

# LUPO

## **Ausbau der Suchfunktionalität der Landesumweltportale und Vernetzung mit dem Umweltportal Deutschland**

*T. Schlachter; W. Geiger; R. Weidemann; G. Zilly  
Karlsruher Institut für Technologie  
Institut für Angewandte Informatik  
Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen*

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

*A. Sawade; K. Zetzmann  
Ministerium für Umwelt, Naturschutz und Verkehr Baden-Württemberg  
Kernerplatz 9, 70182 Stuttgart*

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

*D. Keil  
Thüringer Ministerium für Landwirtschaft, Forsten, Umwelt und Naturschutz  
Beethovenstr. 3, 99096 Erfurt*

*A. Müller  
Ministerium für Umwelt, Forsten und Verbraucherschutz Rheinland Pfalz  
Kaiser-Friedrich-Str. 1, 55116 Mainz*

*Thomas Sattler  
DECON-network Systemhaus & EDV-Vertriebs GmbH  
Bannwaldallee 24, 76185 Karlsruhe*

*U. Keim  
K2 & Partner Managementberatung  
Wörthstr. 8, 65343 Eltville*

<b>1. LUPO LANDESUMWELTPORTALE.....</b>	<b>11</b>
<b>2. NUTZUNG DER LUPO-VOLLTEXTSUCHE IN PORTALU® .....</b>	<b>11</b>
2.1 DER OPENSEARCH-STANDARD.....	12
2.2 ERWEITERUNG DES OPENSEARCH-STANDARDS .....	13
<b>3. TEILANGEBOTE VON WEBSITES.....</b>	<b>15</b>
3.1 WEBGENESIS MIT „SPRECHENDEN URLS“ .....	15
3.2 GENERIERUNG VON SITEMAPS .....	17
<b>4. GLIEDERUNG DER ANBIETERLISTE.....</b>	<b>17</b>
<b>5. GEOPORTAL WASSER IN RHEINLAND-PFALZ .....</b>	<b>18</b>
<b>6. ERFAHRUNGEN BEIM BETRIEB.....</b>	<b>19</b>
<b>7. AUSBLICK.....</b>	<b>20</b>
<b>8. LITERATUR.....</b>	<b>20</b>

## 1. LUPO Landesumweltportale

Die Landesumweltportale (LUPO) der Länder Baden-Württemberg, Sachsen-Anhalt und Thüringen sowie das Geoportal Wasser in Rheinland-Pfalz bieten leistungsstarke Suchfunktionen zu Umweltinformationen dieser Länder an /1/. Damit leisten sie einen wesentlichen Beitrag zu der in den Umweltinformationsgesetzen geforderten aktiven Verbreitung von Umweltinformationen. Um diese Suchfunktionen noch breiter und universeller nutzen zu können, stellen die Umweltportale ihre Dienste nun auch über standardisierte Schnittstellen zur Verfügung. Hierdurch werden in der Zusammenarbeit mit anderen Portalen, wie dem Umweltportal Deutschland PortalU<sup>®</sup>, Synergien genutzt. Neben der Beschreibung dieser Schnittstellen geht dieser Beitrag auch auf die verbesserte, automatisierte Konfiguration der Suchmaschinen über sogenannte sprechende URLs und Sitemaps ein.

## 2. Nutzung der LUPO-Volltextsuche in PortalU<sup>®</sup>

Derzeit wird das deutsche Umweltportal PortalU parallel zu weiteren Umweltportalen der Länder, darunter mehrere LUPO-Portale, betrieben. All diese Umweltportale verfügen jeweils über eine eigene Volltextsuche, die jedoch auf unterschiedlicher technischer Basis realisiert sind. Da viele Informationen sowohl im PortalU als auch in den Landesumweltportalen enthalten sind, werden diese zum Teil mehrfach indiziert.

Dabei entstehen unter anderem folgende Probleme:

- Erhöhte Serverlast durch mehrfaches Crawling der einzelnen Informationsangebote
- Mehrfacher Konfigurations- und Administrationsaufwand für alle Informationsangebote, die in mehreren Portalen bzw. Suchindexen enthalten sind
- Mehrfachdatenhaltung (Speicherbedarf für Indexe, Übertragungskapazitäten, Rechenzeit zur Indizierung der Informationen)
- Ggf. Inkonsistenzen bei unterschiedlichem Stand der verschiedenen Indexe und bei laufzeitbedingten langen Indizierungszyklen

Bereits im Jahr 2008 wurde ein Konzept erarbeitet, um die Suchindexe der LUPO-Portale auch im PortalU verfügbar zu machen und dort für die Volltextsuche zu nutzen. Der dabei erwartete Gewinn erstreckt sich auf folgende Punkte:

- Die verteilte Indexpflege sorgt dafür, dass Meta-Informationen und Konfigurationsdaten an jeweils nur einer Stelle zu pflegen sind
- Entlastung der Informationsserver durch Wegfall der Mehrfachindizierung
- Reduzierter Indizierungsaufwand beim PortalU und dadurch eine insgesamt reduzierte Laufzeit zur Erstellung bzw. zum Update des Volltextindex
- Reduzierung von Inkonsistenzen

