Metainformation bei Umweltinformationssystemen und deren Transfer

E. Ernst, W. Sinowski, M. Knorrenschild[1]

1. Einleitung

Die Thematik Metainformation in Umweltinformationssystemen ist ein aktives und aktuelles Forschungsfeld, wie u.a. auf den Vorgängertagungen (Schütz u.a. 1993, Greve u.a. 1994, Denzer u.a. 1994, Güttler u.a. 1995, Ernst u.a. 1995) mit zahlreichen Vorträgen dokumentiert wurde. Dieser Themenbereich wird in der Zukunft aufgrund der ständig wachsenden Flut an Information, die in zunehmendem Maße nur mit Hilfe der entsprechenden Metainformation weiter verarbeitet werden kann, noch weiter an Bedeutung gewinnen.
Die Entwicklung von Informationssystemen, die Metainformation bereitstellen, erzeugt dabei eigene, spezifische Problemfelder. Diese bieten der Informatik neue Herausforderungen und Problemstellungen, deren Lösungen von wachsender Bedeutung sind.

Da Metainformation stark kontextabhängig ist, ist der Zusammenhang, in dem ein Informationssystem die Information sieht, von entscheidender Bedeutung für die Art und Struktur der Metainformation, die in das System aufgenommen wird. So ist es vorstellbar, daß die Metainformation der gleichen Daten in verschiedenen Informationssystemen grundsätzlich verschieden ist.
Die Auswahl und das Design der Metainformation muß deshalb für jedes Informationssystem neu durchdacht werden und gestaltet sich oft recht schwierig, da meist nicht alle Blickwinkel, aus der die Information gesehen werden kann, berücksichtigt werden können. Eine gute Lösung dieses Problems ist dann besonders wichtig, wenn mehrere Metainformationssysteme miteinander verknüpft werden.

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

Eine weitere wichtige Fragestellung ist, wo die gewünschte Metainformation gefunden werden kann und wie sie in das System gelangt. Meist wird diese Information direkt vom Datenerheber erfragt. Je höher die Integrationsebene eines Metainformationssystems allerdings ist, umso häufiger müssen die Metadaten aus der Datenbank eines schon bestehenden Metainformationssystems entnommen werden. Beim dafür nötigen Transfer der Daten von einem Metainformationssystem in ein anderes treten häufig folgende Probleme auf:

- Es werden dort keine oder nur ungenügende Metainformation verwaltet.
- Die Sicht, aus der die Information betrachtet wird, ist eine andere und die Metainformation, die verfügbar ist, ist deshalb nicht die gesuchte.
- Die gesuchte Metainformation ist zwar vorhanden, wird aber nicht in der gewünschten Form bereitgehalten.
- Die gesuchte Metainformation ist ohne Befragen des zuständigen Datenbankadministrators schwer auffindbar oder gar nicht erhältlich, da die Datenbank nicht für den öffentlichen Zugang konzipiert wurde.

Für den Erfolg eines Informationssystems ist allerdings nicht nur entscheidend, welche Information in der zugehörigen Datenbank gespeichert ist, sondern auch die Bedienbarkeit der Oberfläche und der Abfrageapplikationen des Systems. Da die Zahl der Informationssysteme ständig zunimmt und damit auch die Vielfalt der Oberflächen wächst, über die diese präsentiert werden, wird es für die Benutzer immer schwieriger, sich in den vielen, verschiedenen Systemen zurechtzufinden.
Mit der Einführung des World Wide Web (WWW) wurde eine einheitliche Schnittstelle zu Informationen geschaffen. Es läßt sich damit leicht von Informationssystem zu Informationssystem innerhalb des WWW navigieren und die Benutzer treffen jeweils eine gleichermaßen zu bedienende Oberfläche an. Das WWW ist daher in besonderem Maße für den Zugang zu Metainformation aus entsprechenden Informationssystemen geeignet. Es ist also sinnvoll,

öffentlich zugängliche Inhalte von Datenbanken als HTML(HyperText Markup Language)-Dokumente auf dem WWW anzubieten, auch wenn dies mit einem Mehraufwand verbunden ist, da meist schon eine andere Oberfläche zur Verfügung steht.
Weiter bietet es sich an, auch beim Transfer von verfügbaren Daten zu den Informationssystemen in zunehmendem Maße sich dieser Schnittstelle zu bedienen, da dann stets eine vertraute Umgebung angetroffen wird.

Neben Lösungsvorschlägen zu den oben erwähnten Problemen wird im folgenden ein Beispiel eines solchen Metainformationssystems, das Umweltforschungs-Informationssystem UFIS, vorgestellt.

