

# Open Data!

## Umweltinformatik in Behörden und Verwaltung

Gerlinde Knetsch

Informationssysteme Chemikaliensicherheit  
Umweltbundesamt  
Wörlitzer Platz 1  
06844 Dessau-Roßlau  
Gerlinde.Knetsch@uba.de

**Abstract:** Gesetzliche Regelungen und Anforderungen der Öffentlichkeit an die aktive Bereitstellung von Umweltdaten und Informationen verlangen fachliche und technische Konzepte sowie Umsetzungsstrategien für den Aufbau und die Entwicklung von Umweltinformationssystemen im behördlichen Umfeld. Kooperationsmodelle zwischen Länder- und Bundesverwaltungen bieten die Grundlage für ein abgestimmtes Vorgehen in nationalen und internationalen Gremien. Zunehmend setzen sich Strategien und Konzepte wie Open Data, Linked Open Data und das Prinzip der Service Orientierten Architektur (SOA) im Angebot von Umweltdaten durch. Umweltinformatik ist somit Mittel zum Zweck mit Blick auf die Umsetzung des Regierungsprogramms „Vernetzte und transparente Verwaltung“. Die Initiative der Bundesregierung, eine bundesweite Plattform für Open Data zu schaffen, ist mit der Beta-Version des „GovDATA Datenportal für Deutschland“ in einem ersten Schritt in der Umsetzung.

### 1 Einleitung

Der Begriff Umweltinformationssystem wurde erstmals auf der UNO-Weltkonferenz für menschliche Umwelt (kurz Umweltschutzkonferenz) in Stockholm 1972 im Kontext der Umweltbeobachtungs- und Bewertungsaktivitäten definiert. "The Stockholm Conference (1972) called its action plan to monitor and assess the global environment Earthwatch. [...] Today Earthwatch is the framework for the United Nations system to harmonize and integrate its environmental observing and assessment activities under UNEP's leadership." In der Empfehlung 76 (b) der Deklaration stellt die UNO fest: "That the World Health Organization co-ordinate the development and implementation of an appropriate international collection and dissemination system to correlate medical, environmental and family-history data" (UNEP 1972).

In Deutschland gründete sich zeitgleich der Bund-Länder-Arbeitskreis Umweltinformationssysteme (BLAK UIS) mit dem Auftrag an Experten, ein

Umweltplanungs- und Umweltinformationssystem aufzubauen. Zahlreiche Konzepte mit wegweisendem Charakter und technische Lösungen wurden in den folgenden Jahren entwickelt, die einerseits auf spezifische fachbezogene Informationssysteme ausgerichtet sind, andererseits den Charakter von Verweissystemen und integrierende vernetzten Systemen aufweisen.

Bilanziert man vierzig Jahre später die Entwicklung und Umsetzung derartiger Konzepte und Umsetzungsstrategien von Umweltinformationssystemen, so findet derzeit ein Paradigmenwechsel statt. Der Wert der Ware „Umweltinformation“ erhält im Kontext der Beantwortung umweltpolitischer und gesellschaftlicher Fragestellungen, in der Auseinandersetzung mit Problemen des Umwelt- und Risikomanagement einen zunehmenden wissenschaftlichen und ökonomischen Stellenwert. Die Entwicklung von öffentlichen Datennetzwerken auf der Grundlage des Open Data – Prinzips öffnet Datenbestände, die im Interesse der Allgemeinheit der Gesellschaft sind. Sie werden zur freien Nutzung und zur freien weiteren Weiterverwendung bekannt und öffentlich zugänglich gemacht. Die Open Data Governance birgt somit erhebliche Nutzungspotenziale. Sie fördert die Vernetzung von Daten untereinander und trägt somit zum Wissens- und Informationsmanagement innerhalb der öffentlichen Verwaltung bei [Kl12].

