ScaDS.AI DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

NFDI 4 BIOIMAGE

NATIONAL RESEARCH DATA MANAGEMENT INFRASTRUCTURE FOR MICROSCOPY AND BIOIMAGE ANALYSIS

GloBIAS Global BioImage Analysts' Society

GLOBAL BIOIMAGE ANALYST'S SOCIETY

Kollaboratives Arbeiten und Versionskontrolle mit Git

Robert Haase

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

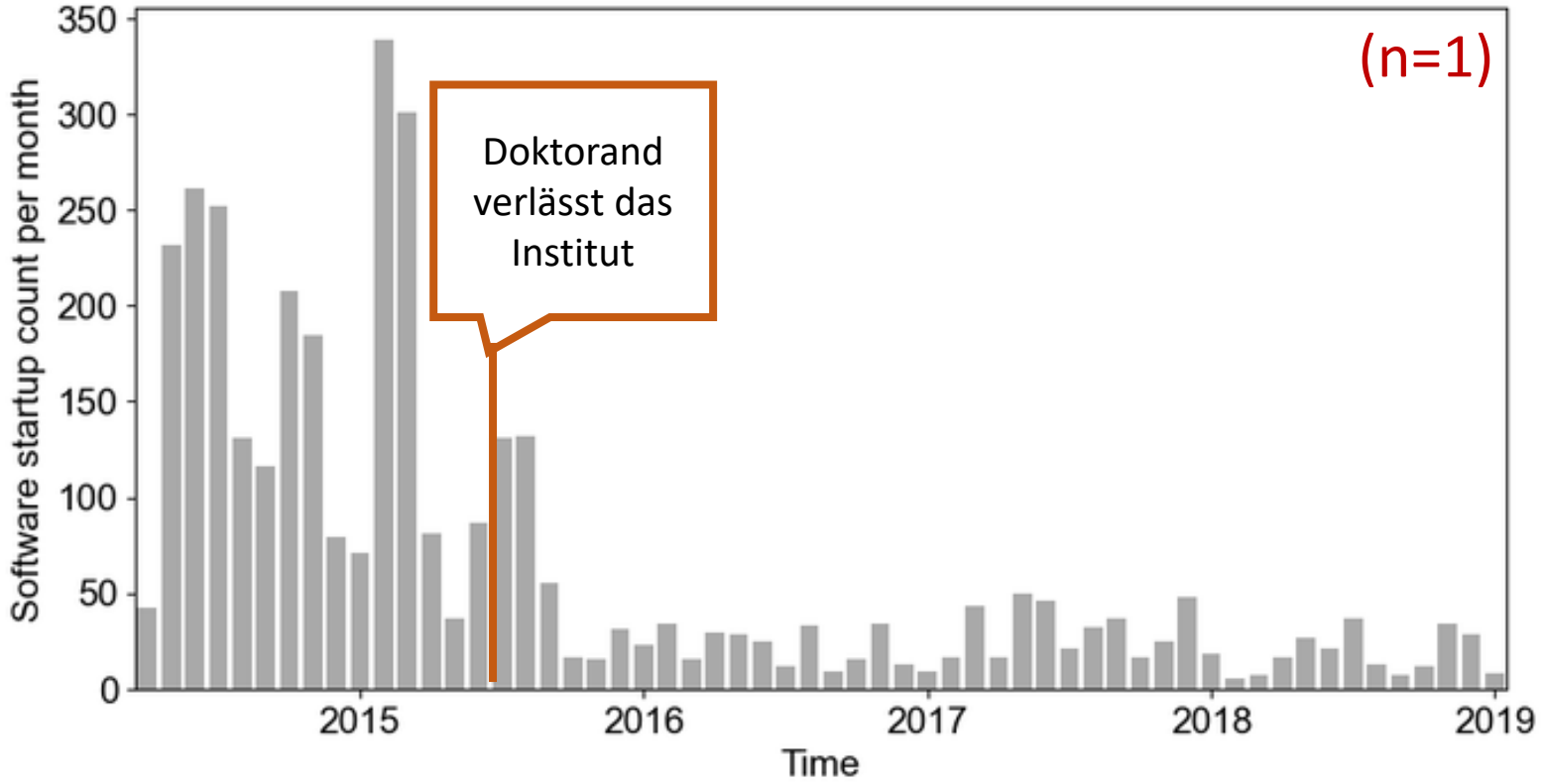
SACHSEN

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.

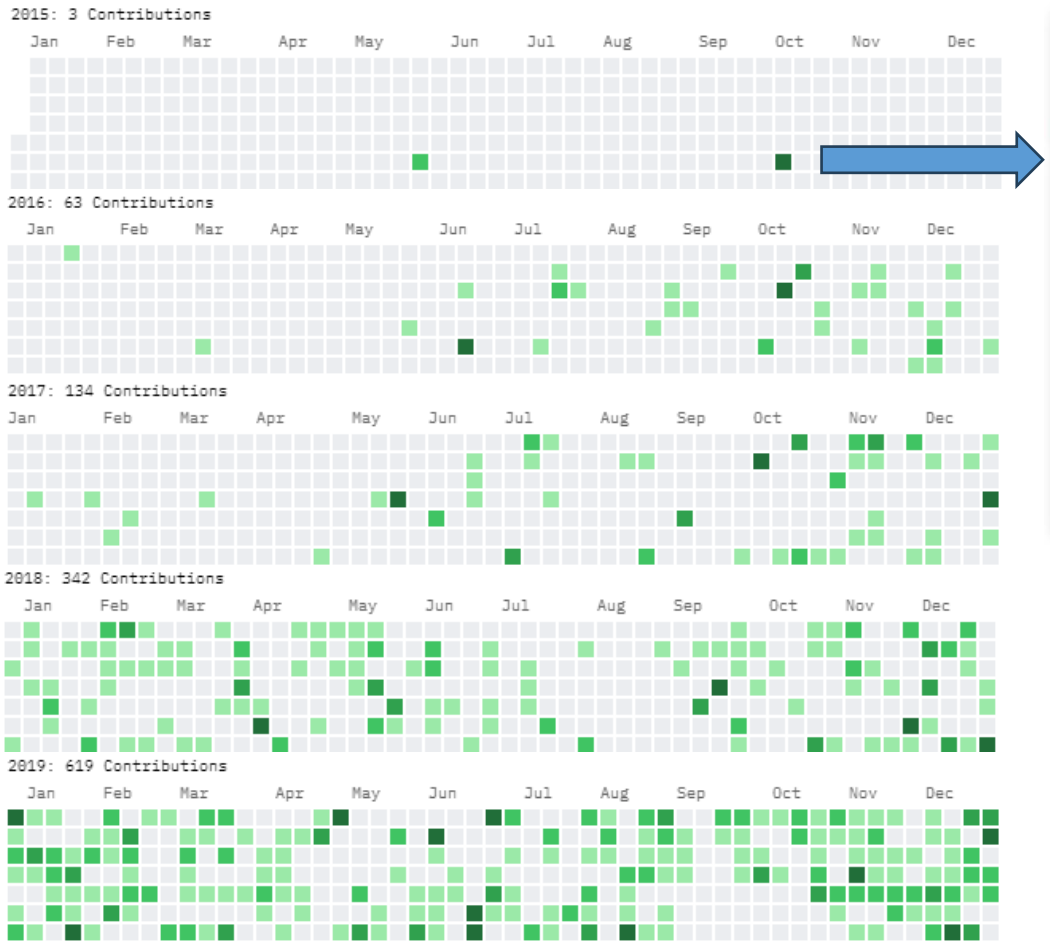
https://zenodo.org/doi/10.5281/zenodo.10966173

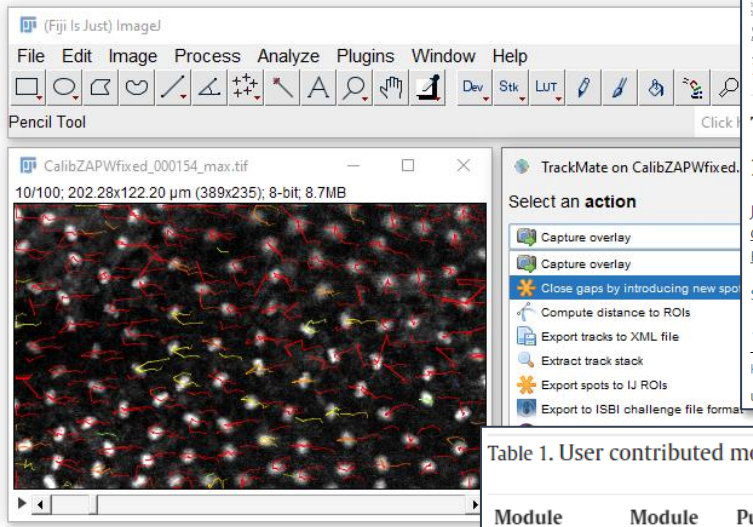
Nachhaltigkeit von Softwareprojekten

- Was passiert wenn der Doktorand/Mitarbeiter das Institut verlässt?



Mitarbeit an Open Source Software





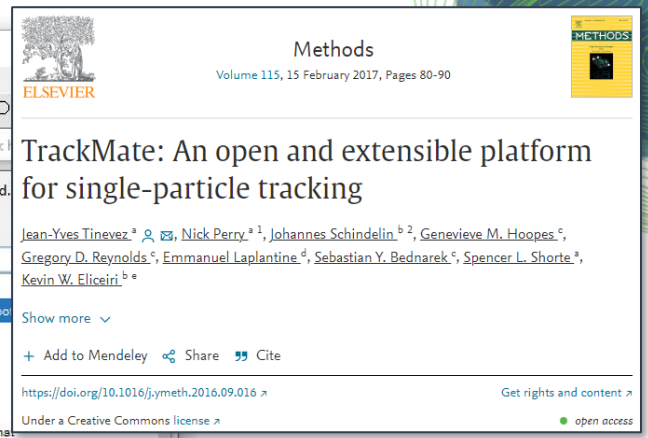


Table 1. User contributed modules of TrackMate v3.4.0.

Module name	Module type	Purpose	Author	Location
Linear tracker	Particle-linking	Linking transported particles by extrapolating their velocity	Ronny Sczech	https://github.com/chicoronny/RonnyTrackMate↗
Batch mode	Plugin	Runs TrackMate in batch, reading		
Close gaps	Generic action	Close gaps in tracks by creating spots in missing frame by linear interpolation of their coordinates	Robert Haase	Integrated into TrackMate v3.4.0

Note: The colour-code is not consistent between the years

Quiz

- Wie oft nutzt Ihr git oder Github?

Nie

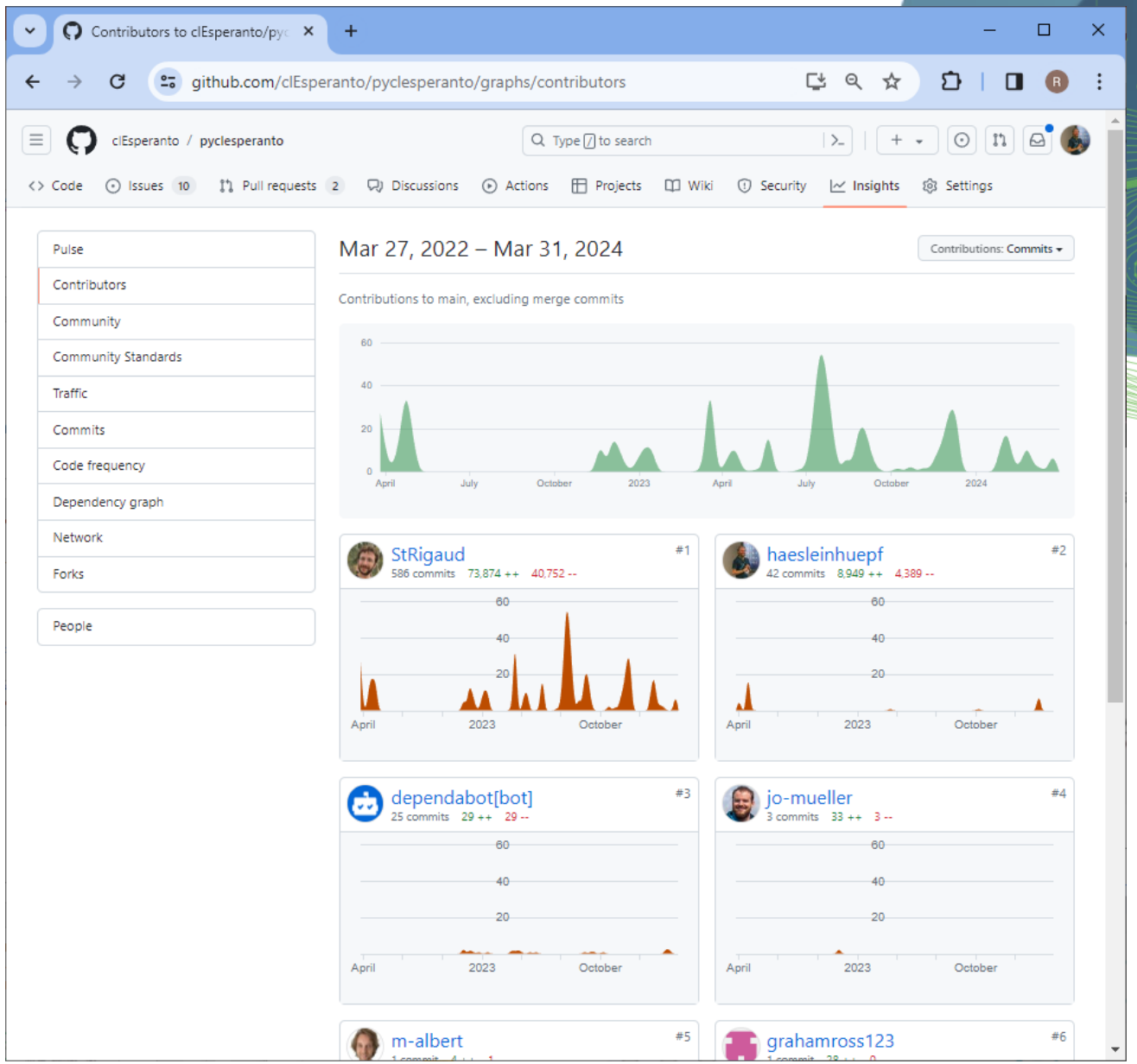
Hin und wieder

1/Woche

Täglich

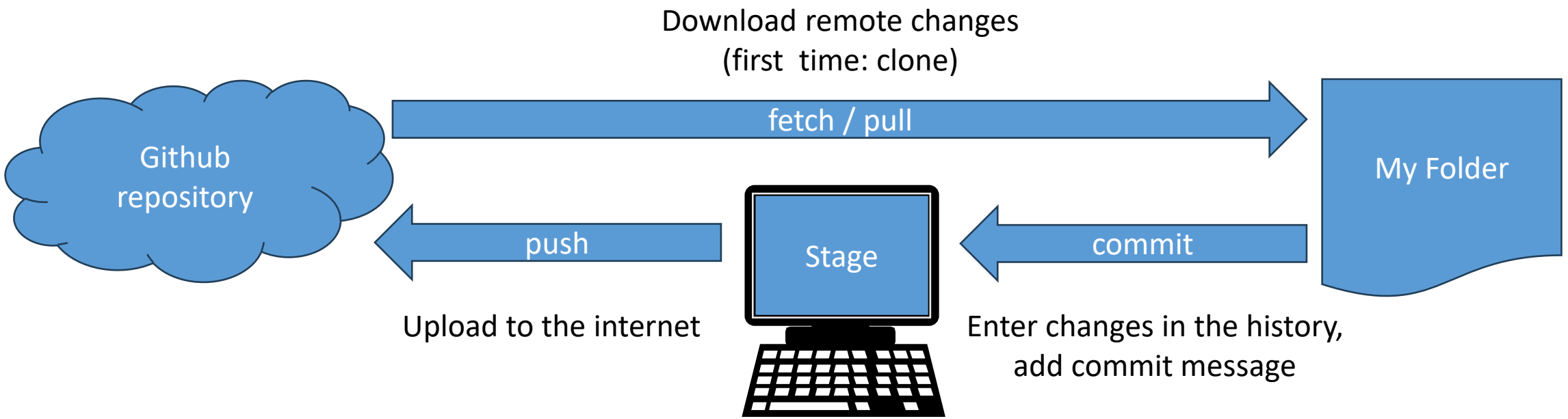
git

- Versionskontrolle ist ein wichtiges Werkzeug fuer Softwareentwickler und Datenwissenschaftler
- Verteiltes Dateisystem mit Logbuchfunktion
- Kontrolle was Teil des Ganzen wird – und was nicht



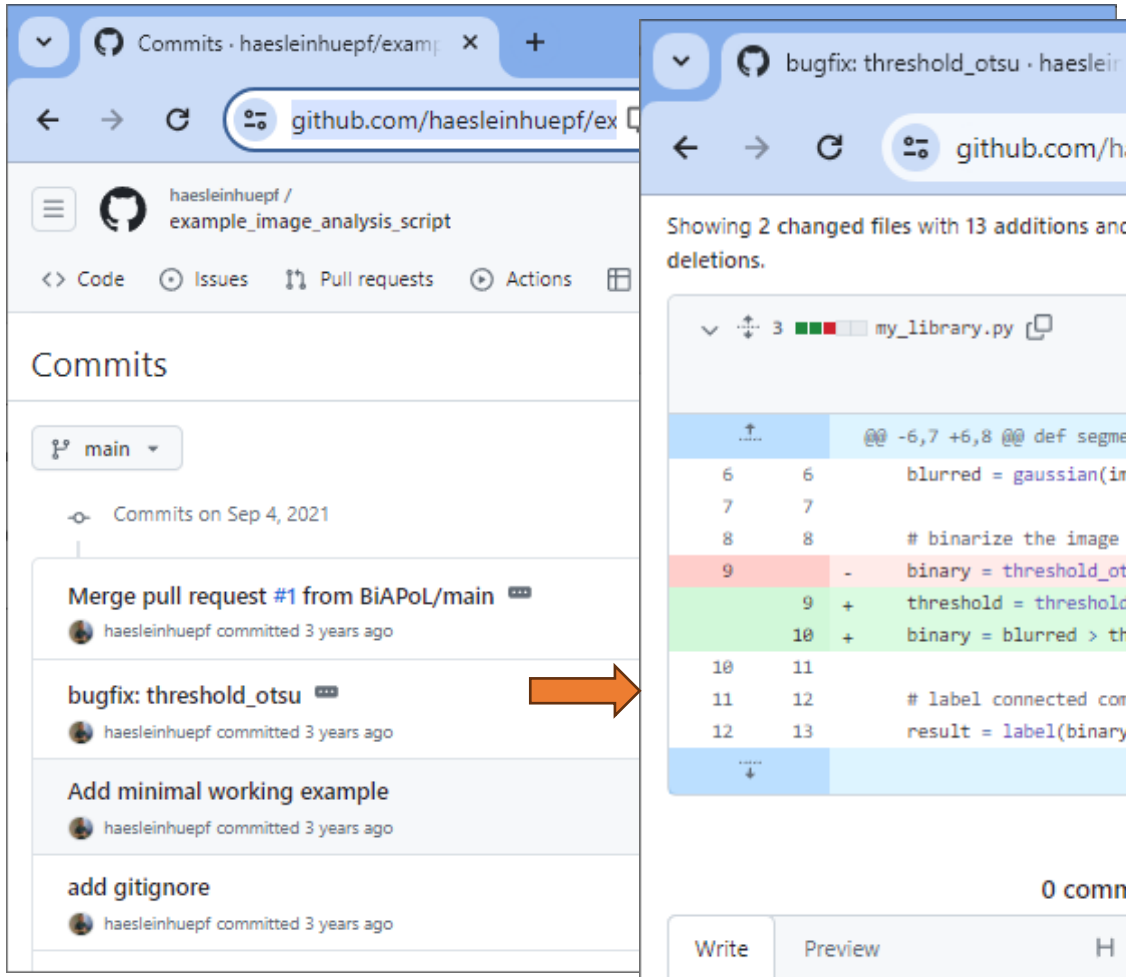
git

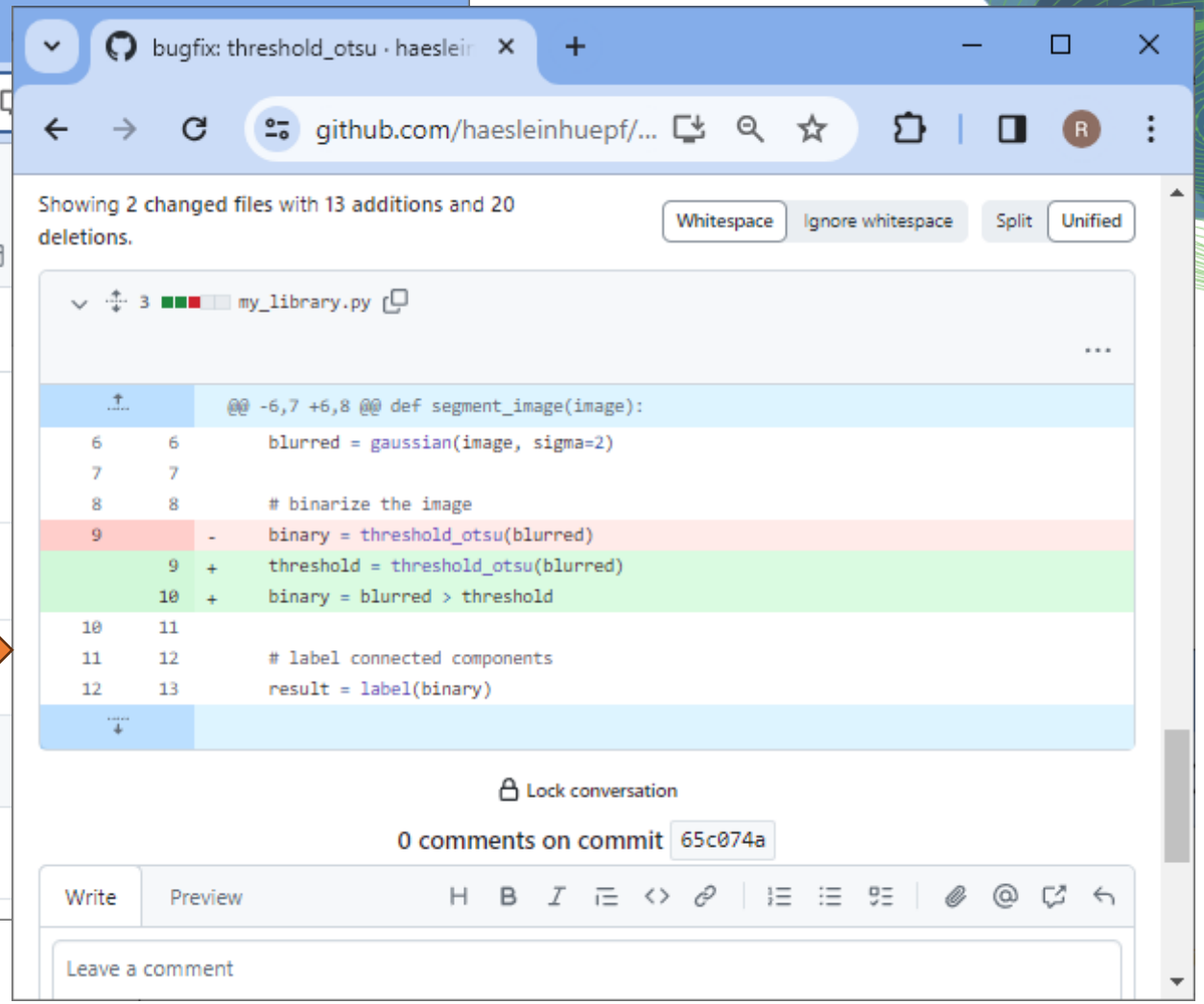
- Git verändert das Arbeiten: Es überführt Änderungen in einen aktiveren Prozess
  (Wodurch wir mehr darüber nachdenken)



git

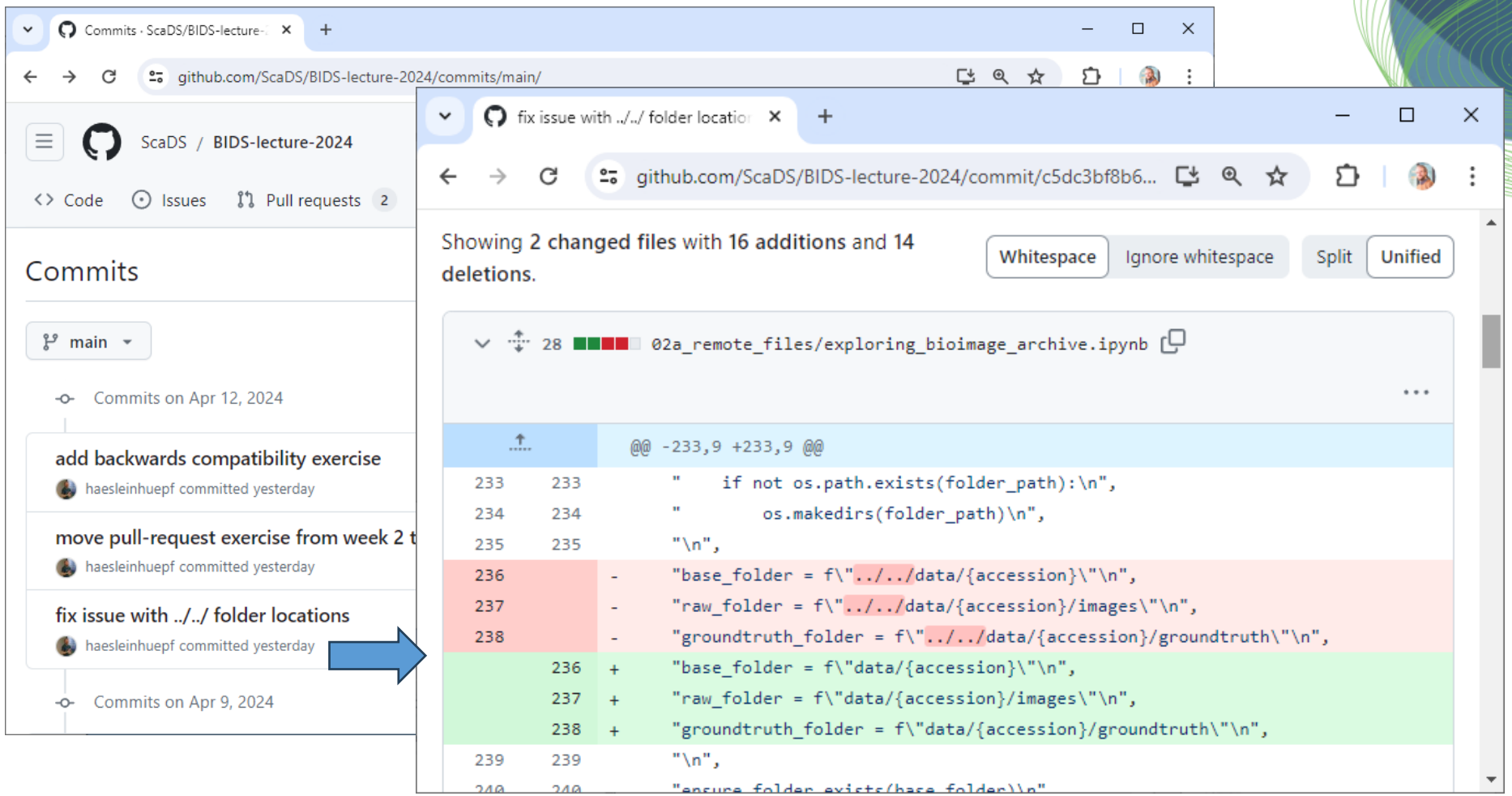
- Wer hat diesen Code geschrieben?
- Wann?
- Wieso?





git

- History
- Track recent changes



Quiz

- Was ist kein FAIR-Prinzip?

