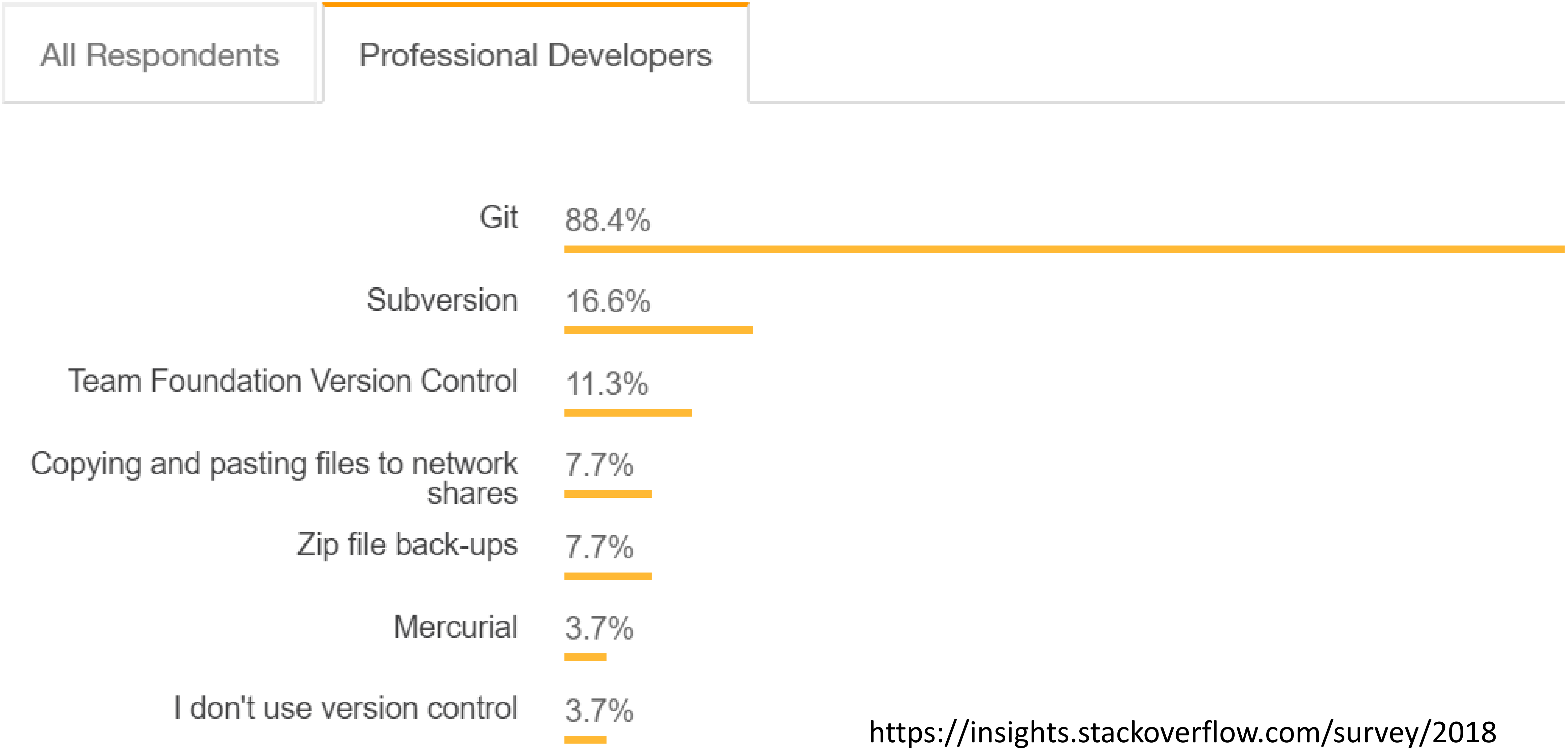
Fachhochschule Kiel
University of Applied Sciences

Tools and Programming for Data Science
Git and GitHub

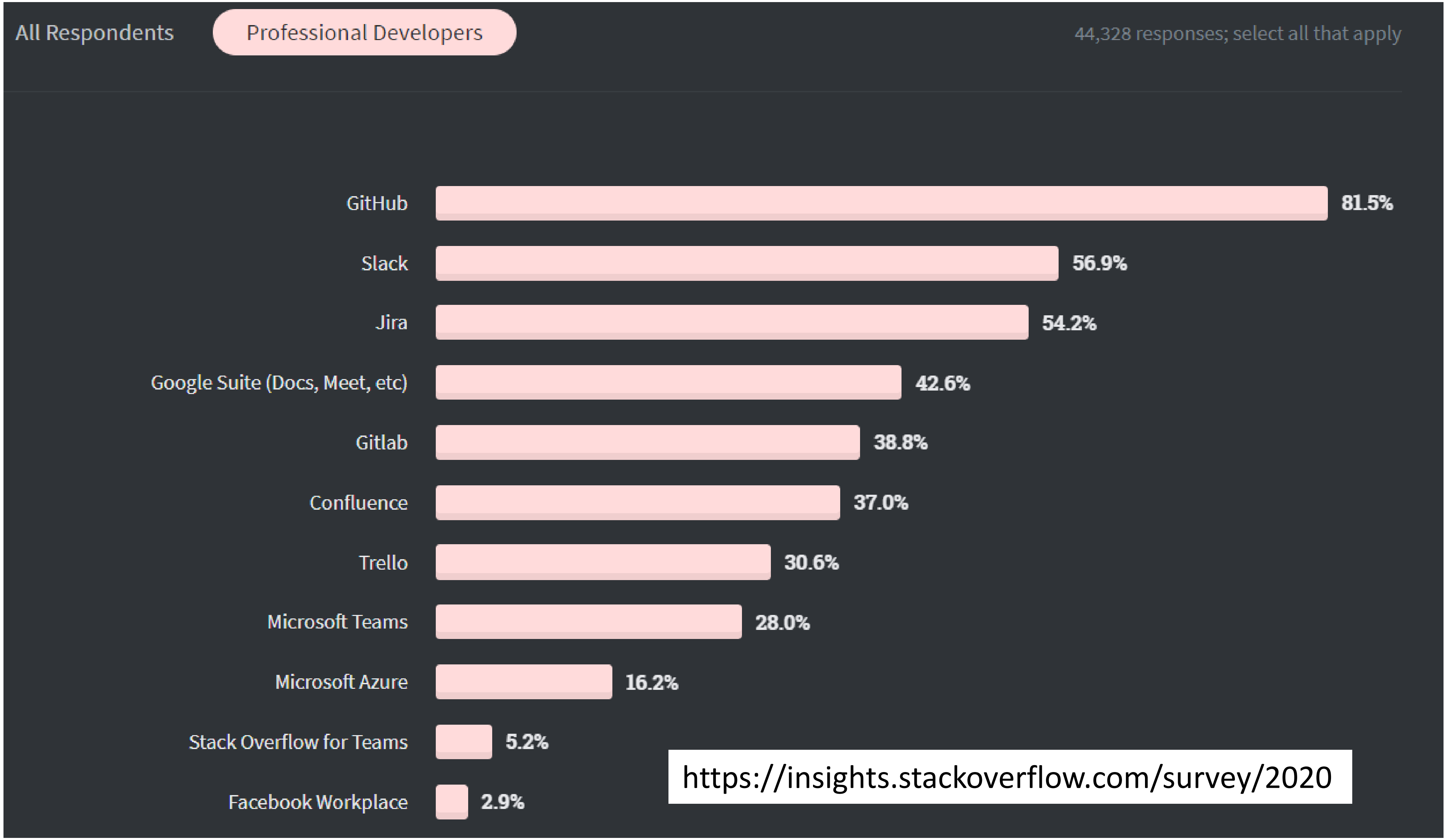
Study Program Data Science
Prof. Dr. Tillmann Schwörer

