# Metainformation zu Meßdaten aus der Sicht von Simulationsmodellen

Zwischenergebnisse aus dem UFIS Projekt zur Entwicklung einer Metadatenbank über Daten

Ekkehard Ernst, Christoph Herderich, Ulrike Weisemann, Walter Sinowski, Roman Lenz, Michael Knorrenschild[1]

## Einleitung

Das Umweltforschungs-Informationssystem (UFIS) (Lenz u.a. 1994) soll Informationen zur ökologischen Forschung, insbesondere über Modelle und dazugehörige Daten bereitstellen. Diese Information umfaßt eine standardisierte Dokumentation von Modellen und von solchen Datensätzen, die für die ökologische Modellierung von Bedeutung sind.
Nutzergruppen sind:

- Wissenschaftler, die in fachübergreifenden, interdisziplinären Programmen (z.B. GAIM, GCTE, IGBP) aktiv sind,
- Modellierer, die Zugang zu Informationen außerhalb ihres eigenen Spezialgebietes suchen,
- Planungs- und Ingenieurbüros, die die Daten und Ergebnisse der ökologischen Forschung nutzen wollen,
- forschungsfördernde Institutionen, die ihre Fördergelder effektiv und gezielt dort ansetzen wollen, wo noch Bedarf und Lücken innerhalb der ökologischen Forschung bestehen.

Seit Februar 1995 ist die Datenbank über Modelle, die Modelldokumentationen enthält, öffentlich zugänglich und wird laufend um weitere Modelle erweitert. Diese Datenbank wird in benutzerfreundlichen Oberflächen, über Telnet oder World Wide Web (URL: http://www.gsf.de/UFIS/ufis) abfragbar, angeboten.
Analog zur Metainformation über Modelle soll eine Datenbank zur Metainformation über Daten, die vor dem Modellauf eingegeben werden bzw. nach dem Modelllauf ausgegeben werden, aufgebaut werden.
Die Datenbank zur Metainformation über Daten wird benötigt, um

- mögliche Modellverknüpfungen aufzuzeigen,
- eine Analyse des Datenbedarfs und Datenoutputs von Modellen zu ermöglichen,

---

[1]GSF-Forschungszentrum für Umwelt und Gesundheit, Projektgruppe Umweltgefährdungspotentiale von Chemikalien, Neuherberg, 85764 Oberschleißheim

- die Verwendbarkeit der in Meßdatenbanken gespeicherten Datensätze als Input von Modellen zu prüfen,
- notwendige Information für den Zugriff auf Meßdatenbanken zu liefern.

Dieser Beitrag handelt von den bisherigen Überlegungen und Konzepten zur Entwicklung dieses Metainformationssystems über Daten.

## Metainformation

Eine verbreitete Definition versteht Metainformation als die Menge an Information, die notwendig ist, einen Datenbestand zu finden und in seiner Bedeutung zu erfassen. Hierbei ist die Fragestellung, unter der der Datenbestand betrachtet wird, von zentraler Bedeutung. Metainformation ist also immer kontextabhängig.
Das Thema Metainformation in Umweltinformationssystemen ist ein aktives Forschungsfeld, wie z.B. auf den Vorgängertagungen mit mehreren Vorträgen (Schütz u. a. 1993; Lessing u. a. 1994; Greve u. a. 1994; Denzer u. a. 1994) dokumentiert wurde.

