

Günter Hipler, Nicolas Prongué und René Schneider


Swissbib und linked.swissbib.ch: Leistung und Potenziale einer offenen Plattform für Schweizer Bibliotheksdaten


Abstract: Der Artikel beschreibt die schweizweiten Initiativen swissbib und linked.swissbib.ch. Swissbib kann als eine Plattform verstanden werden, die auf das Data Processing – insbesondere auf die Dedoublierung – der Katalogdaten der schweizerischen Bibliotheksverbände ausgerichtet ist und verschiedene auf diesen Daten basierende Services anbietet. Die Plattform baut dabei auf einer offenen Infrastruktur auf und erlaubt flexible Erweiterungen. Zu diesen Erweiterungen gehören die im Projekt linked.swissbib.ch entwickelten Komponenten, die die Katalogdaten interoperabel darstellen und eine Verlinkung mit anderen Daten erlauben. Parallel dazu wurden nahezu sämtliche Metadaten unter einer Creative Commons 0-Lizenz der Allgemeinheit zur Verfügung gestellt, was den Rahmen der gemeinschaftlichen Neu- und Nachnutzung bibliographischer Daten noch mehr ausweitet. Mit dem Ende des Projekts linked.swissbib.ch enden die Linked Data-Aktivitäten nicht, sondern werden im Kontext von swissbib kontinuierlich weiterverfolgt.

Einleitung

Bibliotheken sind seit Jahrhunderten Bastionen der Verwaltung von Schriftgut und wenige Berufe können eine derart lange Tradition in der Verwaltung von Metadaten vorweisen wie Bibliothekarinnen und Bibliothekare. Im Zeitalter der Digitalisierung wurde dazu ein äusserst erfolgreiches Format zum Austausch dieser Metadaten entwickelt: MARC21. Allerdings führte dieses Format nicht unbedingt zu einer Homogenisierung bei der Verwaltung der Metadaten; zudem wurde der Austausch über die Grenzen der Bibliotheken hinweg völlig vernachlässigt.

Swissbib zielte von Anbeginn darauf ab, eine flexible und offene Plattform für die Zusammenführung und Verarbeitung der bibliographischen Daten der zahlreichen Verbände der schweizerischen Bibliothekslandschaft zu werden und damit die Öffnung der Metadaten zu unterstützen. Eine solche Öffnung ist durch Linked Open Data möglich geworden, denn damit wird den Bibliotheken die Möglichkeit gegeben, einerseits ihre Erfahrungen bei der Standardisierung der Metadaten in die Gestaltung des Semantic Web einzubringen und andererseits

 Open Access. © Günter Hipler, Nicolas Prongué und René Schneider, publiziert von De Gruyter.

 Dieses Werk ist lizenziert unter der Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 Lizenz.

<https://doi.org/10.1515/9783110553796-008>

Unauthenticated

Download Date | 10/3/18 11:01 AM

den überfälligen Austausch ihrer Metadaten über die Grenzen der Bibliotheksmauern hinweg zu bewerkstelligen.

Der nachstehende Artikel stellt die beiden Initiativen swissbib und linked.swissbib.ch vor.¹

Die swissbib-Plattform als Basis für semantisch verlinkte Datenstrukturen

Die seit Februar 2010 im produktiven Einsatz stehende swissbib-Plattform wird von ihren zahlreichen Nutzern in der Regel lediglich über die Suchoberfläche www.swissbib.ch wahrgenommen. Dabei ist dies nur ein einziger, wenn auch der bekannteste Service neben einer Reihe weiterer, die alle nur möglich sind, weil swissbib eine flexible, leistungsfähige und leicht erweiterbare Data Processing Komponente innerhalb der Gesamtplattform besitzt. Diese Datenkomponente führt die sehr heterogenen Daten von rund 960 Institutionen aus dem Bibliotheksbereich zusammen, reichert diese an und aggregiert sie teilweise in Clustern. Weitgehend homogene Daten im klassischen MARC21-Format sind das Ergebnis dieses Datenworkflows und dienen verschiedenen Benutzer- und Maschinenservices als Grundlage.

Der beschriebene Prozess findet jedoch nicht nur einmalig und statisch statt, sondern wiederholt sich täglich. Die Aktualisierungen der angeschlossenen Datenquellen werden jede Nacht abgeholt und in die swissbib-Plattform integriert, was täglich einen Datenstrom von ca. 100'000 bis zu 600'000 Ereignissen (Metadateninformationen) erzeugt. Die Datenbasis wird regelmässig vergrössert. So erfolgte beispielsweise Ende 2016 die Integration von mehr als 6 Millionen Artikelaufnahmen, die mit hohem Kostenaufwand im Rahmen des nationalen Projekts Nationallizenzen beschafft worden waren. Da alle swissbib-Services die Datenbasis als Ganze oder, je nach Bedarf, auch nur teilweise nutzen können, stehen neu integrierte Daten auch dem linked.swissbib.ch-Service automatisch zur Verfügung.

