

Workshop des Arbeitskreis 'Umweltdatenbanken' 10./11. Juni 1999 in Karlsruhe
Einreichung von erweiterten Kurzfassungen im Umfang von 2-4 Seiten

Thomas Bandholtz (thomas@sema.de)

1 Die verteilte Suche in GEIN 2000 (G2K)

Die Ist-Aufnahme bei den G2K-Informationsanbietern zeigte, daß bereits heute ein umfangreiches und qualitativ ernst zu nehmendes öffentliches *Internet-Angebot* für die hier untersuchte Umweltinformation besteht. Dies wird in der Projektlaufzeit, insbesondere bis zur EXPO 2000, in noch kaum vorhersehbarem Maße ausgedehnt und qualitativ restrukturiert werden. Da auch der Anteil der Internet-Nutzer bei den Anwendern entsprechend steigt, ist der direkte Zugang des "interessierten Laien" zur Umweltinformation dabei, auf einem unvorhergesehenen Weg und in unvorhergesehenem Ausmaß wahr zu werden.

Dieser Zugang zu dieser Information lief bisher an den *Metadatenkatalogen* vorbei. Mit dem UDK 4.0 sind Internet-Verweise bewußt in das Datenmodell (und die Interaktion) aufgenommen worden. Bisher bieten drei Länder den UDK auf ihren öffentlichen Web-Seiten an und etablieren ihn als lokale "Suchmaschine". (Am konsequentesten scheint dies Karlsruhe (LfU) zu betreiben, und das übrigens mit der Version 3.02. Hier gab es auch eine Anwenderbefragung zur Beurteilung des UDK).

Doch der Strukturbaum des UDK und die "Baumstruktur" von HTML-Dateien in WebSites sprechen nicht dieselbe Sprache. Der UDK ignoriert die Suchmuster des Internet und wird im selben Maße vom Internet ignoriert. Dies kann jeder leicht überprüfen, indem er mit einer gängigen Suchmaschine z.B. nach Ozon sucht und allein darauf achtet, ob und an welcher Stelle Bund- oder Länderbehörden in der Trefferliste auftauchen. Das Ergebnis ist eher deprimierend. Ein UDK ist in einer solchen Trefferliste noch nie aufgetaucht.

Das hat seinen Grund: Allgemein geht das Internet vom Zugang zu Informationen aus und versucht diese zu strukturieren, während der UDK von einer Struktur ausgeht und versucht, den Zugang zu ihren Elementen herzustellen.

Der UDK hat ja in den frühen 90ern als reiner Katalog begonnen, der sich in seiner Struktur an die Organigramme von Ministerien orientierte. 1993/4 bemühte er sich um ein "Kommunikationsmodul", noch auf LAN/WAN-Technik basierend, mit dem parametrisierbare Abfragen auf die Fachdatenbanken ermöglicht werden sollten - Visionen, aus denen geradewegs die aktuellen Erwartungen an ein "Z39.50-Profil" hervorgegangen zu sein scheinen (die allerdings heute nicht vom UDK vorgetragen werden).

Der *WWW*-UDK (seit Version 3.0) beschränkt sich auf eine Internet-*Oberfläche*. Dahinter liegt derselbe UDK. Mit der Version 4 werden erstmals

ausdrücklich URLs (Unified Resource Locator, in etwa = "Internetadressen") als Verweise zu UDK-Objekten zugelassen.

Es gibt allerdings auch heute und in Zukunft Umweltinformation, die gar nicht öffentlich zugänglich gemacht wird, sei es aus Personenschutzgründen, sei es wegen der Priorität des Betriebsgeheimnisses (Einleiterdaten), sei es, weil sie nicht hinreichend validiert vorliegt. Hier ist die Position des "Users", daß er erfahren möchte, welche Informationen es gibt und wen er danach befragen kann. Es ist schließlich nicht ausgeschlossen, daß er solche Daten im einzelnen einsehen kann, wenn er ein berechtigtes Interesse nachweist.

Jedes Metainformationssystem verfügt hier über ein *Alleinstellungsmerkmal*: Keine Web-Suchmaschine kann Informationen finden, auf die sie nicht zunächst zugreifen kann. Der UDK (im "klassischen Gewand" der Versionen bis einschließlich 3) kann dies (und *nur* dies), indem er Personen befragt, die von den nicht-allgemein-zugänglichen Informationen wissen. Das kann eine Suchmaschine nicht.

Die Krux des UDK ist seine vollkommene Abhängigkeit von eben diesen Personen (Informationsträgern), wenn es darum geht, Metainformation zu erfassen. Die Internet-Suche arbeitet vergleichsweise provokativ: bevor sie mit einem Informationsanbieter spricht, hat sie bereits die Informationen selbst bis in alle Winkel durchsucht, sie konnte sich den Quellcode einzelner Seiten anschauen und Hyperlinks aufzeichnen, bis sie jede Seite aus allen Richtungen kennt. (Von der eigentlichen Information hat sie allerdings bis dahin noch nichts verstanden). Der Informationsanbieter entwickelt zwangsläufig eine ganz andere Motivation, als wenn man ihn schlicht fragt, über welche Informationen er denn verfügt.

