

FADO BW

Realisierung erster Komponenten für ein verteiltes Fachdokumentenmanagement im Umweltinformationssystem Baden-Württemberg

*R. Weidemann; W. Geiger; C. Greceanu; T. Schlachter; G. Zilly
Forschungszentrum Karlsruhe GmbH
Institut für Angewandte Informatik
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen*

*P. Lautner; K. Türk
Ingenieurgesellschaft für Umwelttechnik und Bauwesen Dr. Eisele mbH
Lindenbergstr. 12
79199 Kirchzarten*

*R. Ebel; R. Hahn; W. Eitel; M. Lehle; M. Theis;
W. Weiß; J. Witt-Hock; R. Zimmermann
Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg
Griesbachstr. 1
76185 Karlsruhe*

1. EINLEITUNG	33
2. PFLEGE UND WEITERENTWICKLUNG DER XFAWEB-SYSTEMFAMILIE.....	33
3. FADO (FACHDOKUMENTE ONLINE)	34
3.1 PROTOTYP-ENTWICKLUNG.....	35
3.1.1 <i>AlfaWeb-neu</i>	35
3.1.2 <i>Zerlegung von Berichten am Beispiel eines Web-Journals Naturschutz-Info</i>	36
3.2 VERTIEFUNG DES FADO-KONZEPTS	37
3.2.1 <i>Schnittstellen zu InGrid-BW</i>	37
3.2.2 <i>Schnittstellen mit Fachsystemen (LUBW-Shop, ZSV, DRS)</i>	38
3.3 ERSTE REALISIERUNGSSCHRITTE	40
3.3.1 <i>XfaWeb-neu</i>	40
3.3.2 <i>InDesign-to-HTML-Konverter</i>	41
3.3.2.1 XML-Tags und XML-Export	42
3.3.2.2 Konvertierung und Zerlegung.....	42
3.3.3 <i>Fachdokumenten-Browser (FDB)</i>	43
4. WEITERES VORGEHEN	43
5. LITERATUR.....	44

1. Einleitung

Im November 2005 wurde in der Studie zum „Fachdokumentenmanagement im Umweltinformationssystem Baden-Württemberg“ (/1,2/) ein dreistufiges Konzept vorgeschlagen, das allgemein zu einer besseren Vernetzung von Fachsystemen auf der Basis von Webservices führen soll und einen umfassenden und übergreifenden Zugriff auf Fachdokumente im Umweltinformationssystem Baden-Württemberg (UIS BW) vorsieht. Gleichzeitig wurde ein Nachfolgesystem für die XfaWeb-Systemfamilie entworfen. Die Umsetzung des Konzepts wird in zwei Projekten verfolgt. Ein zentrales Verzeichnis für Dienste entwickelt das FZI im Rahmen des Projekts „UIS-UDDI“ /3/. Der vorliegende Beitrag beschäftigt sich mit FADO (Fachdokumente Online), das sich aus den Teilprojekten Fachdokumenten-Browser (FDB) für den Zugriff auf Fachdokumente und XfaWeb-neu für die Neukonzeption der XfaWeb-Systeme zusammensetzt (vgl. Abbildung 1). Nach Klärung in der Studie offen gelassener Punkte wurde die Planung für FADO in 2006 in der LUBW abgestimmt und es wurden einige zentrale Komponenten prototypisch implementiert (s. Kap. 3.1). Auf dieser Basis stimmte die XfaWeb-Projektleiterrunde am 17.11.2006 den ersten, bis Mitte 2007 laufenden Umsetzungsschritten zu. In deren Rahmen wurde die Spezifikation bedarfsorientiert vertieft, insbesondere wurden auch Schnittstellen mit anderen Fachsystemen geklärt (s. Kap. 3.2). Der Einstieg in die Realisierung erfolgte mit der Umsetzung des Objektmodells für XfaWeb-neu, der Implementierung einer ersten Version des InDesign-to-HTML-Konverters und der Metadatenbank des FDB (s. Kap. 3.3).

Gleichzeitig wurden die bisherigen XfaWeb-Systeme weiter gepflegt, inhaltlich aktualisiert und ergänzt und bedarfsgetrieben funktional weiterentwickelt (s. Kap. 2).

2. Pflege und Weiterentwicklung der XfaWeb-Systemfamilie

Die bisherige Version der XfaWeb-Systeme wird weiterhin im Internet und, teilweise mit erweitertem Inhalt, im Intranet der Umweltverwaltung BW betrieben. Die Pflege der enthaltenen Materialien, seien es Neueinstellungen, Überarbeitungen oder Löschungen veralteter Dokumente, und ebenso die Pflege der Zugänge ist eine kontinuierliche Aufgabe. Änderungen werden zeitnah auf dem Entwicklungsserver abgelegt und ermöglichen den Auftraggebern und Projektpartnern das Testen. Die Produktionsserver wurden weiterhin im Viermonatsrhythmus aktualisiert. In den zurückliegenden Monaten musste die Sicherheit des Entwicklungsservers erhöht werden, da dieser massiven Angriffen aus dem Internet ausgesetzt war.

Das System **AbfaWeb** (Abfallbehandlung) war im Vergleich zu den übrigen XfaWeb-Systemen nur in geringem Umfang von inhaltlichen Änderungen betroffen. Dahingegen wurden in **AlfaWeb** (Altlastenbearbeitung) das als CD erschienene Kursbuch Altlasten integriert und der Fachzugang Recht komplett überarbeitet und neu gegliedert, wobei auch der Bericht „Rechtsanwendung“ eingebaut wurde. Neuerscheinungen werden inzwischen in der Regel nur noch als PDF-Dokumente eingestellt, teilweise ergänzt um eine Kurzbeschreibung.

In **BofaWeb** (Bodenschutz) wurde ebenfalls der Fachzugang Recht überarbeitet. Im Navigationsmenü wurde die Option 'Bodenschutzrecht' eingefügt und dafür die Option 'DV-

Programme' herausgenommen. Neben zwei neuen Berichten wurden ca. 50 Urteile ergänzt und Änderungen im Menüpunkt „Bodenschutz in Baden-Württemberg“ sowie an 15 vorhandenen Berichten vorgenommen. Nach Verabschiedung der REACH-Verordnung im Dezember 2006 wurde eine Reihe entsprechender Materialien und Dokumente in **ChemfaWeb** (Behördliches Chemikalienmanagement) unter dem Zugang „Umgang mit Chemikalien » Chemikalienpolitik, Grundsätze“ eingestellt. Auch andere Zugänge wie die „Stoffinformationen“ wurden ausgebaut. Im Intranet-Bereich wurde im Zugang "Protokolle/Beschlüsse" die Beschlussdatenbank um die Protokolle einer Sitzung erweitert.