Herausforderungen bei der Realisierung dieser Idee waren:

- Einigung auf ein einheitliches, möglichst standardkonformes Austauschformat
- Konfigurierbarkeit verschiedener Suchmaschinen, insbesondere der in den Ländern verwendeten

- Möglichkeit zur Konfiguration einer Suchkaskade
- Akzeptable Antwortzeiten bei der verteilten Suche
- Integration von Antworten (Trefferlisten) zu einer Ergebnisliste, insb. unter Berücksichtigung der unterschiedlichen Ranking-Strategien der Suchmaschinen

Die Untersuchung von Möglichkeiten zur standardkonformen Umsetzung dieser Anforderungen führte zu einer Entscheidung für das OpenSearch-Format [2], das sowohl die Möglichkeit der Beschreibung des Zugriffs auf Suchmaschinen als auch das entsprechende Format für Ergebnislisten bietet.

Das Konzept sah dabei die Entwicklung eines OpenSearch-iPlugs zur Anbindung an die InGrid-Software von PortalU vor. Auf Seiten der Landesumweltportale war demnach ebenfalls die Kompatibilität zu OpenSearch zu implementieren. Abbildung 1 zeigt die Struktur dieses Konzepts.

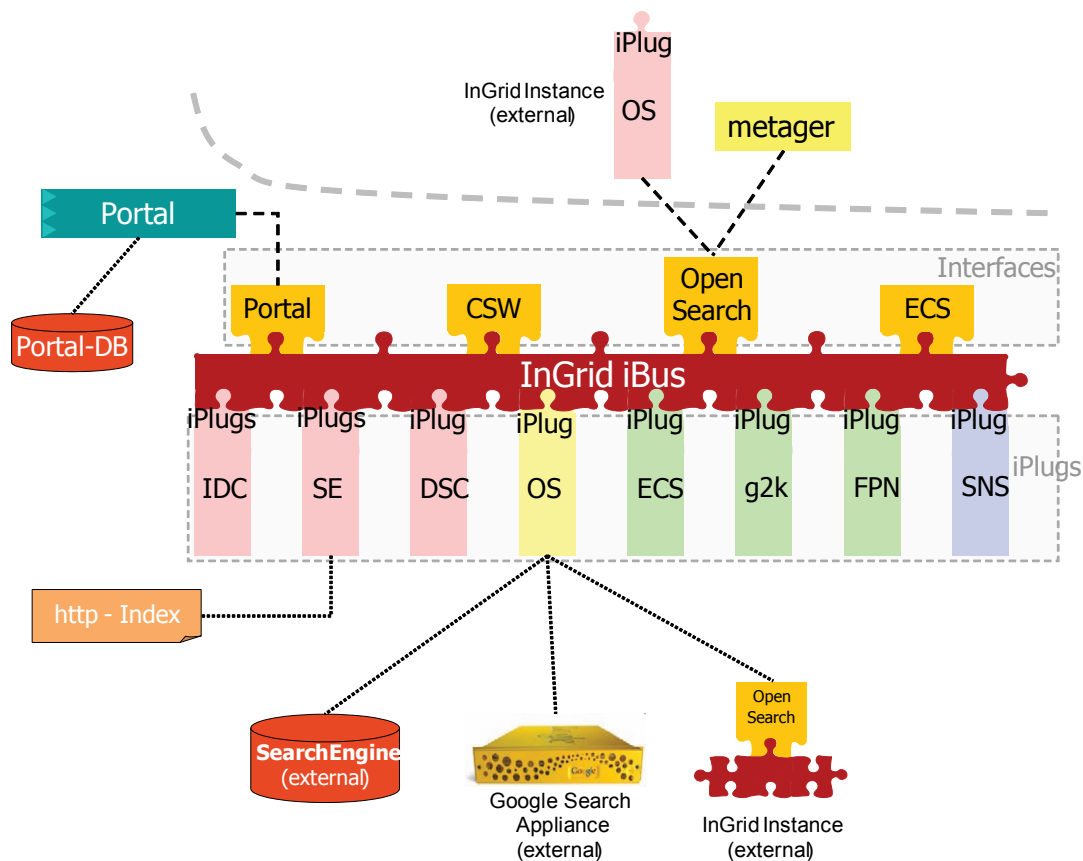


Abbildung 1: OpenSearch-iPlug in PortalU (Grafik der Koordinierungsstelle PortalU)

## 2.1 Der OpenSearch-Standard

Bei OpenSearch (OS) handelt es sich um eine auf XML basierende Sammlung von Technologien, die es ermöglicht, Suchergebnisse von Suchmaschinen und Websites in einem standardisierten und maschinenlesbaren Format abzufragen und auszugeben. OpenSearch wur-

de durch Amazon.com bzw. A9.com entwickelt. Die OpenSearch-Spezifikation ist durch A9 unter der Lizenz Creative Commons Attribution-ShareAlike 2.5 lizenziert worden /3/.

