

# Research Software Management using git and zenodo











Robert Haase



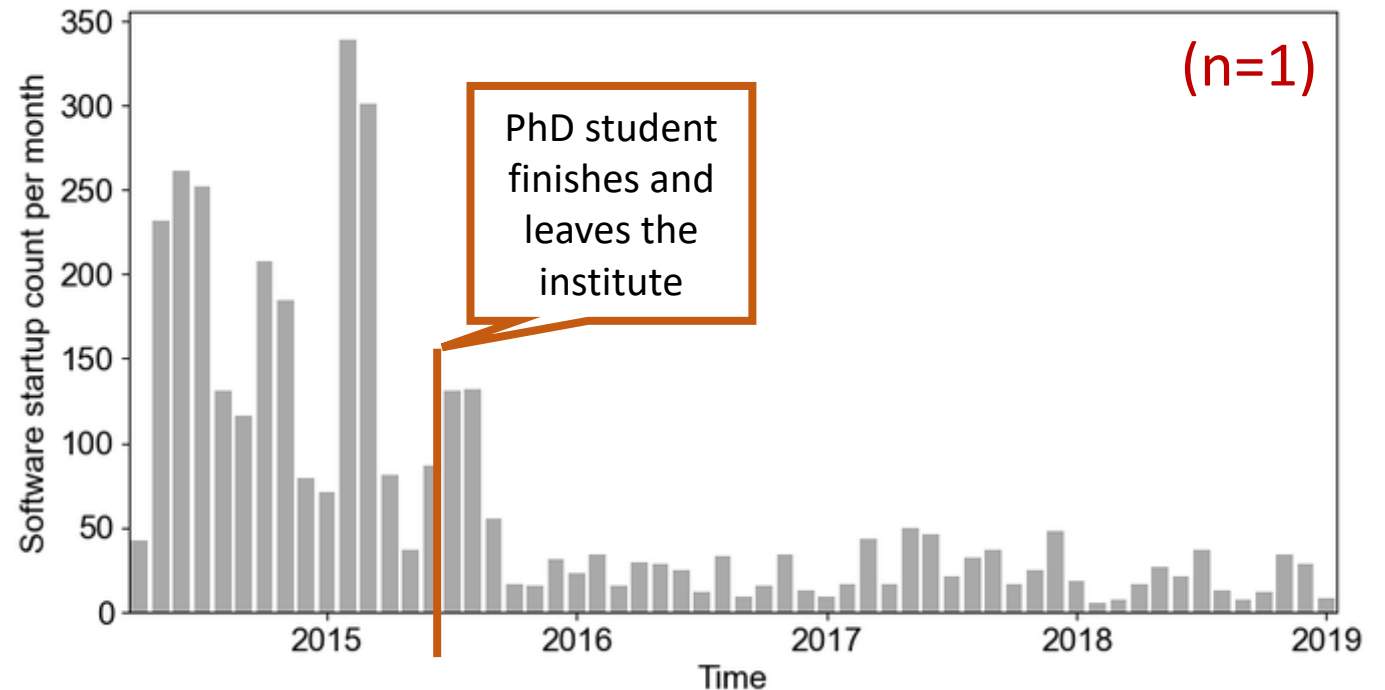
<https://doi.org/10.5281/zenodo.11259495>

# Managing Research Software

- Me, during my Phd:

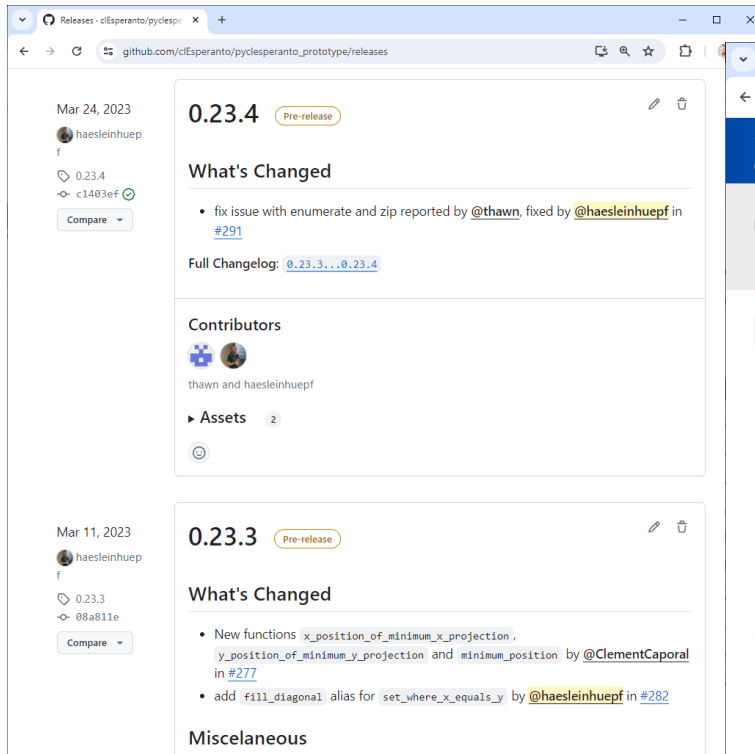
 geisterr - Kopie (2).exe	2/27/2013 4:43 PM
 geisterr - Kopie.exe	3/26/2012 9:53 PM
 geisterr.exe	2/18/2015 3:09 PM
 geisterr_alpha.exe	2/20/2014 8:44 AM
 geisterr_alpha2.exe	1/7/2014 1:41 PM
 geisterr_alpha3.exe	1/8/2014 11:26 AM
 geisterr_alpha4.exe	1/15/2014 2:09 PM
 geisterr2013.exe	12/17/2013 9:55 AM
 geisterrbilder.exe	3/22/2013 4:38 PM
 histogeisterr.exe	4/29/2013 2:19 PM

„Geisterr“ usage statistics



# Managing Research Software

## • Me, today:



Releases · clEsperanto/pyclesperanto-prototype

Mar 24, 2023  
haesleinhuepf

0.23.4 (Pre-release)  
c1403ef

Compare

### What's Changed

- fix issue with enumerate and zip reported by @thawn, fixed by @haesleinhuepf in #291

