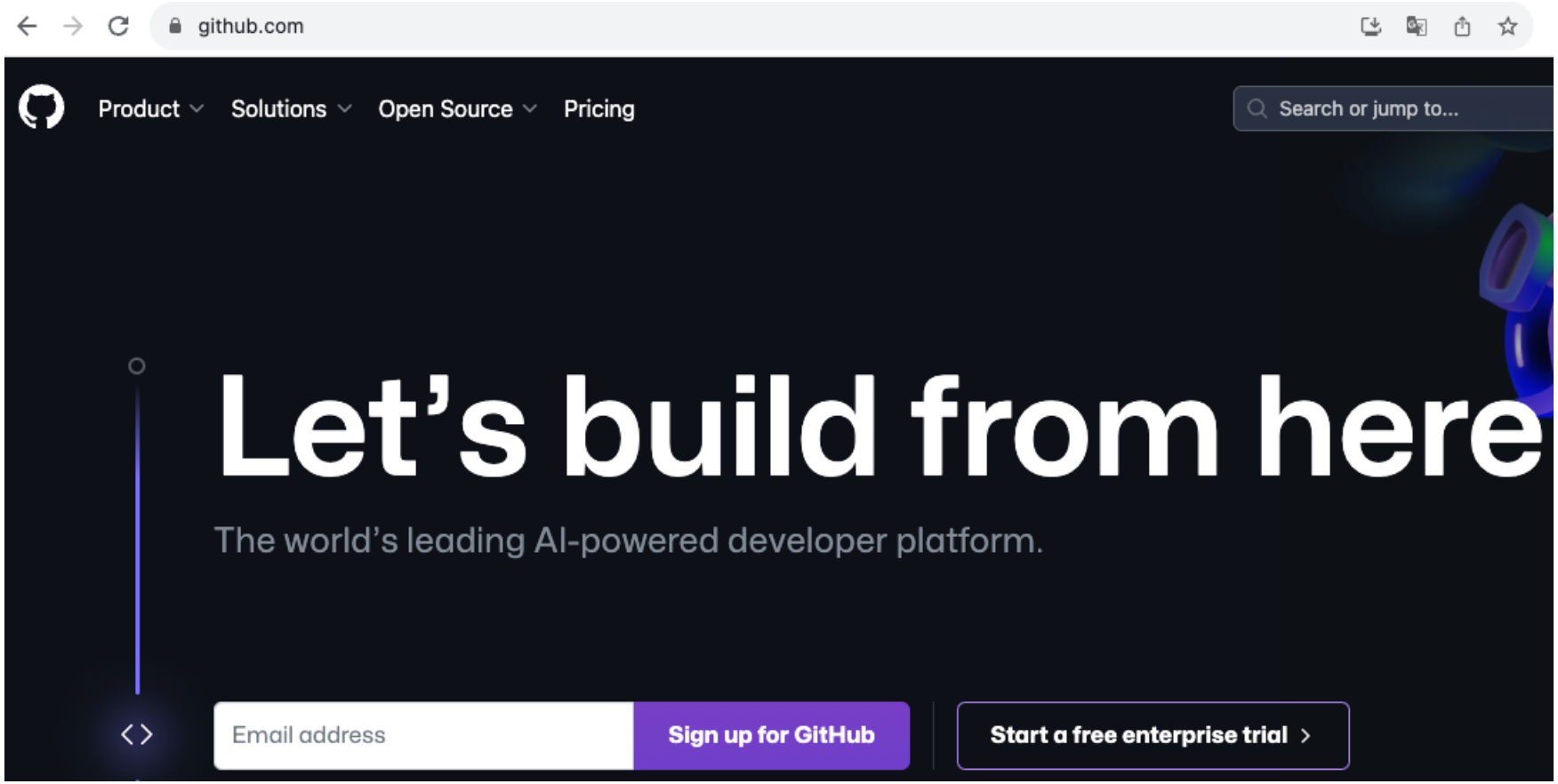
GitHub for Research (Data)



Dr. Renat Shigapov

26.03.2024

Agenda

1. **Introduction**
2. Demo: Repository „ReproResearch“
3. Conclusions

Why „GitHub for research (data)“?

FDZ FORSCHUNGSDATENZENTRUM UB MANNHEIM | BERD @NFDI

- Managing your research (data) and automating your research
  - Ensuring transparency and reproducibility of your research
  - Empowering collaboration across countries and disciplines

Research data

What is Data?
Data are objects that you use and produce during your research life cycle, encompassing datasets, software, code, workflow, models, figures, tables, images and videos, interviews, ~~articles~~.
Data are your research asset.