Es entwickelt sich also die Idee einer Vereinigung des *strukturorientierten* UDK mit der *faktenorientierten* Internet-Suchmaschine.

An dieser Stelle müssen wir uns die Frage gefallen lassen, welche Rolle denn eigentlich die *Fachdatenbanken* für G2K spielen, denn sie sind sozusagen der geplante "Dritte im Bunde" (neben Internet und UDK). Tatsache ist, daß kein einziges Bundesland den direkten Zugriff auf Fachdatenbanken angeboten hat, während alle an einer Integration ihres Internet-Angebots sehr interessiert waren.

Interessiert an direkten Schnittstellen zwischen G2K und Fachdatenbanken zeigte sich allein der Bund, jedenfalls in den Fällen, wo eine eigene Internet-Präsentation (noch) nicht verfügbar ist. Man wünscht sich dort von G2K eine Visualisierung, und dafür nimmt man den Datenbankzugriff eher in Kauf, als das man ihn eigentlich selbst wünscht.

Die Probleme einer Visualisierung, die nicht unmittelbar beim "Datenherren" (wäre "Datenherrin" korrekt?) entsteht, sind im Leserkreis dieses Konzepts hinlänglich bekannt. Es bestehen da sicher auch gewisse Eitelkeiten auf beiden Seiten, tatsächlich ist die fachlich korrekte Interpretation der Daten weder den Daten selbst, noch den heute verfügbaren Metadaten allein zu

entnehmen. Hier liegt die Schnittstelle zwischen Umweltbeobachtung und Informationsmanagement. Das Informationsmanagement arbeitet viel formaler und abstrakter, fixiert auf seine Systemmodelle. Die Umweltbeobachtung steht dagegen immer mit einem Bein in der Realität, weil sie sie *beobachtet*. Sie erkennt auch den startenden Lastwagen (alter Diesel) unmittelbar neben dem Meßcontainer, obwohl der im Systemmodell eigentlich gar nicht vorgesehen war.

G2K kann keine universelle Visualisierungsmaschine sein. Die Visualisierung ist immer speziell. Wenn eine Harmonisierung von Länder- und Bundesdaten in einer einzigen thematischen Karte möglich ist, dann ist dies das Ergebnis einer langjährigen fachlichen und administrativen Abstimmung, aber nicht die Knopfdruck-Leistung¹ eines besonders schlaunen Computer-Programms.

Damit geht konform, daß G2K eigentlich weniger daran interessiert ist, viel Aufwand in die Übertragung von Fachdatensätzen zu stecken, mit deren Präsentation es dann allein gelassen wird. Eine Vergleichbarkeit von z.B. Meßnetzdaten zwischen den Bundesländern ist auf der Ebene der Daten selbst nicht gegeben.

G2K sollte die Entstehung von geeigneten Visualisierungen in Fachregie in jeder erdenklichen Hinsicht fördern. Dies erscheint erfolversprechender als eine technische Allgemeinlösung in G2K selbst.

Dennoch beinhaltet die Kommunikation mit Fachdatensätzen einen besonderen Reiz für die anspruchsvolleren unter den "interessierten Laien". Und nicht zuletzt ist ja auch jeder Informationsanbieter selbst wieder Suchender und möchte mit den Fachkollegen anderer administrativer Einheiten kommunizieren. Dies möchte G2K unterstützen (und davon lernen).

Der Fokus liegt dennoch auf dem interessierten Laien, und der steht einem "ausgewachsenen" Fachdatensatz wahrscheinlich eher hilflos gegenüber.

Dennoch kann die direkte Kommunikation mit Fachdatenbanken nicht aus dem weiteren Konzept ausgespart werden: der Experte soll gleichberechtigt unterstützt werden, und: in gewisser Weise treten uns auch die Metainformationssysteme wie Fachdatenbanken gegenüber.

G2Ks verteilte Suche umfaßt also

- WebSites (Visualisierungen) des Bundes und der Länder,
- Metadaten, insbesondere die Umweltdatenkataloge des Bundes und der Länder,
- ausgewählte Fachdatenbanken ohne heute bestehende eigene Internetpräsentation.

¹ "Press the Button - We Do the Rest" - damit warb Kodak vor Jahrzehnten für den Kauf einer Kamera .



Diese drei Kategorien werden hier als Domänen bezeichnet. Sie unterscheiden sich inhaltlich und methodisch hinsichtlich:

- ihrer Verbreitung und Einheitlichkeit,
- der auf sie anzuwendenden Suchverfahren,
- der in ihnen verbreiteten Protokolle,
- des angebotenen Zugriffs auf die angebotene Information,
- der durch sie vertretenen Betrachtungsebene, indem sie zum Teil aufeinander verweisen.