¹ Zum Projekt linked.swissbib.ch siehe auch Nicolas Prongué, René Schneider: Data Streams in linked.swissbib.ch. The Swiss Metacatalog in the Linked Open Data Cloud. In: Maria Gäde u. a. (Hrsg.): Everything Changes, Everything Stays the Same? Understanding Information Spaces. Proceedings of the 15th International Symposium of Information Science (ISI 2017); Berlin, Germany, 13th–15th March 2017. Glückstadt/Elbe 2017 (Schriften zur Informationswissenschaft 70), S. 359–361.

In den Diskussionen zwischen den beteiligten Projektpartnern von linked.swissbib.ch wurde schnell klar, dass diese Architektur eine ideale Grundlage für ein vor allem datengetriebenes Projekt ist, das zum Ziel hat, die wegen des vorherrschend zeichenkettenbasierten MARC21-Formats wenig oder schwach verlinkten Daten in ein vokabulargestütztes Modell wie RDF (Resource Description Framework) zu überführen, das eine bessere Verknüpfung sowie Maschinenlesbarkeit erlaubt.

In den Vorabklärungen beschäftigten sich die Projektpartner auch mit der Frage, ob die technische Architektur der bestehenden swissbib-Plattform den Anforderungen einer erweiterbaren und skalierbaren Komponente für neue verlinkte Datenstrukturen genügt und die Möglichkeiten zur Nachnutzung auch durch externe Projekte gegeben sind. Die Frage konnte eindeutig bejaht werden. Bereits in der nach GATT-Kriterien durchgeführten weltweiten Ausschreibung für die anfängliche Infrastruktur der swissbib-Plattform im Jahre 2008 lag ein sehr hohes Gewicht auf den Bewertungen der unabhängigen Softwarekomponenten, die im Zusammenspiel und durch die Nutzung von offenen Schnittstellen neuen Nutzerwünschen im sich ständig verändernden digitalen Umfeld gerecht zu werden vermögen. Dabei wurde und wird undogmatisch entweder kommerzielle oder frei verwendbare Open Source-Software nach Best Effort-Kriterien gewählt. Die stetig wachsende Bedeutung von Open Source-Software vor allem in den letzten fünf Jahren spiegelt sich auf der swissbib-Plattform wider, wo (Stand 2017) lediglich ‚eine‘ kommerzielle Komponente im Einsatz ist.

Linked.swissbib.ch

Der eingangs erwähnten Aufgabe der Öffnung in Richtung Interoperabilität stellten sich die Kollegen der Universitätsbibliothek Basel, der Haute école de gestion Genève und der Hochschule für Technik und Wirtschaft (HTW) Chur, als sie sich daran machten, ein Linked Data-Projekt für die bibliographischen Daten von swissbib aufzugleisen, dessen Name mit „linked.swissbib.ch“ schnell gefunden war. Nahezu genauso schnell wurde dann eine Reihe von Prämissen formuliert, die als verbindliche Leitlinien für die gesamte im Projekt zu leistende Arbeit galten und das Projekt während der gesamten Laufzeit begleiteten:

- Linked.swissbib.ch wurde von Anfang an als Ergänzung zur bereits bestehenden swissbib-Plattform angesehen. Auch wenn es sich bei linked.swissbib.ch selbst nur um ein Projekt mit begrenzter Laufzeit handelte, sollten alle Tätigkeiten von Anfang an auf eine möglichst nahtlose spätere Integration in die swissbib-Plattform und die Schaffung einer Infrastruktur zur Konversion von bibliographischen Daten zu Linked Data abzielen.

