

VERNETZTE RECHTSINFORMATION – DER EINSATZ VON LINKED DATA TECHNOLOGIEN AM BEISPIEL DES ÖGB-VERLAGES

Tassilo Pellegrini

FH-Professor, Fachhochschule St. Pölten, Department Medienwirtschaft, Matthias Corvinus Strasse 15, 3100 St. Pölten, Österreich, Tassilo.pellegrini@fhstp.ac.at; <http://www.fhstp.ac.at>

Christian Wachter

Verlag des Österreichischen Gewerkschaftsbundes, Johann-Böhm-Platz 1., A-1020 Wien, Austria, Christian.Wachter@oegbverlag.at; <http://www.oegbverlag.at>

Andreas Blumauer

Geschäftsführer, Semantic Web Company, Neubaugasse 1, 1070 Wien, Austria, A.Blumauer@semantic-web.at; <http://www.semantic-web.at>

Jürgen Paulus

Compass Verlag, Matznergasse 17, 1140 Wien, Austria, Juergen.Paulus@compass.at; <http://www.compass.at/>

Schlagnworte: *Linked Data, Content Management, Content Curation, Dynamic Semantic Publishing, IRIS 2014*

Abstract: *Viele juristische Fachverlage stehen heute vor der Herausforderung, dass in den vergangenen Jahren große Mengen an (teilweise unstrukturiertem) Content angefallen sind, jedoch die aktuelle technische Ausstattung nur bedingt eine kosteneffiziente Weiterverarbeitung und Mehrfachverwertung ermöglicht. Dies trifft auch auf den ÖGB-Verlag zu, der über einen umfassenden Dokumentenbestand zu arbeitsrechtlichen Fragestellungen und insbesondere über eine vollständige Datenbank der österreichischen Kollektivverträge verfügt. Der Beitrag illustriert den Einsatz von Linked Data Technologien zur Verbesserung redaktioneller Workflows sowie Mehrwertpotenziale, die sich durch die Vernetzung und semantische Anreicherung von Content-Beständen ergeben.*

1. Einleitung

Der ÖGB-Verlag verfügt über einen erschöpfenden Bestand an Kollektivverträgen und dazugehörige Gesetze, Verordnungen, EU-Richtlinien, Kommentaren, Monografien, Aufsätze und

Ratgeber im Bereich des Arbeitsrechts. Im Rahmen eines Entwicklungsprojektes¹ soll das bestehende, an Experten gerichtete Kollektivvertrags-System (einsehbar unter <http://www.kvsystem.at/>) erweitert und die komplexe Materie „Kollektivverträge“ für Laien und PraktikerInnen in Unternehmen, im Sinne einer Ermächtigung zur Selbsthilfe (z.B. im Rahmen von Rechtsstreitigkeiten, der Wahrnehmung von Rechten und Pflichten oder bei Veränderungen der Lebenssituation) aufbereitet werden. Das System soll begründete Empfehlung für die Suche nach dem relevanten Kollektivvertrag (aus dzt. insgesamt 600) geben und Kurzdarstellungen der wichtigsten und meist gefragten Inhalte wiedergeben.² Das KV-System soll weiters in ein übergreifendes arbeitsrechtliches Informationssystem, das neben Kollektivverträgen auch andere Rechtsquellen heranzieht, eingebettet und die konzeptionellen Voraussetzungen für eine automatische Verrechnung und Cross-Lizenzierung geschaffen werden. Über leicht bedienbare Schnittstellen (APIs) sollen die Daten Content-Partnern und der Öffentlichkeit für Zeitreihenanalysen oder detaillierte Lohn- & Gehaltstafeln zur Verfügung gestellt werden. Umgekehrt sollen externe Quellen (z.B. aus dem Rechtsinformationssystem des Bundes) eingebunden und sukzessive um relevante Daten aus dem DACH-Raum erweitert werden. Nach Projektende plant der ÖGB-Verlag das Framework neben dem Arbeitsrecht auf weitere Domänen wie Arbeitnehmerschutz und Sozialrecht auszudehnen.

