

# Research Software Management using git and zenodo











Robert Haase



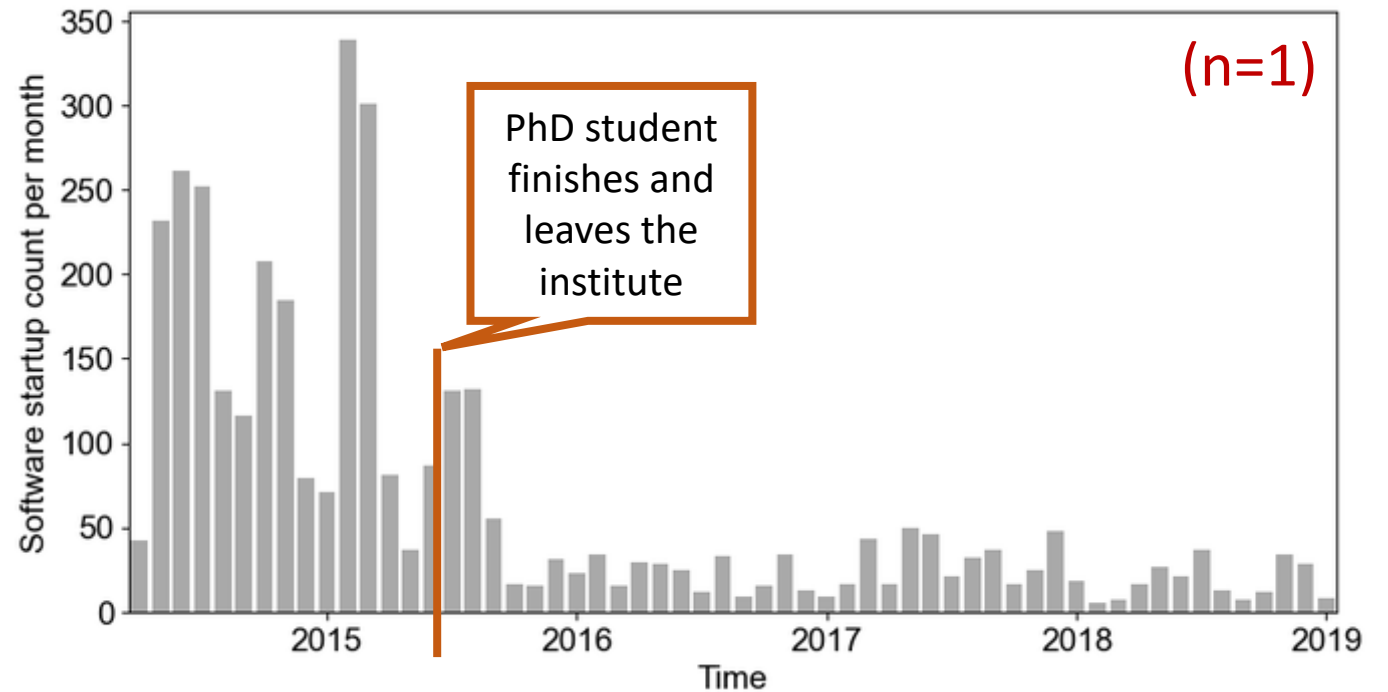
<https://doi.org/10.5281/zenodo.11259494>

# Research Software Management

- 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



# Research Software Management

## • Me, today:

Releases - clEsperanto/pyclesperanto-prototype

Mar 24, 2023  
haesleinhuepf

### 0.23.4 Pre-release

#### 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 Pre-release

#### 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](https://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

Date

# 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

The screenshot shows a GitHub repository page for 'Bio-image\_Analysis\_with\_Python' by user 'BiAPoL'. The repository is public and has 19 watchers, 42 forks, and 181 stars. The main branch is 'main'. The repository contains 10 folders, each representing a lecture slide or topic, with a brief description and the date of the last update (all 'last year').

Folder Name	Description	Last Update
01_introduction_python_pr...	add lecture slides	last year
02_python_algorithms	Correction in name for Credits	last year
03_image_processing	minor slide updates	last year
04_image_segmentation	fix typo	last year
05_feature_extraction	Add files via upload	last year
06_biostatistics_introduction	Introduction statistics	last year
07_descriptive_statistics	Descriptive statistics, lecture, co...	last year
08_hypothesis_testing	Testing statistics	last year
09_machine_learning	minor slide update	last year
10_machine_learning	Machine Learning by Melissa S...	last year

# git for tracking contributions

- Who did what? Who deserves co-authorship?