Full Changelog: 0.23.3...0.23.4

### Contributors

thawn and haesleinhuepf

Assets 2

Mar 11, 2023  
haesleinhuepf

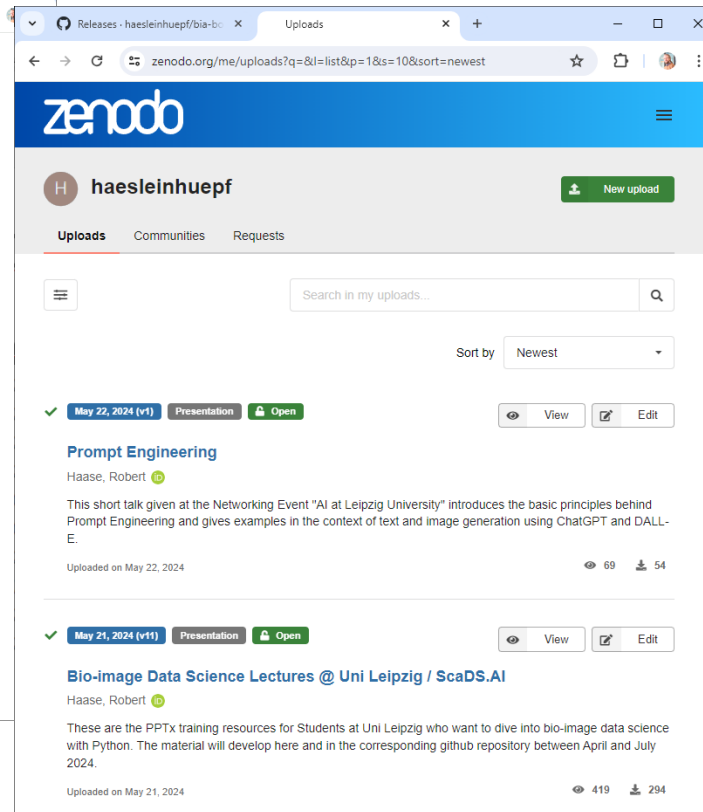
0.23.3  
08a811e

Compare

### What's Changed

- New functions `x_position_of_minimum_x_projection`, `y_position_of_minimum_y_projection` and `minimum_position` by @ClementCaporal in #277
- add `fill_diagonal` alias for `set_where_x_equals_y` by @haesleinhuepf in #282

### Miscellaneous



zenodo

haesleinhuepf

Uploads Communities Requests

Search in my uploads...

Sort by Newest

May 22, 2024 (v1) Presentation Open

View Edit

### Prompt Engineering

Haase, Robert

This short talk given at the Networking Event "AI at Leipzig University" introduces the basic principles behind Prompt Engineering and gives examples in the context of text and image generation using ChatGPT and DALL-E.

Uploaded on May 22, 2024

69 54

May 21, 2024 (v11) Presentation Open

View Edit

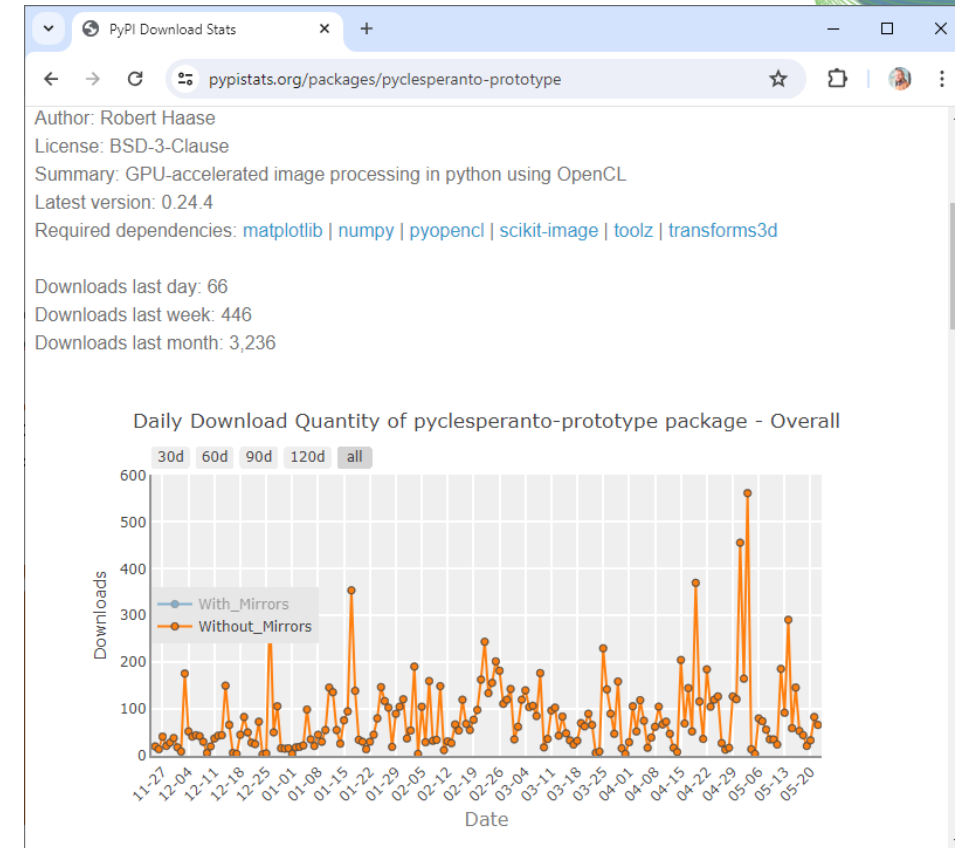
### Bio-image Data Science Lectures @ Uni Leipzig / ScaDS.AI

Haase, Robert

These are the PPTx training resources for Students at Uni Leipzig who want to dive into bio-image data science with Python. The material will develop here and in the corresponding github repository between April and July 2024.

Uploaded on May 21, 2024

419 294



PyPI Download Stats

pypistats.org/packages/pyclesperanto-prototype

Author: Robert Haase  
License: BSD-3-Clause  
Summary: GPU-accelerated image processing in python using OpenCL  
Latest version: 0.24.4  
Required dependencies: matplotlib | numpy | pyopencl | scikit-image | toolz | transforms3d

Downloads last day: 66  
Downloads last week: 446  
Downloads last month: 3,236

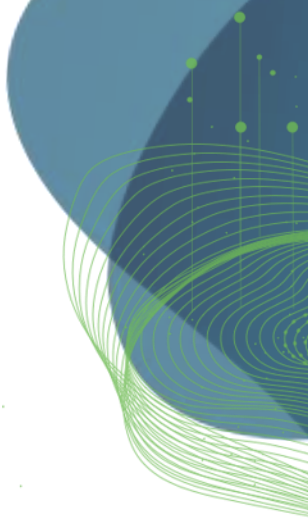
### Daily Download Quantity of pyclesperanto-prototype package - Overall