Findable	Accessible	Interoperable	Reproducible
			

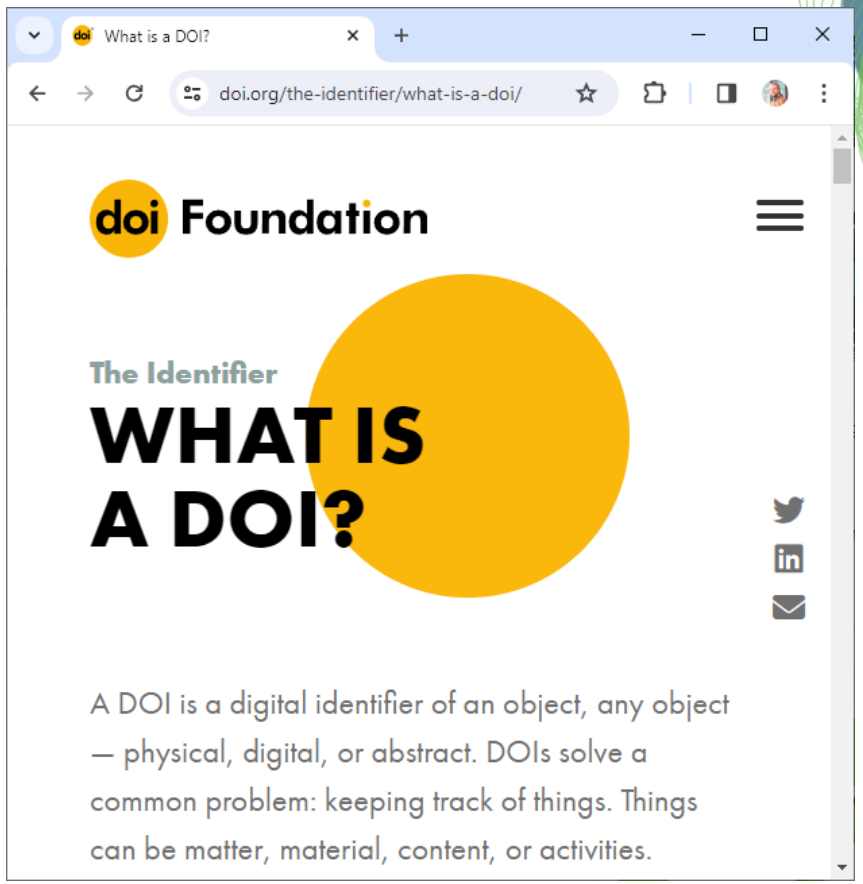
Die FAIR-Prinzipien

- Findable
- Accessible
- Interoperable
- Reusable

Die FAIR-Prinzipien

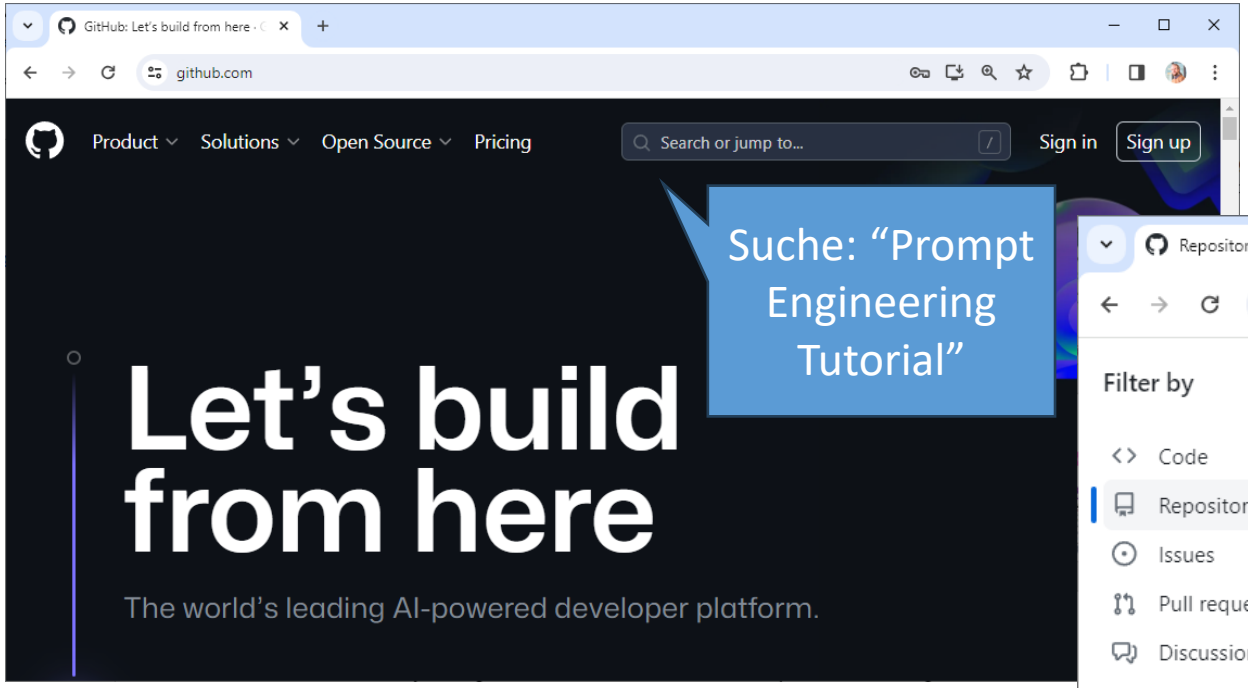
Findable / Findbarkeit

- F1. (Meta)daten sind verbunden mit einem global eindeutigem Identifier (DOI)
- F2. Daten sind mit “reichen Metadaten” beschrieben
- F3. Metadaten beinhalten die DOI, die sie beschreiben
- F4. (Meta)daten sind in einer durchsuchbaren Resource registriert

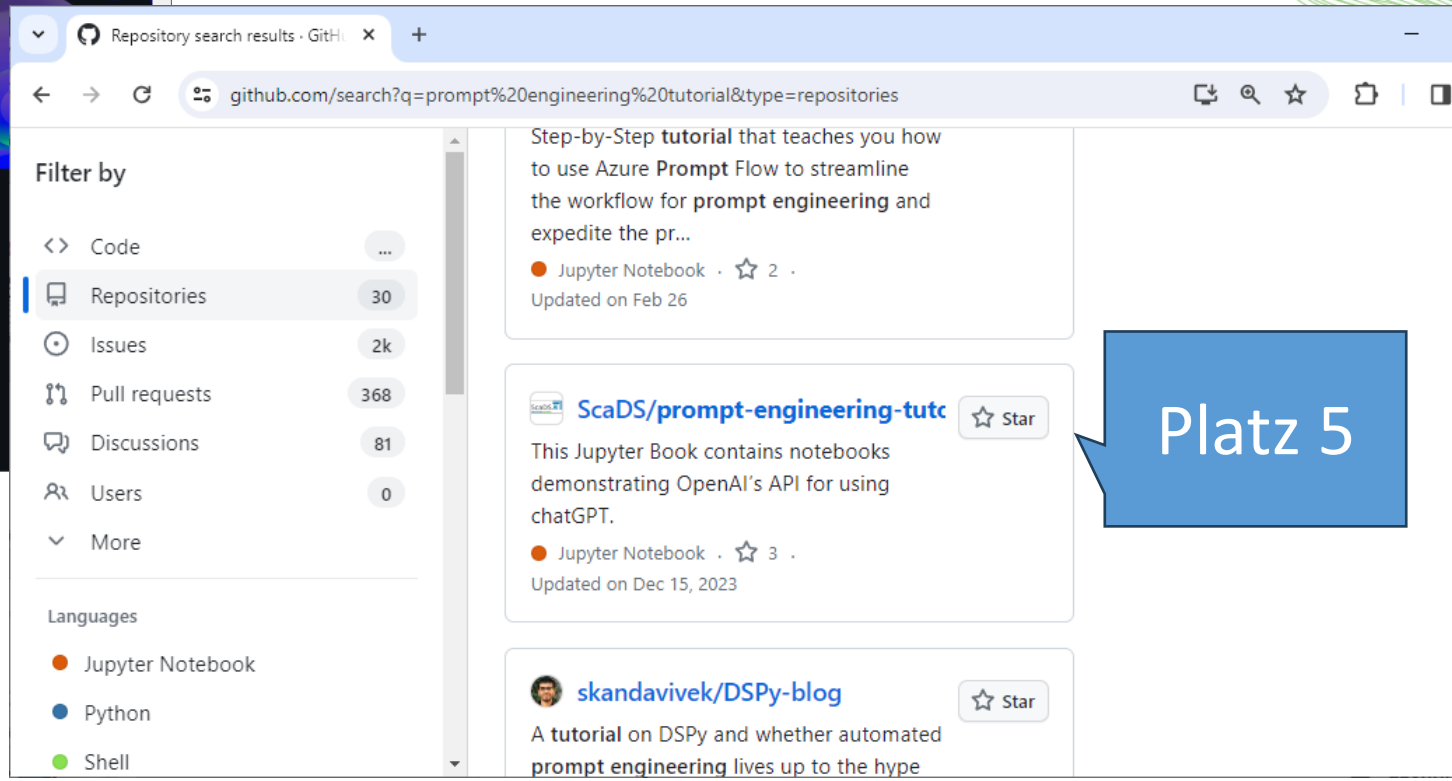


Screenshot source: https://www.doi.org/the-identifier/what-is-a-doi/ licensed CC-BY 4.0
https://www.go-fair.org/fair-principles/
https://biii.eu/introduction-3d-analysis-3d-imagej-suite

Github: Eine durchsuchbare Resource



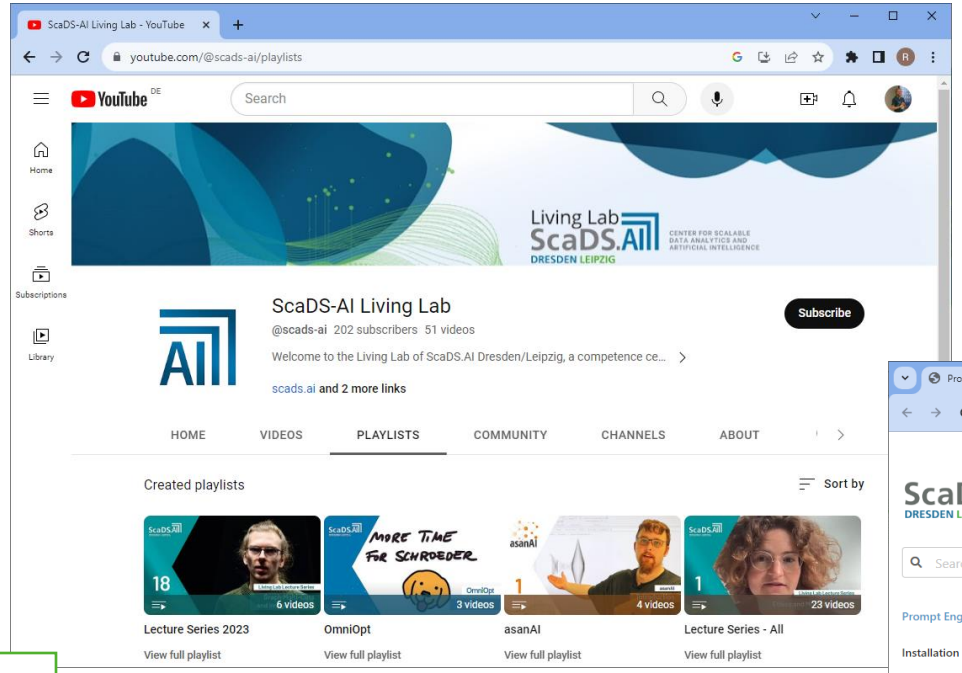
Suche: "Prompt Engineering Tutorial"

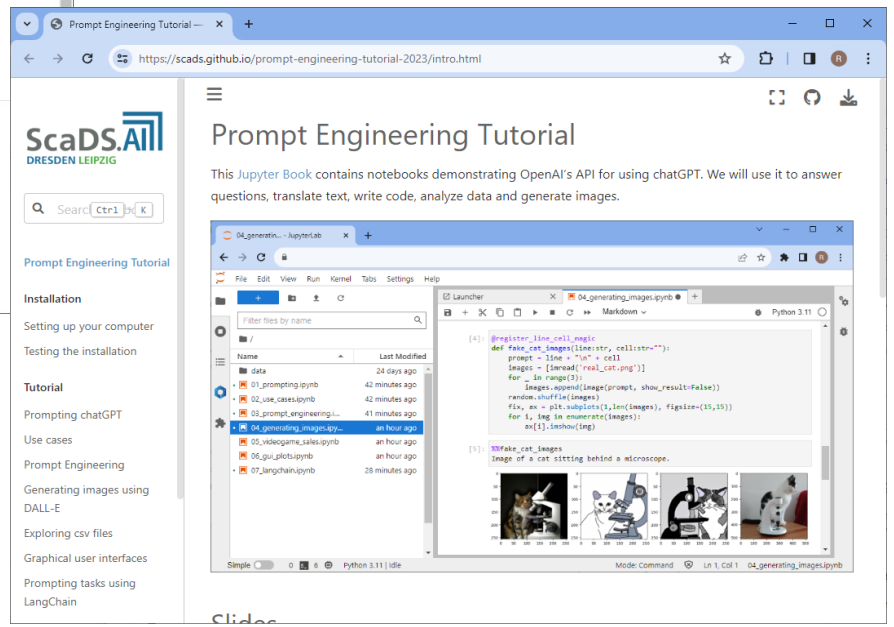


Platz 5

Nutzen: Findbarkeit -> Sichtbarkeit

- YouTube
- Github





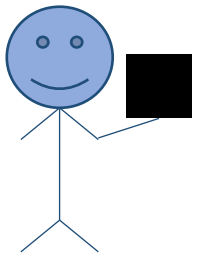
Offenes & FAIRes Teilen
ist ein PR-Instrument

... fuehrt zu

- mehr Softwarenutzer:innen
- neuenw Kollaborationen

Offenheitslevel

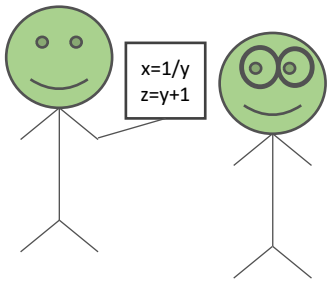
Closed source



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

Hardware device drivers

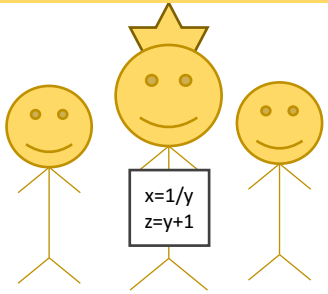
Open source



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

Custom image analysis scripts

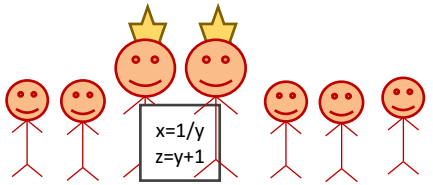
Benevolent dictatorship



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

TrackMate, SNT, MorpholibJ, CLIJ

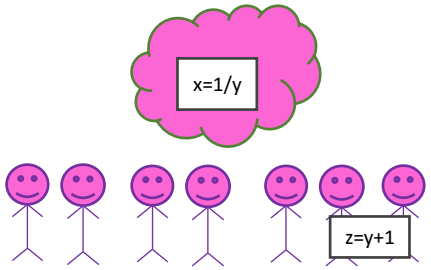
Community driven



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

scikit-image, scipy, OpenCL

Openly extensible

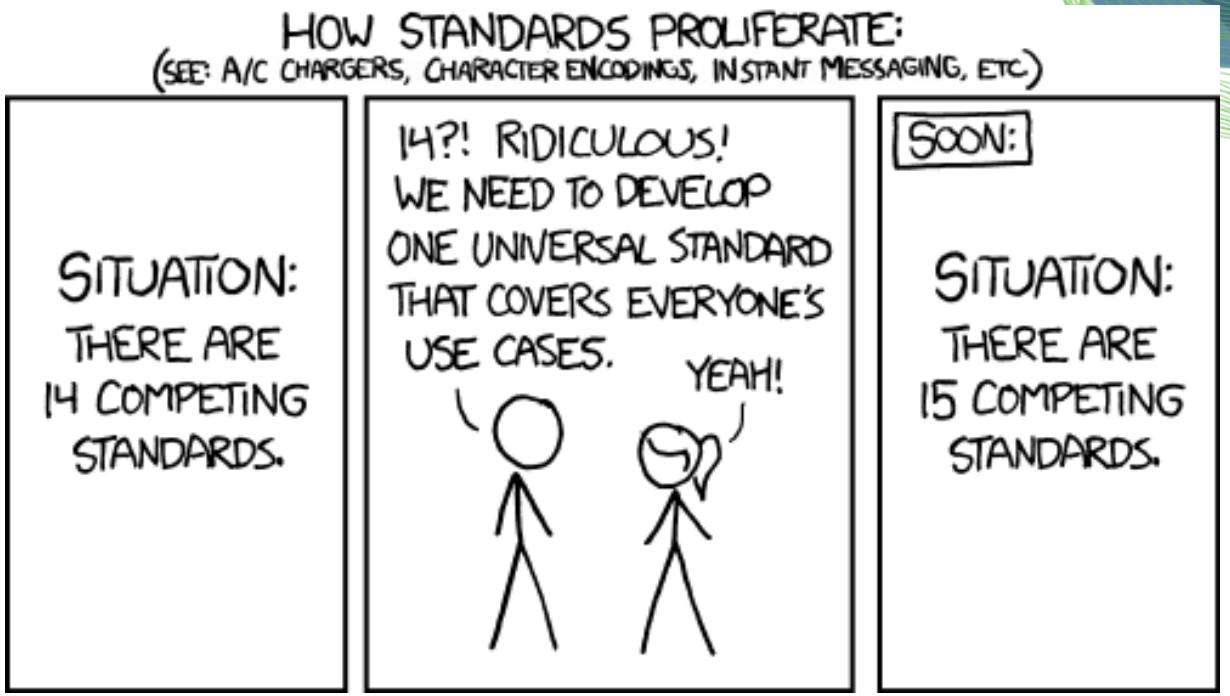


- Openly extensible; without maintainers involved
- Partially community driven

ImageJ, Python, numpy

Die FAIR-Prinzipien

- Accessible / Verfügbar / Zugreifbar
- A1. (Meta)daten koennen ueber ein Standardisiertes Protokoll empfangen werden
  - A1.1 Das Protokoll ist offen, fei und universell implementierbar
  - A1.2 Das Protokoll erlaubt Authentifizierung und Authorisierung, wenn erforderlich
- A2. Metadaten sind verfuegbar, auch wenn die Daten selbst nicht mehr verfuegbar sind



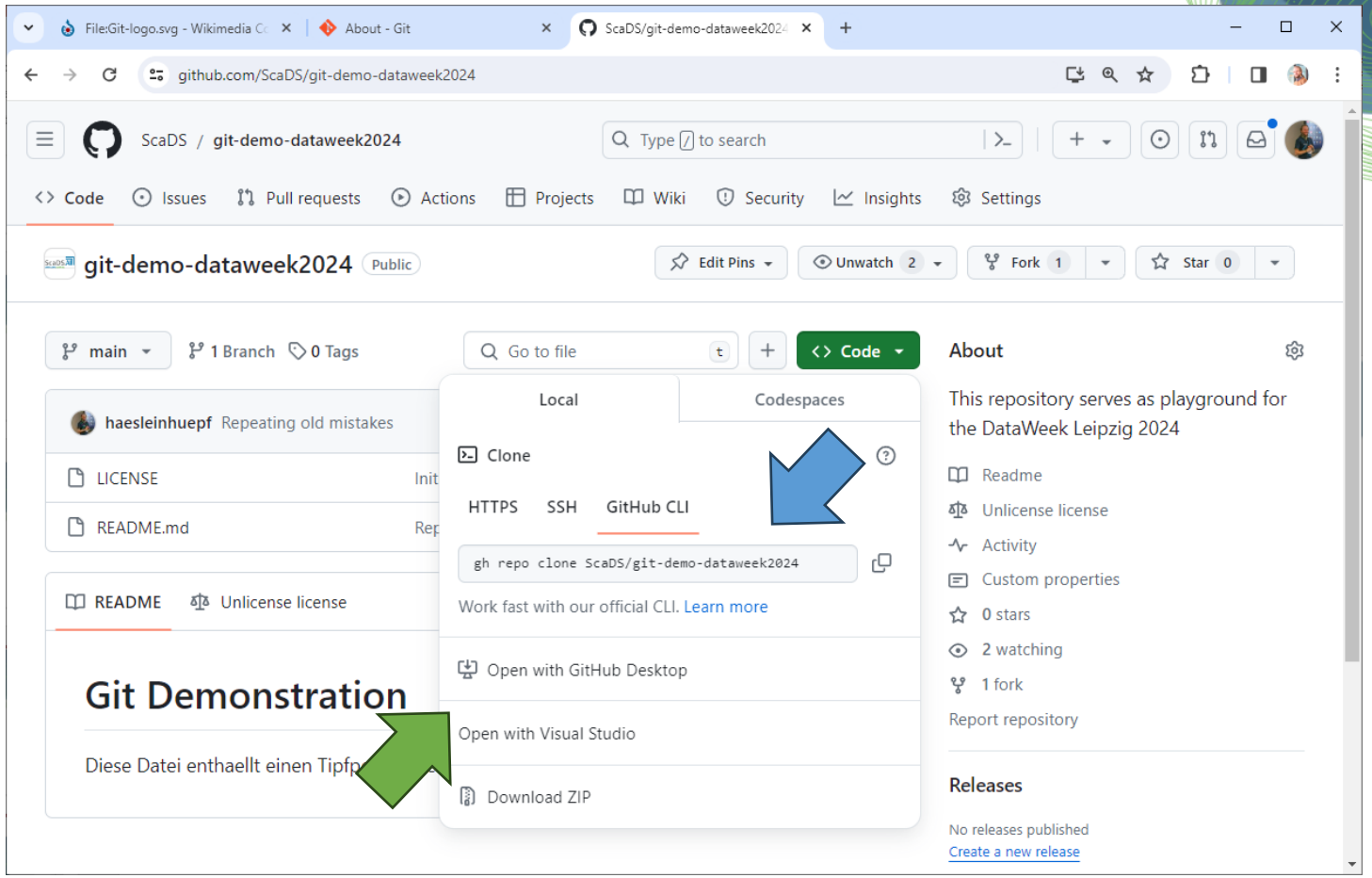
https://www.go-fair.org/fair-principles/

git: standardized communication protocol

- Git nutzt “Branches” um paralleles Arbeiten zu erlauben

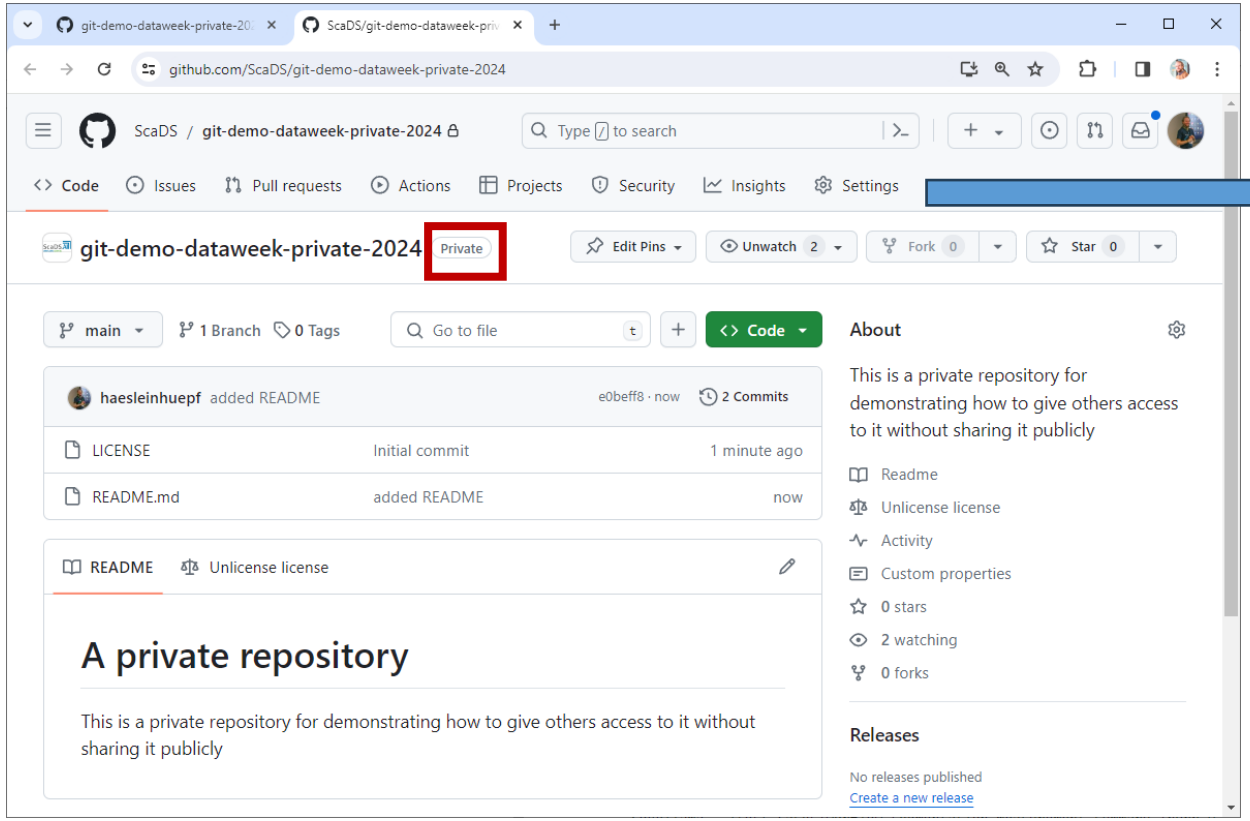
Accessibility

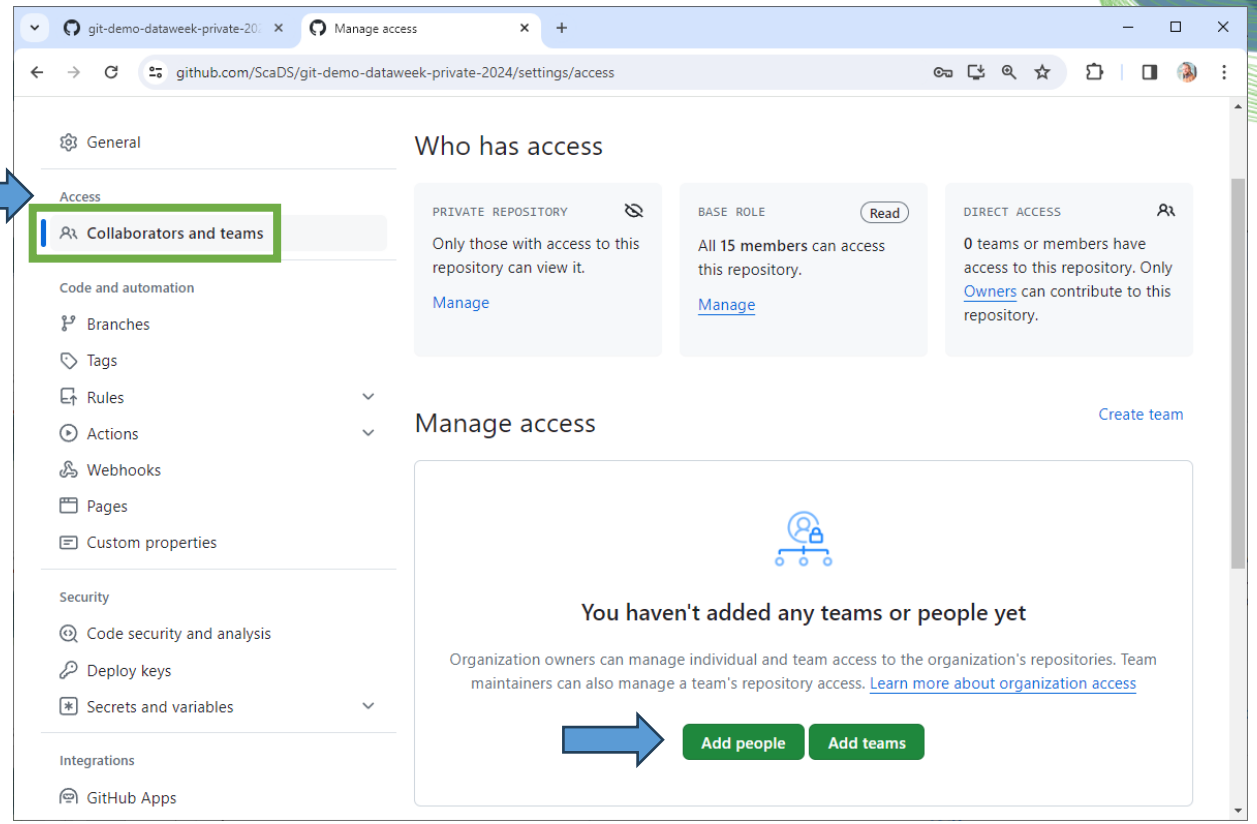
- Möglichkeit fuer Menschen und Computer, Daten herunterzuladen



Restricted Access

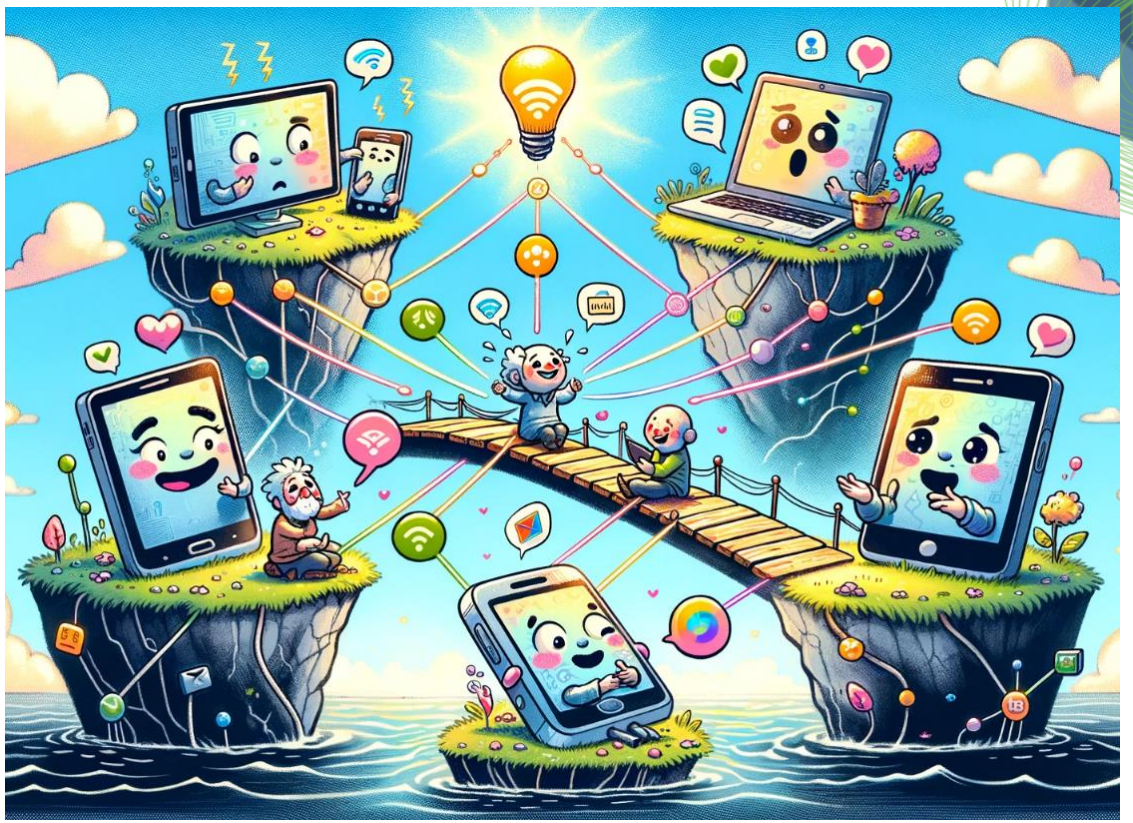
- Das A in FAIR steht nicht zwingend fuer Open-Access





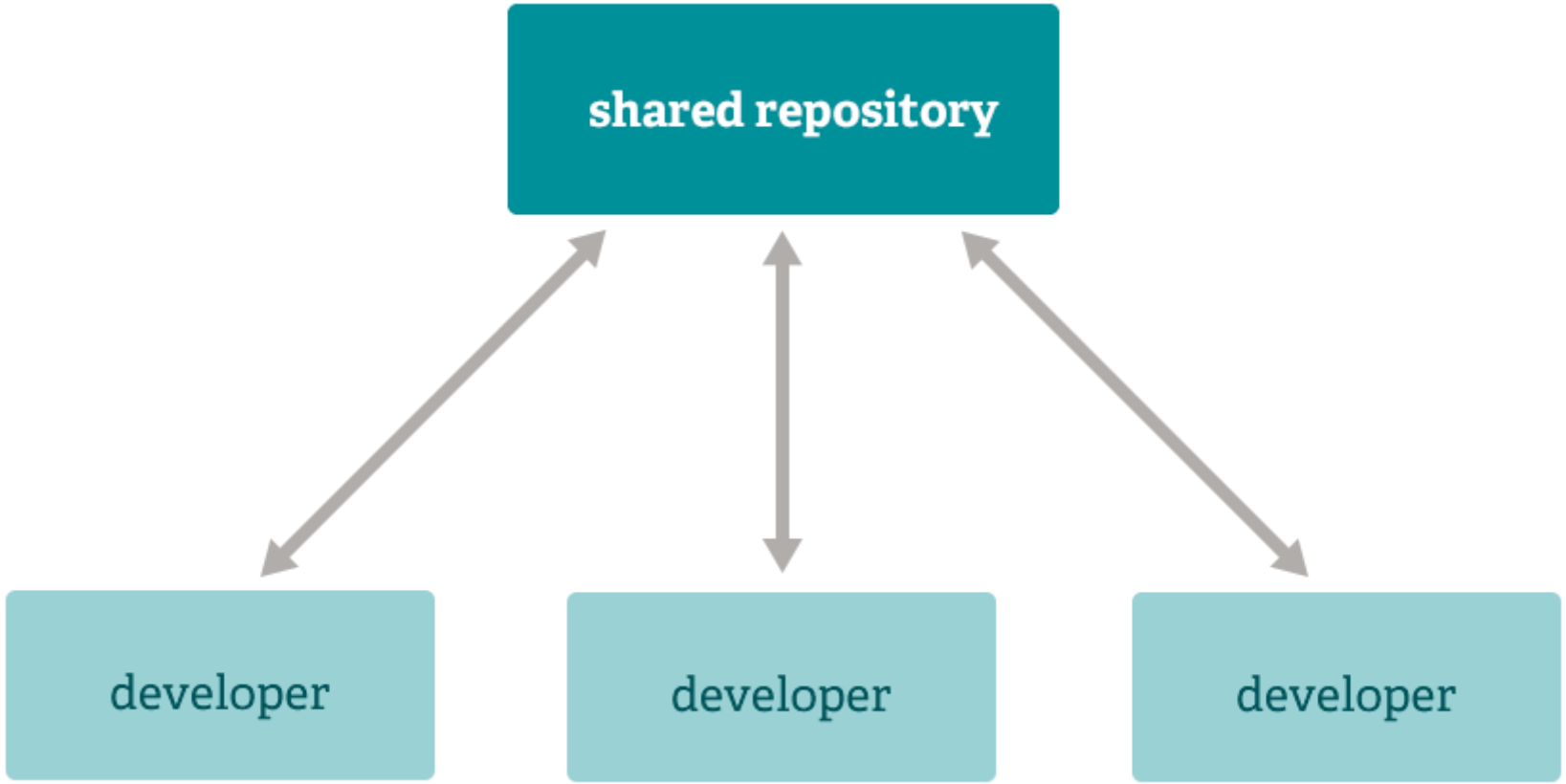
Die FAIR-Prinzipien

- Interoperable
  - I1. (Meta)daten sind formalisiert in einer zugaenglichen, gemeinsamen, breit angewandten Sprache, geeignet fuer Wissensrepraesentation
  - I2. (Meta)daten nutzen ein Vokabular, dass ebenfalls den FAIR-Prinzipien unterliegt
  - I3. (Meta)daten referenzieren andere qualifizierte (Meta)daten