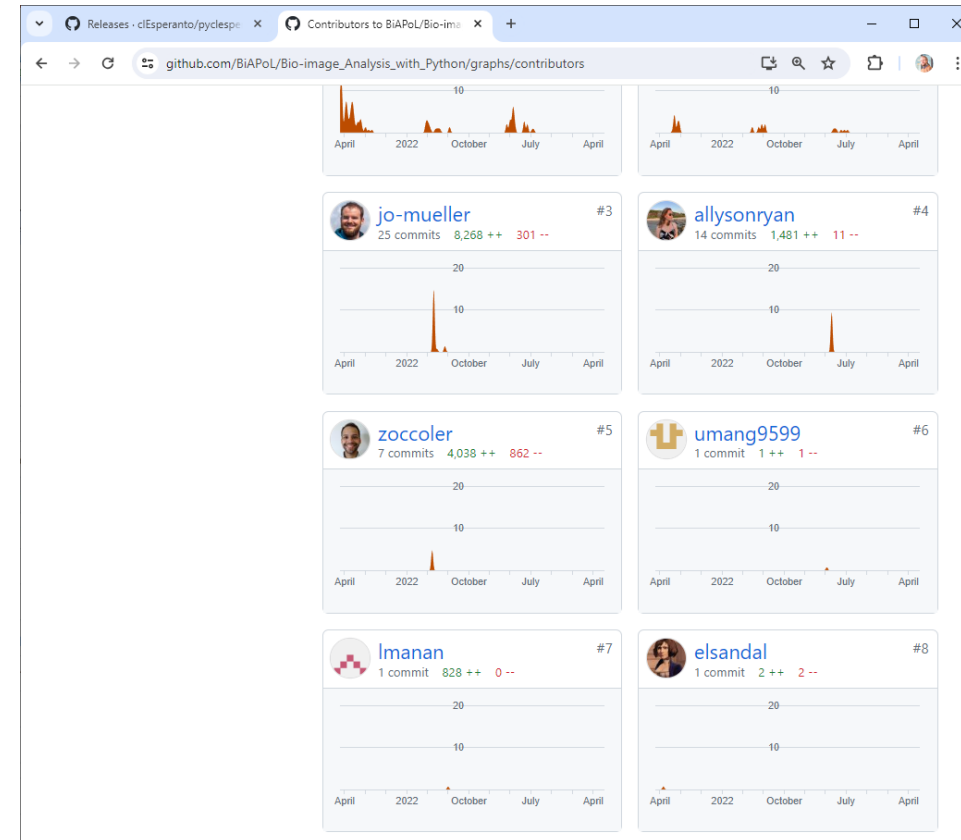
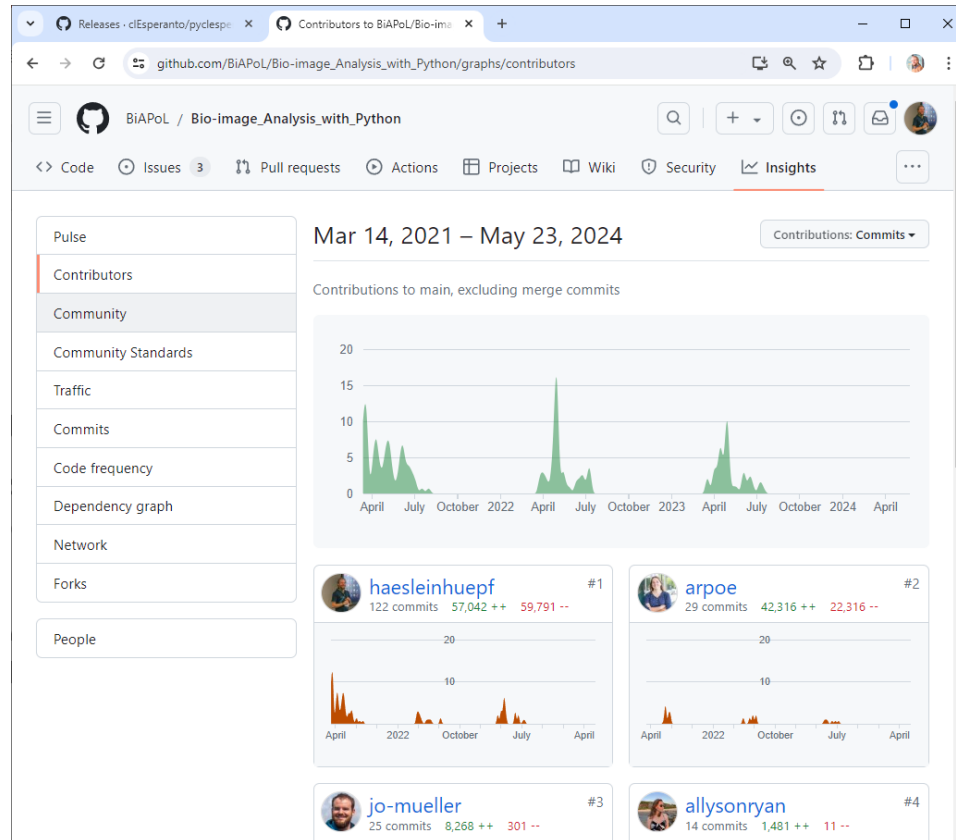
The image displays three browser windows illustrating a Git workflow:

- Left Window:** Shows the commit history for the repository `BiAPoL/Bio-image_Analysis_with_Python`. The history includes commits such as "added missing link to slides", "second part of summary", "added summary slides, links", "Added dimensionality reduction", "update slides", "added DL slides + links", and "deep learning exercises". A blue arrow points from the "deep learning exercises" commit to the middle window.
- Middle Window:** Shows the details of a specific commit (59cab22) titled "deep learning exercises". It lists the files changed: `11_deep_learning/cellpose.ipynb` (467 lines) and `11_deep_learning/stardist.ipynb` (572 lines). A blue arrow points from the `cellpose.ipynb` file to the right window.
- Right Window:** Shows the content of the `cellpose.ipynb` file. The title is "Image Segmentation with CellPose". The content includes a description of CellPose, a list of links (e.g., "Cellpose in Nature Methods"), and a code block for installing Python libraries: 

```
In [1]: !pip install cellpose stackview ipycanvas==0.11
```

