

Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanager*innen

Softwarezentrierter Ansatz zur Langzeitarchivierung von Forschungsdaten

Softwareerhaltung und Emulation als Absicherung für eine langfristige
Nutzbarkeit von Daten

Klaus Rechertⁱ

Dirk von Suchodoletzⁱⁱ

2025

Zitiervorschlag

Rechert, Klaus, und von Suchodoletz, Dirk. 2025. Softwarezentrierter Ansatz zur Langzeitarchivierung von Forschungsdaten. Softwareerhaltung und Emulation als Absicherung für eine langfristige Nutzbarkeit von Daten. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanager*innen* Nr. 3/2025: S. 1-9. DOI: [10.17192/bfdm.2025.3.8854](https://doi.org/10.17192/bfdm.2025.3.8854).

Dieser Beitrag steht unter einer
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

ⁱHochschule Kehl. ORCID: [0000-0002-2454-4374](https://orcid.org/0000-0002-2454-4374)

ⁱⁱUniversität Freiburg. ORCID: [0000-0002-4382-5104](https://orcid.org/0000-0002-4382-5104)

Abstract

Die langfristige Erhaltung von Forschungsdaten erfordert mehr als die Bewahrung von Bits (Bitstream-Preservation), da Forschungsdaten als komplexe digitale Artefakte einzustufen sind und ihren wissenschaftlichen Nutzen erst durch softwarebasierte Verarbeitung entfalten können. Für die langfristige Erhaltung ist daher entscheidend, ob geeignete Software für die Nutzung zur Verfügung steht und praktisch einsetzbar ist. Angesichts der Vielfalt von Datenformaten und der Heterogenität von Datensätzen erscheint eine vollständige (manuelle) Vorauswahl, Standardisierung der Datenformate sowie Vorkuratierung aller anfallender Datensätze weder praktikabel noch finanzierbar. Stattdessen wird eine softwarezentrierte Strategie in Kombination mit Emulation als infrastruktureller Ansatz vorgeschlagen, die disziplinübergreifend implementiert werden kann und als Absicherung in der Breite dienen soll.

1 Ausgangslage

Während die Bitstream-Preservation, d. h. die Bit-genaue Erhaltung von Daten, heute technisch gut beherrschbar ist, reicht diese allein für eine nachhaltige Nutzung von Forschungsdaten nicht aus. Forschungsdaten sind im Sinne der Langzeitarchivierung (LZA) als komplexe Artefakte zu betrachten, deren Wert sich häufig erst durch eine softwarebasierte Verarbeitung erschließt (von Suchodoletz et al., 2021, 2023). Es muss also sichergestellt werden, dass die notwendige Software ebenfalls langfristig verfügbar ist.

Durch eine technische und inhaltliche Vorauswahl aller für die LZA relevanten Datensätze, Begrenzung der möglichen Quellen, Standardisierung der Formate oder eine fachliche Spezialisierung der Archive können Aufwand und Komplexität dieser Aufgabe reduziert werden. Die Bewertung der Relevanz sowie der technischen Tauglichkeit ist insbesondere angesichts langfristiger Zeithorizonte aber sehr herausfordernd, zumal unklar bleibt, wie diese Aufgaben ressourcenseitig bewältigt werden können. Insbesondere die Heterogenität der Formate sowie fachspezifische Entwicklungen führen zu absehbaren Schwierigkeiten und Kosten bei Zugriff und Nachnutzung (Giessler et al., 2021). Im Fall von stark genutzten oder standardisierten Formaten kann das zukünftige Vorhandensein der Software erwartet werden, für weniger gängige oder proprietäre Formate bleibt hingegen in der Regel nur der Rückgriff auf die ursprünglich eingesetzte Software.

Eine breit aufgestellte Forschungsdatenmanagementstrategie sollte also für die langfristige Erhaltung von Daten auch den ursprünglichen softwarebasierten Ausführungskontext berücksichtigen und idealerweise einschließen. Diese kann zwar eine zukünftige Interpretation und Nachnutzung nicht uneingeschränkt garantieren, ermöglicht sie doch mit hoher Wahrscheinlichkeit, vorausgesetzt die ursprünglich eingesetzte Software ist vorhanden und kann genutzt werden.