- **Data is not only a dataset**
- **Data is everything what makes your research (research articles) reproducible and replicable**

"The Turing Way: A handbook for reproducible, ethical and collaborative research"
https://the-turing-way.netlify.app/reproducible-research/rdm/rdm-data.html

Research and data in research

Two-step research:

1. Conducting research

- Data collection, creation or reuse (**dataset**)
- Data processing and analysis (**software, models, algorithms, methods, protocols, etc.**)

2. Publishing or perishing

- Publishing an article in a top-tier journal

Research data and code policies of a journal => Data and code availability statements

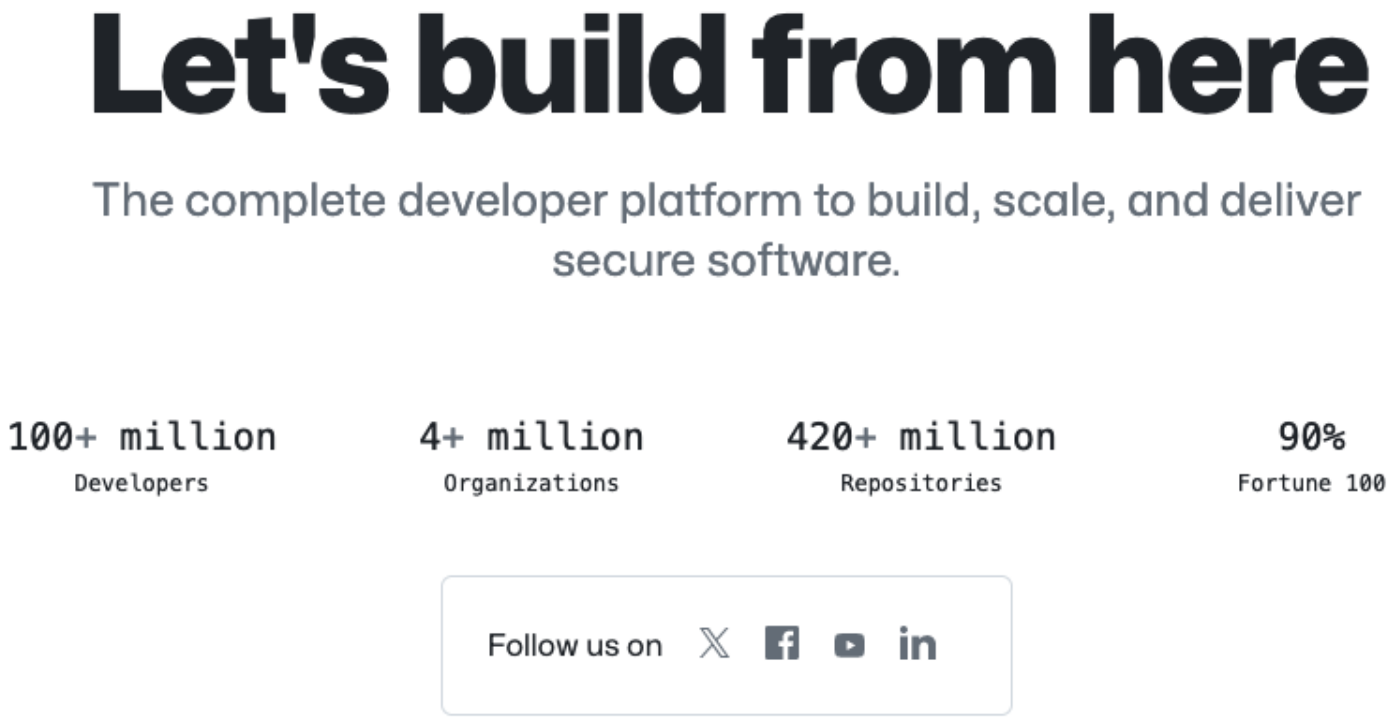
<=> Replication package

- Perishing without any research outputs

Research is not reproducible and not replicable

<=> No replication package

GitHub



A version control and collaboration platform

- Git-based version control
- Pull requests and issue tracking
- GitHub Actions for automated workflows
- GitHub Pages for static website hosting
- Wiki for repositories
- Project management tools

The main concepts at GitHub

FDZ FORSCHUNGSDATENZENTRUM UB MANNHEIM | BERD @NFDI

Repository (Repo): A folder where your project lives. It can be public or private.

Branch: A parallel version of a repository.

Commit: An individual change to a file or files.

Pull request: A method of submitting contributions to a project.

Merge: The process of taking the changes from one branch and integrating them into another.

Fork: A copy of a repository that you manage on your account.

Issue: A way to track of tasks, enhancements, and bugs for your projects.

Clone: A copy of a repository that lives on your computer, and the act of making that copy.

