

## Datenmanagement

Alexandra Weissgerber\* und Niklaus Stettler

# Einsatz von Linked Data in Archivinformationssystemen – Chancen und Herausforderungen

<https://doi.org/10.1515/iwp-2020-2114>

**Zusammenfassung:** Dieser Beitrag zeigt auf, dass sich mit Linked Data wesentliche Schwächen bisheriger Archivinformationssysteme (AIS) überwinden lassen. Es wird ein am Schweizerischen Institut für Informationswissenschaft (SII) der FH Graubünden entwickeltes Pilotsystem vorgestellt, das dazu dient, Lösungen für aktuelle Herausforderungen aus der Praxis anzugehen. Beispielhaft wird gezeigt wie das Datenmodell flexibel für die Use Cases des Archivs erweiterbar ist und wie Materialien, die aufgrund des Datenschutzes für längere Zeit unter besonderen Schutz gestellt werden müssen, in einem Linked Data-AIS verwaltet werden können.

**Deskriptoren:** Archiv, Archivierung, Archivinformationssystem, Informationssystem, Linked Data, Open Access, Open Linked Data

### Use of Linked Data in Archive Information Systems – Opportunities and Challenges

**Abstract:** This article shows that the key weaknesses of previous archival information systems (AIS) can be overcome with linked data. A pilot system developed at the Swiss Institute for Information Science at the Graubünden University of Applied Sciences is presented, which is used to tackle current practical challenges. As an example, it is shown how the data model can be flexibly expanded for archival use cases and how materials that have to be placed under special protection due to data protection can be managed in a Linked Data-AIS.

---

\*Kontaktperson: Dr. Alexandra Weissgerber, Fachhochschule Graubünden, Schweizerisches Institut für Informationswissenschaft, Pulvermühlestrasse 57, 7000 Chur, Schweiz, E-Mail: alexandra.weissgerber@fhgr.ch

Prof. Dr. Niklaus Stettler, Fachhochschule Graubünden, Schweizerisches Institut für Informationswissenschaft, Pulvermühlestrasse 57, 7000 Chur, Schweiz, E-Mail: niklaus.stettler@fhgr.ch

**Descriptors:** Archive, Archiving, Archive information system, Information system, Linked data, Open access, Open linked data

### Utilisation de données liées dans les systèmes d'information des archives – opportunités et défis

**Résumé:** Cet article montre que les principales faiblesses des précédents systèmes d'information archivistique (AIS) peuvent être surmontées avec des données liées (linked data). Un système pilote développé à l'Institut suisse des sciences de l'information de l'Université des Sciences Appliquées des Grisons qui permet de relever les défis pratiques actuels est présenté. A titre d'exemple, le pilote montre comment le modèle de données peut être étendu de manière flexible pour les cas d'utilisation d'archivage et comment les archives qui doivent être placées sous protection spéciale en raison de la période de protection des données peuvent être gérées dans un AIS à l'aide de données liées.

**Descripteurs:** Archives, Archivage, Système d'information sur les archives, Système d'information, Données liées, Accès ouvert, Données liées ouvertes

## Einleitung

Aktuelle Archivinformationssysteme (AIS) werden mit zahlreichen neuen Herausforderungen der Beschreibung von Informationsobjekten konfrontiert. Herausfordernd ist insbesondere, dass laufend neue Typen von Informationsobjekten beschrieben und möglichst viele Daten aus anderen „Welten“ integriert werden sollten. Die dazu notwendige Flexibilität kann mit Systemen auf Basis von relationalen Datenbanken nicht erlangt werden.

Mit Linked Data (LD) ist das Problem dank der Möglichkeit, das Datenmodell flexibel zu erweitern, zu lösen. Anwendungsfälle einer solchen Erweiterung wären z. B.:

- nachträgliches Einfügen von Attributen zu einzelnen Entitäten

- flexibles Einfügen von Relationen zwischen Entitäten (z. B. zwischen Objekten und Akteuren)
- Verweis auf externe Wissensquellen wie Vokabularien oder Normdateien
- Zusammenarbeit mit anderen Gedächtnisorganisationen wie Archiven, Bibliotheken, Forschungsdatenrepositorien oder Museen.

