ScaDS.AI
DRESDEN LEIPZIG
CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE

Research Data Management

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.

Recap quiz

- We write good documentation to enabling others to do an experiment. This is good for …

Repeatability

Reproducibility

Replicability

Reliability

Recap quiz

- "Resolution" in microscopy imaging describes

Camera pixel size

Screen pixel size

Size of differentiable objects

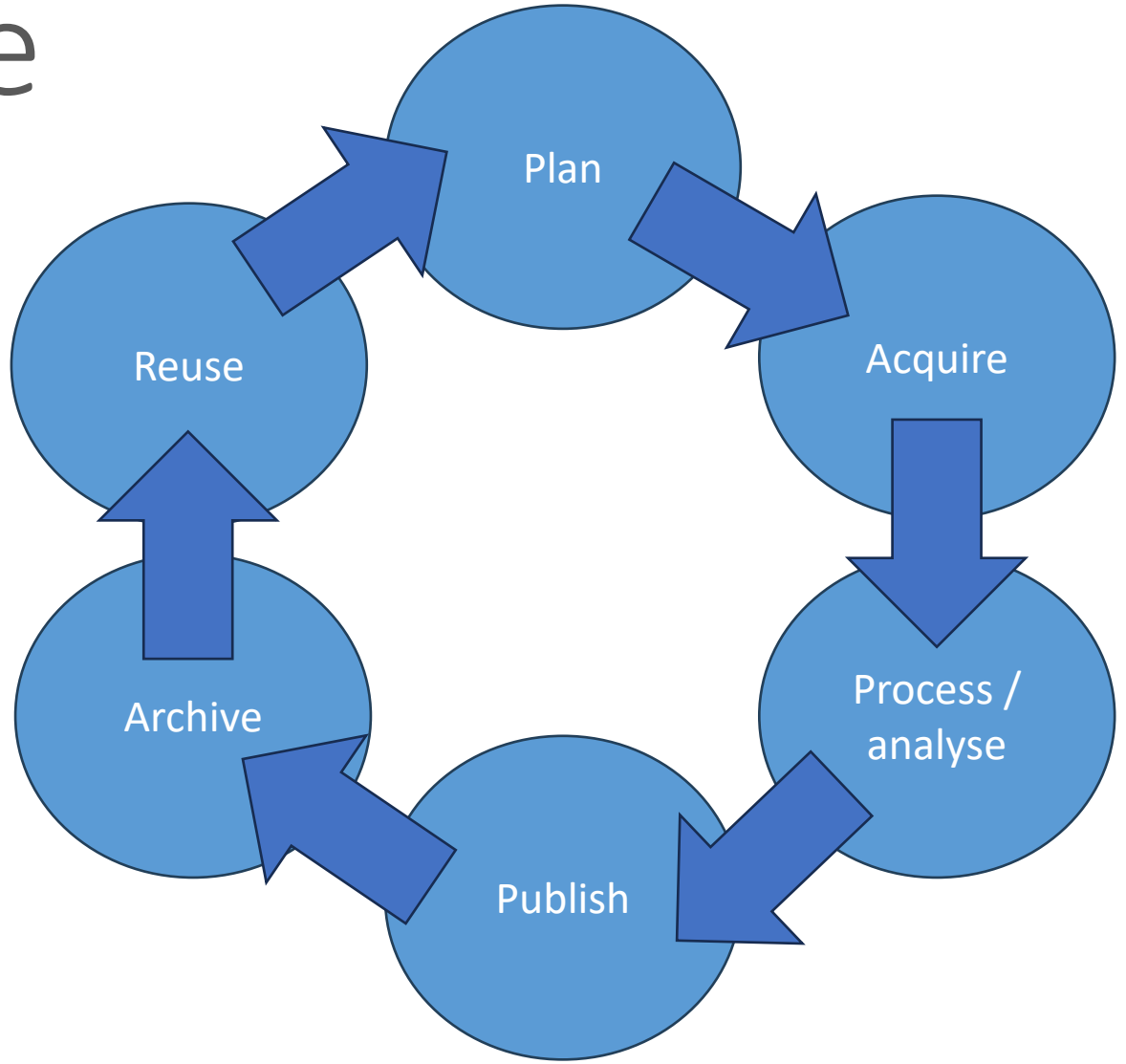
Objective magnification

Research Data Management (RDM)

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

RDM Life Cycle

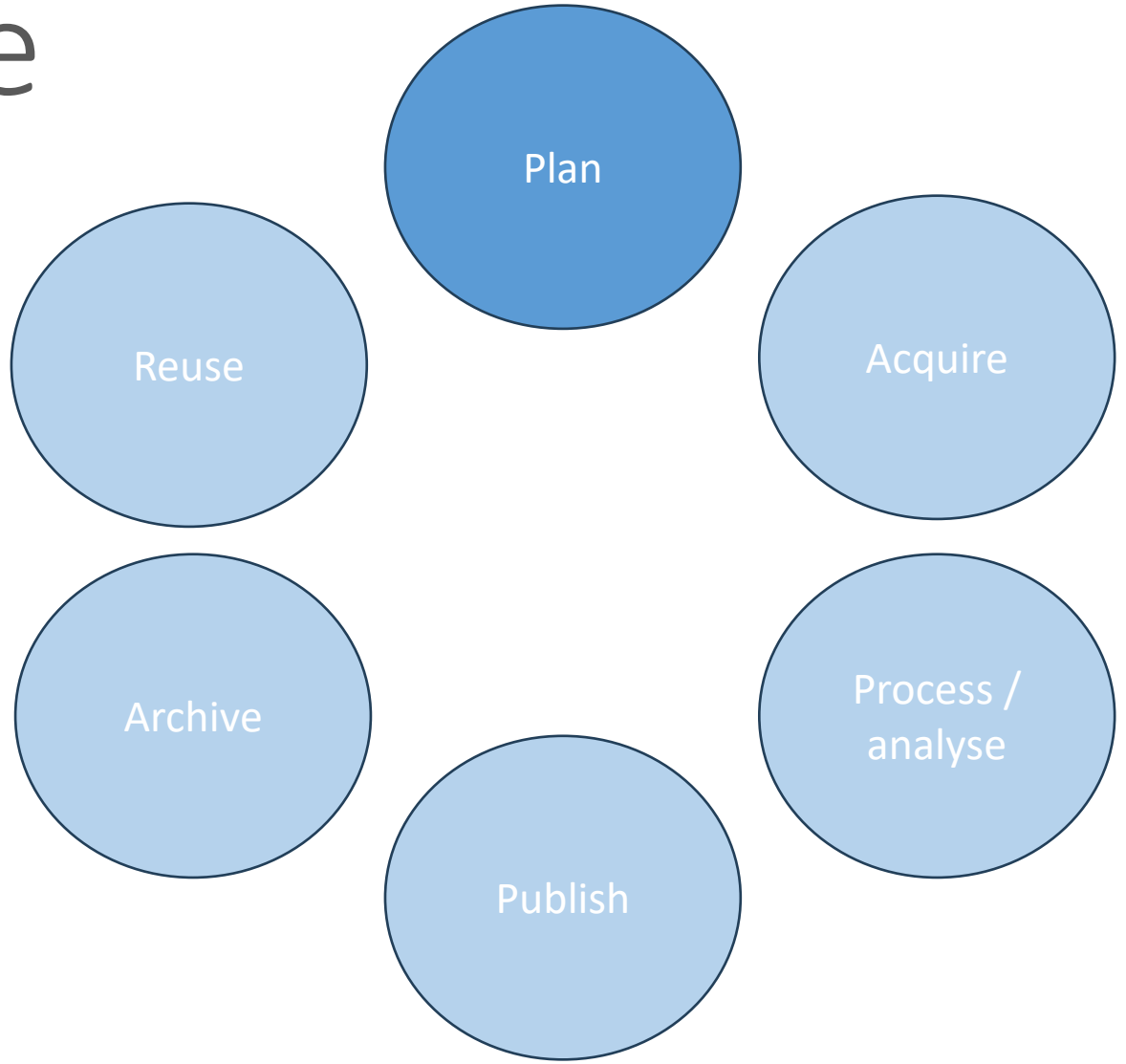
- Processes are ideally cyclic



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

RDM Life Cycle

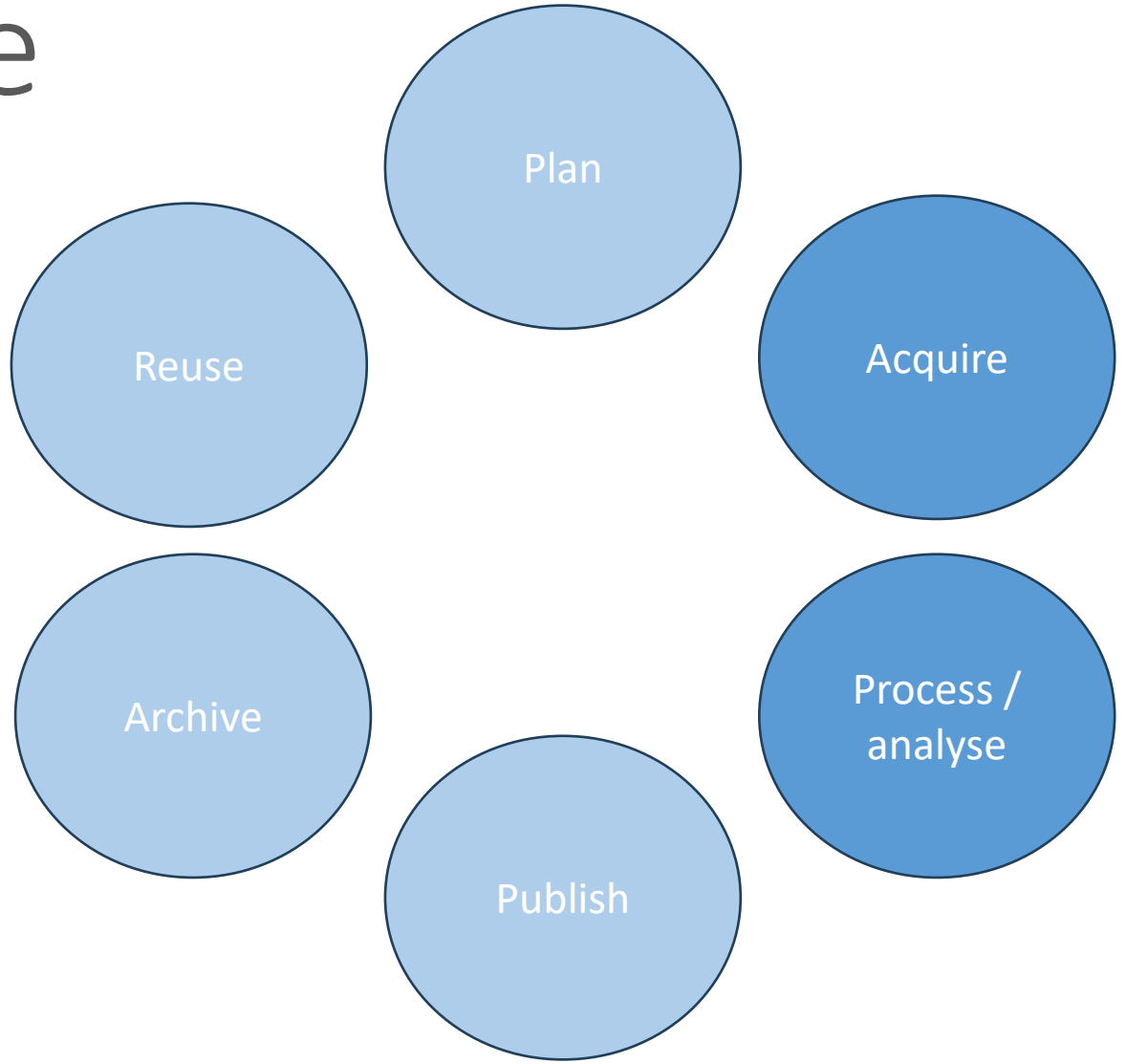
- Cost
- Benefit
- Quality
- Strategic decisions



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

RDM Life Cycle

- Types of data
- Terms and conditions
  - Usage rights
  - Copyright
- IT infrastructure
- Backup



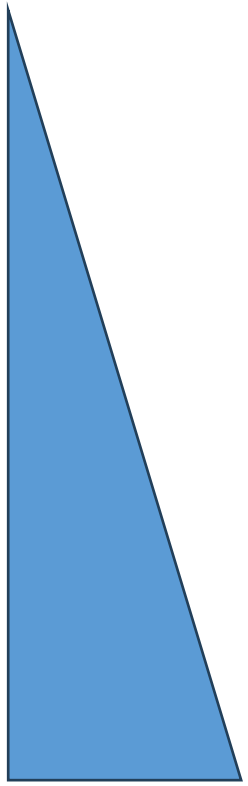
Adaptiert von:
https://forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/

Types of data

- Structured data
  - Tables, databases
- Unstructured data
  - Texte, emails, videos, pictures
- Semi-structured data
  - Frageboegen
  - Scientific images

Types of data

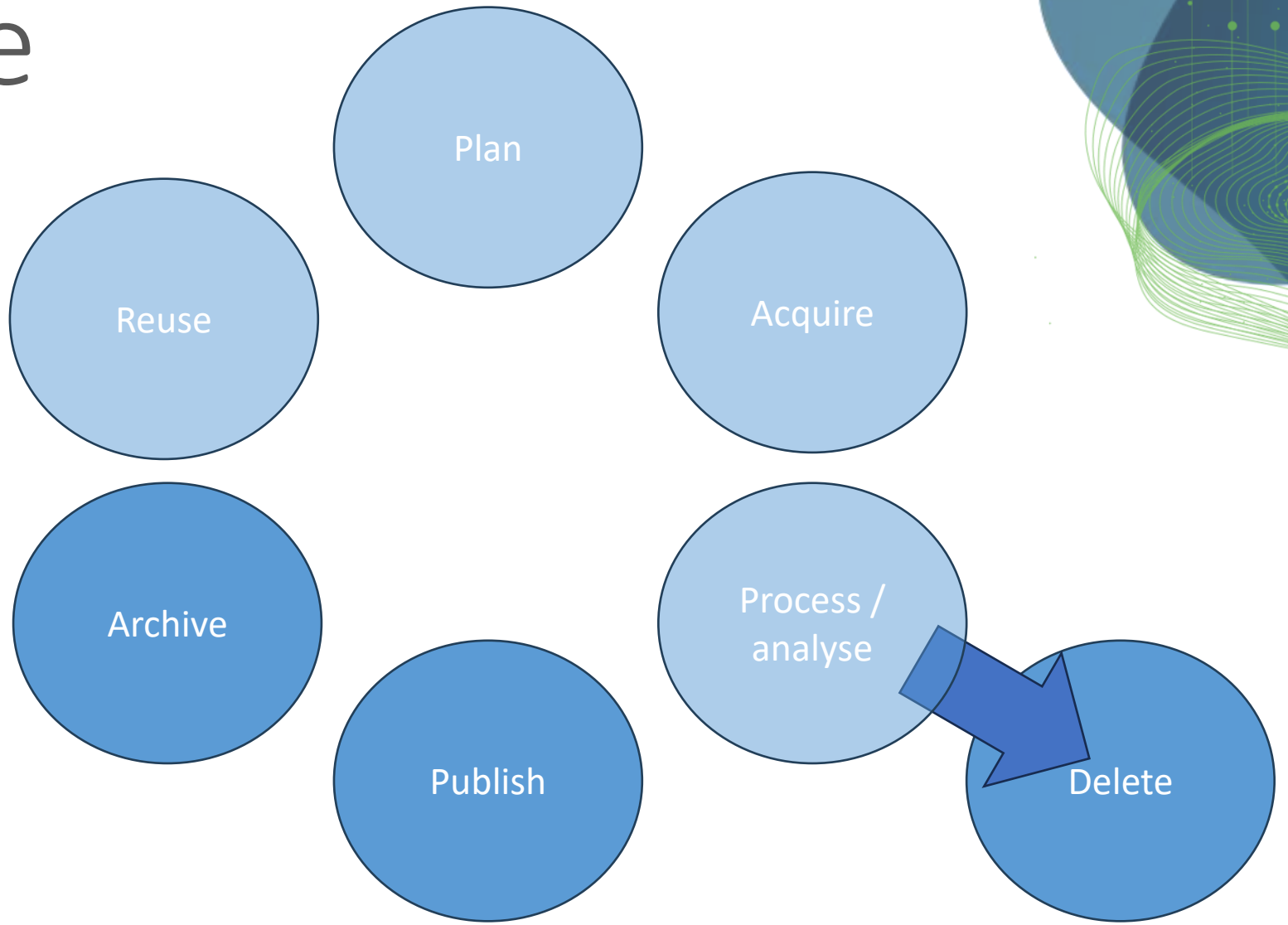
- Openly accessible data
  - „open data“
  - „open source“ software
- Business data
- Research data
  - Hot / cold
- Personal data
- Secret data



In need of protection (schutzbedürftig)

RDM Life Cycle

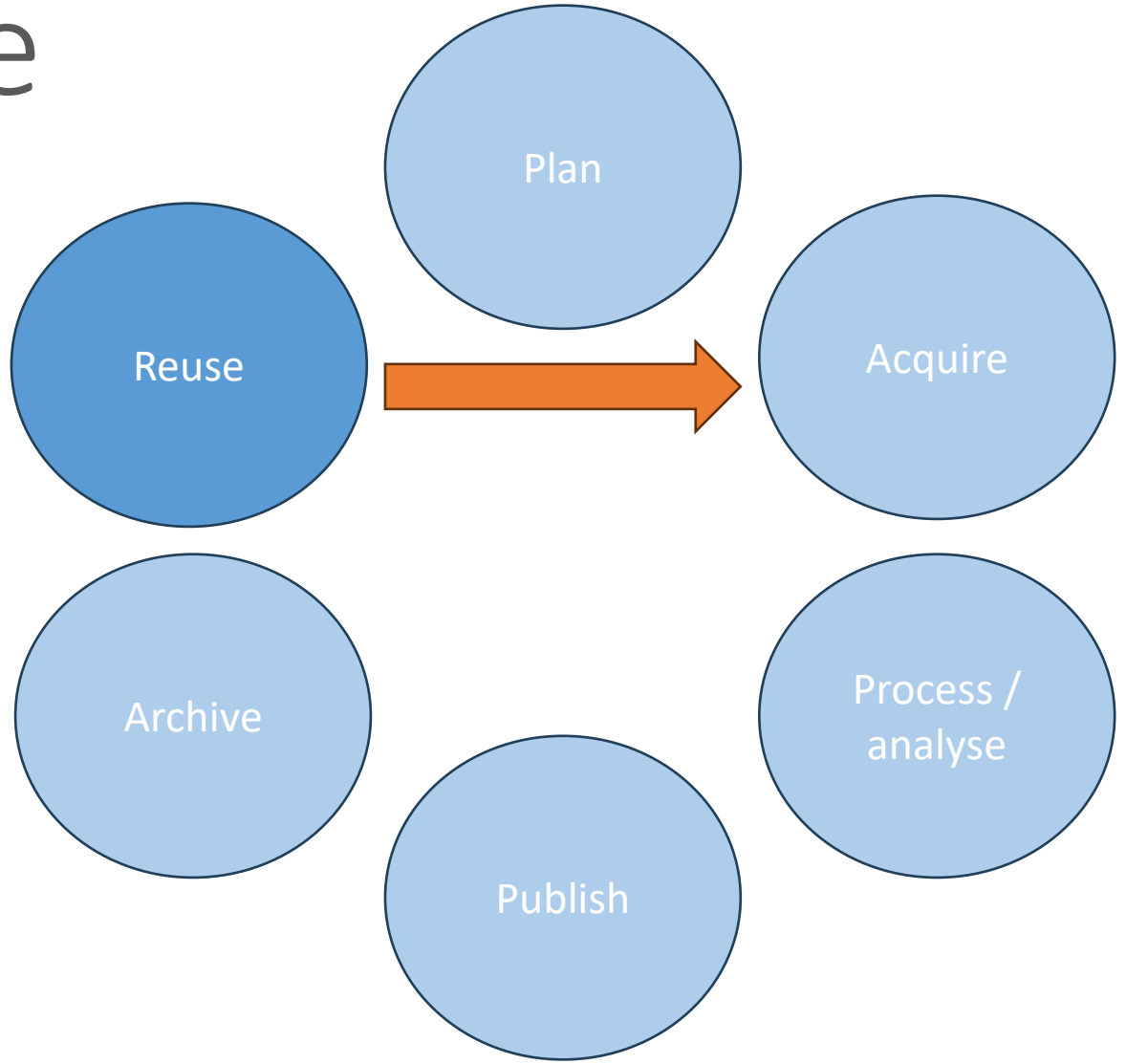
- Right to publish
- Regulatory aspects
  - Research data: archive 15 years
- Authorship
- Registration (-> Findable)



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

RDM Life Cycle

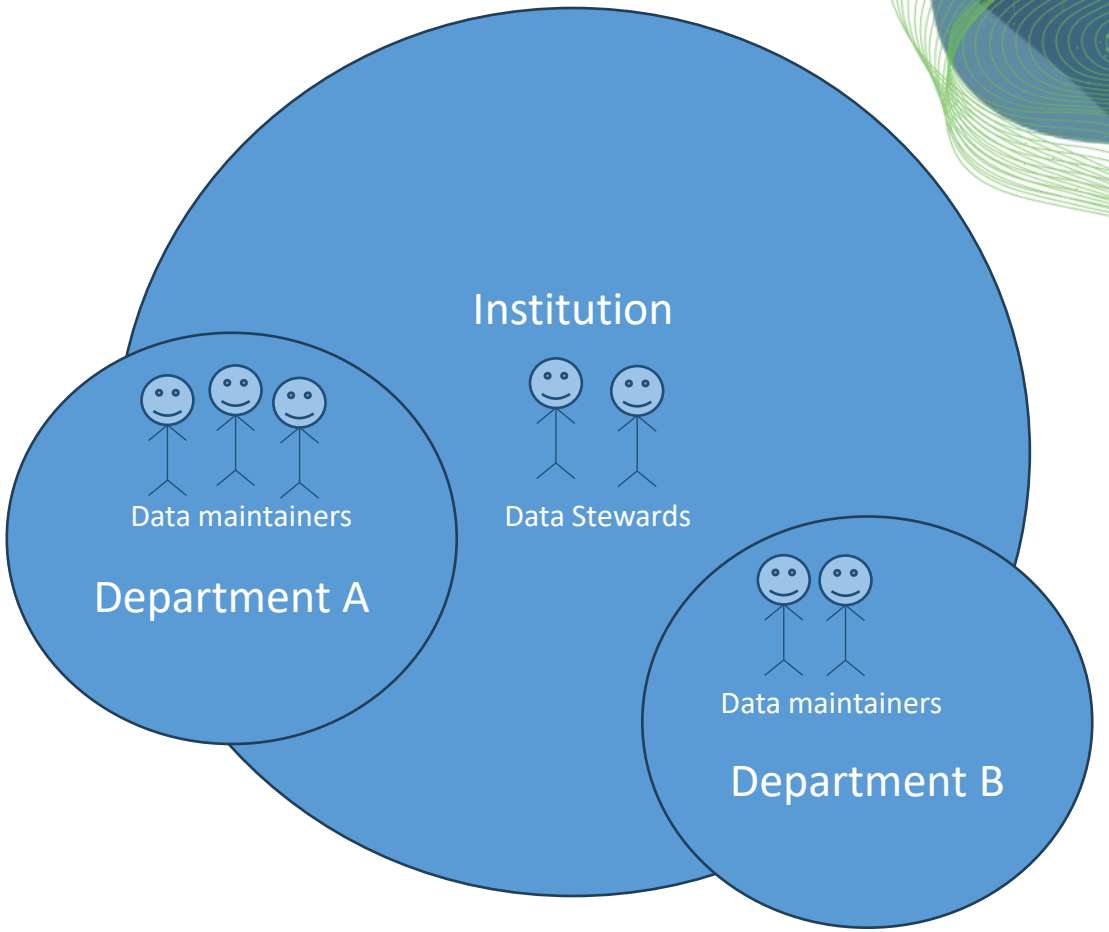
- Potential future benefit
- Sustainability
- Important: Licensing
  - Has impact on next cycle / acquisition



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

What is good RDM?

- Clearly defined responsibilities and processes (Governance)
  - Data Management Plan (DMP)
- Communication of goals, metrics, responsibilities, processes
- Dedicated personnel
  - “Data maintainers”
  - IT infrastructure maintainers
- Expert consultants
  - “Data stewards”



Roles != Job profiles

Domain specialist

- Focuses on scientific question, often related to the physical world
- Requires sound insights and sustainable solutions
- Examples:
  - Biologist
  - Geologist
  - City planner

Data analyst

- Focuses on methods for data processing / visualization
- Gains sound insights
- Examples:
  - Statistician
  - Bioinformatician
  - Data Scientist

IT specialist

- Focuses on IT infrastructure
  - Hardware
  - Software
- Builds sustainable solutions
- Examples:
  - Computer scientist
  - IT specialist

Data Management Plans (DMPs)

- Describes the IS-state of a data environment
  - Which data is acquired / processed?
    (content, format, amount)
  - What meta-data is collected?
  - Which quality standards are targeted?
  - How is data saved, archived, backed-up, shared, published…?
  - Who is responsible for what?
    - Roles, job-profiles
  - What does this all cost?
    (IT infrastructure + human resources)

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!

