

# OD.FMI – Engineering operativ-administrativer Daten für die universitäre Lehre

Hans-Gert Gräbe, Michael Martin, Thomas Riechert  
{graebe|martin|riechert}@informatik.uni-leipzig.de

Universität Leipzig, 04009 Leipzig

**Abstract:** Die Weiterentwicklung bestehender E-Learning-Strukturen erfordert neben der Entwicklung neuer Werkzeuge die Integration vorhandener Datenbestände. Damit verbunden ist ein Paradigmen-Wechsel von einer werkzeug- zu einer datengetriebenen Entwicklungsmethodik und semantischen Technologien der Datenmodellierung. Dieser Artikel stellt am Anwendungsfall, der Open Data Plattform der Fakultät für Mathematik und Informatik der Universität Leipzig (OD.FMI), die formale semantischen Strukturierung operativ-administrativer Daten zur Planung und Abwicklung von Lehrveranstaltungen vor. Im Fokus steht dabei eine leichtgewichtige und agile Knowledge Engineering Methodik für derartige Szenarien.

## 1 Einleitung

E-Learning-Portale wie das OLAT<sup>1</sup>-basierte sächsische Bildungsportal OPAL<sup>2</sup> fokussieren stark auf die interaktive Gestaltung der Lehr- und Lernprozesse *innerhalb eines Kurses* und erlauben es, in diesem Bereich zielgruppengenau Lehrmaterialien zu präsentieren und flexible Lernszenarien umzusetzen. Für die Koordination verschiedener operativ-administrativer Aufgaben auf *kursübergreifender* Ebene bieten derartige Portale dagegen nur rudimentäre Unterstützung an. So beschränken sich die wesentlichen OLAT-Funktionen in dieser Richtung auf den Katalog, in welchem die Struktur des Lehrangebots abgebildet werden kann, sowie die privaten Kalender, die mit Kurs- und Gruppenkalendern verbunden werden können.

Im Gegensatz zum Kursbetrieb, in dem mit Kursleitern (Autoren), Studenten (Nutzer) und Tutoren eine überschaubare Anzahl von weitgehend standardisierten Rollen mit gut abgrenzbaren Verantwortungsbereichen und Interaktionslinien identifiziert werden können, ist Anzahl und Zuschnitt von weiteren Rollen im operativ-administrativen Geschäft einer Fakultät in weiten Bereichen nicht standardisiert und übergreifend nicht standardisierbar. Dies ergibt sich vor allem aus der verschiedenen Größe von Studiengängen und den kulturellen Traditionen der einzelnen Fächer.

Andererseits werden operativ-administrative Prozesse auf Fakultätsebene (wie Prüfungsmanagement, Lehrveranstaltungs- oder Raumplanung) durchaus werkzeuggestützt vorbe-

---

<sup>1</sup>OLAT – Your Open Source Learning Management System. <http://www.olat.org>

<sup>2</sup><http://bildungsportal.sachsen.de>

reitet und abgewickelt. Die eingesetzten Werkzeuge dominieren heute oft die Datenstrukturierung und begrenzen außerdem die Interkommunikationsmöglichkeiten zwischen den verschiedenen Anwendungen, so dass eine Integration – wenn überhaupt auf die entsprechenden Spezifikationen zugegriffen werden kann – nur mit hohem Ressourcenaufwand möglich ist. Einrichtungsübergreifende Anstrengungen, solche Integrationslösungen für einzelne Werkzeuge gemeinsam zu entwickeln, werfen komplexe Fragen auf. Solche Interaktionsfragen auf Werkzeugebene stehen zum Beispiel im Fokus des Projekts *Integrate-IT* des Sächsischen Arbeitskreises E-Learning<sup>3</sup>.

Neue Softwareprojekte im operativ-administrativen Bereich starten heute meist nicht mehr auf der „grünen Wiese“ der Erstbeschaffung wie noch vor einigen Jahren, sondern treffen auf eine bereits existierende Werkzeug- und Datenlandschaft, die erweitert oder fortgeschrieben werden soll. Oft sind dabei alte Werkzeuge durch neue zu ersetzen, aber dabei vorhandene Datenbestände zu erhalten oder zu migrieren, damit diese – möglicherweise in neuen Formaten – weiter operativ verfügbar bleiben oder wenigstens mit geringem Aufwand aus einem Archiv beschaffbar sind. Herausforderungen in Bezug auf Daten-Interoperabilität ergeben sich auch, wenn für die eigenen operativen Prozesse Werkzeuge und Dienstleistungen verschiedener Hersteller und Anbieter zum Einsatz kommen, die auf einen gemeinsamen operativ-administrativen Datenbestand zugreifen sollen.

Eine solche historisch gewachsene Werkzeugvielfalt ist heute in den meisten Einrichtungen Realität und spielt mit Blick auf eine anzustrebende Anbieterflexibilität im Kontext der nachhaltigen Sicherung eigener Handlungsfähigkeit in strategischen Überlegungen eine zunehmend wichtige Rolle. Eine zentrale Herausforderung ist dabei, operative Daten und deren Verknüpfungen untereinander auf flexible Weise so darzustellen und zu formalisieren, dass mit verschiedenen Werkzeugen und Komponenten dieser gemeinsame Datenbestand kontrolliert verwaltet und unter definierbaren Integritäts- und Qualitätssichtspunkten fortgeschrieben werden kann.