2. Transfer von Metainformation

Eine Standardisierung von Metainformation gestaltet sich - auch wenn große Anstrengungen dahingehend unternommen werden - schwierig, da andere Fragestellungen häufig auch ein anderes Design der Metainformation fordern und daher die Administratoren bestehender Datenbanken oft nicht bereit sind, ihr Metainformationsschema einem Standard anzupassen. So sind Lösungen der oben aufgeführten Probleme beim Transfer von Metainformation von einem System zum anderen von großem Interesse.
Ein einfacher Lösungsvorschlag besteht darin, einen Fragebogen zu erstellen, in den der Datenerheber die Metainformation in der gewünschten Form einträgt. Dieser Fragebogen wird als HTML-Formular über WWW ausgefüllt und der Inhalt dann direkt von diesem Fragebogen (evtl. erst nach einer fachlichen Überprüfung) in die entsprechende Datenbank übertragen. Dies vermeidet nochmalige Tipparbeit und hat, da der Fragebogen über WWW angeboten wird, den Vorteil, daß der Ausfüller eine vertraute Oberfläche geboten bekommt und der Fragebogen jederzeit verfügbar ist. Hierbei wird die Übertragung in das entsprechende Metainformations-Schema vom Ausfüller des Fragebogens vollzogen. Daher sollte dieser mit den Daten und der zugehörigen Metainformation gut vertraut sein. Denkbar ist hierfür neben dem Datenerheber auch der Administrator einer Datenbank, in der die Metainfor-

mation gespeichert ist. Auf diese Weise ist auch Metainformation erhältlich, die noch in kein Informationssystem aufgenommen war. Allerdings müssen bei dieser Methode für alle Datensätze eigene Fragebögen ausgefüllt werden und alle Informationen per Hand in den Fragebogen eingegeben werden.
Dies vermeidet der folgende Lösungsvorschlag, der aber ebenfalls nicht auf die Mitarbeit beispielsweise des entsprechenden Datenbankadministrators verzichten kann und darauf angewiesen ist, daß die Information verfügbar und kompatibel ist.
Hierbei ist es nötig (durch Ausfüllen eines Fragebogens), vom Datenbankadministrator der betreffenden Datenbank zu erfahren, wo sich in der Datenbank die benötigte Metainformation befindet, wie sie beschrieben ist und wie auf sie zugegriffen werden kann. Diese Information wird dann verwendet, um einen automatisierten Zugriff auf die Metadaten zu ermöglichen. So können diese dann direkt von der einen in die andere Datenbank übertragen werden und das mühsame Ausfüllen eines Datenfragebogens für jeden einzelnen Datensatz entfällt. Dieses Verfahren hat zusätzlich den Vorteil, daß eine Aktualisierung auf dem selben Wege leicht möglich ist.
Der Aufwand dafür ist allerdings nicht gering und die zugehörigen Programme müssen für jede Datenbank, aus der Metainformation entnommen werden soll, einzeln angepaßt werden. Deshalb lohnt sich dies nur für Datenbanken, von denen aus ein großer Datentransfer stattfinden soll.
Das Prinzip dieser Lösungsvorschläge ist zwar nicht daran gebunden, die jeweiligen Fragebögen als HTML-Formulare anzubieten, aber angesichts dessen, daß sich in Zukunft wohl Informationssysteme zunehmend über WWW präsentieren werden und auch immer häufiger direkte Zugriffe über diese Schnittstelle auf die Datenbanken ermöglicht werden, ist diese einheitliche Oberfläche sehr vorteilhaft.

Unter gewissen Voraussetzungen an die Struktur der Datenbanken, zwischen denen ein Datenaustausch stattfinden soll, ist auch eine Kommunikation zwischen Umweltinformationssystemen über das Integrationswerkzeug SIRIUS (Güttler u.a. 1995) möglich.

