InGeo-IC : Metadateninformationssystem für Geodaten im WWW

Dr. Uwe Jasnoch *Stefan Göbel* *Dirk Balfanz*

Fraunhofer-Institut für Graphische Datenverarbeitung
Rundeturmstr. 6, D-64283 Darmstadt
email: [uwe.jasnoch, sga, balfanz]@igd.fhg.de URL: http://www.ingeoforum.de

Abstract. Das **InGeoForum** [GÖB98a, GÖB98b] versteht sich als Informations- und Kooperationsforum zwischen Geodatenanbietern und Geodatennutzern. Das grundlegende Ziel des InGeoForum besteht darin, basierend auf dem derzeitigen Geodatenmarkt, den Aufbau einer *"Nationalen Infrastruktur für Geodaten"* zu forcieren.
Als zentrale Dienstleistung bietet das InGeoForum ein sogenanntes Information Center (InGeo-IC) an, in dem einerseits Datenanbieter ihr Datensortiment vermarkten können, andererseits Anwender (z.B. GIS-Dienstleister) innerhalb des InGeo-IC nach geeigneten Geodaten suchen können.
Die Grundlage des InGeo-IC bildet das InGeoForum-Metadatenformat (InGeo-MDF), das nicht nur einen standardisierten/harmonisierten Metadatenaustausch in verteilten Umgebungen ermöglicht, sondern auch die Modellierung von Metadatenbanken und implizit den Aufbau der Benutzungsoberfläche des InGeo-IC beeinflußt.
Dieser Beitrag stellt die Architektur und Infrastruktur des InGeo-IC sowie Metadatenformat vor und beschreibt Möglichkeiten, wie themenspezifische/fachbezogene Metadaten - insbesondere Umweltdaten- im InGeo-MDF erfaßt werden können.

1. Einleitung

Das **InGeoForum**[1] versteht sich als Informations- und Kooperationsforum zwischen Geodatenanbietern und Geodatennutzern. Das grundlegende Ziel des InGeoForum besteht darin, basierend auf dem derzeitigen Geodatenmarkt, den Aufbau einer *"Nationalen Infrastruktur für Geodaten"* zu forcieren.

Die aktuelle Situation auf dem Geodatenmarkt in Deutschland läßt sich wie folgt beschreiben:

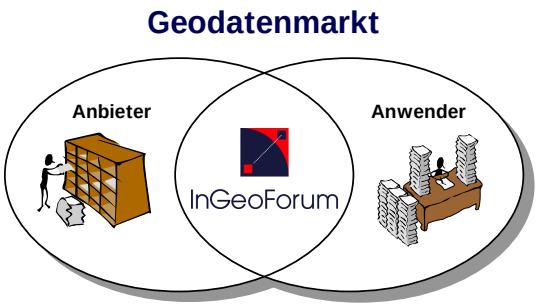

Abb.1: Geodatenmarkt in Deutschland

[1] InGeoForum: Informations- und Kooperationsforum für Geodaten, gegründet am 15.12.1997 als Forum des Zentrum für Graphische Datenverarbeitung e.V. (ZGDV)

Es gibt eine sehr große Menge an *'brachliegenden'* Geodaten, deren Existenz potentiellen Nutzern wie GIS-Dienstleister nicht bekannt ist. Aus Sicht der Nutzer stellt sich vielmehr die Frage, welche Daten existieren, wie diese zu bekommen sind und wie sie letztendlich - z.B. in einem GIS- weiterverarbeitet werden können. Demgegenüber stehen die Datenanbieter vor dem Problem, wie die Geodaten am vermarktet werden können.

Ausgehend von dieser Situation versucht das InGeoForum, mit Hilfe eines umfassenden Informationsservers im WWW die Potentiale zu öffnen, also den Geodatenmarkt transparenter zu machen und allen Beteiligten auf diesem Markt zu helfen: Geodaten-Anbieter profitieren, da sie das InGeo-IC als Plattform zur Vermarktung ihrer Geodaten nutzen können, Anwender können das Center als Instrument bei der Suche nach geeigneten Geodaten einsetzen.