30d 60d 90d 120d all

Downloads

With Mirrors Without Mirrors

Date

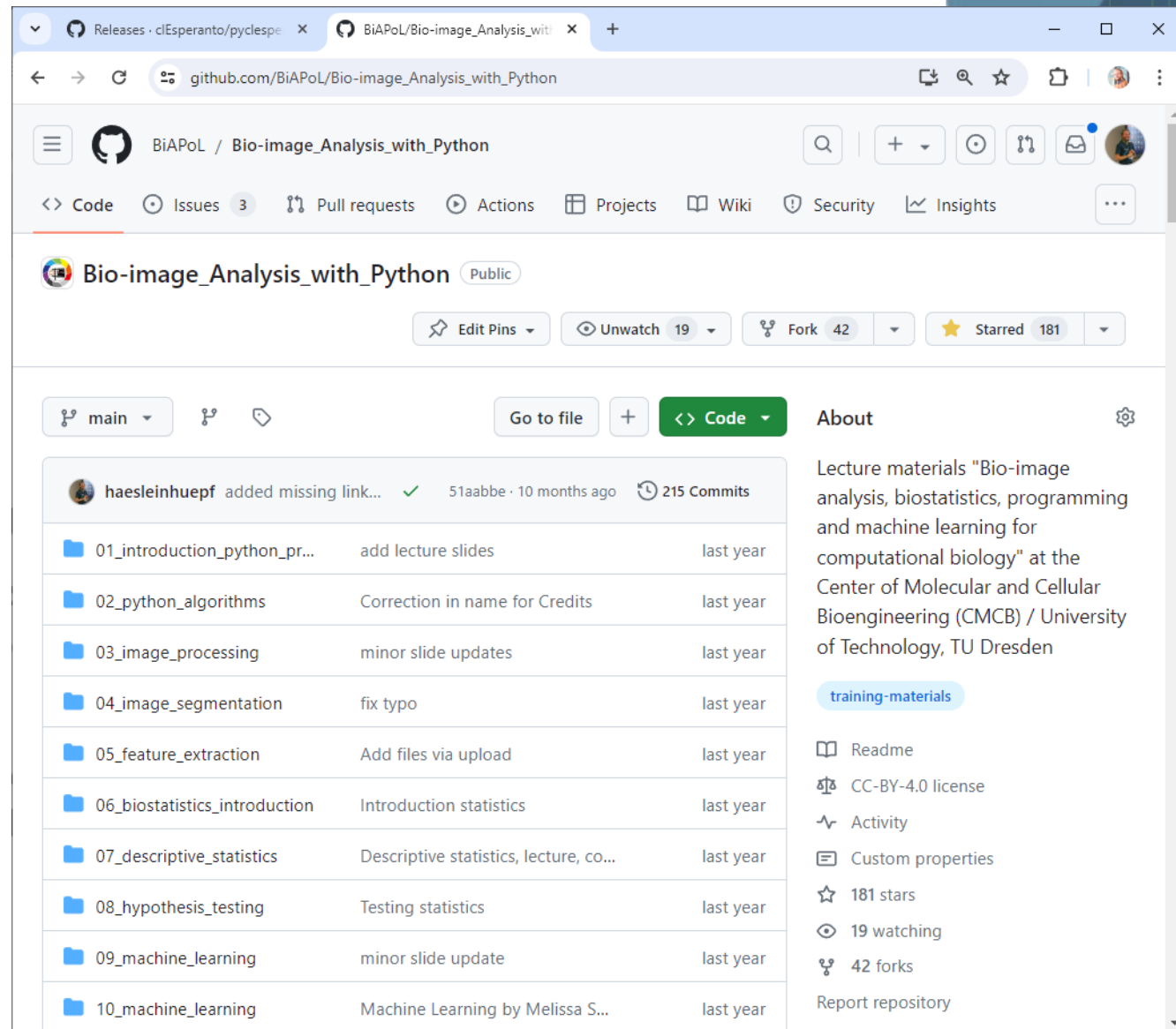


# Github versus Zenodo

- Github
  - Good for collaboration
  - Small changes committed separately
  - Full history of code available
- Zenodo
  - Publicly funded infrastructure
  - Good for Archiving
  - Releases are archived, not every commit



- 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 for tracking contributions

- Who did what? Who deserves co-authorship?

The image displays three screenshots from a GitHub repository named 'BiAPoL / Bio-image\_Analysis\_with\_Python'.

The leftmost screenshot shows the 'Commits' page, listing recent commits. A blue arrow points from the 'deep learning exercises' commit to the middle screenshot.

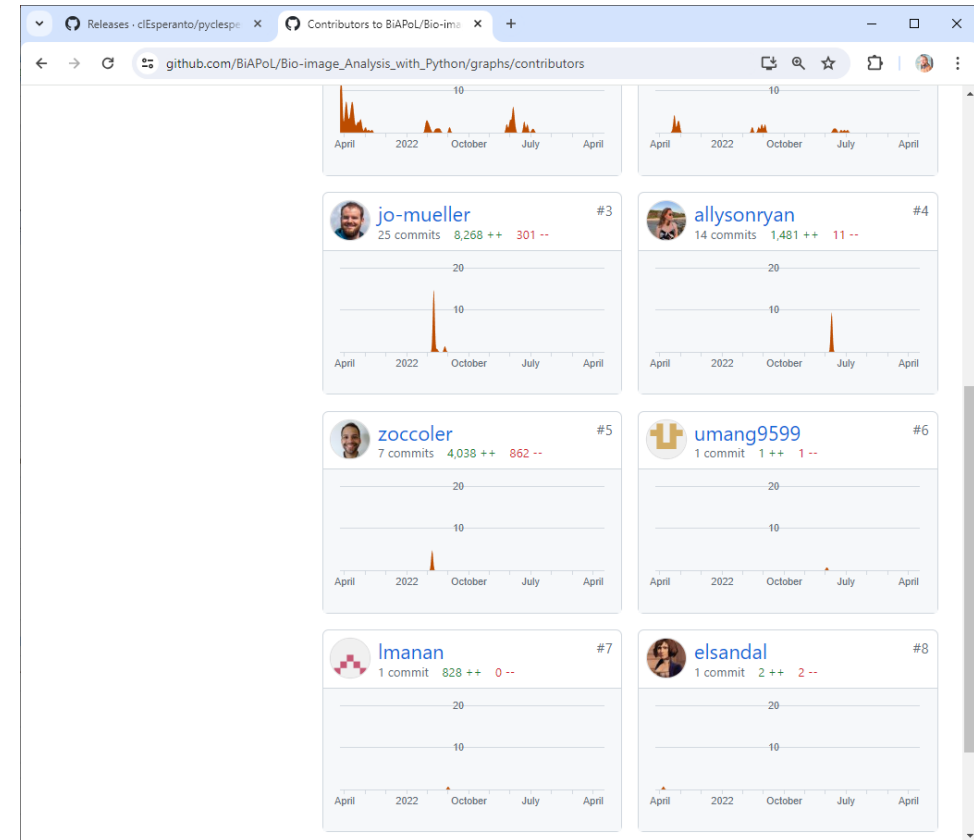
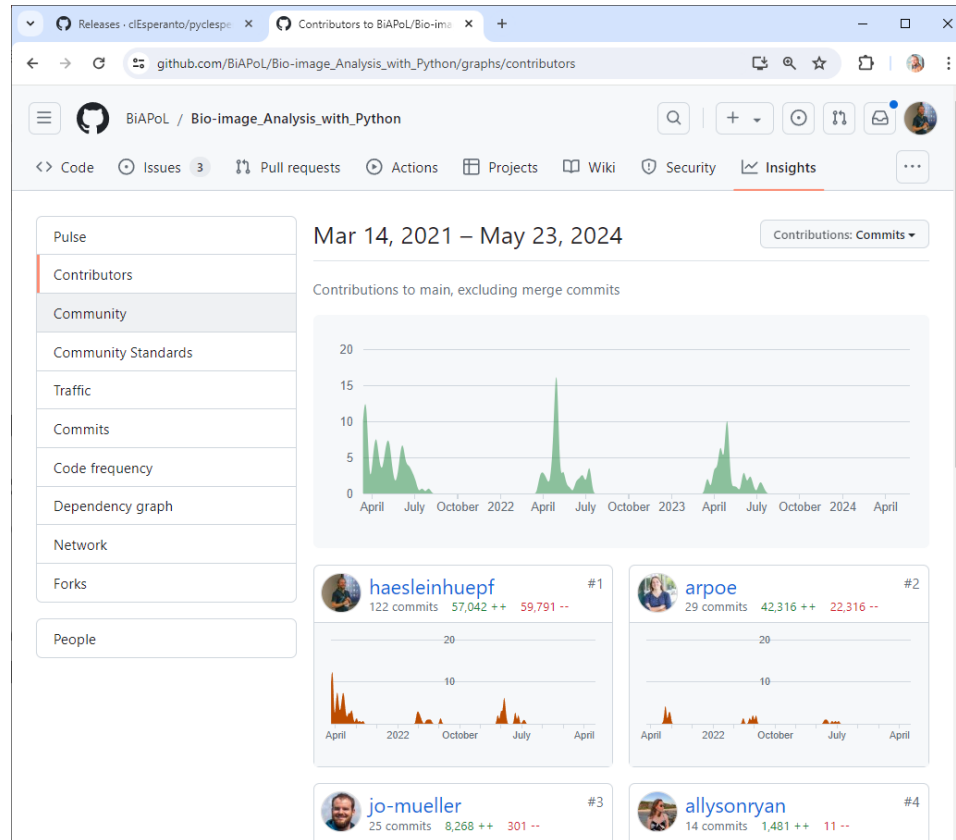
The middle screenshot shows the 'Commit' page for commit 59cab22. It displays the commit message 'deep learning exercises' and a list of changed files. A blue arrow points from the 'cellpose.ipynb' file to the rightmost screenshot.

