SARA

SARA
Software Archiving of Research Artefacts

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

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

SARA-Service: Langzeitverfügbarkeit und Publikation von Forschungsdaten und Softwareartefakten

Motivation

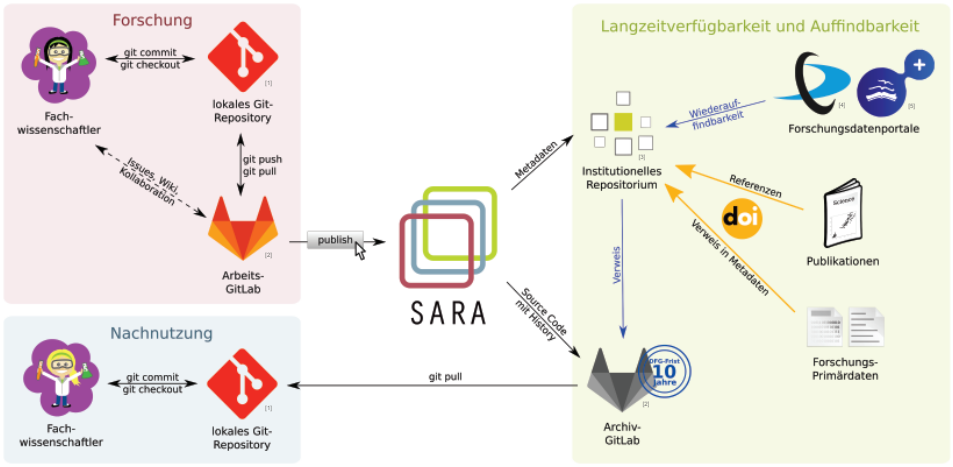
Das Projekt SARA (Software Archiving of Research Artefacts) hat das Ziel, einen neuen wissenschaftlichen Dienst zu entwickeln, mit dem Softwareartefakte langfristig verfügbar gemacht und publiziert werden können. Der Schwerpunkt liegt besonders auf den Softwarewerkzeugen, mit deren Hilfe Forschungsdaten 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.

Beispielhafte Einsatzgebiete und Szenarien

- Informatik / Elektrotechnik
  - Abschlussarbeiten
  - Softwareentwicklung
- Biologie und andere Laborwissenschaften
  - selbst entwickelte Softwareartefakte
  - Weiterentwicklungen an Open Source Software
- Digital Lab Journals mit Git
- Digital Humanities

Architektur



[1] Git-Logo: https://commons.wikimedia.org/wiki/File:Git-logo.svg, Jason Long (https://twitter.com/jasonlong), CC BY 3.0 (https://creativecommons.org/licenses/by/3.0/deed.en). [2] GitLab-Logo: https://gitlab.com/gitlab-org/gitlab-artwork/tree/master/, GitLab, Inc. (https://about.gitlab.com/), CC BY-NC-SA 4.0 (https://creativecommons.org/licenses/by-nc-sa/4.0/). [3] DSpace-Logo: http://www.dspace.org/marcom, DuraSpace (http://www.dspace.org/), CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0/). [4] DataCite-Logo: https://commons.wikimedia.org/wiki/File:DataCite_logo.png, DataCite (https://www.datacite.org/), CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/). [5] OpenAIRE-Logo: https://www.openaire.eu/images/site_images/OpenAIREplus_logo.png, OpenAIRE Consortium (https://www.openaire.eu/), CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/). Cliparts: https://openclipart.org/ (CC0 1.0).

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 der **SARA-Service** aufgerufen werden kann. 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 ein zitierbarer Nachweis der Publikation im **Institutionellen Repositorium** des Nutzers möglich, so dass die Daten dort und in übergreifenden Systemen (z.B. Google Scholar, OpenAIRE) gefunden werden können. Dem Nutzer stehen derzeit zwei Institutionelle Repositorien zur Auswahl: **KOPS** (Universität Konstanz) und **OPARU** (Universität Ulm), die beide auf der Software DSpace basieren.

Alle veröffentlichten Arbeitsprojekte erhalten einen **DOI** (oder ähnlichen Persistent Identifier) 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“** sowie weiteren Empfehlungen und Best Practices.

Metadaten

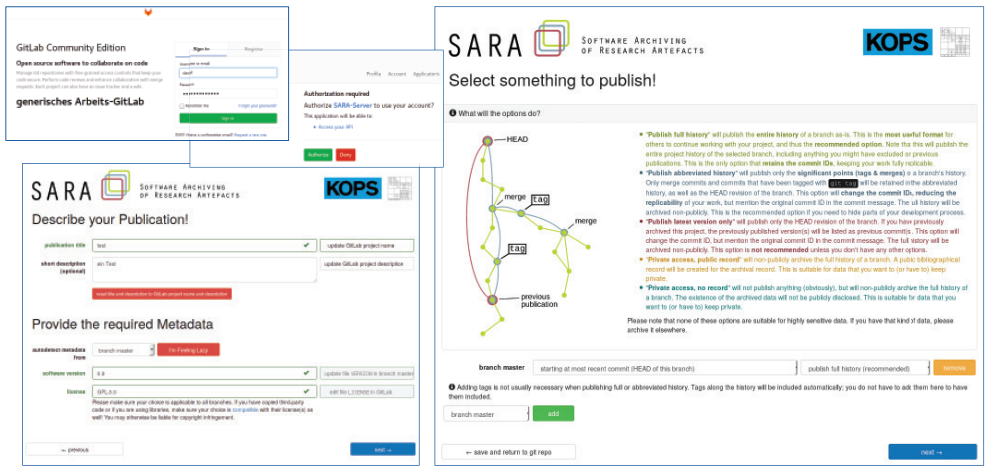
Jede veröffentlichte Version eines Softwareartefakts soll über einen DOI eindeutig referenziert werden. Für jedes Softwareobjekt werden die folgenden Metadaten erfasst und auf der Landing Page des Institutionellen Repositoriums angezeigt:

DataCite Pflichtfelder: Identifier; Creator; Title; Publisher; Publication Year; Resource Type

Weitere Pflichtfelder: Link ins Archiv-GitLab; Person, welche die Veröffentlichung angestoßen hat

Die Lizenz wird, entsprechend der Best Practices in der Softwareentwicklung, direkt im Git Repo hinterlegt.

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.



Universität Konstanz

Universität Konstanz:
AG Verteilte Systeme,
FB Informatik und Informationswissenschaft

ulm university universität ulm

Universität Ulm:
Kommunikations- und Informationszentrum (kiz),
Institut für Organisation und Management von Informationssystemen (IOMI)

Dieses Poster ist lizenziert unter CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-2-1790358amuita1

Gefördert vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg

Baden-Württemberg
MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST