

Research Data Management A Scientific Software Development Perspective – DFG SFB 1194 (Z-INF)



TECHNISCHE
UNIVERSITÄT
DARMSTADT

H. Marschall¹⁾, B. Lambie²⁾, D. Gründing¹⁾,
T. Marić¹⁾, D. Bothe¹⁾, C. Bischof³⁾

- 1) Mathematical Modeling and Analysis,
Department of Mathematics, TU Darmstadt
- 2) Geschäftsführung SFB 1194
(Prof. Peter Stephan, Prof. Dieter Bothe)
- 3) FG Scientific Computing,
FB Informatik, TU Darmstadt



Interaction between
Transport and
Wetting Processes

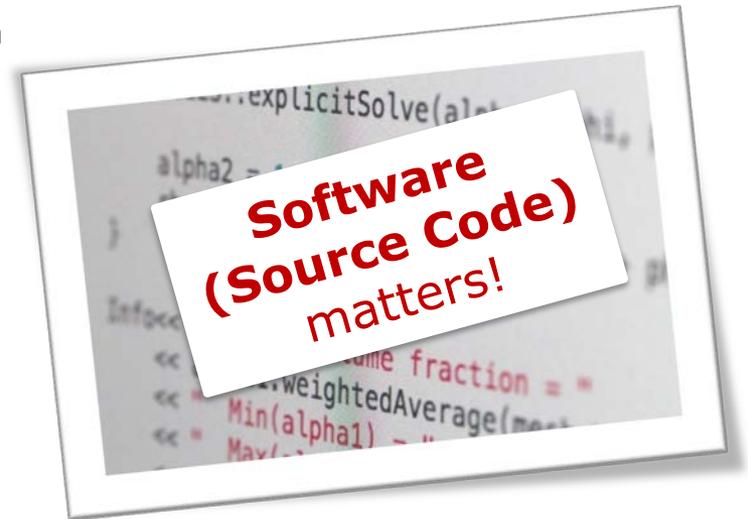
Impulsvortrag, 08.03.2018

**Workshop zum Forschungsdatenmanagement
in den Ingenieurwissenschaften,**
Technische Universität Darmstadt

Typen von Forschungsdaten?

- Messungen/Experimente
- (Feld-/Labor-)Beobachtungen
- Bilder, Videos, Audiodateien
- Textdokumente, Tabellen, Datenbanken
- Quantitative (Umfrage-)Daten
- Umfrageergebnisse, Interviewtranskripte
- Simulationen, Modelle, Software
- Folien, Artefakte, Proben
- Skizzen, Laborbüchern, Notizen

- Rohdaten und deren Verarbeitungsstufen
- Simulationsdaten
- Beobachtungsdaten
- Interviewdaten
- Modelle
- Software
- Persönliche Daten
- Physische Objekte
- Abgeleitete Daten



Definition und Umgang mit Forschungsdaten

Innerhalb des SFB 1194 (Z-INF)



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Erweiterte Definition auf Basis der ISO/IEC-Norm 24765 ^[1]

Relevante, im Rahmen des SFB erhobene Daten umfassen (...) sowohl den **Quellcode der Programme zur Datenauswertung und -erfassung (Experimente) und Datenerzeugung (Modellierung und Simulation)** als auch die hiermit **generierten Mess- bzw. Simulationsdaten (Primärdaten)** einschließlich aller **zur Datenerzeugung notwendigen nicht ausführbaren Programmteile (Sekundärdaten)**.

- Open Source & Open Data Konzept (2 x ½ Stelle für Z-INF!)

^[1] ISO IEC 24765:2010, Zitat „[Software is] 1. all or part of the programs, procedures, rules, and associated documentation of an information processing system 2. computer programs, procedures, and possibly associated documentation and data pertaining to the operation of a computer system 3. program or set of programs used to run a computer“.
<https://www.iso.org/obp/ui/#iso:std:iso-iec-ieee:24765:ed-1:v1:en> (abgerufen am 09.06.2015)