FofaWeb (Umweltforschung in Baden-Württemberg) erhielt einen thematisch strukturierten Zugang zu den inzwischen über 340 Berichten aus den verschiedenen Forschungsbereichen. 30 Forschungsberichte wurden als PDF-Dateien und zusätzlich mit verschlagworteten Kurzbeschreibungen integriert. In **NafaWeb** (Naturschutz und Landschaftspflege) wurden acht Berichte und drei Falblätter komplett mit Kurzbeschreibungen und Druckvorlagen und zum Teil auch als HTML-Version neu eingestellt. Zusätzlich wurden verschiedene vorhandene Berichte und Verzeichnisse aktualisiert bzw. um zusätzliche Formate ergänzt.

3. FADO (Fachdokumente Online)

Die Konzeption und Realisierung von FADO ist ein über mehrere Jahre angelegtes Projekt, das in mehrere Abschnitte unterteilt, sukzessive umgesetzt wird.

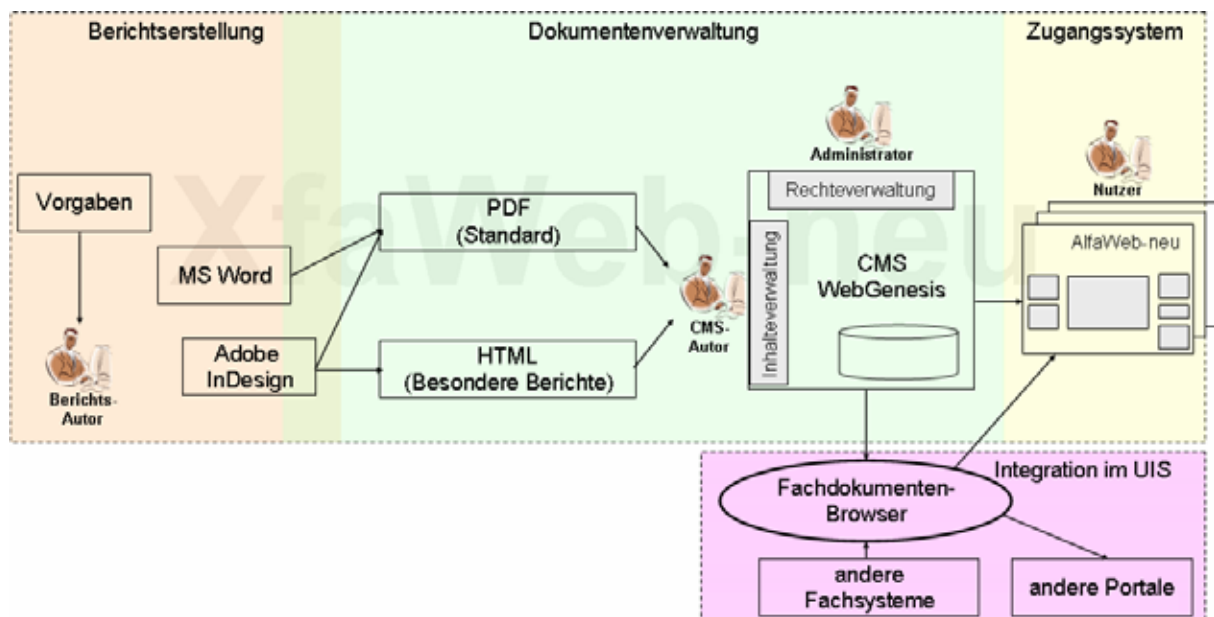


Abbildung 1: Komponenten von FADO (Grobstruktur), mit XfaWeb-neu (oben) und Fachdokumenten-Browser (unten)

Besondere Schwierigkeiten entstehen durch die zahlreichen Schnittstellen zu und Überlappungen mit anderen Fachsystemen und Portalen. Beim FDB ergeben sich die Schnittstellen zwangsläufig aus der Kernaufgabe des FDB, der Indizierung von Fachdokumenten aus Fachsystemen des UIS BW. Bei XfaWeb-neu waren insbesondere das Zusammenwirken mit dem LUBW-Shop und die Positionierung der XfaWeb-Fachportale zu den LUBW-Themenportalen klärungsbedürftig. FDB und XfaWeb-neu sind darüber hinaus vom geplanten Einsatz von InGrid-Komponenten im UIS BW im Rahmen der Neukonzeption von UINBW berührt /4/. Diese zahlreichen Schnittstellen ergeben zwangsläufig einen erhöhten Abstimm-

mungsbedarf. Wegen den vielfältigen Querbeziehungen wird für FADO ein diensteorientierter Ansatz favorisiert, d.h. die Realisierung klar definierter Funktionen (Dienste) und deren Bereitstellung über standardbasierte Schnittstellen, bevorzugt als Webservices.

3.1 Prototyp-Entwicklung

Im Rahmen des Abstimmungsprozesses wurden bis Ende 2006 Prototypen einiger zentraler Komponenten entwickelt und schließlich im November 2006 dem Steuerungsgremium von XfaWeb und FADO auf einer XfaWeb-Projektleitersitzung präsentiert und dort diskutiert. Neben einer Demoversion des InDesign-to-HTML-Konverters (diese wurde bereits in /2/ beschrieben) wurden erste Prototypen eines Fachportals AlfaWeb-neu und eines Web-Journals Naturschutz-Info entwickelt.

3.1.1 AlfaWeb-neu

Anhand eines Prototyps für das Fachportal AlfaWeb-neu wurden verschiedene Aspekte des Konzepts für XfaWeb-neu im Vergleich zu den bisherigen XfaWeb-Systemen verdeutlicht. Ein erster Punkt ist hier, dass die Portale der neuen XfaWeb-Systeme den Anspruch haben, umfassend Arbeitsmittel und Informationsquellen für Fachleute zusammenzuführen und soweit möglich zu integrieren (Fachportale). Die alten XfaWeb-Systeme liefern i. W. nur einen Zugang zu den jeweiligen Materialien im System selbst, eine Vernetzung mit anderen Angeboten ist nur schwach vorhanden.

Die Integration von Inhalten wird dadurch begünstigt, dass die neue XfaWeb-Systemfamilie über eine gemeinsame und einheitliche Dokumentenverwaltung auf Basis des Content-Management-Systems (CMS) WebGenesis verfügen wird. Die Sichtbarkeit von Materialien wird nicht mehr durch die Systemzugehörigkeit bestimmt, sondern kann nach fachlichen, thematischen Gesichtspunkten geregelt werden. Werden die Dienste des Fachdokumenten-Browsers in das Fachportal eingebaut, ist sogar ein UIS-übergreifender Zugriff auf thematisch passende Fachdokumente möglich.