## **2 Zugang zu Umweltdaten Open Data – Linked Data**

Umweltinformatik in der öffentlichen Verwaltung bietet ein breites Entwicklungs- und Anwendungsgebiet für innovative Entwicklungen zur Bereitstellung von Umweltinformationen. Seit Jahren liefert sie einen wichtigen Impuls über die Verwaltungsstrukturen hinweg, Daten für Metadaten-, Fachinformations- und integrierte Umweltinformationssysteme bereitzustellen. Das Prinzip der offenen Daten kann darüber hinaus interne Strukturen der Verwaltung wesentlich vereinfachen und Datenflüsse zwischen verschiedenen Abteilungen einer Behörde optimieren.

Technische Möglichkeiten, die die Umweltinformatik bietet, verstärken die partizipativen Bestrebungen der Mehrfachnutzung von Umweltdaten. Die Umsetzung der Open Data-Architektur erfordert somit neue und übergreifende Strategien zur Bereitstellung und Nutzbarkeit von Daten aus dem behördlichen Umfeld. Ein wesentliches strategisches Ziel ist die Bereitstellung von qualitativ hochwertigen und verlässlichen Daten [Le12].

Erfahrungen zeigen jedoch, dass vorhandene Datenbestände die Anforderungen an Vollständigkeit, Reproduzierbarkeit und Eindeutigkeit nicht immer erfüllen. Dies betrifft nicht nur die Primärdaten, also die Messdaten zum Beispiel eines Umweltbeobachtungsprogramms, sondern auch die Metadaten d.h. Informationen über die Primärdaten selbst. Die Qualität der Metadaten ist somit ein Indikator für die Möglichkeiten und Grenzen der Integration von Umweltdaten in einem fachlichen Kontext.

## 2.1 Voraussetzungen für die Umsetzung von Open Data in der Verwaltung

Voraussetzung für die Umsetzung des Open Data Prinzips in der öffentlichen Verwaltung und von Behörden sind qualitativ hochwertige, gut strukturierte, valide und leicht zugängliche sowie reproduzierbare Umweltinformationen aus verschiedenen Bereichen der Umwelt und des öffentlichen Lebens. Effektive und effiziente Methoden der Datenharmonisierung, der Archivierung und des Datenmanagement stellen eine Herausforderung dar, um über Portale mit intelligenten Such- und Zugriffsmechanismen diese Informationen in einen fragegeleiteten Kontext zu stellen.

Verschiedene Prozessschritte für die Bereiche der Datenhaltung und Qualitätssicherung, der Datenaufbereitung, und Darstellung sind für eine langfristige Verfügbarkeit sowie die angemessene Interpretierbarkeit von Umweltdaten im Kontext der freien Verfügbarkeit notwendig. Abbildung 1 gibt einen Überblick zu wichtigen Metainformationen mit Blick auf die fachlichen, räumlichen und zeitlichen Kriterien für qualitätsgesicherte Daten.

