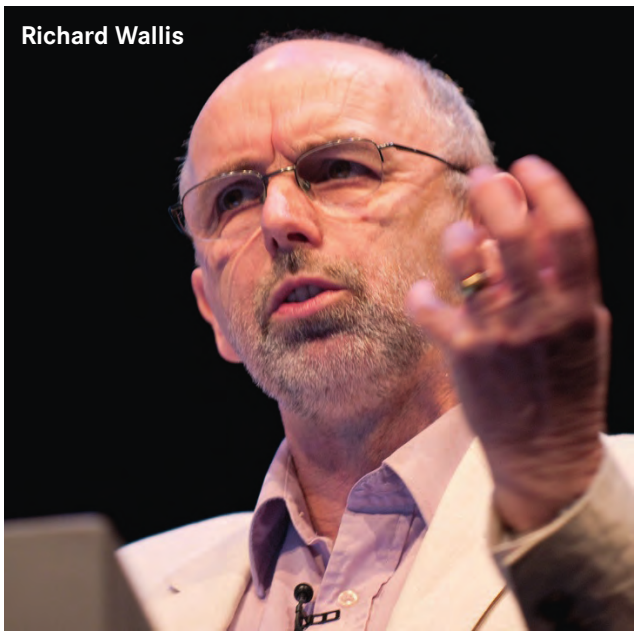
Was Linked Data für Bibliotheken bedeuten – und was man damit machen kann

Richard Wallis und **John O'Donovan**, Hauptredner des nächsten OCLC EMEA Regional Council Annual Meeting, über die Macht und die Möglichkeiten dynamisch geteilter Daten und Informationen.

Richard Wallis, Technologieevangelist und Data Exposure Program Manager der Bibliotheksorganisation OCLC sowie Gründer von Data Liberate[1] und John O'Donovan, Direktor für technische Architektur und Entwicklung bei der größten englischen Multimedia-Nachrichtenagentur „The Press Association"[2] sind fest überzeugt: Wenn Bibliotheken im Netz der Zukunft sichtbar bleiben wollen, müssen sie sich aus ihrer eigenen, gut organisierten und gut vernetzten Datenwelt heraus in den natürlichen Suchraum ihrer Kunden bewegen – in das durch Linked Data und Linked Open Data (LOD) entstehende „Wider World Wide Web". Was die beiden Experten zu ihrer Einschätzung veranlasst, werden sie als Hauptredner beim nächsten OCLC EMEA Regional Council Annual Meeting am 26. und 27. Februar 2013 in Straßburg erklären, auf dem auch der Biophysiker Erez Lieberman Aiden und der Datenforscher Jean-Baptist Michel, beide Harvard University, über die Kraft geteilter Daten, sogenannter Shared Data, sprechen werden. OCLC ist derzeit dabei, die Datensätze im Bibliothekskatalog WorldCat so aufzubereiten, dass die Linked-Data-lesenden Suchmaschinen der Neuzeit die dort nachgewiesenen Ressourcen finden können[3]. Seit Juni 2012 laufen bei OCLC Experimente mit internationaler Beteiligung[4]. „The Press Association" hat unter Anwendung der neuesten Technik der Datenzusammenführung aus vielen verschiedenen Quellen u.a. die offizielle Webseite der Olympischen Spiele 2012 in London[5] gebaut, dynamisch aktualisiert und in 26 Sprachen bereitgestellt.

Im Interview mit B.I.T.online sprachen Richard Wallis und John O'Donovan über die Macht und die Möglichkeiten dynamisch geteilter Daten und Informationen. Die Fragen stellte Vera Münch.


Richard Wallis

„Linked Data befähigen das Internet, Informationen so vorzuschlagen, wie auch ein Mensch sie assoziieren würde." Richard Wallis

Mr. Wallis, Sie sind Technologieevangelist von OCLC Research in EMEA. Was ist Ihre Aufgabe in dieser Position?

Richard Wallis Als Technology Evangelist ist es meine Rolle, neue Technologien voranzubringen und Menschen zu helfen, die neue Technologien zu verstehen und zu erklären, wie sie diese nutzbringend in ihrer täglichen Arbeit anwenden können.

Dürfen wir das in Ihrem Vortrag in Straßburg erwarten?

Richard Wallis *(lacht)*: Sie erwarten jetzt nicht im Ernst von mir, dass ich Ihnen heute schon sagen kann, worüber ich in Straßburg sprechen werde? Straßburg ist Ende Februar. Bei der Weiterentwicklung des Webs, der Anwendung der Linked Data Technologie bewegen sich die Dinge sehr schnell vorwärts. Es ist schwer vorherzusagen, was als nächstes passiert und wann es passieren wird. Vieles, was gerade vor sich

1 http://dataliberate.com/
2 http://www.pressassociation.com/
3 http://www.oclc.org/research/activities/linkeddata.html
4 http://www.worldcat.org/title/harry-potter-and-the-deathly-hallows/oclc/155131850
5 http://www.london2012.com/

geht, hat es in der Geschichte der Welt zuvor noch nicht gegeben.

Aber ja – natürlich werden verlinkte Daten, dynamische Daten und geteilte Daten das Thema meines Vortrags sein. Im Wesentlichen werde ich eine Übersicht über die Linked Data Landschaft geben, berichten, was aktuell in der Welt vor sich geht, dann hinein zoomen in einzelne wichtige Fragen und natürlich auch auf die Aktivitäten von OCLC im Bereich Linked Data eingehen. Und ich werde auch ein wenig in die Glaskugel blicken und versuchen aufzuzeigen, welche Herausforderungen sich für die Zukunft der Bibliotheksgemeinschaft aus Linked Data ergeben.

Dürfen wir Sie bitten, für uns Linked Data, Linked Open Data und Big Data zu definieren?

