



<https://www.forschungsdaten.info/projekte-in-bw/fdm/sara/>

Autoren: Ackermann, F.; Enderle, P.; Fratz, M.; Kushnarenko, V.; Scharon, D.; Schmücker, P.; Waldvogel, M.; Wesner, S.
Kontakt: marcel.waldvogel@uni-konstanz.de

SARA-Service: Langzeitverfügbarkeit und Publikation von Softwareartefakten

Motivation

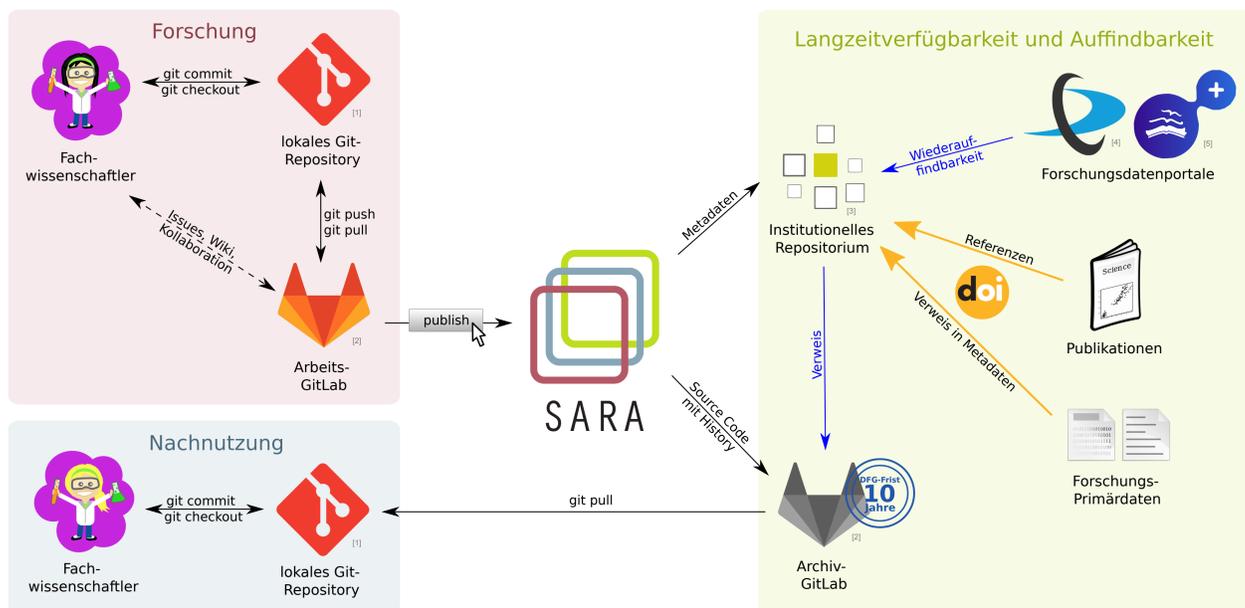
Das Projekt SARA (Software Archiving of Research Artefacts) hat das Ziel, einen neuen wissenschaftlichen Dienst zu entwickeln, mit dem Softwareartefakte langfristig verfügbar gemacht und publiziert werden können. Der Schwerpunkt liegt besonders auf den Softwarewerkzeugen, mit deren Hilfe Forschungsdaten verarbeitet und ausgewertet werden. In der Fachdisziplin Biologie werden beispielsweise Messdaten computergestützt erfasst und weiterverarbeitet. In der Elektrotechnik und Informatik müssen kontinuierlich die unterschiedlichen Versionsstände der entwickelten Software gespeichert werden, um den Fortschritt der Entwicklung zu dokumentieren.

Der Dienst soll die Workflows der Forschenden begleiten und sie anregen, Zwischenstände ihrer Forschungsarbeit auch bei Softwarewerkzeugen bereits prozessbegleitend festzuhalten. Damit kann auf die gesamte Entstehungshistorie der oft von den Forschenden selbst entwickelten oder modifizierten Werkzeuge zurückgegriffen werden. Die erhobenen Forschungsdaten sind zusammen mit den jeweiligen Versionsständen der dazugehörigen Softwarewerkzeuge für die weiterführende Forschung nachvollziehbar. Der im Projekt prototypisch für Informatik / Elektrotechnik und Biologie entwickelte Dienst soll nach Ablauf der Evaluierungsphase grundsätzlich allen Fachdisziplinen zur Verfügung stehen.