Version control systems



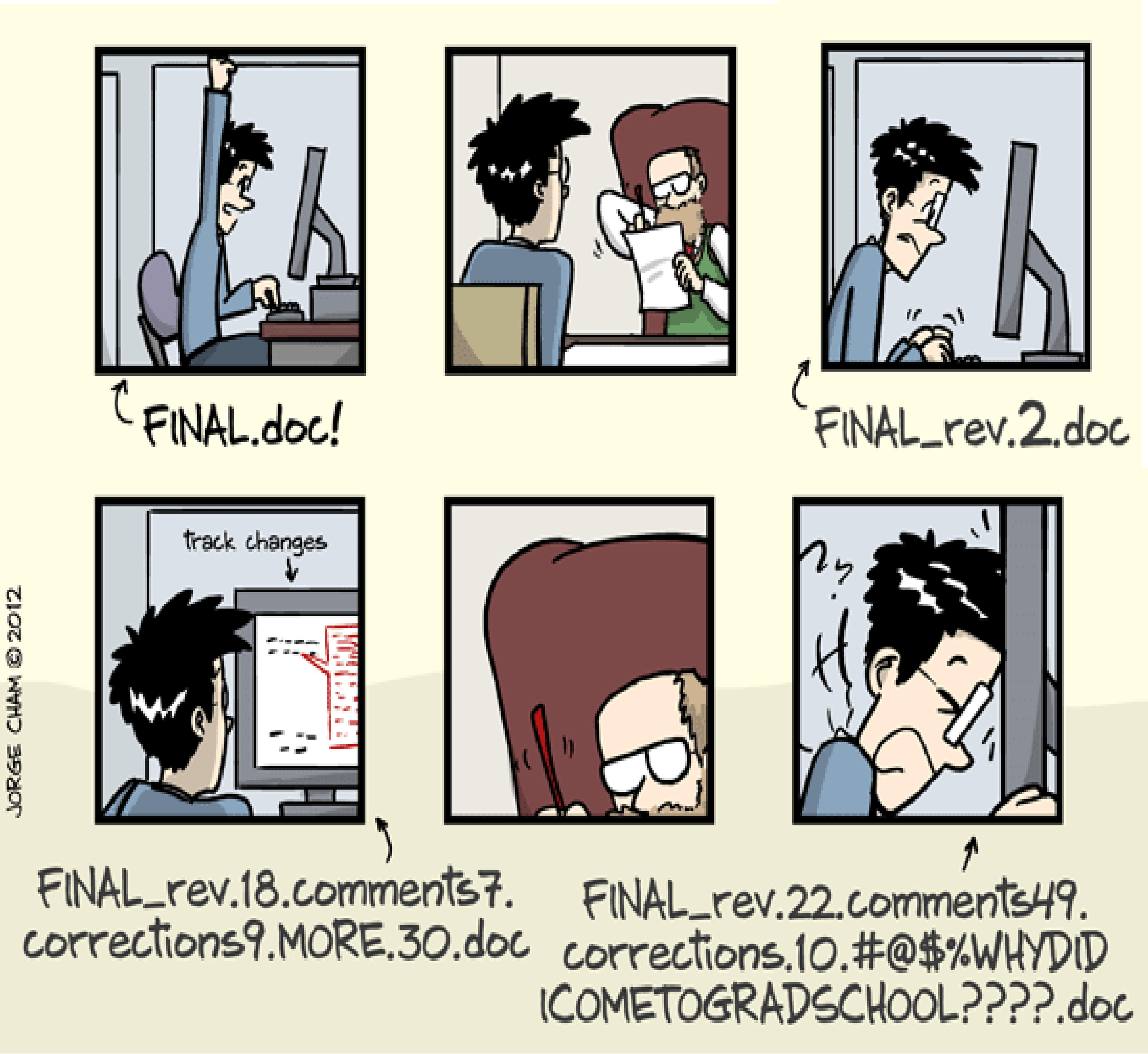
https://insights.stackoverflow.com/survey/2018

Top Collaboration Tools



https://insights.stackoverflow.com/survey/2020

Why Git and GitHub?

