Datenintegration im Umweltmonitoring

- Extended Abstract –

Marcus Briesen[1], Claus Hofmann[1], Arnd Otterstätter[1], Ralf Nikolai[1], Michael Mirtl[2], Herbert Schentz[2], Wilhelm Vogel[2], Karin Weber[2]

1. Einleitung

Um den Zustand der Umwelt nachhaltig zu verbessern hat in den vergangenen Jahren die Forderung nach medien- und fachressort-übergreifendem Umweltmonitoring Einfluss auf die internationale und nationale Gesetzgebung, auf Konventionen und Richtlinien genommen. Um ein solches integriertes Umweltmonitoring leisten zu können, sind zum einen Lücken bei der Datenerhebung zu schließen, zum anderen aber auch die vielfältigen bereits existierenden Dateninseln zusammen zu führen, im größeren Kontext zu analysieren und auszuwerten, sowie als Basis für ein übergreifendes Berichtswesen zu verwenden.

Dieser Beitrag greift gezielt den Aspekt der Datenintegration - also das Zusammenführen von Daten aus unterschiedlichen Quellen in heterogenen Formaten, die an verschiedenen Orten vorgehalten werden - im Umweltmonitoring auf. Im folgenden Abschnitt 2 werden die Notwendigkeit und die Anforderungen einer solchen Datenintegration an konkreten Beispielen einer nationalen Umweltbehörde dargestellt. Mögliche Methoden der Datenintegration und Werkzeuge zu deren Umsetzung bilden den Schwerpunkt des Abschnittes 3. Es wird sowohl der Fall der Integration von neu zu erhebenden Datenbeständen als auch die Integration von Alt-Beständen diskutiert. Dazu werden die Werkzeuge MORIS der Umweltbundesamt GmbH Wien und disy Cadenza der disy Informationssysteme GmbH vorgestellt. Schließlich werden Möglichkeiten der Zusammenfassung dieser Methoden einhergehend mit einer Integration der Werkzeuge zu einem umfassenden Werkzeug für die Datenintegration im Bereich des Umweltmonitoring vorgestellt. Wir schließen mit einem Fazit und Ausblick in Abschnitt 4.

2. Notwendigkeit und Anforderungen der Datenintegration

Unsere Umwelt fordert aufgrund der komplexen Zusammenhänge von Einflussfaktoren und Zustandsänderungen per se eine möglichst umfassende, integrierte Betrachtung die Halt weder an Medien (z.B. Luft, Wasser, Boden), noch Fachressort- oder politischen Grenzen macht. Anhand isolierter Messwerte allgemeinere Aussagen zum Zustand der Umwelt zu machen ist unmöglich, da lokale Verbesserungen durchaus mit Verschlechterungen nicht betrachteter Faktoren einhergehen können. Die aktuelle Gesetzgebung fordert daher zu Recht eine stärker integrierte Betrachtung, als dies bisher erfolgt ist. Die folgenden Beispiele aus dem Zuständigkeitsbereich einer nationalen Umweltbehörde illustrieren dies.

[1] disy Informationssysteme GmbH, www.disy.net
[2] Umweltbundesamt GmbH Wien, www.ubavie.gv.at

2.1. Wasserrahmenrichtlinie

2.1.1. Ursprung und generelle Zielsetzung

Am 22.12. 2000 trat die Wasserrahmenrichtlinie der Europäischen Gemeinschaft in Kraft. Die Wasserrahmenrichtlinie versteht sich als grundlegender Beitrag zu einer staats- und länderübergreifenden Harmonisierung des Gewässerschutzes innerhalb der weiter anwachsenden Gemeinschaft und zu einer Verminderung der Gewässerbelastung.