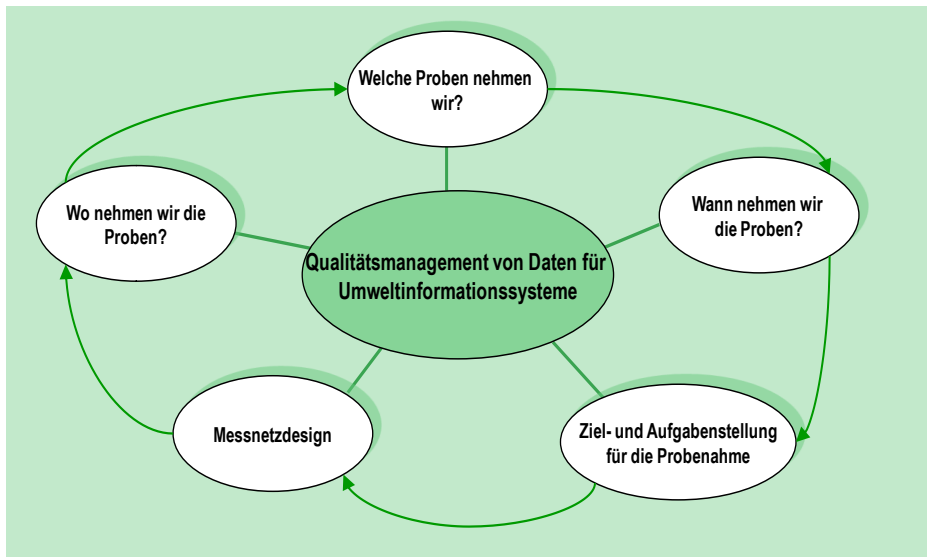


Abbildung 1: Qualitätsmanagement von Umweltdaten

Die Bereitstellung von Open (Government) Daten ist somit ein Glied in der Wertschöpfungskette von Umweltinformationen, die verantwortungsbewusst und unter Einhaltung der Schutzwürdigkeit z.B. von personenbezogenen und sicherheitsrelevanten Informationen stattfinden muss. Die Festlegung von verbindlichen Regeln und der damit verbundenen Prozessschritte stellen ein wesentliches methodisches Design dar für die Organisation, Methodik und Technik im Kontext der Umweltinformatik. Die nachfolgende Übersicht zeigt, dass neben den technischen Regeln auch organisatorische und methodische Regelungen für hochwertige Umweltdaten von Bedeutung sind [Kn11].

Organisatorische Regelungen:

- betrifft nachvollziehbare und nachprüfbare Prozessschritte der Datenerhebung sowie die eindeutige Verantwortlichkeit des Datenhalters,
- umfasst Angaben zur Validität (Gültigkeit), Reliabilität (Zuverlässigkeit) und Reproduzierbarkeit (Nachvollziehbarkeit),
- schließt die Dokumentation von Auffälligkeiten ein, die im Kontext zu dem Datum steht.

Methodische Regelungen:

- betrifft die Validität der Probennahmeplanung und das Messnetzdesign, die Probennahme und die Aufbereitung für das Ziel der Untersuchung,
- umfasst das Qualitätssicherungsmanagement der analytischen Ergebnisse,
- schließt die Art und Weise der statistischen Bearbeitung und Auswertung der Daten ein.

Technische Regelungen:

- betrifft die technische Übermittlung der Daten nach vorgegebenen organisatorischen Regelungen,
- umfasst die Einhaltung von technischen Standards und von Datenformaten,
- schließt die Interoperabilität von Umweltdaten in einen anderen Kontext ein (semantischer Datensatz).

Ein systematischer Ansatz zur Überprüfung dieser Regeln kann z.B. mit technischen Algorithmen vorgenommen werden. Durch die Implementierung von SQL-Abfragen, die spalten-, zeilen- und tabellenübergreifende Wenn-Dann-Beziehungen prüfen, steht ein methodisches Instrumentarium zur Verfügung, um Erwartungshaltungen an die Daten in einem Umweltinformationssystem zu erfüllen.

## **2.2 Umsetzungsbeispiel von Open Data im Umweltbundesamt**

Projekte im Umweltbundesamt realisieren seit einigen Jahren den Aufbau eines öffentlichen Datennetzwerkes auf der Grundlage des Open Data – Prinzips. Datenbestände, die im Interesse der Allgemeinheit der Gesellschaft sind, werden zur freien Nutzung und zur freien weiteren Weiterverwendung öffentlich zugänglich gemacht. Nachdem in vielen Ländern zum Beispiel in Großbritannien bereits eine umfangreiche Bereitstellung von offenen Daten praktiziert wird [Co12], findet eine Umsetzung auch in Deutschland statt.

Ein Beispiel für ein derartiges offene Informationsangebote ist das Instrument der Umweltprobenbank des Bundes. Die Umweltprobenbank des Bundes ist ein Archiv von Proben, mit der die Qualität der Umwelt und die Belastungssituation des Menschen dokumentiert und bewertet werden kann. Diese Proben sollen für einen bestimmten Raum repräsentativ sein und sie müssen regelmäßig erhoben werden, um Veränderungen der (Schad)stoffbelastung im Laufe der Zeit verfolgen zu können [BM08].

Vor Einlagerung der Umwelt- und Humanproben findet eine Charakterisierung auf etwa 60 chemische Stoffe statt. Alle im Routinebetrieb der Umweltprobenbank anfallenden Daten aus dem Umwelt- und Humanbereich werden im Informationssystem Umweltprobenbank (IS UPB) für den aktuellen Zugriff gespeichert, aufbereitet und qualitätsgesichert. Diese Daten stehen der Öffentlichkeit mit allen dazu verfügbaren Metadaten über ein Recherche-Portal online zur Verfügung.