3. Das Umweltforschungs-Informationssystem UFIS

Das Umweltforschungs-Informationssystem (UFIS) (Lenz u.a. 1994), das 1992 vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) ins Leben gerufen wurde, stellt Informationen zur ökologischen Forschung über Modelle und dazugehörige Daten bereit und ist seit Februar 1995 über WWW öffentlich zugänglich (URL: http://www.gsf.de/UFIS/ufis). Zum Einen soll es dem BMBF und anderen forschungsfördernden Institutionen dazu dienen, einen Überblick über die ökologische Forschung zu erhalten, um Fördergelder effektiv und gezielt dort einsetzen zu können, wo noch Bedarf und Lücken innerhalb der ökologischen Forschung bestehen. Zum Anderen kann UFIS von Wissenschaftlern, Modellierern und weiteren Interessenten als Informationsquelle über die ökologische Forschung genutzt werden.
UFIS ist ein reines Metainformationssystem, das aus zwei INGRES-Datenbankteilen besteht, der Datenbank über Metainformation zu Modellen aus der ökologischen Forschung und der Datenbank über Metainformation zu Daten. D.h., es sind weder die Modelle noch die eigentlichen Daten in UFIS vorhanden, sondern lediglich Metainformation dazu. Diese beiden Datenbankteile sind eng miteinander verknüpft und mit komfortablen Suchmöglichkeiten versehen. Dabei ist es mit Hilfe von UFIS möglich, über die Metainformation zu Modellen ein für eine bestimmte Problemstellung geeignetes Modell zu finden sowie mit Hilfe der Metainformation zu Daten die für ein Modell benötigten Datensätze. Desweiteren können mögliche Modellkopplungen festgestellt werden und allgemein nach Datensätzen über Schlagworte gesucht werden. Hierbei spielt die Art und Struktur der bereitgestellten Metainformation für eine erfolgreiche Suche eine entscheidende Rolle.

4. Problemlösungen innerhalb UFIS

Die Metainformation zu Modellen wird von UFIS im allgemeinen direkt von den Modellierern bezogen. Dafür wurde ein standardisierter Fragebogen entwickelt, der über WWW ausgefüllt werden kann. Dieser ist recht ausführlich gehalten und die Auswahl der Metainformation zielt darauf ab, die Modelldo-

kumentation des Modellierers zu ersetzen und künftige Modelle aus der ökologischen Forschung in eben dieser Form zu dokumentieren.
Der Inhalt des Fragebogens, der die Bereiche

- Allgemeine Information zum Modell
- Mathematische Beschreibung des Modells
- Technische Information

umfaßt, wird mittels eines CGI-Skriptes, das in Perl geschrieben ist, per Email direkt an UFIS geschickt und dann (nach einer fachlichen Überprüfung) in die INGRES-Datenbank von UFIS übertragen. Der Ausfüller des Fragebogens erhält einen ausgefüllten Fragebogen als Dokumentation zu seinem Modell zugesandt. Eine Automatisierung des Transfers würde hierbei nur Sinn machen, wenn die Information aus einer Datenbank entnommen werden kann.

Bei der Auswahl der Metainformation zu Daten wurde in erster Linie berücksichtigt, daß diese als Variablen für Modelle verwendet werden können. Aber das Konzept zielt auch darauf ab, daß losgelöst von einer Verwendung für ein Modell nach Datensätzen recherchiert werden kann. Die Auswahl der Metainformation ist in Abbildung 1 ersichtlich.

Für die Metadatenbank der Variablen gestaltet sich der Zugang zur benötigten Metainformation besonders schwierig. Die Anzahl der Variablensätze ist ungleich höher als die Anzahl der Modelle, die in der UFIS-Datenbank gespeichert werden. Bei kleineren Datenmengen, wie z.B. den Kalibrierdatensätzen eines Modells ist es sinnvoll, einen Datenfragebogen zu benutzen, in den die Metadaten zu den Datensätzen eingetragen werden und analog zum Modellfragebogen über ein CGI-Skript an UFIS geschickt werden; sie können dann in die UFIS-Datenbank übertragen werden.