Eine wesentliche Neuerung betrifft die Darstellung von Fachdokumenten. Da bis auf bestimmte Ausnahmen (s. Kap. 3.1.2) die Vollversionen von Berichten nicht mehr in HTML-Seiten zerlegt, sondern nur noch als PDF-Datei angeboten werden sollen, ergibt sich eine deutlich andere Sicht auf die Inhalte. Die alten XfaWeb-Systeme haben den Anspruch, die Einzelberichte in ein Netz aus Informationsseiten aufzulösen, in dem über Links inhaltliche Querbezüge über Berichtsgrenzen hinweg verfolgt werden können und so eine Art Gesamt-Hypertextdokument geschaffen wird. Die dazu erforderliche Aufbereitung der Berichte ist aufwändig und kann in Zeiten knapper Mittel immer weniger erbracht werden. Der neue Ansatz ist sehr viel stärker auf das Einzeldokument zentriert, Querbezüge spielen keine so große Rolle mehr, der Aufbereitungs- und Pflegeaufwand soll merklich geringer ausfallen. Der Prototyp zu AlfaWeb-neu demonstrierte die neue Sicht auf die Inhalte. Nach dem vorliegenden Konzept besteht ein Fachdokument aus Metadaten plus Dokument-Datei(en). Zur Visualisierung eines Fachdokuments werden zuerst die Metadaten in Form einer Metadaten-Beschreibungsseite (MDBS) angezeigt, aus der heraus auf die Dokument-Datei(en) verwiesen wird. Damit können Dokument-Dateien verschiedenen Typs inkl. Mediaformaten einheitlich gehandhabt werden.

Der Prototyp AlfaWeb-neu wurde auch dazu benutzt, unterschiedliche Ansätze zur Verschlagwortung zu diskutieren. Vorab wurde bereits entschieden, dass keine textstellen-

genaue Verschlagwortung mehr erfolgen soll, sondern den Fachdokumenten als Ganzes einige zutreffende Schlagworte zugewiesen werden sollen. Diskutiert wurde über die Vor- und Nachteile einerseits einer freien Verschlagwortung gegenüber einer mit festem Vokabular und andererseits einer automatischen Verschlagwortung oder einer manuellen durch Fachexperten oder durch die Nutzer. Als Ergebnis der Diskussion soll geprüft werden, ob in XfaWeb-neu eine Verschlagwortung durch Fachexperten ermöglicht werden kann, bei der über eine automatische Verschlagwortung Begriffe aus dem UBA-Thesaurus vorgeschlagen werden, aber auch freie Schlagworte vergeben werden können.

Als letzter Aspekt wurde mit dem Prototyp die künftige formularbasierte Autorenschnittstelle skizziert. Die Pflege der Inhalte wird durch CMS-Autoren erfolgen, welche mit dem CMS WebGenesis entwickelte Erfassungsmasken nutzen, um Metadaten zu Fachdokumenten und anderen Objekten einzugeben und die Relationen zwischen den Objekten zu pflegen.

3.1.2 Zerlegung von Berichten am Beispiel eines Web-Journals Naturschutz-Info

Wie weiter oben erwähnt, sollen Berichte in XfaWeb-neu nur noch in Ausnahmefällen in HTML-Seiten zerlegt werden. An einem konkreten Beispiel, dem Prototyp eines Web-Journals Naturschutz-Info wurden Kriterien erarbeitet, die eine Zerlegung sinnvoll erscheinen lassen. Das Web-Journal Naturschutz-Info selbst ist eine elektronische Zeitschrift auf der Basis der Naturschutz-Infos des Fachdienstes Naturschutz. Das Naturschutz-Info wiederum ist eine regelmäßig erscheinende Publikation, in der einzelne Artikel in vorgegebenen, thematischen Rubriken erscheinen. Es drängt sich als Beispiel für eine mögliche Zerlegung geradezu auf, da jeder Artikel isoliert, ohne den Rest des Dokuments verständlich ist, und eine feste Struktur über mehrere Bände hinweg Band-übergreifende Zugänge (z.B. nach Fachthemen oder Autoren) nahelegt (s. Abbildung 2).

Als wichtigstes Kriterium für die Zerlegung eines Berichts in XfaWeb-neu wurde herausgearbeitet, dass das Ziel sein muss, abgeschlossene Informationseinheiten zu bilden, die bereits für sich allein verständlich sind und als solche in verschiedenen Kontexten eingebaut werden können oder sollen. Dafür bieten sich Berichte an, die Sammlungen gleich strukturierter Beschreibungen enthalten, wie Sammlungen von Einzelbeiträgen / Artikeln (z.B. Naturschutz-Infos), Methodensammlungen (z.B. Methoden zur Grundwassererkundung in AlfaWeb), Kataloge (z.B. Baustoffkatalog). Weiterhin könnten Berichte zerlegt werden, die stark vernetzte Informationen enthalten, d.h. bei denen der lineare Textfluss wegen zahlreicher Querverweise nicht die optimale Leserichtung vorgibt, sondern eine hypermediale Aufbereitung die Erschließung erleichtert.

In der XfaWeb-Projektleitersitzung wurde über das Kosten-/Nutzenverhältnis des Web-Journals diskutiert. Eine Entscheidung über die Realisierung steht noch aus.

