

Julia Hauser, Reinhold Heuvelmann

BIBFRAME

Erste Schritte auf dem Weg zu einem neuen Standard

Die bisher bewährten bibliothekarischen Formate für den Datenaustausch können den Anforderungen nicht mehr gerecht werden, die durch die immer stärkere digitale Vernetzung der Informationswelt entstehen. 2011 rief die Library of Congress (LoC) daher die »Bibliographic Framework Transition Initiative« (BIBFRAME) ins Leben mit dem Ziel, die vorhandenen Formate zu einem nachhaltigen Rahmenkonzept weiterzuentwickeln. Dabei sollen diese Formate auf ihre Zukunftsfähigkeit überprüft und die Möglichkeiten berücksichtigt werden, die Semantic-Web-Technologien mit sich bringen.

Die Deutsche Nationalbibliothek (DNB) erklärte im Januar 2012 ihre Unterstützung der BIBFRAME-Initiative¹⁾ und ist seither Mitglied der Early Experimenters Group. Im Zuge der Internationalisierung der deutschen Standards begleitet die DNB also aktiv die Initiative. Erfahrungen im Bereich der Formatumstiege und Konkordanzen sowie der Datenbereitstellung als Linked Open Data²⁾ bilden die Basis für die Mitarbeit.

Ziel der ersten Phase des BIBFRAME-Projektes der DNB ist schwerpunktmäßig die Mitwirkung an den BIBFRAME-Point-Papers sowie der Erkenntnisgewinn aus einer prototypischen Konversion der DNB-Daten in das BIBFRAME-Format.

Die DNB sieht sich im deutschsprachigen Raum als Multiplikatorin und Ansprechpartnerin der Initiative und nutzt bestehende Strukturen und Gremien der Standardisierungsarbeit (z. B. die Expertengruppe Datenformate), um über den aktuellen Stand zu informieren und die bereits vorhandene Expertise in die Initiative mit einzubringen.

Hierbei spielt die Arbeit der DINI AG KIM-Gruppe Titeldaten³⁾ eine wichtige Rolle: Vertreter der Bibliotheksverbände sowie einiger großer Bibliotheken aus dem deutschsprachigen Raum haben

sich zum Ziel gesetzt, eine Harmonisierung der RDF-Repräsentationen von Titeldaten im deutschsprachigen Raum zu erreichen. Im Rahmen der BIBFRAME-Initiative wird nun eine internationale Lösung angestrebt.

Aktuelle Fragen der BIBFRAME-Entwicklung werden in dieser Arbeitsgruppe diskutiert und gegebenenfalls in die Runde der Early Experimenters zurückgemeldet, um die Initiative konstruktiv und kritisch zu begleiten. Hierbei profitiert die Gruppe von den bereits gewonnenen Erfahrungen der Titeldatenmodellierung in RDF und ihrer Vereinheitlichung.

Ausgangssituation: Standards, Regelwerke, Datenformate

Im Laufe der Jahre entstand eine Vielzahl an Regelwerken, Datenformaten und weiteren Standards, um der wachsenden Informationsflut zu begegnen und die Masse an Daten und Informationen nachnutzbar zu halten. Die heutigen technischen Möglichkeiten überholen hierbei oftmals die bestehenden Strukturen. Auch innerhalb der Bibliothekscommunity gehört es zum Tagesgeschäft, mit einer Reihe von Formaten und Standards umzugehen. Über viele Jahre hinweg haben sich Arbeitsabläufe entwickelt, die auf diese historisch gewachsene Struktur zugeschnitten sind. Die Herausforderung liegt heute darin, die Bibliotheksdaten auch für Sparten außerhalb der engen Bibliothekswelt verfügbar zu machen, einen Mehrwert zu schaffen, indem Metadaten unterschiedlichster Bestände zusammengeführt werden, um z. B. weitere Sucheinstiege für die Recherche zu schaffen. Grundlage einer solchen Datenvernetzung ist die Interoperabilität dieser Bestände. Im Folgenden soll ein Überblick über bestehende (bibliothekarische) Standards, Regelwerke und Datenformate gegeben werden, da dies der Ausgangspunkt aller Bemühungen für eine spartenübergreifende interoperable Metadatenumgebung ist.