Abbildung 2: Portalseite der Umweltprobenbank des Bundes<sup>1</sup>

Ein hohes Maß an Qualitätssicherung wird dadurch erreichen, dass alle Schritte von der Probenahme über den Probentransport, die Probenaufarbeitung und Analytik bis zur Langzeitlagerung in Standardarbeitsanweisungen (englisch Standard Operating Procedures (SOP)) für alle Probenarten verbindlich festgelegt wurden.

Alle Daten können nach verschiedenen Parametern und Kriterien über ein frei zugängliches Interface recherchiert und in Diagrammen oder Tabellen angezeigt werden. Über die Funktion Export stehen dem Anwender verschiedene Formate (.xls oder .csv) im Download-Bereich zur Verfügung. Jede Abfrage kann als Lesezeichen (Bookmarks) gesetzt und für spätere Recherchen mit aktualisierten Daten wiederholt werden, die durch eigene echte Web-Adressen referenziert werden.

Das Informationssystem zu Daten aus der Umweltprobenbank des Bundes ist ein Beispiel für die aktive Bereitstellung von Umweltdaten unter Anwendung innovativer Methoden der Umweltinformatik. Es setzt die Anforderungen an die Open Data Strategie um und kann nach Berners-Lee's 5-Star-Rating<sup>2</sup> mit drei Sterne eingestuft werden [Rü12].

<sup>1</sup> <http://www.umweltprobenbank.de/>

<sup>2</sup> <http://www.w3.org/DesignIssues/LinkedData.html>

## 2.3 Linked Open Data Prinzip - Brücke zur Integration und Interoperabilität zwischen Fachdatenbanken

Linked Open Data (LOD) ist eine ganz spezifische Entwicklungsrichtung der Datenbereitstellung. Einerseits betrifft dies die Öffnung von Datenbeständen für die Verwendung ausgerichtet auf verschiedene Nutzergruppen. Andererseits steht die Verknüpfung mit anderen Datenbeständen in einem semantischen Netzwerk (Web of Data) im Mittelpunkt. Im Umweltbundesamt bestehen erste praktische Ansätze in der Implementierung derartiger Vernetzungsstrategien.

Während bei Daten Services der konzeptuelle Zugriff bzw. die Bereitstellung von Daten über eine REST-Schnittstelle im Vordergrund steht, ist bei dem Linked Open Data Prinzip die Abbildung der Daten gemäß eines Resource Description Framework (RDF) vorzunehmen, wobei die eine Ressource (Object, Entity) beschreibenden Eigenschaften/Attribute jeweils in ein Tripel (Ressource, Eigenschaft, Objekt / Wert) aufgelöst wird. Die Adressierung der Tripel-Elemente erfolgt einheitlich über ein URI (Uniform Resource Identifier). Die Speicherung dieser Tripel erfolgt in einem Tripel Store, der mit der Abfragesprache SPARQL abgefragt werden kann. Weitere technische Details sind unter der Informationsplattform des W3C eGov Wiki<sup>3</sup> abrufbar.

Diese Entwicklungsrichtung der Datenverlinkung über einen Uniform Resource Identifier (URI) untersucht ein Forschungsprojekt des Umweltbundesamt seit 2012. Das Pilotprojekt „Linked Environment Data - Informationsmehrwert durch verknüpfte Umweltdaten“ (UFOPLAN 2012 - FKZ 3712 12 100) befasst sich mit der Vernetzung von öffentlichen, allgemein zugänglichen („offenen“) Umweltdaten nach technischen Mustern des W3C<sup>4</sup>. Offenheit hat dabei drei Bedeutungen:

- Diese Daten sind uneingeschränkt frei verfügbar;
- Der Zugang basiert auf offenen Standards - kommerzielle Werkzeuge sind nicht erforderlich;
- Die Daten dürfen weltweit mit Daten anderer Organisationen vernetzt werden.