Richard Wallis Linked Data sind eine Reihe von Grundsätzen; Prinzipien, die Sir Tim Berners-Lee[6] 2006 vorgestellt hat. Sie beschreiben, wie man ‚Web Standards'[7] benutzt, um im Internet publizierte Daten so zu klassifizieren, dass andere[8] sie verstehen können. Diese Regeln lauten verkürzt: 1. Verwende URIs zur Bezeichnung von Dingen. 2. Verwende HTTP URIs, damit jeder nach dem Gegenstand suchen kann. 3. Wenn sich jemand den Gegenstand ansieht, stelle weitere nützliche Informationen zu diesem Gegenstand bereit, indem Du sie per RDF beschreibst. 4. Baue Links zu weiteren URIs ein, so dass der Informationssuchende mehr über den Gegenstand seines Interesses erfahren kann. Die Schlüsselelemente der Linked Data Technik sind a) die URI (Uniform Resource Identifier) als eindeutige Identifikationsmarke für Dinge. b) Das Datenformat RDF (Resource Description Framework), durch das man die Gegenstände so beschreiben kann, dass sie für das erweiterte World Wide Web lesbar und tatsächlich vom Inhalt her gesehen sinnvoll verknüpfbar werden und c) darauf aufbauende weitere Verweise (Links). Linked Data befähigen das Internet, Informationen so vorzuschlagen, wie auch ein Mensch sie assoziieren würde.

(Anm. der Redaktion: Bei den unter c) genannten ‚Links' handelt es sich nicht um Vorschläge von Empfehlungssystemen (Recommender) zum vermeintlichen Interesse des Nutzers, sondern um bidirektionale Verweise auf tatsächlich zum aufgerufenen Gegenstand gehörende weitere Informationen.)

Warum sollten sich Bibliotheken diese ganze Arbeit machen?

Richard Wallis Es geht bei Linked Data immer um die eindeutige individuelle Identifikation von Dingen im Internet – Gegenständen, Informationen, Musikstücken, Büchern in jeder Form, Dokumenten ... Darum, genau dieses eine Ding in seiner elektronischen Repräsentation im Internet so exakt zu bezeichnen, dass es eindeutig identifizierbar ist und man es darauf aufbauend mit wirklich nützlichen weiteren Informationen verbinden kann; Hinweise, die einen echten Mehrwert zu eben diesem Ding liefern, beispielsweise zu einem Buch die Angabe, wo die nächstgelegene Bibliothek ist, in der man es ausleihen kann, ob sie geöffnet hat, ob das Buch da ist oder verliehen usw. Das ist alles noch viel Zukunftsmusik und in Bezug auf die Linked Data aus dem OCLC WorldCat erst im frühen Experimentierstadium[9]. Fest steht: Jeder Gegenstand hat Attribute, die ihn unverwechselbar und durch gute Metadaten unverwechselbar beschreibbar machen. Wir haben jetzt die Technik dafür, diese Informationen aus den Zettelkästen und den Katalogen der Bibliotheken in die Datenwelt des Internets zu übertragen. Die Gemeinschaft der Bibliothekarinnen und Bibliothekare ist wahrscheinlich die Community der Welt, die das Beschreiben anhand von Metadaten am besten kann.

Klassifizieren, Indexieren und Katalogisieren sind ja wesentliche Teile des bibliothekarischen Berufsbildes. Das können Bibliothekarinnen und Bibliothekare ganz sicher. Aber noch einmal zurück zur Informationstechnik, weil das eine sehr wichtige Frage ist: Wie unterscheiden sich Linked Data von Linked Open Data?

Richard Wallis Aus technischer Sicht gibt es keinen Unterschied. Praktisch macht ihn die Lizenz aus, unter der die Daten im Internet publiziert werden: Wenn das eine Open Data Lizenz ist und die Daten offen im Web zur Verfügung gestellt werden, sind es Linked Data. Es ist Aufgabe eines Jeden, der Informationen im Internet als Linked Data veröffentlicht, die Lizenzkonditionen mit zu veröffentlichen. OCLC hat Rahmenrichtlinien[10] für die Zulieferung von Linked Data für den WorldCat erarbeitet. Sie stehen auf der Webseite. Die LOD-Metadatensätze aus dem WorldCat stellt OCLC unter der Open Data Commons-Lizenz ODC-BY 1.0[11] der Open Knowledge Foundation[12] im Web bereit.

6 http://en.wikipedia.org/wiki/Tim_Berners-Lee - http://de.wikipedia.org/wiki/Tim_Berners-Lee

7 Das Wort Standards wird in der Informatik und im Internet als Synonym für „Normierung" benutzt.

8 Mit „andere" sind sowohl Suchmaschinen, als auch Menschen gemeint, die den „Katalogeintrag" interpretieren und weiteren thematisch dazu gehörenden Informationen zuordnen.

9 Siehe dazu Kasten „OCLC Experimente für Jedermann" in diesem Bericht.

10 http://www.oclc.org/data/attribution.html

11 http://opendatacommons.org/licenses/by/summary/

12 http://okfn.org/

Wo kommen Big Data ins Spiel?

Richard Wallis Big Data ist mehr ein Marketingbegriff als ein technischer Terminus. Big Data hat es immer schon gegeben und was Big Data für die eine Organisation ist, kann für die andere etwas ganz anderes sein. Die Größe von Datenbeständen wird von drei V's bestimmt – Volume, Variety, Velocity (Volumen, Vielfalt, Geschwindigkeit). Linked Data Techniken und Werkzeuge zum Umgang mit Big Data unterstützen sich gegenseitig: Big Data Tools helfen den Anbietern von Linked Data, ihre Datenbestände in kurzer Zeit für die Veröffentlichung aufzubereiten. Linked Data Technik kann oft bei der Beschreibung und Kopplung von Big Data helfen.

