

Open-Science-Snack

01/2025

Ein Service der Universitätsbibliothek Freiberg

Forschungssoftware im Kontext von Open Science

Forschungssoftware ist längst ein zentraler Bestandteil wissenschaftlicher Arbeit und spielt eine entscheidende Rolle sowohl als Werkzeug als auch als Forschungsergebnis und -objekt. Vor dem Hintergrund der großen fachlichen, thematischen und qualitativen Varianz, kann sich Software, die während des Forschungsprozesses oder für einen speziellen Forschungszweck erstellt wurde, stark „in ihrem Umfang, Zweck, Kontext oder ihren Eigenschaften unterscheiden“ (siehe [DFG-Definition](#)). Sie ist kein Nebenprodukt, sondern wird vielmehr häufig benötigt, um in Kombination mit den Forschungsdaten [veröffentlichte Forschungsergebnisse nachvollziehen](#) zu können. Umso bedeutender ist es, dass sie gut dokumentiert und unter Open-Science-Prinzipien öffentlich zugänglich, reproduzierbar und transparent gemacht wird.

Vor allem bei großen Projekten oder wenn Software der primäre Projektoutput ist, kann die Erstellung eines [Softwaremanagementplans](#) (SMP) sinnvoll sein. Er umfasst Aspekte wie Entwicklung, Dokumentation, Qualitätssicherung, Lizenzierung, Wartung, Veröffentlichung, Wiederverwendbarkeit und Langzeitarchivierung der Software. Eine Hilfestellung hierzu bietet der [Practical guide to Software Management Plans](#) (2023). Eine [Handreichung](#) in Bezug auf die Herausforderungen im Umgang mit Forschungssoftware mit entsprechenden Handlungsempfehlungen wurde auch im Rahmen der Schwerpunktinitiative Digitale Information der [Allianz der deutschen Wissenschaftsorganisationen](#) erarbeitet. Basierend auf den [FAIR-Prinzipien für Forschungsdaten](#) wurden im Mai 2022 zudem die [FAIR4RS-Principles](#) veröffentlicht, um den spezifischen Anforderungen an Forschungssoftware (Ausführbarkeit, Entwicklung und Versionierung) gerecht zu werden. GitHub oder GitLab sind Beispiele für Plattformen, die die Umsetzung der FAIR-Prinzipien ermöglichen. Für die sächsischen Universitäten und Hochschulen stellt die TU

Chemnitz eine [GitLab-Instanz](#) als gemeinsam nutzbare technische Dienstleistung zur Verfügung. Die gemeinnützige Organisation [Software Heritage](#) hat es sich außerdem zur Aufgabe gemacht, den gesamten Bestand an öffentlich zugänglicher Software weltweit zu sammeln, zu bewahren und zugänglich zu machen. Seit 2017 wird sie dabei von der [UNESCO](#) unterstützt. Das jährliche gemeinsam ausgerichtete [Symposium](#) bietet eine einzigartige Austauschplattform für Wissenschaftler, Industriepartner und politische Entscheidungsträger.

Auch in Deutschland gewinnt das Thema in der Förderpolitik an Bedeutung. Die DFG hat 2024 sowohl eine Handreichung zum „[Umgang mit Forschungssoftware im Förderhandeln](#)“ herausgegeben als auch ein neues unbefristetes Förderprogramm zum Aufbau von [Infrastrukturen für Forschungssoftware](#) aufgesetzt. Ab 2025 können Anträge zweimal jährlich eingereicht werden. Forschungssoftware ist außerdem ein Schwerpunktthema der NFDI-AG [Research Software Engineering](#). Im Juli 2024 wurde mit [nfdi.software](#) auch ein neuer NFDI-Basisdienst bewilligt, der als zentraler Marktplatz den Zugang zu Forschungssoftware erleichtern sowie die nachhaltige Nutzung und Entwicklung unterstützen soll. Darüber hinaus hat die DFG im Dezember 2024 die [Amsterdam Declaration on Funding Research Software Sustainability](#) (ADORE Software) unterzeichnet. Damit unterstützt sie internationale Bestrebungen zur Standardisierung und Förderung von Forschungssoftware sowie die Anerkennung von Softwareentwicklung als wissenschaftliche Leistung.

Weiterführende Literatur und Informationen:

- [Defining Research Software: a controversial discussion](#) (2021)
M. Barker u. a.: [Introducing the FAIR Principles for research software](#) (2022)
- R. v. Nieuwpoort, D.S. Katz: [Defining the roles of research software](#) (Version 2, 2024)



Vorgestellt: OpenGeoSys (OGS)

Beim Begriff „Forschungsinfrastruktur“ denkt man zunächst an die großen Teilchenbeschleuniger der Physiker oder umfangreiche Labore voller komplexer Geräte. Doch wohin mit den Daten? Wie generiert man aus Daten Wissen, testet Hypothesen und Modelle? Wie wird dieses Wissen konkret und transferierbar? Hier kommt eine andere Art von Forschungsinfrastruktur zum Einsatz: Forschungssoftware.