Die Vernetzung erfolgt inhaltlich „punktgenau“ durch direkte Verweise zwischen zwei Datensätzen in verschiedenen Datenbanken. Die Verweise sind Hyperlinks, ihnen kann daher auch im Web-Browser sofort gefolgt werden. Damit entsteht ein hochwertiger Baustein der „globalen Datenbank“ des Semantic Web. Abbildung 3 veranschaulicht einige dieser Bausteine, die derzeit mit dem Forschungsprojekt nach dem LED-Prinzip mit einander vernetzt werden sollen. Die Umweltprobenbank des Bundes (ESB) ist bereits in Kap. 2.2 beschrieben; unter dem Baustein „more“ wird auf ein Informationssystem zu Chemikalien verwiesen, der „Gemeinsame Stoffdatenpool Bund Länder (GSBL)“. Kap. 2.4 befasst sich mit der Verlinkung eines Informationssystems zu persistenten organischen Stoffen<sup>5</sup> mit einem Informationspool (GSBL<sup>6</sup>) zu chemischen Stoffeigenschaften.

---

<sup>3</sup> [http://www.w3.org/egov/wiki/Main\\_Page](http://www.w3.org/egov/wiki/Main_Page)

<sup>4</sup> <http://www.w3.org/standards/>

<sup>5</sup> <http://www.pop-dioxindb.de>

<sup>6</sup> <http://www.gsbl.de>

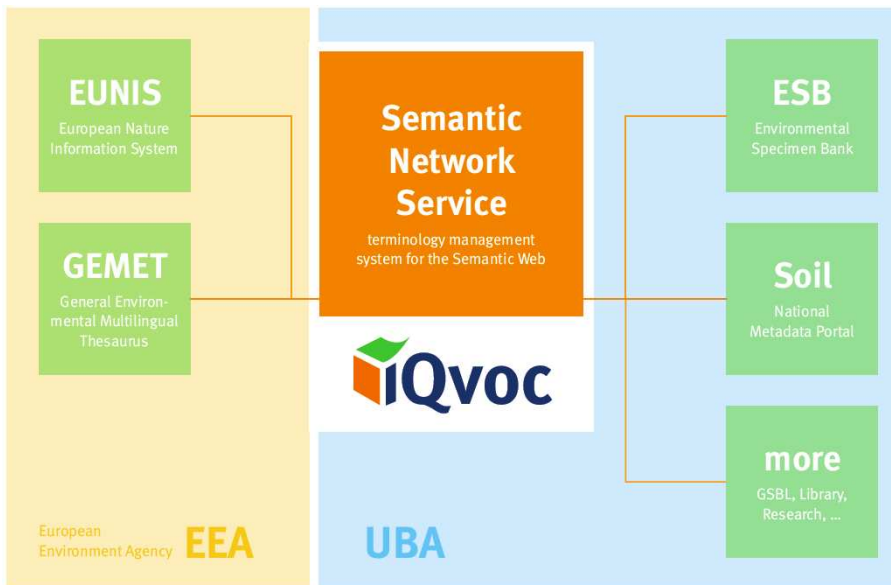


Abbildung 3: Bausteine eines Linked Environment Data Network [Fo12]

## 2.4 Verlinkung von Monitoringdaten mit Chemischen Stoffeigenschaften

Mit dem Informationssystem zu persistenten organischen Stoffen, umgangssprachlich auch POPs<sup>7</sup> genannt, wurde ein zuständigkeits-(Bund/Länder) und ressortübergreifendes (Lebensmittel/Gesundheit/Umwelt) Instrument zum Management und zur Auswertung von Daten aus mehr als 100 Umweltbeobachtungsprogrammen des Bundes und der Länder entwickelt. Die rasante Entwicklung von technologischen Möglichkeiten der Umweltinformatik unterstützte die konzeptionelle Umsetzung eines föderativen Umweltinformationssystems zwischen mehreren Bundes- und Länderbehörden

