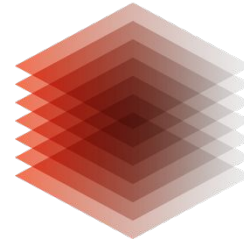


LEIBNIZ-INFORMATIONSZENTRUM  
TECHNIK UND NATURWISSENSCHAFTEN  
UNIVERSITÄTSBIBLIOTHEK




TIB

# Versionskontrolle: Kollaboratives Arbeiten mit Git


Dr. Katrin Leinweber  
Hannover, 3. April 2019  
Erfahrungsaustausch zu FDM in Verbundprojekten

Folien: [is.gd/TB3K6f](https://is.gd/TB3K6f)


# Über Katrin Leinweber: Biochemie & IT

▪ Life Science an  Universität Konstanz



, sowie Arktische Ökologie & Geologie an  UNIVERSITY OF TROMSØ



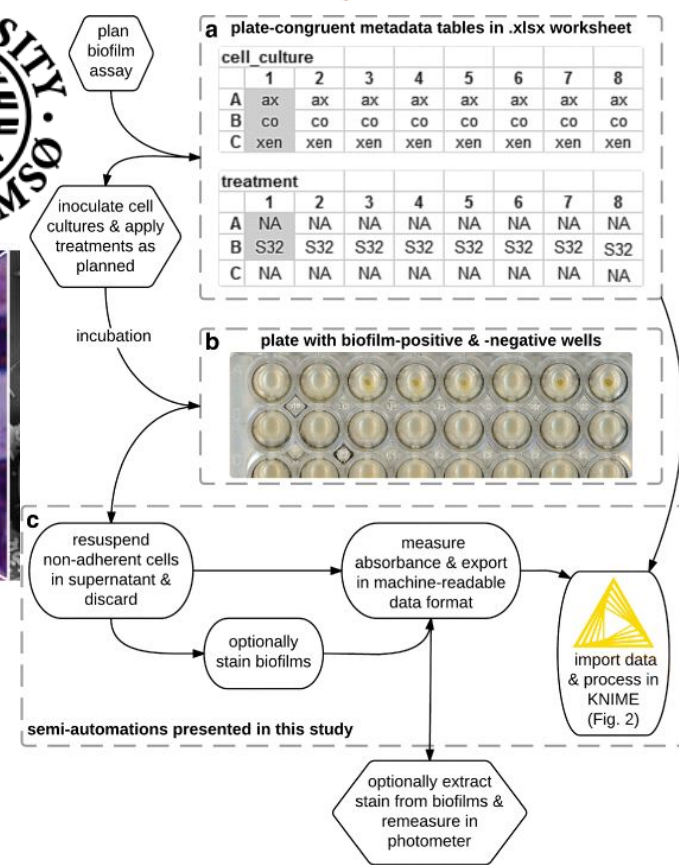
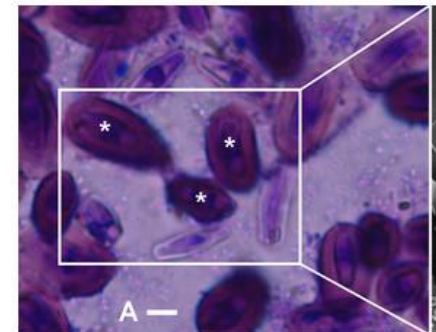
▪ technischer Kundendienst bei  Prezi

▪ Doktorarbeit: Methodenentwicklung für Biofilmforschung

▪  für Datenauswertung;  **git** & **L<sup>A</sup>T<sub>E</sub>X** zum Schreiben

▪ 7 Quartale ein Pharma-Laborinformations- & Managementsystems getestet

▪ seit Nov'17 an TIB: Betreuung wissenschaftlicher Softwareprojekte & Schulungen



## Versionskontrolle: Was ist das?

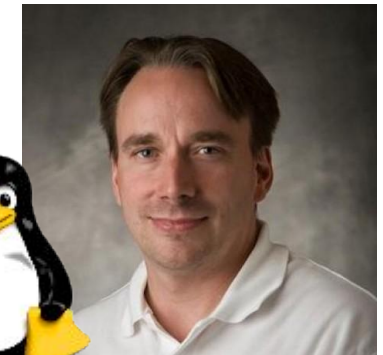
- Assistenzsysteme für Softwareentwicklung seit ≈1970
  - (Quellcode)dateien versionieren & synchronisieren
  - Projektevolution nachvollziehen
  - Zusammenarbeit strukturieren



Margaret Hamilton  
mit AGC-Quellcode



濱野 純  
(Junio Hamano)



