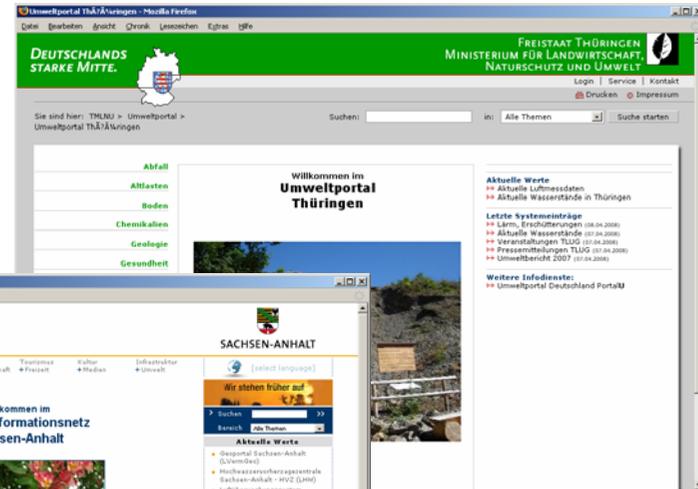
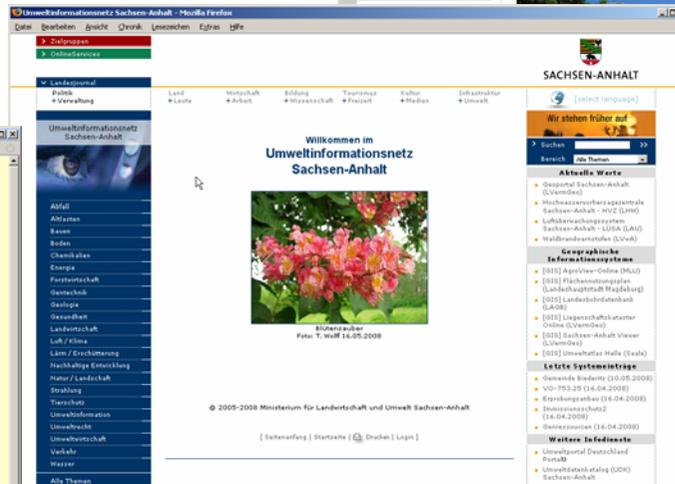


# Erschließen von Datenbank-Inhalten durch die Volltextsuche in Landes-Umweltportalen



# Übersicht

- Landes-Umwelt-Portale
- Suchmaschinen für die Volltextsuche
- Indizierung von Datenbank-Inhalten
- Fazit und Ausblick



Thüringen

Sachsen-Anhalt

Baden-Württemberg

- Umweltinformationsgesetze verpflichten Behörden, Umweltinformationen zugänglich zu machen
  - Verschiedene Formen (Fachdokumente, Mess- und Sachdaten, Geoinformation, Multimedia, ...)
  - Verteilt auf viele unterschiedliche Informationssysteme
- Ziele der Landesumweltportale
  - Singulärer Einstiegspunkt
  - „Leichter Zugang“ (Zielgruppe: Öffentlichkeit)
  - Verschiedene Zugangswege
  - Umfassender Zugang zu behördlichen, auch kommunalen Informationsangeboten

# Bisherige Landes-Umweltportale

- Volltextsuche
- Aktuelles
- Themen
- Geo-Information
- Neueste Einträge
- Anbieter

**Umweltinformationsnetz Sachsen-Anhalt**

**Themen (red box):** Abfall, Altlasten, Bauen, Boden, Chemikalien, Energie, Forstwirtschaft, Gentechnik, Geologie, Gesundheit, Landwirtschaft, Luft / Klima, Lärm / Erschütterung, Nachhaltige Entwicklung, Natur / Landschaft, Strahlung, Tierschutz, Umweltinformation, Umweltrecht, Umweltwirtschaft, Verkehr, Wasser.

