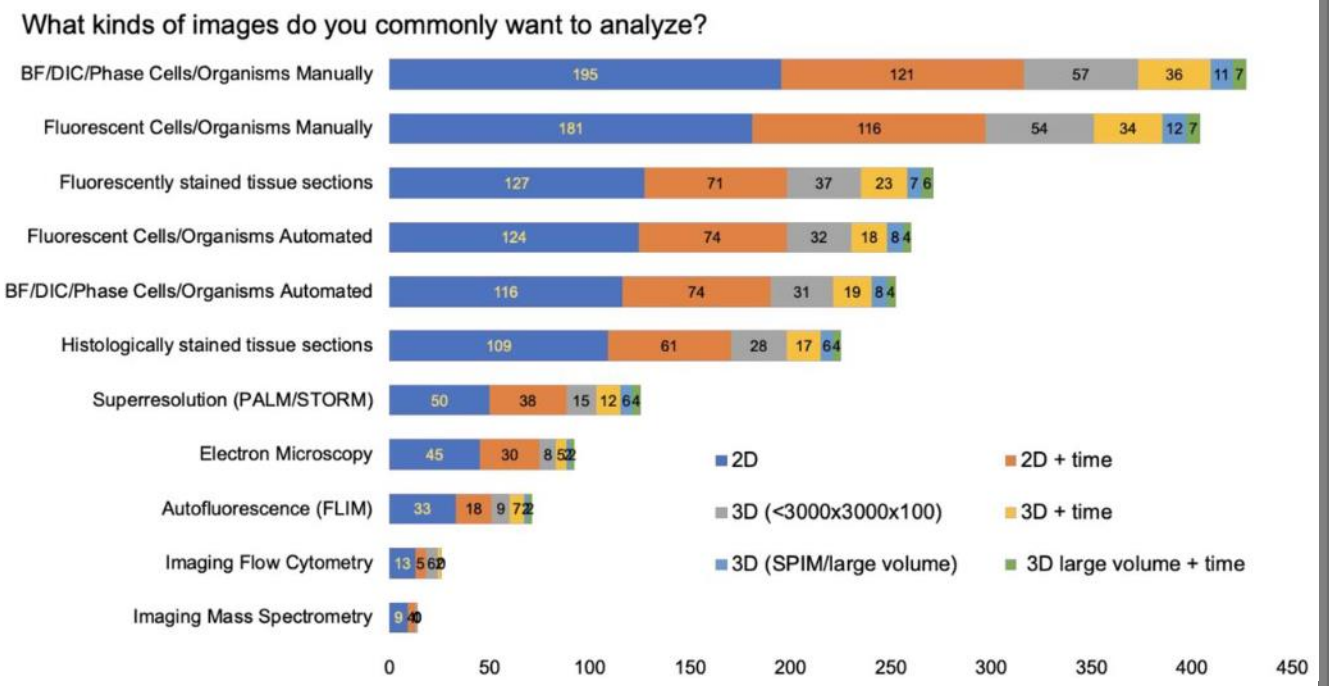


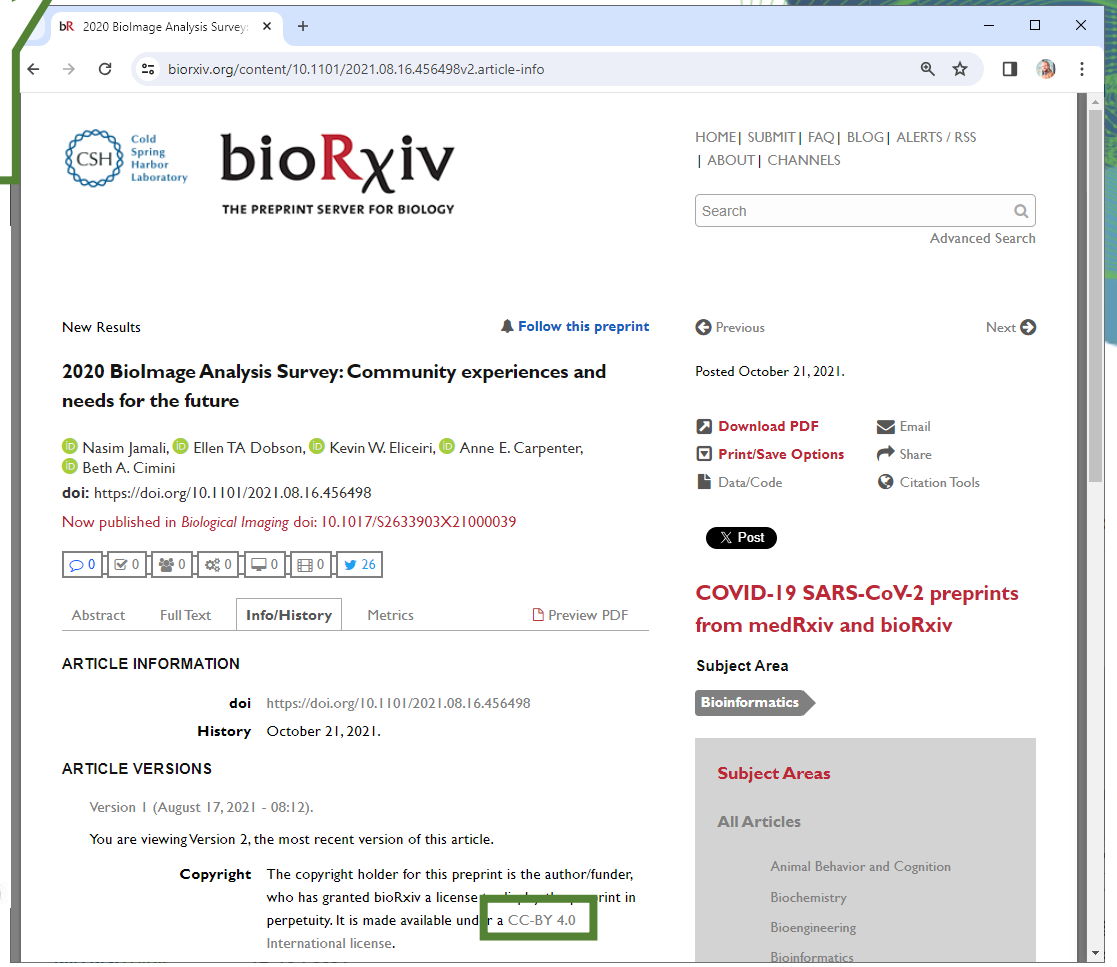


Licensing: Permissive versus restrictive

Example

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