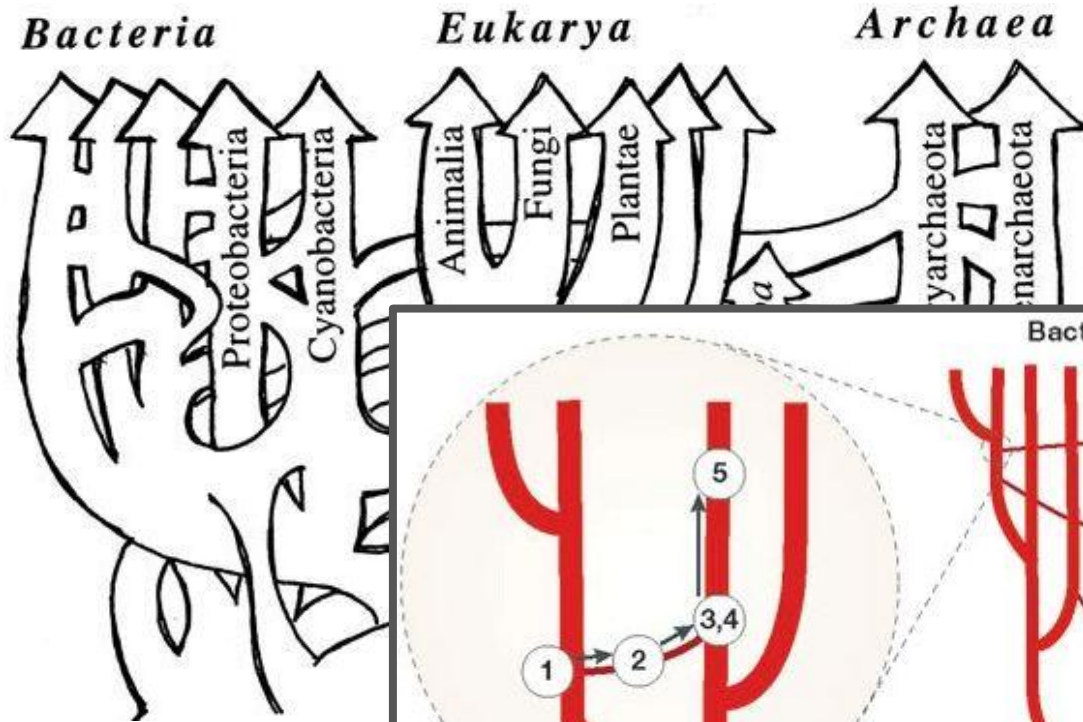
Linus Torvalds

+ viele mehr:  
[GitHub.com/git/git/  
graphs/contributors](https://github.com/git/git/graphs/contributors)

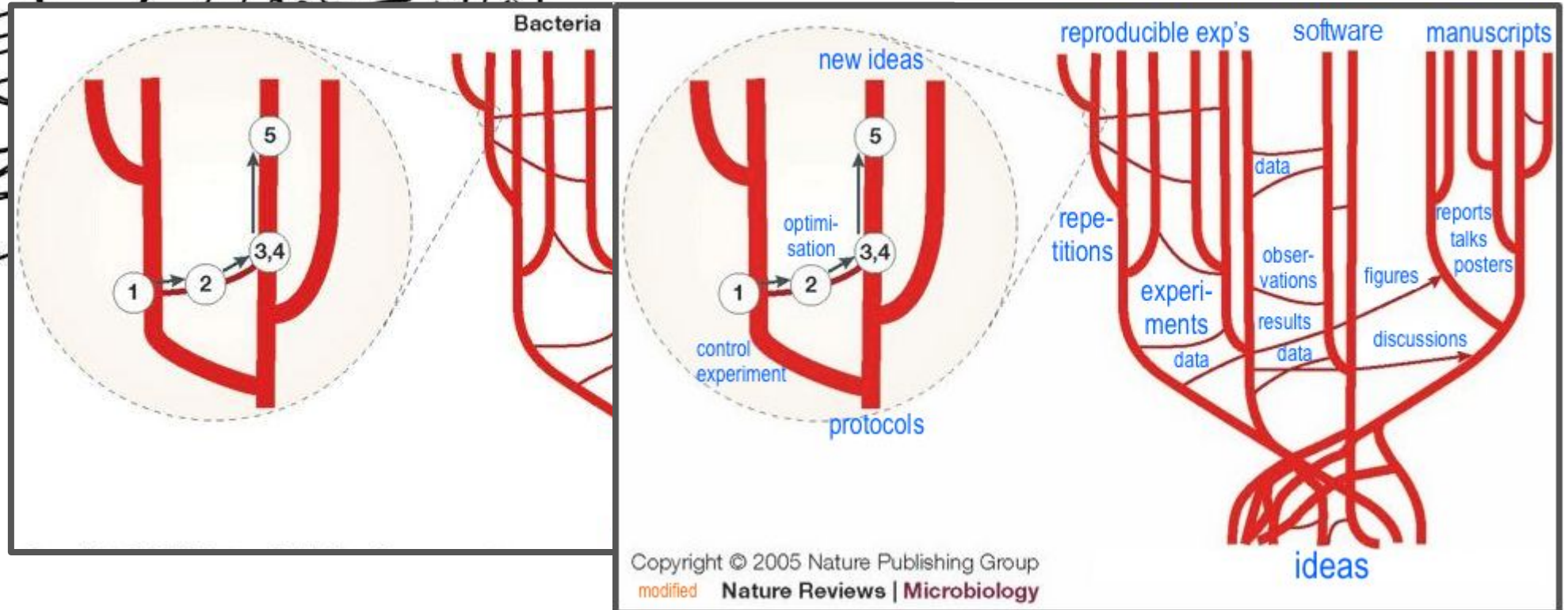


- seit Apr'05 für Linux-Kernel-Entwicklung [GitHub.com/git/git/commits/e497ea2a](https://github.com/git/git/commits/e497ea2a)

