

Editorial

Liebe Fachgruppenmitglieder,

Sie erhalten diese Ausgabe des EMISA FORUM später als gewohnt, weil die umfangreichen Arbeiten zur Fertigstellung des EMISA-Sonderhefts anlässlich des 30-jährigen Bestehens unserer Fachgruppe die reguläre Ausgabe etwas verzögert haben. Aber ich denke, dies hat sich in jedem Fall gelohnt; wir haben bereits viele positive Reaktionen zu diesem Sonderheft bekommen, was uns sehr freut. Nochmals ein herzliches Dankeschön an den Herausgeber und Mitgestalter Manfred Reichert sowie an die Autoren.

Auch das Fachgruppentreffen im Oktober 2010 in Karlsruhe war von dem Jubiläum *30 Jahre EMISA* geprägt. Dem Organisationsteam um Andreas Oberweis gelang es, alle bisherigen EMISA-Sprecher nach Karlsruhe zu holen, lediglich Bernd Meyer musste leider kurzfristig absagen. Dies zeigt auf beeindruckende Art und Weise, welches Zusammengehörigkeitsgefühl die EMISA auch nach 30 Jahren noch besitzt!

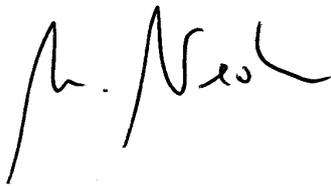
Der Fokus dieses Fachgruppentreffens lag dann auch weniger auf aktuellen wissenschaftlichen Ergebnissen als auf der Fachgruppe selbst. Die Sprecher berichteten über ihre jeweilige Amtszeit; der Schwerpunkt war dabei durchaus unterschiedlich. Den Beginn machte Georg Lausen, der die EMISA-Themen aus den späten 1980er Jahren vorstellte und aus heutiger Sicht kommentierte. Helmut Thoma erzählte, wie alles begann, damals in Tutzing ... Heinrich C. Mayr ergänzte um Anekdoten aus der Gründungs- und Anfangszeit der EMISA. Gottfried Vossen berichtete über seine Amtszeit in den 90er Jahren und beleuchtete auch aktuelle Entwicklungen, etwa zum Thema Web 2.0, bevor Andreas Oberweis die Zeit zu Beginn des neuen Jahrtausends aus Sicht der EMISA kommentierte. Schließlich war es an dem aktuellen Sprecher, den Reigen zu beenden; seine Kernaussage war, dass sich die EMISA-Fachgruppe als Netzwerk von an einem Thema interessierten Personen gegründet hat, zu einer Zeit, als Netzwerke keinerlei elektronisches Pendant hatten. Die Welt hat sich wenigstens in dieser Beziehung seitdem komplett geändert, die EMISA aber ist stabil geblieben.

Dies führt auf die Frage, wie sich die EMISA in Zeiten von sozialen Netzwerken und mannigfaltigen Möglichkeiten zum Austausch von Informationen und Wissen positionieren soll. Themen wie EMISA auf facebook, twitter oder XING stehen im Raum. Diesen Fragen wird sich die EMISA in diesem Jahr annehmen. Wertvollen Input haben wir auch von Peter Lockemann erhalten, der auf seiner Keynote beim Fachgruppentreffen zwar die Themen der EMISA unterstützt, aber eine Strukturierung und ein gemeinsames Forschungsszenario vermisst. Auch diesen Fragestellungen wird sich die EMISA stellen. Das Fachgruppentreffen 2011 in Hamburg bietet dazu einen idealen Rahmen; es wird von Markus Nüttgens am 22. und 23. September an der

Universität Hamburg organisiert, bitte vormerken! Bei Redaktionsschluss stand das wissenschaftliche Programm der EMISA'11¹ noch nicht fest, aber angesichts von fast 40 Beitragseinreichungen kann davon ausgegangen werden, dass die PC Chairs Barbara Weber, Oliver Thomas und Markus Nüttgens ein sehr gutes und spannendes Programm zusammenstellen werden.

Den inhaltlichen Schwerpunkt dieser Ausgabe bilden Fachartikel des EMISA-Arbeitskreises *Entwicklung Digitaler Bibliotheken*. Ein herzliches Dankeschön an die Autoren sowie an den AK-Sprecher Stefan Klink für die Zusammenstellung dieser Beiträge.

Mit den besten Wünschen,

A handwritten signature in black ink, appearing to read 'M. Weske', with a stylized, cursive script.

Mathias Weske
(EMISA-Sprecher)

¹ EMISA 2011: <http://www.wiso.uni-hamburg.de/conferences/emisa2011/>

Fachbeiträge des Arbeitskreises

Entwicklung Digitaler Bibliotheken

- **Benjamin Adrian und Michael Gillmann**

Besser Finden und Verstehen:

Das Semantische Web in Digitalen Bibliotheken

- **Kai Stalman, Reinhard Budde, Robert Mertens, Christoph Tornau, Dennis Wegener, Volker Heydegger, Thorsten Wunderlich, Bernd Ingenbleek und Florian Schulz**

Kultur im Wissensnetz:

Die Architektur der Deutschen Digitalen Bibliothek

- **Jan Brase und Ina Blümel**

Integration von nicht-textuellen Objekten in das GetInfo

Portal der Technischen Informationsbibliothek

Besser Finden und Verstehen

Das Semantische Web in Digitalen Bibliotheken

Benjamin Adrian¹, Michael Gillmann²

¹ Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH,
Knowledge Management, Trippstadter Straße 122, 67663 Kaiserslautern, Germany
Benjamin.Adrian@dfki.de

² Insiders Technologies GmbH, Brüsseler Straße 1,
67657 Kaiserslautern, Germany
M.Gillmann@insiders-technologies.de

Abstract. Bibliotheken erlauben es Benutzern nach Informationen in Form von Dokumenten zu suchen. Im Zeitalter der digitalen Bibliotheken und Information Retrieval wurden Computer genutzt, um mehr Dokumente zu verwalten, diese besser zu indexieren und sie somit auch besser auffindbar zu machen. Das Semantische Web ist ein Ansatz Informationen im WWW miteinander zu verknüpfen und sie mit semantischen Auszeichnungen zu versehen. Dies ermöglicht völlig neuartige Möglichkeiten der Indexierung und Suche nach Informationen. In diesem Papier werden Möglichkeiten erörtert die erfolgreichen Konzepte des Semantischen Web auf die Anwendung in digitalen Bibliotheken zu übertragen um dem Benutzer das Finden relevanter Information erheblich zu vereinfachen.

Keywords: Digitale Bibliotheken, Semantisches Web, RDF, RDFa, Semantische Indexierung, Semantische Annotation, Facettierte Suche.

1 Motivation

Bibliotheken sind ein Wissensschatz und beherbergen die Erkenntnisse der Menschheit über hunderte von Jahren hinweg in Dokumenten unterschiedlichster Form und Struktur. Einzelseiten sind ebenso zu finden wie Sammlungen, Zeitschriften, gebundene Bücher bis hin zu Mikrofiche. Hinzu kommen vermehrt digitale Dokumente, die ausschließlich elektronisch publiziert oder nachträglich digitalisiert wurden. Methoden der automatischen Indexierung und manuellen Kategorisierung versuchen diese Wissensflut zugreifbar und überschaubar zu halten. Basierend auf solchen Strukturen zeigt Abbildung 1 eine typische Benutzeroberfläche zur Recherche, hier von der Universitätsbibliothek Kaiserslautern. Es zeigt, dass der Benutzer versucht, seinen Informationsbedarf in Stichworten auszudrücken in der Hoffnung, dass diese Stichworte im Index oder Kategoriensystems enthalten sind. Zwar gibt es Bestrebungen, die Indexierung von Dokumenten zumindest für bestimmte Gebiete zu vereinheitlichen; dennoch ist die Qualität von Suchergebnissen

unmittelbar davon abhängig, wie gut und mit welchem Verständnis die Indexierung und Kategorisierung durchgeführt wurde.

Universitätsbibliothek Kaiserslautern

Einfache Suche Komplexe Suche Experten-CCL Indexsuche Neuerwerbungen Anschaffungsvorschlag

Einfache Suche

im Bestand ohne Aufsätze

Geben Sie ein Wort oder eine Wortfolge ein

Suchkriterium

Exakte Wortfolge? Nein Ja

Datenbank gesamter Medienbestand
 nur elektronische Medien
 Zeitschriften
 Lehrbuchsammlung

* Hauptfelder enthält u. a. Person, Titel, Körperschaft, Schlagwort, Erscheinungsjahr

Für die ausführlichere und ergänzende Suche - auch nach Aufsätzen - klicken sie [hier](#).

Abbildung 1. Eine typische Suchoberfläche eines OPAC

In dieser Arbeit zeigen wir neue Möglichkeiten auf, wie man Technologien des „Semantic Web“ einsetzen kann, um Benutzer in ihrer Recherche zu unterstützen.

Diese Arbeit gliedert sich wie folgt: Kapitel 2 fasst grob die gängigen Methoden des „Information Retrieval“ zusammen. Kapitel 3 erklärt Verfahren und Methoden des „Semantic Web“, die heute bereits im Internet etabliert sind und das Finden von relevanten Informationen für den Nutzer effizient unterstützen und erleichtern. Kapitel 4 gibt aktuelle und existierende Anwendungsbeispiele aus Wirtschaft und Forschung. In Kapitel 5 zeigen wir anhand exemplarischer Anwendungsfälle auf, wie die gezeigten Methoden auch auf die Domäne der Digitalen Bibliotheken überführt werden können, um auch hier ihren Nutzen zu entfalten. Kapitel 6 schließt die Arbeit mit einer Zusammenfassung und einem Ausblick auf zukünftige Arbeiten ab.

2 Traditionelle Lösungsansätze

Es gibt bereits zahlreiche Ansätze, um Diskrepanzen zwischen dem explizierten Informationsbedarf des Benutzers mit den existierenden Indexierungen zu verringern.

Sprachliches Hintergrundwissen wird verwendet um unterschiedliche Schreibweisen von Wörtern bedingt durch Deklinationsformen, Flexionsformen, Singular-Plural oder Synonymen zu vereinheitlichen.

Basierend auf Relevanzfeedback und/oder Statistiken innerhalb der Suchhistorie eines Benutzers oder einer ähnlichen Benutzergruppe können Benutzeranfragen automatisch angereichert werden, um bessere Suchresultate zu generieren. Dies geschieht oft durch eine erneute Sortierung der Suchergebnisliste für die ursprüngliche Suchanfrage.

Im Unterschied zu diesen Methoden, die auf Statistiken und vor allem auf Worthäufigkeiten aufbauen, stellen wir Methoden vor, wie man durch die Hinzunahmen von Semantik die Beschreibungen von Dokumenten und Suchanfragen sowohl für den Menschen als auch für die Maschine verständlicher gestaltet.

3 Das Semantische Web

Im Jahr 2002 beschreibt Tim Berners-Lee die Erweiterung des WWW, genannt das „Semantic Web“. Anstatt dass Webseiten ihre Daten in natürlicher Sprache, also nach dem aktuellen Stand der Technik lediglich für den Menschen verständlich darstellen, werden im „Semantic Web“ die Bedeutung der Daten explizit und für die Maschine verständlich ausgezeichnet. Dies geschieht mithilfe des „Resource Description Framework“ RDF.

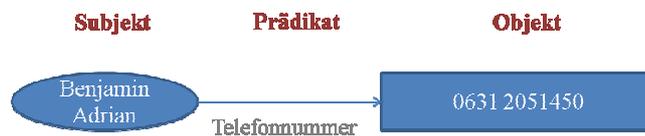


Abbildung 2. Ein einfacher Sachverhalt

Abbildung 2 zeigt, wie man in RDF den Sachverhalt von folgendem Text modelliert:

Benjamin Adrian
Tel.: 06312051450

Abbildung 3. Umgangssprachlicher Ausdruck dieses Sachverhalts

Das Wichtigste an RDF ist, dass all das was wir in unserer Welt erfassen können, in RDF Ressourcen sind. Ressourcen werden über „International Resource Identifier (IRI)“ (internationalisierte Form von URI) identifiziert. Alles kann somit durch eine IRI angesprochen werden. Der Person Benjamin Adrian wird von uns die IRI <http://www.dfki.uni-kl.de/~adrian#me> vergeben. Den Sachverhalt, dass etwas eine Telefonnummer hat, identifizieren wir mit der IRI <http://xmlns.com/foaf/0.1/phone>. Zusätzlich identifizieren wir Namen mit <http://xmlns.com/foaf/0.1/name>. RDF ermöglicht es uns nun Ausdrücke der Form Subjekt-Prädikat-Objekt aufzuschreiben:

```

<http://www.dfki.uni-kl.de/~adrian#me>
<http://xmlns.com/foaf/0.1/phone>      „06312051450“;
<http://xmlns.com/foaf/0.1/name>      „Benjamin Adrian“.

```

Abbildung 4. RDF Ausdruck des Sachverhalts (geschrieben in TURTLE-Syntax)

In Bezug auf die Beschreibung von Dokumentinhalten liegen die Vorteile dieser Repräsentation in der Entkopplung von sprachlichen Phrasen und der vom Autor gemeinten Bedeutung (Semantik). Zudem können Inhalte von Dokumenten, die RDF Metadaten bereitstellen, von Maschinen verstanden werden. RDF-Daten lassen sich in verschiedenen Formaten (RDF/XML, N3, NTriples, Turtle) speichern.

Es verbleibt das Problem, dass identische Dinge mit unterschiedlichen IRIs identifiziert werden. Durch die Verwendung von Ontologien als gemeinsames Verständnis einer Konzeption kann dem begegnet werden. Für die Repräsentation von Ontologien wurden die Sprachen RDF Schema und die „Web Ontology Language“ (OWL) entwickelt. In solchen Sprachen geschriebene Ontologien vereinheitlichen Konzepte wie „hat einen Namen“ oder „hat eine Telefonnummer“. Die „Friend-of-a-Friend“-Ontologie (FOAF) ist ein Beispiel hierfür. Doch Ontologien bieten noch mehr. Sie können Klassen beschreiben, die gemeinsame Eigenschaften von Ressourcen (man spricht in diesem Fall von Instanzen) zusammenfassen.

```

<http://www.dfki.uni-kl.de/~adrian#me>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://xmlns.com/foaf/0.1/Person> .

```

Abbildung 5. Benjamin Adrian ist eine Person im Sinne der FOAF-Ontologie

Abbildung 5 definiert, dass die Ressource mit Namen Benjamin Adrian eine Person im Sinne der FOAF-Ontologie ist. Eine Maschine kann daraus Schlüsse ziehen, welche Attribute (z.B. Vorname, Nachname) oder Beziehungen zu weiteren Instanzen diese Person noch besitzen kann.

```

<http://www.dfki.uni-kl.de/~adrian#me>
<http://xmlns.com/foaf/0.1/knows>
<http://www.insiders.com/~gillmann#me>

```

Abbildung 6. Beziehungen zwischen Instanzen

Die FOAF Ontologie definiert, dass die Beziehungen „kennt“ lediglich zwischen Personen bestehen darf. Falls eine Maschine nun den Ausdruck in Abbildung 6 liest, kann sie daraus schließen, dass es sich bei beiden Instanzen um Personen der FOAF-Ontologie handelt. Darüber hinaus lassen sich in RDFS und OWL Hierarchien von Klassen und Eigenschaften modellieren. OWL bietet in Vergleich zu RDFS eine Vielzahl von Regeln und Restriktionen an, die von Maschinen beim Verarbeiten von RDF-Daten, berücksichtigt werden können.

Im Jahr 2006 schlug Tim Berners-Lee vor, die bestehenden „Semantic Web“ Techniken anzuwenden, um bereits existierende Daten im WWW als „Linked Data“

zu publizieren. Hierbei zeichnet sich „Linked Data“ dadurch aus, dass Ressourcen durch IRIs identifiziert werden die durch das HTTP-Protokoll als URLs interpretiert und de-referenziert werden können. Somit wird es ermöglicht mittels einer IRI Daten über diese Ressource zu bekommen. Diese Daten sollten in RDF vorliegen, die Ressource bestmöglich beschreiben und Verknüpfungen (engl. links) zu weiteren Ressourcen enthalten, die weitere Informationen darüber beinhalten oder in einer Beziehung zu dieser Ressource (z.B. kennt) stehen. Diese einfachen Richtlinien führten zu einer Welle von als „Linked Open Data“ publizierten Datensätze die von DBpedia, einer maschinenverständlichen Version von Wikipedia bis hin zu BBC Programms, Webseiten der BBC mit hinterlegten RDF-Daten über das Fernsehprogramm oder Musikinterpreten reicht.

Im Jahr 2009 wurde die RDFa Arbeitsgruppe im W3C gegründet. RDFa bietet Möglichkeiten RDF-Daten direkt in HTML-Seiten einzubetten. Mittlerweile beschäftigt sich bereits das „Open Document“-Gremium von OASIS mit der Nutzung von RDFa in diesem offenen Dokumentformat.

```
<div about=" http://www.dfki.uni-kl.de/~adrian#me">
  <span property="foaf:name">Benjamin Adrian</span>
  <span>Tel.:</span>
  <span property="foaf:phone">06312051450</span>
</div>
```

Abbildung 7. Bekannter Sachverhalt in RDFa und HTML

Im Folgenden wird anhand von Anwendungen beschrieben, wie die Nutzung von „Semantic Web“-Technologien Möglichkeiten bietet die Suche und das Finden von Informationen in Digitalen Bibliotheken zu unterstützen.

4 Anwendungen

Die folgenden Anwendungen nutzen existierende und für den Computer verständliche Informationen, um für den Benutzer die Informationsrecherche auf Basis von Textdokumenten zu erleichtern.

4.1 Semantische Indexierung

Im Bibliotheksbereich werden mit (inversen) Wortlisten Dokumente Begriffen und Begriffen Dokumente zugeordnet. Durch diese Indexierung wird es ermöglicht mit einfachen Suchbegriffen Dokumente zu finden, in denen diese Begriffe eine signifikante Rolle spielen. Falls die benutzten Begriffe im Index Teil eines kontrollierten Vokabulars sind, spricht man auch von Schlüsselwörtern oder Kategorien.