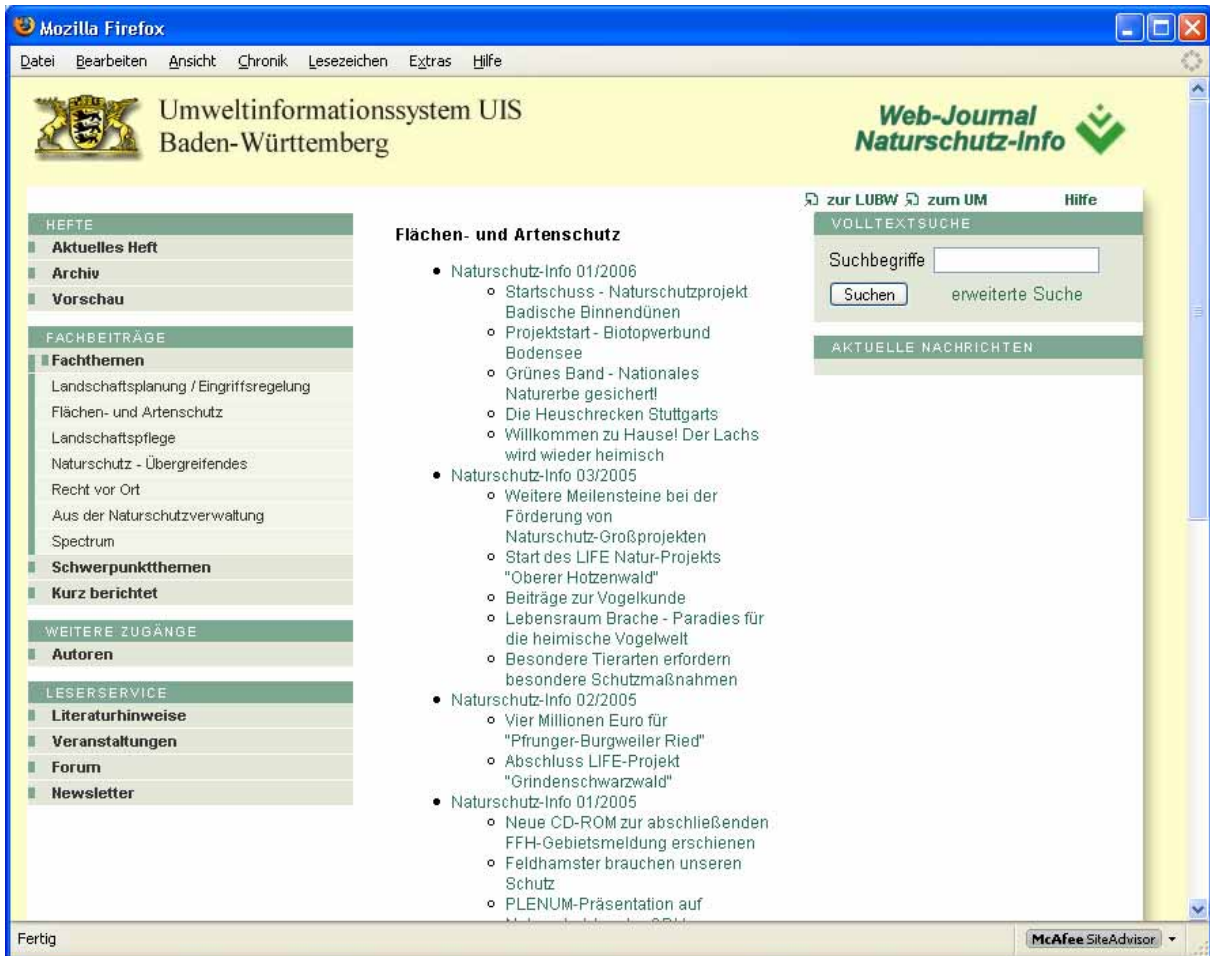


Abbildung 2: Prototyp Web-Journal Naturschutz-Info, Zugang über Fachthema

3.2 Vertiefung des FADO-Konzepts

Das FADO-Konzept wurde gezielt verfeinert. Das Hauptaugenmerk war dabei zum einen darauf gerichtet, ein vollständiges und belastbares Objektmodell zu entwerfen und zum anderen, die Schnittstellen von XfaWeb-neu und FDB zu anderen Systemen zu klären.

3.2.1 Schnittstellen zu InGrid-BW

Im Konzept zum Einsatz von InGrid-Komponenten im UIS BW (siehe /4/) werden zum einen Schnittstellen zu PortalU definiert, die von FADO mit Inhalten zu versorgen sind, und zum anderen wurde geprüft, ob InGrid-Komponenten für FADO einsetzbar wären.

Berichtspflichten zu PortalU und zu InGrid-BW, der geplanten Installation der InGrid-Software im UIS BW als Komponente von UINBW 2.0, bestehen für XfaWeb-neu in der Bereitstellung von Metadaten zu Forschungsprojekten. XfaWeb-neu wird die derzeit noch selbstständige Datenbank der Umweltforschungsprojekte integrieren. Der FDB wird im Rahmen der geplanten Dezentralisierung des Umweltdatenkatalogs (UDK) die Verwaltung der Metadaten zu UDK-Objekten der Klasse „Veröffentlichungen / Fachdokumente“ übernehmen. Dazu werden die Metadatenbank und die Schnittstelle des FDB-Crawlers zu den Fachsystemen um spezifische, meist optionale Metadaten-Attribute erweitert. Schließlich wird die Suchfunktionalität (Volltextsuche) des FDB mit derjenigen der LUBW-Site und der von UINBW in InGrid-BW zusammengeführt. Der FDB versorgt den Crawler von InGrid-BW über seine Metadaten mit den URLs der Fachdokumente (MDBS, Dokument-Dateien). Die Volltextsuche des

FDB wiederum erfolgt über einen Service von InGrid-BW (OpenSearch-Schnittstelle). Genauere Informationen zum Zusammenwirken mit InGrid-BW enthält /4/.

3.2.2 Schnittstellen mit Fachsystemen (LUBW-Shop, ZSV, DRS)

Während XfaWeb-neu alle relevanten Arbeitsmaterialien (Fachdokumente) für Fachleute vorhalten soll, verwaltet der LUBW-Shop alle offiziellen Veröffentlichungen der LUBW. Die Inhalte überschneiden sich im Bereich der Fachveröffentlichungen. Es wurden Verfahren und Schnittstellen festgelegt, welche die vorrangige Bedienung des LUBW-Shops sicherstellen und eine redundante Datenhaltung vermeiden. Neue Fachveröffentlichungen müssen zuerst in den Shop eingestellt werden. XfaWeb-neu wird darüber benachrichtigt (Schwarzes Brett) und übernimmt und ergänzt dann die zugehörigen Metadaten (siehe Abbildung 3) für seine interne Verwendung. Nach außen, insbesondere über den FDB, ist nur die Fachveröffentlichung des LUBW-Shops sichtbar. Entsprechende Szenarien wurden für das Ändern und Löschen von Fachveröffentlichungen und das Einstellen anderer Materialien festgelegt.