OpenSearch besteht grundsätzlich aus drei Elementen:

- einer Beschreibungsdatei
- einem Ausgabeformat für Suchergebnisse (OpenSearch-RSS)
- speziellen Feedreadern bzw. Programmen, die OpenSearch-RSS interpretieren können (die Rolle eines solchen Readers übernimmt in diesem Fall das OpenSearch-iPlug von PortalU)

Bei der OpenSearch-Beschreibungsdatei (OpenSearch-Description) handelt es sich um eine XML-Datei, die den Zugriff auf eine Suchmaschine oder Website beschreibt. OpenSearch-RSS ist ein RSS-Dialekt, in dem Suchergebnisse bereitgestellt werden (Abbildung 2). OpenSearch-RSS kann z.B. von OpenSearch-Feedreadern interpretiert bzw. dargestellt werden.

```
<?xml version="1.0" encoding="utf-8" ?>
<feed xmlns="http://www.w3.org/2005/Atom" xmlns:os="http://a9.com/-/spec/opensearch/1.1/">
  <title>Suchergebnisse: dioxin</title>
  <link rel="alternate" type="text/html" href="http://193.197.158.227:80/search?access=p&entqr=3&output=xml&sort=8&client=BW_OpenSearch_xml&q=dioxin&ud=1&site=BW_UP&oe=UTF-8&proxystylesheet=BW_OpenSearch_xml" />
  <link rel="search" href="search?access=p&output=xml&ie=UTF-8&client=BW_OpenSearch_xmlBW_OpenSearch_xml&site=BW_UP&proxystylesheet=BW_OpenSearch_xmlBW_Op3E&markup=opensearch&alt_markup=atom" type="application/opensearchdescription+xml" title="OpenSearch" />
  <os:totalResults>635</os:totalResults>
  <os:startIndex>1</os:startIndex>
  <os:itemsPerPage>10</os:itemsPerPage>
  <os:Query role="request" searchTerms="dioxin" />
  <subtitle>Results 1 - 10 of about 635</subtitle>
  <id>tag:193.197.158.227,1998-09-07:dioxin</id>
  <author>
    <name>Google Search Appliance</name>
  </author>
  <updated>1998-09-07T09:00:00Z</updated>
  <generator uri="http://www.google.com/enterprise/">Google Search Appliance</generator>
  <link rel="self" type="application/atom+xml" href="http://193.197.158.227:80/search?access=p&entqr=3&output=xml&client=BW_OpenSearch_xml&q=dioxin&alt_markup=atom&ud=1&site=BW_UP&oe=UTF-8&proxystylesheet=BW
  <entry xmlns:relevance="http://a9.com/-/opensearch/extensions/relevance/1.0/">
    <title>Ministerium für Ländlichen Raum, Ernährung und ...</title>
    <link rel="alternate" type="text/html" href="http://www.mlr.baden-wuerttemberg.de/Dioxin_in_irischem_Schweinefl
    <id>tag:193.197.158.227,1998-09-07:dioxin/Iby7gXPX-38J</id>
    <updated>1998-09-07T08:00:00Z</updated>
    <summary>... Dioxin in irischem Schweinefleisch. ... Weitere Informationen, Links und Downloads. Mehr zu »Dioxin in
    ...</summary>
    <relevance:score>10</relevance:score>
    <provider xmlns="">Ministerium für Ernährung und Ländlichen Raum Baden-Württemberg</provider>
    <partner xmlns="">bw</partner>
  </entry>
```

Abbildung 2: Ausschnitt aus einem OpenSearch-XML-Dokument

## 2.2 Erweiterung des OpenSearch-Standards

Um die Integration von Antworten (Trefferlisten) zu einer Ergebnisliste in PortalU zu ermöglichen, mussten die OpenSearch-Antworten der GSA (Google Search Appliance) über den Standard hinaus um weitere Informationen ergänzt werden:

- Ein Ranking-Attribut, das die Relevanz der Treffer innerhalb der Trefferliste reflektiert. Dessen Inhalt wird durch das iPlug in den Wertebereich des internen InGrid-Rankings transformiert. Für das Ranking-Attribut wird eine offizielle OpenSearch-Erweiterung /4/ verwendet.

- Der Name des Informationsanbieters wird zusammen mit dem Suchtreffer ausgeliefert. Da diese Information im Suchindex der GSA nicht explizit vorliegt, wird sie auf Basis der URL und der in LUPO hinterlegten Anbieter-Informationen dynamisch per XSL-Transformation eingefügt. Das notwendige XSLT-Fragment wird dabei aus der Datenbank der Landesumweltportale erzeugt.
- In entsprechender Art und Weise wird auch der „Partner“ in die Suchergebnisse eingefügt. Dabei handelt es sich in der Regel um das Bundesland, aus dem die Informationen stammen.

Abbildung 3 zeigt die Volltextsuche in PortalU mit über OpenSearch eingebetteten Treffern aus der Google Search Appliance (GSA), der Volltextsuchmaschine der LUPO-Portale.