Analog zu den Strategien der Langzeitarchivierung anderer komplexer Objekte (z. B. softwarebasierter Kunst) bietet ein softwarebasierter Ansatz in Verbindung mit späterer Emulation eine LZA-Basisstrategie. Für deren Umsetzung sind derzeit noch organisatorische und infrastrukturelle Maßnahmen erforderlich, die jedoch hinreichend abstrakt und generisch sind, sodass sie disziplinübergreifend anwendbar sind. Auf diese Weise lassen sich sowohl der Aufwand als auch das Risiko eines zukünftigen Zugriffs auf eine sehr breite, heterogene Sammlung minimieren sowie deren Nachnutzung sicherstellen. Die Nationale Forschungsdateninfrastruktur (NFDI) kann hierbei sowohl organisatorische Koordination als auch technisch-infrastrukturelle Unterstützung leisten.

2 Notwendige Erhaltungsvoraussetzung

Wissenschaftler*innen aller Fachdisziplinen aber auch die Anbietenden von Diensten für die Forschung sind bei ihrer Arbeit mit Software in allen Erscheinungsformen konfrontiert. Ein großer Teil davon hat Einfluss auf die Nachnutzbarkeit und Reproduzierbarkeit von Forschung im Sinne der FAIR-Prinzipien.

2.1 Erhaltung von Software und Wissen

Software bildet die technische Grundlage für die Datenverarbeitung und -interpretation und ist somit essenziell für den Langzeitzugriff auf Forschungsdaten. Obwohl Datenformate in der Regel migrierbar (vgl. Neuroth et al., 2016) sind, bedarf auch deren Anpassung an künftige Arbeitsumgebungen wiederum geeigneter Software und kontinuierlicher Überwachung. Der Großteil heutiger Datensätze lässt sich (mit geringem) Aufwand in aktuellen Arbeitsumgebungen nutzen; dennoch sollten Strategien des Forschungsdatenmanagements (FDM) um Maßnahmen zur nachhaltigen Sicherung des Datenzugriffs ergänzt werden, insbesondere bei längeren Zeiträumen. Da Forschung Trends unterliegt, besteht die Gefahr, dass Datensätze längere Zeit ungenutzt bleiben und nicht an aktuelle Umgebungen angepasst werden. Verfügbare Software aus der Zeit der Datenerhebung bildet hierbei eine wichtige Grundlage, weshalb eine strukturierte Sammlung und Archivierung relevanter Software notwendig ist.

Software kann entweder als Quellcode – etwa selbst entwickelte Programme – oder lediglich in binärer Form vorliegen, wie kommerzielle Produkte oder spezifische Kompilate für exakte Reproduktionserfordernisse. Für beide Gruppen ist es notwendig, nicht nur die jeweilige Software selbst zu archivieren, sondern auch sämtliche erforderlichen Laufzeitumgebungen und Abhängigkeiten (z. B. Betriebssysteme, Bibliotheken) zu dokumentieren und zu bewahren. Insbesondere bei Quellcode müssen zudem geeignete Interpreter oder Compiler sowie oft zahlreiche Zusatzwerkzeuge verfügbar

gehalten werden. Eine systematische Softwaresammlung stellt somit eine wesentliche Voraussetzung für die Nachnutzung digitaler Artefakte dar und bildet das Fundament einer generischen Langzeitstrategie. Während mit Initiativen wie der Software Heritage Foundation¹ ein wesentlicher Akteur zur Archivierung von Quellcode existiert, fehlen vergleichbare Sammlungs- und Erhaltungsstrategien für andere Softwareformen weitgehend. Anders als im Bereich von Printerzeugnissen existiert bislang keine strukturierte Sammlung oder sogar umfassendes Pflichtabgabesystem für Software in Deutschland; mit Ausnahme der Internationalen Computerspielesammlung (ICS)² gibt es weder koordinierende noch disziplinspezifische Softwaresammlungen noch übergreifende Initiativen.

Ein weiteres zentrales Problem stellen die Vielzahl und Komplexität von Lizenzen kommerzieller Software dar: Während zu Beginn der Digitalära meist zeitlich unbegrenzte Nutzungsrechte galten, schränken Anbieter heute Rechte hinsichtlich Dauer und Eigentumsübergang zunehmend ein. Zusätzliche Schwierigkeiten ergeben sich durch Firmenübernahmen oder Produktaufgaben; sowohl klassische Lizenzbeschaffung als auch Übertragungen zwischen Institutionen sind häufig aufwändig oder rechtlich unklar geregelt. Mit der Verbreitung kommerzieller Software wurden zudem zahlreiche Kopierschutzmechanismen entwickelt – von physischen Maßnahmen bis hin zu serverbasierten Lizenzschlüsseln –, deren Funktionieren meist an das Fortbestehen von Hardwareschnittstellen oder des Herstellers gebunden ist. Dies erschwert langfristig insbesondere Archiven und Wissenschaftseinrichtungen den Zugriff auf benötigte Programme; geschützte Software behindert damit die Prinzipien offener Wissenschaft sowie unabhängige Überprüfbarkeit wissenschaftlicher Ergebnisse. Daher sind dringend rechtliche Regelungen erforderlich, um Zugänge zu wissenschaftlichen Ergebnissen sowie digitalen Kulturgütern unabhängig von einzelnen Anbietern sicherzustellen.