In diesem Artikel analysieren wir im Abschnitt 2 die Rahmenbedingungen für typische Engineering-Prozesse in heterogenen E-Learning Landschaften. Im Abschnitt 3 werden kursorisch semantische Technologien vorgestellt, die in den folgenden Abschnitten in unserem Anwendungsfall, der formalen semantischen Strukturierung operativ-administrativer Daten zur Planung und Abwicklung von Lehrveranstaltungen an der Fakultät für Mathematik und Informatik der Universität Leipzig, angewendet werden. Im Rahmen des Projekts wurde eine agile Methodik des Knowledge Engineering entwickelt, die in Abschnitt 5 beschrieben ist. Das entstandene und sich stetig verändernde Vokabular kann hier aus Platzgründen nicht im Detail vorgestellt werden. Die darauf aufbauende Datenintegration in die IT-Infrastruktur der Fakultät wird im Abschnitt 6 erläutert.

---

<sup>3</sup>[http://bildungsportal.sachsen.de/e135/e1466/e2507/index\\_ger.html](http://bildungsportal.sachsen.de/e135/e1466/e2507/index_ger.html)

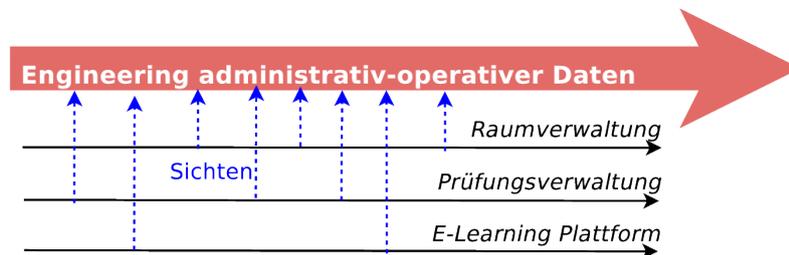


Abbildung 1: Datengetriebene Koevolution von Daten und Werkzeugen

## 2 Engineering heterogener E-Learning Infrastrukturen. Eine Analyse der Rahmenbedingungen

Ausgehend von der klassischen *Drei-Schichten-Architektur* – bestehend aus Datenschicht, Geschäftslogik und Präsentationsschicht – betrachten wir insbesondere das Konzept der Interaktion zwischen Datenschicht und Geschäftslogik in Richtung größerer Flexibilität der Datenstrukturierung. Die Datenschicht muss entsprechend der im letzten Abschnitt hergeleiteten komplexeren Anforderungen periodisch überdacht und umgebaut werden. Dazu bedarf es formalisierter Beschreibungsmittel, also einer Datenabstraktionssprache, um aus der Geschäftslogik heraus die entsprechenden Aufgaben möglichst laufzeitnah *formulieren* zu können.

Aus softwaretechnischer Perspektive wird im Wechselspiel von Werkzeugsicht und Datensicht auch eine Zerteilung von Verantwortlichkeiten sichtbar. Während IT-Firmen oder entsprechende zentrale Abteilungen der Hochschulen als Auftragnehmer für Erstellung und Wartung der entsprechenden Softwaresysteme primär für das reibungslose Funktionieren und die Weiterentwicklung der Werkzeuge zuständig sind, ist der Auftraggeber vor allem an der konsistenten Fortschreibung der Datenlandschaft der von ihm zu verantwortenden operativ-administrativen Prozesse interessiert und für deren Zustand auch verantwortlich. Damit interessiert ihn die Frage nach konkreten Werkzeugen nicht vordergründig. Diese Gesamtverantwortung für eine konsistente Datenlandschaft wird oft in eine Reihe von Teilverantwortlichkeiten heruntergebrochen, die konkreten Mitarbeitern des Auftraggebers zusammen mit *Rollen* zugeordnet werden. Die damit verbundene lokale Verantwortung für Teilbereiche der Datenintegrität *innerhalb* des bestehenden IT-Systems steht im schlechtesten Fall im Konflikt mit der operativ-administrativen Herausforderung, einen konkreten Fall unter Heranziehung von Daten aus *verschiedenen* Verantwortungsbereichen so zu bearbeiten, dass er sachgerecht entschieden wird *und* die dabei anfallenden Daten in das bestehende System integriert werden können. Ein Ansatz, der den lokalen Akteuren dabei nicht auch die Möglichkeit einräumt, sachlich veranlasste Veränderungen an der *Datenstrukturierung* auf kontrollierte Weise zu veranlassen oder selbst vorzunehmen, skaliert nicht.