Um die Kontextualisierung und Accessibility der existierenden Informationsressourcen zu verbessern und um in Folge redaktionelle Workflows zu optimieren und neue Verwertungskanäle zu erschließen, konzentriert sich der ÖGB-Verlag verstärkt auf die automatisierte Verarbeitung von Content (Content Curation) als auch die Adaption neuer Standards für die automatische Syndizierung und Integration von strukturierten Daten (Dynamic Semantic Publishing) für die interne und externe Informationslogistik. Dieses als Linked Data³ bekannte Prinzip hält sukzessiven Einzug in die Redaktionssysteme und Verwertungsstrategien content-verarbeitender Unternehmen wie aktuelle Untersuchungen (Graube et al. 2011; Pellegrini 2012) und Fallstudien von international agierenden Fachverlagen wie Reed Elsevier, Wolters Kluwer, Pearson Publishing, Springer Verlag u.a.m belegen.

Dieser Beitrag stellt die Linked Data Strategie des ÖGB-Verlages im Themenfeld der Kollektivverträge vor. Dazu werden im ersten Schritt die technischen Grundlagen von Linked Data erläutert, die technische Architektur des semantischen Indizierungs- und Suchsystems vorgestellt sowie der Einsatz von Linked Data in den redaktionellen Workflows für Zwecke des Content Enrichments und des Data Publishings illustriert. Zusätzlich werden die Mehrwertpotenziale für Linked Data am Beispiel von vier Use Cases dargestellt.

2. Linked Data – Technologische Grundlagen

Konventionelle Datenbereitstellungsstrategien in Form von (semi-)strukturierten Dokumenten (z.B. HTML, CSV-Dateien) oder proprietären APIs werden nur bedingt den Ansprüchen hoch vernetzter und dynamischer Daten-Ökosysteme gerecht. Mit jeder zusätzlichen Quelle steigen die Integrationsaufwände exponentiell, Veränderungen in der Datenbankstruktur gehen oftmals zu Lasten der Systemintegrität und Aktualisierungen der Datenbasis sind meist nur mit hohen Aufwänden in Echtzeit verfügbar. Hier setzt der Linked Data Ansatz an, der eine höchst-mögliche

¹ Die Entwicklung findet im Rahmen des Projektes „NoLDE – Network of Linked Data Excellence“ statt. Das Projekt wird durch die FFG-Programmlinie „COIN Kooperation und Netzwerke“ gefördert (Projektnummer: 3592880).

² Siehe z.B. die Kurzübersicht zum Kollektivvertrag der Arbeiter der Metallindustrie <http://www.kollektivvertrag.at/kv/eisen-metallerzeugende-und-verarbeitende-industrie-arb> aufgerufen am 4.1.2014.

³ Für eine Beschreibung der „Linked Data“ Methode siehe <http://www.w3.org/standards/semanticweb/data>, aufgerufen am 28.10.2013.

technische Interoperabilität anstrebt und so die kosteneffiziente und zeitkritische Integrierbarkeit, eindeutige Interpretierbarkeit und Wiederverwendbarkeit von Daten ermöglicht (Mitchell & Wilson 2012; Cranford 2009). Linked Data bedient sich sogenannter Semantic Web Standards um existierende Datenbestände hoch strukturiert aufzubereiten und plattformunabhängig für die Integration und Syndizierung bereitzustellen. Hierbei werden Daten mittels eines normierten Datenmodells (RDF = Resource Description Framework; siehe W3C 2004) strukturiert und verfügbar gemacht. Die dabei generierten Daten werden im konventionellen Sinne nicht relational sondern als Graph repräsentiert. Sowohl die Knoten als auch die Kanten des Graphen sind über URIs (Uniform Resource Identifiers) eindeutig identifizierbar und referenzierbar. Dieser semantische RDF-Graph kann mittels der normierten Abfragesprache SPARQL (W3C 2008) feingranular und in hoher semantischer Tiefe abgefragt werden. Dies erlaubt hoch expressive Datenbankabfragen, die mit konventionellen Mitteln entweder gar nicht oder nur mit hohen technischen und personellen Aufwänden machbar wären.