Aufgabe der POP-Dioxin-Datenbank ist es, auf der Grundlage der dort dokumentierten Ergebnisse aus Messprogrammen zeitliche und räumliche Belastungsaussagen zu den Kompartimenten Boden, Wasser, Luft, Abwasser und Abfall, Biota, Stoffen, Zubereitungen und Erzeugnissen zu erstellen. Das Bundesamt für Verbraucherschutz und Lebensmittelsicherheit und das Bundesinstitut für Risikobewertung, beide Partner dieses Informationssystems, führen in ihrem Zuständigkeitsbereich insbesondere zu Lebensmitteln eigene Aus- und Bewertungen durch. Standen zu Beginn der Dioxin-Datenbank die Dokumentation der Programme und ihrer Ergebnisse, mit Blick auf die Berichterstattung zur Belastung der Umwelt im Vordergrund, so wurden die Aufgaben mit Entwicklung eines Webportals in 2002 auch um die Information für die Fachöffentlichkeit und die Bürger erweitert.

<sup>7</sup> <http://www.umweltbundesamt.de/chemikalien/pops.htm>

Ein Alleinstellungsmerkmal dieser Internetanwendung ist die Zusammenführung von Daten der verschiedenen Umweltkompartimente, dem Futter- und Lebensmittelbereich sowie der menschlichen Belastung in einem gemeinsamen Datenbanksystem. Mit einem derartigen integrativen Ansatz können Pfade von Belastungen aufgespürt werden und Ursachen von Beeinträchtigungen insbesondere von Lebensmitteln erforscht werden. Über einen Web-Zugang und ein Rechte- und Rollenkonzept können verschiedene Nutzergruppen in den Primärdaten des Informationssystems recherchieren [Kn07].

Der interessierte Bürger, aber auch die wissenschaftliche Community möchte jedoch nicht nur Informationen zur Belastungssituation der Umwelt mit diesen persistenten Stoffen. Auch deren chemische Eigenschaften, wie die Toxizität, das Bioakkumulationsverhalten in Fischen oder anderen Lebewesen steht dabei im Blickfeld des Interesses. Über eine URI verlinkt das Register „Schadstoffe“ des POP - Dioxin-Informationssystem zu dem Chemischen Stoffdatenpool GSBL direkt in dessen Stoff-Dossier. Abbildung 4 stellt dieses Verlinkungsprinzip (LED) an einem ausgewählten Beispiel des PCB 153 dar.

The diagram illustrates the linking principle (LED) between two information resources. It shows a navigation menu on the left of a 'Schadstoffe' page, with 'PCB 153' highlighted. An arrow points from this entry to a 'public' page showing the 'Stoff 1' dossier for '2,2',4,4',5,5'-Hexachlorbiphenyl'. The dossier includes a search bar, a list of search results, and a table of registrant names.

| CAS-Nr.    | Substanz | Synonym                                  |
|------------|----------|------------------------------------------|
| 25429-29-2 | PentaCB  | -                                        |
| 26601-64-9 | HexaCB   | -                                        |
| 26914-33-0 | TetraCB  | -                                        |
| 28655-71-2 | HeptaCB  | -                                        |
| 31508-00-6 | PCB 118  | 1,1'-Biphenyl, 2,3',4,4',5-pentachlor-   |
| 32598-13-3 | PCB 77   | 1,1'-Biphenyl, 3,3',4,4'-Tetrachlor-     |
| 32598-14-4 | PCB 105  | 1,1'-Biphenyl, 2,3,3',4,4'-Pentachlor-   |
| 32690-93-0 | PCB 74   | 1,1'-Biphenyl, 2,4,4',5-tetrachlor-      |
| 32774-16-6 | PCB 169  | 1,1'-Biphenyl, 3,3',4,4',5,5'-Hexachlor- |
| 33025-41-1 | PCB 60   | 1,1'-Biphenyl, 2,3,4,4'-tetrachlor-      |
| 35065-27-1 | PCB 153  | 1,1'-Biphenyl, 2,2',4,4',5,5'-Hexachlor- |

public

GSBL2012\_1\_public > Liste 1 > Stoff 1 > 2,2',4,4',5,5'-Hexachlorbiphenyl

Suche | Export

Treffermengen | Dossier

Suchbegriff:

☒ Bezeichnung ☐ Kurzname

☒ Stoff 1: 2,2',4,4',5,5'-Hexachlorbiphenyl

☒ IDENTMERKMALE

☒ RECHTSEIGENSCHAFTEN

☒ UMWELTVERHALTEN

☒ Akkumulation und Verteilung in Umweltkomp

☒ Konzentration in der Umwelt (19)

