

Linked Data für die Verwaltung von Forschungswissen in gemeinschaftlichen Domains von Bibliotheken, Archiven und Museen Zusammenfassung

Bibliotheken, Archive und Museen (Library, archive and museum, kurz: LAM) besitzen eine beträchtliche Anzahl von Daten und Ressourcen, die bisher oft in Katalogen und Archiven versteckt und daher nicht nutzbar waren. Die Aktivierung dieses Potentials und die Öffnung dieser bisher nicht sichtbaren Informationen für einen weiteren Nutzerkreis würde das World Wide Web bereichern, eine offenere Wissenskultur fördern und jedem einzelnen Glied der Informationskette zahlreiche Vorteile bringen. Bibliotheken, Archive und Museen hätten die wertvolle Möglichkeit, Brücken zwischen Sammlungen zu bauen und dadurch den Nutzern umfassendere Mittel zur Verfügung zu stellen oder vielmehr eine vorher nicht vorhandene Datenfülle zu bieten. Herausgeben von Daten erschließen sich hingegen neue Formen der Zusammenarbeit. Auf diese Art und Weise kann das weltweite Kulturerbe für zukünftige Generationen weitergegeben und gesichert werden, Forschung kann in den Originalsprachen verwahrt werden und die kulturelle Vielfalt und Energie bleibt erhalten. Das Potenzial für den kompletten geisteswissenschaftlichen Sektor ist beachtlich, gerade für die vielen Bereiche und Nischen, denen jetzt noch stärker als bisher die Marginalisierung droht.

Linked Data ist der Schlüssel zur Realisierung dieser Ziele.

Dieser neue Ansatz für die Daten- und Informationsverarbeitung legt den Grundstein für weitere Verbesserungsmöglichkeiten, wie einzelne Institutionen untereinander zusammenarbeiten können, und ebenso für die Wiederverwendung von Daten in den verschiedensten Kontexten für eine effizientere Erkennung und web-basierende Identifizierung von Entitäten. Gleichzeitig fördert der neue Ansatz den aktiven Nutzen von Kulturerbe-Sammlungen und lädt dazu ein, Daten zu verbreiten und zu teilen. Die Anwendung der Paradigmen von Linked Data eröffnet Bibliotheken, Archiven, Museen und allen im Informationsmedienbereich Tätigen umfassendere Mittel, um an Wissen zu gelangen. Diese Tools werden durch den Zusammenschluss von Institutionen in gemeinsamen Workspaces unabhängig weiterentwickelt und können so auch mit der Webtechnologie Schritt halten.

Gleichzeitig anpassungsfähig und trotzdem nachhaltig erlaubt dieser gemeinschaftliche Ansatz die Schaffung eines wohlstrukturierten Systems der Datenorganisation, das verschiedene kulturelle Traditionen respektiert und technologische Möglichkeiten nutzt. Schlüsselfiguren in diesem Sektor flankieren kleinere Institutionen, die bereit sind, ihre eigenen Erfahrungen und einzigartigen Vermögenswerte zu teilen.

Um die Bedingungen dieser neuen offenen Philosophie der Teilung von Daten und deren Weiterverwendung zu erfüllen, muss auch das Konzept der traditionellen Normdatensätze weiterentwickelt werden. Frühere Diskussionen, ob Normdaten zentral oder lokal erstellt werden sollen, werden sich verändern, wenn sich der Fokus von einem steifen Konzept der Normdaten zu einer flexibleren Idee der Identifizierung von Entitäten und dem Verhältnis der Entitäten untereinander bewegt. Organisationen wie Bibliotheken, Archive, Museen, Verlage und Provider werden sich bei der Generierung der neuen Daten und der Entwicklung neuer Ressourcen miteinander austauschen und die zwischen ihnen liegenden Grenzen überschreiten, um Möglichkeiten der Datenanreicherung zu schaffen, wie sie bisher nie vorstellbar gewesen wären.

Das Datenmodell BIBFRAME ist ebenfalls im Moment Gegenstand vieler Diskussionen und Entwicklungen innerhalb der Bibliotheksgemeinschaft und darüber hinaus. Die Absicht ist, die Möglichkeiten von Linked Data für Bibliotheken, Archive und Museen zu öffnen, um eine größere Kompatibilität, Sichtbarkeit und Auffindbarkeit von Ressourcen jeder Art zu ermöglichen, unter Berücksichtigung der sich entwickelnden Richtlinien und Datenmodelle wie das IFLA Library Reference Model (LRM) und die Resource Description and Access-Richtlinien (RDA).