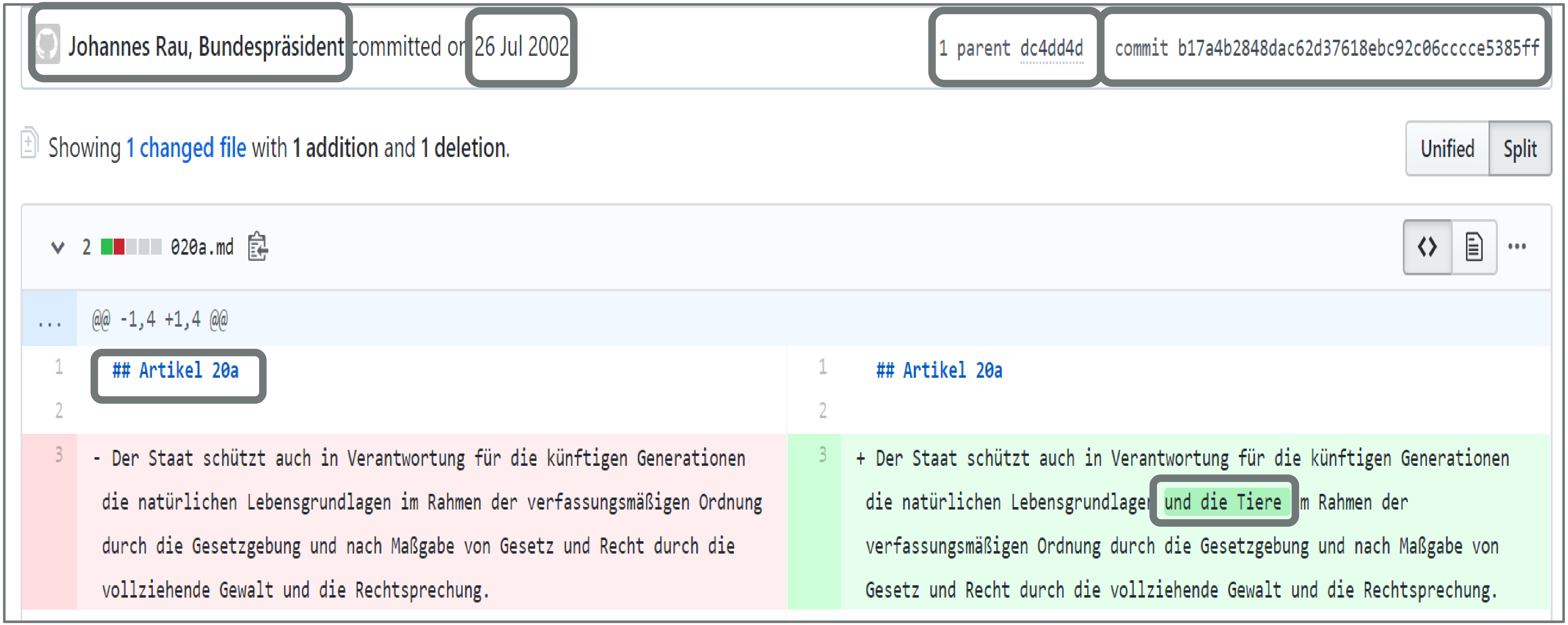

FINAL.doc!
FINAL_rev.2.doc
track changes
FINAL_rev.18.comments7.corrections9.MORE.30.doc
FINAL_rev.22.comments49.corrections.10.#@$%WHYDIDICOMETOGRADSCHOOL????.doc
JORGE CHAM © 2012

Why Git and GitHub?

- **Backup**: undo changes, restore files, safely experiment
- **Transparency:** what was changed? by whom? when?
- **Collaboration:** work simultaneously with coauthors on the same project
- **Job applications:** showcase your version control and data science skills through your own GitHub repository

Example

Chaos Computer Club: Entdecke unsere Verfassung



Johannes Rau, Bundespräsident committed on 26 Jul 2002

1 parent dc4dd4d

commit b17a4b2848dac62d37618ebc92c06cccce5385ff

Showing 1 changed file with 1 addition and 1 deletion.

2 020a.md

```
@@ -1,4 +1,4 @@
```

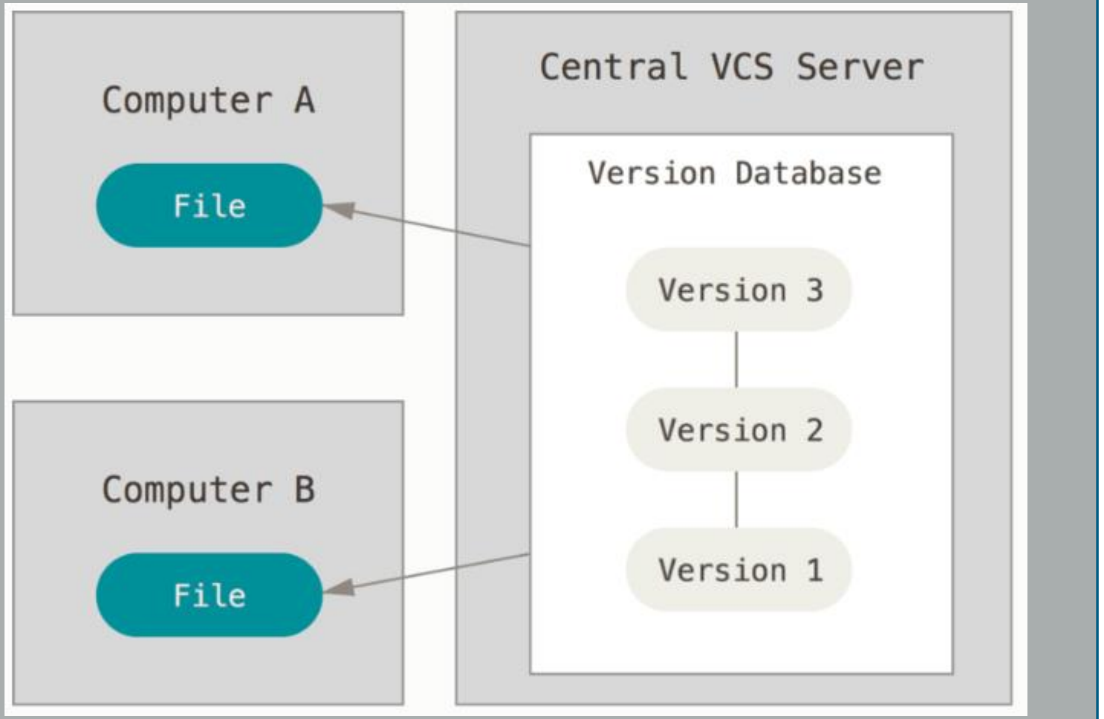
```
1  ## Artikel 20a
2
3  - Der Staat schützt auch in Verantwortung für die künftigen Generationen
     die natürlichen Lebensgrundlagen im Rahmen der verfassungsmäßigen Ordnung
     durch die Gesetzgebung und nach Maßgabe von Gesetz und Recht durch die
     vollziehende Gewalt und die Rechtsprechung.
```

```
1  ## Artikel 20a
2
3  + Der Staat schützt auch in Verantwortung für die künftigen Generationen
     die natürlichen Lebensgrundlagen und die Tiere im Rahmen der
     verfassungsmäßigen Ordnung durch die Gesetzgebung und nach Maßgabe von
     Gesetz und Recht durch die vollziehende Gewalt und die Rechtsprechung.
```

