arbido

2017/3 Metadaten – Datenqualität

Zala Sacha, Direktor Diplomatische Dokumente der Schweiz

Sibille Christiane, Leiterin Digitale Innovation, Diplomatische Dokumente der Schweiz

Tags: Gedächtnispolitik Elektronisch, Erschliessung, Normen und Standards,

Metagrid

Metagrid ist ein Projekt der Schweizerischen Akademie für Geistes- und Sozialwissenschaften (SAGW) für die Online-Vernetzung von geisteswissenschaftlichen Ressourcen, durchgeführt von den Diplomatischen Dokumente der Schweiz (Dodis) mit der Unterstützung des Historischen Lexikons der Schweiz (HLS).

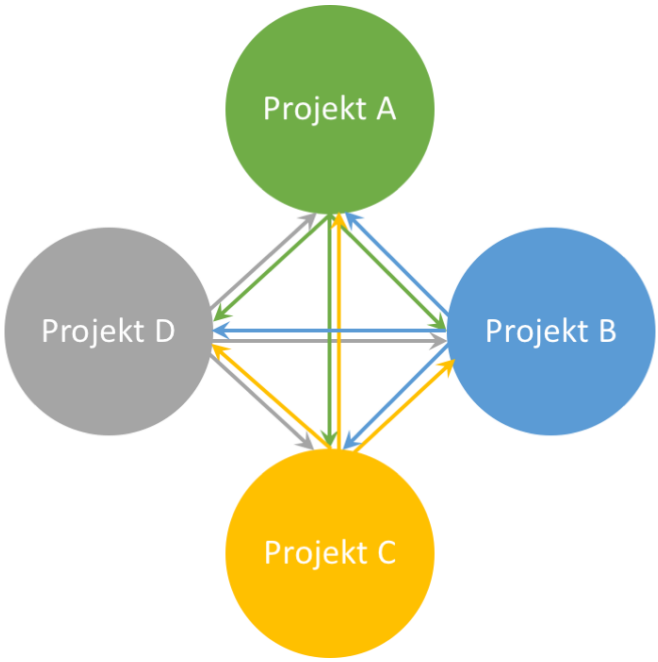
Der Webservice Metagrid (www.metagrid.ch) ermöglicht die Einrichtung, Verwaltung und Analyse von Links zwischen identischen historischen Akteuren von verschiedenen Websites und Datenbanken. Im Mittelpunkt steht die Idee der pragmatischen Vernetzung von Inhalten historisch ausgerichteter Forschungsdaten. Hiervon sollen zunächst die Nutzerinnen und Nutzer profitieren, deren Recherche durch eine Erweiterung des Informationsangebots vereinfacht wird. Gleichzeitig wird aber auch die Visibilität der beteiligten Partnerprojekte dadurch erhöht, dass ihre Inhalte durch eine semi-automatische gegenseitige Verlinkung auf den Websites der Partnerprojekte dargestellt werden können.

Die in Metagrid verwalteten Inhalte können von interessierten Nutzerinnen und Nutzern auf zwei unterschiedlichen Wegen konsultiert werden. Seit Herbst 2016 ist die zentrale Metagrid-Suche (http://www.metagrid.ch/metagrid_search/#/search.html) online. Hier können alle in Metagrid gespeicherten Personen zentral recherchiert und angezeigt werden.
Gleichzeitig werden dort, wo bereits Konkordanzen erstellt wurden, die vernetzten Daten auch auf den jeweiligen Seiten der beteiligten Projektpartner dargestellt, wie dieses Beispiel zeigt: http://dodis.ch/P1047

