

Bericht von der Dublin-Core-Konferenz (DC-2005) in Madrid

Thomas Baker, Thomas Fischer

1. Die Konferenz

Vom 12. bis 15. September 2005 fand in Leganés (Madrid) die „International Conference on Dublin Core and Metadata Applications“ mit dem Thema „Vocabularies in Practice“ statt [DC2005]. Gastgeber war der Fachbereich Bibliothekswesen und Dokumentation der „Universidad Carlos III de Madrid“ zusammen mit dem Institut „Agustin Millares“ für Dokumentation und Wissensmanagement. Den 214 Teilnehmern aus 33 Ländern wurden 14 ausführliche und 18 Kurzpräsentationen geboten sowie zehn „Special Sessions“ [DC2005-PAPERS].

Fünf Einführungsseminare zu Themen der Metadaten und maschinell verarbeitbarer Thesauri wurden abgehalten. Die Hauptreden der vier Konferenztage wurden von Thomas Baker (Staats- und Universitätsbibliothek Göttingen), Ricardo Baeza (University of Chile), Johannes Keizer (Food and Agriculture Organization of the United Nations) und Eric Miller (World Wide Web Consortium) gehalten. Plenarvorträge wurden simultan ins Spanische übersetzt und mehrere Treffen wurden in französischer oder spanischer Sprache abgehalten.

Die Dublin-Core-Konferenz ist auch das zentrale Ereignis des Jahres für die Dublin Core Metadata Initiative (DCMI) als Organisation. Vor und nach der Konferenz tagten das DCMI Board of Trustees, ein Gremium aus Metadatenexperten und nationalen Vertretern („Affiliates“); das „Usage Board“, das den Standard inhaltlich verwaltet, und das „Advisory Board“, das hauptsächlich aus Leitern von DCMI-Arbeitsgruppen besteht. Während der Konferenz haben sich vierzehn Arbeitsgruppen zu speziellen Fragen im Bereich Metadaten getroffen.

2. Von der Kernsemantik zum Modell

„Zehn Jahre Dublin Core“ war der Hintergrund für die Keynote-Präsentation von Thomas Baker, DCMI Director of Specifications and Documentation. März 1995 fand in Dublin (Ohio) der Workshop statt, auf dem die Kernelemente erstmals entworfen wurden – Creator, Subject, Date, usw. – die der Initiative den Namen gegeben haben. Dieser „Dublin Core“ wurde 1998 bei der Internet Engineering Task Force als Request for Comments (RFC 2413) publiziert, 2000 formal als Standard in Europa (CWA 13874/2000 bei CEN), 2001 in den USA (Z39.95 bei NISO) und 2003 international (ISO 15836/2003) anerkannt [DUBLINCORE].

Am Anfang wurde der Dublin Core als Datenformat konzipiert – d.h. als streng festgelegte Vorlage für digitale Karteikarten. Bereits früh wurden die Elemente jedoch als Vokabular aufgefasst, d.h. als Satz prinzipiell rekombinierbarer Elemente für Beschreibungen, die den Anforderungen spezifischer Anwendungsbereiche angepasst werden konnten – kurz, als Bausteine für Anwendungsprofile. Ausgehend von der vermeintlich simplen Aufgabe, Webseiten auf einfache Art zu beschreiben, hat sich ab 1997 in gegenseitiger Beeinflussung mit der sich entwickelnden Webtechnik von HTML bis hin zu XML und RDF ein allgemeines Modell für Metadaten herauskristallisiert.

Im März 2005 hat die DCMI mit der Verabschiedung des so genannten „Abstrakten Modells“ einen wichtigen Meilenstein erreicht. Dieses DC-Modell ist die formale Grammatik der Metadatensprache, die sich im Laufe der Jahre entwickelt hat. Es hat eine gemeinsame Wurzel mit dem Modell des Semantic Web beim W3C und teilt dessen Grundstruktur, bleibt jedoch absichtlich einfacher als die voll entwickelten Ontologiesprachen des letzteren.

Das abstrakte Modell dient als Maßstab für den systematischen Vergleich verschiedenartiger Implementierungstechniken in Bezug auf deren Ausdrucksfähigkeit. Ein hierarchisch aufgebautes XML-Schema kann beispielsweise Metainformationen differenzierter übertragen als ein HTML-Webdokument mit einem flachen Satz eingebetteter META-Tags. Dagegen kann RDF expliziter als ein XML-Schema die Semantik einer Beschreibung in einen größeren semantischen Zusammenhang verankern und somit die Rekombinierbarkeit der Daten erleichtern.

