

UINBW und UINST

Ausbau der Umweltinformationsnetze von Baden-Württemberg und Sachsen-Anhalt; technische Weiterentwicklung

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

*R. Ebel; M. Tauber
Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg
Griesbachstr. 1
76185 Karlsruhe*

*V. Bachmann; B. Köther
Ministerium für Landwirtschaft und Umwelt des Landes Sachsen-Anhalt
Olvenstedter Straße 4
39108 Magdeburg*

*A. Sawade; R. Mayer-Föll
Umweltministerium Baden-Württemberg
Kernerplatz 9
70182 Stuttgart*

| | |
|---|-----------|
| 1. EINLEITUNG | 9 |
| 2. UIN 1.X | 9 |
| 2.1 FUNKTIONALE WEITERENTWICKLUNGEN | 9 |
| 2.2 UINBW | 12 |
| 2.3 UINST..... | 12 |
| 3. UIN 2.0: ENTWICKLUNG EINES KONZEPTS FÜR DEN EINSATZ DER INGRID®- SOFTWARE | 13 |
| 3.1 INFORMATIONSMQUELLEN FÜR PORTALU®..... | 14 |
| 3.2 OPTIONEN..... | 15 |
| 3.3 LÖSUNGSSZENARIEN | 16 |
| 3.4 TESTINSTALLATION | 19 |
| 3.5 WEITERES VORGEHEN..... | 20 |
| 4. LITERATUR..... | 20 |

1. Einleitung

Die Umweltinformationsnetze von Baden-Württemberg (UINBW) und Sachsen-Anhalt (UINST) bieten sowohl Fachnutzern als auch der Öffentlichkeit Zugang zu behördlichen Umweltinformationen im World-Wide-Web. Über mehrere Zugänge, z.B. Umweltthemen, eine Volltextsuche, eine Anbieterliste oder eine Liste von Sites mit aktuellen Messwerten, kann der Nutzer die gewünschten Informationen auffinden.

Bei der Weiterentwicklung der UIN 1.x-Systeme ging es unter anderem darum, die Qualität der bestehenden Zugänge weiter zu verbessern, es sollten aber auch neue Zugangswege entwickelt werden, welche den Suchenden die relevantesten Informationen möglichst schnell und auf einfache Art und Weise auffinden lassen.

Wesentliche Änderungen wird es im geplanten UIN 2.0 geben. Das Konzept hierzu wird ausführlich im Abschnitt 3 beschrieben.

2. UIN 1.x

Die Weiterentwicklung der Umweltinformationsnetze in Baden-Württemberg und Sachsen-Anhalt wurde in zwei Stufen vollzogen (UIN 1.1 und 1.2). Die Versionsbezeichnungen beziehen sich auf die Freischaltung der Unterstützung der Benutzer durch ähnliche Suchbegriffe (1.1) bzw. der Schlagwortsuche (1.2) als Meilensteine. Kleinere Änderungen und Erweiterungen erhielten keine eigenen Versionsnummern.

2.1 Funktionale Weiterentwicklungen

Die wesentlichen Entwicklungen bei den Umweltinformationsnetzen betreffen den Aufbau einer Schlagwortsuche auf Basis der Semantic-Network-Services (SNS), die vom Umweltbundesamt bereitgestellt werden. Kern der SNS ist eine Topic-Map /5/, eine Datenstruktur mit definierten Assoziationen. Bestandteile dieser Topic-Map sind der Umweltthesaurus UmThes®, der Geo-Thesaurus-Umwelt (GTU) sowie eine Umwelt-Chronologie mit aktuellen und historischen Ereignissen, welche die Umweltsituation wesentlich beeinflusst haben. Die SNS boten in ihrer ursprünglichen Fassung, die im Rahmen der Entwicklung des Umweltportals *gein*® gemacht wurden, im Wesentlichen drei Dienste an:

- Automatische Klassifizierung eines Dokumentes (*autoClassify*). Hier werden einem Dokument passende Topics (Deskriptoren aus dem im SNS enthaltenen Umweltthesaurus UmThes) zugeordnet. Im Gegensatz zu einer Volltextindizierung können hier auch Topics vergeben werden, die gar nicht im Text des Dokuments vorkommen.
- Suche nach einem Schlagwort (*findTopics*). Zu einer übergebenen Zeichenkette werden passende Topics zurückgeliefert. Es gibt verschiedene Möglichkeiten, die Qualität der Übereinstimmung (exakte Übereinstimmung, Präfix etc.) zu definieren.
- Suche nach assoziierten Topics zu einem gegebenen Topic (*getPS*).

Alle drei Dienste liefern jeweils ein Fragment der TopicMap zurück, das entsprechend der Anforderung weiterverarbeitet werden kann.

Im Rahmen der Entwicklung des UIN sowie der Entwicklung des PortalU® wurde die Notwendigkeit eines weiteren Dienstes erkannt, der auf Anregung des Projekts im Rahmen eines KoopUIS-Kooperationsprojektes realisiert wurde:

- Der neue Dienst *similarTerms* liefert zu einer gegebenen Zeichenkette verwandte Topics. Dieser Dienst soll insbesondere solche Begriffe liefern, die im Kontext einer Volltextsuche, deren Suchbegriffe als Eingabe dieser Funktion verwendet werden, sinnvoll als Eingabe für einen weiteren Suchlauf der Volltextsuche verwendet werden können.

Auf Basis der automatischen Klassifizierung durch die SNS wurde für die Umweltinformationssysteme ein eigener Schlagwortindex aufgebaut. Kern dieses Indexes ist eine Datenbank, in der für jedes im UIN enthaltene Dokument eine Schlagwortliste verwaltet wird. Hierdurch ist eine Zuordnung eines bestimmten Schlagworts zu passenden Dokumenten möglich, die in verschiedenen Kontexten zur Suche verwendet werden kann.

Um diesen Index aufzubauen ist es erforderlich, jedes Dokument durch die SNS klassifizieren zu lassen und die erhaltenen Schlagworte diesem Index zuzuführen. Dabei wird durch den Indizierer auch eine Bewertung der Schlagworte vorgenommen, zum Beispiel wird das Vorkommen eines Schlagworts im Titel oder der Kurzbeschreibung eines Dokuments mit einem höheren Gewicht belegt. Auch das exklusive Vorkommen eines Schlagworts in einem oder wenigen Dokumenten einer Website erhöht dessen Relevanz. Ein gängiges Problem, insbesondere bei Websites, die von einem Content-Management-System erzeugt werden, ist die auf jeder Seite vorhandene Navigation, die durch die SNS nicht vom eigentlichen Inhalt unterschieden werden kann. Die SNS ordnen Begriffen der Navigation ebenfalls passende Topics zu. Deshalb werden Schlagworte, die in allen Dokumenten einer Website, jedoch nicht im Titel oder der Kurzbeschreibung vorkommen, entsprechend weniger stark gewichtet.