OpenGeoSys (OGS) ist eine solche international genutzte Forschungssoftware, mit der ein interdisziplinärer und kooperativer Ansatz zahlreicher Forschungsgruppen unterstützt wird. Es handelt sich um ein wissenschaftliches Open-Source-Projekt zur Entwicklung numerischer Methoden für die Simulation thermisch, hydraulisch, mechanisch, chemisch/biologisch (THMC/B) gekoppelter Prozesse in porösen und geklüfteten Medien, vorrangig in Locker- und Festgesteinen. Entsprechend sind die klassischen Anwendungsfelder von OGS in den Bereichen [Hydrogeologie](#), [geothermische Energiesysteme](#), [Energiespeicherung](#), [Geotechnik](#), [CO₂-Speicherung](#) sowie in der [Entsorgung radioaktiver Abfälle](#) zu finden. Die Software blickt nunmehr auf eine fast 30-jährige Geschichte zurück, ist aktuell in der 6. Generation, umfasst etwa 330 000 Zeilen C++ Code und [wurde im Laufe der Zeit stets an aktuelle Standards der Softwareentwicklung angepasst](#).

Hauptentwickler von OGS sind das [Department für Umweltinformatik](#) am Helmholtz-Zentrum für Umweltforschung in Leipzig sowie [die AG von Prof. Thomas Nagel](#) am Institut für Geotechnik der TU Bergakademie Freiberg. Ein Team von etwa sechs Kernentwicklern, die für das Softwareengineering verantwortlich sind, wird ergänzt von Domain Scientists aus unterschiedlichen Bereichen, die sich um Entwicklung und Anwendung spezifischer Modelle kümmern.

Die Entwicklung der aktuell vier Releases pro Jahr folgt den Open-Science-Prinzipien. Der Zugang zum Quellcode ist unerlässlich für die Reproduzierbarkeit wissenschaftlicher Ergebnisse. Offene Entwicklungsprozesse – im Falle von OGS über GitLab – wiederum fördern eine größere Gemeinschaft an Nutzen-

den und Mitwirkenden, was zu robusterer, besser getesteter und allgemein anwendbarer Software führt. Zur Qualitätssicherung setzt OGS u.a. auf umfangreiches automatisiertes Testen (über 2 700 Testfälle, die auf mehreren Betriebssystemen und Prozessorarchitekturen ausgeführt werden), intensives Code Review, Continuous Integration sowie die Beteiligung an [internationalen Modellvalidierungsinitiativen](#). Aktuell stehen zudem plattformübergreifende und HPC-taugliche Workflows sowie die Anbindung an das Python-Ökosystem ([OGSTools](#)) im Vordergrund bei der Verbesserung der Anwendbarkeit sowie zur Integration von Methoden des maschinellen Lernens in die Software. Die Software ist in [Docker und Singularity Containern](#) verfügbar und kann auch mit [PiPy](#) oder [Guix](#) installiert werden. Zur Einbindung interessierter Mitwirkender organisiert das OGS-Team neben Community Meetings und Trainingskursen auch Hackathons, bei denen Nutzende von OGS gemeinsam mit Kernentwicklern eigene Ideen gezielt umsetzen können.

Quelloffene Software bedient neben den o.g. Punkten auch die Forderung der Öffentlichkeit nach Transparenz bei der Umsetzung geotechnischer Großprojekte. Ein aktuelles Beispiel ist die Standortauswahl für die Endlagerung hochradioaktiver Abfälle, bei dem sich OGS als ein wissenschaftlicher Referenzcode etwa der [Bundesgesellschaft für Endlagerung](#) sowie der [Bundesanstalt für Geowissenschaften und Rohstoffe](#) etabliert hat. Dabei geht es nicht nur um die Simulationssoftware selbst. Hier spielen offene, transparente, reproduzierbare und qualitätsgesicherte [Workflows eine zentrale Rolle, in denen Daten verarbeitet, Simulationen vorbereitet, durchgeführt, ausgewertet und visualisiert werden](#). Die Entwicklung derartiger Fähigkeiten im Kontext der Endlagersuche ist aktuell Gegenstand des [OpenWorkflow Projekts](#). Daten- und simulationsgestützte Workflows dieser Art ermöglichen auch die systematische Untersuchung des Einflusses von [Ungewissheiten in komplexen geowissenschaftlichen Projekten](#) sowie den [Transfer wissenschaftlicher Entwicklungen in die Praxis](#).

