SARA

SARA
Software Archiving of Research Artefacts

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

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

SARA Service: Long-Term Access and Publishing of Research Data and Software Artefacts

Motivation

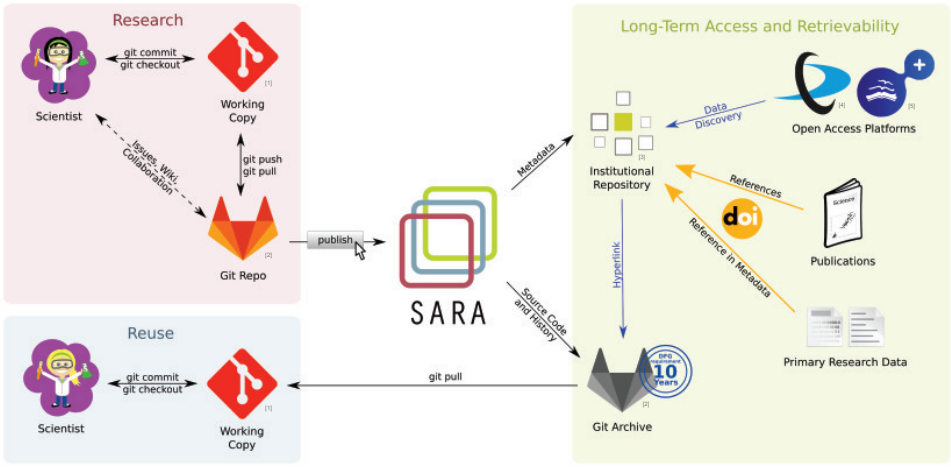
The SARA project (Software Archiving of Research Artefacts) aims to develop a new scientific service that allows long-term access and publishing of research data and scientific software. Its main focus is on software tools that support the processing and analysis of research data. In biological research, for example, measured data is collected and processed with the support of computers. The matching version of this software is required to faithfully reproduce the research results. Within Computer Science and Electrical Engineering, the different versions of newly developed software need to be continually stored in order to document the progress of development.

The service is designed to support the research process and encourage researchers to archive or publish preliminary results during their research, particularly regarding software tools as part of the research process. This allows other researchers to access the whole development history of these tools which are often locally developed, or at least modified, by researchers themselves. Thereby, the collected research data, together with the matching versions of accompanying software tools, are reproducible for further research. The service is being prototypically developed for Computer Science, Electrical Engineering and Biology and is planned to be open to all scientific disciplines after the evaluation phase ends.

Example Use Cases and Scenarios

- Computer Science and Electrical Engineering
  - Bachelor's and Master's Theses
  - general software development
- Biology and other lab sciences
  - locally developed software artefacts
  - derivatives of existing open source software
- Digital Lab Journaling with Git
- Digital Humanities
- Planned: Cooperation with CITAR (Archiving, Citation and Reuse of Virtual Research Environments)

Architecture



[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-com/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).

This diagram shows the components of the SARA service and the dependencies between them. The **Git Repo** is a GitLab instance with which scientists work as usual and from which they can invoke the **SARA service**. At the user's request, this initiates the storage of the working project in the **Git Archive** (a GitLab instance of the University of Konstanz, operated as a state-wide service) for long-term availability. In addition, it is also possible to add a citeable bibliographical reference to the user's **Institutional Repository**, so that the archived project can be found there and in Open Access platforms (e.g. Google Scholar, OpenAIRE) indexing it. Two institutional repositories are currently available for user selection: **KOPS** (University of Konstanz) and **OPARU** (Ulm University), both based on the DSpace application.

All published projects receive a **DOI** (or a similar persistent identifier) for unambiguous and permanent referencing. Descriptive metadata is displayed on a **landing page** of the institutional repository. From there, users can navigate to the central Git Archive. If the publishing user has decided to publish the entire version history, the Git Archive not only allows users to download all intermediate versions, but also to explore and reconstruct the development history on-line. This architecture is based on the **FORCE 11 "Software Citation Group"** ***Software Citation Principles*** and other recommendations and best practices.

Metadata

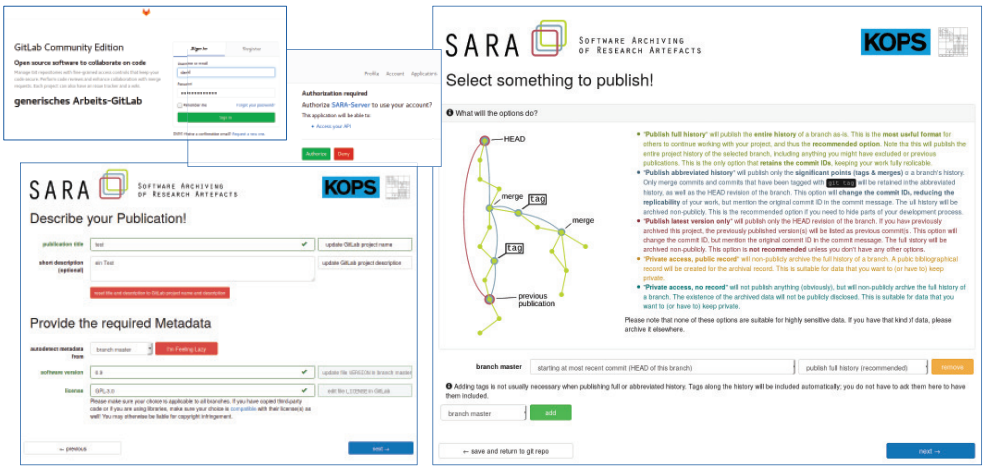
We intend to use DOIs to reference the different versions of software artefacts published using the SARA service. The following metadata are captured and presented on the landing page of the Institutional Repository:

Metadata required by DataCite: Identifier; Creator; Title; Publisher; Publication Year; Resource Type

Other mandatory fields: Link to Git Archive; Person who Triggered the Publishing Process

The license is stored directly in the Git repo, in accordance with common practice in software engineering.

Furthermore, we investigated how metadata that are already available in Git and GitLab can be used for this purpose. We defined strategies for automatic extraction of relevant information to reduce manual entry. One of the challenges we faced is the definition of author / contributor with respect to software and how to best represent them.



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)

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

Baden-Württemberg
MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-2-9j5v39vxqfij3