Quiz

- Regularly copying files to a remote place is …

Archiving

Backup

Quiz

- Data Scientists is a …

Role

Job profile

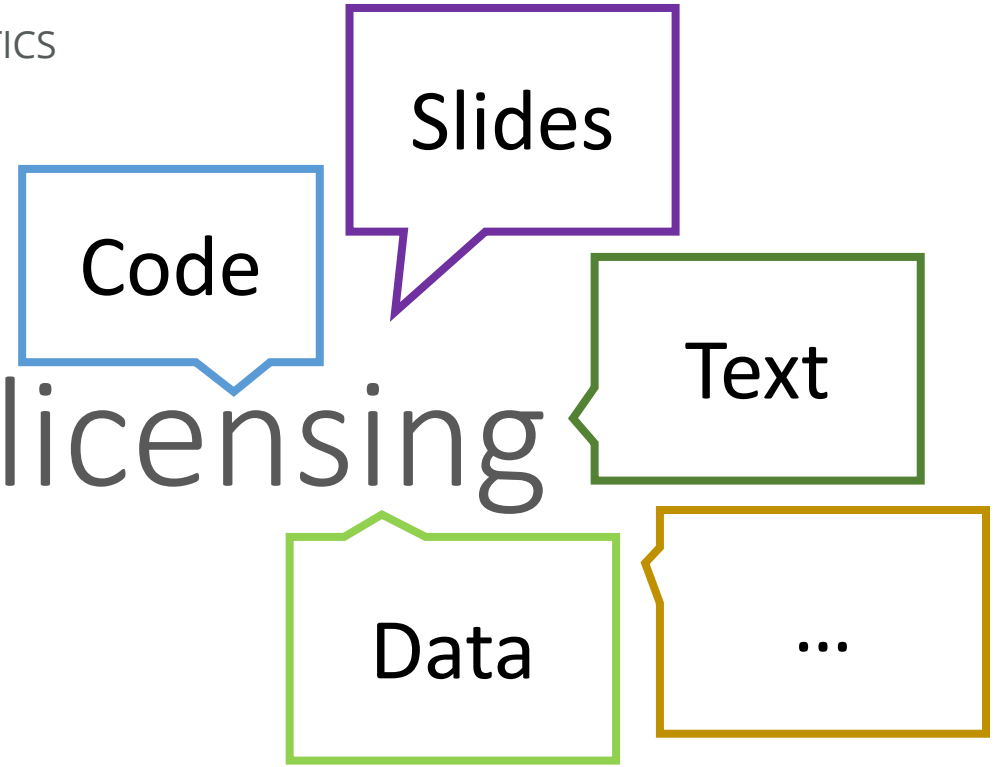
Quiz

- Data Steward is a …

Role

Job profile

ScaDS.AI
DRESDEN LEIPZIG
CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE



Sharing & licensing

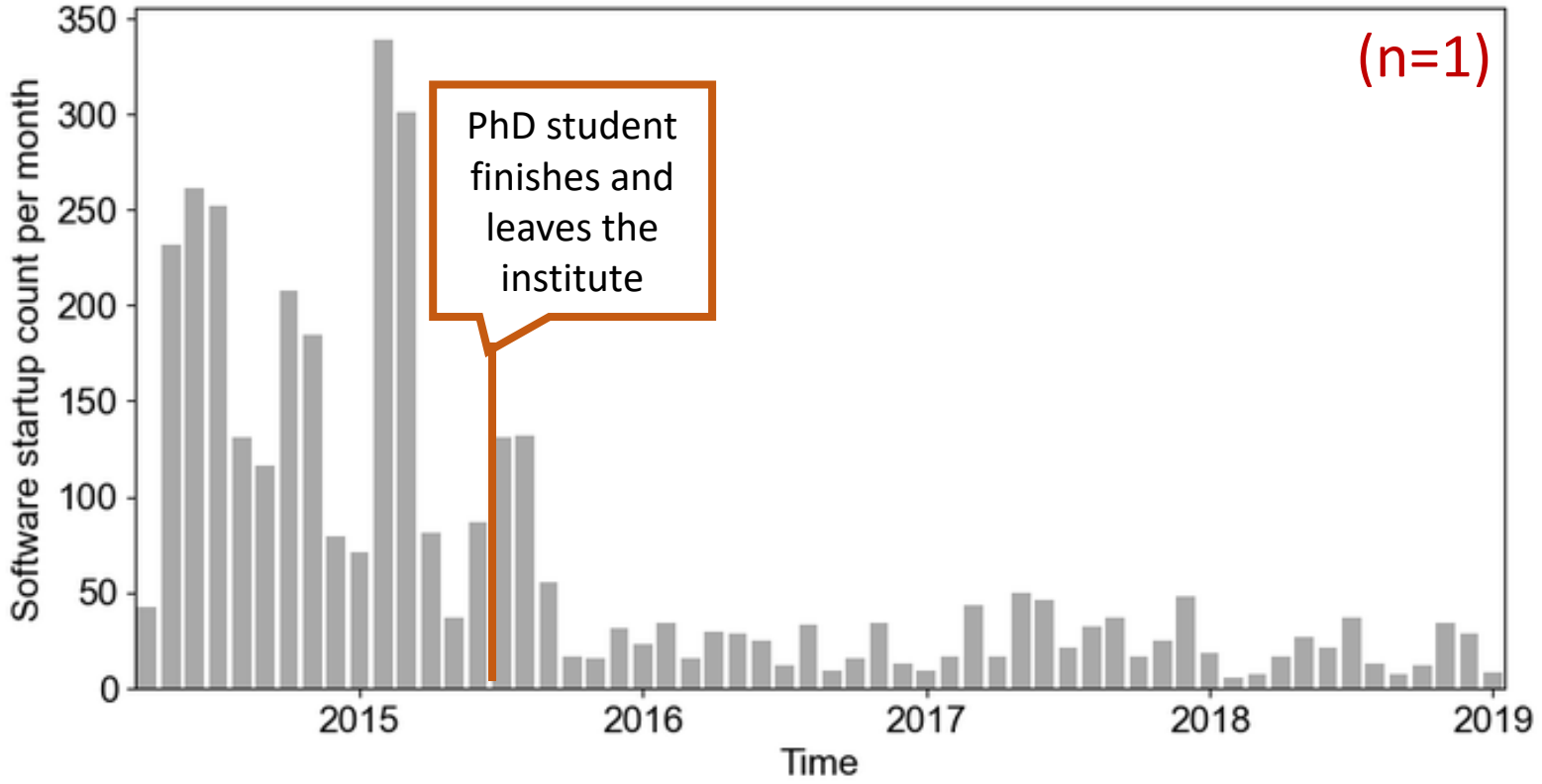
Robert Haase

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

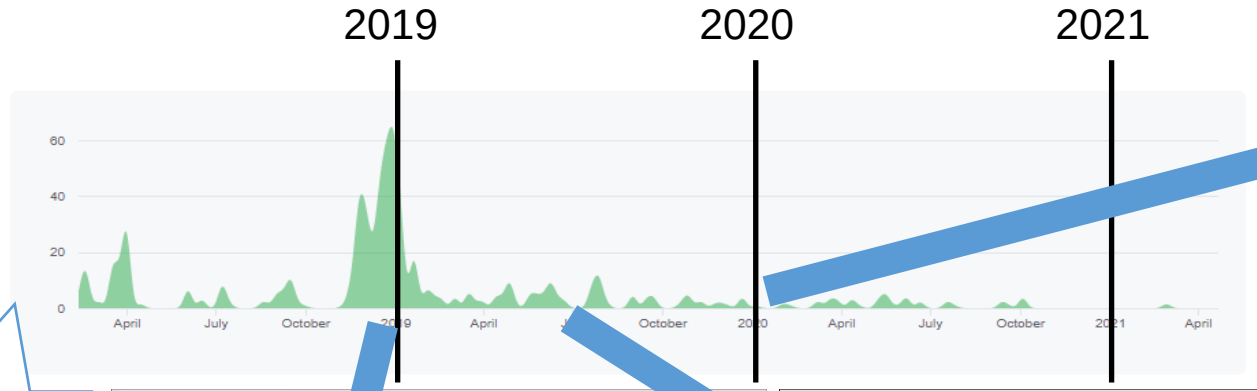
SACHSEN
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages.
Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Sustainability of *my* contribution to science

- What happens to research software once the PhD student leaves the institute / field?

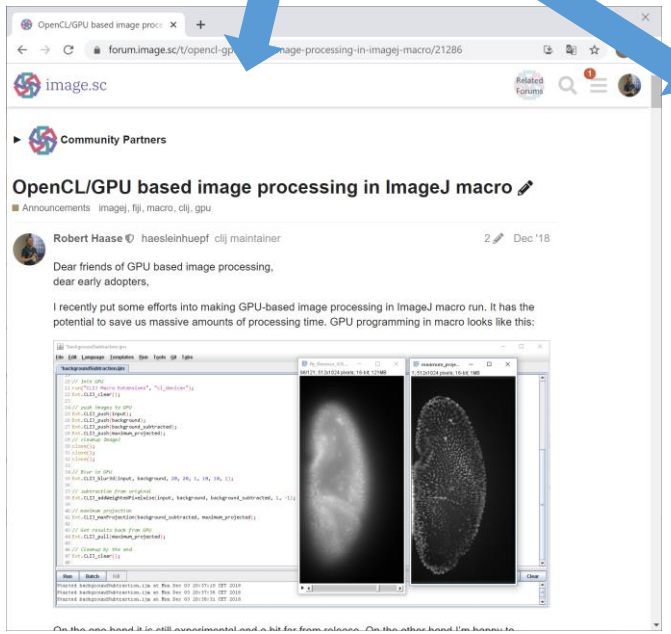


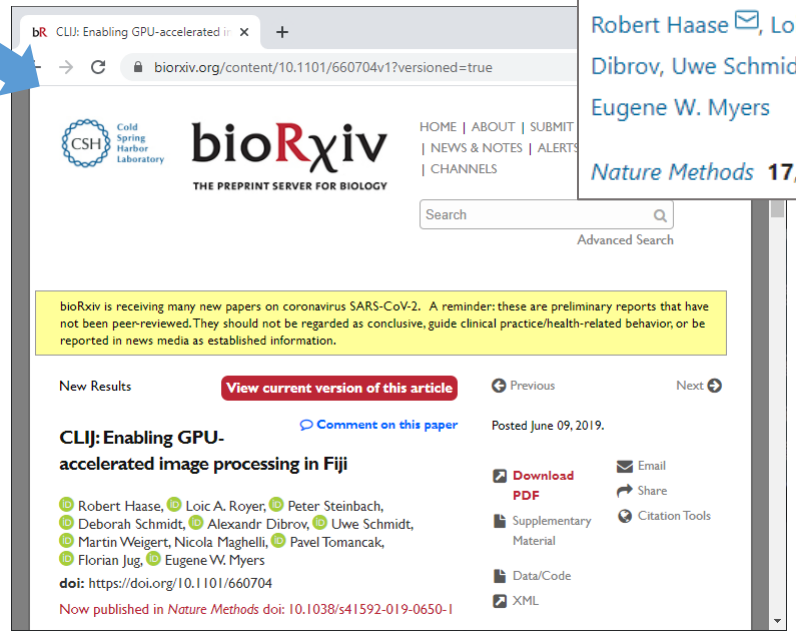
Developing software in the open

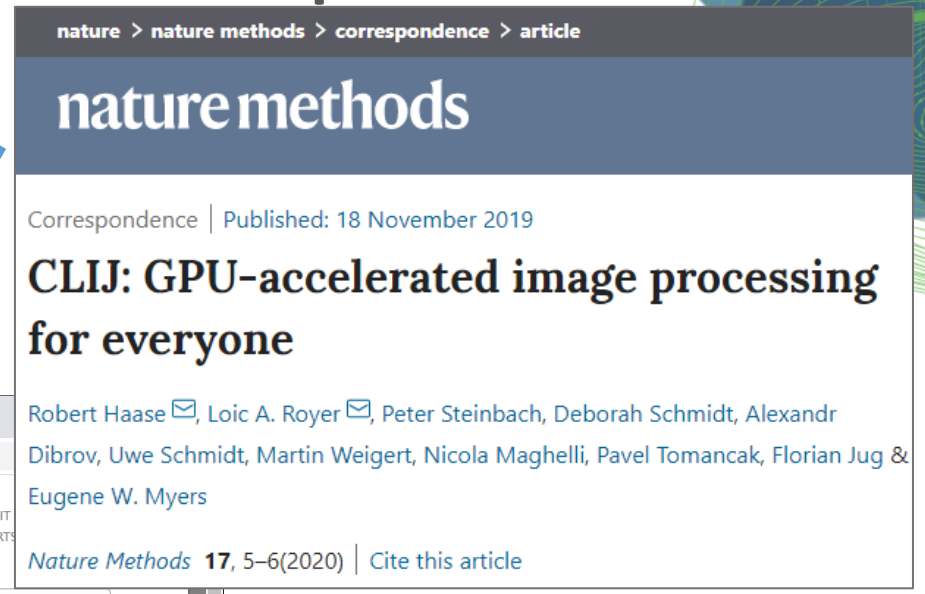


Nov. 2017: I took over microscopy control software from Loic and "found" some GPU-accelerated image processing in there

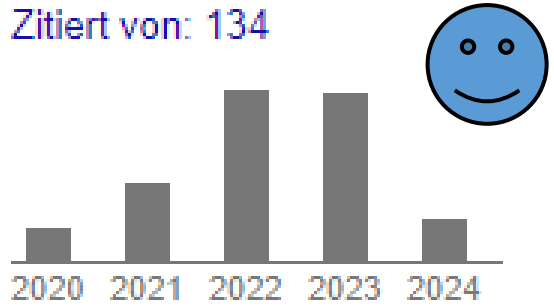
Loic A. Royer
(CZ Biohub)
@loicaroyer





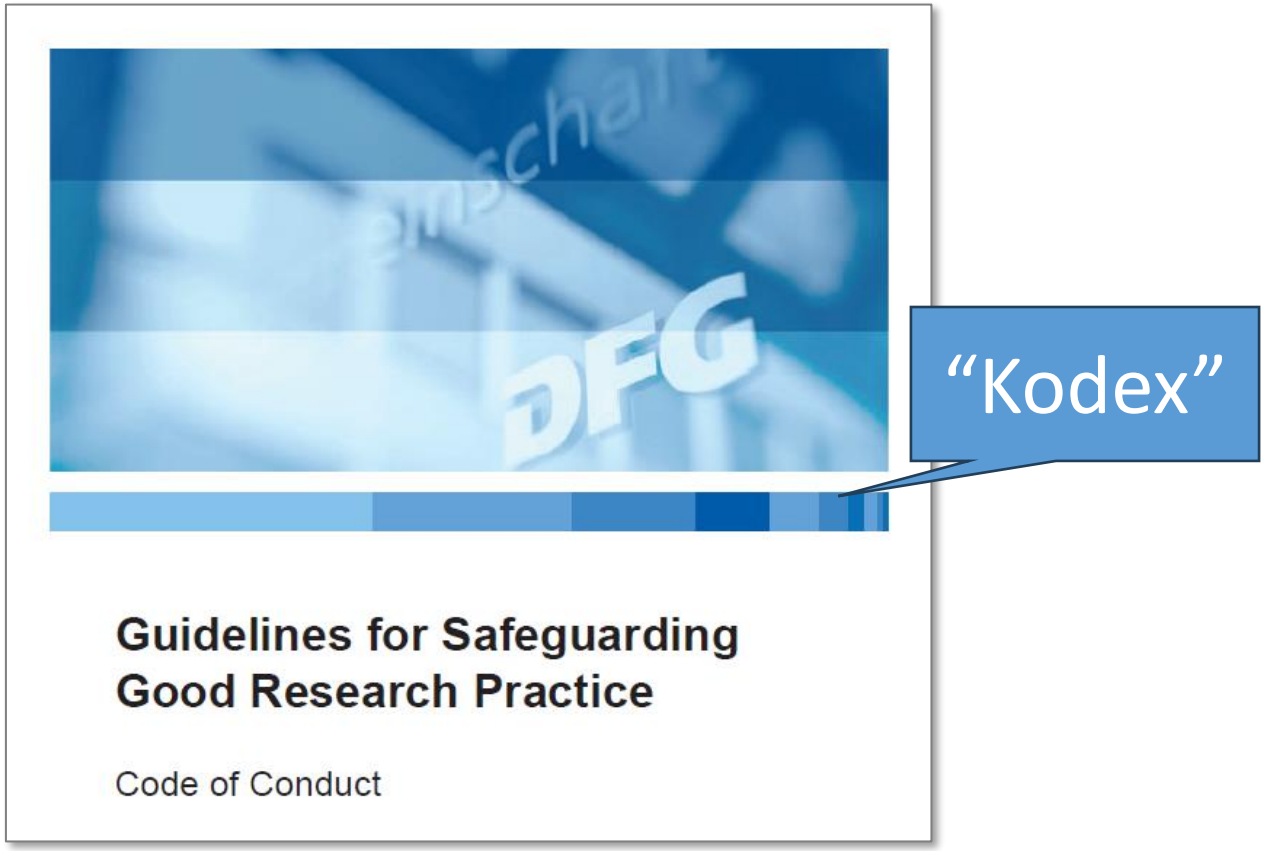


Today: 134 citations
(Google scholar, 2024-03-18)



Scientific culture

Public access to research results -> Reusability



"Kodex"

Guideline 13: Providing public access to research results

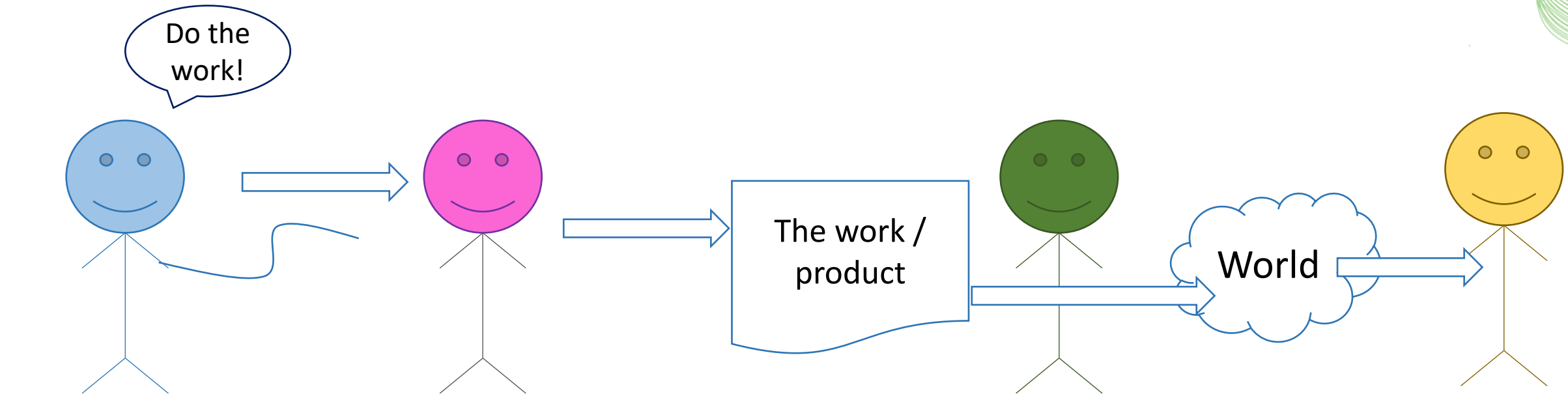
- As a rule, researchers make all results available as part of scientific/academic discourse. In specific cases, however, there may be reasons not to make results publicly available (in the narrower sense of publication, but also in a broader sense through other communication channels); this decision 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

Am I allowed to publish my stuff?

- … it depends… on who is responsible



Copyright holder, e.g. employer of the author

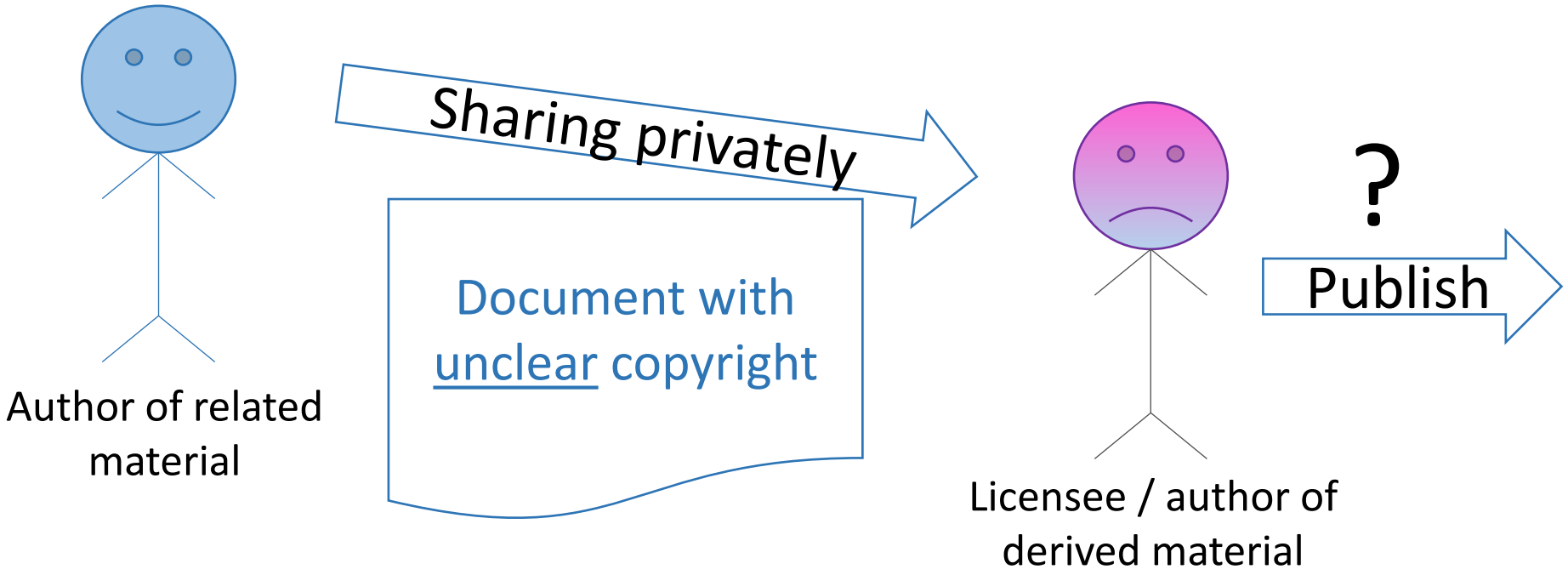
Author
Creator of the work

Publisher
The person who made the work available

Licensee
The person who uses / reuses the material

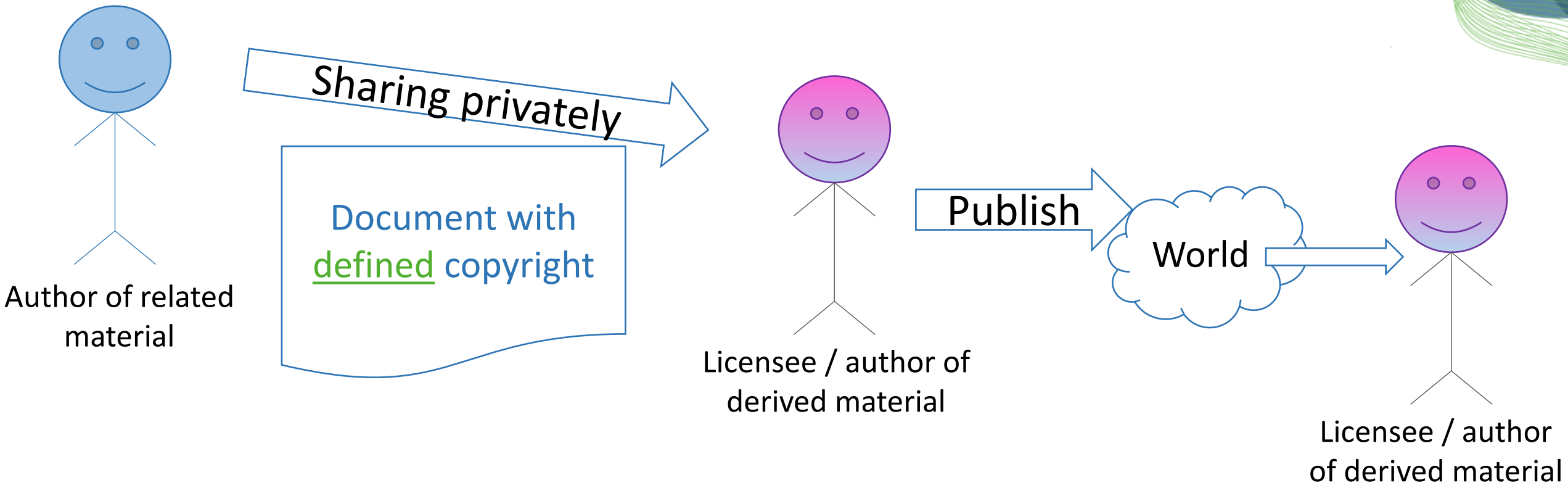
Am I allowed to publish my stuff?

- … it depends… on what materials served as basis



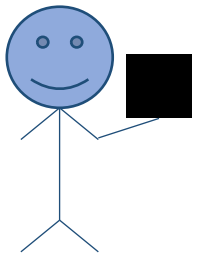
Am I allowed to publish my stuff?

- … it depends… on what materials served as basis



Openness of software / projects

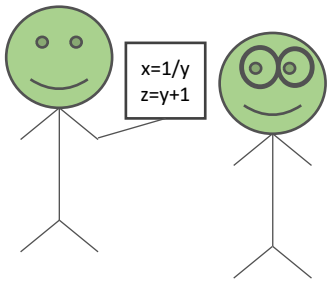
Closed source



- Open to collaborations
- “Black box”
- Compiled code (e.g. C/C++)
- Good for protecting intellectual properties ($$$)

Hardware device drivers

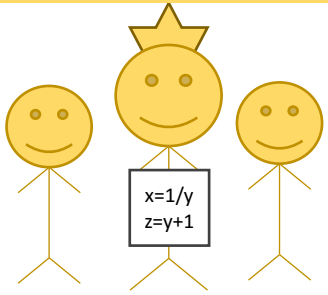
Open source



- Code available to read
- Not necessarily executable code
- No maintenance / support efforts

Custom image analysis scripts

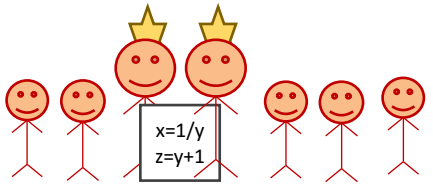
Benevolent dictatorship



- Open to contributions
- Single maintainer, often overwhelmed
- Efficient decision making
- Bus factor ≈1

TrackMate, SNT, MorpholibJ, CLIJ

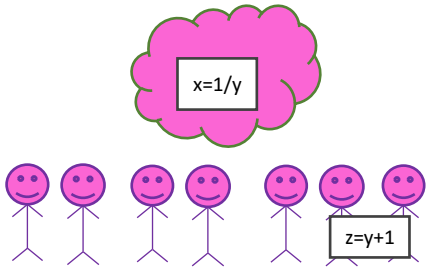
Community driven



- Open to contributions
- Partially democratic
- Board of maintainers (core developers)
- Long-winded decision making

scikit-image, scipy, OpenCL

Openly extensible



- Openly extensible; without maintainers involved
- Partially community driven

ImageJ, Python, numpy

Read more: https://coiled.io/blog/stages-of-openness/

Quiz

- What is the role of Github in the context of publishing open-source code?

Copyright holder

Author

Publisher

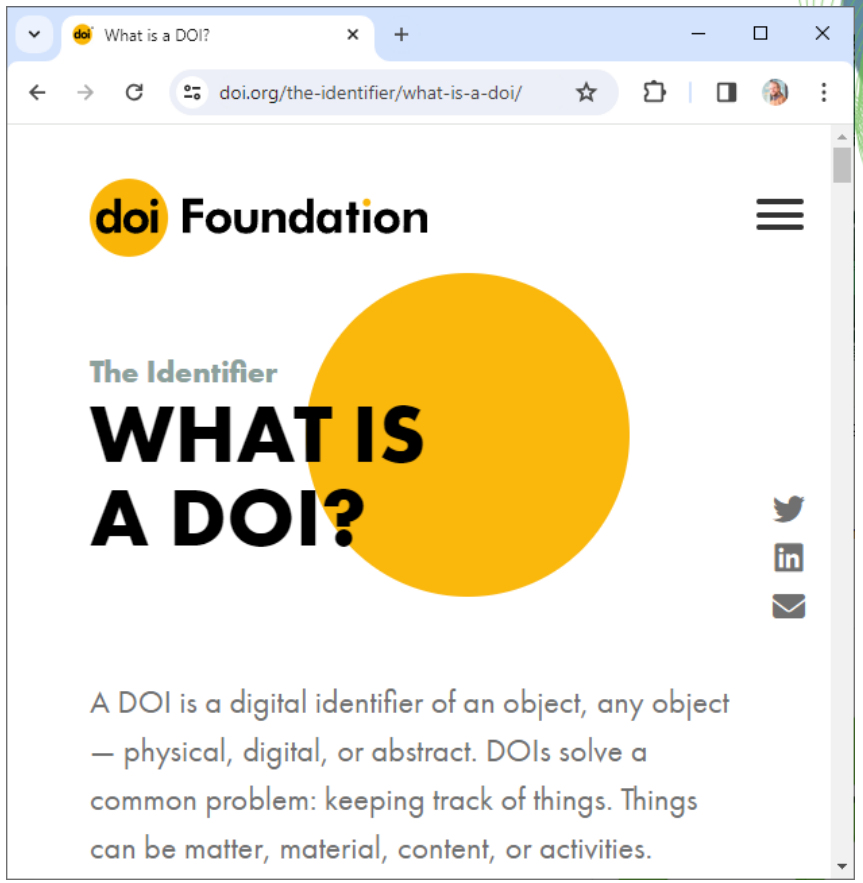
Licensee

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



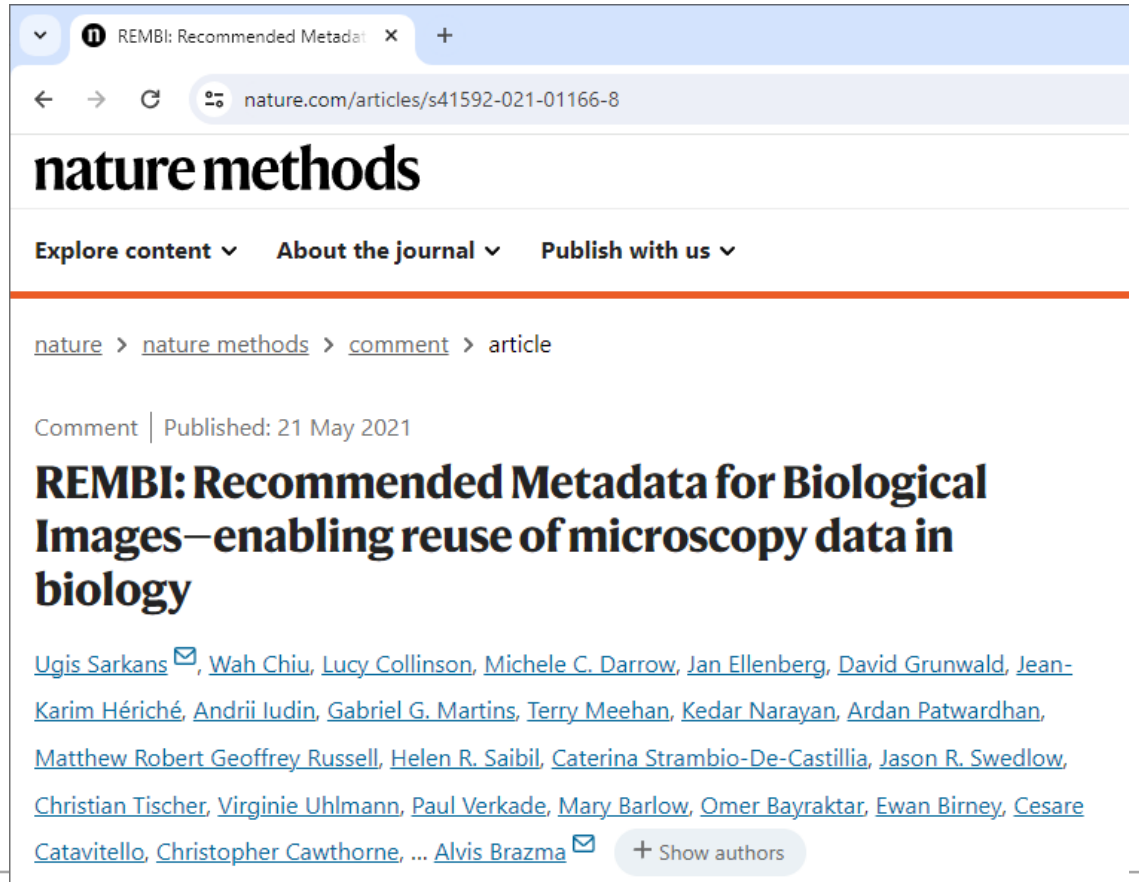
Screenshot source: https://www.doi.org/the-identifier/what-is-a-doi/ licensed CC-BY 4.0
https://www.go-fair.org/fair-principles/
https://biii.eu/introduction-3d-analysis-3d-imagej-suite

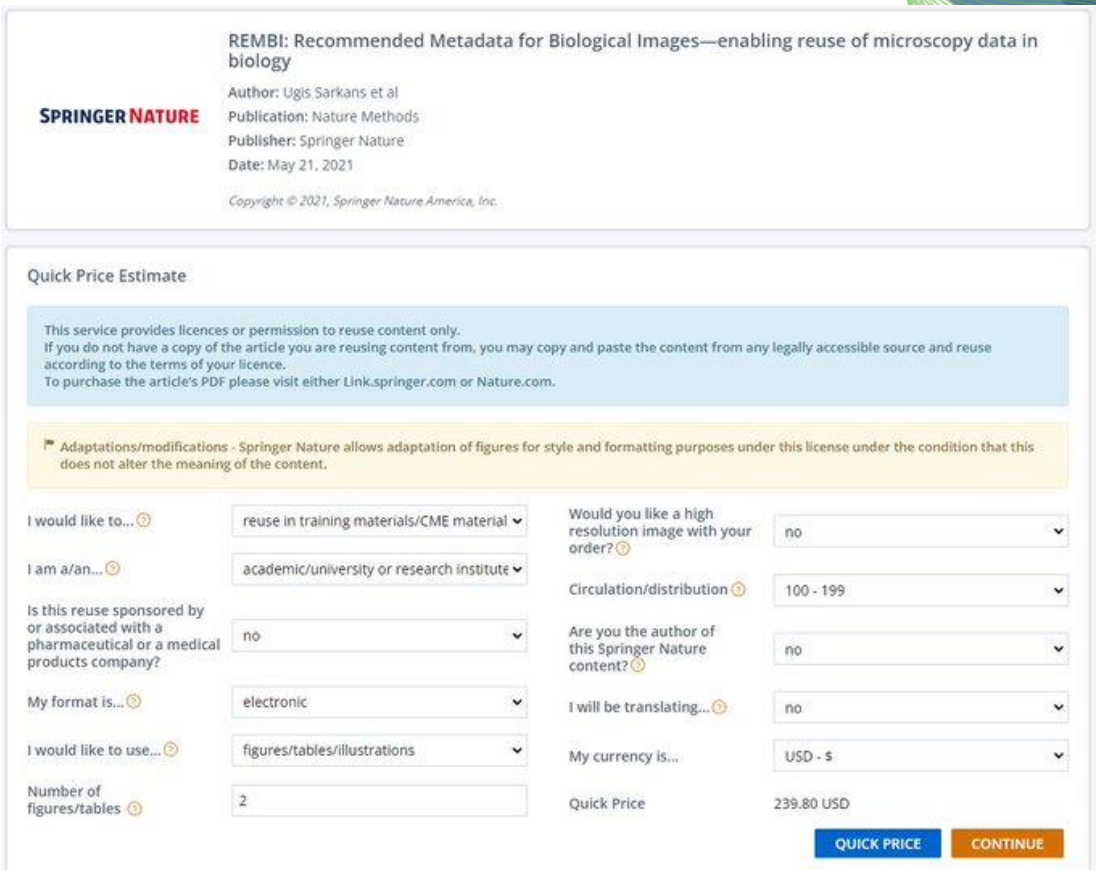
Meta data

- Generic
  - Author
  - Usage license
  - Creation date
- Field-specific (microscopy)
  - Exposure time
  - Wavelength (colour)
  - Microscope type/vendor

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

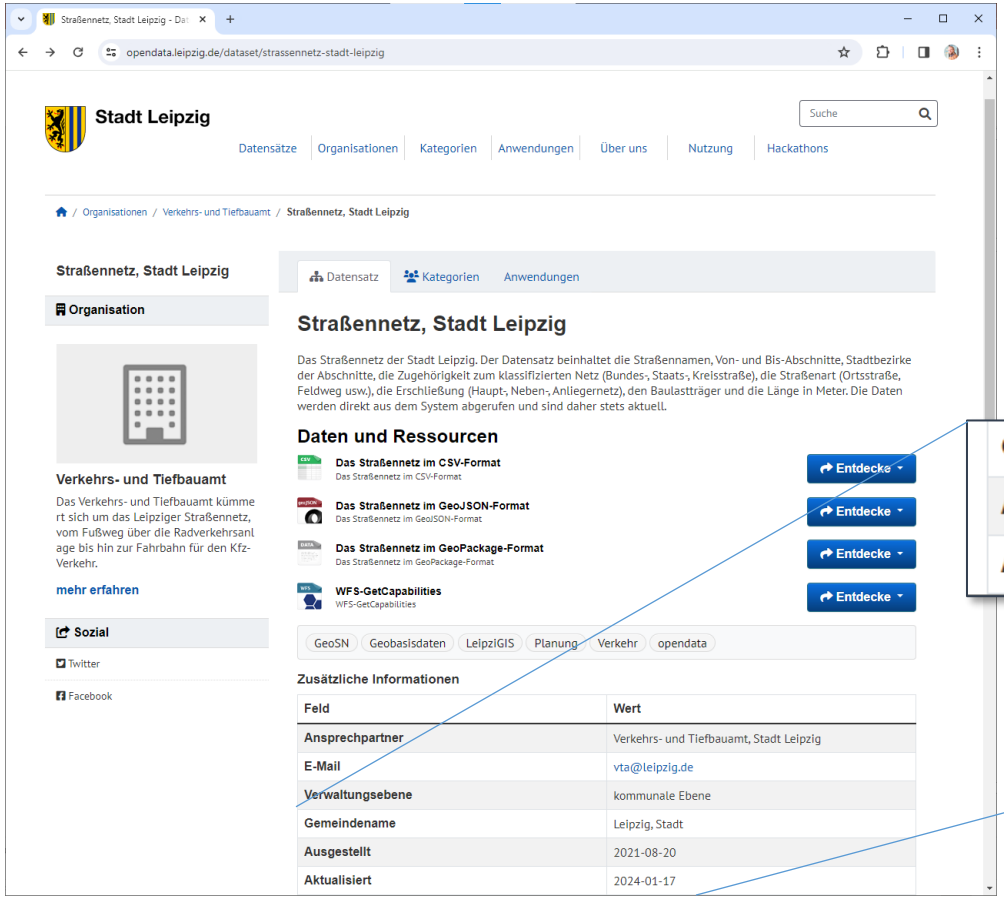
- Read more:





Digital Object Identifiers (DOI)

- DOIs / URIs always point at the same data
- DOIs are centrally registers, URIs not
- Unified Resource Locators (URLs) may point at different things



Gemeindename	Leipzig, Stadt
Ausgestellt	2021-08-20
Aktualisiert	2024-01-17

This no DOI, no URI, it's a URL

Unified Resource Identifiers

- Which of these are URIs?

https://twitter.com/haesleinhuepf/status/891596662782779392

https://doi.org/10.5281/zenodo.28325

https://opendata.leipzig.de/dataset/vornamenstatistik-2023

https://www.leipzig.de/

Digital Object Identifiers

- Which of these are DOIs?

https://twitter.com/haesleinhuepf/status/891596662782779392

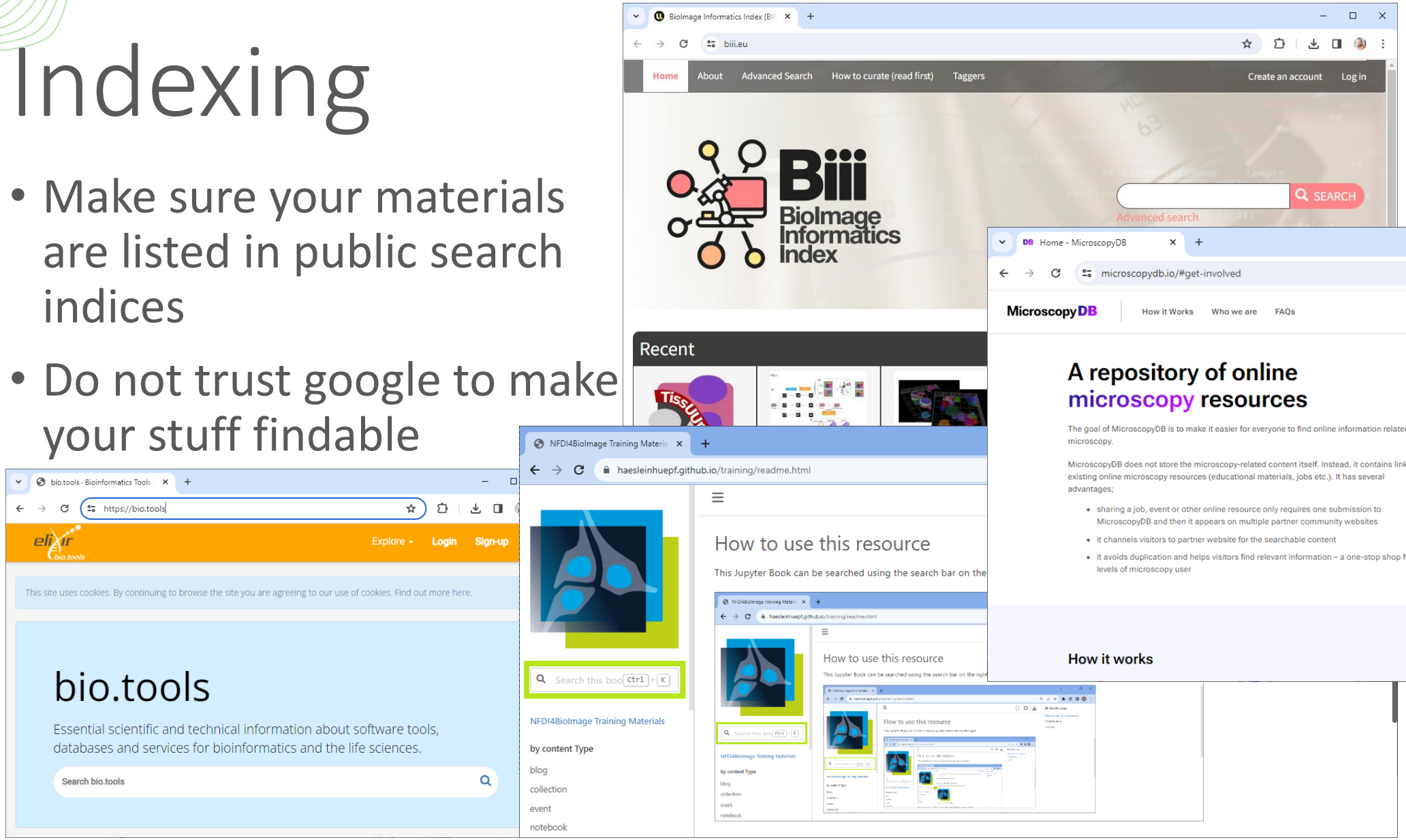
https://doi.org/10.5281/zenodo.28325

https://opendata.leipzig.de/dataset/vornamenstatistik-2023

https://www.leipzig.de/

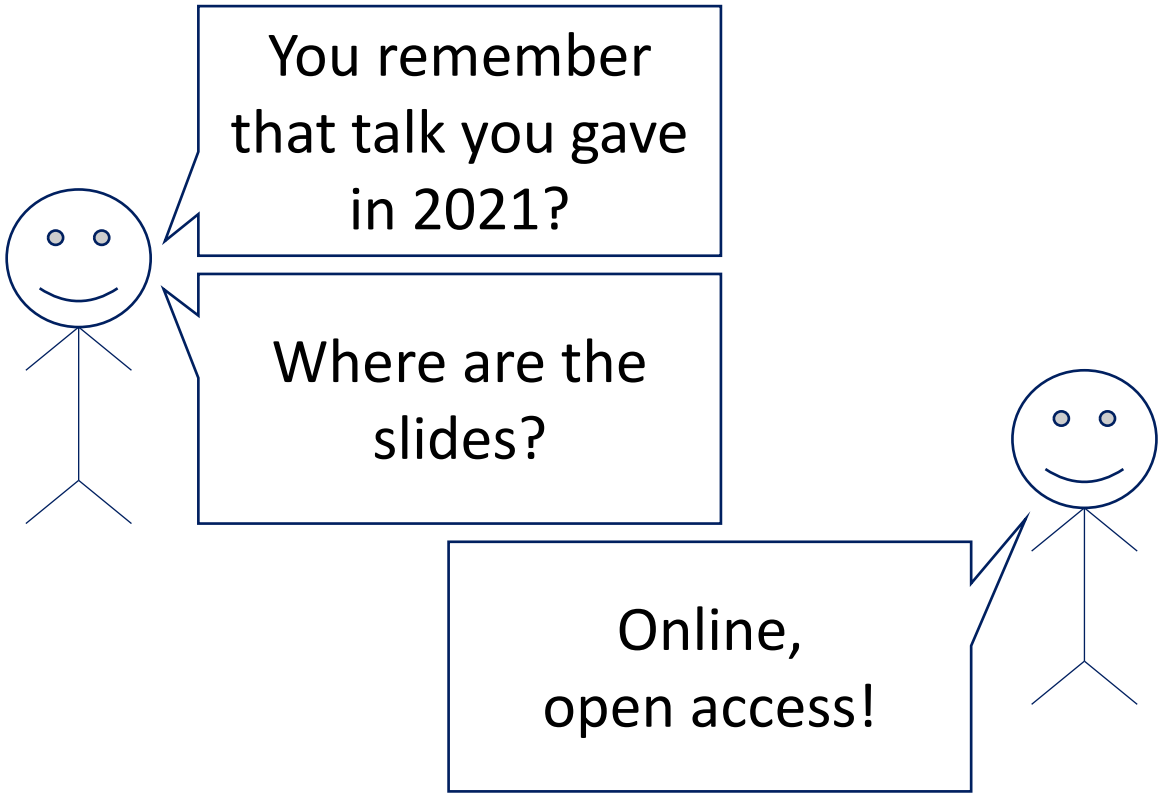
Indexing

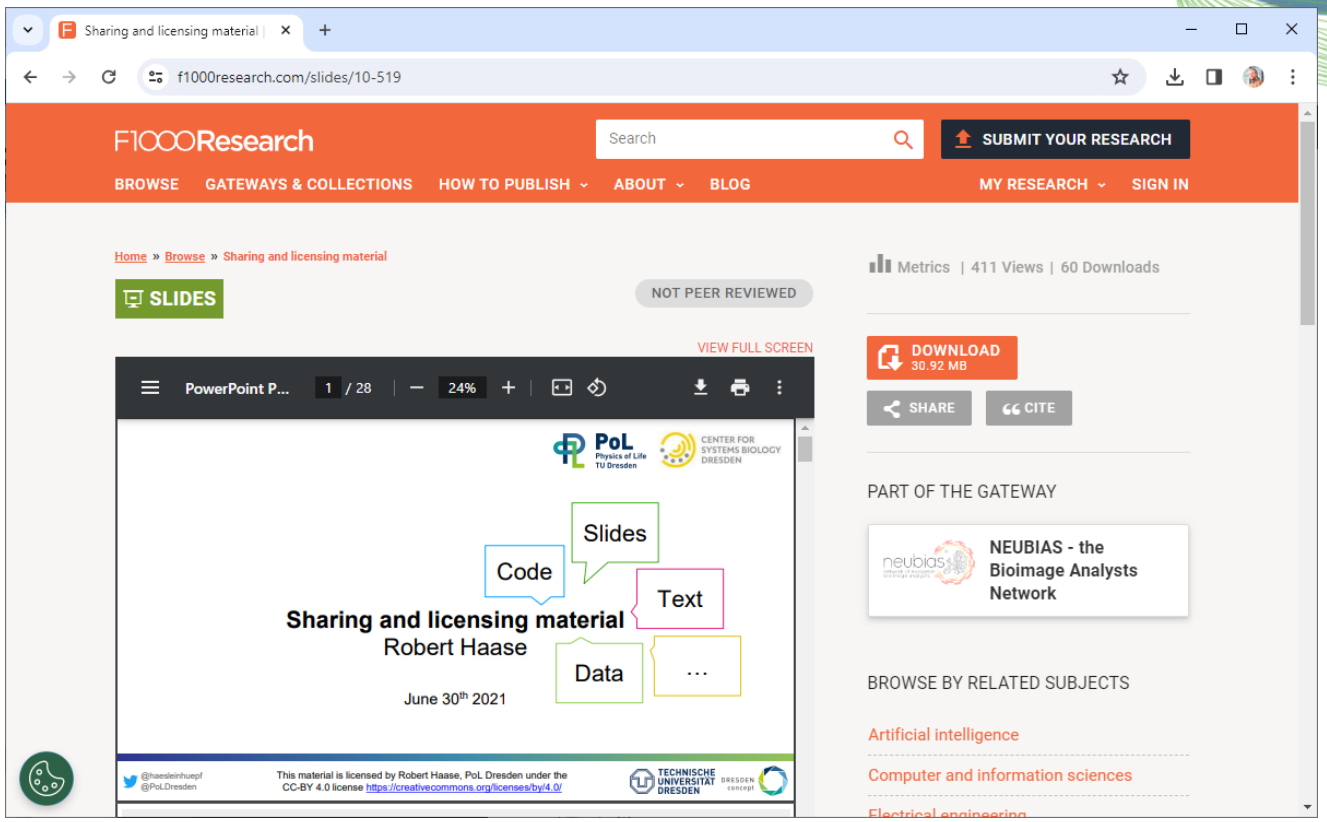
- Make sure your materials are listed in public search indices
- Do not trust google to make your stuff findable



Incentives: Findability

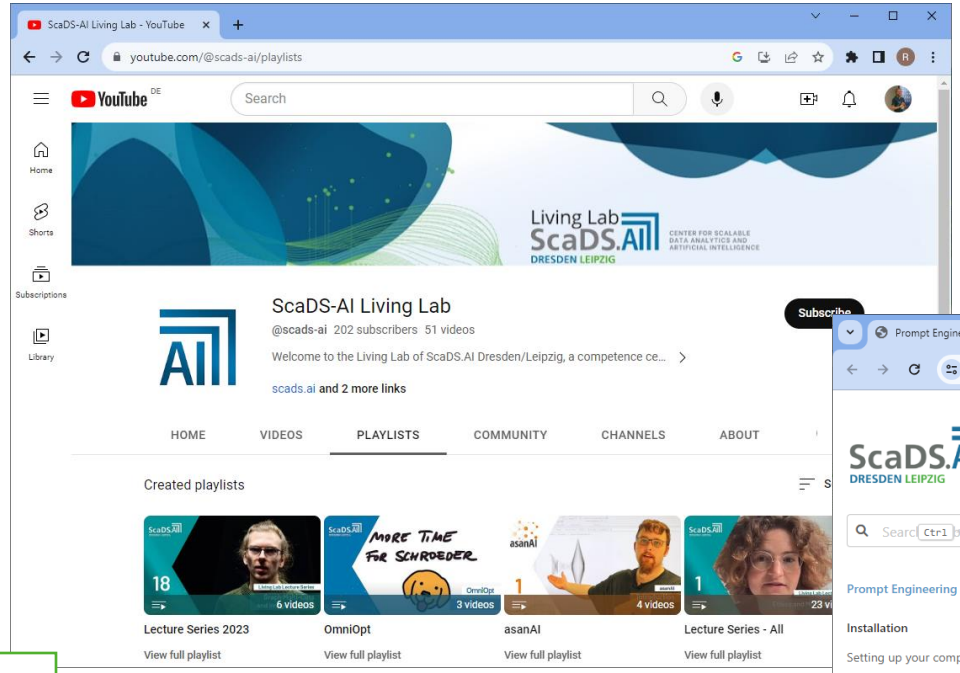
- Your *future-self* will thank you, because they will find your work

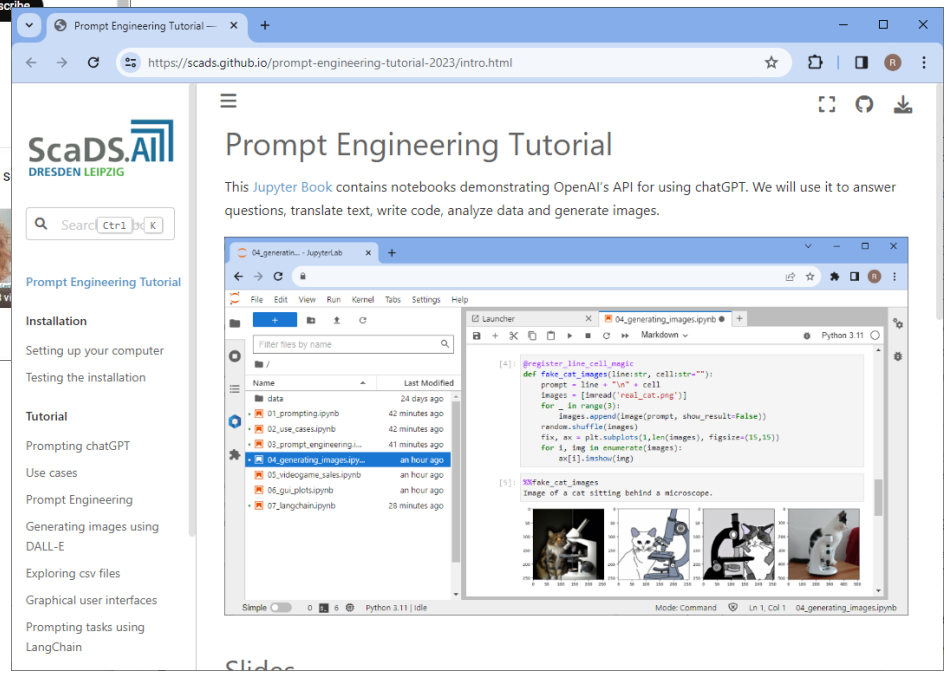




Incentives: Findability -> Visibility

- YouTube
- Github





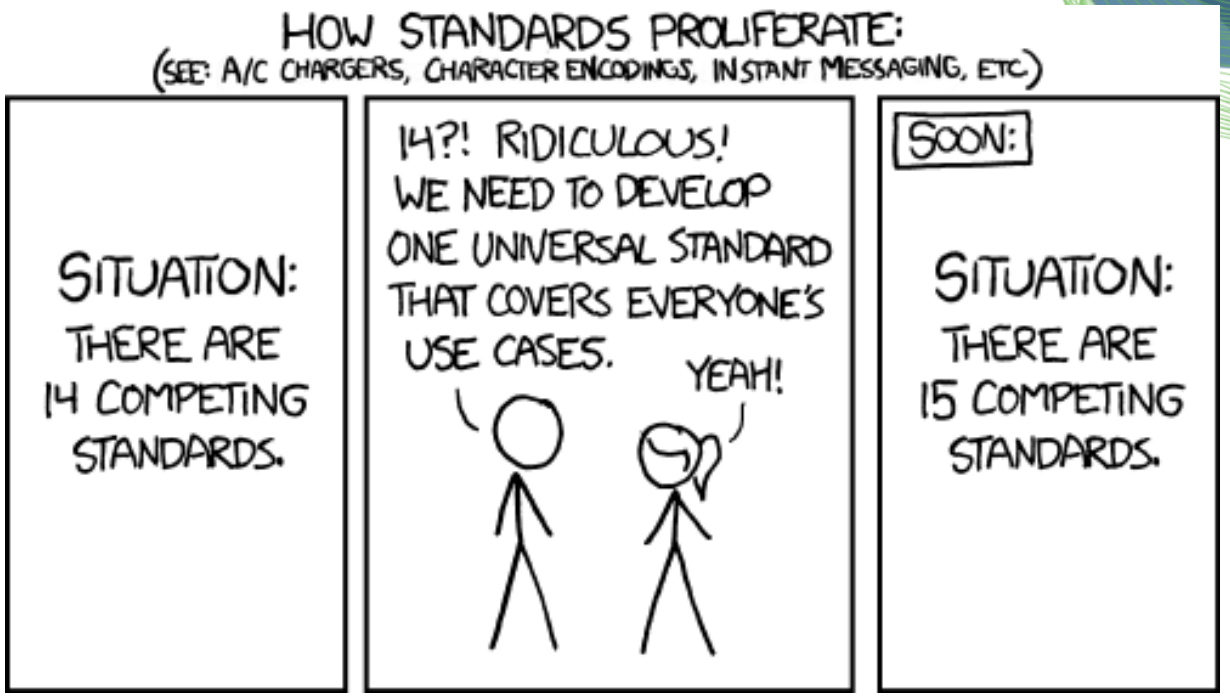
Open & FAIR sharing is a PR instrument

… leading to

- more software users
- new collaborations

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



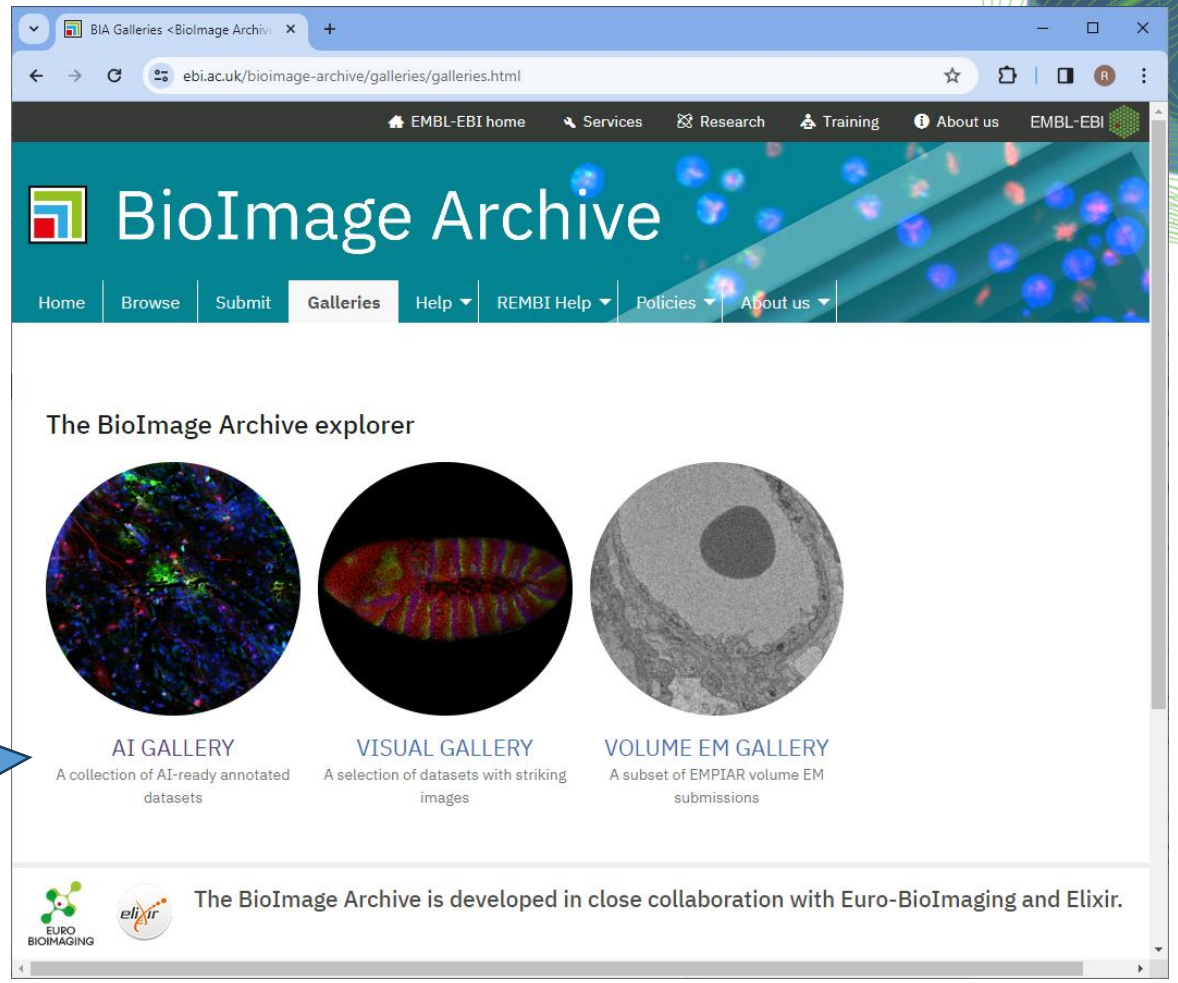
https://www.go-fair.org/fair-principles/