## Standardisierung der Metainformation in UFIS

Ein System wie UFIS faßt Daten aus verschiedenen Modellen und Meßdatenbanken zusammen. Da jedes Modell und jede Meßdatenbank Daten in einem anderen Kontext sieht, ist die dazugehörige Metainformation meist sehr unterschiedlich und muß erst für UFIS in eine standardisierte Metainformation übersetzt werden.
Ein bereits in UFIS integriertes Standardisierungsinstrument für die Modelldatenbank ist ein Suchhhilfesystem, das auf einem systemökologischen Thesaurus für Modelle basiert (Springstubbe 1993). Die Hauptebenen dieses Klassifikationssystems sind Organisationsebene, Umweltmedium, geographische Lage, FAO Bodenklassifikation und Klima-Wasserhaushalt. Es ermöglicht eine allgemeingültige, für alle Modelle einheitliche Beschreibung der umweltrelevanten Aspekte eines Modells.
Zur Dokumentation von Modellen wurde ein UFIS-Dokumentationsschema entwickelt, das sich als allgemeiner Standard eignet. Dieses Dokumentationsschema definiert bereits aus der Sicht von Modellen, wie Größen beschrieben werden und damit, wie die Metainformation zu Daten aussehen soll. Dabei spielen die zeitlichen und räumlichen Geltungsbereiche eine große Rolle.

## Metainformation für Daten im UFIS-Dokumentationsschema

Das UFIS-Dokumentationsschema enthält für Größen folgende Metainformation:
Name der Größe, Kurzbezeichnung, Kurzbeschreibung, Typ (Skalar, Vektor, Matrix, Zufallsgröße), Dimension im Fall von Vektor und Matrix, Einheit, Wertetyp,

Intervall der sinnvollen Wertebereiche, Intervall der verwendeten Wertebereiche, Quelle der verwendeten Werte, zeitliche Aggregation, räumliche Aggregation, ergänzende Bemerkungen.
Hierbei handelt es sich um eine vor allem quantitative Beschreibung einer meist skalaren Größe. Im Fall von vektoriellen und matrixartigen Größen sollen die nachfolgenden Angaben wie "Intervall der sinnvollen Wertebereiche" (z.B. pH-Wert zwischen 3 und 5) für alle Komponenten des Vektors bzw. der Matrix in gleicher Weise gelten.
Qualitative Angaben, wie z.B. methodische Beschreibungen, sind in diesem Schema nur in den Freitextfeldern (Feld: Kurzbeschreibung) möglich.
Die Größen eines Modells können in fünf Kategorien unterteilt werden:

- Die Inputdaten, die im allgemeinen zeitabhängig sind,
- die Parameter, die während der Modellentwicklung kalibriert werden und während des Modellaufs konstant bleiben,
- die Anfangswerte als Startwerte der Simulation,
- die internen Variablen und
- die Outputdaten, die in der Regel zeitabhängig sind.

Aus der Modellsicht wird im UFIS-Dokumentationsschema nur zwischen dem Input oder Datenbedarf eines Modells, der sich aus Inputdaten, Parametern und Anfangswerten zusammensetzt, den internen Variablen und dem Output, der aus den Outputdaten besteht, unterschieden.

## Erweiterung des UFIS-Schemas

Da der Output der Modelle in der gleichen Art beschrieben ist wie der Input, entstehen neue Datensätze, die die Menge des zur Verfügung stehenden Inputs für nachfolgende Modellanwendungen erweitern. Theoretisch entsteht so aus einer gegebenen Menge von Daten und Modellen die Menge der aus diesen Daten und Modellen ableitbaren Daten und die Menge der damit verknüpfbaren Modelle.
Für die Analyse möglicher Modellverknüpfungen allerdings ist dann die Unterscheidung zwischen der zeitlichen Auflösung des Inputs und der zeitlichen Auflösung des Outputs sowie der räumlichen Auflösung des Inputs und der räumlichen Auflösung des Outputs notwendig.
Die Bedeutung der Kenntnis der räumlichen Auflösung soll folgendes Beispiel verdeutlichen:
Ein Modell zum vertikalen Transport von Sickerwasser benötigt die ungesättigte Leitfähigkeitskurve in einer räumlichen Auflösung von 10 cm-Schritten in Richtung der Bodentiefe. Durchschnittswerte über die durchwurzelte Bodentiefe von 1 m hätten somit, bei sonst gleicher Meßmethodik, eine zu geringe räumliche Auflösung.