Zunehmend relevant wird zudem eine neue Form softwarebasierter Anwendungen ohne klar abgrenzbare Objektgrenzen (Espenschied & Rechert, 2018), was Sammlung, Archivierung sowie Zitation erheblich erschwert („Versionslosigkeit“). Diese Problematik betrifft sowohl proprietäre Cloud-basierte Anwendungen als auch quelloffene Lösungen mit dynamisch eingebundenen externen Ressourcen wie GitHub oder Docker Hub. Vor diesem Hintergrund bedarf es weiterentwickelter Konzepte der Softwareerhaltung: Neue Methoden müssen darauf abzielen, verschiedenartige Softwareformen erfassbar, beschreibbar, zitierbar und nachnutzbar zu machen und insbesondere nahtlos in wissenschaftliche Workflows integrierbar und automatisierbar sein, so dass möglichst kein zusätzlicher Aufwand für Archivierungsschritte entsteht. Beispielsweise könnte im Rahmen von CI/CD-Pipelines automatisiert ein Snapshot der jeweiligen Umgebung erstellt und archiviert werden (*continuous preservation*), sodass nachhal-

¹Software Heritage Archiv, <https://archive.softwareheritage.org/>, letzter Zugriff am 29.09.2025.

²Internationale Computerspielesammlung (ICS), <https://www.game.de/die-internationale-computerspielesammlung-ics/>, letzter Zugriff am 29.09.2025.

tige Langzeitarchivierung integraler Bestandteil des wissenschaftlichen Outputs wird – ohne zusätzlichen Aufwand für Forschende.

Ein zentraler Kritikpunkt an einer auf Originalsoftware basierenden Langzeitstrategie ist das potenziell zukünftig fehlende operative Wissen zur Bedienung der Software. Der Erhalt dieses Wissens ist ebenso wichtig wie die Bewahrung der Software selbst. Zwar liegt entsprechendes Wissen häufig in Form von Handbüchern oder anderer Dokumentation vor, bei populärer Software zusätzlich auch durch unabhängige Autor*innen und Communities von Nutzenden, etwa als Online-Texte oder Videos. Dennoch gestaltet sich der Zugriff auf dieses Wissen oft aufwändig, und einzelne Handbücher genügen meist nicht, um komplexe (etwa emulierte) Systeme vollständig zu bedienen; umfassenderes Systemwissen, insbesondere zum Betriebssystem und dessen Interaktionskonzepten, ist erforderlich. Mit Large-Language Models (LLMs) existieren nun leistungsfähige Methoden, große Mengen digitaler Dokumentation kontextbasiert zu durchsuchen und zusammenzufassen (Rechert & Gieschke, 2024; Zhang et al., 2024). Voraussetzung dafür ist jedoch eine digitale, maschinenlesbare Verfügbarkeit der relevanten Informationen. Da viele Softwareprodukte unzureichend dokumentiert sind, wird essenzielles Wissen oftmals von Nutzenden auf informellen Plattformen wie Reddit, Stack Overflow oder YouTube ergänzt und geteilt. Die Webarchivierung dieser informellen Quellen ist daher für die langfristige Sicherung des Wissens über Software und Betriebssysteme zentral, muss jedoch systematisch erfolgen und steht nur in einem engen Zeitfenster zur Verfügung. Nur so kann die Nutzbarkeit von aktuellen und zukünftiger Software gewährleistet werden.

2.2 Erhaltung von Datenzugriff und Interaktion

Mit der Verfügbarkeit der Software ist deren Ausführung der nächste notwendige Schritt, um den Zugriff auf und die Interaktion mit archivierten Daten zu erlauben. Ist die Software auf aktuellen Computersystemen nicht mehr lauffähig, erfolgt die Wiederherstellung des ursprünglichen Systems meist über Emulation (vgl. Neuroth et al., 2016). Emulation ist mittlerweile – zumindest konzeptionell – eine etablierte Zugriffsstrategie in der Langzeitarchivierung, insbesondere für interaktive Softwareobjekte wie CD-ROMs, Computerspiele oder digitale Kunstwerke in Museen, Bibliotheken und Archiven.