Bei der semantischen Indexierung werden anstatt von flachen Wörtern, semantische Konzepte als indexierende Begriffe verwendet. Zum Beispiel stellt die DBpedia eine in RDF formalisierte Konzeptionalisierung von Wikipedia-Inhalten zur

Verfügung. Hierbei wird zu jedem Wikipedia-Artikel eine Instanz angelegt, die mit den weiteren im Artikel vorkommenden Instanzen semantisch verknüpft werden. Der Wikipedia-Artikel über Angela Merkel enthält in der DBpedia somit eine Instanz über Angela Merkel und weitere Konzepte die im Artikel Erwähnung finden (CDU, Kanzler, Deutschland, etc.). Jede Instanz ist hierbei mit beschreibenden Klassen einer bestehenden Hierarchie (z.B. Person, Ort, Partei) typisiert. Verwendet man nun diese Instanzen in einem semantischen Index erhält man die Möglichkeit konkretere Informationsanfragen zu stellen: „Ich bin interessiert an Dokumenten in denen die Kanzler von Deutschland sich in den USA aufhielten.“

Im Fall von Wikipedia und DBpedia existiert dieser Index bereits. Für Dokumente außerhalb dieser beiden Systeme ist es dennoch möglich Konzepte der DBpedia als Indexierungsgrundlage zu verwenden. Eine Ontologiebasierte Informationsextraktion kann mit einem RDF Datensatz initialisiert werden und erkennt fortan, welche Instanzen und Beziehungen zwischen Instanzen innerhalb dieses Datensatzes in Textpassagen vorkommen. Das System SCOOBIE¹ (Service for Ontology-based Information Extraction) erstellt auf Basis von einem gegebenen RDF Datensatzes und einem Dokument einen neuen RDF-Graphen, in dem exakt die Instanzen, Klassen, und Beziehungen enthalten sind, die im Datensatz existieren und die im Dokument erwähnt wurden. Eben dieser RDF Graph kann als ein semantischer Index über das ursprüngliche Dokument betrachtet werden.



```
</music/artists/0383d9df-2a4c-4d10-a46a-e9e041da8eb3#artist>
  rdf:type mo:MusicGroup ;
  rdf:type mo:MusicArtist ;
  foaf:name „Queen“ .
```

Abbildung 8. Die BBC stellt eine semantische Indexierung von eigenen Webseiten über Musikgruppen zur Verfügung. (siehe <http://www.bbc.co.uk/music/artists/0383d9df-2a4c-4d10-a46a-e9e041da8eb3> und <http://www.bbc.co.uk/music/artists/0383d9df-2a4c-4d10-a46a-e9e041da8eb3.rdf>)

4.2 Semantische Annotationen

Die Semantische Indexierung bietet eine maschinenlesbare Beschreibung eines Dokumentes in der formalen RDF Beschreibungssprache an. Diese Informationen sind wertvolle Metadaten über den Dokumentinhalt und können als Grundlage dienen zu weiteren Dokumenten zu gelangen, in denen konkrete Instanzen ebenfalls erwähnt werden, oder weitere Informationen zu einer Instanz innerhalb einer Textpassage zu

¹ Unter <http://projects.dfki.uni-kl.de/scoobie/> werden in nächster Zeit SCOOBIE und seine Funktionalitäten erläutert werden.

erfragen. Hier bieten Semantische Annotationen die Möglichkeit maschinenlesbare Informationen zu Instanzen in Texte einzubetten. Im Fall von Dokumenten in einem XML oder HTML-Format und existierenden Informationen die in RDF repräsentiert sind, kann die Sprache RDFa verwendet werden um den DOM (das in HTML und XML verwendete Dokumentobjektmodell) mit RDF Daten anzureichern. Das System Epiphany² bietet hier die Funktionalität RDF Graphen über Dokumentinhalte dem Dokument in Form von RDFa Annotationen wieder zurückzuführen. Darüber hinaus erlaubt Epiphany für an Textpassagen annotierte Instanzen weitergehende Informationen anzufragen, sofern Informationen über diese Instanzen über „Linked Data Prinzipien“-publiziert wurden, so dass man mit HTTP-Anfragen an die URI der Instanz RDF Inhalte geliefert bekommt.

Tim Burton

Timothy „Tim“ Walter Burton (* 25. August 1958 in Burbank, Kalifornien) ist ein US-amerikanischer Autor, Produzent und Filmregisseur.

Inhaltsverzeichnis [Verbergen]

- 1 Leben
- 2 Privatleben
- 3 Betrachtung zum Werk
 - 3.1 Stil
 - 3.1.1 Wiederkehrende Motive
 - 3.1.2 Stammbesetzung
- 4 Filmografie
 - 4.1 Regisseur
 - 4.2 Produzent
 - 4.3 Schauspieler
 - 4.4 Auszeichnungen
- 5 Weiterführende Informationen
 - 5.1 Literatur
 - 5.2 Weblinks
- 6 Einzelnachweise



Tim Burton (2008)

Leben
[Bearbeiten]

The Island of Doctor Agor Disney-Studios studierte er drei Jahre Trickfilmkunst am California Institute of the Arts *Der Herr der Ringe* mit. Dort wirkte er u. a. bei den Filmen *Cap und Capper* sowie *Taran und der Zauberkesel* und *Tron* Kurzfilm *Vincent*. Tim Burton entwickelte schon als Kind ein zeichnerisches Talent. Seinen ersten Film , , drehte er 1971 mit 13 Jahren. Mit einem Stipendium der und wurde anschließend bei Disney angestellt. Vorher arbeitete Tim Burton jedoch noch als Zeichner von Zwischenphasen beim 1978 erschienenen Fantasy-Zeichentrickfilm mit. 1982 drehte er als Autor und Regisseur in Disneys Auftrags einen ersten erfolgreichen und preisgekrönten

Abbildung 9. Visualisierung von Semantischen Annotationen in einem Wikipedia-Artikel über den Filmregisseur und Produzenten Tim Burton

² Siehe auch <http://projects.dfki.uni-kl.de/epiphany/> um weitere Informationen über Epiphany zu erhalten.

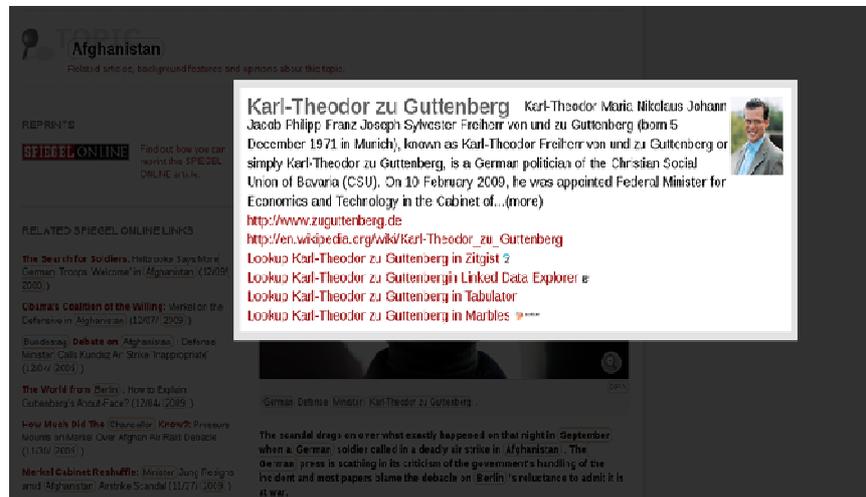


Abbildung 10. Zu einer in einem Dokument gefundenen und semantisch annotierten Instanz über Karl-Theodor zu Guttenberg werden weitere Informationen aus der DBpedia angeboten.

4.3 Facettierte Suche

Ausgehend von Funktionalitäten der Systeme SCOOBIE zur semantischen Indexierung von Dokumenten und Epiphany zur semantischen Annotation von Dokumenten kann man die allgemeine Dokumentsuche in Systemen Google oder aber auch Bibliotheksspezifischen OPAC-Systemen anreichern mit einer Perspektive, die sich auf einen verwendeten RDF-Datensatz begründet. Das am DFKI entwickelte System Sterntaler verwendet beispielsweise in RDF repräsentierte Daten vom Produktkatalog des Internetversandhändlers Amazon, um den Benutzer bei einer Produktsuche in der Suchmaschine Google auch innerhalb von Dokumenten zu unterstützen. Abbildung 11 zeigt hierbei die erweiterte Version der minimalistischen Suchmaske von Google durch das Sterntalersystem. Zu der gegebenen Suche mit den Begriffen „tim burton filme“ lieferte Google eine Liste von gefundenen Dokumenten. Die semantische Indexierung erkannte nun in jedem Dokument eine Reihe von Instanzen aus dem Amazon-Datensatz. Die gefundenen Instanzen haben hierbei eine Reihe von Facetten (hier Preis, Bewertung, Titel, Anzahl von Benutzerkommentaren), die genutzt werden um einen generischen Filter zu erstellen. Durch diesen Filter kann der Benutzer solche Dokumente suchen, in denen Instanzen bestimmte Attributwerte haben (z.B. den Titel „Alice im Wunderland“). Das Sterntalersystem reduziert entsprechend die Resultatliste. Entscheidet sich der Benutzer schließlich für ein Dokument, so wird dieses Dokument mit allen in ihm gefundenen Instanzen semantisch annotiert durch das Epiphany-System aufbereitet

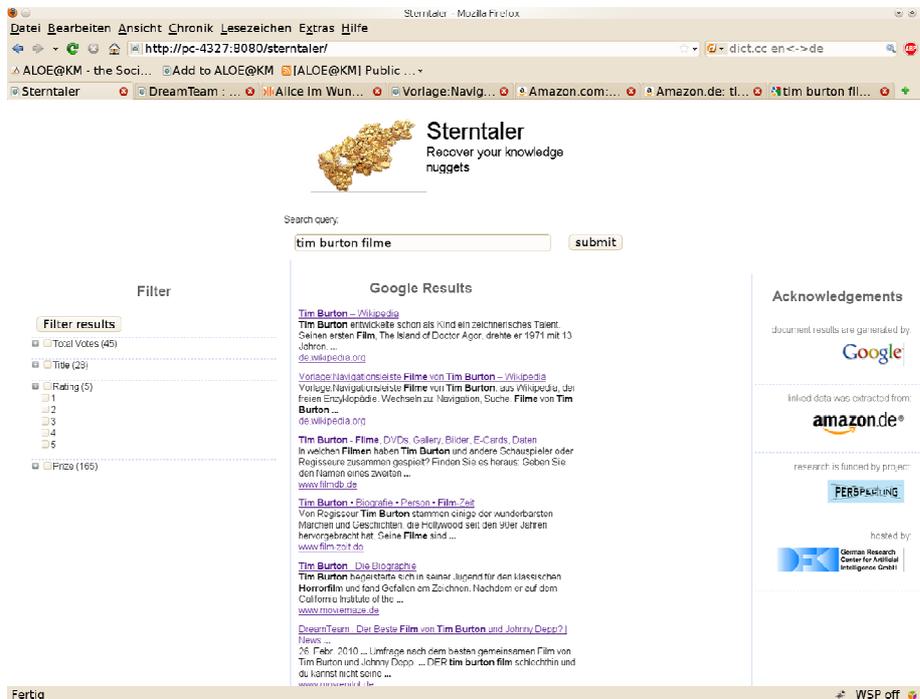


Abbildung 11. Die generische Facettierte Suche von Sterntaler

5 Semantik in Bibliotheken

Wie in den voran gegangenen Kapiteln gezeigt haben sich Ansätze und Techniken wie RDF in anderen Domänen, insbesondere dem Semantischen Web, über Jahre hinweg etabliert. Die Dokumente in digitalen Bibliotheken weisen vergleichbare Charakteristiken auf wie Web-Seiten im Internet:

- Sie enthalten interessante und potentiell relevante Informationen verschiedenster Art in verschiedensten Strukturen und Darstellungsformen.
- Sie wollen gesucht und gefunden werden.
- Sie sind leider nur manchmal explizit untereinander verlinkt (z.B. über Literaturangaben / Citations), obwohl sie oft miteinander „verwandt“ sind, da sie ein ähnliches oder das selbe Thema behandeln.
- Die verwendete Terminologie ist nicht immer einheitlich oder sie sind in unterschiedlichen Sprachen verfasst, wodurch die „Verwandtschaft“ oder Ähnlichkeit nicht ausschließlich über Thesauri oder andere konventionelle Methoden erkennbar ist.

Semantische Methoden lösen diese Probleme bereits für die „Dokumente im Internet“, der „größten Digitalen Bibliothek, die es gibt“. Entsprechend wird auch ihre Anwendung auf die in digitalen Bibliotheken abgelegten Dokumente Mehrwerte bieten.

Insbesondere bei der Online-Recherche versprechen sie Gewinn bringende Vorteile hinsichtlich der Bedienbarkeit und Kundenzufriedenheit durch schnelle und zielgerichtete Sucherfolge. Ein Nutzer der Online-Recherche kann unmöglich alle Eventualitäten und Rahmenbedingungen „erraten“, die zu der Auswahl von Index-Werten aller für ihn relevanten Dokumente geführt haben, insbesondere dann nicht, wenn diese Dokumente über mehrere Bibliotheken, eventuell sogar international verteilt vorliegen. Sind die Dokumente allerdings mit Methoden des Semantischen Web annotiert und miteinander verknüpft, ermöglichen ihm RDFa Annotationen in den Trefferlisten ein zielgerichtetes Hinsteuern zum gesuchten Dokument über Anwendungs- und damit Bibliotheksgrenzen hinweg.

Ein weiteres Problem heute ist, dass der Benutzer gegebenenfalls verwirrt wird, wenn ihm Treffer angeboten werden, die er selbst gar nicht erwartet hätte, oder die aus seiner Sicht nicht relevant sind. Dieses Verständnis ist aber unabdingbare Bedingung, wenn er seine Suche präzisieren oder korrigieren möchte. Nur wenn er versteht, warum aus seiner Sicht falsche Dokumente gefunden werden, kann er beim weiteren Suchen die Präzision erhöhen. Durch die semantische Facettierte Suche kann der Benutzer die Treffermenge zielgerichtet nach seinen Bedürfnissen und seinem Verständnis einschränken beziehungsweise Filtern, ohne dabei semantische Beziehungen zu verlieren.

Schließlich ist es sogar denkbar, dass sich ein Benutzer dank semantischer Verknüpfungen zwischen Dokumenten (oder sogar einzelnen Kapiteln der Dokumente) virtuelle Bücher „bindet“, indem er ein Thema vorgibt (= semantisch beschreibt) und ein System ihm die Dokumente (oder Kapitel), die zu dem Thema in semantischem Bezug stehen, zur Auswahl anbietet.

6 Schlussfolgerung und Ausblick

In dieser Arbeit präsentieren wir, wie im Internet etablierte und erfolgreiche Methoden und Konzepte des Semantischen Web auf die Anwendung in digitalen Bibliotheken übertragen werden können um dem Benutzer das Finden relevanter Information erheblich zu vereinfachen. Insbesondere nehmen wir dabei auch darauf Bezug, wie die Methoden dem Benutzer helfen zu verstehen, wieso die gefundenen Dokumente als relevant einzustufen sind. Dazu beschreiben wir die in Frage kommenden Methoden und Konzepte anhand von konkreten Beispielen, bei denen sie bereits heute im Internet zum Einsatz kommen. Abschließend zeigen wir beispielhaft Use Cases auf, in denen die Anwendung eben dieser Methoden und Konzepte auch in der Domäne der digitalen Bibliotheken Benutzern wie Betreibern einen Mehrwert insbesondere im Sinne einer effizienteren Suche von digitalisierten Dokumenten beschert. Effizienter wird das Arbeiten dadurch, dass es zum einen intellektuell vereinfacht wird und zugleich zum zweiten bessere Ergebnisse im Sinne von mehr und genaueren Treffern verspricht.

Zusammengefasst ist diese Arbeit ein Plädoyer, Methoden und Konzepte des Semantischen Webs in die Entwicklung von Werkzeugen für digitale Bibliotheken mit einzubeziehen. Forschungsunternehmungen, die sich dieser Sache bereits heute angenommen haben, finden sich bei der W3C Library Linked Data Incubator Group³ sowie der Deutschen Nationalbibliothek (DNB)⁴.

Die systematische Analyse von Anforderungen durch das Sammeln und Durchleuchten weiterer Use Cases sowie die darauf hin gerichtete Umsetzung von Werkzeugen auf Basis von Methoden und Konzepten des Semantischen Web bilden entsprechend die nächsten Schritte der zukünftigen Arbeit.

³ <http://www.w3.org/2005/Incubator/ld/>

⁴ Linked Data Services der DNB: <https://wiki.d-nb.de/display/LDS/>

Kultur im Wissensnetz - Die Architektur der Deutschen Digitalen Bibliothek

Kai Stalman, Reinhard Budde, Robert Mertens, Christoph Tornau, Dennis Wegener, Volker Heydegger, Thorsten Wunderlich, Bernd Ingenbleek und Florian Schulz

Fraunhofer Institut für Intelligente Analyse- und Informationssysteme IAIS
Schloss Birlinghoven
53754 St. Augustin
`{vorname.nachname}@iais.fraunhofer.de`

Zusammenfassung Die Deutsche Digitale Bibliothek (DDB) ist ein IT-Leuchtturmprojekt des Bundes und der Länder. Die DDB soll das Kulturerbe und die wissenschaftlichen Informationen von - in letzter Ausbaustufe - bis zu 30.000 Einrichtungen mit geschätzten 300 Millionen Objekten zusammenzuführen und über das Internet öffentlich zugänglich zu machen. Die DDB wird als erste länder-, disziplin- und spartenübergreifende Plattform in Deutschland zeitgemäße Recherche- und adäquate Präsentationstechniken für unterschiedlichste digitale Objekte mit heterogenen Metadaten bieten. Fraunhofer IAIS ist an dem Projekt als technischer Projektleiter beteiligt und verantwortet den Aufbau der zentralen Infrastruktur, in der die Metadaten und Inhalte verwaltet werden. Wenn dieser Schritt Ende 2011 abgeschlossen ist, können die Daten der Einrichtungen nach und nach erschlossen werden. Parallel muss die Dienstplattform ausgebaut werden. Kriterien der Dienstqualität wie Sicherheit, Skalierbarkeit, Performanz, Interoperabilität, Nutzerfreundlichkeit sind für den Erfolg der DDB ebenso essentiell wie die Signifikanz des Gesamtangebots, die von den Alleinstellungsmerkmalen der DDB geprägt ist. Das Papier stellt technische Konzepte und Lösungsansätze für das geplante Informationssystem vor und beleuchtet für die Realisierung relevante Teilaspekte.