Harmonisierung der RDF-Repräsentationen von Titeldaten

Nachnutzbarkeit bibliothekarischer Daten für Institutionen außerhalb der Bibliothekswelt

Datenvernetzung und Interoperabilität

Weiterentwicklung bibliothekarischer Austauschformate

Internationalisierung der Regelwerke

Im Jahr 2004 beschloss der Standardisierungsausschuss den Umstieg auf internationale Formate und Regelwerke (MARC 21, AACR2)⁴⁾. Dieser Entschluss basierte auf den Ergebnissen einer durch die Deutsche Forschungsgemeinschaft (DFG) geförderten Studie und leitete u. a. den Umstieg von MAB auf MARC 21 als einheitliches Austauschformat im deutschsprachigen Raum ein und legte den Grundstein für die aktive Mitarbeit am neuen Erschließungsstandard Resource Description and Access (RDA)⁵⁾.

In der DNB wird zwischen dem Internformat (Pica+), dem Katalogisierungsformat (Pica3), den Austauschformaten (MARC 21, bis Juli 2013 auch MAB, RDF) sowie spezifischen Formaten wie ONIX, xMetaDissPlus u. a. unterschieden (siehe Abb.: Datenformate in der DNB). Die einzelnen Formate werden von unterschiedlichen Interessengruppen für eine Vielzahl von Anwendungsfällen genutzt. Die größte Granularität und Flexibilität ist mit dem Internformat gegeben, es dient als Drehscheibe (turntable) – daher werden Konkordanzen in der Regel auf Basis von Pica+ erstellt (siehe z. B. das Feldverzeichnis der Titeldaten der DNB⁶⁾).



Datenformate in der DNB

Das Internformat besteht unabhängig von Entwicklungen im Bereich des Austauschformates und wird entsprechend der lokalen Anforderungen laufend weiterentwickelt. Demnach gilt es, die Konkordanzen und Konversionen zum Austauschformat gegebenenfalls anzupassen. Die Entwicklungen im Bereich der Austauschformate können von der DNB zunächst relativ unberührt beobachtet werden. Wird auf Basis eines Internformats gearbeitet,

hat eine Änderung der Austauschformate für die lokalen Systeme keine unmittelbaren Konsequenzen. Betroffen ist hingegen der Datenaustausch: Hier gilt es, die Daten auf Grundlage des Internformats in das Austauschformat zu bringen, das von anderen Bibliotheken, und idealerweise auch von anderen Kultur- und Gedächtnisinstitutionen, interpretiert werden kann. In früheren Zeiten war dies MAB2, heute werden bibliothekarische Daten in MARC 21 ausgetauscht, und BIBFRAME könnte das Austauschformat werden, das auch außerbibliothekarischen Institutionen, Informationseinrichtungen und kommerziellen Anbietern die Nutzung unserer bibliothekarischen Daten ermöglicht.

Was wird BIBFRAME haben, was MARC 21 nicht hat?