Bei Tests zur Indizierung einer großen Anzahl von Dokumenten zeigte sich, dass eine rein sequenzielle Klassifizierung der Dokumente eine erhebliche Laufzeit beansprucht. Deshalb wurde ein Indizierer entwickelt, der diese Klassifizierungsanfragen parallelisiert und damit eine erhebliche, durch Umgehung von Timeouts sogar superlineare Beschleunigung erzielen kann. Damit ist die Performance der Indizierung im Prinzip durch die Leistungsfähigkeit der SNS begrenzt. Bei Tests zwischen Mai 2006 und März 2007 konnten teilweise über 100.000 Dokumente (überwiegend HTML- und PDF-Dateien) pro Tag verschlagwortet werden, leider wurde diese Performanz aber nicht bei allen Tests erreicht und die Stabilität der SNS war insbesondere beim gleichzeitigen Zugriff von PortalU® und den UIN noch verbesserungswürdig.

Mit der Verfügbarkeit eines Schlagwortindexes ist es nun möglich, eine Schlagwortsuche aufzubauen. Mehrere Ansätze wurden dafür geprüft: Zunächst wurde eine Schlagwortsuche in Analogie zur Schlagwortsuche der XfaWeb-Systeme betrachtet, bei welcher der Nutzer aus einer alphabetischen Schlagwortliste eine Auswahl treffen kann und dann die passenden Dokumente angezeigt bekommt. Dieser Ansatz nutzt jedoch nicht die Möglichkeiten, welche die Assoziationen des UmThes bzw. der SNS bieten. Hier werden auch Querbezüge zwischen Schlagworten, sowie Ober- und Unterbegriffbeziehungen angeboten. Deshalb wurde für die Schlagwortsuche ein Ansatz gewählt, der genau diese Struktur visualisiert und zu jedem Schlagwort automatisch die passenden Dokumente anzeigt (s. Abbildung 1).

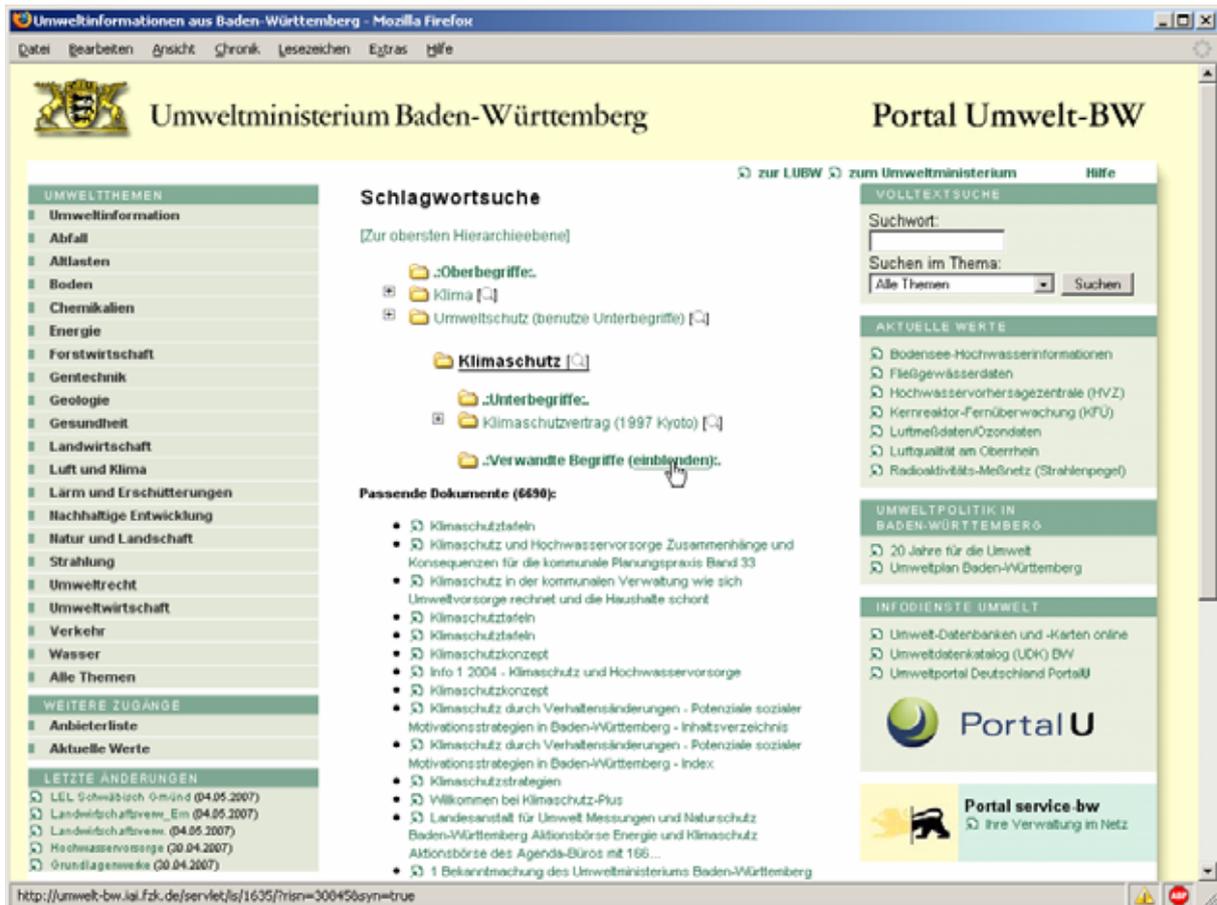


Abbildung 1: Navigation innerhalb der Schlagwortsuche

Auch aus anderen Funktionen soll der Einstieg in diese Schlagwortsuche möglich sein. Einmal kann über die oberste Ebene der Hierarchie von Ober- und Unterbegriffbeziehungen eingestiegen werden. Zum anderen wird bei entsprechenden Suchbegriffen innerhalb der Volltextsuche automatisch auf dazu passende Schlagworte verwiesen. Per Klick ist hier ein Wechsel von der Volltext- zur Schlagwortsuche möglich.