Tim Berners-Lee, Direktor des World Wide Web Konsortiums, fasst die technischen Prinzipien von Linked Data folgendermaßen zusammen (Berners-Lee 2009):

1. Nutze eindeutige Identifikatoren (Uniform Resource Identifiers – URIs) als Name für Dinge.
2. Nutze http-URIs um diese Dinge im World Wide Web auffindbar zu machen.
3. Nutze den RDF-Standard zum Annotieren der URIs mit sinnvoller Kontextinformation.
4. Verknüpfe URIs mit anderen URIs um weitere Informationen auffindbar zu machen.

Das erste Prinzip stellt eine Grundbedingung dar und besagt, dass Ressourcen über einen Uniform Resource Identifier (URI) entsprechend der IETF URI Konventionen (Berners-Lee 1998) ausgezeichnet werden müssen.

Das zweite Prinzip besagt, dass http-URIs als Bezeichnungen für Ressourcen verwendet werden sollen. Wie Berners-Lee anmerkt, wird oftmals übersehen, dass http-URIs im eigentlichen Sinne keine Adressen sondern „Namen“ darstellen, auf denen eine mächtige und evolvierende Infrastruktur in Form von Schreib- und Leseautomatismen (sog. REST-Services) aufbaut.

Das dritte Prinzip besagt, dass die maschinelle Verarbeitung verfügbarer Datenquellen, die bereits die URI-Konventionen erfüllen, durch Anreicherung mit interoperablen Metadaten verbessert wird. Strukturierte Annotation auf Grundlage von Wissensmodellen bzw. Ontologien⁴ setzt hier an. Die Repräsentation der hierbei verwendeten Vokabulare muss der RDF-Norm genügen.

Als viertes und letztes Prinzip sollen die verfügbaren URIs durch gegenseitige Verweise de-referenziert und dadurch vernetzt werden, so wie es im konventionellen „Web of Documents“ auf Basis des Hypertext-Prinzips bereits erfolgt. Denn durch Interoperabilität werden Daten zu Netzwerkütern und steigern ihren Wert mit dem Grad ihrer Konnektivität und Referenzierbarkeit – eine ökonomische Gesetzmäßigkeit, die für den Produktionsfaktor Metadaten bisher unterbelichtet ist. Dieser Aspekt stellt auch den wichtigsten Unterschied zur konventionellen Datenbewirtschaftung dar, wo aufgrund proprietärer Repräsentationsstandards und Schemata kaum Netzeffekte zu erzielen sind. Weitere wertsteigernde Aspekte durch Linked Data im Vergleich zu konventionellen Integrationstechnologien lassen sich laut Auer (2011) folgendermaßen beschreiben:

De-Referenzierbarkeit: Die Verwendung von URIs erlaubt nicht nur Dinge am Web eindeutig zu identifizieren sondern auch diese inklusive der angereicherten Zusatzinformation abzurufen.

⁴ Saumure & Shiri definieren Ontologien folgendermaßen: „Ontologies are being considered valuable to classifying web information in that they aid in enhancing interoperability - bringing together resources from multiple sources.“ (Saumure & Shiri 2008, S. 657)

Kohärenz: Die Verwendung von RDF als universelles Datenmodell erlaubt die kohärente Vernetzung von Informationen aus unterschiedlichen Namensräumen und ermöglicht auf diese Weise die semantische Anreicherung von Information durch sogenannte typisierte Links.

Integrierbarkeit: Das geteilte Datenmodell erlaubt – aus technischer Perspektive – die niedrighschwellige Integration von syntaktischen und semantischen Informationen aus den vernetzten, dispers vorliegenden Datenquellen. Mittels Schema-Mapping (z.B. von RDF Vokabularien) und Instance Matching können in Folge auch schwergewichtige Integrationsmaßnahmen gesetzt und semantisch hoch expressive Informationsbestände aggregiert werden.

Aktualität: Die Datenquellen inklusive ihrer Netzstruktur können aufgrund des geteilten Datenmodells leicht aktualisiert und veröffentlicht werden, ohne das daraus zusätzliche Integrationsaufwände oder Performanceverluste wie sie üblicherweise bei konventionellen Extraktionsmaßnahmen oder Datentransformationen entstehen. Dies garantiert unter anderem eine hohe Aktualität der Daten und darauf aufbauender Dienste.