Das Ziel der einzelnen Komponenten und Tools, die im Rahmen der auf Linked Data basierenden Share-Initiativen Share-VDE (Virtual Discovery Environment), Share-ART (mit den Bibliotheken der Max-Planck-Institute für Kunstgeschichte) und Share-MUSIC (ein experimentelles Projekt im Musiksektor) entwickelt werden, ist es, einen Rahmen zu schaffen, der der Verwaltung von Forschungswissen dient. Fortschrittliche Discovery-Schnittstellen verbessern die Benutzererfahrung und liefern Nutzern von Bibliotheken, Archiven und Museen umfangreichere Suchergebnisse. Bibliothekare, Archivare und Museumskuratoren hingegen werden Verbesserungen bemerken, angefangen bei der Weiterentwicklung der Katalogisierungsfunktionen, die semantischen Webstandards entsprechen, bis hin zur Prozessintegration in lokale Systeme und Tools für die Implementierung in einem gemeinsamen Kontext. Die Absicht ist, die Fülle an Daten in bereits bestehenden Sammlungen zu enthüllen, die in traditionellen Katalogen oft versteckt bleiben.

Im Rahmen der Initiative werden Katalogdaten von Bibliotheken, Archiven und Museen durch die Identifizierung von Entitäten, Datenabgleich und -anreicherung in Linked Data umgewandelt. Attribute werden verwendet, um Personen, Werke oder andere Entitäten eindeutig identifizieren zu können, mit möglichen Ausdrucksvariationen, die in einem Datencluster alles vereinen, was sich auf die gleiche Entität bezieht. Diese werden dann mit traditionellen Datensätzen und anderen externen Quellen abgeglichen, um die Datensätze anzureichern und ein Netzwerk von Informationen und Ressourcen zu bilden. Es wird eine Datenbank von Verbindungen geschaffen, die der gesamten Gemeinschaft offen steht, genauso wie eine gemeinsame Wissensbasis von Clustern, die in RDF (Resource Description Framework) zugänglich ist. Diese Datenbank benutzt das Modell des Semantic

Web, ermöglicht es aber den teilnehmenden Bibliotheken, ihre eigenen Daten so frei und unabhängig wie möglich zu verwenden.

Zu den primären Tätigkeitsbereichen zählen:

- Anreicherung von MARC-Datensätzen mit URI Identifikatoren;
- Umwandlung von MARC-Daten in RDF unter Verwendung des BIBFRAME Vokabulars (und gegebenenfalls anderer zusätzlicher Bezugssysteme);
- Entwicklung einer Discovery-Plattform, die sich auf die Anpassung des BIBFRAME-Modells stützt, mit dem Ziel, einen Discovery service in Linked Data zur Verfügung zu stellen;
- Konzipierung einer Datenbank von Verbindungen und Informationsclustern (Sapientia Cluster Knowledge Base), die in RDF zugänglich ist;
- Implementierung von Komponenten für die direkte, interaktive Arbeit mit Daten, die eine Validierung, Aktualisierung, langfristige Kontrolle und die Instandhaltung der Cluster und Identifikatoren der Entitäten gestattet;
- automatische Prozeduren der Datenaktualisierung oder Datenbatch;
- Verbreitung in batch/automatisch von Daten an Bibliotheken;
- progressive Implementierung von weiteren Arbeitsabläufen gemäß der Ordnung, wie sie von der Gemeinschaft festgelegt wurde, unter anderem: copy cataloguing, original cataloguing, API (Application Programming Interface) für URI, rückwirkende Umwandlung für lokale Erwerbungs- und Verwaltungssysteme basierend auf MARC, Reportistik.

Die teilnehmenden Institutionen werden dazu ermutigt, mitzuarbeiten, damit Entscheidungen über die späteren Schritte, die zu einer vollständigen Testanwendung in den Bibliotheks-, Archiv- und Museumsgemeinschaften führen, auf soliden Grundlagen aufbauen. Diese Datensätze sind dann nicht einfach nur in RDF konvertiert, sondern auch mit Identifikatoren und Verknüpfungen angereichert, eine Voraussetzung für die Nutzung im Linked Data-Umfeld. Diese Prozesse können später dazu beitragen, ein realistisches Modell der Nutzeraktivitäten zu erstellen, wodurch eventuelle Probleme wie auch potentielle Vorteile der Umstellung in das Linked Data-Umfeld angesprochen werden. Funktionalität ist bestimmend Share-VDE, die verschiedenen Umgebungen und Nutzerschnittstellen für die Datenerstellung, Anreicherung und Bereitstellung von Workflows für die verschiedenen Gruppen der teilnehmenden Bibliothekare, Experten, Wissenschaftler, Forscher und Studenten umfassen ein breites Spektrum von Anforderungen.

Casalini Libri (www.casalini.it) ist eine bibliografische Agentur, die Normdaten und bibliografische Datensätze erstellt, ein Lieferant von Büchern und Zeitschriften für Bibliotheken mit zusätzlichen Diensten für Bestandsaufbau und Buchbearbeitung, sowie ein Anbieter von elektronischen Ressourcen in Zusammenarbeit mit Bibliotheken und Verlagen.

@Cult (www.atcult.it) ist ein Softwareentwicklungsunternehmen spezialisiert auf Datenkonvertierung für Linked Data wie ein Anbieter von Integrierten Bibliothekssystemen und Discovery-Tools, die effektive und innovative technologische Lösungen zur Verbesserung von Informationsmanagement und Wissensaustausch liefern.