An dieser Stelle greift auch der neu entwickelte SNS-Dienst *similarTerms*. Zu jeder Volltextsuche können – so vorhanden – passende verwandte Begriffe angeboten werden, auf welche der Nutzer die Suchanfrage abändern, erweitern oder einschränken kann.

Die Arbeiten zur Erweiterung der SNS waren zum Zeitpunkt der Erstellung dieses Berichts (Mai 2007) noch nicht vollständig abgeschlossen, so dass abschließende Tests, insbesondere was die Qualität der durch *autoClassify* gelieferten Topics sowie der durch *similarTerms* gelieferten verwandten Begriffe betrifft, noch nicht durchgeführt werden konnten. Erst danach wird die Schlagwortsuche mit all ihren Komponenten im produktiven Betrieb freigeschaltet.

Weitere kleine Entwicklungen wurden im Bereich der Volltextsuchmaschine [ht://Dig](http://Dig) gemacht, deren Weiterentwicklung inzwischen eingestellt wurde. Zum Beispiel wurde die Unterstützung der UTF-8-Zeichencodierung, welche in den letzten Jahren zunehmend an Bedeutung gewinnt, implementiert. Mittelfristig wird die Ablösung dieser Volltextsuchmaschine unumgänglich sein. Eine entsprechende Untersuchung aus dem Jahr 2006 /6/ empfiehlt die Ablösung von [ht://Dig](http://Dig) durch die Open-Source-Produkte Lucene/Nutch, welche auch integraler Bestandteil des SE-iPlug (Search Engine) der InGrid®-Software im PortalU® sind. Szenarien zur möglichen Nutzung dieses iPlug finden sich im Abschnitt 3.

Einige Anregungen und Erweiterungen betrafen die als Kern der UIN eingesetzte Software WebGenesis. Hier wurde prototypisch die Unterstützung von RSS-Feeds entwickelt, die inzwischen durch das IITB ausgebaut wurde und in WebGenesis eingeflossen ist. Eine weitere kleine Entwicklung betrifft die Implementierung einer Schnittstelle zu Google-Sitemaps, welche künftig die Indizierung von Teilbereichen von WebGenesis-Servern – auch durch das UINBW selbst – vereinfachen soll.

2.2 UINBW

Im November 2006 wurde der Umzug des Produktivsystems des UINBW mit seinem Portal Umwelt-BW vom Forschungszentrum Karlsruhe ins ITZ der LUBW nach Stuttgart vollzogen. Gleichzeitig wurde nach einer entsprechenden Schulung der redaktionelle Betrieb vollständig in die Hände der LUBW bzw. eine durch die LUBW beauftragte Firma übergeben.

Im UINBW sind inzwischen rund 170 Websites mit insgesamt ca. 500.000 Dokumenten enthalten, darunter auch die neun größten Kommunen des Landes Baden-Württemberg. Eine wesentliche inhaltliche Hürde stellten die durch die Verwaltungsreform und durch die Einführung von CMS bedingten Änderungen in vielen Webangeboten des Landes und seiner Behörden dar. Hier mussten viele zeitlich teilweise weit auseinander liegende Änderungen nachvollzogen werden. Die Qualitätssicherungswerkzeuge, die unter anderem die Nichtverfügbarkeit von Websites erkennen, waren hierbei eine große Hilfe.

Die im PortalU® gegenüber *gein*® verfeinerten Umweltthemen wurden im UINBW inzwischen nachgezogen. Die Volltextsuche des UINBW wird nun auch im Themenpark Umwelt /7/ verwendet. Für die Websites der LUBW und des UM soll künftig ebenfalls die Volltextsuche des UINBW zum Einsatz kommen. Damit wird das UINBW als zentraler Suchdienst im UIS Baden-Württemberg positioniert.

Für das UIS Baden-Württemberg und das UINBW wurde ein Konzept zum Einsatz der InGrid®-Software und zum Aufbau eines UINBW 2.0 erstellt (s. Abschnitt 3).

2.3 UINST

Das Umweltinformationsnetz Sachsen-Anhalt ist seit Januar 2006 im produktiven Betrieb. Im Herbst 2006 wurde die redaktionelle Arbeit vollständig durch das MLU Sachsen-Anhalt bzw. eine durch das MLU beauftragte Firma übernommen. Auch das UINST wurde auf die im PortalU® verwendeten Umweltthemen umgestellt. Im UINST sind derzeit rund 450 Websites mit ca. 20.000 Dokumenten enthalten; die Erfassung von Websites ist hier deutlich feingranularer als im UINBW und geht hinab bis zur Erfassung einzelner Dokumente.

Das Layout des UINST wird inzwischen durch eine spezielle Schnittstelle direkt aus dem Landesportal Sachsen-Anhalt importiert, so dass dort gemachte Änderungen transparent im Umweltportal erscheinen. Die vom MLU Sachsen-Anhalt kommende Anforderung zur Bereitstellung wechselnder aktueller Bilder mit Erklärungstext auf der Startseite des UINST wurde kurzfristig umgesetzt (s. Abbildung 2).

Auch für das UINST wurde ein Konzept zum Einsatz der InGrid®-Software und zum Aufbau eines UINST 2.0 erstellt. Dazu mehr im folgenden Abschnitt.

