



Schweizerische Gesellschaft für Geschichte
Société suisse d'histoire
Società svizzera di storia
Societad svizra d'istorgia

Datenstandards für die historische Forschung

Ein White-Paper der SGG

Stand: 10.11.2023, Version 1.0

Standards pour les données numériques de la recherche historique. Livre blanc de la SSH

Dieses Dokument basiert auf einer Vorlage, welche durch die Arbeitsgruppe «Nachhaltige Datennutzung» der [Abteilung Grundlagenerschliessung](#) der [Schweizerischen Gesellschaft für Geschichte](#) (SGG) erarbeitet wurde. Dazu gehörten (teilweise zu unterschiedlichen Zeitpunkten) Simon Gabay, Tobias Hodel, Moritz Mähr, Stefan Nellen, Barbara Roth-Lochner, Pascale Sutter, Andrea Voellmin und Karin von Wartburg. Das Papier wurde in Workshops und durch Freiwillige in einem *open peer review* erweitert und kritisch diskutiert.

Wir danken allen Beteiligten für das grosse Engagement und die konstruktiven Diskussionen.

Das vorliegende Dokument wurde am 27.2022 von der Abteilung «Grundlagenerschliessung der SGG» verabschiedet und am 22.11.2022 vom Vorstand der SGG genehmigt.

Grundsätzliches

Das Dokument ist interaktiv und verweist mit Links auf die Überschriften, welche verwandte Themen abhandeln.

Zielgruppe

Die Zielgruppe dieses Papiers sind **Historiker:innen**, die mit Daten(-beständen) arbeiten und forschen, Dokumente erschliessen und online publizieren wollen sowie **Informatiker:innen**, die in geschichtswissenschaftlichen Projekten (mit-)arbeiten.

Ziel

Das Papier stellt ein **Glossar** und einen **Katalog** zur Verfügung mit den wichtigsten Standards zur Erzeugung von Interoperabilität, sodass digitale Geschichtsprojekte möglichst einfach von Dritten nachgenutzt und verlinkt werden können. Das Papier fokussiert auf technische Standards.

Interoperabilität bezeichnet die Anschlussfähigkeit und Austauschbarkeit von Daten und Informationen über unterschiedliche Systeme und Anbieter:innen hinweg. Dank der Ausrichtung an Standards und der Weitergabe eigener Daten(-stämme) wird eine Nachnutzung aktiv gefördert.

Die Orientierung an Standards erlaubt den Anschluss an inter- und transnationale Forschungsarbeiten, indem Forschungsdaten und -resultate weiterverwendet werden können. Gleichzeitig vereinfacht die Verwendung von Standards die Archivierung der Datenstämme.¹

Das Dokument hat zum Ziel, die Standards aus geschichtswissenschaftlicher Perspektive (in Anschluss an geisteswissenschaftlichen Perspektiven) widerzuspiegeln, ohne dass Wertungen bezüglich der verwendeten Technologien abgegeben werden.

Das Dokument besteht aus einer Einleitung, welche die grundlegenden Begriffe einführt, einem Glossar, das die geläufigsten Begriffe bestimmt, und einem Katalog, der die heute gebräuchlichen Standards beschreibt. Aufgrund des technischen Wandels ist weder das Glossar noch der Katalog vollständig oder langfristig gültig. Die datierte und publizierte Version ist ein Produkt ihrer Zeit. Überarbeitungen werden mit Anpassungen von Formaten und Technologien notwendig werden.

Die Unabgeschlossenheit des Dokuments soll gleichzeitig aufzeigen, dass nur der gelebte Umgang mit digitalen Methoden und Daten produktive Diskussionen und Weiterentwicklungen

¹ Wir verwenden den Begriff «Standard» in Abgrenzung zu «Norm». Unter Norm verstehen wir eine vergleichsweise einheitliche oder vereinheitlichte, weithin anerkannte und meist angewandte (oder zumindest angestrebte) Art und Weise, etwas herzustellen oder durchzuführen, die sich gegenüber anderen Arten und Weisen durchgesetzt hat. Eine Norm ist definiert als eine weithin faktisch (manchmal auch rechtlich) anerkannte und durch ein Normungsverfahren beschlossene, allgemeingültige sowie veröffentlichte Regel zur Regelung eines Sachverhaltes. Eine Norm wird beschlossen und veröffentlicht, nachdem sie alle Instanzen eines Normungsverfahrens durchlaufen hat. Eine solche Norm ist nach [EN 45020](#) ein «Dokument, das mit Konsens erstellt und von einer anerkannten Institution angenommen wurde und das für die allgemeine und wiederkehrende Anwendung Regeln, Leitlinien oder Merkmale für Tätigkeiten oder deren Ergebnisse festlegt, wobei ein optimaler Ordnungsgrad in einem gegebenen Zusammenhang angestrebt wird» (<<https://de.wikipedia.org/wiki/Normung>>, Stand: 23.12.2021). Mit der Normung sind oftmals weitere Ziele verbunden wie [Rationalisierung](#), Verminderung der Vielfalt, [Kompatibilität](#), [Gebrauchstauglichkeit](#) und [Sicherheit](#) (o.A.: Normung, in: Wikipedia. Online: <<https://de.wikipedia.org/wiki/Normung>>, Stand: 23.12.2021).

Ein Standard kann in einem formalisierten oder nicht formalisierten Regelwerk (in einer oder mehreren [Regeln](#) oder einer Norm) beschrieben sein oder sich ungeplant ergeben. In dieser Bedeutung ist der Begriff Standard insbesondere in den Bereichen [Technik](#) und [Methodik](#) üblich (<<https://de.wikipedia.org/wiki/Standard>>, Stand 21.03.2022). Im deutschen Sprachgebrauch ist in den letzten Jahren eine Begriffsverwirrung eingetreten, indem «Standard» analog zum englischen Begriff *standard* auch für Normen verwendet wird. Aus dem englischen Sprachgebrauch kommt der Begriff *de-jure-Standard*, der sich mit dem deutschen Begriff Norm deckt. Von De-facto- oder Quasi-Standards spricht man, wenn sich Methoden oder Regeln in der Praxis durchsetzen und nicht infolge von Vereinbarungen, Gesetzen, Verordnungen oder Ähnlichem gesetzt sind (<<https://de.wikipedia.org/wiki/Standard>>, Stand 21.03.2022).

von Theorie und Praxis ermöglichen. Entsprechend befürworten wir eine offensive, offene und häufige Publikation (selbst) von («dirty») Daten und einen kritischen Umgang mit Daten.

Einleitung

Die folgenden Seiten sollen als Wegweiser für geschichtswissenschaftliche Projekte mit digitalen Ansätzen dienen, die sich an forschungsfragen-geleiteten Workflows orientieren. In zwei Teilen wird eine Handreichung für technische Standards und aktuelle Herangehensweisen an Datenaufbereitung, -management und -auswertung gegeben. Damit möchten wir den Dialog zwischen Historiker:innen sowie Computerwissenschaftler:innen und Informationswissenschaftler:innen erleichtern.

Der erste Teil erklärt in Form von Definitionen zentrale Technologien und Ansätzen, die in den digitalen Geisteswissenschaften Verwendung finden. Im zweiten Teil werden in einem Katalog technische Standards beschrieben, an denen sich Projekte orientieren können.

Das vorliegende Papier beschäftigt sich mit **technischen Datenstandards** im Zusammenhang geschichtswissenschaftlicher Projekte. «Standard» bezeichnet hier eine vergleichsweise einheitliche oder vereinheitlichte, weithin anerkannte und meist angewandte (oder zumindest angestrebte) Art und Weise, etwas herzustellen oder durchzuführen, die sich gegenüber anderen Arten und Weisen durchgesetzt hat. Dabei spielt es keine Rolle, ob diese Standards *de jure* oder *de facto* zustande gekommen sind.

Daten

Der Plural Daten leitet sich aus dem lateinischen *datum* ab, was sich mit «Gegebenes» übersetzen lässt. Der Begriff Daten bezeichnet streng reglementierte, in Zeichen(-ketten) der sogenannten Syntax kodierte digitale Repräsentationen. Da der Begriff in unterschiedlichen Formen und Kontexten Verwendung findet, einige Bemerkungen dazu.

Durch semantische Anreicherung werden aus Daten Informationen. Die Vernetzung von Informationen führt zu (historisierbarem) Wissen. Daten werden hergestellt und aufbereitet. Die Herstellung und Aufbereitung von Daten ist folglich ein eigener, subjektiver Prozess. Daten sind so gesehen vielmehr *capta* (Erwishtes) oder *gar facta* (Gemachtes) als *data* (Gegebenes).²

Der Prozess der Anreicherung lässt sich als Wissenspyramide beschreiben:

- **Zeichen** werden mittels **Syntax** zu einer **Aussage** → Dies ergibt ein **Datum**
- **Daten** werden über **Semantik(en)** eine/mehrere Bedeutung(en) zugewiesen → Dies ergibt **Informationen**
- Informationen werden **verknüpft** und **gewertet** → Dies ergibt **Wissen**

Eine besondere Form von Daten sind **Forschungsdaten**. Darunter verstehen wir sämtliche Daten, die in einem Forschungsprozess hergestellt werden, z.B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen. Die

² Drucker, Johanna: Humanities Approaches to Graphical Display, in: Digital Humanities Quarterly 5 (1), 2011. Online: <<http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>>, Stand: 11.03.2016.

Verfügbarkeit und Nachnutzung digitaler Informationen schliesst den möglichst kostenfreien und barrierearmen Zugang zu Forschungsdaten ein. Entsprechend gehören sowohl Primär- als auch Metadaten in diese Kategorie.

Der Begriff **Primärdaten** ist in der Geschichtswissenschaft nicht gebräuchlich, vielmehr wird von Quellen gesprochen, unabhängig davon, ob diese digital oder analog vorliegen. Aus Gründen der Konsequenz wird der Begriff Primär- bzw. Urdaten auf den Seiten dieses Dokuments mit dem Begriff «Quellen» gleichgesetzt. Folglich kann darunter ein Bild, eine Karte, ein Gegenstand oder ein Text verstanden werden. Grundsätzlich gibt es keine materielle Form, die nicht als Quelle und entsprechend als Primärdatum verstanden werden kann.

Wichtig ist die Unterscheidung zwischen analogen und digitalen Quellen. *Digital-born* Daten sind historische Quellenstücke, die für ein geschichtswissenschaftliches Forschungsthema ausgewertet werden sollen und bereits in digitaler Form hergestellt wurden.³ Im Gegensatz dazu stehen analoge (= physisch, nicht binär abgespeicherte Daten), die erst durch Digitalisierung (Scanning, Texterkennung, Bildverarbeitung etc.) elektronisch verfügbar gemacht werden.

Das Gegenstück zu den Primärdaten bilden **Metadaten**, worunter wir mehr verstehen als technische Metadaten, die sich auf digitale Dokumente beziehen. Metadaten sind für uns Beschreibungen von Daten, die sich auch auf analoge Dokumente beziehen.⁴ Insbesondere für Metadaten ist die Nutzung von Standards zentral, da dadurch ein reibungsloser Austausch über Plattformen und Institutionen hinweg gewährleistet werden kann. Metadatenstandards sind Spezifikationen, die Metadaterme detailliert beschreiben. Sie dienen der Festlegung von Elementen und von deren Struktur und Bedeutung und ermöglichen die semantische Interoperabilität zwischen Anwendungen, die diesen Standard verwenden. Ziel von Metadatenstandards sind die einheitlichen, maschinenverarbeitbaren Beschreibungen von Ressourcen und der plattform- und institutionsübergreifende Austausch dieser Beschreibungen.⁵

Zur Vernetzung von Daten und Wissensbeständen werden im Bibliotheks-, Informations- und Dokumentationswesen **Normdaten** angelegt und verwendet (als Deskriptor in der Dokumentation). Eine Normdatei ist damit eine Form eines kontrollierten Vokabulars, in dem festgelegt wird, welche Ansetzung bei der Erschliessung zu verwenden ist. Im Englischen werden Normdateien als «authority files» bezeichnet. Normdaten dienen zur Identifikation einer Entität (Person, Ort etc.).

Daten können aufbereitet, verarbeitet, angereichert oder strukturiert bzw. modelliert werden. Unter **Datenaufbereitung** versteht man die Beschreibung, Transkription oder andere Formen der Zuschreibung von Bedeutung. Dazu zählen wir auch die **Datenanreicherung** bzw.

³ Beispiele für digital-born Daten: Videospiele; Protokolle, die als Worddokumente abgespeichert wurden; Fahndungsdatenbanken.

⁴ Im Gegensatz dazu, das Verständnis von Metadaten im DaSCH. Dort wird ein grundlegendes Set an Informationen zu einem Projekt als Metadaten verstanden.

⁵ Hartmann, Sarah: Metadatenstandards. Teil 2 der Einführung in die Interoperabilität von Metadaten und Metadatenformaten, Göttingen 05.11.2010. Online: <http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Schulungen/InteroperabilitaetVonMetadaten/teil21.pdf?__blob=publicationFile>, Stand: 11.05.2017. Beispiele für Metadatenstandards: → [METS/MODS](#), → [MARCXML](#), → [EAD](#).

Datenstrukturierung, die (bestenfalls) gemäss Standards erfolgt und zum Ziel hat, Daten um Informationen anzureichern.