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

Accessibility

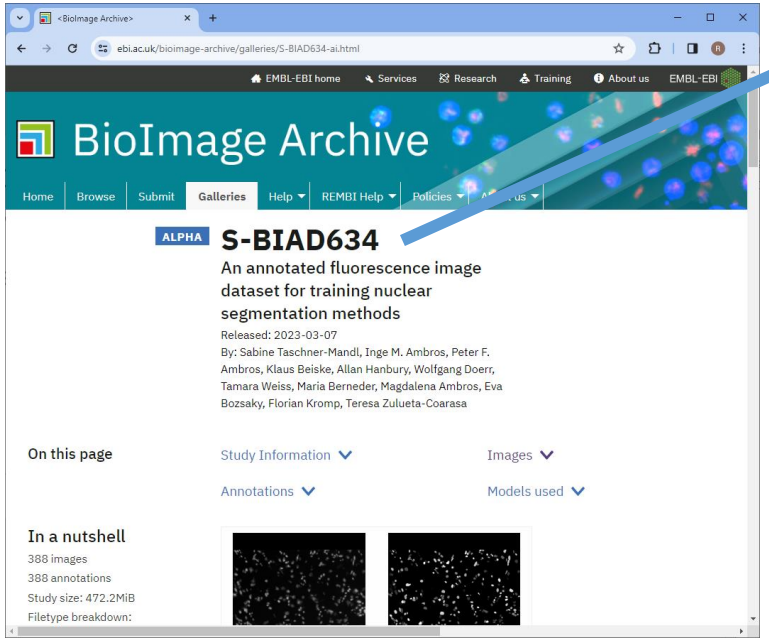
- The ability to download data, for humans and computers

Essential for AI developers =-)



Accessibility

- The ability to download data, for humans and computers



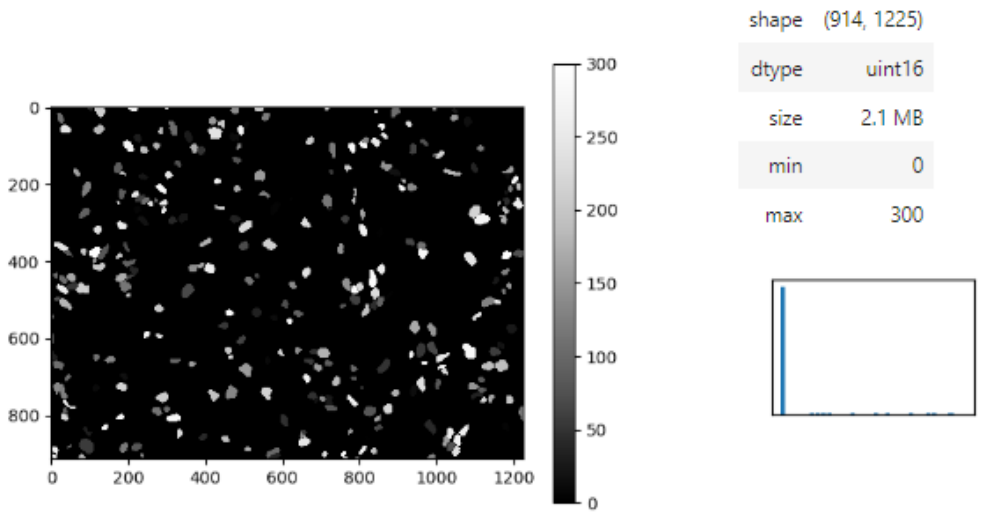

```
[1]: from bia_explorer import io, biostudies
     from skimage.io import imread
     import stackview

     accession = 'S-BIAD634'
     study = io.download_bia_study(accession)
     image = study.images[0]
```

Displaying images using stackview

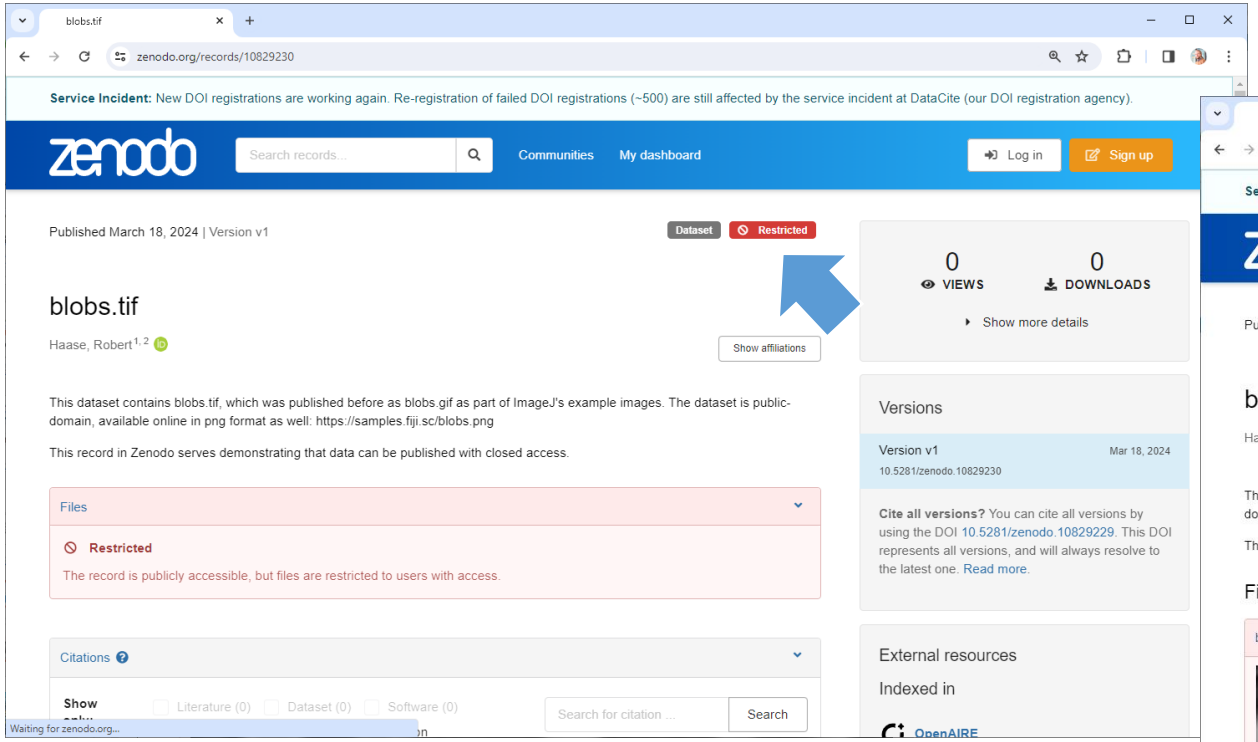
```
[2]: uri = image.uri.replace("\\", "/")
     image_data = imread(uri)
     stackview.insight(image_data)
```

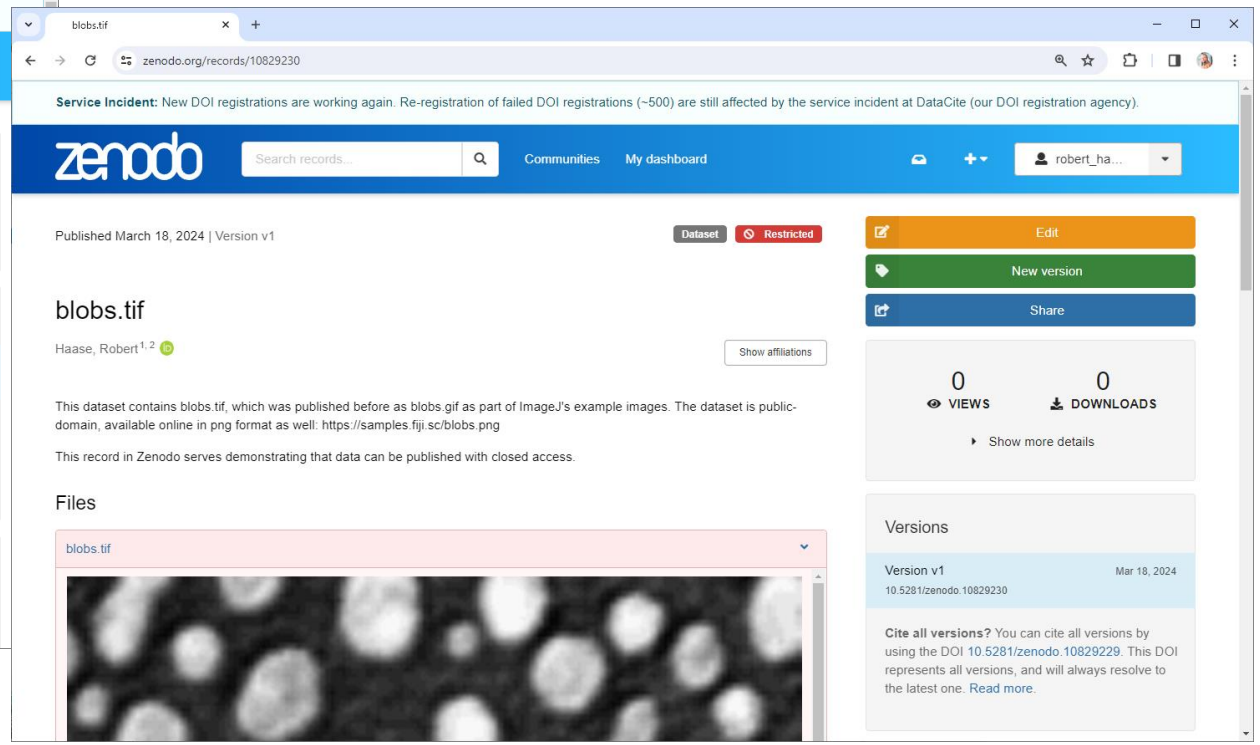
[2]:



Restricted Access

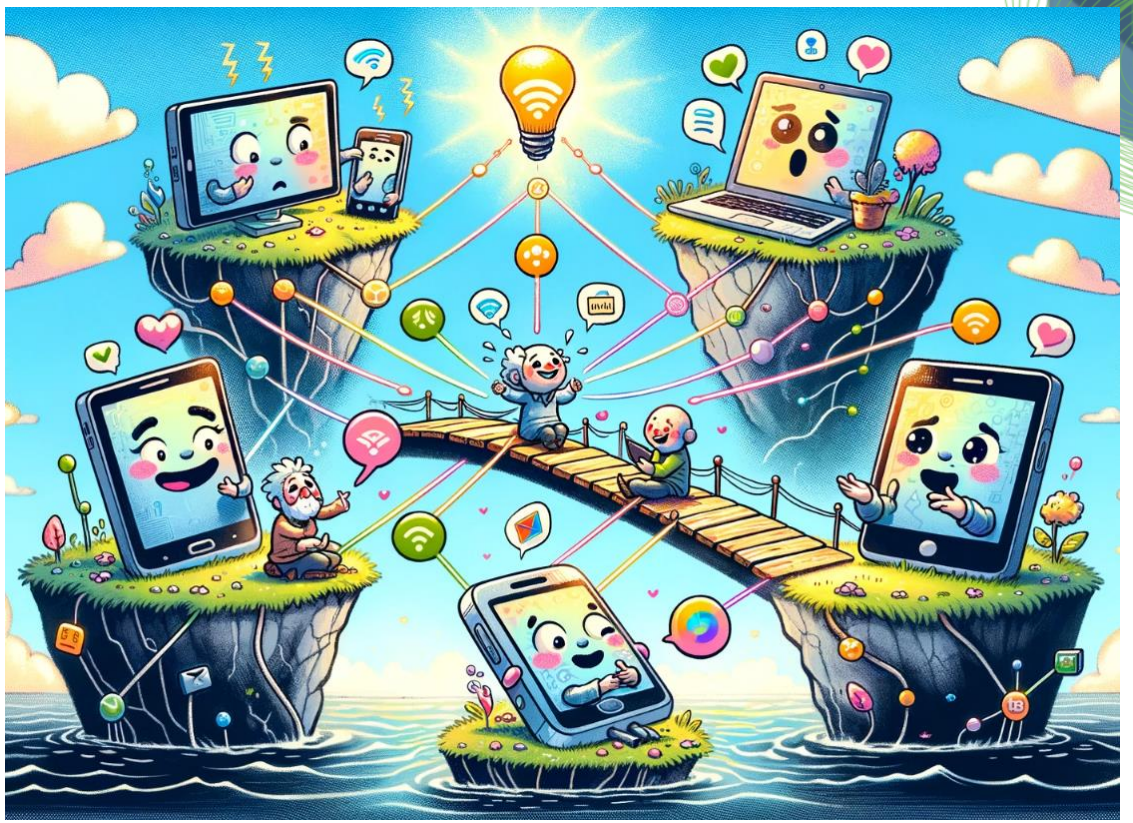
- The A in FAIR does not necessarily stand for Open Access





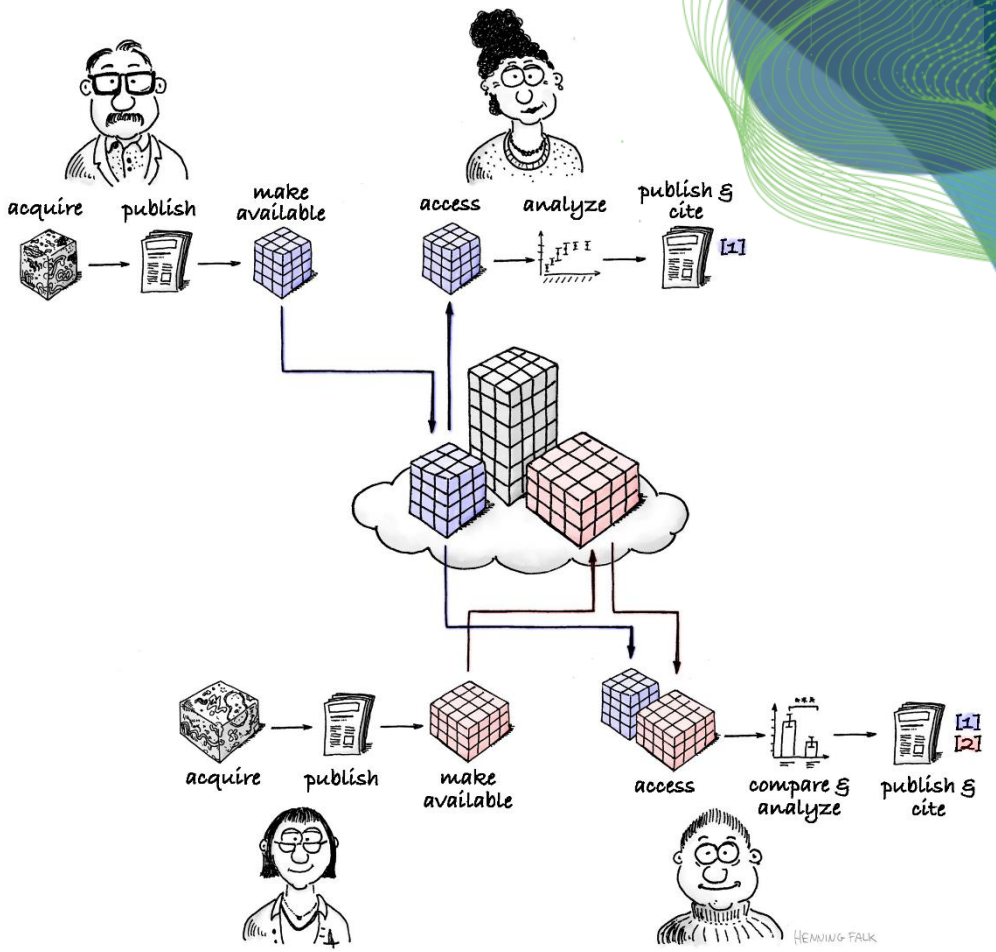
The FAIR-principles

- 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

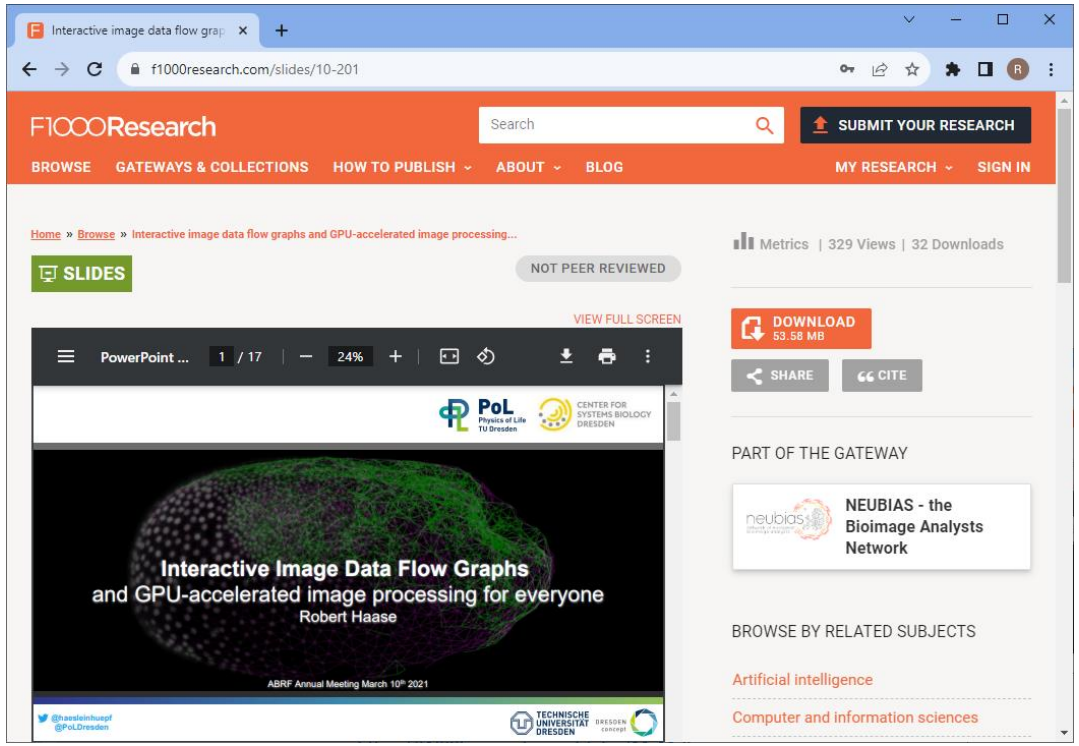


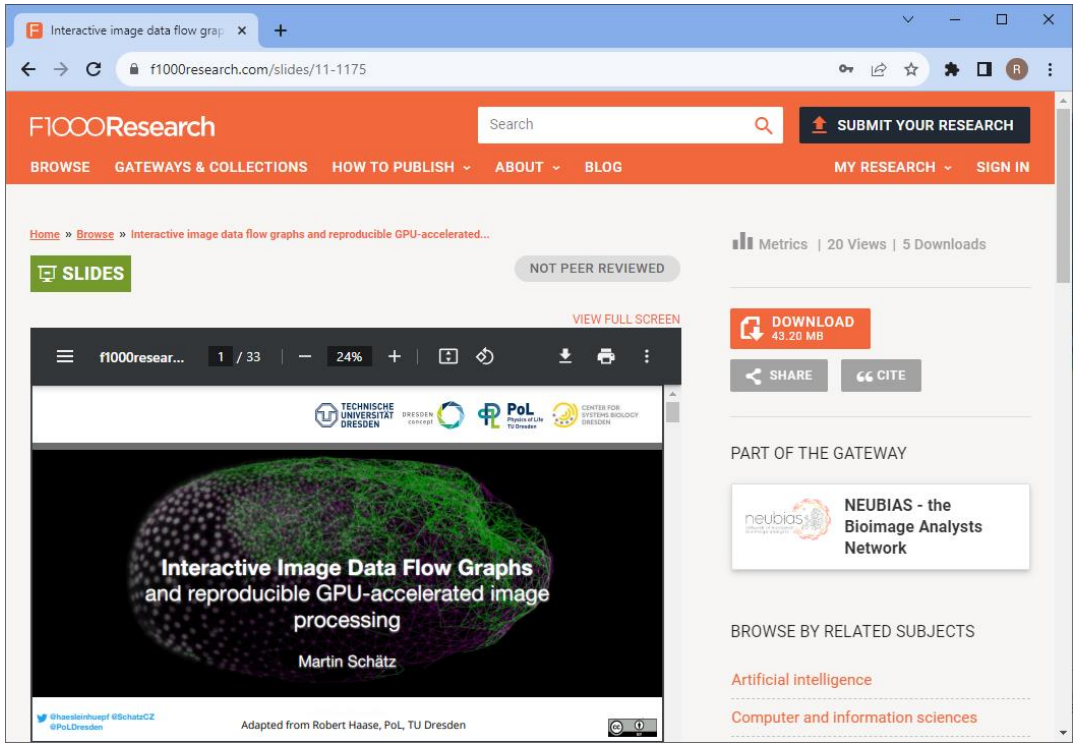
https://www.go-fair.org/fair-principles/

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

Incentives: Reusability

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





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)

Quiz

- Where might open source code be most *visible*?

Git server of the university

Zenodo.org

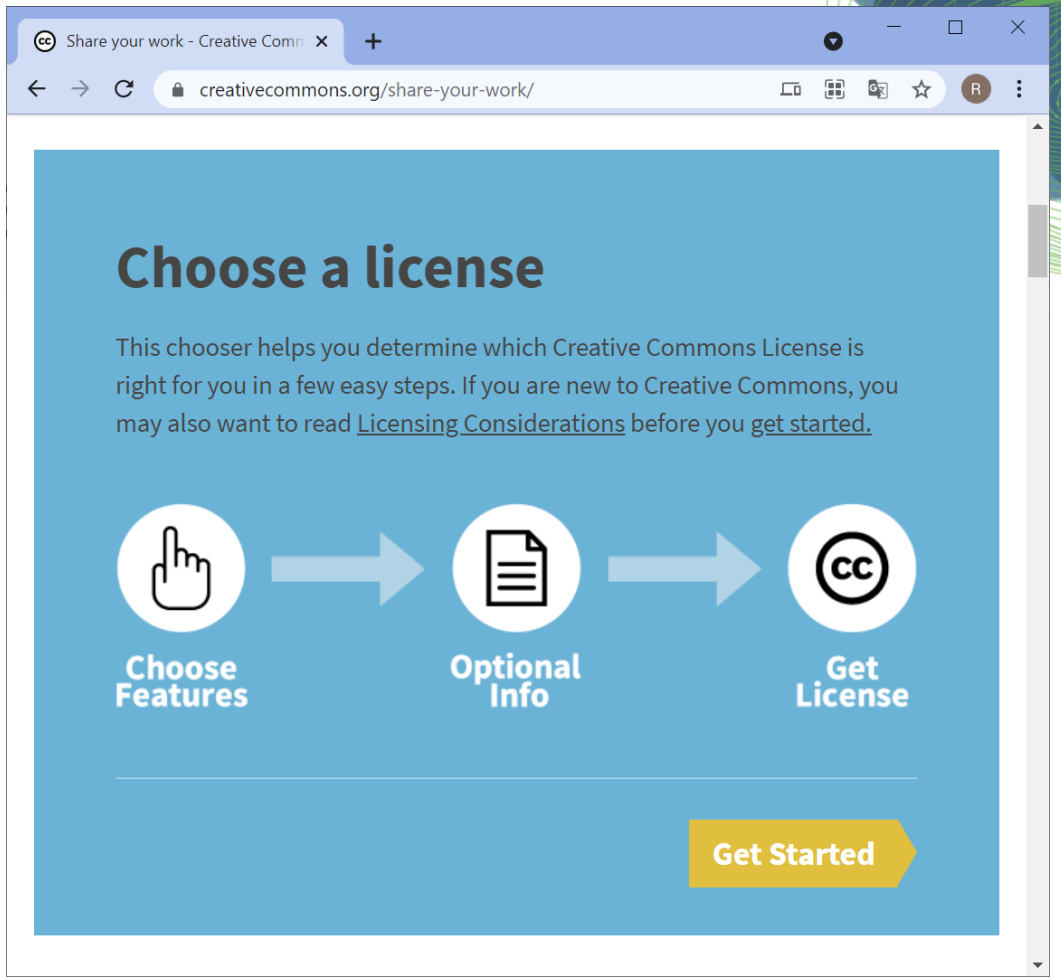
Github.com

Group / institute / personal website

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



Content on this site is licensed under a Creative Commons Attribution 4.0 International license. Icons by The Noun Project.

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* you work.

Content on this slide was adapted from https://creativecommons.org/about/cclicenses/ which is licensed under a Creative Commons Attribution 4.0 International license. Icons by The Noun Project.

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.

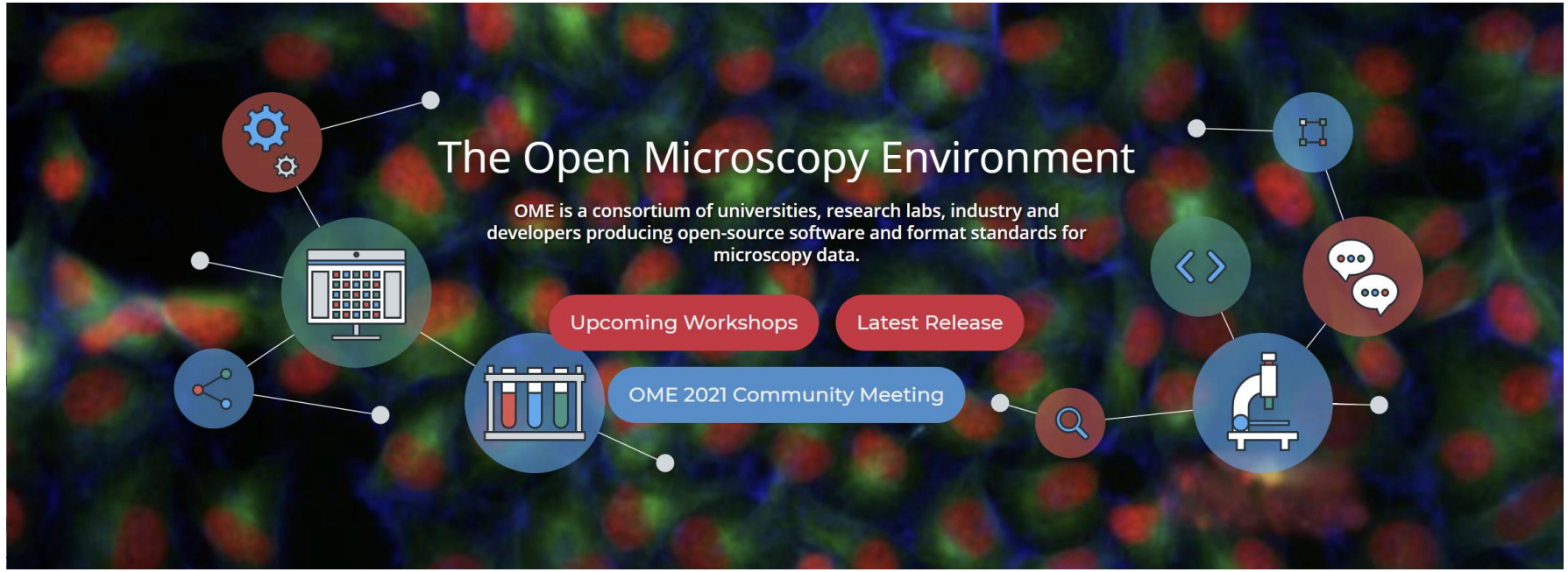
CC BY includes the following elements:

BY – Credit must be given to the creator

Content on this slide was adated from https://creativecommons.org/about/cclicenses/ which is licensed under a Creative Commons Attribution 4.0 International license. Icons by The Noun Project.

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

Content on this slide was adapted from https://creativecommons.org/about/cclicenses/ which is licensed under a Creative Commons Attribution 4.0 International license. Icons by The Noun Project.

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

Content on this slide was adapted from https://creativecommons.org/about/cclicenses/ which is licensed under a Creative Commons Attribution 4.0 International license. Icons by The Noun Project.

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

Content on this slide was adapted from https://creativecommons.org/about/cclicenses/ which is licensed under a Creative Commons Attribution 4.0 International license. Icons by The Noun Project.