- ‚Linked Data‘ bedeutet hauptsächlich eine Interoperationalisierung von Daten, um eine a priori nicht absehbare Nachnutzung in ähnlichen oder auch ganz anderen Kontexten zu ermöglichen. Das heisst konkret, dass robuste Schnittstellen zum Datenaustausch zwischen Maschinen geschaffen werden müssen, die den anfallenden Datenmengen standhalten. In diesem Kontext stehen demnach die Daten im Vordergrund und es kann von einer datenorientierten bzw. datengetriebenen Vorgehensweise gesprochen werden.
- Gleichzeitig erhalten Daten und die um sie herum erstellten Systeme ihre Bedeutung aufgrund der Nutzung durch die Benutzerinnen und Benutzer. Somit gilt Wittgensteins berühmtes Diktum „Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache“² (PhU §43) nicht nur für die menschliche Sprache, sondern – mit den notwendigen kontextuellen Anpassungen – auch für Daten. Tim O’Reilly³ drückt dies ähnlich, dafür etwas salopp und dennoch äusserst prägnant in seiner Web 2.0 MemeMap mit dem Motto „Data as the new Intel inside“ aus. Von daher war es den Projektverantwortlichen ein Bedürfnis, der Nutzerin bzw. dem Nutzer auch den Mehrwert von Linked Data zu zeigen. Dies kann insbesondere durch die Verlinkung und die Anreicherung mit anderen Daten geschehen. Von daher wurde parallel zur datenorientierten eine benutzerorientierte Vorgehensweise gewählt, die auf eine Veränderung der Benutzerschnittstelle abzielte. Interessanterweise ist diese Vorgehensweise wiederum sehr stark abhängig von der Qualität der anzureichernden Ausgangsdaten und der Daten, die bei der Anreicherung hinzugefügt werden, so dass erkennbar ist, dass im Grossen und Ganzen die Daten bei der Projektrealisierung immer stärker im Vordergrund standen, was möglicherweise als allgemeine Tendenz und sich immer stärker manifestierender Trend festzuhalten ist.
- Bewusst ausgeklammert wurde bei den ersten Überlegungen die Frage nach der Lizenzierung der Daten, da dies aufgrund der eher technischen Ausrichtung des Projekts und des gewählten Zeitraums von zweieinhalb Jahren als ein nachgeordneter Aspekt eingeschätzt wurde. Diese Annahme erwies sich jedoch als irrig, da sich das Thema einer Lizenzierung der bibliographischen Daten als ‚Public Domain‘ oder unter einer Creative Commons 0-Lizenz mit zunehmender Projektdauer immer stärker in den Vordergrund drängte.

² Ludwig Wittgenstein: Philosophische Untersuchungen. Kritisch-genetische Edition. Herausgegeben von Joachim Schulte. Frankfurt 2001.

³ Tim O’Reilly: What is Web 2.0? Design patterns and Business Models for the Next Generation of Software. Sebastopol CA 2009 (<http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>). Alle Links in diesem Beitrag wurden am 26.03.2018 überprüft.

- In der Konzeptionsphase des Projekts wurde auch darauf geachtet, dass mit dem Projekt linked.swissbib.ch nicht ‚das Rad neu erfunden‘ wird, sondern möglichst vorhandene und bewährte Komponenten, die von vergleichbaren Institutionen für ähnliche Projekte bereits eingesetzt werden, genutzt werden. Für die Architektur des neuen Services hatten Softwareartefakte aus den Projekten Culturegraph der Deutschen Nationalbibliothek sowie lobid.org des Hochschulbibliotheksentrums Köln Beispielcharakter.

Anhand dieser fünf Kerngedanken sollen im Folgenden die Ergebnisse des Projekts und die weitere Verfolgung des Linked Data-Gedankens im Kontext von [swissbib](http://swissbib.org) vorgestellt werden. Diese Ergebnisse lassen sich durch die Metapher ‚data lake‘ (Datensee) und ‚data streams‘ (Datenströme) veranschaulichen und werden in Abbildung 1 erläutert. Darin stellen die ursprünglichen bibliographischen Datenquellen der in [swissbib](http://swissbib.org) integrierten Bibliotheksverbünde und weitere Quellen die oberen Wasserläufe (upstreams) dar. Diese werden täglich (genauer jede Nacht) aktualisiert, dedoubliert, indexiert und transformiert. Die wichtigste Transformation findet dabei im Übergang von MARC/XML (dem altbewährten Datenformat der bibliographischen Datensätze) hin zu SOLR/XML (dem Datenformat der Open Source-Suchmaschine, die in [swissbib](http://swissbib.org) Verwendung findet) statt. Dabei handelt es sich um eine Datenmenge von etwa 30 Millionen Datensätzen. Die Nutzerinnen und Nutzer können über die [http-URI www.swissbib.ch](http://www.swissbib.ch) auf diese Daten zugreifen, für Maschinen steht unter sru.swissbib.ch eine Schnittstelle zur Verfügung. Dabei handelt es sich um die Datenströmung innerhalb des Datensees, die sich – vom Betrachter aus gesehen – auf der linken Seite der Abbildung befindet; die Schnittstellen für Benutzer und Maschinen, die sich am Ende dieses Streams befinden, wurden aus Platzgründen nicht visualisiert.