2.1.2. Konkrete Anforderungen

Die Implementierung der Wasserrahmenrichtlinie stellt hohe Anforderungen an ein Informationsmanagement. Daten, die aus völlig unterschiedlichen Fachbereichen stammen wie z.B. sozioökonomische Daten (wie demographische Daten, Informationen über unterschiedliche Wirtschaftsaktivitäten), Emissionsdaten (Punktquellen und diffuse Quellen mit Informationen betreffend Schadstoffmengen und Saisonalitäten) sowie Daten über den Zustand der Umwelt (geologischer Background, Bodenzustand usw.) und deren Belastung (chemischer und ökologischer Zustand von Oberflächengewässern sowie chemischer und mengenmäßiger Zustand von Grundwässern) sind auf Flussgebietsebene zu aggregieren. Hinzu kommen Informationen über den Schutzzustand einzelner Gebiete, d.h. Gesetze, Verordnungen und deren Geltungsbereich.

Diese Informationen müssen einerseits den Partnern im Flusseinzugsgebiet kommuniziert werden können, um die Erstellung eines einheitlichen Flussgebietsbewirtschaftungsplanes zu ermöglichen, andererseits müssen wesentliche Elemente im Rahmen der Erfüllung der Berichtspflichten an die Europäische Kommission nach Brüssel übermittelt werden.

2.2. Integrated Monitoring

2.2.1. Ursprung und generelle Zielsetzung

Die Langzeit-Umweltbeobachtungsgebiete des „Integrated Monitoring“-Programmes der UN-ECE repräsentieren wichtige Naturräume des jeweiligen Staates. In einem europaweiten Netzwerk werden Ökosysteme mit standardisierten Methoden untersucht, um den augenblicklichen Zustand, die Belastungssituation sowie die Stoffflüsse und deren Veränderungen zu dokumentieren. Dabei werden die Stoffeinträge (Schad- und Nährstoffe) durch Luft und Niederschläge gemessen, die Wirkungen und das Verhalten dieser Stoffe im Ökosystem umfassend festgestellt und die Austräge durch Oberflächenwässer und ins Grundwasser erhoben.
Die Ergebnisse dienen u. a. der Überprüfung der Wirksamkeit von Maßnahmen zur Schadstoffreduktion und von bereits unterzeichneten Abkommen, z. B. der Protokolle für die Reduktion der Emissionen von SO_2 und NO_2 sowie des Göteborg-Protokolls aus dem Jahr 2000 (Emissionsobergrenzen für SO_2, NO_x, NH_3 und VOC für die integrierte Behandlung von bodennahem Ozon, Versauerung und Eutrophierung). Sie werden als Grundlage für weitere internationale Protokolle zur europaweiten Reduktion von Schadstoffbelastungen durch weiträumige Schadstoffverfrachtung verwendet.

2.2.2. Konkrete Anforderungen

In diesem Rahmen sind Daten aus einem breiten Spektrum von Fachbereichen in Langzeitreihen in ein generisches Informationssystem abzubilden. Aus diesem Primärdatenpool sind Schnittstellen zur internationalen Projektdatenbank zu bedienen und selektierte Datensets für Auswertungen zur Verfügung zu stellen. Daneben soll

das Informationssystem die operative Projektabwicklung (Codierungssystem...) erleichtern.

3. Methoden und Werkzeuge der Datenintegration

3.3. Integration bestehender Datenbanken

Die Anforderungen der o.g. Gesetzgebungen treffen häufig auf eine Situation, in der Daten für das Umweltmonitoring von verschiedenen Stellen bereits erhoben wurden, eine Integration und gemeinsame Auswertung oder ein übergreifendes Berichtswesen aber bisher nicht stattgefunden hat.

3.3.1. Herausforderungen

Um bisher isoliert erhobene und verwaltete Datenbestände in der Zusammenschau betrachten zu können, gilt es eine Reihe von Herausforderungen zu lösen:

- Überwindung der räumlichen Entfernung
- Schaffung eines einheitlichen Zugangs
- Überwindung der syntaktischen (Formate) und semantischen Heterogenität (Inhalte, z.B. durch Anwendung verschiedener Messverfahren)
- Berücksichtigung des Raumbezugs, der bei beinahe allen Umweltdaten vorhanden ist, durch Integration der entsprechenden Geodatenbestände
- Bedienung von unterschiedlichsten Benutzergruppen mit unterschiedlichen Aufgaben und Rechten
- Verteilte Verantwortung bei der Verfügbarmachung der Daten