Anforderungen und Ziele

Forschungsdatenmanagement

Anforderungen

- 1998/2013 **DFG** Denkschrift *Sicherung guter wissenschaftlicher Praxis*
- 2009 **DFG** *Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten*
- 2010 **Allianz der deutschen Wissenschaftsorganisationen** *Grundsätze zum Umgang mit Forschungsdaten*
- 2014 **Hochschulrektorenkonferenz (HRK)** *Management von Forschungsdaten – eine zentrale strategische Herausforderung für Hochschulleitungen*
- 2015 **DFG** *Leitlinien zum Umgang mit Forschungsdaten*
- 2017 **EU** Horizon 2020 – *open research data*

Ziele

- Durchsuchbarkeit und Wiederauffindbarkeit
- Nutzbarkeit, Nachnutzbarkeit, Verfügbarmachung
- Reproduzierbarkeit und Vergleichbarkeit
- Hohe Software-Qualität
(Portabilität, Testbarkeit sowie Wartbarkeit und Erweiterbarkeit)
- Effizienz in der wissenschaftl. Arbeit und Zusammenarbeit

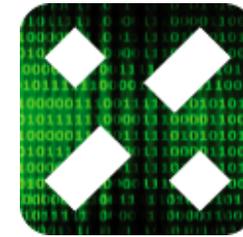
Aktuelle Tätigkeit / Mitwirkung

Aufbau und Etablierung eines zentralen Forschungsdatenmanagements an der TU Darmstadt (SFB als Prototyp-User)

Informationskompetenz, insb. zu FDM bzgl. Forschungssoftware:

- Repositorien, Release-Management
- Langzeitarchivierung
- Persistente Identifikation
- Metadaten

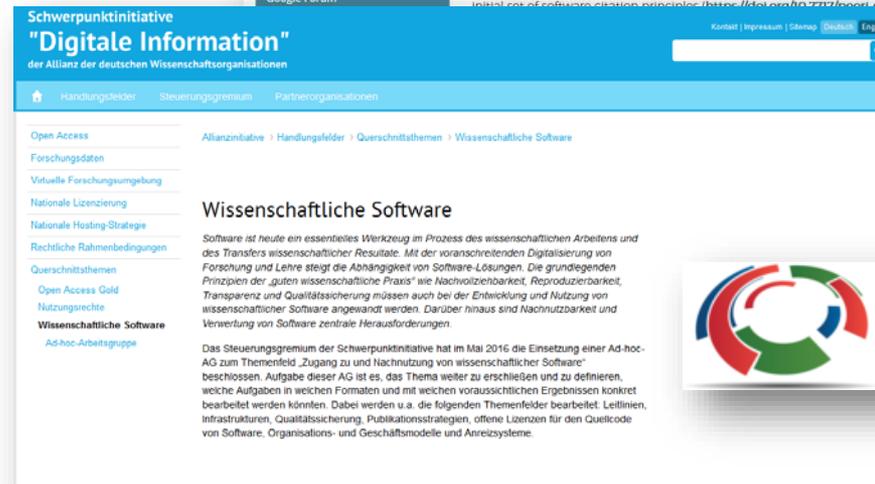
Einbettung und Etablierung des FDM in wissenschaftl. Arbeit und Arbeitsabläufe (Schulungen, Trainings)