Um eine Größenordnung zu nennen: Mit dem neuen Big Data Werkzeug Hadoop[13] können die rund zwei Milliarden Bibliotheksbestandsnachweise des WorldCat in ungefähr 20 Minuten abgearbeitet werden. Das dauerte früher zwei Wochen. Das gibt uns die Möglichkeit, viel mehr auszuprobieren. Wenn etwas nicht klappt, lässt man es einfach noch einmal durchrechnen.

„Die Bibliotheken haben keine Wahl: Wenn sie für die Gesellschaft relevant bleiben wollen, müssen sie sich den Einflüssen aus dem Wider Web aktiv stellen."

Vielen Dank für die Erklärungen. Aber warum sollten sich Bibliotheken mit diesen neuen Datentechnikfragen auseinandersetzen? Bibliotheken haben seit vielen Jahren ein sehr gut funktionierendes, weltweites Netzwerk.

Richard Wallis Linked Data ist sehr viel größer als die internationale Bibliotheksgemeinschaft. Wir setzen Linked Data ein, um die Welt zu vernetzen. In bestimmten Domänen wie in der Öffentlichen Verwaltung, in der Wissenschaft und in der Medienlandschaft werden sie bereits in einem viel größeren Umfang angewendet als wir das in der Bibliotheksgemeinschaft tun. Regierungen öffnen Informationsquellen in großem Maße, um die verfügbaren öffentlichen Vermögensgegenstände besser für wissenschaftliche Forschung und Studium verfügbar zu machen. Dahinter steht der Gedanke: Je mehr Menschen Zugang zur Information und zum Wissen haben, desto mehr Innovationen gibt es.

Wenn die Bibliotheken weiterhin ihren traditionellen Weg gehen und exzellente Standards produzieren, die außerhalb der bibliothekarischen Gemeinschaft niemand versteht, wäre das ein wenig, als würden sie sich ihr eigenes Web entwickeln. Das funktioniert für Bibliotheken. Aber niemand würde es benutzen, weil es im Wider Web nicht sichtbar wäre und damit auch die Daten und Informationsangebote der Bibliotheken nicht.

Wenn die Bibliotheken den Grundsätzen von Linked Data nicht folgen und damit beginnen, ihre Daten in dieser Form zu publizieren, wird der Rest der Welt ihr Informationsangebot nicht konsumieren, zumindest nicht in größerem Ausmaß.

Was macht OCLC im Bereich Linked Data?

Richard Wallis Bei Linked Data innerhalb von OCLC dreht sich derzeit alles um das Veröffentlichen der bibliografischen Datensätze aus dem WorldCat; um die Fähigkeit, sie als Linked Data über den WorldCat bereitzustellen, der ja als Suchmaschine vorhanden ist. Zur Zeit ziehen wir die Linked Data aus den MARC Records heraus, den standardisierten maschinenlesbaren bibliografischen Daten. Wir hoffen, dass es irgendwann möglich sein wird, Linked Data mit Werkzeugen aus der OCLC WorldShare Plattform direkt zu erzeugen, aber das ist auch noch Zukunftsmusik. Zur Kennzeichnung verwenden wir die Markups von Schema.org[14], dem Verzeichnis, das die größten Suchmaschinen des Webs – Google, Microsoft, Yahoo! und Yandex – 2011 gemeinsam entwickelt und veröffentlicht haben.

Mit den Markups von Schema.org zu allen in WorldCat.org auffindbaren Büchern, Zeitschriften und anderen bibliografischen Quellen werden die kompletten Daten des WorldCat, soweit sie öffentlich zugänglich gemacht sind, für die Suchmaschinen nutzbar. Die Suchmaschinen können die Metadaten in Suchindizes und anderen Anwendungen verwerten. WorldCat.org ist damit die größte Sammlung verlinkter bibliografischer Daten im Web. OCLC hat dazu im Sommer eine Pressemitteilung[15] herausgegeben und in seinen Newslettern[16] darüber berichtet.

Jetzt, drei Monate später, kann ich schon sagen, dass sich bereits die ersten Auswirkungen auf die Erkennung von WorldCat Ressourcen zeigen.

Google, Microsoft, Yahoo, Yandex – das hört sich wirklich nach einer Vernetzung der Bibliothekswelt mit der ganzen Welt an. Was haben die Bibliotheken davon, wenn sie ihre Datenschätze in dieses Welt-

13 http://hadoop.apache.org/

14 http://schema.org/

15 http://www.oclc.org/us/en/news/releases/2012/201238.htm

16 http://www.oclc.org/de/de/enews/2012/29/de_01.htm

netz einbringen? Worin liegt der Gewinn, die Honorierung, die Gegenleistung?

Richard Wallis Der größte Gewinn, oder, wie Sie es nennen, die größte Gegenleistung, liegt darin, dass die Bibliothek die Information darüber, welchen Buchbestand und welche lizenzierten Medien sie anzubieten hat, im Wider Web sichtbar machen kann – dort, wo die Informationsinteressenten und Informationskonsumenten zu finden sind. Über Linked Data kann man die Menschen, die auf der Suche nach bestimmten Informationen sind, direkt zum Bibliotheksangebot führen. Darüber hinaus haben Linked Data aber auch weiteren Nutzen, der sich über alle Bereiche bibliothekarischer Arbeit erstreckt. Zum Beispiel können autoritative Systeme, die über Linked Data eingebunden werden, die Klassifizierung und Katalogisierung unterstützen. Wenn sich die Technologie und die Datenmodelle über das derzeitige reine Veröffentlichen von Linked Data aus dem WorldCat weiterentwickeln, werden sich weitere Optionen und Chancen ergeben, die es dann zu untersuchen gilt.

Führt das nicht zu einer Konzentration, die viele Bibliotheken überflüssig macht?