3. Die Linked Data Architektur des ÖGB-Verlages

Zur Erreichung der oben beschriebenen Ziele kommen beim ÖGB-Verlag zwei Software-Frameworks zum Einsatz: 1) das ONTEASY Framework zur automatischen Strukturierung und Analyse deutscher Texte und 2) das Poolparty Framework zur Wissensorganisation und Herstellung von Mehrwertdiensten auf Basis von Linked Data.

3.1. Information Preprocessing mit dem ONTEASY Framework

Seit dem Jahr 2010 entwickelt der Compass-Verlag das System ONTEASY, das der feingranularen semantischen Analyse von deutschsprachigen Texten dient. Das System ist in der Lage automatisch Bedeutungszusammenhänge aus Texten zu extrahieren und diese in Form von lexikalischen und grammatikalischen Ontologien formalisiert abzubilden. Dadurch ist es möglich Freitexte nachträglich mit geringen Kostenaufwänden und hoher semantischer Präzision zu strukturieren und einer weiterführenden maschinellen Verarbeitung zuzuführen. Das Wissen der dabei entstehenden Ontologie wird durch verschiedene Arten von Relationen beschrieben, die die jeweiligen Begriffe / Konzepte miteinander semantisch verbinden. Abbildung 1 veranschaulicht exemplarisch die generische Architektur und dazugehörigen Prozesse einer Ontology-Based Information Extraction (OBIE).

Die ONTEASY-Methode verbessert folgende Aspekte der Textanalyse: 1) Klassifizierung der Textsorte (z.B. Gesetzestext, Kommentar, Meldung, Bericht, etc.); 2) Mehrfache inhaltlich-kategoriale Zuordnung eines Textes; 3) Verbesserung der Eigennamenerkennung (Named Entity Recognition); und 4) Extraktion der Kerninhalte eines Textes bzw. Zusammenfassung (Automatic Summarization).

Damit diese Aufgaben von der Ontologie unterstützt werden, werden folgende Relationen in der Ontologie gespeichert:

- Hierarchische Substantiv-/Instanz-/Teil-Von- Relationen (z.B. Institution => Unternehmen => Ges.m.b.H.)
- Synonym-Relationen zu den einzelnen Wortarten (z.B. Weihnachtsgeld - Weihnachtsremuneration)
- Antonym-Relationen zu den einzelnen Wortarten (z.B. Anstellung <=> Kündigung)

- Prädikats-Relationen (Subjekt, Prädikat, Objekt, Adverbialbestimmung, z.B. Belegschaft wählt Betriebsrat in geheimer Wahl.)
- Ereignis-Relationen (Handlungen, z.B. Arbeitgeber spricht Kündigung aus => Arbeitnehmer empfängt Kündigung => Kündigungsfrist beginnt zu laufen.)
- Zustands-Relationen (Ausgangs- und die Endsituation eines Ereignisses, z.B. Anstellung: arbeitssuchend => beschäftigt)

Die Erstellung der Ontologie mittels ONTEASY erfolgt semi-automatisch. Dazu werden große Textkorpora zu einem Themenbereich analysiert, um die jeweiligen allgemeinen Muster der Textsorten zu erkennen (spezielle Satzstrukturen, spezielles Vokabular etc.). Beschrieben werden dann die Konzepte durch die im Text verwendeten Wörter sowie durch die verschiedenen Relationen, die diese mit den anderen Wörtern im Satz/Satzgefüge haben. Dabei sollen die Ergebnisse der automatischen Textanalyse auf der Wort- und Satzebene vom Anwender verwendet werden, um den Aufwand bei der Erstellung der Ontologie auf ein Minimum zu begrenzen.