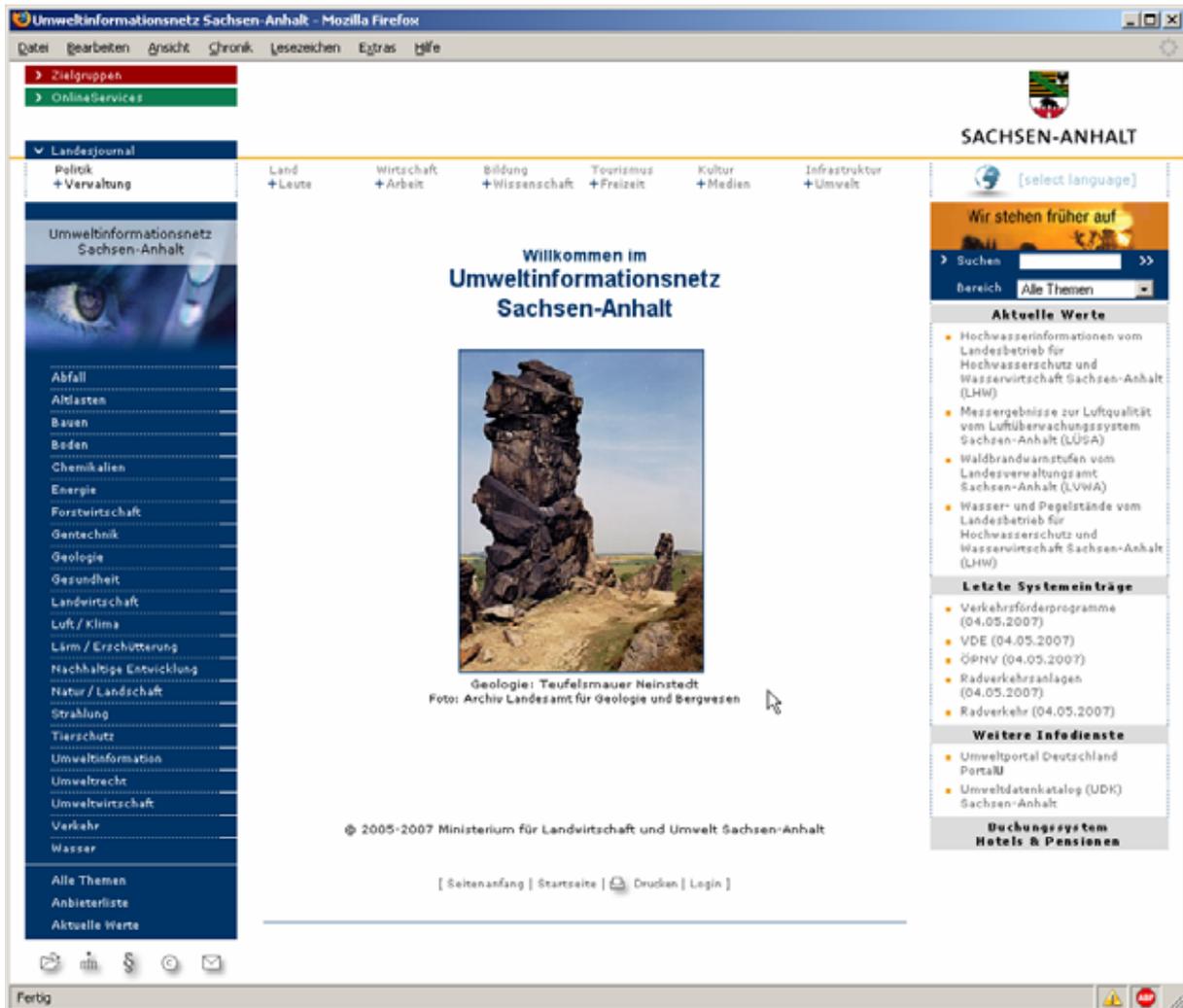


Abbildung 2: Das UIN Sachsen-Anhalt mit wechselndem Foto auf der Startseite.

3. UIN 2.0: Entwicklung eines Konzepts für den Einsatz der InGrid®-Software

Im Rahmen einer Bund-/Länderkooperation wird als Nachfolger von *gein*® das PortalU® inkl. der darunter liegenden Softwareschicht InGrid® entwickelt. Ab Version 1.1 wird InGrid® auch die Rolle des Umweltdatenkatalogs (UDK), insbesondere bei der Erfassung von Metadaten, übernehmen. Der Produktionsbetrieb von PortalU® begann im Mai 2006. Gleichzeitig entwickeln und betreiben, wie in Kapitel 2 beschrieben, Baden-Württemberg (BW) und Sachsen-Anhalt (ST) in einem gemeinsamen Projekt Landesumweltportale (Umweltinformationsnetze Baden-Württemberg und Sachsen-Anhalt). In beiden Bundesländern besteht starkes Interesse, die Entwicklungen stärker mit PortalU® zusammenzuführen und insbesondere Konzepte und Komponenten von InGrid® in die Landessysteme zu übernehmen, soweit dies fachlich sinnvoll, technisch möglich und wirtschaftlich vorteilhaft ist.

Daher wurden zuerst in zwei getrennten Untersuchungen die Einsatzmöglichkeiten von InGrid®-Komponenten in den Umweltinformationssystemen Baden-Württemberg (im Herbst 2006) und Sachsen-Anhalt (Frühjahr 2007) betrachtet und bewertet. Die Analysen basierten auf dem DV-technischen Feinkonzept PortalU® / InGrid® 1.0 / 1.1 (insbesondere Teile iBus, iPlug) und der jeweils aktuellen Version von PortalU® /2, 3/.

Die beiden Untersuchungen wurden nahezu gleich strukturiert, um einen unmittelbaren Abgleich zu ermöglichen. Vorangestellt wurde ein kurzer technischer Überblick zu InGrid® und PortalU®. InGrid® ist ein modularer Anwendungsrahmen mit definierten Serviceschnittstellen zwischen den eigenständigen Komponenten, konzipiert für die Suche in verteilten, unterschiedlich strukturierten Datenquellen. Kernkomponenten von InGrid® sind der iBus, der eine Anfrage des Front-End an die im Einzelfall relevanten Datenquellen verteilt und die zusammengeführten Ergebnisse zurückliefert, und die iPlugs, das sind netzwerkfähige, auf verschiedenartige Datenquellen spezialisierte Abfragemodule als Schnittstelle zwischen iBus und Datenquelle. Es wird zwischen iPlugs mit Quellenadapter und solchen mit Serviceadapter unterschieden. Während im ersten Fall InGrid® die Datenquelle indizieren und damit voll in die Suche integrieren kann („linke“ Ergebnisliste), werden im zweiten Fall zur Laufzeit Suchanfragen an die Datenquelle weitergereicht und lediglich Verweise auf die separate Ergebnisliste der Datenquelle zurückgegeben („rechte“ Ergebnisliste). Das mit InGrid® realisierte PortalU® bietet im Wesentlichen eine allgemeine Suchfunktion über per Volltextcrawler indizierte Seiten und über UDK-Objekte, eine spezialisierte Suche in Adressen und Daten zu Forschungsprojekten, verschiedene Kataloge mit Einstiegsseiten zu Informationsangeboten, u.a. zu Messwerten und Umweltthemen sowie raum- (auch karten-) und zeitbasierte Suchfunktionen.

