

InGrid[®] 2.0

Die Technologie des Umweltportal Deutschland PortalU[®]

Version 1.1 | 10.11.2008

Inhalt

1	Einleitung	3
2	Technischer Aufbau von InGrid®	4
2.1	<i>Der iBus</i>	4
2.2	<i>Die iPlugs</i>	5
2.3	<i>Offene Schnittstellen</i>	6
2.4	<i>Portaloberfläche</i>	6
2.5	<i>InGrid® Catalog</i>	7
3	Verwendete Open Source Software für InGrid®	9
4	Einsatzmöglichkeiten von InGrid®	10
4.1	<i>Einsatz im Umweltbereich in Deutschland</i>	10
4.2	<i>Einsatz als Geoportal</i>	10
4.3	<i>Einsatz als allgemeines Verwaltungsportal</i>	10
4.4	<i>Einsatz als Umweltportal in anderen europäischen Staaten</i>	11
4.5	<i>Einsatz als europäisches Umweltportal</i>	11

1 Einleitung

Die Technologie des Umweltportal Deutschland PortalU[®] (www.PortalU.de), InGrid[®], basiert auf einer modular aufgebauten Software zur zentralen Bereitstellung von Informationen aus dem Internet sowie aus Datenbanksystemen. Der Name InGrid steht hierbei für *Information Grid* – Informationsnetz. In einem InGrid[®]-Portal können sowohl Webseiten als auch Daten und Metadaten eingebunden werden (Abb. 1). InGrid[®] wurde für den Betrieb des Umweltportal Deutschland PortalU[®] entwickelt und ist vielseitig einsetzbar.

Wichtigste Komponenten von InGrid[®]

- Nutzerfreundliche und flexibel konfigurierbare Portaloberfläche
- Metadatenkatalog InGrid[®]Catalog
- Karten-Visualisierungskomponente
- Diverse An- und Abfrageschnittstellen zur Recherche in angeschlossenen Systemen bzw. für den Transfer von Informationen

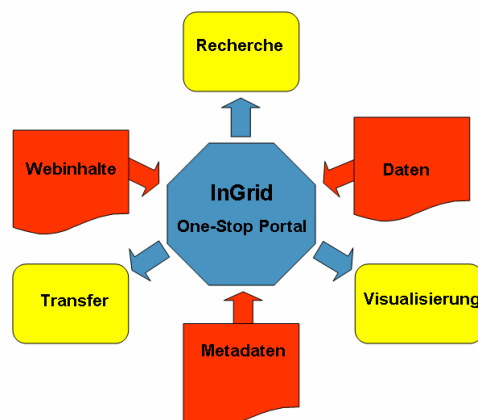


Abbildung 1: Schematische Darstellung eines InGrid[®]-Portals

Funktionen von InGrid[®]

- Portal- und Recherchefunktionen wie Volltextsuche, fach- und geothesausbasierte Suche, Anzeige aktueller Nachrichten, Messwerten und Themenseiten
- Visualisierung digitaler Karten über integrierten Kartenklienten
- Bereitstellung von Karten über integrierten Kartenserver
- Räumliche Suche nach Karten, Diensten und anderen Informationsobjekten
- Bestimmung des Raumbezugs von nicht georeferenzierten Inhalten über semantische Analyse der im Dokument benannten Ortsnamen
- ISO 19115 / 19119-konforme Beschreibung und Verwaltung von Geodaten und Geo-Diensten im integrierten Metadatenkatalog InGrid[®]Catalog
- Einbindung von externen Geodatenkatalogen über OGC-konforme Catalog Service Web Schnittstellen (CSW-iPlug)
- Transfer von InGrid[®]-Metadaten in Geodateninfrastrukturen wie die GDI-DE oder INSPIRE über CSW-Schnittstelle

2 Technischer Aufbau von InGrid®

InGrid® basiert auf dem iPlug-Konzept, das im Bereich der Suchmaschinen bereits seit längerem erfolgreich eingesetzt wird. Bei diesem Konzept werden Arbeitsaufgaben in Teilaufgaben unterteilt und parallel bearbeitet. Dadurch ergibt sich ein erweiterbares, flexibles und damit auch skalierbares System, mit dem eine hohe Rankingqualität und eine hohe Geschwindigkeit gewährleistet werden kann. InGrid® besteht aus unterschiedlichen Modulen (vgl. Abb. 2). Die wichtigsten sind der Informationsbroker (iBus), die Schnittstellen zu den Datenquellen (iPlugs), Schnittstellen zur Portaloberfläche und offene Schnittstellen zur Weitergabe von Informationen an andere Systeme, die Portaloberfläche sowie der InGrid® Catalog.