# git for tracking contributions

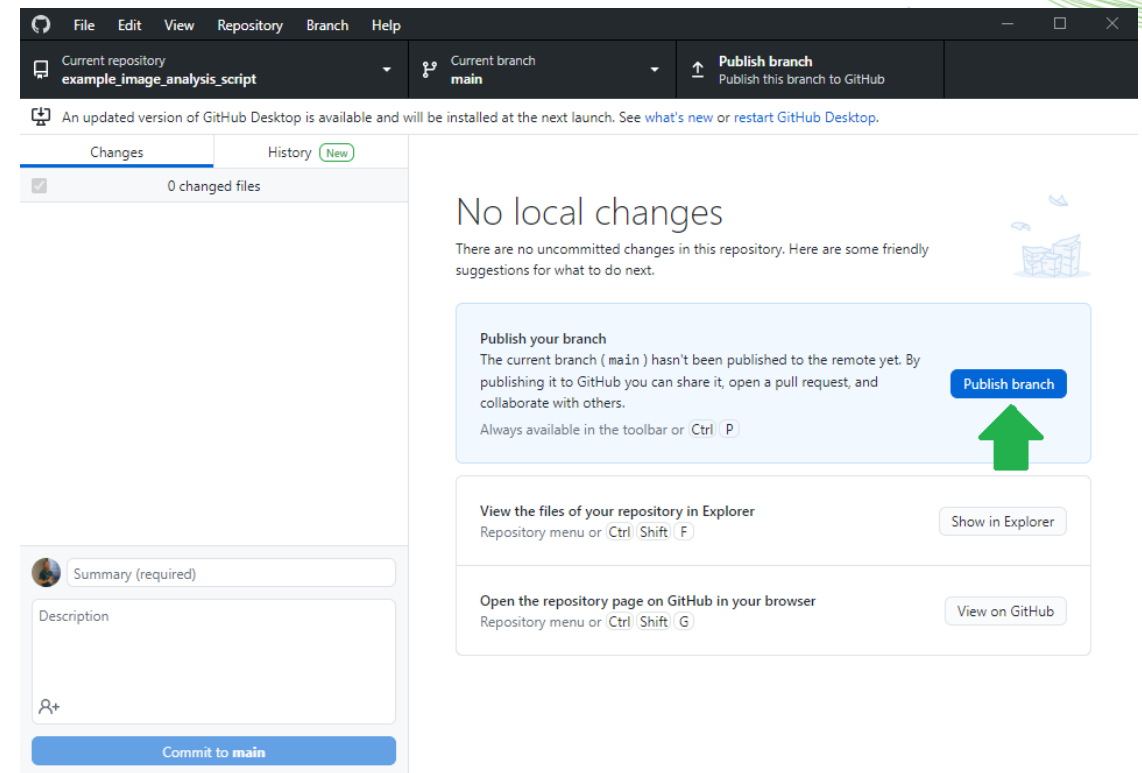
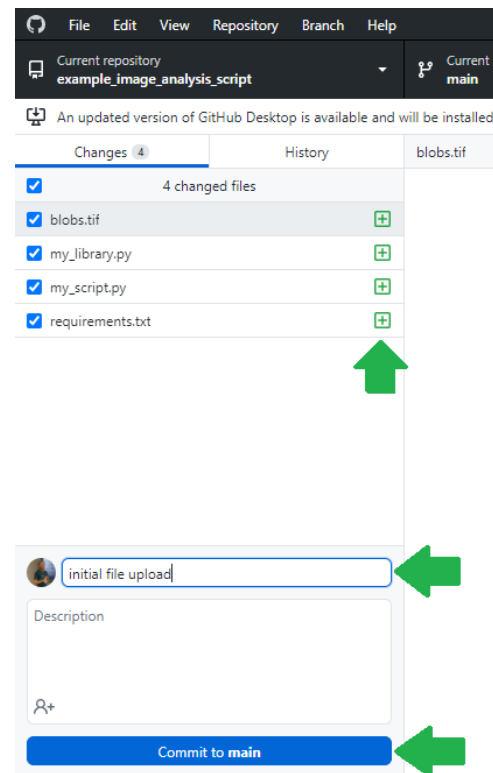
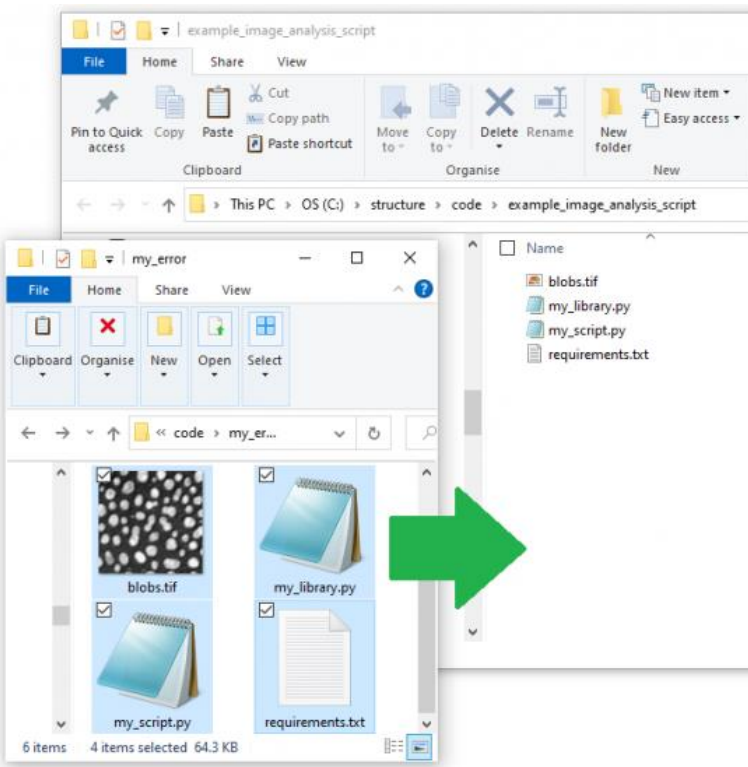
- Who did what? Who deserves co-authorship?