In der Praxis müssen Systementwickler sich für die eine oder andere Implementierungstechnik entscheiden, dabei liefern die Metadaten nur eines von mehreren Kriterien. Das DC-Modell bietet eine Basis für den Vergleich der möglichen Lösungen in Hinblick auf die Unterstützung von Metadaten und dient somit als Vorbereitung für deren spätere Integration.

Die Interoperabilität der Metadaten ist aber nicht nur eine Frage einer gemeinsamen Semantik mit einem gemeinsamen Modell. Wie auch bei menschlichen Sprachen wird die Interoperabilität umso vollkommener, je besser die Sprachgebräuche verstanden werden – d.h. die Katalogisierungsregeln, die den Metadaten zugrunde liegen. Die Metadatenregeln, die innerhalb einer Anwendungsgemeinschaft benutzt werden, sind Gegenstand eines so genannten Anwendungsprofils. Viele DCMI-Arbeitsgruppen sehen ihre Hauptaufgabe darin, ein Anwendungsprofil für ein bestimmtes Arbeitsgebiet zu erstellen (siehe Abschnitt 6, unten). Diesem Trend zufolge orientiert sich das DCMI Usage Board zunehmend auf die Überprüfung ganzer Anwendungsprofile auf Übereinstimmung mit dem DCMI-Modell.

3. Ein SKOS Core für maschinenverarbeitbare Thesauri

Alistair Miles (Rutherford Appleton Laboratories, UK) hat in Madrid durch einen Plenarvortrag, ein Einführungsseminar und die Beteiligung an mehreren Panels den so genannten SKOS Core (Simple Knowledge Organisation System Core) vorgestellt [SKOS]. SKOS Core war auch Bezugspunkt einiger Forschungsarbeiten, die in der Konferenz präsentiert wurden.

Wo der Dublin Core sich für die Beschreibung dokumentähnlicher Objekte eignet, eignet sich der SKOS Core für die Beschreibung von abstrakten Begriffsstrukturen wie Thesauri und Klassifikationssystemen. Wie auch Dublin Core beruht der SKOS Core auf einer gemeinsamen Metadatengrammatik und stellt dabei einen kleinen Satz allgemein gültiger Kernbegriffe in den Mittelpunkt.

Der SKOS Core wurde in enger Kooperation zwischen Teilnehmern der Semantic-Web-Arbeitsgruppe beim W3C und Experten in der Thesauruskonstruktion entwickelt. Anlass dieser Zusammenarbeit war die Erkenntnis, dass sich existierende Thesauri nicht ohne weitere Interpretation und Nachbearbeitung in die neuen, streng logischen Sprachen der Ontologien (wie z.B. OWL, die Web Ontology Language) übersetzen lassen. Würde man zum Beispiel in einem herkömmlichen Thesaurus die semantische Beziehung zwischen einem „engeren Begriff“ und einem „weiteren Begriff“ in OWL als „sub-class“ bezeichnen, dann würde man der Begriffsbeziehung eine logische Genauigkeit zuschreiben, die mit den losen Begriffen „enger“ und „weiter“ vielleicht nicht gemeint ist. Indem er die Begriffe „enger“ („narrower“) und „weiter“ („broader“) direkt übernimmt, ist der SKOS Core also gut dafür geeignet, herkömmliche Thesauri mit all ihren gemeinten und letztlich auch hilfreichen Ungenauigkeiten sinngetreu in Semantic-Web-fähiger Form abzubilden. Im DCMI-Zusammenhang bietet sich der SKOS Core an, kontrollierte Vokabulare für die Nutzung mit dem Dublin-Core-Element Subject zu definieren.

4. Heterogenitätsbehandlung innerhalb großer Institutionen

Robina Clayphan and Bill Oldroyd haben die Rolle Dublin-Core-basierter Anwendungsprofile bei der Behandlung heterogener Informationsquellen in der British Library beschrieben [CLAYPHAN]. Verteilt auf siebzehn Standorte zwischen London und Yorkshire besteht die British Library weitgehend aus Abteilungen, die historisch als unabhängige Institutionen entstanden und erst im Nachhinein zusammengewachsen sind.