Der operativ-administrative Datenkorpus *lebt* also nicht nur in dem Sinne, dass sich Dateninhalte ändern, sondern – über einen längeren Zeitraum betrachtet – sind auch Datenstrukturierungen weiterzuentwickeln (vgl. Abbildung 1). In klassischen, über relatio-

nale Datenbanken realisierten Softwareprojekten bedeutet eine solche Änderung der Datenstrukturierung meist einen tiefen Eingriff auf der Ebene von Datenbankschemata und löst umfangreiche Arbeiten auf Werkzeugebene aus. Dies kann bei hochgradig monolithischen Lösungen nur im Rahmen von einschneidenden Release-Wechseln ausgerollt werden, während bei komponentenbasierten Lösungen die entsprechenden Änderungen vielleicht auch schrittweise eingespielt werden können. Für eine primär werkzeuggetriebene Entwicklungsmethodik in jedem Fall eine gravierende Änderung.

Der *lebende Datenkorpus* ist jedoch ein Abbild der *einen großen realweltlichen Erzählung* über die operativ-administrativen Vorgänge im entsprechenden Einzugsbereich, die mit verteilten Rollen und Stimmen vorgetragen wird. Könnte diese Erzählung ausreichend detailliert formalisiert erfasst werden, so wäre dies eine erstklassige Quelle wenigstens für alle Akteure, die in dieser Domäne betroffen sind. Deren Werkzeuge könnten an dieser Formalisierung andocken und die für sie relevanten Sichten herausfiltern. Eine solche primär datengetriebene Entwicklungsmethodik bietet also viele Vorteile. Genau dieser Weg wird mit modernen semantischen Ansätzen beschrrieben.

### 3 Technologien des Semantic Webs zur Wissenmodellierung

Die in den letzten Jahren entwickelten Ansätze im Bereich der sematischen Datenstrukturierung können hier nicht im Detail diskutiert oder auch nur einigermaßen umfassend dargestellt werden. Die wichtigsten Elemente sind

- die Einführung von Unicode und UTF zur plattform-übergreifenden einheitlichen Zeichendarstellung,
- die Standardisierung von XML als Markup-Sprache, mit der einzelne semantisch bedeutsame Textteile markiert werden können,
- RDF<sup>4</sup> und RDF-S<sup>5</sup> [MM04, BG04] als (XML-basierte) Beschreibungssprache, um diese semantische Bedeutung im Rahmen von *Ontologien* formalisiert auszudrücken und damit entsprechende *Wissensbasen* aufzubauen,
- OWL<sup>6</sup> [BvHH<sup>+</sup>04] als (ebenfalls XML-basierte) Beschreibungssprache, um diese Ontologien selbst formal zu beschreiben und damit einem maschinengestützten *logischen Reasoning* zugänglich zu machen und
- Reasoning-Konzepte, um in Wissensbasen implizit vorhandene Informationen durch Verknüpfungen zu explizieren.

Insbesondere die Abtrennung der Ebene der Ontologien von der Ebene der Datenhaltung eröffnet neue Möglichkeiten der verteilten Verwaltung und Fortschreibung von Datenbeständen entsprechend operativer Verantwortlichkeiten, wenn Einigkeit über eine gemeinsame Ontologie hergestellt werden kann.

<sup>4</sup>Resource Description Framework: <http://www.w3.org/RDF/>

<sup>5</sup>RDF-Schema: <http://www.w3.org/TR/rdf-schema/>

<sup>6</sup>Web Ontology Language: <http://www.w3.org/TR/owl-features/>

Webgestützte Zugriffskonzepte gestatten es, Wissensbasen dezentral im Zuständigkeitsbereich der jeweiligen Verantwortlichen anzusiedeln. Über Lese- und Schreibrechte entsprechend der verschiedenen Rollenbilder kann so der Zugriff auf die operative Datenbasis feingranular gesteuert werden. Eine solche Infrastruktur setzt allein eine zentrale Authentifizierungsstruktur voraus, mit der realweltlichen Akteuren ihre digitalen Identitäten sowie die zugehörigen Rechte zugeordnet werden können. Eine solche Infrastruktur existiert aber in den meisten Einrichtungen bereits zusammen mit der Vernetzung der Arbeitsplatzrechner und der zentralen Verwaltung dieses Netzzugangs.

Auf der Ebene des logischen Reasoning hat sich in den letzten Jahren der Fokus von schwergewichtigen Konzepten der künstlichen Intelligenz hin zu stärker praktischen Fragen der Übertragung klassischer SQL-Anfrageszenarien verschoben. Mit der hierfür entwickelten SPARQL-Anfragesprache<sup>7</sup> [PS08] können verteilte Wissensbasen mit gemeinsamen Ontologien oder Ontologiefragmenten wesentlich flexibler durchsucht werden. Dezentrale Anfrageszenarien lassen sich dazu mit *Mashup-Techniken* zu verschiedenen Zwecken zusammenführen.

So wie es im Bereich der klassischen Datenbanken eines entsprechenden Servers mit einer Datenbank-Engine bedarf, um Antworten auf Datenbankabfragen zu erhalten, muss ein Webserver einen *SPARQL-Endpunkt* anbieten, um SPARQL-Anfragen an eine dort verwaltete Datenbasis verarbeiten und Antworten verschicken zu können.

## 4 Anforderungen an das Projekt OD.FMI

Ausgangspunkt des Open Data Projekts OD.FMI an unserer Fakultät waren

- Konzeptionelle Erfahrungen in der Strukturierung semantischer Daten aus verschiedenen Bereichen und die Entwicklung entsprechender Werkzeuge.