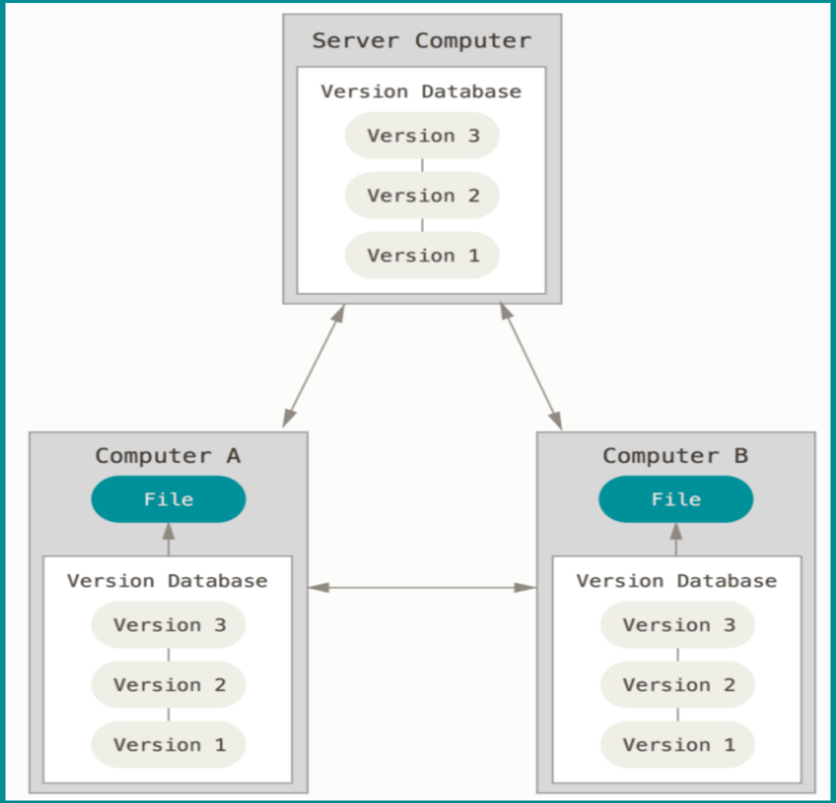
Clarifying terms

- **Version control system** (VCS): keeps track and manages changes in your / your team's documents
- **Git:** an open source, distributed version control system (VCS)
- **GitHub:** a platform for hosting and collaborating on Git repositories