Da Daten eine zentrale Grösse in digitalen Projekten darstellen, sind **Datenformate** von immenser Bedeutung, um Anschlussfähigkeit und Datenaustausch zu gewährleisten. In den Anfängen des digitalen Zeitalters stellten Datenformate ein Hindernis dar, da vielfach auf proprietäre Formate zurückgegriffen wurde. Seit einigen Jahren erfolgt eine Verständigung auf freie und/oder maschinenlesbare Formate (PDF → [XML](#)). In diesem Papier werden vorwiegend die freien und maschinenlesbaren Formate/Systeme XML und RDF beschrieben.⁶

Der Entscheid für freie Datenformate erleichtert die langfristige Aufbewahrung und vereinfacht die Datennachnutzung.

Offene, strukturierte Daten sind ein Schritt zu einem intelligenteren World-Wide-Web, das nach Tim Berners-Lee als **Semantic Web** beschrieben werden kann. In diesem Netz sollen Informationen in einer Art und Weise strukturiert werden, die eine bessere Verarbeitung durch Maschinen ermöglicht, als dies bislang der Fall ist. So soll die automatische Verarbeitung von Informationen vereinfacht und verbessert werden. → RDF, → [RDFS](#) und → [OWL](#) bilden grundlegende Bestandteile des Semantic Web. Semantic Web Technologien versprechen die Extraktion von Informationen anstelle von extrahierten Texten oder Textteilen. Entsprechend sind potentere Abfrage- und Auswertungsmechanismen denkbar als etwa durch relationale Datenbanken möglich sind.⁷

Datenaufbereitung und Daten(re-)präsentation

Die Trennung von Datenaufbereitung (Datenstrukturierung /-modellierung und Erfassung) und Datenpräsentation hat sich etabliert. Da Erwartungen in Bezug auf die Darstellung von Daten stark im Fluss sind, erfordert sie eine regelmässige Anpassung der Präsentation, die jedoch nichts an der Aufbereitung und Strukturierung der Daten ändern soll. Umgekehrt sollen auch Anpassungen im Datenmodell keine Auswirkungen auf die Präsentation haben.

Die strikte Trennung von Präsentation und Daten führt dazu, dass aufgrund derselben Datengrundlage unterschiedliche Präsentationen aufgesetzt werden können, die sich jeweils an ein unterschiedliches Publikum richten.⁸

Im Buchzeitalter wurde auf die visuelle Wiedergabe von Informationen (Textsatz, Tabellen etc.) fokussiert. Digitale Daten zeichnen sich jedoch dadurch aus, dass sie mit ihren

⁶ Die vorgeschlagenen/beschriebenen Datenstandards entsprechen nicht den Vorgaben der [KOST](#) zur langfristigen Aufbewahrung von Daten.

⁷ Kuczera, Andreas: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi., Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte, 05.05.2015, <<http://mittelalter.hypotheses.org/5995>>. Sogenannte Friend-of-A-Friend (FOAF) Abfragen sind in relationalen Datenbanken nur mit grossem Aufwand möglich und selten tatsächlich realisiert. Kuczera, Andreas: Das Deutsche Textarchiv in der Graphenwelt, Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte, 04.04.2017, <<http://mittelalter.hypotheses.org/10025>>.

⁸ Analog zu: Sahle, Patrick: Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 2: Befunde, Theorie und Methodik., Bd. 2 / 3, Norderstedt 2013 (Schriften des IDE 8). Online: <<http://kups.ub.uni-koeln.de/5352/>>. Dort bezogen auf digitale Editionen: Es handelt sich bei jeglichen Ausgaben jeweils nur um eine mögliche Darstellungs-/Präsentationsart.

Verknüpfungen und in ihrer Gänze nur verlustbehaftet visuell wiedergegeben werden können: Jede Präsentation ist nur eine mögliche Form der Datenwiedergabe.

Bei allen Hinweisen auf Standardisierung und Strukturierung bleibt der Hinweis auf einen kritischen Umgang mit digitalen Daten und Infrastrukturen jeglicher Art zentral. Standards und Infrastrukturen widerspiegeln epistemologische Grundannahmen und sehen (implizite) Methoden und Praktiken vor. Entsprechend gilt es nach Machtstrukturen zu fragen, die durch die Daten, die Datenstrukturierung und -wiedergabe gestärkt oder unterstützt werden. Die intersektionale Forschung in den Digital Humanities zielt aktuell vor allem auf die Sichtbarmachung und damit offensive Diskussion solcher Probleme.⁹

Glossar

Im Glossar werden zentrale Begriffe kurz definiert. Die Relevanz für Projekte in den Geschichtswissenschaften sind kursiv in der rechten Spalte platziert. Grau unterlegt folgen jeweils Beispiele.

⁹ Siehe insbesondere D'Ignazio, Catherine; Klein, Lauren F.: Data Feminism, Cambridge, Massachusetts 2020. Online: <<https://mitpressonpubpub.mitpress.mit.edu/data-feminism>>, Stand: 18.03.2022. Jarrett, Kylie: The Digital Housewife. Feminism, Labour and Digital Media, 2016. Losh, Elizabeth; Wernimont, Jacqueline: Bodies of Information: Intersectional Feminism and the Digital Humanities, 2018 (Debates in the Digital Humanities). Online: <<https://dhdebates.gc.cuny.edu/projects/bodies-of-information>>, Stand: 18.03.2022. Russell, Legacy: Glitch Feminism. A Manifesto, 2020 und Ortolja-Baird, Alexandra; Nyhan, Julianne: Encoding the haunting of an object catalogue: on the potential of digital technologies to perpetuate or subvert the silence and bias of the early-modern archive1, in: Digital Scholarship in the Humanities, 19.10.2021, S. fqab065. Online: <<https://doi.org/10.1093/llc/fqab065>>, Stand: 18.03.2022.

<h1>Definitionen</h1>	
<h2>Daten</h2>	
<p>Daten sind einzelne Fakten, Zeitpunkte, kalendarische Angaben etc., die durch Beobachtung, Messung oder andere Aufzeichnungsformen erfasst worden sind und als zeichenförmige Informationen vorliegen.</p> <p>Daten können in unterschiedlichen Formen und Strukturierungsgraden erfasst werden. Das Verständnis von Strukturierung variiert dabei stark und es gibt keine genauen Abgrenzungen zwischen den verschiedenen Graden (typischerweise unstrukturierte, semistrukturierte und strukturierte Daten).</p>	
<h2>Dirty/noisy und clean Data</h2>	
<p>Daten entstehen nicht aus dem nichts, sondern werden erhoben. Die Resultate solcher Prozesse sind häufig sog. <i>dirty/noisy data</i>, also «unsaubere» Daten. Auch diese Daten(-stämme) können sich als wertvoll und auswertbar für die Forschung erweisen. Datensäuberungsprozesse können genutzt werden, um die erhobenen Daten in eine Form zu bringen, die für eine Forschungsfrage zielführender/relevanter ist, sog. <i>clean data</i>. Prozesse der Datenreinigung führen jedoch immer zu Verlusten und Umwandlungen, die in der Forschung intensiv diskutiert werden.¹⁰</p>	
<h2>Unstrukturierte Daten</h2>	
<p>Unstrukturierte Daten liegen in einer nicht formalisierten Struktur vor, wie etwa Texte aus Digitalisierung und Texterkennung oder Textverarbeitungsprogrammen sowie die meisten Bild-, Audio- und Videodateien. Sie verfügen über kein Datenmodell, das über technische Metadaten (Anlage und Verarbeitung der Datei) hinausgeht. Im Rahmen der → Datenaufbereitung wird mittels verschiedener Formen versucht, den Daten Struktur(en) zuzuordnen. Automatische und halbautomatische Verfahren (häufig mit computerlinguistischem Hintergrund) können bei der Strukturierung unterstützend eingesetzt werden.</p>	<p><i>Unstrukturierte Daten stammen häufig aus umfangreichen Digitalisierungsvorgängen und sind Steinbrüche für das Auffinden von Bildern oder Texten. Quantität nicht Qualität ist das Hauptmerkmal der Datenform.</i></p>

¹⁰ Siehe insbesondere Rawson, Katie; Muñoz, Trevor: Against Cleaning (Chapter 23), in: Gold, Matthew K.; Klein, Lauren F. (Hg.): Debates in the Digital Humanities 2019. Online: <<https://doi.org/10.5749/j.ctvg251hk>>, Stand: 21.03.2022.

<p>OCR gelesene Texte liegen in einer un- bzw. wenig strukturierten Form vor. TXT-Dateien sind ebenfalls unstrukturierte Textablagen. Ein weiteres Beispiel unstrukturierter Daten sind Texte in Google Books.</p>	
<p>Semi-strukturierte oder schwach strukturierte Daten</p>	
<p>Semistrukturierte Daten sind nicht strukturiert, aber im Gegensatz zu unstrukturierten Daten können sie entweder Metadaten enthalten, etwa Bilder (Ort, Aufnahmedatum usw.) oder E-Mails (Empfänger, Datum usw.), und/oder andererseits Tags, wie XML-codierte Texte oder JSON-Dateien.</p> <p>Zwischen strukturierten und unstrukturierten Daten gibt es ein weites Feld an Ausprägungen. Die Struktur der Daten kann nur teilweise definiert sein, z.B. in einem Schlüssel/Wert-Paar, d.h. es ist nur der Sinn des Schlüssels oder in einer Tabelle nur die Bedeutung der Spalte bekannt, aber nicht die des Inhalts bzw. dieser liegt in reiner Text-form vor. Solche Daten sind typischerweise XML-kodierte Texte, Tabellen oder Metadaten zu digitalen Ressourcen.</p>	
<p>Halbstrukturierte Daten: Textedition der Jahrrechnungen der Stadt Basel 1535 bis 1610 (https://gams.uni-graz.at/context:srbas, Stand: 23.12.2021)</p>	
<p>Strukturierte Daten</p>	
<p>Strukturierte Daten liegen in einer ausserhalb der Daten beschriebenen sinngebenden Struktur vor und verlinken möglichst eindeutig identifizierte Entitäten. Typischerweise handelt es sich um relationale Datenbanken und <i>linked data</i> (teilweise basierend auf Thesauri und → Ontologien).</p>	<p><i>Strukturierte Daten lassen sich schnell und häufig ohne grösseren Aufwand weiterverarbeiten und auswerten. Zentral an strukturierten Daten ist die Dokumentation der Strukturierung, sodass nachvollzogen werden kann, welche Strukturierungsentscheide aufgrund welcher Annahmen/Gegebenheiten getroffen wurden.</i></p>
<p>Hochgradig strukturierte Daten: Amtliches Gemeindeverzeichnis (https://www.bfs.admin.ch/bfs/de/home/grundlagen/agvch.html, Stand: 23.12.2021)</p>	

Zeichensätze/Unicode	
<p>Unicode ist ein internationaler Standard, in dem langfristig für jedes sinntragende Schriftzeichen oder Textelement aller bekannten Schriftkulturen und Zeichensysteme ein digitaler Code festgelegt wird. Ziel ist es, die Verwendung unterschiedlicher und inkompatibler Kodierungen in verschiedenen Ländern oder Kulturkreisen zu beseitigen. Unicode wird ständig um Zeichen weiterer Schriftsysteme ergänzt.¹¹</p> <p>Zur Wiedergabe der Sonderzeichen, z.B. der kombinierten Zeichen, wird die Unicode Normalization Form D (NFD) verwendet.</p>	<p><i>Die korrekte und nachhaltige Abbildung von Zeichen ist für Texteditionen zentral. Um Austauschbarkeit zu ermöglichen, sind (universelle) Standards anzuwenden.</i></p>
<p>Fehlende Sonderzeichen können beispielsweise via MUFi (Medieval Unicode Font Initiative) codiert werden. Es gibt sehr viele Fonts, die frei zugänglich sind, z.B. Andron Scriptor https://mufi.info/m.php?p=mufi&i=968. Weitere umfangreiche Schriften mit grossem Zeichensatz und zusätzliche Werkzeuge zur Anpassung von Tastaturbelegungen findet man auf sil.org.</p>	

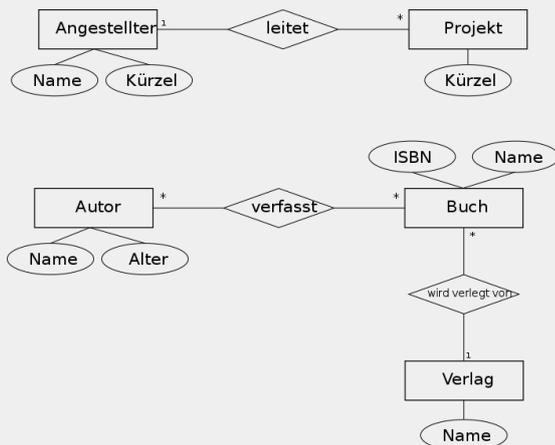
¹¹ o.A.: Unicode, in: Wikipedia, o.D., <<https://de.wikipedia.org/wiki/Unicode>>, Stand: 28.04.2017.