The rightmost screenshot shows the 'Preview' view of the 'cellpose.ipynb' file. The notebook content is titled 'Image Segmentation with CellPose' and includes a code cell with the command `!pip install cellpose stackview ipycanvas==0.11`.



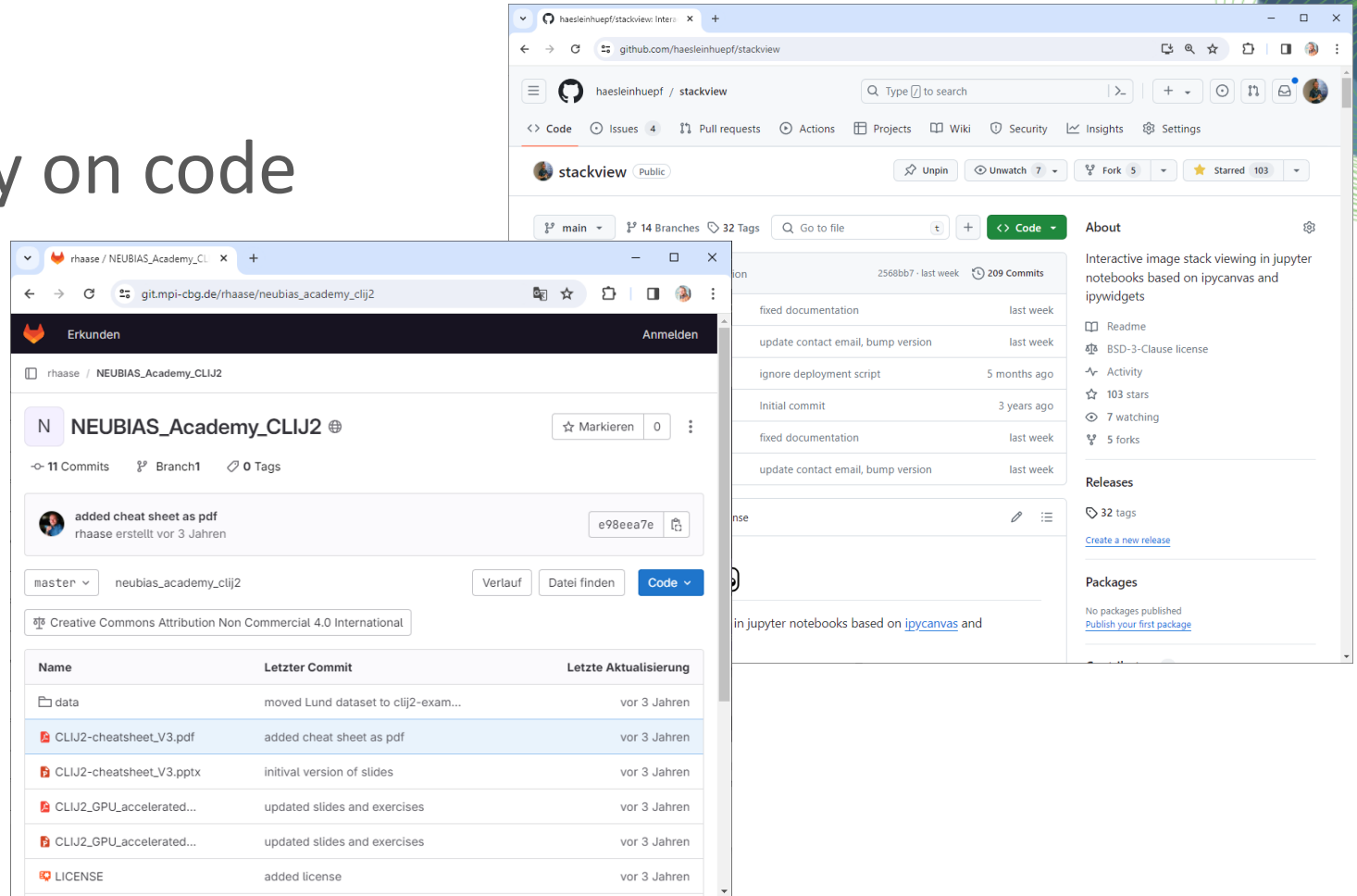
# git for tracking contributions

- Who did what? Who deserves co-authorship?