# Versionskontrolle, Git: Wieso, weshalb, warum?

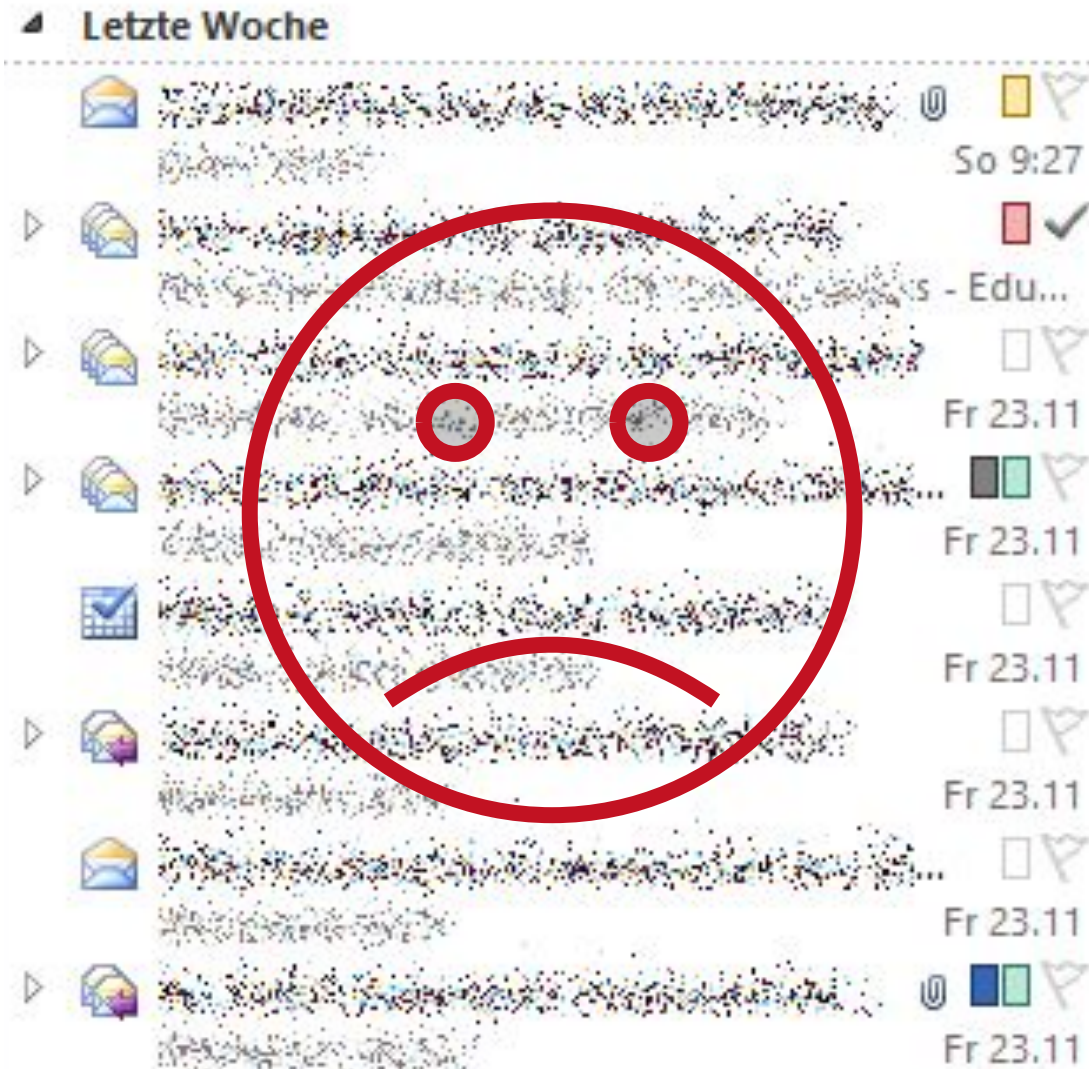


- Nachvollziehbarkeit & Durchblick in digitalen Projekten

