# Ein Metadatenkonzept für das Umweltinformationssystem Baden-Württemberg

Renate Ebel[1], Andree Keitel[1], Roland Mayer-Föll[2],
Thomas Dombeck[3], Albrecht Schultze[3],
Jörg Westbomke[3], Michael Haase[3]

## 1. Einführung und Zielsetzung

Daten in Umweltinformationssystemen (UIS) liegen üblicherweise in strukturierter Form in Datenbanken vor. Parallel zu dieser klassischen Form der Datenhaltung werden zunehmend auch Umweltinformationen in Form von Websites und Web-basierten Fachinformationssystemen angeboten. Neben den reinen Umweltdaten werden meist auch Metadaten, also beschreibende Informationen über Struktur und Inhalt der Umweltdatenobjekte, vorgehalten. Diese Metadaten vermitteln dem Anwender einerseits Zugang zu den Daten, ermöglichen aber auch die Integration von Datenbeständen und Diensten in übergeordneten Nachweis- und Recherchesystemen.

In Baden-Württemberg dient der Umweltdatenkatalog (UDK, 2002) als zentrales Nachweissystem des UIS. Der Nachweis von Umweltdatenbeständen im UDK durch manuelle Erfassung von Metadaten ist allerdings mit hohem Aufwand verbunden. Bisher konnte nur ein teilautomatischer Austausch von Metadaten zur Aufwandsreduktion durch SGML-basierte Schnittstellen, z. B. in der XfaWeb-Familie (XfaWeb, 2002), realisiert werden. Suchmaschinen, die insbesondere Web-Inhalte automatisch indizieren, sind der manuellen bzw. teilautomatischen Metadatenerfassung in Leistungsfähigkeit und Flexibilität zwar weit überlegen, liefern jedoch nicht die hohe Qualität der Metadaten des UDK-Standards.

Neben dem UDK sind Umweltinformationen aus dem UIS Baden-Württemberg auch über das Umweltinformationsnetz Deutschland (GEIN, 2002) recherchierbar. Hierbei handelt es sich um ein reines Recherchewerkzeug für öffentliche Umweltin-

[1] Landesanstalt für Umweltschutz Baden-Württemberg (LfU), Bannwaldallee 24, D-76157 Karlsruhe, E-Mail: Andree.Keitel | Renate.Ebel@lfuka.lfu.bwl.de
[2] Ministerium für Umwelt und Verkehr Baden-Württemberg (UVM), Kernerplatz 9, D-70182 Stuttgart, E-Mail: Roland.Mayer-Foell@uvm.bwl.de
[3] Forschungsinstitut für anwendungsorientierte Wissensverarbeitung (FAW), Helmholtzstraße 16, D-89081 Ulm, E-Mail: Dombeck | Schultze | Westbomke | Haase@faw.uni-ulm.de

formationen im Internet. Um einen umfassenden Zugriff auf alle Daten des UIS zu ermöglichen, wird die Einrichtung eines zentralen Umweltportals der Landesverwaltung Baden-Württemberg angestrebt. Grundlage dafür ist die Etablierung eines einheitlichen Metadaten-Standards im UIS Baden- Württemberg (Mayer-Föll et al. 2002).

## 2. Zur Problematik der Vereinheitlichung von Metadaten in Umweltinformationssystemen

Neben dem Aspekt des Nachweises von Umweltdatenbeständen sind Metadaten auch für den Zugang zu den Informationsangeboten von entscheidender Bedeutung. Dabei sind zwei grundsätzliche „Informationstypen“ klar zu unterscheiden: hoch strukturierte (DB-gestützte) Umweltinformationen werden zumeist über Navigationsfunktionen, i.d.R. basierend auf systeminternen Metadaten erschlossen, während in gering strukturierten Dokumentbeständen (Online-Dokumentensammlungen) neben dem navigatorischen Zugang, i.d.R. basierend auf mehr oder minder umfangreichen Metadaten, oft auch Katalog- und Suchfunktionen (z.B. eine Volltextsuche) zum Einsatz kommen, die den Zugang erleichtern.

Umweltfachdatenbanken erlauben i.d.R. eine Identifikation auf Objektebene (z.B. ein bestimmtes Biotop), d.h. eine Verknüpfung zwischen unterschiedlichen Umweltdatenbanken kann auf dieser Objektebene erfolgen. Unterschiedliche Attribute zu einem Umweltobjekt können so relativ leicht über vereinheitlichte Objektidentifikatoren Datenbank-übergreifend zusammengestellt werden. Im Gegensatz hierzu ist dieser Sachverhalt bei schwach oder nur gering strukturierten Dokumentbeständen häufig nicht gegeben, weil eine eindeutige Zuweisung zu einem Umweltobjekt nicht möglich ist. Vielmehr gilt es hier, über Metadaten bzw. Teile des Contents auf mögliche Beziehungen zu Umweltobjekten rückzuschließen.