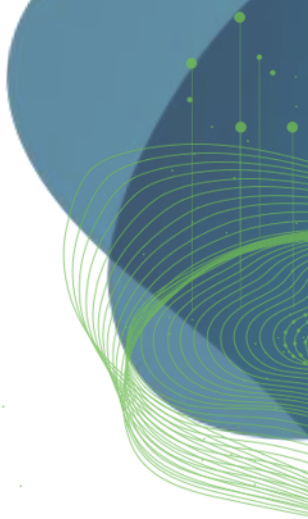


# Where to share? Gitlab / github

- Version control
- Working collaboratively on code
- Gitlab: commonly available at research institutions
- Github: Owned by Microsoft, ideal for collaborating with externals





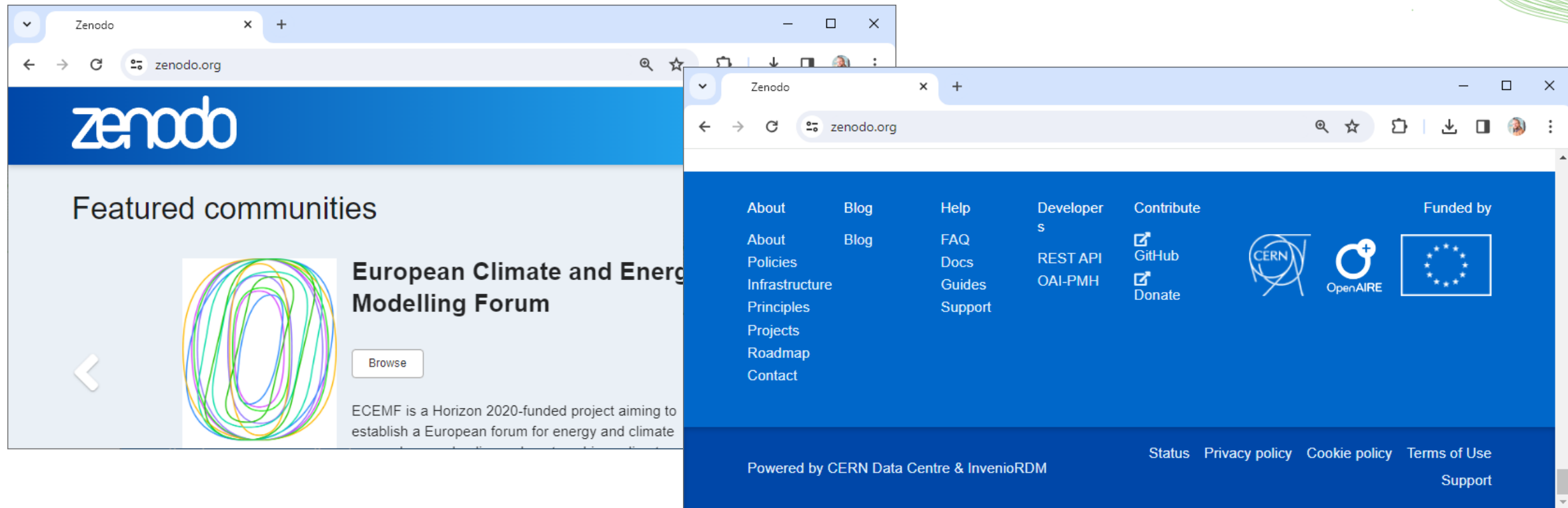


# Github versus Zenodo

- Github
  - Good for collaboration
  - Small changes committed separately
  - Full history of code available
- Zenodo
  - Publicly funded infrastructure
  - Good for Archiving
  - Releases are archived, not every commit

# Zenodo

- Great for archiving code, slides, data
- Publicly funded infrastructure @ CERN / Switzerland



# Zenodo

- Sharing and citing made easy

The image displays three sequential screenshots of the Zenodo website interface:

- Left Screenshot:** The 'New upload' page. It features a search bar, 'Upload' and 'Community' buttons, and a 'Delete' button. A large text area prompts users to 'Drag and drop files here' or 'Choose files'. A note at the bottom states '(minimum 1 file required, max 50 GB per dataset) - contact us for larger files'. The 'max 50 GB per dataset' text is highlighted with a red box.
- Middle Screenshot:** The 'Deposit' page for a dataset titled 'Strausberg\_Tribolium\_LA-GFP\_tailpole\_run (Excerpt)'. It shows the author 'Daniela Vorkel' and affiliations 'CSBD / MPI CBG' and 'Robert Haase' and 'Eugene Myers' and 'CSBD / MPI CBG'. A 'Description' field contains text about 'Tribolium castaneum, life-actin GFP (transgenic), imaged with a microscope' and 'This is an excerpt of time points 291-340 showing a single mitotic wave'. A 'Description' field is highlighted with a red box.
- Right Screenshot:** The 'Record' page for the dataset, showing a list of files including 'ClearControl meta header files', 'ClearControl program', 'Background subtracted / maximum projected thumbnails as video', and 'Single time point 3D stack'. A 'Preview' section shows a thumbnail of a mitotic wave. A 'Cite as' section is highlighted with a red box, displaying the citation: 'Daniela Vorkel, Robert Haase, & Eugene Myers. (2020). Strausberg\_Tribolium\_LA-GFP\_tailpole\_run (Excerpt timepoints 291-340) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3981193'.

# Restricted Access

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

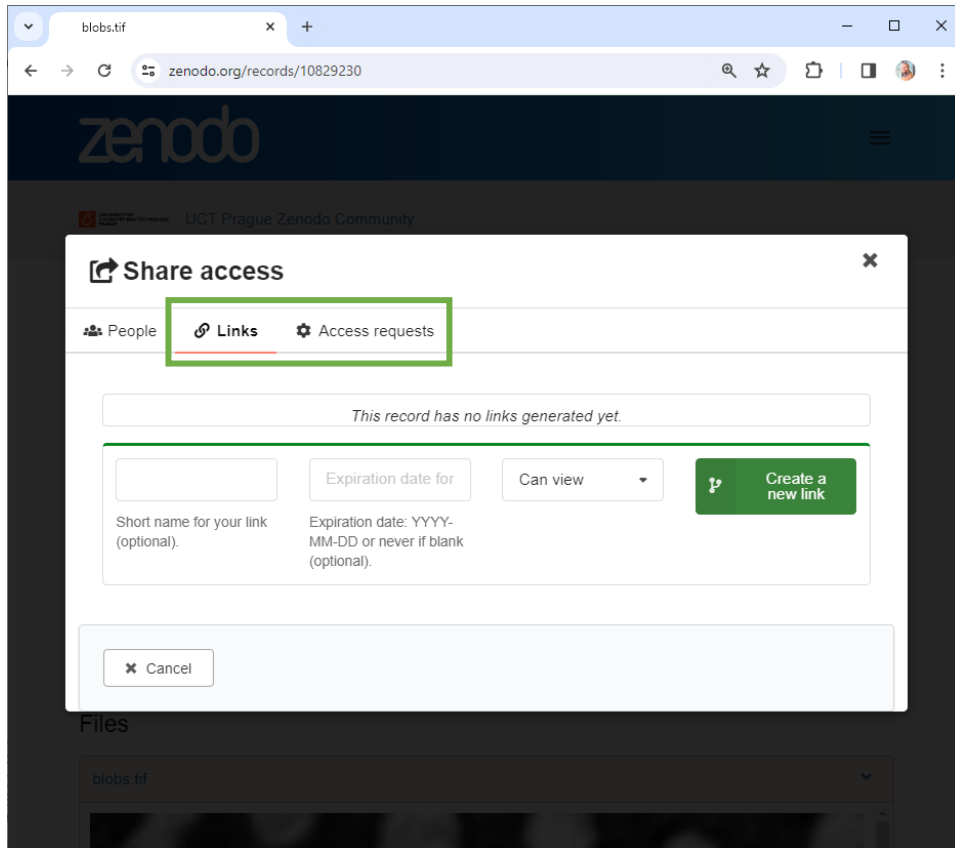
The screenshot shows the Zenodo record page for 'blobs.tif' (DOI: 10.5281/zenodo.10829230). The page indicates it was published on March 18, 2024, as Version v1. A red 'Restricted' label is visible next to the 'Dataset' tag. The 'Views' and 'Downloads' statistics are both 0. A blue arrow points to the 'Restricted' label. The 'Files' section shows a single file, 'blobs.tif', with a red 'Restricted' icon and the text: 'The record is publicly accessible, but files are restricted to users with access.' The 'Citations' section shows filters for Literature (0), Dataset (0), and Software (0).