Analog dazu wurden die Digitalisierungsaktivitäten der British Library weitgehend dezentral und auf Initiative einzelner Abteilungen oder im Rahmen von Drittmittelprojekten mit externen Partnern unternommen. Typischerweise wurde für den vorgegebenen Zweck ein Content Management System mit eigener Datenstruktur beschafft. Metadaten wurden nicht auf die übliche Weise

durch MARC21-Katalogisierung erstellt, sondern auf der Basis projektspezifischer Anforderungen. Oft fehlte dabei eine Anforderung für die Interoperabilität mit anderen Systemen der British Library. Resultat ist eine Vielzahl isolierter Datenbestände, die zum Teil als Insellösungen entwickelt wurden.

Um eine elementare Interoperabilität und einheitliche Suche über diese existierenden Bestände hinweg zu ermöglichen, hat die British Library eine begrenzte Zahl von Zielformaten definiert, in die die Metadaten der verschiedenen Projekte exportiert werden sollen. Eine zentrale Rolle spielt dabei das „British Library Application Profile“ (BLAP), ein Dublin-Core-basiertes Anwendungsprofil in XML.

5. Die Nachbearbeitung heterogener Metadaten

Diane Hillmann von der Cornell-Universität hat die Rolle der Metadaten in der US-amerikanischen „National Science Digital Library“ (NSDL) beschrieben [HILLMANN]. Die NSDL ist bemüht, die verteilten Ressourcen dutzender Institutionen im Bereich Wissenschaft und Technik durch einheitliche Schnittstellen effizient verfügbar zu machen. Die Erfahrungen der NSDL in diesem Bereich enthalten wichtige Lehren für eine neu entstehende Informationslandschaft „post-OPAC“, in der das Suchverzeichnis der NSDL nicht mehr wie ein zentraler Bibliothekskatalog verwaltet wird, sondern erst durch das ständige Abfragen der angeschlossenen Internet-Server erzeugt und dynamisch erneuert werden muss.

Die zentralen Dienste von NSDL funktionieren als „aggregators“ (Sammler) der Metadatenbestände Dutzender Mitgliedsinstitutionen. Die abgefragten Metadaten sind jedoch von sehr unterschiedlicher Qualität und lassen sich nur mit Aufwand in einen kohärenten Zentralkatalog einbringen. Mangelnde und gar fehlende Angaben verschlechtern die allgemeine Qualität der Suchergebnisse. Es hat sich aber früh als unrealistisch erwiesen, eine flächendeckende Verbesserung der Qualität der Metadaten bei den Mitgliedsinstitutionen als Dienstleistung einer zentralen Stelle anzustreben. Vielmehr wurde auf die Entwicklung von Werkzeugen und Arbeitsvorgängen zur Nachbearbeitung und Verbesserung der abgefragten Metadaten gesetzt.

Diese Nachbearbeitung besteht aus Verfahren, die sich als Geflecht spezialisierter Dienste entwickelt haben, die bei Bedarf über das Internet abgerufen werden können. Solche Dienste können beispielsweise fehlende Werte (z.B. Medientyp) aus dem Zusammenhang ergänzen (z.B. dem Dateinamen) oder zusammenhangslose Stichwörter mit Quellenangaben versehen (d.h. Simple ins Qualified Dublin Core verwandeln). Manche Nachbearbeitungsdienste wenden dabei Methoden der Suchmaschinen an, indem sie Volltexte herunterladen, um daraus bestimmte Metadatenwerte algorithmisch zu extrahieren.

Andere spezialisieren sich darauf, mehrere Beschreibungen eines gleichen Objektes miteinander zusammenzufassen oder die Beschreibung eines Objektes mit zugehörigen Anmerkungen (z.B. Rezensionen) zu verknüpfen. Einige Dienste bieten sich den Aggregators als Ersatzquellen verbesserter Metadaten an oder gewährleisten die kontinuierliche Verfügbarkeit bestimmter Inhalte durch einen Cache.

Diese Nachbearbeitungsprozesse lassen sich nicht so weit automatisieren, dass alles einfach den Maschinen überlassen werden kann. Die Maschinen sind vielmehr dafür da, den Menschen die Übersicht zu gewähren, damit sie effizienter eingreifen können. Es gilt dabei die Metapher des Orchesterdirigenten, der die verschiedenen Musiker der Reihe nach im richtigen Moment durch Fingerzeig zum Beitrag auffordert. Die dahinter stehende Steuerungssoftware sorgt dafür, dass die Abläufe korrekt nacheinander abfolgen.