Software X



ELSEVIER



SOFTWARE CITATION IMPLEMENTATION WORKING GROUP

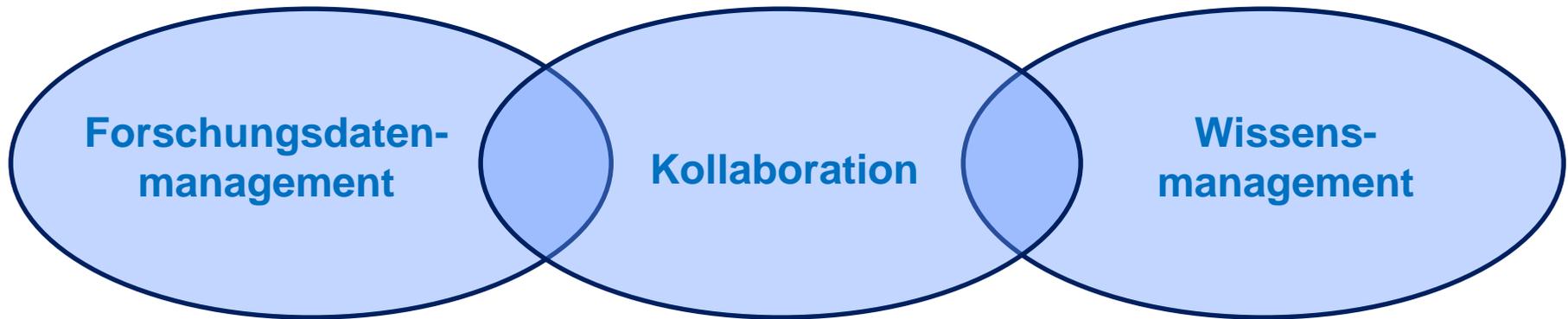
Description

This group builds on the previous [Software Citation Working Group](#), which developed and publicized an initial set of [software citation principles](#) (<https://doi.org/10.7717/peercomm.96>). The activities of the Software Citation Implementation Working Group (publishers, librarians, related working groups, etc.) to:

may also be updated based on feedback



Software
Sustainability
Institute



- Rohdaten
- Simulationsdaten
- Quellcode
- Skripte
- ...

- Austausch
- Protokolle
- Besprechungsnotizen
- ...

- Anleitungen (Howtos)
- Datei-Listen (dig. Laborbücher)
- Berichte
- ...

Plattform für netzbasierte Kollaboration basiert auf

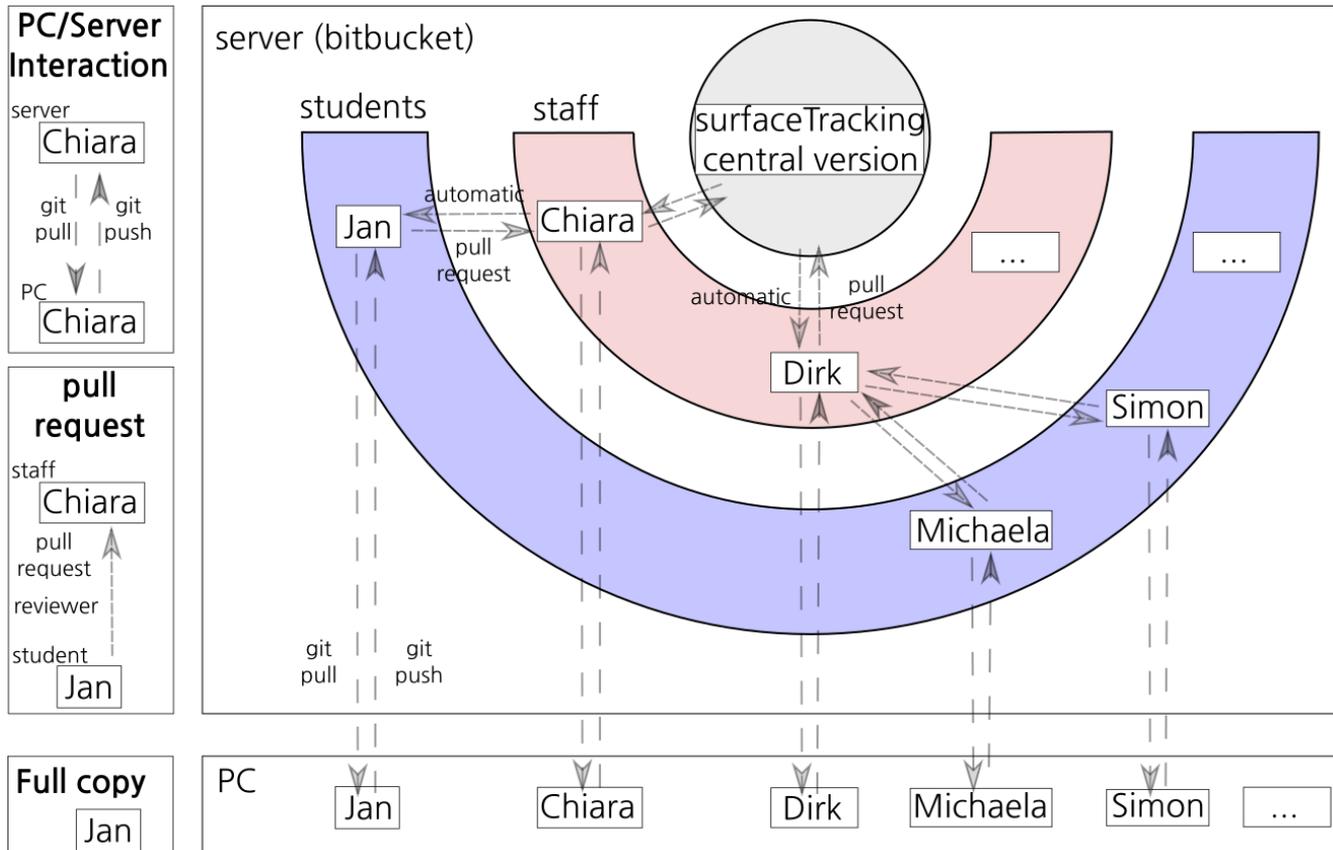
- git (Versionierung, gemeinsames Repository) ✓
- Jenkins / CDash (Tests, wiss. Qualitätssicherung)
- docker (Container, Verfügbarmachung, Reproduzierbarkeit)