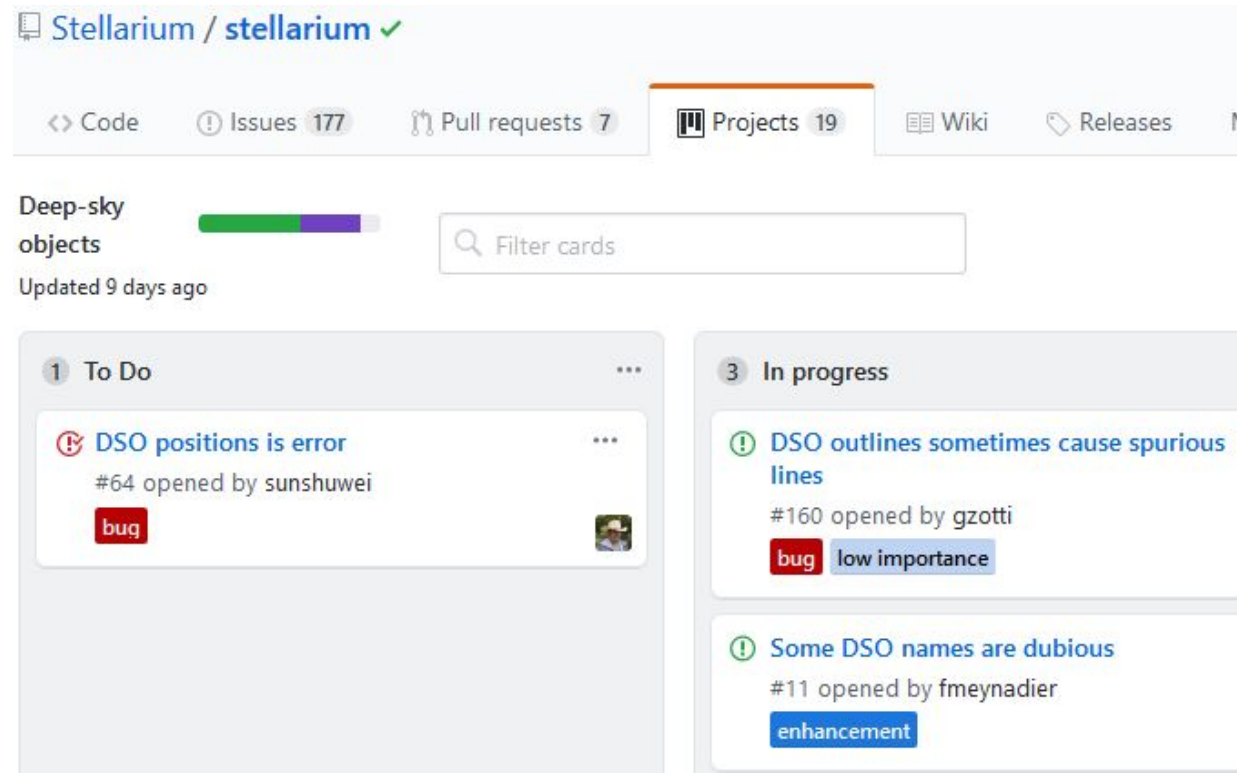




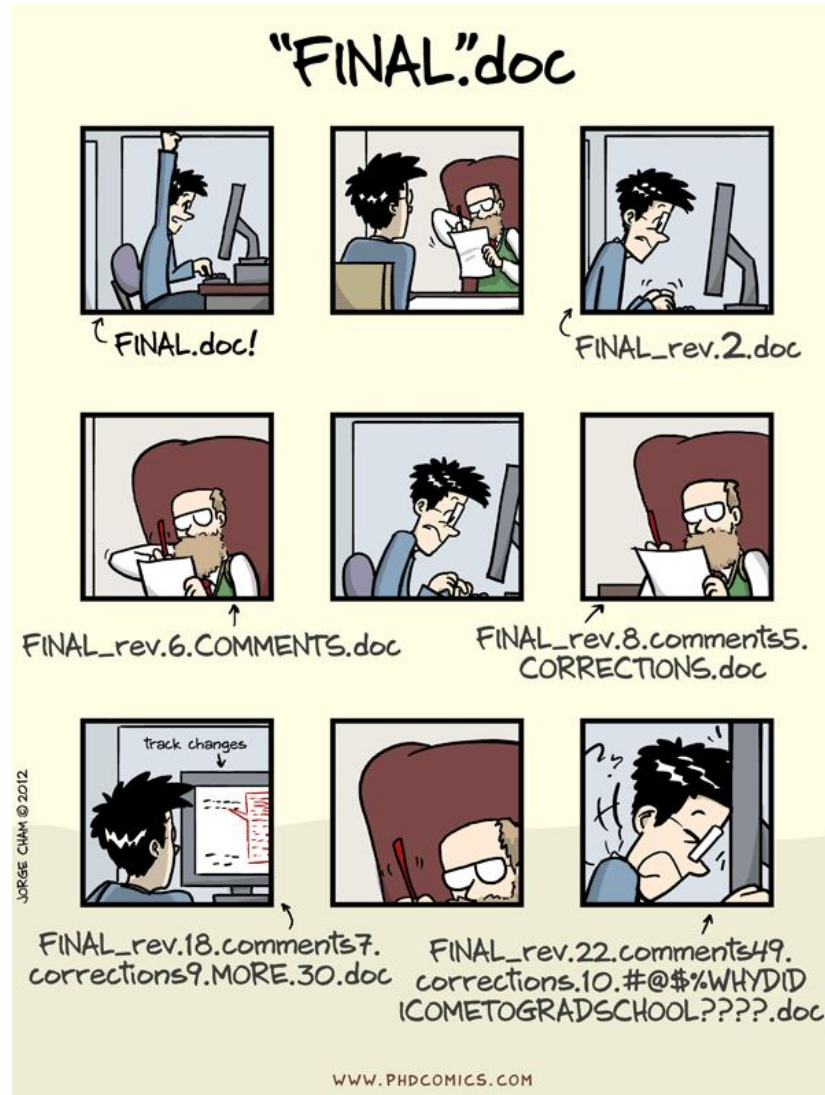
# Versionskontrolle, Git: Wieso, weshalb, warum?