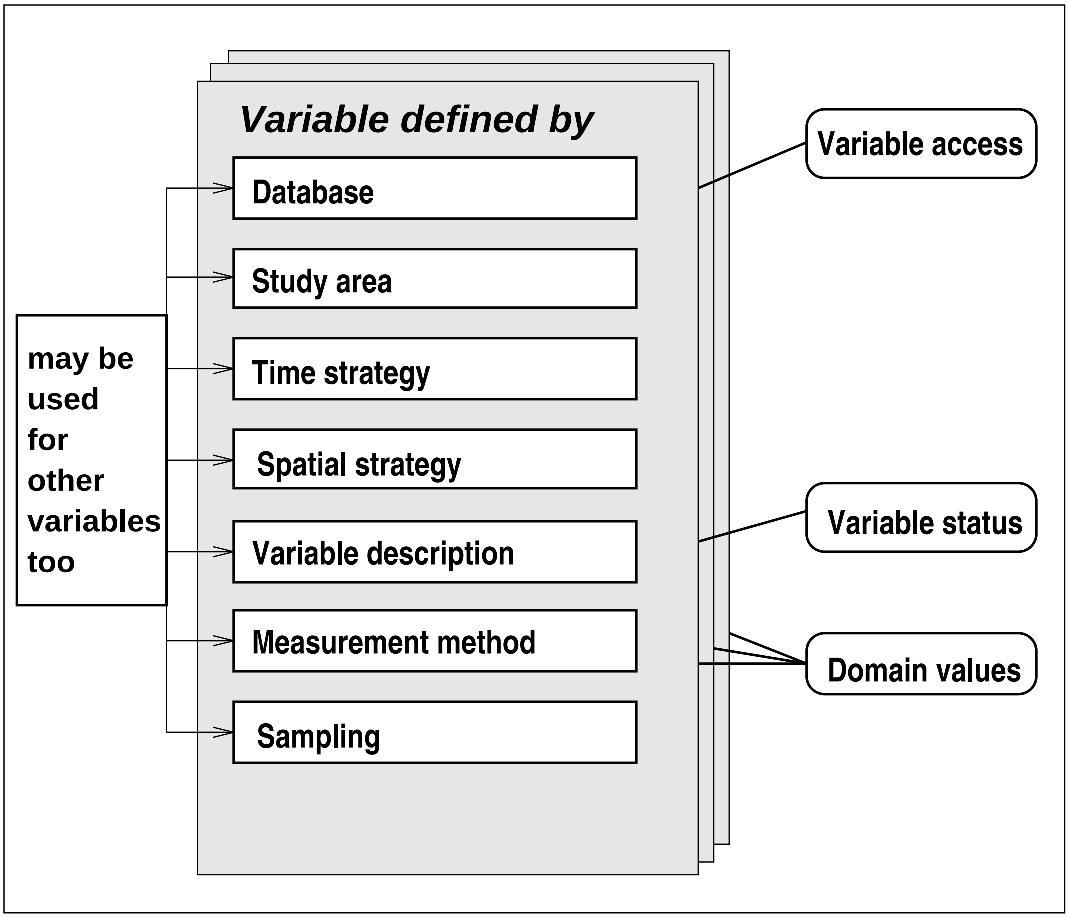

Abbildung 1: Metainformation zu Daten innerhalb UFIS

Bei einer größeren Meßdatenbank, aus der eine große Zahl von Datensätzen entnommen werden soll, ist diese Methode kaum praktikabel, da für jede Variable ein eigener Fragebogen ausgefüllt werden müßte. Hier wird der Transfer der Metainformation folgendermaßen vollzogen:
In einer solchen Datenbank ist meist die Metainformation vieler Variablen gleichermaßen abgespeichert. Daher wird vom Datenbankadministrator mittels eines Fragebogens erfragt, wo und wie welche Metainformation auffindbar ist. In einer Tabelle wird diese Information dann gespeichert und mit einem Programm darauf zugegriffen. Damit kann dann nacheinander für alle Variablen die gesuchte Metainformation an der entsprechenden Stelle der Meßdatenbank geholt werden und in die UFIS-Datenbank übertragen werden. Dieser automatisierte Datentransfer hat zudem den Vorteil, daß eine Aktualisierung der

Daten leicht auf dem selben Wege vollzogen werden kann. Dies geht allerdings nur mit Metainformation, die in der Datenbank gespeichert ist. Metadaten, die nicht gespeichert sind, können, falls sie überhaupt bekannt sind, nur über den Datenfragebogen erfragt werden.

5. Schlußfolgerungen

Abschließend läßt sich feststellen, daß die Art, Menge und Struktur der Metainformation abhängig vom Charakter der Zielgruppe gewählt werden muß und die Form der Erhebung der Metadaten wiederum dem angepaßt werden muß. Als Ein- und Ausgabeschnittstelle von Informationssystemen eignet sich das WWW in besonderem Maße, da eine standardisierte, benutzerfreundliche Oberfläche verwendet werden kann, die einem großen Nutzerkreis zugänglich ist.

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.

Ernst E./Herderich C./Weisemann U./Sinowski W./Lenz R./Knorrenschild M. (1995): Metainformation zu Meßdaten aus der Sicht von Simulationsmodellen - Zwischenergebnisse aus dem UFIS Projekt zur Entwicklung einer Metadatenbank über Daten, in: Kremers u.a. "Raum und Zeit in Umweltinformationssystemen", 9. Internat. Symposium on Computer Science in Environmental Protection CSEP '95, Berlin (vol. 1), Metropolis, Marburg, S. 408-413.

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

Güttler R./Denzer. R. (1995): SIRIUS - Ein Integrationswerkzeug für inselhafte Umweltinformation, in: Kremers u.a. "Raum und Zeit in Umweltinformationssystemen", 9. Internat. Symposium on Computer Science in Environmental Protection CSEP '95, Berlin (vol. 1), Metropolis, Marburg, S. 422-430.

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.

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