So entwickelten wir im Projekt *SoftWiki*<sup>8</sup> [ALLR09] eine Methodologie zur verteilten Anforderungserfassung in Software-Projekten. Im Projekt *Catalogus Professorum Lipsiensis*<sup>9</sup> [RMA<sup>+</sup>10, MR10] realisierten wir eine semantische Wissensdatenbank in der Geschichtswissenschaft.

- Erfahrungen aus dem Betrieb einer eigenen OLAT-basierten E-Learning-Plattform<sup>10</sup> für die Fakultät, über die neben der Begleitung von Lehrveranstaltungen insbesondere das Prüfungsanmeldungsmanagement<sup>11</sup> abgewickelt wird.
- Übernahme der Verantwortung für die semesterweise Planung der Lehrveranstaltungen auf Fakultätsebene durch den ersten Autor zum Sommersemester 2010.

---

<sup>7</sup>SPARQL Query Language <http://www.w3.org/TR/rdf-sparql-query/>

<sup>8</sup>Siehe auch: <http://projekte.softwiki.de>

<sup>9</sup>Professorenkatalog der Universität Leipzig: <http://catalogus-professorum.org>

<sup>10</sup><http://olat.informatik.uni-leipzig.de>

<sup>11</sup><http://olat-pa.informatik.uni-leipzig.de>

Besonders mit der Übernahme der letzteren Aufgabe war das weitgehend nur informell vorhandene Wissen über die Vorgänge und Zusammenhänge der für die Planung relevanten operativ-administrativen Daten neu zu erschließen und über den Umgang mit Inkonsistenzen in diesem Datenmaterial – etwa der ewigen Frage nach der Aktualität der veröffentlichten Modulbeschreibungen – nachzudenken. Es bot sich an, bekannte theoretische Konzepte und in anderen Projekten erprobte semantische Beschreibungstechniken auf diesen Anwendungsfall zu übertragen.

Zu Beginn wurde das Verhältnis der zu entwickelnden neuen Datenbasis zum „Altdatenbestand“ im Rahmen des BIS-OLAT-Portals und dort vor allem zur Übersicht, die im Katalog aggregiert ist, analysiert. Diese Kataloginformationen werden in der Version OLAT 6.3 nach einem fest vorgegebenen Schema aus den Kursen extrahiert. Neben den damit verbundenen Einschränkungen, welche Informationen überhaupt kursübergreifend aggregierbar sind, wird auch der entsprechende operativ-administrative Prozess nur ungenügend abgebildet, weil der auf Einrichtungsebene Verantwortliche für die Lehrveranstaltungsplanung nicht einbezogen ist. Das bei diesem Verantwortlichen *strukturiert* (wenn auch nicht unbedingt schon formalisiert) vorhandene Gesamtwissen wird durch das partielle Wissen der Kursverantwortlichen über diesen Prozess ersetzt, die zudem in ihrer Mehrzahl das OLAT-Portal nicht nutzen. Entsprechend unvollständig geben die Kataloginformationen das volle Spektrum des Lehrangebots wieder.

Für die vollständige und konsistente Darstellung des Lehrveranstaltungsangebots unserer Einrichtung innerhalb von OLAT müsste dieser Workflow umgekehrt werden: aus Quellen über das gesamte Lehrangebot, die bei der Planung oder zur digitalen Darstellung des Lehrangebots der Einrichtung auf unseren Webseiten sowieso erstellt werden und wenigstens teilformalisiert vorliegen, wäre zunächst der Katalog zu generieren. In einem zweiten Schritt hätten die Kursleiter Gelegenheit, ihre eigenen Kurse mit den entsprechenden Katalogeinträgen zu verlinken.

Hierfür muss nicht der komplette Prozess der Stundenplanverwaltung in OLAT integriert sein; insbesondere muss keine weitere OLAT-Rolle „Stundenplanverwalter“ geschaffen werden. Es reicht aus, wenn eine solche Rolle auf Einrichtungsebene besetzt ist, die eine öffentlich zugängliche Datenquelle verwaltet und aktualisiert, in der die entsprechenden Informationen zu den Lehrveranstaltungen in einer bekannten Weise strukturiert abgelegt sind. Dann kann OLAT um einen Modul erweitert werden, der auf diese externe Quelle mit Stundenplaninformationen *lesend* zugreift, um daraus eine eigene Sicht auf die Stundenplandaten der Einrichtung zu generieren und diese Informationen mit anderen Informationen zu verknüpfen. Beim Entwurf einer solchen OLAT-Erweiterung ist zu bedenken, dass das Änderungsmanagement dieser Quelle – etwa durch ein tägliches Update der primären Stundenplandaten – geeignet propagiert werden muss.

Es ist klar, dass sich dieses Konzept auf andere Bereiche der operativ-administrativen Verwaltung öffentlicher Daten der Einrichtung (Mitarbeiter der Einrichtung, Modulbeschreibungen, Raumsituation usw.) ausdehnen lässt.

