

Flexible und modulare Systemarchitektur für InGrid 1.0 und Portal-U

Thomas Vögele

Fred Kruse

Martin Klenke

Koordinierungsstelle UDK/GEIN

im Niedersächsischen Umweltministerium

Archivstrasse 2, 30169 Hannover

kug@numis.niedersachsen.de

Einleitung

Der Umweltdatenkatalog UDK und das Umweltinformationsnetz Deutschland *gein*[®] durchlaufen derzeit eine grundlegende Veränderung. Aus zwei getrennten und technisch völlig unterschiedlich aufgebauten Softwaretools entsteht ein neues, integriertes System. Die neue Software (InGrid 1.0) vereint die Vorteile beider Einzelsysteme, also effizientes Metadatenmanagement (UDK) mit komfortablen Zugriffsmethoden (*gein*[®]). Mit InGrid 1.0 wird primär ein neues Portal für das vom Bund und den Ländern gemeinsam betriebene Umweltinformationsnetz Deutschland aufgebaut. Dieses neue Portal – Portal-U – ersetzt das bisherige *gein*[®] und wird von der Koordinierungsstelle UDK/GEIN (KUG) betrieben. Die Software InGrid 1.0 ist jedoch so konzipiert, dass damit im Bedarfsfall auch andere Umweltportale, z.B. auf Länder- und kommunaler Ebene aufgesetzt werden können.

Die technische Feinkonzeption des Systems wurde von der KUG in Zusammenarbeit mit den Firmen GISTec GmbH (www.gistec-online.de), wemove (www.wemove.com), MediaStyle (www.media-style.com), und dem Fraunhofer Institut Graphische Datenverarbeitung (www.igd.fhg.de) erarbeitet. Die Umsetzung der Konzeption erfolgt bis Ende 2005. Ein erster Prototyp wird Anfang 2006 ans Netz gehen.

InGrid 1.0 setzt auf eine modulare Softwarearchitektur auf Basis lizenzfreier OpenSource Software auf, die das System auf neue Anforderungen und

Einsatzgebiete vorbereiten soll. So basiert das derzeitige Umweltinformationsnetz Deutschland auf einer zentral verwalteten Instanz der *gein*[®] Software, die als Informationsbroker das behördliche Informationsangebot beim Bund und den Ländern zugänglich macht. Auch der Umweltdatenkatalog nutzt zur Verknüpfung der dezentral beim Bund und den Ländern betriebenen UDK Kataloge eine zentrale Broker-Instanz, den V-UDK. Die Aufgaben dieser beiden Systeme werden in Zukunft von Portal-U als dem von der KUG zentral betriebenen Umweltportal des Bundes und der Länder übernommen. Zusätzlich wird es mit InGrid 1.0 jedoch möglich sein, bei Bedarf weitere Informationsknoten auf verschiedenen Ebenen der Verwaltungshierarchie zu etablieren. Damit können z.B. landesbezogene Informationsportale aufgebaut oder aber zusätzliche Informationsanbieter (z.B. auf Grund der Einbindung der Kommunen in das Informationsnetz) verwaltet werden.

In Portal-U wird mit Hilfe einfach zu konfigurierende Datenschnittstellen die Einbindung von Fachinformationssystemen und Datenbanken, die über reguläre Suchmaschinen nicht zugänglich sind (das sog. „invisible web“), erleichtert. Außerdem kann Portal-U über standardisierte Austauschschnittstellen Daten und Informationen mit anderen Datenkatalogen und Dateninfrastrukturen austauschen. Geplant ist z.B. eine Verknüpfung mit nationalen und internationalen Geodateninfrastrukturen (GDI-DE bzw. INSPIRE). Damit kann die Bereitstellung von Umweltinformationen, wie sie in der geplanten INSPIRE Richtlinie gefordert wird, sichergestellt werden.