3.1 Informationsquellen für PortalU®

Im Zusammenspiel der Umweltinformationssysteme von Baden-Württemberg und Sachsen-Anhalt einerseits und von InGrid® und PortalU® andererseits sind zwei miteinander verknüpfte Aspekte zu berücksichtigen: Zum einen sollen die behördlichen Umweltinformationen der Länder umfassend in PortalU® eingebracht werden, wozu bereits InGrid®-Komponenten benötigt werden, und zum anderen sollen InGrid®-Komponenten genutzt werden, um eigene Umweltportale zu realisieren. Dabei sollen Doppelerfassungen von Umweltinformationen vermieden und Suchfunktionen vereinheitlicht und zusammengeführt werden, um Synergieeffekte zu nutzen. In den durchgeführten Untersuchungen wurde zuerst analysiert, welche Informationen aus den Landesumweltinformationssystemen an PortalU® angeschlossen werden sollen. Es wurde dabei unterschieden nach Informationsseiten, Fachdokumenten, UDK-Metadaten, Daten zu Forschungsprojekten, Karten, Sachdaten und Katalogeinträgen. In beiden Bundesländern liegt der Schwerpunkt im Bereich Informationsseiten (Volltextsuche), UDK-Metadaten und Forschungsprojekten, in BW zusätzlich bei Fachdokumenten und Karten in ST bei Katalogeinträgen. Da mit der derzeit in Entwicklung befindlichen Version 1.1. von InGrid® zusätzlich noch die Ablösung des Windows-basierten UDKs und dessen Internetversion WWW-UDK durch eine in InGrid® integrierte voll web-basierte Lösung möglich oder erforderlich wird, wurde auch die Metadatenstrategie der Länder hinterfragt. Während ST den UDK 1:1 mit InGrid® ersetzen will, plant BW ein differenziertes Vorgehen. Hier soll die Metadatenverwaltung dezentralisiert und näher an die Datenbesitzer herangeführt werden. Durch die Schnittstellenphilosophie von InGrid® kann die Pflege der UDK-Metadaten in dedizierten Systemen wie z.B. dem Fachdokumenten-Browser für Daten zu Fachdokumenten erfolgen, um dann die Metadaten über geeignete iPlugs in PortalU® oder einer anderen InGrid®-Installation zusammenzuführen.

Da die betroffenen Informationen auch derzeit schon in existierenden Fachsystemen und Portalen in den Umweltinformationssystemen der Länder erschlossen werden oder in geplanten erschlossen werden sollen, wurden auch diese mit in die Überlegungen einbezogen,

um zu entscheiden, ob Zugänge überflüssig werden oder mehrfach nutzbare Dienst zur Verfügung gestellt werden können.

3.2 Optionen

Breiten Raum in den Untersuchungen zum Einsatz von InGrid®-Komponenten nimmt die Beschreibung und Bewertung möglicher Einsatzoptionen ein. Durch die modulare Struktur von InGrid® und die Kopplung der Module über netzbasierte Schnittstellen ist eine äußerst flexible Konfiguration von Systemlösungen möglich. Das Spektrum reicht dabei von einer oder sogar mehreren Vollinstallationen in den Ländern bis zu einer weitgehenden Nutzung von PortalU®-Funktionen unter Verzicht auf eine eigene Installation. Dabei wurde versucht, bestimmte Komponenten zuerst isoliert zu betrachten, um frühzeitig auf sinnvolle Szenarien einschränken zu können und nicht die ganze kombinatorische Vielfalt an Szenarien bearbeiten zu müssen.

Zur ersten Komponente, der Benutzeroberfläche, existieren zwei Hauptalternativen. Entweder wird das Portal-Modul von InGrid® genutzt oder es wird eine eigene Portaloberfläche entwickelt, welche die OpenSearch-Schnittstelle von InGrid® verwendet, um Abfragen und Suchergebnisse mit InGrid® auszutauschen. Das zweite Verfahren ist aufwändiger, dafür aber flexibler, sodass auch von PortalU® abweichende Funktionalitäten angeboten und Oberflächen im vorgeschriebenen Landeslayout erstellt werden können. Über die OpenSearch-Schnittstelle können Indexfelder aus der InGrid®-Datenverwaltung angesprochen werden, die über PortalU® nicht verfügbar sind. Im Gegenzug sind über OpenSearch die per Serviceadapter angeschlossenen iPlugs nicht verfügbar.

Die zentrale Komponente iBus ist das Kernstück einer InGrid®-Installation. Wird auf die Installation des iBus im Landes-UIS verzichtet, bedeutet dies, dass PortalU® als alleiniges Sammelbecken für die Informationen des Landes dienen muss. Das Landesumweltportal ist dann eine alternative Benutzeroberfläche zu PortalU®, die Suchergebnisse landesspezifisch filtert. Ein eigener iBus ermöglicht von PortalU® abweichende Inhalte und landesspezifische Anpassungen, z.B. beim Ranking der Suchergebnisse. Der unmittelbare Zugriff auf das System wird dabei mit erhöhtem Pflegeaufwand erkauft. Die teilweise redundante Datenhaltung mit PortalU® ist insofern kein Problem, da die iPlugs an mehrere Systeme gleichzeitig angeschlossen werden können, so dass diese parallel mit den gleichen Informationen versorgt werden, oder das Landes-InGrid® als Drehscheibe fungieren und PortalU® nachgeordnet mit Informationen versorgen kann.

Beim Vorgehen bezüglich Pflege und Bereitstellung von UDK-Metadaten ist eine kurzfristige Phase, bei der noch der bisherige UDK eingesetzt wird, von der mittel- bis langfristigen Phase, in der der UDK abgelöst wird, zu unterscheiden. Kurzfristig kann der UDK über eine Export-/Import-Schnittstelle oder über iPlugs mit Quellen- bzw. Serviceadapter an InGrid® angeschlossen werden. Das mögliche Vorgehen in Phase II hängt von der Metadatenstrategie des Landes ab.

Eine relativ komplexe Fragestellung ist die Realisierung der Volltextsuche. Basis ist die Indizierung von Webseiten über einen SE-iPlug (SE = Search Engine) und die Suchmaschine Nutch/Lucene. Zusätzliche Anforderungen sind, dass ein Teil der Informationsseiten nur im Landesumweltportal verfügbar sein darf, es verschiedene Suchebenen (Suchräume) gibt und die Suchfunktionalität von UIN 1.x nachgebildet werden soll. UIN 1.x filtert und bearbeitet regelbasiert die URLs der Inhalte der integrierten Informationsangebote und stellt speziali-

sierte Schnittstellen für Angebote bereit, die ansonsten nicht einzubinden wären. Außerdem wird eine thematische Volltextsuche angeboten, indem für jedes Thema ein eigener Volltextindex aufgebaut wird. Die hierzu betrachteten Optionen unterscheiden sich im Wesentlichen in der Anzahl der SE-iPlugs, deren Anbindung und in der Frage, ob und wie massiv diese zu modifizieren sind. Eine Out-of-the-box Lösung für alle Anforderungen gibt es nicht.