Keywords: Deutsche Digitale Bibliothek, Architektur der Dienstplattform, Metadaten Mapping, Digitales Repository, Linked Data, Ontologie für Kulturgut

1 Einleitung

Die zahlreichen Digitalisierungsprojekte, die auf nationaler Ebene beispielsweise von der Deutschen Forschungsgemeinschaft (DFG)¹, dem Bundesministerium für Bildung und Forschung (BMBF)², dem Bundesbeauftragten für Kultur und

¹ <http://www.dfg.de/>

² <http://www.bmbf.de/>

Medien (BKM)³ und auf europäischer Ebene etwa von dem EFRE Programm⁴ und dem 7. Forschungsrahmenprogramm⁵ getragen wurden und werden, schlagen die Brücke von der Ära des Buches zur Linked-Data-Cloud. Für Nutzer stellt die ortsunabhängige Recherchemöglichkeit, die häufig den direkten Zugriff auf digitale Informationsobjekte wie gescannte Bücher, Audiofiles, Bilder und Filme einschließt, eine Bereicherung dar. Für institutionelle Informationsanbieter bietet die Veröffentlichung der Inhalte im Internet ebenfalls signifikante Chancen. Nutzer könnten mit Bibliothekaren, Archivaren oder Kuratoren über das Netz direkt bzw. über Annotationen am Content in Kontakt treten und Korrekturen und Ergänzungen zu eingestellten Inhalten beitragen. Welche Dynamik in diesem Modell liegt, zeigt Wikipedia oder - weniger bekannt, doch enorm erfolgreich - das Portal der Australischen Zeitungen[HO09], das nicht nur gestattet, die digitalisierten Inhalte auch weit zurückliegender Jahrgänge zu durchsuchen, zu lesen und als PDF herunterzuladen, sondern Nutzer ermutigt, Artikel anzureichern, zu kommentieren, zu korrigieren. In Deutschland hat beispielsweise das Bundesarchiv die informationstechnischen Möglichkeiten ebenfalls sehr weitsichtig genutzt. Bestandsübersichten, Findbücher und teilweise die Archivmaterialien selbst sind auf der Website des Bundesarchives recherchierbar. Ressourcen werden außerdem über Kooperationen mit Wikipedia der Öffentlichkeit zugänglich gemacht[BA08]. Einige Inhalte sind teilweise volltextindexiert.

In Kapitel 2 werden verwandte Arbeiten und Projekte aus dem Umfeld der digitalen Bibliotheken vorgestellt. Kapitel 3 beschreibt funktionale und nicht-funktionale Anforderungen an die DDB, während in Kapitel 4 auf Lösungsansätze eingegangen wird. Kapitel 5 beschäftigt sich mit der technischen Architektur und dem Datenfluss. Schließlich ist eine Zusammenfassung und ein Ausblick in Kapitel 6 zu finden.

2 Verwandte Arbeiten

Bibliotheken, Archive und Museen wachsen und passen sich an eine sich verändernde Welt an. Diese Institutionen sind die wichtigsten Aktionäre des kulturellen Erbes. Im Jahr 2004 startete Google das „Google Book Project“, welches die Portierung jedes Buches in jeder Sprache in die Google-Digital Library zum Ziel hat. In Folge dieser andauernden Tätigkeit wurden viele öffentlich gegründete Digitalisierungsprojekte und digitale Bibliotheken auf den Weg gebracht. Neuere Beispiele sind die Max Planck Digital Library [RO07] oder die Biodiversity Heritage Library[HS10]. Eines der ersten digitalen Bibliotheks-Projekte ist das Internet Archive[JK09], welches privat gegründet wurde und dessen Ursprung

³ <http://www.bundesregierung.de/Webs/Breg/DE/Bundesregierung/BeauftragterfuerKulturundMedien/beauftragter-fuer-kultur-und-medien.html>

⁴ http://europa.eu/legislation_summaries/employment_and_social_policy/job_creation_measures/l60015_de.htm

⁵ http://cordis.europa.eu/fp7/home_de.html

zurück geht bis ins Jahr 1996. Nationalbibliotheken verfolgt nach und starteten das crawlen von digitalen Inhalten von Webseiten. Eines der prominentesten europäischen digitalen Bibliotheks-Projekte ist Europeana[CGS09], welches im Jahr 2005 initiiert wurde und derzeit Metadaten aus Aggregatoren mit Links zu rund 10 Millionen digitalen Objekten aus ganz Europa enthält.

Die Fachinformationszentren in Deutschland, von denen einige Technologie- und Dienstleistungsanbieter mit wettbewerbsfähigen Rechenzentren und Entwickler-Teams haben, entwickelten gemeinsam vascoda.de [PT03], ein Portal mit rund 75 Millionen Datensätze aus dem Bereich wissenschaftlicher Literatur im Internet. Vascoda, technisch gehostet bei der Technischen Informationsbibliothek und Universitätsbibliothek Hannover, trägt zur WorldWideScience.org [HJ08] bei, eines US-Web-Gateway, welcher einen föderierten Such-Dienst auf weltweit verteilten Datenbanken mit mehr als 200 Millionen Datensätzen zur Verfügung stellt. Ein erster Versuch, unterschiedliche Datenquellen von Kulturinstitutionen zu integrieren, ist das BAM-Portal [MA02]. Es ermöglicht die Suche in den Sammlungen der teilnehmenden Bibliotheken, Archive, Museen und anderen Quellen in Deutschland. Bis heute verwaltet dieses Portal mehr als 45 Millionen Objekte, etwa 1 Million dieser Objekte sind in digitaler Form verfügbar.

Projekte, die einen ähnlichen technischen Fokus haben sind Fedora [PL98], ein OpenSource-Repository zur Verwaltung von digitalem Inhalt, und eSciDoc [RA09], eine e-Research Kollaborations-Umgebung für wissenschaftliche Communities darstellt, die ihrerseits auf Fedora basiert. Der wesentliche Unterschied zum technischen Ansatz der DDB ist, dass diese Projekte zwar eine Ablage von Daten ermöglichen, deren Metadaten jedoch nicht zu einem Wissensnetz zusammengesetzt werden. Weiterhin sind sie teilweise auf kleinere Datenmengen ausgelegt, weshalb ihre Systemarchitektur nicht skalierbar gestaltet wurde.

3 Anforderungen

Eine Digitalisierung kann zur Sicherung von Wissen dienen. Für das DDB Projekt steht jedoch zunächst das Ziel der Erreichbarkeit und Nutzbarkeit der Inhalte im Vordergrund. Daher ist die Qualität der DDB-Suchmaschine von größter Bedeutung. Aus Nutzersicht spielt die Bedienbarkeit und insbesondere die Aufbereitung der Such-Ergebnislisten einschließlich der Filter – auch Facetten oder Navigatoren genannt – eine wichtige Rolle. Für eine wissenschaftliche Nutzung müssen die von der DDB verwalteten Quellen vor allem zitierfähig sein, was die Frage nach persistenten Kennungen aufwirft. Aus Anbietersicht erscheint eine den jeweiligen Inhalten angemessene Darstellung als vorrangiges Ziel.

Im Folgenden sind weitere Anforderungen an die DDB im Detail aufgestellt (vgl. auch [MC08]).

3.1 Sicherheit, Zuverlässigkeit und hohe Skalierbarkeit

Neben einer Fülle funktionaler Anforderungen wird von der DDB Plattform vor allem Sicherheit, Zuverlässigkeit und hohe Skalierbarkeit erwartet. Ein Men-

gengerüst kann gegenwärtig nur durch Vergleiche mit ähnlichen Projekten erschlossen werden. Unter geeigneten Voraussetzungen kann die DDB nach drei Betriebsjahren 20 Millionen Objekte enthalten, die an eine Million Nutzer pro Tag ausgeliefert werden. Für das Storage-System lässt sich daraus eine durchschnittliche Last von ca. 2.500 IO Operationen pro Sekunde ableiten, eine Zahl, die in Spitzen auf mehr als das Hundertfache anwachsen kann. Das Gleiche gilt für den Durchsatz, der im Mittel ca. 60 MB pro Sekunde betragen könnte, aber in Spitzen auch bei einem Gigabyte pro Sekunde liegen kann. Mit Blick auf die anzunehmenden konstanten Wachstumsraten bei Datenmenge, Durchsatz und Nutzerzahlen ist davon auszugehen, dass alle relevanten Systemkomponenten für eine elastische Skalierung durch Verteilung auf mehrere Rechner, eine Cloud bzw. einen verteilten Storage ausgelegt sein müssen. Suchmaschinen, bzw. Indexierer wie Solr unterstützen Verteilung und sind für parallele Abfragen ausgelegt. Problematisch ist, dass die DDB konstant in großer Menge Daten importieren muss, während das System im Betrieb ist. Das System muss daher so modular aufgebaut sein, dass eine zentrale Komponente wie der Indexer ausgetauscht werden, wenn in einigen Jahren Datenmengen und Dateizuwachsraten erreicht werden, die Solr überfordern.

3.2 Datenquellen

Die Dienste der DDB beruhen auf Metadaten zu digitalen Objekten und Digitalisaten bzw. Derivaten der Digitalisate. Binärer Content (Bilder, Audiodateien, Filme etc.) wird in der DDB alternativ aus einem eigenen Repository an den Browser des Nutzer geliefert oder über Verlinkung auf ein Repository der Einrichtung, die diese Daten hält, bereit gestellt. Die Einrichtungen, aus denen die DDB Daten bezieht, liefern Metadaten sehr unterschiedlicher Auszeichnungstiefe. Die Metadaten sind in verschiedenen Formaten kodiert, zu den populäreren zählen Marc [MARC], EAD [EAD], museumdat [MUDA], METS [METS], MODS [MODS], Lido [LIDO], DC [DC08] und CIDOC CRM [CCRM]. XHTML mit RDFa[RDF] Annotationen kann in diesem Zusammenhang ebenfalls angeführt werden. Probleme bereitet teilweise der Informationsverlust, wenn Eingangsdaten zum Beispiel im Falle von DC als Feldinhalte Literale – bei `dc:creator` etwa den Namen des Autors – senden, obwohl intern in den Ursprungsdaten Verweise auf eine Normdatei vorhanden sind wie beispielsweise die Personennormdatei der Deutschen Nationalbibliothek. Innerhalb der DDB werden die aus den Metadaten extrahierten Informationen über die digitalen Objekte zweckmäßig vereinheitlicht und als RDF Triples gespeichert, auch um Queries über transitive Hüllen zu berechnen. Zum Zweck der besseren Indexierung werden die Properties der Triples auf ein kompaktes Vokabular reduziert.

3.3 Metadaten

Die skizzierten Metadatenformate dienen vor allem der semantischen Beschreibung der Inhalte. Neben deskriptiven Aspekten spielen eine Reihe anderer Aspekte eine Rolle. Die von Einrichtungen erzeugten Informationsobjekte unterliegen

oft lizenzrechtlichen Beschränkungen⁶. Solche Informationen werden mit den Metadaten übertragen und müssen von den Einrichtungen eingepflegt werden. Die Beschränkungen können sich auf ein gesamtes Informationsobjekt beziehen⁷, sie können aber auch eingebettet sein und sich z.B. auf einzelnen Seiten eines Buches beziehen⁸. Ein Buch kann als gemeinfrei lizenziert sein, während bestimmte Abbildungen in dem Buch durch Bildrechte geschützt sind. Gegenwärtig sind abgelegte lizenzrechtliche Beschränkungen oft unvollständig. Zum Teil ist die Rechtesituation unklar. Bei sogenannten verwaisten Werken geht man davon aus, dass eine Beschränkung nicht besteht, ohne letzte Gewissheit zu haben. Für die DDB wird zurzeit auf politischer Ebene versucht, eine Lösung hierfür zu finden.

Eine maschinelle Interpretation z.B. der lizenzrechtlichen Auszeichnung setzt voraus, dass Inhalte auch strukturell beschrieben werden, damit klar ist, worauf sich Rechteangaben genau beziehen (das ganze Buch, die hochauflösenden Scans, alle oder einzelne Abbildungen etc.). Die populärste Form der maschinellen Auswertung von strukturellen Metadaten ist die Nutzung eines Viewers oder Players, der binäre Datei-Inhalte präsentiert, indem er sie hörbar oder sichtbar macht. Mets ist ein z.B. für Bibliotheksinhalte häufig verwendeter Container, der dem Nutzer eine sinnvolle Navigation über gescannte Seiten und Scan-Derivate verschiedener Auflösung erst ermöglicht. Viewer wie der browserbasierte DFG Viewer interpretieren Mets direkt und navigieren über den in einem Repository liegenden Digitalisat und Derivat z.B. eines gescannten Buchs. Bei der Anzeige kann unter ggf. vorliegenden, verschiedenen Auflösungen gewählt werden. In Mets können als inneres Xml andere Container eingebettet sein, die deskriptive oder administrative Metadaten enthalten. Mets eignet sich trotz der eher fragwürdigen systematischen Trennung von administrativen, strukturellen und deskriptiven Metadaten dazu, ein Submission Information Package (SIP)⁹ zu kodieren.

Für das Mapping der insgesamt sehr heterogenen, teilweise komplexen Dateinobjekte auf ein anderes Format ist zu berücksichtigen, dass ein rein schematisches Vorgehen nicht immer zu befriedigenden Ergebnissen führt. Die Semantik der Daten bedarf häufig einer sachgerechten Interpretation, die wiederum eine Anpassung des Mappings beziehungsweise der Vorverarbeitungsschritte zur Bereinigung und Normalisierung der Daten erforderlich machen kann. Prinzipiell kann im Einzelfall wegen konkurrierender Feldverwendungskonventionen eine einrichtungsspezifische Anpassung des Mappings notwendig werden. Um bei der Bearbeitung komplexer Mappings den Überblick zu behalten, wird der fachlich geschulte Nutzer auf statistische Auswertungen der Feldbelegungen und Messungen zur Vollständigkeitsgrad einer Abbildung zurückgreifen wollen.

⁶ Intellectual Property Rights (IPR)

⁷ Item Level

⁸ Page Level

⁹ ein komplettes Informationsobjekt zum Ingest in die Digital Library

3.4 Diversität von Inhaltstypen

Über die Suchmaschine der DDB greifen Nutzer auf digitale Inhalte vor allem aus Bibliotheken, Archiven, Museen und wissenschaftlichen Einrichtungen zu. Die Inhalte sind so heterogen wie die Metadaten, die sie beschreiben. Die Inhalte unterscheiden sich in der Präsentation des Contents, im Medientyp und im Gegenstand. Während Museen Objekte praktisch beliebiger Art und Größe digitalisieren, werden bei Archiven im Allgemeinen nur die sogenannten Erschließungskontexte in Findbüchern dokumentiert. Museen haben also Gegenstände, während Archive Suchhilfen für Dokumente im weitesten Sinne enthalten, die Aktenlagen dokumentieren. Bibliotheken dagegen digitalisieren in großem Umfang Bücher, bei denen meist aber nicht der Gegenstand Buch, sondern der Inhalt im Zentrum des Interesses steht, und stellen - ähnlich wie Google Books - die Digitalisate bzw. Derivate der Digitalisate auch ins Netz. Neben diesen naheliegenden Inhalten sind für die DDB relevante Informationsobjekte auch Personen, Orte, Einrichtungen, Konzepte, Tierarten, Pflanzen, Gebäude, Nachrichten, etc.

Bei der DDB findet sich das kulturelle Erbe in seiner Vielfalt unter einem Dach vereint. Eine Herausforderung liegt darin, den Content sinnvoll, angemessen und ansprechend zu präsentieren bzw. zu visualisieren sowie die Ergebnisse der Suche in benutzerfreundlicher Weise aufzubereiten. Dass manche Einrichtungen bereit sind, ihre Inhalte in die DDB zu übertragen und von dort ausliefern zu lassen, andere aber den Nutzer auf ihre eigenen Websites und Gateways leiten möchten, muss ebenso berücksichtigt werden wie die vor allem bei Büchern relevante Tatsache, dass Publikationen eine schwer zu fassende Identität haben und dass es verlässliche institutionsübergreifende persistente Kennungen (persistent identifiers) für kulturelle Objekte nicht generell gibt. Dieser letzte Aspekt hat einen Einfluss auf die dritte wesentliche Herausforderung der DDB - die Vernetzung der Objekte - die ihrerseits voraussetzt, dass Objekte zuverlässig dauerhaft identifiziert sind.

4 Lösungsansätze

4.1 Wissensnetz

Die Gegenstände des kulturellen Erbes sind in vielfältiger Weise mit Entitäten verbunden, seien es Personen, Orte, der Geschichte und so fort. Diese Entitäten können ebenso wie der Gegenstand selbst als Ressourcen im Sinne von Linked Data verstanden werden. Sie werden daher ebenfalls als Inhalte¹⁰ einer DDB angesehen, die das Wissen über unsere Kultur modellhaft abbildet. Der Autor eines Buches, ein Staatsratsvorsitzender, dessen Akten im Bundesarchiv liegen und der Künstler, der ein Werk erschaffen hat, sind hier nicht mehr Metadaten, die als Literal in einem Metadatenfeld gespeichert werden können, sondern selbst Inhalte, die über Verlinkung referenziert werden. In einem solchen Wissensnetz kann durch explorative Erforschung neues Wissen generiert werden, wenn durch

¹⁰ link targets

Anfragen an das Wissensnetz unerwartete Zusammenhänge erschlossen werden. Voraussetzung für die Realisierung eines umfassenden Wissensnetzes ist die Modellierung der vorliegenden Daten in Form einer Ontologie, die explorative Queries bzw. Reasoning erlaubt und die den Gegenständen des kulturellen Erbes hinreichend entspricht. Desweiteren ist ein Bezugssystem zur dauerhaften Identifizierung der Objekte in der Plattform unabdingbar.