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

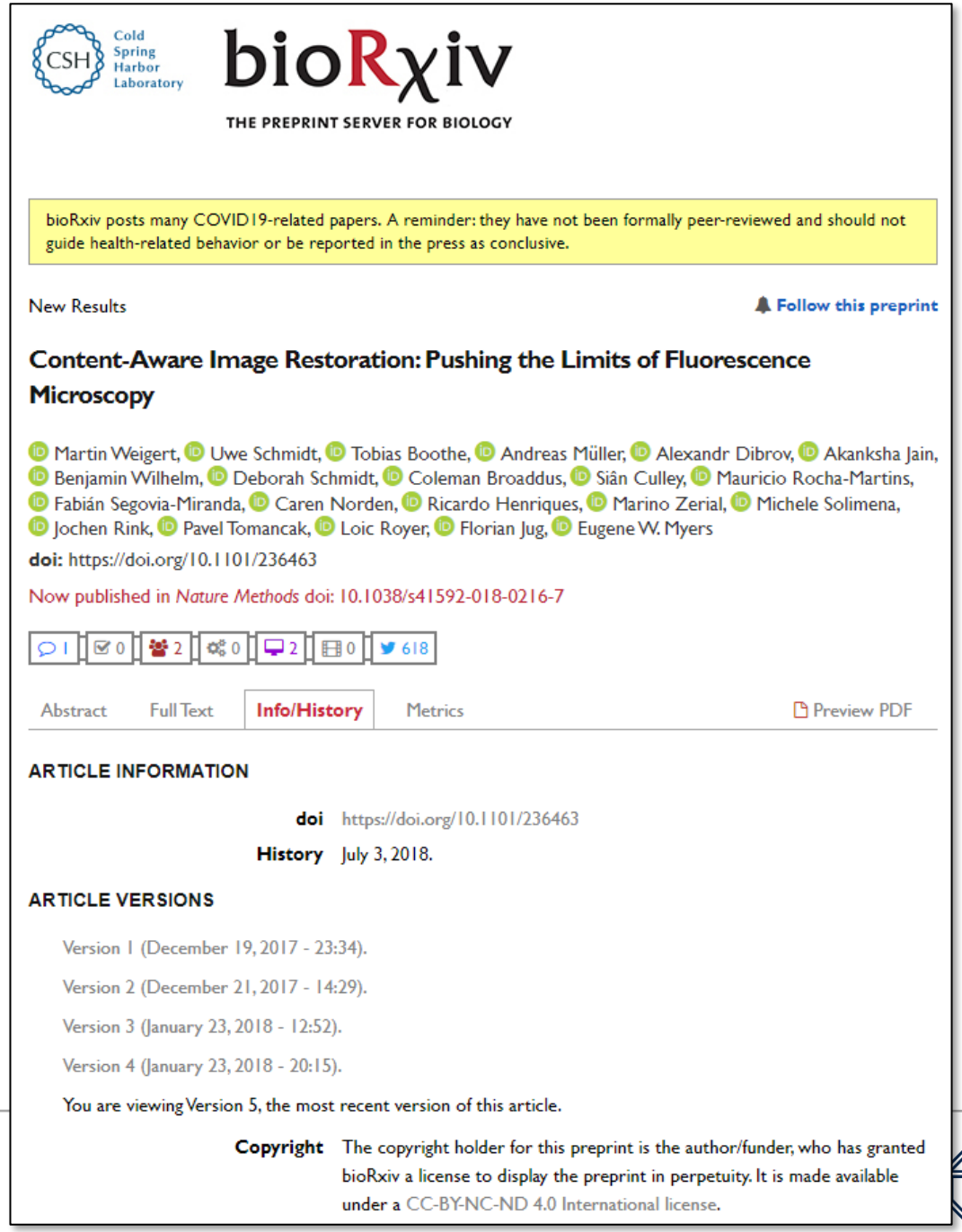
"Restrictive" licensing

Content on this slide was adapted from https://creativecommons.org/about/cclicenses/ which is licensed under a Creative Commons Attribution 4.0 International license. Icons by The Noun Project.

Quiz

- May I use one of the Figures from this preprint?

- May I download and redistribute this preprint to students of a course for free?



Cold Spring Harbor Laboratory

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results

Follow this preprint

Content-Aware Image Restoration: Pushing the Limits of Fluorescence Microscopy

Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, Mauricio Rocha-Martins, Fabián Segovia-Miranda, Caren Norden, Ricardo Henriques, Marino Zerial, Michele Solimena, Jochen Rink, Pavel Tomancak, Loic Royer, Florian Jug, Eugene W. Myers

doi: https://doi.org/10.1101/236463

Now published in *Nature Methods* doi: 10.1038/s41592-018-0216-7

1 0 2 0 2 0 618

Abstract Full Text **Info/History** Metrics Preview PDF

ARTICLE INFORMATION

doi https://doi.org/10.1101/236463

History July 3, 2018.

ARTICLE VERSIONS

Version 1 (December 19, 2017 - 23:34).

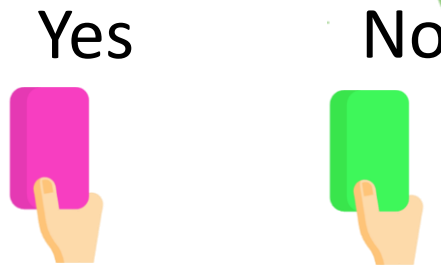
Version 2 (December 21, 2017 - 14:29).

Version 3 (January 23, 2018 - 12:52).

Version 4 (January 23, 2018 - 20:15).

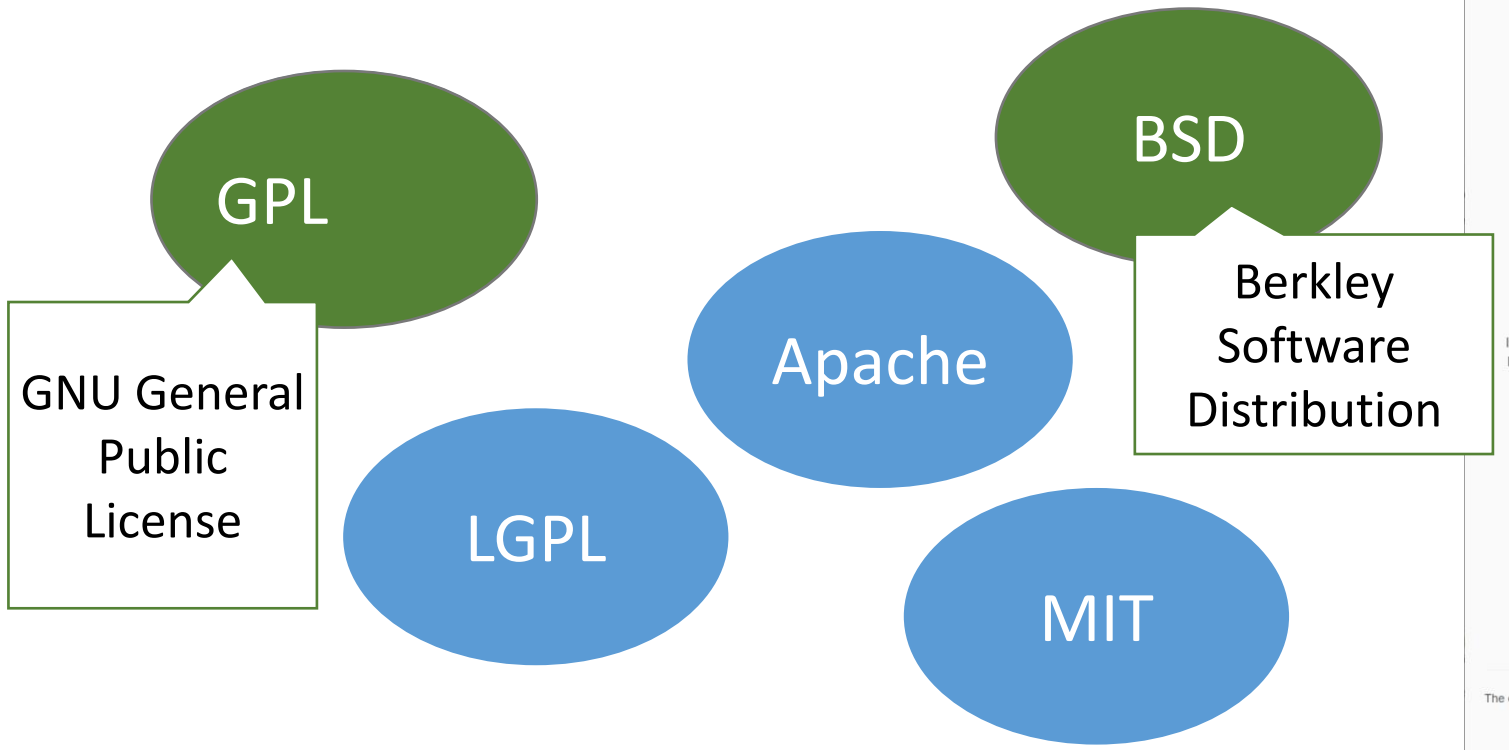
You are viewing Version 5, the most recent version of this article.

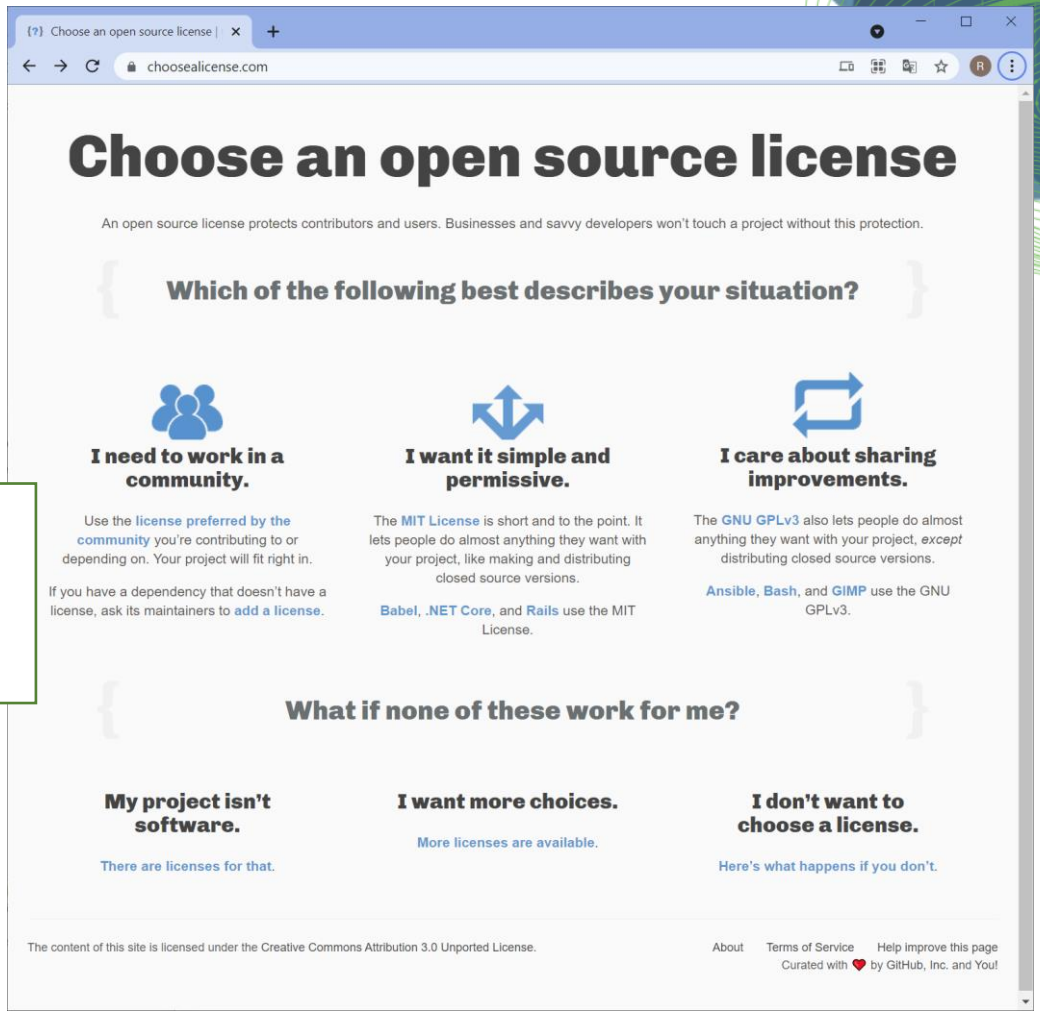
Copyright The copyright holder for this preprint is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



Licensing Software

In the software world, other licenses are more popular, historically grown.

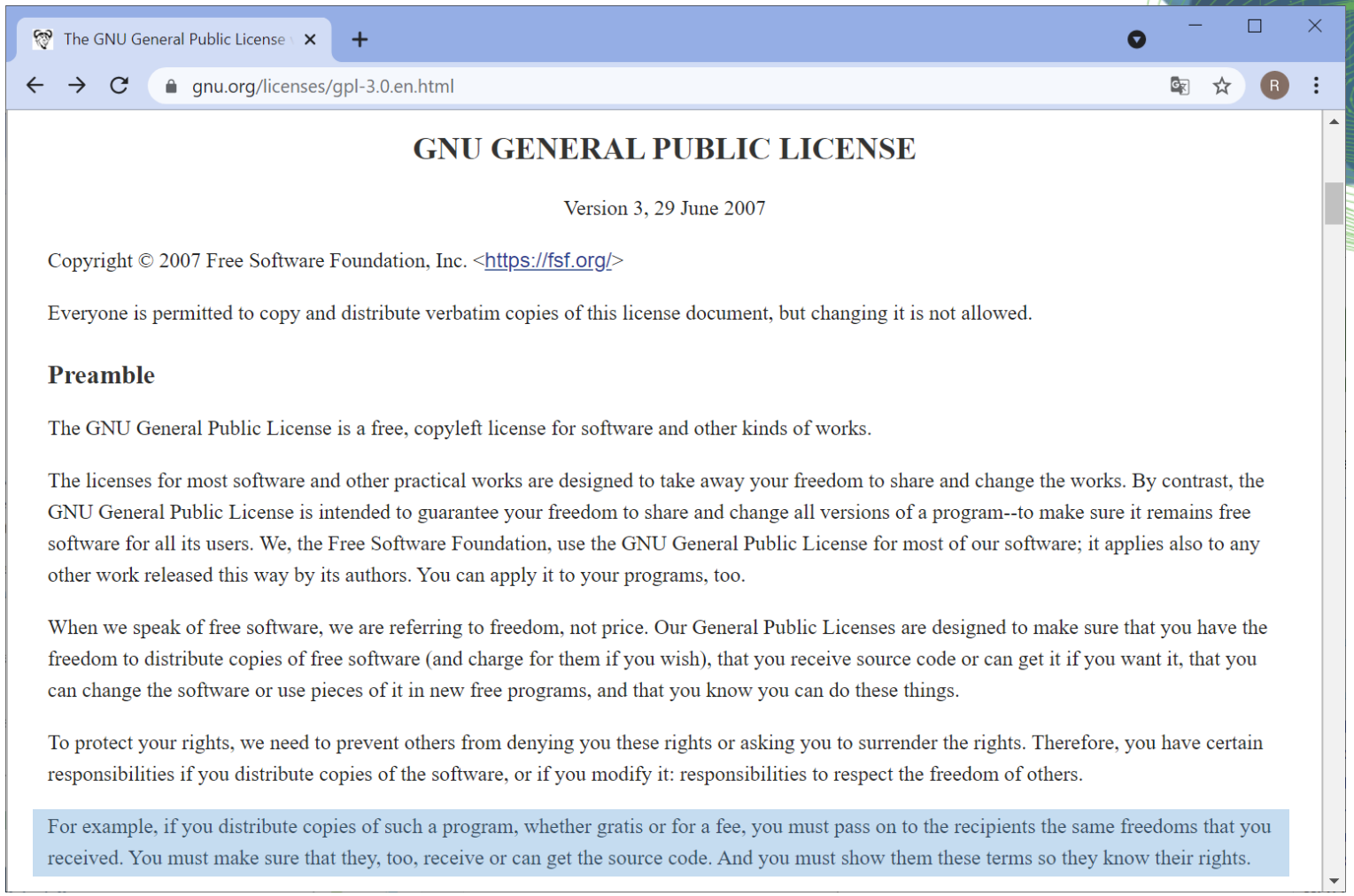




Licensing Software: GPL

GPL-derivatives must also be GPL-licensed

"Restrictive" licensing



Quiz

Can I build a commercial product on the basis of GPL-licensed code?

Yes

No

Do I have to release the code openly for this commercial product?

Yes

No

Licensing Software: BSD0

Copyright (C) [year] by [copyright holder] <[email]>

Permission to use, copy, modify, and/or distribute this software for any purpose with or without fee is hereby granted.

Similar to CC0 (public domain)

THE SOFTWARE IS PROVIDED "AS IS" AND THE AUTHOR DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY SPECIAL, DIRECT, INDIRECT, OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE

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

Licensing Software: BSD0

Copyright (C) [year] by [copyright holder] <[email]>

Permission to use, copy, modify, and/or distribute this software for any purpose with or without fee is hereby granted.

THE SOFTWARE IS PROVIDED "AS IS" AND THE AUTHOR DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY SPECIAL, DIRECT, INDIRECT, OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE

Disclaimer

Whatever you do with it, we [the authors] are not liable

Licensing Software: BSD2

Copyright (c) <year>, <copyright holder>
Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE

Similar to CC-BY

Licensing Software: BSD3

Copyright <year> <copyright holder>
Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:
1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.
THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE

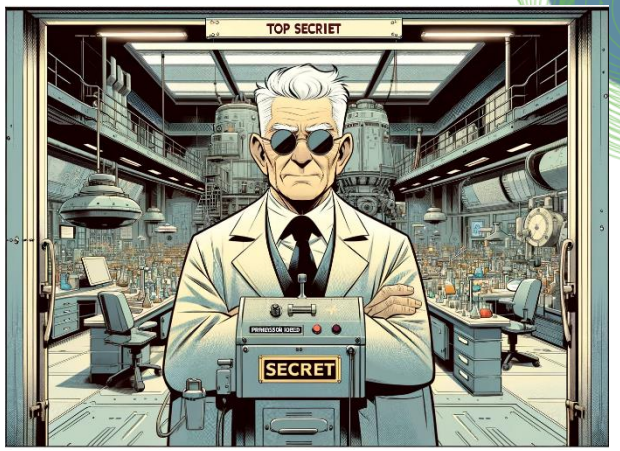
Similar to CC-BY

You must not use the copyright holder's name to endorse your derivative of the work.

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



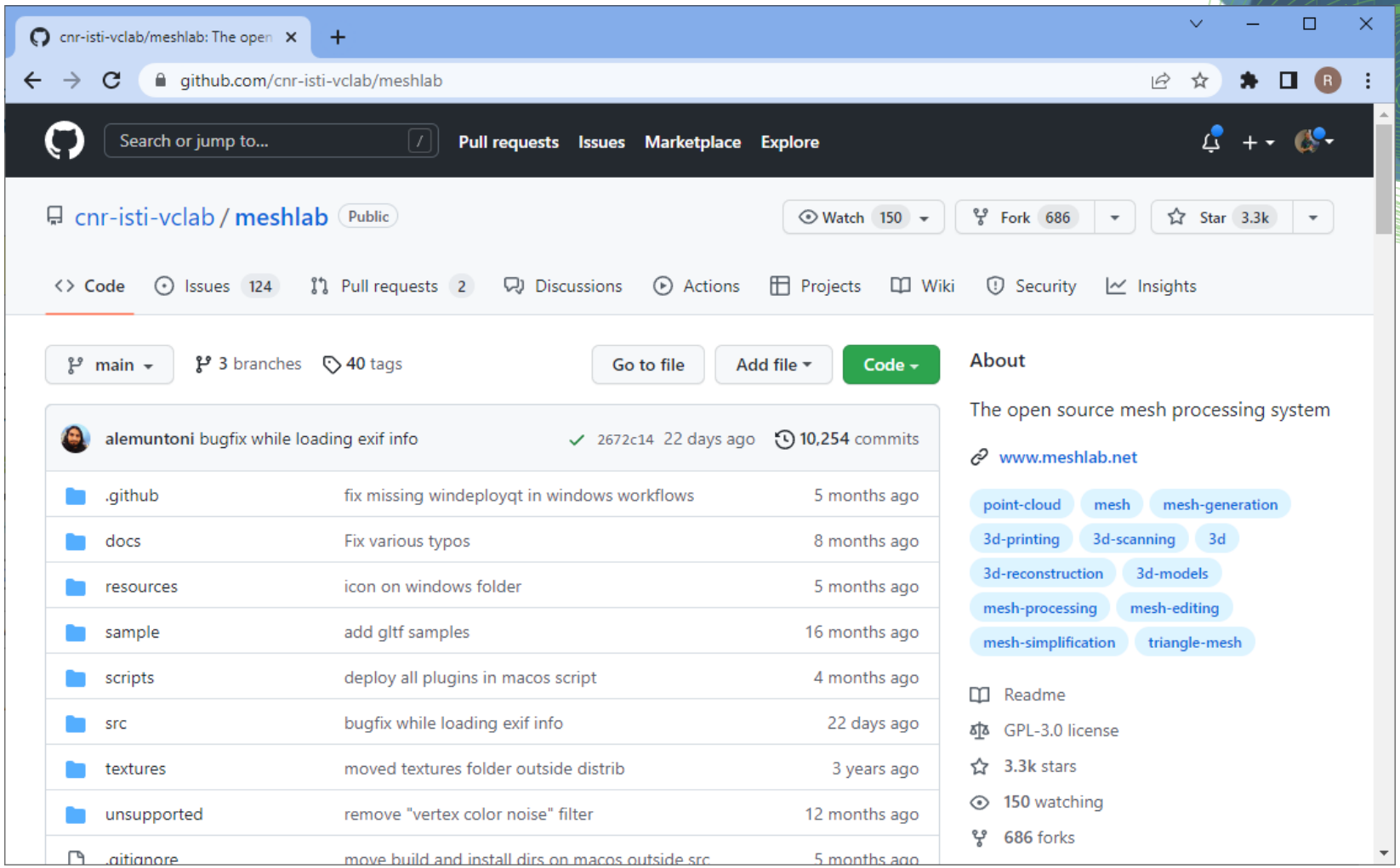
- Permissive licensing:
  - Do whatever you like with our stuff, just make sure to mention / cite us …
  - Examples: BSD, MIT, Apache, CC-BY

Quiz

May I reuse code from this repository in my own BSD-licensed work?

Yes

No

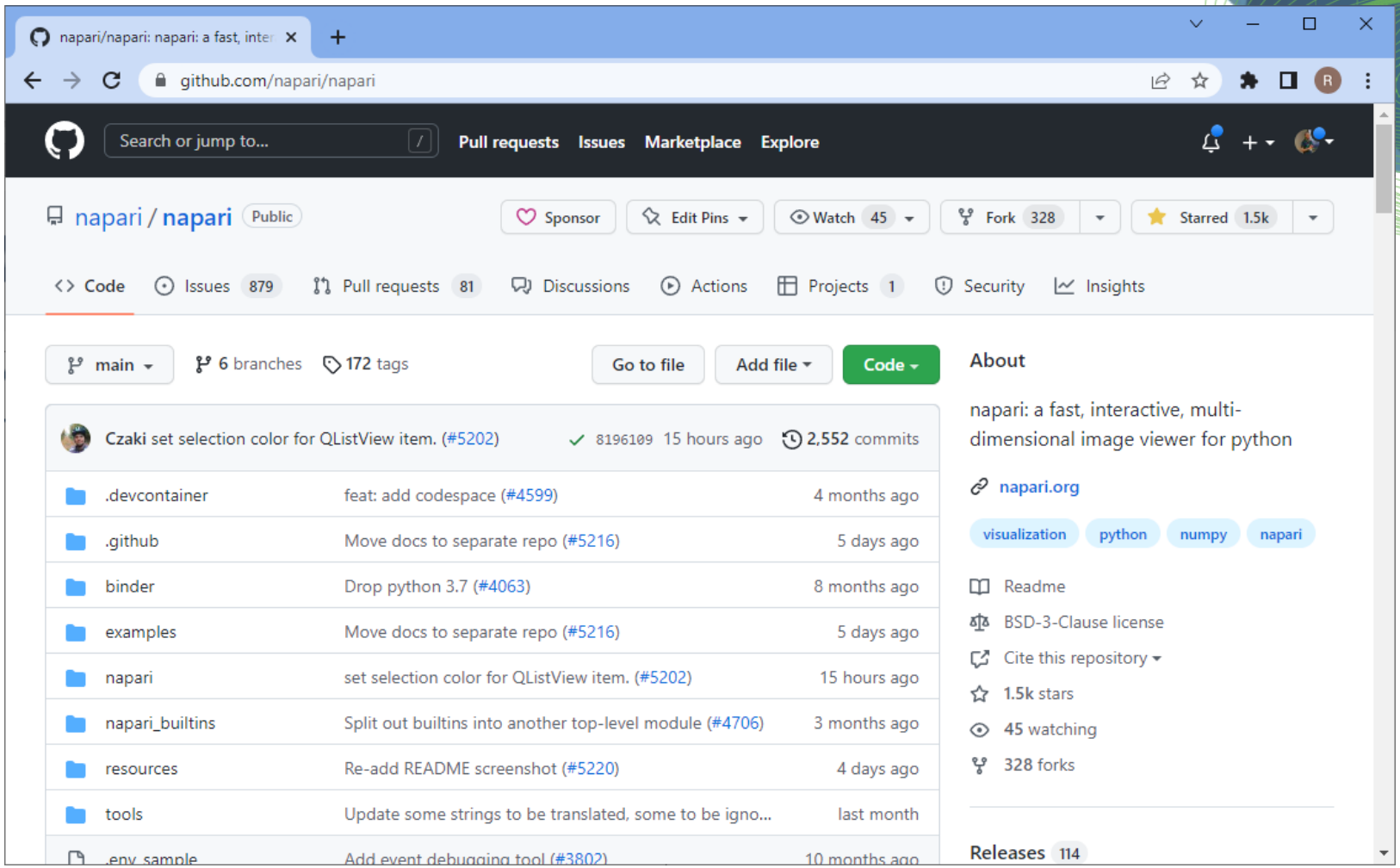


Quiz

May I reuse code from this repository in my own GPL-licensed work?

Yes

No



ScaDS.AI
DRESDEN LEIPZIG
CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE

Software environments

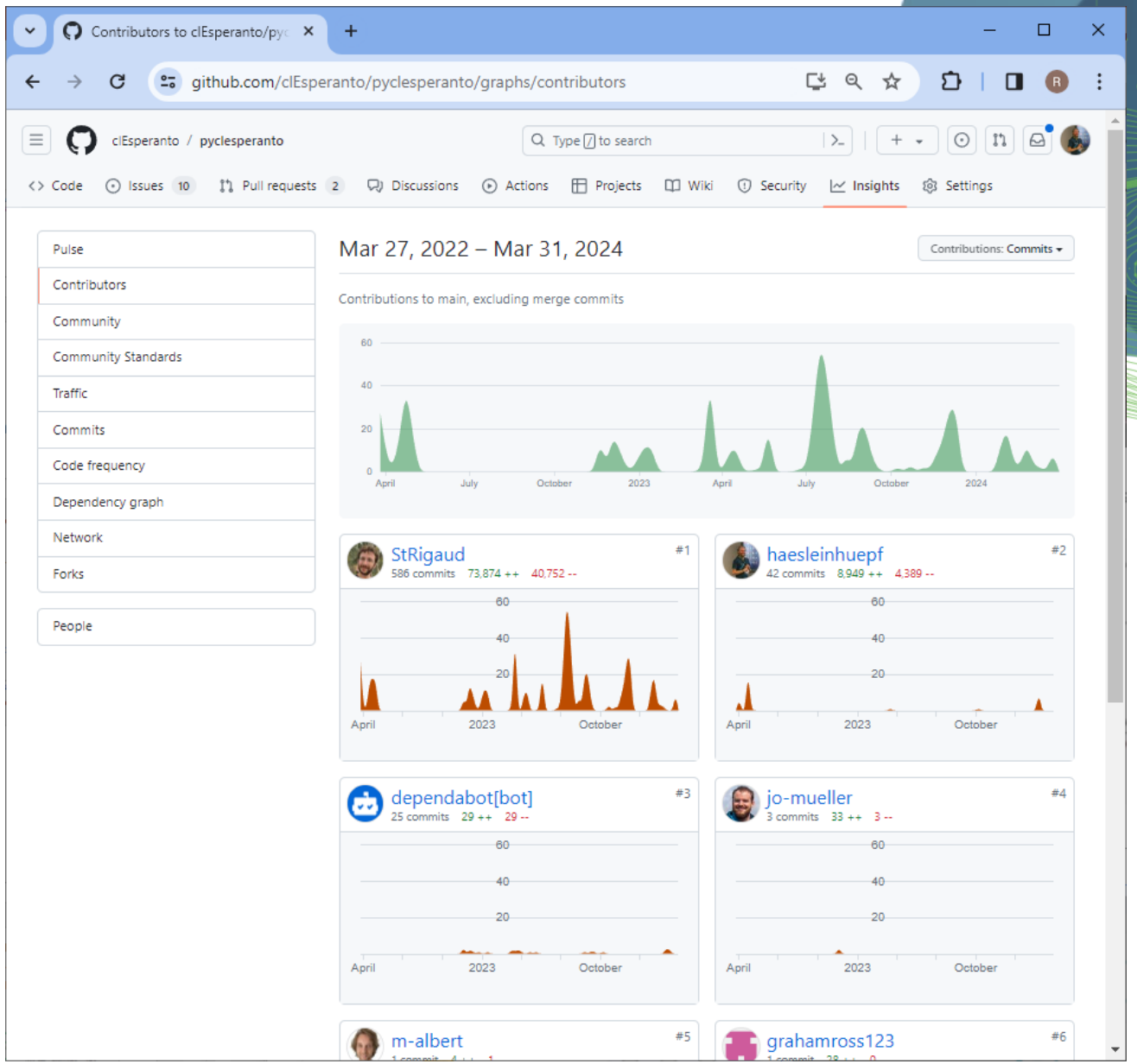
Robert Haase

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.