3.3.2. Einsatz und Nutzen des Systems disy Cadenza

Aktuelle Internet-Technologien ermöglichen die für den Nutzer transparente Überwindung der räumlichen Entfernung und bilden die Basis moderner Softwaresysteme. disy Cadenza setzt daher von Anfang an auf Internet-Technologien – sowohl für den Internet- als auch den Intranet-Einsatz. Die Informationssichten auf Datenbestände können über ein sogenanntes Repository in einer eigenen, XML-basierten Beschreibungssprache durch die Anwendungsbetreuer definiert werden. Das ermöglicht das gezielte Eingehen auf das Bedürfnis verschiedener Fachnutzer bei gleichzeitiger Unabhängigkeit von der Entwicklungsstelle. Ein datenbankübergreifendes Rechtekonzept, das wiederum Benutzer- und Rollenkonzepte von Datenbanken verwenden kann, ermöglicht es, den Zugriff auf Daten und Dienste selektiv freizugeben. Die Forderung der einheitlichen Benutzungsoberfläche wird durch die Integration von Modulen in den Cadenza Desktop sowie ein generisches Konzept für das Auffinden, Selektieren (Auswählen), Aggregieren, Aufbereiten (Geschäftsgrafiken, Kartendiagramme, Berichte) von Daten umgesetzt. Der modulare Ansatz ermöglicht die Verarbeitung sowohl von Meta-, Sach- als auch von Geodaten durch das „Einklinken“ der entsprechenden Werkzeuge (z.B. Catalog, Selector, Visualizer, Reporter, GISterm) in das System. Während Formate in den Anfragen konvertiert werden können, stellt die semantische Integration weiterhin im wesentlichen eine intellektuelle Aufgabe dar, die durch menschliche Experten gelöst wird. Als zielführend hat sich dabei etwa die Definition von Fachobjekten sowie das Einführen gemeinsamer Schlüssellisten erwiesen.

3.4. Informationssystem für (fast) jeden Umweltdatenbestand

3.4.1. Herausforderungen

Ausgehend von der Problemstellung des Integrated Monitorings musste für Österreich ein Informationssystem geschaffen werden, das im Stande ist, das anfallende äußerst heterogene Datenmaterial gut dokumentiert und kommentiert abzubilden. Es

galt Luft und Klimadaten, Waldzustandserhebungen, Artenerhebungen, Bodendaten, Gewässerdaten, physiologische Daten, ... und bei diesen nicht nur Untersuchungsergebnisse, sondern zu deren Qualitätssicherung auch die Methodik der Erhebung und andere Metadaten zu importieren, speichern, verwalten, exportieren und auszuwerten. Bereits in der Analysephase zeigte sich, dass dazu mehr als ein rein relationales Datenmodell und mehrere Werkzeuge zur Strukturierung und Gliederung der Inhalte notwendig sind.

3.4.2. Einsatz und Nutzen des Systems MORIS

Ein an das objektrelationale Konzept angelehntes Datenmodell erlaubt dem Endnutzer, Klassen und deren Attribute zu definieren und zur Laufzeit auch zu erweitern. Die Basisstruktur ist mit 5 Bereichen fixiert:

- **Objekte –** das, woran untersucht, gemessen, ... wird
- **Parameter –** das, was gemessen, untersucht, ... wird
- **Methoden –** wie untersucht, gemessen, ... wird
- **Datenpunkte –** welche Parameter an welchen Objekten untersucht werden
- **Werte –** Ergebnisse der Untersuchungen
- **Proben –** Ablauf der Untersuchung

Innerhalb dieser Bereiche führt der Endnutzer nähere Definitionen durch. Darüber hinaus kann an Einträge der Basisstruktur Zusatzinformation angefügt werden:

- **Akteure –** Personen und Institutsverzeichnis
- **Archiv –** Dokumente, Dateien, Schlagworte
- **Projekte**