Das vorliegende Papier gibt einen Überblick über die wichtigsten Komponenten der für InGrid 1.0 bzw. Portal-U vorgesehenen Systemarchitektur.

1 Architektur

Die Systemarchitektur von InGrid 1.0 sieht insgesamt 4 Hauptmodule vor: das Modul „Portal“, das Modul „WMS“, das Modul „Suchmaschine“ und das Modul „iBus“ (Abbildung 1). Im Folgenden werden die wichtigsten Merkmale dieser Module kurz skizziert.

1.1 Modul „Portal“

Zur Realisierung der Portaloberfläche setzt InGrid 1.0 die OpenSource Portal-Engine „Jetspeed“ ([ASF, 2005]) ein. Jetspeed ermöglicht die flexible Gestaltung und

Verwaltung von Internetportalen. Spezifische Portalfunktionen werden über einzeln zu verwendende „Portlets“ implementiert. Dies sind kleine (in JAVA implementierte) Applikationen, die auf der Basis einer gängigen Servlet-Engine (Tomcat) serverseitig aufgesetzt werden. Je nach gewünschter Funktionalität des Gesamtportals können einzelne Portlets aktiviert oder deaktiviert werden. Für Portale, die mit InGrid 1.0 betrieben werden (z.B. Portal-U) bedeutet das, dass zum einen die standardmäßig angebotenen Portalfunktionen vom Systemadministrator kontrolliert werden können und zum anderen jeder Nutzer über eine Personalisierung des Portals einzelne Funktionen ausblenden kann. Insgesamt kann die Funktionalität eines Portals beliebig und flexibel erweitert werden indem in dem einzelne Portlets neu erstellt bzw. direkt von Drittanbietern hinzugekauft werden.

1.2 Modul „WMS“

Die OpenSource Produkte UMN Mapserver ([UMN, 2005]) und MapBender ([CCGIS, 2005]) werden zur Umsetzung der Geoportalfunktionen in InGrid 1.0 eingesetzt. Der WMS bzw. der MapBender Map-Viewer erfüllen dabei zwei Aufgaben: Zum einen bilden sie die Grundlage für die kartenbasierte Suche in Portal-U. Durch Aufziehen eines Such-Rechtecks oder die Auswahl administrativer Einheiten per Mausklick kann dort der räumliche Aspekt einer detaillierten Suchanfrage spezifiziert werden. Zum anderen werden WMS und Map-Viewer zur Visualisierung digitaler Karten genutzt, sofern diese ebenfalls als WMS angeboten und über die Suchfunktionen des Portals bzw. alternative Zugangswege (z.B. Themenseiten) zugänglich sind. Dabei können Attributdaten abgefragt und mehrere Karten aus verschiedenen Datenquellen gemeinsam dargestellt werden.

1.3 Modul „Suchmaschine“

Die effiziente Indexierung und Suche in Webseiten und Datenbankinhalten ist eine der Schlüsselfunktionalitäten des Portal-U bzw. InGrid 1.0. Entsprechend werden hierfür moderne, leistungsfähige und gut skalierbare Softwaretools eingesetzt. Als Indexierer dient das OpenSource Produkt Lucene ([AJP,2005]), während das Crawling und Ranking der Informationsbestände mit Hilfe von Software aus dem (ebenfalls OpenSource) Projekt nutch ([AI,2005]) und Eigenentwicklungen durchgeführt wird. Für ein integriertes Ranking und Darstellung der an Portal-U angeschlossenen Datenquellen ist es notwendig, dass alle Datenquellen über

Lucene indiziert werden. Um diesem Ziel möglichst nahe zu kommen wird ein sog. Data-Source Client bereitgestellt werden. Über den Data-Source Client kann auf einfache Weise ein individueller Volltextindex für jede externe Datenquelle erstellt werden.