Richard Wallis Nicht notwendigerweise. Es gibt viele neue Aufgaben. Aber die neu entstehende Technik wird sicherlich dazu führen, dass viel weniger Doppelarbeit gemacht wird, als das heute in der Bibliothekswelt der Fall ist; beispielsweise beim Katalogisieren. Muss man die bibliografischen Basisdaten von Harry Potter wirklich -zig Mal aufnehmen? Das kann wesentlich effizienter organisiert werden. An den verschiedenen Standorten werden dann nur noch die lokalen Nuancen, Informationen, die lokal relevant sind, hinzugefügt. Es gibt jede Menge lokale, regionale Werte. Durch die Verknüpfung wird die Beschreibung dessen, was verfügbar ist, viel reichhaltiger werden als das heute der Fall ist, weil die Metadaten aus allen Quellen zusammengeführt und mit weiteren zum Gegenstand des Interesses gehörenden Informationen direkt verknüpft werden können. In der Idealvorstellung einer vollkommenen Welt würde alles, was über einen Gegenstand, eine Sache bekannt ist, zusammengeführt.

Auf der anderen Seite bieten dynamisch verlinkte Daten wie gesagt den Bibliotheken neue Möglichkeiten, ihre Stärken und Kompetenzen weltweit anbieten zu können und darauf aufbauende neue Informationslieferdienste zu entwickeln. Zu historischen Fragen über ein verschwundenes Dorf in Nordengland z.B. wird man wahrscheinlich in der Britischen Nationalbibliothek mehr vertrauenswürdiges Material erwarten als anderswo. Die Deutsche Nationalbibliothek wieder-

Ankündigung
Voraussichtlicher Erscheinungstermin im Februar 2013
Neue Deutsche Biographie
Herausgegeben von der Historischen Kommission bei der Bayerischen Akademie der Wissenschaften
Das biographische Standardlexikon
Band 25: Stadion-Tecklenborg
<978-3-428-11206-7>
Leinenausgabe: ca. € 138,–
<978-3-428-11294-4>
Halblederausgabe: ca. € 159,–
Bd. 1–25
Leinenausgabe: ca. € 2.700,–
Nähere Informationen zur NDB erhalten Sie in Kürze unter: www.duncker-humblot.de
Duncker & Humblot · Berlin
www.duncker-humblot.de

Duncker & Humblot eLibrary
Die E-Book-Plattform für Bibliotheken
Breites Angebot
✓ Bereits über 2000 E-Books verfügbar
✓ Laufend neue Front- und Backlisttitel
Ganz nach Ihren Bedürfnissen
✓ Einzeltitel ohne Mindestabnahme
✓ Preiswerte Print & E-Book-Bundles
✓ Fach- und Jahres-Kollektionen
z. B. 550 Titel der Kollektion Recht 2004–2006
✓ Pick & Choose mit attraktiven Rabatten
Bequem und komfortabel
✓ Volltextsuche über das gesamte E-Book Programm
✓ Seiten ausdrucken und herunterladen
✓ Unbegrenzte Zahl simultaner Nutzungen
✓ Zeitlich unbegrenzte Nutzungsrechte
✓ Metadaten u. a. im MARC 21-Format
✓ COUNTER-konforme Nutzungsstatistiken
www.duncker-humblot.de/elibrary

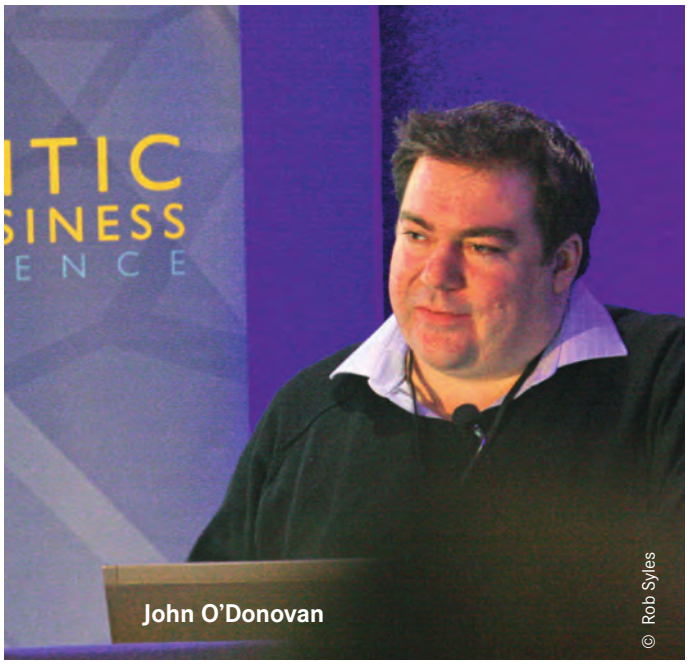
um hat ihre Stärken bestimmt rund um die deutsche Sprache, aber auch zu ihren besonderen Sammelgebieten ... Übrigens wird die Frage nach dem Vertrauen in die Informationsquelle und die bereitgestellten Ressourcen in einer dynamisch vernetzten Welt zu einem wichtigen Thema werden, das Bibliotheken besetzen können. Bibliotheken vertraut man.

Eine abschließende Frage, die uns noch einmal tiefer in die Technik zurückbringt: Können Sie sich vorstellen, dass in Zukunft Linked-Data-Metadaten und Volltexte bei Suchanfragen im Web gleichzeitig ausgewertet werden? Also Linked Data, Textmining und weitere mathematischen Analysen verknüpft eingesetzt werden.