Im folgenden Kapitel werden die Schwierigkeiten beschrieben, die bei der Errichtung eines branchenunabhängigen Metadaten-Informationssystems zu beachten sind. Insbesondere wird dabei die Errichtung eines Systems in einer verteilten Umgebung mit mehreren Metadatenservern bzw. Metadatenbanken aus den verschiedensten Geodaten-Anwendungsgebieten (z.B. Umweltdatenbereich, Verkehrswesen, Satellitendaten, etc.) diskutiert.

Als Basis für ein solches branchenunabhängiges IS wird dann in Kapitel 3 das InGeoForum-Metadatenformat vorgestellt, das einerseits -in Form von Metadatenfeldern- die Schnittmenge aller Metadatenbeschreibungen aus den einzelnen Branchen darstellt, andererseits aber auch flexibel und erweiterbar ist und somit themenspezifische Ergänzungen vorsieht.

In einer Zusammenfassung werden abschließend weitere Entwicklungsmöglichkeiten des InGeo-IC angedeutet.

2. InGeo-IC

Je mehr Informationen und Daten in der heutigen Informationsgesellschaft produziert werden, desto wichtiger werden Mechanismen und Systeme, die diese Daten ordnen und Auskunft darüber geben, wo welche Daten zu finden sind. Ausprägungen solcher Auskunfts-/Informationssysteme stellen Metadaten-Informationssysteme (MIS) und Katalogsysteme (KS) dar.

Unterscheiden lassen sich diese Systeme in ihrem Umfang, d.h. in diesem Zusammenhang durch das Informationsspektrum, das sie abdecken. Während z.B. Warenhauskataloge oder Telefonbücher typische Stellvertreter für KS darstellen und ein relativ klar umrissenes Informationsspektrum abdecken (Adresse, Telefonnummer, Artikelnummer, Preise, etc.), so umfassen andere Systeme ein wesentlich breiteres Spektrum.

Beispiele hierfür sind - bezogen auf den dt. Geodatenmarkt- u.a. umweltbezogene MIS wie der UDK (eine webbasierte Version des UDK wird derzeit entwickelt und prototypisch in Niedersachsen und Baden-Württemberg eingesetzt, s.[UDK98]), GEIN [RIE97] oder das Landesumweltinformationssystem Brandenburg (German Environmental Information Network; LUIS, Beschreibung/Analyse beider Systeme, s. [FRE98]). Ein typischer Vertreter eines Katalogsystems auf nationaler Ebene stellt das vom BKG erstellte System dar, in dem vor allem Datensätze aus dem Vermessungs-/Katasterwesen katalogisiert werden.

Auf internationaler/europäischer Ebene sind diesbezüglich im Bereich Umweltdaten die webbasierte Version des Catalogue of DataSources (WebCDS), GELOS (Global Environment Information Locator Service, [CDS]), GILS (Governmental Information Locator System, [CDS]) oder das EIONET (European Environment Information and Observation Network) als typische Beispiele zu nennen. Im Bereich der Satellitendaten und Luftbildaufnahmen stellt das seitens des CEO (Center for Earth Observation) entwickelte EO-Informationssystem INFEO (ehemals unter EWSE: European Wide Service Exchange bekannt) dar.

Die Schwierigkeit bei der Errichtung eines branchenunabhängigen MIS wie dem InGeo-IC, das zum Aufbau einer nationalen Infrastruktur für Geodaten (... aus allen Branchen) beitragen soll, besteht nun darin, all diesen Informationsspektren in einem System gerecht zu werden. Einerseits stellen die oben aufgelisteten MIS in den einzelnen Branchen in sich geschlossene, fachspezifische Informationssysteme dar und erlauben das Auffinden von branchenbezogenen Datensätzen, andererseits werden bei der Durchführung von Projekten im allgemeinen Datensätze aus verschiedenen Branchen benötigt. Dies kann folgendes Fallbeispiel veranschaulichen:

Bei der Planung eines Autobahnkreuzes benötigen Architekten, Ingenieure und GIS-Dienstleister unter anderem Luftbildaufnahmen des entsprechenden Gebiets, geologische Informationen zur Bodenbeschaffenheit, Katasterdaten oder weitere umweltbezogene Sachdaten.