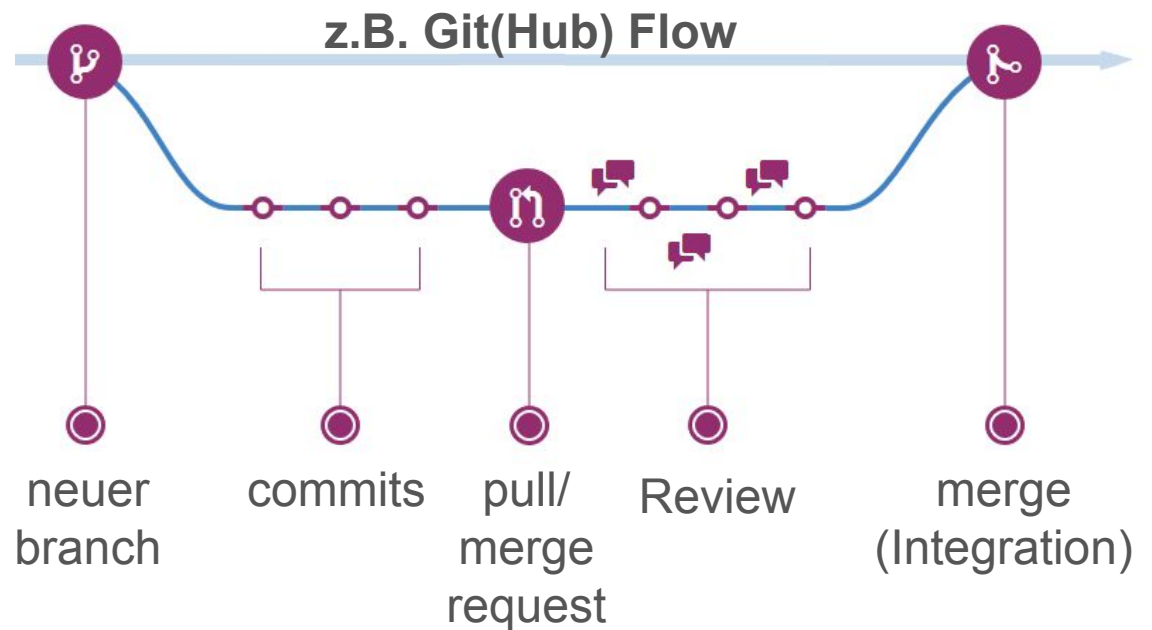
- Nachvollziehbarkeit & Durchblick
- Kommunikation via “Issues” im Projekt
- Fokus & Wissenskontinuität



# Versionskontrolle, Git: Wieso, weshalb, warum?



- Nachvollziehbarkeit (digitales Laborbuch)
- Fokus & Wissenskontinuität
- **strukturierte Zusammenarbeit, auch asynchron & offline**





# Versionskontrolle, Git: Auch darum

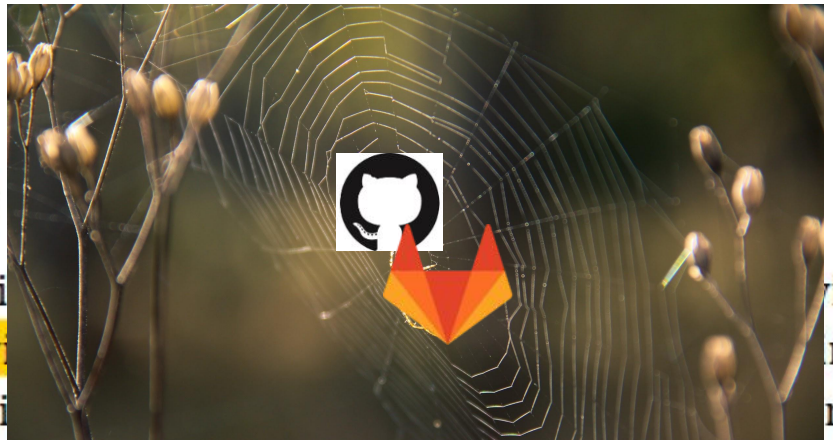
- Wilson et al. (2014) Best / Good enough (2017) ... practices for Scientific Computing [doi:10.1371/journal.pbio.1001745](https://doi.org/10.1371/journal.pbio.1001745)
- Jiménez et al. (2017) Four simple recommendations to encourage best practices in research software [doi:10.1371/journal.pcbi.1005510](https://doi.org/10.1371/journal.pcbi.1005510)
- [doi:10.12688/f1000research.11407.1](https://doi.org/10.12688/f1000research.11407.1)

Handreichung zum  
**Umgang mit  
Forschungssoftware**



## Entwickler-Verantwortung

Software-Entwickler im akademisch-technischen als auch wissenschaftlich-technisches als auch wissenschaftlich-technisches auf die Grundprinzipien Guter Wissenschaft auf die Grundprinzipien Guter Wissenschaft Best-Practice-Regeln der Qualitätskontrolle bei der Software-Entwicklung einzuhalten. Dazu zählen neben der Veröffentlichung des Quellcodes auch die Nutzung von Versionsverwaltungssystemen, eine umfassende Dokumentation des Codes, die Nutzung entsprechender Infrastruktursoftware für das Qualitätsmanagement, standardisierte Tests sowie idealerweise die unabhängige Begutachtung des Quellcodes.