☒ Biokonzentration (1)

| GSBL-RN  | Stoffart           | Struktur |
|----------|--------------------|----------|
| 1 132166 | Einzelinhaltsstoff |          |

Registriernamen (20)

| Registriernamen                             | Namensart      | Sprachkennung | GSBL-STOFF | Quelle   |
|---------------------------------------------|----------------|---------------|------------|----------|
| 1 1,1'-Biphenyl, 2,2',4,4',5,5'-hexachloro- | Hauptname      | unbekannt     |            | CHEMIS - |
| 2 Biphenyl, 2,2',4,4',5,5'-hexachloro-      | CAS-Index Name | unbekannt     |            | CHEMIS - |

Abbildung 4: Darstellungsprinzip der Verlinkung zwischen zwei Informationsressourcen

### 3 Schlussfolgerungen und Ausblick

Mit der Entwicklung von webbasierenden Informationsangeboten und deren Vernetzung untereinander stehen für Bürger und Wissenschaft verlässliche und verständliche Informationen zur Verfügung. Die Darstellung von Belastungssituationen und zeitliche Trendentwicklungen geben dem Bürger die Möglichkeiten, sich zu speziellen Themen zu informieren. Hierbei stehen zumeist aggregierte Informationen im Vordergrund, die bewertet sind und somit eine Einordnung für den Bürger zulassen. Die in Abbildung 7 ausgewerteten Daten des deutschen Lebensmittelmonitorings zu spezifischen Lebensmittelgruppen werden zu den in der EU geltenden Höchstwerten in Beziehung gesetzt.

### Literaturverzeichnis

- [Kl12] Klessmann, J. et al.: Open Government Data Deutschland. Studie im Auftrag des Bundesministeriums des Inneren. Berlin, 2012.
- [Le12] Leißner, A.; van Nouhuys, J.: Open Data Governance. In (Umweltbundesamt, Hrsg.): Umweltinformationssysteme – Frühwarn- und Informationssysteme für den Hochwasserschutz. UBA-Text 41/2012, Dessau-Roßlau, 2012, S. 81-105.
- [Kn11] Knetsch, G.: Auswahl und Bewertung von Daten aus Umweltbeobachtungsprogrammen zur Bilanzierung von polychlorierten Biphenylen Modelltheoretische Ansätze der Integration. Dissertation, Freie Universität Berlin, 2011.
- [BM08] Konzeption der Umweltprobenbank des Bundes. Bonn. 2008. [http://www.umweltprobenbank.de/upb\\_static/fck/download/Konzeption\\_Okt\\_2008\\_de.pdf](http://www.umweltprobenbank.de/upb_static/fck/download/Konzeption_Okt_2008_de.pdf)
- [Co12] Improving the transparency and accountability of government and its services. Open Data White Paper. GB, 2012. <https://www.gov.uk/government/publications/open-data-white-paper-unleashing-the-potential>
- [Rü12] Rüther, M.; Bandholtz, T.; Schulte-Coerne, T.: A Common Tool for Managing Environmental Monitoring Data. In (Arndt, H.K.; Knetsch, G.; Pillmann, W., Eds.): Proc. 26th International Conference on Informatics - Informatics for Environmental Protection, Sustainable Development and Risk. Shaker Verlag, Aachen, 2012; pp 467-472.
- [Fo12] Fock, J.; Bandholtz, T.: Linked Environment Data – Getting Things Connected. In (Arndt, H.K.; Knetsch, G.; Pillmann, W., Eds.): Proc. 26th International Conference on Informatics - Informatics for Environmental Protection, Sustainable Development and Risk. Shaker Verlag Aachen, 2012; pp 451-457.
- [Kn07] Knetsch, G.; Weihs, E.: Web-Services as an Instrument for data Retrieval and Data Mapping – The Dioxin-POP Database of Germany. In (Hryniewicz, O.; Studzinski, J.; Romaniuk, M.; Eds.): Environmental Informatics and Systems Research Warschau, Shaker Verlag Aachen, 2007; pp 585-592.