Richard Wallis Textverarbeitung, maschinelle Sprachverarbeitung und Analyse sind komplementäre Techniken zu vielen anderen Technologien wie eben auch zu Linked Data. Linked Data spielen bei der Kopplung und beim Publizieren von Informationen aller mit einer Ressource verbundenen Konzepte eine Rolle, unabhängig davon, welche Technik eingesetzt wird, um sie zu extrahieren. In gleichem Maße wie Linked Data Werkzeuge im Wider Web heranreifen, werden Text Mining und Analyse folgen. Wie schnell das gehen wird, ist sehr schwer vorherzusagen, aber es gibt bereits Werkzeuge, die auf diesen Prozess angewendet werden können. Einige meiner Kollegen bei OCLC Research untersuchen in Zusammenarbeit mit der Europeana einige dieser Techniken und experimentieren bereits damit. Das Problem an dieser Stelle ist kein technisches; es ist ein kommerzielles: Es geht um Urheber- und andere Rechte – ein massives Copyright-Minenfeld. Aber die andauernde Weiterentwicklung des Webs wird sowohl Daten wie Text umfassen.

Das klingt nach einer sehr spannenden Zukunft für die Bibliotheken dieser Welt. Mr. Wallis, wir bedanken uns für das Gespräch.

Richard Wallis Ich möchte gerne noch etwas hinzufügen: Die Linked Data Prinzipien sind erst 2006 herausgekommen und ihre Anwendung auf das Wider Web reift erst heran. Für ihre Anwendung in der Bibliothekswelt ist es noch sehr früh, obwohl ihr Einfluss bereits auf breiter Ebene spürbar und erlebbar wird. Ich glaube, dass es in den nächsten Jahren bedeutende Entwicklungen geben wird, die auf diesen Prinzipien aufbauen. Die Bibliotheken haben keine Wahl: Wenn sie für die Gesellschaft relevant bleiben wollen, müssen sie sich den Einflüssen aus dem Wider Web aktiv stellen.



„Da draußen ist eine Welt voller neuer Möglichkeiten für bibliothekarische Angebote. Es kommt darauf an, wie man sich darauf einlässt."

John O'Donovan

Mr. O'Donovan, „The Press Association" hat die Website[17] der Olympischen Spiele 2012 in London aufgesetzt, diese während der Spiele dynamisch betrieben und tut das heute noch. Die Website besteht aus rund 18.000 Einzelseiten, die automatisch aktualisiert und in 26 Sprachen reproduziert wurden. Die Inhalte sind multimedial; Texte, Video, Bilder, Grafiken ... Wie haben Sie das gemacht?

John O'Donovan Manuell wäre das nicht möglich gewesen. Wir haben Semantik und Metadaten eingesetzt, um die Abläufe zu automatisieren und haben die Workflows völlig neu organisiert. Durch den Einsatz von semantischen Konzepten und Linked Data Techniken ist es uns gelungen, Funktionsfähigkeit zu schaffen, mit der der Inhalt eindeutig automatisch zugeordnet und an der entsprechenden Stelle der Webseite präsentiert werden kann. Das Eintragen aller Informationen aus den verschiedenen Quellen und Systemen des Internationalen Olympischen Komitees (IOC), der Verantwortlichen vor Ort, von Presseagenturen usw. basiert darauf. Wir konnten mit Hilfe der semantischen Beschreibung den Inhalt für die rund 18.000 Einzelseiten automatisch einziehen, managen und ihn in 26 verschiedenen Sprachen reproduzieren.

17 http://www.london2012.com/about-us/the-people-delivering-the-games/locog/

Die Webseite wird dynamisch gepflegt und aktualisiert.

Sie finden mich sprachlos. Wie geht das, ohne Nonsens zu aggregieren?

John O'Donovan Der große Unterschied bei Semantik ist: Man geht mit der Bedeutung der Dinge um, nicht mehr mit den Worten zu ihrer Bezeichnung. Das ist ein fundamentaler Schritt vorwärts. Ein Klavier ist überall auf der Welt ein Klavier, auch wenn die Begriffe dafür unterschiedlich sind. Unser System hat zum Beispiel Namen chinesischer Athleten eindeutig erkannt. Dabei war sogar egal, wie herum Vor- und Nachname geschrieben waren. Die Idee ist, vom Inhalt, dem zentralen Informationsgegenstand auszugehen und alle weiteren Informationen darum herum zu gruppieren; das System zu benutzen, um zu schauen, wo der Inhalt hingehört.

In ihrem Vortrag in Straßburg werden Sie darüber sprechen, wie man Metadaten benutzen kann, um Arbeitsabläufe neu zu organisieren. Verraten Sie dann, wie Sie das genau gemacht haben?

John O'Donovan Das würde zu weit führen. Aber es ist natürlich ein Teil meines Vortrags. Die Olympia-Webseite ist das aktuelle Anwendungsbeispiel zu meinem Thema, wie man neue Technologien einsetzen kann, um Probleme mit großen Datenmengen – Big Data – zu lösen. Das Problem mit Big Data, das die gesamte Wirtschaft hat, ist, wie man vernünftige Erkenntnisse daraus gewinnen kann, wenn man so große Mengen von Daten hat. Wie findet man heraus, was bestimmte Vorgänge bedeuten; wie stellt man die richtigen Anfragen, um vernünftige, nutzbare Antworten aus dem System zu bekommen; zeigen sich Muster in den Daten, kann man sie automatisch präsentieren und wenn ja, wie, usw.

Dieses Problem hat sich die IT selbst gemacht. Wenn man eine Suche in einem großen Archiv durchführt, bekommt man Millionen von Ergebnissen zu so gut wie jedem Thema, das man sich vorstellen kann. Keine Antwort zu bekommen ist genau dasselbe wie Millionen Antworten zu bekommen. Beides ist nicht wirklich hilfreich. Hier hat Semantik bewiesen, dass sie bei der Lösung zweckdienlich ist. Mit Hilfe von Semantik kann man einen Gegenstand, der mit einem weltweit einzigartigen Identifikationscode gekennzeichnet ist – einem Globally Unique Identifier (GUID)[18] – so in Beziehung zu weiteren Informationen stellen, dass seine Bedeutung von Computern eindeutig interpretiert werden kann. Es ist eine völlig neue Art von Content-

18 http://de.wikipedia.org/wiki/Globally_Unique_Identifier

NESCHEN

BÜCHER SCHÜTZEN – WERTE ERHALTEN!