Emulation ermöglicht die zukünftige Nutzung von Objekten im Originalzustand, ohne dass zwischen Archivierung und Nutzung zusätzliche Arbeiten erforderlich sind. Erst beim Zugriff entsteht ein Aufwand durch die Bereitstellung einer emulierten Hardwareumgebung, die üblicherweise ein breites Spektrum an Software- und Systemumgebungen abdeckt. Der zukünftige Aufwand, einen Emulator zu implementieren (sofern notwendig), ist im Verhältnis zum potenziellen Nutzen – gemessen an der Anzahl der Softwareanwendungen und Objekte, die auf dieser Hardware genutzt werden können – sehr gering und muss pro Computergeneration wahrscheinlich nur einmal

erbracht werden. Aufgrund dieses generischen Charakters müssen Gedächtnisinstitutionen solche Umgebungen nicht selbst bereitstellen; zentrale Ansätze – etwa cloud-basierte Lösungen (von Suchodoletz et al., 2013) – ermöglichen derzeit eine effiziente Umsetzung.

Individueller Aufwand entsteht jedoch durch die Notwendigkeit angepasster Softwareumgebungen: Spezifische Datensätze und Nutzungsszenarien erfordern individuell konfigurierte Softwareinstallationen. Eine Zuordnung von Datenformaten zur jeweils erforderlichen Software ist bislang – mangels umfänglicher Formatregistries – nur eingeschränkt möglich; vielmehr erfordert dieser Prozess manuelle Expertise bei Auswahl und Installation. Häufig wird dieses Problem mittlerweile durch den Einsatz sogenannter Container (z. B. Docker, Podman etc.) gelöst, welche portabel vorinstallierte und vorkonfigurierte Softwareumgebungen bereitstellen. Geht dieses Wissen verloren, gestaltet sich der Prozess noch aufwändiger und wird risikoreicher. Daher ist es entscheidend, solches Wissen bereits heute sowohl für Menschen als auch für Maschinen nutzbar zu dokumentieren und aktiv genutzte Umgebungen zu archivieren (vgl. Rechert et al., 2024). Ziel muss es sein, mithilfe dieser Wissensbasis die individuell notwendigen Prozesse möglichst vollständig zu automatisieren.

Trotz technischer Fortschritte wie bwFLA (Rechert et al., 2012) bleiben Emulationslösungen technisch anspruchsvoll sowie organisatorisch und rechtlich herausfordernd. Entsprechend wurden bisher nur wenige Emulationsumgebungen in Gedächtnisorganisationen etabliert (Cochrane et al., 2022; Moya et al., 2024; Steinke et al., 2016), was sowohl an fehlender technischer Expertise und IT-Kapazitäten als auch an bislang geringer Nachfrage liegt. Für Forschungsdaten ist die technische Umsetzung von Emulation derzeit noch nicht vorrangig. Zusammen mit der Softwareerhaltung erscheint jedoch mittelfristig eine zentrale Koordination sowie der Aufbau entsprechender Kompetenzen und Strukturen erforderlich, um die neuen technisch herausfordernden Softwaresystem abbilden zu können. Die Aktivitäten der NFDI im Bereich der Langzeitarchivierung, die Erfahrungen von nestor und die Einbindung der Open-Source-Community bieten Potenzial, diese Lücke zu schließen. NFDI und nestor können geeignete Institutionen identifizieren und verfügen über Expertise in Marktbeobachtung, rechtlicher Begleitung und Abstimmung. Darüber hinaus können durch Arbeitsgruppen Akteure gefunden werden, die gezielt LZA-spezifische Weiterentwicklungen dieser Strategie verfolgen.

3 Ausblick

Viele Forschungseinrichtungen haben sich in den letzten Jahren auf Open Science Policies verpflichtet, die transparente, reproduzierbare und nachnutzbare Forschungsergebnisse zum Ziel haben. Diese Policies umfassen damit implizit oder explizit auch Software in ihrer vielfältigen Form. Mit der Nationalen Forschungsdateninfrastruktur entstand in den letzten fünf Jahren ein übergreifender organisatorischer Rahmen, der

es erlaubt, ausgehend von der fachspezifischen Sicht eine gemeinsame Abstraktionsebene für die Umsetzung erfolgreichen Langzeitzugriffs auf Forschungsdaten schrittweise zu realisieren. Dies ergänzt den nationalen Rahmen von nestor, welcher sich dem nachhaltigen Langzeitzugriff auf digitale Artefakte der traditionellen Gedächtnisorganisationen verschrieben hat. Allerdings besteht das Risiko, dass durch die rasanten Veränderungen technischer Mittel – insbesondere in der Forschung – deren Nachnutzung, insbesondere in der Breite, langfristig nicht sichergestellt werden kann.