**Suche (blue and green boxes):** Suchen [input], Bereich: Altlasten

**Aktuelle Werte (orange box):**

- Geoportal Sachsen-Anhalt (LVermGeo)
- Hochwasservorhersagezentrale Sachsen-Anhalt - HVZ (LHW)
- Luftüberwachungssystem Sachsen-Anhalt - LÜSA (LAU)
- Waldbrandwarnstufen (LVwA)

**Geographische Informationssysteme (purple box):**

- [GIS] AgroView-Online (MLU)
- [GIS] Flächennutzungsplan (Landeshauptstadt Magdeburg)
- [GIS] Landesbohrdatenbank (LAGB)
- [GIS] Liegenschaftskataster Online (LVermGeo)
- [GIS] Sachsen-Anhalt Viewer (LVermGeo)
- [GIS] Umwelatlas Halle (Saale)

**Anbieterliste (yellow box):**

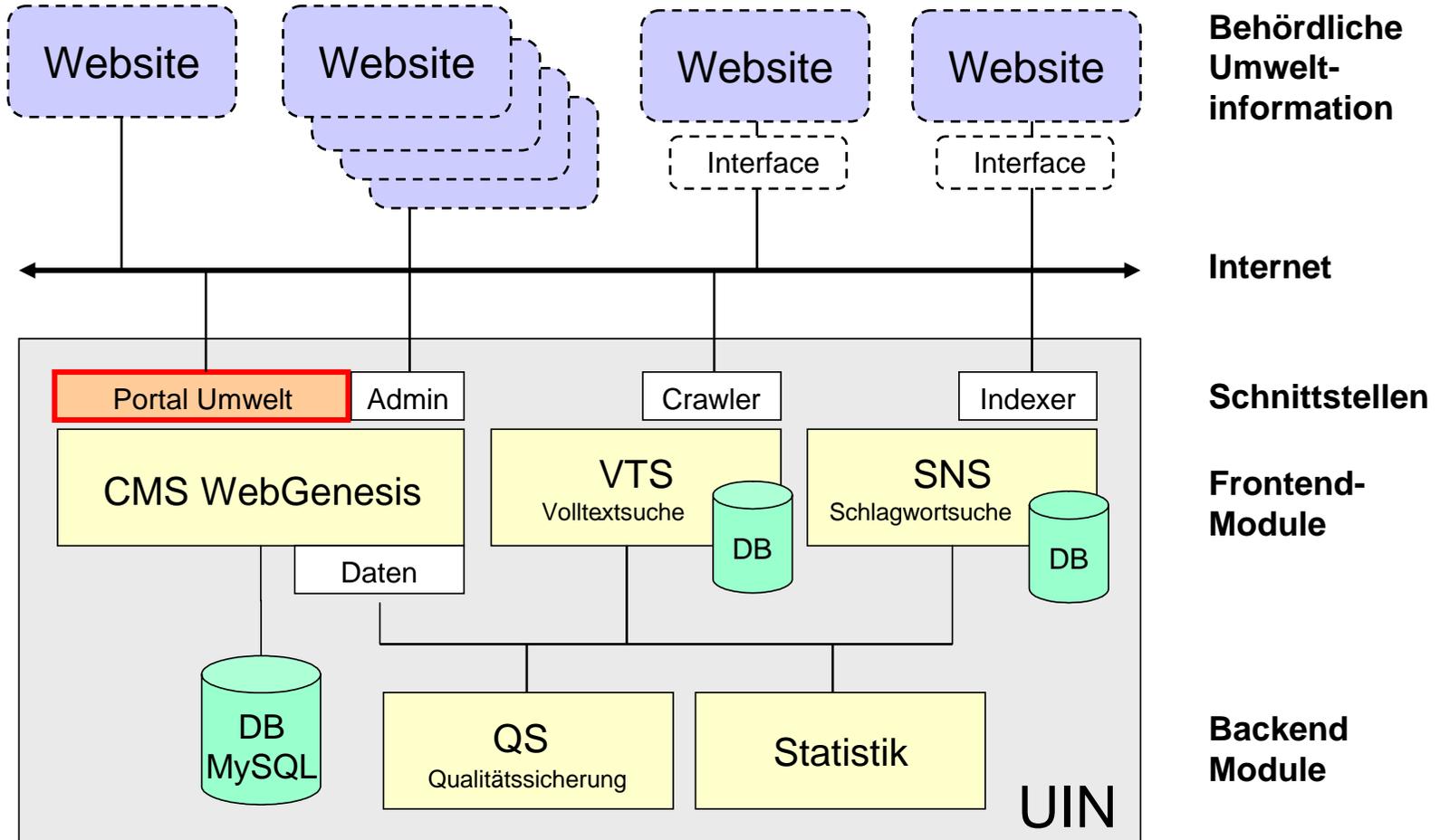
- Gemeinde Biederitz (10.05.2008)
- VO-753.25 (16.04.2008)
- Erprobungsanbau (16.04.2008)
- Immissionschutz2 (16.04.2008)
- Genressourcen (16.04.2008)