# Github Desktop

- An app for managing files locally and online syncing



# Github pages

- Project documentation / websites

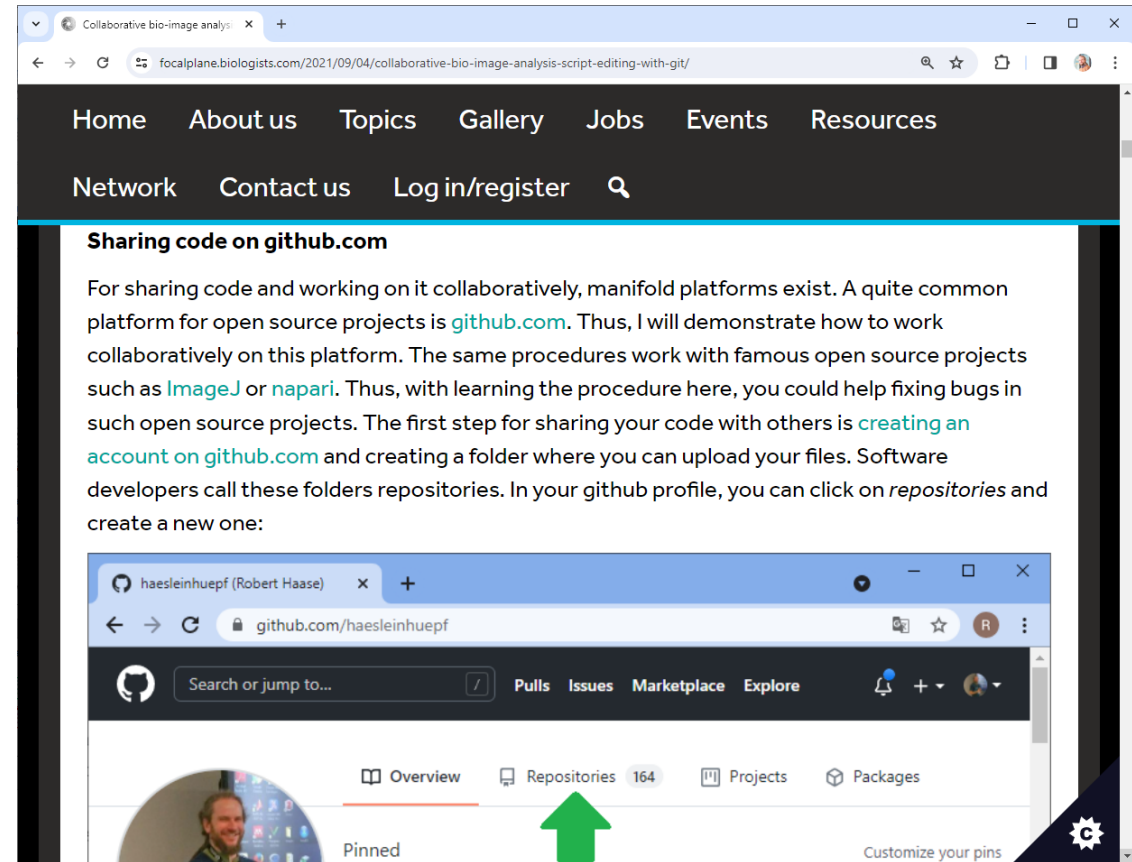
The screenshot shows the GitHub repository page for ScaDS/BIDS-training-2024. The repository is public and has 7 stars and 3 forks. The main branch is selected. The file list shows a directory structure with folders for workflows, data, and training days (day0\_preparation, day1.1\_python\_basics, day1.2\_file\_handling, day1.3\_image\_processing\_..., day2.1\_image\_segmentation, day2.2\_feature\_extraction, day2.3\_machine\_learning). The commit history shows a merge pull request by haesleinhuepf with 137 commits.

The screenshot shows the website for the Bio-Image Data Science Training Schools 2024. The website is hosted on GitHub Pages and features the ScaDS.AI logo. The main heading is "Bio-Image Data Science Training Schools 2024". The text states: "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." The website includes sections for "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), and "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). The website also includes a link to the training materials: <https://scads.github.io/BIDS-training-2024/>.