Ohne eine Sammlung von Software besteht eine erhebliche Gefahr, dass ein signifikanter Teil der heute produzierten Artefakte künftig nur noch mit großem Aufwand nachgenutzt werden kann – insbesondere dann, wenn Daten und Software miteinander verknüpft sind. Selbst wenn mittelfristig keine strukturierte Lösung für die Softwarearchivierung (beispielsweise durch Pflichtabgabe) umsetzbar erscheint, sollte zumindest eine zentrale Koordination dezentraler Sammlungen durch Fachbereiche und individuelle Initiativen in Erwägung gezogen werden. Ohne den Aufbau entsprechender Organisationsstrukturen, Expertise und Infrastruktur riskieren wir nicht nur, den langfristigen Zugang zu bereits bestehenden Artefakten nicht mehr sichern zu können; vielmehr werden wir auch den vielfältigen zukünftigen Herausforderungen im Bereich der Softwarearchivierung – wie etwa Cloud-Computing, Software-as-a-Service oder Abonnementlizenzen – nicht mehr angemessen begegnen können.

Gleiches gilt für den Erhalt operativen Wissens, der bislang im Wesentlichen von Bibliotheken durch die Sammlung von Fachbüchern und gegebenenfalls Handbüchern übernommen wurde. Mit dem Wandel von klassischen Publikationen hin zu vielfältigen, teils informellen Online-Formaten für technische Dokumentationen ist der Aufbau beziehungsweise Ausbau von Archivierungsstrategien für diese Formate zwingend erforderlich. Erhaltung von Software und Wissen stellen die Basis für eine funktionale Zugriffstrategie mittels Emulation, die auch dank neuer technischer Entwicklungen zunehmend automatisierbar und damit effektiver und nutzerfreundlicher wird.

Langzeitarchivierung und Sicherstellung der späteren Nachnutzbarkeit sollten gemeinsame Aufgaben aller Konsortien in der NFDI sein. Der Einrichtung einer entsprechenden Arbeitsgruppe in der Sektion „Common Infrastructure“ sollte nun die konzeptionelle Entwicklung und praktische Umsetzung der hier skizzierten Aufgaben folgen. Die Bereitstellung einer entsprechenden Basisinfrastruktur muss über eine reine Bitstream-Preservation hinausgehen und ein integriertes Ökosystem schaffen, das strukturierte Softwaresammlungen, Emulation, systematische Wissensdokumentation und rechtssichere Verfahren für proprietäre Software umfasst.

Danksagung

Im NFDI-Konsortium DataPLANT (7/1 – 42077441) werden die Herausforderungen des Langzeitzugriffs auf Forschungsdaten auf verschiedenen Ebenen adressiert. So setzt das Konsortium auf ein breit angelegtes „open contribution and collaboration model“, welches alle Beiträge zu Forschungsdaten, Annotation, Software für die Prozessierung und Datenhaltung öffentlich nachvollziehbar macht. Zudem wurden die Voraussetzungen für eine vollständige Reproduzierbarkeit von Workflows, wie beispielsweise in Galaxy untersucht und im Annotated Research Context (ARC) spezifiziert. Zudem ist DataPLANT direkt in die Arbeitsgruppe Langzeitarchivierung der Sektion Common Infrastructure der NFDI involviert. In der NFDI4BIOIMAGE (46/1 – 501864659) werden verschiedene Formen von Virtual Desktop Infrastructures entwickelt, die es erlauben, sowohl zeitlich eingefrorene Snapshots von Arbeitsumgebungen zu erstellen als auch den Fernzugriff auf eine Vielzahl von emulierten Hardwareumgebungen effizient bereitzustellen. Beide Konsortien werden im Rahmen der NFDI durch die Deutsche Forschungsgemeinschaft DFG auf Basis der Bund-Länder-Vereinbarung zum Aufbau einer nationalen Forschungsdateninfrastruktur vom 26. November 2018 gefördert.