Ebenso kann es beispielsweise für ein Modell wichtig sein, daß der Niederschlag für jeden einzelnen Tag eines Jahres bekannt ist und es nicht genügt, daß der Niederschlag über ein Jahr gesehen zur Verfügung steht, aus dem nur der Durchschnittswert pro Tag errechnet werden kann.
Die hier beschriebene Erweiterung des UFIS-Schemas zielt einmal darauf ab, die Navigation auf der Menge der Größen zu verbessern und zum Anderen wird versucht, die Unterschiede von Modellen und ihrer Größen bezüglich ihrer räumlichen und zeitlichen Gültigkeit bzw. ihrer räumlichen und zeitlichen Auflösung besser zu fassen. Dabei muß berücksichtigt werden, daß der Datenbedarf eines Modells sowie der Output sich innerhalb gewisser Spannweiten der räumlichen und zeitlichen Auflösung bewegen.
Für die Analyse des Datenbedarfs eines Modells bzw. für mögliche Modellverknüpfungen wird also zusätzliche Metainformation über Modellgrößen benötigt. Diese muß je nach Modell unterschiedlich detailliert sein. Um auf diese unterschiedlichen Detaillierungsgrade reagieren zu können, wird für das Metainformationssystem über Daten eine Darstellung der Metainformation in Ebenen entwickelt, in denen der Informationsgehalt jeweils zunimmt. Zentrale Bedeutung hat dabei die Bezeichnungsebene zur Navigation auf der Menge der Größen.
Es lassen sich so je nach Verfügbarkeit von Metainformation Aussagen von z.B. folgender Art speichern: "Es wird die Größe Wassergehalt des Bodens benötigt" in mehreren Stufen bis zu einer detaillierten Aussage wie: "Die Größe Wassergehalt des Bodens muß in der Einheit Vol.-% mit der Methode nach Anonymous für die Bodentiefe 0-30 cm über einen Zeitraum von 3 Jahren ereignisabhängig (mind. 2 Messungen/Jahr) in einer räumlichen Dichte von 10 regelmäßig verteilten Messungen/Hektar verfügbar sein."
Desweiteren sind bei der Beschreibung der Größen die Standardisierungsansätze bisher noch nicht ausreichend, um zwei Größen als synonym (oder nicht) zu erkennen.

## Realisierungsansätze für eine bessere Standardisierung der Größen

Der oben aufgezeigte Bedarf einer standardisierten Größenbeschreibung läßt sich nur über eine technische Realisierung von dazu geeigneten Werkzeugen abdecken.
Deshalb ist geplant, folgende Instrumentarien in UFIS zu implementieren:

- Eine standardisierte Darstellung der Einheiten, um einen Vergleich der Daten zu ermöglichen.
- Synonymtabellen, um voneinander abweichende Größenbezeichnungen, die sich auf ein und dieselbe Modellgröße beziehen, auf einen in der gesamten Datenbank auftretenden einheitlichen Namen zurückzuführen.
- Ausbau des Thesaurus für Modelle, so daß er auch als Suchhilfe für die Modellgrößen und für die Modellgleichungen verwendet werden kann.

Erste Konzepte zur Umrechnung der Einheiten auf einen standardisierten Eintrag und zu den Synonymtabellen wurden von Weisemann (1994; 1995) entwickelt.
Synonymtabellen beispielsweise sind nötig, um bei der Verwendung von verschiedenen Namen für die gleichen Größen, wie z.B. pH-Wert und Säuregrad, zu erreichen, daß die Datenbank bei der Eingabe solcher synonymer Namen als Suchbegriffe gleiche Resultate liefert.