4.2 Das Linked-Data-Konzept

Datenmodell Das Datenmodell bildet die Grundlage einerseits des Indexierungsprofils und andererseits der Verfahren zu Berechnung von Ähnlichkeitsmaßen, die für Deduplikations- und Clustering-Ansätze benötigt werden. Als Datenmodell der DDB wurde das CIDOC CRM [CCRM] gewählt, da es die Voraussetzungen für die Realisierung eines Wissensnetzes in einem hohen Grad erfüllt. Das CRM ist zunächst ein ontologisches Modell, das die Beschreibung von Objekten über sog. Entitäten und Properties erlaubt, die in ihrer Verknüpfung zunächst einfache Relationen, in ihrer Gesamtheit ein System aus Graphen bzw. Clustern bilden. Dieses Netz kann in idealer Weise durch die Verwendung von RDF [RDF] in eine XML-konforme Darstellung abgebildet werden. Im Gegensatz zu den einschlägigen Metadatenformaten, die die semantischen Beziehungen der Beschreibungseinheiten eines Objekts nur implizit enthalten und sich an der Metapher des Karteikastens orientieren, werden Beschreibungseinheiten durch die Modellierung in CRM Graphen mit einem hohen Maß an semantischer Präzision explizit gemacht.

Mapping Das CRM erlaubt die Abbildung von Objektdaten aus verschiedenartigsten Darstellungsarten (Datenbanktabellen, Metadatenformate). Dieser Prozess wird als Mapping bezeichnet. Die im Rahmen des Entwicklungsprojekts entworfenen Mappings für Metadatenformate orientieren sich an Empfehlungen, die unter Beteiligung von Entwicklern des CRM veröffentlicht wurden.

Eines der Grundprinzipien des CRM besteht darin, einzelne Beschreibungseinheiten eines Objekts (z.B. bei einem Buch: der Autor, der Publisher), über Ereignisse (Events) zu vermitteln.

Die Tatsache, dass ein Autor ein Buch geschrieben hat, wird in CRM durch eine ereignisbezogene Entität modelliert, in diesem Fall über die Entität *E65 Creation* (vgl. Abbildung 1). Events ergeben interessante Kandidaten für eine Deduplikation: zwei identische Creation Events verweisen offensichtlich auf zwei identische Bücher. Auch wenn diese Bücher von verschiedenen Einrichtungen eingebracht werden und uneinheitlich erfasst wurden, besteht eine Chance, die Identität anhand der Events zu erkennen.

Für die Übertragung der flachen Formate in den CRM Graphen werden Sets von Feldern auf Cluster gemappt, wobei die Gesamtmenge der Cluster ein komplettes Mapping ergibt. Für diese Art der Übertragung gibt es die bereits erwähnten konzeptionellen Vorlagen.

Wenn das Ausgangsformat beispielsweise Dublin Core (DC) ist, entstehen Cluster wie in Abbildung 1 gezeigt und in [DCCRM] beschrieben.

1. Für den Endnutzer wird die DDB in Form eines Webportales zur Verfügung gestellt, damit der Nutzer mit einem beliebigen internetfähigen PC ohne Installationsaufwand Zugriff auf das System erhält. Hier werden Daten im Wesentlichen gelesen; eigener Content, Annotationen, Arbeitsbereiche für Arbeitsgruppen werden erst im Lauf der Zeit größere Bedeutung bekommen.
2. Die Kultur und Wissenschaftlichen Einrichtungen, kurz als Partner bezeichnet, erstellen mit Hilfe des Augmented SIP Creators (ASC) SIPs, welche in das System importiert werden. Die SIPs können über einen Harvester automatisiert oder über ein Webportal in manueller Weise zur DDB gelangen. Hier steht die Verarbeitung der Daten in der DDB im Vordergrund.

Für diese zwei Verwendungsarten der DDB gibt es nicht zwei verschiedene Systeme. Es existiert kein abgeschlossenes Portal, das die Suche und Präsentation für den Endbenutzer übernimmt und kein eigenständiges System (z.B. ein CRM), das die Partner bedient. Der Kern der DDB ist als eine Menge von Services gebaut, die zwar geschützt, aber offen zugänglich sind (siehe Abbildung 2). Die Kern-Services sind:

- der Ingestservice, der für die Ablage und Indexierung sowohl der Metadaten als auch der Binärdaten sorgt. Dabei werden digitale Objekte (und ihre Metadaten), Beschreibungen der Partner über ihre Einrichtung und Endbenutzerdaten (wie z.B. Annotationen) gleich behandelt: Sie müssen als SIP verpackt in die DDB gelangen und alles, was autorisiert als SIP an die DDB übergeben wird, wird importiert.
- der Searchservice, der der Suche nach Objekten in der DDB dient. Mögliche Suchergebnisse sind Bücher, Archivalien, museale Objekte oder Beschreibungen über Institutionen.
- der Accessservice, der den Zugriff auf die eigentlichen Inhalte des Repository realisiert. Alle Funktionalitäten, die Endbenutzer und Partner nutzen, werden durch Kombinationen der Services zur Verfügung gestellt. Die DDB präsentiert sich nach außen primär als eine Menge von Ressourcen. Diese Ressourcen können digitale Objekte, Beschreibungen der Partner, Suchen oder auch Suchergebnisse sein.

5 Technische Architektur

5.1 Offene Standards und Schnittstellen

Die Verwendung von offenen Standards wurde angestrebt. Die Anbindungen der Komponenten erfolgt über HTTP / REST. Andere offene Standards, die in diesem Projekt zum Einsatz kommen, sind XML, JSON, RDF, CIDOC CRM, SPARQL, XSLT, XACML und OAuth.

Die Architektur basiert auf Java-EE. Für die Umsetzung werden OpenSource-Implementierungen eingesetzt. So kommen Jetty, Jersey, Xalan, Solr, Joseki und Jena zum Einsatz. Das System ist modular aufgebaut, so dass es an dynamische

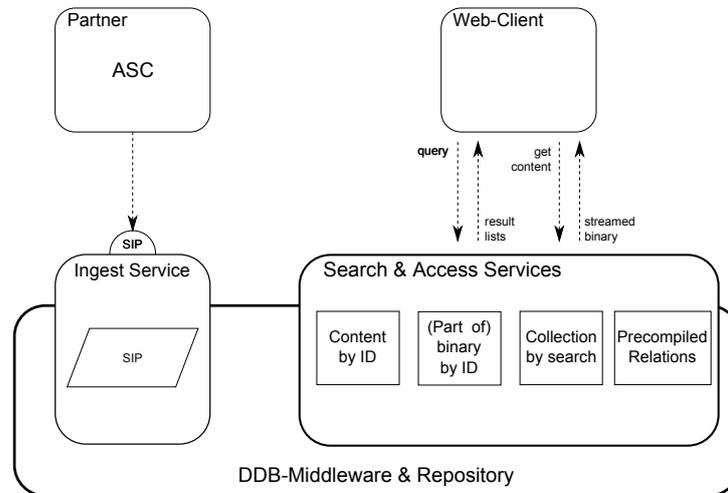


Abbildung 2. Überblickssicht auf die Architektur.

Anforderungen in der Entwicklungsphase optimal angepasst werden kann. Da auf offene Standards gesetzt wurde, können grundsätzlich Implementierungen ersetzt werden.

Dies ermöglicht auch einen Austausch der Basiskomponenten, so dass auch kommerzielle Komponenten, die Standards implementieren, in das System eingefügt werden können. Ein solcher Austausch ist zu erwarten in Zusammenhang mit Cloud-Computing, einem verteilten Storage-System, dem Triplestore oder dem dem Triplestore zugrunde liegenden Datenbankmanagementsystem. Welche Basiskomponenten zum Betrieb tatsächlich verwendet werden, liegt letztlich in der Hand des Betreibers der DDB.

5.2 Modelle

Neben den genannten offenen Standards lehnt sich das Konzept der Plattform an das OAIS und DELOS-DLRM-Modell [OAIS,DELOS] an.

Die DDB implementiert einen Teil des OAIS-Modells, insbesondere Ingest, Access, Storage. Die Langzeitarchivierung, auch als Preservation Planning bezeichnet, ist dagegen nicht Bestandteil des Projekts. Eine spätere Umsetzung ist jedoch angedacht.

Das DLRM führt eine Reihe von Domänen zur Modellierung der Entitäten und Funktionen einer Digital Library ein, darunter vor allem Begriffe wie Resource, Content, User, Functionality und Policy. Diese wurden für die Deutsche Digitale Bibliothek übernommen.

5.3 Funktionale Bestandteile

Ein Überblick über die funktionalen Bestandteile ist in Abbildung 3 zu finden.

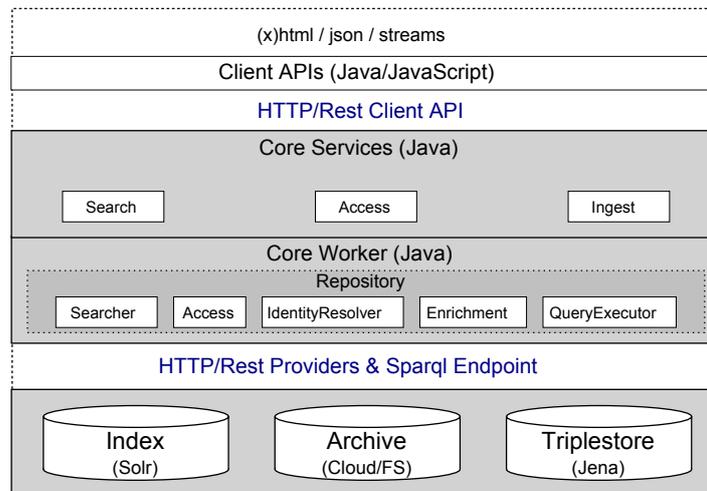


Abbildung 3. Core Services und deren Umfeld.

Harvesting Schnittstellen Die Deutsche Digitale Bibliothek (DDB) führt ein Harvesting auf anderen Systemen durch. Der entsprechende Service bedient sowohl OAI-PMH-Schnittstellen als auch FTP-Server. Weitere Schnittstellen zu Anbietersystemen, die möglich wären, sind ATOM (SWORD), SRU und Z39.50.

Datenaufbereitung (Pre-Ingest) Die in der Plattform verarbeiteten Daten werden vor dem Ingest für die Plattform aufbereitet. Die Aufbereitung muss sehr unterschiedliche Metadaten und Formate in ein geeignetes Modell übertragen und so normalisieren. Nur so können Informationen über die Objekte des kulturellen Erbes in sinnvoller Weise für ein Retrieval genutzt werden können. Die Eingangsdaten müssen außerdem in Dokumente umgeformt werden, die es Menschen (HTML) und Maschinen (RDF) ermöglichen, die Daten zu verstehen. Diese Aufbereitung, bestehend aus der Metadaten transformation und dem Rendering von Sichten auf die Daten, wird im Folgenden als Pre-Ingest bezeichnet. Die Aufbereitung erfolgt unabhängig davon, ob bei einem Ingest zusätzlich auch binäre Inhalte in die DDB Plattform transferiert werden.

Datenimport (Ingest) Der Ingest ist Teil des OAIS Modell und stellt den Prozess der Aufnahme des SIPs und die Umwandlung des SIP in ein Archival Information Package (AIP) inklusive der Ablage des AIP im Archival Storage dar. Funktionale Bestandteile des Prozesses sind entsprechend der OAIS Nomenklatur Receive Submission, Quality Assurance, Generate AIP, Generate Descriptive Information und Coordinate Updates. Diese Funktionen werden innerhalb des Ingestservice der Plattform implementiert.

Ergänzend zu den gerade genannten Funktionen umfasst der Ingest die folgenden weiteren Teilschritte:

- Innerhalb der Plattform müssen persistente Identifier für die digitalen Objekte vergeben werden, die einen Zugriff auf die Objekte und deren Vernetzung mit anderen Objekten ermöglicht. Während des Pre-Ingest werden zunächst lokale Identifier für die Entitäten erzeugt, mit deren Hilfe der Metadatengraph eines Objekts aufgebaut wird. Die lokalen Identifier werden dann beim Ingest durch persistente ersetzt.
- Entitäten, die oft implizit in den Ursprungsdaten enthalten sind und durch das Mapping auf das Modell der DDB explizit als Entitäten ausgewiesen werden, müssen als Ressourcen angelegt werden. So entsteht zu einer Entität eines Buches weitere Entitäten, z.B. das Ereignis seiner Publikation.
- Die Objekte werden in einem Triplestore vernetzt. Speziell dazu werden persistente Identifier benötigt.
- Für den Ingest ist ein Indexing Profile definiert. Ausgehend von einer Entität z.B. einem Museumsobjekt enthält das Profil Transitionsketten, die wichtige Informationen über das Objekt enthüllen. Jede dieser Transitionsketten entspricht letztlich eine Facette, die dem Nutzer bei der Suche als Filter angeboten wird. Diese Transitionsketten werden beim Ingest im Triplestore für jedes digitale Objekt berechnet und im Index für die Suche abgelegt. Als Folge dieser „Vorbereitung“ der Facetten wird der Triplestore zwar beim Ingest benötigt, ist aber bei der Suche und beim Zugriff nicht notwendig und damit kein Verursacher von Performance-Problemen des Web-Portals.
- Während des Pre-Ingests und Ingests können Anreicherungen der Metadaten erfolgen. Dafür sind Integrationspunkte vorgesehen, z.B. für Contentus Dienste, etwa Named Entity Recognition (NER), automatische Extraktion von Metadaten aus Bildern, Audio- oder Videostreams zu sehen.

Datenauslieferung (Access) Die im OAIS in der Access Entität zusammengefassten Funktionen sind in der DDB Plattform als zwei Services angelegt: Access für den Zugriff auf Ressourcen im Storage, Search für den Zugriff über die Suchmaschine. Search entspricht bei OAIS den Query Requests des Data Managements, das Results Sets zur Präsentation an den Nutzer ausliefert. OAIS kennt außerdem Report Requests, die eine Reihe von Abfragen bündeln und formatierte Reports an einen Consumer liefern, und Orders, die entweder an Data Management oder an Archival Storage oder beide gerichtet werden, um ein formales DIP aufzubereiten und auszuliefern. Orders bleiben vorläufig ohne Entsprechung in der DDB. Report Requests werden von Providerseite an die Plattform gestellt, um die Ergebnisse eines Ingests abzufragen.

Anfragen an die Plattform erfolgen über die bereitgestellten Schnittstellen. In der Plattform löst eine solche Anfrage einen Aufruf des entsprechenden Services aus. Dieser ermittelt dann das Ergebnis und sendet dies an den Client zurück.

Stellt z.B. ein Client (Webbrowser) eine Suchanfrage an die Plattform, dann ruft diese intern den Search-Service auf. Dieser bereitet die Anfrage auf und leitet diese weiter an die Suchmaschine. Die Suchmaschine bearbeitet die Anfrage und gibt ein Ergebnis an den Search-Service zurück. Dieser wiederum sendet die Daten zurück an den Client.

5.4 Technische Architektur und Systemkomponenten

Teilsysteme Das Gesamtsystem der DDB besteht aus zwei funktionalen Teilsystemen und einem Authentifizierungs- & Authorisierungssystem (AAS). Die funktionalen Teilsysteme sind der Augmented SIP Creator (ASC) zur Aufbereitung der Daten (Pre-Ingest) und der DL Core, der die Plattform realisiert, auf der die Deutsche Digitale Bibliothek betrieben wird.

Im Folgenden ist nur der Aufbau des DL Cores beschrieben.

Schichten und Komponenten des DL Cores Die Architektur der DDB wird von Komponenten bestimmt, die über einen IoC-Container (IoC = Inversion of Control) eingebunden werden. Sämtliche Komponenten sind gegen Interfaces implementiert. Als Container kommt Spring zum Einsatz.

Bei Systemstart initialisiert der Starter eine von mehreren Spring-Konfigurationen und fährt eine Reihe von Servern hoch. Die Server können je nach gewählter Konfiguration und je nach Deployment auf eine oder mehrere Maschinen verteilt sein.

In der aktuellen Version entstehen keine weiteren Abhängigkeiten, d.h. das zum Deployment zur Verfügung gestellte System ist self-contained und initialisiert sich automatisch. Eine Ausnahme hiervon bildet das optionale Authentifizierungs- und Authorisierungssystem (AAS), das eine eigenständige Anwendung ist.

Die Komponenten der DDB sind auf drei Schichten verteilt, diese werden im Folgenden beschrieben.

Das Webportal, aber auch alle anderen Nutzer kommunizieren über REST mit der Plattform. Das REST API definiert, wie Datenobjekte abgefragt und geschrieben werden. Die Datenobjekte werden als XHTML, mit JSON kodiert bzw. als Streams an den Nutzer geliefert.

Diese REST Schnittstelle wird auf einem REST-Server zur Verfügung gestellt, der CS (Core Server) genannt wird. Damit man die Implementierungen der Services leicht austauschen kann, macht die CS-REST-Schnittstelle nichts anderes, als über Interfaces die Service-Implementierungen aufzurufen, die als Spring Beans in den CS injiziert worden sind. Mit verschiedenen Spring-Konfigurationen lassen sich so leicht verschiedene DDB-Varianten herstellen. Die (wirklichen) Implementierungen der Services werden CW (Core Worker) genannt.

Die Worker erhalten Daten als XML bzw. JSON kodiert. Auf eine Modellierung der Daten in Java, also z.B. den Aufbau eines DOM, wurde weitestgehend verzichtet. Die Kombination Core Server/Core Worker kann auf einer oder in Kopie auf vielen Servern zur Verfügung gestellt werden.

Die Worker kommunizieren mit den drei Backends, dem Index Server für die Suche, dem Archive für den Inhalt der DDB und dem Triple Store für den RDF-Graphen wieder über http. Je nach gewähltem Backend können über die http-Verbindung REST, Webservices oder ein Sparql Endpoint angesprochen werden. Als Sparql Endpoint ist zurzeit die Joseki Implementierung im Einsatz.