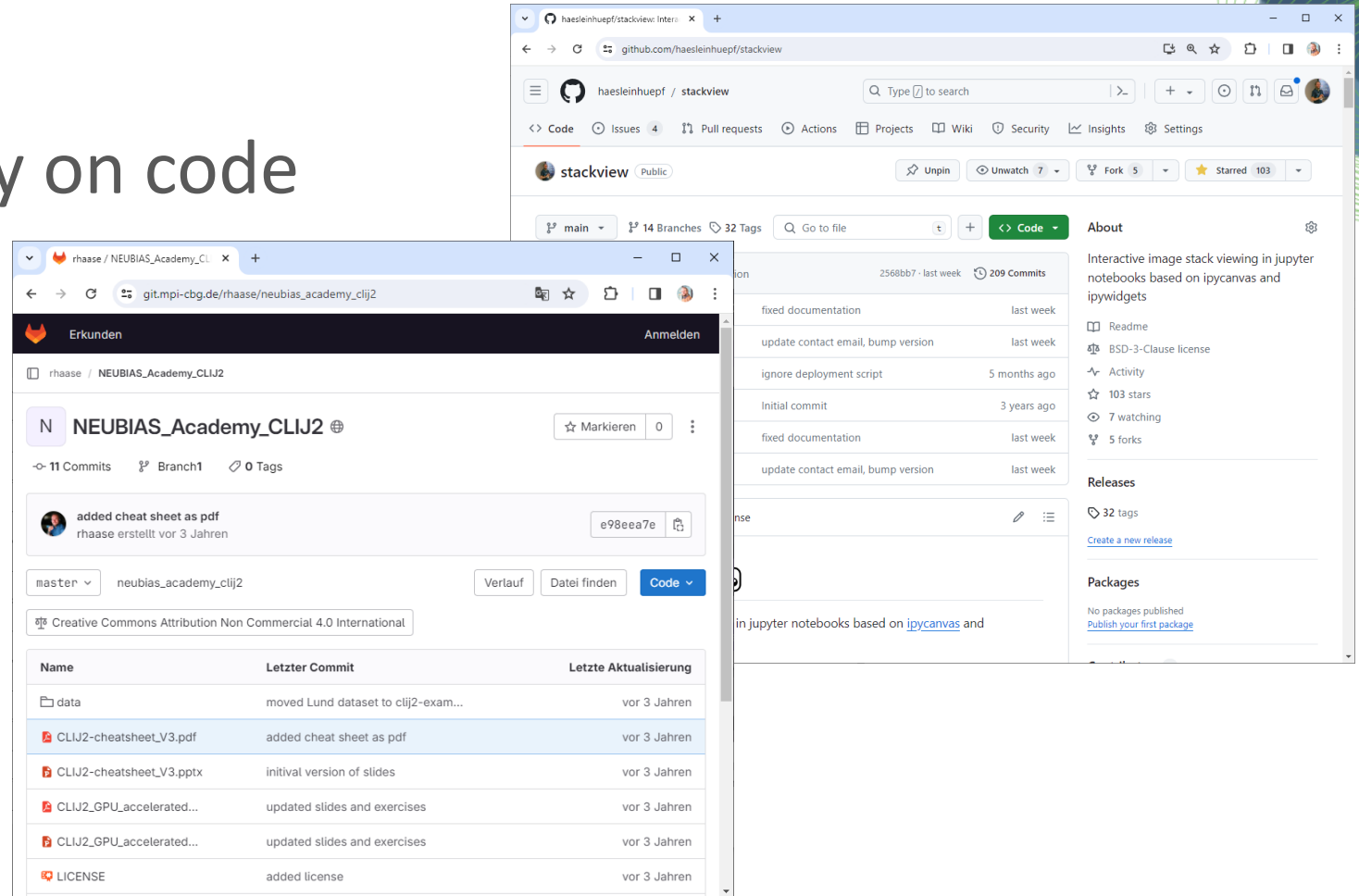
# Git

- Ideal for working on code together online



# 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

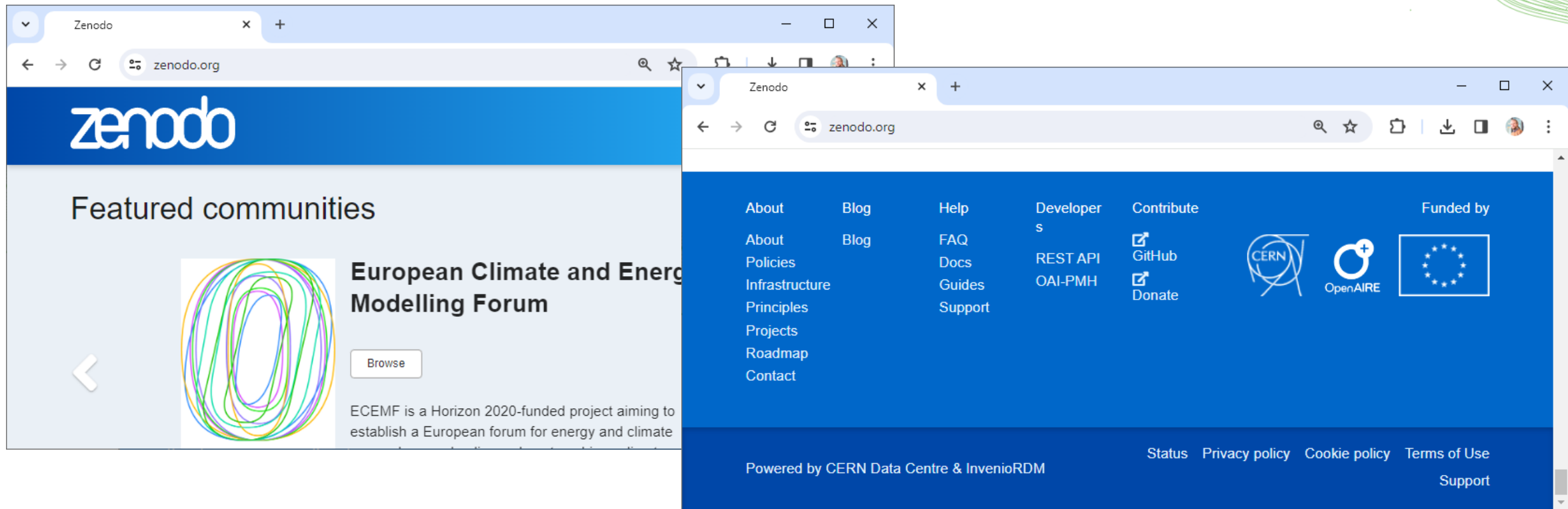


# Github versus Zenodo

- Github
  - Good for collaboration
  - Small changes committed separately
  - Full history of code available
- Zenodo
  - Publicly funded infrastructure
  - Good for Archiving
  - Only substantial releases are archived

# Zenodo

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



# Zenodo-Github integration

The screenshot shows a web browser window with the URL `sandbox.zenodo.org/account/settings/github/`. A yellow warning banner at the top states: "This is the Zenodo Sandbox instance, used for testing purposes. DOIs created in this instance are not real and will not resolve. You can find the production instance of Zenodo at zenodo.org". The Zenodo logo and navigation menu are visible. The main content area is titled "My account" and features a sidebar with settings options: Profile, Change password, Security, Notifications, Linked accounts, Applications, and GitHub (selected). The main content is titled "GitHub Repositories" and includes a "Get started" section with three steps: 1. Flip the switch (with an "ON" toggle), 2. Create a release (with a link to "create a release"), and 3. Get the badge (with an example DOI: `DOI 10.5281/zenodo.8475`). Below the steps are sections for "Enabled Repositories" and "Repositories".



# 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, a 'Delete' button, and a large 'Choose files' button. A note at the bottom indicates a 'max 50 GB per dataset' limit.
- Middle Screenshot:** The 'Deposit' page for a record titled 'Strausberg\_Tribolium\_LA-GFP\_tailpole\_run (Excerpt)'. It shows the author list (Daniela Vorkel, Robert Haase, Eugene Myers) and a rich-text description of the data set.
- Right Screenshot:** The 'Record' page for the deposit. It lists the files uploaded, including ClearControl meta header files, program files, and image stacks. It also shows the 'Versions' section (Version 1, Aug 13, 2020) and a 'Cite as' section with 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

zenodo.org/records/10829230

Service Incident: New DOI registrations are working again. Re-registration of failed DOI registrations (~500) are still affected by the service incident at DataCite (our DOI registration agency).

zenodo

Published March 18, 2024 | Version v1

**Dataset** **Restricted**

**blobs.tif**

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

0 VIEWS 0 DOWNLOADS

Show more details

Files

**Restricted**

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

Citations

Show Literature (0) Dataset (0) Software (0)

Waiting for zenodo.org...