Auszeichnungs- und Strukturierungsformen

Datenmodell

Ein Datenmodell ist ein formales Modell der zu beschreibenden und zu verarbeitenden Daten eines Anwendungsbereichs (z.B. einer Datenbank, einer Edition oder aller für ein Forschungsprojekt relevanten, zu erhebenden Daten) und ihrer Beziehung zueinander.¹² Datenmodelle ermöglichen komplexe maschinelle Operationen mit den Daten. Wenn sie dabei Standards folgen, ermöglichen sie darüber hinaus auch den Austausch oder das Zusammenführen von Daten und tragen so zur Nachhaltigkeit bei.¹³

Datenmodelle operationalisieren Forschungsfragen. Sie ermöglichen es, Forschungsfragen digital zu prozessieren und sichern eine höhere Qualität der Daten, indem sie Bedingungen formulieren, denen Daten bei der Eingabe entsprechen müssen.



Quelle: Körner, Peter: Zwei (kleine) Beispiele für Entity Relationship Diagramme als SVG, online: <https://de.wikipedia.org/wiki/Datei:Er-diagramm.svg>

Datenbank

Datenbanken und Datenbankmanagementsysteme sind Systeme zur Datenverwaltung. Je nach Anwendungsszenario unterscheiden sich ihre Aufgaben. Dazu gehören die effiziente, widerspruchsfreie, dauerhafte Speicherung und Darstellung, bzw. Auswertung von Daten. Wichtige Formen sind relationale Datenbanken und graphbasierte Datenbanken. Viele Datensysteme garantieren für die basalen Transaktionen (Create, Read, Update, Delete) gewisse Eigenschaften (Atomicity, Consistency, Isolation, Durability).

¹² o.A.: Datenmodell, in: Wikipedia, o.D., <<https://de.wikipedia.org/wiki/Datenmodell>>, Stand: 21.12.2021.

¹³ Jannidis, Fotis; Kohle, Hubertus; Rehbein, Malte (Hg.): Digital Humanities. Eine Einführung, Stuttgart 2017, S. 100.

<h2>Relationale Datenbank</h2>	
<p>Relationale Datenbanken zeichnen sich dadurch aus, dass sie aus verknüpften, zweidimensionalen Tabellen gebildet werden. Relationale Datenbanken sind seit längerer Zeit erprobt und stabil, aber anders als graphbasierte Datenbanken typischerweise unflexibler bei der Erweiterung des Datenbankmodells.</p> <p>Relationale Datenbanken weisen eine hohe Datenkonsistenz auf. Normalisierte relationale Datenbanken ermöglichen eine widerspruchsfreie Datenhaltung und tragen somit zur Datenkonsistenz bei. Relationale Datenbanksysteme bieten zudem Funktionen, mit denen sich Integritätsbedingungen definieren und automatisch überprüfen lassen. Bis heute wird die elektronische Datenverwaltung vom relationalen Datenbankmodell dominiert.</p>	<p><i>Forschungsdatenbanken sind typischerweise auf konkrete Fragestellungen ausgerichtet und erlauben die Befüllung von klar definierten Feldern. Die Anschlussfähigkeit und die Interoperabilität müssen bei der Anlage berücksichtigt werden.</i></p>
<p>Ein langjähriges Datenbankprojekt ist das Repertorium Academicum Germanium (http://www.rag-online.org/, Stand: 23.12.2021), das elaborierte Abfragemöglichkeiten innerhalb der Datenbank bietet.</p>	
<h2>Graphbasierte Datenbanken</h2>	
<p>Graphbasierte Datenbanken definieren sich über die Vernetzung von einzelnen Objekten (und nicht über Tabellen).</p> <p>Informationen können in Form von Graphen etwa als Triple in RDF gespeichert werden, also in Form vieler simpler Gleichungen bestehend aus den drei Teilen Subjekt–Prädikat–Objekt.</p> <p>Ein Graph meint also eine Sammlung von Aussagen in Form von Vernetzungen, womit mehrere Eigenschaften oder Beziehungen zu einem Objekt beziehungsweise zwischen Objekten ausgedrückt werden können.</p> <p>RDF (Resource Description Framework) ist ein von der W3C verabschiedetes Modell zur flexiblen Repräsentation von Daten als Triple (Dreiheiten), bestehend aus Subjekt (zu beschreibende Ressource), Prädikat (Eigenschaft) und Objekt (Wert).¹⁴ Mit RDF können einfach Beziehungen zwischen Ressourcen abgebildet werden (Verweise). Mit RDF repräsentierte Daten bilden Netzwerke. Um → Vokabulare und → Ontologien</p>	<p><i>Die Bündelung zu Graphen erlaubt gleichzeitig mehrere Aussagen über Entitäten analog zu Datenblättern in Datenbanken.</i></p> <p><i>Mit RDF lassen sich flexibel Wissensstrukturen wiedergeben, ohne dass Normalisierungen/Anpassungen an vorgegebene Datenstrukturen (wie etwa in relationalen Datenbanken) zwangsläufig notwendig sind.</i></p>

¹⁴ Siehe die Website der W3C: <<https://www.w3.org/RDF/>>, Stand: 21.03.2022. Aktuell wird bei der W3C an RDF-star and SPARQL-star gearbeitet. Dadurch können Aussagen, z.B. hinsichtlich Provenienz von Informationen/Annotationen, über ein Triple gemacht werden. Siehe https://w3c.github.io/rdf-star/cg-spec/editors_draft.html, <https://www.ontotext.com/knowledgehub/fundamentals/what-is-rdf-star/>, Stand: 21.03.2022.

<p>zu definieren und abzubilden, existiert → RDFS (RDF-Schema für einfache Fälle) bzw. die Web Ontology Language (→ OWL zur Bildung komplexer Vokabulare).</p> <p>Triplestore nennt man eine Datenbank, die zur Aufbewahrung von (RDF-)Triples gebaut ist. Ähnliche Aufgaben erfüllen Graphendatenbanken, die jedoch weniger spezifisch auf RDF-Technologien zugeschnitten sind. Die Auslieferung der Triples erfolgt über «Query Languages» (z.B. → SPARQL).</p> <p>Graphdatenbank sind gegenüber optimierten relationalen Datenbanken weniger performant und entsprechend langsamer. Die Modellierung erfolgt explizit und wird dadurch komplexer.</p>	
<p>Subjekt–Prädikat–Objekt: Person X – ist Autor von – NN. Damit wird der Aussagesatz gebildet: «Max Frisch» (Subjekt) «ist der Autor von» (Prädikat) «Homo Faber» (Objekt). Ein Editionsprojekt, das RDF als Technologie nutzt: Euler Edition der «Christian Apocryphal Literature»: <https://www.unil.ch/irsb/fr/home/menuinst/recherche/editer-la-litterature-apocryphe-chretienne-1.html>, Stand: 18.03.2022.</p>	
<h3>XML (eXtensible Markup Language)/Textauszeichnung</h3>	
<p>XML ist eine Auszeichnungssprache, die verwendet wird, um den Austausch von Daten zu erleichtern. In XML können die Daten unterschiedliche Formen annehmen, etwa als Metadaten (z.B. bibliografische Informationen) oder auch als Daten selbst (z.B. Editionen).</p> <p>Mit Textauszeichnung ist die strukturierte Auszeichnung (häufig nach Schemata) von digitalen Texten gemeint, meist in Form von XML. XML (Extensible Markup Language) ist eine Metasprache, mit der Markup-Sprachen definiert werden können, insbesondere ihre Elemente, deren Attribute und deren Beziehungen zueinander.</p> <p>Mit XML können Texte strukturiert, mit Informationen angereichert und maschinell lesbar gemacht werden; dabei bleiben die Dateien auch für Menschen lesbar. Geschieht die Auszeichnung nach definierten Standards, können die annotierten Dokumente ausgetauscht und vernetzt werden. Mit XML-Schemata lassen sich Markup-Sprachen formal definieren. XML-basierte Markup-Sprachen bzw. unter ihrer Verwendung erstellte Dokumente sind geordnete Hierarchien (Baumstrukturen). XML ist eine offene Sprache, weshalb die Daten nur mit Schemata bzw. der Verwendung von Standards verstanden und nachgenutzt werden können.</p>	<p><i>Das freie Format und die Maschinenlesbarkeit machen XML zu einer sinnvollen Aufbereitungsform von strukturierten Daten, die als Texte vorliegen.</i></p>
<p>Texteditionen werden heute typischerweise in XML umgesetzt und nach Vorgaben spezifischer Auszeichnungssprachen (→ TEI, → CEI, → MEI) getaggt.</p>	

<p>Struktur einer typischen XML-Datei nach TEI:</p> <pre> <TEI> <teiHeader> <!-- Metadaten--> </teiHeader> <body> <text> <!-- Editionstext--> <p> <orgName> <!-- ausgezeichnete Entität--></orgName> </p> </text> </body> </TEI> </pre>	
Schema	
<p>Ein Schema ist eine formale Beschreibung der Struktur von Daten. Schemata sind sprachlich artikulierte Definitionen, zum Aufbau einer Datei und definieren, wie unterschiedliche Teile der Auszeichnung einer → Ontologie miteinander zusammenhängen. «Ein Schema erlaubt es [...], bei [...] der Erfassung von Daten gemeinsame Regeln zu verwenden, und damit sicherzustellen, dass bestimmte Informationen auch erfasst werden oder bestimmte Strukturen ausgeschlossen sind.»¹⁵</p> <p>Schemata dienen der → Interoperabilität.</p> <p>In XML und RDF lassen sich Schemata anwenden.</p> <p>Das Schema validiert, dass die Daten wohlgeformt (<i>well formed</i>), aber nicht unbedingt gültig (<i>valid</i>) sind, d. h. das Schema kontrolliert nicht, ob die Daten den Regeln des Modells (z. B. der Semantik der TEI) entsprechen.</p>	<p><i>Erst durch die Anwendung von Schemata wird (automatisiert) nachvollziehbar, welche Auszeichnungen und Kombinationen von Auszeichnungen erlaubt, beziehungsweise möglich sind.</i></p>
XML Schema	
<p>Zur Kontrolle und Validierung von XML-Dateien können Schemata verknüpft werden. Dadurch wird es möglich, gewisse Auszeichnungen nur an klar definierten Stellen zu erlauben und damit Austausch und Interpretation der Daten zu erleichtern. Schemata sind restriktiv und bilden die Syntax von XML-Codes. Schemata unterstützen darüber hinaus die Datenerfassung, indem XML-Editoren kontextabhängig erlaubte Werte anbieten.</p>	

¹⁵ Jannidis, Fotis; Kohle, Hubertus; Rehbein, Malte (Hg.): Digital Humanities. Eine Einführung, Stuttgart 2017, S. 135.

<p>Es ist zwischen den Sprachen zur Validierung der Grammatik (XML Schema, Relax NG) und Schematron zu unterscheiden, wobei Letztere als Ergänzung zu Ersteren zu verstehen ist.</p> <p>Im Falle der TEI wird das Schema automatisch aus ODD (One Document Does it all) erzeugt, das es ermöglicht, dem Schema eine in natürlicher Sprache verfasste Dokumentation zuzuordnen.</p>	
<p>RDF Schema</p>	
<p>RDFS (RDF-Schema) ist eine Erweiterung von → RDF zur Bildung von (einfachen) → Vokabularen und → Ontologien. Damit lassen sich Klassen (Typen) von Ressourcen und ihre möglichen Eigenschaften definieren. Die Notation basiert auf RDF.¹⁶</p>	
<p>Im Kontext von Metadaten bedeutet die Anwendung von Schemata, deren maschinellen systemübergreifenden Austausch unter möglichst geringem Aufwand.¹⁷</p>	
<p>Ontologie</p>	
<p>Ontologien sind strukturierte Daten, die anhand eines eindeutig definierten und dokumentierten → Datenmodells den historischen Diskurs über die Wirklichkeit, d.h. Informationen über Entitäten (Orte, Konzepte, Berufe etc.), deren Merkmale und Beziehungen (Äquivalenz, Hierarchie etc.) samt den Metadaten zu ihrer Definition und Entstehung abbilden.</p> <p>Einzelne Informationen/Merkmale/Beziehungen können auf → Vokabularen bzw. → Thesauri beruhen.</p>	<p><i>Mit Hilfe von Ontologien sind Informationen explizit (ohne verborgene Annahmen), unifiziert und maschinenlesbar.</i></p>
<p>Vokabular</p>	
<p>Ein (kontrolliertes) Vokabular ist eine Schlagwortliste mit einer Anzahl von Wörtern, die definiert vorliegen bzw. zu Projektbeginn definiert werden.</p>	

¹⁶ Siehe dazu die Definition der W3C: <<https://www.w3.org/TR/rdf-schema/>>, Stand: 11.05.2017.

¹⁷ Hartmann, Sarah: Metadatenstandards. Teil 2 der Einführung in die Interoperabilität von Metadaten und Metadatenformaten, Göttingen 05.11.2010. Online: <http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Schulungen/InteroperabilitaetVonMetadaten/teil21.pdf?__blob=publicationFile>, Stand: 11.05.2017.