Version control systems

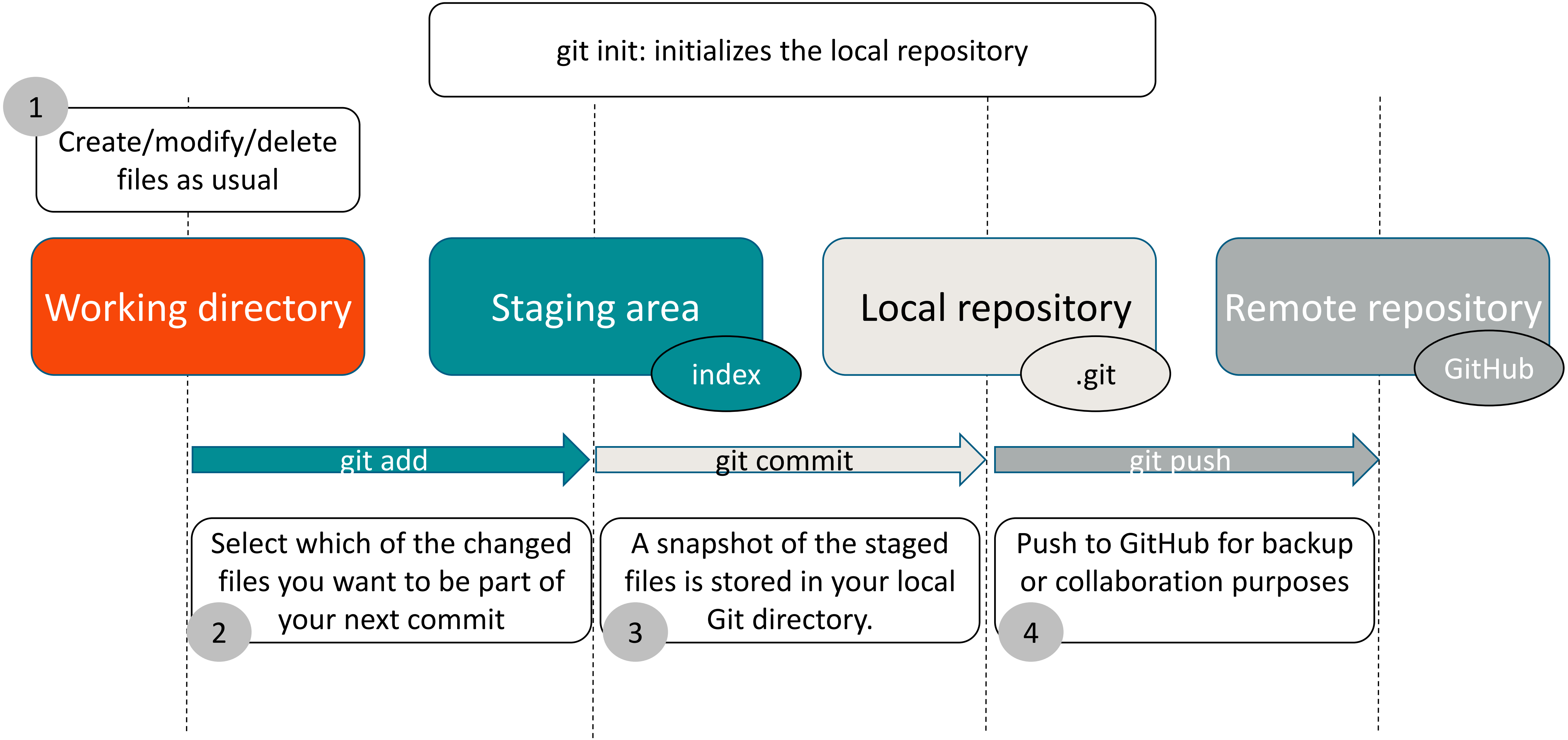


Centralized: single files are pulled and pushed from/to central repository (e.g. **Subversion**)



Distributed: Each client has full copy of entire repository (→ **git**)