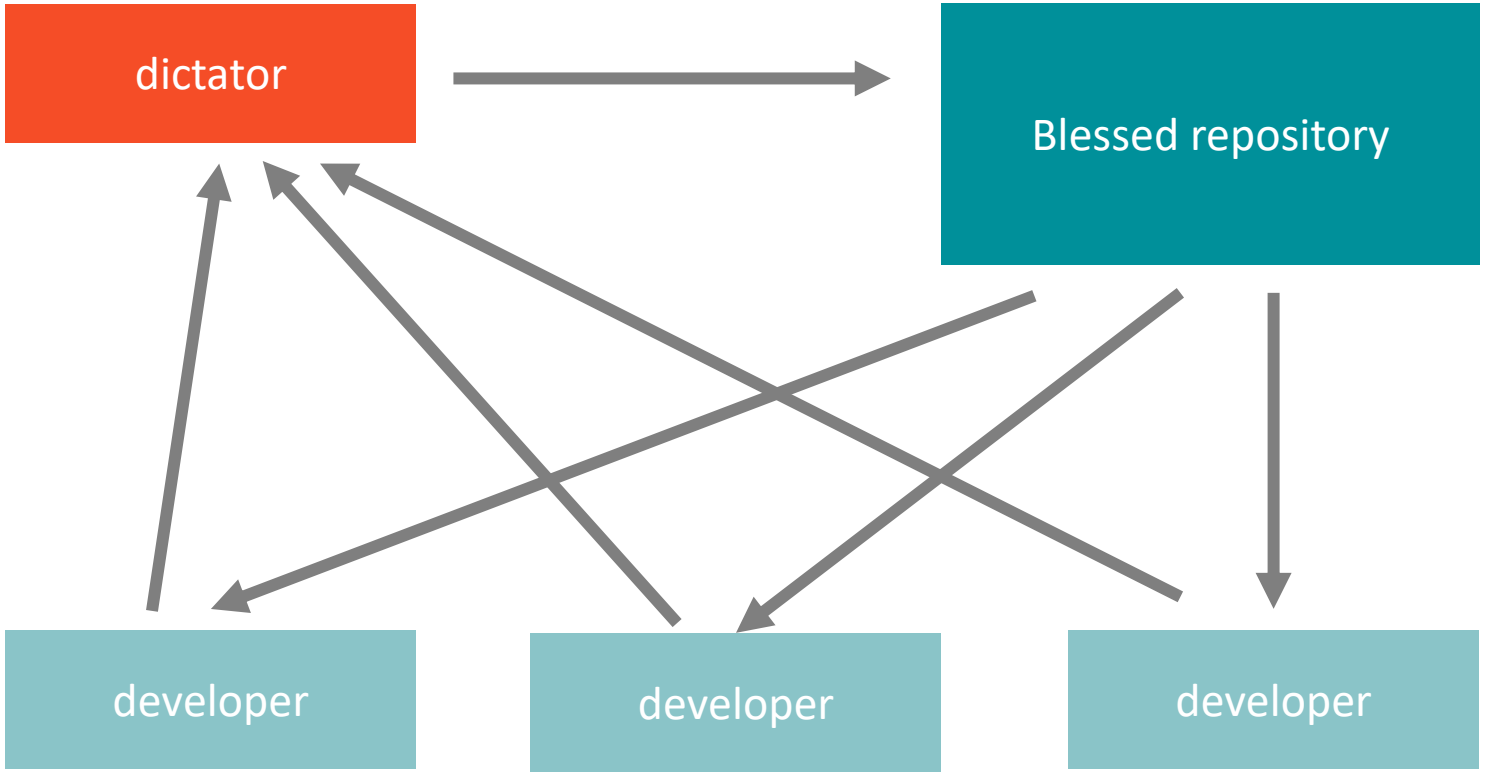
git: Workflows

- Subversion-style



git: Workflows

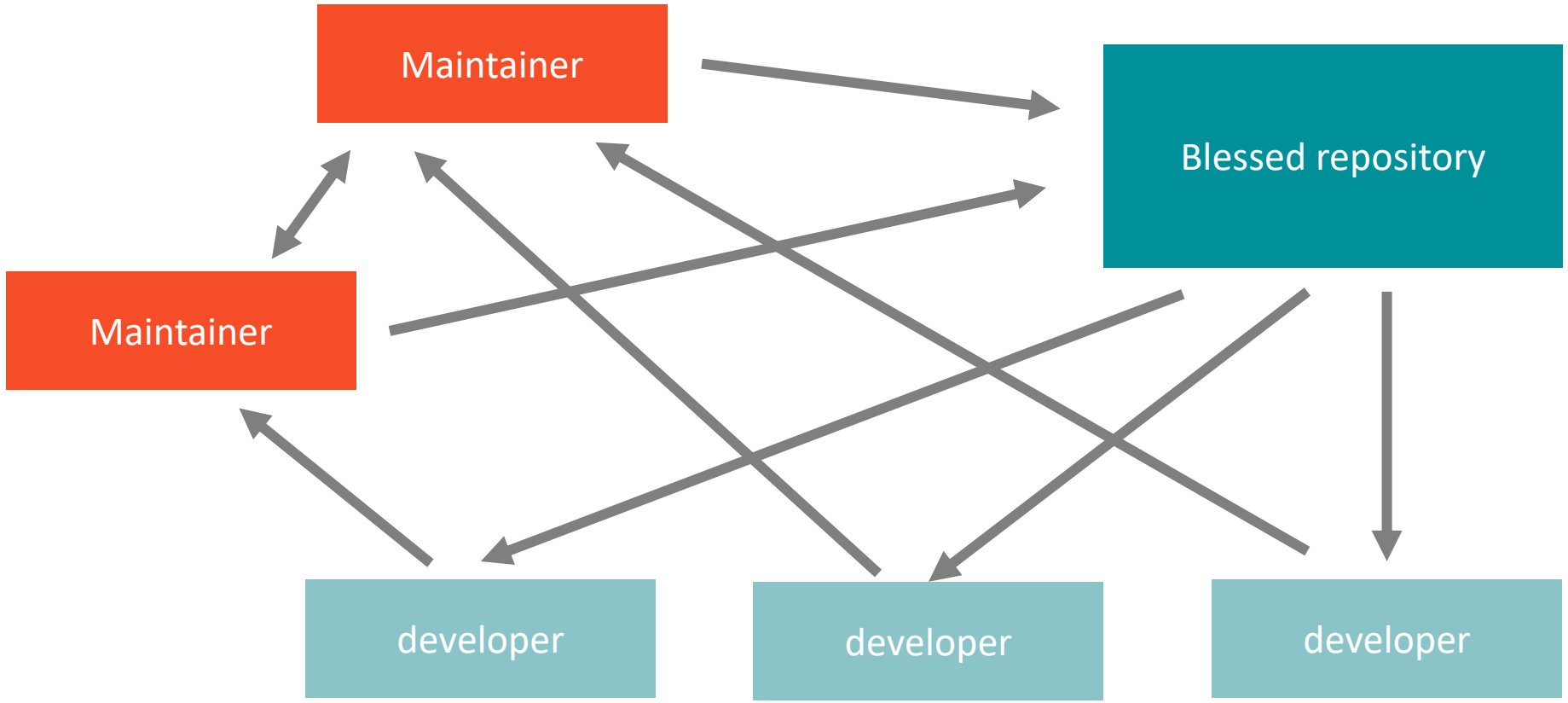
- Benevolent dictatorship workflow



Read more:
https://en.wikipedia.org/wiki/Benevolent_dictator_for_life

git: Workflows

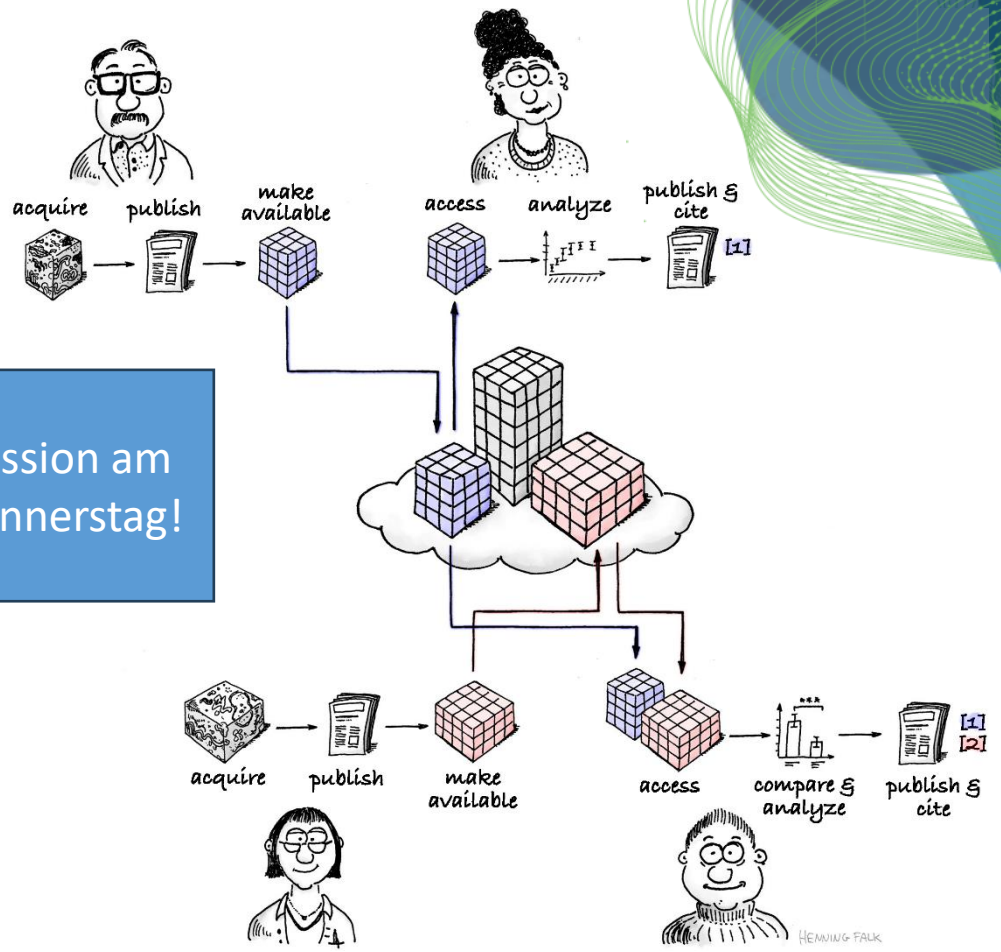
- Joint maintainers (may implement some form of democracy)



The FAIR-principles

- Reusable / Wiederverwendbar
  - R1. (Meta)daten sind reich an vielfaeltigen, akkuraten und relevanten Attributen
  - R1.1. (Meta)daten werden mit einer klaren und verfuegbaren Nutzungslizenz versehen
  - R1.2. (Meta)daten sind stets mit der detailierten Herkunft erfasst
  - R1.3. (Meta)daten folgen gemeinschaftlich definierten Standards

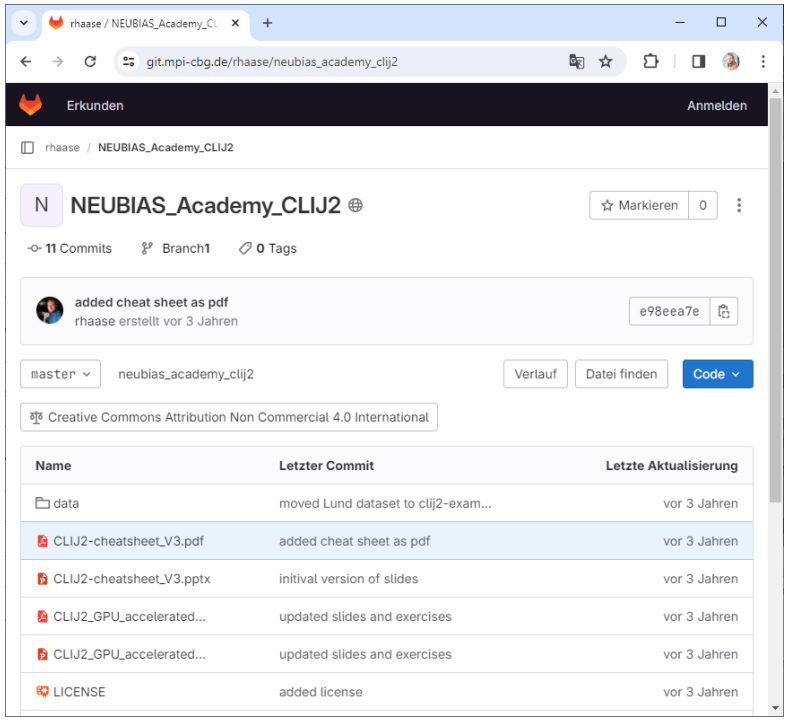
Session am Donnerstag!

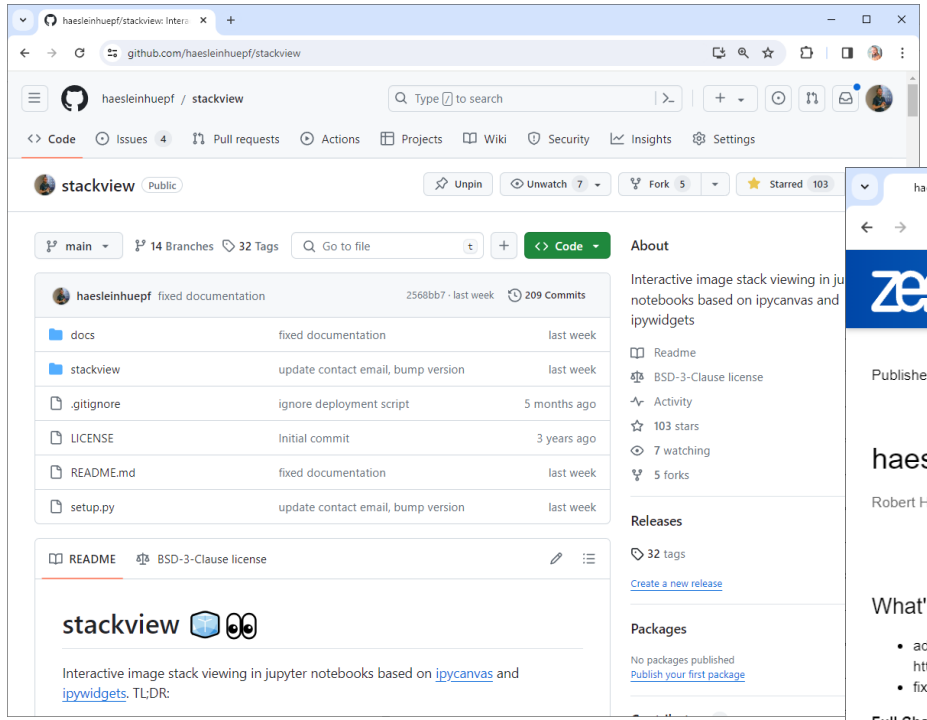


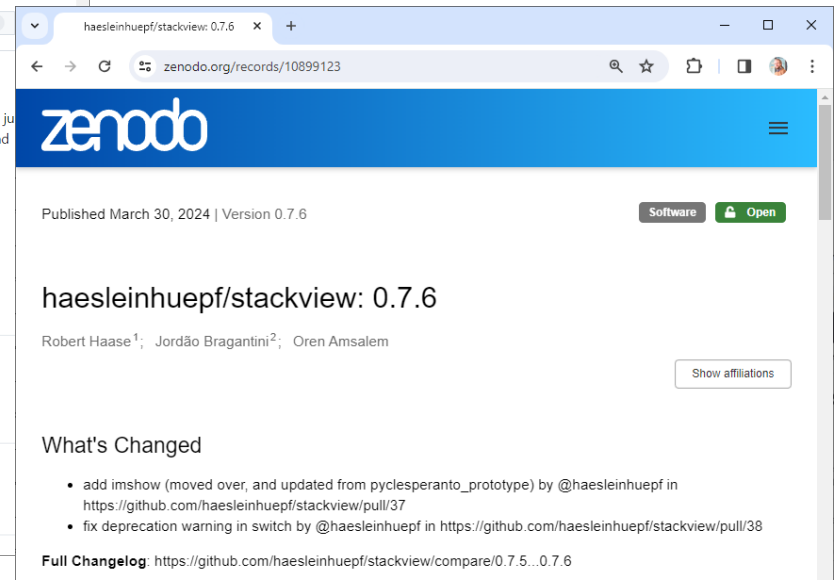
Teilen, aber wo?

- Gitlab / Github:
  Ideal fuer Kollaboratives Arbeiten,
  bei Open Source Projekten

- Zenodo:
  archiving / citing

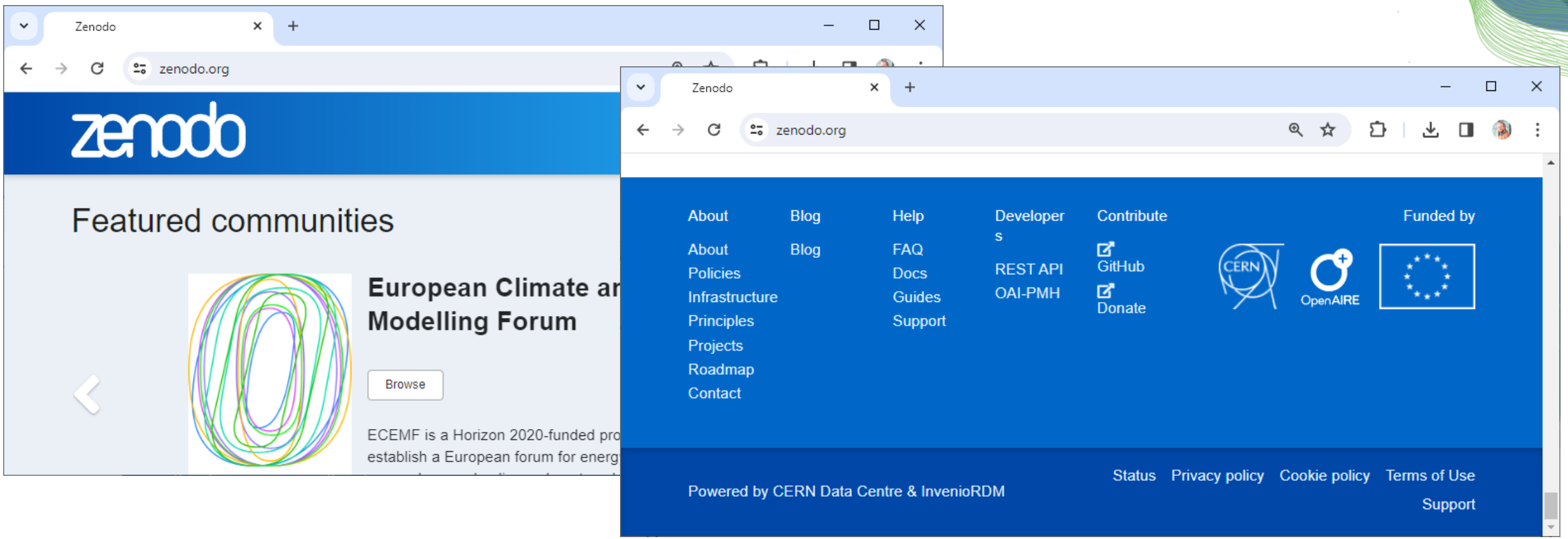






Zenodo

- öffentlich geförderte Infrastruktur @ CERN / Schweiz



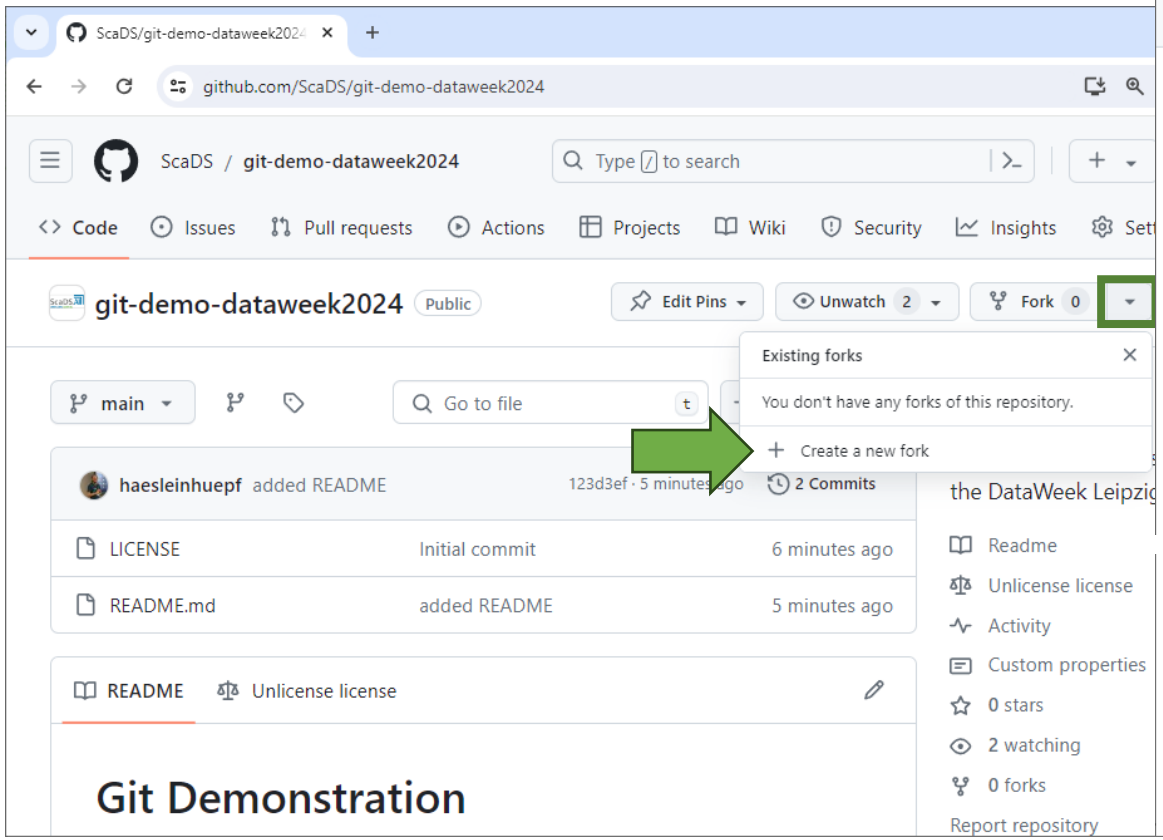
Quiz

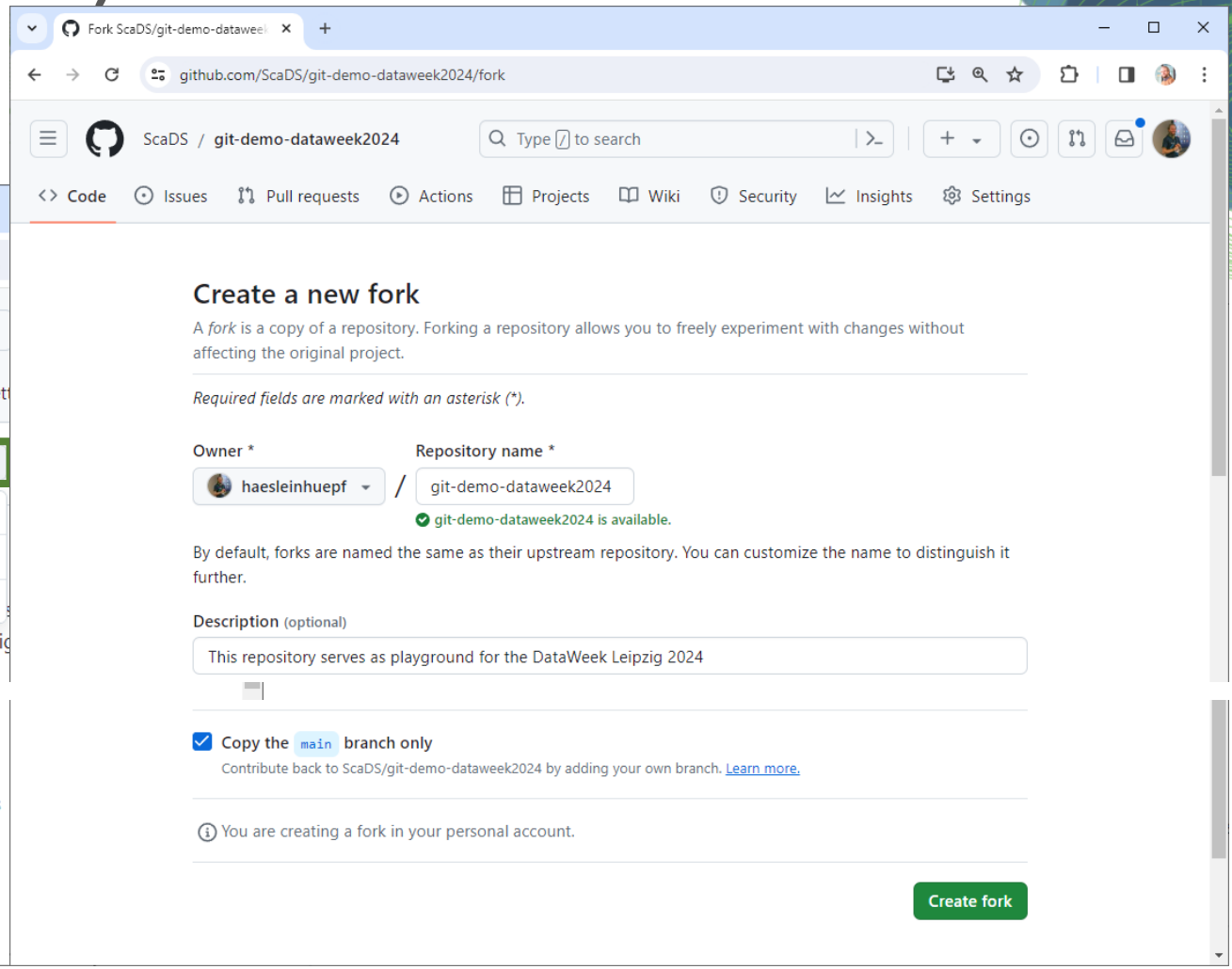
- Wo ist Open-Source Software am besten sichtbar?

Git server an der Universität	Zenodo.org	Github.com	Firmen- / Instituts-Webseite
			

Quick start: git(hub)

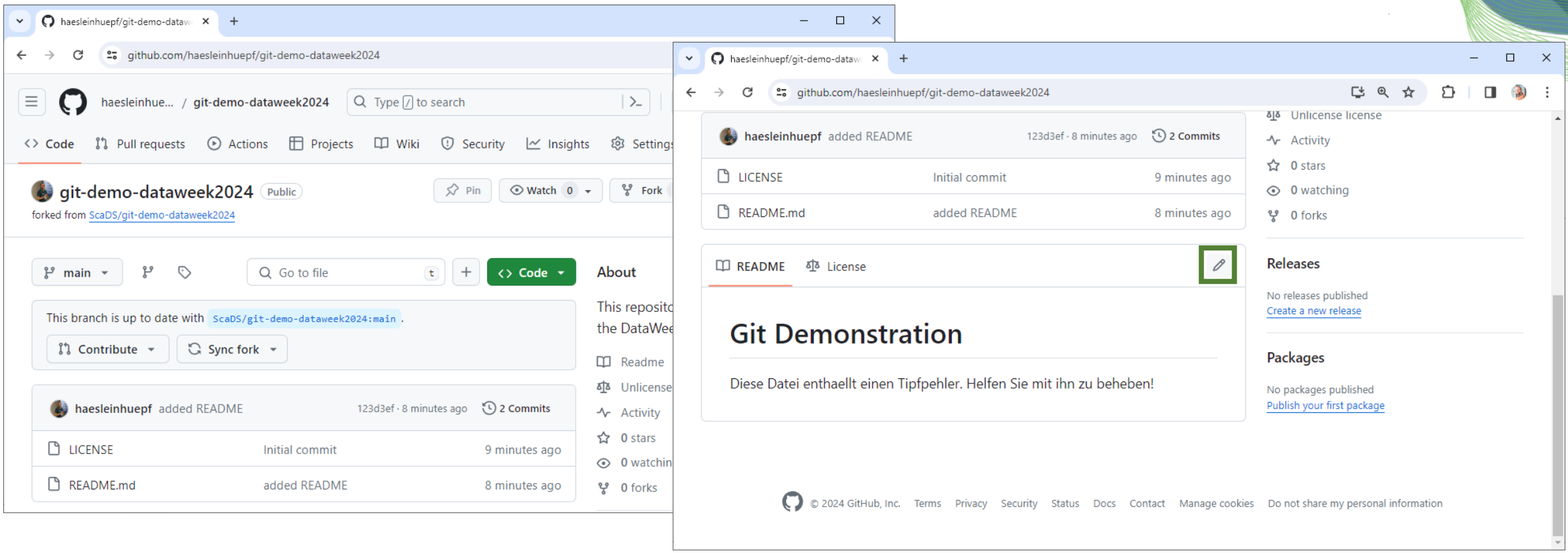
- *Forken* einer repository





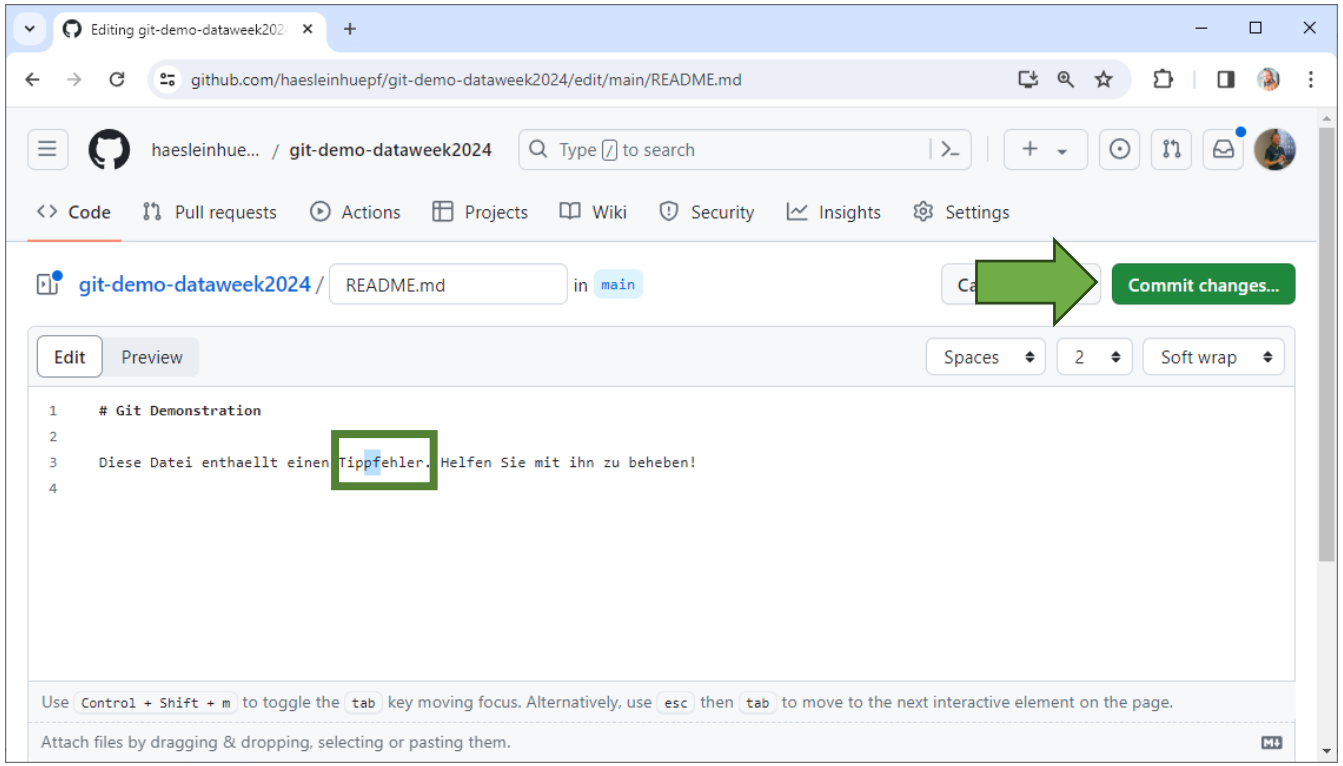
Quick start: git(hub)