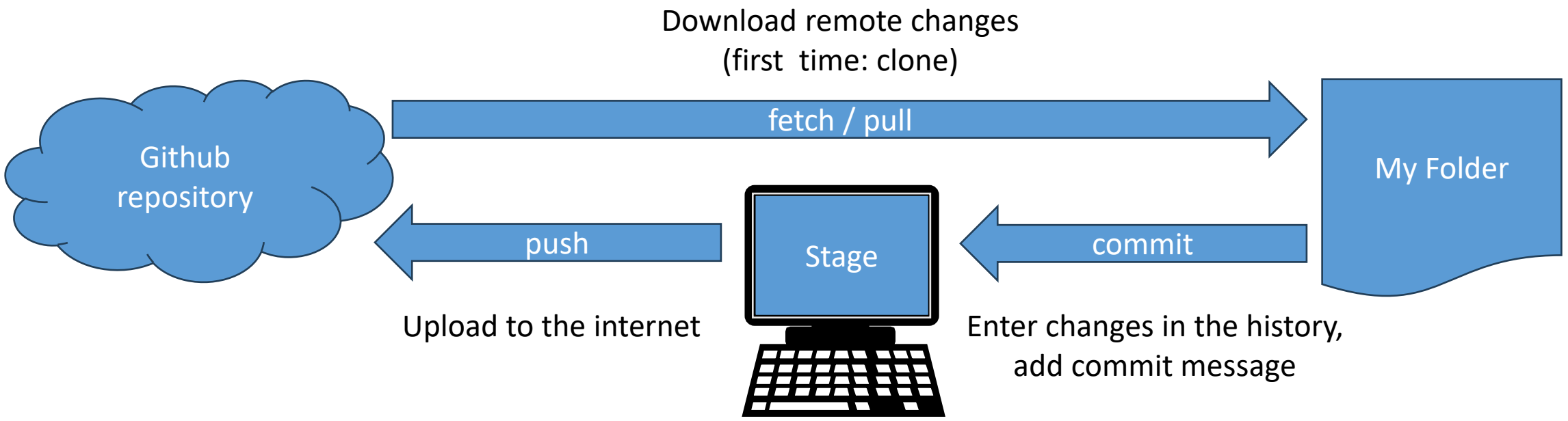
git

- Version control is key element of data scientist's toolbox
- Distributed file system with sophisticated logging mechanisms
- Control about what becomes part of a repository and what not



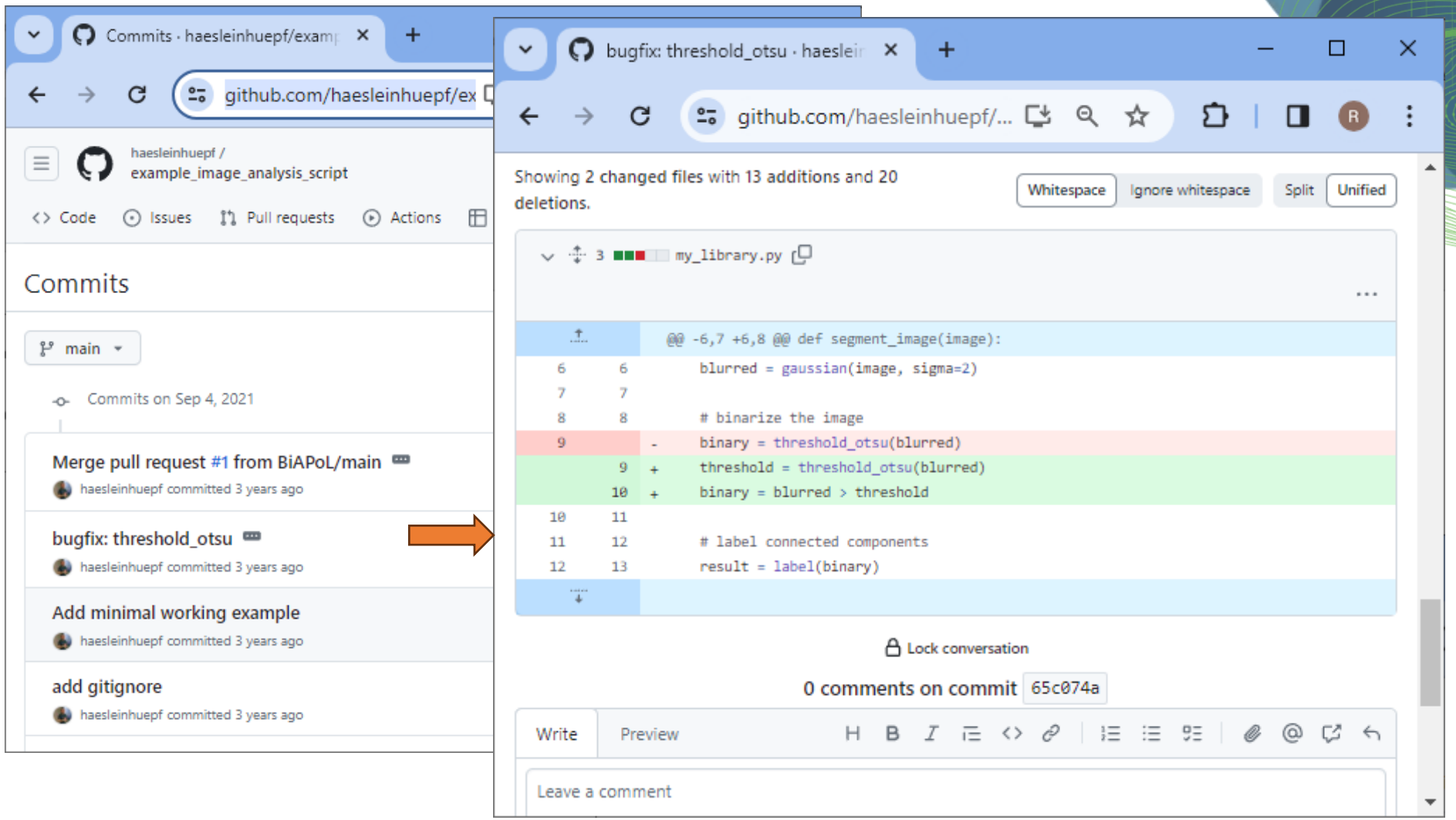
git

- Git makes file modifications a more active / involved process (making people think about)



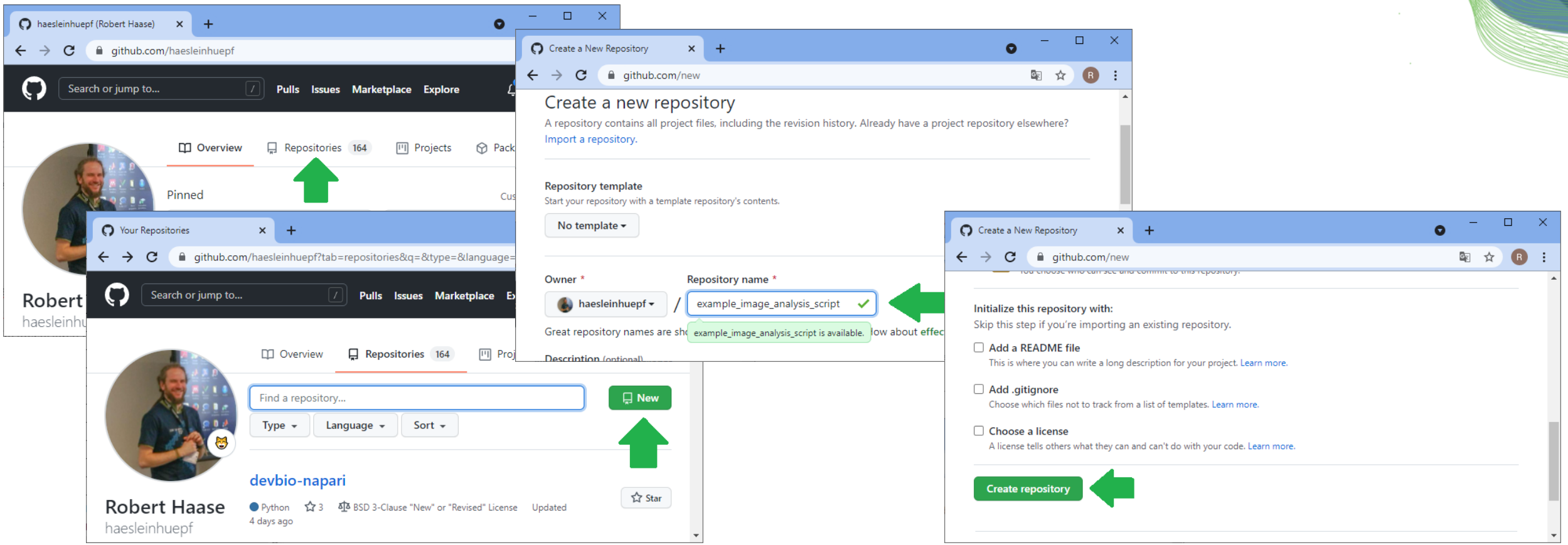
git

- Who wrote this code
- when and
- why?



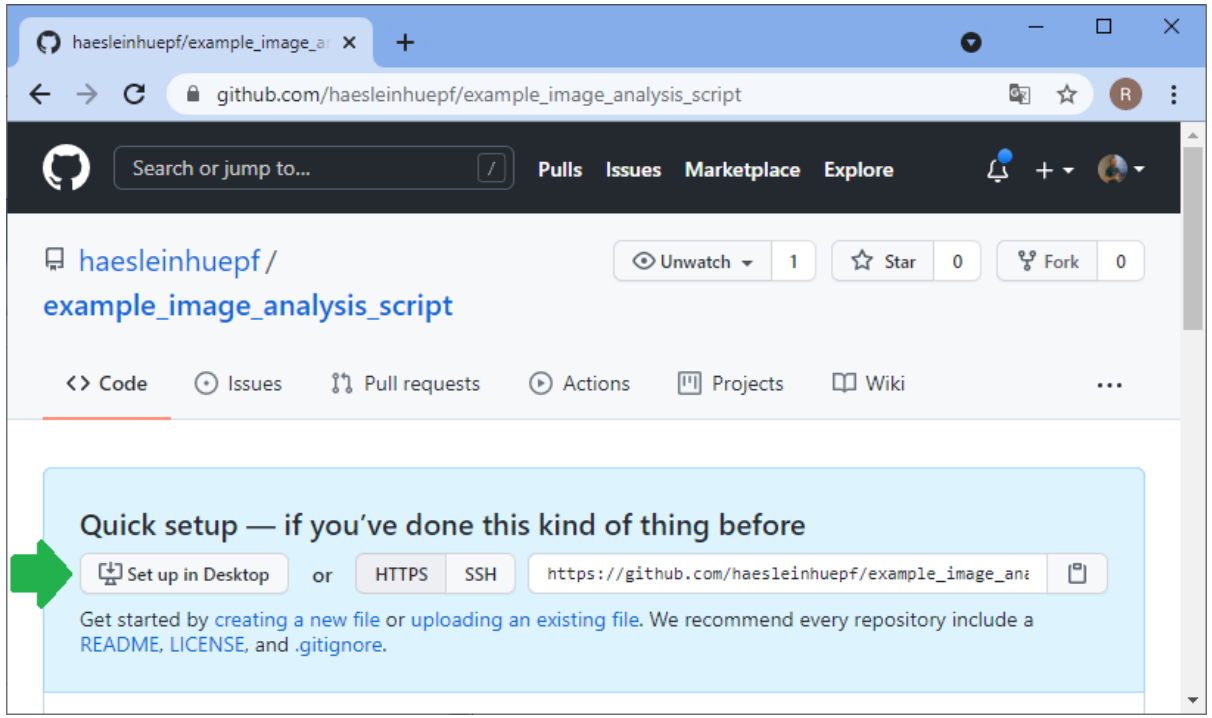
github – creating repositories

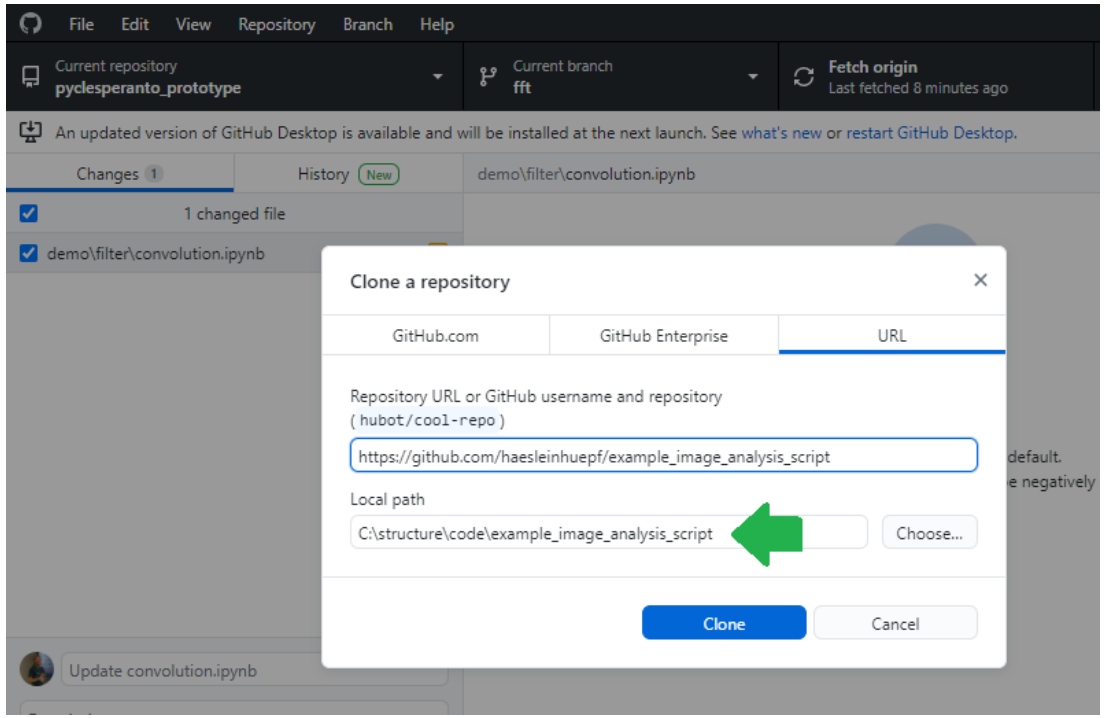
- Add a new, empty repository



github – clone repositories

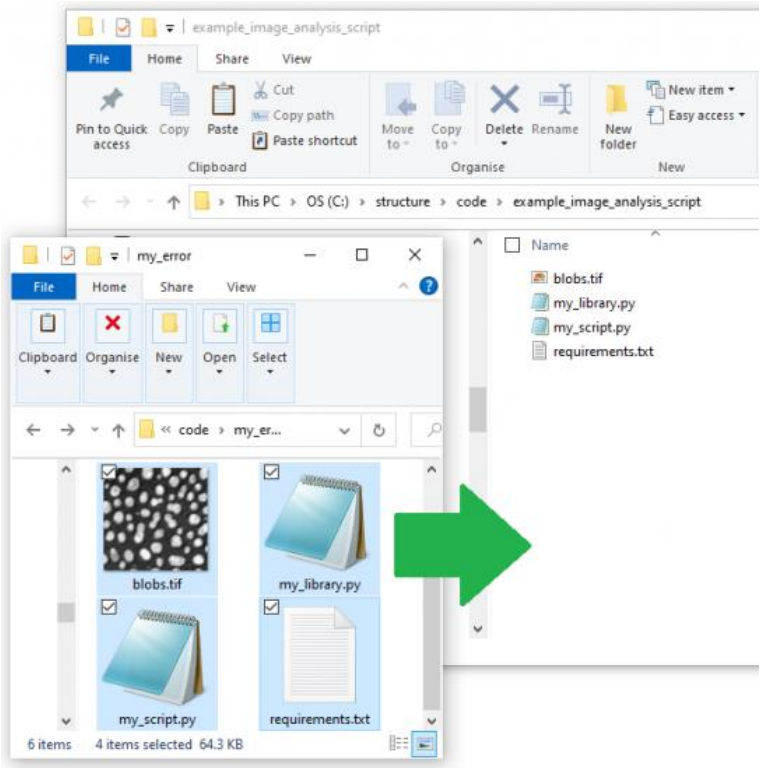
- `git clone https://github.com/organization/repository`
- Or: Use the Github Desktop app

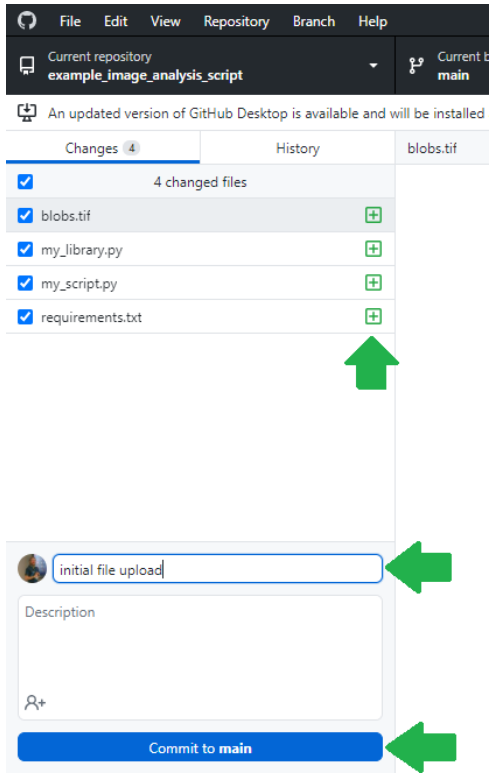


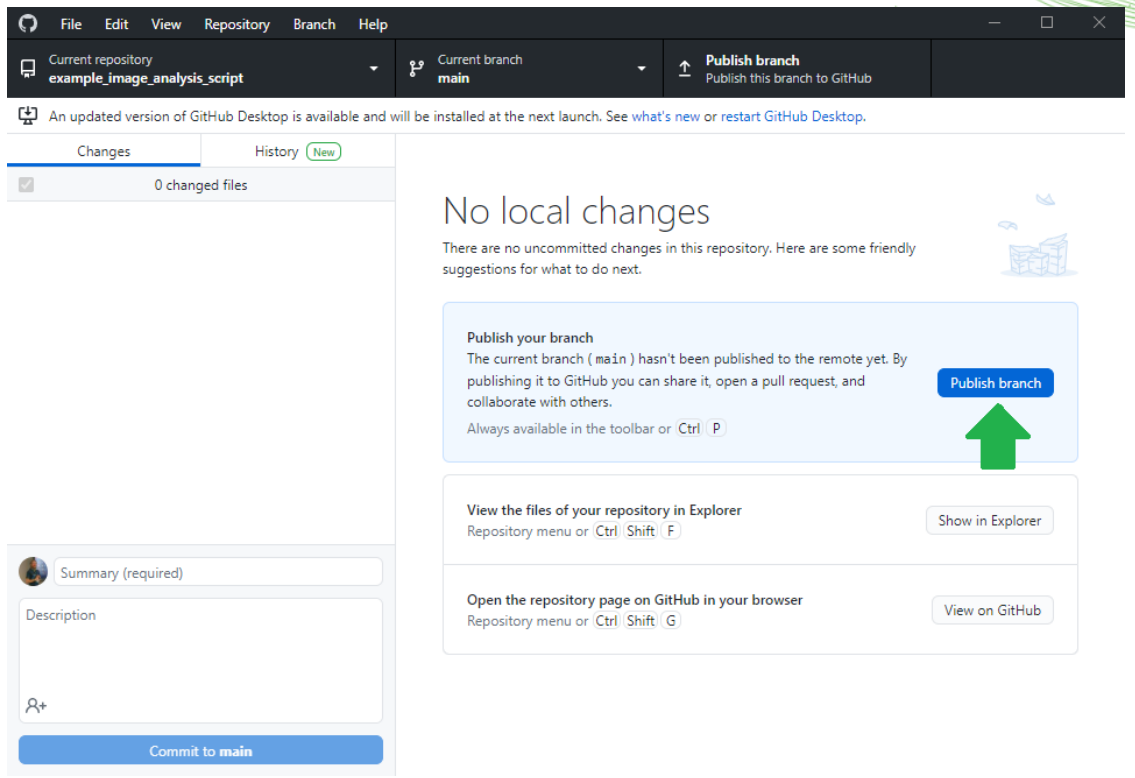


github - uploading

- git [add], commit, push

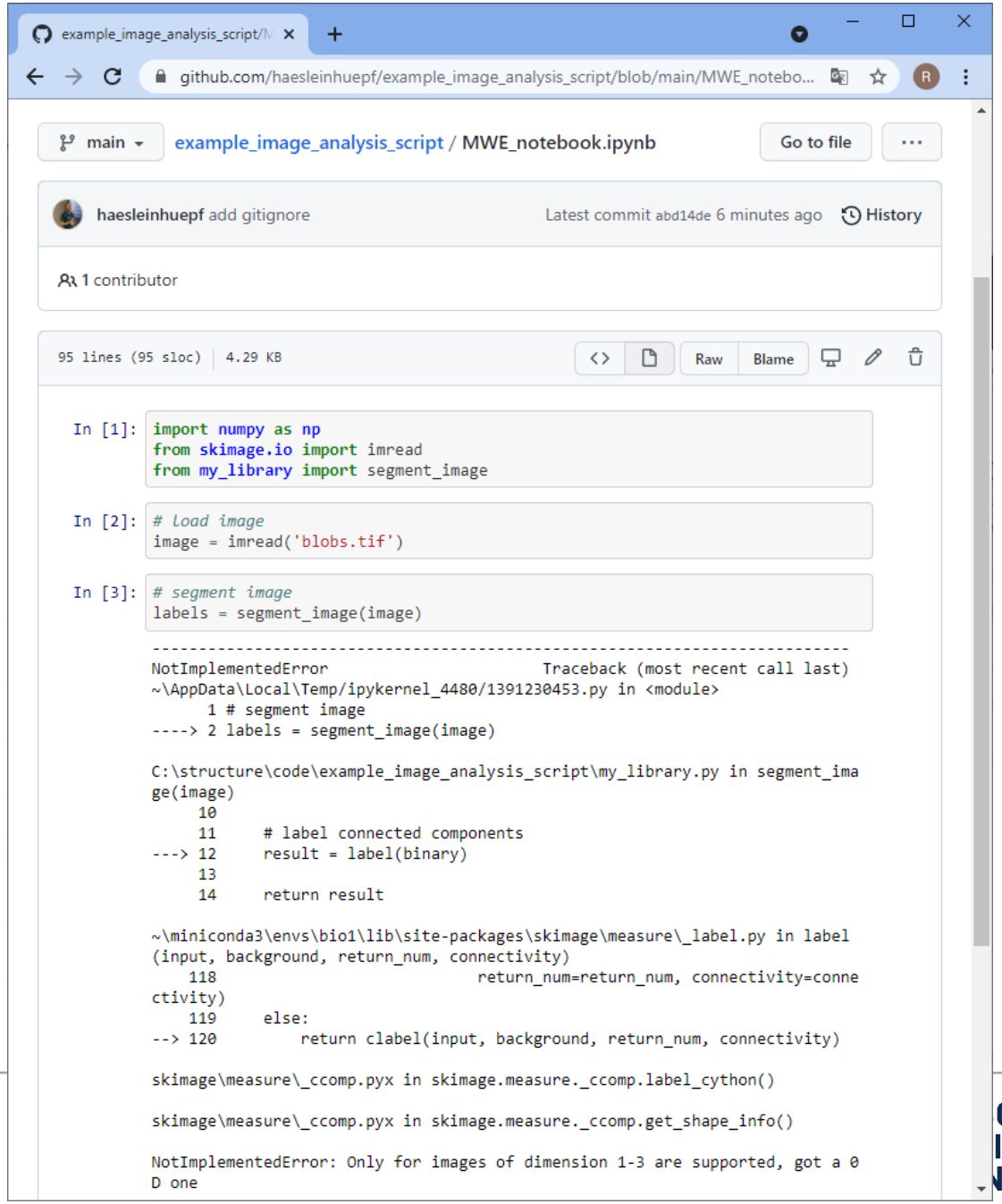






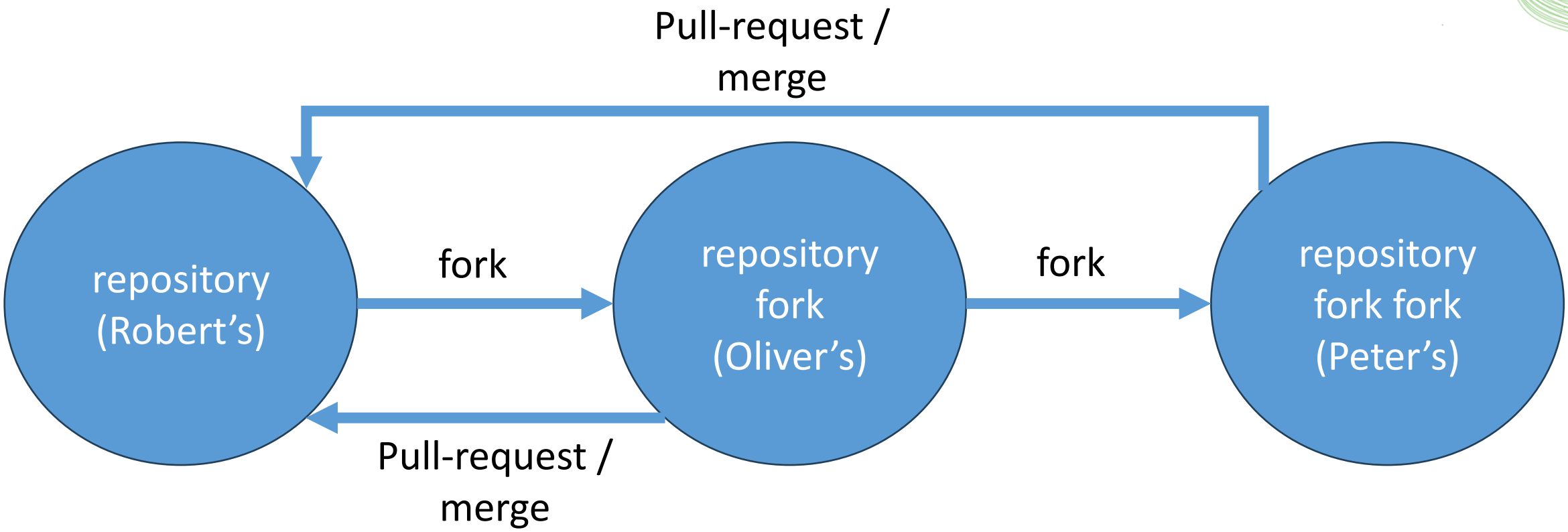
github

- Ease of reading notebooks online
- No need to download and execute code



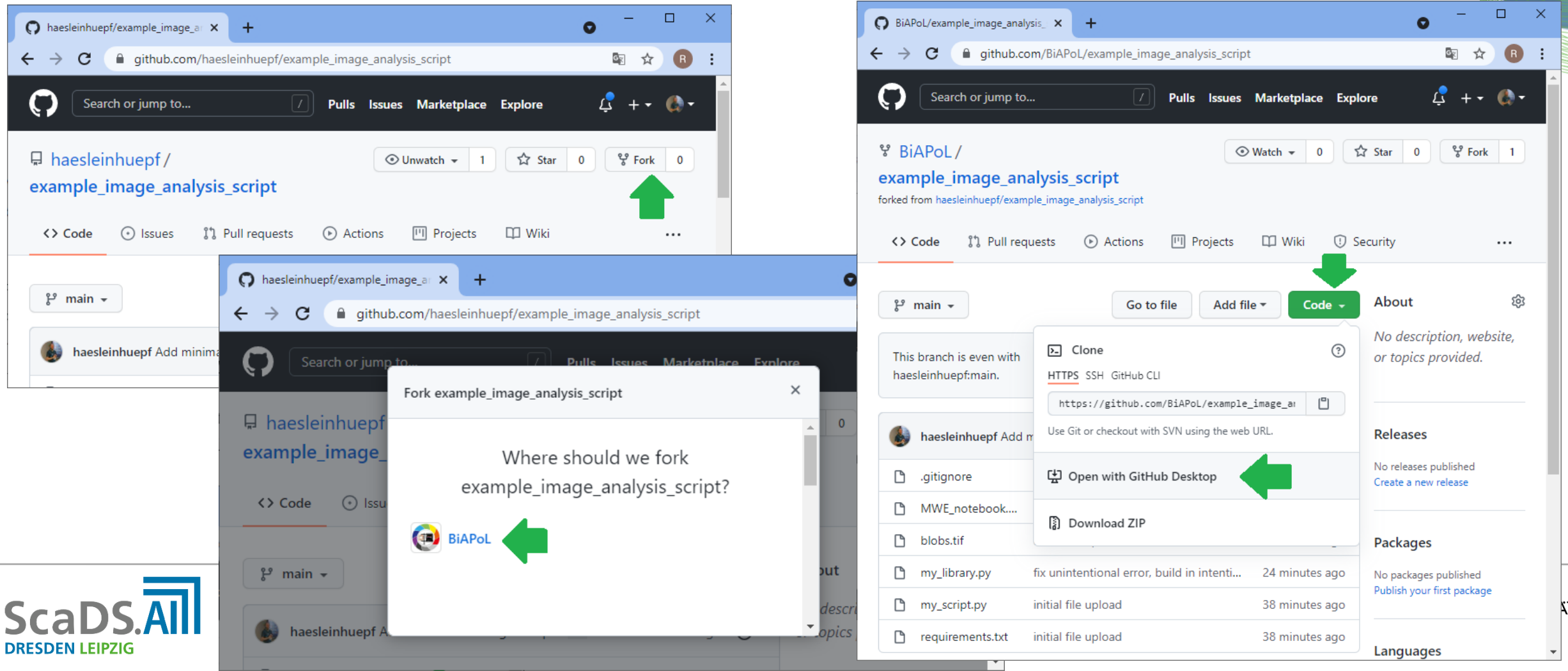
git - forking

- Making a copy where we have edit rights



github - forking

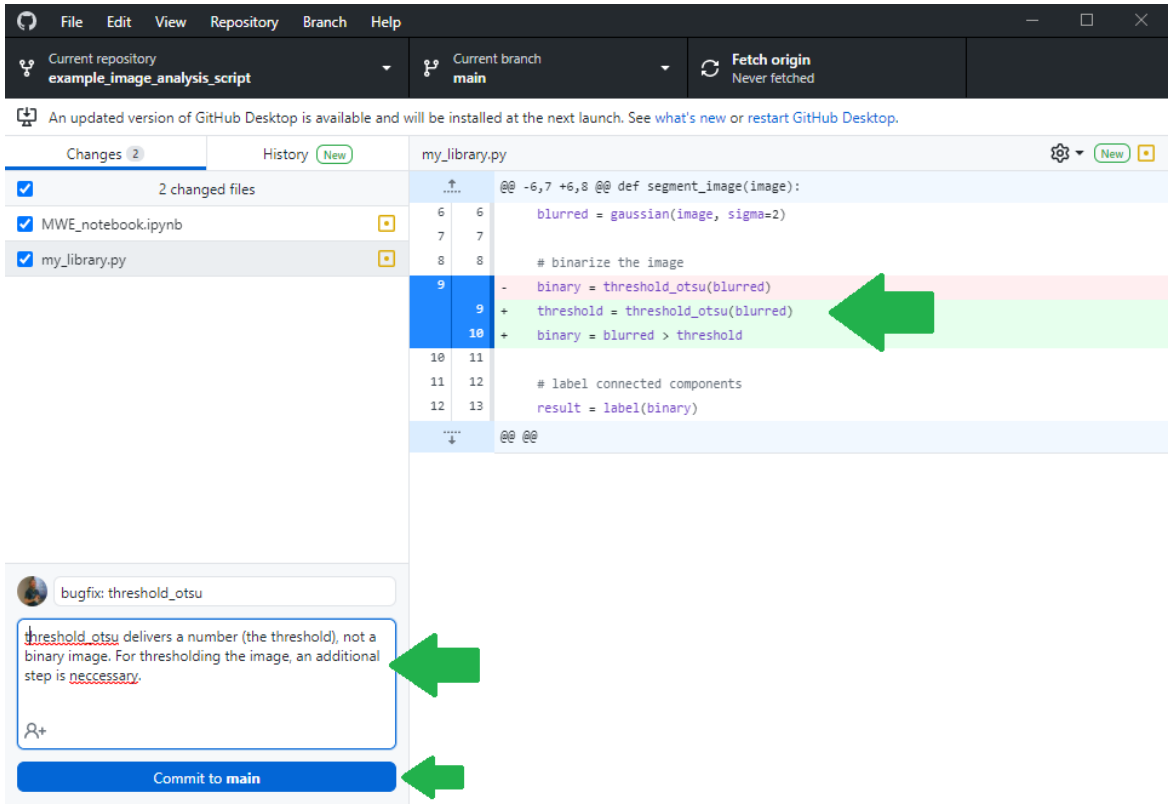
- Making a copy where we have edit rights