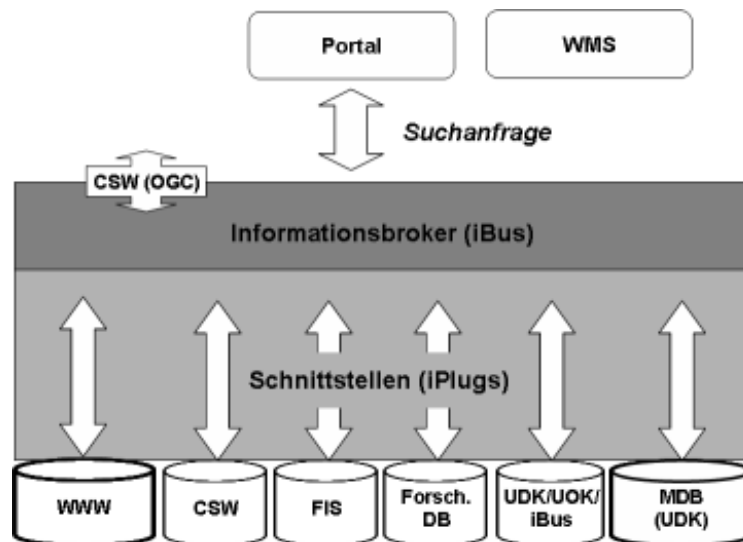


Abbildung 1: Systemarchitektur InGrid 1.0

1.4 Modul „iBus“

Der iBus ist das Kernmodul eines InGrid 1.0 Portals. Hier werden Anfragen aus den Modulen „Portal“ und „WMS“, so wie aus externen Datenkatalogen, Suchmaschinen oder Portalen, angenommen und an die verschiedenen an das Portal angeschlossenen Datenquellen weitergeleitet. Der iBus bedient sich dabei einer Reihe von Schnittstellen und Schnittstellenadaptern. Als Anfrageschnittstelle werden neben einer InGrid 1.0 spezifischen Portalschnittstelle eine OGC kompatible CSW 2.0 Schnittstelle und eine UDK Schnittstelle zur Verfügung gestellt. Über die Portalschnittstelle wird die Kommunikation des Moduls „iBus“ mit dem Modul „Portal“ abgewickelt. Die CSW Schnittstelle implementiert das derzeit dem Open Geospatial Consortium (OGC) als Recommendation Paper vorliegenden *Application Profile for CSW 2.0 Catalog Services*. Über diese Schnittstelle können alle OGC CSW-2.0 kompatiblen Katalogdienste bzw. Informationsbroker (wie z.B. das GeoMIS.Bund) auf diejenigen Metadaten in Portal-U zugreifen, die in den Standardformaten ISO 19115 (Geodaten) und ISO 19119 (Dienste) abgelegt sind. Das sind z.B. alle in der Metadatenkomponente des Portal-U (also dem UDK) beschriebenen Geodaten und

Dienste. Die technische Umsetzung des iBus erfolgt u.a. mit den OpenSource Werkzeugen Java Management Extensions (JMX) ([SUN,2005]) und Hibernate ([SF, 2005]).

Vom iBus aus wird der Zugang zu verschiedenen Datenquellen durch eine flexible Plug-In Architektur realisiert: Jede Schnittstelle wird als individueller Schnittstellenadapter, oder „iPlug“, implementiert. Derzeit sind iPlugs für die interne (UDK)Metadatenbank, für externe UDK-kompatible Datenkataloge bzw. den bayrischen Umweltobjektkatalog (UOK), für CSW-kompatible Geodatenkataloge, für Fachinformationssysteme, für die Semantic Network Services (SNS) des Umweltbundesamtes (zur semantischen Erweiterung der Suchanfragen), und für reguläre Web-Sites geplant. Je nach Bedarf können aber weitere Schnittstellenadapter entwickelt und in den iBus integriert werden.