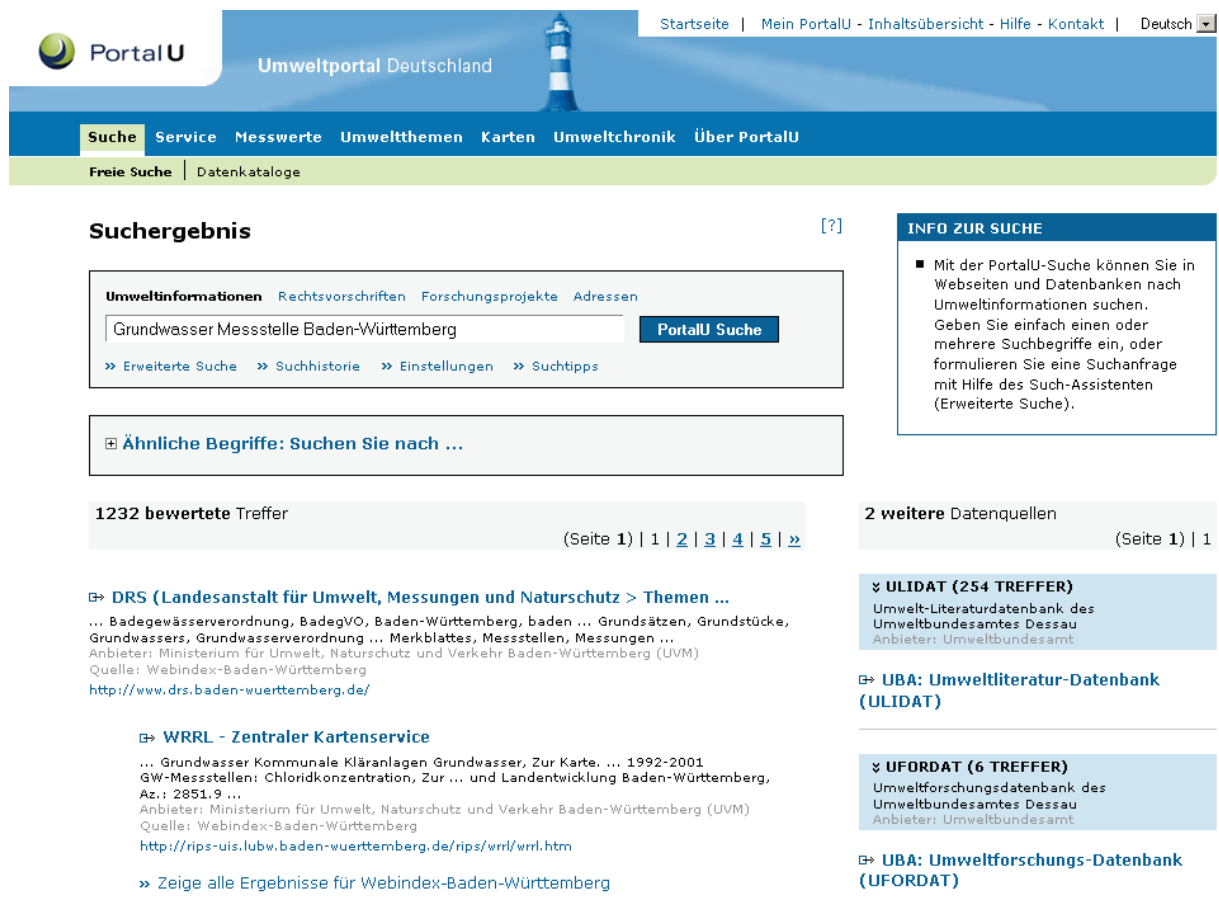


Abbildung 3: Darstellung von GSA-Suchergebnissen im PortalU



Abbildung 4: Neben vielen anderen Suchmaschinen lässt sich auch die Volltextsuche der Landesumweltportale mit einem Klick in die Firefox-Suchmaschinenleiste eintragen

Durch die Verwendung der OpenSearch-Beschreibung der Suchschnittstelle ist auch eine Einbindung der LUPO-Suche in einen Webbrowser möglich (Abbildung 4). Die Beschreibung ist in Form einer XML-Datei auf dem Server hinterlegt. Sie kann im Suchbereich des Webbrowsers mit einem Klick des Nutzers per JavaScript-Schnittstelle installiert werden. Als Trefferseiten werden hier die normalen Ansichten aus den LUPO-Portalen verwendet.