Important concepts / commands



git init: initializes the local repository

1 Create/modify/delete files as usual

Working directory

Staging area

index

Local repository

.git

Remote repository

GitHub

git add

git commit

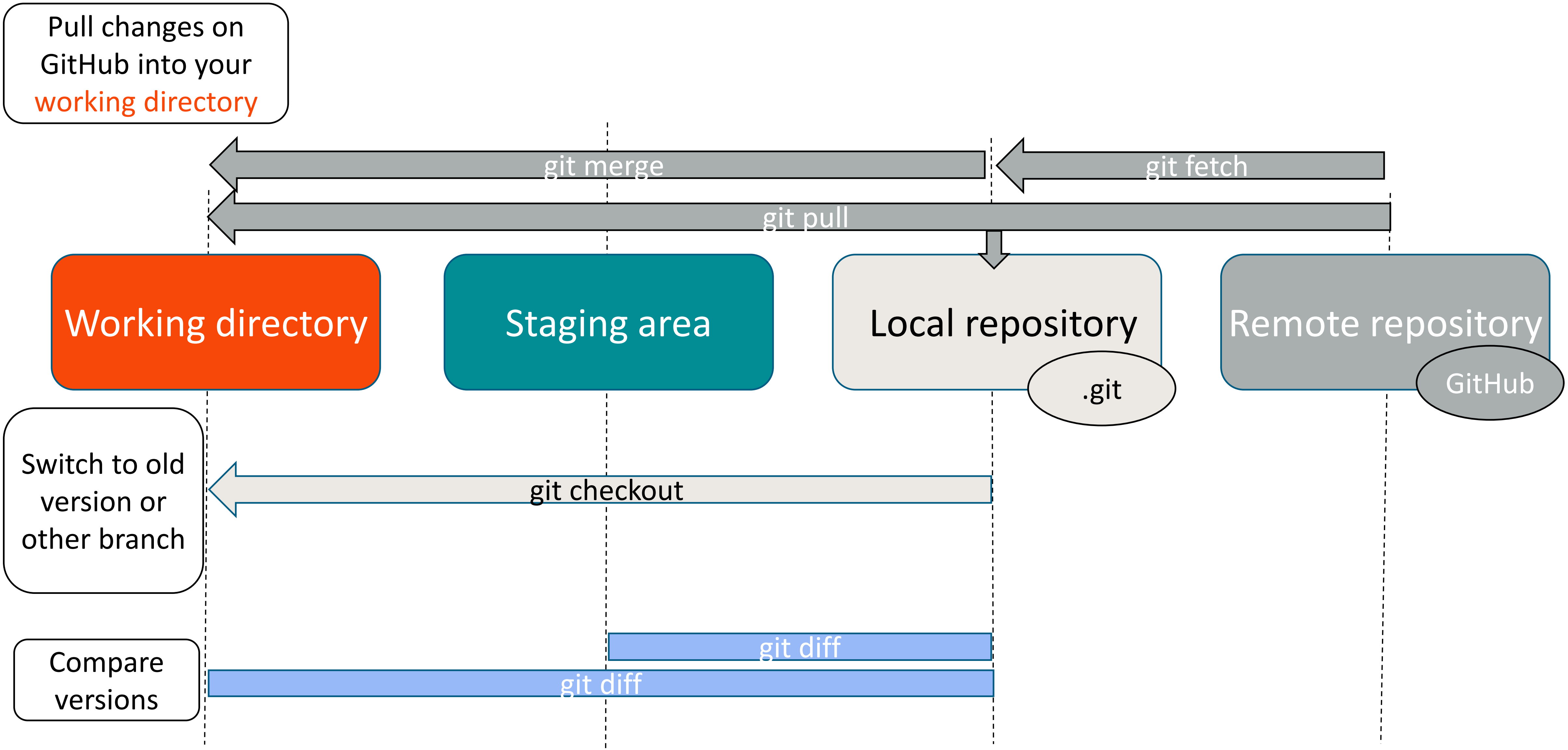
git push

2 Select which of the changed files you want to be part of your next commit

3 A snapshot of the staged files is stored in your local Git directory.

4 Push to GitHub for backup or collaboration purposes

Important concepts / commands



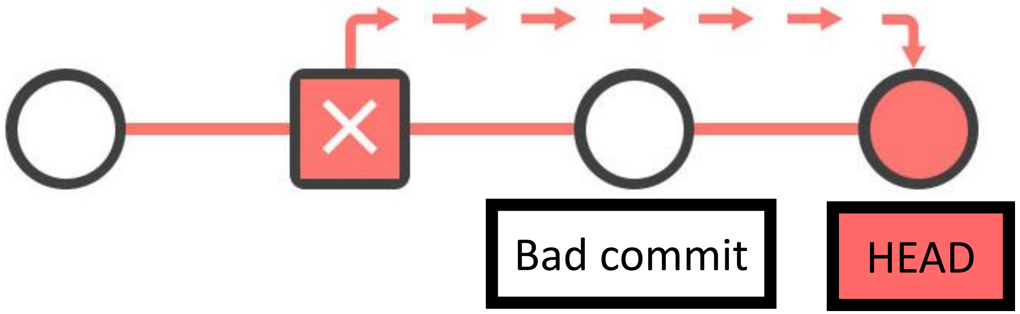
Undo commits

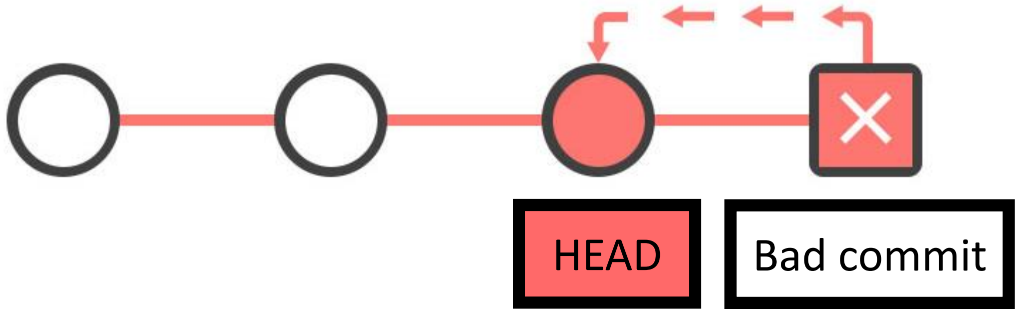
Revert (*#bad-commit*)

- Adds a new commit which reverses the „bad commits“
- Leaves the commit history intact
- Can be safely done, even if you have already pushed to GitHub

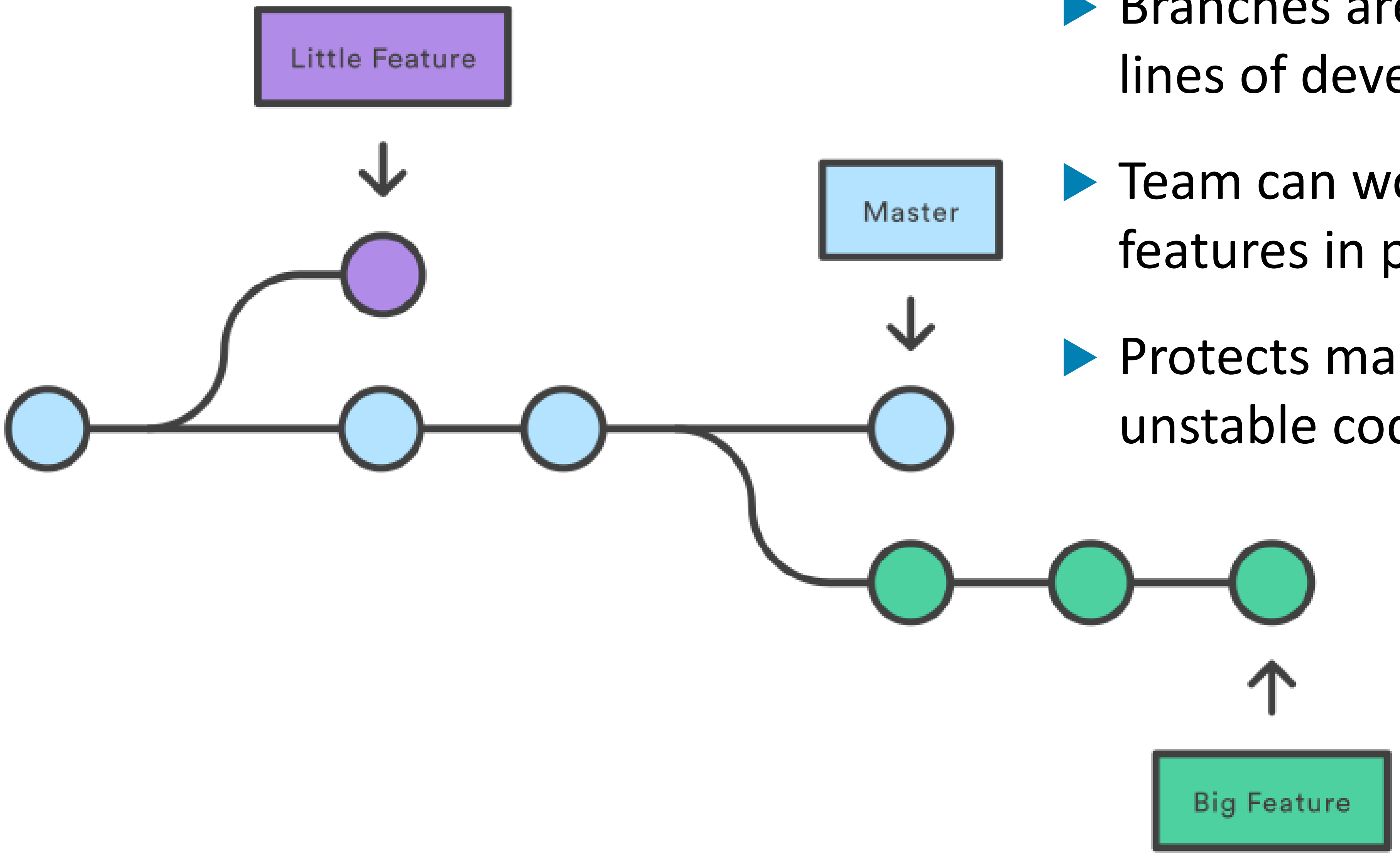
Reset (*#last-good-commit*)

- Discard the „bad commit“
  - Mixed (default): moves HEAD to the desired old commit, but keeps changes in working directory
  - Hard: entirely resets to desired old commit (danger!)
- Use Reset only for local changes!