Momentan ist eine Verknüpfung dieser beiden „Informationstypen“ bei der Datenbereitstellung und Recherche kaum feststellbar. Ansätze hierzu sind beispielsweise mit der Einbindung von Datenbanken des Umweltbundesamtes Berlin in das Informationsangebot von GEIN realisiert. In Baden-Württemberg wird nun die Realisierung eines zentralen Portals angestrebt, das die hoch strukturierten (DB-gestützten) Umweltinformationen mit den gering strukturierten (text- und bildbasierten) Informationen verbindet. Zur Realisierung eines solchen Portals ist jedoch die Verknüpfung der Informationen über einen einheitlichen Satz von Metadaten von entscheidender Bedeutung.

## 3. Vorgehensweise bei der Erstellung des Metadatenkonzeptes

Im Rahmen des Projektes UIS-Meta wurde die bisher praktizierte Metadatenverwaltung in 10 Systemen des UIS Baden-Württemberg analysiert. Ziel war es, An-

satzpunkte für eine mögliche Vereinheitlichung und Optimierung der Metadatenstrukturen aufzuzeigen. Die in den einzelnen Systemen eingesetzten Metadaten wurden identifiziert, analysiert und in Tabellen, gruppiert nach unterschiedlichen Metadatenklassen wie z.B. Raumbezug, Zeitbezug oder Verschlagwortung, dargestellt. Davon ausgehend wurde ein einheitliches Metadatenkonzept entwickelt, das den Anforderungen aller untersuchten Systeme genügt.

Das Spektrum der analysierten Systeme umfasst sowohl fachspezifische digitale Dokumentensammlungen (wie z.B. das Document Retrieval System (DRS) des Zentralen Fachdienstes Wasser, Abfall, Altlasten, Boden (ZFD) in der LfU) als auch komplexere, Datenbank-gestützte Systeme (wie z.B. das UIS-Berichtssystem) sowie den Umweltdatenkatalog als übergeordnetes Katalogsystem. Besondere Berücksichtigung fand auch die Verwendung von Metadaten im Umweltinformationsnetz Deutschland (GEIN, 2002).

## 4. Das Metadatenkonzept für das UIS Baden-Württemberg

Aufgabe des Metadatenkonzeptes ist es, einen Satz von Metadaten zu bestimmen, welcher einerseits die in den Systemen des UIS geführten Datenobjekte hinreichend genau beschreibt, andererseits aber auch auf alle Systeme angewendet werden kann. Um dieser Anforderung gerecht zu werden, wurde ein Kernsatz und ein erweiterter Satz von Metadaten definiert. Der Kernsatz wurde so konzipiert, dass die dort beschriebenen Metadaten für jedes der UIS-Systeme erhoben werden können. Dagegen deckt der erweiterte Satz speziellere Aspekte ab und kann je nach Anforderungen dem jeweiligen System angepasst werden. Während die Erfassung der Kernmetadaten verpflichtend ist, können die erweiterten Metadaten optional angegeben werden. Dieses Konzept eines erweiterbaren Kernsatzes findet u.a. auch im Metadatenschema des Catalogue of Data Sources (WebCDS 2002, Pick 1997), Anwendung. Das Metadatenkonzept für das UIS Baden-Württemberg beschreibt die einzelnen Metadatenfelder und deren Formate in Form einer Tabelle, die auszugsweise in Tabelle 1 wiedergegeben ist.

Das Metadatenkonzept für das UIS Baden-Württemberg unterscheidet mehrere Klassen bzw. Kategorien von Metadaten, die sich aus der Analyse der einzelnen Systeme ableiten. So finden sich beispielsweise in der Kategorie „Quelle“ Angaben zur Herkunft der Daten. Die Kategorien „Raumbezug“ und „Zeitbezug“ beschreiben die geografische bzw. zeitliche Dimension (Datumsangabe nach ISO 8601) der Daten, während unter „Verschlagwortung“ Thesauri (z.B. UBA-Thesaurus) definiert werden, aus denen zutreffende Schlagworte zu entnehmen sind. Das Metadatenkonzept erlaubt aber darüber hinaus auch eine zusätzliche freie Verschlagwortung. In der Kategorie „Administrative Informationen“ aus dem erweiterten Satz können ergänzende Informationen über die Datenobjekte wie etwa das Datenformat oder der Status angegeben werden. Zentral in dieser Kategorie ist das Metadatum „Fachobjekt-