**Artikelübersicht:**

- Altlasten**
  - [\[Altlasten\] - Altlasten in Sachsen-Anhalt](#)
  - [\[Altlasten\] - Altlastensituation im Land Sachsen-Anhalt](#)
  - [\[Altlasten\] - Informationen zur Altlastenfreistellung und -sanierung im Land Sachsen-Anhalt](#)
  - [\[Altlasten\] - Schwermetalle und spezielle Elementparameter](#)
  - [\[Altlasten\] - Was sind Altlasten?](#)
  - [\[Boden & Altlasten\] - Umweltamt Lankreis Harz](#)
  - [\[Boden\] - Organische Spurenanalyse in Sachsen-Anhalt](#)
  - [\[Bodenschutz\] - Aufgaben Bodenschutz/Altlasten](#)
  - [\[Bodenschutz\] - Quellenangaben und Download-Texte zum Thema Bodenschutz/Altlasten](#)

Umweltinformationsnetz  
Sachsen-Anhalt

# Architektur der Landes-Umweltportale



- Austausch durch funktional gleichwertige Suchmaschine
  - Kein wesentlicher Mehrwert zu erwarten
  - Lucene/Nutch, dtSearch
- Einsatz einer kommerziellen Suchmaschine
  - Leistungsfähigere Kategorie (Performanz, DBs, Feeds)
  - Google Search Appliance, Oracle Secure Enterprise Search
- Einsatz der InGrid-Software
  - Problemlose Integration mit PortalU
  - Viele weitere Schnittstellenkomponenten
  - Mehrfachnutzung von Daten
  - Produktreife?

# Entscheidung für Google Search Appliance

- Hard-/Softwarelösung
  - Einfaches Lizenzmodell (URLs)
  - DB-Indizierung, Feeds
  - Produktreife, Nutzerakzeptanz
  - Skalierbarkeit
- 
- Mankos:
    - Fehlende API
    - Nicht komplett mandantenfähig
    - Keine Schnittstelle zu InGrid/iBus