An diesem Beispiel wird der Wunsch bzw. die Notwendigkeit nach einem branchenunabhängigen, fachübergreifenden MIS deutlich, das die Suche nach verschiedenen fachbezogenen Datensätzen erlaubt und -durch

Metadatenangaben wie Datenformat, Datengenauigkeit, Objektkataloge etc.- wichtige Schlüsse im Hinblick auf die Integration dieses hetrogenen Datenmaterials zuläßt.

Aus diesem Grunde gibt es sowie auf europäischer Ebene (ESMI: European Spatial Metadata Infrastructure, [SCH98]) als auch auf nationaler Ebene in Form des InGeoForum sowie des IMAGI (Interministerieller Ausschuß für Geo-Informationen) Bestrebungen, derartige branchenunabhängige MIS zu errichten.

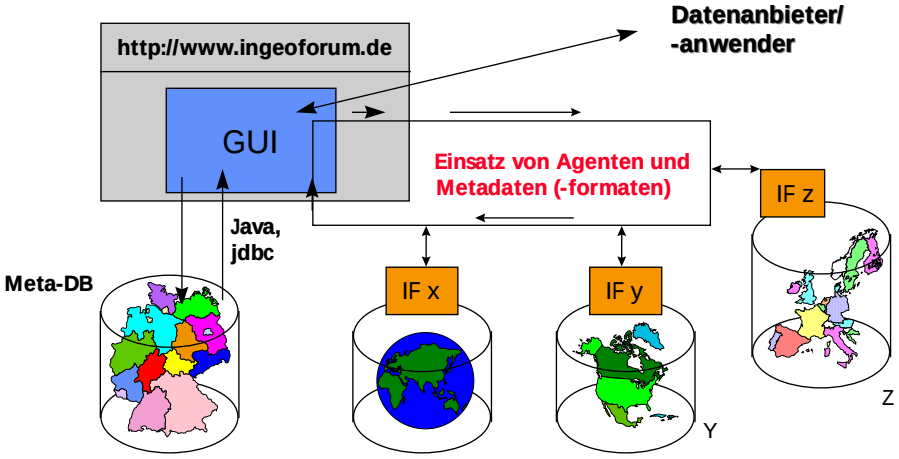

Abb.2: (Geplantes) Szenario des InGeo-IC

Abbildung 2 zeigt das geplante Szenario des InGeo-IC, das nicht nur auf eine eigene/lokale Metadatenbank zugreifen will, sondern auch beabsichtigt bereits existierende (branchenbezogene) MIS (z.B. WebCDS, WWW-UDK, etc.) und Metadatenbanken (einzelner Institutionen) einzubeziehen und somit die Suche nach (Meta-) Datensätzen aus den diversen Branchen in einer verteilten, vernetzten Umgebung ermöglichen will.

Entscheidend sind hierbei folgende Punkte:

- Metadatensätze sollen nicht redundant (zusätzlich in der InGeo-MDB) gehalten werden. Geplant ist auch der Zugriff auf Metadatenserver/-banken, die auf anderen Metadatenformaten (z.B. UDK, CDS) basieren
- Es gibt ein Suchsystem mit einer Benutzungsoberfläche, das Anwendern die Suche nach Metadatensätzen der verschiedensten Branchen erlaubt und Ergebnisse der Suche, d.h. die Auflistung und Präsentation der gefundenen Metadaten, in einem Ausgabefenster gemeinsam darstellt.
- Datenanbieter, die ihre Daten bisher nur in fachspezifischen MIS/KS vermarkten, steigern ihren Bekanntheitsgrad durch die Präsenz in weiteren, branchenübergreifenden MIS.

Zur technischen Realisierung eines solchen Szenarios sind spezielle Schnittstellen und Datenaustauschprotokolle erforderlich. Abbildung 3 skizziert den Zusammenhang zwischen Datenaustauschprotokoll Z39.50 [Z39.50], Profiles (z.B. CIP, “Geo”-Profile, s.[BES98]), Attributsätzen - die beim Einsatz von Protokollen für eine einheitliche Terminologie und die Codierung von Begriffen für den Datenaustausch sorgen - und den Metadatenstandards, die auf dem Geodatensektor eine entscheidende Rolle spielen.