### 3. Teilangebote von Websites

Häufig können aus Sicht der Landesumweltportale<sup>1</sup> Teilbereiche von Webservern als eigenständige Informationsangebote aufgefasst werden. Für solche Angebote liegt der Wunsch nahe, diese auch separat per Volltextsuche durchsuchen zu können bzw. über die Volltextsuche der Landesumweltportale eine eigene Suchfunktion für diese Angebote zu realisieren. Viele Web-Content-Management-Systeme (WCMS) verwenden jedoch zur Adressierung ihrer Inhalte Nummern (Primärschlüssel-Attribute), so dass sich die Struktur einer Website nicht wie bei statischen Webservern über die URL ausdrückt. Statt mit „sprechenden“ Adressen wie

```
http://servername/bereich/thema/unterthema/seite.html
```

wird auf viele Server mit Adressen der Art

```
http://servername/4711/
```

zugegriffen, wobei „4711“ für einen solchen Primärschlüssel steht. Hierdurch ist eine Unterscheidung von Teilbereichen einer Website anhand der URL alleine nicht möglich. Damit ist auch die Konfiguration einer Suchmaschine, die nur einen bestimmten Teil der Website erfassen soll, leider unmöglich.

Zur Lösung dieses Problems gibt es zwei mögliche Ansätze, die jedoch beide Eingriffe auf Seiten des WCMS notwendig machen:

1. Erweiterung des WCMS um die Möglichkeit „sprechende URLs“ zu produzieren
2. Entwicklung einer Möglichkeit, der Suchmaschine mitzuteilen, welche Seiten (URLs) zu einem spezifischen Teilangebot gehören

Für das in Baden-Württemberg und Rheinland-Pfalz häufig eingesetzte WCMS WebGenesis (auf dem auch die LUPO-Portale selbst basieren) wurden beide Strategien umgesetzt.

#### 3.1 WebGenesis mit „sprechenden URLs“

Die Inhaltsobjekte von WebGenesis werden über URLs adressiert. Diese beinhalten in den meisten Fällen das Primärschlüsselattribut der Inhaltsobjekte, die sog. WebGenesis-ID. Viele WebGenesis-URLs haben die Form<sup>2</sup>:

```
http(s)://<Server-Adresse>/servlet/is/<WebGenesis-ID>/?<Parameterliste>
```

---

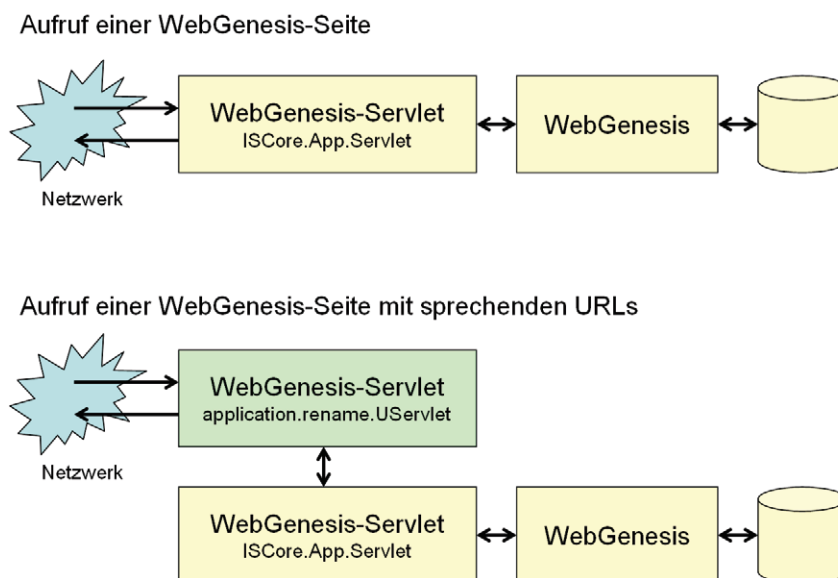
<sup>1</sup> Natürlich auch aus Sicht der Betreiber bzw. Informationsanbieter

<sup>2</sup> Es gibt weitere Formate. Der volle Umfang ist im Handbuch zur Erweiterung „Sprechende URLs“ beschrieben.

Die implementierte Lösung für „sprechende URLs“ besteht darin, dem WebGenesis-Servlet ein weiteres Servlet vorzuschalten, das

- einen separaten Adressbereich für „sprechende URLs“ verwaltet,
- Adressen der Form `/servlet/is/<WebGenesis-ID>/?<Parameterliste>` in sprechende URLs umwandelt und dem Client (Browser, Suchmaschine) über einen HTTP-Status-Code 301 (moved permanently) mitteilt, statt dieser physikalischen URL künftig eine sprechende URL zu verwenden<sup>3</sup>,
- sprechende URLs in physikalische Adressen umwandelt und dem WebGenesis-Servlet transparent mitteilt,
- nicht unterstützte (physikalische) Adressen transparent dem WebGenesis-Servlet weiterleitet,
- alle Antworten des WebGenesis-Servlets transparent an den Client weitergibt,
- die Implementierung individueller Abbildungen (Mappings) zulässt, z.B. für komplexe Fachsysteme wie FADO /5/ oder den Themenpark Umwelt /6/.