# Versionskontrolle, Git: Was tut es? Durchblick!

- normale Datei- & Ordneransicht
- History (`git log`) & Blame
- Detailansicht: `git show ...`
- Zeitmaschine: `git checkout / bisect ...`



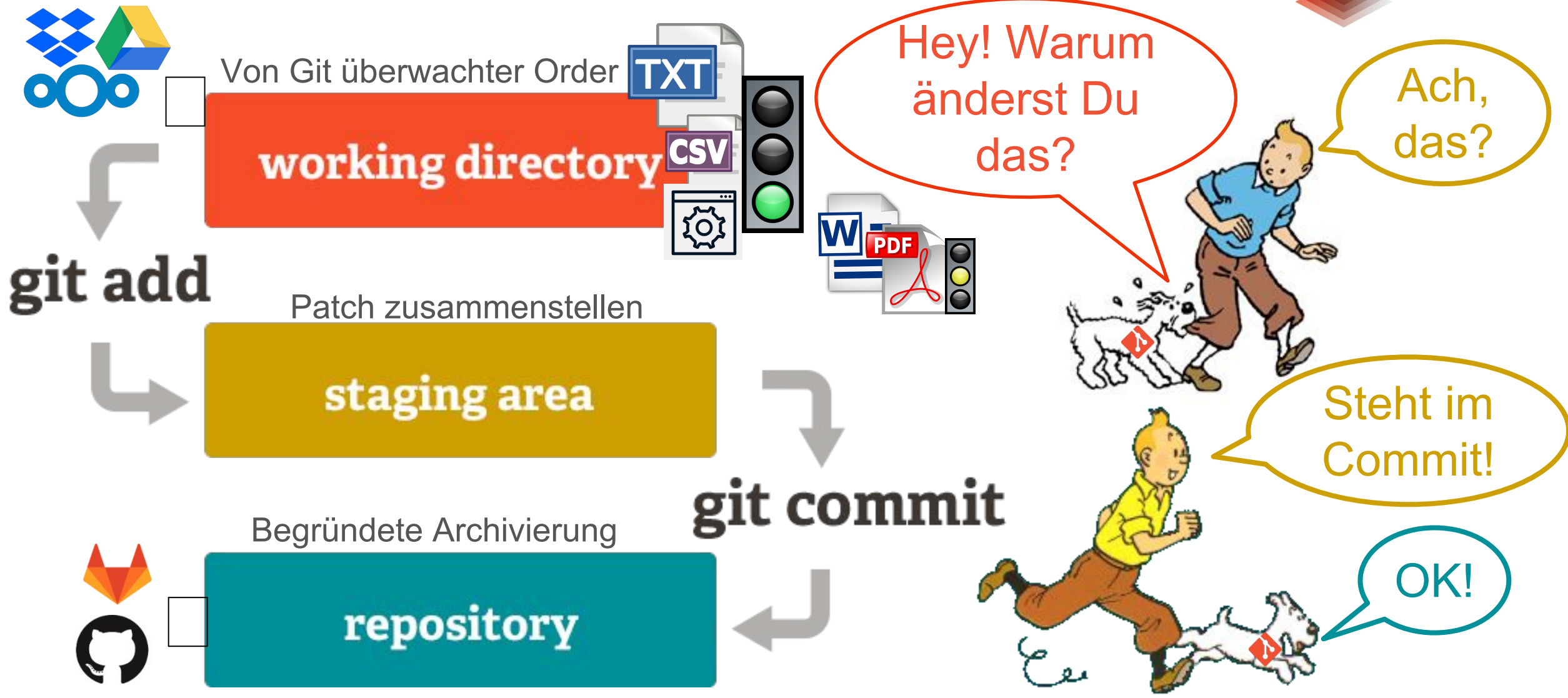
[pxhere.com/en/photo/757871](http://pxhere.com/en/photo/757871)  
CC0 Public domain



By Chad Anderson, staff photographer for SFO Museum [CC BY-SA 2.0](https://creativecommons.org/licenses/by-sa/2.0/)  
via [commons.Wikimedia.org/w/index.php?curid=45625745](https://commons.wikimedia.org/w/index.php?curid=45625745) (cropped)



# Versionskontrolle, Git: Was tut es? Modify, Add, Commit!



## DEMO

1. Inkonsistenter Datensatz: BacDiveR.rproj > vignette > [logic-checking-bacdive-datasets.Rmd](#)
2. Nachvollziehbarkeit der Ausarbeitung des Dokuments: Commit-History & Blame-Ansicht
3. Modify-Add-Commit-Zyklus

# DEMO

## Example of a data inconsistency

Just as the correctness of data analysis code should be tested automatically, the consistency of data should be evaluated and monitored as well. Using [BacDive's advanced search](#) and [BacDiveR's retrieve\\_search\\_results\(\)](#) several examples of geographic inconsistencies have been found. Presumably due to an overly strict location-to-country-to-continent mapping, several samples collected from seas neighbouring Russia (like the [Sea of Japan](#)), were assigned to Europe.

species name, culture col. no., sequence r  
[API® test finder](#)
[TAXplorer](#)
[Advanced search](#)
[Isolation sources](#)
[Download](#)