3. InGeo-MDF

Vielen fachspezifischen MIS/KS liegen spezielle Metadatenformate bzw. Metadatenmodelle zugrunde. Jedoch basieren sie nur in den seltensten Fällen auf internationalen (Metadaten)-Standards. Somit sind für den Datenaustausch zwischen diesen Formaten oftmals sehr aufwendige Konvertierungen notwendig oder eine Metadatenbeschreibung eines Geodatensatzes in Format x ist überhaupt nicht in Format y abzubilden, da Format x Metadatenfelder enthält, die Format y nicht besitzt.

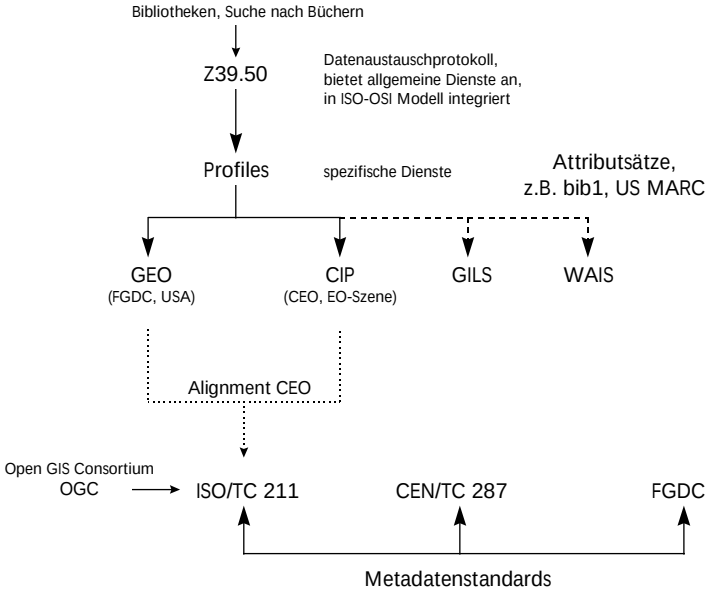

Abb. 3: Metadatenstandards/-formate und Protokolle

Aus diesem Grunde gibt es auf internationaler Ebene verschiedene Institutionen und Gremien (ISO/TC 211, CEN/TC 287, OGC, etc. [ISO, CEN, OGC]), die eine Harmonisierung und Standardisierung auf dem Gebiet von digitaler geographischer Information zum Ziel haben. Sie entwickeln bzw. spezifizieren Methoden, Tools und Services zur Datenverwaltung (inkl. der Definition und Beschreibung der Daten), zur Datenerfassung, Datenverarbeitung, Datenanalyse, für den Datenzugriff, die Darstellung/Präsentation der Daten und den Datentransfer in digitaler, elektronischer Form zwischen verschiedenen Benutzern, Systemen und (physikalischen) Orten.

Besonders interessant im Hinblick auf die Entwicklung von Metadatenbanken und MIS/KS im WWW sind hierbei die Arbeitsgruppen zu den Themen Metadaten und Katalogisierung.

- ISO/TC 211 Geographic Information: part 10 - Cataloguing, part 15 - Metadata
- OGC Abstract Specification: topic 11 - Metadata, topic 13 - Catalog Service
- CEN/TC 287 Work item 287009: Metadata.

Einige der in Deutschland bereits existierenden MIS/KS (z.B. das in Kapitel 2 erwähnte Katalogsystem des BKG) basieren auf dem vom CEN/TC 287 erarbeiteten Standard (prENV12657, Stand Oktober 1998, noch nicht offiziell verabschiedet); andere MIS/KS wie z.B. WebCDS oder WWW-UDK basieren auf eigenen Metadatenformaten (CDS, UDK), jedoch liegen keine allgemeinen Metadatenmodelle wie die des CEN oder des ISO (CD15046-15, ebenfalls noch nicht offiziell verabschiedet) zugrunde.