This block contains three overlapping screenshots of the Zenodo record page for 'blobs.tif', illustrating the change in access status and statistics over time:

- April 7<sup>th</sup> 2024:** The top screenshot shows 26 VIEWS and 0 DOWNLOADS. A 'Show more details' link is visible.
- May 23<sup>rd</sup> 2024:** The middle screenshot shows 56 VIEWS and 0 DOWNLOADS. A 'Show more details' link is visible.
- Later date:** The bottom screenshot shows 0 VIEWS and 0 DOWNLOADS. A yellow arrow points to the 'Restricted' label, indicating that access was restricted after May 23rd.

# Restricted Access

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



# Restricted Access

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

**Share access**

People Links **Access requests**

☒ Allow authenticated users to request access to restricted files.

☒ Allow non-authenticated users to request access to restricted files.

Enable users and guests to request access to your record's files. When access is requested by someone, you will get an e-mail asking for approval. After you approve a request, users will be granted access and guests will receive a secret link.

Accept conditions

Optional. Specify conditions for access to this record.

Paragraph

Advanced options

**Recommendation: Give access to people, because they cannot easily forward the right to download, and you can revoke it for individuals.**

Published March 18, 2024 | Version v1

**blobs.tif**

Haase, Robert<sup>1,2</sup>, Schätz, Martin

Dataset **Restricted**

26 VIEWS 0 DOWNLOADS

Show more details

Versions

Version v1  
10.5281/zenodo.10829230 Mar 18, 2024

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.10829229](https://doi.org/10.5281/zenodo.10829229). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

External resources

Indexed in

[OpenAIRE](#)

Communities

UCT Prague Zenodo Community

Details

DOI  
[10.5281/zenodo.10829230](https://doi.org/10.5281/zenodo.10829230)

**Files**

**Restricted**

The record is publicly accessible, but files are restricted to users with access.

**Request access**

If you would like to request access to these files, please fill out the form below.

**You are currently not logged in.** Do you have an account? [Log in here](#)

Your email address \*  
Email address

Your full name \*  
Full name

**Request message**

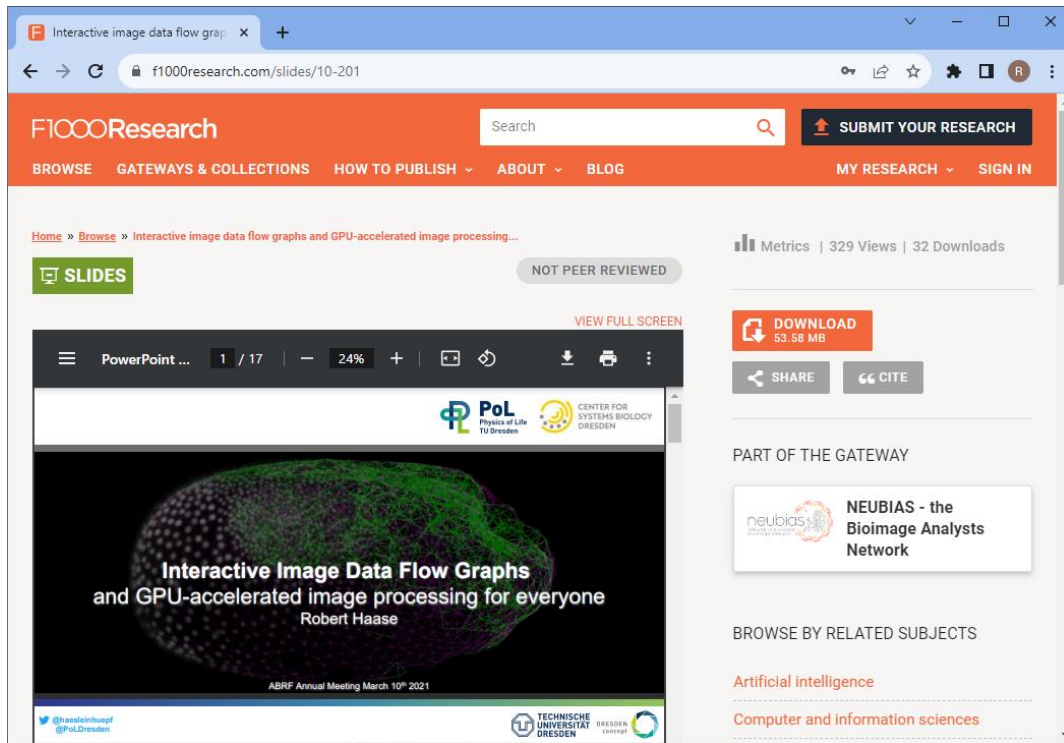
☐ I agree to that my full name and email address is shared with the owners of the record

[Request access](#)




# Open Access

- Others teach using your shared materials




This screenshot shows the F1000Research website interface for slide 10-201. The slide title is "Interactive Image Data Flow Graphs and GPU-accelerated image processing for everyone" by Robert Haase. The slide content features a 3D visualization of a brain with a green network overlay. The slide is part of a presentation titled "PowerPoint ... 1 / 17". The right sidebar shows metrics (329 Views, 32 Downloads), a download button (53.58 MB), and a "PART OF THE GATEWAY" section for NEUBIAS - the Bioimage Analysts Network. The slide is marked as "NOT PEER REVIEWED".



This screenshot shows the F1000Research website interface for slide 11-1175. The slide title is "Interactive Image Data Flow Graphs and reproducible GPU-accelerated image processing" by Martin Schätz. The slide content features a 3D visualization of a brain with a green network overlay. The slide is part of a presentation titled "f1000resear... 1 / 33". The right sidebar shows metrics (20 Views, 5 Downloads), a download button (43.20 MB), and a "PART OF THE GATEWAY" section for NEUBIAS - the Bioimage Analysts Network. The slide is marked as "NOT PEER REVIEWED".

# Open Access

- Others teach using your shared materials



## Quantitative Bio-image Analysis with Python

Search this book...