Authentifizierungs- und Authorisierungssystem Um die Verwaltung der Nutzerdaten (Name, Passwort) von der DDB unabhängig durchzuführen, werden sowohl die Authentifizierung als auch die Autorisierung in einem separaten System (AAS=Authentication/Authorisation Server) verwaltet. Plattform und AAS kommunizieren über Token miteinander. Die Abbildung 4 zeigt die beteiligten Komponenten.

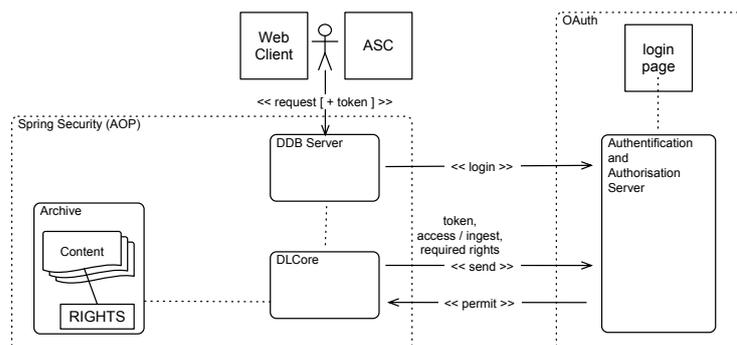


Abbildung 4. Authentifizierungs- und Authorisierungssystem Überblick.

Geschützte Digitalisate liegen in der Regel bei den Providern; ihr Schutz ist damit außerhalb der Verantwortung der DDB Plattform. Prinzipiell können jedoch geschützte Digitalisate auch innerhalb der Plattform abgelegt werden. In solchen Fällen sorgt die DDB Plattform selbst für den Schutz vor unerlaubtem Zugriff.

Ergebnislisten von Suchabfragen werden nicht dynamisch auf Rechte überprüft, da davon auszugehen ist, dass ausschließlich solche Inhalte indiziert werden, die im öffentlichen Netz sichtbar sein sollen. Der Ingest Service benötigt als einziger Service Schreibzugriff. Jeder Ingest erfordert deshalb immer eine entsprechende Authentifizierung des Nutzers und Autorisierung zum Zugriff auf die Datenbestände. Die Implementierung der Authentifizierung entspricht der OAuth Spezifikation¹¹. Sie ermöglicht es, den Login Prozess durch Indirektion komplett von der DDB Plattform zu separieren.

Die Implementierung des Permission Control Systems des AAS folgt einem OASIS Standard¹², der eXtensible Access Control Markup Language (XAML)¹³. Die Implementierung des AAS wird beim Betreiber erfolgen.

5.5 Datenaufbereitung und Datenfluss

Ein Überblick über den Datenfluss ist Abbildung 5 zu entnehmen. Die Metadaten werden durch den Augmented SIP Creator (ASC) zu einem SIP transformiert.

¹¹ <http://oauth.net/core/1.0/>

¹² <http://www.oasis-open.org>

¹³ http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml

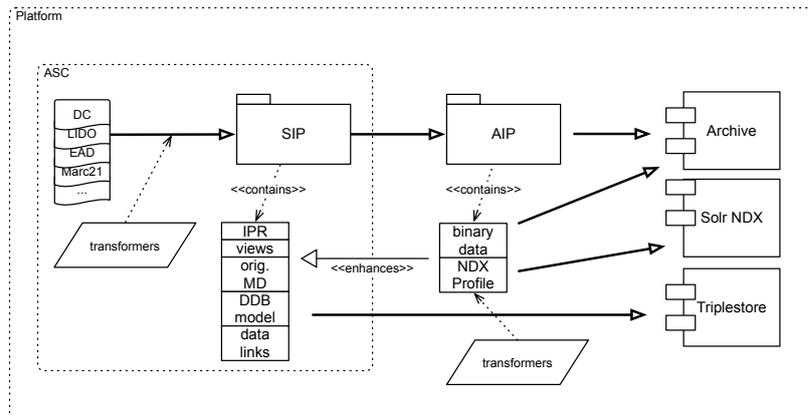


Abbildung 5. Überblick über den Fluss der Metadaten.

Während des Ingestprozesses wird aus dem SIP ein Archival Information Package (AIP) erzeugt. Mit dessen Daten wird das Datei-Archiv, der Solr-Index und der Triplestore befüllt:

ASC Je nach Einrichtung unterscheiden sich die eingesetzten Metadatenformate. Diese Ursprungsmetadatenformate bilden den Ausgangspunkt für ein Mapping auf ein DDB konformes Datenmodell. Optional können neben den reinen Metadaten auch Binärdaten, Digitalisate und Derivate in die DDB übertragen werden.

Die Datenaufbereitung beginnt mit den Ursprungsmetadaten. Die Umwandlung in das DDB konforme Datenmodell erfolgt im Augmented SIP Creator (ASC). Transformationsskripte im ASC wandeln die Ursprungsmetadaten in das Datenmodell der DDB um. Zusätzlich erzeugt der ASC statische, templatebasierte XHTML Ansichten der Metadaten. Diese dienen später in der Plattform zur Anzeige im Browser (z. B. Listen- und Detailansichten). Die normalisierten Metadaten werden zusammen mit den erstellten XHTML Ansichten und den Ursprungsmetadaten in einem Submission Information Package (SIP) zusammengefasst und mittels eigener Metriken auf Konsistenz hin geprüft. Abschließend wird das SIP durch den ASC an die Plattform zur weiteren Verarbeitung übergeben. Alle Funktionen des ASC sind als graphisch modellierte Prozesse organisiert.

Ingest aus Sicht der Plattform Aus Sicht der Plattform erfolgt der Import der SIPs durch den Ingest. Das SIP und die darin enthaltenen Metadatencluster enthalten lokal gültige Identifier. Diese werden im Pre-Ingest durch plattformspezifische, persistente Identifier ersetzt. Aus den Metadatenclustern werden die Facetten und die zu indexierenden Daten ermittelt. Die Facetten und die zu indexierenden Daten werden als Indexing Profile zusammengefasst. Der Metada-

tencluster, das Indexing Profile und die Ursprungsmetadaten bilden das Archival Information Package (AIP).

Der Ingest-Service übernimmt die Persistierung des AIP in der DDB. Die in den Metadatenclustern enthaltenen Beziehungen zwischen Subjekten, Prädikaten und Objekten/Literalen werden im Triplestore angelegt. Die im Indexing Profile enthaltenen Facetten und die zusätzlich zu indexierenden Daten werden in den Indexer übernommen. In einem letzten Schritt erfolgt die Ablage des AIPs im Archiv.

6 Zusammenfassung und Ausblick

Für das DDB Projekt steht in der ersten Phase auf die Erreichbarkeit und Nutzbarkeit der Inhalte von Kultur- und Wissenschaftseinrichtungen im Vordergrund. In diesem Papier wurden die technischen Konzepte und Lösungsansätze für das geplante Informationssystem dargestellt. Die DDB Architektur besteht aus einem Tool zum Importieren von Daten in die Plattform und einem Portal für Recherche und Zugriff auf die Daten. Der DDB Kern bündelt in Services die Funktionalität der Plattform. Die Services ermöglichen das Ingestieren von Daten, die Ablage und Indexierung der Metadaten und Binärdaten, das Suchen nach Objekten in der DDB sowie zum Zugriff auf die eigentlichen Inhalte des Repository.

Kommerzielle Partner wie Verlage sind interessiert an Integrationsschnittstellen und sicheren Transaktionsverfahren. Kleineren Einrichtungen, denen die technischen Möglichkeiten zur Langzeitarchivierung fehlen, kann die DDB entsprechende Dienstleistungen anbieten. Wenn die DDB sich darüber hinaus von vorhandenen Online-Katalogen abheben will, die digitalen Content auflisten, können die Nutzer über geeignete Anreize wie Mashups und Verlinkungen auf im jeweiligen Kontext relevante andere Inhalte, aber nicht allein durch die schlichte Auflistung von Suchergebnissen à la Google stärker an die Plattform gebunden werden. Nutzer müssten auch eingeladen werden, aktiv zu der DDB und zu sozialen Netzwerken innerhalb der DDB beizutragen. Schließlich ergibt sich eine Reihe von Anforderungen aus der Notwendigkeit, dass die DDB Funktionen zur Verwaltung und Organisation der beteiligten Einrichtungen unterhalten muss.

Literatur

- MC08. Christen, Michael: Machbarkeitsstudie zum Aufbau und Betrieb einer „Deutschen Digitalen Bibliothek“, http://www.deutsche-digitale-bibliothek.de/pdf/machbarkeitsstudie_20080723.pdf, 2008.
- JK09. Elliot Jaffe and Scott Kirkpatrick: Architecture of the Internet Archive, 2009.
- CGS09. Cesare Concordia, Stefan Gradmann, Sjoerd Siebinga: (just) a Portal: A Portrait of Europeana as an API, IFLA 2009.
- HO09. Holley, Rose: A success story - Australian Newspapers Digitisation Program. Online Currents, Volume 23, Issue 6, December 2009 pp 283-295 ANDP.

- BA08. Deutsche Geschichte : Bundesarchiv-Bilder bei Wikipedia. Stuttgarter Zeitung. 04.12.2008. http://www.stuttgarter-zeitung.de/stz/page/1891925_0_2147_deutsche-geschichte-bundesarchiv-bilder-bei-wikipedia.html.
- RO07. Romary L.: OA@MPS - a colourful view Max Planck Digital Library, Zeitschrift für Bibliothekswesen und Bibliographie, 2007.
- HS10. Hoffmann, & H. Scholz: BHL-Europe: Biodiversity Heritage Library for Europe-Nimis, In P.L. & R. Vignes Lebbe (eds.). Tool for identifying biodiversity: Progress and problems, pp. 43-48, 2010.
- MA02. Maier, G.: Gemeinsames Internetportal für Bibliotheken, Archive und Museen-BAM-Portal. World Library and Information Congress: 68th IFLA Council and General Conference, 2002.
- PT03. Pianos, T.: Vascode - a Portal for Scientific Resource Collections Created by German Libraries and Information Centres. World Library and Information Congress: 69th IFLA General Conference and Council, 2003.
- RA09. Razum, M. et. al.: eSciDoc Infrastructure: A Fedora-Based e-Research Framework. Research and Advanced Technology for Digital Libraries. Volume 5714, p227-238, 2009.
- HJ08. Hitson, B. A.; Johnson, L. A.: WorldWideScience.org: Bringing Light to Grey. Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society, 8-9 December 2008.
- PL98. Payette, S.; Lagoze, C.: Flexible and Extensible Digital Object and Repository Architecture (FEDORA). Research and Advanced Technology for Digital Libraries, Volume 1513, p. 517, 1998.
- OAIS. Reference Model for an Open Archival Information System (OAIS), CCSDS Secretariat, Washington, DC, 2002, <http://public.ccsds.org/publications/archive/650x0b1.pdf>.
- DELOS. The DELOS Digital Library Reference Model, http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_096.pdf.
- DC08. Dublin Core Metadata Element Set, Version 1.1. <http://dublincore.org/documents/dces/>.
- METS. METS, Metadata Encoding & Transmission Standard, The Library of Congress, <http://www.loc.gov/standards/mets/>.
- MODS. MODS, Metadata Object Description Schema, The Library of Congress, <http://www.loc.gov/standards/mods/>.
- MARC. Marc Standards, Library of Congress, <http://www.loc.gov/marc/>.
- EAD. Encoded Archival Description, Library of Congress, <http://www.loc.gov/ead/>.
- MUDA. Museumdat, <http://www.museumdat.org>.
- LIDO. Lightweight Information Describing Objects, CIDOC, <http://cidoc.icom.museum/>.
- CCRM. CIDOC Conceptual Reference Model, <http://www.cidoc-crm.org/>.
- RDF. Resource Description Framework (RDF), W3C, <http://www.w3.org/RDF/>.
- DCCRM. DC.type mapping to CIDOC/CRM, Konstantia Kakali, Martin Doerr, Christos Papatheodorou Thomais Stasinopoulou, Department of Archives and Library Science / Ionian University, 2007.

Integration von nicht-textuellen Objekten in das GetInfo Portal der Technischen Informationsbibliothek

Jan Brase, Ina Blümel

jan.brase@tib.uni-hannover.de, ina.bluemel@tib.uni-hannover.de

Abstract: Nicht-textuelle Objekte wie Forschungsdaten aber auch 3D-Modelle sind mittlerweile eine immer wichtigere Quelle des Wissenserwerbes. Die Technische Informationsbibliothek (TIB) als Deutsche Zentrale Fachbibliothek für Technik sowie Architektur, Chemie, Informatik, Mathematik und Physik hat in verschiedenen Projekten begonnen über ihr Portal GetInfo auch solche Objekte suchbar und verfügbar zu machen. Das vorliegende Paper stellt die aktuellsten Beispiele dieser Integration von nicht-textuellen Objekten in ein Bibliotheksportal vor.

1 Einleitung

Wissenschaftliche Information wird immer komplexer und basiert immer häufiger auf großen Datensammlungen. Die übliche Verbreitungsform für Wissen ist immer noch der Wissenschaftliche Artikel, der aber nur den letzten Schritt in einem Prozess darstellt, an dessen Anfang Forschungsdaten gestanden haben. In ihrem 2007 Bericht *Cyberinfrastructure Vision for 21st Century Discovery* schrieb die National Science Foundation aus den USA:

Science and engineering research and education have become increasingly data-intensive as a result of the proliferation of digital technologies, instrumentation, and pervasive networks through which data are collected, generated, shared and analyzed. Worldwide, scientists and engineers are producing, accessing, analyzing, integrating and storing terabytes of digital data daily through experimentation, observation and simulation. Moreover, the dynamic integration of data generated through observation and simulation is enabling the development of new scientific methods that adapt intelligently to evolving conditions to reveal new understanding. ([8]).

Im dem kürzlich veröffentlichten Buch *The Fourth Paradigm: Data-Intensive Scientific Discovery* [5] von Microsoft Research wird ebenfalls darauf hingewiesen, wie sich die Wissenschaft generell in den letzten Jahrhunderten von einer Empirische zu einer Datenintensiven Disziplin gewandelt hat. Aus der Sicht von Bibliotheken werden dort ebenfalls neue Aspekte identifiziert:

Scientific communication, including peer review, is also undergoing fundamental changes. Public digital libraries are taking over the role of holding publications from conventional libraries because of the expense, the need for timeliness, and the need to keep experimental

data and documents about the data together.

Wissenschaftliche Information tritt heutzutage nicht mehr nur in der Form eines Artikels oder eines Buches in Erscheinung. Daher müssen auch Bibliotheken ihre Kataloge für andere Inhaltsformen öffnen, wenn sie ihren Aufgaben der Informationsversorgung gerecht werden wollen. Dieses bedeutet aber auch gerade im Umgang mit teilweise sehr großen Datensätzen, dass Bibliotheken nicht mehr alle Inhalte selber vorhalten müssen. Entscheidend ist vielmehr den Katalog der Zukunft als ein Portal zu verstehen. Externe Inhalte wie große Datensätze aber auch Filme oder beliebige Lernobjekte werden katalogisiert, um dem Nutzer passende Ergebnisse zu seinen Suchen anzubieten. Der Zugang zu den Inhalten erfolgt dann aber über eine stabile Verlinkung zu den externen, vertrauenswürdigen Archiven und Datenbanken, die auf die Speicherung dieser Inhalte spezialisiert sind. (siehe Abbildung 1).

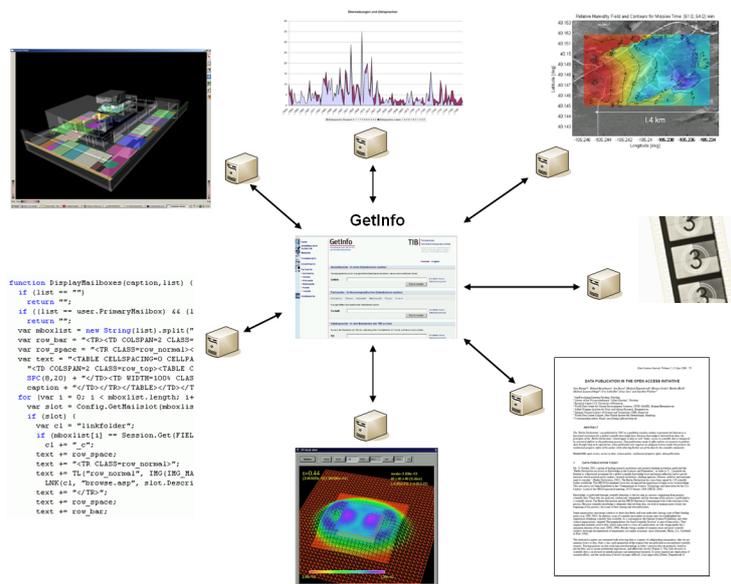


Abbildung 1: Das Bibliotheksportal der Zukunft, beispielhaft für GetInfo der TIB, basierend auf einem Netzwerk von vertrauenswürdigen Inhaltsanbietern

2 Zugang zu Forschungsdaten durch GetInfo: Beispiele

Die Technische Informationsbibliothek (TIB) ist die Deutsche Zentrale Fachbibliothek für Technik sowie Architektur, Chemie, Informatik, Mathematik und Physik. GetInfo ist das Portal für technisch- naturwissenschaftliche Fach- und Forschungsinformationen, es bündelt den Zugang zu führenden Fachdatenbanken, Verlagsangeboten und Bibliothekskatalogen mit integrierter Volltextlieferung. GetInfo bietet damit einen weltweit einzigar-

tigen Bestand an technisch-naturwissenschaftlicher Fachinformation. Derzeit ist GetInfo das einzige große Bibliotheksportal in Europa, das auch den Zugang zu Forschungsdaten bietet. Anbei drei Beispiele:

- Bibliotheksportale sind die klassische Quelle für Information [7]. Auf der Suche nach bestimmten Themen mag den Nutzer nicht nur Interesse an allen Publikationen zu dem Thema haben, sondern auch an den Datensätzen, die ein bestimmter Wissenschaftler gesammelt hat. Durch die Vergabe von persistenten Identifiern werden diese Daten direkt über den Katalog zugänglich. Derzeit sind bereits über 2.000 Datensätze, die Grundlage einer Wissenschaftlichen Publikation sind, direkt als eigenständige Objekte zugänglich über GetInfo und den Gemeinsamen Bibliotheksverbund (GBV) [3]. Abbildung 2 zeigt Daten zu einem Bohrkern als Ergebnis einer Suche in GetInfo. Durch Auflösen des persistenten Identifier, gelangt man auf eine Vorschauseite des Datenzentrums, die weitere Metadaten zu dem Datensatz enthält, sowie Download-Links zu einzelnen Teilen des Datensatzes oder den gesamten Daten (Abbildung 3). Dieser Workflow entspricht der traditionellen Verwendung von DOI Namen bei Wissenschaftlichen Zeitschriften.
- Ein weiteres Beispiel nicht-textueller sind geologische Karten. Auch solche sind über GetInfo verfügbar. Die Auflösung des DOI-Namen führt hier ebenfalls auf eine Vorschauseite beim Datenzentrum (Abbildung 4), von der dann die eigentlich Karte zugänglich ist (Abbildung 5)
- Innerhalb des Bereichs der nicht-textuellen Information ist in den letzten Jahren ein starker Anstieg von multimedialen Inhalten zu beobachten. Beispielsweise wird die Verwendung von 3D-Modellen im Bereich der Ingenieurwissenschaften immer wichtiger. Mittlerweile sind auch 3d-Modelle über GetInfo suchbar (Abbildungen 6 und 7). Die Darstellung und Suche erfolgt über die Datenbank PROBADO3D (siehe nächster Abschnitt) als eingebundene externe Quelle bei GetInfo (siehe [2])

3 Inhaltsabhängig Indexierung

Nicht-textuelle Materialien enthalten in der Regel keine beschreibenden Metadaten, wenn Sie nicht extra katalogisiert werden. Heutzutage werden nicht-textuelle Objekte, wenn überhaupt, basierend auf textuellen Metadaten in Bibliotheken indiziert und präsentiert. Eine von der TIB durchgeführte Umfrage unter Architekten ergab, dass Architekten überwiegend mit textuellen Schlagwörtern nach 3D-Modellen suchen und nicht mit visuellen Interfaces. Daher ist es wichtig solche textuellen Informationen aus den Modellen zu extrahieren. Im DFG geförderten Projekt PROBADO wurden an der TIB verschiedene Verfahren untersucht und implementiert ([1]).

