



Conquaire: Datenqualität durch Kontinuierliche Integration

Kolloquium Wissensinfrastruktur
12.04.2019

Vidya Ayer¹, Christian Pietsch², Fabian Herrmann¹, Andreas Rempel¹,
Cord Wiljes¹, Jochen Schirrwagen², Vitali Peil², Philipp Cimiano¹

¹ Exzellenzclusters Kognitive Interaktionstechnologie (CITEC)

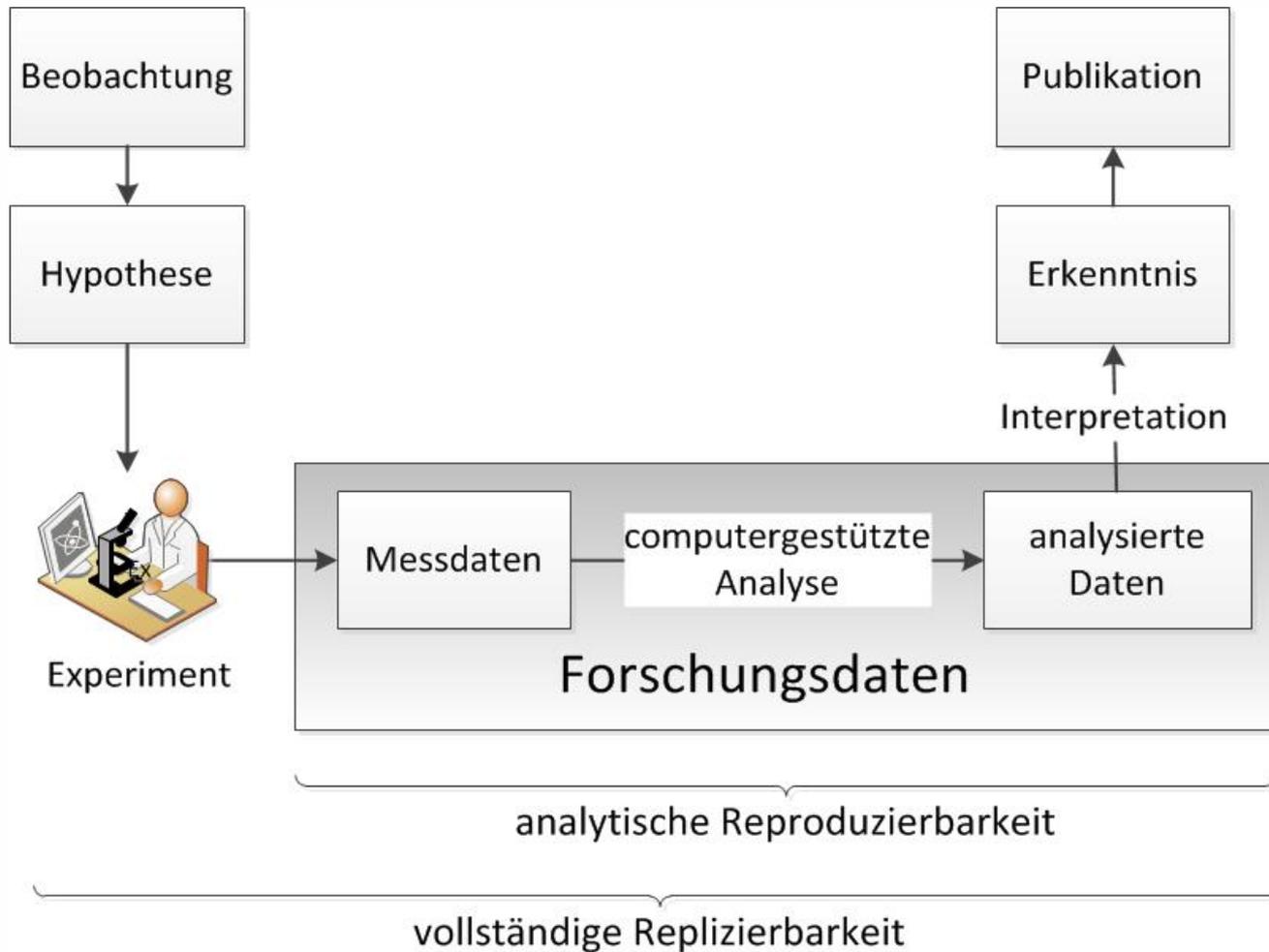
² Universitätsbibliothek Bielefeld

<http://uni-bielefeld.de/conquaire/>

Conquaire

- ▶ “Continuous quality control for research data to ensure reproducibility”
- ▶ Antrag: [doi:10.5281/zenodo.31298](https://doi.org/10.5281/zenodo.31298)
- ▶ 02/2016 bis 04/2019
- ▶ Ziele:
 - ▶ Verbesserung der Qualität von Forschungsdaten
 - ▶ Stärkung der analytischen Reproduzierbarkeit von Forschungsergebnissen

Forschungsdaten



Datenqualität

1. Syntaktische Korrektheit
2. Semantische Plausibilität
3. Übereinstimmung mit FAIR Data Prinzipien
4. Wissenschaftliche Qualität

FAIR Prinzipien

- ▶ Auffindbarkeit (**F**indable)
- ▶ Zugänglichkeit (**A**ccessible)
- ▶ Interoperabilität (**I**nteroperable)
- ▶ Wiederverwendbarkeit (**R**eusable)

<https://www.force11.org/group/fairgroup/fairprinciples>

Findable

F1. (Meta)-Daten wird ein global eindeutiger und dauerhaft persistenter Identifier zugewiesen.

F2. Daten werden mit umfangreichen Metadaten beschrieben.