<h2>Thesaurus</h2>	
<p>Thesauri sind kontrollierte Vokabulare, die so strukturiert sind, dass sowohl hierarchische als auch netzwerkartige Beziehungen ausgedrückt werden (können). Ein Thesaurus soll ein Themengebiet beschreiben und repräsentieren.</p>	<p><i>Thesauri helfen der wissenschaftlich fundierten inhaltlichen Erschliessung von Texten und erlauben die Suche und Auszeichnung mit heutigen Begrifflichkeiten.</i></p>
<p>Jurivoc: Dreisprachiger Thesaurus des Schweizerischen Bundesgerichts und des ehemaligen Eidgenössischen Versicherungsgerichts.</p>	
<h2>Adresse/Adressierbarkeit</h2>	
<p>Um Daten gezielt aufrufen zu können, muss eine Adressierbarkeit gegeben sein. Insbesondere für → Semantic Web Anwendungen ist die (langfristige) Aufrufbarkeit von Daten nötig. Darauf aufbauend ist es möglich, die Dateninhalte regelmässig zu extrahieren und damit andere Inhalte aktualisiert anzureichern. Notwendig dafür sind → Identifizier, → URI/IRI oder → DOI. Durch den Einsatz eines → Permalinks kann somit ein Datensatz langfristig aufgerufen werden.</p>	<p><i>Erst durch die Adressierbarkeit wird es möglich, Ressourcen zielgenau und wiederholt aufzurufen.</i></p>
<p>Im Rahmen des Projekts → Metagrid werden dank der Adressierbarkeit der Informationen Lebensdaten der aufgenommenen Personen unter den Partnern ausgetauscht.</p>	
<h2>Identifizier/Persistent Identifizier</h2>	
<p>Ein Zeichenstring gilt als Identifikationsinstrument, wenn dieser ein Objekt eindeutig zuordnet. Um Änderungen in Identifiern möglichst zu vermeiden, wird empfohlen, eine zufällige Kombination aus Zahlen und Buchstaben zu verwenden und keine Klarnamen oder ähnliches.</p> <p>Persistent Identifizier ist ein permanenter Identifikator, d.h. die Institution, die ihn zur Verfügung stellt, garantiert, dass er langfristig erhalten bleibt.</p>	<p><i>Über Identifizier werden Informationen zu Entitäten adressier- und abrufbar.</i></p>
<p>Die Identifikationsnummer einer Person in einer Datenbank ist ein Identifizier.</p>	

URL	
URL (Uniform Resource Locator) ist eine Internetadresse, die eine Ressource, z.B. eine Webseite, lokalisiert und identifiziert sowie das Kontaktprotokoll (häufig → http / https) definiert.	
URI/IRI	
URI/IRI (Uniform Resource Identifier/Internationalized Resource Identifier) ist ein eindeutiger Identifikator für eine Ressource. Eine URI/IRI kann aus einer URL bestehen.	<i>Die Vergabe von URI ermöglicht die Adressierbarkeit, bspw. über Browser.</i>
URN	
URN (Uniform Resource Name) ist ein URI mit dem Schema urn. Es handelt sich um einen dauerhaften, ortsunabhängigen Bezeichner für eine digitale Ressource.	
ARK	
Archival Resource Keys (ARKs) dienen als Persistent Identifier oder stabile, vertrauenswürdige Referenzen für Informationsobjekte. Das ARK-System ist offen, und jede Einrichtung kann ARKs vergeben, sobald sie sich registriert hat und eine sogenannte Name Assigning Authority Number (NAAN) besitzt.	
DOI	
DOI (Digital Document Object Identifier) ist ein eindeutiger und dauerhafter digitaler Identifikator für physische, digitale oder abstrakte Objekte. Er wird momentan vor allem für Online-Artikel von wissenschaftlichen Fachzeitschriften und Datenpublikationen (bspw. → ZENODO) verwendet.	

<p>Während URL und URI frei durch den jeweiligen Webmaster vergeben werden können, sind DOI kostenpflichtig zu beziehen.¹⁸ URN werden in der Schweiz durch die Nationalbibliothek vergeben.</p>	
<p>Link/Permalink</p>	
<p>Link ist die Kurzform von Hyperlink, d.h. ein Verweis auf einen Hypertext (vgl. auch → URL).</p> <p>Ein Permalink bezeichnet eine → URL, die dauerhaft bestehen bleibt. Damit wird ein digitales Objekt zitierfähig, da der Link immer auf das Dokument/die Ressource verweist.</p>	<p><i>Mit Permalinks kann garantiert werden, dass eine Information langfristig aufgefunden werden kann.</i></p>
<p>Datenaustausch/Kommunikation/Vernetzung/Identifikation</p>	
<p>Datenabruf und Schnittstellen</p>	
<p>Angebote zum Datenaustausch sind zentral, um eigene Daten, Texte und Dokumente der Öffentlichkeit zur Verfügung zu stellen. Dafür existieren unterschiedliche Möglichkeiten: Über Websites, Download (FTP) oder im Idealfall durch die Freigabe über → Schnittstellen (→ RESTful-API/ SPARQL, SPARQL Protocol and RDF Query Language).</p> <p>Die Freigabe über Schnittstellen erlaubt live Abfragen. Die Daten werden bei Abfragen über Schnittstellen nur bei entsprechenden Berechtigungen verändert/angepasst.</p>	<p><i>Standardisierte Schnittstellen ermöglichen das Aggregieren von Daten und Metadaten aus unterschiedlichen Quellen/von unterschiedlichen Anbietern.</i></p>
<p>HTTP</p>	
<p>Hypertext-Transfer-Protocol (HTTP) ist ein Datenübertragungsprotokoll und wird hauptsächlich eingesetzt, um Websites in Webbrowser zu laden. Über das Protokoll können Daten versandt werden. HTTPS ist die verschlüsselte Variante des Protokolls, damit kann ein sicherer Datenverkehr gewährleistet werden.</p>	

¹⁸ Zu den Kosten für DOI, siehe bspw. den Blog von DataCite zum Thema: <https://datacite.org/blog/the-new-datacite-membership-and-fees-model/>.

API/RESTful-API	
<p>Eine Programmierschnittstelle (englisch <i>application programming interface</i>, wörtlich «Anwendungs-programmier-schnittstelle»), genauer <i>Schnittstelle zur Anwendungsprogrammierung</i>, häufig nur kurz API genannt, ist ein Programmteil, der von einem Softwaresystem anderen Programmen zur Anbindung an das System zur Verfügung gestellt wird.¹⁹</p> <p>Eine RESTful-API (Representational State Transfer Application Programming Interface) ist eine webbasierte Schnittstelle für den Zugriff auf eine Datenbank durch eine Applikation wie einen Webbrowser. Die Anfragen an eine RESTful-API erfolgen über die verschiedenen Methoden von HTTP auf URLs mit vordefinierten Pfadsegmenten (GET, DELETE etc.). Dadurch können Daten gelesen, erzeugt, modifiziert oder gelöscht werden. Ausgaben können als JSON und → XML erfolgen.²⁰</p> <p>Aus den Abfragen ist es möglich, gewünschte Daten/Informationen zu extrahieren, wenn die Daten nach Vorgaben strukturiert vorliegen und entsprechend an eine Auswertung weitergegeben werden.</p>	<p><i>REST erlaubt den Zugriff und die Anpassung von eindeutig identifizierten Ressourcen (bspw. über URI), aber auch von Suchabfragen, die über die URL mitgegeben werden können.</i></p>
SPARQL	
<p>SPARQL ist die Abfragesprache für RDF-Datenbanken, damit können Anfragen an einen → Triplestore formuliert werden (SPARQL-Endpoint), um gezielt nach Aussagen/Datensätzen (Triples) zu suchen. Ausserdem kann SPARQL genutzt werden, Daten (Triples) in einem Triplestore zu erzeugen (SPARUL: SPARQL/Update).²¹</p> <p>Typischerweise werden die Abfrageresultate als XML oder Turtle-Datei mit einer Zeile pro Resultat ausgegeben.</p>	<p><i>SPARQL sorgt für eine gezielte Extraktion der in Triple-Form hinterlegten Informationen.</i></p>

¹⁹ o.A.: Programmierschnittstelle, in: Wikipedia, o.D., <<https://de.wikipedia.org/wiki/Programmierschnittstelle>>, Stand: 21.12.2021.

²⁰ Siehe für eine ausführliche Einführung: Rodriguez, Alex: RESTful Web services. The basics, in: IBM. Developer Works, 09.02.2015, <<http://www.ibm.com/developerworks/library/ws-restful/index.html>>, Stand: 19.04.2017.

²¹ Siehe für eine ausführliche Dokumentation: W3C SPARQL Working Group: SPARQL 1.1 Overview, W3C, 21.03.2013, <<https://www.w3.org/TR/sparql11-overview/>>, Stand: 19.04.2017.

Export und Datenaustausch	
Open Access/Open Source/Open Data	
<p>Open Access bedeutet permanenten, freien, kostenlosen und elektronischen Zugang zu wissenschaftlichen und anderen Publikationen.²²</p> <p>Um den Zugriff auf die Datenbasis (unabhängig davon, ob es sich um Softwarecode oder wissenschaftliche Arbeiten handelt) langfristig und barrierefrei zu gewährleisten, verfolgt eine Vielzahl von namhaften Institutionen Open Access/Open Source Strategien.²³</p> <p>Frei verfügbare und ohne gewichtige Einschränkung nutzbare Daten werden als Open Data bezeichnet. Bekannt ist das Fünf-Sterne-Modell von Tim Berners-Lee, das unterschiedliche Stufen offener Daten definiert: von einem Stern (offene Lizenz) bis zu fünf Sternen (offene Lizenz, strukturiertes, offenes/nicht proprietäres Format, Daten mit → URIs verlinkt).²⁴</p> <p>Die Open Data Bewegung (in der Schweiz: https://opendata.ch/) ist der Ansicht, dass frei zugängliche Daten Effizienz und Innovation fördern und zu mehr Transparenz (in Forschung und Verwaltung) führen.²⁵</p>	<p><i>Ermöglicht die freie Nachnutzung von Erschliessungs- und Forschungsarbeiten.</i></p> <p><i>Die Offenlegung von Codes und Texten erlaubt die einfache und korrekte Nachnutzung. Weiterentwicklungen können auf bereits bestehenden Erkenntnissen/Systemen aufbauen.</i></p>
<p>Open Data Das Portal für offene Daten der Verwaltung der Schweiz, betrieben vom Schweizerischen Bundesarchiv: <https://opendata.swiss/>, Stand: 23.12.2021.</p>	
Linked (Open) Data (LD/LOD)	
<p>Als Linked Open Data (LOD) werden über das Web frei verfügbare vernetzte und strukturierte Daten bezeichnet, die über → URIs/IRIs identifiziert und dereferenziert (aufgelöst) werden können.²⁶ LOD wird heute meistens mit → Semantic Web-Technologien wie → RDF umgesetzt (unter Verwendung von → Ontologien, Abfrage über → SPARQL-Endpunkte).</p>	<p><i>Linked Open Data erlaubt die Vernetzung mit bereits bestehenden und andernorts befindlichen Daten(beständen) und somit die Nutzung von bereits angehäuften Wissen.</i></p>

²² Nach SAGW, siehe dort auch das Glossar zum Thema: <<https://www.sagw.ch/sagw/sagw/themen/wissenschaftskulturen/open-science/open-access/open-access-strategie-der-sagw>>, Stand: 18.03.2022.

²³ Siehe dazu auch die Berliner Erklärung zum Open Access: Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 22.10.2003, <<https://openaccess.mpg.de/Berliner-Erklaerung>>, Stand: 19.04.2017.

²⁴ Vgl. <<http://5stardata.info/de/>>, Stand: 23.12.2021.

²⁵ Siehe die Website: <<https://opendata.ch/>>, Stand: 12.5.2017.

²⁶ Siehe die Website: <<http://linkeddata.org/>>, Stand 12.05.2017.

<p>swissTopo stellt geographische Informationen zu einzelnen Gemeinden und anderen geographischen Merkmalen als Linked Open Data zur Verfügung: https://www.geo.admin.ch/en/geo-services/geo-services/linkedata.html, Stand: 23.12.2021.</p>	
Wikidata	
<p>Wikidata ist der Name einer frei bearbeitbaren Datenbank, die unter anderem das Ziel hat, Wikipedia zu unterstützen. Das Projekt wurde von Wikimedia Deutschland gestartet und stellt als gemeinsame Quelle bestimmte Datentypen für Wikimedia-Projekte bereit, zum Beispiel Geburtsdaten oder sonstige allgemeingültige Daten, die in allen Artikeln der Wikimedia-Projekte verwendet werden können.²⁷</p>	<p><i>Wikidata stellt Linked Open Data zur Verfügung, die durch die Wikipedia-Community gepflegt wird.</i></p>
<p>Henker als Familienname oder als Konzept vgl. https://www.wikidata.org/wiki/Q37547302, Stand: 23.12.2021.</p>	
Nachhaltige Daten (und Code)	
<p>Digitale Daten (und Code) sind oft «nur» Arbeitsmittel oder Zwischenresultate und haben temporären Charakter. Im Gegensatz zu Papier lassen sich viele Dateiformate jedoch ohne die Umgebung, in der sie entstanden sind, weder lesen noch rekonstruieren. Um digitale Daten und ihre Repräsentationen langfristig nachnutzbar zu machen, müssen sie deshalb in Formaten mit offenen Spezifikationen gespeichert werden, die langfristig genutzt werden können.</p>	
<p>Prinzipien für langlebige Daten wurden beispielsweise im Rahmen des Endings Project entwickelt.</p>	

²⁷ o.A.: Wikidata, in: Wikipedia, o.D., <https://de.wikipedia.org/wiki/Wikidata>, Stand: 23.12.2021.