**Filters**

Geographic location (country and/or sea, region): contains "Sea of Japan" ✕

Continent: "Europe" ✕

**Advanced search**

expand / minimize all ▼ NOT

Name and taxonomic classification			
Morphology and physiology			
Culture and growth conditions			
Isolation, sampling and environmental information			
Sample type/isolated from	contain ▼	insert * to show all	<input type="checkbox"/>
Geographic location (country and/or sea, region)	contain ▼	Sea of Japan	<input type="checkbox"/>
Country	contain ▼	insert * to show all	<input type="checkbox"/>

hits: 2

Select for download

[de-/select all displayed strains](#)

- [Pseudoalteromonas ruthenica Ivanova et al. 2006](#)  
 type strain: ✓  
 strain number: DSM 15912, KMM 300  
 Geographic location (country and/or sea, region): Sea of Japan  
 Continent: Europe
- [Salegentibacter flavus Ivanova et al. 2006](#)  
 type strain: ✓  
 strain designation: Fg 69  
 strain number: DSM 22702, CIP 107843, KMM 6000  
 Geographic location (country and/or sea, region): Sea of Japan  
 Continent: Europe



## DEMO

While one may debate where exactly in Russia the Asian-European border is located, it is clear that its Eastern shoreline is located well within Asia. These and other datasets with East Russian locations have been reported to the BacDive team and a portion of those was corrected in [BacDive's 04.07.2018 release](#).

```
library(BacDiveR)

mismatched <- list(location = "Japan", continent = "Europe")

inconsistent_data <- retrieve_search_results(
  paste0(
    "https://bacdive.dsmz.de/advsearch?advsearch=search&site=advsearch",
    "&searchparams[20][typecontent]=contains",
    "&searchparams[20][searchterm]", mismatched$location,
    "&searchparams[17][searchterm]", mismatched$continent
  )
)
#> Warning: 'retrieve_search_results' is deprecated.
#> Use 'bd_retrieve_by_search' instead.
#> See help("Deprecated")
#> Downloading BacDive IDs 131115 to 139987 (but not necessarily contiguously):
#> 131115 139987
```

As long as this specific inconsistency is not corrected in the datasets, some should be found and downloaded here.

## DEMO

While one may debate where exactly in Russia the Asian-European border is located, it is clear that its Eastern shoreline is located well within Asia. These and other datasets with East Russian locations have been reported to the BacDive team and a portion of those was corrected in [BacDive's 04.07.2018 release](#).

```
library(BacDiveR)

mismatched <- list(location = "Japan", continent = "Europe")

inconsistent_data <- retrieve_search_results(
  paste0(
    "https://bacdive.dsmz.de/advsearch?advsearch=search&site=advsearch",
    "&searchparams[20][typecontent]=contains",
    "&searchparams[20][searchterm]", mismatched$location,
    "&searchparams[17][searchterm]", mismatched$continent
  )
)
#> Warning: 'retrieve_search_results' is deprecated.
#> Use 'bd_retrieve_by_search' instead.
#> See help("Deprecated")
#> Downloading BacDive IDs 131115 to 139987 (but not necessarily contiguously):
#> 131115 139987
```

As long as this specific inconsistency is not corrected in the datasets, some should be found and downloaded here.