PROFI-BUCHSCHUTZFOLIEN VON NESCHEN

- ✓ erhöhen die Lebensdauer von Büchern
- ✓ schützen vor Verschmutzung
- ✓ auch mit reduzierter Anfangsklebkraft erhältlich
- ✓ lösungsmittelfreie alterungsbeständige Acrylatkleber

Fragen? Sie erreichen uns unter 05722-207-0 oder Sie senden uns eine E-Mail an neschen@neschen.de.

NESCHEN AG
Hans-Neschen-Str. 1 | D-31675 Bückeburg
Telefon: 0 57 22/2 07-0 | Fax: 0 57 22/2 07-197
E-Mail: neschen@neschen.de
www.neschen.com

Besuchen Sie uns!
5. Leipziger Kongress für Information und Bibliothek
11.–14.03.2013, Leipzig

Management. Man muss Inhalt und Produktion komplett anders denken. Der Informationsgegenstand steht im Mittelpunkt der Informationsorganisation. Seine GUID ist der Schlüssel für die sinnvolle automatische Verknüpfung von Informationen durch Maschinen. Ich werde in Straßburg auch darauf eingehen, mit welchen Schwierigkeiten wir konfrontiert waren, als wir die Daten für Olympia so aufbereitet haben.

Was meinen sie damit, „Inhalte und Produktion anders denken"?

John O'Donovan Anstatt viele Menschen dafür einzusetzen, die Website zu managen, haben wir eine grundlegend neue Systemarchitektur aufgesetzt. Jetzt schreiben viele Menschen die Inhalte und wir könnten tatsächlich das System dazu benutzen, um zu entscheiden, wo die Information platziert werden soll, zu welchem individuellen Informationsgegenstand sie gehört; hier also z.B. zu welchem Athleten, welchem Veranstaltungsort, welcher Sportart, ist es ein Ergebnis, eine Nachricht usw. Eine spannende Detailfrage war übrigens, wie man es verhindert, dass Seiten leer bleiben.

„Im Web gab es nie Metadaten, die beschreiben, was drin ist. Die Forschung hierzu versetzt Bibliothekarinnen und Bibliothekare in eine ganz neue Rolle."

Wir gehen also vom Schreiben des Inhalts aus und nehmen die Technik für die Zuordnung. Ich denke, das ist es, wohin die Zukunft geht.

Unser System kann das, weil wir zu jedem wichtigen Informationsgegenstand rund um Olympia semantische Konzepte erarbeitet oder aus vorhandenen Systemen herausgezogen haben. Das ist ganz viel Arbeit, diese Konzepte aufzusetzen, Informationsflüsse und Strukturen zu planen und alles zu modellieren. Aber wenn man das geschafft hat, sinkt die Menge der anfallenden Arbeit massiv.

Können Sie Zahlen nennen?

John O'Donovan Der große Unterschied sind um die 80 % weniger Pflegeaufwand bei den Webseiten. Aus wirtschaftlicher Sicht ist das Problem doch immer die Frage: Wie baut und betreibt man solche Webseiten mit hohem Service und hoher Qualität und hält die Kosten niedrig. Bei dieser Herangehensweise wird alles dynamisch aufgebaut und die Aktualisierung kommt mit dem Beitrag, dem Bericht, aus dem die Information für das dynamische Update anhand der Konzepte herausgezogen wird.

Was glauben Sie: Werden Linked Data die Wirtschaft beeinflussen?

John O'Donovan Sie beeinflussen die Wirtschaft, weil sie Probleme in der IT lösen können, die bisher ignoriert wurden, die schwierig zu umgehen waren. Deshalb sind Linked Data und die Technologie, die es möglich macht, Linked Data zu managen, von großem Wert, um unsere Systeme effektiv zu verbinden und die Dinge zusammenzuführen, die zusammen gehören. Grundlegend betrachtet ist es eine Technologie, die eine Menge an Einsichten ermöglicht, wie ein Geschäft läuft und wie ein Unternehmen seine Informationen verwaltet und nutzt. Linked Data Technologien werden einen großen Einfluss auf die Wirtschaft haben, weil es die einzige Lösung ist, die derzeit funktioniert, um Big Data Probleme zu lösen. Es gibt keine andere.

Sie denken also, Metadatenaggregation in ganz großem Stil ist wünschenswert?

John O'Donovan Ja. Sie wird auf die Wirtschaft einen signifikanten Einfluss haben, abhängig davon, wie Unternehmen einsteigen, wie groß ihr Appetit darauf ist. Man kann sich entweder dafür entscheiden, die Nase vorn zu haben oder abzuwarten und zu sehen, was die neue Technologie mit einem macht. Das mag komfortabel sein – aber, und hier geht es nicht nur um Linked Data, sondern um alle Neuerungen im Umfeld von IT und Globalisierung – wenn man sich die Unternehmen ansieht, die nur Informationen reproduzieren ohne Mehrwert zu schaffen, dann sieht man, dass sie schwer mit der großen Konkurrenz im neuen, stark fragmentierten Markt zu kämpfen haben. Diejenigen, die neue Technologien aus vollem Herzen annehmen und mit neuen Ideen darauf aufbauenden Mehrwert schaffen, diesen Unternehmen geht es gut.

Sie treten auf einer Versammlung von Bibliotheksvertreterinnen und -vertretern auf. Wo sehen Sie den Berührungspunkt?