Das WMS-Modul von InGrid® ist für die karten-basierte Suche zuständig. Dies könnte nur bei einer InGrid®-Vollinstallation im Landes-UIS genutzt werden, wenn auch das Portal-Modul eingesetzt wird, da beide gekoppelt sind. Dem kommt entgegen, dass es für BW ausreicht, aus Metadaten (z.B. UDK o.a.) auf ein vorhandenes Portal mit eigenen WMS-Diensten (UDO Umweltdatenbanken und –karten online) zu verlinken und ST vorerst auf WMS-Dienste verzichtet, sodass auf das WMS-Modul keine Rücksicht genommen werden muss.

Einträge in Kataloge müssen manuell über die Autorenoberfläche von InGrid® vorgenommen werden. Ein automatisierter Abgleich zwischen Systemen wie er sonst durch die iPlugs erfolgen kann, ist hier nicht vorgesehen. Entsprechend müssen die Kataloge der verschiedenen InGrid®-Systeme entweder parallel gepflegt werden oder die Portale nutzen den Katalog einer ganz bestimmten Installation, am besten die des PortalU® über die OpenSearch-Schnittstelle. Die möglichen Optionen sind u.a. davon abhängig, ob die Inhalte der Kataloge gleich sind.

3.3 Lösungsszenarien

Basierend auf den vorliegenden Randbedingungen und den beschriebenen Optionen wurde für BW und ST jeweils ein Hauptvorschlagsszenario entworfen und zusammen mit einigen plausiblen Varianten näher ausgeführt. Die Eckpunkte der miteinander abgestimmten Szenarien sind wie folgt:

- Es wird eine eigene InGrid®-Installation in den Umweltinformationssystemen der Länder vorgenommen.
- Statt des Portal-Moduls werden jedoch eigene Portale nach den Landesdesignvorgaben unter Verwendung der OpenSearch-Schnittstelle realisiert.
- Die Portale verfügen über eine eigene Metadatenverwaltung (Basis WebGenesis) zur Parametrisierung der Volltextsuche (iPlugs).
- Die Suchfunktionalität von InGrid® wird dahingehend erweitert, dass eine thematische und anbieterspezifische Suche (leichter) möglich wird.

InGrid®-Installation und Umweltportal bilden zusammen UIN 2.0, d.h. die nächste Generation der Umweltinformationsnetze der Länder Baden-Württemberg und Sachsen-Anhalt. Die folgenden Abbildungen 3 bis 5 visualisieren die geplanten Strukturen.

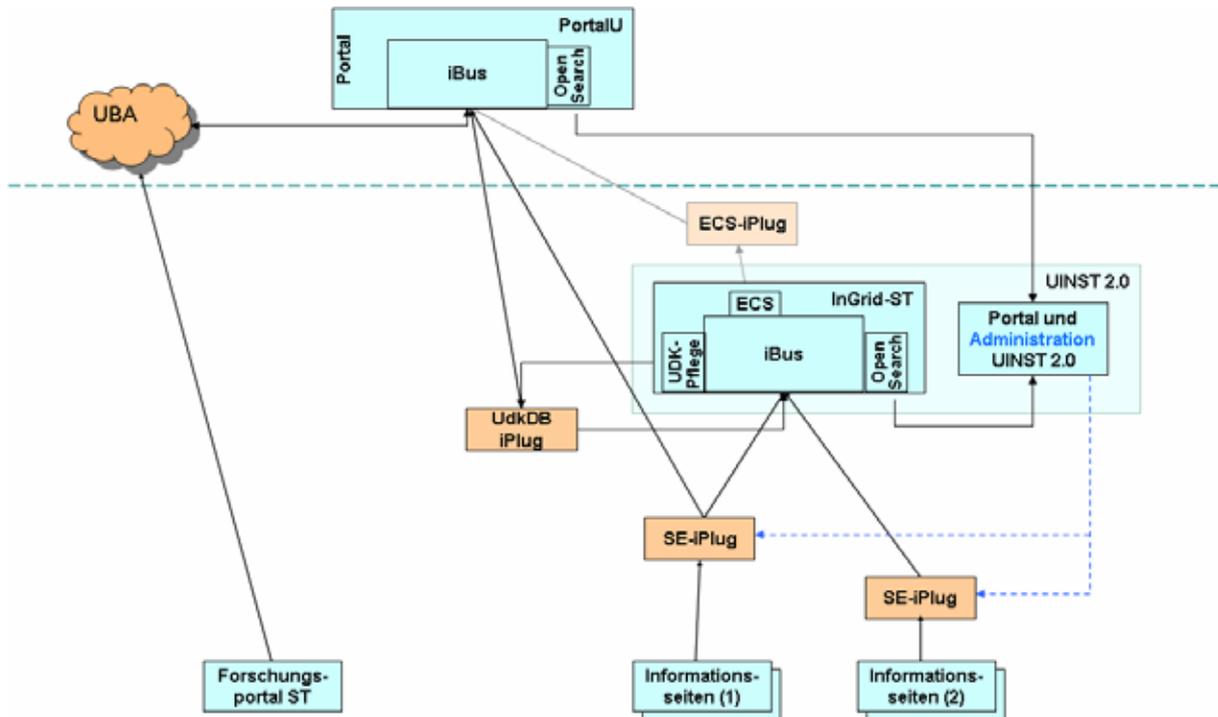


Abbildung 3: Geplanter Einsatz von InGrid®-Komponenten im UIS ST

Die Strukturen für Sachsen-Anhalt (Abbildung 3) sind relativ überschaubar. Zur Indizierung für die Volltextsuche werden zwei getrennte SE-iPlugs eingesetzt, um Informationsseiten wahlweise nur in UINST bzw. zusätzlich auch in PortalU® einbinden zu können. Die Portal-Komponente von UINST 2.0 steuert und parametrisiert die Indizierung. Die Erfassung der UDK-Metadaten erfolgt über die entsprechende Pflegekomponente von InGrid®-ST. Dadurch, dass hier ein UdkDB-iPlug zwischengeschaltet ist, können die Metadaten gleichzeitig in UINST 2.0 und PortalU® eingebunden werden. Etwas blasser ist ein ECS-iPlug (ECS = Environmental Catalogue Service) eingezeichnet, der alternativ zur direkten Anbindung des UdkDB-iPlugs an PortalU® eingesetzt werden kann. Je nach Wahl könnten dadurch Suchergebnisse aus UDK-Metadaten entweder in die gerankte Gesamtergebnisliste (linke Ergebnisliste) von PortalU® oder als Block in die rechte Ergebnisliste eingebaut werden. Forschungsprojekte aus dem Forschungsportal Sachsen-Anhalt sind bereits auf anderem Weg über das Umweltbundesamt in PortalU® integriert und sollen nicht in UINST erscheinen.