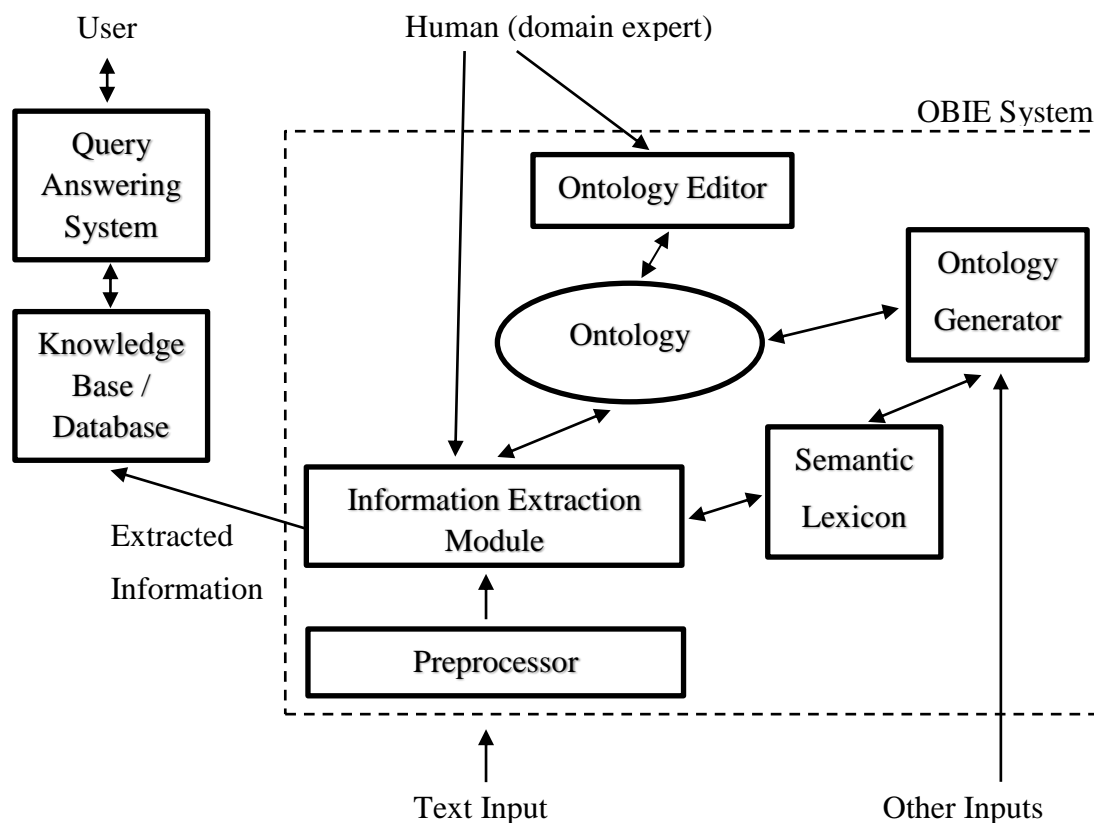


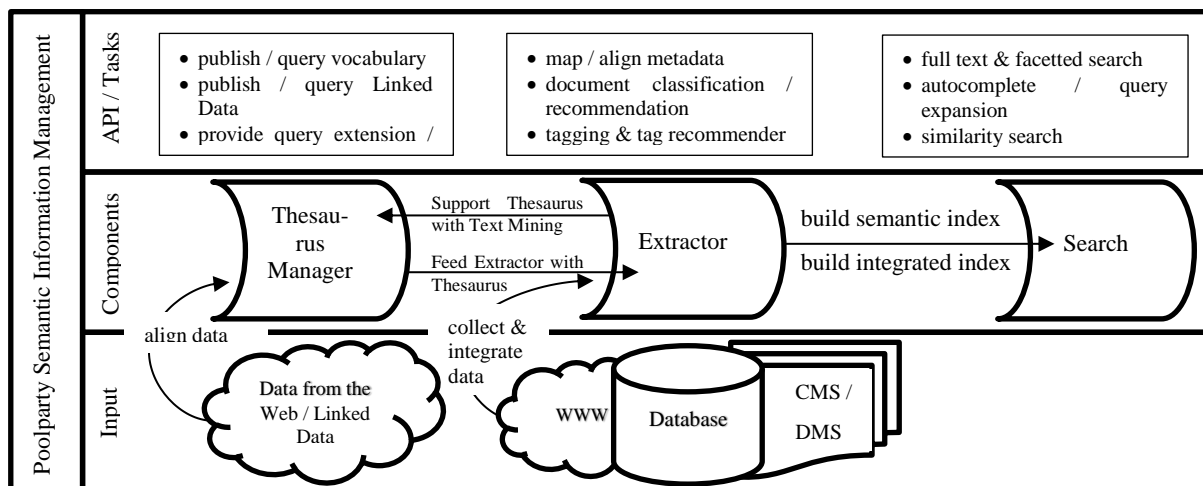
Abbildung 1: Schematische Darstellung des ONTEASY Frameworks