1.5 Datasource Client

Als Grundbaustein eines iPlugs dient der sog. Datasource Client (DSC). Er übersetzt aus dem spezifischen Datenformat einer Datenquelle (z.B. JDBC im Falle eines FIS, oder HTML im Falle von Webseiten) in das einheitliche und performante Datenformat einer Indexdatenbank. Die Datenquellen werden zunächst über einen konfigurierbaren Quellen-Adapter an den DSC angeschlossen (Abbildung 1). Ein auf der Open-Source Suchmaschine Lucene ([AJP,2005]) basierender Indexer erzeugt aus den so aufbereiteten Daten eine hochperformante, mit einem Ranking versehene Indexdatenbank. Das iPlugInterface schließlich macht diesen Index für InGrid Portale bzw. iBus-Instanzen zugänglich, wobei ggf. ein und derselbe Datasource Client von mehreren Portalen gemeinsam genutzt werden kann. Um die Administration der InGrid Datenquellen zu vereinfachen und eine performante Kommunikation des iBus mit den Datenquellen sicherzustellen wird auf Basis des OpenSource Produktes JXTA ([SUN,2004]) eine Peer-to-Peer Kommunikationsinfrastruktur implementiert. Dabei wird jeder Datasource Client bzw. jede über ein iPlug erschlossene Datenquelle bei einem zentralen Repository Server angemeldet. Über die auf dem Repository Server gespeicherten Verbindungsdaten kann die Datenquelle dann an jeden beliebigen iBus angeschlossen werden.

Zusammenfassend erhält also jede an ein InGrid Portal angeschlossene Datenquelle über den Datasource Client einen eigenen Lucene-Index. Unter „Datenquellen“

werden hier vorwiegend Datenbanken und Fachinformationssysteme verstanden. Die Gesamtheit der über Portal-U direkt eingebundenen Webseiten entspricht einer einzigen Datenquelle, d.h. es wird ein zentraler Index für alle „Start-URLs“ der Informationsanbieter erstellt. Jeder Datasource Client kann die Suchanfragen des iBus hochperformant abarbeiten und eine gerankte Trefferliste zurückliefern. Da die Rankingparameter aller Datasource Clients miteinander kompatibel sind kann im Portal eine einheitliche und gerankte Gesamt-Ergebnisliste mit Treffern aus verschiedenartigen Datenquellen erzeugt werden.

Praktisch stellt ein Datasource Client ein kleines, plattformunabhängig implementiertes Softwaremodul dar, das entweder direkt beim Informationsanbieter bzw. der jeweiligen Datenquelle, oder aber zentral auf dem Rechner des InGrid Portals installiert werden kann. Da die Software nur Teile der OpenSource Suchmaschine Lucene und InGrid-Eigenentwicklungen beinhaltet, kann sie im Rahmen der VwV UDK/GEIN lizenzfrei an beteiligte Informationsanbieter weitergegeben werden. Jeder Datasource Client verfügt über eine Administrationsoberfläche die es gestattet, das Modul vor Ort (d.h. durch den Informationsanbieter) oder von einem anderen Rechner aus (d.h. durch den Portal-Administrator) zu konfigurieren und zu warten.

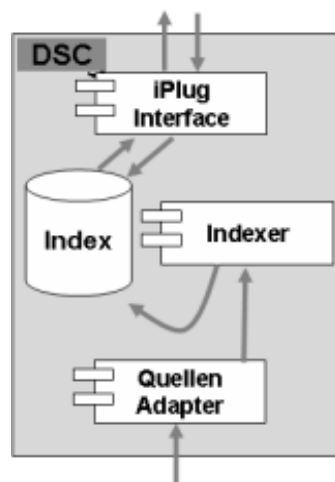


Abbildung 2: Schematischer Aufbau eines Datasource Clients

Webseiten, die wie in gein auch in Portal-U eine der wichtigsten Datenquellen darstellen, werden über ein spezielles iPlug eingebunden (siehe oben). Dieses verknüpft einen Datasource Client mit einem effizienten Web-Crawler, der die Internetsites der angeschlossenen Informationsanbieter durchläuft und deren Inhalt

in einem geeigneten Format zurückliefert. Der Web-Crawler wird über Software aus dem (ebenfalls OpenSource) Projekt nutch ([AI,2005]) umgesetzt. Über die automatische Verschlagwortung mit SNS erhält jede Webseite u.a. einen Raumbezug und steht damit den räumlichen Suchfunktionen des InGrid Portals zur Verfügung.