Damit ist der Rahmen abgesteckt, in welchem die Datenerfassung und -strukturierung des OD.FMI Projekts erfolgt. Als „Proof of Concept“ und um das Zusammenspiel einer solchen Datenbasis mit Werkzeugen zu erproben, sollten zugleich Werkzeuge entwickelt wer-

den, mit denen sich die folgenden Aufgaben abdecken lassen:

- (A) Transparentere Darstellung des aktuellen Stands der Planung der Lehrveranstaltungen für das kommende Semester, so dass sich die Mitarbeiter der Fakultät stets aktuell informieren können (Anforderung an die Entwickler im Projekt selbst),
- (B) Einführung eines neuen Stundenplans, der auf die neue Datenbasis zugreift (Anforderung an den Webmaster der Fakultät),
- (C) Prototypische Implementierung einer OLAT-Sicht auf die Stundenplandaten (Anforderung an ein studentisches Team im Rahmen des Softwaretechnik-Praktikums im Sommersemester 2010).

Die für diesen Ansatz erforderliche Rollen- und Rechtesituation ist – wie für die meisten Open Data Projekte – übersichtlich. Da es in diesem Projekt primär um die Strukturierung öffentlich verfügbarer operativ-administrativer Daten geht, steht die Frage der Beschränkung des lesenden Zugriffs auf diese Daten nicht. Im Zuge der Projektrealisierung stellte sich im Gegenteil heraus, dass das öffentliche Abwickeln insbesondere der Planung der Lehrveranstaltungen, wo sich die Kollegen stets online über den aktuellen Planungsstand informieren können, die Transparenz und Qualität der Planung und damit die Zufriedenheit der Betroffenen deutlich gesteigert hat.

Auch die Frage des Managements von Schreibrechten stand nicht, da der explorative Charakter des Ansatzes sowie Fragen der Datenqualität in dieser frühen Phase des Projekts nahe legen, dass nur eine Person schreibenden Zugriff auf die Daten hat, der zudem ein komplexes Editierwerkzeug<sup>12</sup> zugemutet werden kann. Zum Einsatz kam eine Instanz des im Rahmen des Softwiki Projekts [ALLR09] entwickelten OntoWiki [ADR06]. Damit bestand die Möglichkeit, sich auf die Erstellung der Datenbasis und die Ausgestaltung einer entsprechenden Struktur zu konzentrieren und dabei die bisher an der Abteilung entwickelten semantischen Werkzeuge zum Einsatz zu bringen.

## 5 Leichtgewichtige und agile Datenmodellierung

Um Informationen einer Domäne auf der Basis von RDF und OWL abzubilden und diese mit anderen Wissensbasen im Semantic Web zu verlinken, wird ein Vokabular zur Abbildung des Schemas der Informationen benötigt. Klassische Ansätze zur semantischen Strukturierung realweltlicher Daten starten deshalb mit einer Datenmodellierung, mit der ein solches Vokabular relevanter Begrifflichkeiten und Beziehungen (ER-Modellierung) erstellt wird, die in einer zweiten Phase dann mit den konkreten Beschreibungsdaten zu füllen ist.

Sinnvollerweise wird dabei versucht, bestehende domänenspezifische Vokabulare zu verwenden, um die Daten-Interoperabilität auch über die Projektgrenzen hinaus zu sichern.

---

<sup>12</sup><http://od.fmi.uni-leipzig.de/>

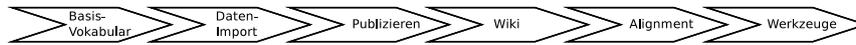


Abbildung 2: Entwicklungsprozess im OD.FMI Projekt

Daher erfolgte in der ersten Phase des Projektes eine Recherche. Im Ergebnis dieser Recherche konnten wir feststellen, dass bereits viele solcher Vokabulare existieren. Beispielfähig seien hier AIISO<sup>13</sup> für die Beschreibung der Infrastrukturen von Hochschulen und DOAP<sup>14</sup> zur Beschreibung von Projekten genannt.

Ein erster Versuch der Datenerfassung startete daraufhin mit dem Vokabular AIISO und einem Satz initial digital verfügbarer Daten vergangener Semester aus der bisherigen datenbankbasierten Stundenplanung. Dazu wurden die verfügbaren Daten entsprechend transformiert und initial importiert. Dieser Ansatz scheiterte allerdings, da mit dem verwendeten Vokabular die speziellen Gegebenheiten und die Struktur der operativ-administrativen Prozesse der Fakultät nicht ausreichend ausdrucksmächtig darstellbar waren.

Mehr noch ließen sich diese Gegebenheiten am Anfang des Projekts nicht einmal genau spezifizieren, da das Wissen über die Prozesse nicht ausreichend formalisiert vorlag und erst im Laufe des Fortschreitens Gestalt annahm, so dass eine agilere Entwicklungsmethodik zum Einsatz kommen musste, welche in der Lage war, Anforderungspräzisierungen und -erweiterungen unter Einhaltung von Qualitäts- und Integritätsstandards schnell bis zur Datenebene zu propagieren. Derartige Probleme mit Wasserfall-Ansätzen sind aus anderen Bereichen der Softwaretechnik hinreichend bekannt. Wir haben bereits in anderen Projekten Erfahrungen mit dem agilen Anforderungsmanagement in Softwareprojekten [ALLR09, S. 39 - 48] sammeln können, die an dieser Stelle in das neue Projekt einfließen.