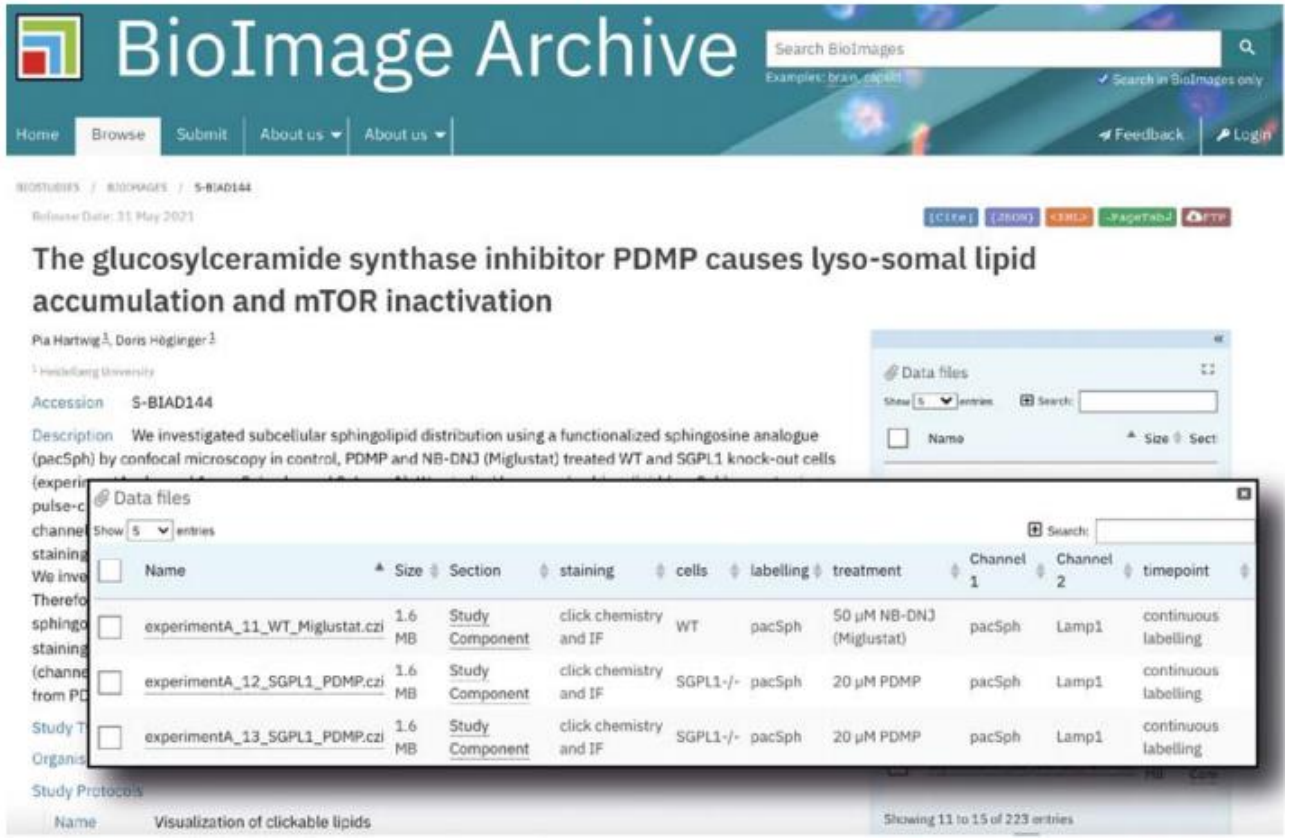




Licensing: Permissive versus restrictive

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 $$

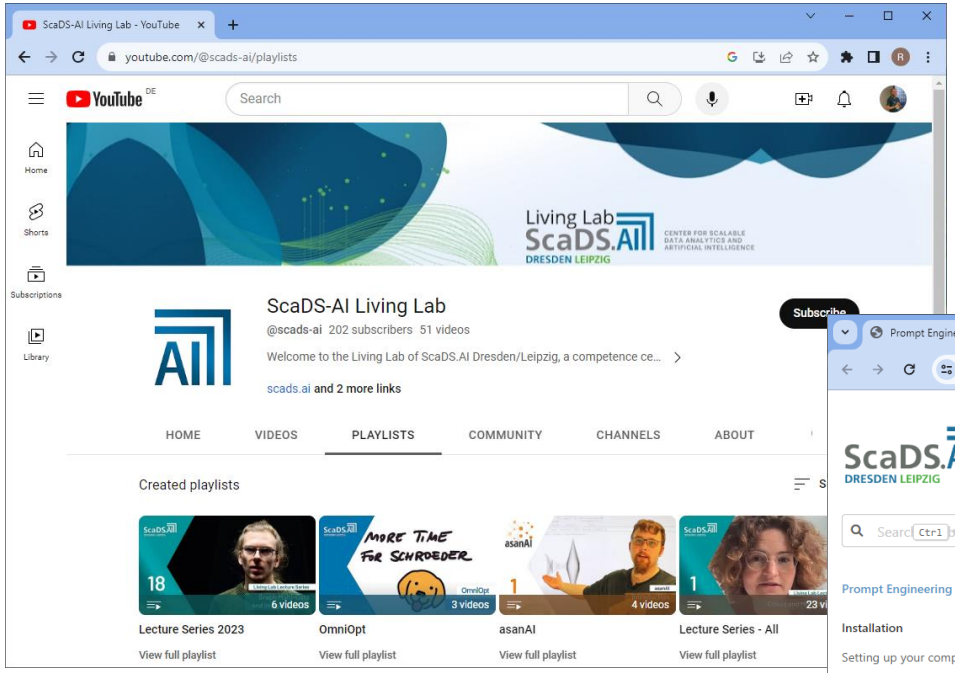