Die Idee von Metagrid - Pragmatische Vernetzung statt Normierung

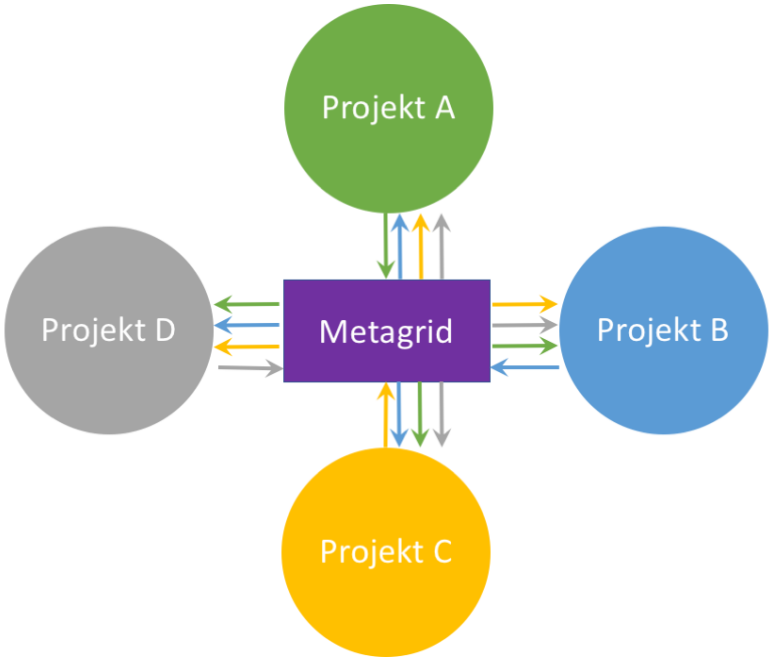
Das Ziel von Metagrid ist die Vernetzung von Informationen, die im Rahmen historischer Forschungsprojekte oder Infrastrukturen gewonnen und digital zur Verfügung gestellt wurden. Der Ansatz hierbei ist partnerschaftlich und föderal. Das heisst, dass die Daten jedes Projekts auf gleicher Ebene miteinander verlinkt werden können, ohne dass eine übergeordnete Norminstanz vorhanden sein muss. Das Kernelement von Metagrid ist die «Konkordanz». Sie speichert die Information darüber, dass eine Person X aus Projekt A der Person Y aus Projekt B entspricht und ermöglicht somit die Verlinkung der beiden Personenseiten. Die von Metagrid erstellten Konkordanzen haben explizit keinen normativen oder normalisierenden Anspruch. Die Pflege und Verantwortung für die Ausgangsdaten bleibt in der Kompetenz der einzelnen Projekte, da sie dazu sowohl die wissenschaftlichen Kenntnisse als auch die mit den Projektzielen verbundenen und Fragestellungsorientierten Interessen mitbringen.Dementsprechend pflegt Metagrid selbst keine Informationen zu den vernetzten Personen und greift auch explizit nicht in die Daten der beteiligten Projektpartner ein. Bildlich gesprochen hat Metagrid die Funktion eines Relais': Wenn zu einer Person in mehr als einem beteiligten Projekt Informationen vorliegen, können diese vernetzt und dann bei allen beteiligten Partnern angezeigt werden, wenn die jeweilige Personenseite aufgerufen wird.

Während Metagrid selbst keine normalisierende Zielsetzung verfolgt, stellt es jedoch eine wichtige Schnittstelle zu existierenden Normdatensätzen dar. So bietet Metagrid die Möglichkeit, die Daten der Partnerprojekte mit den Personendatensätzen der GND abzugleichen und zu vernetzen. Die so entstandenen Konkordanzen können als Beacon-Dateien (https://de.wikipedia.org/wiki/Wikipedia:BEACON) ausgegeben und auf diese Weise auch ausserhalb von Metagrid vernetzt werden.



Metagrid 1

Während Metagrid selbst keine normalisierende Zielsetzung verfolgt, stellt es jedoch eine wichtige Schnittstelle zu existierenden Normdatensätzen dar. So bietet Metagrid die Möglichkeit, die Daten der Partnerprojekte mit den Personendatensätzen der GND abzugleichen und zu vernetzen. Die so entstandenen Konkordanzen können als Beacon-Dateien (https://de.wikipedia.org/wiki/Wikipedia:BEACON) ausgegeben und auf diese Weise auch ausserhalb von Metagrid vernetzt werden.