Es stellt sich die Frage, warum MARC 21 den Anforderungen nicht gerecht werden kann, die durch die immer stärker vernetzte Informationswelt entstehen. MARC 21 ist ein in sich geschlossenes System. Für Institutionen und Nutzer außerhalb der MARC-21-Community bleiben die Daten unverständlich und nicht interpretierbar. Das neue Regelwerk RDA soll u. a. den Weg für die Nutzung der Bibliotheksdaten im Kontext des Semantic Web ebnen. Jedoch fehlen einige Strukturen, mithilfe derer die in Zukunft nach RDA erfassten Daten ausgetauscht werden können: So können die FRBR-Ebenen, die von RDA bedient werden, nicht vollständig mittels MARC-21 transportiert werden. Ein weiterer Aspekt ist die Datenbereitstellung als Linked (Open) Data. Ein Mehrwert entsteht durch die Veröffentlichung von Linked Data erst, wenn die hiermit geschaffenen Strukturen tatsächlich genutzt werden, um den Grad der Datenvernetzung zu erhöhen und zu nutzen. Um die Datensilos aufzubrechen und zu vernetzen, ist es neben technischen Voraussetzungen auch erforderlich, entsprechende Kooperationen einzugehen und neue Standards zu finden, die die Interoperabilität zwischen unterschiedlichen Datenquellen steigern bzw. herstellen. Auch dies ist ausdrücklich genanntes Ziel der BIBFRAME-Initiative: Es soll ein Rahmenkonzept geschaffen werden, das von einer Viel-

Trennung zwischen Intern- und Austauschformat

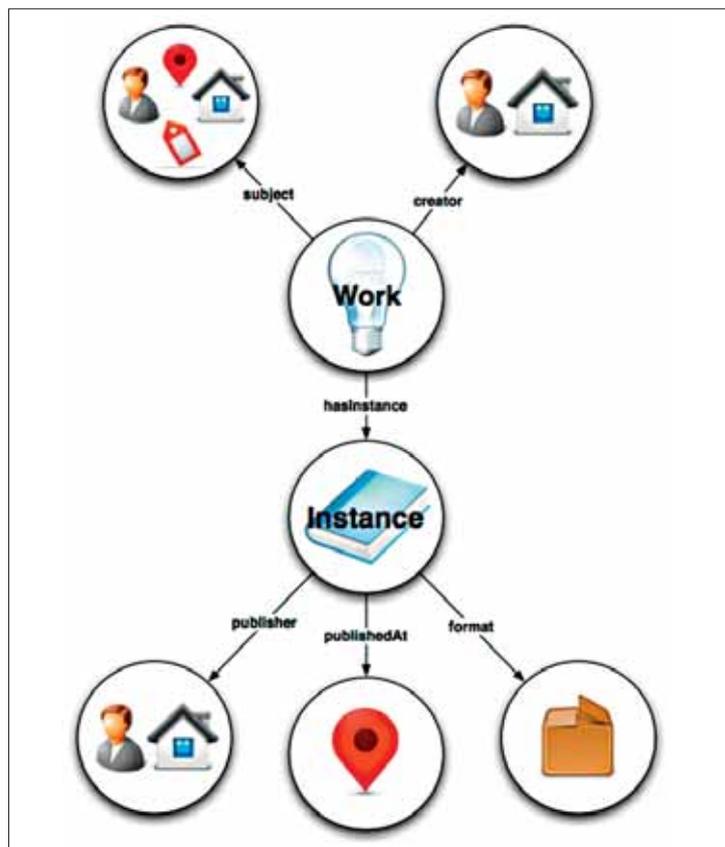
Datennutzung im Semantic Web

zahl heterogener Datenquellen verwendet und somit von Datenbereitstellern unterschiedlicher Sparten genutzt und interpretiert werden kann.

Das BIBFRAME-Modell

Im November 2012 veröffentlichte die LoC in Zusammenarbeit mit der Firma Zepheira das Arbeitspapier »Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services«⁷⁾, das einen ersten Stand des grundlegenden BIBFRAME-Modells beschreibt. Es geht zunächst darum, das MARC-21-Format zu dekonstruieren, um zu analysieren, welche Informationen hiermit transportiert werden, welche unter Umständen verzichtbar sind und an welchen Stellen die Daten in anderer Weise aufbereitet sein müssen, um sie im Kontext des Semantic Web optimal nutzen und auswerten zu können. Das BIBFRAME-Modell sieht vier Kernklassen vor: (Creative) Work, Instance, Authority und Annotation (siehe Abb.⁸⁾).