Katalog für Standards mit Bezug zu den Geschichtswissenschaften

Gedächtnisinstitutionen haben Standards für die Erfassung von (Meta-)Daten definiert, um den Datenaustausch zu ermöglichen. Kenntnisse der wichtigsten Standards helfen, die Metadaten der Gedächtnisinstitutionen zu interpretieren, insbesondere wenn die Daten (automatisiert) nachgenutzt werden.

Es gibt Standards in den Bereichen formale Beschreibung, normierter Sucheinstieg (Autorität, Normdateien) und Datenformat.

Regeln zur formalen Beschreibung von Dokumenten und anderen Ressourcen in Gedächtnisinstitutionen

<p>Archive</p> <p>Der Verein Schweizerischer Archivar:innen (VSA-AAS) hat seit 2005 eine eigene Arbeitsgruppe «Normen und Standards», die sich intensiv mit dem Themenbereich auseinandersetzt. Ihre Tätigkeit ist auf der Website der Arbeitsgruppe dokumentiert.</p> <p>Expertengruppen der internationalen Nichtregierungsorganisation «International Council on Archives (ICA)» setzen sich für Standards in der archivischen Erschliessung ein.</p> <p>Trotz dieser Bemühungen existieren in der Schweiz bisher noch keine verbindlichen Standards zur Erfassung von Archivmaterial, sodass sich die in den Archivinformationssystemen befindenden Informationen leider nicht einfach vergleichen oder maschinell nachnutzen lassen.</p>	
<p>ISAD(G) = International Standard Archival Description (General)</p>	
<p>ISAD(G) ist der internationale Anwendungsstandard zur Verzeichnung archivischer Unterlagen der ICA mit einheitlichen Pflichtfeldern und einer mehrstufigen Verzeichnungshierarchie.</p> <p><i>Wird wie andere Archivstandards aktuell von → RIC abgelöst.</i></p>	<p><i>Aufgrund individueller Handhabung in den Archiven ist ISAD(G) nur beschränkt hilfreich beim Datenaustausch.</i></p>
<p>ISAAR(CPF) = International Standard Archival Authority Record for Corporate Bodies, Persons, and Families</p>	
<p>ISAAR(CPF) ist der internationale Standard für Normdatensätze von Archivalien von Organisationen/Institutionen, Personen und Familien. Er ermöglicht die standardisierte Erfassung der Urheber von Schriftgut und somit die Beschreibung des Kontexts der Produktion (Herkunft). Seine Verwendung ist weniger verbreitet als ISAD(G), weil die Erfassung zeitaufwendig ist.</p>	<p><i>ISAAR(CPF) ermöglicht den Zugriff auf die gesamte Überlieferung eines Aktenbildners.</i></p>

<p><i>Wird wie andere Archivstandards aktuell von → RIC abgelöst.</i></p>	
<p>RiC = Records in Context</p>	
<p>RiC ist aktuell in Entwicklung. Der Standard versteht sich als neue Norm zur umfassenden Beschreibung von Archivalien und ihrer Beziehungen untereinander. RiC versucht die archivischen Standards (ISAD[G], ISAAR[CPF], ISDF, ISDIAH) mit Hilfe einer Ontologie zu <i>einem</i> Standard zusammenzufassen.</p> <p>Siehe auch → RiC-O, → EAD.</p>	<p><i>RiC will die Möglichkeiten von Linked Open Data nutzen und unterschiedliche Relationen und Verbindungen zwischen Archivalien aufzeigen.</i></p>
<p>Bibliotheken</p> <p>Bibliosuisse (https://bibliosuisse.ch/), vormals Bibliothek Information Schweiz, ist der nationale Verband der Bibliotheken, Informationszentren und ihrer Mitarbeitenden. Auf internationaler Ebene vertritt die IFLA (International Federation of Library Associations and Institutions) die Interessen der Informationseinrichtungen. Die IFLA hat mehrere Standards im Bereich der Informationswissenschaften herausgegeben.</p> <p>Die Metadaten der Bibliotheken, insbesondere die bibliografischen Aufnahmen, sind häufig frei zugänglich und können ohne Einschränkungen genutzt werden. Die bibliografischen Aufnahmen sind im MARC-Format standardisiert erfasst und enthalten häufig Verweise auf Normdaten.</p>	
<p>Die Metadaten des elektronischen Katalogs der Schweizerischen Nationalbibliothek, des Helveticat und der Bibliographie der Schweizergeschichte (BSG) sind im Format MARC21 über die Schnittstelle Z39.50 und auf Anfrage über OAI-PMH erhältlich. Eine Auswahl der Daten kann vom Portal opendata.swiss heruntergeladen werden.</p>	
<p>ISBD</p>	
<p>Die ISBD (International Standard Bibliographic Description = Internationaler Standard zur Beschreibung bibliographischer Daten) ist ein Standard, der von der IFLA betreut wird. Die ISBD regelt die einheitliche Beschreibung von unterschiedlichen bibliothekarischen Ressourcen und ermöglicht den Datenaustausch.</p> <p>Die ISBD schreibt vor, welche Elemente für die Beschreibung des Dokuments herangezogen werden. Sie legt die Reihenfolge der Elemente in der Beschreibung fest und sie bestimmt die Zeichensetzung (Deskriptionszeichen). Die ISBD ist in Überarbeitung. Sie soll künftig den Vorgaben von Linked Data genügen.</p>	<p><i>Die Interpunktion zwischen den verschiedenen Elementen wie Haupttitel und Untertitel werden beim Datenexport mitgeliefert und müssen daher je nach Nachnutzung automatisiert gelöscht werden.</i></p>

<p>Beispiel für die ISBD-Zone Sachtitel und Urheberangabe: Sachtitel: erster Zusatz: weiterer Zusatz / Verfasserangabe Zwingli: widerständiger Geist mit politischem Instinkt / Franz Rueb; Bilder: Hans Rueb Das entsprechende Katalogisat in der BSG ist hier zu finden.</p>	
AACR2	
<p>Die Anglo-American Cataloguing Rules sind Regeln zur formalen Beschreibung von Werken. Sie stammen aus den 1960er-Jahren und wurden mehrmals revidiert. Es war das international am weitesten verbreitete Regelwerk und wurde von über 30 verschiedenen Ländern inkl. der Schweiz übernommen.²⁸ AACR2 ist veraltet und wurde durch →RDA ersetzt.</p>	<p><i>In den Katalogisierungsregeln ist bestimmt, wie die formalen Elemente, wie z.B. der Titel, die Autoren oder der Reihentitel eines Werks beschrieben werden.</i></p>
<p>Bei Werken mit mehr als drei Autor:innen werden alle ausser der ersten Person weggelassen. Die Weglassung wird durch den Text «et al.» angezeigt. Zudem werden mit den AACR2 viele immer wiederkehrende Begriffe abgekürzt, wie z.B. Zusammenfassung für Zusammenfassung.</p>	
RDA	
<p>Das Regelwerk Resource Description and Access (RDA) ist der Nachfolger der AACR2. Es wurde 2010 eingeführt. 2021 wurde eine revidierte Ausgabe publiziert. Die RDA basieren auf dem Datenmodell → IFLA LRM und den Internationalen Katalogisierungsprinzipien (ICP). Zunächst für die Anwendung in Bibliotheken entwickelt, ist das Regelwerk jedoch auch für einen weiteren Anwenderkreis in Bereichen wie Museen und Archive geeignet. Der Text des Regelwerks wird laufend aktualisiert und ausschliesslich online im «RDA Toolkit» publiziert. Die Benutzung des «RDA Toolkit» ist kostenpflichtig. Im deutschsprachigen Raum wurden die RDA im Jahr 2016 eingeführt. Die wichtigen Bibliothekskataloge in der Schweiz (SLSP, NB etc.) erschliessen nach RDA.²⁹</p>	

²⁸ o.A.: Anglo-American Cataloguing Rules, in: Wikipedia, o.D., <https://de.wikipedia.org/wiki/Anglo-American_Cataloguing_Rules>, Stand: 21.12.2021.

²⁹ Weiterführende Informationen unter: o.A.: RDA-Info, in: DNB, 06.12.2021, <<https://wiki.dnb.de/display/RDAINFO/RDA-Info>>, Stand: 23.12.2021.

<p>Für die Beschreibung des einzelnen Werks ist die Transkription ein wichtiges Merkmal der RDA. Im Unterschied zu den AACR2 wird mit den RDA die Informationen so übertragen, wie sie in der Vorlage zu finden sind. Bei den Autorenangaben werden beispielsweise – wenn in der Vorlage vorhanden – die aufgelisteten Titel und die affilierten Universitäten und der Wirkungsort angeführt. Beispiel einer wortwörtlichen Übertragung in Helveticat: Prof. Dr. Andreas Furrer (RA, Universität Luzern), Prof. Dr. Markus Müller-Chen (RA, Universität St. Gallen).</p>	
RSWK	
<p>Die Regeln für den Schlagwortkatalog sind in wissenschaftlichen Universalbibliotheken, in öffentlichen Bibliotheken oder in Spezialbibliotheken Deutschlands, Österreichs und der deutschsprachigen Schweiz das Regelwerk für die intellektuelle inhaltliche Erschliessung. Sie regeln die Anwendung der verwendeten Schlagworte, die der GND entnommen sind.</p> <p>Die RSWK ist eine präkoordinierte Indexiersprache: Reicht ein Schlagwort zur Beschreibung des Gegenstandes eines Werks nicht aus, werden die einzelnen Termini zu einer Schlagwortkette zusammengeführt.</p> <p>In der französischsprachigen Schweiz wird → RAMEAU für die Verschlagwortung in den Bibliothekssystemen verwendet.</p>	<p><i>Immer mehr Bibliotheken erschliessen inzwischen ohne Schlagwortketten und verwenden stattdessen die GND-Termini als Einzelschlagworte.</i></p>
<p>Das Werk von Thomas Maissen «Verweigerte Erinnerung. Nachrichtenlose Vermögen und die Schweizer Weltkriegsdebatte 1989–2004» wurde mit folgender Schlagwortkette versehen: Schweiz – Neutralität – Weltkrieg (1939–1945) – Vergangenheitsbewältigung – Geschichte 1989–2004. Die Zeitdimension ist mit der RSKW nicht strukturiert abfragbar (Zeitschlagwort: «Geschichte 1989–2004»). Ist das Thema eines Werks zeitlich begrenzt, wird das Zeitschlagwort Geschichte mit der exakten Zeitdauer erfasst.</p>	
FRBR/FRBRoo	
<p>Die Functional Requirements for Bibliographic Records (FRBR; deutsch «Funktionale Anforderungen an bibliographische Datensätze») sind ein Datenmodell für bibliografische Metadaten. Sie sind heute die wichtigste theoretische Grundlage zur Erstellung von bibliothekarischen Regelwerken.³⁰ Die FRBR-Ontologie basiert auf dem Entity-Relationship-Modell und definiert eine Reihe von grundlegenden Konzepten für die Katalogisierung.</p> <p>Die RDA basieren auf dem IFLA Library Reference Model (IFLA LRM).</p> <p>Das Modell FRBR wurde durch → IFLA LRM abgelöst.</p>	<p><i>Während sich herkömmliche Regelwerke vor allem auf Details der Beschreibung konzentrieren und wesentliche Konzepte (bspw. «Werk», «Ausgabe», «Exemplar») als gegeben vorausgesetzt werden, versucht FRBR genau diese Begriffe zu definieren und in</i></p>

³⁰ o.A.: Functional Requirements for Bibliographic Records, in: Wikipedia, o.D., <https://de.wikipedia.org/wiki/Functional_Requirements_for_Bibliographic_Records>, Stand: 21.12.2021.

	Beziehung zueinander zu setzen.
<p>Mit FRBR werden beispielsweise verschiedene Ausgaben und Übersetzungen eines Werks zusammengeführt, was zu einer besseren Übersicht für die Benutzer:innen führt. Beispiel: Dürrenmatt, Friedrich: Das Versprechen (Übersetzung mit Autoritätseintrag für das Werk).</p>	
IFLA LRM	
<p>Das IFLA Library Reference Model (IFLA LRM) ist ein konzeptionelles Modell, das alle Aspekte der bibliografischen Erschliessung abdeckt. Zudem unterstützt es die Einbindung von bibliografischen Daten in Linked-Data-Anwendungen. Es beruht auf den Vorgänger-Modellen Functional Requirements for Bibliographic Records (→FRBR), Functional Requirements for Authority Data (FRAD) und Functional Requirements for Subject Authority Data (FRSAD).³¹</p>	
Museen und Sammlungen	
<p>Aufgrund der Heterogenität von Sammlungen und Museen ist die Einigung auf verbindende Standards viel komplexer. Hier genannt sind aus diesem Grund einige wenige Thesauri, Vokabulare und Metadatenstandards.</p>	
Getty Vokabulare: AAT, TGN, ULAN, CONA, IA	
<p>Die Vokabulare von Getty beschreiben unterschiedliche Aspekte von Objekten, Künstler:innen und Orten, die typischerweise in Museen gefunden werden. Alle Vokabulare sind strukturiert und nicht abgeschlossen.</p> <p>AAT = Arts and Architecture Thesaurus TGN = Getty Thesaurus of Geographic Names ULAN = Union List of Artist Names CONA = Cultural Objects Name Authority IA = Iconography Authority</p>	
LIDO = Lightweight Information Describing Objects	
<p>LIDO ist ein Metadatenformat zur Bereitstellung von Museums-Objektdaten in Internet-Portalen.</p>	
CIDOC-CRM (CIDOC Conceptual Reference Model)	
<p>CIDOC-CRM ist ein theoretisches und praktisches Werkzeug für die Informationsintegration im Bereich des kulturellen Erbes.</p>	

³¹< <https://www.ifla.org/resources/?oPubId=11412>>, Stand: 21.03.2022.