Um dies zu ermöglichen, erarbeitet aktuell die Experts Group on Archival Description (EGAD) des International Council on Archives (ICA) ein neues Datenmodell beziehungsweise einen neuen Standard für die archivische Erschließung. Beteiligt an diesem Entwicklungsprozess ist auch eine Arbeitsgruppe des Vereins Schweizerischer Archivarinnen und Archivare, die nicht zuletzt darauf fokussiert, RiC<sup>1</sup> konkret umsetzbar zu machen. Tatsächlich erweist sich Linked Data – Archivinformationssysteme als vielversprechend, jedoch sind viele Detailprobleme der Praxis noch nicht gelöst. Um diese anzugehen, konnte das SII einen Piloten eines Archivinformationssystems erstellen. Im Folgenden werden kurz die Herausforderungen bei der Erschließung von Informationsobjekten im Archivkontext mit Linked Data vorgestellt. Sodann wird das Pilot-AIS kurz umrissen, um aufzuzeigen, welche Probleme gelöst werden konnten und wo noch weiterer Entwicklungsbedarf besteht.

## Aktuelle Herausforderungen

Die wohl meist beklagte Beschränkung klassischer AIS ist, dass sie die Zuordnung der einzelnen Objekte zu einer Position in der hierarchischen Struktur voraussetzen. Die monohierarchische Ordnung der Unterlagen entspricht oft nicht der Realität. Informationsobjekte (Records) können in verschiedenen Kontexten erarbeitet und genutzt werden, beziehungsweise unterschiedlichen Ordnern (Dossier, Akte) gleichzeitig zugehören. Die Lösung der Querverweise auf andere Ordner und Informationsobjekte ist umständlich und wenig befriedigend. Daher wird sie auch sehr selten genutzt.

Eine weitere Beschränkung von AIS ist die Verwaltung von Akteuren. Oft sind mehrere Akteure an Dossiers und Records in verschiedenen Rollen beteiligt. In aktuellen Systemen werden diese häufig in unterschiedlichen Feldern erschlossen, zum Beispiel Erwähnungen als Bestandesbildner im Titel oder in Feldern wie *Urheber* oder

*Thema*. Eine Suche nach allen Objekten, die eine Verbindung zu einem Akteur haben, ist nur bedingt möglich, diejenige nach Objekten, bei denen der Akteur eine bestimmte Rolle einnimmt, meist gar nicht. Oftmals sind auch die Beziehungen zwischen den Akteuren, wie z. B. Verwandtschaftsbeziehungen, nicht dokumentiert. Ein Datenmodell, in dem die Akteure (Agent im Datenmodell) eigene Entitäten darstellen, erlaubt es solche Beziehungen zu modellieren.

Unbefriedigend ist auch, dass aktuell kaum Wissen anderer Informationsdienstleister genutzt werden kann. Ein Bezug solcher Daten aus Informationssystemen anderer Gedächtnisorganisationen (Archive, Bibliotheken, Museen etc.) wäre wünschenswert. Die Verwendung existierender Normdaten (GND, VIAF) würde Datenkonsistenz sichern. Besonders angezeigt ist es für die Verschlagwortung von Objekten auf bereits vorhandene Vokabularien (z. B. AAT oder die in Entwicklung befindliche LD-Version von Termdat). Reiches Wissen ist in der Linked Open Data (LOD)-Cloud (Wikidata) auch zu den Akteuren vorhanden und sollte so weit als möglich genutzt werden.

Die erwähnten Anwendungsfälle sind die drei häufigsten, für die in diesem Beitrag eine Lösung vorgestellt wird. Auf zahlreiche andere kann hier aus Platzgründen nicht eingegangen werden. Daher sei nur nochmals kurz erwähnt, dass traditionelle AIS oft an ihre Grenzen stoßen, wenn es darum geht, neue Typen von Informationsobjekten zu beschreiben. Ein LD-Datenmodell erlaubt es, die Objekte in eine Netzstruktur einzubinden, und kann sehr flexibel mit neuen Beziehungen und Objekttypen erweitert werden.