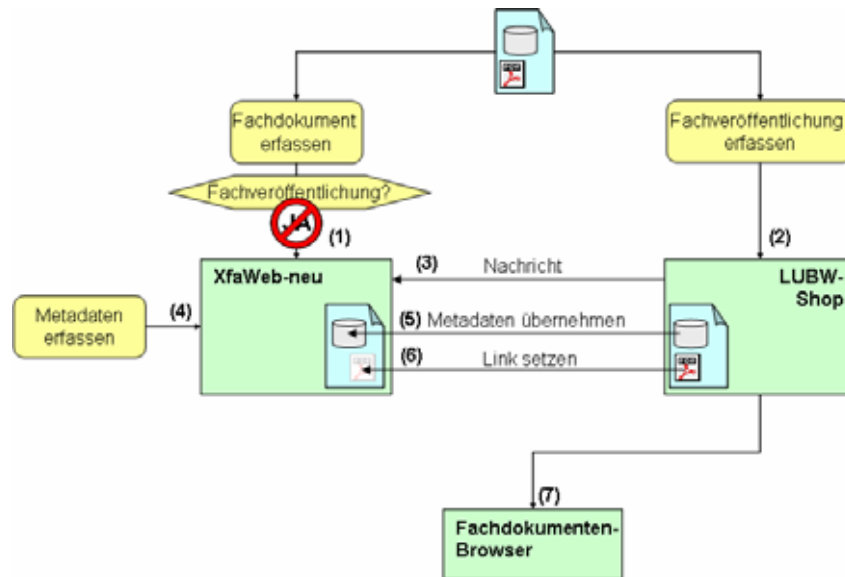


Abbildung 3: Bereitstellung neuer Fachveröffentlichungen

Der FDB ist eine Komponente, die besonders zahlreiche Schnittstellen besitzt bzw. nutzt (s. Abbildung 4). Um Metadaten zu Fachdokumenten einsammeln zu können, wurde festgelegt, dass Fachsysteme, die Fachdokumente zur Verfügung stellen, zwei Webservices implementieren, die eine Liste vorhandener Dokumente bzw. definierte Metadaten (Minstdatensatz plus optionale UDK-Metadaten) zu jedem Dokument liefern können. Der FDB wird die Services regelmäßig abfragen und auf dieser Basis seine Metadatenbank pflegen. Auf der anderen Seite soll der FDB wiederum Dienste bereitstellen, ohne selbst eine eigene Benutzeroberfläche zu besitzen, sieht man einmal von der Administrationsoberfläche ab.

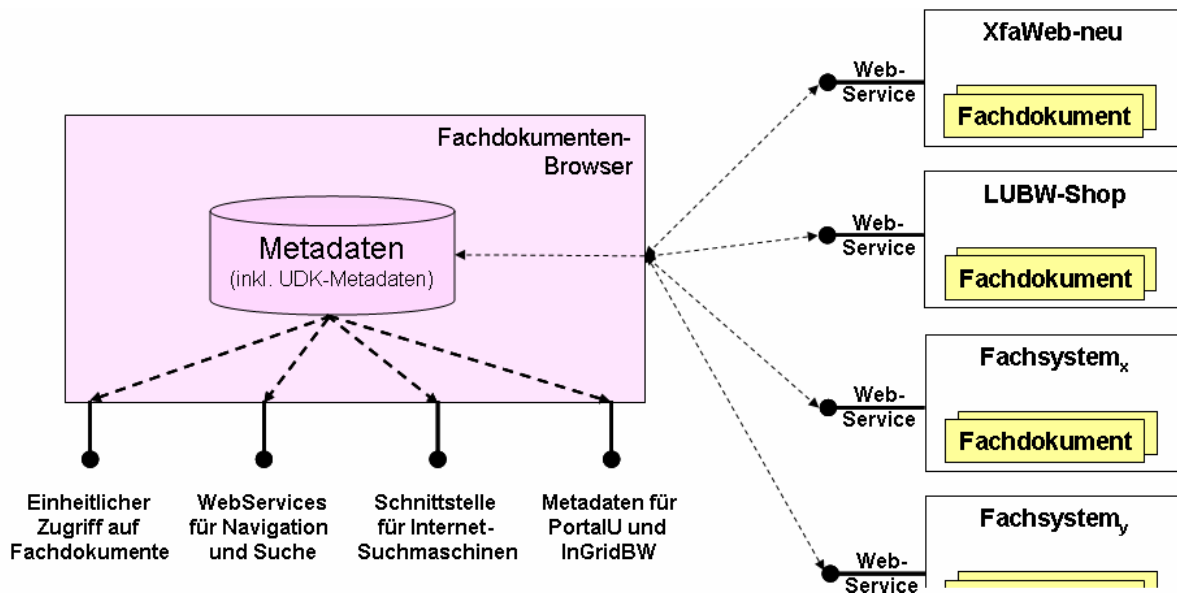


Abbildung 4: Schnittstellen des Fachdokumenten-Browsers

Nun ist die Bereitstellung der Webservices für die Fachsysteme keineswegs trivial. Zum einen ist rein technisch die Realisierung und Pflege von Webservices mit gewissem Aufwand verbunden und zum anderen müssen auch die erforderlichen Metadaten im Fachsystem vorhanden oder mit vertretbarem Aufwand zu ergänzen sein. Im Gegensatz zu XfaWeb-neu, wo die Anforderungen bereits bei der Erstellung des Systems berücksichtigt werden können, handelt es sich bei den übrigen Fachsystemen meist um teilweise schon seit Jahren laufende Applikationen. Aus diesem Grund wurde für eine Reihe von Fachsystemen analysiert, ob die für den FDB benötigten Dienste bereitgestellt werden können. Der LUBW-Shop verwaltet zahlreiche Metadaten zu Fachdokumenten, so dass dieser nicht nur die Mindestdaten, sondern auch eine Reihe der optionalen UDK-Metadaten liefern kann. Im Rahmen einer Diplomarbeit wurde bereits eine prototypische Implementierung der Webservices angefertigt. Die ZSV (Zentrale Stelle für die Vollzugsunterstützung der Gewerbeaufsicht beim RP Tübingen) stellt Vorschriften, Publikationen und andere Arbeitshilfen für die Gewerbeaufsicht bereit. Der Webauftritt ist mit WebGenesis realisiert, ein Teil der Materialien wird mit einer auf Perl basierenden Verwaltungskomponente erfasst. Nicht alle Materialien sind für den FDB relevant, die ZSV definiert ein Auswahlkriterium. Das Fachsystem enthält anwendungsspezifische Metadaten. Der minimal erforderliche Metadatensatz kann bereitgestellt werden, weitere UDK-Daten sind nicht vorhanden. Ähnlich verhält es sich mit dem DRS (Document Retrieval System) des ZFD (Zentraler Fachdienst Wasser, Boden, Abfall, Altlasten), das Rechtsvorschriften in einem speziellen Dokumentenverwaltungssystem vorhält. Auch hier sind umfangreiche Metadaten vorhanden, die sich mit denen des FDB aber nur zum kleinen Teil überlappen, es reicht jedoch für den minimalen Metadatensatz.