Also this seems to be changing thanks too new business models…

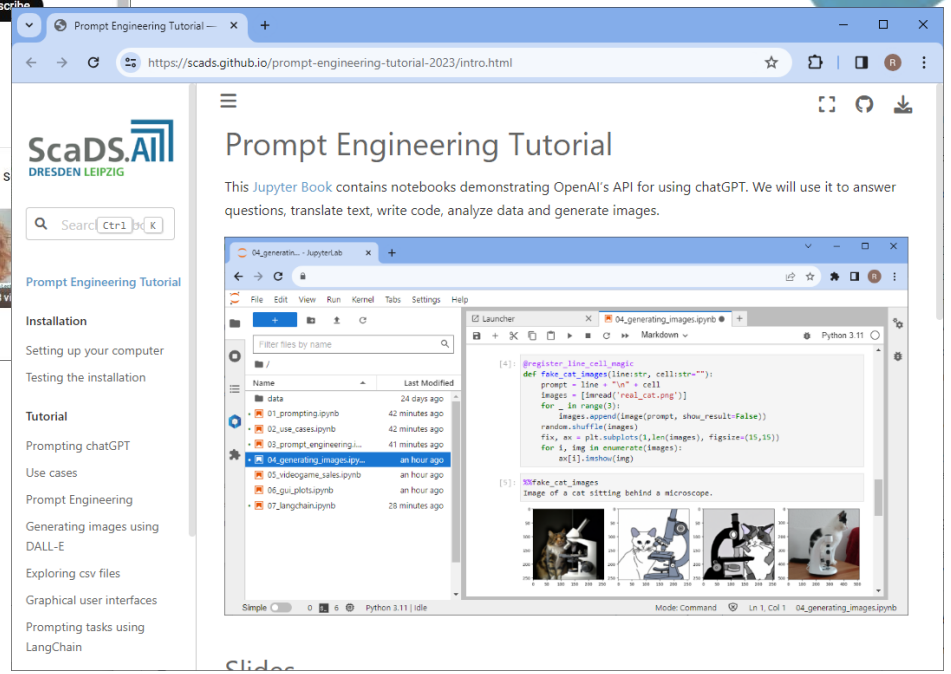
* Note: this may differ depending on the country. In the US, lecturer is a career path, in Germany not really.