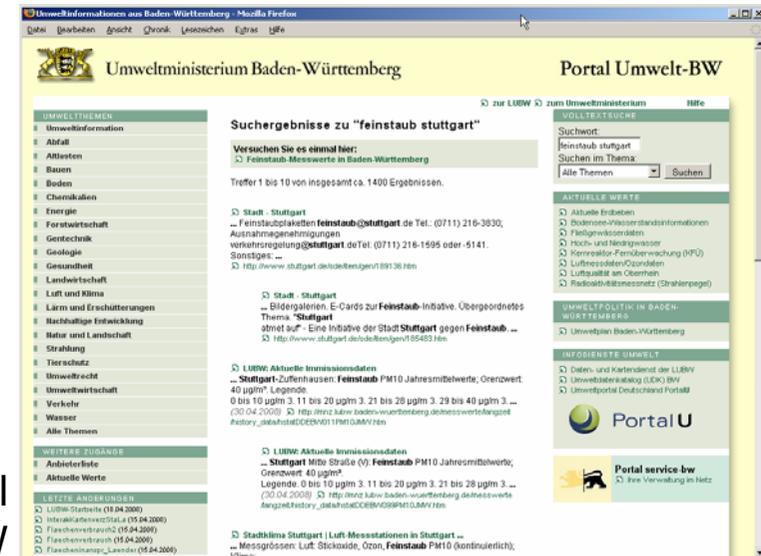


Google Search Appliance; © picasa.google.de

# Umsetzung der Volltextsuche mit der GSA

- Evaluierung seit Januar 2008
- Aufbau Volltextindex innerhalb weniger Wochen
- Collections bieten Teilindexe
- Integration Umweltthesaurus (UBA)
- Abfrage-Schnittstelle für CMS WebGenesis
- Im produktiven Einsatz:
  - LUBW, UM-BW (April 2008)
  - Portal Umwelt-BW (Mai 2008)
  - UIN Sachsen-Anhalt (in Kürze)
  - $\Sigma = \sim 1.000$  Websites
  - $\Sigma = \sim 400.000$  Dokumente