Derzeit verwendete inhaltsbasierte Indexierung-Werkzeuge für 3D-Modelle basieren entweder auf der geometrischen Form ([6]) der Objekte, wodurch eine schnelle Suche durch große Datenbestände möglich ist; oder auf Verbindungsgraphen für die einzelnen Räume. Letzteres Verfahren ist hervorragend für den Einsatz im Architektonischen Kontext geeig-



Abbildung 3: Präsentation des Datensatzes beim Datenzentrum mit entsprechendem Download-link

- Hochladen von eigenen Modellen und 3D-Skizzieren für Query-by-example Suchen
- Skizzieren und Suche mit Raumverbindungsgraphen in Gebäudemodellen

Für die Möglichkeit über nach den in GetInfo verfügbaren Forschungsdaten visuell zu suchen hat 2010 das Projekt *Visueller Zugang zu Forschungsdaten* gestartet. (<http://www.tib-hannover.de/de/die-tib/projekte/visueller-zugang-zu-forschungsdaten/>) Ziel ist die Entwicklung und prototypische Umsetzung von innovativen Ansätzen für den interaktiven, graphischen Zugang zu Forschungsdaten, um diese optimal im Information Retrieval Prozess darstell- und suchbar zu machen. Im Projekt werden hierzu Verfahren zur Datenanalyse sowie für visuelle Suchsysteme untersucht und weiterentwickelt, sowie deren prototypische Umsetzung in das Fachportal GetInfo evaluiert. Ziel ist neben der metadatenbasierten Suche auch die unmittelbare Suche in den Forschungsdaten. Die Abbildung 10 visualisiert die graphisch-interaktive Suche nach einem vom Nutzer vorgegebene Kurvenverlaufsmuster innerhalb eines großen Bestandes von Verlaufsmustern. Eine visuelle Clusteranalyse des Gesamtbestandes bildet die Ausgangsbasis. Die Suchanfrage wird über die interaktive Skizzierung des gesuchten Kurvenverlaufs gestellt (siehe links oben). Ein hell-dunkles Colormapping über dem Gesamtbestand visualisiert den Grad der Übereinstimmung des vorgegebenen Verlaufsmusters mit den Mustern innerhalb des gesamten Datenbestandes.

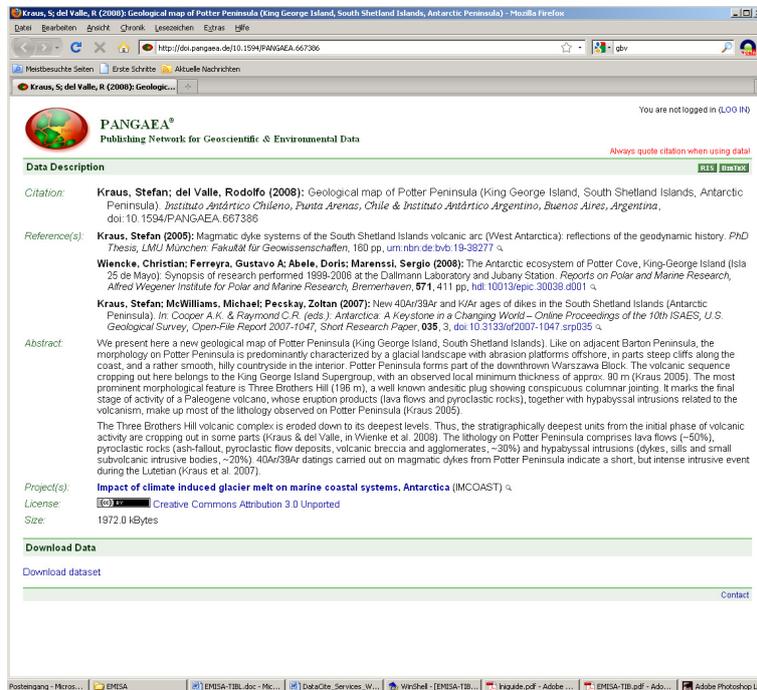


Abbildung 4: Vorschau einer geologischen Karte beim Datenzentrum PANGAEA

5 Persistente Identifizierung

Integration von Daten in Publikationen ist ein wichtiger Bestandteil wissenschaftlicher Zusammenarbeit. Es ermöglicht die Verifizierung wissenschaftlicher Ergebnisse und aktiven Wissensaustausch von Forschern. Im wissenschaftlichen Bereich besteht zwar grundsätzlich Bereitschaft, Daten für eine interdisziplinäre Nutzung zur Verfügung zu stellen, aber es ist zur Zeit unüblich, dass die erforderliche Mehrarbeit für Aufbereitung, Kontextdokumentation und Qualitätssicherung im Wissenschaftsbetrieb anerkannt wird. Die klassische Form der Verbreitung wissenschaftlicher Ergebnisse ist ihre Veröffentlichung in Fachzeitschriften, normalerweise ohne Veröffentlichung der zugrunde liegenden Daten. Diese klassische Publikation wird im "Citation Index" erfasst. Dieser Index wird zur Leistungsbewertung von Wissenschaftlern herangezogen. Datenveröffentlichungen werden darin bisher nicht berücksichtigt. In dem an der Technischen Informationsbibliothek (TIB) Hannover durchgeführten Projekt Publikation und Zitierfähigkeit wissenschaftlicher Forschungsdaten wurde eine Infrastruktur zur Registrierung von DOI-Namen und URNs für wissenschaftliche Datensätze geschaffen und erfolgreich getestet. Mit dem System wurden an der TIB bereits über 650.000 Datensätze aus dem Bereich der Geowissenschaften mit persistenten Identifiern versehen.

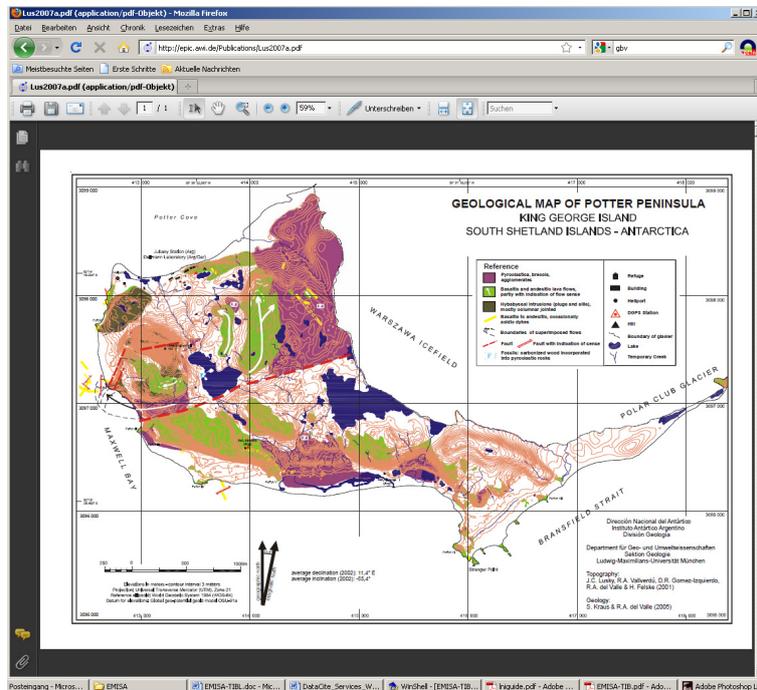


Abbildung 5: Anzeige der eigentlichen Karte

Die Verwendung von DOI-Namen als Identifier ermöglicht eine elegante Verlinkung zwischen einem wissenschaftlichen Artikel und den im Artikel analysierten Forschungsdaten. Artikel und Datensatz sind durch ihre jeweiligen DOI Namen in gleicher Weise eigenständig zitierbar. Diese Form der Zitierung und Verlinkung bietet sich insbesondere bei Forschungsdaten an, die in direkter Beziehung zu Wissenschaftlichen Artikeln stehen, sogenannte supplementary data. So wird beispielsweise der Datensatz:

Kuhlmann, H et al. (2009):

Age models, iron intensity, magnetic susceptibility records and dry bulk density of sediment cores from around the Canary Islands.

PANGAEA, Bremen

doi:10.1594/PANGAEA.727522

in folgendem Artikel verwendet:

Kuhlmann, Holger; Freudenthal, Tim; Helmke, Peer; Meggers, Helge (2004):

Reconstruction of paleoceanography off NW Africa during the last 40,000 years: influence of local and regional factors on sediment accumulation.

Marine Geology, 207(1-4), 209-224,

doi:10.1016/j.margeo.2004.03.017

Der im Dezember 2009 gegründete Verein DataCite [4] hat sich zum Ziel gesetzt, Wissen-

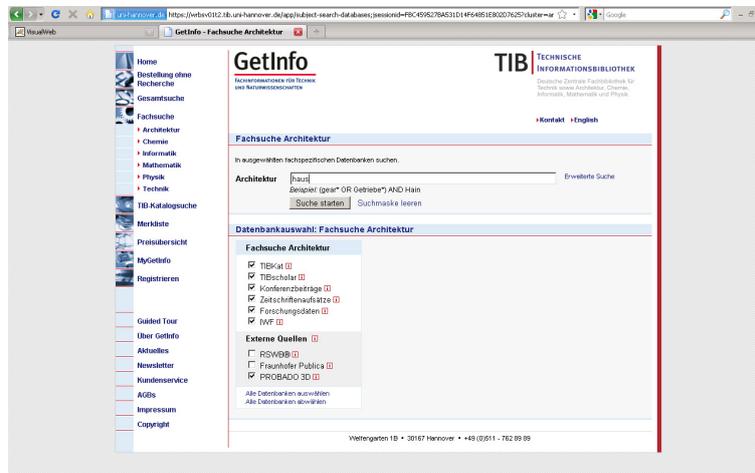


Abbildung 6: Auswahl der Datenbank PROBADO3D als externe Quelle bei GetInfo

schaftlern den Zugang zu Forschungsdaten über das Internet zu erleichtern, die Akzeptanz von Forschungsdaten als eigenständige, zitierfähige wissenschaftliche Objekte zu steigern und somit die Einhaltung der Regeln guter wissenschaftlicher Praxis zu gewährleisten. Bis heute haben sich 12 Partner aus 9 Ländern unter dem Dach von DataCite zusammengefunden: die British Library, das französische Institut de l'Information Scientifique et Technique (INIST), das Technical Information Center of Denmark, die TU Delft Bibliothek aus den Niederlanden, das Canada Institute for Scientific and Technical Information (CISTI), die California Digital Library (USA), der Australian National Data Service (ANDS) die Purdue University (USA) und die Eidgenössische Technische Hochschule in Zürich. Deutsche Mitglieder sind neben der Technischen Informationsbibliothek (TIB), die Goportis Partner Deutsche Zentralbibliothek für Medizin (ZB MED) und voraussichtlich ab dem 01.01.2011 die Deutsche Zentralbibliothek für Wirtschaftswissenschaften, sowie das Leibniz-Institut für Sozialwissenschaften GESIS.

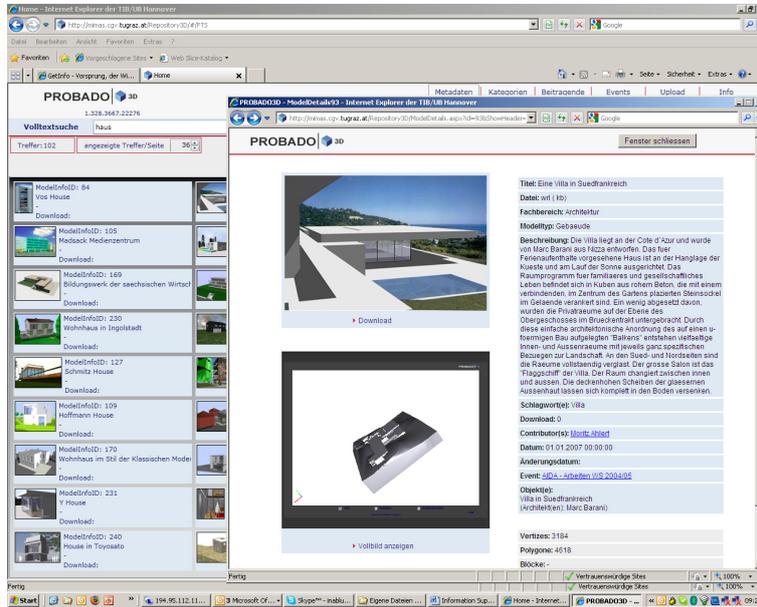


Abbildung 7: Anzeige eines 3D Modells als Suchergebnis

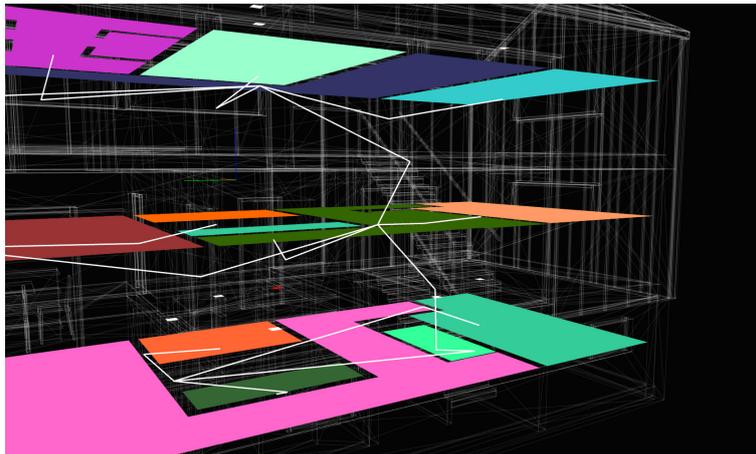


Abbildung 8: Automatisch extrahierter Raumverbindungsgraph eines Gebäudes

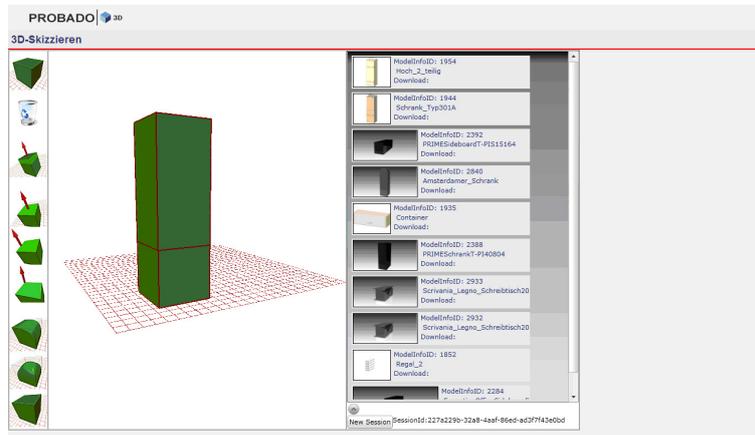


Abbildung 9: Online Zeichenoberfläche für Query-by-example Suchen mit Ergebnissen

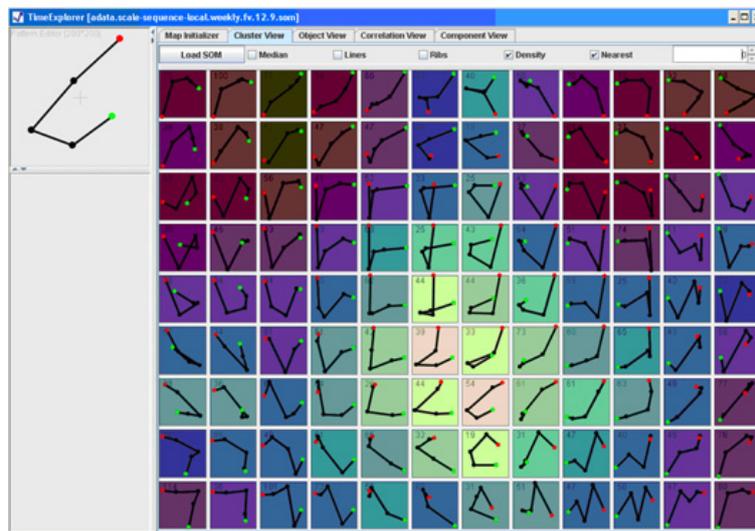


Abbildung 10: Prototyp einer graphisch-interaktiven Suchoberfläche über Forschungsdaten