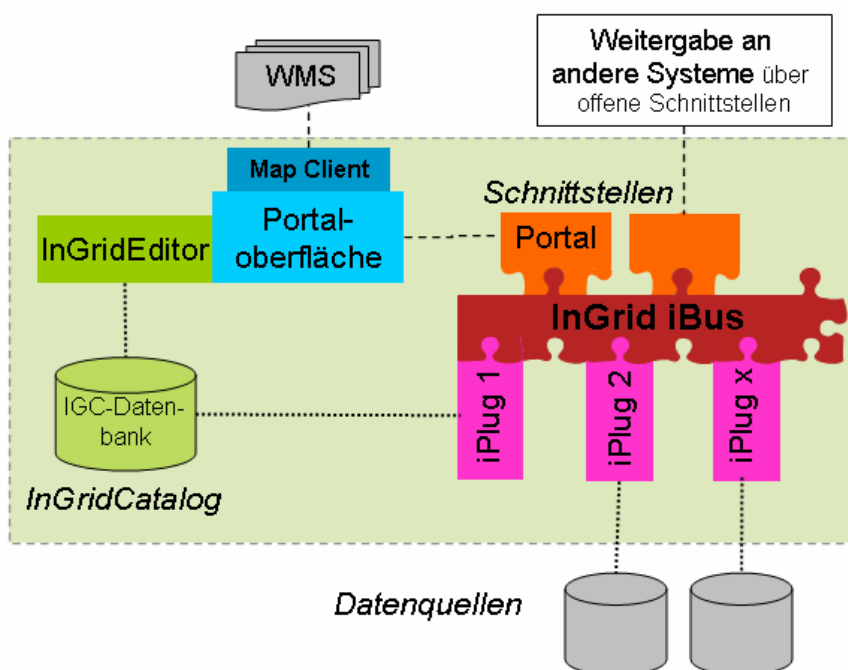


Abbildung 2: Modularer Aufbau von InGrid®

2.1 Der iBus

Der iBus (information bus) bildet in einem InGrid®-Portal das zentrale Element. Er fungiert als Verteilungsstation zwischen Datenquellen und Suchanfragen. So nimmt der iBus eine Suchanfrage von der Portaloberfläche oder einer anderen übergeordneten Schnittstelle entgegen, bereitet die Anfrage auf und gibt sie über iPlugs an die angeschlossenen Datenquellen weiter. Die Anfrageergebnisse der einzelnen Datenquellen werden vom iBus eingesammelt und als Ergebnislisten an das anfragende System zurückgegeben.

2.2 Die iPlugs

Ein iPlug (information plug) ist eine Schnittstelle, die das Gesamtsystem mit einzelnen Datenquellen verbindet. Die iPlugs werden hierbei in den iBus eingehängt. Das System ist flexibel erweiterbar, d.h. neue iPlugs können nachträglich entwickelt und angeschlossen werden, ohne dass das bestehende System geändert werden muss. Die iPlugs können entweder, soweit dies die Gegebenheiten der Infrastruktur zulassen, dezentral direkt beim Datenanbieter oder zentral am iBus eingerichtete werden. Somit kann alternativ zu einer weitgehend zentralen Installation, ein performanteres voll verteiltes System erzeugt werden. Jede Datenquelle erhält ein eigenes iPlug, so dass die Datenhalter volle Kontrolle über die von Ihnen für ein InGrid[®]-Portal bereitgestellten Informationen behalten. Die wichtigsten iPlugs werden im Folgenden vorgestellt:

DSC-iPlug: Über das Datasource-Client-iPlug können Datenbanken mit beliebiger Datenstruktur angeschlossen werden. Als Datenbanktypen werden MySQL, Oracle, MS-SQL-Server, PostgreSQL und Tamino unterstützt. Mit dem Datasource Client wird u.a. ein Index erstellt, der für die Sortierung der Ergebnislisten (suchanfragenspezifisches Ranking) und für ein performantes Antwortverhalten erforderlich ist. Das DSC-iPlug wird in der Regel beim Datenhalter installiert und verbindet sich mit dem iBus über tcp-sockets.