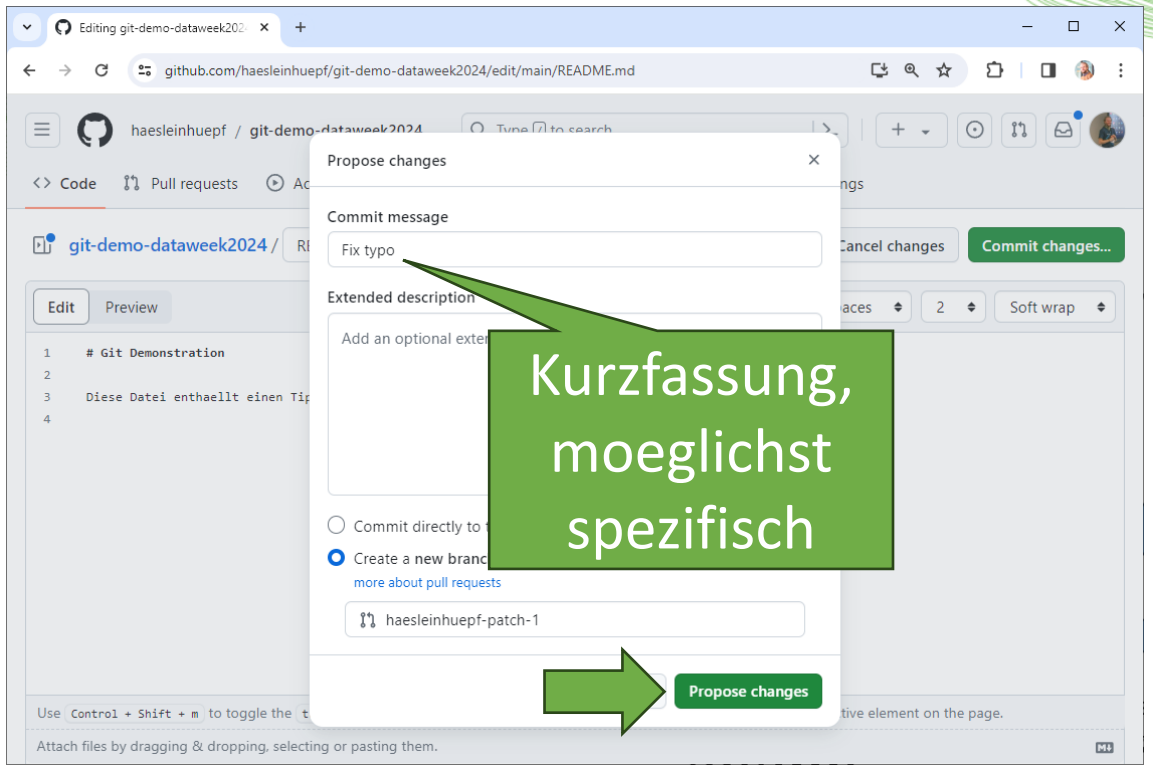
- In einen Fork haben wir Schreibrechte…



Quick start: git(hub)

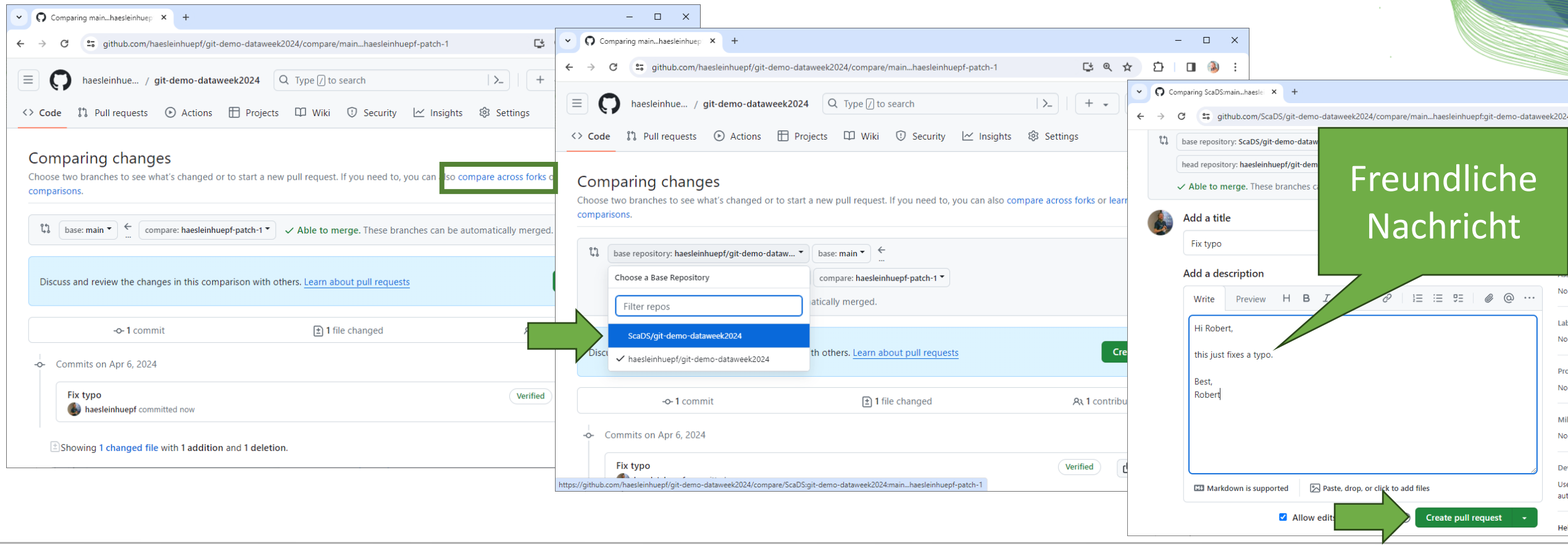
- In einen Fork haben wir Schreibrechte…





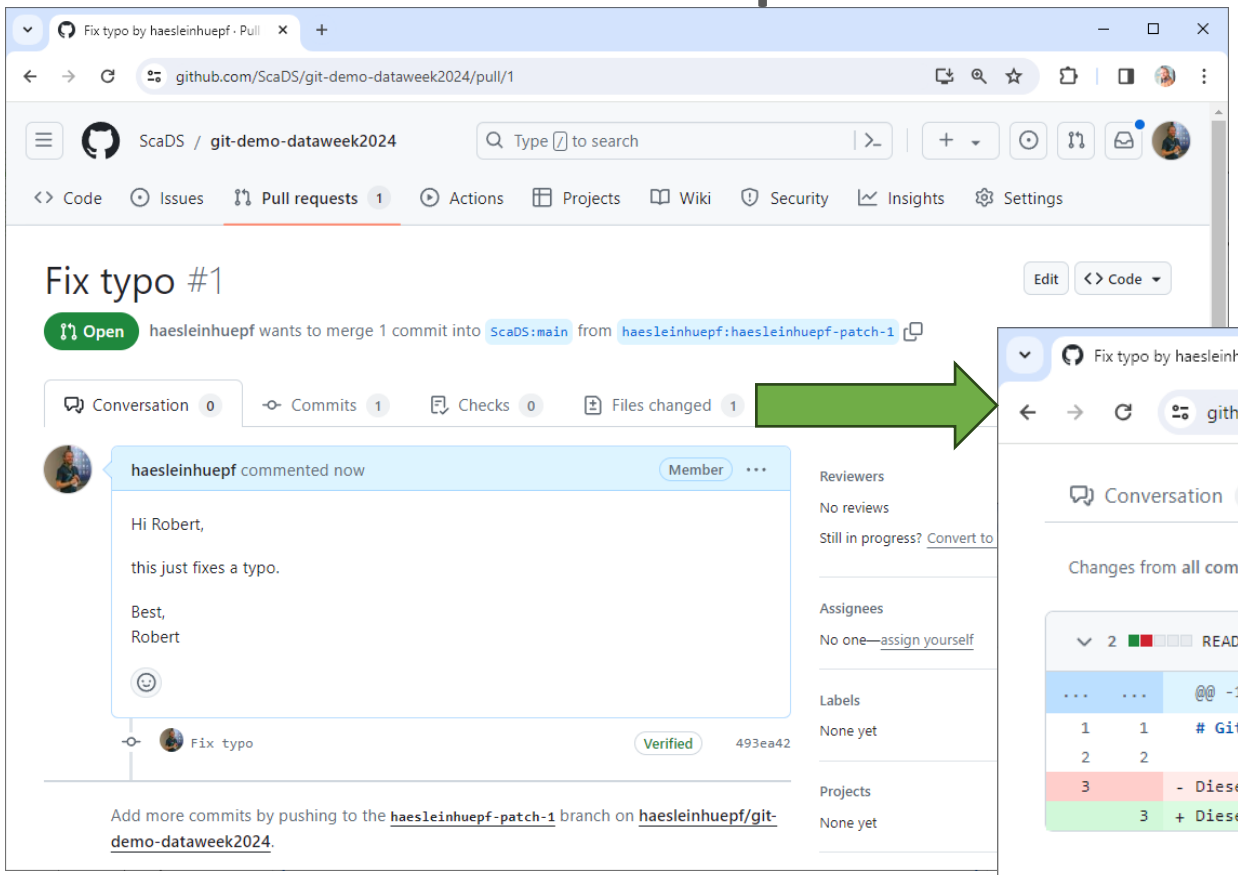
Quick start: git(hub)

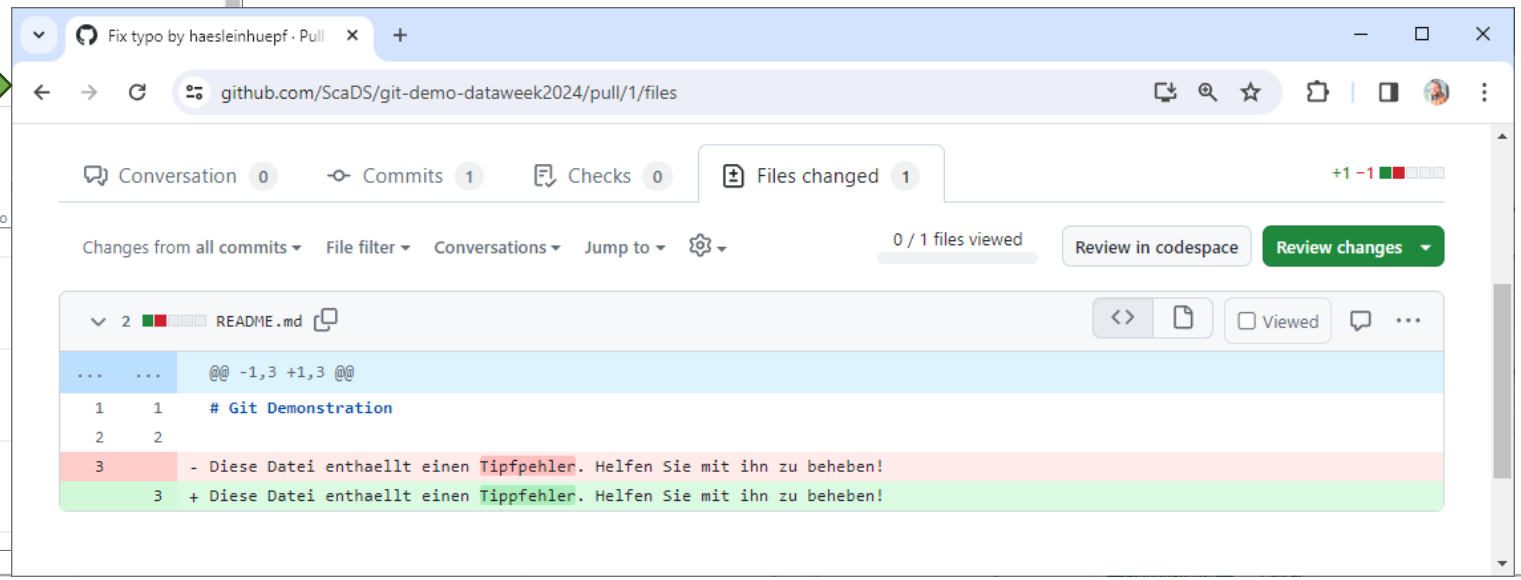
- *Pull-requests* sind *der* Mechanismus zur Zuammenarbeit



Quick start: git(hub)

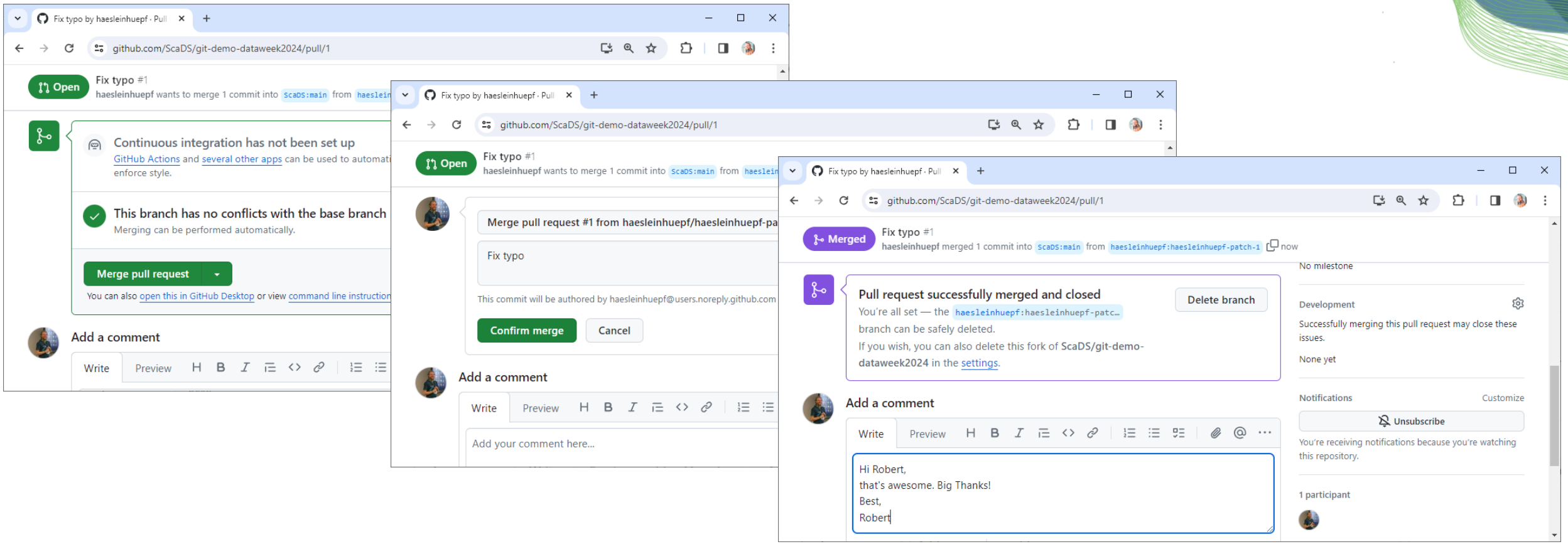
- Reviewer-Perspektive





Quick start: git(hub)

- Reviewer-Perspektive



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

NFDI 4 BIOIMAGE
NATIONAL RESEARCH DATA MANAGEMENT INFRASTRUCTURE FOR MICROSCOPY AND BIOIMAGE ANALYSIS

GloBIAS Global BioImage Analysts' Society
GLOBAL BIOIMAGE ANALYST'S SOCIETY

Praktikum

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.

Praktikum

- Helfen Sie den Tippfehler zu beheben
  - Forken Sie die Repository

    https://github.com/ScaDS/git-demo-dataweek2024

  - Editieren Sie readme.md
  - Senden Sie einen pull-request

Quiz

- Die Bitte Änderungen zu übernehmen ist ein…

Fork

Pull-request

Push-request

Commit

ScaDS.AI DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

NFDI 4 BIOIMAGE

NATIONAL RESEARCH DATA MANAGEMENT INFRASTRUCTURE FOR MICROSCOPY AND BIOIMAGE ANALYSIS

GloBIAS Global BioImage Analysts' Society

GLOBAL BIOIMAGE ANALYST'S SOCIETY

Kollaboratives Arbeiten und Versionskontrolle mit Git II

Robert Haase

GEFÖRDERT VOM

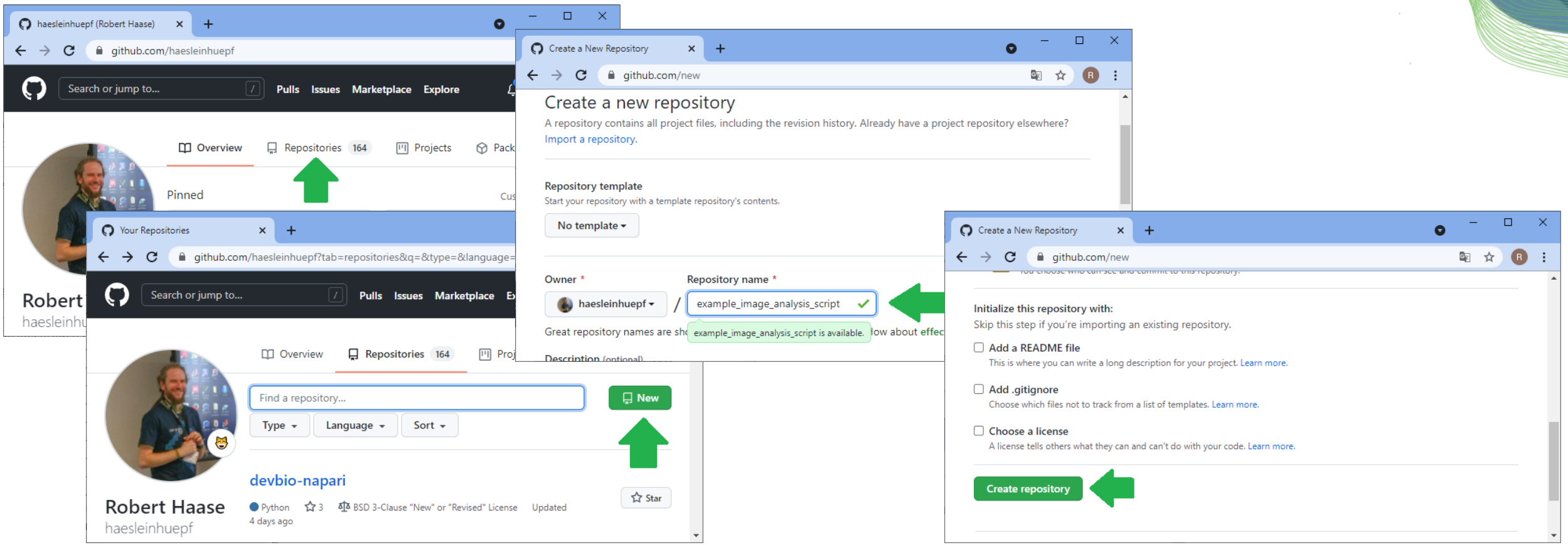
Bundesministerium für Bildung und Forschung

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.

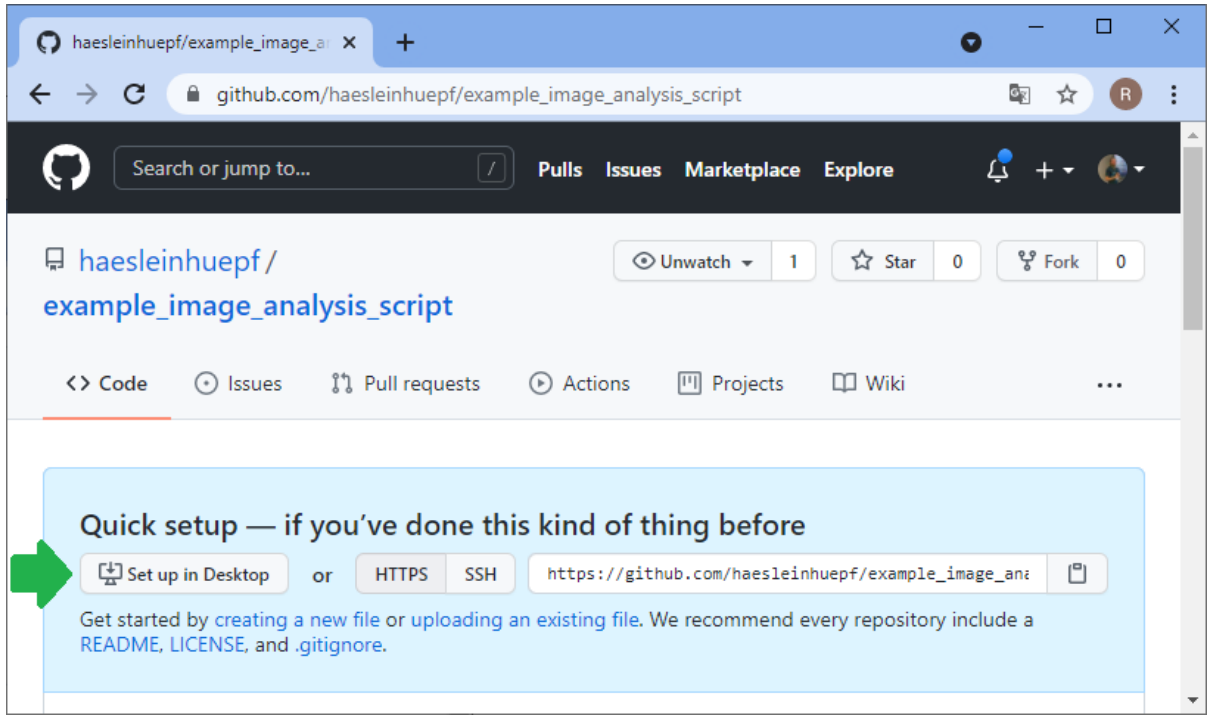
Github – Repositories anlegen

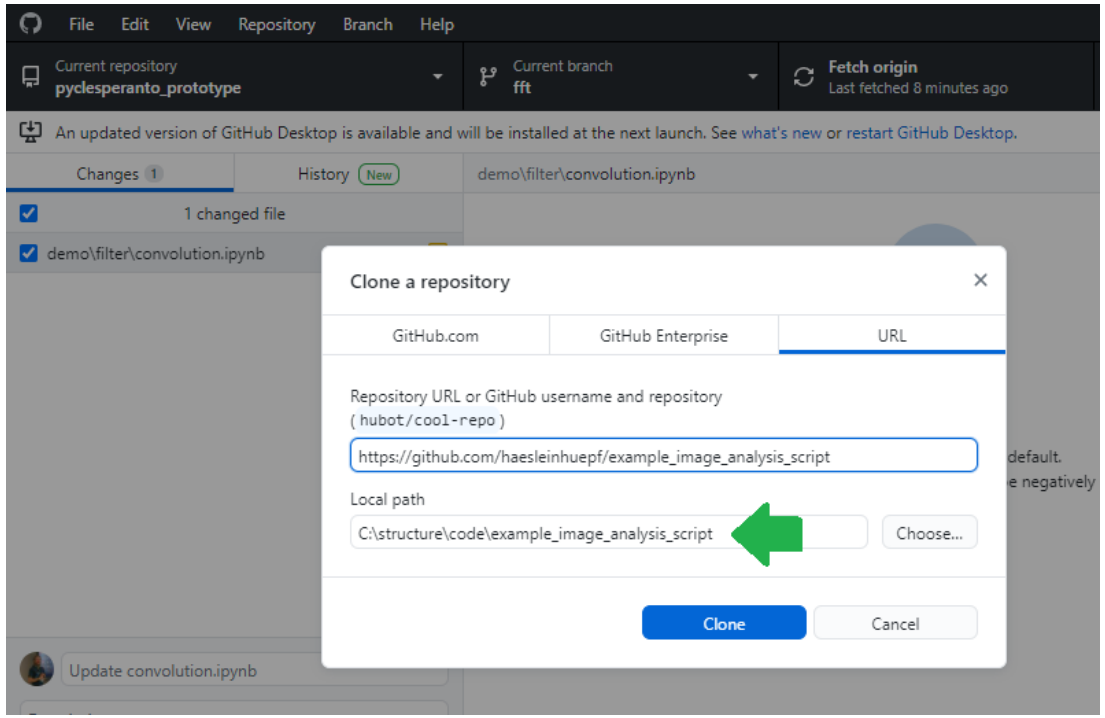
- Anlegen einer neuen Repository



Github – repositories clonen

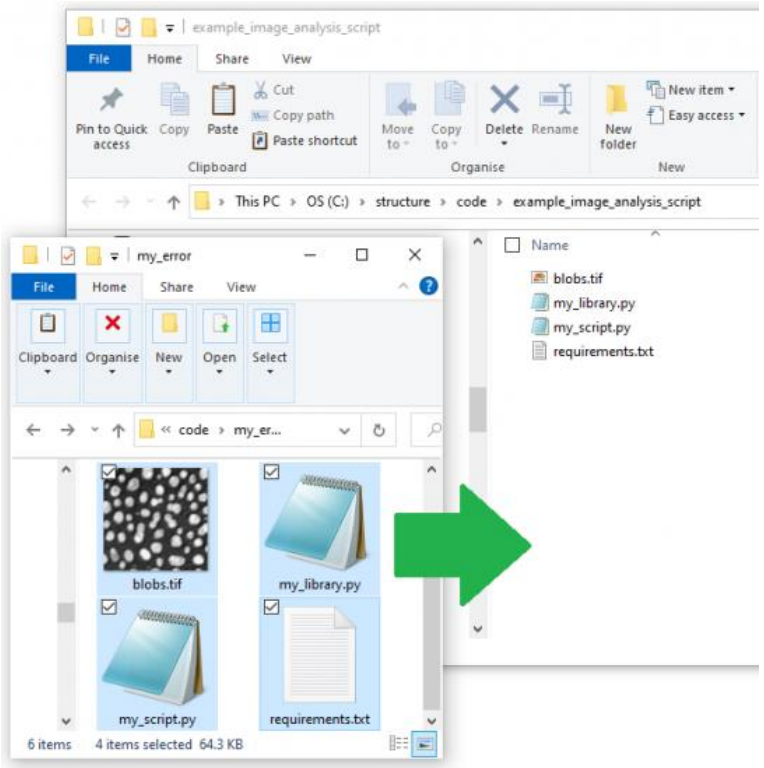
- `git clone https://github.com/organization/repository`
- Alternativ: Github Desktop