Dieser alternative Entwicklungsansatz geht davon aus, Datenerfassung, Fortschreibung des Datenmodells und Weiterentwicklung der Werkzeuge in einem koevolutiven Prozess zu verzahnen und in einer relativ schnellen Abfolge zu definierten Zeitpunkten neue Generationen auszurollen. Ein solches Vorgehen hat sich mit dem *Rational Unified Process* (RUP) [Kru98] in der Softwaretechnik hinreichend bewährt, so dass wir hier auch methodisch Anleihen nehmen konnten. Als Abstand zwischen dem Ausrollen zweier Generationen hat sich eine Frist von etwa sechs Monaten bewährt – die Dauer eines Semesters gibt hier eine natürliche Taktung vor.

Die verwendete Methodik setzt sich dabei aus folgenden Schritten zusammen:

- (1) Entwicklung und Weiterentwicklung eines Basis-Vokabulars für Lehrveranstaltungen und damit zusammenhängende operativ-administrative Gegebenheiten.
- (2) Import und Anpassung bestehender Daten aus früheren Semestern in unser semantisches Datenwiki, basierend auf diesem Vokabular, und Aktualisierung dieser Daten.
- (3) Publizieren der Daten für alle beteiligten Akteure.
- (4) Wikibasierte Fortschreibung von Datenbasis und Datenmodell im Zuge der Abwicklung der erfassten operativ-administrativen Prozesse im laufenden Semester,

<sup>13</sup>Academic Institution Internal Structure Ontology: <http://vocab.org/aiiso/>

<sup>14</sup>Description of a Project Vokabular: <http://usefulinc.com/doap>

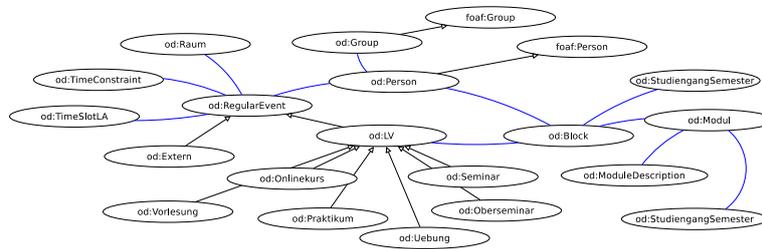


Abbildung 3: Klassenstruktur des OD.FMI Vokabulars<sup>15</sup>

ohne dies auf die Werkzeugebene zu propagieren. Dabei werden neue Vokabular-Elemente entwickelt und bereits verwendet.

- (5) Konzentrierte Alignment-Phase, in der das etablierte Vokabular konsolidiert und die Werkzeuge entsprechend den neuen Ausdrucksmöglichkeiten angepasst werden. Diese neue Generation wird in der jeweiligen Semesterpause ausgerollt.
- (6) Bugfixing auf Werkzeugebene erfolgt für kleinere Fehler im operativen Betrieb im laufenden Semester, grundlegende Probleme werden in der folgenden Alignment-Phase aufgegriffen.

Im Fortschreiten zu neuen Generationen werden operativ-administrative Prozesse in immer größerem Umfang einbezogen. Initial startete das Projekt mit Daten zur Personal- und Raumplanung von Lehrveranstaltungen, worin auch die neuen Gegebenheiten modularisierter Studiengänge abgebildet wurden. Dabei stellte sich von Anfang an heraus, dass die Strukturierung nach Modulen zwar aus studentischer Sicht konkreter Studiengänge plausibel, aber für die Planung auf operativ-administrativer Ebene der Fakultät ungeeignet ist, da sich das Verhältnis zwischen Lehrveranstaltungen, Modulen, Studiengängen und empfohlenen Semestern aus dieser Sicht als deutlich komplexer darstellt. Als Lösung wurde eine Strukturierung von Lehrveranstaltungen in *Blöcken* entsprechend der administrativen Verantwortlichkeit für die verschiedenen Lehrgebiete eingeführt.

Den Kern für die erste Generation des Datenmodells von OD.FMI bildete das Zusammenspiel von Blöcken, Lehrveranstaltungen, Personen, Räumen, Studiengängen und empfohlenen Semestern (vgl. Abbildung 3). Damit war eine initiale Klassenmodellierung vorgegeben, auf deren Basis die Erfassung von Daten für die LV-Planung zum Sommersemester 2010 als realweltlicher Sätze mit weitgehend freien Prädikaten (Verben) erfolgen konnte. Dazu wurden entsprechende Daten vergangener Semester aus der bisher verwendeten Stundenplan-Datenbank extrahiert, in das neue Format transformiert und über die als Editierwerkzeug eingesetzte OntoWiki-Instanz weiter konsolidiert. Daten zu Personen, Räumen und Studiengängen wurden zunächst als Literale erfasst. Der Schwerpunkt der Datenmodellierung dieser Phase lag auf der Konsolidierung des Vokabulars der entsprechenden Prädikate, um werkzeuggestützte Sichten auf die Datenrelationen zu ermöglichen.