Es soll eine gemeinsame Sprache für Fachleute sein, um Anforderungen an Informationssysteme zu formulieren und als Leitfaden für eine gute Praxis der konzeptionellen Modellierung dienen. Auf diese Weise kann es den «semantischen Klebstoff» liefern, der für die Vermittlung zwischen verschiedenen Quellen von Informationen für kulturelles Erbe benötigt wird, wie etwa die von Museen, Bibliotheken und Archiven veröffentlichten Informationen.

Allgemeine Normdateien (Autoritätsfiles, engl. authority files)

Normdaten haben folgende Zielsetzungen:

- Eindeutige Identifikation von Entitäten,
- Zusammenführen von Synonymen und Schreibvarianten,
- Unterscheidung von Homonymen sowie
- Einheitlicher Zugang mittels einer Ansetzungsform (Labels/Heading/Deskriptor).

Bei den Normdaten handelt es sich um Verzeichnisse von normierten Begriffen, die für die formale wie auch für die thematische Erschliessung verwendet werden.

In durch Bibliotheken geschaffenen Normdateien kommen nur diejenigen Begriffe und Namen vor, die für die Beschreibung von in Bibliotheken und anderen Dokumentationseinrichtungen vorhandenen Werken nötig sind (sog. «literary warrant»). Normdateien sind dementsprechend keine Thesauri, die ein ganzes Wissensgebiet abdecken.

Im Folgenden werden in der Schweiz verbreitete und international wichtige bibliothekarische Normdateien vorgestellt.

GND

Die Gemeinsame Normdatei (GND) ist eine Normdatei für Personennamen/Namen von Personen, Körperschaften, Konferenzen, Geografika, Sachschlagworte und Werktitel, die vor allem der Katalogisierung von Literatur in Bibliotheken dient, zunehmend aber auch von Archiven, Museen, Projekten und in Webanwendungen genutzt wird.³² Die [Lobid GND](#) bietet eine erweiterte Suche. Sie wird von der Deutschen Nationalbibliothek, allen deutschsprachigen Bibliotheksverbänden mit den angeschlossenen Bibliotheken und zahlreichen weiteren Einrichtungen gemeinschaftlich geführt. Die einzelnen Entitäten können mittels des Teilbestandskennzeichens separat abgefragt werden.

Die den Normdatensätzen zugrunde liegenden Regelwerke sind zum einen die «Resource Description and Access» (→[RDA](#)) für die Entitäten,

Die Mitarbeit in der GND erfolgt entweder über Verbände, die Schweizerische Nationalbibliothek oder nach direkter Absprache mit der Deutschen Nationalbibliothek. Auch Geschichtsprojekte (z.B. HLS) können – nach einer entsprechenden Schulung – selbst neue Einträge für Personen in der GND erstellen.

³²https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd_node.html, Stand: 18.03.2022.

<p>die für die Formal- und Inhaltsermittlung genutzt werden, wie beispielsweise Personen und Körperschaften. Zum anderen definieren die «Regeln für den Schlagwortkatalog» (→ RSWK) die Erstellung der Normdatensätze für diejenigen Entitäten, die ausschliesslich von der Inhaltsermittlung genutzt werden, wie beispielsweise Sachbegriffe.</p> <p>Die GND-Normdaten stehen allen Interessenten in den Formaten MARC 21 Authority, MARCXML und RDFxml über verschiedene Bezugswege kostenfrei unter CC0 1.0 zur Verfügung.</p>	
<p>Für die Erfassung eines neuen Begriffs gibt es Minimalkriterien, damit eine eindeutige Identifizierung gewährleistet werden kann. Bei Personennamen beispielsweise sind dies die Namensform und entweder das charakteristische Betätigungsfeld oder die Lebensdaten. Die einzelnen Datensätze können aber beliebig mit strukturierten und innerhalb der GND verlinkten Informationen ausgebaut werden, vgl. z.B. Albert Einstein oder Wilhelm Tell.</p>	
<p>RAMEAU</p>	
<p>RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) ist eine Sacherschliessungssprache. Die Dokumentationsform wird in Frankreich von der Bibliothèque nationale de France, den Universitätsbibliotheken, zahlreichen öffentlichen Lese- und Forschungsbibliotheken sowie mehreren privaten Organisationen verwendet.</p>	
<p>Beispiele in RAMEAU: Albert Einstein, Wilhelm Tell.</p>	
<p>LCSH</p>	
<p>Die Normdatei Library of Congress Subject Headings (LCSH) wird von der Library of Congress in Washington gepflegt. Sie wird für die inhaltliche Erschliessung von Bibliotheksbeständen im englischsprachigen Raum verwendet.</p>	<p><i>Neueingaben und Änderungen des LCSH können nur via eines «proposals» an die Policy and Standards Division (PSD) beantragt werden.</i></p>
<p>Beispiel in LCSH: Albert Einstein.</p>	
<p>MACS</p>	
<p>Das Projekt Multilingual Access to Subjects (MACS) hatte die Erstellung eines multilingualen Recherchevokabulars auf Normdatenebene zum</p>	<p><i>Es gibt nur Verknüpfungen für Sachschlagworte.</i></p>

<p>Ziel.³³ Aktuell sind Sachschlagworte der GND, RAMEAU und Library of Congress Subject Headings verknüpft und in der GND abfragbar. Zurzeit werden in der GND keine neuen Links erstellt. Das Projekt wird nicht mehr weitergeführt.</p>	<p><i>Die Mappings sind als Linked Data frei zugänglich.</i></p>
<p>GND-Mul</p>	
<p>Die Deutsche Nationalbibliothek ist die Initiatorin des Projekts «GND mul». Ziel ist es, die GND sowohl mit anderssprachigen Normdateien als auch mit Fachthesauri zu verknüpfen. Das datenstrukturgebende Grundlagenprojekt soll eine Referenz für alle Arten von Crosskonkordanzen zur GND bieten.</p>	
<p>Normdaten für einzelne Entitäten (Autoritätsfiles)</p>	
<p>Personen</p>	
<p>ULAN</p>	
<p>Das Getty Research Institute stellt ihre Datenbank mit Künstler:innen frei zur Verfügung: Getty Union List of Artist Names (ULAN).</p>	
<p>GND</p>	
<p>In der GND sind viele Namen von Personen (Autor:innen, historische Persönlichkeiten) bereits erfasst. Die GND verzeichnet Personen aus der ganzen Welt und aus sämtlichen Sprachregionen. Sie beschränkt sich nicht auf Personen aus dem deutschsprachigen Raum.</p>	<p><i>Die GND, die bei Namen von Personen sehr stark ist, hat gerade bei dieser Entität einige Dubletten. Je nach Verwendung der Daten lohnt sich die Einschränkung auf die Katalogisierungslevel Tp3 oder Tp1.</i></p>
<p>ISNI (International Standard Name Identifier)</p>	
<p>ISNI ist ein Metastandard. ISNI wird von der ISNI International Agency Ltd betrieben. INSI alimentiert sich zum einen aus bestehenden Identifikator-Systemen und andererseits aus den verschiedenen ISNI-Agenturen. Personennormdaten stammen zurzeit aus 52 Quellen, unter anderem aus der GND. Die ISNI dient der weltweit eindeutigen Identifizierung von Urheber:innen und allen Akteur:innen</p>	

³³ o.A.: MACS. Multilingual Access to Subjects, in: DNB, 30.07.2020, <https://www.dnb.de/DE/Professionell/Metadatendienste/Metadaten/Voclink/voclink_node.html#doc58264bodyText2>, Stand: 21.12.2021.

<p>(Wissenschaftler:innen, Erfinder:innen, Künstler:innen, Grafiker:innen, Interpret:innen, Produzent:innen, Verleger:innen, Übersetzer:innen etc.), die an der Veröffentlichung eines Werkes beteiligt sind. Auch Organisationen, Gruppen und Körperschaften können eine ISNI erhalten. Namensvarianten werden auch erfasst. Für die Vergabe von ISNI in der Schweiz ist MVB zuständig (https://german-isbn.de/isni/die-isni).</p>	
<p>ORCID (Open Researcher and Contributor ID)</p>	
<p>ORCID verknüpft die Identität von Wissenschaftler:innen mit ihren Werken. Mit ORCID können Dokumente und Datensätze nicht nur mit Namen, sondern einem Identifikator verbunden werden, der einer Person zugeschrieben wird. ORCID eignet sich nicht für die Identifikation von historischen Personen.</p>	
<p>data.bnf.fr</p>	
<p>Data.bnf.fr ist eine Datenbank, welche die Daten der Bibliothèque nationale de France (BnF) mit Hilfe von Technologien des Semantic Web (→RDF) fördert, um ihre Wiederverwendung durch Dritte zu erleichtern. Da es sich um die Daten einer Bibliothek handelt, sind die meisten registrierten Personen Autor:innen, für die ein bibliografischer Eintrag vorhanden ist. Links zu externen Ressourcen (wie → VIAF oder wikidata) werden ebenfalls gemacht.</p>	
<p>Vernetzungsdienste für Personen</p>	
<p>VIAF</p>	
<p>In dem Projekt Virtual International Authority File (VIAF) werden die GND und andere Normdateien mit Personennamen über eine Konkordanz zu einer <i>virtuellen internationalen Normdatei</i> verbunden. VIAF selbst ist keine Normdatei, es können keine Autoritätsdatensätze in der VIAF erfasst werden. Die Daten werden online angeboten und stehen für Recherchen, den Datenaustausch und die gemeinsame Datennutzung zur Verfügung. Zur Referenzierung erhält ein VIAF-Datensatz eine eigene Normdatennummer als Identifikator, aus der eine Form als URI für Linked-Data-Anwendungen hervorgeht.</p>	<p><i>Um eine VIAF-Nummer zu erhalten, muss die Entität immer in einer der beteiligten nationalen Normdateien erfasst sein.</i></p>
<p>Beispiel in VIAF: Albert Einstein.</p>	

Metagrid	
<p>Der Webservice Metagrid ermöglicht die Einrichtung, Verwaltung und Analyse von Links zwischen identischen Personen verschiedener Websites und Datenbanken.³⁴ Metagrid ist keine Normdatei. Dank der Metagrid-Suche können gespeicherte Konkordanzen zu Personen aus den an Metagrid beteiligten Datenbanken durchsucht werden. Metagrid ist insbesondere von Vorteil, wenn für Personen keine GND Einträge existieren. Der Austausch erfolgt nicht hierarchisch, sondern föderal.</p>	<p><i>Der Webservice leistet einen wichtigen Beitrag bei der Suche nach relevanten historischen Informationen zu Personen, die oftmals keine GND haben.</i></p>
<p>Jemand möchte wissen, auf welchen Websites (angeschlossener Institutionen) Informationen zu einer bestimmten, identifizierten historischen Person zu finden sind. Beispiel: Friedrich Traugott Wahlen bei Metagrid, damit verknüpft (Auswahl) Diplomatische Dokumente der Schweiz, Historisches Lexikon der Schweiz, BSG, Archiv für Agrargeschichte.</p>	
Körperschaften/Organisation/Institutionen	
<p>Bisher existiert unseres Wissens noch keine umfassende Zusammenstellung von historischen Körperschaften, Organisationen und Institutionen und deren Funktionen.</p>	
GND	
<p>In der GND sind verschiedene Namen von Körperschaften erfasst.</p>	<p><i>In der GND wird zum Teil nicht zwischen Körperschaft und Ort unterschieden.</i></p>
Orte	
Getty Thesaurus of Geographic Names	
<p>Der Thesaurus of Geographic Names (TGN) ist eine Datenbank von circa 1'000'000 Bezeichnungen für rund 900'000 Orte. Er wird vom Getty Research Institute herausgegeben. Der Thesaurus deckt die gesamte Welt in aktuellen und historischen, sowohl in geophysischen als auch geopolitischen Hierarchien ab. Neben den Bezeichnungen werden Koordinaten, Ortstypen, Ortsbeschreibungen und weitere Informationen gespeichert. Seit 2014 wird der Thesaurus als Linked Open Data bereitgestellt.</p>	<p><i>Der TGN ist eine ausgesprochen grosse Sammlung, jedoch nach unterschiedlichen Bedürfnissen und daher nicht immer konsequent strukturiert.</i></p>

³⁴ <<https://www.metagrid.ch/>>, Stand: 23.12.2021.