Einsatzszenarien (Beispiele)

- Informatik / Elektrotechnik
 - Abschlussarbeiten
 - Softwareentwicklung
- Biologie und andere Laborwissenschaften
 - selbst entwickelte Softwareartefakte
 - Weiterentwicklungen an Open Source Software
- Digital Lab Journals mit Git
- Digital Humanities (mögliche Kooperation mit dem Projekt RePlay-DH)

Architektur



Die Abbildung stellt die beteiligten Komponenten und die Abhängigkeiten zwischen ihnen dar. Das „Arbeits-GitLab“ bezeichnet eine GitLab-Instanz, mit der Wissenschaftler wie gewohnt aktiv arbeiten und über die sie den **SARA-Service** aktivieren können. Dieser veranlasst auf Wunsch des Nutzers eine langfristige Archivierung des Arbeitsprojekts in einem **Archiv-GitLab** (eine GitLab-Instanz der Universität Konstanz als Landesdienst). Zusätzlich ist der Nachweis der Publikation in **Institutionellen Repositorien** möglich, so dass die Daten dort und in übergreifenden Systemen (z.B. Google Scholar, OpenAIRE) gefunden werden können.

Alle veröffentlichten Arbeitsprojekte erhalten beispielsweise einen **DOI** zur eindeutigen und dauerhaften Referenzierung. Beschreibende Metadaten werden auf einer **Landing Page** des Repositoriums angezeigt. Von dort erfolgt ggf. der Sprung ins zentrale Archiv-GitLab. Entscheidet sich der Nutzer für die Publikation der gesamten Versionsgeschichte, kann im Archiv-GitLab nicht nur jede spezifische Version heruntergeladen, sondern auch die Versionsgeschichte online exploriert und nachvollzogen werden. Diese Umsetzung orientiert sich an den **Software Citation Principles der FORCE 11** „Software Citation Working Group“ und weiteren Empfehlungen und Best Practices.

Soll-Workflows

Den Fachwissenschaftlern sollen Workflows für den Umgang mit Softwareartefakten zur Verfügung gestellt werden, um einerseits den Arbeitsaufwand für deren Verwaltung gering zu halten und andererseits bereits bei der täglichen Arbeit einen Mehrwert zu bieten. Beispielsweise ist es bei sauberer Versionierung problemlos möglich, vorübergehend zu einer älteren oder alternativen Version zu wechseln und eine Auswertung mit dieser zu wiederholen.

Diese Soll-Workflows zum Umgang mit Softwareartefakten orientieren sich an bewährten Best Practices der Softwareentwicklung, die entsprechend den jeweiligen Anforderungen der Fachdisziplinen angepasst werden.

Metadaten

Jede veröffentlichte Version eines Softwareartefakts soll über einen DOI eindeutig referenziert werden. Daher werden für jedes Softwareobjekt die folgenden Metadaten erfasst und auf der Landing Page des Institutionellen Repositoriums angezeigt:

- | | |
|---------------------|------------------|
| 1. Identifizier | Kardinalität 1-n |
| 2. Creator | Kardinalität 1 |
| 3. Title | Kardinalität 1 |
| 4. Publisher | Kardinalität 1 |
| 5. Publication Year | Kardinalität 1 |
| 6. Resource Type | Kardinalität 1 |

Außerdem wurden die in Git und GitLab standardmäßig verfügbaren Metadaten untersucht. Für die einzelnen Metadatenfelder wurde analysiert, welche automatisiert extrahiert werden können und Strategien dafür definiert. Eine Herausforderung besteht hier zum Beispiel in der Definition und Nennung der Autoren / Contributors.

Technologie

- Git für die Versionierung, GitLab als Oberfläche
- Ruby on Rails für die Erweiterung von GitLab
- „Continuous Integration“ oder GUI-Integration für die Anbindung des Arbeits-GitLab an den SARA-Service
- JAVA für die Programmierung des SARA-Servers
- REST als Schnittstelle zu den Repositorien

Im Projekt handelt es sich bei den Institutionellen Repositorien um KOPS (Universität Konstanz) oder OPARU (Universität Ulm), die beide auf der Software DSpace basieren. Folglich liegt der Fokus auf der Anbindung von DSpace-Repositorien, wobei über einen modularen Ansatz die Anbindung weiterer Institutioneller Repositorien grundsätzlich möglich sein soll.