Im Hinblick auf die Entwicklung des InGeo-MIS und die Integration bereits existierender MIS/KS stellte sich die Frage, wie die verschiedenen Metadatenmodelle und -formate am besten zu integrieren sind, bzw. welcher allgemeine Metadatenstandard sich am besten eignet, um die fachspezifischen Metadatensätze aller Branchen darin zu beschreiben. Eine Analyse der Standardisierungsvorschläge von CEN und ISO hat ergeben, daß das ISO-Metadatenmodell eher als Grundlage zum Aufbau der InGeo-MDB und dem gesamten InGeo-MIS geeignet ist, da es gegenüber CEN/prENV12657 eine detailliertere Metadatenbeschreibung zuläßt, klarer strukturiert und auch erweiterbar ist.

Dementsprechend wurde zur Errichtung des InGeo-MIS ein Metadatenformat (InGeo-MDF) entwickelt, das sich sehr stark an den ISO-Standard anlehnt, jedoch nicht alle Metadatenfelder dieses Standards enthält und desweiteren in verstärktem Maße fachspezifische Erweiterungen vorsieht.

Analog dem ISO-Standard läßt sich das InGeo-MDF als mehrstufiges Metadatenformat auffassen: In Stufe 1 sind allgemeine Metadatenfelder enthalten, die eine grobe Beschreibung von Daten aus den verschiedensten Geodaten-Branchen zulassen. Diese Stufe entspricht bis auf einige geringfügige Änderungen dem entsprechenden Level des ISO-Standards, wodurch eine Konformität erreicht wird. Stufe 2 enthält weitere Metadatenfelder, die eine möglichst

präzise, fachgerechte Beschreibung der einzelnen Datensätze erlauben, jedoch nicht allen Branchen gemein sind und somit nicht mehr zur Schnittmenge aller 'Geodaten-Branchen' gehören. Es ist hierbei zu überprüfen, ob diese Erweiterungen gegenüber dem Level 2 des ISO-Standards echte Erweiterungen darstellen. Ziel ist es, soweit wie möglich Abbildungen direkt in die Metadatenfelder des ISO-Standards vorzunehmen.

Im Hinblick auf die Entwicklung des InGeo-IC, das nicht nur die Suche nach allgemeinen Metadatenfeldern (z.B. Raum- und Zeitbezug des Datensatzes), sondern auch eine fachspezifische, thematische Suche (auch nach speziellen qualitativen Merkmalen von Datensätzen aus den einzelnen Branchen, z.B. eingesetzte Brennweiten bei Luftbildaufnahmen aus dem Bereich EO-Daten) erlaubt, bieten beide Stufen des InGeo-MDF Ansätze für dementsprechende Suchmöglichkeiten:

In Stufe 1 können spezielle, fachspezifische Thesauri (z.B. im Bereich von Umweltdaten der multilinguale Thesaurus GEMET oder der nationale Thesaurus GenThes) in Kombination mit dementsprechenden Schlüsselwörtern angegeben werden, die somit eine themat. (Schlagwort-)Suche ermöglichen.

In Stufe 2 können sogenannte branchenbezogene Profiles angegeben werden, d.h. das InGeo-MDF wird bezogen auf die jeweilige Branche um einige spezielle Metadatenfelder angereichert. Somit können Datensätze sehr detailliert beschrieben werden und eine Suche nach diesen fachspezifischen Details wird ermöglicht.

Bei beiden Ansätzen ist festzuhalten, daß für eine Realisierung dem InGeo-IC als Einstiegspunkt für eine Suche in einer verteilten Umgebung (s. Kapitel 2) entsprechende Thesauri und Profiles zur Verfügung gestellt werden müssen.

Selbstverständlich bleiben trotz der Einführung des InGeo-MDF Konvertierungen zwischen Metadatensätzen bereits existierender MIS unerläßlich. Aus diesem Grunde werden seitens des Fraunhofer-IGD als technisch ausführender Kraft des InGeoForum derzeit neben einem allgemeinen Eingabetool für neue Metadatensätze entsprechend dem InGeo-MDF auch Konvertierung-Tools entwickelt, die z.B. das Abbilden von CEN-Metadatensätzen oder im UDK bzw. CDS-Datenmodell vorliegenden Metadatensätzen in das InGeo-MDF automatisieren oder zumindest erleichtern.