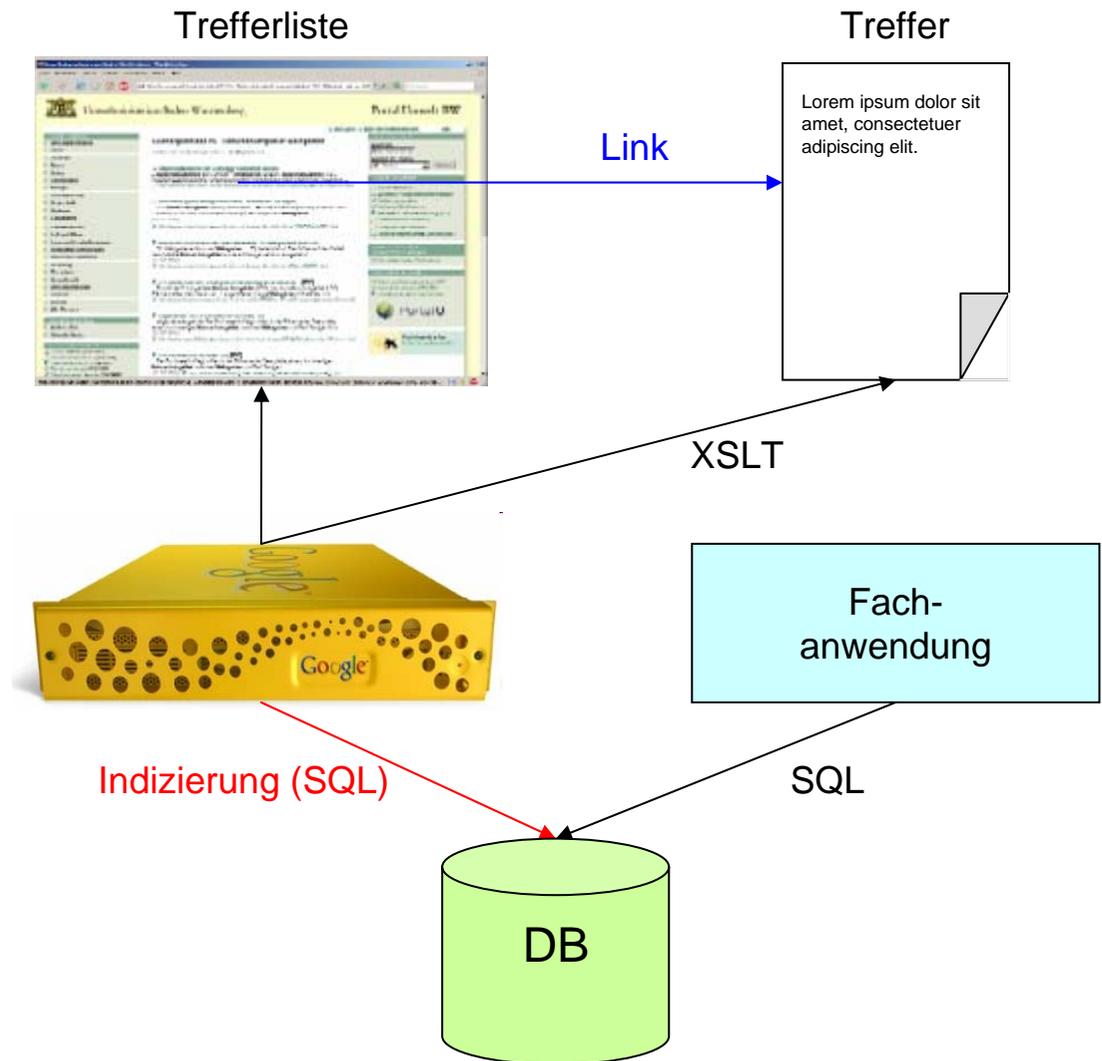
Portal  
Umwelt-BW



1. die ausschließlich über eine **Datenbankschnittstelle** (z.B. per SQL-Abfrage) verfügbar sind (keine Web-Schnittstelle)
2. die als **Basis von Webanwendungen** dienen, welche die Daten ausschließlich über einen **Abfragemechanismus** (z.B. Formularauswahl) zur Verfügung stellen
3. die als **Basis von Webanwendungen** dienen, welche über Links in entsprechenden **Menüstrukturen** alle Daten zugänglich machen (✓ → Crawler)

# Indizierung von Datenbank-Inhalten

- Indizierung unabhängig von der Fachanwendung
- Keine Webanwendung notwendig
- Darstellung der Inhalte auf Stellvertreterseite per XSL-Transformation



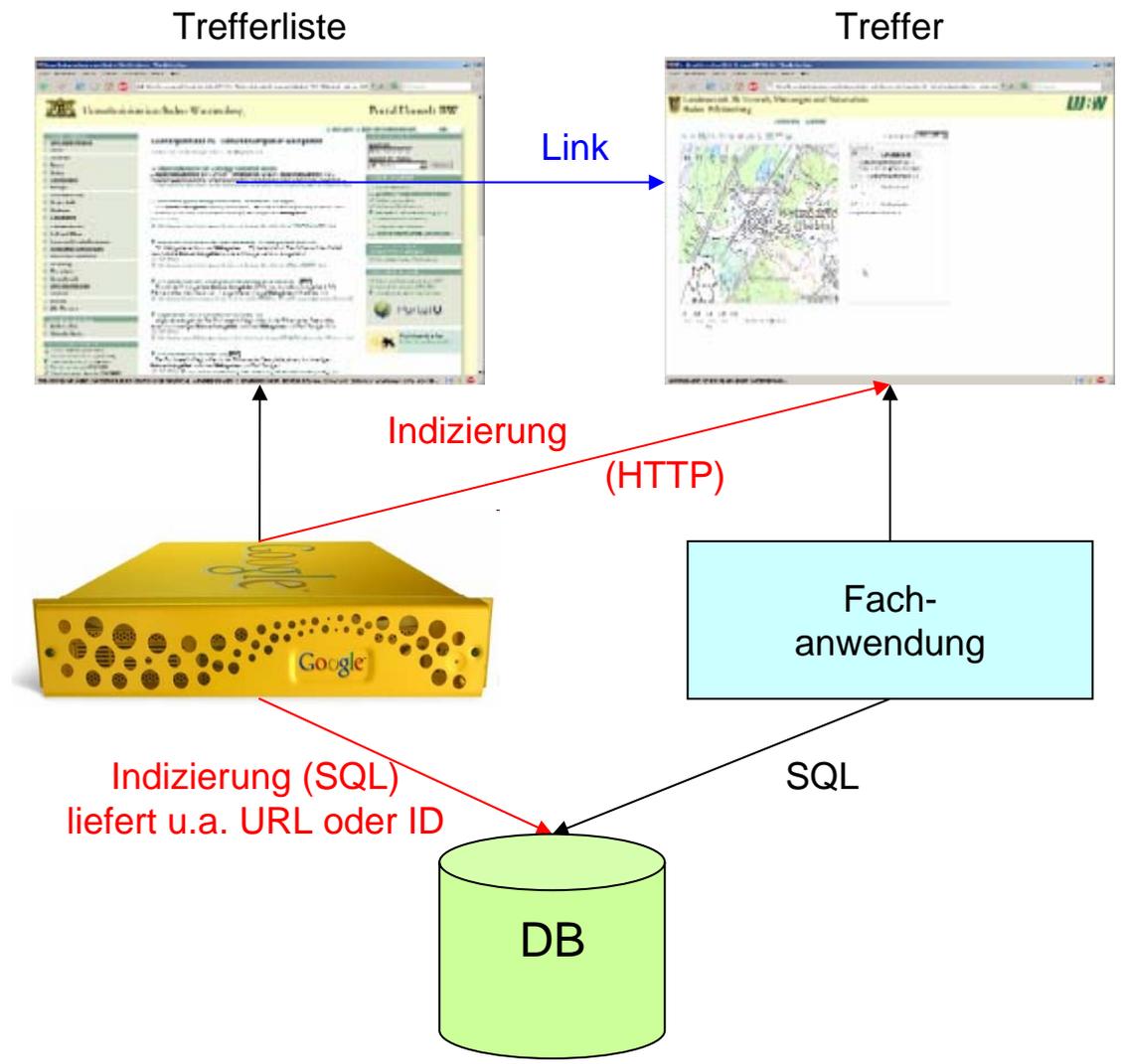
# Konfiguration einer DB-Abfrage in der GSA