Metagrid 2

Den Metadaten sei Dank - Algorithmusbasierte semi-automatische Vernetzung

In der Projektpraxis lebt Metagrid von den Daten der beteiligten Partner und den ebenfalls in Metagrid integrierten Normdaten. Um den Aufwand für den Betrieb möglichst gering zu halten wurden wichtige Schritte im Workflow soweit wie möglich automatisiert und vereinfacht. Die Qualität der verwendeten Metadaten spielt dabei eine wichtige Rolle. Prinzipiell lassen sich folgende Arbeitsschritte unterscheiden: Import, Analyse, Vernetzung, redaktionelle Bearbeitung, Darstellung.
Der Pilotbetrieb von Metagrid hat gezeigt, dass es für die beteiligten Projekte am besten ist, wenn der Aufwand für die Datenübergabe gering ist. Daher erfolgt der Import der Daten soweit wie möglich automatisch. Hierfür wird für jedes Projekt ein Crawler entwickelt, der die Daten auf der Ursprungsseite ausliest. Hierfür ist es notwendig, dass jede Personenentität über eine eigene permanent referenzierbare Seite verfügt (URI). Die Datenanalyse wird durch die strukturierte Auszeichnung von Metadaten, beispielsweise mit Hilfe von schema.org, beschleunigt und insbesondere die Erkennung von Namensbestandteilen und Lebensdaten vereinfacht. Neben diesen Informationen ist der Crawler ausserdem in der Lage bereits vorhandene manuelle Links zu Metagrid-Partnern zu erkennen. Diese Links können im nächsten Schritt ausgewertet und basierend auf diesen Informationen automatisiert neue Konkordanzen erstellt werden. Nachdem die Daten in Metagrid abgelegt wurden, haben die Projekte die Möglichkeit, ihre Daten mit den bereits vorhandenen Daten zu vergleichen und auf mögliche Übereinstimmungen hin zu überprüfen. Für die Zukunft ist die Anreicherung von eigenen Daten aufgrund der in Metagrid gespeicherten Informationen geplant.
Für diese semi-autoamtische Überprüfung kann auf ein eigenes Redaktionssystem zurückgegriffen werden, das variable Überprüfungskriterien ermöglicht. So können in einem ersten Schritt solche Datensätze überprüft werden, deren Namens- und Lebensdatenfelder eine sehr hohe Übereinstimmung haben und anschliessend durch Anpassung der Präzision die Treffermenge schrittweise erweitert werden. Daneben verfügt das Redaktionssystem über eine Editiermöglichkeit, die jederzeit eine manuelle Bearbeitung der Konkordanzen (Erstellen, Erweitern, Löschen) ermöglicht. Die auf diese Weise in Metagrid erstellten Konkordanzen werden sofort freigeschaltet und können von den Projektpartnern mit Hilfe eines konfigurierbaren Widgets auf ihren Seiten dargestellt werden.

Metagrid 3

Ausblick - histhub.ch

Metagrid wird künftig seine Vernetzungstechnologien im Rahmen des Projekts histHub zur Verfügung stellen. Um das Potenzial der Strukturierung und Vernetzung digitaler Daten künftig gemeinsam zu nutzen und weiterzuentwickeln haben sich die Rechtsquellenstiftung des Schweizerischen Juristenvereins mit ihrer Sammlung Schweizerischer Rechtsquellen (SSRQ), die Forschungsstelle der Diplomatischen Dokumente der Schweiz (DDS) mit ihrer Datenbank Dodis, das Historische Lexikon der Schweiz (HLS), das Schweizerdeutsche Wörterbuch (Idiotikon) und das Portal der schweizerischen Ortsnamenforschung (ortsnamen.ch) im Consortium Historicum zusammengeschlossen und das Projekt histHub gestartet. histHub entwickelt eine Plattform für vernetzte und normierte Daten für die historischen Wissenschaften in der Schweiz. In diesem Kontext werden weitere Werkzeuge zur Vernetzung historischer Entitäten entwickelt werden. Neben den Aktivitäten im Bereich Vernetzung wird sich ein weiterer Projektteil mit der Normierung historischer Forschungsdaten beschäftigen. Durch den Einsatz semantischer Datenmodellierung soll der histHub-Normdatensatz einfach adaptierbar und in der Lage sein, feinste historische Unterschiede abzubilden.