Die „Architektur“ dieses Mechanismus sieht (vereinfacht) wie in Abbildung 5 gezeigt aus:



**Abbildung 5: Funktionsweise der WebGenesis-Erweiterung „Sprechende URLs“**

Die Abbildung der physikalischen WebGenesis-URLs auf „sprechende URLs“ und zurück muss eindeutig sein und drückt auf Seiten der sprechenden URLs die Hierarchie der Daten aus. Der gewählte Ansatz implementiert dabei eine Art Plug-In-Konzept: Für jeden Server lässt sich eine solche Abbildungsvorschrift individuell implementieren. Die Basis-Implementierung nutzt die WebGenesis-Verzeichnis-Hierarchie.

Die Erweiterung für „sprechende URLs“ wurde einige Monate auf den Servern mehrerer LU-PO-Portale erfolgreich getestet. Für die Inhalte anderer Server, z.B. der LUBW Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg und des Ministeriums für Umwelt, Naturschutz und Verkehr Baden-Württemberg (UVM), hätte eine Umstellung auf

<sup>3</sup> Suchmaschinen, insb. die GSA nutzen dieses „moved permanently“ tatsächlich im Sinne von „verwende stattdessen“ und indizieren nur die Seite, auf die durch dieses Kommando verwiesen wird



diese Technologie jedoch einen erheblichen Aufwand bei der Migration der Inhalte und der Templates bedeutet, da dort die Adressierung – im Gegensatz zu den absoluten URLs der LUPO-Portale – sehr häufig über relative URLs erfolgt, die in den getrennten Adressräumen von physikalischen und logischen/sprechenden URLs nicht aufgelöst werden können. Die Umsetzung der Teilbaumsuche über „sprechende URLs“ wurde für diese Server deshalb nicht weiter verfolgt. Das Problem soll nun auf eine andere Art gelöst werden, die im folgenden Abschnitt beschrieben wird.

## 3.2 Generierung von Sitemaps

Die Lösung der Teilbaumsuche durch die Generierung von „Sitemaps“ wurde ebenfalls als WebGenesis-Dienst realisiert. Sie besteht darin, dass für einen beliebigen WebGenesis-Eintrag eine Liste der „unterhalb“ (im hierarchischen Sinne) dieses Eintrags befindlichen Inhalte erzeugt werden kann. Diese Liste entspricht einem echten Teilangebot des WebGenesis-Servers, in das (konfigurierbar) auch Layout- und Inhalts-Dateien einbezogen werden können. Aus dieser zunächst abstrakten Inhaltsliste lassen sich nun Sitemaps in verschiedenen Formaten generieren:

- XML-Sitemaps nach dem von Google, Yahoo! und Microsoft verabschiedeten De-facto-Standard<sup>4</sup>
- RSS-Feeds (Versionen 0.91 und 2.0)<sup>5</sup>
- GSA-Collection-Konfigurationen, die mit Hilfe der GSA-Admin-API<sup>6</sup> zeitgesteuert an eine Google Search Appliance übertragen werden können

Mit diesem Dienst ist es nun möglich, Teilbereiche von WebGenesis-Servern explizit als Sammlungen/Collections in der GSA anzulegen und automatisch zu aktualisieren. Darüber hinaus kann per XML-Sitemaps auf die Indizierung der WebGenesis-Server durch Internet-Suchmaschinen direkt Einfluss genommen werden.

Die Sitemap-Erweiterung für WebGenesis ist derzeit (Stand Anfang Juni 2010) in Erprobung.

## 4. Gliederung der Anbieterliste

Durch den inhaltlichen Ausbau der Landesumweltportale wurde die Anbieterliste in diesen Systemen zunehmend unübersichtlich. Daher wurde auf Autoren-Seite die Möglichkeit zur Gliederung der Informationsanbieter innerhalb von Verzeichnissen geschaffen.

Diese Gliederung drückt sich auf Nutzerseite durch eine Gruppierung von gleichartigen Informationsanbietern, z.B. „EU“, „Ministerien“ oder „Landkreise“, unter einer entsprechenden Überschrift aus (Abbildung 6). Damit wurde die Auffindbarkeit bestimmter Informationsanbieter und deren Informationsangebote für den Nutzer deutlich erleichtert und vereinfacht.

---

<sup>4</sup> <http://googleblog.blogspot.com/2005/06/webmaster-friendly.html>

<sup>5</sup> <http://www.rssboard.org/rss-specification>

<sup>6</sup> Die GSA unterstützt im Gegensatz zur Google-Internet-Suche derzeit keine Sitemap-Formate.



Abbildung 6: Gegliederte Anbieterliste (Ausschnitt aus LUPO Sachsen-Anhalt)

## 5. Geoportal Wasser in Rheinland-Pfalz

Seit Anfang des Jahres 2010 ist auch das Land Rheinland-Pfalz mit seinem Geoportal Wasser<sup>7</sup> am Projekt LUPO beteiligt. Im Gegensatz zu den anderen LUPO-Portalen kommt dort jedoch nicht die komplette LUPO-Software, sondern nur die GSA-basierte Volltextsuche sowie die Schnittstelle zur Einbindung der Volltextsuche in WebGenesis zum Einsatz (Abbildung 7). Die übrigen Komponenten des Geoportals Wasser bestehen aus einer eigenständigen Implementierung auf Basis von WebGenesis.