Sobald die Ontologie über genügend Begriffe und Relationen verfügt, wird dieses Wissen für die Textanalyse genutzt, indem die Wörter eines Textes und die jeweilige Grammatik der Sätze mit den gespeicherten Strukturen und Konzepten der Ontologie verglichen werden.

3.2. Wissensorganisation & Linked Data Services mit dem Poolparty Framework

Die durch ONTEASY gewonnenen Daten bilden die inhaltliche Grundlage für die nachfolgende Wissensorganisation und darauf aufbauende Services. Dazu entwickelt die Semantic Web Company im Rahmen des Poolparty Frameworks ein Set von Software-Tools, das die Anreicherung, die

Integration und die Veröffentlichung der gewonnenen Daten als Linked Data ermöglicht. Das Metadatenmanagement basiert auf dem offenen W3C Standard SKOS/RDF⁵. Das SKOS-Modell fungiert als semantische Integrationsschicht zwischen unterschiedlichen Datenquellen und bildet die Basis für Content Curation Services, Knowledge Discovery & Recommendation Services und Dynamic Semantic Publishing Strategien. Die PoolParty API's erlauben weiters die Integration in andere Systeme wie Intranet, CMS, DMS, Web Shops oder Enterprise Wikis. Abbildung 2



veranschaulicht die Architektur und Funktionsweise des Poolparty Frameworks.

Abbildung 2: Schematische Darstellung des PoolParty Frameworks

Das System besteht aus drei Komponenten:

Mittels eines Thesaurus Managers werden kontrollierte Vokabulare auf Basis von W3C Semantic Web Standards wie SKOS, RDF und SPARQL erstellt und gewartet. Das System erlaubt zusätzlich die Publikation von Thesauri als Linked Data. Mittels einer SPARQL-Schnittstelle lassen sich die kontrollierten Vokabulare mit CMS, ERP oder Wikis integrieren. Die Installation beinhaltet einen Editor, ein Wiki-Frontend und eine individualisierbare Linked Data Oberfläche.

Der Extractor bietet präzise Text Mining Algorithmen basierend auf SKOS-Ontologien, wie sie etwa durch das ONTEASY Framework bereitgestellt werden. So können Dokumente und Datensätze automatisiert analysiert, Phrasen extrahiert und neue Metadaten generiert werden. Weiters werden strukturierte wie auch unstrukturierte Informationen in einen einzigen, umfangreichen und semantischen Suchindex überführt und verschiedene Metadaten-Schemata auf ein vereinheitlichtes Wissensmodell übertragen.

Die Suchplattform liefert präzise Suchergebnisse durch die komplementäre Orchestrierung des Text Minings mit den semantischen Wissensmodellen. Die dazugehörige Such-API erlaubt Funktionalitäten wie Volltextsuche, facettierter Suche, Autocomplete-Funktionen, Ähnlichkeitssuche und andere Assistenten.

4. Anwendungsfelder von Linked Data im ÖGB-Verlag

Im Folgenden werden die avisierten Anwendungsfelder von Linked Data im ÖGB-Verlag illustriert.

⁵ Siehe <http://www.w3.org/2004/02/skos/> (aufgerufen Dezember 20, 2013)

4.1. Content Augmentation

„Content Augmentation“ bezeichnet jenen Vorgang, in dem Inhalte, die von Autoren oder Mitarbeitern z.B. im Rahmen redaktioneller Workflows oder eines Enterprise Content Management erstellt werden, um andere Inhalte bzw. Kontextinformationen angereichert werden. Diese können aus internen und externen Quellen (z.B. News-Feeds, Content Repositories, Open Data) bezogen werden. Im Enterprise-Umfeld ergeben sich daraus Anwendungsmöglichkeiten wie Semantic Tagging⁶ und multimodale Mashups⁷. *Mashups* mit aktuellen Inhalten aus definierten internen und externen Quellen: Interne Dossiers können um semantisch ähnliche Inhalte angereichert werden.