Incentives: Visibility

YouTube, Github,…

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable





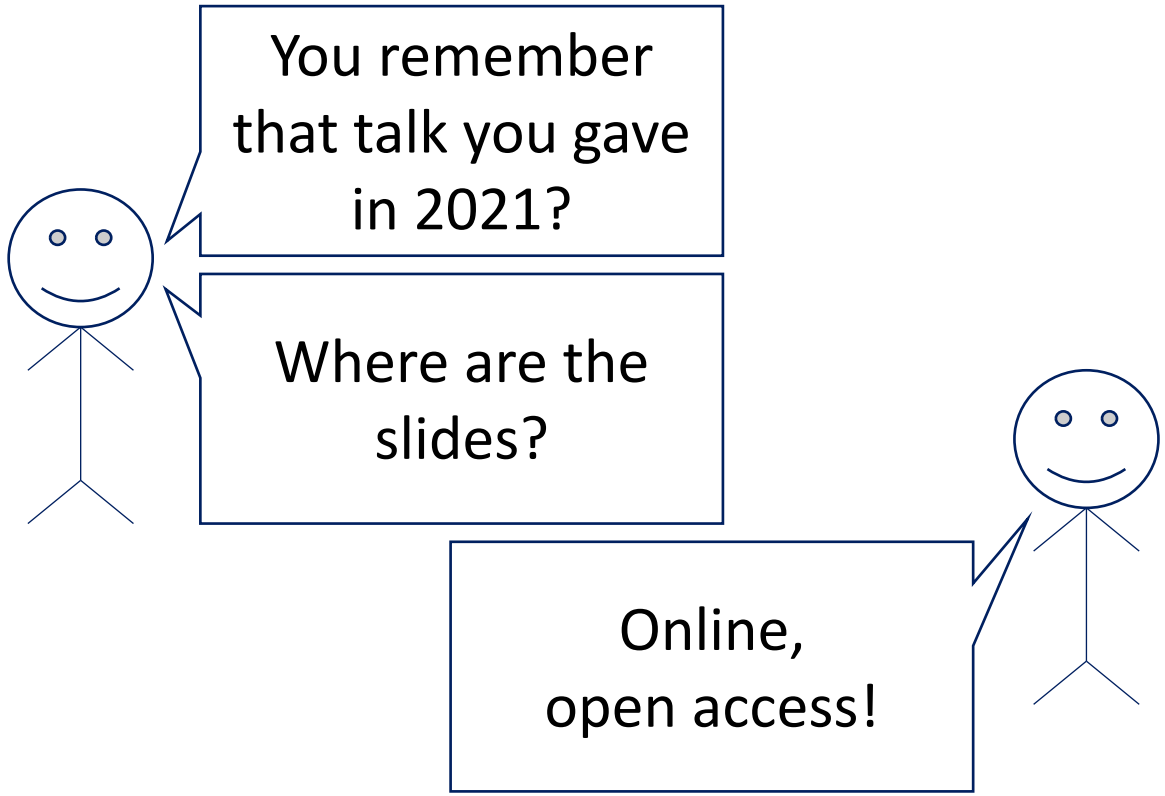
Open & FAIR Training is a PR instrument

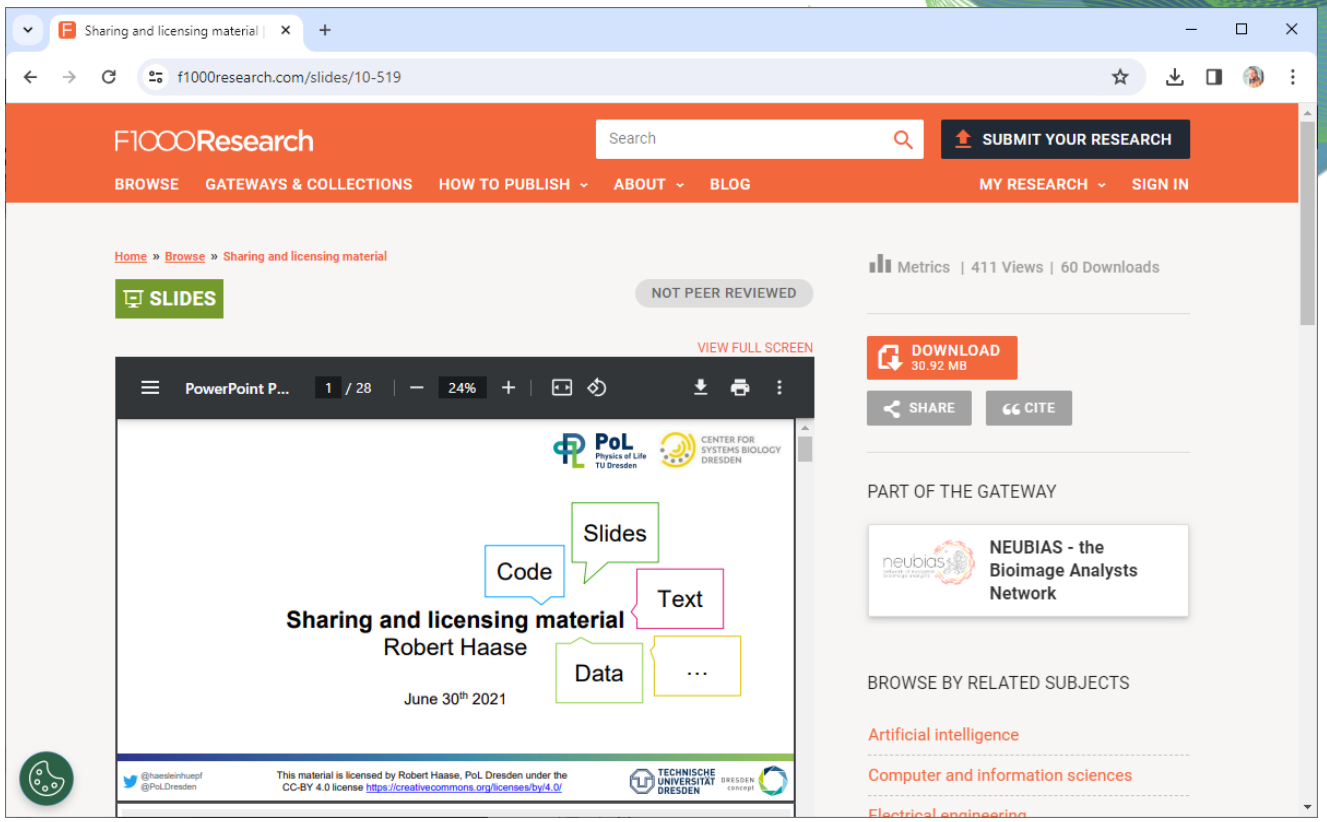