Geonames	
Geonames ist eine Datenbank zur Identifizierung und Referenzierung geografischer Entitäten. Unterschiedliche Personen aus verschiedenen Ländern stellen ihr Wissen Geonames zur Verfügung.	<i>Geonames verfügt nur über eine flache Ortstypologie (Buchstabencodes).</i>
GND	
In der GND sind Namen von Geografika verfügbar. Die hierarchischen Beziehungen für die Geografika sind nicht durchgehend erfasst. Auch fehlen bei historischen Orten häufig die Beziehungen zu den vorherigen und nachfolgenden territorialen Entitäten.	<i>Die Namen der Geografika sind in der GND mit dem Teilbestandskennzeichen Tg1 gekennzeichnet und können entsprechend extrahiert werden. Teilweise wird nicht zwischen Körperschaft (Bürgergemeinde) und Ort unterschieden, weshalb die GND im Bereich Orte nur mit Vorsicht zu gebrauchen ist.</i>
Pleiades	
Pleiades ist ein gemeinschaftlich erstelltes Verzeichnis und ein Graph antiker Orte. Es veröffentlicht massgebliche Informationen über antike Orte und Räume und bietet Dienste zum Auffinden, Anzeigen und Wiederverwenden dieser Informationen unter einer offenen Lizenz. ³⁵	

³⁵ <<https://pleiades.stoa.org/>>, Stand: 21.03.2022.

data.bnf.fr	
<p>Unter den von Data.bnf.fr angebotenen Ressourcen (Autor:innen, Werke, Daten, Aufführungen, Periodika) finden sich Einträge zu Orten. Die Daten stammen aus → RAMEAU und der Abteilung für Karten und Pläne.</p> <p>Neben Links zu den übrigen Daten von Data.bnf.fr findet man zusätzliche Informationen wie geografische Koordinaten.</p>	
Ortsnamen.ch	
<p>Das Portal der schweizerischen Ortsnamenforschung stellt Resultate der schweizerischen Ortsnamenforschung der Öffentlichkeit zur Verfügung und orientiert über den Stand der Namenforschung in der Schweiz.</p> <p>In der Onlinedatenbank kann ein Ortsname gesucht und die Treffer auf einer Karte angezeigt werden. Zu den einzelnen Treffern finden sich weitere Informationen wie Mundart, Phonetik, Beschreibung des Ortstyps, Kanton, Gemeinde, Ort, Teilgebiet, Quellenbelege, Koordinaten, Deutung etc.</p>	<p><i>Mit Ortsnamen.ch können Orts- und Flurnamen, die nicht auf der Landeskarte verzeichnet sind, lokalisiert werden.</i></p>
TOPOterm	
<p>TOPOterm ist ein Webservice, der die Anreicherung von Suchabfragen nach Ortsbezeichnungen durch historische Synonyme ermöglicht. Eine Suchabfrage führt entsprechend zu Treffern mit unterschiedlichen Schreibweisen. Überdies können über eine Suchmaske auch direkte Abfragen im Browser ausgeführt werden.</p>	<p><i>Die Daten stehen als Opendata über Schnittstellen zur Verfügung.</i></p>
SwissNAMES3D/swisstopo	
<p>SwissNAMES3D ist eine Namensdatenbank der swisstopo. Sie enthält Namen von Ortschaften, Bergen, Flüssen und anderen Lokalitäten, die auf den Schweizerischen Landeskarten im Massstab 1:25'000 (und grösser) vermerkt sind. Daneben sind die Einträge gruppiert in unterschiedliche Ortstypen wie Berge, Ortschaften etc., so dass auch dieses Kriterium in der Suche eingeschränkt werden kann.</p> <p>Insgesamt umfasst die SwissNames-Datenbank georeferenzierte Einträge von mehr als 190'000 Lokalitäten. Zu jedem Eintrag gehören auch die genaue Position, Gemeinde, Kanton und – wo sinnvoll – auch Höhenangaben.</p> <p>Die Abfrage als Linked Open Data kann über einen SPARQL-Endpoint erfolgen, bspw. <https://sparql.geo.admin.ch/sparql>, Stand: 2103.2022..</p>	<p><i>Mit SwissNAMES3D können Orte auf der Landeskarte lokalisiert werden.</i></p>

Konzepte (Schlagworte und Lemmata)

GND

Die [GND](#) beinhaltet zahlreiche Schlagworte/Sachbegriffe, die abweichende Benennungen und Relationen zu quasi Synonymen und verwandten Begriffen enthalten. Zudem gibt es Links zu den äquivalenten Normdatensätzen der → [LCSH](#) und → [RAMEAU](#) wie auch zu den →DDC-Notationen. Eine eigene Systematik erlaubt einen systematischen Zugang auf die GND. Diese wird nach Bedarf erweitert bzw. erstellt, d.h. sie ist nicht allumfassend.

Die Mappings zu den Schlagworten der LCSH und RAMEAU sowie zu den DDC-Notationen sind als Linked Data frei zugänglich.

Der Beruf Henker oder Nachrichtler ist in der GND ein Synonym des Sachbegriffs [Scharfrichter](#). In der Systematik gehört der Begriff zu 7.7b «Strafvollzug» und 9.4ab «Einzelne Berufe, Tätigkeiten, Funktionen; Religionszugehörigkeit, Weltanschauung». Ein thematischer Bezug wird mit dem Link zum Begriff «Hinrichtung» gegeben. Es wird zudem auf → [RAMEAU](#) «Exécutions capitales et exécuteurs» verwiesen etc.

Klassifikationen

Bibliotheken verwenden nebst der verbalen Sacherschliessung Klassifikationen, um einen thematischen Zugang zu den Beständen zu ermöglichen. Die Klassifikationen sind meist hierarchisch strukturiert, wobei die Relationen ausgewiesen werden.

<p>DDC</p> <p>Die Dewey-Dezimalklassifikation (DDC) ist die international am weitesten verbreitete Klassifikation für die inhaltliche Erschliessung von Bibliotheksbeständen. Sie ist nicht frei zugänglich, sondern lizenzpflichtig.</p> <p>In der →GND gibt es Verlinkungen zwischen den Schlagworten und der entsprechenden DDC-Notation.</p>	<p><i>Die Mappings zwischen den DDC-Notation und den GND-Schlagworten sind als Linked Data frei zugänglich.</i></p>
<p>UDK</p> <p>Die Universal-Dezimalklassifikation beruht auf der →DDC. Sie wurde Ende des 19. Jahrhunderts in Belgien als Alternative zur amerikanisch-lastigen DDC entwickelt. Sie wird hauptsächlich in Bibliotheken ausserhalb des englischsprachigen Raums verwendet.</p>	<p><i>Die UDK wird bspw. durch die ETH-Bibliothek verwendet.</i></p>
<p>Regensburger Verbundklassifikation (RVK)</p> <p>Die Regensburger Verbundklassifikation (RVK) ist eine weitverbreitete Klassifikation mit Anwendern in Deutschland, Österreich, Italien und der Schweiz. Sie ist ursprünglich als Aufstellungssystematik entwickelt worden, sie dient aber auch als Sacherschliessungsinstrument.</p>	<p><i>Im universitären Bereich in der Schweiz spielt die RVK keine wichtige Rolle.</i></p>
<p>Nachschlagewerke, Vokabulare, linguistische Standards und Wörterbücher</p>	
<p>Rechtshistorisches Vokabular</p>	
<p>Ausgehend von den Schlagworten der deutschsprachigen «Materien der Policeyordnungen» entsteht zur Zeit ein mehrsprachiges, mit → SKOS erstelltes rechtshistorisches Vokabular.</p>	

CoNLL (Organisation linguistischer Daten)	
<p>CoNLL ist eine jährliche Konferenz, die <i>Shared Tasks</i> (gemeinsame Probleme) bearbeitet. Die Form der Daten, die für diese Aufgaben vorgeschlagen werden, hat sich als Standard in der maschinellen Sprachverarbeitung (NLP) durchgesetzt. Jede Zeile steht für ein Token, dessen verschiedene linguistische Annotationen streng geordnet und durch Tabulatoren getrennt sind, die ebenso viele Spalten bilden: ID für die Position im Satz, FORM für das Token, LEMMA für das Lemma, POS für den Satzteil, FEAT für die Morphologie, HEAD für den syntaktischen Elternteil etc. Nullwerte werden durch einen Unterstrich (_) gekennzeichnet. Es gibt Empfehlungen für den Wert von Annotationen: Es gibt Standard-Etikettierungssätze und -Referenzsysteme, die jedoch von der annotierten Sprache abhängen können.</p>	
Nationale Wörterbücher der Schweiz	
<p>In der Schweiz werden folgende Wörterbücher im Auftrag der SAGW bearbeitet und online mit unterschiedlichen Suchfunktionen zur Verfügung gestellt:</p> <p>Deutsch: Schweizerdeutsches Wörterbuch oder Schweizerisches Idiotikon. Für Recherchen zur deutschen Sprache vgl. auch das Wörterbuchnetz.</p> <p>Französisch: Das Glossaire des patois de la Suisse romande (GPSR) lässt sich online durchsuchen. Für Recherchen zur französischen Sprache vgl. auch ATLIF und Lexilogos.</p> <p>Italienisch: Das Vocabolario dei dialetti della Svizzera italiana (VSI) wird seine Online-Version per Ende 2022-Anfang 2023 aufschalten.</p> <p>Romanisch: Das Dicziunari Rumantsch Grischun (DRG) dokumentiert die bündnerromanischen Sprachen seit dem 16. Jahrhundert und ist mit DRG-online durchsuchbar.</p>	
<p>Weitere Nachschlagewerke, aufgenommen durch die DNB: Liste der Nachschlagewerke der DNB</p>	

<h2>Datenformate</h2>	
METS/MODS (Metadata Encoding and Transmission Standard, Metadata Object Description)	
<p>Der Metadata Encoding & Transmission Standard (METS) ist ein mit XML Schema definiertes XML-Format zur Beschreibung von digitalen Sammlungen von Objekten mit → Metadaten.</p> <p>METS wird häufig in Kombination mit MODS (Metadata Object Description) verwendet, das ebenfalls von der Library of Congress als XML-Spezifikation verantwortet wird. Während in METS die technischen Metadaten und Pfade zu den Daten hinterlegt wird. Nutzt man MODS, um inhaltlicher/objektbezogene Metadaten zu kodieren.</p>	
MARC21	
<p>Die MARC-Formate sind Standards für die Repräsentation und den Tausch von Daten in maschinenlesbarer Form. Verantwortlich für die Pflege und Entwicklung von MARC 21 ist das Network Development and MARC Standards Office (NDMSO), das vom MARC Advisory Committee unterstützt wird. Es steht auch eine XML-Variante des Formats zur Verfügung. In der Schweiz wird MARC21 von allen grossen Bibliotheken verwendet.</p> <p>Die Anwendung und die Definition der einzelnen Felder ist zu finden unter: http://www.loc.gov/marc/bibliographic/</p> <p>Nachfolger von MARC21 ist BIBFRAME, dieses Format ist aber noch nicht eingeführt.³⁶ Eine Alternative zu MARC ist das globaler verwendbare → MODS.</p>	<p><i>Während der MARC-Standard definiert, welche Information in welchem Feld vermerkt wird, schreiben die Formalschlussregeln die Art, wie die Informationen in den Feldern wiedergegeben werden, vor.</i></p>
DC	
<p>Dublin Core (DC) ist eine Sammlung von einfachen und standardisierten Konventionen zur Beschreibung von Dokumenten und anderen Objekten.</p> <p>Innerhalb von Dublin Core wird zwischen dem Dublin Core Metadata Element Set und den</p>	<p><i>Nach DC strukturierte Daten in Headern von HTML-Seiten helfen Suchmaschinen, Webseiten zu indexieren, zu</i></p>

³⁶ Siehe <<https://www.loc.gov/bibframe/>>, Stand: 22.12.2021.

<p>DCMI Metadata Terms unterschieden. Das Erstere bezeichnet fünfzehn <i>terms</i> zur Beschreibung von Metadaten. Das Zweitere zielt darüber hinaus auf die Verknüpfung der <i>terms</i> mit Klassen und properties, Datentypen und Vokabularen.</p>	<p><i>hierarchisieren und auffindbar zu machen.</i></p>
<h2>XML-Auszeichnungssprachen</h2>	<p><i>Die semantischen Auszeichnungen der Texte sind maschinenlesbar und interoperabel.</i></p>
<h3>ALTO XML</h3>	
<p>ALTO XML (Analyzed Layout and Text Object) ist ein Standard zur Beschreibung von automatisch erkannten Textdaten, der häufig im Kontext von Optical Character Recognition genutzt wird.³⁷</p>	
<h3>PageXML</h3>	
<p>Page XML ist ein Standard zur Beschreibung von automatisch erkannten Textdaten, der häufig im Kontext von Optical Character Recognition genutzt wird. Page XML wird häufig zur Erstellung von Ground Truth (Trainingsmaterial für Texterkennsysteme) genutzt.³⁸ Im Gegensatz zu →ALTO XML erlaubt PageXML semantische Information <i>stand-off</i> zu annotieren.³⁹</p>	
<h3>HTML</h3>	
<p>HTML (Hypertext Markup Language) ist ein Markup (Auszeichnungssprache) für Webseiten, um deren Struktur zu definieren (z.B. Abschnitte, Titel etc.). HTML wird durch den Webbrowser dargestellt. HTML5 ist die neueste Version von HTML und bietet viele vordefinierte interaktive Elemente an. Das XHTML-Format (Extensible HyperText Markup Language) stellt eine HTML 4.01-kompatible XML-</p>	

³⁷ <<https://github.com/altxml/documentation/wiki/Versions>>, Stand: 21.03.2022.