Dekonstruktion und Analyse des MARC-21-Formats



Das BIBFRAME-Modell

Eine 1:1-Entsprechung zu den FRBR-Entitätentypen Work, Expression, Manifestation und Item ist in dem Modell nicht vorgesehen. Eine Abbildung wird dennoch über Application Profiles möglich sein.⁹⁾ Somit ist auch der Austausch von auf Grundlage der RDA erfassten Daten im BIBFRAME-Modell vorgesehen.

Das BIBFRAME-Vokabular¹⁰⁾ definiert RDF-Elemente, welche BIBFRAME-Klassen zueinander relationieren. Wenn möglich, werden – den Linked Data Prinzipien¹¹⁾ entsprechend – sämtliche Ressourcen, Eigenschaften und Individuen als URIs angegeben.

Das BIBFRAME-Projekt der DNB

Neben der ständigen Beobachtung der Listendiskussionen und der bereits erwähnten Einbeziehung der im deutschsprachigen Raum etablierten Gremien stehen folgende Punkte derzeit im Fokus des DNB-BIBFRAME-Projektes:

– Point Papers

In Positionspapieren werden unterschiedliche Fragestellungen erörtert (sogenannte Point Papers). Bisher veröffentlichte Point Papers finden sich auf der BIBFRAME-Webseite¹²⁾. DNB unterstützt diesen Ansatz, indem eigene Point Papers verfasst sowie hausintern die Papiere besprochen und gebündelt Rückmeldungen in die Early Experimenters Group gegeben werden, die dann teilweise in spätere Versionen der Point Papers einfließen. Auf diese Weise nimmt die DNB eine wesentliche Rolle in den Diskussionen und Entwicklungen wahr.

Rückmeldungen an die Early Experimenters Group

– Prototypische Umsetzung von Pica+ nach BIBFRAME

Wie zuvor erwähnt bringt die Stabilität des DNB-Internformats Flexibilität und Unabhängigkeit von anderen Systemen und Formaten mit sich. Dies ermöglicht, die Daten granular und den Anforderungen entsprechend abzubilden. Daher wird im BIBFRAME-Projekt der DNB ein Mapping des DNB-Internformats auf das BIBFRAME-Modell entwickelt, um Möglichkeiten und Grenzen des neuen Ansatzes auszuloten und in Zukunft die DNB-Daten in BIBFRAME darstellen und zur Verfügung stellen zu können. Das konzeptuelle und

Mapping des DNB-Internformats auf das BIBFRAME-Modell

technische Mapping erfolgt iterativ – das bedeutet, Beobachtungen bezüglich des aktuellen Vokabulars sowie seine Anwendung und Anforderungsspezifikationen werden in mehreren Schleifen in die Early Experimenters Group zurückgespielt, sodass eine schrittweise Annäherung an den angestrebten Standard erfolgt.

Bei dem ersten Entwurf eines Pica+-BIBFRAME-Mappings wurde versucht, die Elemente, die bereits in der DNB-Linked Data-Repräsentation vorhanden sind, auf das BIBFRAME-Vokabular abzubilden. Dabei wurde laufend dokumentiert, an welchen Stellen eine Abbildung nicht 1:1 bzw. nicht ohne Informationsverlust möglich ist, welche weiteren Anforderungen an die Spezifikation des Vokabulars gestellt werden und wo Unklarheiten bezüglich der Umsetzung bestehen.

Die Beobachtungen wurden in folgende Bereiche unterteilt:

- Anforderungen an die DNB-Daten,
- Vorschläge und Erweiterungswünsche für die Webseite bibframe.org,
- Fragen / Anmerkungen zum Vokabular,
- Fragen / Anmerkungen zur Modellierung,
- Allgemeine Fragen / Anmerkungen.