… leading to

- more course attendees
- new collaborations

Incentives: Findability

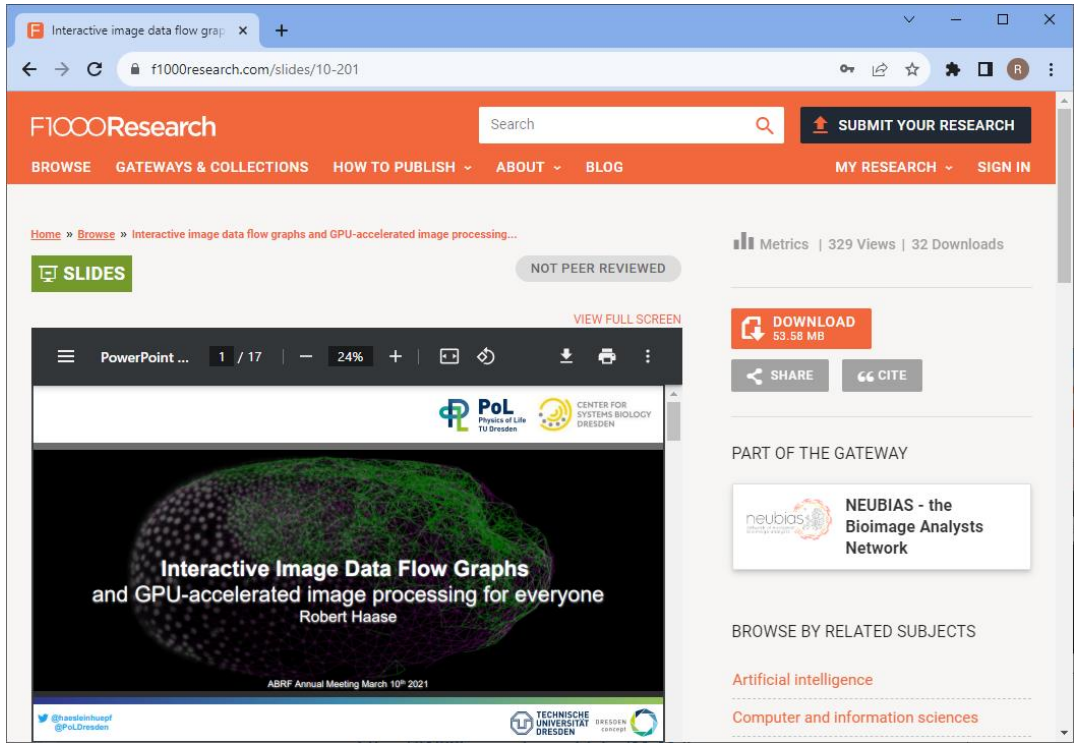
Your *future-self* will thank you, because they will find your work