2 Zusammenfassung

Die Software InGrid 1.0 wird zur Umsetzung von Portal-U eingesetzt, das als integriertes Werkzeug zur Metadatenhaltung und Portal für behördliche Umweltinformationen das German Environmental Information Network *gein*[®] und langfristig auch den Umweltdatenkatalog UDK ablösen soll. Diese Ablösung soll schrittweise und unter Einbeziehung bereits vorhandener Systeme und Datenbestände erfolgen. Technisch wird mit InGrid konsequent eine modulare und flexible Systemarchitektur auf der Basis von OpenSource Technologie umgesetzt. Insbesondere die Module „Portal“ und „iBus“ sind zusätzlich in sich für eine flexible Skalierung und Anpassung an neue bzw. sich ändernde Anforderungen ausgelegt. Damit wird es möglich bei Bedarf, neben dem zentralen Portal-U, weitere Informationsknoten auf verschiedenen Ebenen der Verwaltungshierarchie anzulegen und so ein System von vernetzten Umweltportalen zu schaffen. Diese können durch Zu- oder Abschalten von Modulen an den jeweils gewünschten Funktionsumfang angepasst werden. Durch den ausschließlichen Einsatz von OpenSource Produkten und Eigenentwicklungen ist sichergestellt, dass beim Betrieb von InGrid Informationsknoten keine Kosten für Softwarelizenzierungen entstehen. Da InGrid im Rahmen einer Bund/Länder Verwaltungsvereinbarung entwickelt wurde kann das Produkt von den Partnern der Verwaltungsvereinbarung und deren nachgeordneten Behörden kostenfrei genutzt werden. Mit InGrid und Portal-U wurden somit die technischen und organisatorischen Voraussetzungen zum weiteren Ausbau der im Bereich der deutschen Umweltverwaltung bereits bestehenden und bewährten informationstechnischen Infrastruktur geschaffen. Damit soll insbesondere die Umsetzung einschlägiger EU Richtlinien, wie der EU Richtlinie 2003/4/EG (Umweltinformationsrichtlinie) und der INSPIRE Richtlinie unterstützt werden.

3 Literatur

[ASF, 2005]

ASF: Jetspeed, an Overview, URL: <http://portals.apache.org/jetspeed-1/>, Apache Software Foundation, 2005.

[AI, 2005]

AI: Welcome to Nutch, <http://incubator.apache.org/nutch/>, Apache Incubator, 2005.

[AJP, 2005]

AJP: Lucene, an Overview. <http://lucene.apache.org/java/docs/index.html>, Apache Jakarta Project, 2005.

[CCGIS, 2005]

CCGIS: Mapbender – WebGIS mit freier Software, <http://www.mapbender.org/>, Consulting Center für geografische Informationssysteme, 2005

[UMN, 2005]

UMN: UMN Mapserver Version 4.4.1, <http://mapserver.gis.umn.edu/>, University of Minnesota, USA, 2005.

[SF, 2005]

SF: Hibernate - Relational Persistence for Idiomatic Java, <http://www.hibernate.org/>, SourceForge.net, 2005

[SUN, 2004]

SUN: JXTA Technologies – Creating Connected Communities, <http://myjxta2.jxta.org/project/www/docs/JXTA-Exec-Brief.pdf>, SUN Microsystems, January 2004.

[SUN, 2005]

SUN: Java Management Extensions (JMX), <http://java.sun.com/products/JavaManagement/>, SUN Microsystems, 2005.