<sup>15</sup>Verfügbar unter: <http://od.fmi.uni-leipzig.de/model/>

Auf dieser Basis konnten bereits vier statische Websichten<sup>16</sup> über ein kleines Perl-Skript (500 LoC) erzeugt werden, in denen die geplanten Lehrveranstaltungen nach Blöcken, Lehrenden, Studiengängen und Räumen sortiert dargestellt sind (Anforderung (A)). Die Daten für die entsprechenden Sichten werden über SPARQL-Anfragen direkt aus der OntoWiki-Instanz extrahiert.

Damit wurde der Einsatz der Semantic Web Technologien für viele beteiligte Akteure plausibel. Weitere Anforderungen resultierten vorwiegend aus dem Feedback der Anwender und führten in folgenden Generationen zur Erweiterung des Vokabulars um Modulbeschreibungen und weitere Planungsinformationen auf Block-Ebene. Das entwickelte Vokabular kann hier aus Platzgründen nicht im Detail vorgestellt werden. Wir verweisen statt dessen auf unsere operativen Webseiten zum Vorgehen der LV-Planung<sup>17</sup> sowie zur Darstellung des OD-Vokabulars<sup>18</sup>. Das Alignment zu anderen Vokabularen ist vor allem durch Übernahme von Teilvokabularen etwa aus `foaf:Person` sowie auf der Ebene von RDF und RDF-S im Blick der Entwickler. Schwerpunkte der Datenrestrukturierung in neuen Generationen waren vor allem

- die Umwandlung der Daten für Personen, Räume, Studiengänge, empfohlene Semester usw. in Ressourcen mit eigenen ID's sowie
- die Aufteilung der Daten in mehrere Datenbestände entsprechend der „Halbwertszeit“ (nur für ein Semester relevant oder semesterübergreifend) sowie operativer Verantwortlichkeiten (für Personal, für Räume, für Studiengänge usw.).

Es stellte sich heraus, dass mit Blick auf die gute Strukturiertheit der Daten eine Vokabularanpassung mit entsprechenden Datentransformationen relativ leicht zu bewerkstelligen ist.

## 6 Integration in die IT-Struktur der Fakultät

Zentral für die Integration der Entwicklung in die IT-Struktur der Fakultät ist die in einem agilen Prozess (vgl. 5) entwickelte Wissensbasis mit aktuellen Planungs- und Zustandsinformationen über Lehrveranstaltungen und das dazugehörige Vokabular. Das Vokabular der Wissensbasis hat sich über zwei Iterationen (Sommersemester 2010 und Wintersemester 2010/11) evolutionär stabilisiert. Strukturell wurde die Wissensbasis in der zweiten Iteration in einen generellen Teil mit semesterübergreifenden Informationen über Lehrkräfte und Räume und einen semesterspezifischen Teil mit Informationen über die Lehrveranstaltungen gegliedert. Die Datenbasis wird über die verwendete OntoWiki-Instanz editiert (vgl. Abbildung 4) und in Form mehrerer RDF-Wissensbasen persistiert. Änderungen daran können direkt mit Ontologie-Editoren ausgeführt werden. Im Fall der Persistierung als RDF-Datei ist dies auch mit einem Texteditor möglich. Die Fortschreibung der Wissensbasis ist nicht an das OntoWiki als Werkzeug gebunden.

<sup>16</sup><http://www.informatik.uni-leipzig.de/~graebe/stdplan/dozent.html>

<sup>17</sup><http://bis.informatik.uni-leipzig.de/OLAT/LVPlanung>

<sup>18</sup><http://bis.informatik.uni-leipzig.de/OLAT/LVPlanung/ODQuelle>

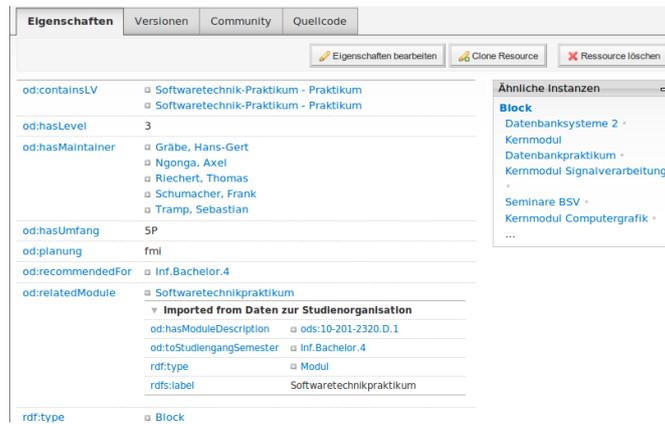


Abbildung 4: Editoransicht zum Block Softwaretechnik-Praktikum<sup>19</sup>

Die verwendete OntoWiki-Instanz stellt zudem einen SPARQL-Endpunkt bereit, über den andere Werkzeuge lesend auf die Daten zugreifen und daraus spezifische Sichten erzeugen können. Wie bereits beschrieben werden auf dieser Basis die aktuellen Daten zur Lehrveranstaltungs-Planung regelmäßig veröffentlicht (Anforderung (A), Abschnitt 4).