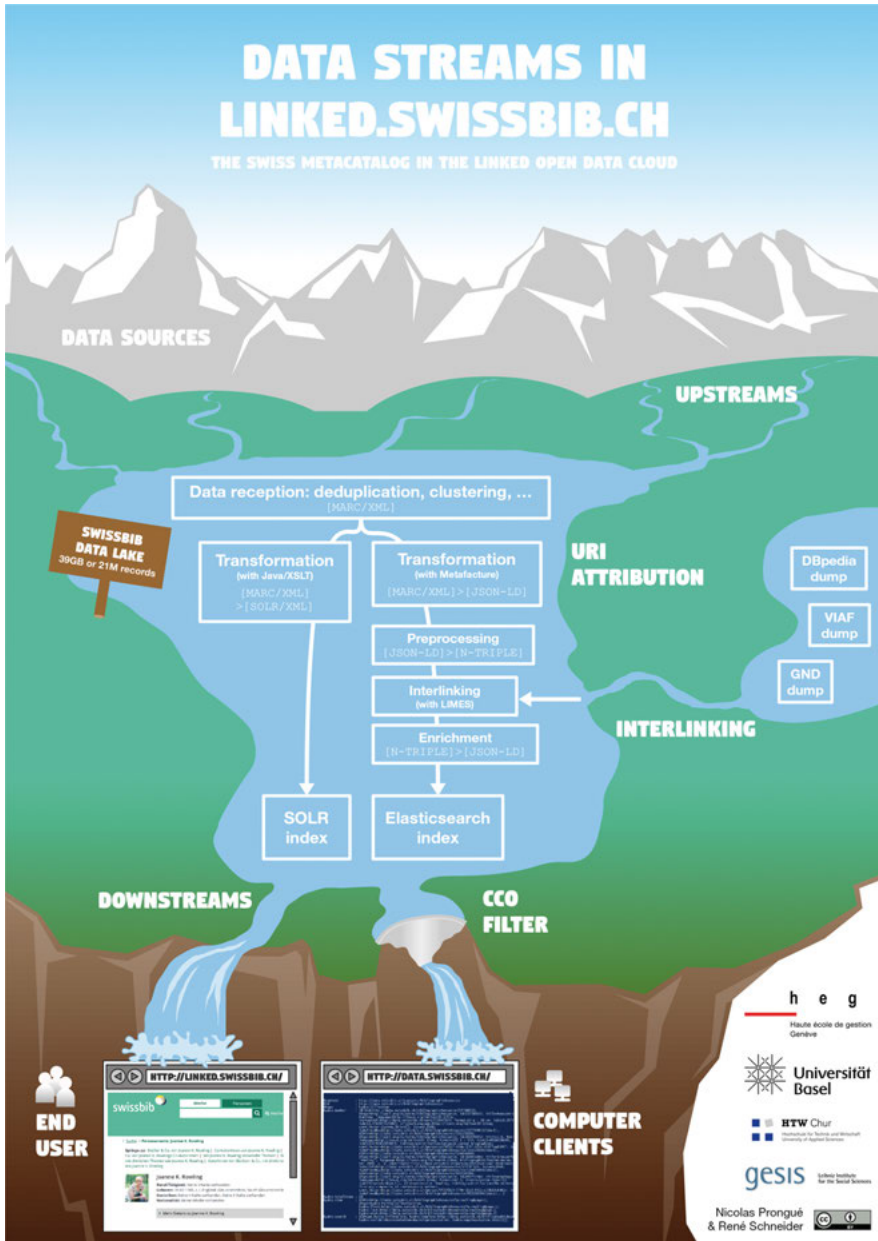


Abb. 1: linked.swissbib.ch als Datenströme und Datensee (<https://hesso.tind.io/record/1973/files/Poster.pdf>)

Die – vom Betrachter aus gesehen – rechte Strömung innerhalb des Datensees stellt die eigentliche Transformationskette oder den Workflow für die Erstellung verlinkter Daten dar. Entsprechend den anfänglichen Überlegungen wird deutlich, dass ein paralleler Workflow in swissbib aufgebaut wird. Die Erstellung dieses Datenkomplements beginnt mit einer Transformation der bibliographischen MARC/XML-Records nach JSON-LD, einem Datenformat der JSON-Familie, das auf die Bedürfnisse von Linked Data (LD) angepasst wurde.

So unscheinbar und spröde dieser Schritt anmutet, umso wichtiger ist er für alle weiteren Schritte, da er auf einem eigens erstellten Modell basiert, das für die Datentransformation der bibliographischen Daten aller schweizerischen Verbände in die Linked Data-Welt erstellt wurde. Genauer gesagt bezeichnet dies eine Transformation nach RDF, einer Art Grammatik für Linked Data, wobei JSON-LD als die eigentliche Sprache, in der sich diese Grammatik manifestiert, zu verstehen ist. Dieser Modellierungs- und Transformationsprozess ist die genuin bibliothekarische Tätigkeit, wie sie in Zukunft auch zum Alltagsgeschäft der Bibliothekarinnen und Bibliothekare gehören könnte (und sollte!).