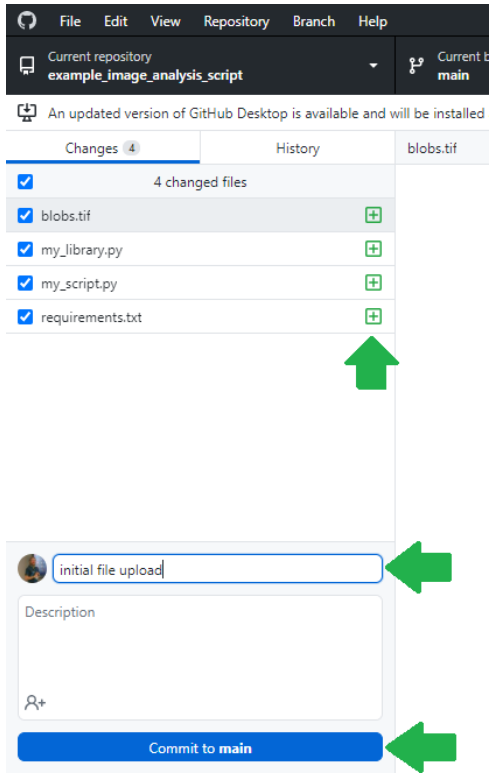


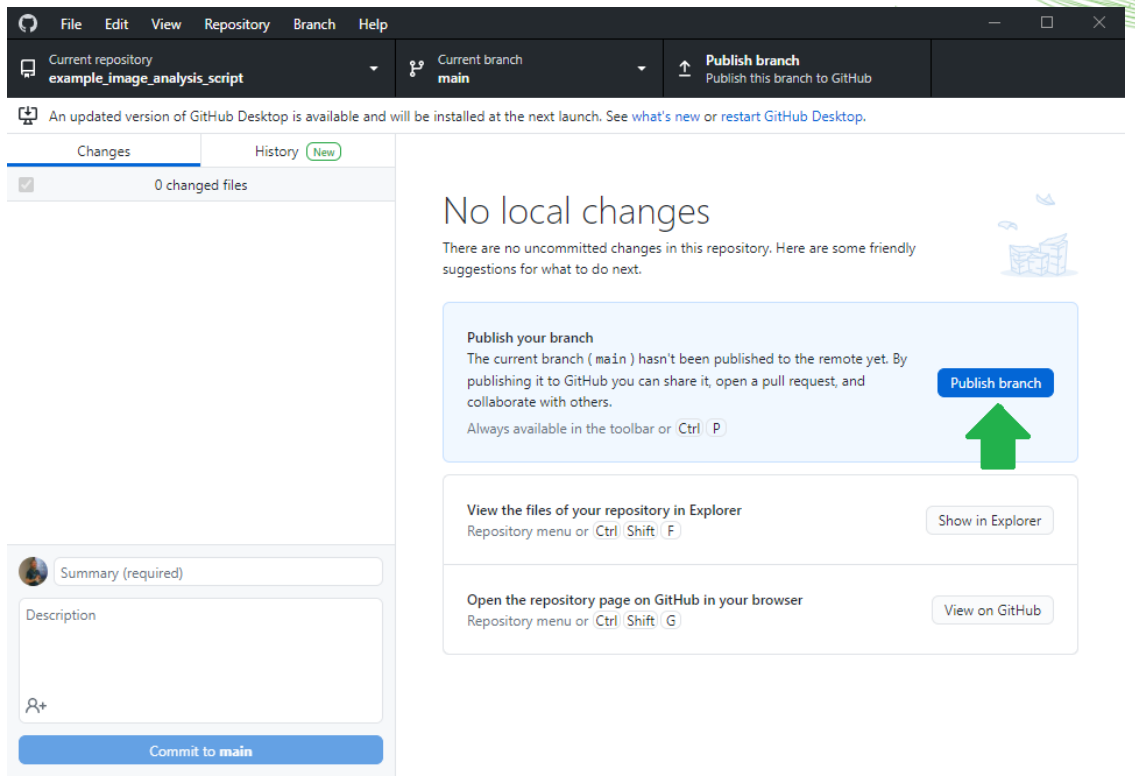


Github - Upload

- git [add], commit, push

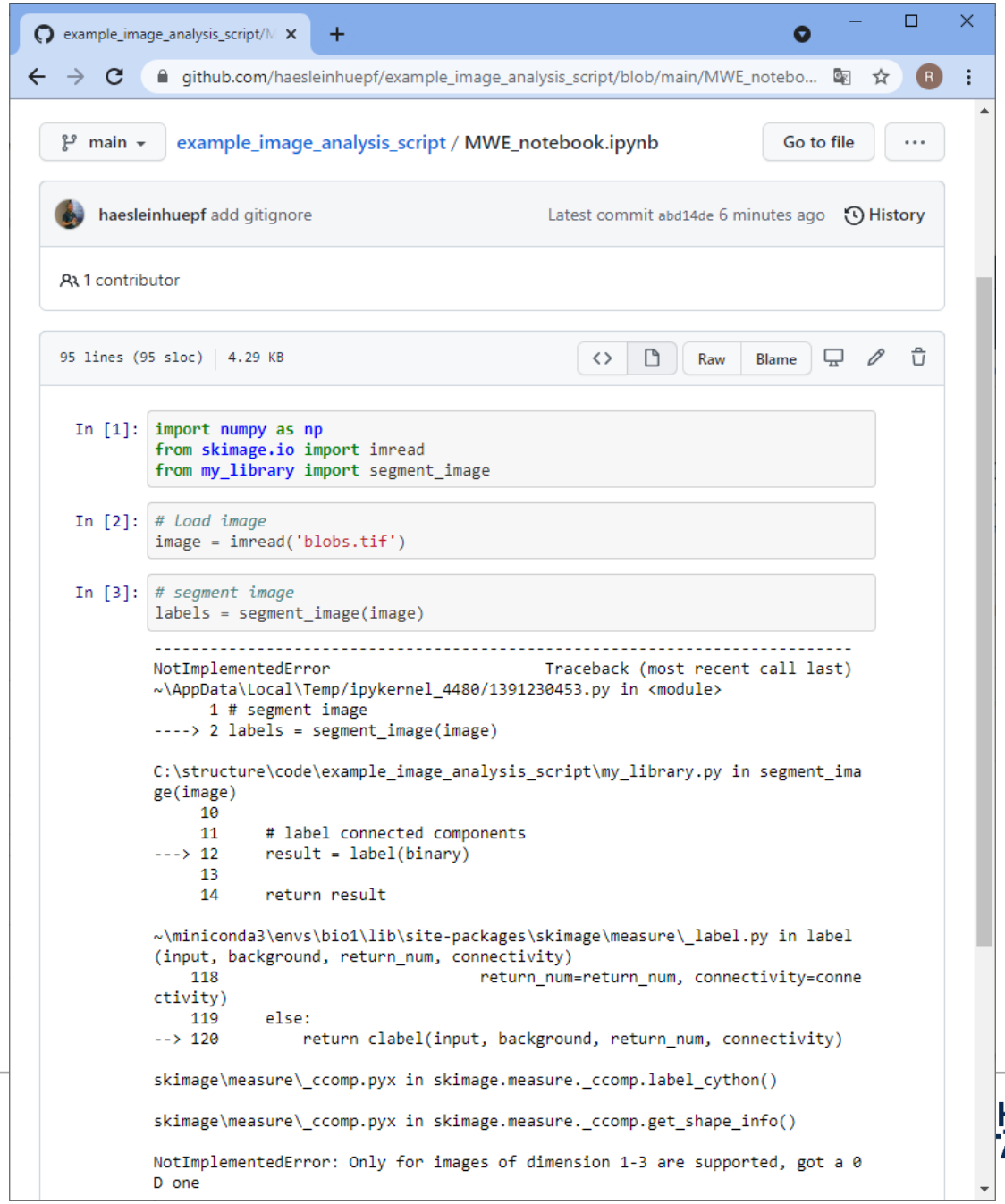






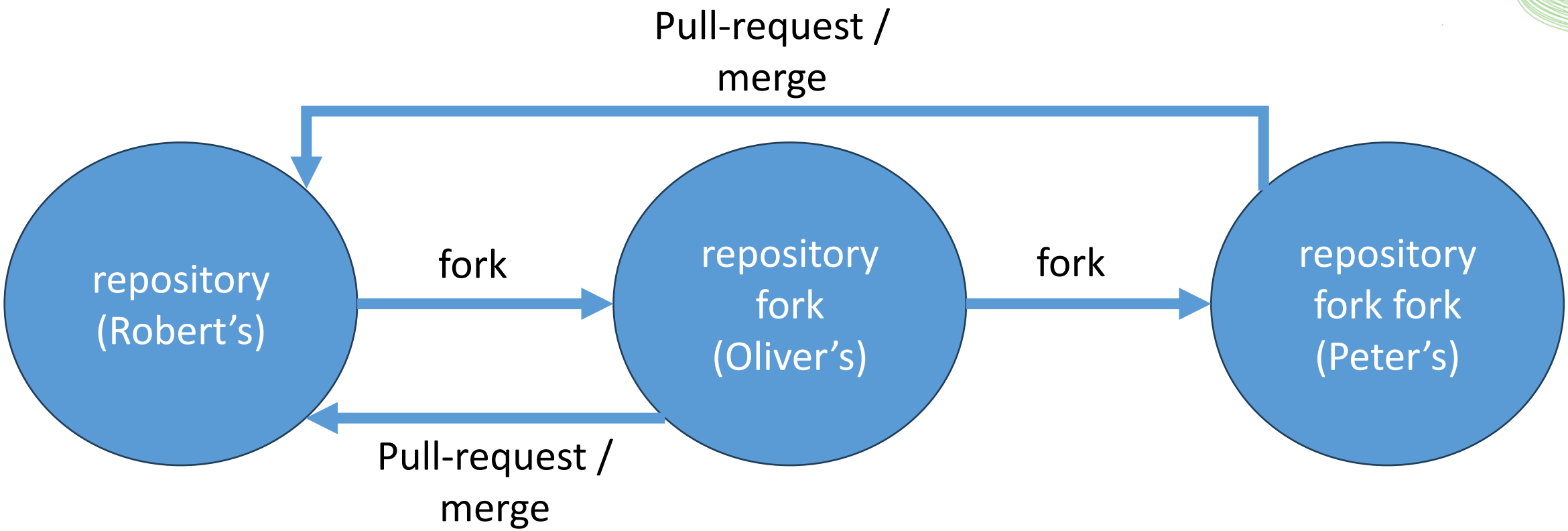
Github

- Python Notebooks online lesbar
- Fehlersuche einfach, ohne etwas herunter laden zu müssen



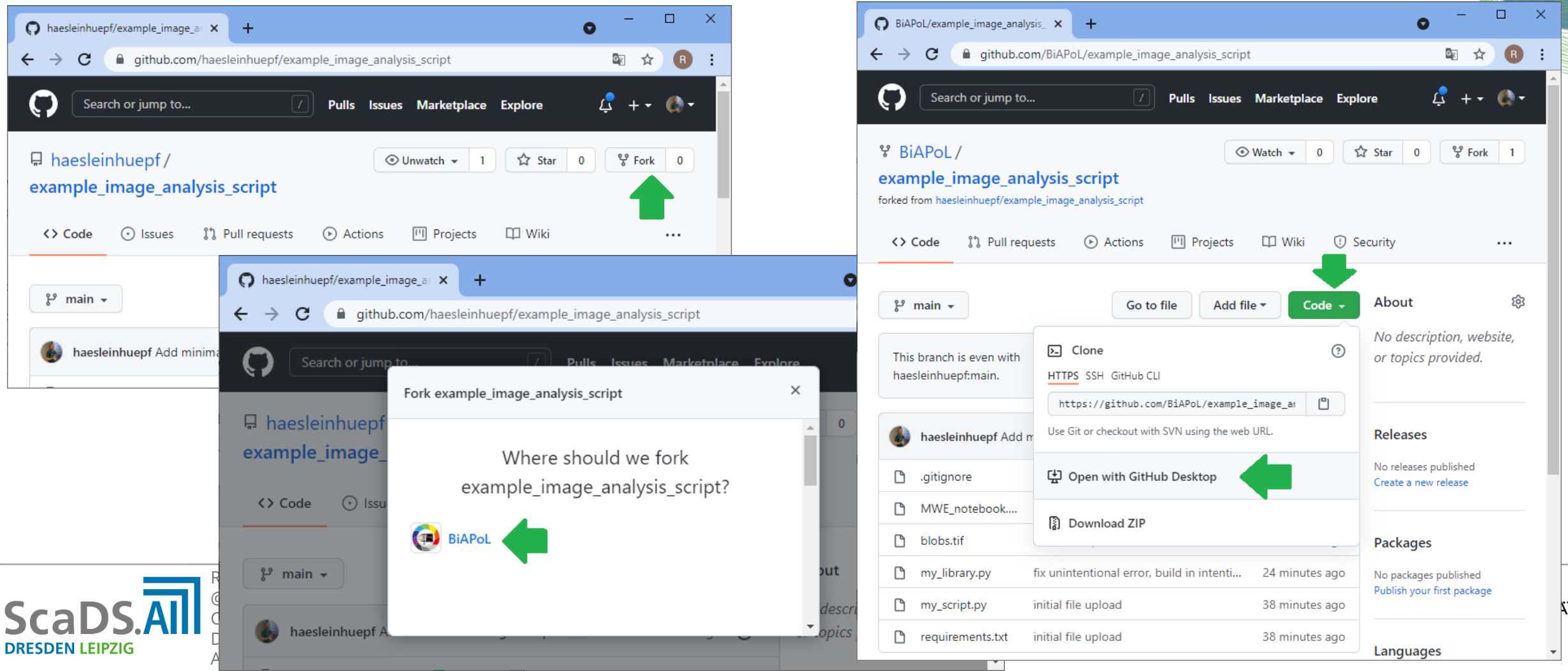
git - forking

- Ein Fork ist eine Kopie (die wir editieren können)



github - forking

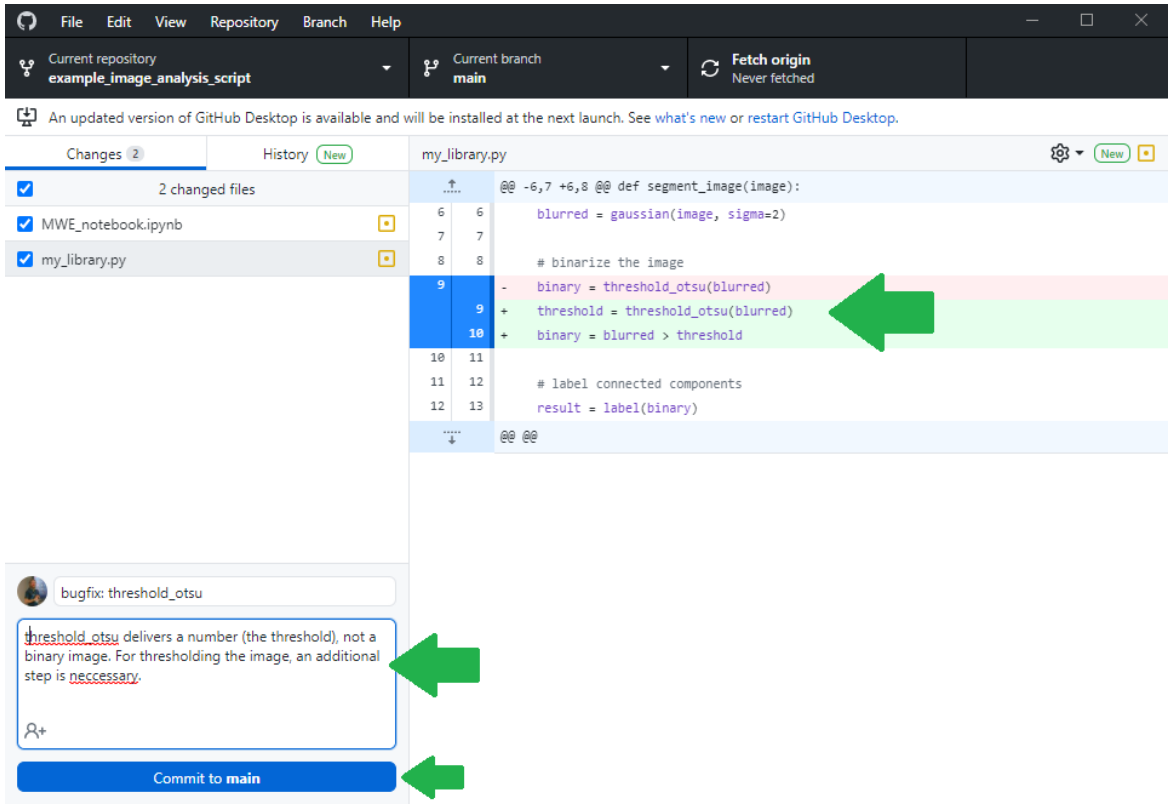
- Ein Fork ist eine Kopie (die wir editieren können)

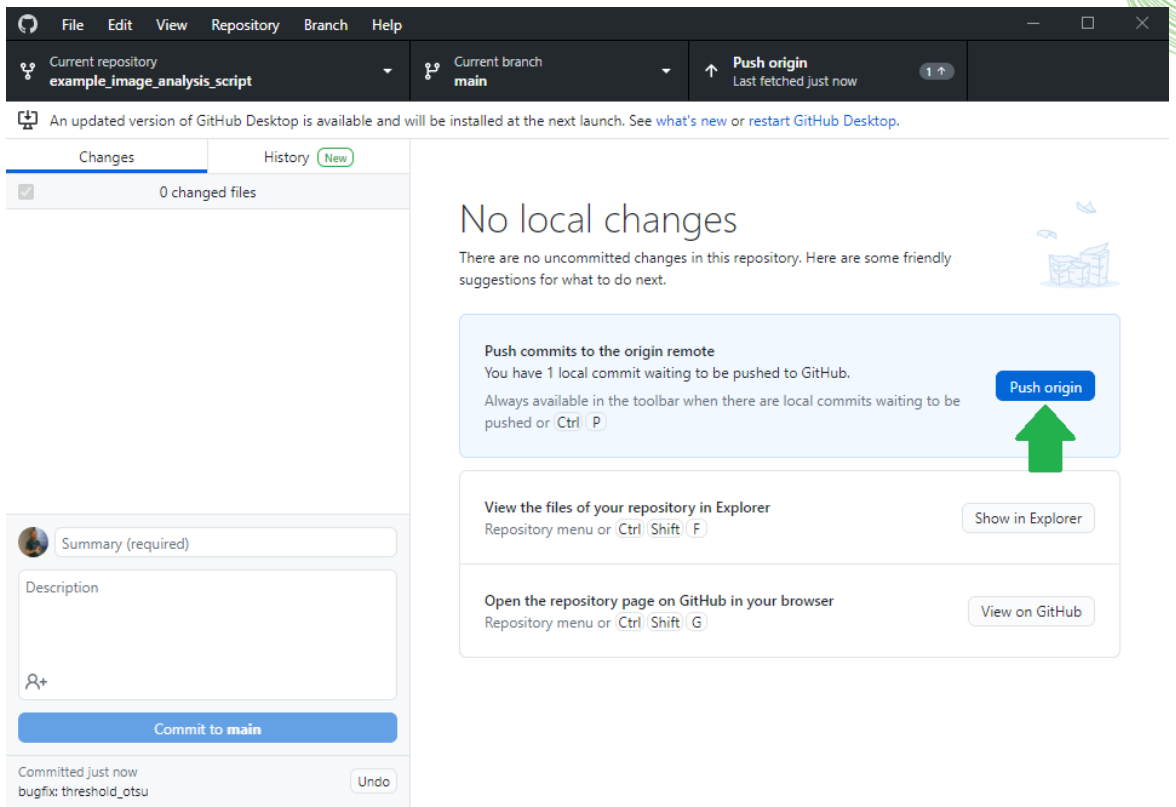


ScaDS.AI
DRESDEN LEIPZIG

github – Upload

- Unsere Änderungen werden zu unserem Fork hochgeladen

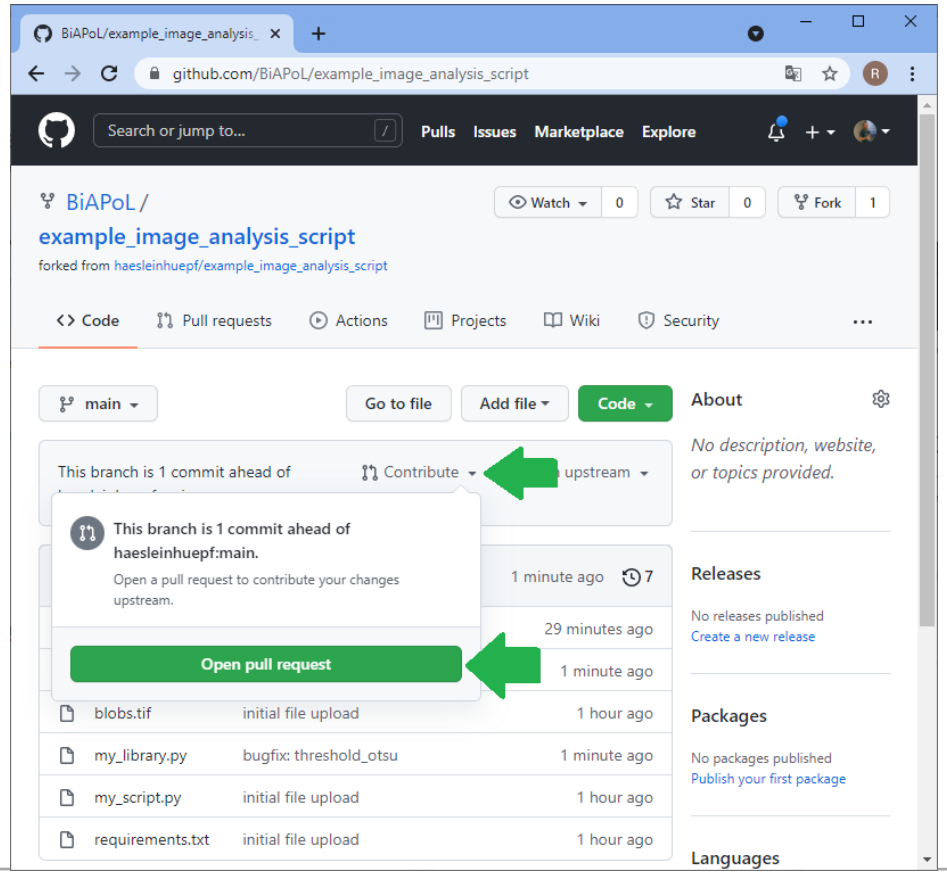


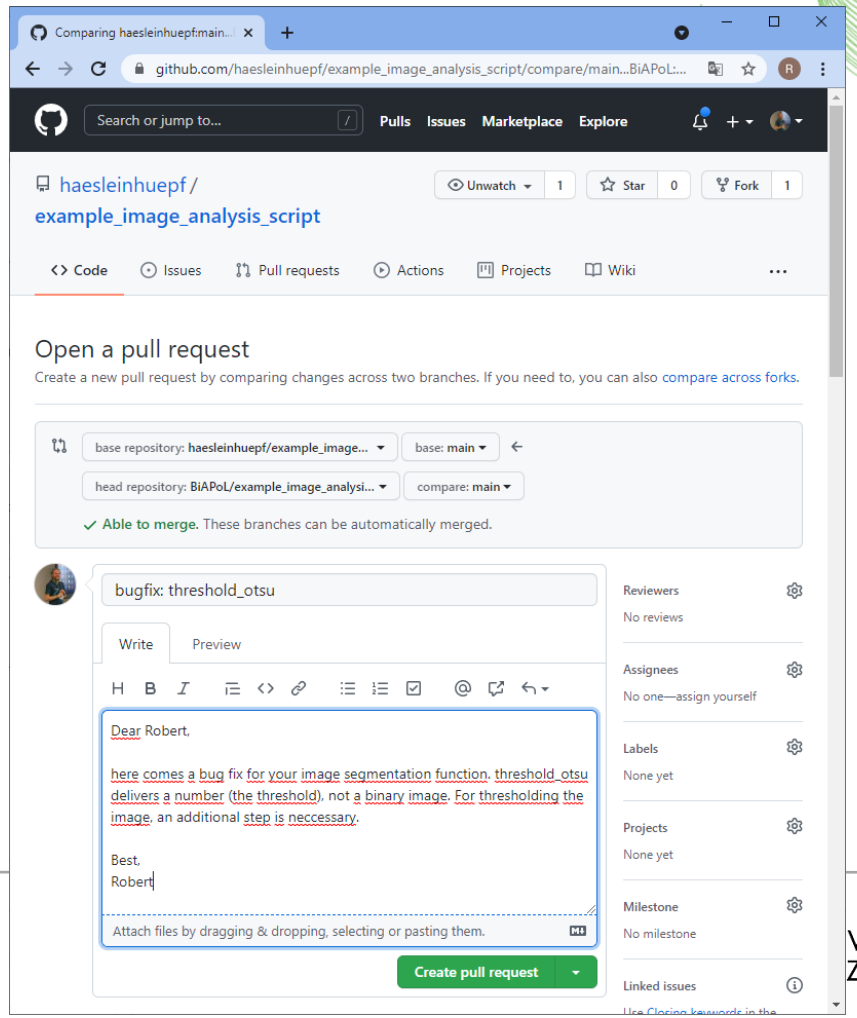


Github – Pull-requests

- Der formale Akt etwas beizutragen

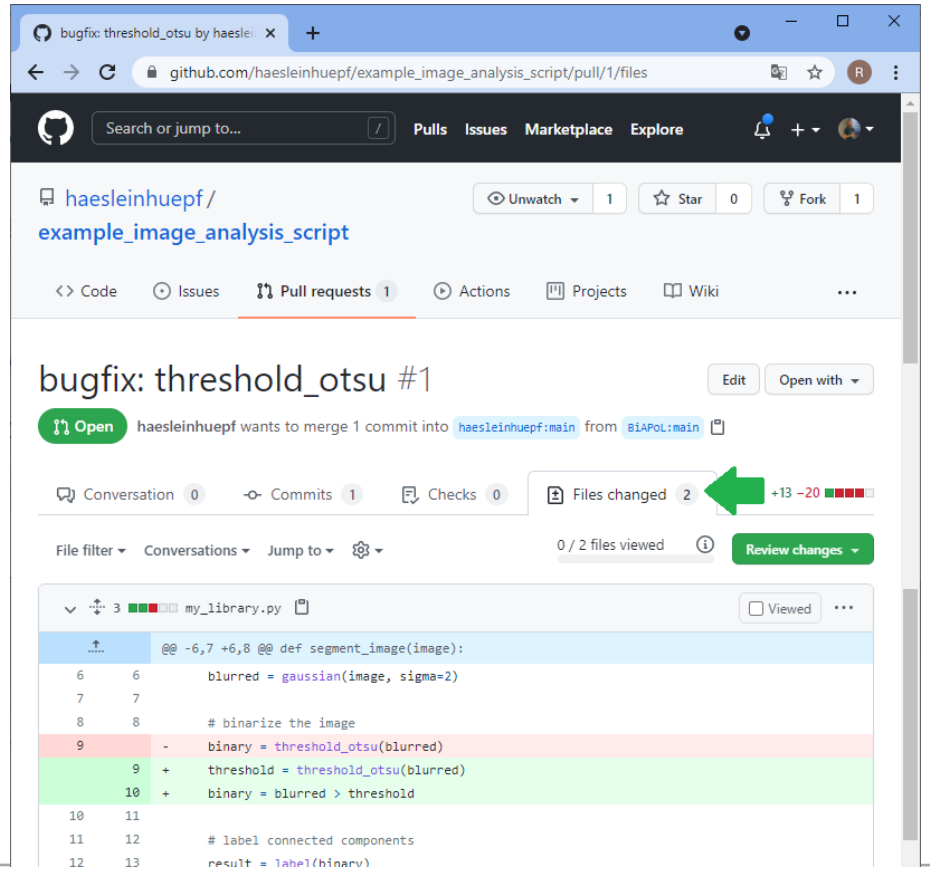
Implizit: Ich stimme Euren Nutzungbedingungen (Lizenz, Copyright) zu