## Prototypische Umsetzung

Die Autoren haben versucht, typische Anwendungsfälle mit der Graphdatenbank GraphDB prototypisch umzusetzen und dabei auftretende Probleme zu lösen. Für den Pilot wurde ein Datenmodell entwickelt, auf dessen Basis die Anwendungsfälle durchgeführt werden können. Die im Prototypen realisierten Entitäten sind am konzeptionellen Datenmodell von RiC orientiert. Zugleich galt es, das Modell möglichst einfach zu halten. Das Datenmodell soll die Aufnahme, der in bestehenden AIS enthaltenen Metadaten erlauben, verzichtet aber bewusst auf die Darstellung von Entitäten, die dort keine Entsprechung haben. Ziel war auch, dass das Modell das Potential für einfache Erweiterungen für allfällige weitere Umsetzungsversuche hat.

Abbildung 1 zeigt einen Ausschnitt des Datenmodells mit exemplarischen Daten, die für die Lösung der drei o. g.

<sup>1</sup> Records in Context: <https://www.ica.org/en/egad-ric-conceptual-model> [11.7.2020].

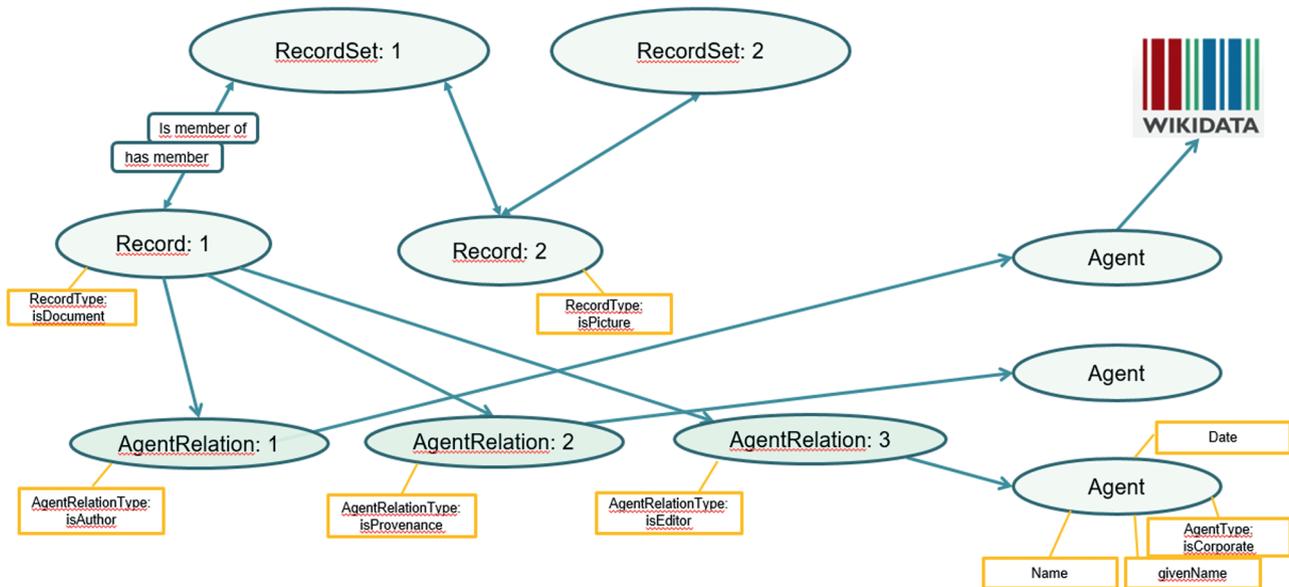


Abbildung 1: Ausschnitt des Datenmodells mit exemplarischen Daten.

Anwendungsfälle geeignet sind. Die Entitäten (RecordSet, Record, Agent, AgentRelation) werden mit Attributen beschrieben. Hier wird nur ein kleiner Ausschnitt davon wiedergegeben.