Abbildung 2 zeigt einen Ausschnitt dieses Datentransformationsmodells.⁴ Dieses Modell impliziert eine Reihe von Umwandlungen der Ausgangsdaten, die im konkreten Fall von linked.swissbib.ch mit dem Werkzeug Metafactory realisiert wurden. Die damit zusammenhängende konkrete Arbeit nach Erstellung des Datenmodells bestand darin, die Transformationsregeln so zu kodieren, wie es den Anforderungen von Metafactory entsprach. Dabei ist festzuhalten, dass es sich um eine sowohl iterative als auch inkrementelle Arbeit handelt: iterativ im Sinne von mehreren Bearbeitungsschleifen, die das Modell sukzessive verfeinern, inkrementell im Sinne von fortlaufenden Erweiterungen der Transformationsregeln.

Allerdings hat eine reine Überführung von Datenpaketen aus einem geschlossenen Datenfonds nach RDF (so umfangreich sie auch im Fall von swissbib sein mag) nichts mit verlinkten Daten im eigentlichen Sinn zu tun. Man könnte dem entgegenhalten, dass die Akkumulierung und Dedoublierung der Daten aller Verbände schon einen Mehrwert generiert, jedoch wird dieser bereits durch die bestehenden swissbib-Komponenten herbeigeführt. Entscheidend bei Linked Data ist die Verlinkung mit anderen, d. h. externen Daten, die auf anderen Servern des World Wide Web verstreut liegen, aber im gleichen Format (eben RDF) zur Verfügung stehen. Dieser Vorgang wird gemeinhin als ‚Interlinking‘ bezeichnet und erzeugt im konkreten Fall Verweise auf externe Datenquellen, bspw. die GND (Gemeinsame Normdatei), VIAF (Virtual Authority Files) oder eine Anreicherung

⁴ Eine genaue Auflistung aller Entitäten, Klassen und Relationen findet sich unter <https://linked-swissbib.github.io/datamodel/>.

mit diesen Daten, zum Beispiel mit Daten der DBPedia, d.h. der Linked Data-Version von Wikipedia.

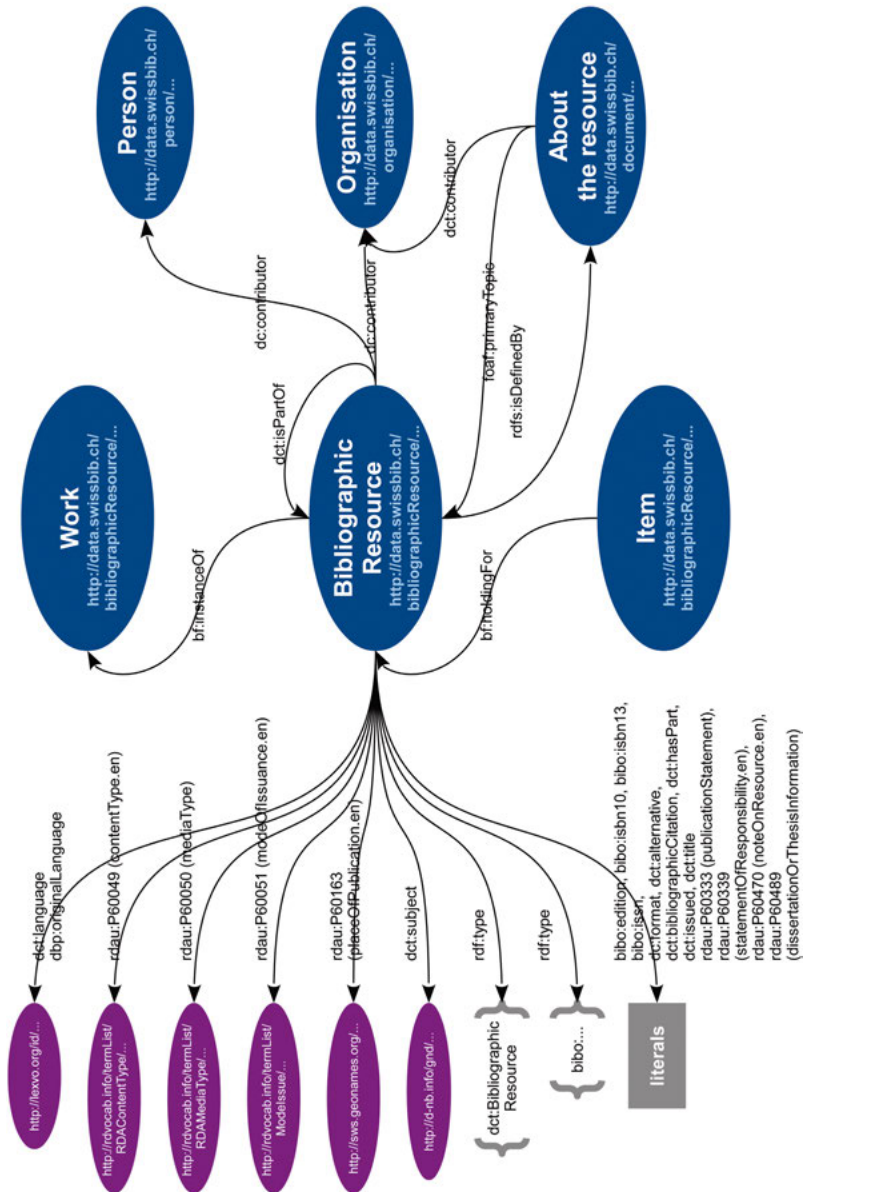


Abb. 2: RDF-Modell für linked.swissbib.ch (Ausschnitt)