³⁸ <<https://ocr-d.de/de/gt-guidelines/trans/trPages>>, Stand: 21.03.2022.

³⁹ <<https://github.com/PRImA-Research-Lab/PAGE-XMLs>>, Stand: 21.03.2022.

<p>Variante von Hypertext dar, die inzwischen angesichts von HTML5 an Bedeutung verloren hat.</p>	
<p>SIARD</p>	
<p>SIARD (Software Independent Archival of Relational Databases) ist ein Format zur Archivierung relationaler Datenbanken unabhängig von ihrem Management-System. SIARD kann Schemata, Relationen, Metadaten und den Inhalt von Datenbanken in XML abbilden. Da SIARD mit dem Standard SQL:2008 konform ist, kann es die meisten neueren Datenbanksysteme archivieren. Die Konformität mit XML und SQL:2008 Standards verspricht eine nachhaltige und interoperable Langzeitspeicherung von Datenbanken sowie optional einer breiten Palette zusätzlicher Informationen zu ihrem Nutzungskontext. Durch die kostenlose Software SIARD Suite ist sowohl das Archivieren von Datenbanken als auch das Überführen bereits archivierter Datensätze in neue Datenbankanwendungen einfach und userfreundlich möglich.</p>	
<p>TEI</p>	
<p>TEI (Text Encoding Initiative) ist ein Konsortium, das einen Standard zur elektronischen Codierung von Texten entwickelt hat. TEI bezeichnet auch eine Markup-Sprache bzw. eine Menge von Grammatiken (XML-Schemata), die kombiniert werden können, um eine solche zu definieren (hier als TEI/XML bezeichnet).</p> <p>Digitale Editionen, die mit TEI arbeiten, verwenden in der Regel ein TEI-Schema zur strukturierten Erfassung der archivischen Metadaten der Quellen sowie zur textkritischen und semantischen Auszeichnung der Primärdaten. Die Editions- und Transkriptionsrichtlinien müssen ausführlich dokumentiert werden.</p> <p>Im Umfeld der Text Encoding Initiative entwickeln sich eine Vielzahl von (Sub-)Initiativen, die sich am Standard orientieren oder diesen erweitern. Darunter fällt insbesondere correspsearch (<https://correspsearch.net/de/start.html>, Stand: 21.03.2022), EpiDoc für Epigraphika (<https://epidoc.stoa.org/gl/latest/intro-intro-fr.html>, Stand: 21.03.2022), TEI Lex-0 für Wörterbücher (<https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>, Stand: 21.03.2022), Vanilla TEI.⁴⁰</p> <p>CEI (Charters Encoding Initiative), die sich an die TEI anlehnt bzw. diese erweitert mit Auszeichnungen von Urkunden nach Vorgaben der Diplomatik (bspw. Auszeichnung von Arenga, Eschatokoll etc.).</p>	

⁴⁰ TEI XML Vanilla ist ein neues «minimal Format», getragen durch den Verein [e-editiones](#) und auch in den [TEI-Publisher](#) integriert, siehe <<https://e-editiones.org/news/tei-vanilla-a-scoop-of-tei-for-everyone/>>, Stand 21.03.2022.

<p>MEI (Music Encoding Initiative) lehnt sich an den Vorgaben und der Logik von TEI an, ist aber eine speziell auf die Auszeichnung von Musik zugeschnittene XML-Auszeichnungssprache.</p> <p>Der TEI-Publisher ist auf die Publikation von TEI XML kodierten Dokumenten und die Indexierung in einer existDB ausgelegt. Der Publisher folgt dem Single-Source Publishing Ansatz, der aus einem Dokument unterschiedliche Ausprägungen herstellen kann.</p>	
<p>Eine Streichung in einem gedruckten oder handschriftlichen Text wird mit dem Tag ausgezeichnet. Für eine Maschine und einen Menschen wird somit klar, dass die von den Tags umfassten Zeichen gestrichen sind.</p>	
<p>EAD</p>	
<p>EAD (Encoded Archival Description) ist ein dokumentarischer XML-Standard zur Beschreibung von Archivalien und Findhilfen in Archiven, Museen und Bibliotheken, der von der Library of Congress herausgegeben wird.⁴¹</p>	
<p>Ontologien</p>	
<p>OWL</p>	
<p>OWL (Web Ontology Language) ist eine Erweiterung von RDF zur Bildung komplexer Vokabulare und Ontologien (schliesst →RDFS ein). Die Notation basiert auf RDF. Durch Ontologien lässt sich eine Domäne (ihre Gegenstände und deren Beziehungen zueinander) formalisiert beschreiben. Bestehende Ontologien können wiederverwendet werden.</p>	
<p>SKOS</p>	
<p>SKOS (Simple Knowledge Organization Systems) ist eine standardisierte formale Sprache zur Ordnung von Wissen, beispielsweise in Form von Thesauri.</p>	

⁴¹ Einführung auf französisch: <https://www.ead-bibliotheque.fr/>

CIDOC-CRM	
<p>CIDOC-CRM (Conceptual Reference Model von CIDOC) ist eine Ontologie, die implizite und explizite Konzepte und Beziehungen von Institutionen modelliert, die das Kulturerbe bewahren und dokumentieren (z. B. Museen). Das Comité International pour la Documentation (CIDOC) ist ein internationales Komitee für Museumsdokumentation. Einzelne Fachgruppen erweitern CIDOC-CRM fachspezifisch, z.B. FRBR. Im Rahmen des Data for History Konsortium (http://dataforhistory.org/) werden auf die Bedürfnisse der Historiker:innen angepasste Version von CIDOC-CRM erarbeitet.</p>	
DOLCE	
<p>DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) ist eine Ontologie, die den menschlichen Diskurs modelliert.</p>	
Ric-O	
<p>RiC-O (Records in Contexts-Ontology) ist eine → OWL-Ontologie zur Beschreibung von Archivbeständen und damit zusammenhängende Einheiten.</p>	
Schema.org	
<p>Initiative zur einheitlichen Strukturierung von Daten auf Webseiten nach einer von schema.org entwickelten Ontologie.</p> <p>Eine Ontologie, die auf Schema.org gehostet wird, ist friend-of-a-friend, um Beziehungen zwischen Personen auszudrücken.</p>	
<p>Einfache RDF/S-OWL-Ontologien: → Dublin Core für Metadaten oder → Friend-of-a-friend für Personen</p> <p>Domain-Ontologien: → FRBR oder SPAR-Serie (FRBR wiederverwendet)</p> <p>Anwendung für diese Ontologien: DBPedia, eine RDFisierte Version der Wikipedia oder Europeana, mit entsprechenden Abfragesystemen.</p>	

Austauschformate/Austauschprotokolle

Web ARChive (WARC)

Mit dem [WARC](#) Archivformat lassen sich digitale Ressourcen in einer aggregierten Archivdatei mit zugehörigen Metadaten speichern. Es stellt eine Weiterentwicklung des ARC-Dateiformats des [Internet Archive](#) dar und gilt als das am meisten verwendete und unterstützte Format. Es gibt ein grosses Ökosystem von Analysewerkzeugen etc. unter <https://github.com/iipc/awesome-web-archiving>, Stand: 21.03.2022.

IIIF

Das Image Interoperability Format ([IIIF](#)) wird gesteuert durch ein Konsortium und legt einen Standard zur Übernahme von Bilddaten, inklusive Metadaten fest.⁴² Pfad und Angaben zu Metadaten werden in sogenannten Manifesten bereitgestellt. Ziel von IIIF ist, dass Bilder, Bildteile, Audio, Video und Metadaten auf Seiten von Dritten eingebunden und weiterverwendet werden können, ohne dass eine Duplizierung der Bilddateien notwendig ist.

⁴² Siehe die Website: o.A.: International Image Interoperability Framework, o.D., <https://iiif.io/>, Stand: 23.12.2021.

[e-codices](#) präsentiert mittelalterliche und frühneuzeitliche Handschriften über einen eigenen IIIF-fähigen Viewer und publiziert alle Inhalte mit IIIF-Manifesten (Metadaten und Bilddateien).

Distributed Text Services (DTS)

[DTS](#) ist ein sich in Entwicklung befindender Standard, der eine Schnittstelle/→API definiert, um Texte und Textteile abzurufen. In Analogie zu → [IIIF](#) soll DTS Texte schematisch adressiert bereitstellen. Mehrere Endpoints erlauben das Ansprechen von Sammlungen ([Collections](#)), der [Navigation](#) und [Dokumente](#).

Canonical Text Services (CTS)

Die Spezifikation für kanonische Textdienste definiert einen Netzdienst zur Identifizierung von Texten und zum Abrufen von Textfragmenten unter Verwendung von Begriffen wie «Werk» und «Zitat», die in den Antikenwissenschaften und anderen Disziplinen üblich sind.

Die CTS-Hauptseite (<http://cite-architecture.org/cts/>) ist aktuell nicht erreichbar. Die Sourceforge-Seite wurde 2015 letztmals aktualisiert: <http://cts3.sourceforge.net/>.

Die [Perseus Digital Library](#) nutzt die CTS Architektur. Die Perseus CTS API/End-point ist verfügbar unter: <http://cts.perseids.org/api/cts/>. Ein browser-kompatibles Interface kann hier aufgerufen werden (<http://cts.perseids.org/>).

IPIF/ProsopogrAPhI International Prosopographical Interchange Framework)

Mit [IPIF/ProsopogrAPhI](#) schlägt die Österreichische Akademie der Wissenschaften einen Standard zum Austausch prosopographischer Daten vor.

Datenablage	
Git	
<p>Das aus der Softwareentwicklung bekannte System zur Versionskontrolle Git bietet Sicherheit und Flexibilität bei der Archivierung von Quellcode und Daten.</p> <p>GitLab ist eine Open Source git-Plattform, die von SWITCH und vielen Universitäten (bspw. Genf und Bern) in der Schweiz betrieben wird. GitHub ist ein häufig genutzter privatwirtschaftlicher Webservice.</p> <p>Bei der Entwicklung von Git stand die Integrität des verwalteten Quellcodes an erster Stelle. Der Inhalt der Dateien sowie die tatsächlichen Beziehungen zwischen Dateien und Verzeichnissen, Versionen, Tags und Commits – all diese Objekte im Git-Repository sind mit einem kryptografisch sicheren Hash-Algorithmus (SHA1) gesichert. Dies schützt den Code und den Änderungsverlauf sowohl vor versehentlichen als auch vor böswilligen Änderungen und stellt sicher, dass der Verlauf vollständig nachvollziehbar ist.</p> <p>Eines der wichtigsten Ziele bei der Entwicklung von Git ist die Flexibilität. Git ist in mehrfacher Hinsicht flexibel: in der Unterstützung verschiedener Arten von nicht linearen Entwicklungsabläufen, in seiner Effizienz bei kleinen und grossen Projekten und in seiner Kompatibilität mit vielen bestehenden Systemen und Protokollen.</p>	
DLCM/OLOS	
<p>Das Data Life Cycle Management (DLCM) in Genf hat die Entwicklung von OLOS im Auftrag der Schweizerischen Hochschulkonferenz vorangetrieben. Mit OLOS steht auf nationaler Ebene eine Lösung für das Datenmanagement von Forschungsdaten (im weitesten Sinn) über den gesamten Lebenszyklus hinweg zur Verfügung.</p>	
DaSCH	
<p>Das Swiss National Data & Service Center for the Humanities (DaSCH) ist eine nationale Forschungsinfrastruktur, welche den langfristigen Zugang zu komplexen Forschungsdaten aus dem Bereich der Geisteswissenschaften sicherstellt und auch als Projektarbeitsumgebung dienen kann. Sie fördert die Vernetzung der Daten mit anderen Datenbeständen (Linked Open Data) und schafft einen Mehrwert für die</p>	

<p>Forschung und die interessierte Öffentlichkeit. Die bereits bestehende DaSCH Service Platform (DSP) dient für die Speicherung, den Austausch und die Arbeit mit geisteswissenschaftlichen Primärquellen und Daten. Über generische Webapplikationen sind die Daten suchbar und auch aktualisierbar. Jedes Objekt in DSP besitzt einen eigenen permanenten Identifikator (ARK), der es erlaubt, gezielt einzelne Dateneinträge zu referenzieren. Die Plattform wird in den kommenden Jahren sukzessive ausgebaut und um Funktionalitäten erweitert.</p>	
<p>ZENODO</p>	
<p>Zenodo ist ein Repository des CERN, das durch die Europäische Union mitfinanziert wird und für die Ablage wissenschaftlicher Daten, aber auch für Publikationen und Präsentationen geeignet ist. Auf GitHub gehostete Projekte können via ZENODO als Abzug langfristig gespeichert werden. Aktuell besteht eine Einschränkung, da ein ZENODO-Datenset immer nur von einem GitHub-Konto gefüttert werden kann, was aus Nachhaltigkeitsperspektive nicht ideal ist. Eine Verbindung nach Zenodo wurde auch für Gitlab schon vor längerer Zeit angeregt, aber es tat sich bisher nicht viel (immerhin gibt es zweckdienliche Skripte im frühen Beta-Stadium).</p>	