SE-iPlug: Die Suchmaschine (Search-Engine-iPlug) besteht aus einer Such- und einer Indexierungskomponente. Die Suchkomponente hat die Aufgabe zu einer Suchanfrage die zugehörigen Treffer zu liefern. Dazu wird der von der Indexierungskomponente erstellte Index analysiert. Teil der Indexierungskomponente ist außerdem ein leistungsfähiger Web-Crawler. Die indexierten Dokumente werden über den Semantic Network Service (SNS) des Umweltbundesamts zusätzlich mit Metadaten für Thema, Raum und Zeit angereichert. Die Pflege der zu indexierenden Inhalte kann durch Informationsanbieter rollenbasiert dezentral über das Internet vorgenommen werden. Hierfür ist das SE-iPlug mit einer Benutzeroberfläche ausgestattet, über die der Volltextindex sowie Themen-, Messwert- und Serviceseiten gepflegt werden können.

SNS-iPlug: Das SNS-iPlug ist eine spezielle Schnittstelle zum Semantic Network Service (www.semantic-network.de) des Umweltbundesamtes. Über das iPlug werden die für das Portal notwendigen SNS-Informationen für die Umweltchronik, die räumliche und die semantische Suche abgefragt.

CSW-iPlug: Das CSW-iPlug stellt eine OGC-konforme CSW-2.0-Schnittstelle (DE-Profil, zukünftig CSW AP ISO 1.0) zur Verfügung. Über das CSW-iPlug können insbesondere geographische Metadatensysteme an ein InGrid[®]-Portal angeschlossen werden.

IGC-iPlug: Über das InGrid[®]Catalog-iPlug werden InGrid[®]-Kataloge eingebunden. Das iPlug stellt die im IGC gepflegten Metadaten zur Verfügung.

MSE-iPlug: Der Meta-Search-Engine-iPlug ist eine spezielle Schnittstelle zur Abfrage anderer Suchmaschinen. Der MSE-iPlug wertet hierbei wie eine Metasuchmaschine html-Seiten aus und gibt die Informationen an den iBus weiter. In PortalU[®] wird dieses iPlug zur Einbindung von forschungportal.net verwendet.

Für Datenquellen, die über das DSC-, SE- oder IGC-iPlug angeschlossen sind, werden Indexdokumente erstellt, um die Suchergebnisse mit suchanfragebezogener Gewichtung in der Ergebnisliste zu sortieren. Alle sortierten Ergebnisse werden in der Hauptergebnisliste auf der Portaloberfläche angezeigt. Die Ergebnisse, die aus aufgrund von Schnittstelleneigenschaften nicht zu indexierenden Datenquellen stammen, erscheinen in der Nebenergebnisliste.

2.3 Offene Schnittstellen

Für die Kommunikation eines InGrid[®]-Portals mit externen Informationssystemen stehen Schnittstellen zur Verfügung, die auf offenen Standards beruhen und HTTP- sowie XML/SOAP-Technologien für den Datenaustausch nutzen. Die wichtigsten Schnittstellen sind hierbei die OpenSearch-Schnittstelle und die CSW-Schnittstelle.

OpenSearch-Schnittstelle: Über diese XML-basierte Schnittstelle kann ein InGrid[®]-Portal über einfache http-GET-Requests abgefragt werden. Die Suchergebnisse werden in einem standardisierten XML-Format ausgegeben. Es können sowohl Trefferlisten als auch Detailergebnisse recherchiert werden. OpenSearch wurde durch die Firma Amazon.com im Rahmen der Suchmaschinenentwicklung A9 definiert und wird mittlerweile in verschiedensten Projekten verwendet. PortalU[®] wird z.B. über diese Schnittstelle von der deutschen Meta-Suchmaschine MetaGer (<http://www.metager.de/>) abgefragt. Die Schnittstelle kann zudem für parametrisierte Suchanfragen an ein InGrid[®]-Portal genutzt werden. Die Ergebnisse einer solchen Anfrage können z.B. in einem eigenen Portal mit individuellem Layout dargestellt werden.

CSW-Schnittstelle: Die OGC-konforme CSW-Schnittstelle basiert, analog zum CSW-iPlug, im Wesentlichen auf den ISO-Standards 19115, 19119 und 19139. Die Schnittstelle ist für den Austausch von geografischen Metadaten konzipiert. Über sie können externe Systeme auf geografische Metadaten eines InGrid[®]-Portals zugreifen. Die Schnittstelle spielt im Rahmen der Geodateninfrastruktur Deutschland (GDI-DE) sowie von INSPIRE eine zentrale Rolle.