Genau diese Verlinkungen und Anreicherungen sollten im Projekt linked.swissbib.ch vorgenommen werden bzw. hat man versucht zu übernehmen, denn einschränkend ist zu sagen, dass diese Operationen und die daraus resultierenden Ergebnisse von der Qualität der damit verbundenen Daten abhängen. Konkret wurde nur mit DBPedia und VIAF verlinkt, Verknüpfungen zur GND werden im Moment von den Verbänden geliefert und teilweise auch in swissbib verwendet. Vor dem Hintergrund der eingeschränkten Laufzeit des Projekts wurde nach recht kurzer Zeit vereinbart, sich bei der Verlinkung und Anreicherung der Daten auf Personenentitäten zu beschränken, auch wenn dies die bekannten Probleme der eindeutigen Zuweisung der Personennamen zum Autor mit sich bringt.

Für eine nicht ambigue, d. h. eindeutige Zuweisung von Identifikatoren für Personen bietet sich das Geburts- bzw. das Todesdatum an. Diese liegen jedoch nur für etwa 10% der Personennamen innerhalb der Katalogdaten der schweizerischen Verbände vor. Daher ist ein solches Unterfangen alles andere als trivial. Nach Beendigung all dieser Operationen wurden sämtliche RDF-Daten in einen Elasticsearch-Index überführt, so dass im aktuellen Zustand der swissbib-Plattform zwei Suchmaschinen komplementär existieren.

Dies wirft die Frage nach den in Abbildung 1 genannten ‚downstreams‘ auf, d. h. den Schnittstellen für die Benutzerin bzw. den Benutzer und die Maschinen, was die Problematik der Lizenzierung der Daten und eine kleine, aber weitreichende Unterscheidung mit sich bringt:

1. Die unter linked.swissbib.ch aufgeschaltete Benutzerschnittstelle ist letztlich eine veränderte Ansicht der schon in swissbib gezeigten Daten (mit dem Unterschied der dahinter liegenden Verlinkung und Anreicherung) und wird aus beiden Suchmaschinen gespeist. Diese neue Ansicht der Daten impliziert keine direkte Nachnutzung durch den Benutzer, insofern stellt sich auch die Frage nach einer Lizenzierung der Katalogdaten unter einer Creative Commons 0-Lizenz nicht.
2. Die unter data.swissbib.ch erreichbare, für Maschinen erstellte RESTful-Schnittstelle (representational state transfer) erlaubt einen direkten Zugang zu den Daten von linked.swissbib.ch im RDF-Format und damit eine (theoretisch gesehen, massenhafte) Nachnutzung der dahinter befindlichen Daten. Aufgrund der Möglichkeit dieser Nachnutzung sind eine a priori durchgeführte eindeutige Lizenzierung der Daten sowie eine Ausfilterung der mit dieser Lizenz nicht kompatiblen Datensätze notwendig.

Aufgrund der Konsequenzen des letztgenannten Punkts wurden während der Laufzeit des Projekts mit sämtlichen Institutionen, die swissbib mit bibliographischen Daten beliefern, Verhandlungen über eine Lizenzierung der Daten unter CC0 geführt. Ausnahmen bildeten Helveticat, der Katalog der Schweizerischen

Nationalbibliothek, sowie RERO, der Westschweizer Bibliotheksverbund, die ihre Daten schon vor Projektbeginn für eine freie Nachnutzung zugänglich gemacht hatten.

Gemeinsam mit einem auf digitales Recht spezialisierten Anwalt wurden Textvorschläge erarbeitet, die die Lizenzierung auf den Webseiten der Verbände und die Markierung der Daten als ‚frei‘ oder ‚unfrei‘ erlauben (vgl. data.rero.ch als beispielgebende Vorgehensweise). Aufgrund dieser ursprünglich nicht vorgesehenen Initiative haben zum Zeitpunkt der Veröffentlichung dieses Artikels sämtliche schweizerischen Verbände (mit Ausnahme eines Unterverbands des IDS) ihre Daten für jegliche Form der Nachnutzung (ob kommerziell oder nicht) zur Verfügung gestellt.