Die Graphik zeigt die Zuweisung von zwei Records zu ihren jeweiligen RecordSets, wobei der „Record 2“ beiden RecordSets zugehörig ist. Die Relationen zwischen RecordSet und Record sind sehr einfach realisiert, indem jedes RecordSet die Attribute „isMemberOf“ und „hasMember“ zugeschrieben wird. Damit werden Zusammenhänge zwischen Dossiers für die Nutzenden transparent. Der Recordtyp erlaubt es, unterschiedliche Objektarten mit spezifischen Attributen genau zu beschreiben.

Record 1 ist mit unterschiedlichen Akteuren in unterschiedlichen Rollen verbunden. Das Attribut „AgentType“ erlaubt es z. B. zwischen Agents, die eine Organisation repräsentieren, und Personen zu differenzieren. Die Beziehung des Agents zum Objekt wird mit dem Knoten „AgentRelation“ beschrieben. Dieser gibt nicht nur Auskunft über Zugänglichkeit und Zeitraum der Beziehung, sondern auch über deren Art. Das Attribut „AgentRelationType“ hat diese Erweiterung möglich gemacht. Dies erlaubt mehr Wissen zu einer Person zusammen zu tragen und damit z. B. zwischen den Rollen als Bestandesbildner, Autor, abgebildete Person, Beitrag über eine Person zu differenzieren. Diese Erschließung ermöglicht sehr differenzierte Abfragen nach allen Materialien zu einer Person, die nach den Rollen gefiltert werden können.

Im Piloten wird die Nutzung externer Quellen repräsentiert, indem Daten aus Wikidata verlinkt werden. Diese Erschließung erlaubt den Zugriff auf das immense Wissen, das in der LOD-Cloud vorhanden ist (analog zur Verlin-

kung mit Wikidata hier), ebenso wie die Verknüpfung mit Materialien anderer Gedächtnisorganisationen.

Der Pilot vermag so die häufigsten Herausforderungen in einem AIS abzudecken. Ein auf LOD aufbauendes AIS kann also eine echte Alternative zu den bisherigen Systemen darstellen, vorausgesetzt allerdings, dass auch eine Lösung für die Verwaltung von sogenannten Schutzfristen gefunden wird – eine Anforderung, die gerade in Archiven ein bisschen speziell ist.

## Herausforderung Schutzfristen

Archive sind im Besitz von vielen Unterlagen, die nicht öffentlich zugänglich sein dürfen. Typischerweise betrifft das Unterlagen, die dem Datenschutz unterworfen sind und daher gelöscht werden müssten. Archive können solche Daten jedoch vorhalten, wenn sie garantieren können, dass diese während der Schutzfrist nicht eingesehen werden können.

Wie können Unterlagen, die noch einer temporären Schutzfrist unterworfen sind, geschützt werden? Schutzfristen werden in Archivinformationssystemen typischerweise auf einer sogenannten Verzeichnungsstufe (Knoten) vergeben. Alle Objekte unterhalb dieses Knotens sind damit geschützt. Mit der Schaffung der Möglichkeit, dass Records mehreren RecordSets zugeordnet sein können (Netzstruktur), ist diese einfache Vererbung der Schutzfristen nicht mehr gegeben. Die Schutzfrist muss jedem Objekt mitgegeben werden.

In vielen Fällen ist es möglich, Schutzfristen auf Ebene der Entitäten zu definieren und so in der Applikation