Agenda

1. Introduction
2. **Demo: Repository „ReproResearch“**
3. Conclusions

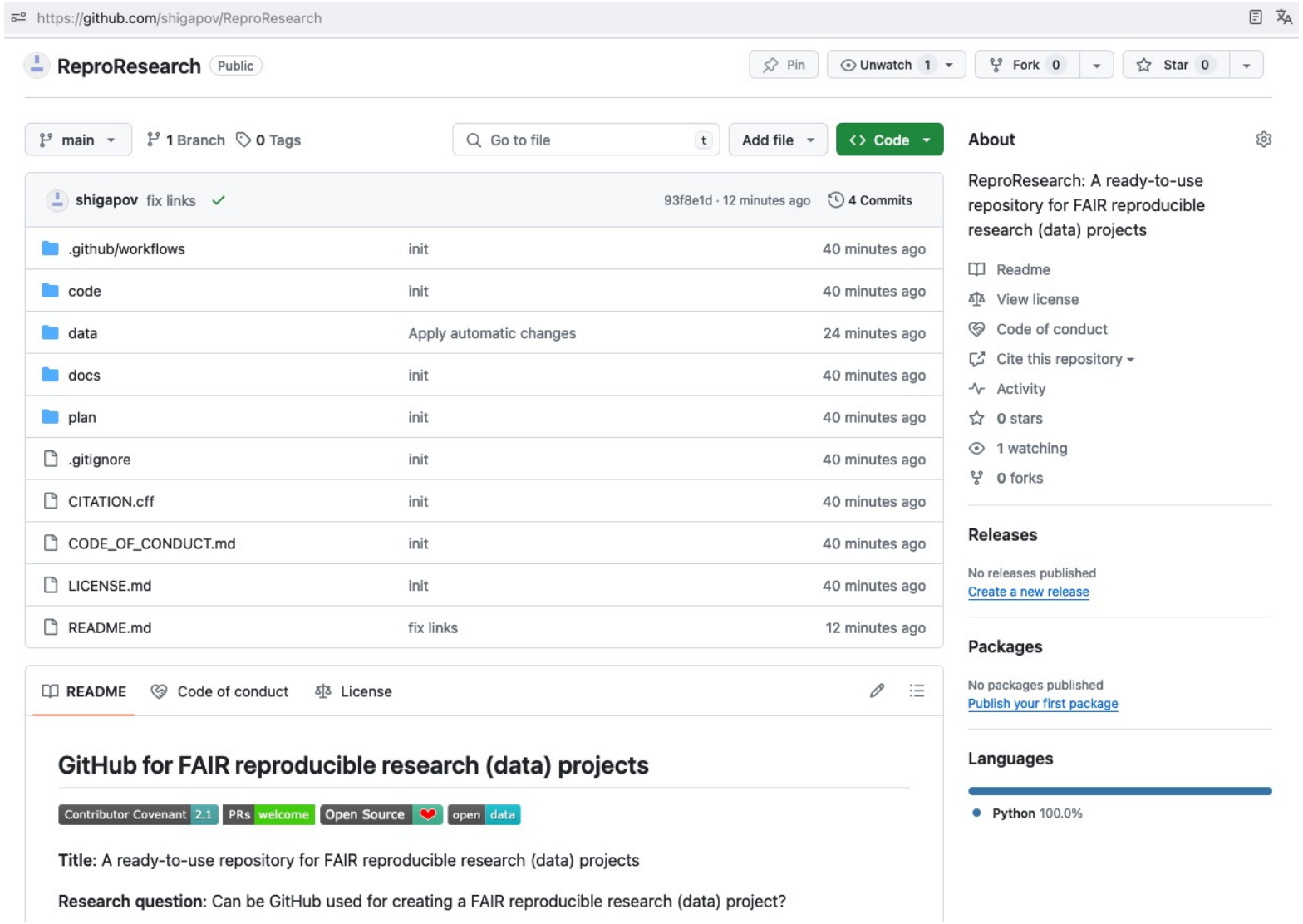
A ready-to-use repo for your research project

FDZ FORSCHUNGSDATENZENTRUM UB MANNHEIM | BERD @NFDI

https://github.com/shigapov/ReproResearch

ReproResearch

- A ready-to-use or -fork or -clone
- A template for creating a replication package
- GitHub Actions for automated reproducible analysis
- GitHub Pages for static website hosting
- Wiki for repositories
- Project management board



Agenda

1. Introduction
2. Demo: Repository „ReproResearch“
3. **Conclusions**

What we have:

- A ready-to-use GitHub repository structured according to the best practices in research data management: https://github.com/shigapov/ReproResearch
- An automatic workflow using GitHub Actions for processing raw data, saving it, and making commit and push: https://github.com/shigapov/ReproResearch/tree/main/.github/workflows
- A project website using the Minimal Jekyll theme with GitHub Pages deployed via GitHub Actions: https://shigapov.github.io/ReproResearch
- A project wiki using GitHub Wiki (it needs seperate cloning): https://github.com/shigapov/ReproResearch/wiki
- An open Kanban board for project management: https://github.com/users/shigapov/projects/1

Getting started with your research project at GitHub

FDZ FORSCHUNGSDATENZENTRUM UB MANNHEIM

BERD @NFDI

1. You can use the ReproResearch repository as a template for your research project repository
2. Make your changes locally
3. Create your new repo at GitHub
4. Follow the instructions at the creation page.
5. Commit and push your local repo to your new repo at GitHub.