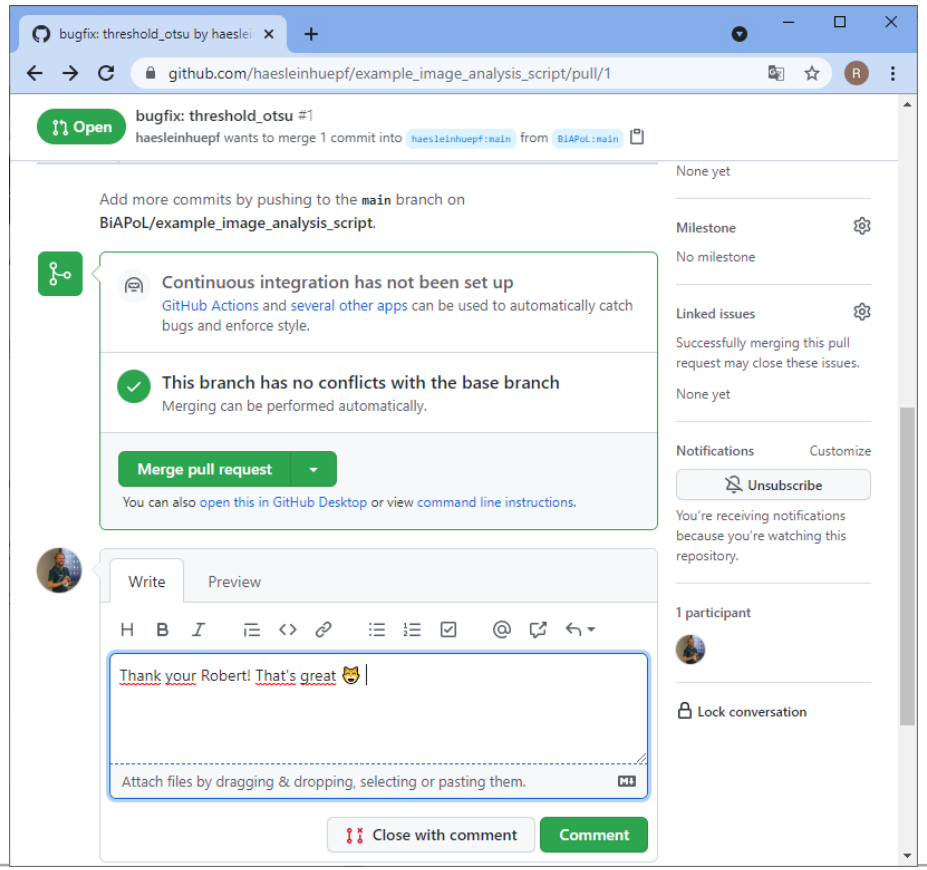




Github – Pull-requests

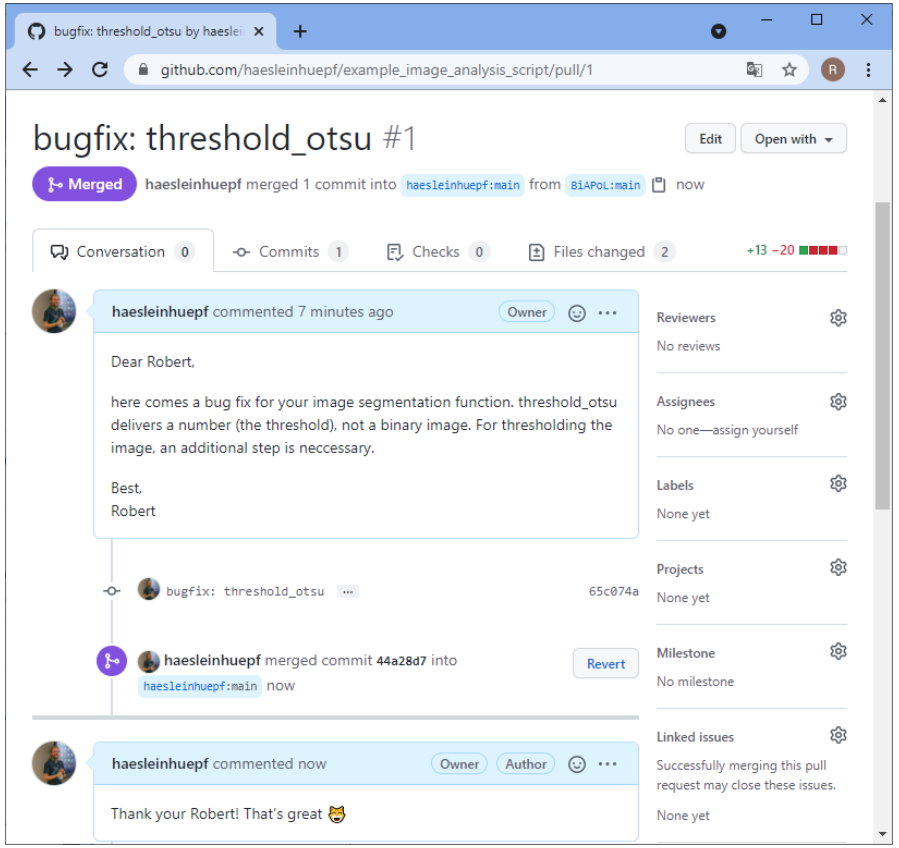
- Reviewer-Perspektive

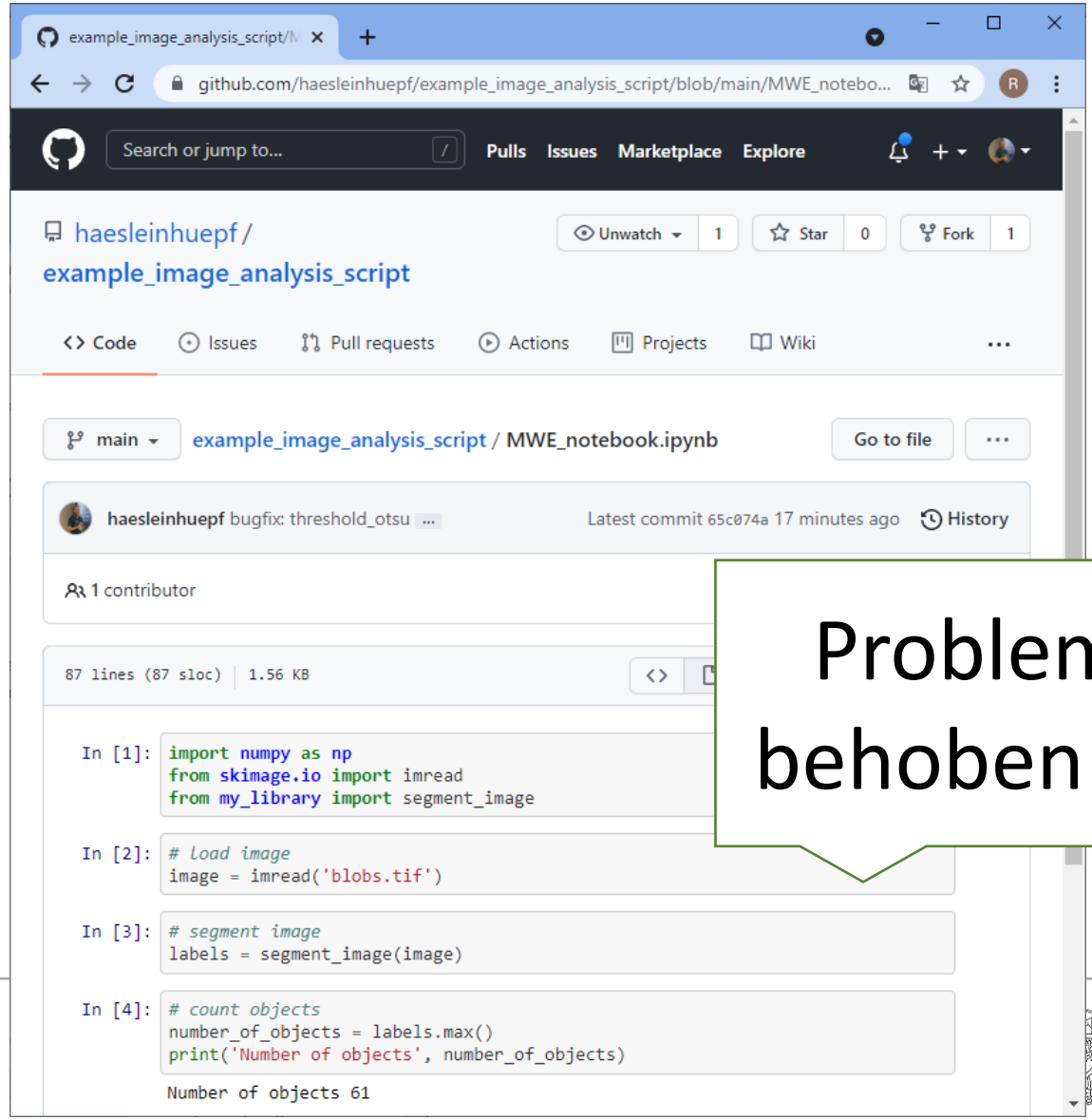




Github – pull requests

- Reviewer perspective

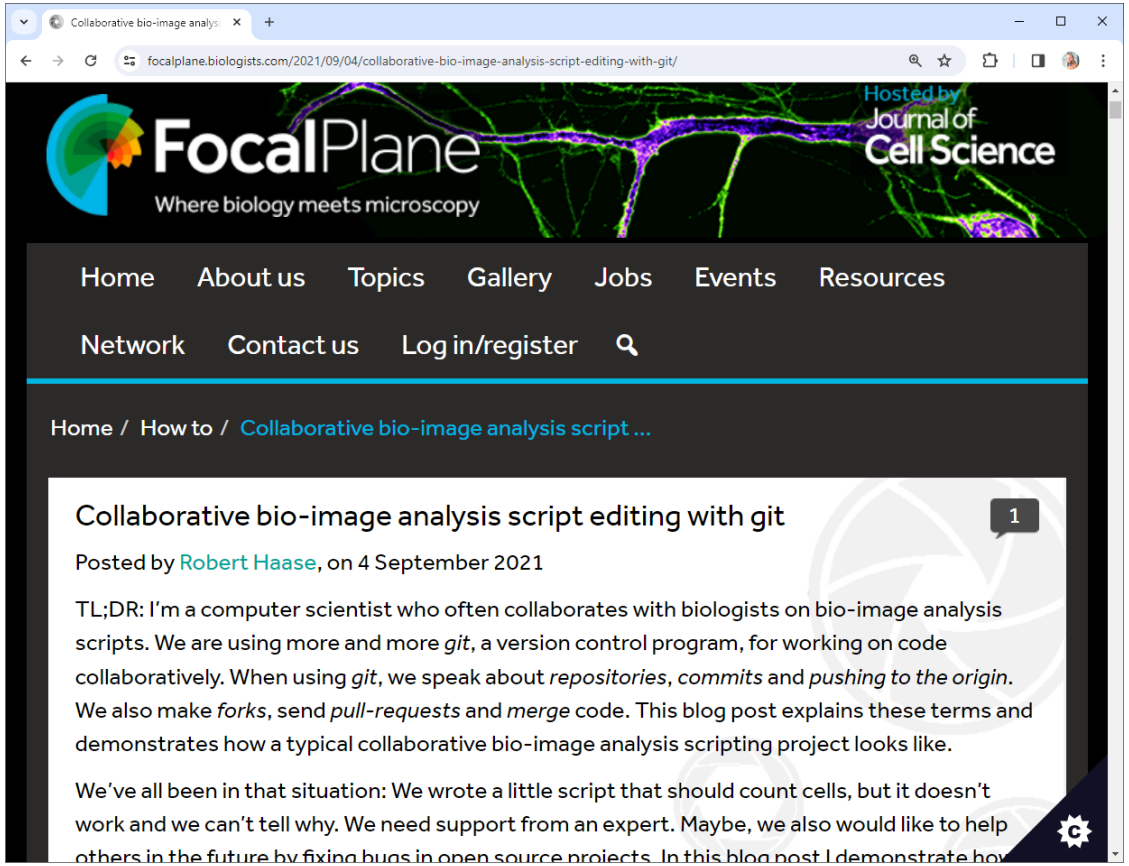


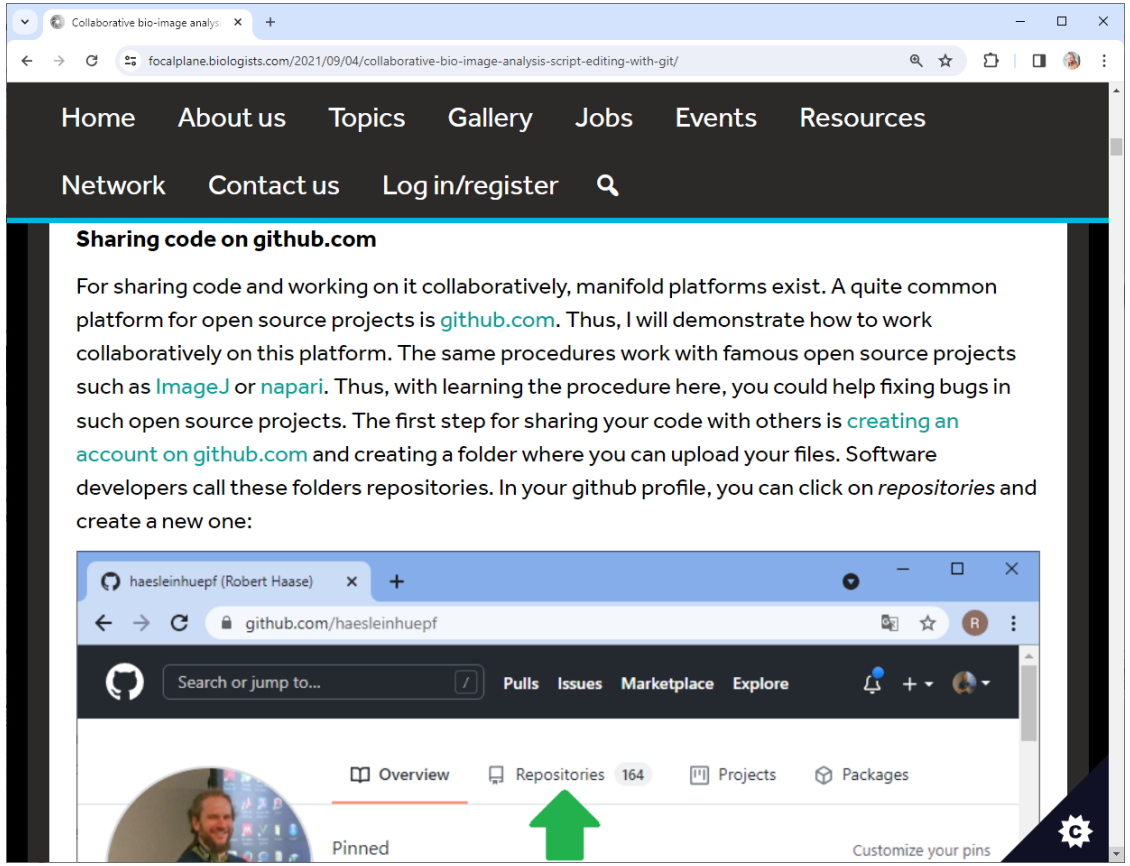


Problem behoben :-)

Github

- Falls das zu schnell war…





Quiz

- Ein Fork ist …

Eine Kopie

Ein Ordner

Eine private Repository

Eine Datei

Mittagspause!

13:30 geht es weiter

ScaDS.AI DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

NFDI 4 BIOIMAGE

NATIONAL RESEARCH DATA MANAGEMENT INFRASTRUCTURE FOR MICROSCOPY AND BIOIMAGE ANALYSIS

GloBIAS Global BioImage Analysts' Society

GLOBAL BIOIMAGE ANALYST'S SOCIETY

Praktikum

Robert Haase

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

SACHSEN

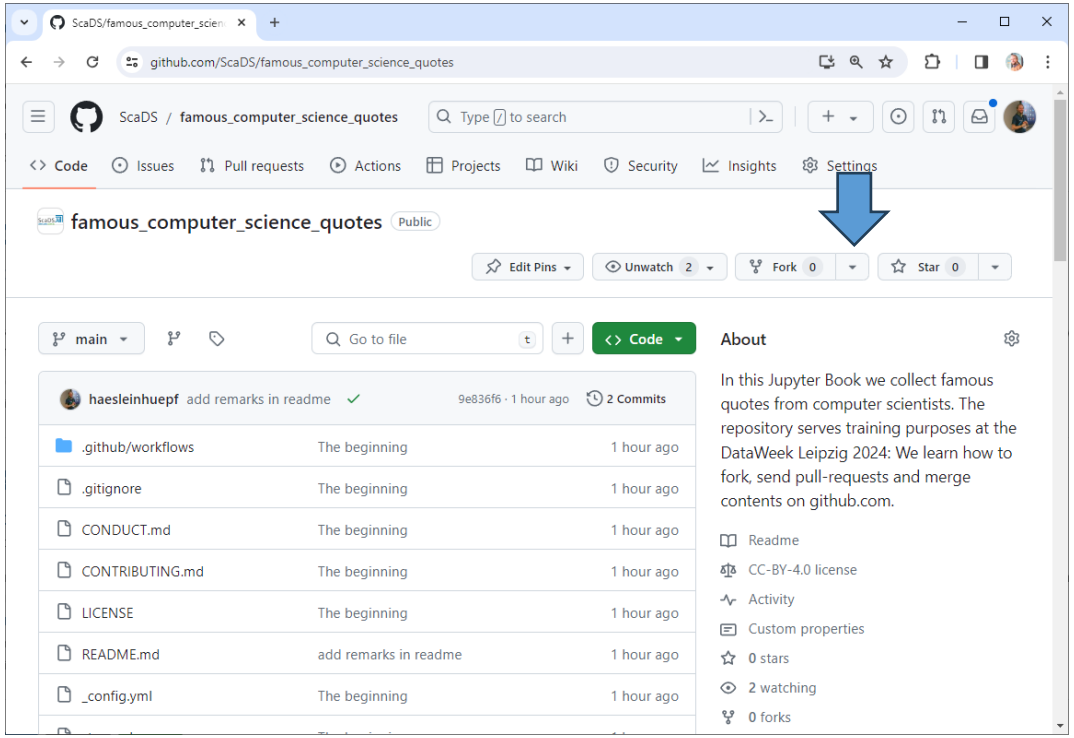
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

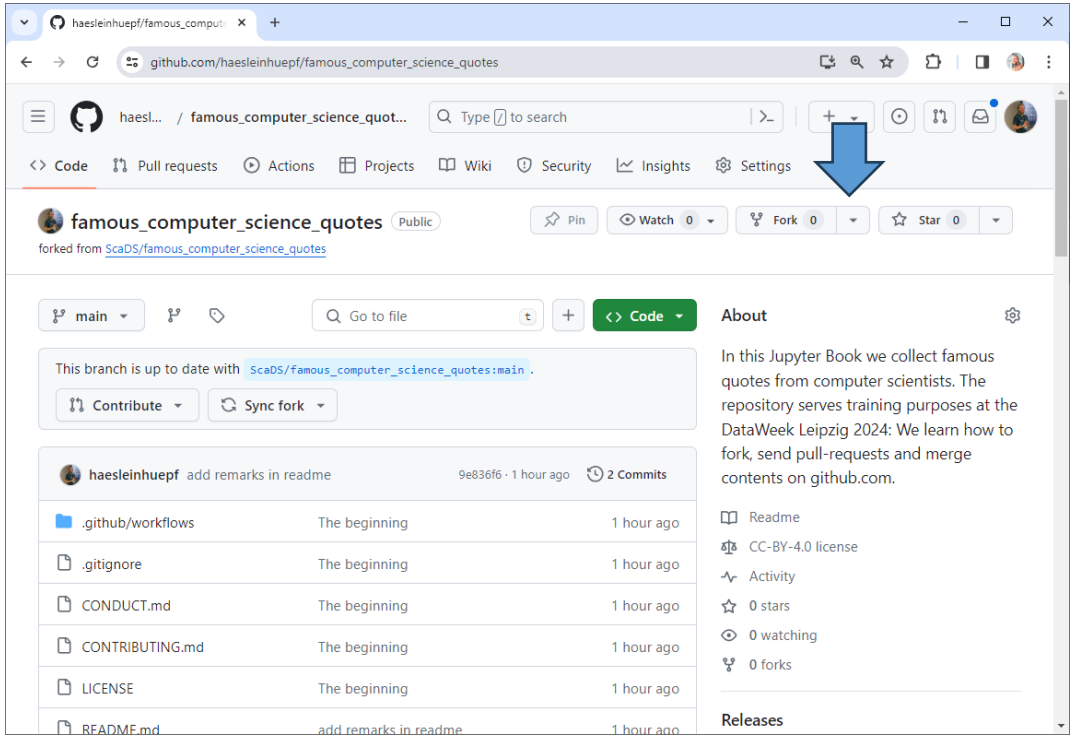
Praktikum: Merge-Konflikte

- Ziel: Erzeugt und löst einen Merge-Konflikt
- Arbeitet zu zweit oder in kleinen Gruppen

Praktikum: Merge-Konflikte

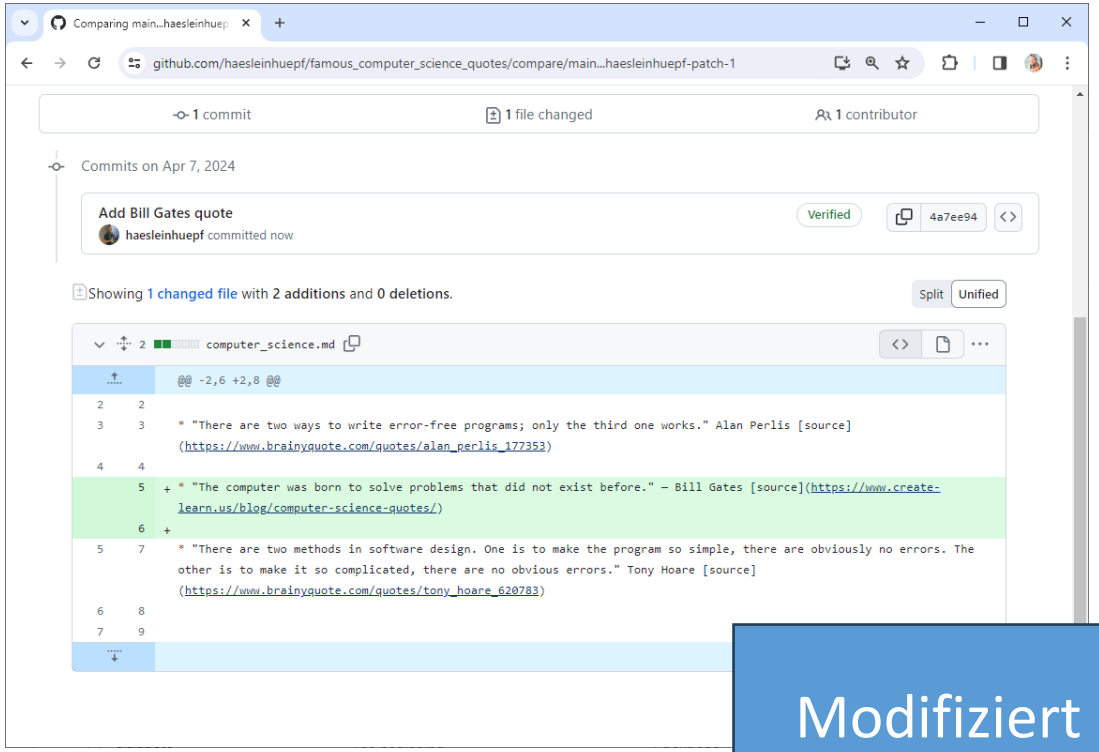
- Schritt 1: Forkt diese Repository
- Schritt 2: Forkt den Fork Eurer Nachbarn

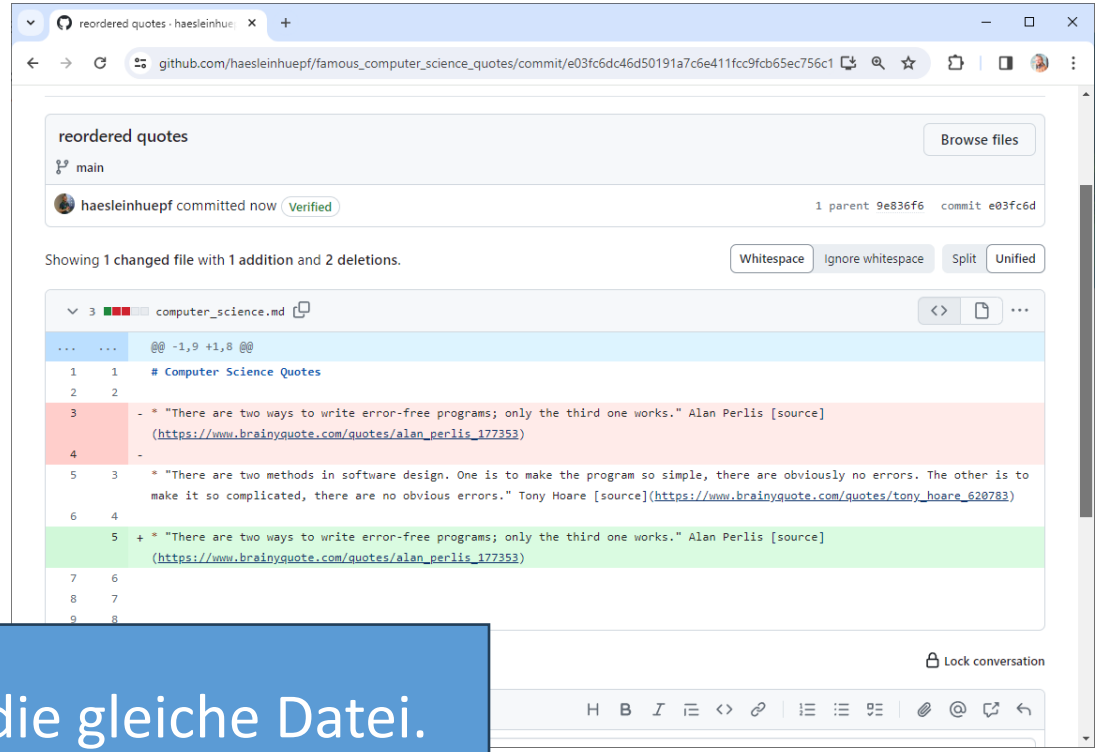




Praktikum: Merge-Konflikte

- Euer task: Ändert eine Datei in Eurem Fork
- Task der Nachbarn: Ändert die gleiche Datei in deren Fork

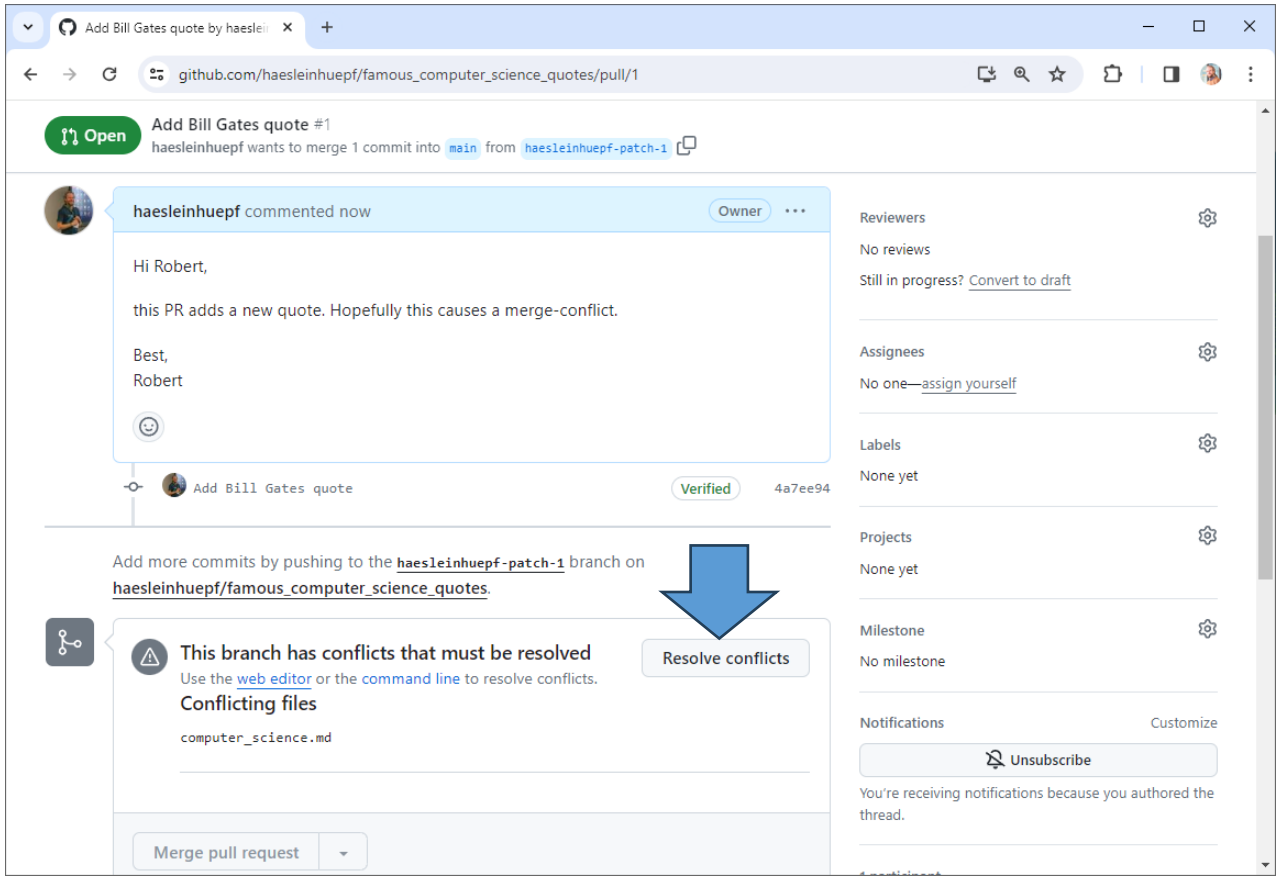




Modifiziert die gleiche Datei.
Erzeugt einen Konflikt!

Praktikum: Merge-Konflikte

- Sendet einen Pull-request zu Euren Nachbarn
- Löst den Konflikt



ScaDS.AI DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

NFDI 4 BIOIMAGE

NATIONAL RESEARCH DATA MANAGEMENT INFRASTRUCTURE FOR MICROSCOPY AND BIOIMAGE ANALYSIS

GloBIAS Global BioImage Analysts' Society

GLOBAL BIOIMAGE ANALYST'S SOCIETY

Github-Zenodo-Integration & Github Pages

Robert Haase

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

SACHSEN

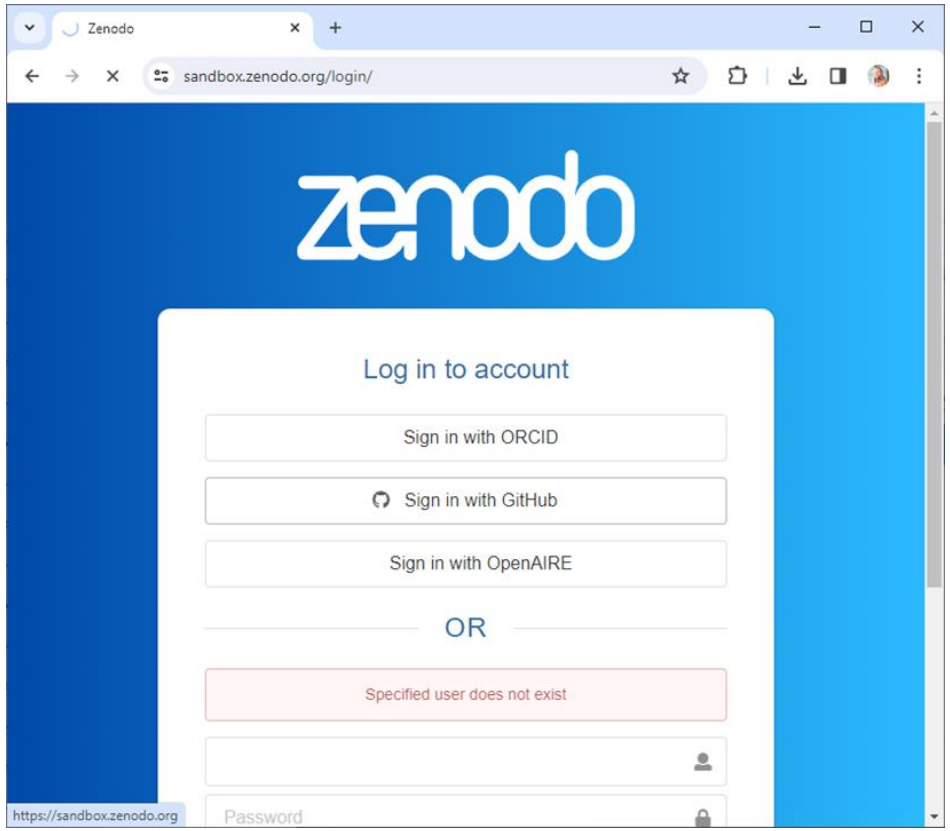
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

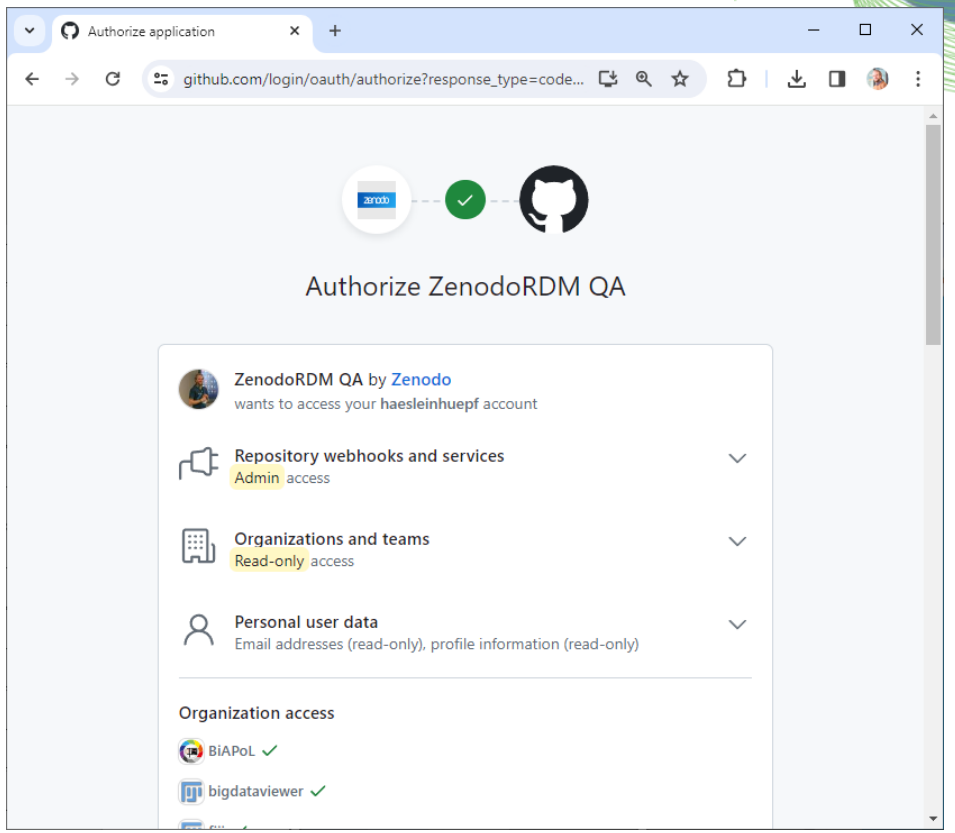
Github versus Zenodo

- Github
  - Gemeinsames Arbeiten
  - Kleine Änderungen werden einzeln erfasst
  - Exaktes Logbuch
- Zenodo
  - öffentlich finanzierte Infrastruktur
  - Archivlösung
  - Nur Releases werden archiviert.

