
SARA: Open Source Projekt zur langfristigen Verfügbarkeit und Zitierbarkeit von Software

Franziska Rapp¹, Daniel Scharon², Matthias Fratz², Stefan Kombrink¹, Volodymyr Kushnarenko¹, Pia Schmücker¹, Marcel Waldvogel² und Stefan Wesner¹

¹ Universität Ulm;

² Universität Konstanz

1. Motivation

Softwareartefakte (von kleinen Skripten bis hin zu komplexen Software-Frameworks) spielen im wissenschaftlichen Forschungsprozess eine tragende Rolle. Sie werden z.B. zur Erhebung, Verarbeitung, Visualisierung oder Modellierung von Forschungsdaten eingesetzt, sollten langfristig verfügbar sein und nach den FAIR Prinzipien¹ behandelt werden. Das Projekt SARA zielt darauf ab, Softwareartefakte langfristig verfügbar [1] und referenzierbar [2] zu machen. Dabei richtet sich SARA disziplinübergreifend an sämtliche Forschende, die Software selbst entwickeln oder bestehende Software nachnutzen und weiterentwickeln.

2. Der SARA Webservice

Im Bereich der Softwareentwicklung hat sich das Versionierungssystem Git als de-facto-Standard etabliert. Zur Veröffentlichung von Daten und Software sind Repositorien bewährte, nachhaltige Infrastrukturen. Dabei handelt es sich um Publikationsplattformen, die zentral für Mitglieder und Angehörige einer wissenschaftlichen Einrichtung oder einer Fachcommunity angeboten werden. Sie verfügen in der Regel über eine Landing Page mit beschreibenden Metadaten wie Titel, Autoren, Abstract, Lizenz etc. Zur eindeutigen Referenzierbarkeit und Zitierbarkeit wird ein persistenter Identifier, wie z.B. ein Digital Object Identifier (DOI), vergeben. Die Auffindbarkeit von in Repositorien beschriebenen Objekten wird durch die Indexierung in Suchmaschinen und -portalen wie Google und OpenAIRE² erhöht.

Im Projekt wurde ein Webservice entwickelt, der diese bestehenden Infrastrukturen verbindet. Der SARA Webservice extrahiert Metadaten und Dateien aus Projekten in GitHub und angebundenen GitLab-Installationen. Über eine intuitive graphische Oberfläche können Forschende den Umfang der extrahierten Dateien nach ihren Vorstellungen

Das hier beschriebene Poster ist in der Open Access-Plattform der Universität Heidelberg heiDOK unter der DOI <https://doi.org/10.11588/heidok.00026854> veröffentlicht.

¹ <https://www.go-fair.org/fair-principles/>, zuletzt aufgerufen am 23.04.2019

² <https://www.openaire.eu/>, zuletzt aufgerufen am 23.04.2019

festlegen, indem sie z.B. nur bestimmte Branches für die Archivierung auswählen. Die Entwicklungshistorie kann komplett oder in gekürzter Version mitarchiviert werden. Für die Nachnutzung der Software wird vom Forschenden eine Lizenz festgelegt. Ist bereits eine LICENSE Datei vorhanden, werden die gängigen Lizenzen von SARA automatisch erkannt. Die Dateien werden anschließend in ein Git-Archiv (read-only) überführt, das die langfristige Verfügbarkeit nach den Empfehlungen der DFG sicherstellt. Die aus den Projekten extrahierten Metadaten können durch den Forschenden editiert und ergänzt werden, bevor sie an das Repositorium übergeben werden. Im SARA Projekt wurden als Repositorien exemplarisch DSpace-Instanzen angebunden. Die konkrete Anbindung ist hierbei flexibel konfigurierbar und passt sich den Anforderungen von Repositorienbetreibern an [3]. Über die Registrierung einer DOI werden die archivierten Softwareartefakte referenzierbar und zitierbar und durch die Indexierung in Suchmaschinen wird schließlich die Auffindbarkeit und Sichtbarkeit erhöht.

3. Vorteile

3.1. Wissenschaftler

- Wissenschaftler können den Empfehlungen der DFG zur Aufbewahrung der Softwareartefakte für mind. 10 Jahre nachkommen (gute wissenschaftliche Praxis)
- Zitierbarkeit der eigenen Arbeit, bessere Auffindbarkeit
- Nachnutzung von Softwareartefakten
- Zugriffsgeschützte Archivierung möglich (dark archive)

3.2 Infrastrukturanbieter

- Individueller oder kooperativer Betrieb
- Git mit dem eigenen Repositorium verbindbar
- Strategische Erweiterung der eigenen FDM-Infrastruktur

4. Nachnutzung

Der Quellcode des SARA Webservice ist auf GitHub frei verfügbar³. Es sind verschiedene Einsatzszenarien denkbar. SARA kann beispielsweise an einer Einrichtung betrieben und mit den dortigen lokalen GitLab-Installationen, GitHub und dem institutionellen Repositorium verbunden werden. Auch ein kooperativer Einsatz ist denkbar, bei dem eine Einrichtung den SARA Webservice hostet und GitLab-Installationen verschiedener Einrichtungen sowie deren jeweilige institutionelle Repositorien angebunden sind. Dasselbe Szenario ist für Fachcommunities statt für einzelne Einrichtungen denkbar, was SARA

³ <https://github.com/sara-service>, zuletzt aufgerufen am 16.04.2019

auch in den Kontext der sich bildenden Konsortien der Nationalen Forschungsdateninfrastruktur⁴ setzt.

Literaturverzeichnis

- [1] Deutsche Forschungsgemeinschaft. Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“. Ergänzte Auflage. Weinheim: Wiley-VHC, 2013. <https://doi.org/10.1002/9783527679188.oth1>.
- [2] Smith, Arfon M., Daniel S. Katz, Kyle E. Niemeyer und FORCE11 Software Citation Working Group. „Software Citation Principles.“ PeerJ Computer Science 2 (2016): e86. <https://doi.org/10.7717/peerj-cs.86>.
- [3] Rapp, Franziska; Kombrink, Stefan; Kushnarenko, Volodymyr; Fratz, Matthias; Scharon, Daniel (2018): SSARA-Dienst: Software langfristig verfügbar machen", o-bib - Das offene Bibliotheksjournal, Bd. 5/Nr. 2 (2018). <https://doi.org/10.5282/o-bib/2018H2S92-105>.

⁴ <https://www.dfg.de/nfdi,zuletztaufgerufenam24.04.2019>