Zusammenfassend kann man feststellen, dass in allen bisher untersuchten Fällen die benötigten Schnittstellen realisiert werden können und die Systembetreiber auch Interesse an einer solchen Vernetzung bekundet haben. Da eine zunehmende Anzahl der Fachsysteme auf WebGenesis basiert (XfaWeb-neu, LUBW-Shop, ZSV-Angebot der Gewerbeaufsicht) soll dafür möglichst eine generische Lösung bereitgestellt werden.

3.3 Erste Realisierungsschritte

Mit der Umsetzung des FADO-Konzepts wurde Anfang 2007 begonnen. Es wird gleichzeitig an XfaWeb-neu, dem InDesign-to-HTML-Konverter und am FDB gearbeitet.

3.3.1 XfaWeb-neu

Die XfaWeb-neu-Systeme werden auf Basis des Content-Management-Systems (CMS) WebGenesis erstellt. Hierzu werden gemäß dem Corporate-Design des Landes Baden-Württemberg entsprechende Templates genutzt, die das Layout des Landes umsetzen (vgl. Abbildung 5). Die XfaWeb-neu-Portale setzen dafür auf derselben WebGenesis-Instanz auf wie die Dokumentenverwaltung, haben jedoch komplett getrennte Benutzer- bzw. Redaktions-Schnittstellen.

Kern der XfaWeb-neu-Systeme sind die strukturierte Navigation über die vorliegenden Fachinformationen sowie die Suche danach. Bei der Navigation zu Fachdokumenten über Berichtsreihen bzw. Verzeichnisstrukturen wird dabei grundsätzlich der Weg über die Metadatenbeschreibungseite des Dokuments gewählt. Über diese ist dann jeweils der Volltext eines Fachdokuments über einen oder mehrere Links erreichbar. In der Regel liegt der Volltext in Form eines oder mehrerer PDF-Dokumente vor. Es kann darüber hinaus auch eine Zerlegung in HTML-Dateien erfolgen (s. Abschnitt 3.3.2), was in den XfaWeb-neu-Systemen jedoch die Ausnahme bleiben soll. Auch die Einbindung der HTML-Version von „alten“ XfaWeb-Berichten ist möglich. Diese werden – wie auch neu konvertierte Berichte – jeweils über einen entsprechenden Importfilter innerhalb der XfaWeb-neu-Seiten angezeigt. Dokumente im PDF- und anderen Formaten werden in externen Fenstern angezeigt.

Wesentlicher Unterschied zwischen XfaWeb-neu und den alten XfaWeb-Systemen ist, dass Fachinformationen nicht nur innerhalb genau eines Fachportals, sondern – wo dies inhaltlich sinnvoll ist – auch innerhalb mehrerer Systeme bereitgestellt werden können. Die Entscheidung über die Sichtbarkeit von Fachinformationen treffen dabei die CMS-Autoren der einzelnen XfaWeb-neu-Systeme.

Die Suche in Fachdokumenten kann einerseits über den Volltext, andererseits über die Metadaten, je nach Art der Fachinformation in übergreifenden oder spezifischen Feldern, erfolgen. Für die Volltextsuche ist dabei die Integration des SE-iPlugs (Search-Engine-iPlug) von InGrid-BW geplant (s. Abschnitt 3.2.1), während die Suche innerhalb der Metadaten über die zu diesem Zwecke erweiterten Suchfunktionen von WebGenesis erfolgt. Fachinformationen, die keine Fachdokumente sind, z.B. Adressen, werden nicht über die Volltext-, sondern lediglich über die Metadatensuche gefunden.

Neben der Navigation über Verzeichnisse oder Berichtsreihen ist es in den XfaWeb-neu-Portalen möglich, einen oder mehrere hierarchisch gegliederte Fachzugänge anzulegen und diese über die Autorenschnittstelle direkt im CMS zu bearbeiten. Dabei wird die Navigation innerhalb eines Fachzugangs automatisch erzeugt und kann durch manuelle Querverweise auf Fachzugangsseiten oder Fachinformationen ergänzt werden. Ein weiterer Zugang bietet jeweils die zuletzt eingestellten Fachinformationen an (neue Materialien).