2.4 Portaloberfläche

Die Portaloberfläche stellt die für den Nutzer sichtbare Oberfläche des Internetportals dar. Sie bietet eine einfache und eine erweiterte Suche, Themen-, Messwert- und Serviceseiten, RSS-Feeds, eine Umweltchronik, einen integrierten OGC-konformen Kartenklienten und administrative Komponenten.

In der erweiterten Suche können Suchfilter gesetzt werden, die weit über die einfache logische Verknüpfungen („UND“, „ODER“ und „UND NICHT“) hinausgehen. Über den semantischen Suchfilter kann gezielt nach Begriffen aus dem UMTHESES, einem umfangreichen Umweltthesaurus des Umweltbundesamtes, gesucht werden. Über den räumlichen Filter kann der Suchraum eingeschränkt werden. Hierbei können sowohl geografische Begriffe wie z.B. Berlin oder Harz als auch Koordinaten, die anhand einer Karte ausge-

wählt werden können, als räumliche Filter verwendet werden. Über den zeitlichen Filter kann die Suche zudem auf einen Zeitraum oder Zeitpunkt eingeschränkt werden.

Die Rubrik Themenseiten bietet Informationen zu einzelnen Themen, die als besonders relevant eingestuft wurden. Die Themenliste schließt hierbei alle Themen ein, die von der Richtlinie 2003/4/EG als Umweltinformationen definiert werden. Bei den Messwerten werden Luft, Wasser, Strahlung und Andere unterschieden. Über die Auswahl einer dieser Kategorien werden anbieterübergreifend die entsprechenden Messwertseiten angezeigt.

Unter der Rubrik Services sind Pressemitteilungen, Publikationen und Informationen zu Veranstaltungen der Anbieter zu finden. Zudem können auf der Startseite eines InGrid[®]-Portals RSS-Feeds verschiedener Anbieter angezeigt werden. Die Feeds werden anbieterübergreifend nach Aktualität sortiert angezeigt. Zudem werden täglich umweltrelevante Ereignisse der Vergangenheit angeboten. Die Daten dafür werden vom SNS geliefert. Eine Recherche dieser Umweltereignisse ist zudem über die Rubrik Umweltchronik möglich. Über den Kartenklienten können Karten angezeigt werden. Er arbeitet sowohl mit dem in InGrid[®] integrierten Kartenserver, als auch mit externen, OGC-konformen Kartenservern zusammen und ermöglicht so die Zusammenstellung und integrative Analyse von Kartenquellen unterschiedlicher Anbieter.

Neben diesen Hauptfunktionalitäten der Portaloberflächen bietet ein InGrid[®]-Portal allen Nutzern unter der Rubrik „Mein PortalU“ die Möglichkeit, sich einen personalisierten Zugang einzurichten und das Portal damit auf seine Bedürfnisse anzupassen (Ansicht und Inhalte der Startseite, Suchpräferenzen, Speichern komplexer Suchen, etc.). Informationsanbietern bietet diese Rubrik einen Zugang zum Administrationsbereich, in dem die URL-Pflege der Informationsangebote (Webseiten für die Suche sowie Themen-, Messwert- und Serviceseiten) vorgenommen wird.

2.5 InGrid[®] Catalog

Mit dem InGrid[®] Catalog, dem Nachfolgesystem der Umweltdatenkatalog-Software, können Metadaten erfasst, aktualisiert und recherchiert werden. Die Erfassung und Aktualisierung der Metadaten erfolgt hierbei im InGrid[®] Editor über eine einfach zu bedienende Weboberfläche. Der InGrid[®] Catalog geht auf ein F&E-Projekt der Jahre 1991-1995 zurück. Software und Datenmodell wurden seither kontinuierlich weiter entwickelt und kommen seit mehr als 15 Jahren in zahlreichen Umwelt- und Vermessungsverwaltungen von Bund und Ländern zum Einsatz.