Workflow im Kern um
Open-Source Software !

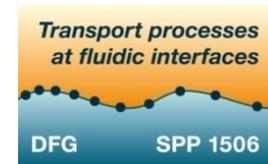
- Nachhaltigkeit
- Verbreitung/Funktionalität
- Sicherheit

Beispiel 1: git-basierter Workflow (Versionierung, Kollaboration)

MMA git model for surfaceTracking



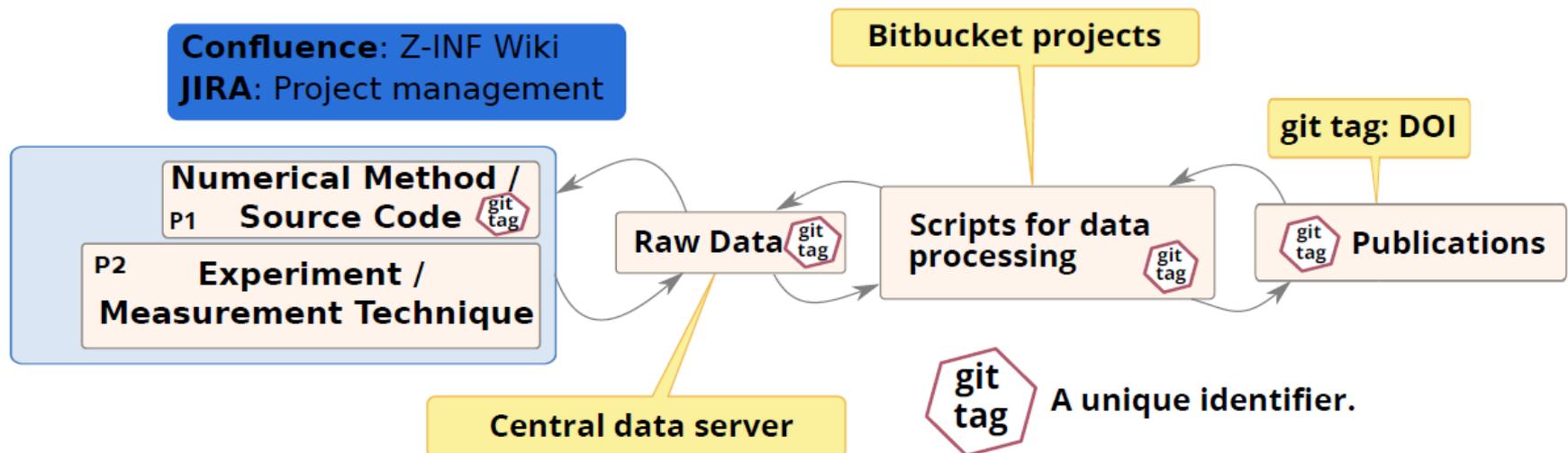
TP B02
Holger Marschall (PI)
Dirk Gründung (WiMi)



Dieter Bothe (PI)
Chiara Pesci (WiMi)

Beispiel 1: git-basierter Workflow (cont'd)