4. Zusammenfassung

Das InGeoForum will zum Aufbau einer nationalen Infrastruktur für Geodaten beitragen. Hierfür wird das InGeo-IC entwickelt, das nicht nur eine eigene lokale Metadatenbank, sondern auch weitere Metadatenserver und bereits existierende MIS/KS integrieren soll. Insbesondere auf dem Umweltdatensektor existieren in Deutschland schon dementsprechende fachspezifische MIS/KS, z.B. WebCDS oder (WWW-)UDK. Jedoch sind die zugrundeliegenden (Meta-) Datenmodelle und -formate nicht ohne weiteres ineinander überführbar bzw. es ist keine Harmonisierung dieser Formate möglich. Deshalb wurde das InGeo-MDF entwickelt, das auf dem ISO-Metadatenstandard basiert, sehr flexibel ist und themenspezifische Erweiterungen -z.B. für den Umweltdatensektor als eine Branche von Geodaten- zuläßt.

Somit können umweltbezogene Metadatensätze nicht nur direkt im InGeo-MDF (ggf. in der InGeo-MDB) abgelegt werden, sondern auch unter Zuhilfenahme von geeigneten Mapping-Tools und Profiles über das InGeo-IC in bereits existierenden MIS und MDB's gesucht und gefunden werden.

Literatur

[GÖB98a] Göbel, S.; Lutze, K.; Dr. Giger, C.:
InGeoForum: Information and Cooperation Forum for Geospatial Data
Proceedings EOGEO'98, Salzburger Geographische Materialien, Heft 27

[GÖB98b] Göbel, S.; Lutze, K.:
Development of meta databases for geospatial data in the WWW
ACM-GIS'98, Proceedings of the 6th International Symposium on Advances in Geographic Information Systems. November 6-7, 1998, Washington, D.C., USA

[UDK98] Unterlagen zum Prototyp-Workshop WWW-UDK 4.0 am 19./20.November 1998, FZI Karlsruhe

[RIE97] Riekert, W.-F.; Tochtermann, K.; Wiest, G.; Gaul, M.; Seggelke, J.; Mohaupt-Jahr, B.:
Fach-, raum- und zeitbezogene Katalogisierung und Recherche von Umweltinformationen auf dem Internet
Hypertext –Information Retrieval- Multimedia'97, Proceedings HIM'97.

[FRE98] Freitag, U.; Kutsche, R.-D.; Busse, S.:
Internet-Technologien als Basis für föderierte Online-Umweltinformationsdienste am Beispiel des LUIS Brandenburg
Hypermedia im Umweltschutz. 1. Workshop 1998. Metropolis-Verlag, Marburg, 1998

[CDS] ETC/CDS: European Topic Centre on Catalogue of Data Sources, inc. Document:
Comparing the ETC/CDS Core Data Model to GILS, GELOS and DUBLIN CORE
http://www.mu.niedersachsen.de/cds/start_documents.html

[SCH98] Scholten, H.; LoCashio, A.; Bonn, B.:
ESMI, towards a European Spatial Metadata Infrastructure
Proceedings EOGEO'98, ESMI webserver: http://www.geodan.nl/esmi/

[Z39.50] ANSI/NISO Z39.50-1995 Information Retrieval (Z39.50):
Application Service Definition and Protocol Specification, Maintenance Agency Official Text
http://lcweb.loc.gov/z3950/agency/ Profiles: /agency/profiles/profiles.html

[BES98] Best, C.; Hayes, L.; Nebert, D.; Reich, L.; Percival, G.; Hylton, J.; Smith, S.:
Alignment of CIP and GEO Profiles
Proceedings EOGEO'98, http://www.sbg.ac.at/geo/eogeo/index.html

[ISO] ISO/TC 211: Geographic Information/Geomatics,
http://www.statkart.no/isotc211/

[CEN] CEN/TC 287: Geographic Information
http://forum.afnor/WORK/AFNOR/GPN2/A13C/PUBLIC/WEB/ENGLISH/index.htm

[OGC] OGC: Open GIS Consortium
Status of OGC, presented by Kurt Buehler at EOGEO'98
http://www.opengis.org, Proceedings EOGEO'98