Nach diesem Vorgehen wurde durch den Webverantwortlichen der Fakultät auf der Basis von PHP und XSL-Transformationen eine Darstellung des offiziellen Stundenplans entwickelt, in die Änderungen der primären Quelle durch regelmäßige Aktualisierungen ebenfalls automatisiert übernommen werden (Anforderung (B), Abschnitt 4).

Im Rahmen des Softwaretechnik-Praktikums 2010 hat schließlich eine studentische Gruppe eine OLAT-Erweiterung entwickelt, mit der personalisierte Stundenplaninformationen in einen eigenen OLAT-Stundenplankalender übernommen werden können (vgl. Anforderung (C), Abschnitt 4). Dazu werden über die SPARQL-Schnittstelle Informationen über die einzelnen Lehrveranstaltungen angefragt und daraus Kalenderevents generiert, welche die Teilnehmer in ihre privaten Kalender übernehmen können. Ein Update-Management sichert eine regelmäßige Aktualisierung der Daten aus der Quelle, wobei Änderungen lokalisiert und auf kontrollierte Weise in die privaten Kalender der Abonnenten propagiert werden. Diese Lösung wurde zum Wintersemester in die BIS-OLAT-Produktivinstanz unserer Fakultät übernommen. Der Quellcode kann aus unserem SVN-Repository<sup>20</sup> übernommen und (wie OLAT insgesamt) unter den Bedingungen der Apache Lizenz Version 2 verwendet werden.

Der hier beschriebene Ansatz der Modellierung operativ-administrativer Daten auf Fakultätsebene lässt Erweiterungen in mehreren Richtungen zu:

- Es können (und werden in künftigen Generationen) weitere Bereiche des operativ-administrativen Geschäfts der Fakultät erfasst werden.

<sup>19</sup>Ressource verfügbar unter: <http://od.fmi.uni-leipzig.de/view/r/BIS.SWT-Praktikum>

<sup>20</sup>Siehe <http://bis.informatik.uni-leipzig.de/de/Forschung/EEE/OLAT>

- Auf Basis der vorgegebenen Datenspezifikation können interessierte Parteien eigene Views auf die Daten (Apps), insbesondere auch mobile Smartphone-Applikationen erstellen.
- In den Wiki-basierten Prozess des Verwaltens der Informationen können weitere Akteure auf der Basis eines zu präzisierenden Rollenkonzepts einbezogen werden.
- Die Datenquelle kann mit weiteren Linked Open Data Quellen vernetzt werden.

## Literatur

- [ADR06] Sören Auer, Sebastian Dietzold und Thomas Riechert. OntoWiki - A Tool for Social, Semantic Collaboration. In Isabel F. Cruz, Stefan Decker, Dean Allemang, Chris Preist, Daniel Schwabe, Peter Mika, Michael Uschold und Lora Aroyo, Hrsg., *The Semantic Web - ISWC 2006, 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9, 2006, Proceedings*, Jgg. 4273 of *Lecture Notes in Computer Science*, Seiten 736–749, Berlin / Heidelberg, 2006. Springer.
- [ALLR09] Sören Auer, Kim Lauenroth, Steffen Lohmann und Thomas Riechert, Hrsg. *Agiles Requirements Engineering für Softwareprojekte mit einer großen Anzahl verteilter Stakeholder*, Jgg. 18 of *Leipziger Beiträge zur Informatik*. Leipziger Informatik Verbund (LIV), Leipzig, Germany, September 2009.
- [BG04] Dan Brickley und R. V. Guha. RDF Vocabulary Description Language 1.0: RDF Schema. Bericht, 2004. <http://www.w3.org/TR/rdf-schema/>.
- [BvHH<sup>+</sup>04] Sean Bechhofer, Frank van Harmelen, Jim Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider und Peter F. Patel-Schneider. OWL Web Ontology Language Reference. Bericht, World Wide Web Consortium, <http://www.w3.org/TR/owl-ref/>, 2004.
- [Kru98] P. Kruchten. *The Rational Unified Process. An Introduction*. Addison-Wesley, 1998.
- [MM04] Frank Manola und Eric Miller. RDF Primer. W3C recommendation, W3C, February 2004. Published online on February 10th, 2004 at <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>.
- [MR10] Ulf Morgenstern und Thomas Riechert, Hrsg. *Catalogus Professorum Lipsiensis – Konzeption, technische Umsetzung und Anwendungen für Professorenkataloge im Semantic Web*, Jgg. 21 of *Leipziger Beiträge zur Informatik*. Leipziger Informatik-Verbund (LIV), Leipzig, Germany, Februar 2010.
- [PS08] Eric Prud'hommeaux und Andy Seaborne. SPARQL Query Language for RDF. Bericht, 1 2008.
- [RMA<sup>+</sup>10] Thomas Riechert, Ulf Morgenstern, Sören Auer, Sebastian Tramp und Michael Martin. Knowledge Engineering for Historians on the Example of the Catalogus Professorum Lipsiensis. In *Proceedings of the 9th International Semantic Web Conference (ISWC2010)*, Lecture Notes in Computer Science, Shanghai / China, 2010. Springer.