Literaturverzeichnis

- Cochrane, E., Rechert, K., Oberhauser, J., Anderson, S., Fox, C., & Gates, E. (2022). Useable Software Forever. *iPRES 2022, Glasgow, 12–16 September 2022*. <https://doi.org/10.17605/OSF.IO/AY48C>
- Espenschied, D., & Rechert, K. (2018). Fencing Apparently Infinite Objects. In N. McGovern & A. Whiteside (Hrsg.), *Proceedings of the 15th International Conference on Digital Preservation, iPRES 2018, Boston, MA, USA, September 24–28, 2018*. <https://hdl.handle.net/11353/10.923620>
- Giessler, J., Gieschke, R., Rechert, K., & Cochrane, E. (2021). Automating the Selection of Emulated Rendering Environments for Born-Digital Data-Sets. In G. Berget, M. M. Hall, D. Brenn & S. Kumpulainen (Hrsg.), *Linking Theory and Practice of Digital Libraries – 25th International Conference on Theory and Practice of Digital Libraries, TPDL 2021, Virtual Event, September 13–17, 2021, Proceedings* (S. 106–111, Bd. 12866). Springer. https://doi.org/10.1007/978-3-030-86324-1_12
- Moya, C., de Vries, D., Stuckey, H., & Swalwell, M. (2024). AusEaaS and Archiving Australian Media Arts. *iPRES 2024 – International Conference on Digital Preservation*. <https://doi.org/10.5281/zenodo.13361955>
- Neuroth, H., Oßwald, A., Scheffel, R., Strathmann, S., & Jehn, M. (Hrsg.). (2016). *Nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung (Version 2.3, 2010)*. Verlag Werner Hülsbusch.

- Rechert, K., & Gieschke, R. (2024). Preserving Secondary Knowledge: Using Language Models for Software Preservation. *International Journal of Digital Curation*, 18(1). <https://doi.org/10.2218/IJDC.V18I1.930>
- Rechert, K., Valizada, I., von Suchodoletz, D., & Latocha, J. (2012). bwFLA – A Functional Approach to Digital Preservation. *PIK – Praxis der Informationsverarbeitung und Kommunikation*, 35(4), 259–267. <https://doi.org/10.1515/pik-2012-0044>
- Rechert, K., Espenschied, D., Gieschke, R., & Hagenmaier, W. (2024). Preserving Users' Knowledge of Contemporary and Legacy Computer Systems: Automated Analysis of Recorded User Activity. *iPRES 2024 Papers - International Conference on Digital Preservation*. <https://ipres2024.pubpub.org/pub/tc52q3z1>
- Steinke, T., Rechert, K., Schoger, A., & Padberg, F. (2016). Project EMiL – Emulation of Multimedia Objects. In B. Signori, S. Knight & N. Milic-Frayling (Hrsg.), *Proceedings of the 13th International Conference on Digital Preservation, iPRES 2016, Bern, Switzerland, October 3–6, 2016*. <https://hdl.handle.net/11353/10.503170>
- von Suchodoletz, D., Brettschneider, P., Axtmann, A., Heber, M., Oberländer, L., Leendertse, J., Schumm, I., Brandt, O., Schmidt, K., Gertis, L., Selzer, M., Ulrich, R., Iglezakis, D., & Boda, V. (2023). *Sicherstellung der Reproduzierbarkeit von Forschungsergebnissen durch Bewahrung des Zugriffs auf Forschungssoftware*. <https://doi.org/10.17192/BFDM.2023.5.8555>
- von Suchodoletz, D., Leendertse, J., Rechert, K., Gieschke, R., Semaan, S., Goldammer, B., & Tolkatsch, D. (2021). Developing a Holistic Research Data Management Strategy for a University: Making Preservation Planning and Long Term Access First Grade Citizens. *iPRES 2021 – 17th International Conference on Digital Preservation, Beijing*. <https://doi.org/10.17605/OSF.IO/ACNSR>
- von Suchodoletz, D., Rechert, K., & Valizada, I. (2013). Towards Emulation-as-a-Service: Cloud Services for Versatile Digital Object Access. *International Journal of Digital Curation*, 8(1), 131–142. <https://doi.org/10.2218/ijdc.v8i1.250>
- Zhang, W., Huang, J.-H., Vakulenko, S., Xu, Y., Rajapakse, T., & Kanoulas, E. (2024). Beyond Relevant Documents: A Knowledge-Intensive Approach for Query-Focused Summarization Using Large Language Models. *International Conference on Pattern Recognition*, 89–104.