John O'Donovan Im Web hat es nie Metadaten gegeben, die beschreiben, was wirklich drin ist. Es gab bisher im Internet kein Contentmanagement. Die Forschung hierzu versetzt Bibliothekarinnen und Bibliothekare in eine ganz neue Rolle. Sie sind die Leute, die Inhalte katalogisieren und verwalten können. Darin liegt ein großer Wert in Bezug auf die Reduktion der Gesamtkosten, wenn man mehr darüber nachdenkt, wie man Inhalte organisiert.

Was denken sie über die globale Kooperation der Bibliotheken bei Linked Data?

John O'Donovan Ich fürchte, dazu habe ich keine

ausgeprägte Meinung, weil ich mir auch nicht vorstellen kann, wie das gehen soll. Das Problem liegt in der öffentlichen Förderung der Bibliotheken durch die nationalen Regierungen. Es wird ihnen sicherlich nicht erlaubt werden, die Informationen kommerziell zu verwerten. Sie können diese also nur frei weggeben und hier entsteht ein interessantes Problem, weil es ja viele Unternehmen gibt, die davon leben, Informationen zu verkaufen. Es will ja auch keiner, dass sie sie verschenken.

Aber da ist auch noch eine ganz praktische Herausforderung in Bezug auf globale Massen-Metadaten-Aggregation von Bibliotheken: Wann immer öffentliche Organisationen an einem Datenprojekt beteiligt sind, haben sie alle ihre eigenen Standards und Systeme mit großer Komplexität. Es scheint, zumindest derzeit noch, als wären sie nicht in der Lage, auf einer einfacheren Ebene zu denken. Wenn man über Semantik spricht, spricht man nicht über die Zusammenführung gewaltiger taxonomischer[19] Kollektionen, sondern über Ontologien[20] (u.[21]). Das ist eine ganz andere Ebene, ein viel einfacherer Versuch,

19 http://en.wikipedia.org/wiki/Taxonomy

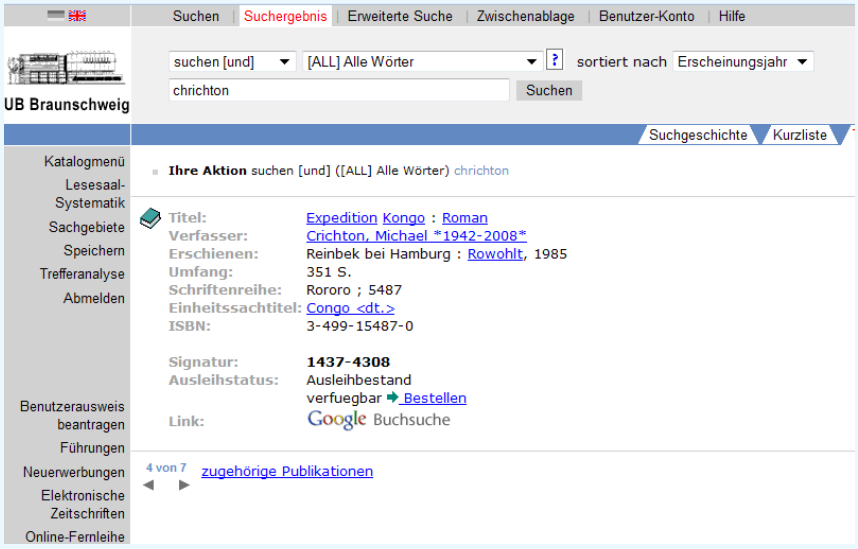
20 Anm. d. Red.: Auf dieser Webseite findet man eine gut verständliche Grafik zu Formalen Konzepten der Ontology – http://www.jfsowa.com/ontology/index.htm

21 Homepage zu Verweis[9] J. F. Sowa http://www.jfsowa.com/ http://www.jfsowa.com/pubs/index.htm

OCLC Experimente für Jedermann

Eine WorldCat Facebook App mit Linked Data und ein Bookmarklet zu URI-Metadatenextraktion – https://apps.facebook.com/worldcat/

Die Universitätsbibliothek Braunschweig, rund 50 Kilometer von meinem Heimatort entfernt, ist nach Auskunft der neuen WorldCat Facebook App mit Linked Data der nächste Standort, an dem ich mir Bücher von Michael Crichton ausleihen kann. Geprüft habe ich das Angebot nicht. Ich habe ja auch keinen Bibliotheksausweis von dort. Aber es ist schon sehr interessant, Linked Data im Experimentierstadium selbst ausprobieren zu können. OCLC Research macht es mit der WorldCat Facebook App mit Linked Data möglich. Mit dieser Anwendung erprobt OCLC, wie man durch die Verfügbarkeit von Linked Data aus dem WorldCat Informationen verknüpfen und anbieten kann.



Klickt man in der WorldCat Facebook App auf ein Buch, bekommt man die Detailseite mit Titel, bibliografischen Angaben mit allen Details bis hin zum Herausgeber des Datensatzes und im Feld „Permalink“ die URI, mit der dieses „Ding“ im Worldcat unverwechselbar markiert ist, in meinem Fall mit http://www.worldcat.org/oclc/53970812.

Darüber hinaus gibt die Anzeige Informationen, wo man das Buch kaufen oder leihen kann, weist auf Rezensionen (im Beispiel von Goodreads) hin, informiert über ähnliche Produkte stellt zu guter Letzt – ganz unten auf der Seite zum Aufklappen – die Linked Data Beschreibung an. Wenn man sie öffnet, erscheint die Anzeige der beschreibenden Metadaten, hergestellt mit Schema.org Markup. Ein Blick lohnt sich.