Um die DNB-Daten im BIBFRAME-Format abbilden zu können, ist eine Ausweisung der Werkebene erforderlich. Mit Ausnahme der DMA-EST-Daten als Teil der Gemeinsamen Normdatei¹³⁾ existieren in den DNB-Daten derzeit keine Datensätze für die Werkebene. Das BIBFRAME-Modell setzt diese jedoch mit seiner Core-Klasse »Work« voraus. Für die erste prototypische Umsetzung wurde daher die Übergangslösung gewählt, zu jeder

Instance einen temporären Werksatz zu generieren. Die werkbezogenen Daten werden hierbei aus dem Datensatz extrahiert.

Im Zuge der Einführung der RDA muss ebenfalls eine Lösung für die Abbildung der Werkebene gefunden werden. Der hier vorgestellte Ansatz entspricht dem Vorgehen im RDA-Kontext. Zu beachten ist jedoch, dass ein BIBFRAME-Work nicht exakt dem FRBR-Work gleichzustellen ist: Nach BIBFRAME ist ein Work »a resource reflecting a conceptual essence of the cataloging resource«, es umfasst nach dieser Definition FRBR-Werke und FRBR-Expressionen. Die eng verwandte Terminologie könnte an dieser Stelle eine andere Erwartungshaltung wecken.

Das BIBFRAME-Vokabular basiert bisher stark auf der Feld- und Unterfeldstruktur des MARC-Formats. Der Aspekt der Abwärtskompatibilität zum MARC-Format spielt derzeit bei der Entwicklung des BIBFRAME-Vokabulars eine zentrale Rolle. Andere Formate sowie eine formatunabhängige Betrachtungsweise sind bisher unterrepräsentiert. Für eine Übergangszeit ist dieser Akzent gewiss sinnvoll und für zeitnahe Ergebnisse pragmatisch, sollte allerdings um die Anforderungen zukünftiger Informationsstrukturen ergänzt werden. Mit einer Abbildung von MODS (Metadata Object Description Schema)¹⁴⁾ auf BIBFRAME zeichnet sich ein erster Schritt in diese Richtung ab. Im Vergleich zur derzeitigen Umsetzung der DNB-Daten nach RDF bedeutete die Umsetzung des BIBFRAME-Vokabulars auf dem aktuellen Stand jedenfalls einen Rückschritt bezüglich einer breiten Ressourcenreferenzierung.

Ein weiterer wichtiger Aspekt ist die Interoperabilität des BIBFRAME-Vokabulars mit bestehenden und etablierten Vokabularen. Wie eingangs erwähnt, kann ein Mehrwert bei der Bereitstellung der Daten als Linked Data nur dann entstehen, wenn die Voraussetzung für eine solide Vernetzung gegeben ist.

Die Entwickler des BIBFRAME-Modells sprechen sich ausdrücklich gegen die Nachnutzung existierender Vokabulare aus, da ihre Nachnutzung einen hohen Pflegeaufwand seitens der Datenbezieher mit sich bringe (siehe FAQ¹⁵⁾). Außerdem solle eine nachhaltige Infrastruktur für die kommenden

Abbildung der Werkebene

Loslösung von bestehenden Formaten und Erhaltung der Abwärtskompatibilität

Beispiel:

```
<rdf>
  <bf:Instance about="http://d-nb.info/123456789">
    <bf:instanceOf resource="http://
      d-nb.info/_bf_temp/work_123456789">
  </bf:Instance>

  <bf:Work about="http://d-nb.info/_bf_temp/
    work_123456789">
    <bf:hasInstance resource="http://d-nb.info/
      123456789/>
  </bf:Work>
</rdf>
```