May 23<sup>rd</sup> 2024:

56 VIEWS 0 DOWNLOADS

April 7<sup>th</sup> 2024:

26 VIEWS 0 DOWNLOADS

Show more details

zenodo.org/records/10829230

Service Incident: New DOI registrations are working again. Re-registration of failed DOI registrations (~500) are still affected by the service incident at DataCite (our DOI registration agency).

zenodo

Published March 18, 2024 | Version v1

**blobs.tif**

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

0 VIEWS 0 DOWNLOADS

Show more details

Files

blobs.tif

0 VIEWS 0 DOWNLOADS

Show more details

Versions

Version v1 Mar 18, 2024

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.10829229. This DOI represents all versions, and will always resolve to the latest one. Read more.

External resources

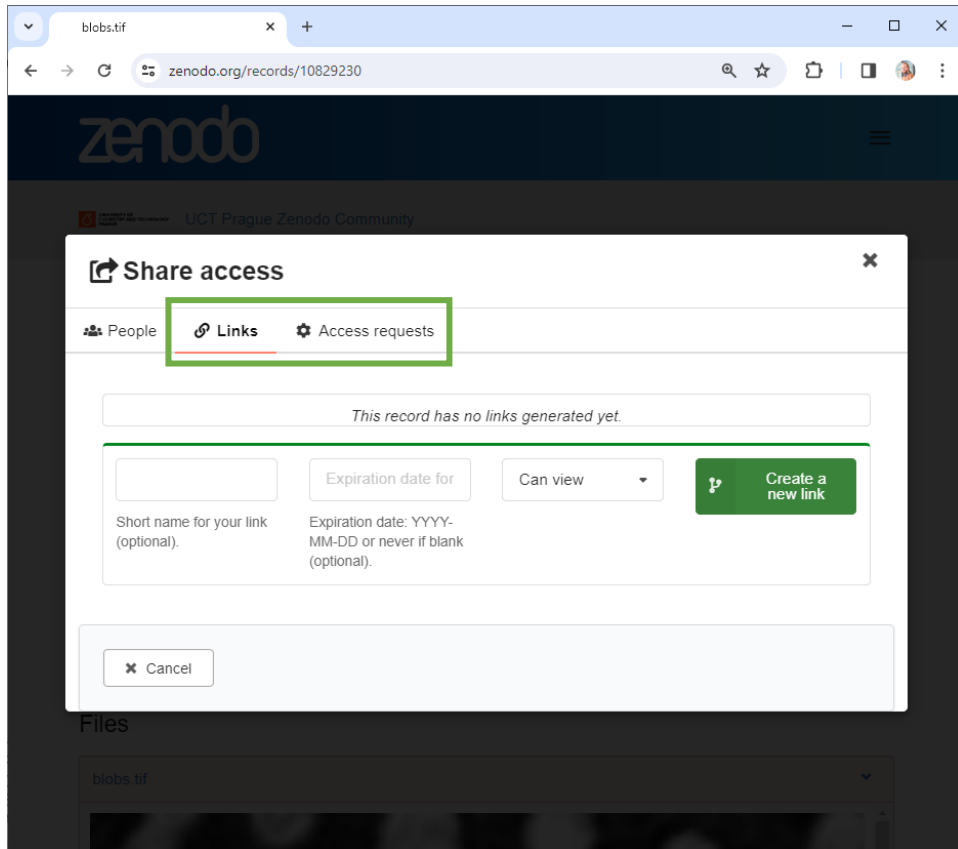
Indexed in

OpenAIRE

Share

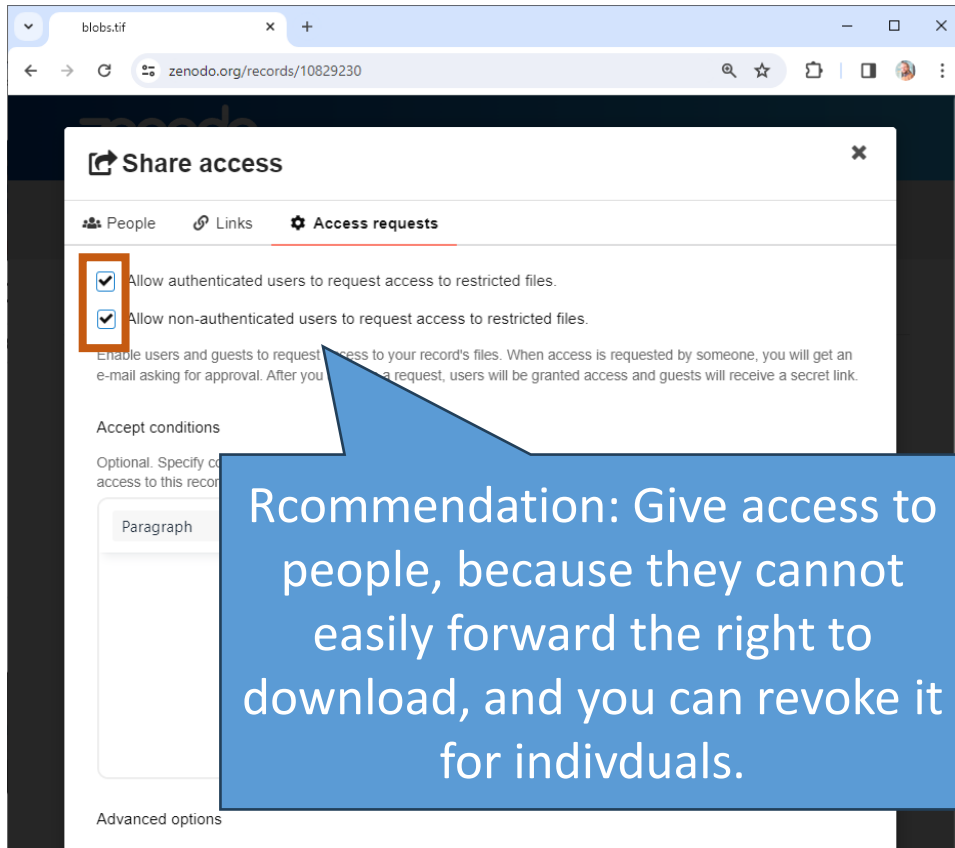
# Restricted Access

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

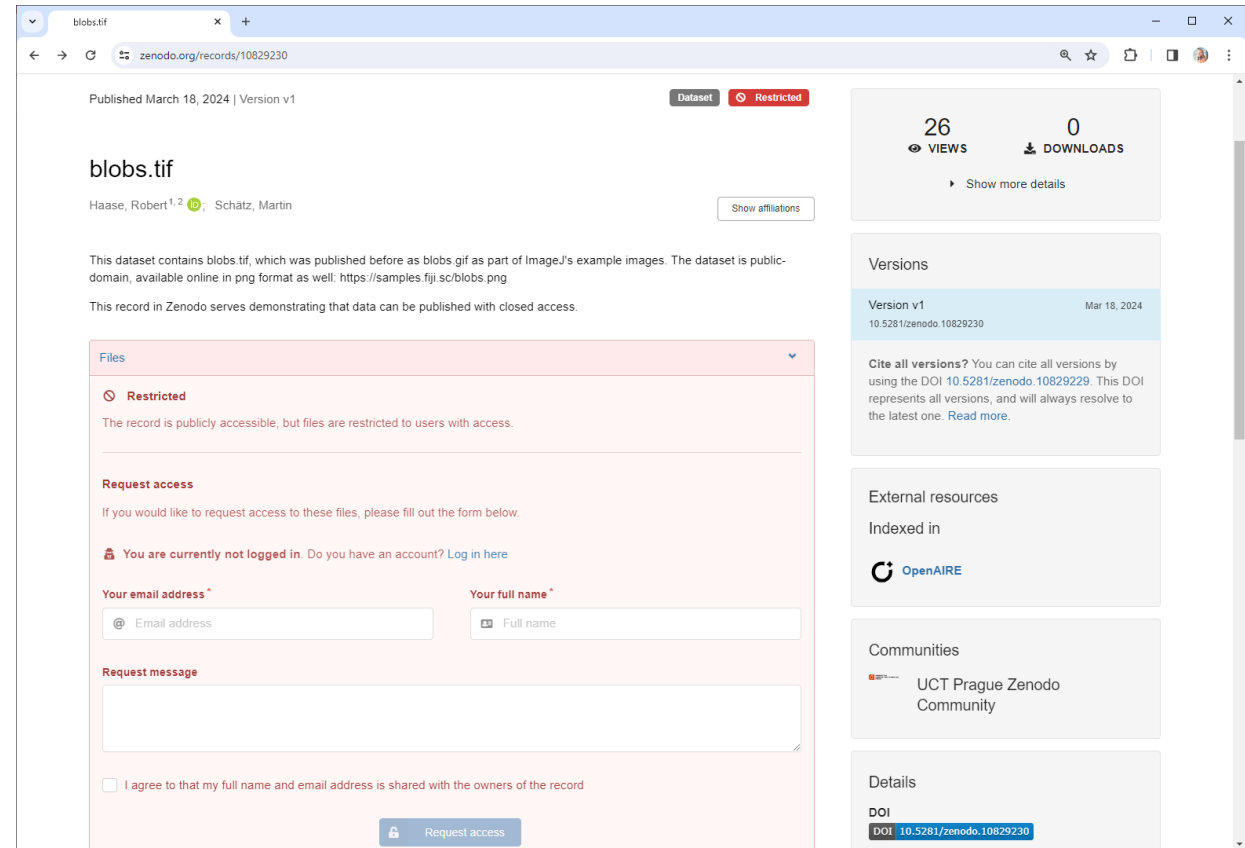


# Restricted Access

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



The screenshot shows the 'Share access' dialog box in Zenodo. The 'Access requests' tab is active, showing two checked options: 'Allow authenticated users to request access to restricted files.' and 'Allow non-authenticated users to request access to restricted files.' A blue callout box points to these options with the text: 'Recommendation: Give access to people, because they cannot easily forward the right to download, and you can revoke it for individuals.'