OCLC benutzt nach eigener Angabe für die Bereitstellung der verlinkten Informationen zwei weitere OCLC Linked Data Services, VIAF[1] für die Personenbeschreibung und FAST[2] für Sachthemen. Die Kopplung der URIs wird über DBPedia[3] und die WikiMediaAPI[4] hergestellt. Die Entwickler haben zudem ein Bookmarklet[5] (ein Java-Makro für Browser) geschrieben, das zeigt, wie man aus Schema.org Markups Daten extrahieren kann, um sie z.B. an ein Patron Driven Acquisition (PDA)-System zu schicken, das OpenURL akzeptiert. Das Bookmarklet extrahiert Autoren-URIs und benutzt diese, um VIAF abzufragen und zu DBPedia zu verlinken.

1 http://www.oclc.org/developer/services/virtual-international-authority-file-viaf

2 http://www.oclc.org/developer/services/fast-linked-data-api

3 http://blog.dbpedia.org/

4 http://www.mediawiki.org/wiki/API:Main_page

5 http://www.oclc.org/developer/prototypes/schemaorg-markup-extractor

Dinge zu beschreiben. Wenn man so viele verschiedene Sprachen, so viele verschiedene Dinge, so viele verschiedene Menschen, so viele verschiedene Themen und so viele verschiedene Systeme zusammenführen will, wie sie in der internationalen Bibliothekswelt vertreten sind, braucht man sprachagnostische

„Wir gehen vom Schreiben des Inhalts aus und lassen dann das System schauen, wo dieser Inhalt hingehört. Ich denke, das ist es, wohin die Zukunft geht."

Systeme, die in der Lage sind, mit dieser Komplexität umzugehen und eine der Optionen würde eine Ontologie sein. Man kann das versuchen, aber dafür braucht man Leute, die bereit sind, Wege zu suchen, um die Komplexität zu reduzieren. Projekte mit Bibliotheken zur Metadatenabstimmung dürften Jahre dauern.

Haben Sie einen Tipp, was Bibliotheken beachten sollten, wenn sie ihre Ressourcen in Linked Data Technologie offen im Web bereitstellen?

John O'Donovan Das Wichtigste ist, sorgfältig darüber nachzudenken, was sie freigeben wollen. Und sie brauchen eine API[22] und Mechanismen, um die Interaktionen kontrollieren zu können. Alles andere ist nachrangig. Es gibt eine große Auswahl an Möglichkeiten.

Möchten Sie unseren Lesern noch etwas sagen, bevor die Teilnehmerinnen und Teilnehmer des OCLC EMEA Regional Council Meeting 2013 in Straßburg die Chance haben werden, Sie persönlich zu hören?

John O'Donovan Da ist eine Sache, die man sich vor Augen führen sollte: Wenn man Bibliothekar ist und jemand, der es gewöhnt ist, mit Metadaten umzugehen, dann ist jetzt wahrscheinlich mehr denn je zuvor die Zeit gekommen, den Wert dessen zeigen zu können, was man Unternehmen anbieten kann. Die meisten Unternehmen versuchen gerade, mit großen Datenmengen, Big Data, umzugehen, Informationen zu analysieren, Information zu organisieren. Wenn man die richtigen Fähigkeiten hat und den Willen, mit dieser Art der Information zu arbeiten, dann gibt es eine Menge Beschäftigungschancen.
Die alte Garde derer, die Bibliotheken führen und Bücher indexieren, Bibliotheken nach alter Schule betreiben, für die gibt es ganz große Möglichkeiten, sich in die neue Welt hinein zu bewegen. Die Probleme, die Unternehmen derzeit im Bereich des Informationsmanagements haben, drehen sich alle um ein starkes und strukturiertes Herangehen an die Aufgaben. Ich glaube, es ist eine sehr interessante Zeit für Bibliotheken und Menschen, die in diesem Bereich arbeiten, weil es so viele Möglichkeiten gibt, etwas sehr Interessantes zu tun.

Das klingt, als würden wir künftig nicht mehr so viele Bibliotheken brauchen?

John O'Donovan Ich glaube, Menschen wollen Bibliotheken haben. Es wird immer einen Platz für Bibliotheken geben. Die Frage ist aber tatsächlich, wie viele und welche Services diese dann anbieten. Der Schlüssel liegt in der Veränderung und die ist, herauszufinden, was die Menschen in der neuen Welt wollen. Firmen wie Amazon haben es herausgefunden. Das war beispielsweise, dass die Kunden die Möglichkeit haben wollten, Bücher herunterzuladen oder Bücher schnell und unkompliziert zu kaufen.
Im Zentrum der IT und Big Data Diskussion ergeben sich viele Chancen, weil die Informationen, die Bibliotheken haben, wie diese verwaltet werden, ihre Struktur, große Möglichkeiten beinhaltet, Unternehmen etwas anzubieten, was sie wirklich brauchen. Während die Firmen sich in der Linked Data Cloud umsehen nach Leuten, die ihnen helfen können, ihre Systeme aufzubereiten, könnte zum Beispiel eine Liste aufgelegt werden mit vertrauenswürdigen Menschen und Organisationen, die diese Arbeit als Dienstleistung anbieten. Die Menschen würden die Dienstleistung kaufen. Sie würden nicht frei verfügbare Ressourcen benutzen, sondern die Dienstleistung kaufen. Ganz einfach, weil sie die Arbeit von kompetenten Personen erledigt haben möchten und den freien Quellen nicht vertrauen.
Da draußen ist eine Welt voller neuer Möglichkeiten für bibliothekarische Angebote. Es kommt darauf an, wie man sich darauf einlässt.

Mr. O'Donovan, wir bedanken uns für diesen spannenden Einblick in diese Zukunftstechnologien, die Daten- und Informationsorganisation ganz offensichtlich auf den Kopf stellen und freuen uns auf Ihren Vortrag.

22 http://en.wikipedia.org/wiki/Application_programming_interface

Vera Münch
Freie Journalistin und
PR-Beraterin/PR+Texte
vera-muench@kabelmail.de