

# Repositorien als digitale Gedächtnisträger zwischen Evolution und Langzeitplanung

## Steiner, Elisabeth

elisabeth.steiner@uni-graz.at  
Karl-Franzens-Universität Graz, Austria

## Vasold, Gunter

gunter.vasold@uni-graz.at  
Karl-Franzens-Universität Graz, Austria

## Scholger, Martina

martina.scholger@uni-graz.at  
Karl-Franzens-Universität Graz, Austria

## Einleitung

Seit beinahe 20 Jahren verwaltet und publiziert das Repository GAMS Forschungsdaten aus den Geisteswissenschaften und dem Kulturerbebereich. Derzeit enthält das Langzeitarchiv etwa 115.000 annotierte digitale Objekte aus mehr als 60 Projekten. Dies reicht von digitalen Editionen oder Textsammlungen (z.B. Briefe, mittelalterliche Rechnungsbücher) über Bildsammlungen (z.B. historische Fotografien und Postkarten) bis hin zu digitalisierten musealen Sammlungen (z.B. archäologische Artefakte oder andere Museumsobjekte) (GAMS 2021). Dabei liegt das Augenmerk sowohl auf der langfristigen Sicherung und Zugänglichkeit von Ressourcen wie auch auf dem nachhaltigen Umgang mit den bearbeiteten Forschungsdaten.

Domänenspezifische Repositorien müssen einen Ausgleich zwischen langzeitauglicher und somit eher konservativer Entwicklung und dem Einsatz aktueller Technologien finden. Das Poster wird den konzeptionellen Rahmen dieser Forschungsdateninfrastruktur vorstellen und Lösungsansätze aus diesem Widerspruch diskutieren, die besonders auf die Migration der Kernkomponenten fokussieren.

## Grundlegende Konzepte

Das OAIS-konforme (Space Data Systems 2012) Repository folgt einer weitgehend XML-basierten Inhaltsstrategie, die domänenspezifische Datenmodelle und bewährte Standards nutzt. Alle kooperativen Forschungsprojekte werden innerhalb der gleichen Infrastruktur realisiert, wobei eine begrenzte Anzahl von bevorzugten Formaten und Technologien verwendet wird, um die Wartbarkeit über einen langen Zeitraum zu gewährleisten. Das System stützt sich ausschließlich auf Open-Source-Software (als Kernelement Fedora Commons), da es sich den Prinzipien von Open Access und Open Science verpflichtet. Die Designprinzipien zielen auf die Einhaltung internationaler Best Practices wie

dem COAR Community Framework for Good Practices in Repositories (COAR 2020) und Zertifizierungsrichtlinien (das Repository ist derzeit mit dem CoreTrustSeal und als CLARIN B-Centre zertifiziert) und folgen den FAIR-Prinzipien (Force 11 2016).

Zentrales Merkmal ist die Verwaltung digitaler Ressourcen als komplexe digitale Objekte auf der Grundlage von Inhaltsmodellen (Klassen). Jedes Inhaltsmodell definiert für einen Forschungsdatentyp eine beliebige Anzahl von Datenströmen (Eigenschaften), ergänzt durch teilweise automatisch generierte Metadatensätze und Methoden, über die spezielle Sichtweisen auf die Objekte realisiert werden. Die digitale Repräsentation einer mittelalterlichen Handschrift besteht beispielsweise aus deskriptiven, technischen, strukturellen und administrativen Metadaten, einer Anzahl von Faksimiles, einer TEI-kodierten Transkription und RDF-Repräsentationen zur Unterstützung von Linked Data. An das Objekt gebundene Methoden ermöglichen synoptische Darstellungen mehrerer Varianten und bieten projektspezifische Such-, Visualisierungs- und Analysefunktionen. Alle Datenströme und Methoden werden innerhalb eines einzigen, vollständig selbstbeschreibenden Objekts gespeichert und verwaltet, das über einen persistenten Identifikator adressierbar ist, was erhebliche Vorteile für die Langzeitarchivierung bietet.

## Herausforderung Langzeitplanung

Eine Herausforderung hinsichtlich der langfristigen technischen Erhaltung jedes Repositoriums liegt einerseits in der sich laufend ändernden IT-Umgebung (Betriebssysteme, neue Software-Versionen), andererseits in sich verändernden Ansprüchen an die Nutzbarkeit eines solchen Systems aus BenutzerInnen-sicht; aus organisatorischer Sicht erweist sich die dauerhafte Finanzierung oft als schwierig.