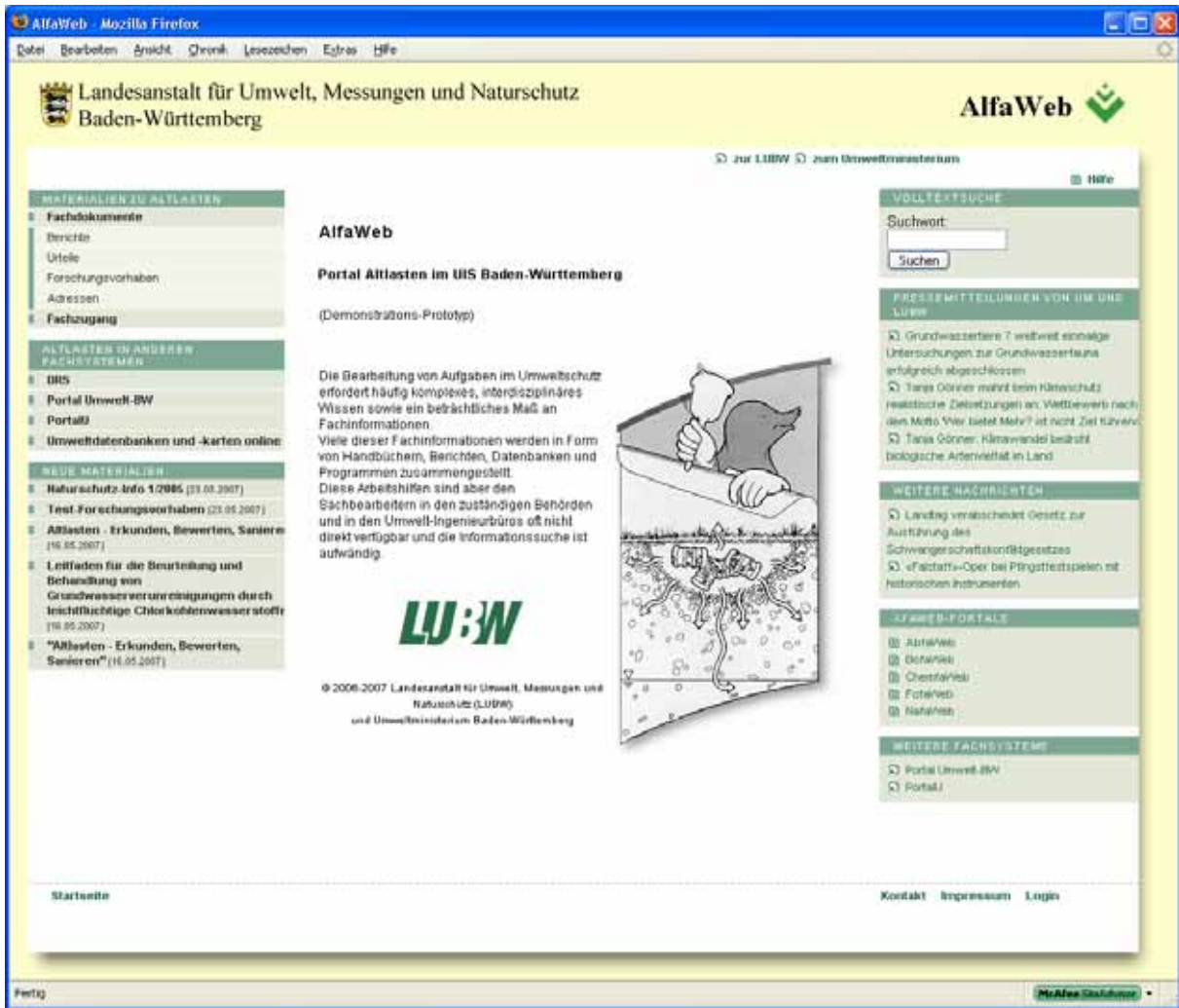


Abbildung 5: Prototyp AlfaWeb-neu

Auch können zu jedem System spezifische Links in andere Fachsysteme bereitgestellt werden sowie über alle Fachsysteme gültige „globale“ Links, zum Beispiel zu den Umweltportalen Portal Umwelt-BW und PortalU. Selbstverständlich ist auch der direkte Einstieg in die anderen XfaWeb-neu-Systeme per Klick möglich.

In die XfaWeb-Portale lassen sich auch direkt Informationen aus fremden Systemen einbinden, sofern diese Systeminhalte als sogenannte RSS-Feeds anbieten. Zu diesem Zweck wurde in XfaWeb-neu eine RSS-Komponente zur Content-Syndication (= Mehrfachverwendung von Inhalten) integriert, die grundsätzlich beliebig viele Nachrichten aus RSS-Feeds importieren kann. Die Nachrichten innerhalb solcher Feeds bestehen in der Regel aus Titel und Kurzbeschreibung sowie einem Link zur vollständigen Meldung. RSS-Feeds werden z.B. über das Landesportal Baden-Württemberg, das UM oder die LUBW angeboten.

3.3.2 InDesign-to-HTML-Konverter

Die Neukonzeption der XfaWeb-Systeme sieht vor, dass Fachdokumente standardmäßig als PDF-Dokumente in XfaWeb-neu eingestellt werden. Eine Konvertierung nach HTML, verbunden mit einer Zerlegung in einzelne Kapitel oder Artikel, wird nur noch in wenigen Ausnahmefällen vorgenommen. Voraussetzung für die Zerlegung ist in XfaWeb-neu die Erstellung und Strukturierung der Berichte mit Adobe InDesign. Der aus InDesign ins XML-Format

exportierte Bericht wird mit einem nachgeschalteten Programm (XSLT-Transformation) zerlegt und nach HTML konvertiert /2/.

3.3.2.1 XML-Tags und XML-Export

Basis der Strukturierung von InDesign-Dokumenten sind die so genannten Tags, mit denen Inhaltsblöcke funktional charakterisiert, d. h. entsprechend ihrer Bedeutung markiert werden können. Da sich Tags schachteln lassen, kann man damit eine hierarchische Gliederung eines Dokuments aufbauen. Auf unterer Ebene (Textabschnitte) können die für die Gestaltung des Textes ohnehin notwendigen Formatvorlagen verwendet und auf Tags abgebildet werden. Die Zuordnung der Tags zu Formatvorlagen erfolgt über die Funktion „Formate zu Tags zuordnen“. Am einfachsten geht das, wenn man Tags definiert, die paarweise gleiche Namen wie die Formatvorlagen von InDesign haben, denn dann kann man die Zuordnung automatisch vornehmen lassen. Es ist aber auch eine manuelle Zuordnung von Tags möglich.

Da InDesign keinerlei Möglichkeiten zur Parametrisierung (z.B. Formatierung) der Export-Datei zur Verfügung stellt, muss die exportierte Datei nachbearbeitet werden. In erster Linie betrifft dies die Darstellung der Datei zwecks Lesbarkeit (Stichwort „Pretty Print“) sowie die Umsetzung bzw. Beseitigung von Sonderzeichen, die in HTML nicht dargestellt werden können.

Für die erste Version des InDesign-to-HTML-Konverters von XfaWeb-neu wurden zwei Tag-Sätze definiert, mit denen sich zum einen ein Bericht mit einer Rubrik-Artikel-Gliederung wie die Naturschutz-Infos und zum anderen ein „klassischer“, mit Kapiteln und Unterkapiteln ausgestatteter Bericht strukturieren lassen. Für diese Tags wurden im Konverter Umsetzungsregeln implementiert (s. Kap. 3.3.2.2) und es wurde eine Anleitung zur Berichtserstellung angefertigt. Im nächsten Schritt werden diese Gliederungen mit der LUBW, insbesondere auch mit dem Handbuch Printmedien, abgestimmt, um verbindliche Vorgaben für künftige, mit XfaWeb-neu kompatible InDesign-Berichte bereitstellen zu können.

3.3.2.2 Konvertierung und Zerlegung

InDesign kennt neben dem InDesign-Dokument, das in genau einer Datei abgelegt wird, auch das InDesign-Buch, in dem mehrere Dokumente zusammengefasst werden. Die XML-Exportfunktion von InDesign beschränkt sich z. Zt. (InDesign CS 2.3) allerdings auf einzelne Dateien, so dass die Dokumente eines Buchs einzeln nach XML exportiert und dann auf XML-Ebene zu einem neuen "XML-Buch" zusammengeführt werden müssen.

InDesign stellt keine Funktion zur Verfügung, mit der die Reihenfolge der Dokumente eines Buchs extern nutzbar gemacht werden könnte, etwa indem die Dateinamen der Einzeldokumente eines Buches in der richtigen Reihenfolge exportiert und als Datei gespeichert werden. Um dieses Problem zu umgehen, verlangt der InDesign-to-HTML-Konverter, dass die Dateinamen der Einzeldokumente in definierter Weise aufgebaut werden und die alphabetische Sortierreihenfolge der Dateinamen die Reihenfolge der Dokumente im Buch widerspiegelt. Der Basis-Name der InDesign-Book-Datei wird als Hauptdokument-Name für die zukünftige XML-Datei übernommen. Die Inhalte der InDesign-Einzeldokumente werden in das Hauptdokument eingearbeitet, welches anschließend wie ein Einzeldokument weiterverarbeitet werden kann.