Innerhalb der meisten dieser Bereiche kann der Inhalt polyhierarchisch strukturiert werden und diese Struktur auf Klassenebene vordefiniert werden. Import und Exporttools sowie ein Selektionswerkzeug und eine große Anzahl an Surfhilfen in der Oberfläche erleichtern den Zugang zu den Daten, was bei der Vielfalt der Daten auch essentiell ist. Eine Zeitreihendarstellung, die direkte Anbindung von EXCEL und des Statistikpaketes SPSS, sowie die geplante GIS-Integration sollen den Endnutzer im Datenmanagement entlasten und ihm helfen, sich in der Auswertung auf Inhalte und Fragestellungen zu konzentrieren. Auf Grund der Vielfalt der Anforderungen an die Oberfläche wurde MORIS als Client / Server Tool entworfen, soll nun aber in Teilen mit einer Internetoberfläche versehen werden.

Obwohl MORIS zunächst nur als Werkzeug für das Integrated Monitoring geplant war, zeigte sich auf Grund seiner Flexibilität bald die Eignung für andere Projekte der Umweltbundesamt GmbH: Schwermetalle in Moosen, Regen, Dioxindatenbank.

3.5. Integration von bestehenden Datenbeständen mit neuen Datenbeständen

3.5.1. Herausforderungen

Sowohl für die Erfüllung bestehender als auch zukünftiger regionaler, staatlicher und europäischer Aufgaben im Berichtswesen, als auch für die kontrollierende und / oder wissenschaftliche Analyse der Umweltsituation, stellen sich folgende **allgemeine Aufgaben:**

- Zusammenfassung von umweltrelevanten Daten aus ihren unterschiedlichen bestehenden Informationssystemen
- Einheitliche Präsentation der Daten in Tabellen, Grafiken, Karten u.ä. mehr
- Regelung der Zugriffsrechte zu den integrierten Daten
- Speicherung, Verwaltung und Auswertung von Daten, für die es noch keine Informationssysteme gibt.
- Gliederung des entstehenden heterogenen großen Datenbestandes

Der Wunsch, aus bestehenden bereits recht allgemein gehaltenen Werkzeugen ein ökologisches Universaltool zu entwickeln, das man mit wenigen Adaptierungsarbeiten einsetzen kann und so vermeidet, Räder neu zu erfinden, ist weit verbreitet.

3.5.2. Vorteile der Integration von MORIS und disy Cadenza

Wie bereits dargestellt, hat sich disy Cadenza als gutes Werkzeug für die Integration bestehender Datenbestände, deren einheitliche Präsentation in Tabellen, Grafiken und Karten sowie der Regelung der Zugriffsrechte zu den integrierten Daten erwiesen. MORIS seinerseits zeigt sich als sehr brauchbares Tool, wenn es darum geht, beliebige ökologische Daten, die noch nirgends verwaltet werden, unterzubringen und andererseits das heterogene Datenmaterial nach mehreren Gesichtspunkten zu gliedern, sodass ein beliebiger Nutzer "seinen Weg" zu den Daten findet.

Die Kombination beider Werkzeuge ermöglicht einem Anwender, sowohl alle bisher bereits "untergebrachten" Daten, als auch beliebige "neue ökologische" Datenbestände ohne großen Aufwand zu integrieren und unter einer einheitlichen, polysemantischen Oberfläche zu sehen.

3.5.3. Konzeptuelle Integration der MORIS- und Cadenza-Ansätze

Hinweis: Dieser Abschnitt wird für die Langfassung des Beitrages ergänzt.

4. Fazit und Ausblick

Es gibt sicherlich noch etliche Anforderungen, die immer wieder an Informationssysteme aus dem ökologischen Bereich gestellt werden, wie Unterstützung von Harmonisierungsprozessen, Integration von Multimedia-Darstellungen für die „Allgemeinheit" u.v.a., die allgemein gelöst werden sollen, doch hat man mit der Kombination von disy Cadenza und MORIS bereits einen Werkzeugsatz in der Hand, den man bei sehr vielen Problemstellungen von der Stange nehmen und einsetzen kann.

5. Literaturverzeichnis

Hinweis: Das Literaturverzeichnis wird – wie Abbildungen und weitere Ausführungen - für die Langfassung des Beitrages ergänzt.