F3. Metadaten enthalten klar und eindeutig den Identifier, der die Daten referenziert.

F4. (Meta)daten werden in einem durchsuchbaren Dienst registriert oder indiziert.

Accessible

A1. (Meta)daten sind über ihren Identifier und ein standardisiertes Kommunikationsprotokoll auffindbar.

A1.1. Das Protokoll ist offen, frei und universell implementierbar.

A1.2. Das Protokoll unterstützt Authentifizierung und Rechteverwaltung, wo notwendig.

A2. Metadaten sind verfügbar, auch für den Fall, dass die Daten nicht mehr vorhanden sind.

Interoperable

- I1.** (Meta)-Daten nutzen eine formale, zugängliche, geteilte und breit anwendbare Sprache für die Wissensrepräsentation.
- I2.** (Meta)-Daten benutzen Vokabular, das den FAIR Prinzipien folgt.
- I3.** (Meta)-Daten enthalten qualifizierte Referenzen auf andere (Meta)-Daten.

Re-usable

R1. (Meta)-Daten enthalten eine Vielzahl von genauen und relevanten Attributen.

R1.1. (Meta)-Daten werden mit einer eindeutigen und zugänglichen Nutzungslizenz freigegeben.

R1.2. (Meta)-Daten sind mit ihrer Provenienz verbunden.

R1.3. (Meta)-Daten entsprechen den Domain-relevanten Community Standards.

Versionierung mit Git

- ▶ ältere Versionen können jederzeit wiederhergestellt werden
- ▶ lückenloser Nachweis, wer wann welche Änderungen durchgeführt hat
- ▶ Highlighting der Änderungen
- ▶ revisionssichere Historie der Änderungen