Quantitative Bio-Image Analysis using Python

Course preparation

Day 1: Introduction to Python and Bio-image Analysis

Day 2: Image Filtering, Segmentation and Feature Extraction

Day 3: Biostatistics and data science

Day 4: Best practices in scientific programming and developing Napari plugins

Day 5: Surfaces and bring your own data

Powered by Jupyter Book


## Quantitative Bio-Image Analysis using Python

This [Jupyter book](#) contains training resources for scientists who want to dive into image processing with Python. It specifically aims for students and scientists working with microscopy images in the life sciences. We presume the attendees have some basic programming and image analysis knowledge. To get everyone on the same level, we start with Python programming basics, and image analysis basics, we then dive into descriptive statistics for working with measurements and [matplotlib](#) and [seaborn](#) for plotting results. We will process images using [numpy](#), [scipy](#), [scikit-image](#) and [cEseranto](#). We will explore [Napari](#) for interactive image data analysis. Finally, we will use [scikit-learn](#) and [StarDist](#) to process images using machine learning and deep learning techniques.

### Timetable

[Course schedule](#)

	Monday	Tuesday	Wednesday	Thursday	Friday
Morning	<ul style="list-style-type: none"><li>Python basics</li><li>Introduction to Bio-image analysis</li></ul>	<ul style="list-style-type: none"><li>Image Filtering</li><li>Image Segmentation</li></ul>	<ul style="list-style-type: none"><li>Working with tabular data</li><li>Plotting</li></ul>	<ul style="list-style-type: none"><li>Writing good code</li><li>Licensing</li><li>Collaborative script editing</li></ul>	<ul style="list-style-type: none"><li>Working with points and surfaces</li></ul>
Afternoon	<ul style="list-style-type: none"><li>Working with image data</li><li>Introduction to Napari</li></ul>	<ul style="list-style-type: none"><li>Machine learning</li><li>Deep learning</li><li>Feature</li></ul>	<ul style="list-style-type: none"><li>Descriptive statistics</li><li>Unsupervised machine learning</li></ul>	<ul style="list-style-type: none"><li>Modular programming</li><li>Making your own Napari plugin</li></ul>	<ul style="list-style-type: none"><li>Bring your own data!</li></ul>



## Bio-Image Data Science Training Schools 2024

Bio-Image Data Science Training Schools 2024

This is the website of the BIDS Training School @ [ScaDS.AI](#) 2024. We provide all training materials and information about the training school on this page.

### General info

- Date: May 13th-15th 2024
- Place: [ScaDS.AI](#) / Uni Leipzig
- Target audience: Life scientists with interest in learning Bio-image Analysis and Data Science using Python. No prior programming experience required.
- Registration is closed.

This course focuses on bio-image analysis and data science with a focus on fluorescence microscopy imaging data and downstream analysis. Our training includes these python libraries:

- [aicsimageio](#)

### Course Preparation

Setting up your computer

#### Monday

Python basics

File handling and working with images

Image processing basics

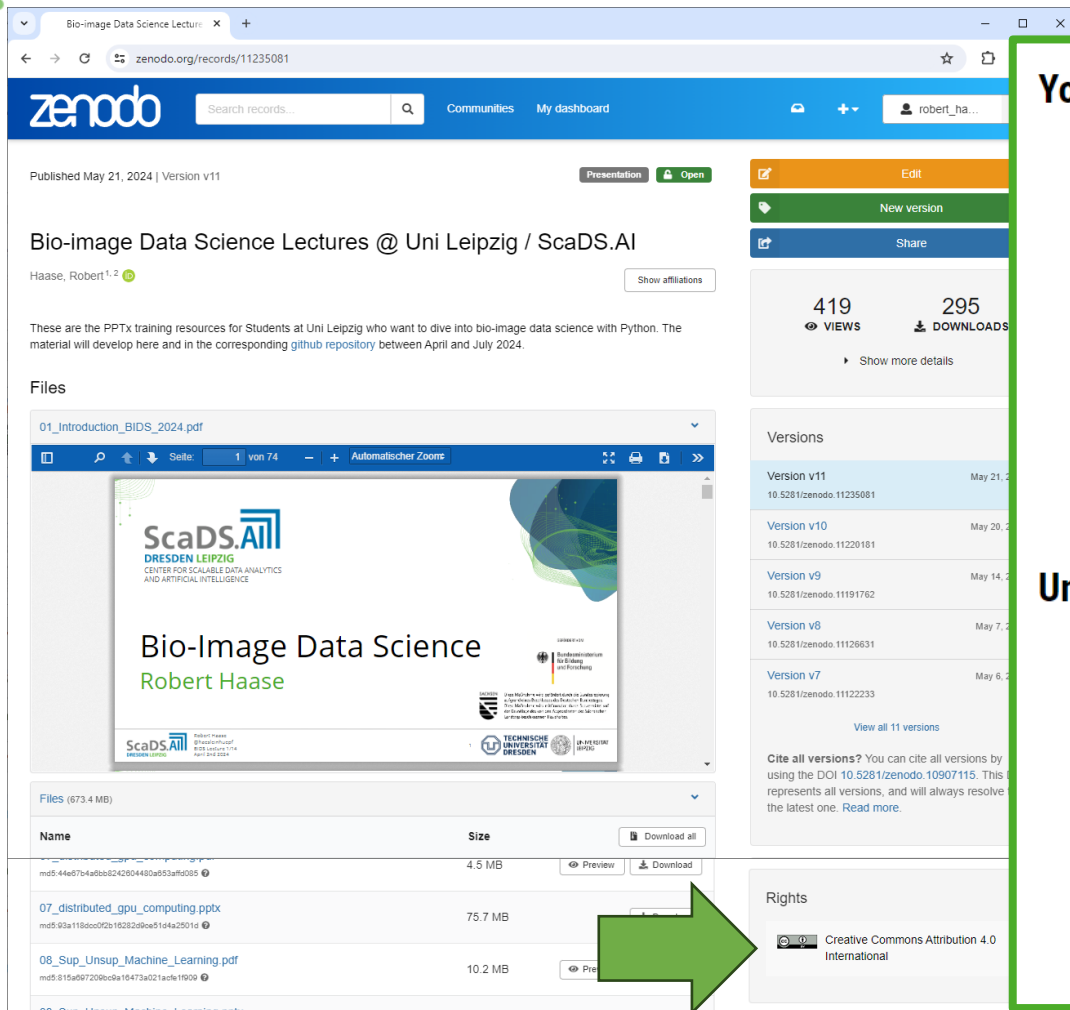
#### Tuesday

Image segmentation

Feature extraction

Machine Learning

# Reusability: Licensing




## You are free to:

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

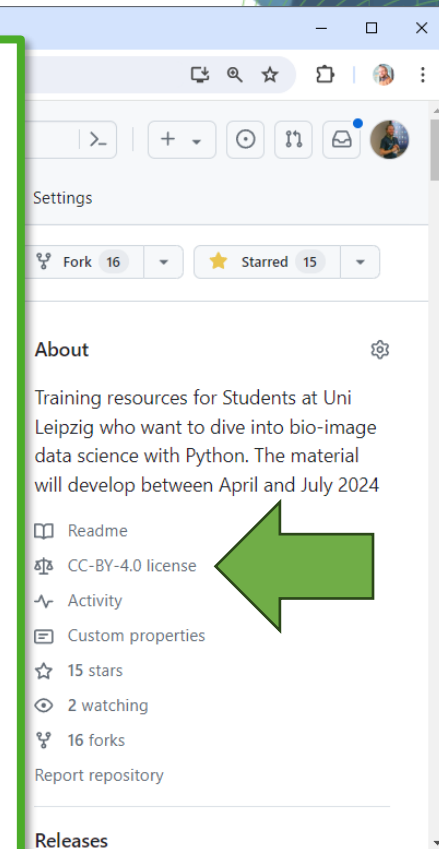
**Adapt** — remix, transform, and build upon the material for any purpose, even commercially.