Im Kontext des ÖGB-Verlages können z.B. eigene Contents und Informationen aus den Webportalen, Blogs, Social-Media-Kanälen und Publikationen der Gewerkschaften um Informationen aus diversen externen Quellen (Nachrichtendienste, Informationen von Bundes- und EU-Institutionen, statistische und wissenschaftliche Quellen) ergänzt, ausgewertet und strukturiert dargestellt werden.

Ein konkretes Projekt in diesem Bereich ist die Entwicklung eines Tools zur Unterstützung der Redaktion bei der Aufbereitung von Kollektivverträgen. Die dabei gewonnenen Erkenntnisse und Methoden sollen schrittweise auf weitere Domänen angewendet werden.

4.2. Agile Datenintegration und integrierte Sichten auf Geschäftsobjekte

Integrierte Sichten auf Geschäftsobjekte wie Produkte, Technologien, Kunden und Mitbewerber sind ein wesentlicher Schlüssel für die strategische Produktplanung und flexible Servicierung. Integrierte Sichten manifestieren sich für den Wissensarbeiter als 'Dashboards', auf denen wesentliche Workflows, Fakten, Neuigkeiten, Trends und Relationen zu einer Entität aufbereitet und aussagekräftig visualisiert werden.⁸

So können im Verlagsumfeld z.B. eigene Verlagsprodukte nach Thema, Produktkategorie, Zielgruppe, Umsatz, Aktualität usw. selektiert und mit dem Angebot der Mitbewerber verglichen werden – eine Routine, die mit herkömmlichen Methoden erhebliche Ressourcen bindet.

4.3. Mitarbeiterportal

Mitarbeiterportale sind ein wesentlicher Bestandteil jedes betrieblichen Wissensmanagement-Systems und bieten einen zentralen Anlaufpunkt für jeden Mitarbeiter bei der Informationsbeschaffung (Enterprise Search), beim Austausch (Enterprise Wiki) und der Veröffentlichung (Blog und Content Management) von Informationen im Intranet.

Auch in einem Kleinunternehmen wie dem ÖGB-Verlag mit ca. 50 MitarbeiterInnen kann sich niemand sicher sein, dass er über alle zu einer Fragestellung relevanten und im Unternehmen vorhandenen Informationen verfügt. Wird in der einen Datenbank von „Kunde“ gesprochen, in der anderen aber vom „Klient“, so beziehen sich zwar beide Bezeichner auf dasselbe „Konzept“ bzw. „Geschäftsobjekt“, jedoch bleibt der Maschine diese Beziehung ohne entsprechend formalisierte

⁶ Inhalte werden um Tags angereichert, die als vernetzte Ressource aus der Linked Data Cloud wiederum selbst über weiterführende Kontextinformation verfügen können.

⁷ Auskoppelung von Dossiers entlang unterschiedlicher Darstellungsweisen z.B. als Tabelle, als Zeitreihe, als Geo-Visualisierung etc.

⁸ Zahlreiche Medien, wie z.B. BBC oder der Spiegel Online, machen von diesem Prinzip bereits Gebrauch, und fassen News, Medien und Fakten zu Personen, Themen oder Orte zusammen, z.B. in Form so genannter 'Topic Pages' für Tiere Siehe z.B. http://www.bbc.co.uk/nature/life/Red_Squirrel (aufgerufen Dezember 20, 2013). Für die technische Umsetzung siehe Rayfield (2012).

Auszeichnung verborgen. Eine übergreifende Suche nach allen Kunden oder die ganzheitliche Sicht auf einen Kunden ist damit oft nicht möglich. Dieses Problem wird zusätzlich verschärft, falls auf Instanzebene ein und derselbe Kunde mehrfach repräsentiert wird und somit keinen „Unique-Key“ zugewiesen bekommen hat.

