KIM WS 2016

Wann: 04. und 05. April 2016

Wo: Universität Mannheim, Fuchs-Petrolub-Festsaal (O 138), Anfahrt

Veranstalter: DINI AG KIM, UB Mannheim

Teilnahmegebühr: 30 Euro

Twitter: Hashtag #kimws16

Abendprogramm: Restaurant und Bar Enchilada (Selbstzahler), ab 20:00 Uhr

Programm

Montag, 04.04.2016

11:00-12:30	**Öffentliche Sitzung der DINI-AG KIM** Protokoll
12:30-13:30	**Mittagspause**
13:30-14:00	**Begrüßung** Christian Benz, Leitender Bibliotheksdirektor der UB Mannheim Stefanie Rühle, Jana Hentschke (DINI-AG KIM Sprecherinnen)

14:00-15:00	**Keynote: Linked Open Data und Strukturiertes Markup im Web (Heiko Paulheim, Uni Mannheim) (Folien)** Um Daten im Web nicht nur menschen-, sondern auch maschinenlesbar darzustellen, wurden in der jüngeren Zeit verschiedene Verfahren vorgeschlagen. Mit Linked Open Data, Microdata, Microformats oder RDFa sowie dem von großen Suchmaschinen propagierten Standard schema.org stehen diverse Möglichkeiten bereit, Daten so aufzubereiten, dass die automatisierte Verarbeitung erleichtert werden kann. In seinem Vortrag gibt Heiko Paulheim einen Überblick über diese Verfahren sowie deren Adaption und zeigt mögliche Anwendungen unter anderem im Bereich wissenschaftlicher Metadaten auf.
15:00-15:30	**Pause**
15:30-17:00	**Session Identifier (Moderation: Stefanie Rühle, SUB Göttingen)** **CHE – Persistente Identifikatoren für Kulturobjekte (Michael Büchner, DDB/DNB) (Folien)** Weltweit eindeutige und persistente Identifikatoren (PI) bilden die zentrale Grundlage für das zuverlässige Referenzieren und Verlinken im Internet. Bestehende PI-Systeme sind dazu speziell auf die Anforderungen einer bestimmten Nutzergruppe ausgelegt und setzen die strikte Unveränderlichkeit der referenzieren Ressourcen voraus. Auch aus den Anforderungen der Datenpartner der Deutschen Digitalen Bibliothek heraus entwickelt, möchte das Fachkonzept von CHE einen neuartigen PI-Dienst entwickeln, der die eindeutige und stabile Identifizierung unterschiedlichster Ressourcen aus allen Kultursparten ermöglicht und zudem eine fortschreitende Entwicklung der Ressourcen erlaubt. **Der Einsatz von Persistenten Identifikatoren und Normdaten bei wirtschaftswissenschaftlichen Forschungsdateninfrastrukturen (Sven Vlaeminck, ZBW) (Folien)** Persistente Identifikatoren sind eine wichtige Voraussetzung für Auffinden, Nachnutzung und Zitation relevanter Forschungsdaten. Der Vortrag beginnt daher mit einem Blick auf die Praxis der Forschungsdatenregistratur für Sozial- und Wirtschaftsdaten da|ra, die Teil des internationalen Data Cite Konsortiums ist. Vorgestellt werden übergeordnete Policies und Verfahrensweisen, aber auch konkrete Möglichkeiten zur Registrierung von Forschungsdaten. Anhand von zwei aktuellen Forschungsdateninfrastrukturen (dem ZBW Journal Data Archive sowie einem Forschungsdatenrepositorium, SowiData Net) wird im Anschluss konkret vorgestellt, wie DOIs aus diesen Applikationen heraus vergeben werden können. Ergänzend wird ein Ansatz vorgestellt, wie durch die Integration von Normdaten der nötige Aufwand für Datengeber/innen beim Self-Upload von Forschungsdaten beträchtlich reduziert werden kann. **ORCID und GND (Sarah Hartmann, DNB) (Folien)** Im Rahmen des von der DFG-geförderten Projektes ORCID DE, das die koordinierte Implementierung der Open Researcher and Contributor ID (ORCID) zur persistenten Identifizierung von Autorinnen und Autoren an wissenschaftlichen Einrichtungen in Deutschland zum Ziel hat, werden ORCID und die Gemeinsame Normdatei (GND) miteinander verzahnt werden. Zum einen wird es darum gehen, wie die Verlinkung zwischen bereits existierenden Identifiern erfolgen kann und zum anderen wie alle mit einer Person verlinkten Publikationen, die in der ORCID-Registry erfasst sind, mit solchen, die in Nachweissystemen wissenschaftlicher Bibliotheken in Deutschland verzeichnet sind, aggregiert und miteinander verknüpft werden können.
17:00-17:30	**Pause**
17:30-18:15	**Session Datenanalyse (Moderation: Philipp Zumstein, UB Mannheim)** **Autoren-Disambiguierung in bibliographischen Metadaten: Der Fall dblp (Marcel R. Ackermann, Schloss Dagstuhl, Leibniz-Zentrum für Informatik) (Folien)** Die dblp computer science bibliography ist die größte Open-Data Sammlung bibliographischer Metadaten in der Informatik. Besonderes Augenmerk legt die Datenbank dabei auf die korrekte Disambiguierung von Autorennamen, d.h., die Urheberschaft wissenschaftlicher Publikationen an Hand weniger Metadaten eindeutig zu erkennen und zuzuordnen. Dieser Vortrag gibt einen kurzen Einblick in den methodischen Ansatz, welcher der Autoren-Disambiguierung in dblp zu Grunde liegt. **Die RapidMiner Linked Open Data Extension (Heiko Paulheim, Uni Mannheim) (Folien)** Um Daten sinnvoll analysieren zu können, ist in vielen Fällen umfassendes Hintergrundwissen notwendig. Dieses manuell bereitzustellen und aufzubereiten ist häufig aufwändig. Die Linked Open Data Extension für die Datenanalyseumgebung RapidMiner ermöglicht mit wenigen Handgriffen eine vollautomatische Verknüpfung von lokalen Daten mit Daten aus der Linked Open Data Cloud, wodurch vielfach neue Einsichten gewonnen werden können.