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

## 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.

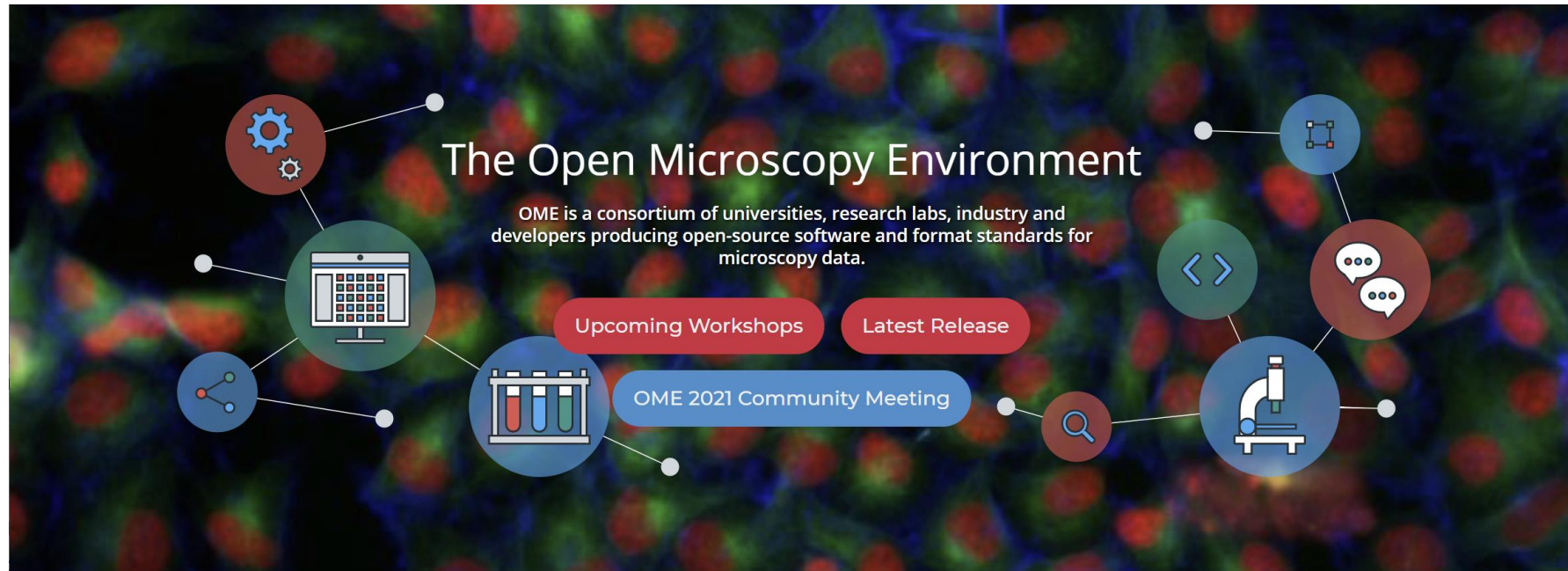
**No additional restrictions** — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.





# Licensing: Creative Commons (CC)

## Example

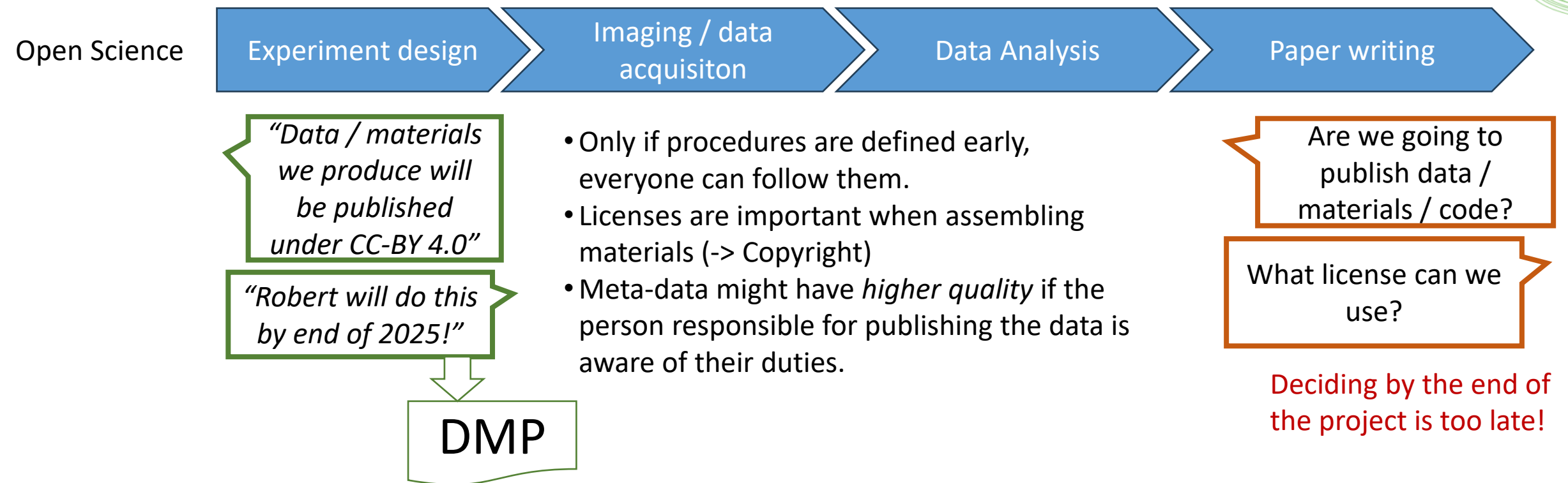


You *must* put such a sentence and keep the link to CC-BY

Figure cropped from <https://www.openmicroscopy.org/> licensed by University of Dundee & Open Microscopy Environment under [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

# Data Management Plans (DMPs)

- Define responsibilities and procedures early!



# Acknowledgements

## Communities & platforms



## BiAPoL team

- Marcelo Zoccoler
  - Johannes Soltwedel
  - Maleeha Hassan
  - Stefan Hahmann
- Former lab members:
- Ryan George Savill
  - Laura Zigutyte
  - Mara Lampert
  - Allyson Ryan
  - Conni Wetzker
  - Somashekhar Kulkarni
  - Till Korten



## 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 



Download  
slides