Literatur

- [1] BLÜMEL, I. DIET, J. KROTTMAIER, H. (2008), *Integrating Multimedia Repositories into the PROBADO Framework*. Third International Conference on Digital Information Management (ICDIM), 2008.
- [2] BLÜMEL, I. SENS (2009), *Das PROBADO-Projekt: Integration von nichttextuellen Dokumenten am Beispiel von 3D-Objekten in das Dienstleistungsangebot von Bibliotheken*. S. 79 ZfBB, Heft 2, 2009, Klostermann, Frankfurt am Main.
- [3] BRASE, J. (2004), *Using Digital Library Techniques - Registration of Scientific Primary Data*. Lecture Notes in Computer Science, No 3232, pp 488-494.
- [4] BRASE, J. (2010), *DataCite - A global registration agency for research data*, Working Paper 149/2010 German Council for Social and Economic Data (RatSWD)
- [5] HEY, T. TANSLEY, S, TOLLE, K. EDS. (2009), *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, ISBN 978-0-9825442-0-4
- [6] JOHNSON, A. (1997), *Spin-Images: A Representation for 3-D Surface Matching*, PhD thesis, Robotics Institute, Carnegie Mellon University,
- [7] INGER, S. GARDNER, T. (2008), *How Readers Navigate to Scholarly Content*, available at <http://www.sic.ox14.com/howreadersnavigatetoscholarlycontent.pdf> accessed 1.9.10
- [8] NSF. (2007), *Cyberinfrastructure Vision for 21st Century Discovery*, Arlington/VA: National Science Foundation (NSF), Cyberinfrastructure Council (CIC).
- [9] WESSEL, R. BLÜMEL, I. AND KLEIN, R. (2008), *The Room Connectivity Graph: Shape Retrieval in the Architectural Domain*, The 16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, Feb. Available at <http://cg.cs.uni-bonn.de/en/publications/paper-details/wessel-2008-the-room/> Accessed 1.9.10

Für Sie gesurft – Neue (und alte) Tipps aus dem WWW

EMISA-Edition, Folge 23

Gottfried Vossen, Universität Münster

Ich möchte Ihnen an dieser Stelle wieder ein paar Websites vorstellen, die mir (und anderen) in letzter Zeit aufgefallen sind. Dabei geht es diesmal um das Thema „Enterprise 2.0“, das sich inzwischen ja auch bereits zu einem geflügelten Wort (und nicht nur zu einem Titel zahlreicher neuer Bücher) entwickelt hat. Grob gesagt versteht man darunter die Anwendung von Techniken, Tools und Konzepten des Web 2.0 auf die Enterprise, also auf Unternehmen. Allerdings lässt sich das auf unterschiedliche Weisen tun: Man kann einerseits Dinge wie Wikis oder soziale Netze *innerhalb* eines Unternehmens anwenden gemäß der Überlegung, dass Mitarbeiter, die diese Dinge privat gerne nutzen, auf am Arbeitsplatz nicht darauf verzichten müssen (und natürlich in der Hoffnung, dadurch eine Produktivitätssteigerung erzielen zu können). Andererseits kann man sich fragen, wie sich die Techniken *nach außen* einsetzen lassen, also etwa für die Kommunikation mit Kunden oder anderen Unternehmen. Die im Folgenden beschriebenen Seiten beleuchten vorwiegend den ersten dieser beiden Aspekte, da sich der zweite hat vor allem auf Blogs und Werbung im Rahmen öffentlicher sozialer Netze bezieht und weniger spezieller Software bedarf.

Wie immer weise ich vorab darauf hin, dass die Inhalte sämtlicher hier beschriebenen Webseiten urheberrechtlich geschützt sind, allerdings ist nach einschlägiger Meinung das Copyright nur relevant für die Verwendung in anderen Webseiten bzw. wenn Gestaltungselemente für andere Designs übernommen würden, was beides nicht der Fall ist. Insofern betrachte ich das Beschreiben von Seiten in der hier vorgenommenen Form weiterhin als Werbung für diese.

Mein erster Tipp dieser Ausgabe betrifft Yammer, das wie auch andere der hier beschriebenen vor einem Jahr in Folge 21 schon einmal erwähnt wurde: „Yammer is revolutionizing internal corporate communications by bringing together all of a company’s employees inside a private and secure enterprise social network. Although Yammer is as easy to use as consumer products like Facebook or Twitter, its enterprise-grade software is built from the ground up to drive business objectives. Yammer enables users to communicate, collaborate, and share more easily and efficiently than ever before. It reduces the need for meetings, increases communication across silos, surfaces pockets of expertise and connects remote workers.“

Yammer

What is Yammer? Contact Sales Log In

Connect with your coworkers

Yammer is the private social network for your company

Only people with a verified company email address can join your company network.

Customer Quotes | "Yammer has proved to be a great way to seek input, test ideas, and gather feedback quickly." - Deloitte

Used at more than 100,000 businesses worldwide:

AMD CISCO Nationwide GROUPON PitneyBowes SUNGARD

<https://www.yammer.com/>

Entsprechend umfangreich sind die Funktionen: Microblogging, Direct Messaging, File- oder Bilder-Upload und -Sharing, private sowie öffentliche Groups und Teams, Profile, Company oder Knowledge Base; zusammen also so umfassend, wie man es sonst nur von mehr als einem Werkzeug kennt. Im ReadWriteWeb wurden ferner die „Communities“ besonders hervorgehoben: „Yammer is opening up its microblogging platform. In "Yammer Community" people may now create a community without the requirement that an email address be associated with a particular domain. This is a big change for Yammer. Many companies do not have their own domains. Opening up the platform means that the service is open to a much larger audience - and has created a much wider place for itself in the enterprise.” Yammer verspricht ferner einfache Administration, hohe Sicherheit, Tagging sowie Integration von Applikationen Dritter. Mobilität wird unterstützt über Apps iPhone, BlackBerry, Android und Windows Mobile.

Mein nächster Hinweis ist Socialcast, schon etwas älter; der Hauptspruch ist hier „Socialcast is a powerful Enterprise Activity Stream Engine: A collaborative microblogging platform that unites your company’s people, data, and applications in real-time.” Darüber hinaus wird natürlich noch vieles mehr versprochen: „Collaboration has never been easier or more effective. Socialcast syncs real-time data-sharing features with enterprise-grade security. Activity Streams unite people, data and processes. Enterprise Microblogging - share status updates, questions, comments and ideas. Mobile Accessibility with iPhone & Blackberry, Outlook 2010, 2007 and 2003 integration, SharePoint Web Part, Desktop App, ferner Messaging, Polls, Profiles und Analytics, Synchronisation von User Accounts über LDAP.

<http://www.socialcast.com/>

Der Dritte im Bunde dieser neuen Generation von Software-Werkzeugen ist Socialtext, auf das ich bereits in Folge 17 einmal hingewiesen habe (aber das liegt drei Jahre, also eine Internet-Ewigkeit zurück): „Chances are you only know what’s going on in your company when someone else decides you should. You probably spend 1/3 of your day in your e-mail

client, and 500 hours a year hunting for stuff. You can't find someone with a particular expertise, leadership and employees are out of touch, and people re-create work. Socialtext 4.0 addresses these problems. With Socialtext, everyone knows what's going on. People and teams are synchronized, engaged and informed. Socialtext provides a broad social software platform that has the ease of SaaS and the security of an on-site appliance. You get a quick deployment that focuses on your people, not on the software code. And you get real business value, fast. Socialtext unlocks the knowledge, expertise, ideas and data that drives superior business performance. Socialtext accelerates business performance by making it easier for employees to find the colleagues and information they need to solve challenges new and old. By simplifying people's ability to share expertise, ideas and corporate data, Socialtext removes knowledge silos that have traditionally hampered companies' ability to respond to change and serve customers efficiently. With Socialtext, people collaborate openly around key business processes in a secure, internal environment, and work together to drive new business opportunities.”

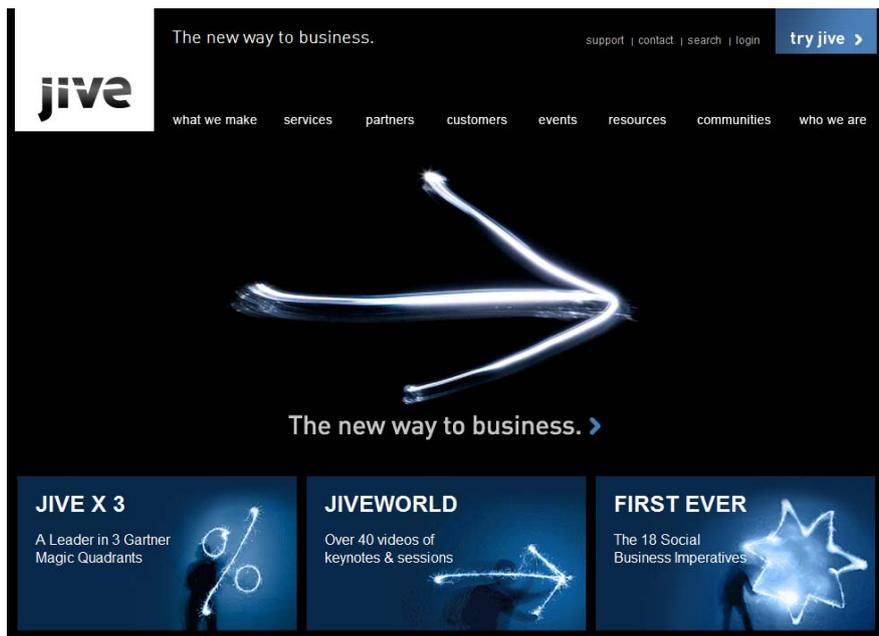
<http://www.socialtext.com/>

Auch hier findet man weitgehend gleichartige Funktionalität:

<p>Social Networking People profiles with live activity streams ensure people discover expertise across the organization, get to know each other much better, and stay highly engaged.</p>	<p>Microblogging Secure microblogging with group channels ensures everyone knows what's going on. People and teams move fast, and leadership and staff stay connected.</p>	<p>Groups Tools for group productivity, that let teams form instantly, stay in sync, and work fast. Groups get energized, cycle times get shorter, and business results come faster.</p>	<p>Social Spreadsheets Distributed spreadsheets that stay in sync. Info is updated in one place, and everyone sees the latest. Spreadsheet information is discoverable, findable and secure.</p>	<p>Dashboard Social intranet home page. IT gets a dynamic intranet, users get a home page they can personalize, and leadership gets a vehicle for the voice of the company.</p>
<p>Wiki Workspaces Group-editable web pages that let people work openly. Expertise and work output are visible across the organization, discovered by others, and highly leveraged.</p>	<p>Internal Blogs Used for two-way, open dialog between leadership and employees, and to create hubs of knowledge, ideas and expertise on important topics.</p>	<p>Desktop App A dynamic desktop application built on Adobe AIR technology that provides a rich and intuitive interface into Socialtext from any desktop platform.</p>	<p>Mobile Stay productive and effective, using your mobile device. Get updates and fast answers from colleagues just like you do when you're in the office.</p>	<p>Integration Integrate your CRM, ERP, document management and other enterprise applications with Socialtext, and get more business value from them.</p>

Erst Anfang November wurde im ReadWriteWeb darauf hingewiesen, dass Socialtext in der neuen Version 4.5 u. a. einen Connector für Salesforce enthält sowie Socialtext Explore zum Browser von Status-Updates mittels Metadaten.

Der nächste Kandidat in dieser Reihe ist Jive: „We're the largest and fastest growing independent Social Business Software company in the world. Our next-generation software is the only truly enterprise-scale social business solution on the market. Our platform integrates the best features of social networking and collaboration software, community software, and social media monitoring to extend a layer of social capabilities across every level of your business.“



<http://www.jivesoftware.com/>

Die Features von Jive sind denen der zuvor genannten ähnlich, aber die Webseite lässt erkennen, dass man wohl auch weiterdenkt:

<p>Collaboration & Corp Comm</p> <p>Tear down hierarchies, connect, and produce results in a fun and productive way.</p> <p>▶ Learn how</p>	<p>Customer Support</p> <p>Give customer product discussions a home. Improve satisfaction as a result.</p> <p>▶ Learn how</p>	<p>Social Media Monitoring</p> <p>Listen, measure, and engage with conversations across the social web.</p> <p>▶ Learn how</p>
<p>Social Marketing</p> <p>Create communities of interest that translate engagement into revenue.</p> <p>▶ Learn how</p>	<p>Sales & Channel Enablement</p> <p>Arm your team with community-enabled tools, support, intelligence.</p> <p>▶ Learn how</p>	<p>Open Government</p> <p>Foster citizen participation, reduce wasted effort, make expertise easy to find.</p> <p>▶ Learn how</p>

Themen wie Kundensupport oder das Monitoring sozialer Medien, ferner Hinweise darauf, wie auch z. B. Regierungsabteilungen diese Software einsetzen können, finden man in dieser Form bei den anderen noch nicht.

Mein nächster Tipp in der immer noch gleichen Kategorie ist SimplyBox: „SimplyBox was originally created as a free Web 2.0 service, with a focus on usability and visual simplicity. After the recommendation of multiple Enterprise analysts, the company made the decision to re-focus its solutions towards the Enterprise space. Today SimplyBox is a company that sees the Enterprise as its natural present and future. Our passion for simplicity brings a unique value proposition to enterprise software, which we have observed over many decades of cumulative experience to be insufficiently user-centric. At SimplyBox, we believe in the power of Human Intelligence. We create solutions that take the friction out of sharing and harnessing the knowledge that traditional enterprise applications and collaboration systems don't adequately deal with, because they are not people-centric in their design. Communication must be fast, ubiquitous, and simple. SimplyBox develops technology that empowers human beings with simpler, more effective communication in the context of their activities within the enterprise.”

Zur Arbeitsweise der Software gibt SimplyBox folgende Hinweise: “The reality is that e-mail still dominates most of the collaboration that takes place in the Enterprise. Don't blame e-mail. The blame should be on the shoulders of collaboration vendors that have not come up with something that truly provides the user with a better alternative. SimplyBox changes enterprise communications by merging ease of use with a decentralized approach to collaboration. During a regular day, users in the Enterprise interact with multiple Enterprise systems, e.g. HR, ERP, CRM, Portals, Dashboards, Reports. Most of these systems provide virtually no collaboration mechanisms. So, when it is time to collaborate, users go to e-mail or to a centralized collaboration application. And as they move to collaborate somewhere else, they lose the context the systems they are interacting with provide. The result: collaboration without context, which equals to slower and less effective collaboration. SimplyBox empowers the user to communicate impulsively, anywhere. Allowing them to collaborate from within the multiple systems they interact with daily.

SimplyBox *Think Inside the Box*  [About](#) [Contact](#) [Product Overview](#) [Videos](#)

What's all the noise about? *SimplyBox announces "App Remix", as featured on ReadWriteWeb.* [Learn More](#)

Enterprise Users work with many apps...
Don't force them to leave to Communicate!

Dashboard
Portal
Service & Support
HR
Sales
Business Intelligence
CRM

Watch  Try [Free 30 Day Trial](#)

Analysts
[Latest Gartner Hype Cycle: Activity Streams take top spot on 'Expectations' curve.](#)
[SimplyBox is "Company to watch".](#)
Klint Finley, ReadWriteWeb Enterprise

Latest News
["Silo Smasher" ...](#) [ReadWriteWeb](#)
 Ecenta showcases SimplyBox at Expo in Nuremberg, Germany.
 SimplyBox now part of Oracle's Fusion CRM Partner Program.
 "Ideal for educational use"... [Tech&Learning](#)

What Customers Are Saying
 "SimplyBox is an exciting complement to our CRM implementation."

 "SimplyBox is so user-friendly that user adoption is greatly enhanced."

<http://www.simplybox.com/>

Putting communication in context. SimplyBox's "Boxes" are key to enabling decentralized contextual collaboration. A "Box" includes conversations and content about a topic. And a box can be shared with as many people as needed. SimplyBox users tend to have multiple boxes, about different topics, and shared with different groups. SimplyBox 2.0 allows users to place boxes on top of pages of Enterprise systems. A user will have some boxes on top of ERP, others on top of a Portal or a Report, and so on. And as they place these boxes on top of these systems they are bringing content and collaboration within the context of the different systems they are interacting with. And as they re-visit these systems, their boxes and their content will be there available to enable collaboration."

Der letzte Hinweis in dieser Reihe ist Yakabod mit ihrem zentralen Tool Yakabox: „The Yakabox™ is a secure knowledge-sharing appliance that incorporates four applications: collaboration, social networking, content management and search. A nimble, robust alternative to traditional enterprise knowledge management systems, the Yakabox™ deploys in one-tenth the time, scales from a few dozen to tens of thousands of users with no downtime, and is so easy to navigate that users are productive immediately. The Yakabox™ is delivered with four pre-installed, ready-to-use secure applications. Simply plug it in and go to work. Create user profiles, build teams, assign tasks, track projects and workflow, create blogs and microblogs, schedule events and send reminders, manage content, search for information, send messages, upload and download links, build your own custom applications (without waiting in line for the IT department), leave comments, conduct polls and more. The Yakabox™ accepts your structured and unstructured information of any type from virtually any source you permit, including assets from Microsoft Office and SharePoint®, shared drives, scanned ink-on-paper archives, electronic repositories...you name it. Since the Yakabox™ is user-centric rather than data-centric, it also captures the good stuff— ideas, opinions and lessons learned—that traditional systems don't recognize.”