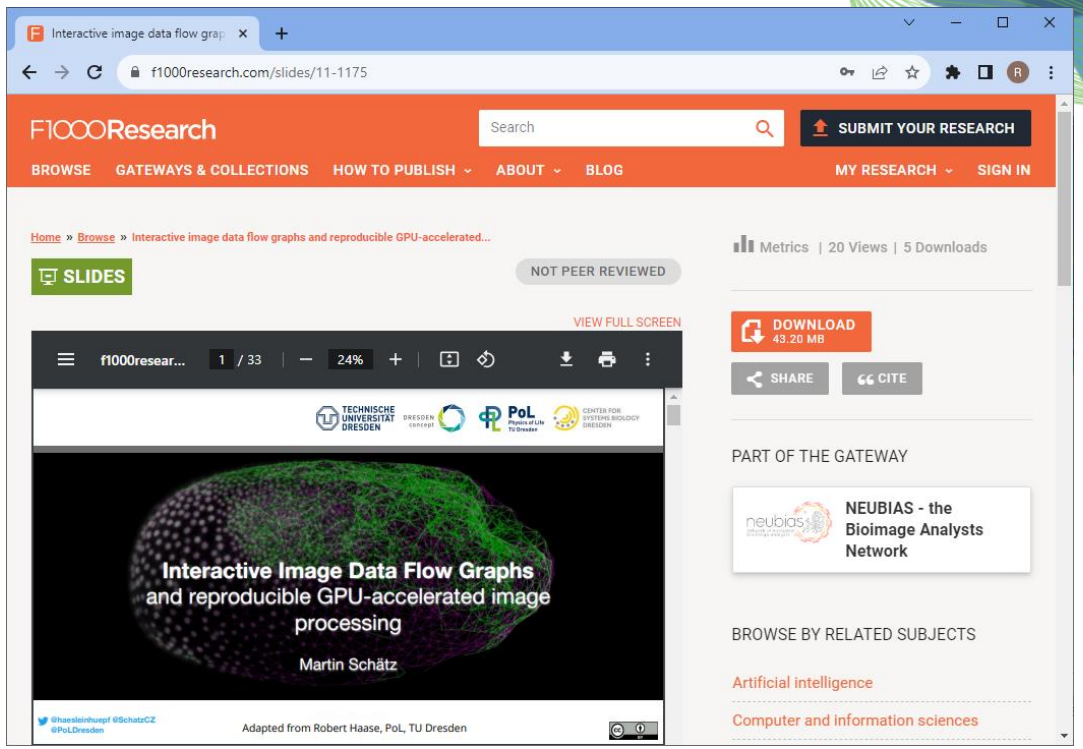




Incentives: Reusability

- Open Access -> Others teach how to use <u>your</u> *tools & methods*



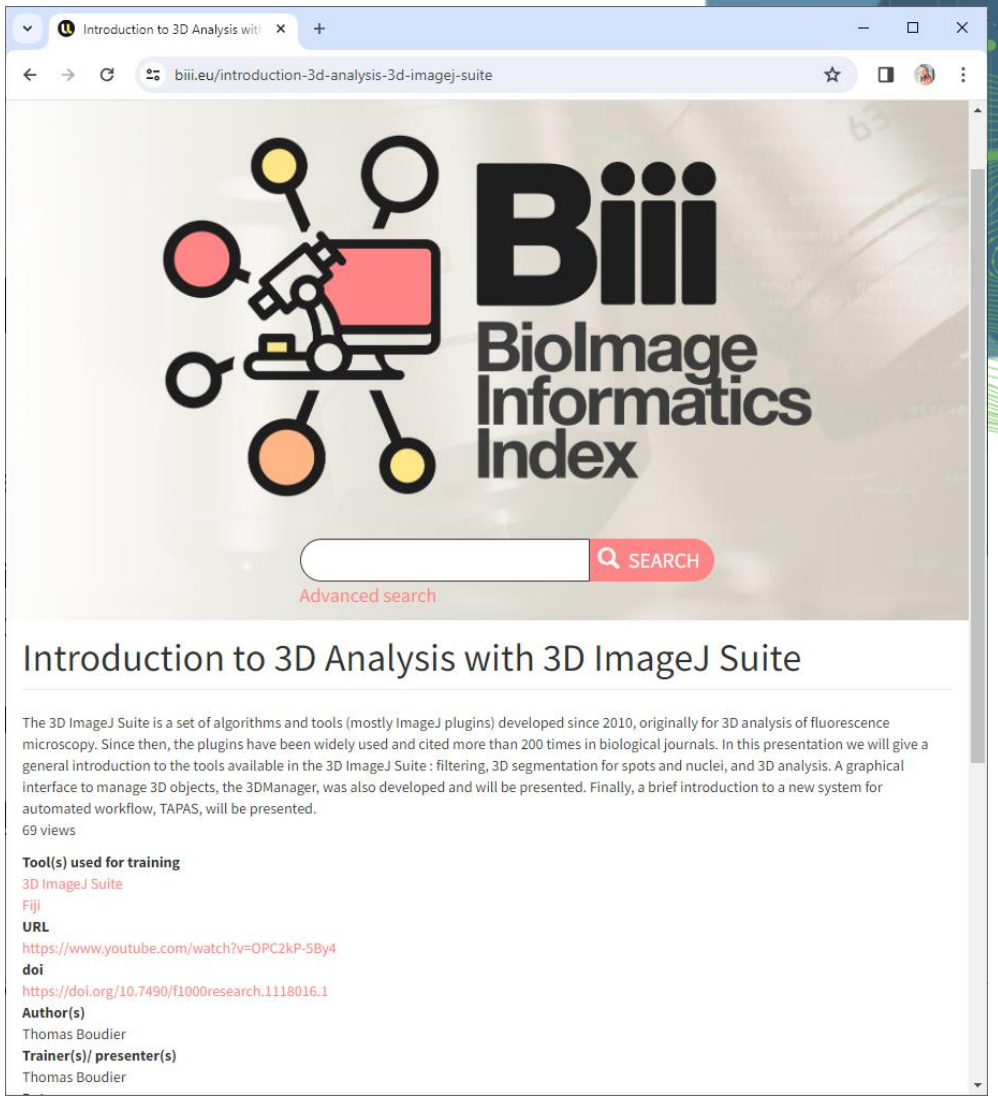


The FAIR-principles

Findable

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

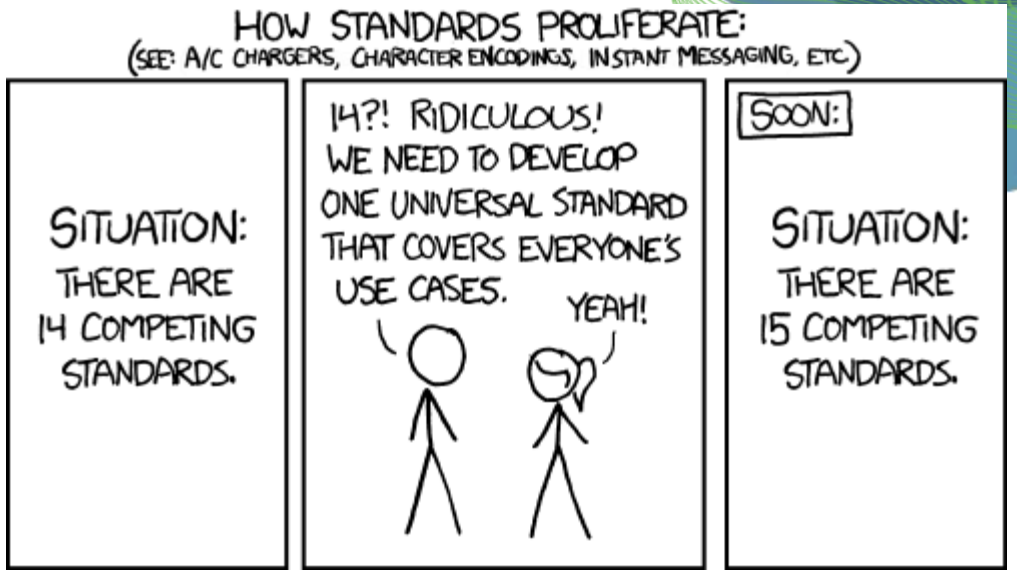
Hint: Do not invent a new search-engines for this ;-)



The FAIR-principles

Accessible

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
  - A1.1 The protocol is open, free, and universally implementable
  - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available



https://www.go-fair.org/fair-principles/
Comic taken from https://xkcd.com/927/ licensed under a Creative Commons Attribution-NonCommercial 2.5 License.