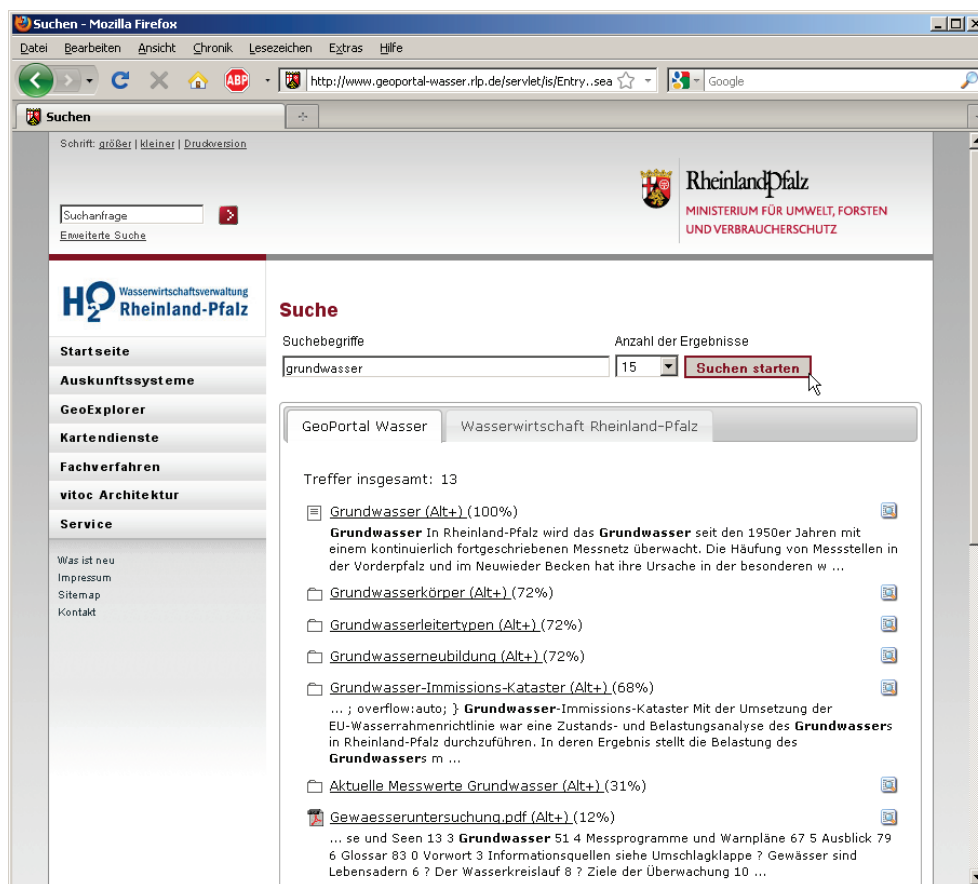


Abbildung 7: GSA-basierte Volltextsuche im Geoportal Wasser (Rheinland-Pfalz)

<sup>7</sup> <http://www.geoportal-wasser.rlp.de>

Innerhalb kürzester Zeit konnte das Geoportal Wasser mit seiner leistungsfähigen Volltextsuche in den produktiven Betrieb gehen. Die Suchoberfläche bietet dabei mit mehreren Registerkarten (Tabs) ein innovatives Konzept für die Darstellung der Suche in verschiedenen Teilindexen.

## 6. Erfahrungen beim Betrieb

Seit Anfang 2009 wird für die Server des UVM und der LUBW die neue Webstatistik-Software WiredMinds<sup>8</sup> eingesetzt. Auch das Umweltportal Baden-Württemberg wurde in diese Webstatistik integriert, während die Zugriffe auf die anderen Landesumweltportale nach wie vor mit der in LUPO integrierten Webalizer-Software<sup>9</sup> ausgewertet werden können. Bei WiredMinds kommen, im Gegensatz zum Webalizer, keine Tools zur Analyse von Log-Dateien der Webserver zum Einsatz, sondern es werden beim Seitenaufruf über JavaScript-Aufrufe bzw. sogenannte Zählpixel Zählimpulse auf dem Statistik-Server ausgelöst, welche anschließend in einer Datenbank des WiredMinds-Servers abgelegt werden. Statistische Auswertungen können auf Basis dieser Daten jederzeit in Echtzeit abgerufen werden.

Zur Analyse der bei der Volltextsuche verwendeten Suchbegriffe steht auf der Google Search Appliance zusätzlich zur Webstatistik eine Möglichkeit zur Auswertung solcher Queries zur Verfügung. Ein kleines, auf Basis der GSA-API implementiertes Tool kann seit April 2010 diese Auswertungen für die verschiedenen Server und Teilindexe automatisiert generieren. Dabei interessieren insbesondere auch solche Suchbegriffe, die dem Nutzer keine Treffer lieferten (sog. Zero-Results), z.B. weil dieser Begriffe der Umgangssprache statt Fachbegriffe verwendet hatte. Gegebenenfalls kann hier über weitere Mechanismen (Thesaurus, ähnliche Suchbegriffe, Keymatches /1/) die Volltextsuche für künftige Suchen optimiert werden.