Das URI-System ist die Basis zur Entwicklung kontextsensitiver, „mitdenkender“ Widgets. Inhalte können z.B. über ein Enterprise Wiki von Mitarbeitern eingestellt und über ein Tagging-System, das auf einen SKOS-basierten Thesaurus zugreift, annotiert werden. Damit lassen sich Intranet-Inhalte intelligent verknüpfen. Somit kann z.B. die Suche nach „ähnlichen“ Inhalten realisiert werden, um Zusammenhänge zu erkennen, Doppelarbeiten zu vermeiden oder weiterführende Quellen zu erschließen.

5. Conclusio und Ausblick

Verlage (aber auch die Publikations- und Informationsabteilungen der öffentlichen Hand!) werden heutzutage nicht nur von der Entwicklung des Marktes angetrieben, sondern immer mehr von einer stürmischen Entwicklung der Technologien. Dies betrifft die gesamte Wertschöpfungskette vom Contentwerb über Verarbeitung, Bündelung, Präsentation, Auslieferung bis zur Abrechnung. Durch diese Technologieentwicklung werden auch die Rahmenbedingungen am Markt - in Bezug auf Mitbewerber, Zulieferer und Kunden - ständig neu gesetzt.

Die Herausforderung für Verlage besteht nicht nur darin, sich diese neuen Technologien anzueignen, sondern vielmehr darin, funktionierende Geschäftsmodelle zu entwickeln. Die ökonomischen Netzwerkeffekte interoperabler, referenzierbarer Daten eröffnen hier neue Horizonte. Strategische Aufgabe der Technik ist in diesem Zusammenhang die durchgängige Einführung zukunftsorientierter, robuster Standards, wie den Semantic Web Standards RDF, SPARQL oder SKOS, sowie die Verankerung grundlegender Prinzipien, wie URI und Linked Data. Daran sollten dann vielfältige, bisher noch gar nicht angedachte Applikationen anknüpfen können.

6. Literatur

Auer, Sören (2011). Creating Knowledge Out of Interlinked Data. In: Proceedings of WIMS'11, May 25-27, 2011, p. 1-8

Berners Lee, Tim (1998). Uniform Resource Identifiers (URI): Generic Syntax. In: IETF Network Working Group, Request for Comments: 2396, See also: <http://www.ietf.org/rfc/rfc2396.txt>, accessed February 20, 2011

Berners-Lee, Tim (2009). Linked Data Design Issues. In: <http://www.w3.org/DesignIssues/LinkedData.html>, accessed 25.05.2010

Cranford, Steve (2009). Spinning a Data Web. In: Price Waterhouse Coopers (Ed.). Technology Forecast, Spring 2009. <http://www.pwc.com/us/en/technology-forecast/spring2009/index.jhtml>, accessed September 20, 2013

Graube, Markus; Pfeiffer, Johannes; Ziegler, Jens; Urbas, Leon (2011). Linked Data as integrating technology for industrial data. In: 2011 Int. Conference on Network-Based Information Systems, 7-9 Sept. 2011, p. 162-167

Mitchell, Ian; Wilson, Mark (2012). Linked Data. Connecting and exploiting big data. Fujitsu White Paper, March 2012. <http://www.fujitsu.com/uk/Images/Linked-data-connecting-and-exploiting-big-data-%28v1.0%29.pdf>, accessed September 12, 2013

Pellegrini, Tassilo (2012). Semantic Metadata in the News Production Process. Achievements and Challenges. In: Lugmayr, Artur et al. (Eds). Proceeding of the 16th International Academic MindTrek Conference 2012. ACM SIGMM, p. 125-133

Rayfield, Jem (2012). Sports Refresh: Dynamic Semantic Publishing. In: BBC Internet Blog, http://www.bbc.co.uk/blogs/bbcinternet/2012/04/sports_dynamic_semantic.html, visited May 5, 2012

Saumure, Kristie; Shiri, Ali (2008). Knowledge organization trends in library and information studies: a preliminary comparison of pre- and post-web eras. In: Journal of Information Science, 34/5, 2008, p. 651-666

W3C (2004). RDF – Resource Description Framework. In: <http://www.w3.org/RDF/>, accessed December 10, 2013

W3C (2008). SPARQL Query Language for RDF. in: <http://www.w3.org/TR/rdf-sparql-query/>, accessed December 20, 2013