Neben dem Aspekt der Nachhaltigkeit sowie des produktiven Einsatzes sollte in ein Projekt innerhalb des universitären Umfelds auch der Innovationsgedanke bei einzelnen Softwarekomponenten mit einfließen. Beispielhaft wird für die Datenschnittstelle data.swissbib.ch ein Open Source-Framework eingesetzt, welches ein zum Zeitpunkt der Erstellung von linked.swissbib.ch völlig neues Protokoll mit dem Namen Hydra verwendet. Hydra ist Bestandteil des im Semantic Web mittlerweile sehr bekannten Ansatzes Linked Data-Fragment. Dieser wurde im Rahmen der Library Science Talks an der Zentralbibliothek Zürich von seinem Initiator vorgestellt.⁵

Systemarchitektur von linked.swissbib.ch

Flexibilität und die Möglichkeit zur permanenten Anpassung einer Softwarearchitektur sind heute entscheidende Matchkriterien im hoch dynamischen und kompetitiven digitalen Umfeld. Kaum minder wichtig für das Gelingen eines Projekts sind jedoch die Rahmenbedingungen für das Zusammenwirken von Personen in unterschiedlichen Rollen und mit divergierendem Wissenshintergrund (Informationswissenschaftler und Informationswissenschaftlerinnen, Entwickler und Entwicklerinnen von Software sowie Projektmanager und -managerinnen), die noch dazu an unterschiedlichen Orten tätig waren (Genf, Chur, Basel). Auch diesbezüglich konnte die [swissbib](http://swissbib.ch)-Plattform ihre Stärken ausspielen. Da im Rahmen des linked.swissbib.ch-Projekts ausschliesslich Open Source-Komponenten eingesetzt werden, welche mittels dokumentierter und offener Schnittstellen untereinander kommunizieren, konnten diese einfach und schnell auf der heute all-

⁵ Ruben Verborgh: „Linked happily ever after“. Library Science Talk. Genf/Zürich 05.–06. Dezember 2016 (<https://cds.cern.ch/record/2239081>).

gegenwärtigen Github-Plattform kollaborativ entwickelt und ausgetauscht werden. Für die Transformationen in das RDF-Modell wurde die im Culturegraph-Projekt (siehe www.culturegraph.org) entwickelte Metafactory-Software verwendet, die es mit einer relativ leicht erlernbaren Definitionssprache ermöglicht, die Grenzen zwischen Domain und Software-Experten aufzuweichen. Diesen im Projekt gelebten Prozess des Aufeinanderzugehens von Personen in unterschiedlichen Rollen wurde von allen Beteiligten als massgeblich für die Zusammenarbeit im wissenschaftlichen bibliothekarischen Umfeld der Zukunft angesehen.

Eine grobe Skizze der Architektur der swissbib-Plattform nach Integration der linked.swissbib.ch-Komponenten gibt Abbildung 3 wieder:

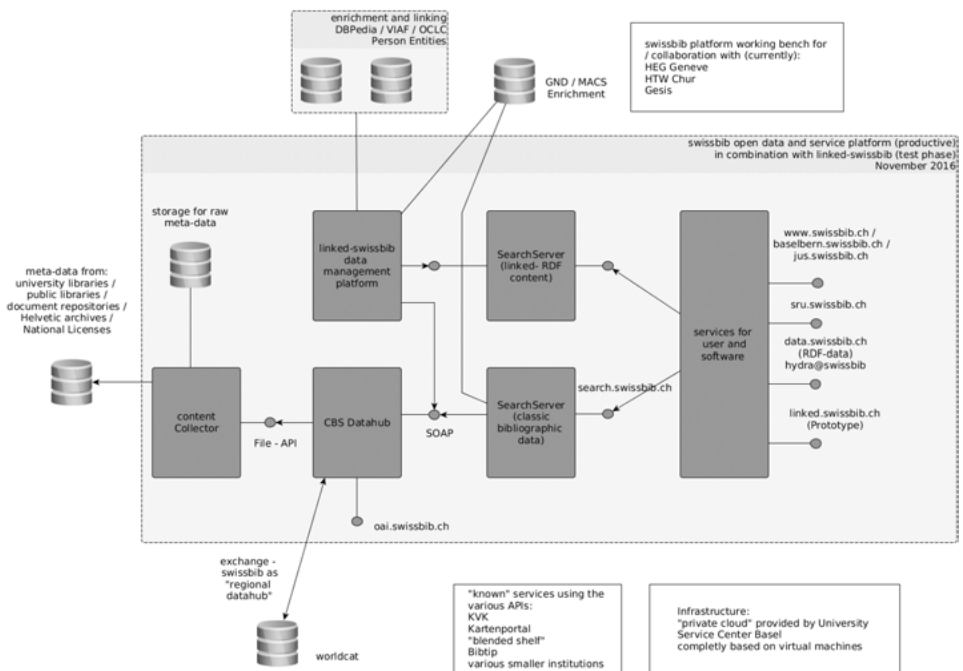


Abb. 3: swissbib-Systemarchitektur nach Integration der linked.swissbib.ch-Artefakte (© Günter Hipler)

Flexibles und mächtiges Data Processing als Grundlage zukünftiger (Benutzer-)Services

Der vorhergehende Abschnitt sollte dem Leser die Vorgehensweise datengetriebener Projekte sowie die Voraussetzungen für ihren Erfolg vermittelt haben. Im

Zentrum steht dabei die Möglichkeit für Institutionen, die Rahmenbedingungen für Datenworkflows sowie deren Instrumente zu steuern und selbst einzusetzen.