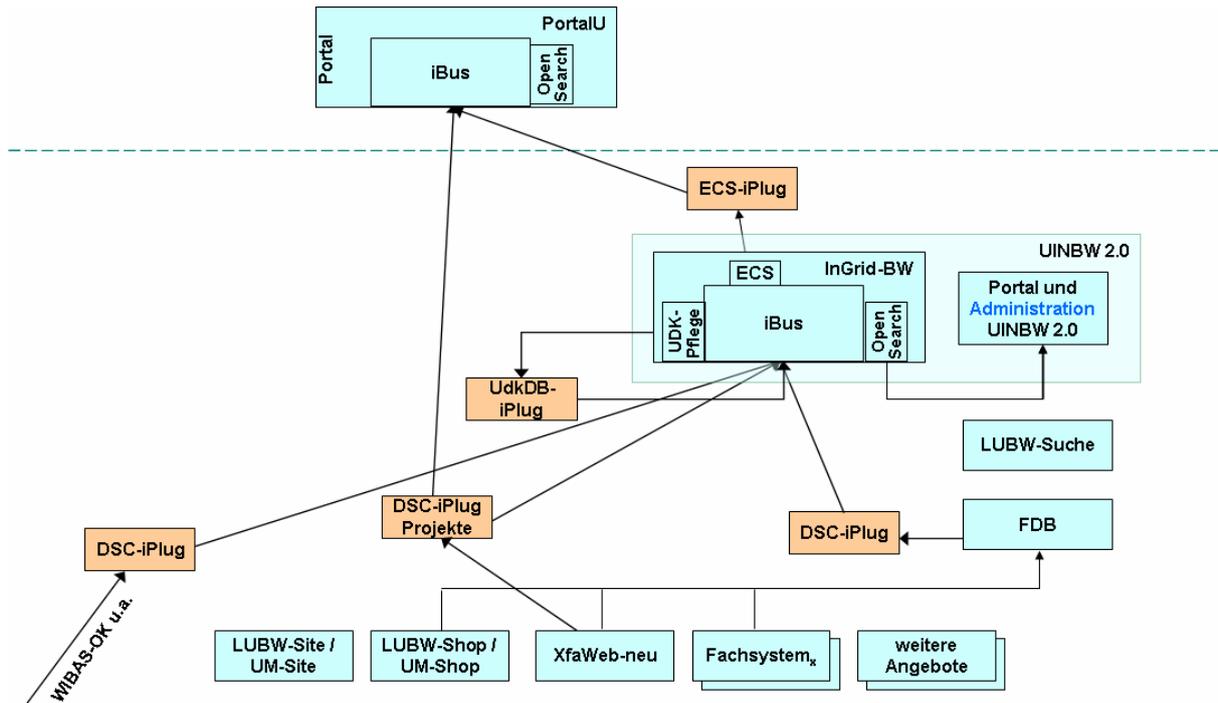


Abbildung 4: Geplanter Einsatz von InGrid®-Komponenten im UIS BW – Metadaten

Da die Verhältnisse in Baden-Württemberg etwas komplexer sind, werden zwei getrennte Schaubilder verwendet. Abbildung 4 enthält den Metadatenteil. Neben den wie in Sachsen-Anhalt über die Pflegekomponente von InGrid® verwalteten UDK-Objekten werden über weitere iPlugs Metadaten eingesammelt. Fach- und Geodaten könnten über den WIBAS-OK bzw. eine in Entwicklung befindliche Preludio-Installation (Metadatenverwaltungssystem der Firma disy) eingebracht werden. Metadaten zu Fachdokumenten soll der im Rahmen des Projekts FADO geplante Fachdokumenten-Browser (FDB) von allen Fachsystemen des UIS BW, die mit Fachdokumenten umgehen, aufsammeln und über einen DSC-iPlug (DSC = Data Source Client) an InGrid®-BW weitergeben. InGrid®-BW fungiert so als Sammelstelle für alle UDK-Metadaten aus Baden-Württemberg. Recherchen sind über UINBW 2.0 und über das per ECS-iPlug (Serviceadapter) angebundene PortalU® möglich. Alternativ zum ECS-iPlug könnten alle anderen iPlugs auch zusätzlich direkt an PortalU® angedockt werden (nicht eingezeichnet), um PortalU® die Indizierung zu erlauben und dadurch in die gerankte Ergebnisliste aufgenommen zu werden. Neben den UDK-Metadaten werden auch noch Metadaten zu Forschungsprojekten an InGrid®-BW und PortalU® geliefert. Die Schnittstelle realisiert ein DSC-iPlug, der die Metadaten von XfaWeb-neu (siehe /4/) entgegennimmt.