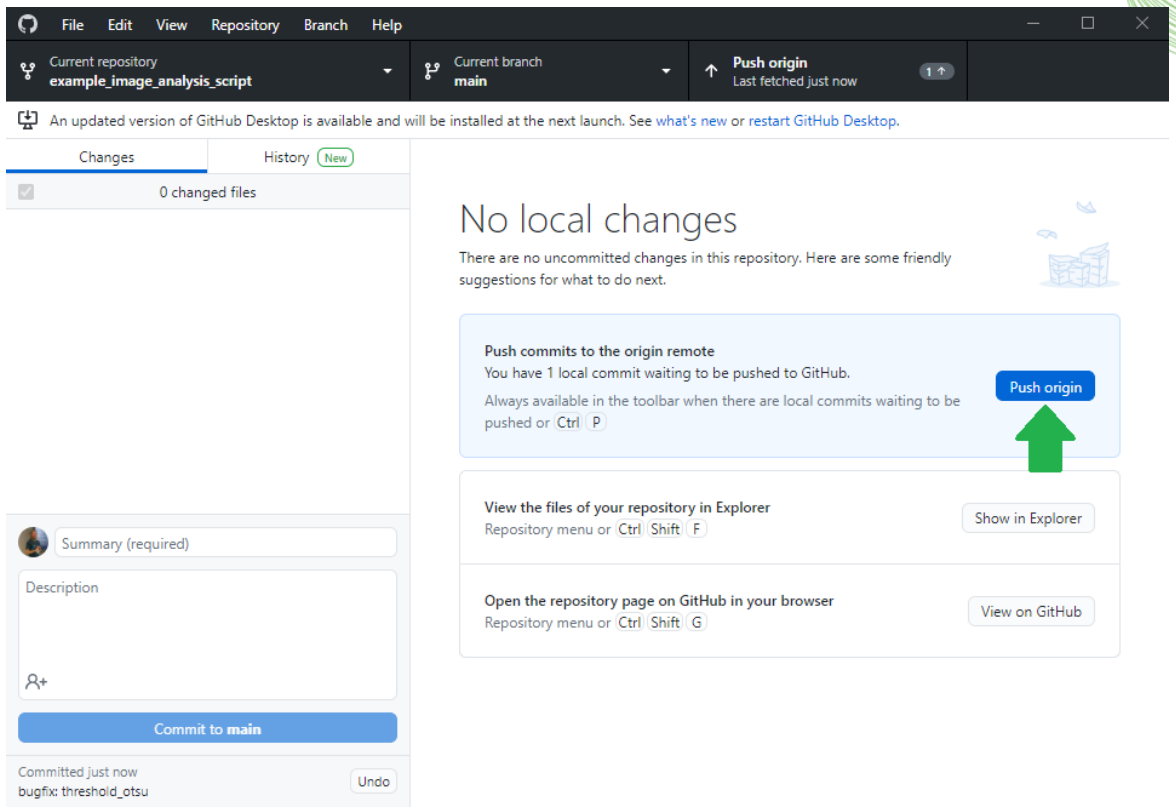


ScaDS.AI
DRESDEN LEIPZIG

github – uploading (again)

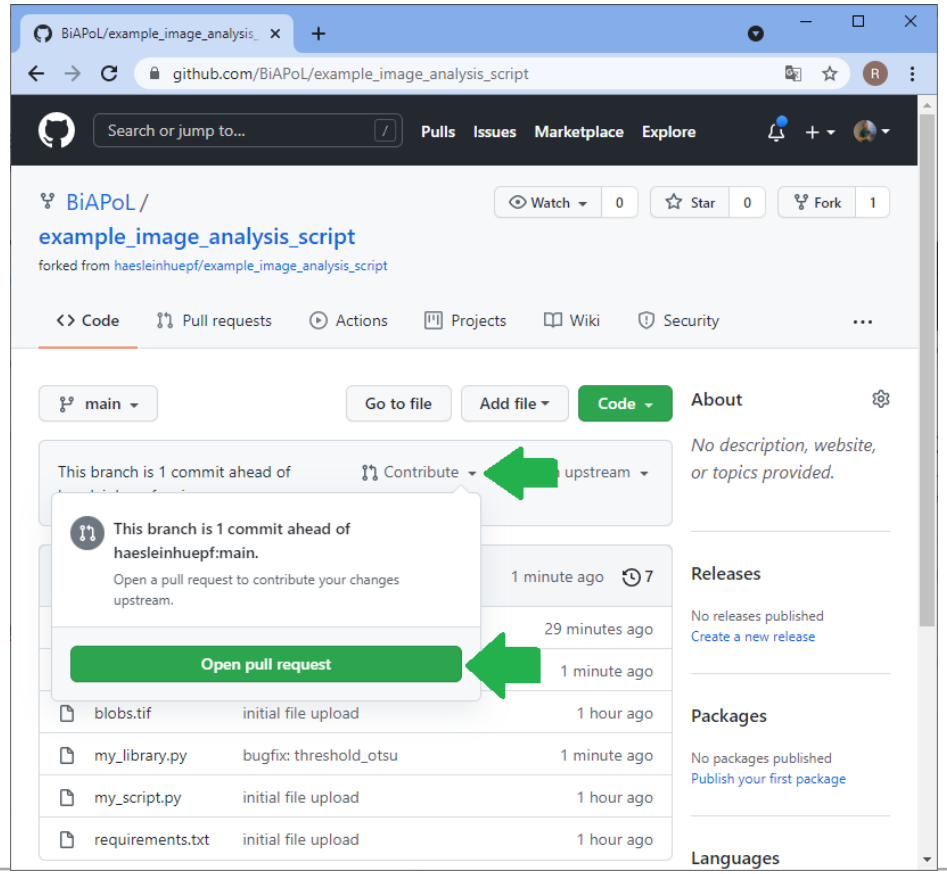
- After fixing a bug, we upload the changes to our fork

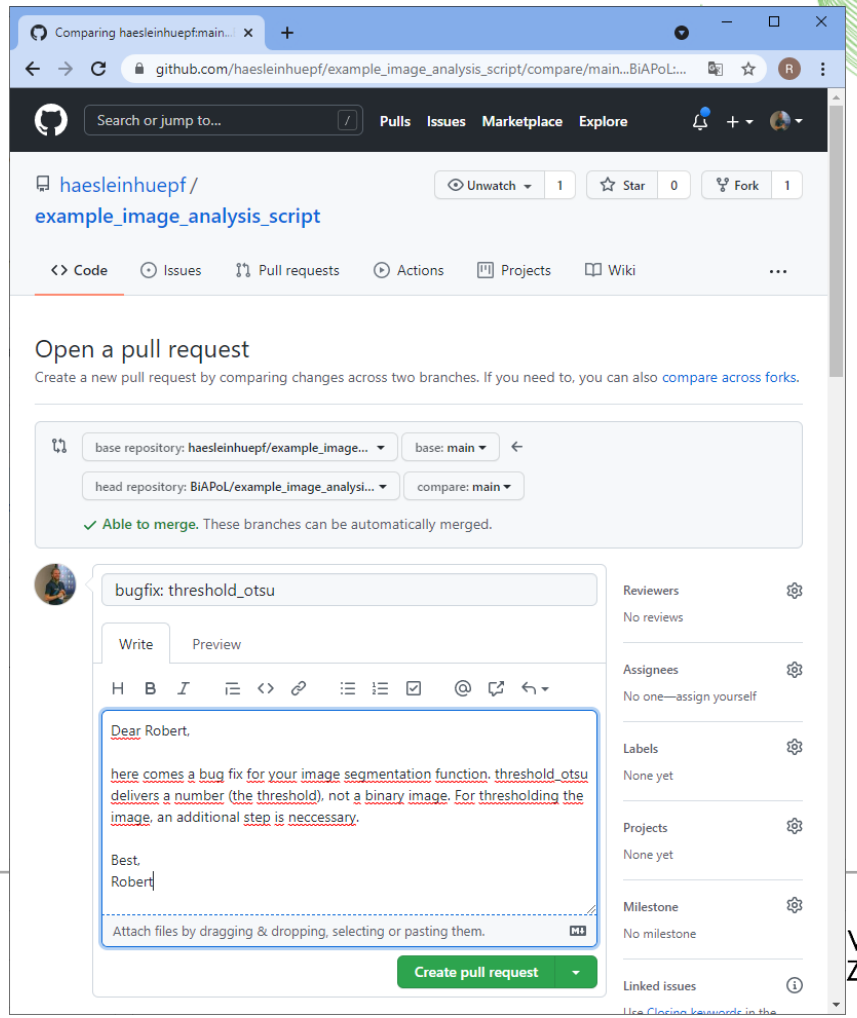




Github – pull requests

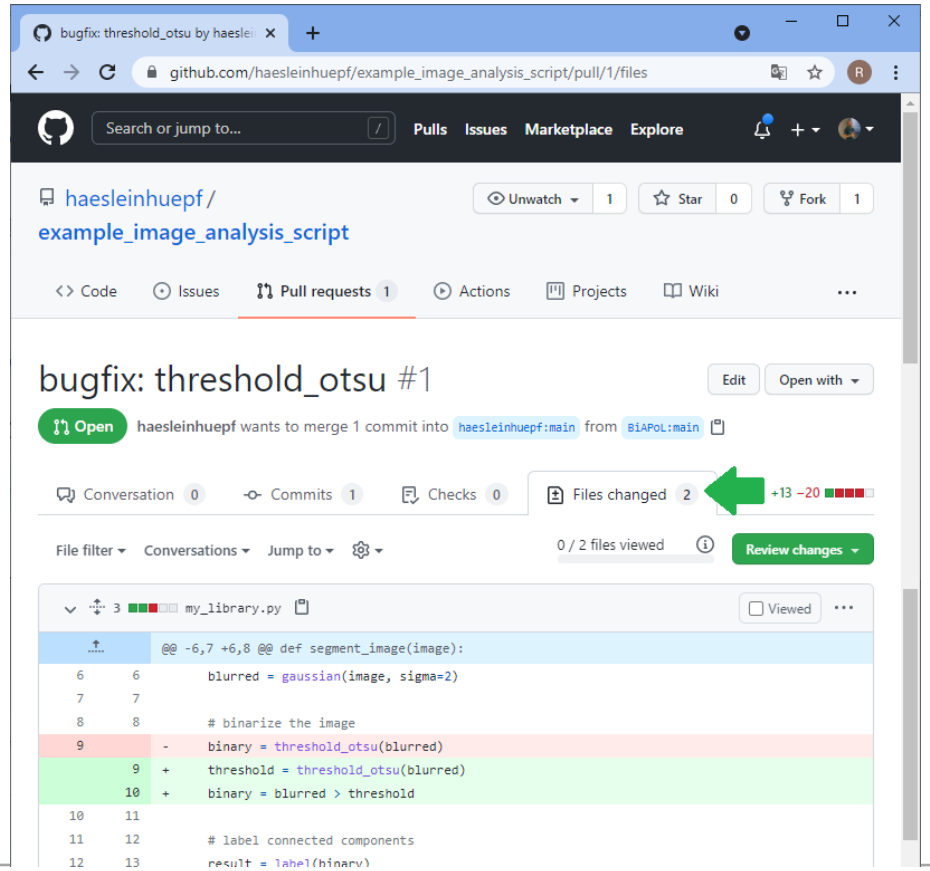
- Contribute to open-source projects

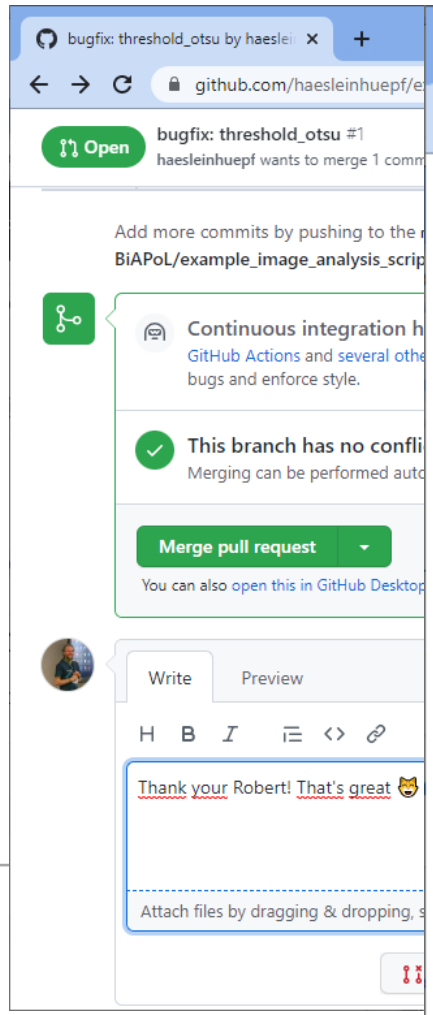


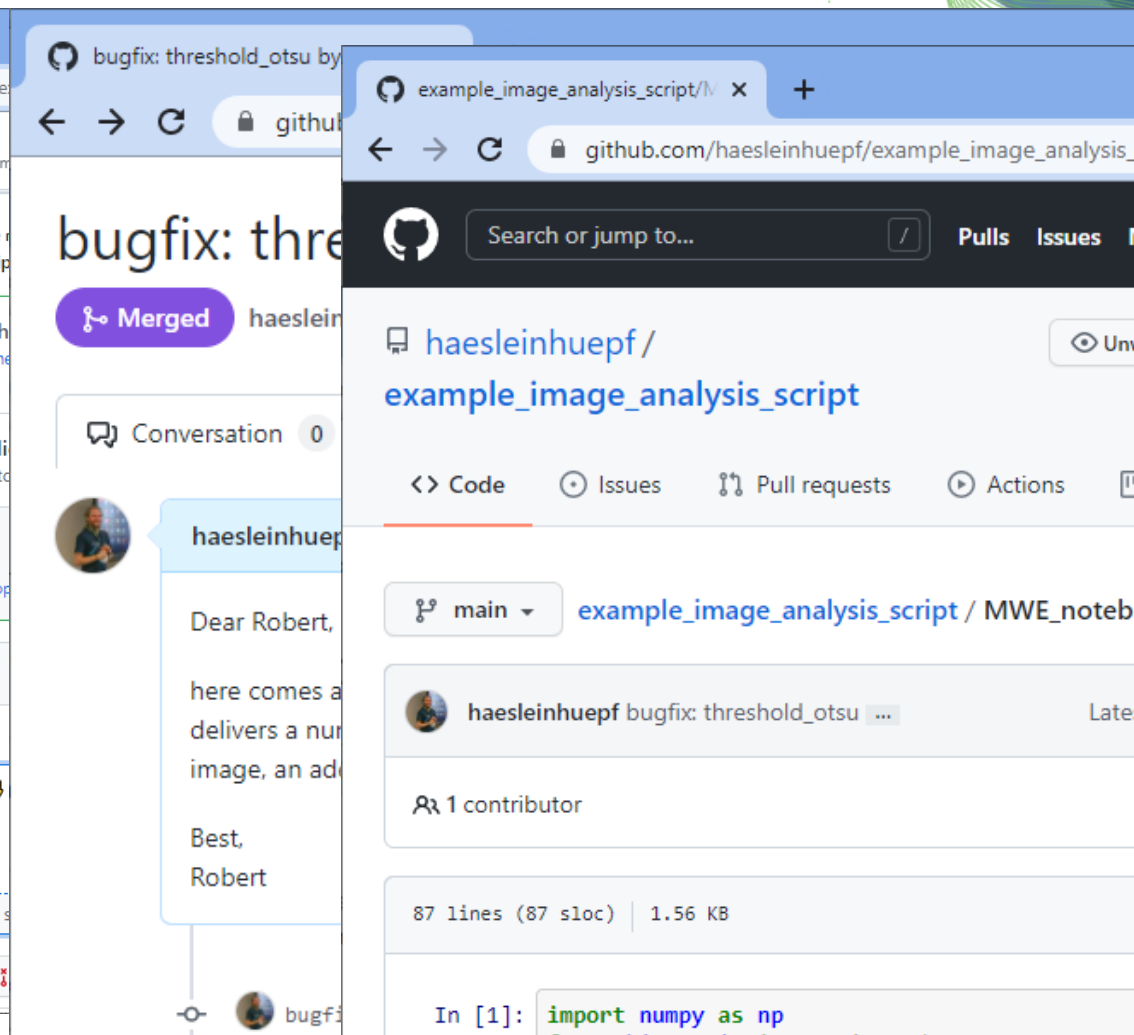


Github – pull requests

- Reviewer perspective



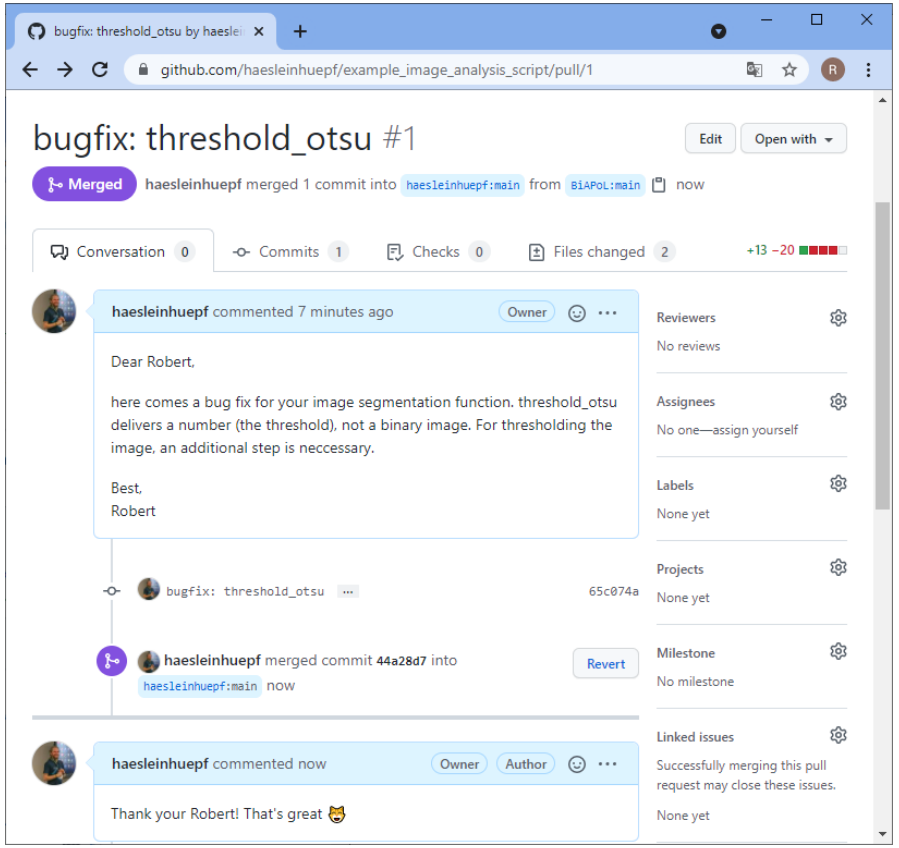


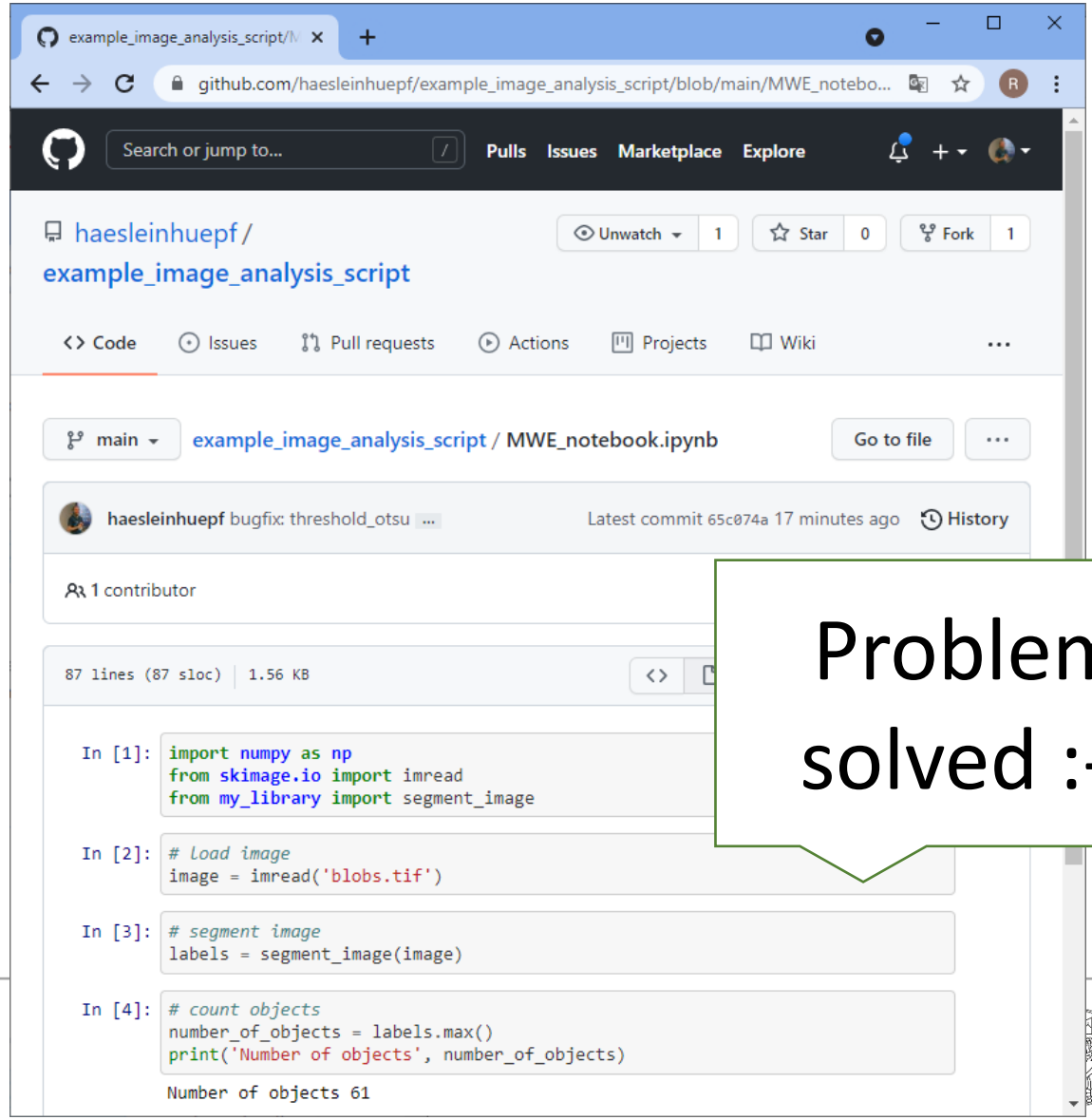


ScaDS.AI DRESDEN LEIPZIG
Robert Haase
@haesleinhuepf
BIDS Lecture 2/14
April 9th 2024