Mit dem InGrid[®] Editor wird die langjährige Desktop-Metadatenerfassungs- und Pflegeanwendung, der Windows-UDK, durch eine leistungsfähige Webanwendung ersetzt, die eine dezentrale Metadatenpflege über Standard-Browsertechnologie via Internet ermöglicht. Der InGrid[®] Catalog basiert auf dem ISO 19115- und ISO 19119-konformen UDK-Datenmodell und erfüllt die Anforderungen der INSPIRE-Richtlinie. Das Datenmodell geht mit seinen sechs Objektklassen fachlich erheblich über die Möglichkeiten anderer gängiger ISO 19115- und ISO 19119-basierter Metadaten-Erfassungsanwendungen hinaus. So können neben Geoinformationen und Geodiensten, Informationen zu Organisationseinheiten, Fachaufgaben, Dokumenten, Berichten und Literatur sowie Vorhaben, Projekten, Programmen, Datensammlungen und Datenbanken erfasst werden.

Besondere Merkmale des InGrid®Editors

- Alle Metadatenobjekte und -adressen werden in einem Hierarchiebaum abgebildet, so dass eine gute Übersichtlichkeit und ein einfacher Zugriff auf die Objekte gegeben sind. Die Hierarchie kann sowohl an organisatorischen als auch an fachlichen Strukturen angelehnt sein.
- Adressen werden in einem separaten Bereich eingegeben und gepflegt und anschließend mit den entsprechenden Metadatenobjekten lediglich verlinkt, was die Pflege erleichtert. Für neue Objekte werden die Adressangaben vom übergeordneten Objekt übernommen.
- Zur Sicherung der Qualität der Metadatensätze können einzelne Arbeitsabläufe überprüft sowie Verfallsdaten für Objekte festgesetzt werden. Ist diese Qualitätssicherung aktiviert, werden alle neu angelegten, bearbeiteten und gelöschten Metadatensätze vor Veröffentlichung einer zweiten Person zur Kontrolle vorgelegt. Bei der Aktivierung einer Verfallszeitspanne für Objekte, werden dem verantwortlichen Metadatenerfasser nach Ablauf einer definierten Zeitspanne (z.B. nach einem Jahr) die jeweiligen Objekte erneut zur Überprüfung vorgelegt, um eine angemessene Aktualität der Kataloginformationen sicher zu stellen.
- Die Metadatenerfassung von Webservices (WMS, WFS, WCS, WCTS), Internetangeboten sowie fachlichen und räumlichen Schlagworten wird durch verschiedene Assistenten unterstützt. Als Fachthesaurus ist hierbei der umfangreiche UMTHEs des Umweltbundesamtes integriert.
- Alle Metadaten eines InGrid®Catalogs können über die OGC-konforme CSW-Schnittstelle von InGrid® in andere Geodateninfrastrukturen eingebettet werden.

3 Verwendete Open Source Software für InGrid®

Bei der Umsetzung von InGrid® wurden als Basis ausschließlich Open Source Technologien verwendet. Die Nutzung und Weitergabe der Software an Partner der Verwaltungsvereinbarung ist damit ohne weitere Lizenzkosten möglich.

Portal Oberfläche	→ Jetspeed (Apache Lizenz, Apache Software Foundation)
Suchmaschine	→ Nutch (Apache Lizenz, ASF) & Lucene (Apache Lizenz, ASF)
Mapserver	→ UMN-Map-Server (eigene Open Source Lizenz)
Kartenklient	→ MapBender (GPL Lizenz)
Indexierung von Datenbanken	→ Lucene (Apache Lizenz, ASF)
Unterstützte Datenbanken	→ MySQL (GPL Lizenz), PostgreSQL (BSD Lizenz) (Proprietär: Oracle, Microsoft SQL-Server, Tamino)

4 Einsatzmöglichkeiten von InGrid®

In Deutschland wurde das Umweltportal Deutschland PortalU® als zentraler Zugang zu Umweltinformationen von Bund und Ländern für Bürger und Fachanwender aufgebaut. Hierfür wurde die Software InGrid® entwickelt, die so konzipiert ist, dass sie auch auf anderen Ebenen und in unterschiedlichen Einsatzszenarien genutzt werden kann.

4.1 Einsatz im Umweltbereich in Deutschland

InGrid® kann für den Aufbau regionaler- bzw. Landesumweltportale sowie kommunaler Umweltportale verwendet werden. Einfache Änderungen der Portaloberfläche können durch Anpassung von Stylesheets und den Austausch von Bilddateien vorgenommen werden. Weitergehende Anforderungen können durch die Entwicklung eigener Ansichtsprofile realisiert werden.