Datenbanktyp:	Oracle	} DB-Verbindungsdaten
Hostname:	10.40.1.2	
Port:	1521	
Datenbankname:	uis	
Nutzername:	udk_lesen	
Passwort:	<small>(Gespeichertes Passwort wird nicht angezeigt.)</small>	
Crawling-Anfrage:	<pre>select obj_id, obj_name, obj_descr, 'http://www2.lubw.baden-wuerttemberg.de/wwwudk /UDKDetailServlet?Type=Data&amp;PK='  obj_id as URL from udks.t01_object where publish_id = 1</pre>	} Abfrage
<input checked="" type="radio"/> Standard-Stylesheet		} Auswahl eines XSLT-Stylesheets zur Darstellung der Ergebnisse
<input type="radio"/> Angepasstes Stylesheet		
Stylesheet hochladen:	<input type="text"/>	Durchsuchen...

**Ergebnis:** 1763 rows crawled. 0 rows failed.  
com.google.enterprise.database.TableCrawler crawl  
INFO: total time = 84755

- Problem: Seiten nicht für Web-Crawler erreichbar
- Lösungsmöglichkeiten
  - Jump-Page
    - + einfach realisierbar; auch für Internet-Suchmaschinen; inhaltliche Kontrolle
    - Implementierung in Webanwendung
  - XML-Sitemap
    - + einfach realisierbar; volle inhaltliche Kontrolle
    - nicht für alle Suchmaschinen geeignet, Implementierung in Webanwendung
  - Parallele Indizierung von DB und Webanwendung
    - Generierung von URLs in DB/Abfrage
    - Generierung von URLs in Suchmaschine
    - + einfach realisierbar; volle inhaltliche Kontrolle; unabhängig von Webanwendung; Indizierung zusätzlicher Inhalte möglich
    - nicht für Internet-Suchmaschinen geeignet

- Indizierung der Datenbank und der Webanwendung
- Treffer verweisen auf Seiten der Webanwendung
- Es werden auch Seiten gefunden, die nicht per Crawler erreichbar sind



# Auffinden von Kartendarstellungen

- **Beispiel:**  
Umweltdaten und -karten online (UDO)  
auf Basis von disy cadenza

Trefferliste im Portal Umwelt-BW

## Suchergebnisse zu "naturschutzgebiet weingarten"

Treffer 1 bis 10 von insgesamt ca. 295 Ergebnissen.

### [Naturschutzgebiete der Gemeinde Weingarten \(Baden\)](#)

... **Naturschutzgebiete** der Gemeinde **Weingarten** (Baden), **Naturschutzgebiet**, NSG, Naturschutzgebietsname: Ungeheuerklamm, Weingartener Moor-Bruchwald Grötzingen, ...