Github – pull requests

- Reviewer perspective

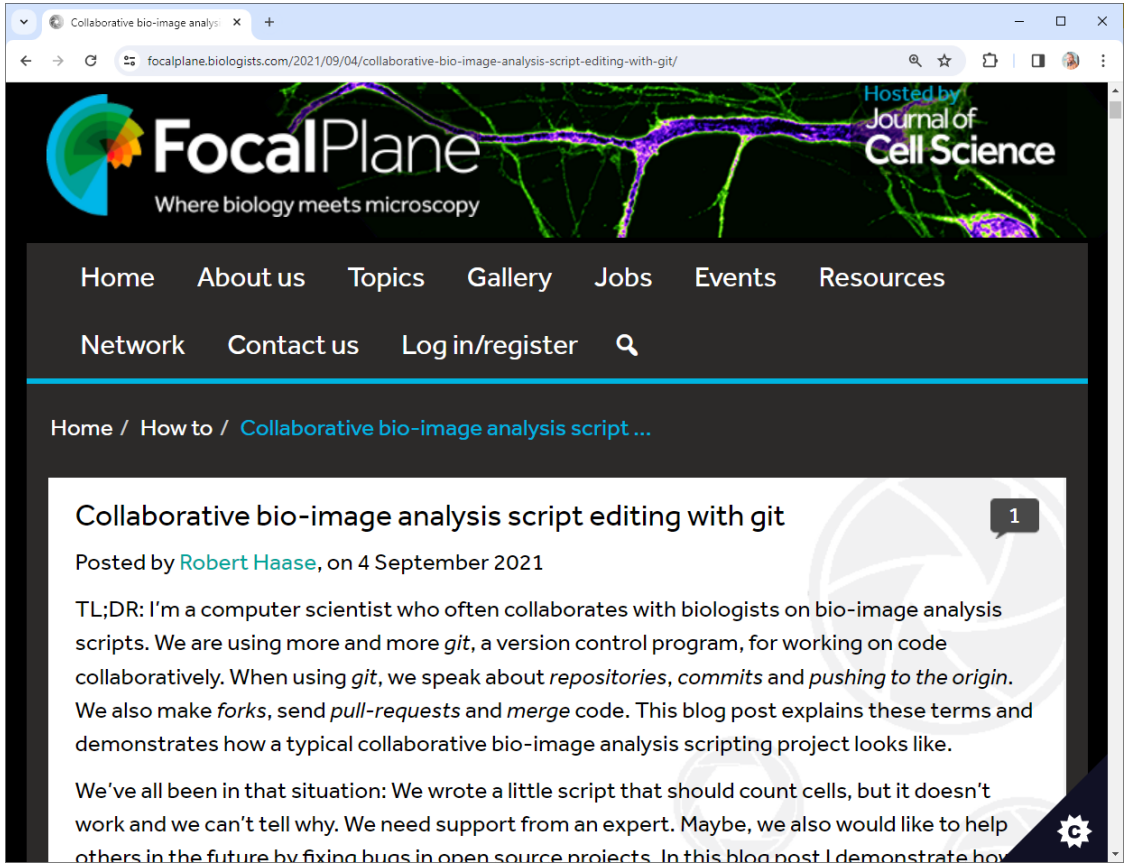


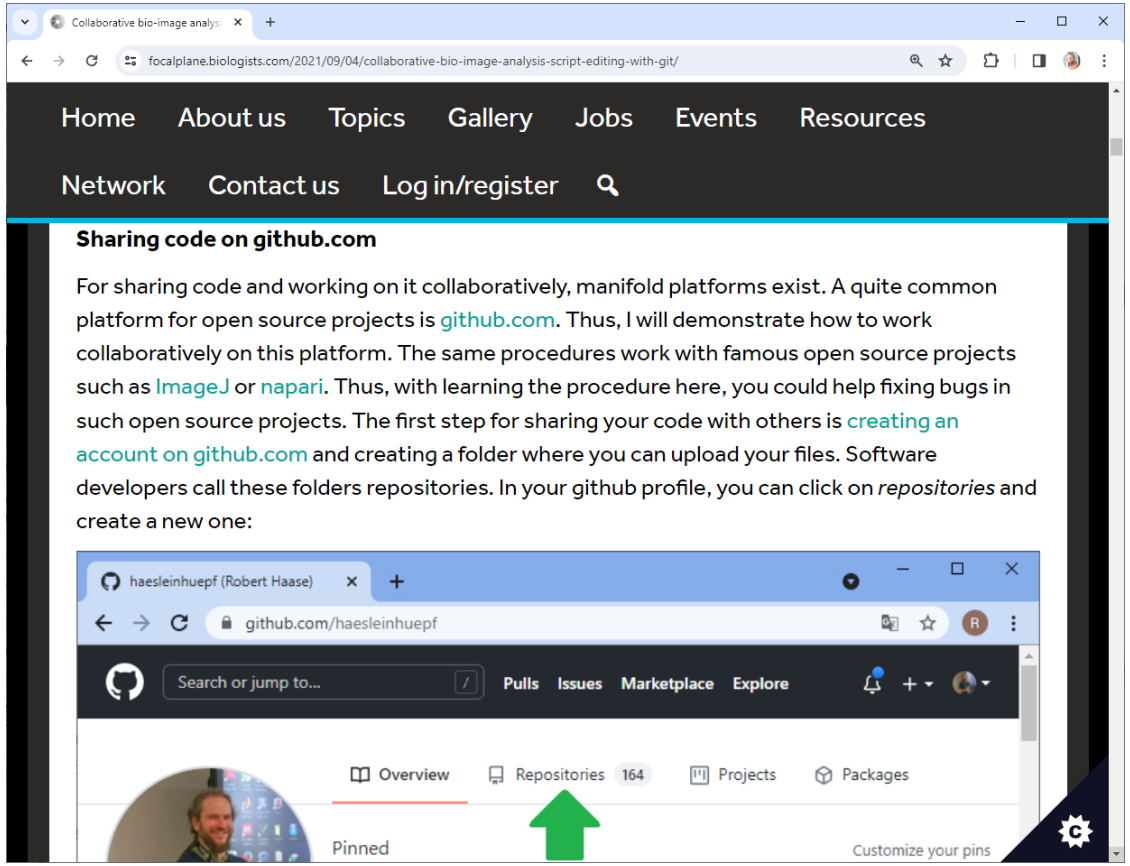


Problem solved :-)

Github

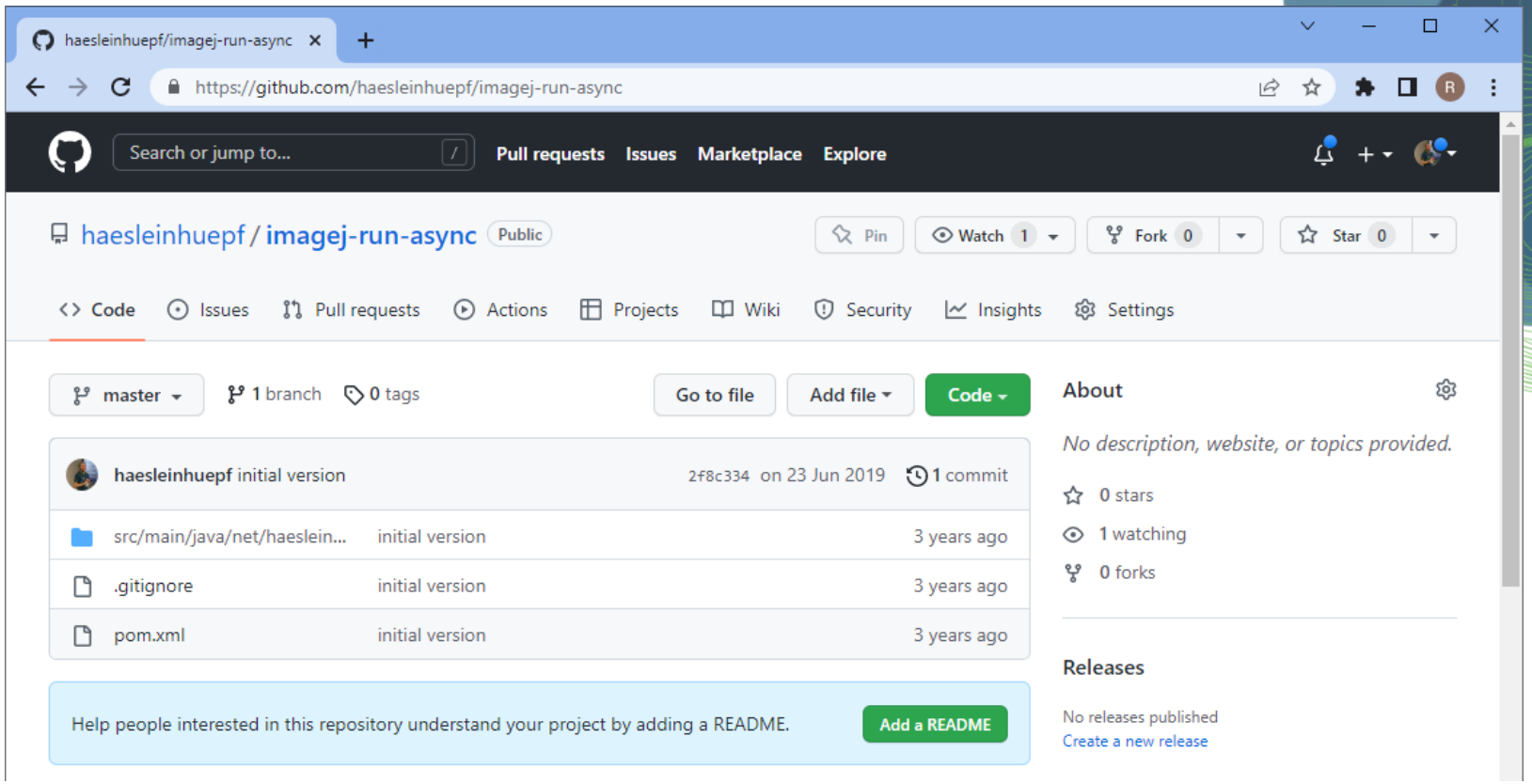
- If this was too fast…





Quiz

It’s ok to reuse this code if …



Mention author

Ask the authors

Link to the license

Copy the copyright statement

ScaDS.AI
DRESDEN LEIPZIG
CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE

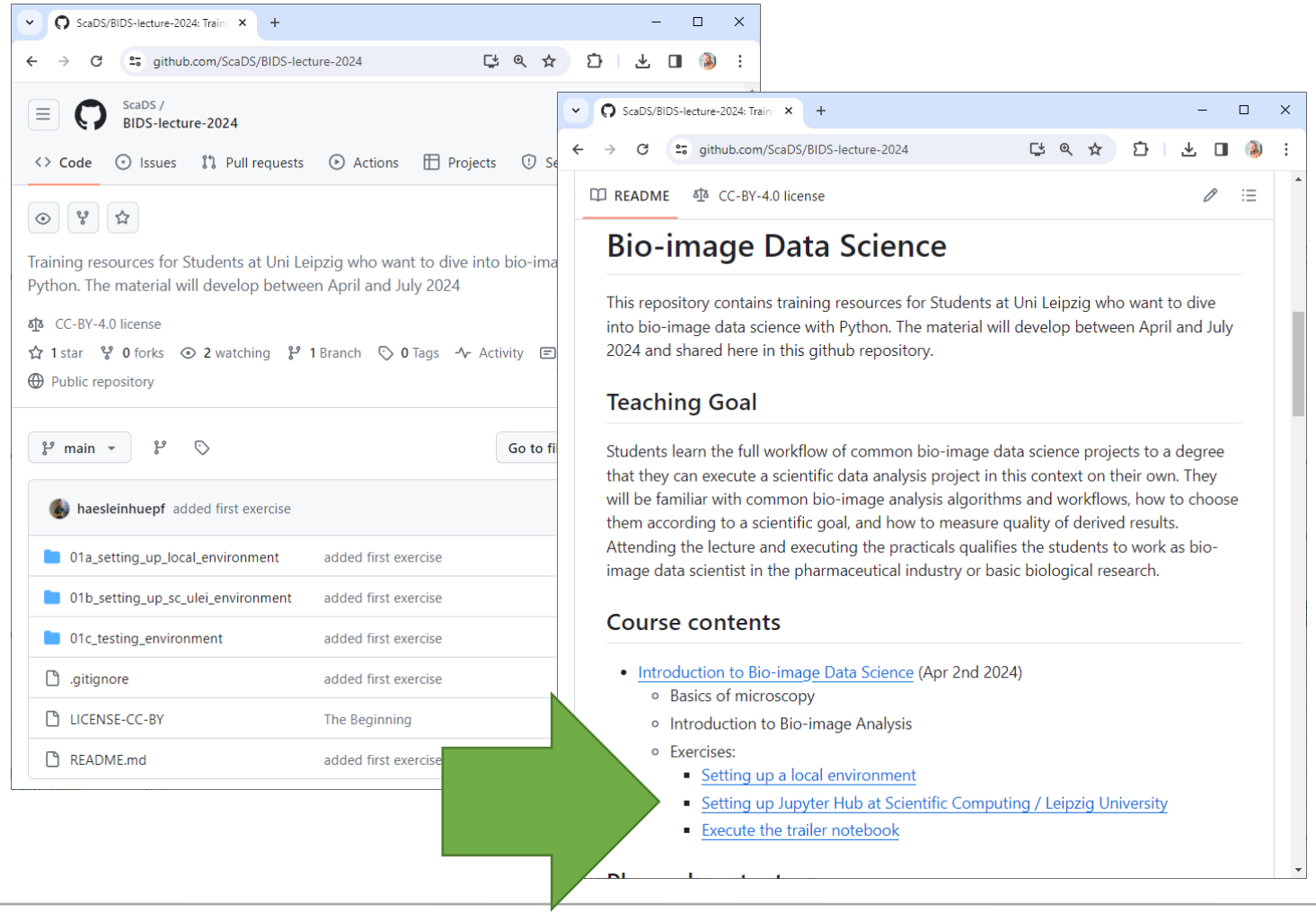
Exercises

Robert Haase

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

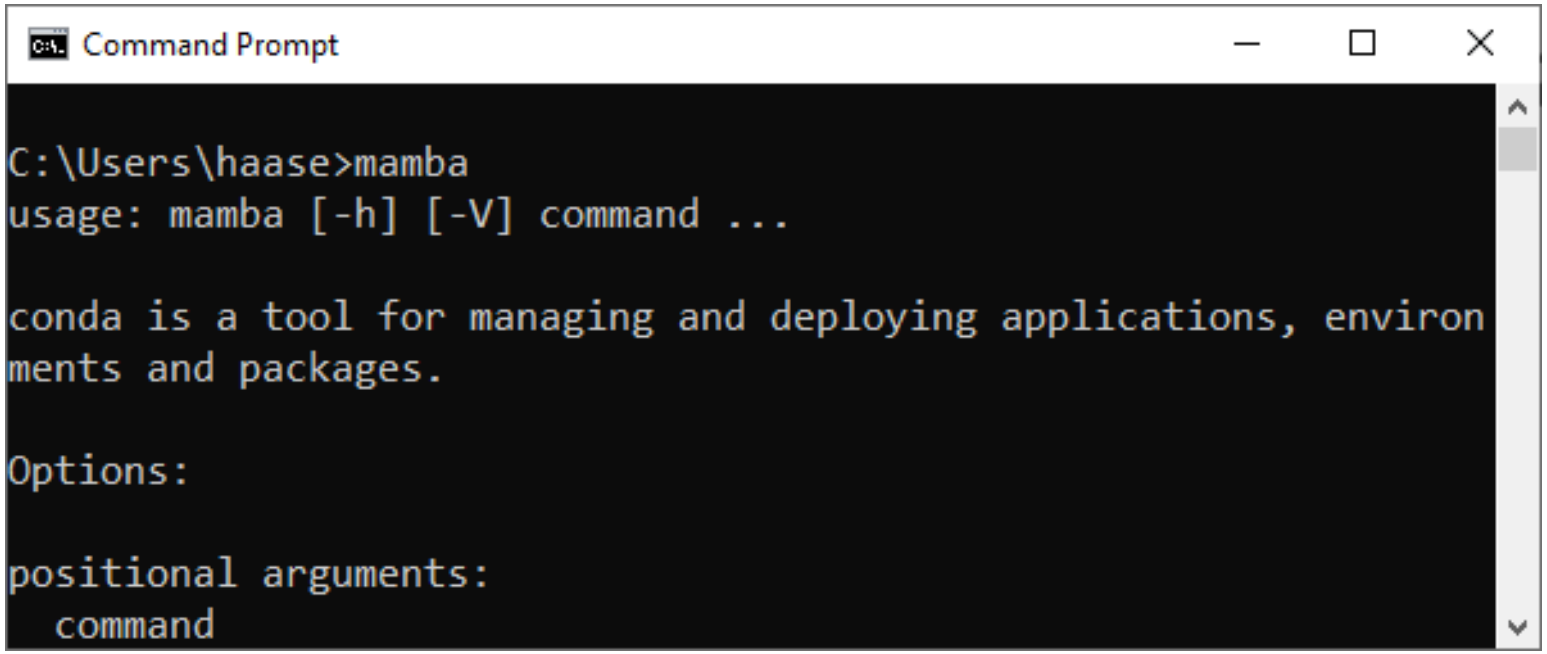
SACHSEN
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages.
Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Exercises


ScaDS/BIDS-lecture-2024: Train
github.com/ScaDS/BIDS-lecture-2024
ScaDS /
BIDS-lecture-2024
Code
Issues
Pull requests
Actions
Projects
Training resources for Students at Uni Leipzig who want to dive into bio-ima
Python. The material will develop between April and July 2024
CC-BY-4.0 license
1 star
0 forks
2 watching
1 Branch
0 Tags
Activity
Public repository
main
Go to fi
haesleinhuepf added first exercise
01a_setting_up_local_environment added first exercise
01b_setting_up_sc_ulei_environment added first exercise
01c_testing_environment added first exercise
.gitignore added first exercise
LICENSE-CC-BY The Beginning
README.md added first exercise
README
CC-BY-4.0 license
Bio-image Data Science
This repository contains training resources for Students at Uni Leipzig who want to dive into bio-image data science with Python. The material will develop between April and July 2024 and shared here in this github repository.
Teaching Goal
Students learn the full workflow of common bio-image data science projects to a degree that they can execute a scientific data analysis project in this context on their own. They will be familiar with common bio-image analysis algorithms and workflows, how to choose them according to a scientific goal, and how to measure quality of derived results. Attending the lecture and executing the practicals qualifies the students to work as bio-image data scientist in the pharmaceutical industry or basic biological research.
Course contents
Introduction to Bio-image Data Science (Apr 2nd 2024)
Basics of microscopy
Introduction to Bio-image Analysis
Exercises:
Setting up a local environment
Setting up Jupyter Hub at Scientific Computing / Leipzig University
Execute the trailer notebook

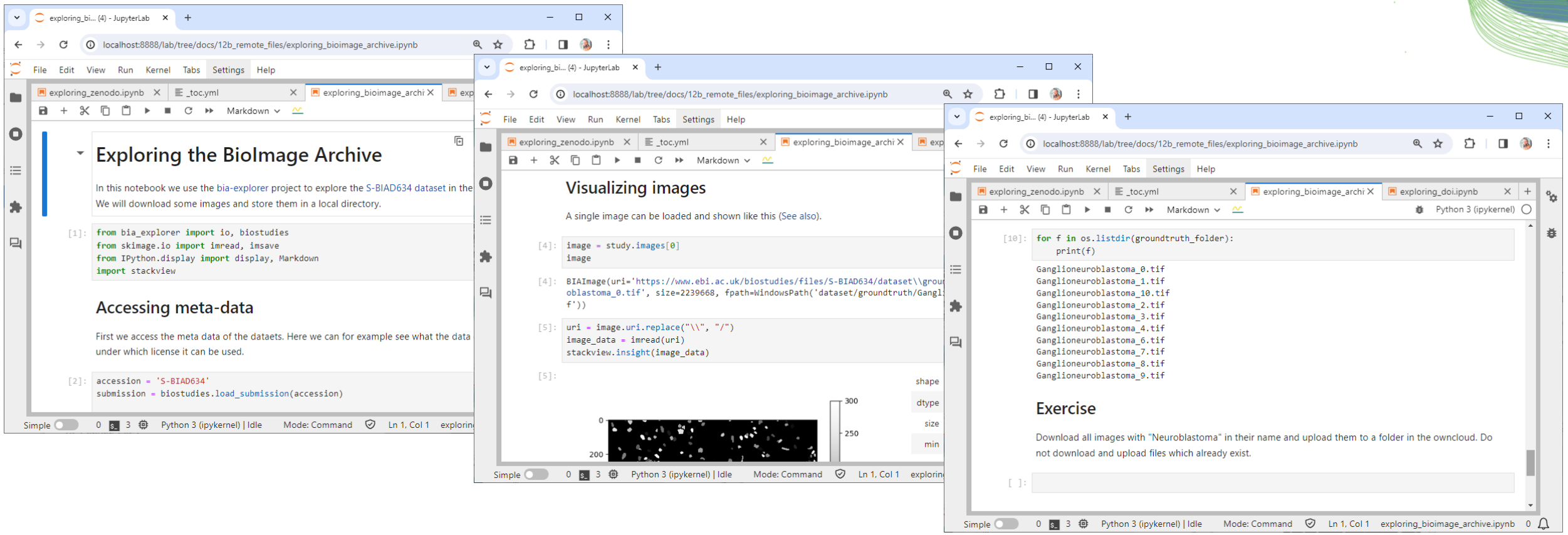
Exercise (recap)

- Make sure `mamba` is installed on your computer (see instructions from last week)



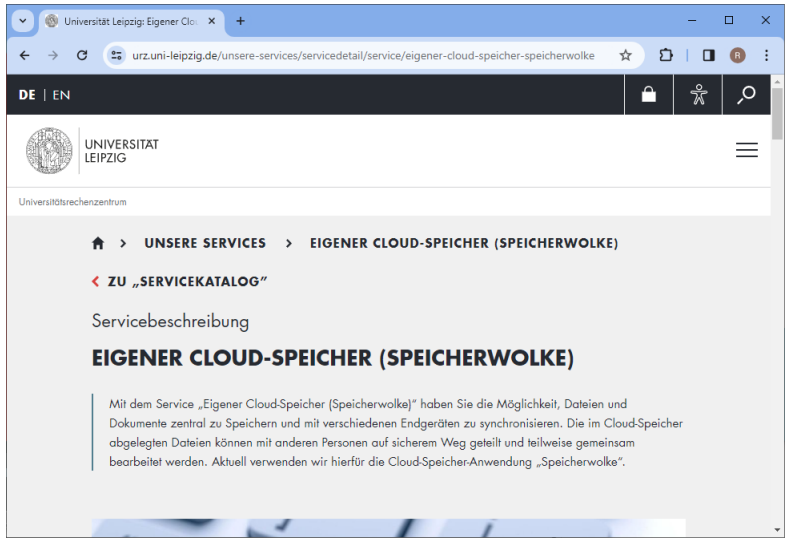
Exercise (BioImage Archive)

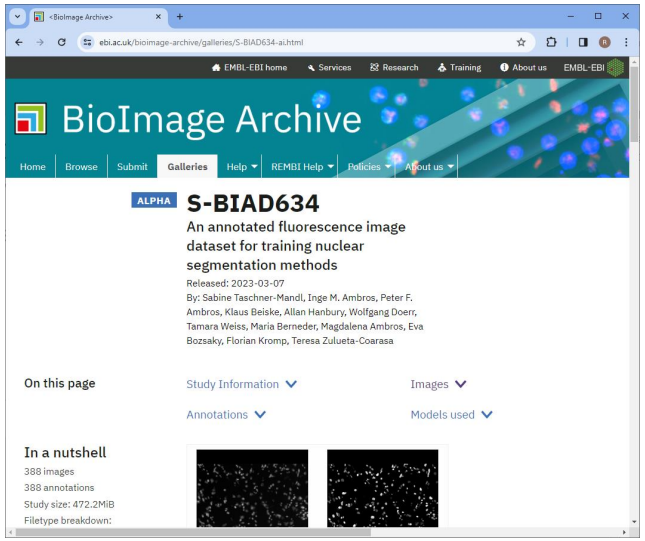
- Download a dataset from the BioImage Archive

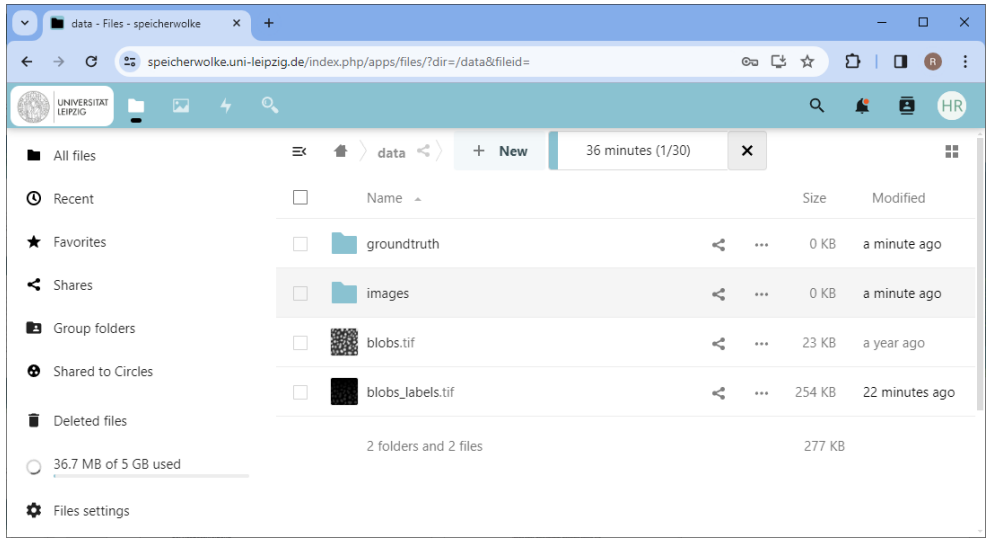


Exercise (nextcloud)

- Register at Speicherwolke @ Uni Leipzig,
- Upload the images from the BioImage Archive to a folder in the Speicherwolke.

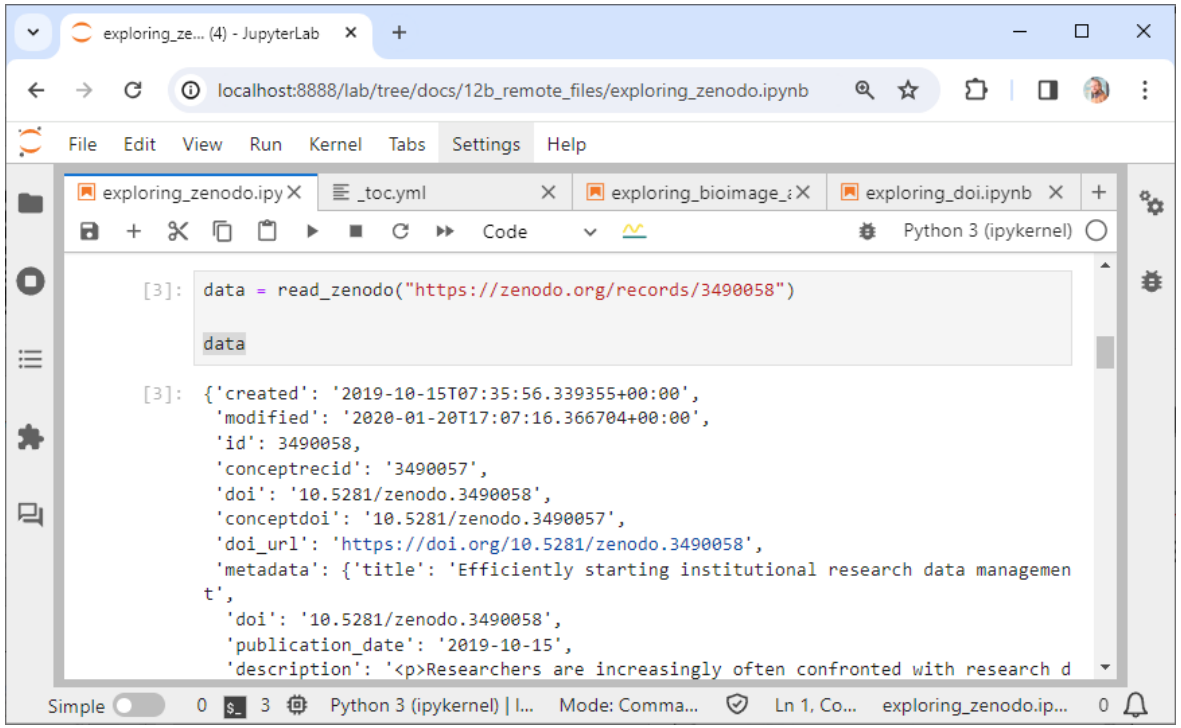


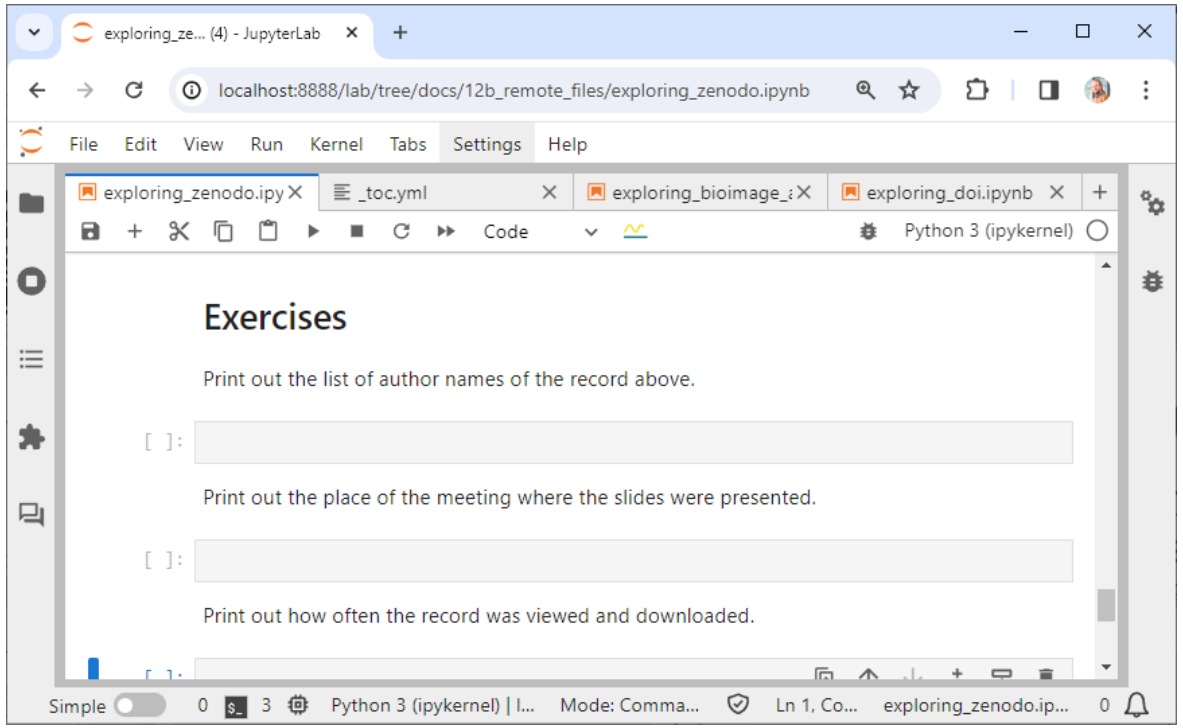




Exercise (Zendo and DOI)

- Explore the DOI and Zenodo APIs to find out the author of online records





Exercise

- Clone the training materials repository
- Fix the typo on this page, send a pull-request