Zenodo-Integration

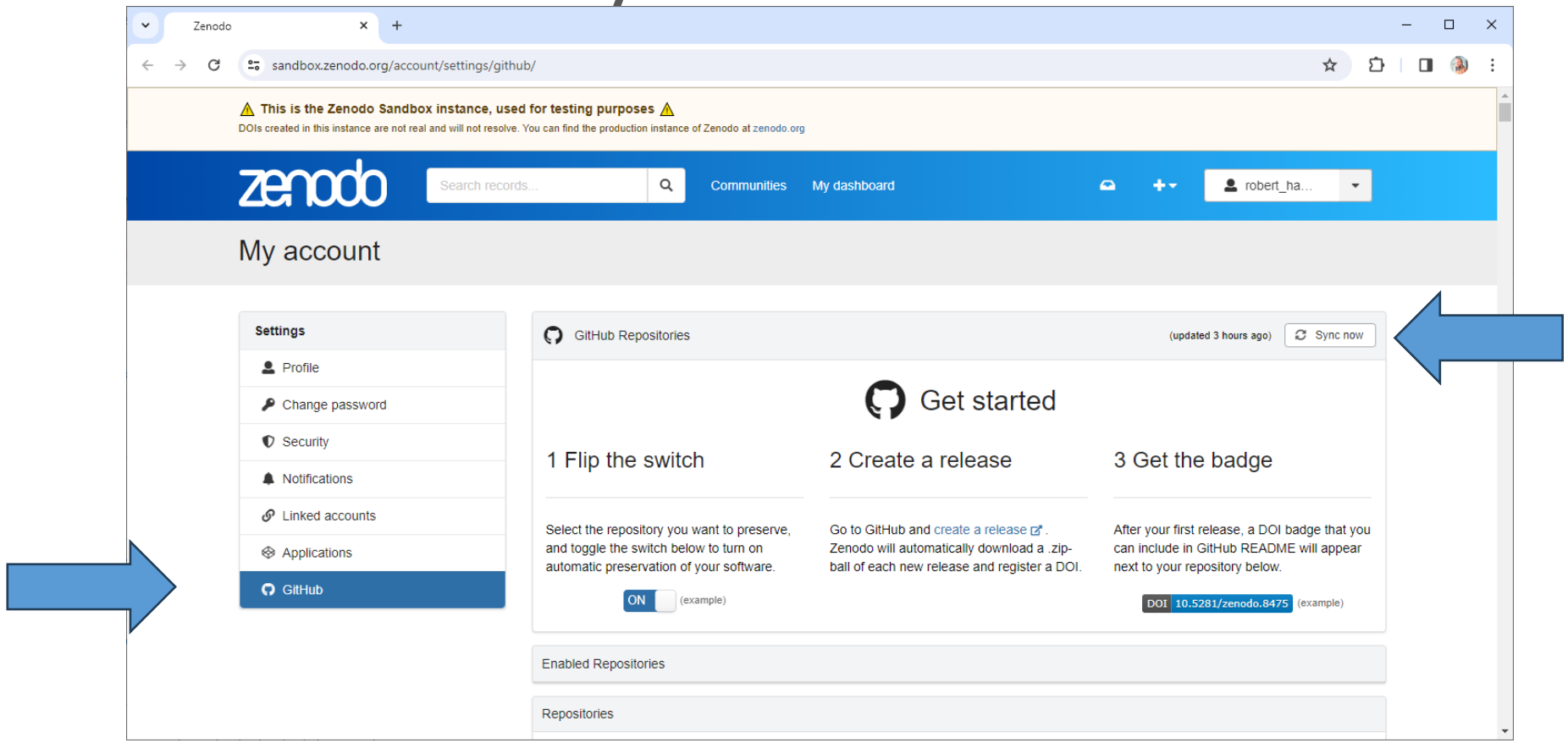
- Login bei Zenodo mit dem Github-Account





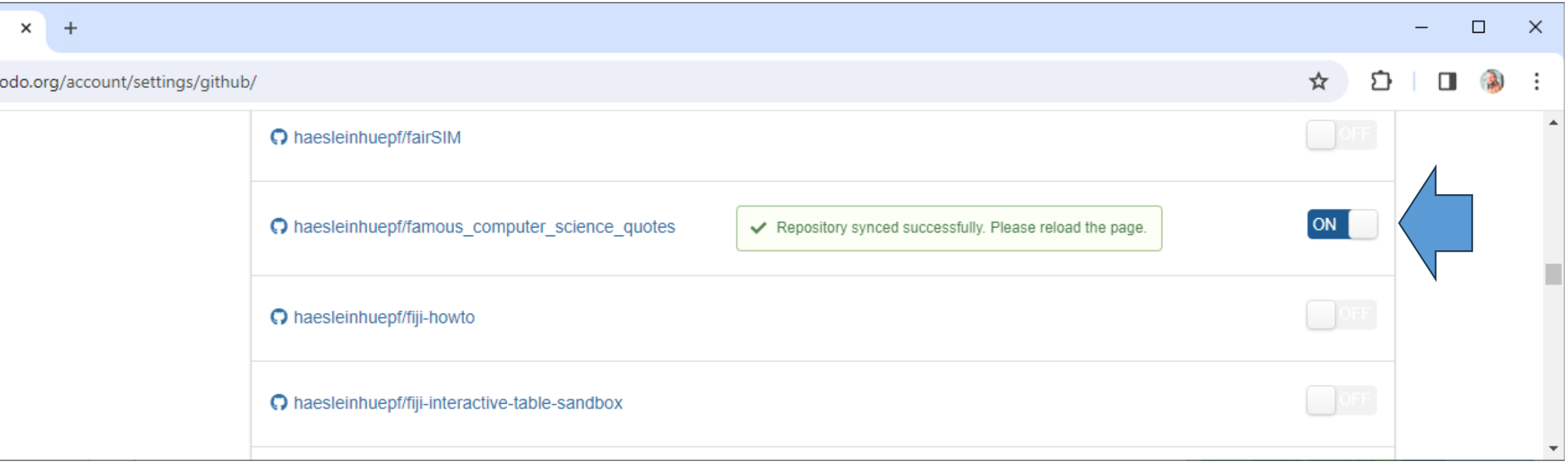
Zenodo-Integration

- Github mit Zenodo synchronisieren



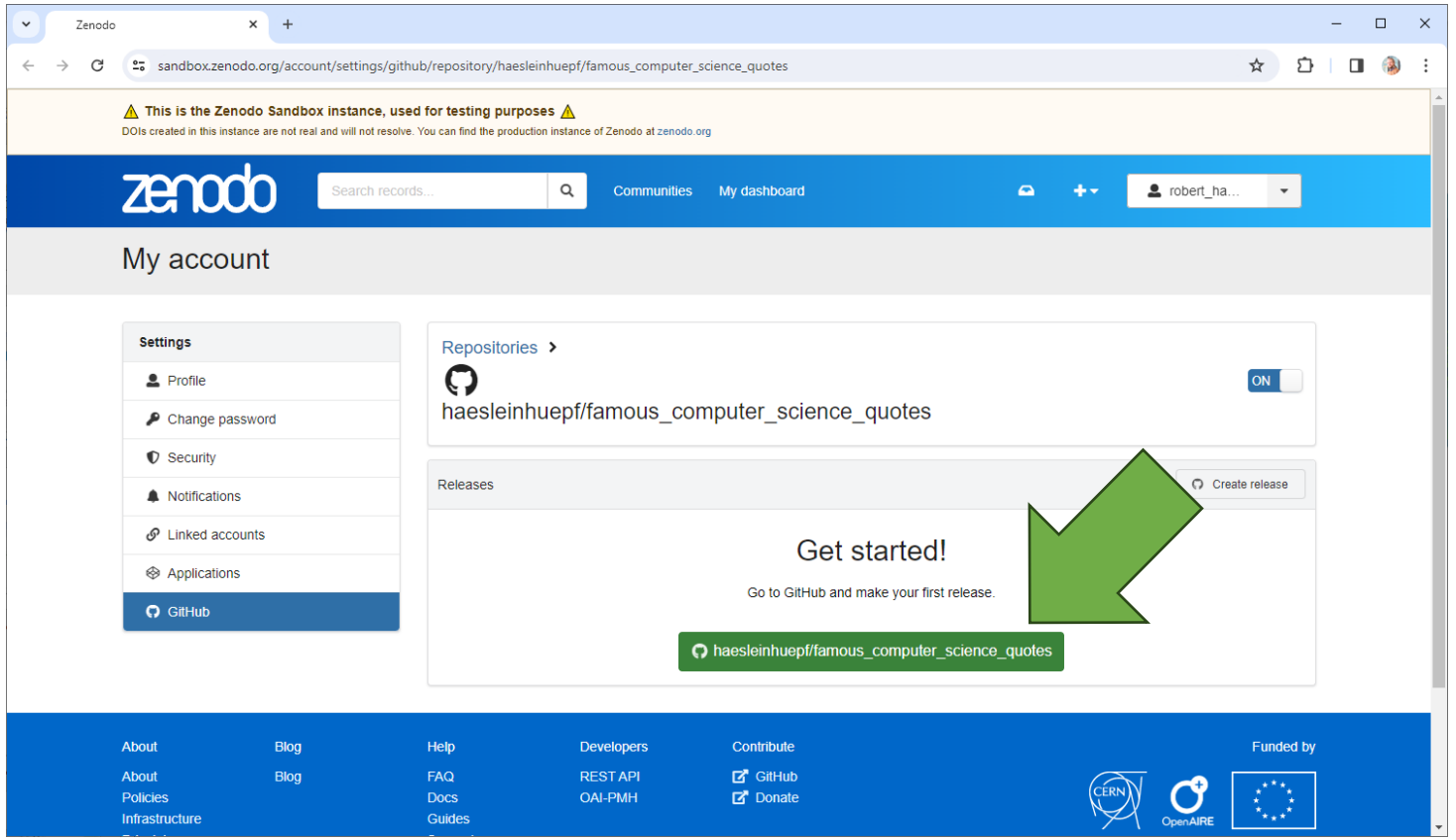
Zenodo integration

- Suche nach dem Projekt, das archiviert werden soll
- Schalte Archivierung ein



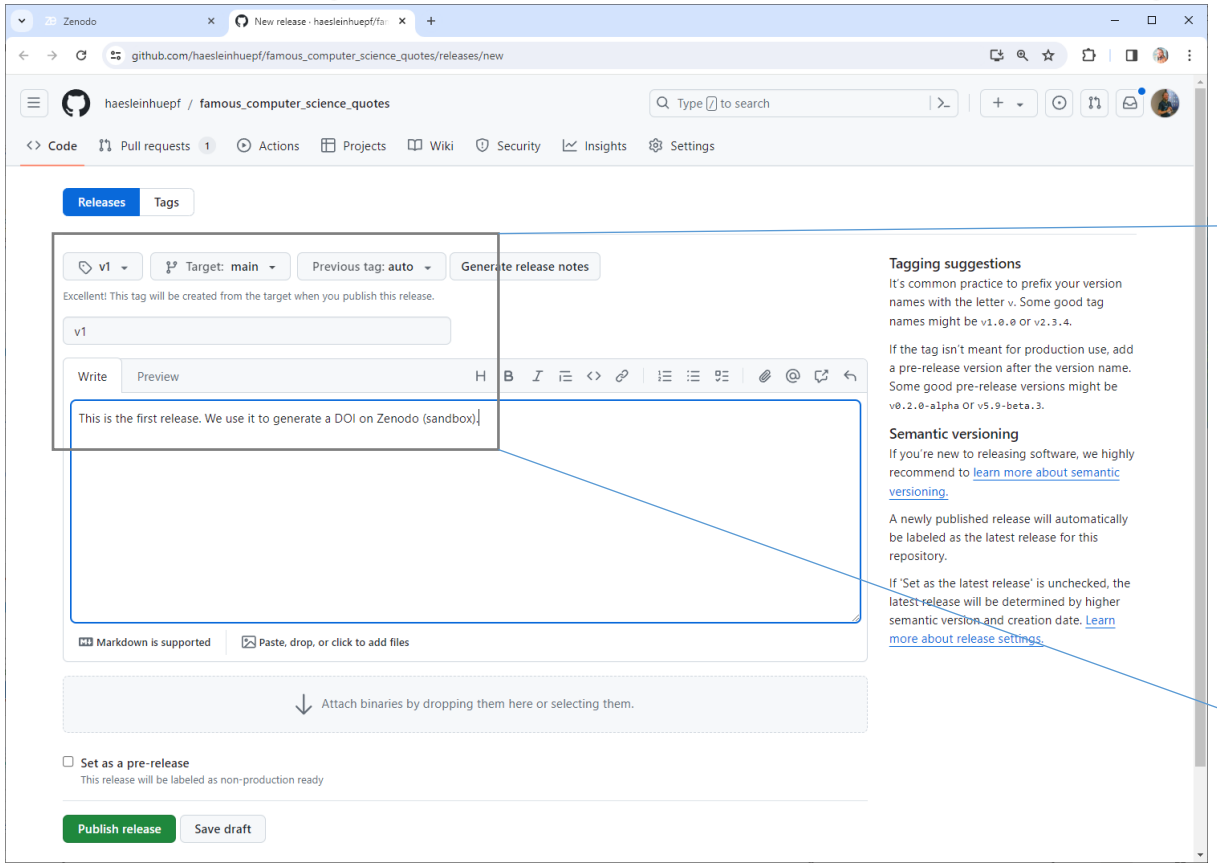
Zenodo integration

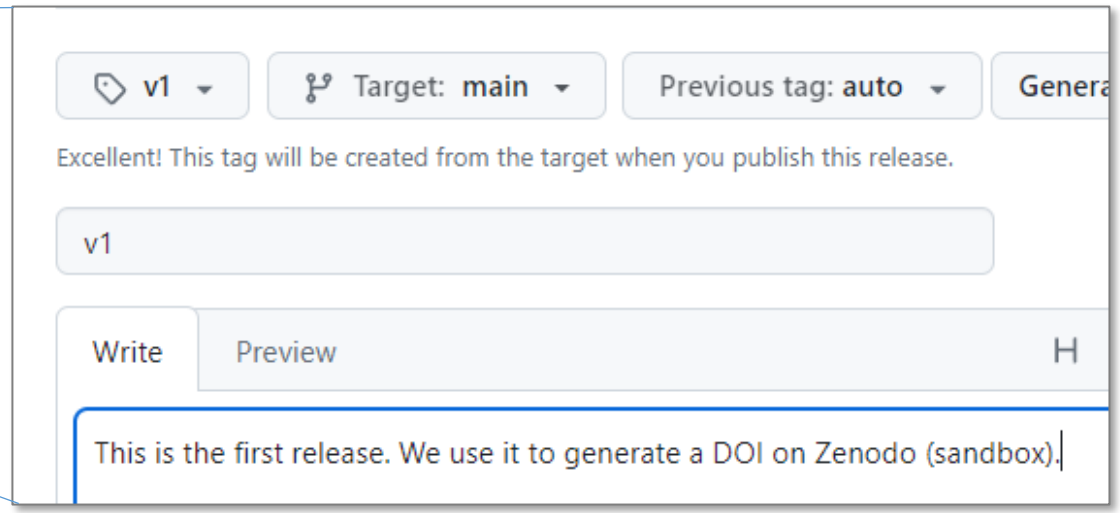
- Klick auf das Repository um Details zu sehen



Zenodo integration

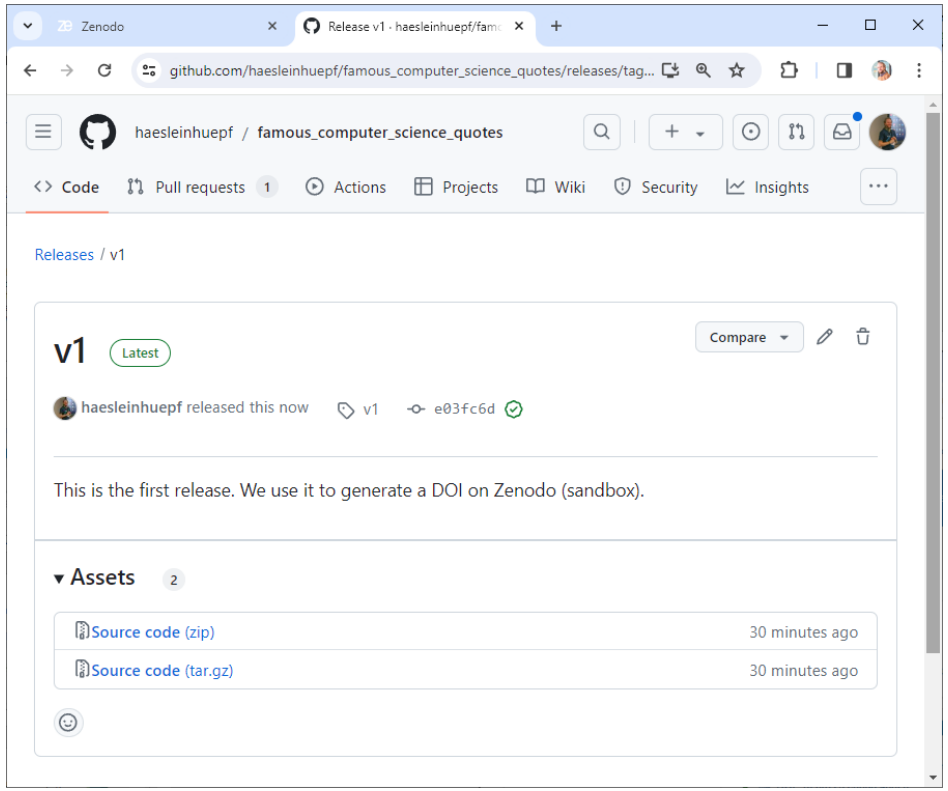
- Erzeugt einen Versionstag (bspw. “v1”)

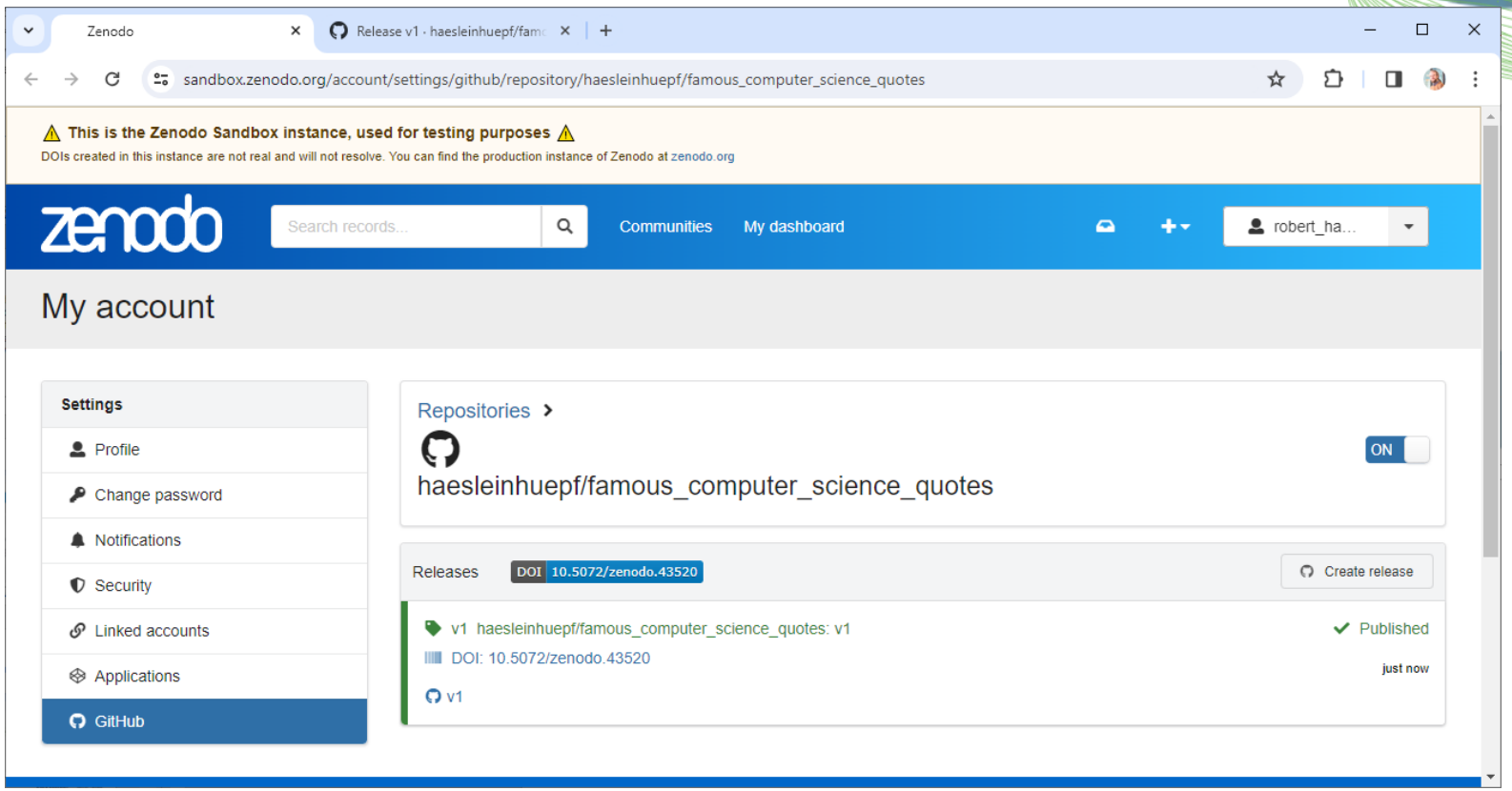




Zenodo integration

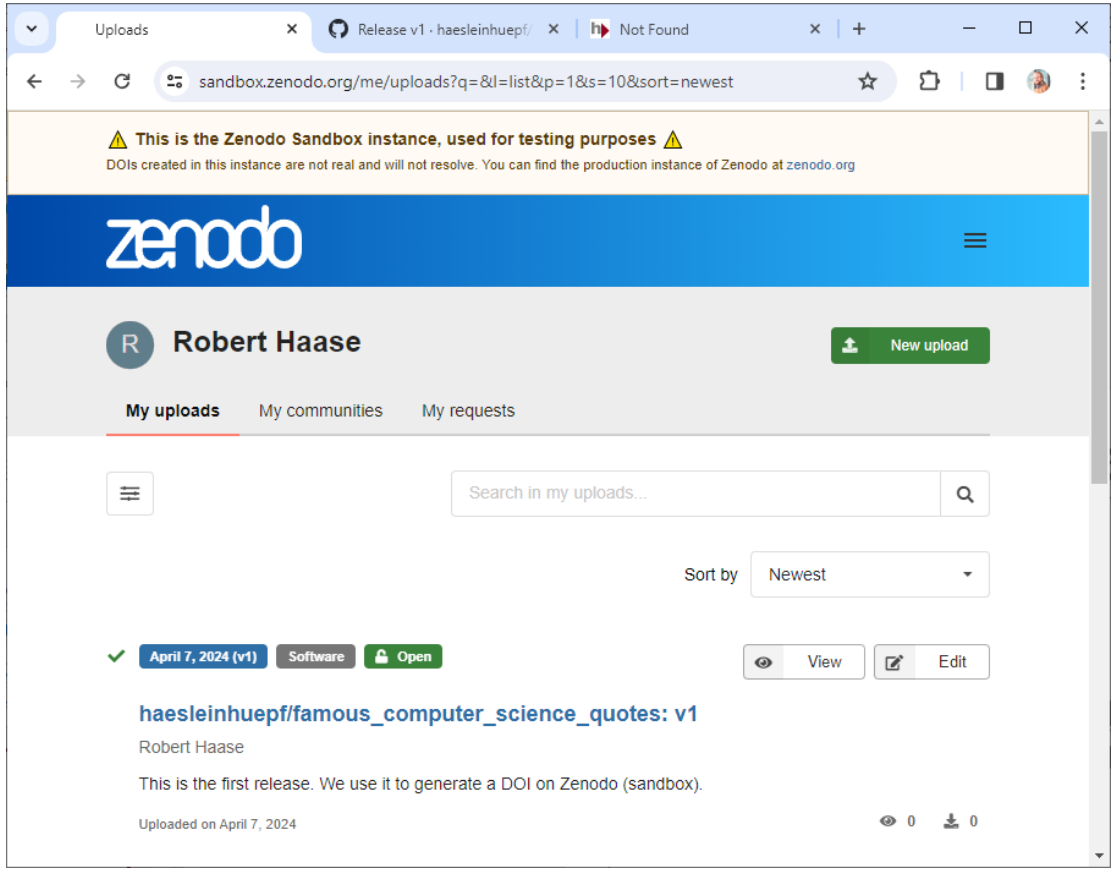
- Erkundet den Release auf Github und Zenodo

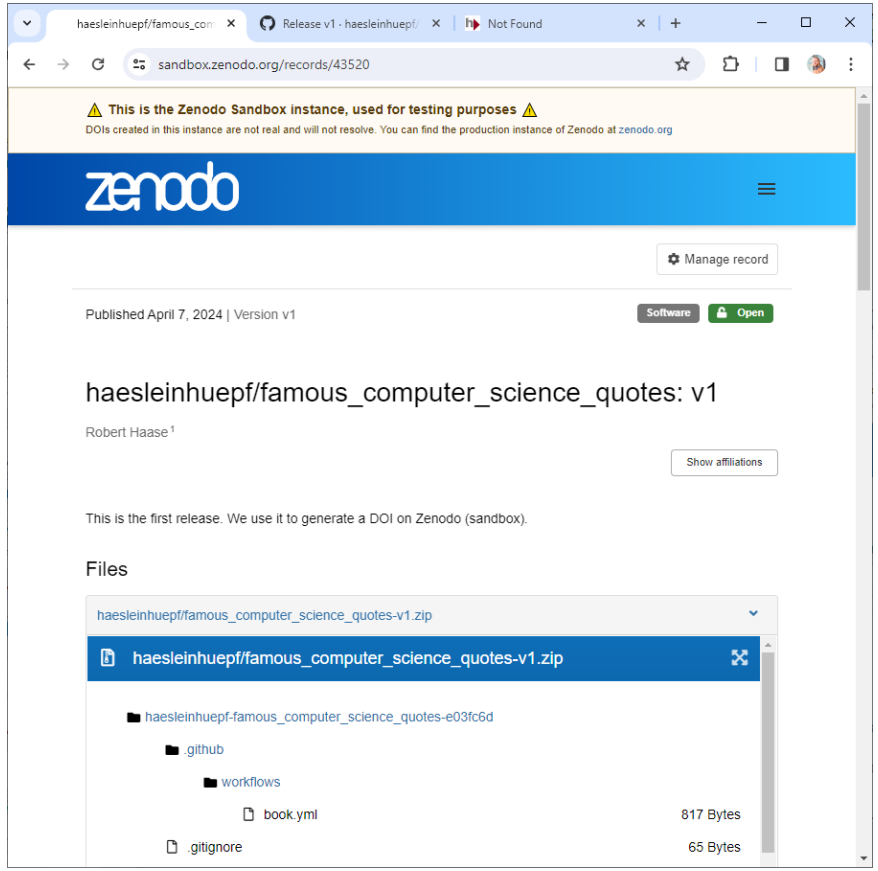




Zenodo integration

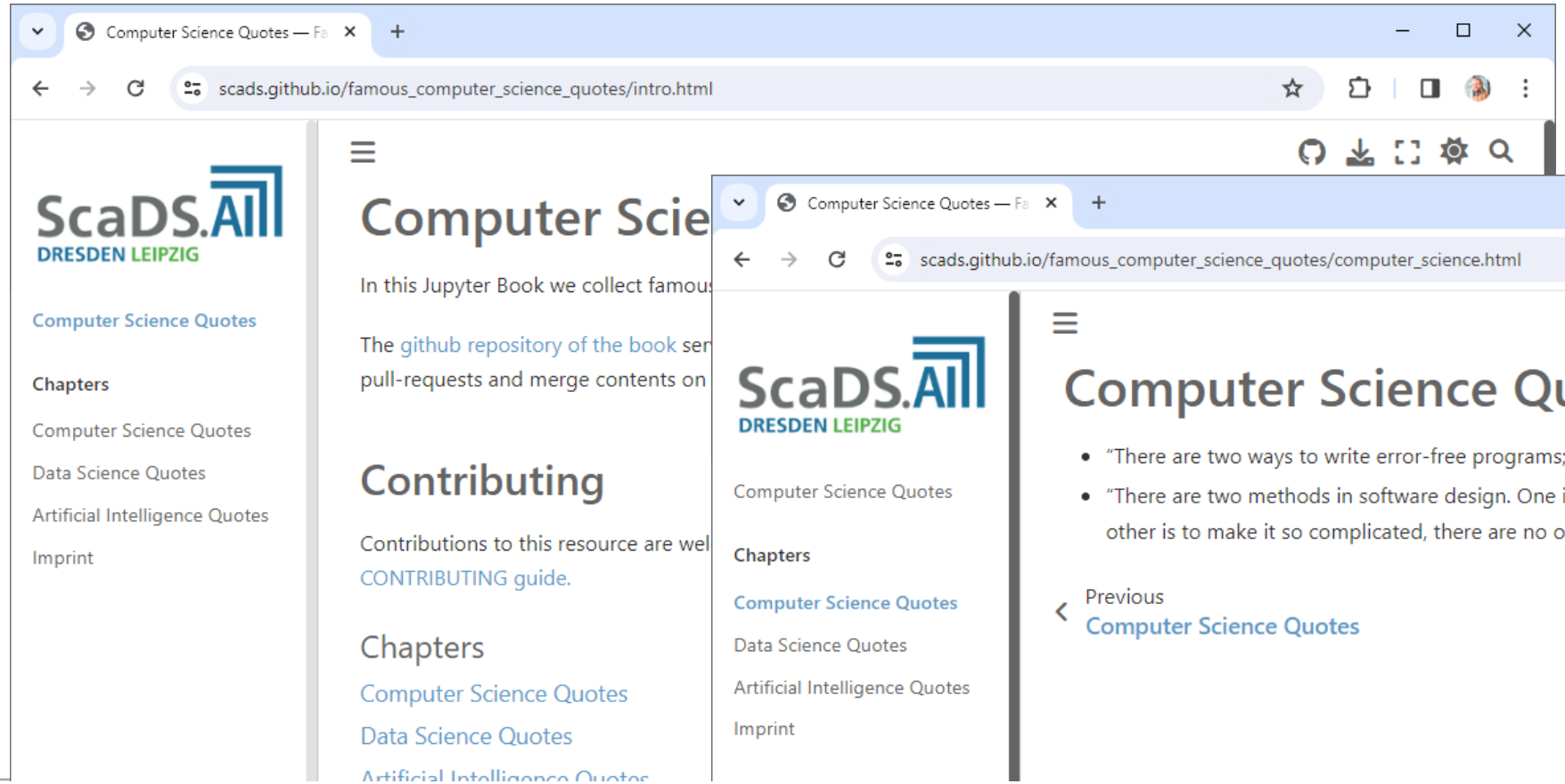
- Zurück auf dem Dashboard wird der neue Eintrag angezeigt

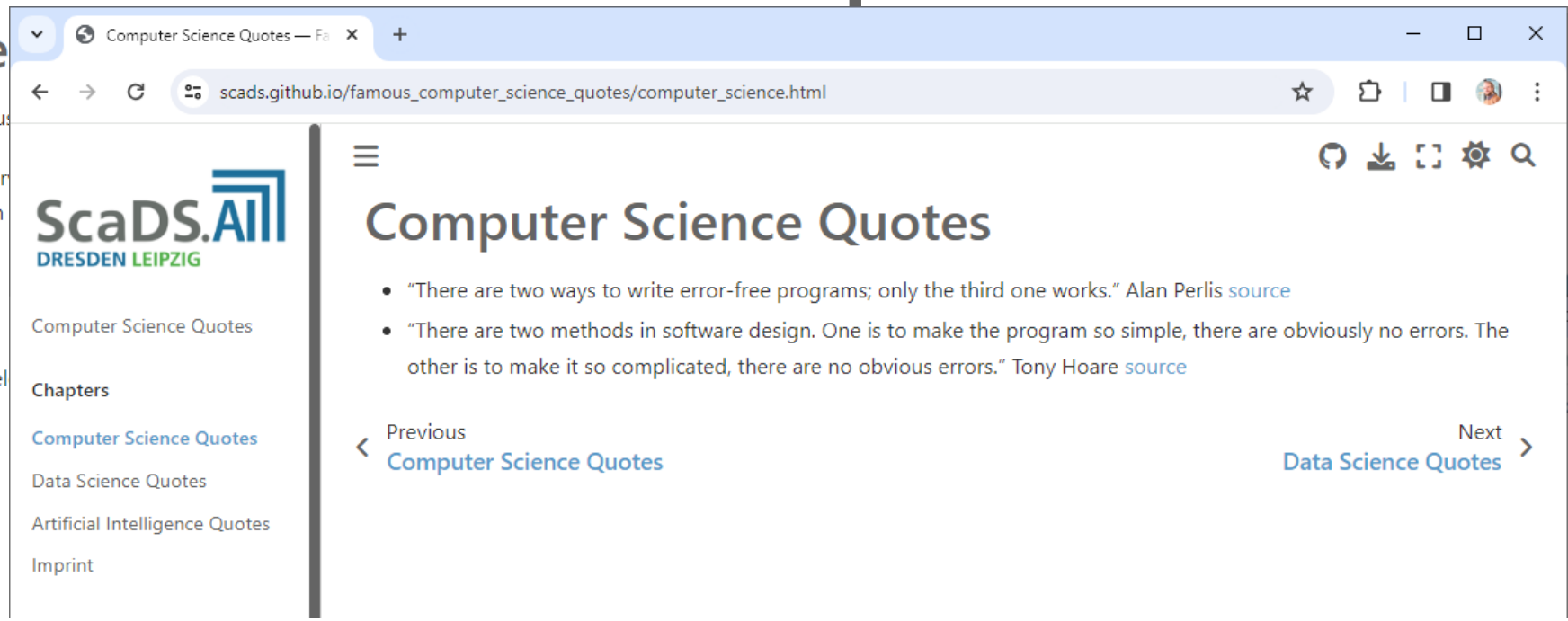




Github pages

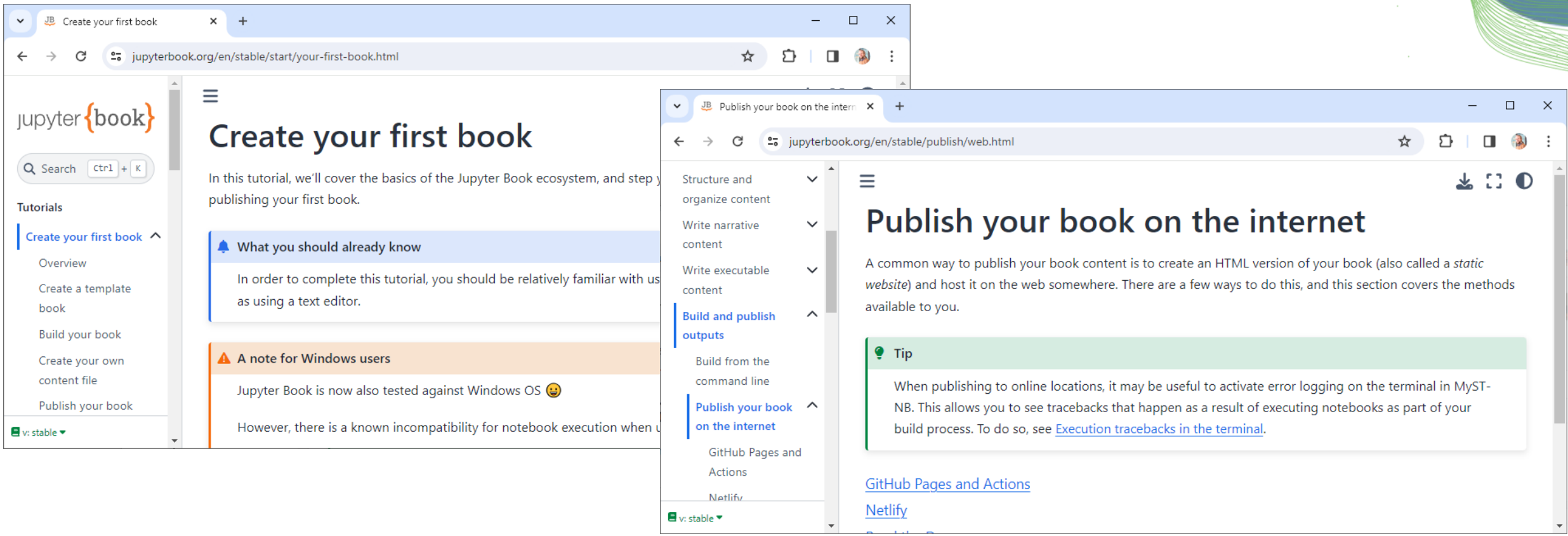
- Kostenloses Webhosting, weitgehend automatisch