Mitmachen

Projekte, die sich für eine Kooperation mit Metagrid interessieren, können sich via info@metagrid.ch oder auf der Website http://www.metagrid.ch/coopera... anmelden. Weitergehende Informationen und Best-Practice Informationen findet man hier: http://www.metagrid.ch/best-pr...

Sacha Zala

Sacha Zala, Dr. phil., studierte Geschichte, Politikwissenschaft und Staatsrecht an der Universität Bern und an der University of North Carolina at Chapel Hill. Er ist Historiker und Direktor der Forschungsstelle Diplomatische Dokumente der Schweiz (Dodis), ein Forschungsunternehmen der Schweizerischen Akademie der Geistes- und Sozialwissenschaften (SAGW). In dieser Funktion hat er etliche Dodis-Projekte im Bereich der Digital Humanities initiiert, darunter die Vernetzungsinitiative Metagrid. Seit 2014 ist Sacha Zala Präsident der Schweizerischen Gesellschaft für Geschichte (SGG).

Christiane Sibille

Dr. Christiane Sibille hat Geschichte und Musikwissenschaft in Heidelberg studiert und arbeitet heute an den (digitalen) Schnittstellen zwischen Forschung, Archiven und Bibliotheken. Zu ihren beruflichen Stationen zählen die Universitäten Heidelberg und Basel, die Zürcher Hochschule der Künste sowie die Forschungsstelle Diplomatische Dokumente der Schweiz (Dodis) und das Projekt Metagrid.ch.

Seit Juni 2021 ist Christiane Sibille in der Sektion Sammlungen und Archive der ETH-Bibliothek Verantwortliche für die Fachstelle «Digital Scholarship Services». Ausserdem unterrichtet sie regelmässig zu digitalen Themen an der Universität Basel und ist Präsidentin des Vereins «Geschichte und Informatik».

Abstract

Deutsch
Metagrid ist ein Projekt der Schweizerischen Akademie für Geistes- und Sozialwissenschaften (SAGW) für die Online-Vernetzung von geisteswissenschaftlichen Ressourcen, durchgeführt von den Diplomatischen Dokumente der Schweiz (Dodis) mit der Unterstützung des Historischen Lexikons der Schweiz (HLS). Der Webservice ermöglicht die Einrichtung, Verwaltung und Analyse von Links zwischen identischen historischen Akteuren von verschiedenen Websites und Datenbanken. Im Mittelpunkt steht die Idee der pragmatischen Vernetzung von Inhalten historisch ausgerichteter Forschungsdaten.

Français
Metagrid est un projet de mise en réseau de ressources dans le domaine des sciences humaines lancé par l'Académie suisse des sciences humaines et sociales (ASSH) et mis en œuvre par les Documents Diplomatiques Suisses (Dodis) avec le soutien du Dictionnaire historique de la Suisse (DHS). Le système permet de créer, gérer et analyser des liens entre des acteurs historiques identiques de différents sites web et bases de données.

Italiano
Metagrid è un progetto dell'Accademia svizzera di scienze umane e sociali (ASSU) per la messa in rete di risorse nell'ambito delle scienze sociali realizzato dai Documenti Diplomatici Svizzeri (Dodis), con il sostegno del Dizionario Storico della Svizzera (DSS). Il web service consente di stabilire, gestire e analizzare link tra entità uguali di diversi siti Internet e banche dati.