Alignments zu bestehenden Ontologien sind notwendig

Jahrzehnte geschaffen werden, und die LoC wolle sich ihre Unabhängigkeit bewahren und selbst ein persistentes Vokabular bereitstellen. Diese Argumentation ist nachvollziehbar. Umso dringender sollte jedoch auf die Einführung von Alignments zu bestehenden Ontologien insistiert werden. Im ersten Entwurf des BIBFRAME-Modells wird nicht auf das Thema Vocabulary Alignment eingegangen. Des Weiteren sollten bibliotheksferne Sparten möglichst zeitnah in die Entwicklung einbezogen werden, um ihre Anforderungen frühzeitig mit einfließen zu lassen. Mit Ausnahme der DNB kommen außerdem alle bisherigen Mitglieder der Early Experimenters Group aus dem angloamerikanischen Sprachraum. Um die Internationalität und Sprachunabhängigkeit der BIBFRAME-Initiative zu stärken, sollten weitere Partner aus Ländern einbezogen werden, deren Landessprache nicht oder nicht ausschließlich Englisch ist.

Ausblick

Das BIBFRAME-Format befindet sich in der Findungsphase, in der Mängel erlaubt, Kritik erwünscht und spezifische Anforderungen an das Modell erforderlich sind. Weiterhin ist geplant und notwendig, zu untersuchen, inwiefern die nach

RDA erschlossenen Daten in BIBFRAME abgebildet werden können.

Konkret sollen bis Ende 2013 die von Pica+ prototypisch konvertierten DNB-Daten in BIBFRAME im Linked Data Service der DNB testweise zur Verfügung gestellt werden. Somit besteht die Möglichkeit, sich die konkreten Daten direkt anzusehen und mit bestehenden Repräsentationen zu vergleichen. Auf diese Weise bleibt der Evaluationsprozess transparent, und die (nicht nur) deutschsprachige Bibliothekscommunity ist eingeladen, Zwischenergebnisse zu kommentieren, Anregungen einzubringen und weitere Aspekte aufzuwerfen, die seitens der DNB wiederum an die BIBFRAME-Early Experimenters Group gebündelt zurückgegeben werden können. Damit soll gewährleistet werden, dass die Entwicklung auf einer vielfältigen Expertise basiert und in kommenden Iterationsschritten konstruktive Kritik umgesetzt werden kann.

Mit der Bereitstellung der nach BIBFRAME konvertierten DNB-Daten wird die erste Phase des DNB-BIBFRAME-Projektes abgeschlossen sein. Die DNB wird sich auch weiterhin an der Entwicklung beteiligen, aktiv in der Early Experimenters Group mitwirken und hierbei flexibel auf aktuelle Entwicklungen reagieren. Weitere Schritte auf dem Weg zu dem neuen Standard BIBFRAME werden folgen.

Abschluss der ersten Projektphase: DNB-Daten im BIBFRAME-Format

Anmerkungen

- 1 <http://files.dnb.de/pdf/bfti_statement_dnb.pdf>
- 2 <<http://www.dnb.de/lds>>
- 3 <<https://wiki.dnb.de/display/DINLAGKIM/Titeldaten+Gruppe>>
- 4 <<http://www.dnb.de/DE/Standardisierung/International/afsProjekt.html>>
- 5 <<http://www.dnb.de/DE/Standardisierung/International/rda.html>>
- 6 <https://wiki.dnb.de/download/attachments/43090988/titeldaten_dnb.pdf?version=2&modificationDate=1369848387000>
- 7 <<http://www.loc.gov/bibframe/pdf/marclD-report-11-21-2012.pdf>>
- 8 <<http://bibframe.org/static/images/bibframe.png>>
- 9 <<http://de.slideshare.net/zepheiraorg/alabibframe-lc20130630>>
- 10 <<http://bibframe.org/vocab/>>
- 11 <<http://www.w3.org/DesignIssues/LinkedData.html>>
- 12 <<http://bibframe.org/documentation/>>
- 13 <<http://www.dnb.de/gnd>>
- 14 <<http://www.loc.gov/standards/mods/>>
- 15 <<http://bibframe.org/faq>>