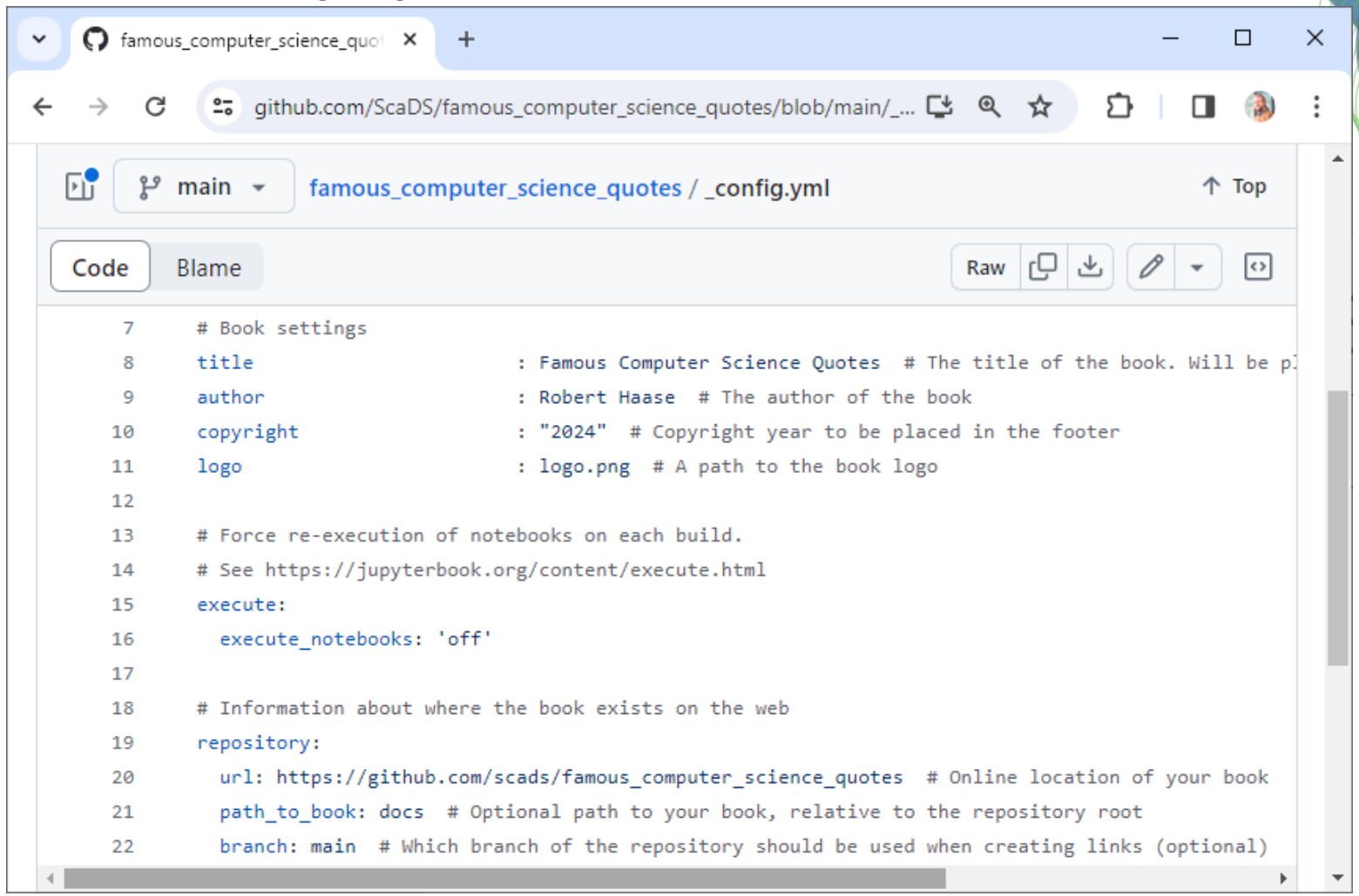
Github pages: Jupyter Books

- Basiert auf dem Jupyter Books Projekt



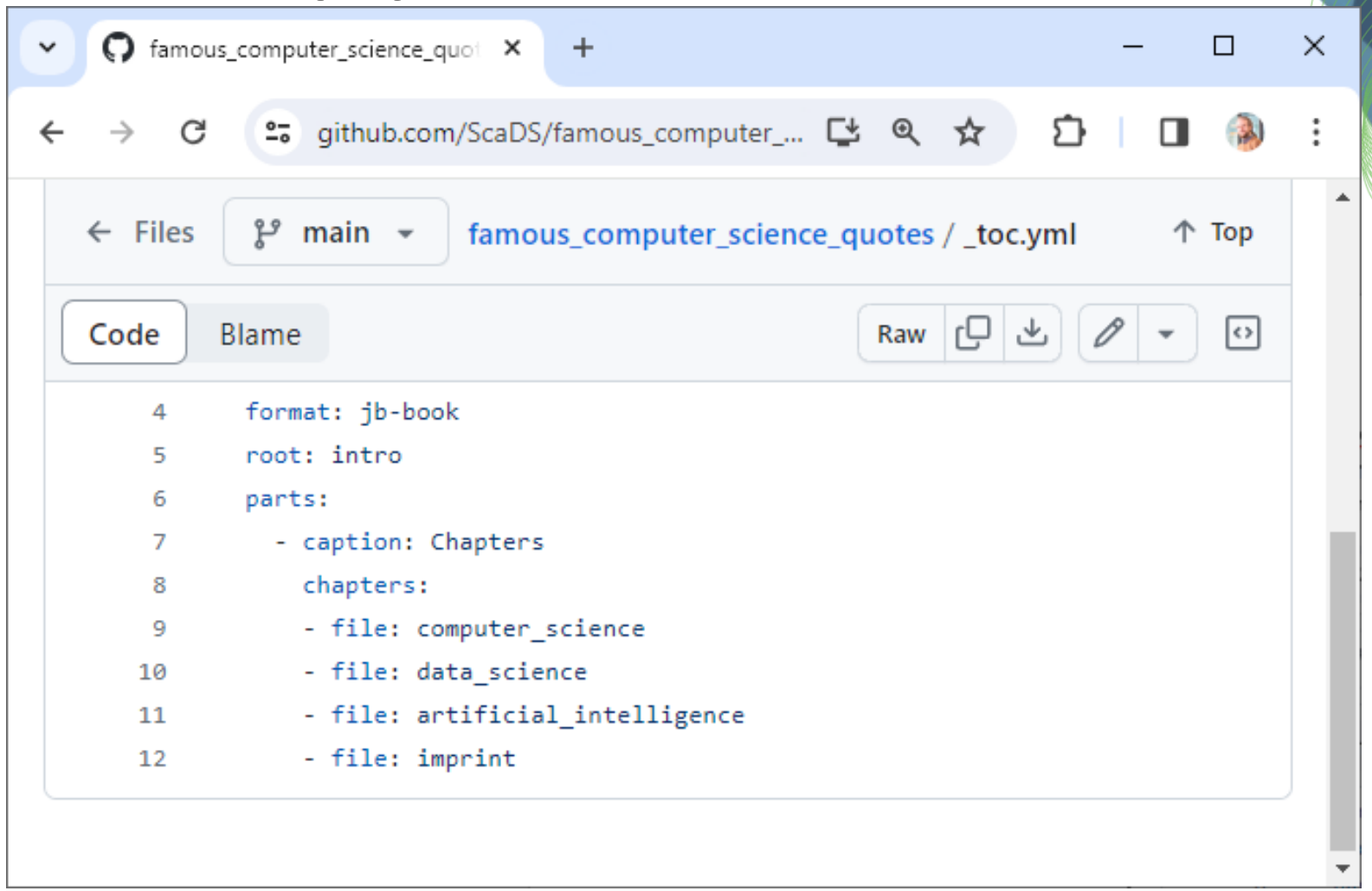
Github pages: Jupyter Books

- Configuration: _config.yml



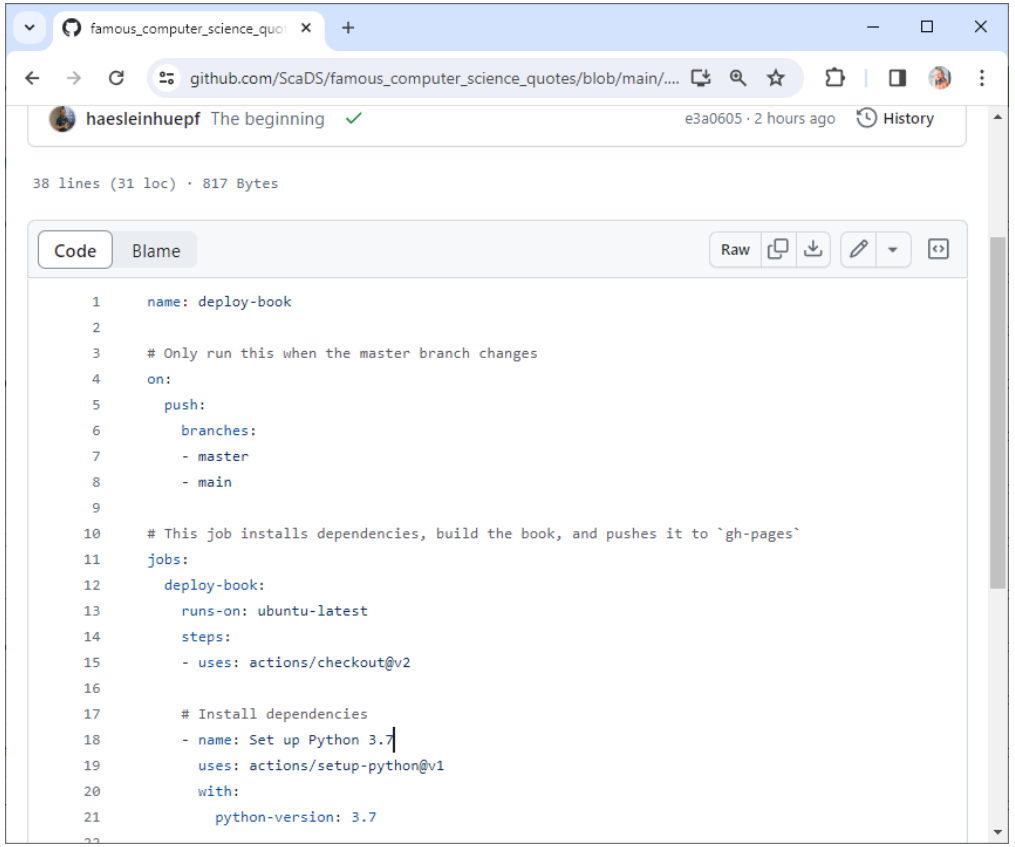
Github pages: Jupyter Books

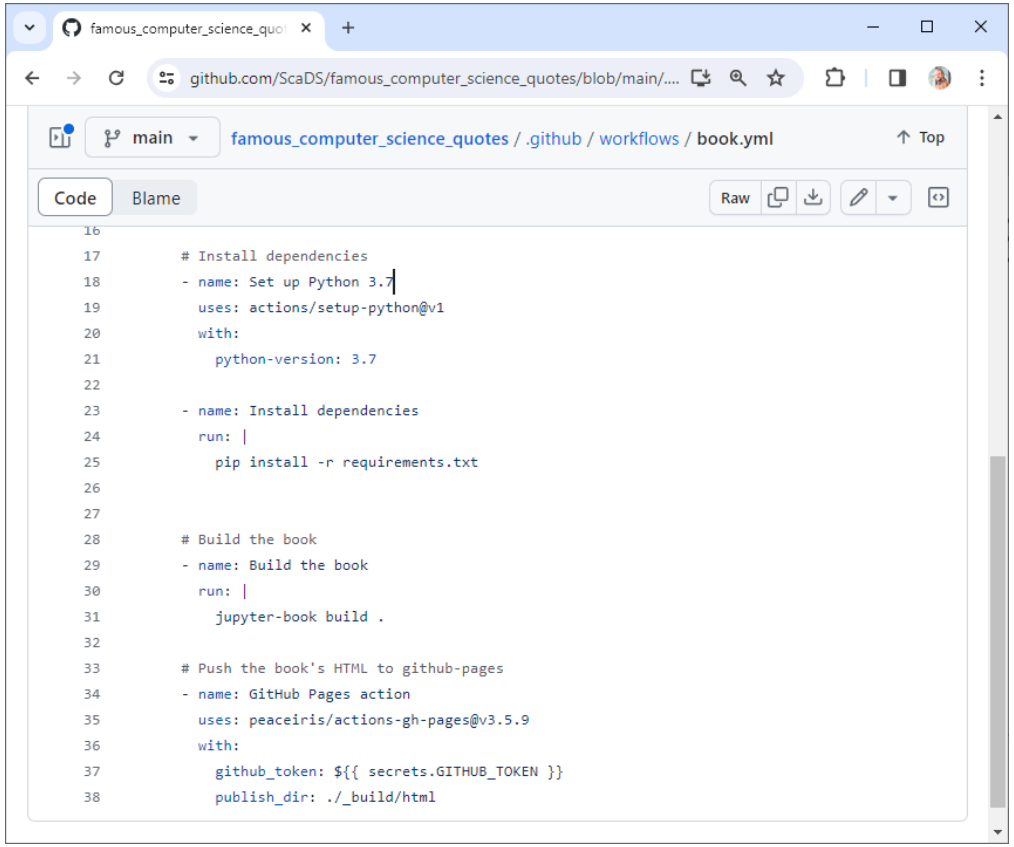
- Configuration: _toc.yml



Github pages: Jupyter Books

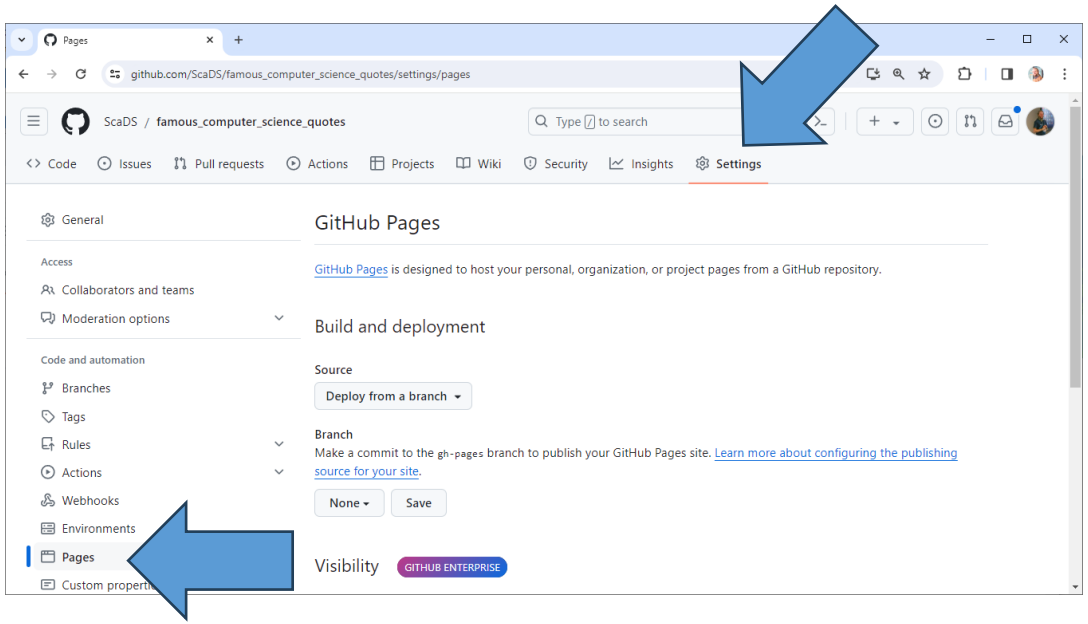
- Behind the scenes: Github workflows

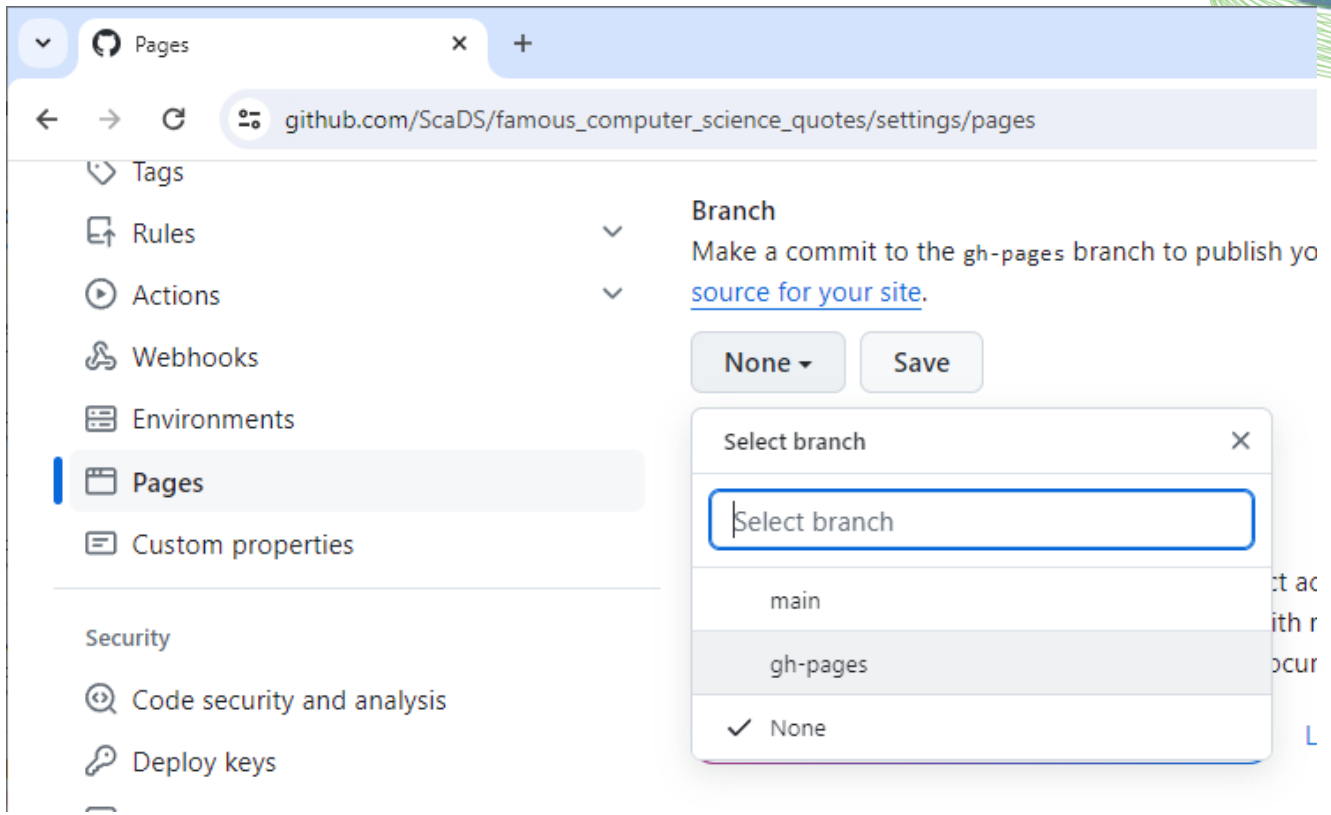




Github pages

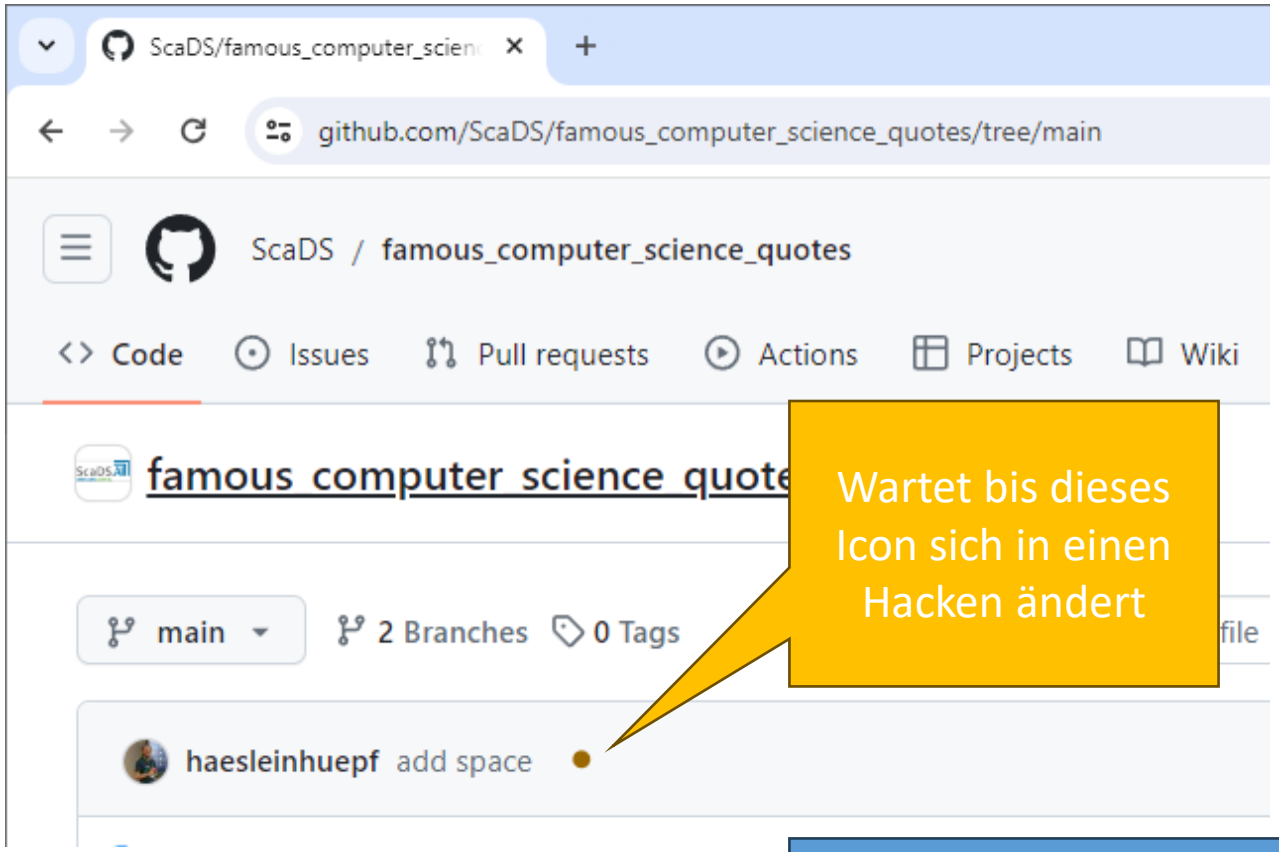
- Konfiguration in Github: “gh-pages” bei Pages / Branch auswaehlen

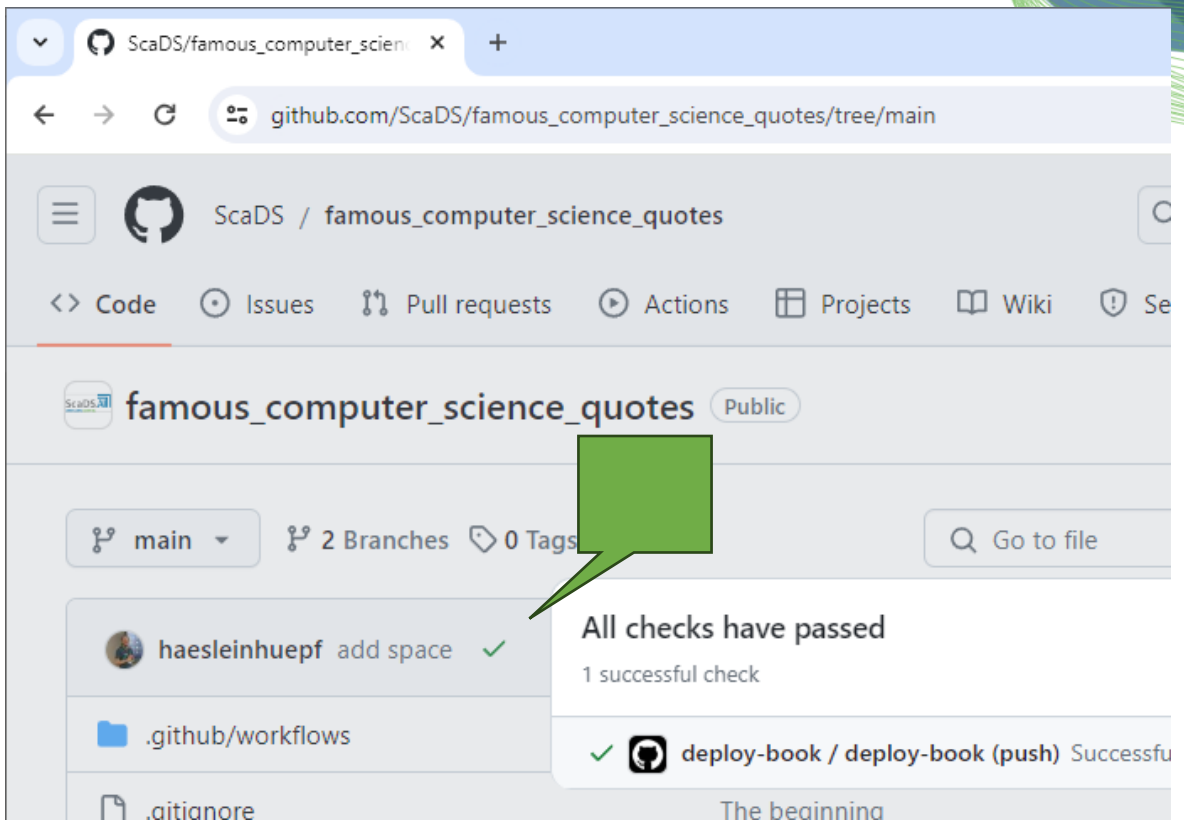




Github pages

- Warten ☺

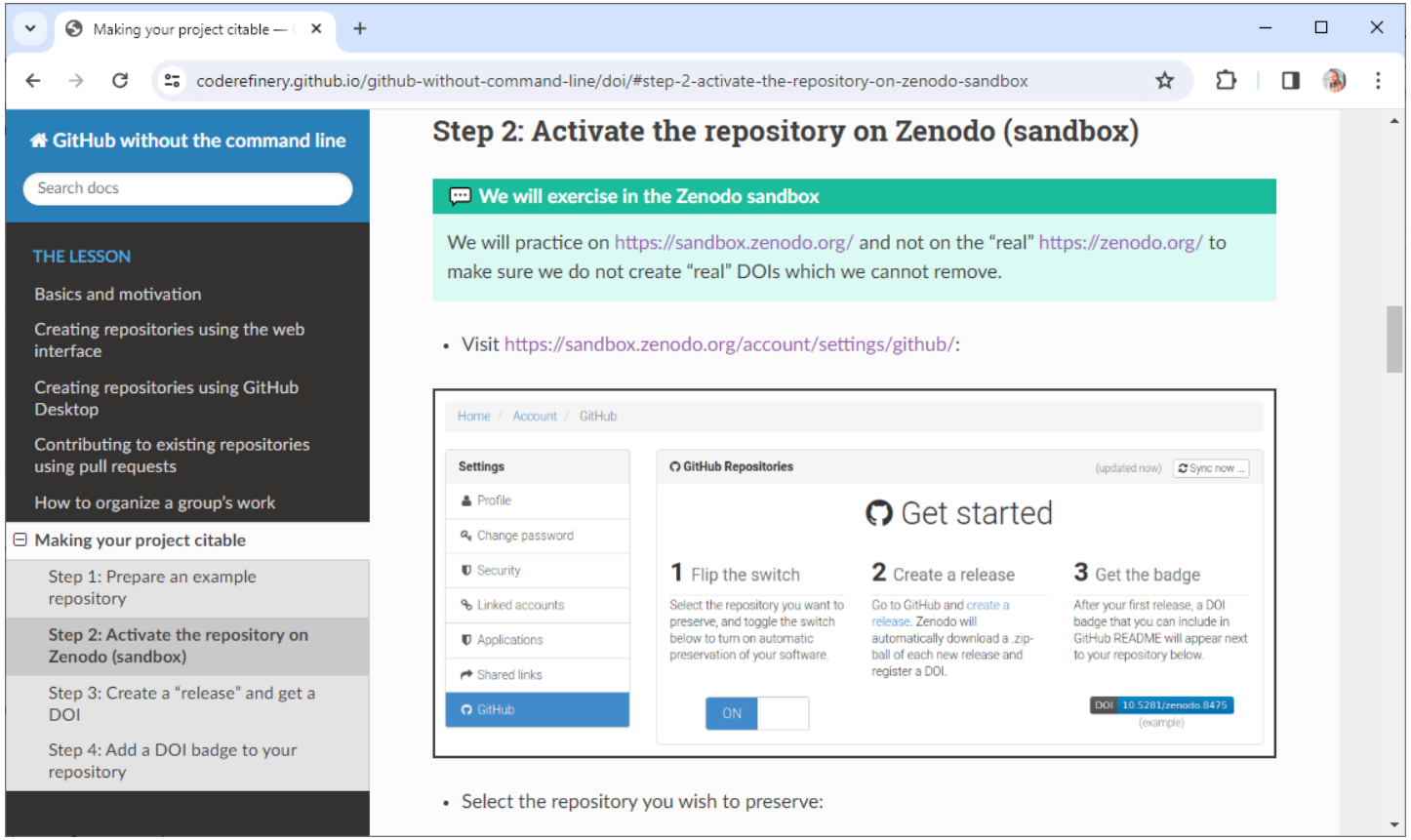




Das wird automatisch passieren, wenn man eine Datei ändert

Zenodo integration & Github pages

- In case this was too fast…



ScaDS.AI DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

NFDI 4 BIOIMAGE

NATIONAL RESEARCH DATA MANAGEMENT INFRASTRUCTURE FOR MICROSCOPY AND BIOIMAGE ANALYSIS

GloBIAS Global BioImage Analysts' Society

GLOBAL BIOIMAGE ANALYST'S SOCIETY

Praktikum

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.

Exercise: Zenodo integration

- Legt eine neue Repository an
  - Ladet die Dateien der Famous Computer Science Quotes Repository hoch
  - Schaltet die Archiv-Funktion in der Zenodo(sandbox) frei
  - Erzeugt einen Release auf Eurer Github Repository

Exercise: Github pages

- Arbeitet weiter in der gleichen Repository
  (Es wird nicht mit einem Fork funktionieren)
- Aktiviert die github-pages in den Einstellungen
- Ändert Titel, Logo, Autoren und die Inhalte.
  (hint: _config.yml and _toc.yml)

Acknowledgements

Communities & platforms

image.sc

NFDI4 BIOIMAGE

GloBIAS Global BioImage Analysts' Society

neubias network of european bioimage analysts

FocalPlane Where biology meets microscopy

SaxFDM

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

DFG Deutsche Forschungsgemeinschaft

Some Figures were generated using DALL·E