The screenshot shows the Zenodo record page for 'blobs.tif'. The record is published on March 18, 2024, and is marked as 'Restricted'. It has 26 views and 0 downloads. The record is publicly accessible, but files are restricted to users with access. A 'Request access' form is visible, requiring an email address and full name. The DOI is 10.5281/zenodo.10829230.

# Open Access

- Others teach using your shared materials

The image displays two browser windows side-by-side, showcasing open access training materials.

**Left Window:** The browser address bar shows `biapol.github.io/Quantitative_Bio_Image_Analysis_with_Python_2022/intro.html`. The page title is "Quantitative Bio-Image Analysis using Python". The main content area contains a paragraph: "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 cEsperanto. 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." Below this is a "Timetable" section with a "Course schedule" table.

	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>

**Right Window:** The browser address bar shows `scads.github.io/BIDS-training-2024/intro.html`. The page title is "Bio-Image Data Science Training Schools 2024". The main content area contains a paragraph: "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." Below this is a "General info" section with a list of details:

- 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.

Below the list, it states: "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](#)

# Reusability: Licensing

The screenshot shows the Zenodo record page for 'Bio-image Data Science Lectures @ Uni Leipzig / ScaDS.AI' by Robert Haase. The page includes a search bar, navigation links, and a list of files. A green arrow points from the '07\_distributed\_gpu\_computing.pptx' file to the 'Rights' section.