The FAIR-principles

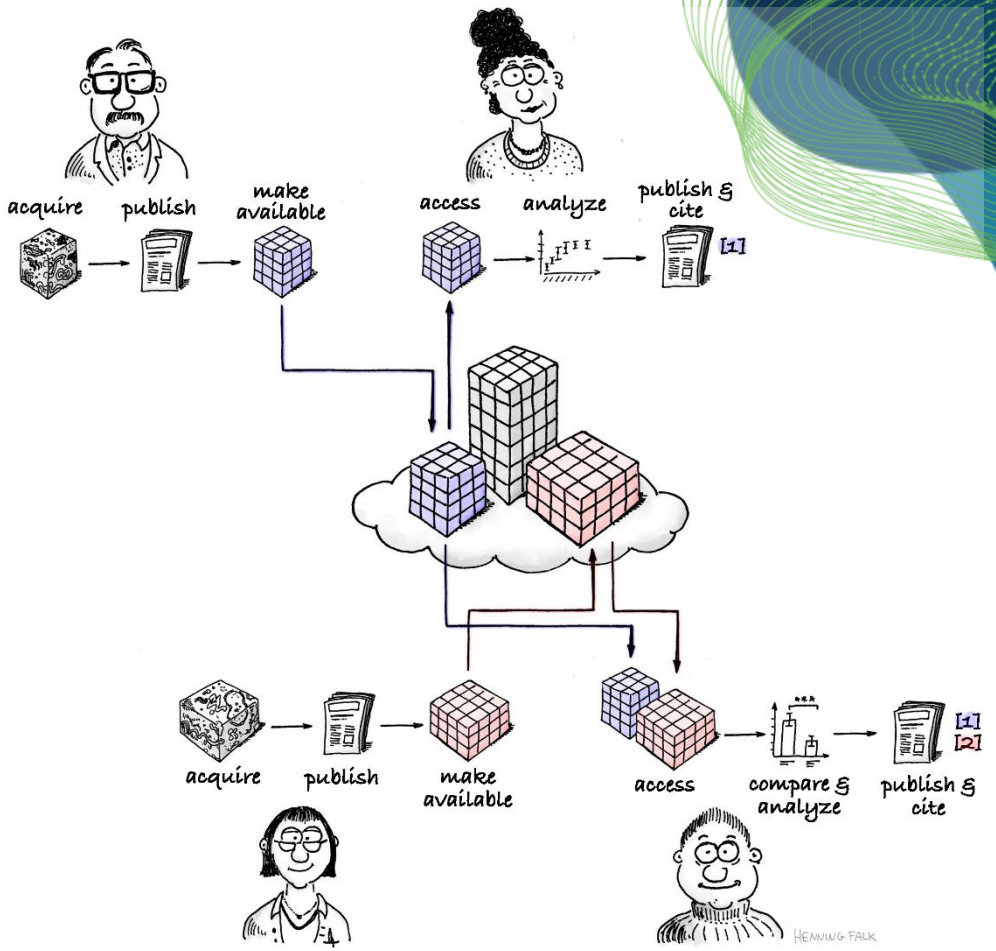
Interoperable

- 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

The FAIR-principles

Reusable

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



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

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

Please use

CC-BY

as license for your materials to make them reusable.

Summary

- If you want to make your stuff reusable:
  - Use permissive licenses
  - Share it on community-wide platforms (not institutional servers)
  - Register them 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/

https://doi.org/10.5281/zenodo.10654775

Acknowledgements

Communities & platforms

image.sc

NFDI4 BIOIMAGE

GloBIAS global bioimage analysts' society

neubias network of european bioimage analysts

FocalPlane Where biology meets microscopy

The Company of Biologists

BiAPoL team

- Marcelo Zoccoler
- Johannes Soltwedel
- Maleeha Hassan
- Till Korten
- Stefan Hahmann
- Somashekhar Kulkarni

Former lab members:

- Ryan George Savill
- Laura Zigutyte
- Mara Lampert
- Allyson Ryan
- Conni Wetzker

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