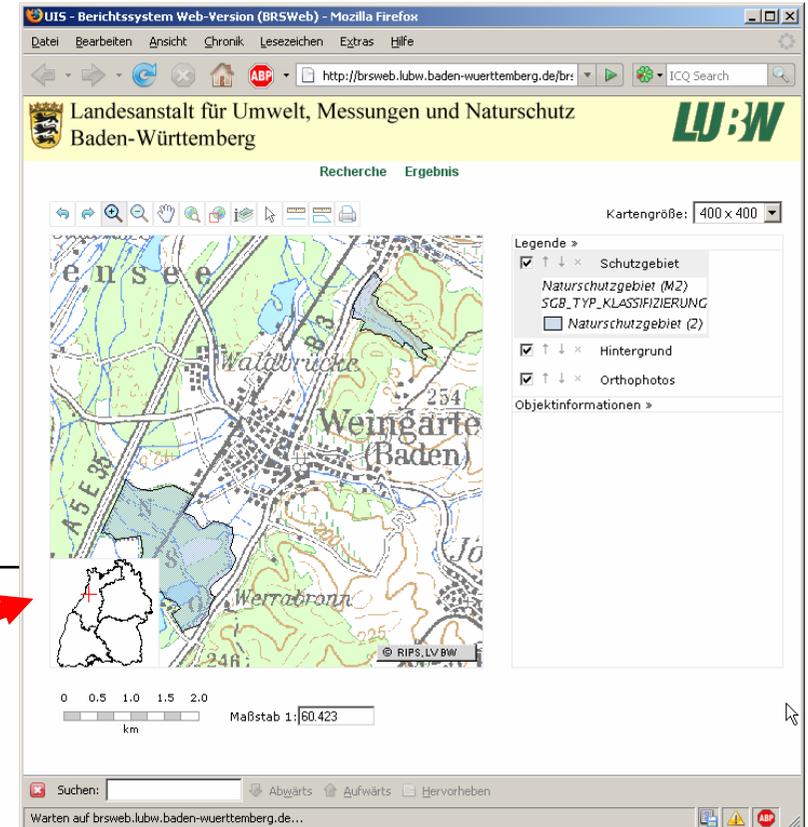
[http://brsweb.lubw.baden-wuerttemberg.de/brs-web/show\\_cweb?repositoryId=.Schutzgebiete.iukbnl](http://brsweb.lubw.baden-wuerttemberg.de/brs-web/show_cweb?repositoryId=.Schutzgebiete.iukbnl)

### [Naturschutzgebiet Weingartener Moor - Bruchwald Grötzingen ...](#)

... Das **Naturschutzgebiet** Weingartener Moor – Bruchwald Grötzingen liegt im Stadt- und Landkreis Karlsruhe auf den Gemarkungen Grötzingen und **Weingarten**. ...

(04.10.2004)

[http://www.xfaweb.baden-wuerttemberg.de/nafaweb/berichte/pfsw\\_02/NSGWein0001.html](http://www.xfaweb.baden-wuerttemberg.de/nafaweb/berichte/pfsw_02/NSGWein0001.html)



Treffer in UDO

- GSA löst Volltextsuchmaschine ht://Dig ab
- Erstmals Indizierung von DB-Inhalten innerhalb der Landes-Umweltportale
- Vereinfachung (bzw. Möglichkeit) der Indizierung von Web-Anwendungen
  - insb. Auffinden von nicht-textuellen Darstellungen wie Multimedia oder GIS
- Synergie durch Nutzung von Inhalten für Realisierung verschiedener Suchen (Suchräume = Collections)
- Keine Einbindung der GSA in die InGrid-Software

- Vervollständigung der Indizierung von UDO
  - „generischer Selektor“ zur Einbindung aller gewünschten Inhalte
- Anschluss weiterer Datenbanken und Webanwendungen
- Automatische Erweiterung der Suchanfrage (Thesauri, Semantic Network Service SNS, UIS-Ontologie, ...)
- Einbindung von Suchergebnissen in „OneBoxes“, d.h. parallele Trefferlisten, z.B. mit Luftmessdaten bei Eingabe eines Ortsnamens
- GSA-iPlug für InGrid?

- Vielen Dank für Ihre Aufmerksamkeit!
  
- „Umwelt-Google“ ausprobieren:
  - <http://www.umwelt.baden-wuerttemberg.de>
  - <http://www.umwelt.sachsen-anhalt.de>
  - <http://www.um.baden-wuerttemberg.de>
  - <http://www.lubw.baden-wuerttemberg.de>