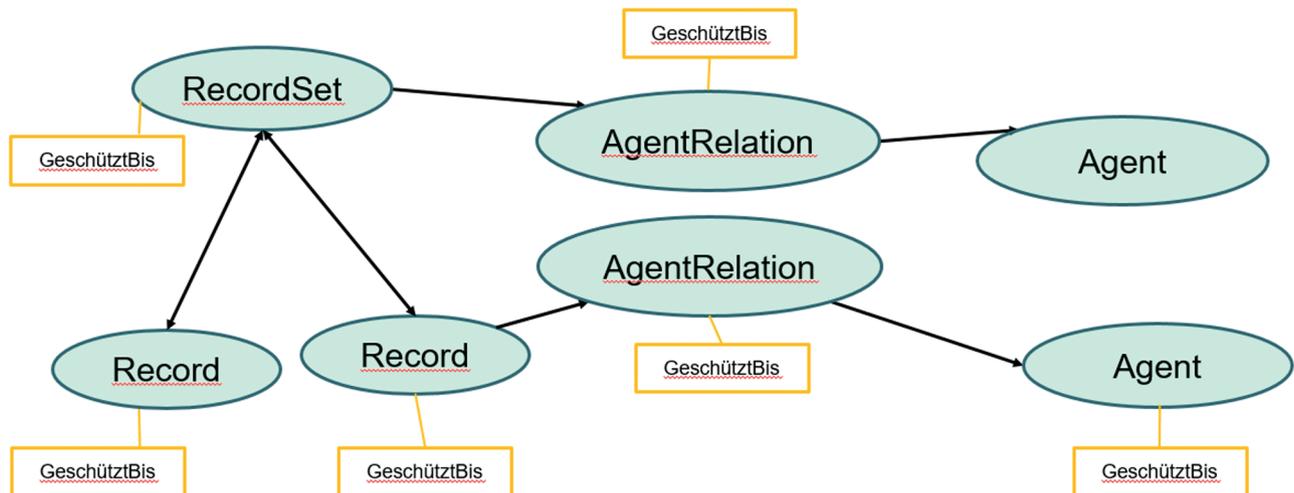


Abbildung 2: Schutzfristen als Attribute sowohl der Entitäten wie auch der Beziehungen.

zu steuern, welche Objekte öffentlich gezeigt werden können und welche nicht. Nicht selten betrifft die Schutzfrist aber nicht die Entitäten selbst, sondern die Beziehung zwischen denselben. Beispielsweise kann ein Objekt öffentlich sein, aber während der Schutzfrist darf im Katalog nicht nachgewiesen werden, dass dieses mit einem bestimmten Akteur verbunden ist.

Das im Piloten hinterlegte Datenmodell erlaubt daher jeder Entität (RecordSet, Record, Akteur wie auch die Beziehung zwischen Akteur und Record) das Attribut „geschütztBis“ zu vergeben.

Da Objekte in einer LD-Umgebung mit zahlreichen Entitäten verbunden sind, sind solche Einschränkungen häufig.

In GraphDB werden für die Datenablagen Datenbanken kreiert. Diese werden *Repository* genannt. GraphDB bietet eine Administrationsfunktion an, um Zugriffsrechte zu verwalten. Die Zugriffsrechte können auf Repository-Ebene angegeben werden, jedoch nicht auf Objekt- oder Beziehungsebene innerhalb eines Repository. Um Archivalien Suchenden den Zugriff auf geschützte Daten zu verweigern, müssen geschützte Daten in einem eigenen Repository gespeichert werden. Abbildung 3 zeigt die Datenbankstruktur mit dem öffentlichen und dem geschützten Repository.

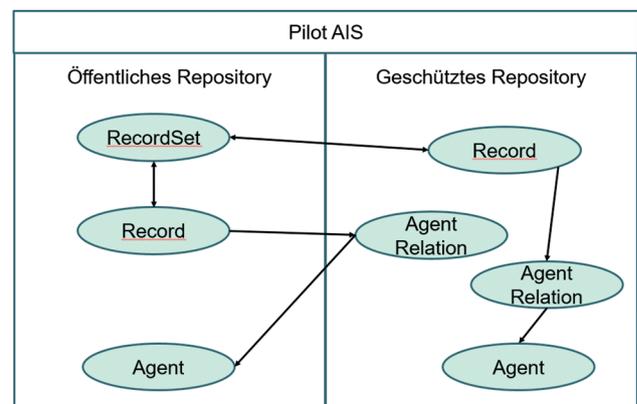


Abbildung 3: Organisation der Datenbank / Repositories.

Das öffentliche Repository umfasst alle RecordSets und Records, die keiner Schutzfrist unterliegen. Außerdem sind im öffentlichen Repository alle Personen- und Körperschafts-Datensätze (Agent) abgelegt, sowie weitere öffentliche Entitäten, die im Beitrag nicht thematisiert werden (z. B. Orte, Schlagworte, Zeiten, etc.).