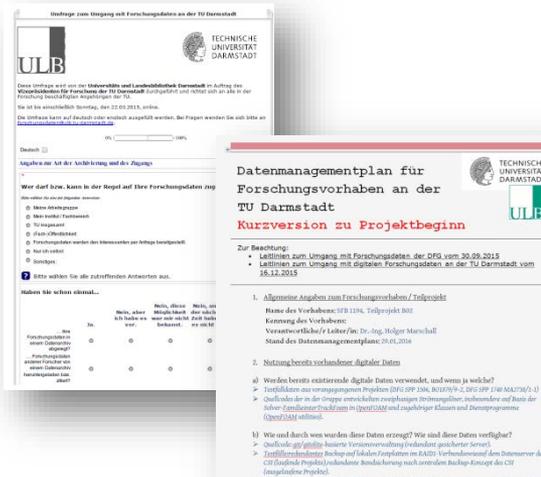
(Versionierung, persistente Identifikation)



Beispiel 2: Datenmanagementpläne als „lebende Dokumente“

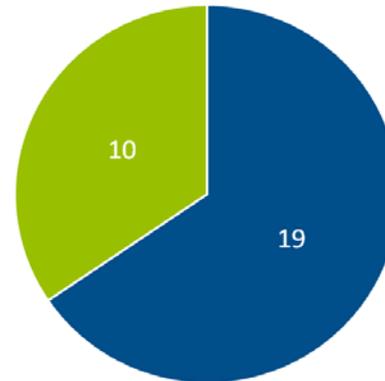
Schärfung entlang des tats. Bedarfs!

- Datenmanagementpläne
- Umfragen
- Gespräche

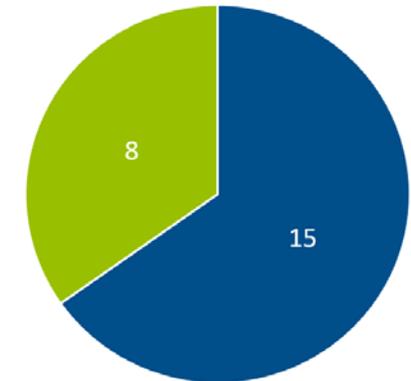


VOR

Speicherbedarf Gesamt in TB

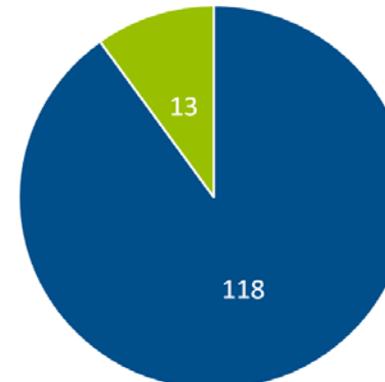


Speicherbedarf für Archivierung in TB

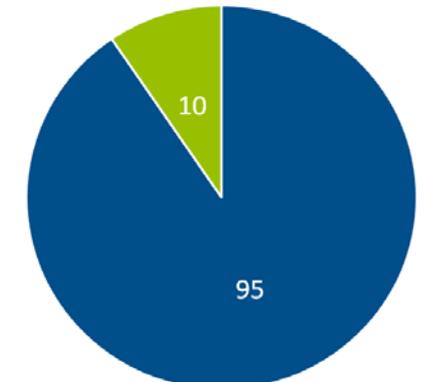


nach

Speicherbedarf Gesamt in TB



Speicherbedarf für Archivierung in TB



■ Experimentelle Projekte ■ Simulative Projekte

■ Experimentelle Projekte ■ Simulative Projekte

Thank You for Your Attention!

Fazit

- Quellcodes (Software) sind Forschungsdaten!
- FDM-Workflows im Kern um Open Source Software
- Erprobte Vorgehensweise
 - Anfangen !
 - Basis mit Grundfunktionalität + Schulungen
 - kontinuierliche Analyse (agiler Ansatz)
Schärfung der Bedürfnisse des Bedarfs
zusammen mit den WissenschaftlerInnen
während Ihrer Forschung



Dr.-Ing. Holger Marschall

Tel.: +49 6151 16-21465

Fax: +49 6151 16-72022

marschall@mma.tu-darmstadt.de



DFG

Deutsche
Forschungsgemeinschaft



Interaction between
Transport and
Wetting Processes