## Besonderheiten bei der Beschreibung der Metainformation von Daten aus Meßdatenbanken

Da Meßdaten häufig als Input von Modellen dienen, soll in die UFIS-Datenbank auch Metainformation von Daten aus Meßdatenbanken aufgenommen werden. Dazu ist innerhalb UFIS eine identische Beschreibung der Metainformation des Datenbedarfs von Modellen und der Metainformation der Daten, die aus Meßdatenbanken stammen, notwendig.
Allerdings müssen Größen aus Meßdatenbanken über zusätzliche navigatorische Metainformation beschrieben werden, um den Zugriff zu gewährleisten. Information über die Struktur des Datenbanksystems, der Zugriffsrechte, den Ansprechpartnern und der zur Verfügung stehenden Metainformationsverwaltung einer Meßdatenbank muß verfügbar sein.
Da die Größen in einer Meßdatenbank nicht notwendigerweise unter dem Gesichtspunkt gesehen werden können, als Eingangsgrößen für ein Simulationsmodell zu dienen, kann im allgemeinen nicht davon ausgegangen werden, daß die zu dieser Größe verfügbare Metainformation in Informationsgehalt und -struktur kompatibel zu der für UFIS benötigten Metainformation ist.
Probleme, die bei der Extraktion von Metainformation aus Meßdatenbanken auftreten können, sind:

- Es werden keine oder nur ungenügende Metainformation mit den Daten verwaltet.
- Die Meßdatenbank ist auf eine deutliche andere Sicht der Daten und ihrer Metainformation konzipiert.
- Die Struktur der Daten- und Metainformationsbeschreibung kann auch von Meßdatenbank zu Meßdatenbank sehr verschieden sein.
- Es gibt keine Möglichkeit mit Hilfe von Suchabfragen die Metainformation von Daten zu finden.
- Die Meßdatenbank ist nicht für einen öffentlichen Zugang konzipiert.

Dies bedeutet in der Regel einen hohen Aufwand an meist fachlicher Übersetzungsarbeit bei der Übertragung der verfügbaren Metainformation einer Meßdatenbank in die für UFIS geeignete Form. Konzepte, diesen Aufwand über geeignete Softwareunterstützung zu erleichtern, sind bisher noch wenig vorhanden.

## Literatur


Denzer R./Güttler R. (1994): Integration von Umweltdaten, in: Page B./Hilty L.M. (Hrsg.), Umweltinformatik, Informatikmethoden für Umweltschutz und Umweltforschung, Handbuch der Informatik, Band 13.3, Oldenbourg/München/Wien, S. 227-240.

Greve K./Häuslein A. (1994): Metainformation in Umweltinformationssystemen, in: Hilty u. a., Informatik für den Umweltschutz 8. Symposium Hamburg 1994, Band I, Metropolis, Marburg 1994, S. 169-178.

Lenz R./Knorrenschild M./Herderich C./Springstubbe O./Forster E./Benz J./Assoff W./Windhorst W. (1994): An information system of ecological models, GSF-Bericht 27/94, ISSN 0721-1694, 92 S.

Lessing H./Schütz T. (1994): Der Umwelt-Datenkatalog als Instrument zur Steuerung von Informationsflüssen, in: Hilty u. a., Informatik für den Umweltschutz 8. Symposium Hamburg 1994, Band I, Metropolis, Marburg 1994, S. 159-168.

Schütz T./Lessing H. (1993): Metainformation von Umwelt-Datenobjekten - Zum Datenmodell des Umwelt-Datenkatalogs Niedersachsens, in: Jaeschke u. a., Informatik für den Umweltschutz 7. Symposium Ulm, 31.3. - 2.4. 1993, Springer, Berlin/Heidelberg 1993, S. 19-28.

Springstubbe O. (1993): Grundzüge der Klassifikation und systematische Beschreibung ökologischer Systeme als Gliederungsprinzip für ein Informationssystem ökologisch relevanter Daten und Modelle, Diplomarbeit, TU München-Weihenstephan, Lehrstuhl für Landschaftsökologie, 146 S.

Weisemann U. (1994): Entwurf eines Datenbankschemas zur Dokumentation des Datenbedarfs und des Outputs von Simulationsmodellen in der Ökologie, Studienarbeit, TU Berlin, Inst. f. Angew. Informatik, Computergestützte Informationssysteme, 61 S.

Weisemann U. (1995): Implementierung eines Datenbankschemas zur Datendokumentation von Simulationsmodellen in der Ökolgie, Diplomarbeit, TU Berlin, Inst. f. Angew. Informatik, Computergestützte Informationssysteme, 33 S. + Anhang.