Zum geschützten Repository haben nur Archivmitarbeitende Zugriff. Hier befinden sich die RecordSets, Records und Datensätze über Akteure, die Schutzfristen unterliegen. Außerdem sind im geschützten Repository auch die Datensätze der Beziehungen (AgentRelation) zwischen Records und Agents abgelegt, die einer Schutzfrist unterliegen.

Mit einer einfachen SPARQL-Abfrage können Objekte und ihre Beziehungen, deren Schutzfrist kürzlich abgelaufen ist, identifiziert werden. So können sie alle zusammen freigegeben werden, indem sie vom geschützten Repository in das öffentliche verschoben werden. Dieser Vorgang kann mit einem Massenmigrationskript, z. B. als

SPARQL-Abfrage, automatisiert und in regelmäßigen Abständen gestartet werden.

## Fazit und Ausblick

Mit dem Piloten konnte nachgewiesen werden, dass grundlegende Funktionalitäten eines AIS auch aufbauend auf der Technologie von LD umgesetzt werden können, und dass eine solche Umsetzung wesentlichen Mehrwert für Archiv und Nutzenden bringt. Die Herausforderung der Schutzfristen konnten mittels Zugriffsrechteverwaltung zufriedenstellend gelöst werden.

Sein volles Potential wird ein LD-AIS bei Neuzugängen entfalten können. Hier wird sich die Flexibilität des Datenmodells schnell auszahlen, wenn wesentlich mehr Metadaten aus den Produktivsystemen übernommen werden können bzw. die Erschließung von komplexeren Objekten strukturiert vorgenommen werden kann.

Bei den Altdaten werden – wie bei jedem Systemwechsel – die Vorteile des neuen AIS weniger zu Tage treten, da die Daten in alten Systemen oft in Freitextfeldern ‚versteckt‘ sind und daher nur unstrukturiert vorliegen. Der wohl wichtigste Schritt in dieser Richtung wäre die Aufbereitung der Altdaten, sodass z.B. Akteure, Orte, Schlagworte etc. identifiziert und konform zum neuen Datenmodell abgelegt werden können.

Weitere Arbeiten sollten insbesondere die folgenden zwei Punkte ins Zentrum rücken:

- Es sind Verfahren zu entwickeln, wie aus Altdaten möglichst automatisiert strukturierte Daten gewonnen werden können. In unterschiedlichen Projekten des SII erprobte Technologien wie Text Mining, Named Entity Recognition, Named Entity Linking und GeoTagging versprechen hier eine erfolgreiche Datenaufbereitung.
- Die Flexibilität der Datenmodelle in LD-Umgebungen ist zweifellos eine der großen Errungenschaften dieser Technologie für AIS. Damit diese voll zum Tragen kommen kann, ist es aber wichtig, dass die Modell-Erweiterungen in einem strukturierten Prozess ablaufen.



**Dr. Alexandra Weissgerber**  
 Fachhochschule Graubünden  
 Schweizerisches Institut für  
 Informationswissenschaft  
 Pulvermühlestrasse 57  
 7000 Chur  
 Schweiz  
[alexandra.weissgerber@fhgr.ch](mailto:alexandra.weissgerber@fhgr.ch)

Seit Dezember 2017 ist Alexandra Weissgerber als Dozentin an der FH Graubünden tätig. Sie unterrichtet Projektmanagement, Wirtschaftsinformatik und Big Data im Studiengang Digital Business Management. Sie studierte Informatik und Computerlinguistik. Ihre Hauptforschungsschwerpunkte sind Wissensmanagement und der Einsatz von humanoiden Robotern.



**Prof. Dr. Niklaus Stettler**  
 Fachhochschule Graubünden  
 Schweizerisches Institut für  
 Informationswissenschaft  
 Pulvermühlestrasse 57  
 7000 Chur  
 Schweiz  
[niklaus.stettler@fhgr.ch](mailto:niklaus.stettler@fhgr.ch)

Prof. Dr. Niklaus Stettler ist seit 2003 Dozent, seit 2005 Professor für Archivwissenschaft und Records Management. Von 2009 bis 2019 leitete Niklaus Stettler das Institut für Informationswissenschaft SII. Seine Arbeitsschwerpunkte sind Records Management sowie Digitale Langzeitarchivierung.