ID“. Mit Hilfe dieses Metadatums, welches ein Umweltobjekt eindeutig kennzeichnet, können Querbezüge auf Objektebene formuliert werden, sofern das Datenobjekt einer Objektart zugeordnet werden kann.

| | Bezeichnung | Format | Beschreibung |
|---|---|---|---|
| **Quelle** | Autor | Freitext | Bezeichnung der erstellenden Stelle des Datenobjektes |
| | Institution | kontrolliertes Vokabular | Typ der erstellenden Institution |
| | Titel | Freitext | Titel der Ressource |
| | URL | Freitext | URL, unter der die Ressource zugreifbar ist |
| **Verschlag-wortung** | Thesaurus-Begriff | Begriff aus Thesaurus | Suchbegriff aus einem Thesaurus (bevorzugt ist der UBA-Thesaurus zu verwenden) |
| | Thesaurus-Name | Freitext | Bezeichnung des verwendeten Thesaurus |
| | Thesaurus-URL | Freitext | URL des verwendeten Thesaurus |
| | Thesaurus-Sprache | ISO 639 Sprachencode | Kürzel der Sprache des verwendeten Thesaurus gemäß ISO Code 639 |
| **Raumbezug** | **Bounding Box** | | |
| | Rechts1; Hoch1 Rechts2; Hoch2 | Zahl | Koordinaten der linken unteren und der rechten oberen Ecke der Bounding Box |
| | Koordinaten-system | kontrolliertes Vokabular | Angabe des für die Bounding Box benutzten Koordinatensystems |
| | **Höhenangabe** | | |
| | Höhe | Zahl | Angabe der Höhe |
| | Maßeinheit | kontrolliertes Vokabular | Maßeinheit, in der die Höhe angegeben ist |
| | **Fachliche Gebietseinheit** | | |
| | Gebiet | Begriff aus Geothesaurus | Beschreibung des Raumbezuges als Begriff aus einem Geothesaurus |
| | Geothesaurus-Name | Freitext | Bezeichnung des verwendeten Geothesaurus |
| | **...** | | |

Tab.1: Ausschnitt aus dem Kernsatz des Metadatenkonzeptes

## 5. Ausblick

Mit dem vorgestellten Metadatenkonzept und dem dort definierten Kernsatz von Metadaten ist nun die erste Voraussetzung für die Integration verschiedenster Datenquellen des UIS in einem einheitlichen Portal geschaffen. Momentan befindet sich das Konzept und seine Einsatzmöglichkeiten in der Verifikationsphase in den beteiligten Behörden. Im Rahmen neu zu entwickelnder Systeme des UIS Baden-Württemberg können die Vorgaben des Metadatenkonzeptes umgehend und ohne größeren Zusatzaufwand berücksichtigt werden.

Wesentlich aufwendiger ist die Umsetzung des Konzeptes in den bereits bestehenden Systemen. Um nicht die bestehenden Datenmodelle im Kern ändern zu müssen, empfiehlt es sich, die betreffenden Systeme mit Schnittstellen zum Metadatenaustausch auszustatten, die dem neuen Metadatenstandard entsprechen. Solche spezifischen Schnittstellen könnten zukünftig auch die Erfassung der Metadaten im UDK wesentlich effizienter gestalten. Ziel ist dabei ein möglichst weitgehender Nachweis der Daten und Dienste des UIS im UDK in geeigneter Granularität. Der UDK erfüllt dabei weiterhin seine Aufgabe als zentrales Nachweissystem unabhängig von dem neuen Umweltportal.

## Literatur

GEIN (2002): http://www.gein.de/.

Mayer-Föll, R., Schulz, K.-P., Keitel, A., Ebel, R., Schultze, A., Dombeck, T., Westbomke, J., Haase, M. (2002): Erstellung eines Meadetnkonzepts für das Umweltinformationssystem Baden-Württemberg (RK UIS/META), Ulm.

Pick, T. (1997): CDS Core Data Model, European Topic Centre on Catalogue of Data Sources, http://www.mu.niedersachsen.de/cds/etc-cds_neu/download-general.html

UDK (2002): http://www.lfu.baden-wuerttemberg.de/udkservlets/UDKServlet.

WebCDS (2002): http://www.mu.niedersachsen.de/system/cds/

XfaWeb (2002): http://www.uvm.baden-wuerttemberg.de/xfaweb/.