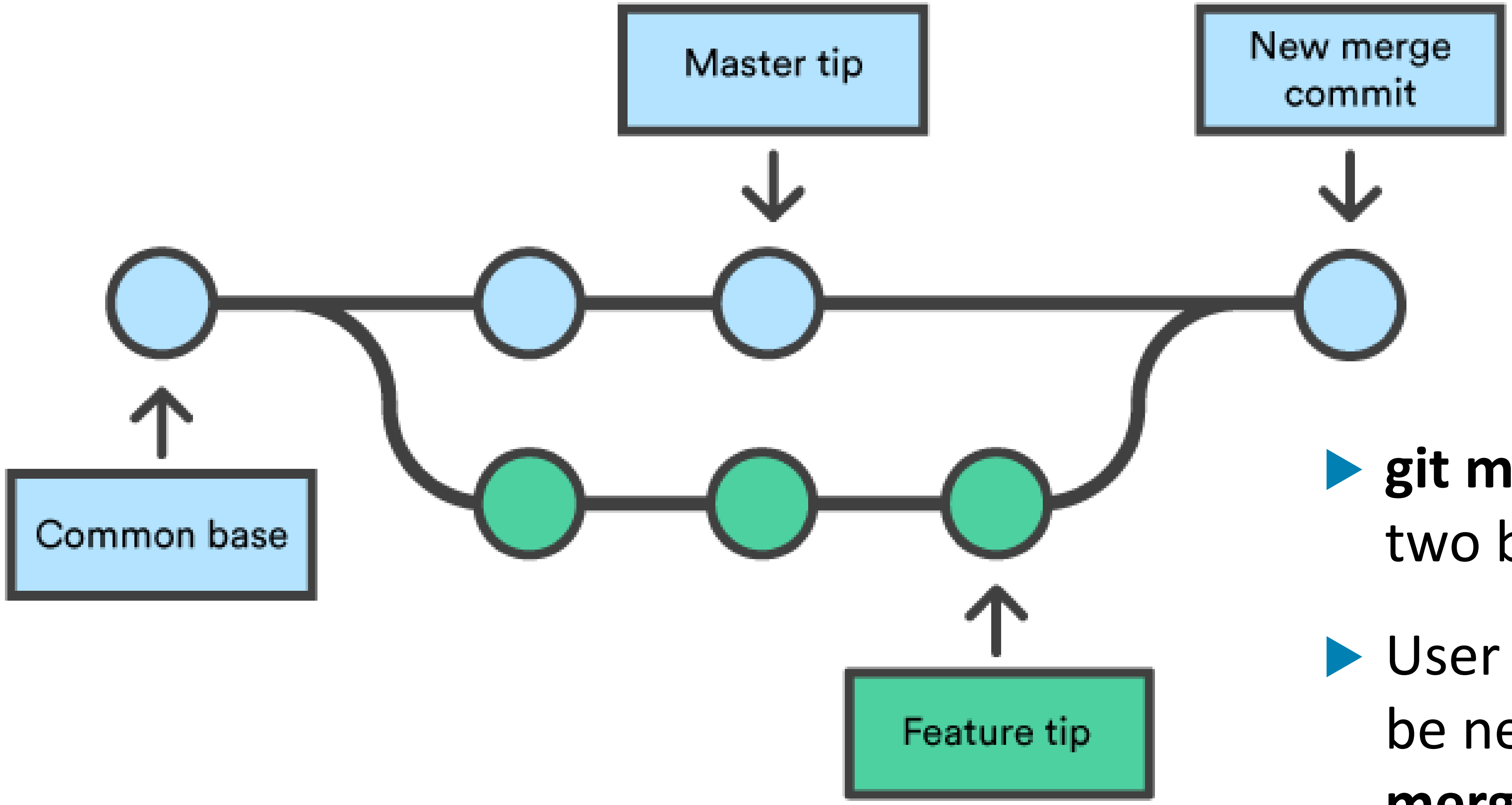


Branching



- Branches are independent lines of development
- Team can work on separate features in parallel
- Protects main branch against unstable code

Merging



- **git merge** combines two branches into one.
- User intervention may be needed to resolve **merge conflicts**

Git Workflows

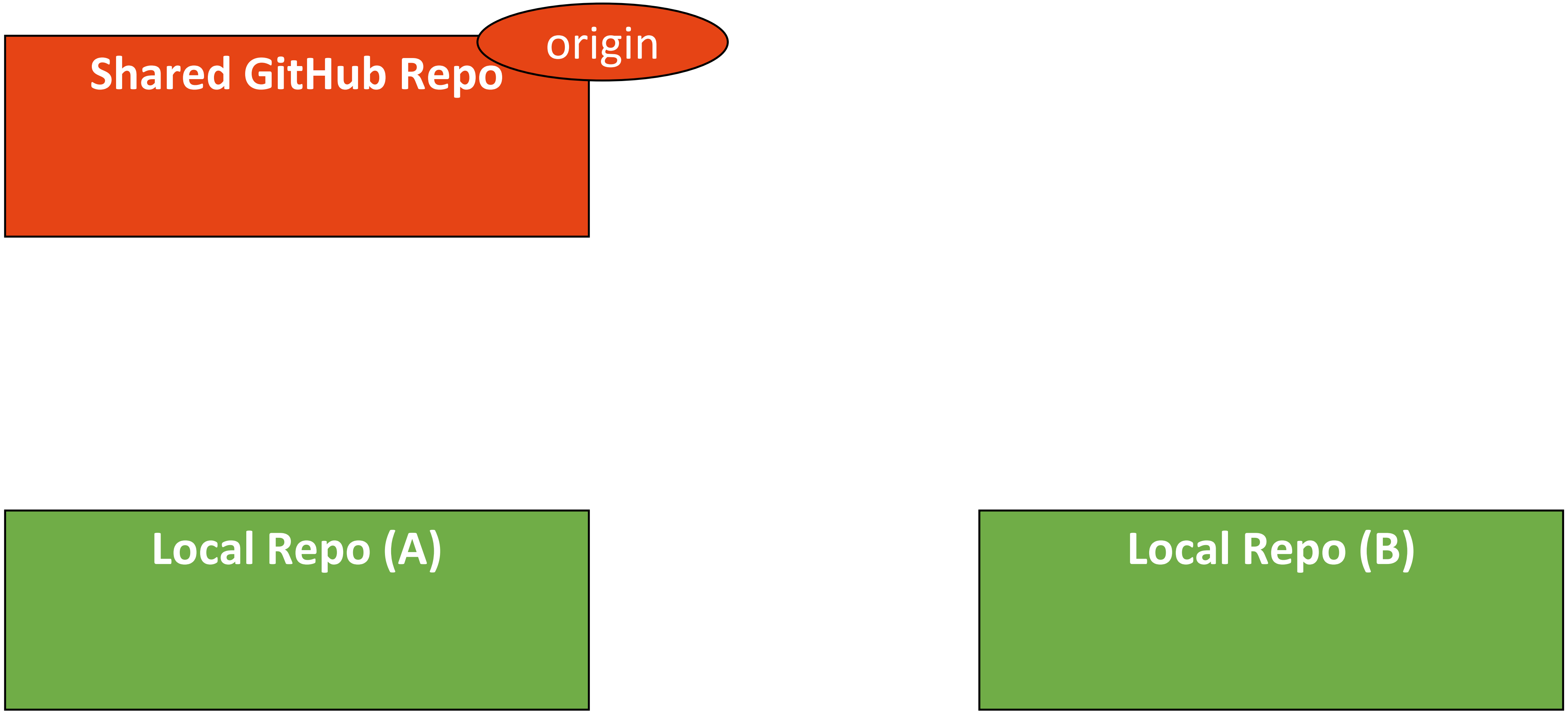
There are several Git workflows: recipes for how to use Git productively.