Die Konvertierung und Zerlegung des XML-Dokuments geschieht in mehreren Arbeitsschritten (vgl. Abbildung 6). Im Konvertierungsschritt 1 wird zunächst aus der XML-Datei mittels des XSLT-Prozessors „Saxon“ die HTML-Startseite generiert. Diese enthält ein Inhaltsverzeichnis mit den Überschriften aller Rubriken bzw. Kapitel. Des Weiteren wird die XML-Datei entsprechend ihrer Gliederungsstruktur in mehrere XML-Dateien zerlegt, welche inhaltlich bereits den zu erzeugenden HTML-Seiten entsprechen. Die generierten Dateinamen werden für die weitere Verwendung im folgenden Konvertierungsschritt abgespeichert. Zusätzlich wird eine XML-Datei erzeugt, welche die Struktur des Gesamtdokuments widerspiegelt. Diese Strukturinformationen werden von der Dokumentenverwaltung von XfaWeb-neu (CMS WebGenesis) verwendet, um die HTML-Version auf die internen Objektstrukturen von XfaWeb-neu abzubilden.

Im Konvertierungsschritt 2 werden die der Startseite nachgeordneten HTML-Dateien generiert, indem die in Schritt 1 erzeugten XML-Dateien, wie vorher die Startseite, per XSLT-Prozessor in HTML-Seiten transformiert werden. Zuletzt wird eine ZIP-Datei mit allen HTML-Dateien eines Dokuments sowie den dazu gehörigen Bildern generiert. Die ZIP-Datei kann dann in die Dokumentenverwaltung von XfaWeb-neu eingelesen werden.

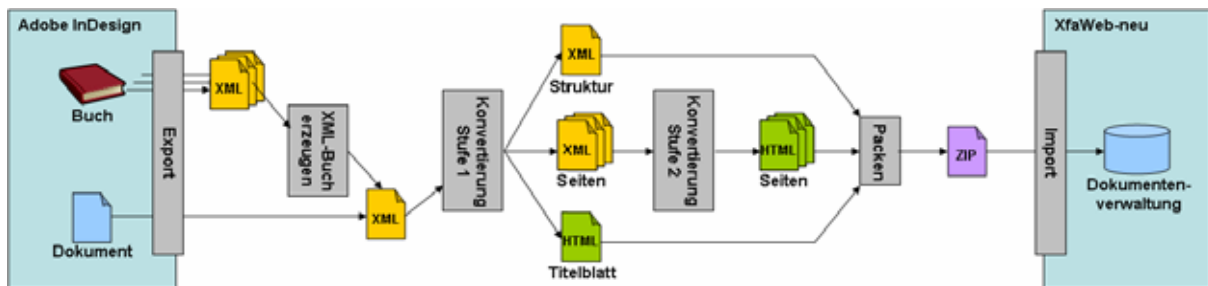


Abbildung 6: Ablaufschema für die Konvertierung von InDesign-Berichten

3.3.3 Fachdokumenten-Browser (FDB)

Der Fachdokumenten-Browser besteht i.W. aus den drei Komponenten Metadatenverwaltung, Crawler und Dienste. Auf der Basis eines Objektmodells für das Gesamtsystem wurde eine erste Version der Metadatenverwaltung implementiert. Die in Java realisierten Objektstrukturen werden über eine Zwischenschicht auf eine über JDBC angekoppelte relationale Datenbank abgebildet. Die Datenbank ist mit dem Open Source Produkt MySQL implementiert.

4. Weiteres Vorgehen

Die Realisierung der verschiedenen Komponenten soll fortgesetzt werden. Bei XfaWeb-neu sollen die Autorenkomponente ausgebaut, Suchfunktionen realisiert und die diversen Schnittstellen (Webservices des FDB, Schnittstellen zum LUBW-Shop, Schnittstelle zu PortalU / UINBW) implementiert werden. Für eine teilautomatisierte Altdatenübernahme müssen Werkzeuge bereitgestellt werden. Neue Funktionen werden zuerst ins Portal AlfaWeb-neu integriert und dort erprobt. Anschließend können erste Prototypen weiterer Fachportale aufgebaut werden. Der InDesign-to-HTML-Konverter soll an konkreten Berichten erprobt und weiterentwickelt werden. Beim FDB werden nach der vollständigen Realisierung der Crawlerfunktionalität diejenigen Fachsysteme angeschlossen, welche die erforderlichen Webservices bereitstellen. Dabei wird mit dem Anschluss von XfaWeb-neu und dem LUBW-Shop be-

gonnen, um mit den gewonnenen Erfahrungen die anderen Systembetreiber besser bei der Implementierung ihrer Webservices unterstützen zu können. In weiteren Schritten werden sukzessive die Dienste des FDB entwickelt.

5. Literatur

- /1/ Weidemann, R.; Ebel, R.; Mayer-Föll, R. (Hrsg.): Fachdokumentenmanagement im Umweltinformationssystem Baden-Württemberg, Forschungszentrum Karlsruhe, Wissenschaftliche Berichte, FZKA-7200, 2005. <http://bibliothek.fzk.de/zb/berichte/FZKA7200.pdf>
- /2/ Weidemann, R. et al.: FADO BW – Konzept für das Fachdokumentenmanagement im Umweltinformationssystem Baden-Württemberg und zur Weiterentwicklung der XfaWeb-Fachinformationssystemfamilie. In Mayer-Föll, R.; Keitel, A.; Geiger, W. (Hrsg.): Kooperative Entwicklung wirtschaftlicher Anwendungen für Umwelt und Verkehr in neuen Verwaltungsstrukturen, Phase I 2005/06, Forschungszentrum Karlsruhe, Wissenschaftliche Berichte, FZKA 7250, S. 7 - 17, 2006. <http://www2.lubw.baden-wuerttemberg.de/lfu/uis/kewa1/03-fzk-fado/kewa1-fzk-fado.html>
- /3/ Paoli, H.; Holtmann, C.; Ebel, R.: UIS-UDDI - Entwicklung eines Web Service-Verzeichnisses für das UIS Baden-Württemberg. In diesem Bericht.
- /4/ Schlachter, T. et al.: UINBW und UINST - Ausbau der Umweltinformationsnetze von Baden-Württemberg und Sachsen-Anhalt; technische Weiterentwicklung. In diesem Bericht.