DO STUFF THAT MATTERS
YAKABOD
Secure Knowledge Sharing Systems

THE YAKABOX APPS COMPARE INSTALL COMPANY RESOURCES NEWS

blogtalkradio
Collaboration Pizza
Who's that scoring on this conference?
Tweets Share RSS iTunes

Tune in to Collaboration Pizza Thursdays at 3:30 pm ET.

ANALYST NAMES YAKABOD "COOL VENDOR"
The 2010 High-Performance Workplace list.
Available here.

NEWEST YAKABOX FEATURES
★ Framework for slick custom apps, microblogs, search shortcuts, task options.
Read more.

SUBSCRIBE TO THE ROI REPORT
Curious about how other organizations collaborate? Get The ROI Report. [Subscribe.](#)

Meet the Yakabox, the knowledge sharing system that can keep a secret.

So robust you can't outgrow it. So straightforward you're productive day one. So secure your secrets are safe. Nothing comes close to the Yakabox. Collaborate with co-workers and business partners. Manage events and workflow. Find, share and create stuff that matters *without* plowing through stuff that doesn't. And do it all on the platform certified by the U.S. government to handle top secret information. The industry's top analyst calls the Yakabox "the secure workspace for high-performance teams" for good reason.

<http://yakabod.com/>

Die Funktionalität erklärt sich auch hier fast von selbst:

This is what you can do with a Yakabox.



Interessant ist hier ein Vergleich, den Yakabod selbst anbietet und der vor allem die o. g. etablierte Konkurrenz einbezieht:

	Yakabox 3.3	SharePoint 2010	Lotus Connections 2.5	Jive SBS 4.0
7-10 day custom app turnaround	✓			
Browser support for IE, Firefox, Chrome, Safari	✓			Partial
Native web 2.0 architecture	✓			✓
Customization and extensibility	REST/SOAP	SP-proficient developers required	Websphere App Server, plug-ins to IBM & MS	REST/SOAP
Mobile access	✓	✓	✓	✓
Content management, collaboration, social networking, search	✓	✓	✓	✓
Limited/no training needed for routine tasks	✓			✓
Internal & external community integration				✓
Project management	Basic	✓	Basic	Basic
Workflow/task management	Basic	✓	Files & tasks only	Add-on
Follow favorites (topics, people, etc.)	✓			✓
Integrated IM, chat, presence, microblogs, tags	✓		Extra	
Live, unified search: find anything in the repository from a single interface	✓		People only	✓
Filter, refine and save searches by date range, content type, teams, people, tags	✓			
Keyword search generates suggestions for related material	✓		People only	✓
User-defined open and closed workspaces	✓	✓	Limited access control	✓

Man darf gespannt sein, wann deutsche Unternehmen diese Art von Software umfassend einsetzen. Übrigens gibt es nicht nur Werkzeuge, welche die gesamte Bandbreite sozialer Software für den Enterprise-Bereich abdecken, sondern manche bieten nur bestimmte Funktionen, so z. B. StatusNet nur Microblogging: „The StatusNet platform enables

communities, brands and organizations to incorporate micro messaging into their own web domain. StatusNet's plug-in architecture provides a feature-rich platform with functionality that you control. Flexible and customizable, StatusNet delivers a micro messaging solution that helps you maintain brand identity and a unique user experience. StatusNet offers two product options: a hosted version and a software download. The Status.net hosted service enables organizations to get up and running quickly on our secure servers, while the software download version provides the maximum in control and customization.”

<http://status.net>

Bei den sozialen Netzwerken für den eher privaten Gebrauch tut sich auch etwas, insbesondere durch die Neuentwicklungen Path und Diaspora: Path ist derzeit nur als App erhältlich. Die Webseite ist dementsprechend wenig informativ; man muss schon auf andere Quellen ausweichen um zu erfahren, was hier Sache ist. Gründer sind der ehemalige Facebook Platform Manager Dave Morin sowie Napster-Gründer Sean Fanning. Aus dem Stealth-Modus hervorgetreten entpuppt sich Path als mobile soziale Foto-App, die laut ReadWriteWeb erheblich weniger kann als erwartet.

<https://www.path.com>

So war denn dort auch nachzulesen, was Path nicht kann:

1. You can't comment on or Like photos.
2. You can't publish photos from your phone's camera roll.
3. You can't publish to other networks like Flickr (for archiving), Facebook (for broader social sharing) or Foursquare (for location check-ins).
4. You can't find friends via your accounts elsewhere, like Facebook, Twitter or Gmail.
5. You can't navigate by the tags you've added to photos - it's unclear whether you can even add more than one Thing tag to a photo.
6. You can't see tags or captions once you've got a photo expanded.
7. You can't opt-out of email notifications from Path - and though emails say you can "pause" individual users, it's not clear how to do that.
8. You can't visit a person's Path profile page if you're not already friends with them.
9. You can't apply any filters or crop photos.
10. You can't add friends on the web site, not even if they've already added you.

Man fand die App dennoch interessant; für uns gilt, dass man nur mit amerikanischer Kreditkarte im App-Store Path herunterladen kann. ☹

Ein „Facebook-Killer“ könnte Diaspora werden, das soziale Netzwerk, das als Open Source daherkommt und Mitte September seine Software freigegeben hat. Am 16.09.2010 war im ReadWriteWeb unter der Überschrift “Startup Versus Goliath: When Your Competition is a Giant” zu lesen: „As promised, open source social network and Facebook alternative Diaspora released its source code yesterday. And while it's a developer release meant to be hacked on and is by no means a finished, there are already of plenty of predictions that Diaspora will fail – or at the least, that the project represents no threat to Facebook. Diaspora is hardly unique as a startup that faces major challenges by entering into a market or an industry where big companies are well-established.

DIASPORA* ALPHA @joindiaspora github Blog Anmelden



Teile was du willst, mit wem du willst.

Entscheidungsfreiheit <p>Diaspora ermöglicht es dir, deine Kontakte in sogenannten Aspekten zu verwalten. Das Besondere an Diaspora ist, dass du mit Aspekten kontrollieren kannst, mit welchen Menschen du deine Fotos, Geschichten und Witze teilst.</p>	Eigentum <p>Deine Bilder gehören dir und du solltest das nicht aufgeben müssen, um sie zu teilen. Bei Diaspora gehört dein Bild immer dir und du hast die volle Kontrolle, wie es vervielfältigt wird.</p>	Unkompliziert <p>Diaspora macht das Teilen einfach und strukturiert - das gilt ganz besonders für deine Privatsphäre. Diaspora ist von sich aus privat; es gibt deswegen keine komplizierten Einstellungen um dein Profil sicher zu machen.</p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<https://joindiaspora.com/>

On a recent TechZing podcast, for example, hosts Jason Roberts and Justin Vincent asked Gabriel Weinberg, founder of the search engine DuckDuckGo if being in the same realm as Google "isn't that, like, crazy?" But as Weinberg points out, he's able to "do things that Google can't copy easily." For example, DuckDuckGo is able to deliver more satisfactory search results than Google, contends Weinberg, because he can actively address questions of spam without facing charges of censorship or anti-trust - something Google can't do because of its size. . . . Going up against major companies can be daunting for a startup. And perhaps, as TechZing suggests, it's crazy. But too often, when we pit startups against their big business competitors, we assess them on whether or not they're a Facebook or Google "killer." DuckDuckGo needn't "kill" Google. Diaspora needn't "kill" Facebook. They need to fill a niche, and do it in a way that their giant competitors cannot."

Zum Abschluss noch ein Hinweis auf einen Blog-Eintrag meines St. Galler Kollegen Hans-Dieter Zimmermann, den man unter <http://networkedblogs.com/9khJm> findet: Das berühmte Schweizer Messer gibt es mittlerweile auch in elektronischer Form (aber auch die anderen Blog-Einträge sind lesenswert).



Rezension

Schönthaler, Frank; Vossen, Gottfried; Oberweis, Andreas; Karle, Thomas: Geschäftsprozesse für Business Communities, Modellierungssprachen, Methoden, Werkzeuge, München, 2011, Oldenbourg Verlag, ISBN 978-3-486-59756-1, Preis: 39,80 Euro

In den letzten zwei Jahrzehnten sind zahlreiche Bücher, Aufsätze und sonstige Beiträge zum IT-gestützten Geschäftsprozessmanagement verfasst worden. Bei der Vielzahl der Titel stellt sich natürlich die Frage, ob überhaupt noch Raum für ein weiteres umfassendes Buch vorhanden ist? Von daher ist es durchaus spannend, einen Blick in das neue Werk der langjährigen EMISA-Mitglieder Oberweis und Vossen und deren Co-Autoren zu werfen, um diese Frage zu beantworten.

Schon das Vorwort weckt Interesse: Ein Blick in die berufliche „Modellierungs-Vergangenheit“ der Autoren zeigt, das in den 90er Jahren die damaligen „Jungforscher“ in etwa folgende Gedanken gehabt haben: Man erstelle ein mathematisch fundiertes Modell eines Geschäftsprozesses, setze es in ein lauffähiges Konstrukt um (z.B. mit Petri-Netzen), simuliere das Modell bis das betriebswirtschaftliche „Optimum“ für den Prozess gefunden ist (Kriterien sind beispielsweise: Durchlaufzeit, Kosten, Auslastung von Ressourcen) und schon steht eine leicht zu wartende lauffähige Applikation für den Einsatz bereit.

Heute wissen wir es besser: Die reale Welt weigert sich beständig in ein „Modell-Korsett“ gezwängt zu werden. Die verfügbaren Informationen sind typischerweise unscharf. Die Änderungsfrequenz von Prozessen ist zum Teil sehr hoch. Sobald ein Prozessmodell entwickelt wurde, ist die Realität wieder einen Schritt weitergegangen. Trotzdem gibt es viele Beispiele in Unternehmen, bei denen modellgesteuerte Prozesse „gefahren“ werden. Daher wird es auch in Zukunft wieder Forscher geben (müssen), die sich der Aufgabe der modellgestützten Prozessgestaltung und -steuerung stellen, die bestehenden Methoden verfeinern oder neue Methoden entwerfen. Das Buch möchte hierzu motivieren, indem es ein praxiserprobtes Konzept vorstellt.

Die Gliederung folgt einem gut nachvollziehbaren Top down-Ansatz. Ausgehend von einführenden Aspekten steht eine spezielle Methode des Geschäftsprozessmanagements im Vordergrund:

- 1 Einführung
- 2 Praktische Einführung in Business Process Management
- 3 Konzepte und Modellierungssprachen
- 4 Die Horus Methode
- 5 Anwendungsbereiche
- 6 Zukunft des Business Process Engineerings

Im ersten Kapitel führen die Autoren mit Hilfe eines realistischen Fallbeispiels in die Modellierungsthematik ein und zeigen auf, wie wichtig die Visualisierung von Prozessen ist. Das zweite Kapitel



kann isoliert gelesen werden, es ist eine sehr kompakte Einführung in das „Business Process Engineering“ basierend auf Petri-Netzen nebst Hinweisen auf typische weiterführende Literatur. Das dritte Kapitel widmet sich den Modellierungssprachen. Behandelt werden grundlegende Aspekte, Sichten, einfache und komplexe Modellierungselemente der Petri-Netze. Dieses eher lehrbuchorientierte Kapitel wird ergänzt durch mehrere Übungsaufgaben, die sich primär an den akademischen Nachwuchs richten. Anschließend behandeln die Autoren im deutlich umfangreicheren 4. Kapitel die von Ihnen mitentwickelte Horus(R)-Methode™. Sie besteht aus einem umfangreichen Werkzeugkasten, der u.a. ein Phasenmodell, Modellierungs- und Simulationskonzepte sowie eine Werkzeugunterstützung beinhaltet.

Dieser integrative Ansatz unterscheidet das Buch von anderen Monographien zum Geschäftsprozessmanagement, die meist Schwerpunkte auf ausgewählten Bereichen des Geschäftsprozessmanagements setzen (Modellierung, Praxis, Einführungskonzepte u.a.). Abschließend behandelt die Autoren verschiedene Anwendungsbereiche (Kapitel 5) und die Zukunft des Business Process Engineering (Kapitel 6).

Fazit: Insbesondere fortgeschrittene Studierende profitieren von zahlreichen Übungen und Literaturverweisen, BPM-Praktiker werden an den umsetzbaren praxisorientierten Konzepten Interesse finden.

Die eingangs thematisierte Frage kann einfach beantwortet werden: Das Buch gehört in den Bücherschrank in die Rubrik „IT-gestütztes Geschäftsprozessmanagement“

Andreas Gadatsch, Sankt Augustin

Sprecher der mit der EMISA assoziierten Arbeitskreise und Fachgruppen

Prof. Dr. Jörg Desel

FG Petrinetze und verwandte Systemmodelle

FernUniversität in Hagen
Lehrgebiet Softwaretechnik und Theorie der Programmierung
Universitätsstraße 1
58097 Hagen
Tel. +49 (0)2331 987-2609
E-Mail: joerg.desel@fernuni-hagen.de

Prof. Dr. Stefan Klink

AK Entwicklung Digitaler Bibliotheken

Duale Hochschule
Baden-Württemberg Karlsruhe
Erzbergerstr. 121
76133 Karlsruhe
Tel.: +49 (0)721-9735951
E-Mail: klink@dhw-karlsruhe.de

Horst Kremers

Comittee on Data for Science and Technology (CODATA-Germany)

Postfach 200548
D-13515 Berlin

Tel: +49 (0)172 3211738
E-Mail: office@horst-kremers.de

Prof. Dr. Markus Nüttgens

AK Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten

Universität Hamburg
WISO Fakultät, Wirtschaftsinformatik
Von-Melle-Park 9
D-20146 Hamburg
Tel: +49 (0)40-428382792
E-Mail: Markus.Nuettgens@wiso.uni-hamburg.de

Prof. Dr. Bernd Müller

AK Informationssysteme mit Open Source

Fachhochschule Braunschweig/Wolfenbüttel
Fachbereich Informatik

Am Exer 2
D-38302 Wolfenbüttel
Tel: +49 (0)5331-9396313
E-Mail: Bernd.Mueller@fh-wolfenbuettel.de

Prof. Dr. Klaus Turowski

FG Modellierung betriebl. Informationssysteme

Universität Augsburg
Lehrstuhl für Wirtschaftsinformatik und Systems Engineering
Universitätsstr. 16
D-86159 Augsburg
Tel. +49 (0) 821 598-4431
E-Mail: Klaus.Turowski@wiwi.uni-augsburg.de

Priv.-Doz. Dr. Barbara Weber

AK Flexible service- und prozessorientierte Informationssysteme

Universität Innsbruck
Institute of Computer Science
Quality Engineering Group
Technikerstrasse 21a
A-6020 Innsbruck
Tel.: +43 (0)512-5076474
E-Mail: Barbara.Weber@uibk.ac.at

EMISA-Leitungsgremium

Das im September 2009 gewählte EMISA-Leitungsgremium besteht aus den Sprechern der assoziierten Arbeitskreise und Fachgruppen (siehe vorige Seite) sowie folgenden gewählten Mitgliedern:

<p>Prof. Dr. Mathias Weske (<i>Sprecher des Leitungsgremiums</i>)</p> <p>Universität Potsdam Hasso-Plattner-Institut für Softwaresystemtechnik Prof. Dr. Helmert-Str. 2-3 D-14482 Potsdam Tel: +49 (0)331-5509191 E-Mail: Mathias.Weske@hpi.uni-potsdam.de</p>	<p>Prof. Dr. Manfred Reichert (<i>Stellv. Sprecher, Redaktion EMISA FORUM</i>)</p> <p>Universität Ulm Institut für Datenbanken und Informationssysteme James-Franck-Ring, Geb. O27 D-89081 Ulm Tel: +49 (0)731-5024135 E-Mail: Manfred.Reichert@uni-ulm.de</p>
<p>Fernand Feltz</p> <p>Centre de Recherche Public - Gabriel Lippmann Département Informatique, Systèmes et Collaboration (ISC) 41, rue du Brill, L-4422 Belvaux, Luxembourg Tel. +352 (0)470261-600 E-Mail: Feltz@lippmann.lu</p>	<p>Prof. Dr. Ulrich Frank</p> <p>Universität Duisburg-Essen Lehrstuhl für Wirtschaftsinformatik und Unternehmensmodellierung Universitätsstr. 9 D-45141 Essen Tel: +49 (0)201-1834042 E-Mail: Ulrich.Frank@uni-duisburg-essen.de</p>
<p>Prof. Dr. Andreas Gadatsch</p> <p>Fachhochschule Bonn-Rhein-Sieg Fachbereich Wirtschaft Grantham-Allee 20 D-53757 Sankt Augustin Tel. +49 (0)2241-865129 E-Mail: Andreas.Gadatsch@fh-bonn-rhein-sieg.de</p>	<p>Prof. Dr. Jan Mendling</p> <p>Humboldt-Universität zu Berlin Wirtschaftswissenschaftliche Fakultät Institut für Wirtschaftsinformatik Spandauer Straße 1, 10178 Berlin, Germany Tel. +49 (0)30 2093 5805 Email: Jan.Mendling@wiwi.hu-berlin.de</p>
<p>Prof. Dr. Andreas Oberweis</p> <p>Universität Karlsruhe Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB) D-76128 Karlsruhe Tel. +49 (0)721-6084516 E-mail: Oberweis@aifb.uni-karlsruhe.de</p>	<p>Dr.-Ing. Hansjürgen Paul</p> <p>Institut Arbeit und Technik Forschungsschwerpunkt WISDOM Munscheidstraße 14 D-45886 Gelsenkirchen Tel. +49 (0)209-1707229 E-Mail: Paul@iat.eu</p>
<p>Prof. Dr. Stefanie Rinderle-Ma</p> <p>Universität Wien Fakultät für Informatik Workflow Systems and Technology Rathausstraße 19/9 A - 1010 Wien Tel. : +43 1 4277 39517 E-mail: Stefanie.Rinderle-Ma@univie.ac.at</p>	<p>Prof. Dr. Gottfried Vossen</p> <p>Westfälische Wilhelm-Universität Münster Database Group Leonardo-Campus 3 D-48149 Münster Tel.: +49 (0)251-8338150 E-Mail: Vossen@uni-muenster.de</p>