Das Problem der Metadatenqualität stellt generell das Modell des Katalogs als Sammlung einzelner Objektbeschreibungen in Frage und führt bei der strategischen Planung zentraler Dienstleistungen zu einem Paradigmenwechsel. Auch diejenigen Informationsanbieter, die ihre Metadaten nur in mangelhafter Qualität abliefern können, haben in mindestens einem Sinne einen wichtigen Dienst erbracht: sie haben eine Auswahl beschreibungswürdiger Ressourcen getroffen. Diese Auswahl bietet schon einen soliden Anhaltspunkt für Nachbearbeitungsdienste an. In der vernetzten Welt des Internets werden somit herkömmliche Rollen – etwa die Rolle des Ressourcenbesitzers im Vergleich zu der Rolle des Ressourcenbeschreibers – in ihre Bestandteile zerlegt und auf neue Weise zusammengestellt.

6. Weitere Themen der Konferenz

In Madrid hat sich eine gemeinsame Arbeitsgruppe von DCMI und dem IEEE (Institute of Electrical and Electronic Engineers), DCMI Education Working Group, gebildet, um die Elemente des IEEE-Standards Learning Object Metadata (LOM) auf das DC-Modell abzubilden und somit eine bessere Interoperabilität der zwei gängigsten Metadatenstandards im Bildungsbereich zu erzielen [EDUCATION].

Seit 1997 haben die Workshops der NKOS-Initiative, Networked Knowledge Organization Systems (NKOS), als Treffpunkte für Entwickler und Nutzer von Thesauri und Klassifikationssystemen im Internet gedient. In einem Special Session in Madrid wurden Gemeinsamkeiten zwischen DCMI, NKOS und dem Semantic Web des W3C thematisiert [NKOS].

„Semantic Web In Practice“ war der Titel des Beitrags von Eric Miller, Leiter des W3C-Arbeitsbereichs „Semantic Web“. Er zeigte einerseits einen engen Zusammenhang zwischen den Entwicklungen des Dublin-Core-Umfeldes mit

denen der Semantic-Web-Initiative auf, und andererseits, wie vom Titel versprochen, erste praktische kommerzielle Anwendungen der Semantic-Web-Technologien [SEMANTIC WEB].

Der DCMI-Arbeitskreis der Wirtschaft, DCMI Corporate Circle, befasst sich mit dem Einsatz von Metadaten und Taxonomien im Bereich firmeninterner Intranets [CORPORATE].

Die DCMI-Arbeitsgruppe für Bibliotheksanwendungen, DCMI Libraries Working Group, diskutierte in Madrid vorrangig die Beteiligung von Experten aus den Reihen der DCMI an der geplanten Neubearbeitung des großen Regelwerks der „Anglo-American Cataloguing Rules“ (AACR), das mit Hinblick auf die Anwendung im Internet nunmehr Resource Description and Access (RDA) heißen soll [LIBRARIES].

Die Beschreibung einer Sammlung als Objekt an sich ist Gegenstand eines Anwendungsprofils, das von einer DCMI-Arbeitsgruppe, DCMI Collection Description Working Group, entwickelt wird. Das Anwendungsprofil wird als Basis einer „Metasearch“-Initiative des US-Standardisierungsgremium NISO angenommen [COLLECTION].

Die DCMI-Arbeitsgruppe „Architektur“, DCMI Architecture Working Group, befasst sich mit der fortwährenden Klärung des Metadaten-Modells, das sich durch die Praxis in den letzten zehn Jahren herauskristallisiert hat, sowie mit Fragen der Implementierung von Metadaten mittels neuer Internet-Techniken [ARCHITECTURE].

Bibliographie

[ARCHITECTURE] <http://dublincore.org/groups/architecture/>

[CLAYPHAN] <http://purl.oclc.org/dcpapers/2005/paper06>

[COLLECTION] <http://dublincore.org/groups/collections/>

[CORPORATE] <http://dublincore.org/groups/corporate/>

[DC2005-PAPERS] <http://dcpapers.lib.sfu.ca/>

[DC2005] <http://dc2005.uc3m.es/>

[DUBLINCORE] <http://dublincore.org/documents/dces/>

[EDUCATION] <http://dublincore.org/groups/education/>

[HILLMANN] <http://purl.oclc.org/dcpapers/2005/paper03>

[LIBRARIES] <http://dublincore.org/groups/libraries/>

[NKOS] <http://nkos.slis.kent.edu/>

[SEMANTIC WEB] <http://www.w3.org/2005/Talks/0915-semweb-em/>

[SKOS] <http://purl.oclc.org/dcpapers/2005/paper01>