Feature Branch Workflow:

1. Add, commit andWork on feature in a **dedicated branch**
2. **Push to GitHub repository**
3. Open **pull request** on GitHub
4. Team reviews code and fixes are pushed into the same branch
5. **Merge** into main branch
6. (Feature branch can be deleted both locally and remotely)

Feature Branch Workflow

Company Setting



Shared GitHub Repo

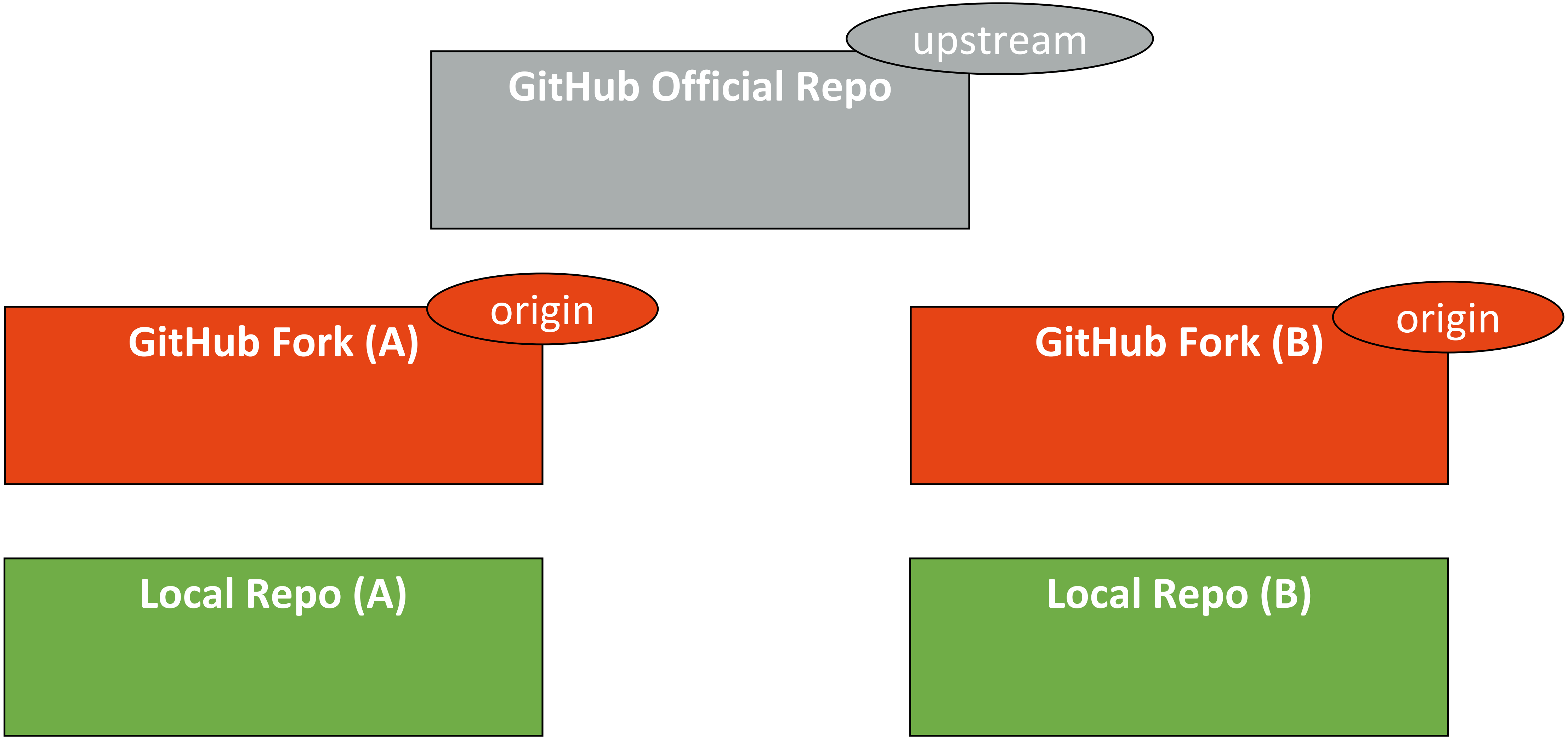
origin

Local Repo (A)

Local Repo (B)

Forking Workflow

Open Source Setting



Further Resources

Fachhochschule Kiel
University of Applied Sciences

- Tutorial on Git (Workflows) (available in German)
- Git reference book (available in German)
- Git commands: cheat sheet
- 10-minute reads mostly on GitHub topics (by GitHub)
- Glossary of Git and GitHub terms (by Github)
- Tutorial for R users