Name	Size	Download
01_Introduction_BIDS_2024.pdf	4.5 MB	Preview Download
07_distributed_gpu_computing.pptx	75.7 MB	Preview Download
08_Sup_Unsup_Machine_Learning.pdf	10.2 MB	Preview Download

## 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.

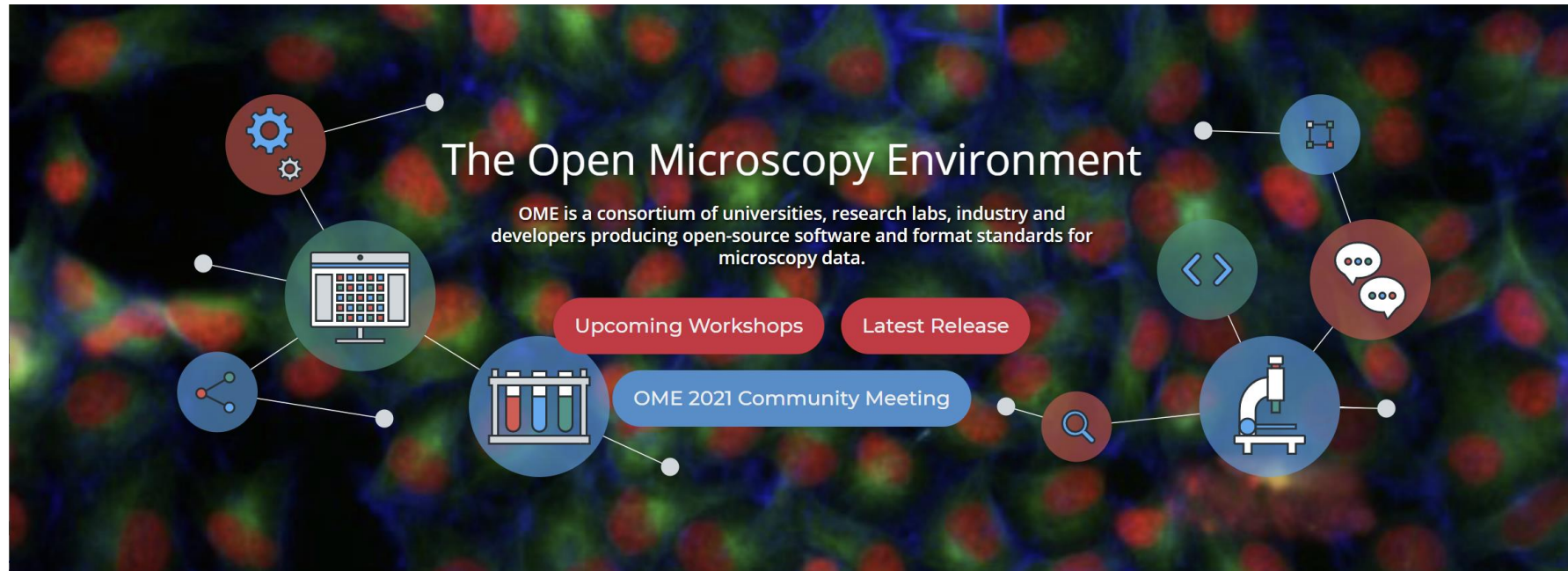
**No additional restrictions** — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

The screenshot shows the GitHub repository page for 'ScaDS/BIDS-lecture-2024: Training resources for Students at Uni Leipzig'. It includes a settings menu, a 'Fork' button (16 forks), a 'Starred' button (15 stars), and an 'About' section. A green arrow points to the 'CC-BY-4.0 license' link in the 'About' section.



# Licensing: Creative Commons (CC)

## Example



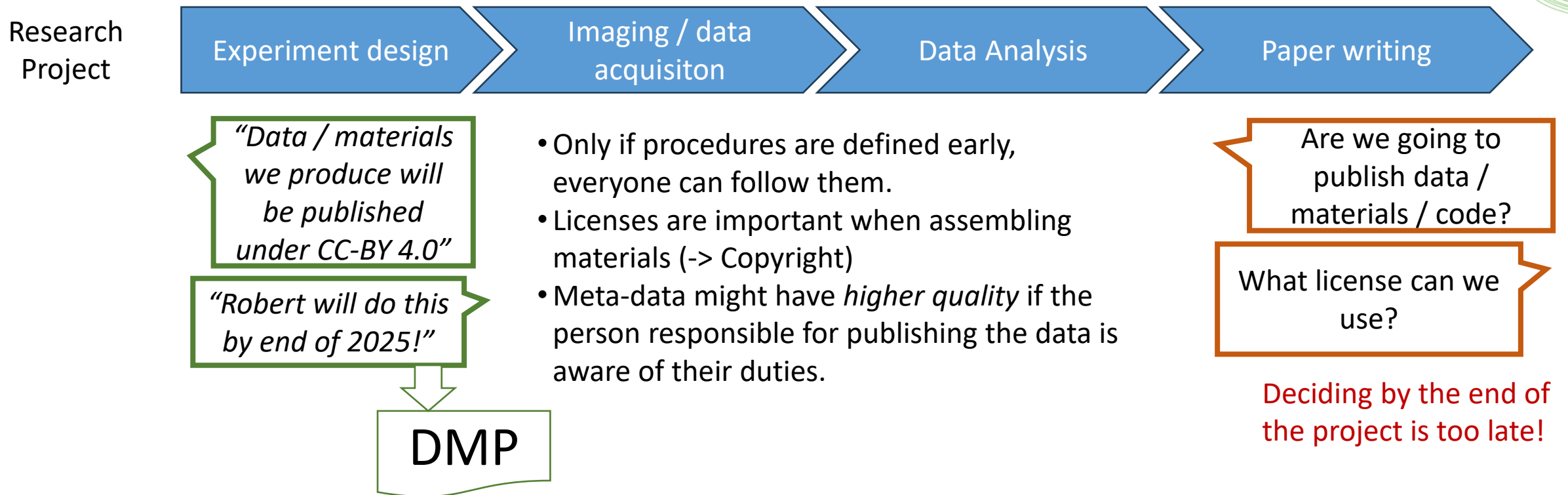
You *must* put such a sentence and keep the link to CC-BY

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

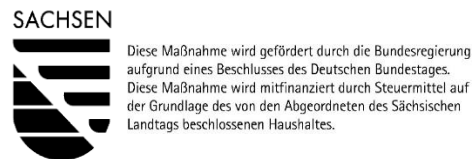


BiAPoL team

- 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



Download slides