Die technische Evolution machte den Austausch der zentralen Komponente des Systems notwendig. Der Schritt von Fedora 3 auf Fedora 6 war kein simples Upgrade, sondern bedeutete einen grundlegenden Wechsel vom objektorientierten Paradigma zu einer Linked-Data-Architektur auf Basis der Linked-Data-Plattform (LDP)-Spezifikation (Speicher/ Arwe / Malhotra 2015). Die Abwendung vom objektorientierten Paradigma und damit das Verschwinden der für uns zentralen Inhaltsmodelle (DuraSpace 2016) stellte uns vor erhebliche Herausforderungen, da der Verzicht auf dieses Prinzip eine massive Überarbeitung aller bestehenden Daten und Zugriffsmethoden bedingt hätte. Als Lösungsstrategie wurde eine Kompatibilitätsschicht eingebracht, die die bestehende objektbasierte API auf die neue graphenbasierte API abbildet. Diese Schicht wurde als Service implementiert, wie auch allgemein darauf geachtet wurde, das neue System als Sammlung einzelner, Docker-basierter Services zu realisieren, die nun in einem Kubernetes-Cluster laufen.

Diese Kompatibilitätsschicht ermöglicht die Übernahme aller bestehenden Projekte und Forschungsdaten ohne Änderungen an diesen notwendig zu machen. Dies wurde ausschließlich durch die einheitliche Abwicklung in den Jahren davor ermöglicht und zeigt die Notwendigkeit dieser Begrenzung mit Hinblick auf die nachhaltige Verfügbarkeit. Zurückblickend auf das Beispiel der mittelalterlichen Handschrift bedeutet dies, dass sich an der grafischen Oberfläche des Projektes für NutzerInnen nichts verändert, alle Links und Disseminatoren weiterhin funktionieren und auch die persistente Identifikation gewährleistet bleibt, obwohl im dahinter liegenden System kaum ein Stein auf dem anderen geblieben ist.

## Fazit

Der Betrieb jedes Repositoriums erfordert die periodische Erneuerung des Technologiestacks. Um eine solche Migration reibungslos zu ermöglichen, müssen die enthaltenen Daten und Projekte einem eng spezifizierten Rahmen genügen. Zusätzlich zu diesem praktischen Erfordernis muss das Repositorium als organisatorische Einheit ausreichend finanzielles und institutionelles Commitment beweisen, um eine solche massive und personalintensive Umstellung bewältigen zu können. Das Ersetzen von Kernkomponenten der Software-Infrastruktur kann damit als ultimativer Beweis für eine erfolgreiche Langzeitstrategie des Betreibers angesehen werden. Dies unterstreicht, dass ein Repositorium in erster Linie keine technische Lösung, sondern eine organisatorische Einheit ist, die bereit ist, aktiv Verantwortung für die enthaltenen Ressourcen zu übernehmen. Vertrauenswürdige Repositorien tragen diese Verantwortung nicht nur für die Wissenschaft, sondern für das kollektive kulturelle Gedächtnis und bilden somit den Grundstein eines "digitalen Gedächtnisses".

## Bibliographie

**Geisteswissenschaftliches Asset Management System (GAMS).** <https://gams.uni-graz.at> [letzter Zugriff 14. Juli 2021].

**Brickley, D. / Guha, R. V.** (2014): RDF Schema 1.1. <https://www.w3.org/TR/rdf-schema> [letzter Zugriff 14. Juli 2021].

**Confederation of Open Access Repositories (COAR)** (2020): COAR Community Framework for Good Practices in Repositories. <https://doi.org/10.5281/zenodo.4110829> [letzter Zugriff 14. Juli 2021].

**DuraSpace** (2016): Design - API Extension Architecture. <https://wiki.duraspace.org/display/FF/Design++api+extension+architecture> [letzter Zugriff 14. Juli 2021].

**Fielding, R.T.** (2000): Architectural Styles and the Design of Network-based Software Architectures. Ph.D. thesis, University of California, Irvine.

**Force 11** (2016): The Fair Data Principles. <https://www.force11.org/group/fairgroup/fairprinciples> [letzter Zugriff 14. Juli 2021].

**Lagoze, C. / Payette, S. / Shin, E.** (2005): Fedora: An architecture for complex objects and their relationships. Berlin/Heidelberg. <http://arxiv.org/ftp/cs/papers/0501/0501012.pdf> [letzter Zugriff 14. Juli 2021].

**Speicher S. / Arwe J. / Malhotra, A.** (2015): Linked Data Platform 1.0. <https://www.w3.org/TR/ldp/> [letzter Zugriff 14. Juli 2021].

**TEI Consortium** (2021): TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 4.2.2. <http://www.tei-c.org> [letzter Zugriff 14. Juli 2021].

**The Consultative Committee for Space Data Systems** (2012): Reference Model for an Open Archival Information System (OAIS). <https://public.ccsds.org/pubs/650x0m2.pdf> [letzter Zugriff 14. Juli 2021].