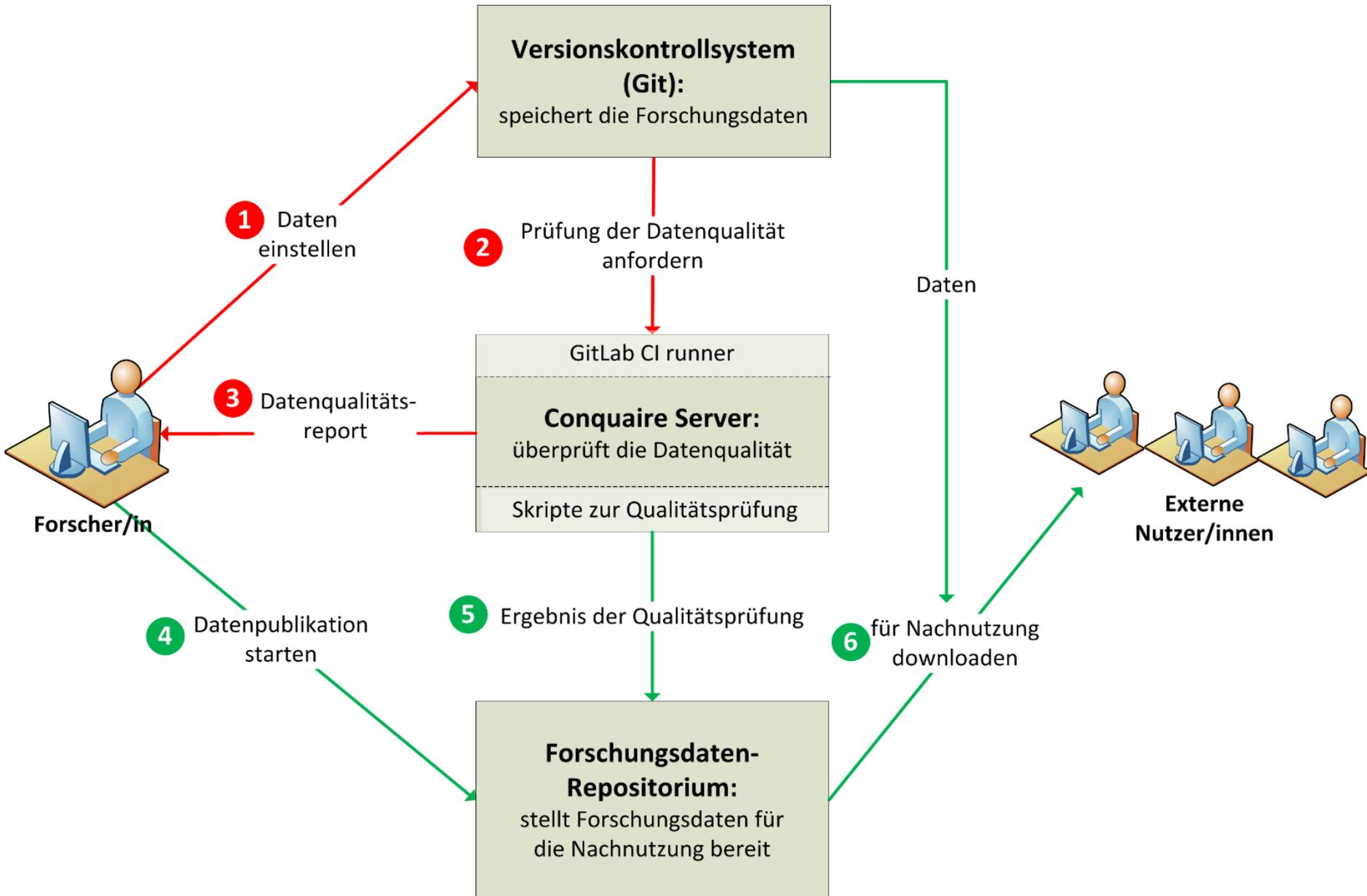
GitLab

- ▶ versionierte Speicherung von Dateien
- ▶ Wiki
- ▶ Issue Tracking
- ▶ Code-Review
- ▶ Continuous Integration (CI/CD), Auto DevOps

Continuous Integration

- ▶ Technologie aus der Softwareentwicklung
- ▶ Ziel: Steigerung der Softwarequalität
- ▶ Beim Einchecken neuer Dateien in die Versionsverwaltung wird automatisch eine Aktion durchgeführt
- ▶ Typische Aktionen:
 - ▶ Übersetzen und Linken der Anwendungsteile
 - ▶ automatisierte Funktions-Tests
 - ▶ Softwaremetriken zur Messung der Softwarequalität
 - ▶ beliebige andere Operationen

Conquaire Systemarchitektur



Weitere Informationen

▶ **Projekt-Website:**

<http://uni-bielefeld.de/conquaire/>

Kontakt

Kompetenzzentrum Forschungsdaten Universität Bielefeld

Tel.: 0521-106-3967

E-Mail: data@uni-bielefeld.de

<https://data.uni-bielefeld.de>