Von Dezember 2009 bis April 2010 wurden die Server aller LUPO-Systeme auf einen mit der LUBW und dem Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (Fraunhofer IOSB) abgestimmten Software-Stand (WebGenesis 7.13 inkl. benötigter Basis-Software) gebracht. Dabei wurde auch die interne Struktur der Systeme angepasst, um alle WebGenesis-basierten Systeme in gleicher Art und Weise administrieren zu können. Inklusive der Entwicklungsmaschinen werden derzeit von der LUBW und dem KIT/IAI acht LUPO-Server auf VMware-Clustern betrieben. Im Zuge der Konsolidierung der LUPO-Server konnten auch die bisher verteilten und auf verschiedener technischer Basis implementierten OneBox-Dienste reimplementiert und auf die LUPO-Produktionsserver übernommen werden. Für die Entwicklung weiterer OneBoxen steht nun ein kleines Java-Framework zur Verfügung.

Wegen des Auslaufens des Lizenz-Supports zum Jahresende 2009 wurde Ersatz für die bisherige Google Search Appliance beschafft. Die neue Maschine basiert auf einer noch leistungsfähigeren Hardware (Modell 7007) und bietet mit ihrem Softwarestand (Version 6.2) verbesserte Möglichkeiten bei der Suche (u.a. automatische Rechtschreibkorrektur) und beim Betrieb. So ist nun die Replikation des Suchindex auf eine Backup-Maschine in Echt-

---

<sup>8</sup> <http://www.wiredminds.de>

<sup>9</sup> <http://www.webalizer.org>

zeit möglich. Bisher musste auch die Backup-Maschine ihren Index selbst aktuell halten, was zu einer doppelten Belastung bei den indizierten Systemen sorgte.

## 7. Ausblick

Die seit Version 6.0 der Google Search Appliance zur Verfügung stehende Programmierschnittstelle soll zur Vereinfachung der Administration der Umweltportale genutzt werden. Hier soll eine Verbindung zwischen der Administration innerhalb von WebGenesis und der Administration der Volltextsuche auf der GSA geschaffen werden. Derzeit erfolgt eine teilweise redundante Pflege von Metadaten in beiden Systemen.

In der näheren Zukunft steht der weitere inhaltliche Ausbau der Umweltportale auf der Agenda. Schwerpunkt sind dabei insbesondere solche Daten, die nicht über den Volltext-Crawler erschlossen werden können, z.B. weil sie in Datenbanken verborgen sind. Außerdem sollen weitere aktuelle Messwerte in Form von OneBoxen (s. /1/) angeboten werden, z.B. aktuelle Luftmessdaten.

Im Zusammenhang mit dem Projekt zur semantischen Suche nach Umweltinformationen (SUI) ist die Integration verschiedenartiger Datenformate wie tabellarischer Daten, Geodaten/Karten und Multimedia-Daten in die Trefferlisten der Umweltportale geplant /7/. Dort wird unter anderem eine einheitliche Beschreibung von Zielsystemen entwickelt, die eine weitgehend automatisierte Anbindung verschiedener Informationssysteme an die Umweltportale ermöglichen soll.

## 8. Literatur

- /1/ Schlachter, T. et al. (2009): LUPO – Fortgeschrittene Suchfunktionen in den Landesumweltportalen von Baden-Württemberg, Sachsen-Anhalt und Thüringen In: Mayer-Föll, R., Keitel, A., Geiger, W.; Hrsg.: Kooperative Entwicklung wirtschaftlicher Anwendungen für Umwelt, Verkehr und benachbarte Bereiche in neuen Verwaltungsstrukturen, Phase IV 2008/09, Forschungszentrum Karlsruhe, Wissenschaftliche Berichte, FZKA 7500, S. 149-156.
- /2/ OpenSearch in der deutschen Wikipedia: [http://de.wikipedia.org/wiki/Open\\_search](http://de.wikipedia.org/wiki/Open_search)
- /3/ Creative Commons License: <http://creativecommons.org/>
- /4/ OpenSearch Relevance Extension  
<http://www.opensearch.org/Specifications/OpenSearch/Extensions/Relevance/1.0>
- /5/ Weidemann, R. et al. (2010): FADO – Funktionale Konsolidierung des Fachdokumentenmanagements im Umweltinformationssystem Baden-Württemberg und Erschließung neuer Themenbereiche. In diesem Bericht.
- /6/ Grießmann, B. et al. (2010): Themenpark Umwelt – Konzept für die Erweiterung um den Themenkomplex Klima. In diesem Bericht.
- /7/ Bügel, U. et al. (2010): SUI II – Weiterentwicklung der diensteorientierten Infrastruktur des Umweltinformationssystems Baden-Württemberg für die semantische Suche nach Umweltinformationen. In diesem Bericht.