18:15-18:45	**Lightening Talk** GND Sachbegriffe als SKOS-Thesaurus (Joachim Neubert, ZBW) (Folien)
ab 20:00	**Abendprogramm** Restaurant und Bar Enchilada

Dienstag, 05.04.2016

09:00-10:30	**Session Anforderungen an Normdatendienste (Moderation: Jakob Voss, GBV) (Folien)** Zur praktischen Nutzung von Normdaten müssen diese in geeigneter Weise bereitgestellt werden. In dieser Session sollen gemeinsam Anforderungen an die Bereitstellung gesammelt werden, die über vorhandene Best-Practice (SKOS & LOD) hinausgehen. Kurz gesagt geht es darum, wo und wie a) Informationen über Normdaten b) Inhalte von Normdaten und c) Mappings zwischen Normdaten so zugänglich gemacht werden können, dass sie sich so einfach wie möglich in eigene Anwendungen einbinden lassen. Die Teilnehmer sind dazu aufgerufen, Beispiele und Ideen zur Normdaten-Nutzung mitzubringen. **Report der Ergebnisse**
10:30-11:00	**Pause**

11:00-12:30

Hands-On-Workshop 1, Teil I	Hands-On-Workshop 2, Teil I
Catmandu & Linked Data Fragments (Leitung: Johann Rolschewski und Carsten Klee, Staatsbibliothek zu Berlin) "Catmandu" ist ein Framework für das Importieren, Speichern, Suchen, Exportieren und Transformieren von Metadaten. Das "Linked Data Fragments" Projekt bietet Tools, um RDF-Daten zu publizieren. In Kombination bieten beide Projekte eine einfache Möglichkeit (bibliographische) Metadaten in RDF zu transformieren und per Weboberfläche oder API durchsuchbar anzubieten. Im Rahmen des Workshops werden alle dazu notwendigen Tools vorgestellt und von den Teilnehmern ausgeführt. Die Teilnehmer werden gebeten eine "Virtual Machine" als Entwicklungsumgebung vorab zu installieren. Nähere Informationen zur Installation sind hier zu finden: https://librecatproject.wordpress.com/2014/12/01/day-1-getting-catmandu/.	**OpenRefine (Leitung: Maike Kittelmann, SUB Göttingen) (Folien)** OpenRefine ist ein Tool zum Transformieren, Durchsuchen, Reinigen, Exportieren und Anreichern von Daten. Das Programm kommt aus dem Open Data Bereich und bietet vielfältige Funktionen, um Dateninkonsistenzen ("messy data") aufzuspüren und zu beheben. Die Oberfläche erinnert an vertraute Tabellenprogramme (Excel, Calc), daher ist der Einstieg leicht. Fortgeschrittene Nutzer können Daten in OpenRefine direkt mit Python oder Clojure bearbeiten. **Vorbereitung:** Die folgenden Dateien sollte beim Workshop auf dem Rechner vorliegen. Download, Version 2.5 und 2.6: openrefine.org OpenRefine läuft im Browser, alle Daten bleiben aber lokal. Da das Programm ursprünglich von Google entwickelt wurde, läuft es besonders gut in Google Chrome: http://portableapps.com/de/apps/internet/google_chrome_portable Daten für Übungen: https://www.dropbox.com/sh/tfg6bb2vzavtr7b/AACMYHoxX_LU_fPA09Zux4N0a?dl=0 Download RDF Extension (optional): http://openrefine.org/download.html Showcases zur RDF Extension (deri.ie via Wayback-Machine): http://web.archive.org/web/20150424012501/http://refine.deri.ie/sparqlRecon Download NER Extension (optional): http://software.freeyourmetadata.org/ner-extension/

12:30-13:00 **Pause**

13:00-14:30

Hands-On-Workshop 1, Teil II	Hands-On-Workshop 2, Teil II
Catmandu & Linked Data Fragments	**OpenRefine**

14:30-15:00 **Closing**

Stefanie Rühle, Jana Hentschke