4.2 Einsatz als Geoportal

InGrid® unterstützt den OGC-Standard CSW 2.0 für geografische Metadaten. Über den eingebauten Kartenklienten können OGC-konforme Kartendienste dargestellt und kombiniert werden. So können die über die Metadatensuche gefundenen Kartendienste direkt angezeigt und analysiert werden. Mit dem InGrid®-Editor verfügt die InGrid®-Architektur über eine leistungsfähige, ISO 19115 / 19119- und INSPIRE-konforme Metadatenerfassungs- und -pflegekomponente. Damit erfüllt InGrid® alle Kriterien zum Einsatz als GDI-DE- oder INSPIRE-Geoportal.

4.3 Einsatz als allgemeines Verwaltungsportal

Über InGrid® lässt sich zudem ein allgemeines Verwaltungsportal aufbauen. Im Unterschied zu PortalU® ist dabei der thematische Fokus nicht auf den Bereich Umwelt eingeschränkt, sondern deckt alle verwaltungsrelevanten Themenbereiche ab. Folgende Anpassungen wären hierfür notwendig:

1. Umweltthemen müssen durch allgemeine Themen wie z.B. Verwaltungsbereiche oder Lebenslagen ersetzt werden.
2. Messwertseiten müssen durch allgemeine Themen ersetzt werden. Denkbar wäre z.B. „Messwerte“ durch „Fakten“ zu ersetzen, hinter denen sich Zahlen zu verschiedenen Politikfeldern verbergen.
3. Der Umweltthesaurus ist für ein allgemeines Verwaltungsportal zu speziell. Er sollte deshalb mittelfristig durch einen allgemeinen Verwaltungs- oder Politik-Thesaurus ersetzt werden.
4. Die Umweltchronik ist ebenfalls zu speziell und sollte daher ausgeschaltet werden.
5. Die Rubrik „Aktuelles“ sollte alle Pressemitteilungen von Bund und Ländern beinhalten.

Grundsätzlich ist nur ein geringer Aufwand notwendig, um InGrid[®] als Software für ein allgemeines Verwaltungsportal zu nutzen.

4.4 Einsatz als Umweltportal in anderen europäischen Staaten

Beim Einsatz von InGrid[®] als Umweltportal außerhalb Deutschlands sind folgende Anpassungen notwendig:

1. Landessprache: Die Oberfläche von InGrid[®] ist zweisprachig (in PortalU[®] derzeit Deutsch und Englisch). Weitere Sprachen können durch den Austausch bzw. die Erweiterung von Konfigurationsdateien realisiert werden.
2. Layout: Änderungen des Layouts sind durch Anpassung von Stylesheets und Austausch von Bilddateien sowie durch die Entwicklung gänzlich neuer Ansichtsprofile möglich.
3. Mehrsprachigkeit: Wenn mehr als zwei Sprachen berücksichtigt werden sollen, muss neben den Übersetzungsarbeiten auch der derzeitige Umschaltmechanismus zwischen Sprache 1 (z.B. Deutsch) und Sprache 2 (z.B. Englisch) durch einen Mechanismus für mehrere Sprachen ersetzt werden.
4. Thesaurusunterstützung: Derzeit wird über SNS der UMTHEs des Umweltbundesamtes als Umweltthesaurus genutzt. Dieser Thesaurus ist zweisprachig (deutsch / englisch). Für Staaten mit anderen Landessprachen muss entweder die Thesaurusunterstützung abgeschaltet werden oder der „deutsche SNS“ durch einen „europäischen SNS“ auf Basis eines multilingualen Thesaurus, z.B. GEMET, ersetzt werden.
5. Raumbezug: Derzeit werden über SNS nur Raumbezüge innerhalb Deutschlands unterstützt. Für andere europäische Staaten muss entweder die Unterstützung des Raumbezugs bei der Suche abgeschaltet werden oder der „deutsche SNS“ muss durch einen „europäischen SNS“ ersetzt werden.

4.5 Einsatz als europäisches Umweltportal

Technisch gesehen unterscheidet sich der Einsatz von InGrid[®] auf europäischer Ebene nur geringfügig von dem auf nationaler Ebene. Die Software ist grundsätzlich so organisiert, dass auch Hierarchiestufen oberhalb der nationalen Ebene abgedeckt werden können. Die PortalU[®]-Auswahl nach Bund/Ländern kann durch die Auswahl der europäischen Staaten ersetzt werden.