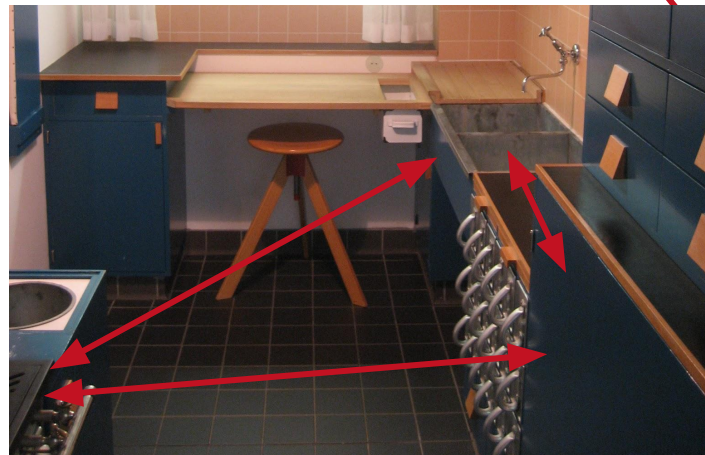
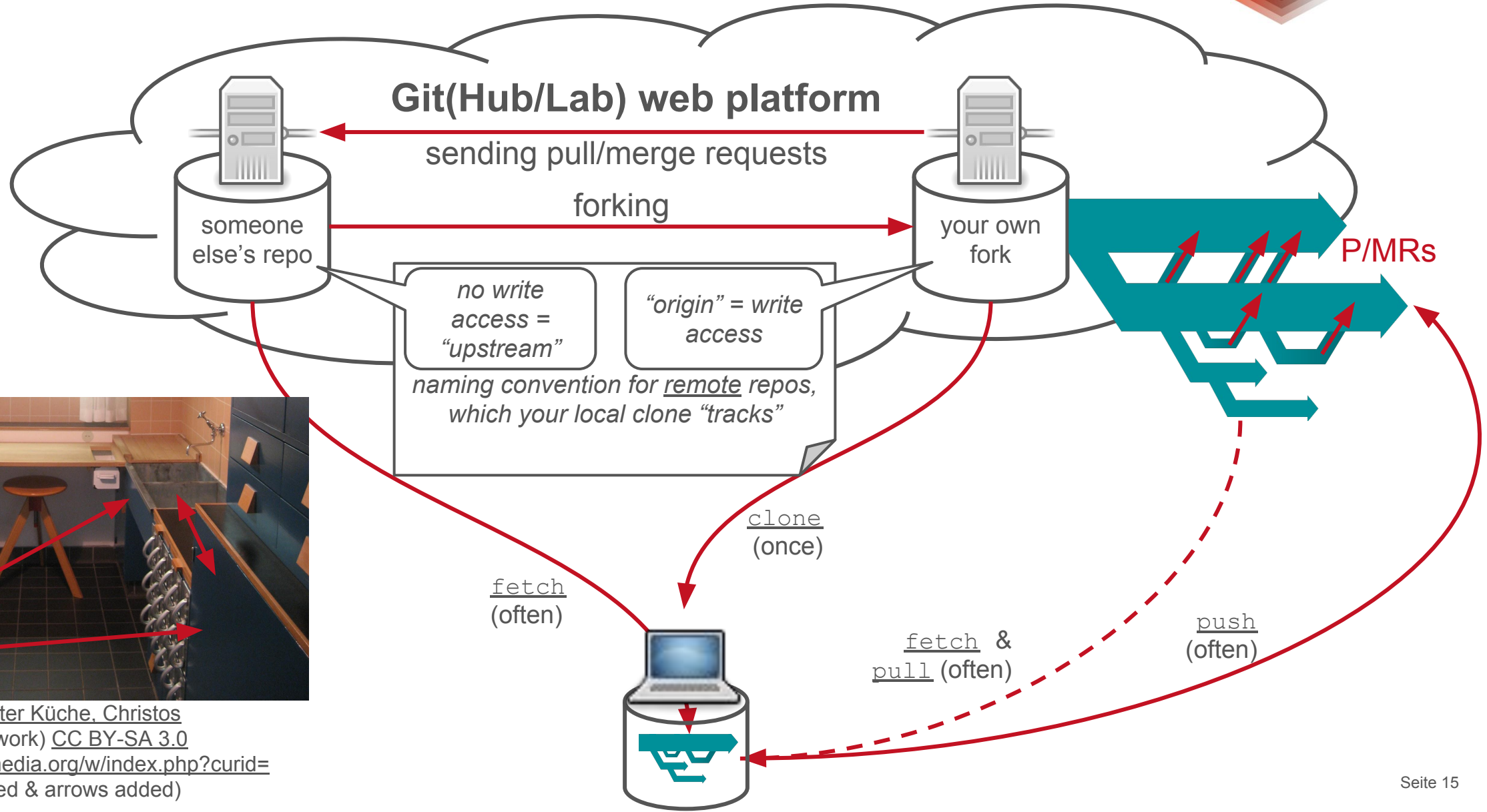
## How to test datasets

If a BacDive user finds an inconsistency within the datasets they use, BacDiveR's `retrieve_search_results()`





# Git workflow

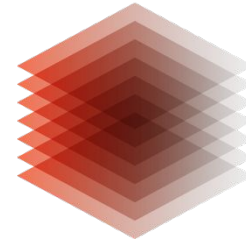


8linden Frankfurter Küche, Christos Vittoratos (own work) CC BY-SA 3.0  
[Commons.wikimedia.org/w/index.php?curid=4201238](https://commons.wikimedia.org/w/index.php?curid=4201238) (cropped & arrows added)

## Ressourcen

- TIBs FAIR Data & Software-”Aufnahmen: [av.TIB.eu/series/530](https://av.tib.eu/series/530)
  - Findability via GitHub & Zenodo: [av.TIB.eu/media/37823#t=01:08:45](https://av.tib.eu/media/37823#t=01:08:45) (40min)
  - Accessibility via GitHub, plus Workflow-Demo: [av.TIB.eu/media/37825#t=12:15](https://av.tib.eu/media/37825#t=12:15) (90min)
- [GitHub.com/topics/research-data-management](https://github.com/topics/research-data-management) & [/data-management](https://github.com/topics/data-management)
- auch für große Daten: [git-LFS.GitHub.io](https://git-lfs.github.io)
  
- Dokumentation der Git-Kommandozeile: [git-SCM.com/doc](https://git-scm.com/doc)
- Anleitungen für die Web-Interfaces: [guides.GitHub.com](https://guides.github.com) & [docs.GitLab.com/ce/user](https://docs.gitlab.com/ce/user)
- interaktive Tutorials: [lab.GitHub.com](https://lab.github.com)
- Videos von [GitHub](https://github.com) & [GitLab](https://gitlab.com)

LEIBNIZ-INFORMATIONSZENTRUM  
TECHNIK UND NATURWISSENSCHAFTEN  
UNIVERSITÄTSBIBLIOTHEK



TIB

**Danke für Ihre Aufmerksamkeit!**  
**Folien: [is.gd/TB3K6f](https://is.gd/TB3K6f)**

**[Katrin.Leinweber@TIB.eu](mailto:Katrin.Leinweber@TIB.eu) oder 0511 762 14693**  
**Git{Hub|Lab}.com: @KatrinLeinweber**



Creative Commons Namensnennung 3.0 Deutschland  
<http://creativecommons.org/licenses/by/3.0/de>