Dieser Gedanke durchzieht das swissbib-Projekt und die entwickelte Plattform mit ihren verschiedenen Services. Die digitale Welt gestattet jedoch keinen Stillstand. Gegenwärtig vielleicht noch erfolgreiche Organisationen werden in Zukunft nur konkurrenzfähig sein, wenn sie neue Verfahren und Methoden des digitalen Wandels integrieren. An dieser Stelle muss das heute in aller Munde befindliche Schlagwort ‚Big Data‘ fallen. Ein Data Processing Verfahren wird in Zukunft nur bestehen können, wenn es die neuen Möglichkeiten nutzbar macht. Swissbib geht hier bereits seit zehn Jahren erfolgreich den Weg der Evolution. Die Möglichkeiten der Weiterentwicklung einer flexiblen Plattform, wie sie mit der Durchführung des linked.swissbib.ch-Projekts bewiesen wurden, sollen weiterhin genutzt werden, entweder im Rahmen von swissbib selbst oder durch Nutzung der Artefakte durch Dritte. Wie die technische Architektur im Rahmen des nächsten Entwicklungsschritts aussehen kann, zeigt die nachfolgende Abbildung.

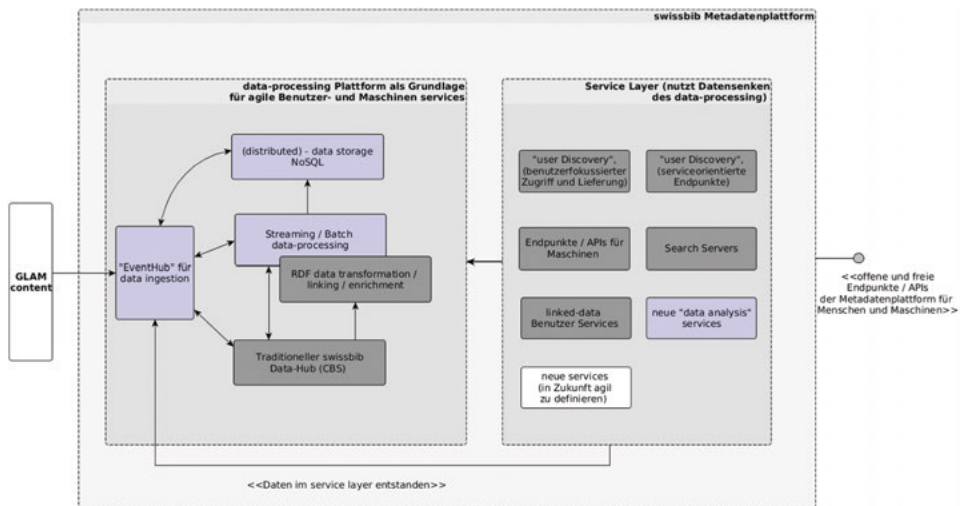


Abb. 4: swissbib Data Processing-Plattform mit Integration von Big Data-Verfahren (© Günter Hipler)

Grau unterlegt sind die bereits vorhandenen Komponenten innerhalb der bestehenden produktiven swissbib-Plattform, blau unterlegt sind neue Verfahren aus der Big Data-Welt, die zu grösseren Teilen bereits als Prototypen auf der swissbib-Plattform vorliegen. Zusammenwirken können diese Teile wie bis anhin über die offenen Schnittstellen. Dies kennzeichnet den evolutionären Charakter des Ansatzes. Stärker zum Ausdruck kommt jetzt die Unterteilung in den Data Proces-

sing-Teil und den Service Layer, der auf den Möglichkeiten der Datenbereitstellung aufbaut. Das datengetriebene linked.swissbib.ch-Projekt nutzt dabei die Möglichkeiten zur Verlinkung, Anreicherung oder Aggregation einer weiterentwickelten Data Processing-Komponente. Die Nutzungsmöglichkeiten semantisch vernetzter Informationen werden menschlichen Benutzern oder Maschinen im Service Layer über verschiedene Kanäle bereitgestellt. Diese Architektur ermöglicht es in optimaler Weise, den aktuell diskutierten Prozess zur Gewinnung relevanter Informationen und neuer Erkenntnisse aus (Meta-)Daten umzusetzen.⁶

⁶ Guido Biland: Wissen verwalten im Zeitalter von Big Data. 29.11.2016 (https://www.switch.ch/de/stories/big_science_data).