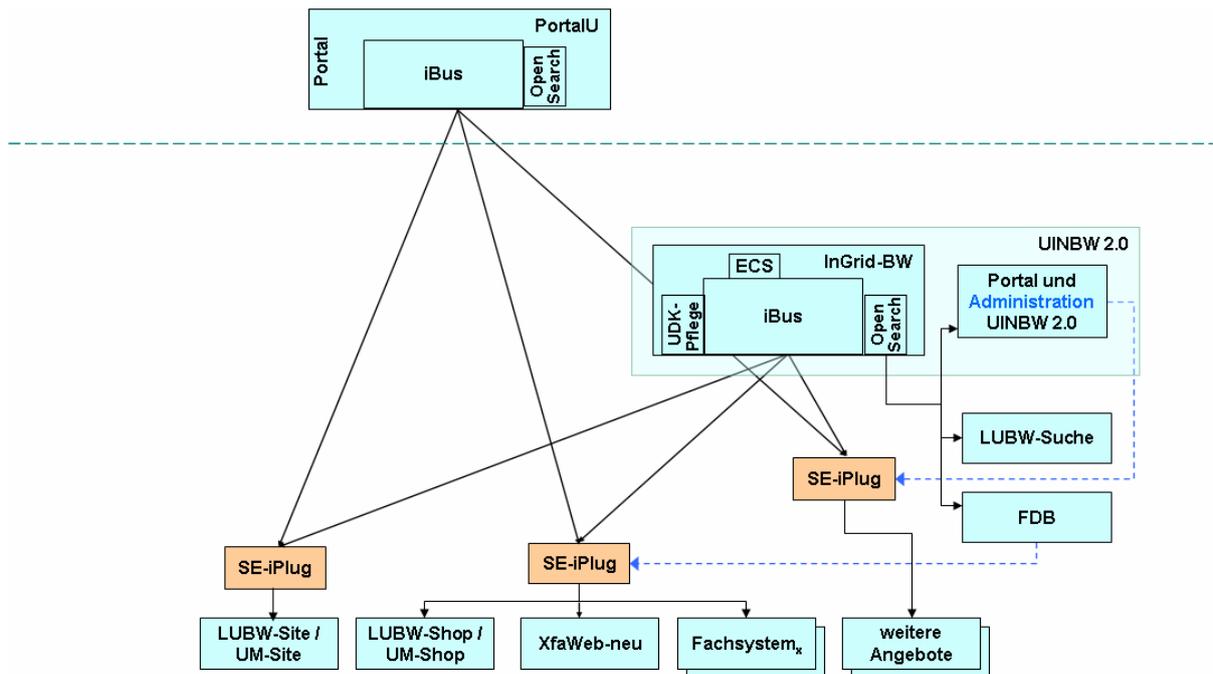


Abbildung 5: Geplanter Einsatz von InGrid®-Komponenten im UIS BW - Volltextsuche

Zur Realisierung der Volltextsuche unter den gegebenen Randbedingungen sind mehrere (mindestens drei, wie in Abbildung 5 eingezeichnet) SE-iPlugs notwendig. Die Verteilung der Indizierung auf mehrere Schnittstellenkomponenten erlaubt unterschiedliche Inhalte für PortalU® und InGrid®-BW (derzeit in Baden-Württemberg nicht vorgesehen), aber auch unterschiedliche Suchebenen/-räume. Der Suchindex von InGrid®-BW soll nicht nur zur Realisierung der Suche in UINBW 2.0, sondern auch für Suchfunktionen des FDB und der LUBW-Website genutzt werden. Man erspart sich dadurch die Mehrfachindizierung der Informationsseiten und die Wartung und Pflege mehrerer Instanzen der Suchmaschine oder sogar verschiedener Suchwerkzeuge. Da der Suchindex zu jeder Informationsseite auch den Bezeichner des SE-iPlugs speichert, können Suchanfragen über die Angabe dieser iPlug-Bezeichner auf den gewünschten Suchraum beschränkt werden. Zudem können SE-iPlugs angepasst werden, um besondere Suchstrategien zu realisieren. Im UIS BW kommt das an zwei Stellen zum Tragen. Zum einen wird ein SE-iPlug benötigt, mit dem die bisherige UINBW-Strategie nachgebildet werden kann, d.h. eine regelbasierte Filterung von URLs um z.B. Dubletten (gleiche Seite, verschiedene URLs) zu beseitigen und spezielle Schnittstellen für ansonsten nicht indizierbare Angebote (z.B. StaLA). Die Parametrisierung dieses iPlugs soll über die Autorenumgebung und die Metadatenverwaltung der Portal-Komponente von UINBW 2.0 erfolgen. Zum anderen wird für die Indizierung von Fachdokumenten ein spezieller SE-iPlug benötigt, dem man explizit die Liste der zu indizierenden URLs vorgeben kann (keine Crawlerfunktion). Dies soll über den FDB erfolgen, der in seiner Metadatenverwaltung alle relevanten URLs bereits kennt.

3.4 Testinstallation

Um die auf der Basis der Spezifikation von InGrid® und Aussagen der Koordinierungsstelle PortalU® getroffenen Annahmen zu verifizieren und dadurch das Konzept auf eine solidere Basis stellen zu können, soll eine Testinstallation der InGrid®-Software vorgenommen werden. Da die Fertigstellung und Abnahme von InGrid® mehrfach verschoben wurde, kann die Testinstallation erst während der zweiten Jahreshälfte 2007 vorgenommen werden.

3.5 Weiteres Vorgehen

Im nächsten Schritt soll ein gemeinsames Konzept für UIN 2.0 erstellt werden, das sowohl die Untersuchungen in Baden-Württemberg und Sachsen-Anhalt berücksichtigt, als auch die Erkenntnisse aus der Evaluierung der InGrid®-Software mit einbezieht. Ziel ist, auf einer gemeinsamen und flexiblen Softwarebasis die individuell ausgeprägten Länderumweltportale realisieren zu können. In der Zwischenzeit bis zur Fertigstellung des UIN 2.0 werden die bisherigen Versionen von UINBW und UINST bedarfsorientiert weiterentwickelt, inhaltlich ausgebaut und als Internetportale angeboten.

4. Literatur

- /1/ Schlachter, T. et al.: UINBW und UINST – Entwicklung, Betrieb und Ausbau erster Produktionsversionen der Umweltinformationsnetze von Baden-Württemberg und Sachsen-Anhalt. 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. 53 - 62, 2006.
<http://www2.lubw.baden-wuerttemberg.de/ifu/uis/kewa1/08-fzk-uin/kewa1-fzk-uin.html>
- /2/ Klenke, Martin; Kruse, Fred; Lehmann, Hanno; Riegel, Thomas; Vögele, Thomas: InGrid® 1.0 - The Nuts and Bolts of PortalU®. In Tochtermann, Klaus; Scharl, Arno (Hrsg.): Managing Environmental Knowledge, Shaker-Verlag, Aachen, 2006.
- /3/ Projektdokumente zu PortalU® / InGrid® 1.x. <http://www.kst.portalu.de/portalu/dokumente.html> (besucht am 21.03.2007).
- /4/ Weidemann, R. et al.: FADO BW - Realisierung erster Komponenten für ein verteiltes Fachdokumentenmanagement im Umweltinformationssystem Baden-Württemberg. In diesem Bericht.
- /5/ Smolnik, S.: Wissensmanagement mit Topic Maps in kollaborativen Umgebungen - Identifikation, Explikation und Visualisierung von semantischen Netzwerken in organisationalen Gedächtnissen, Shaker-Verlag, Aachen, 2006.
- /6/ Schlachter, T. et al.: Volltextsuchmaschinen für die Umweltinformationsnetze Baden-Württemberg und Sachsen-Anhalt, interner Bericht, FZK, Mai 2006.
- /7/ Döpmeier, C. et al.: Fortentwicklung des Themenparks Umwelt, Erprobung von Web 2.0 Technologien. In diesem Bericht.