
Preprint aus:

Strobl, J., Blaschke, Th, Griesebner, G. [Eds.]

Angewandte Geoinformatik

Wichmann, Heidelberg (2006)

S. 41-46

Metadateneditoren für geowissenschaftliche Daten

Knut BEHREND, Wolfgang CZEGKA und Stephan BRAUNE

Zusammenfassung

Dieser Beitrag ist ein Vergleich verschiedener Softwaretools zum formularbasierten Editieren von Metadaten für geowissenschaftliche Daten.

Einleitung

Für die Nutzung wissenschaftlicher Daten sind Kenntnisse über ihr Zustandekommen, die Urheberschaft, Format und Qualität unerlässlich. Daher sind Metadaten bei der Langzeit-speicherung in Datenzentren und Archiven mit abzulegen. Geoinformationen verlieren an Qualität wenn die Dokumentation mit Metadaten unvollständig erfolgt oder verloren geht. Metadaten spielen eine wichtige Rolle bei der Bereitstellung von Geodaten im Internet und bei der Vernetzung von Informationssystemen (SIETMANN 2005, SENKLER et al. 2003, SLIWINSKI & SENKLER 2003).

Für Geodaten gibt es mehrere Dokumentationsstandards wie z.B. Dublin Core (DC), FGDC-CSDGM oder ISO 19115, wobei die Anzahl der zu dokumentierenden Elemente variiert. DC beschränkt sich auf etwa 15 Elemente, während FGDC und ISO 19115 über 300 Elemente spezifizieren, von denen die meisten jedoch optional sind. Diese Standards geben vor, mit welchen Attributen die Daten zu beschreiben sind, nicht aber, wie die Speicherung technisch umgesetzt wird (vgl. MOELLERING et al. 2005, NOGUERAS-ISO et al 2005, ISO/TR 19120).

Metadaten können von den eigentlichen Daten separat oder im Datenfile selbst, z.B. in Dateiheadern oder speziellen Bereichen, abgelegt werden. Solche Header enthalten z.B. die EXIF Daten der JPEG-Digitalfotos oder die globalen Attribute bei NetCDF-Files. In diesem Beitrag geht es um den Vergleich verschiedener Softwaretools zum formularbasierten Editieren von separat gehaltenen Metadaten, also z.B. in Metadaten-Files oder Datenbanken.

1. Drei Anwendungsszenarios

Damit Fachwissenschaftler zu ihren Datensätzen eine aussagekräftige Beschreibung entsprechend eines Metadatenstandards anlegen können, sollte ihnen ein benutzerfreundlicher Metadaten-Editor zur Verfügung gestellt werden.

In dieser Publikation können die Anforderungen an so einen Editor nur skizziert werden: Der Editor sollte mehrere Metadatenstandards korrekt und vollständig unterstützen. Eine flexible Vorlagenverwaltung sollte dem User die Eingabe von gleich bleibenden Elementen erleichtern. Das Erscheinungsbild der Software (z.B. Menüs, Farbschemas), ihr Verhalten (Input-Formatierung und Validierung) und die Dateien und Datenstrukturen (Konfigurationsdateien, gebundene Datenquellen, etc) sollten entsprechend der Zugriffsprivilegien und der technischen Gegebenheiten von Nutzern und Administratoren flexibel angepasst werden können. Um die Metadatenverwaltung von Massendaten automatisiert zu betreiben oder um die Software in bestehende Informationssysteme zu integrieren, sollte sie über Programmierschnittstellen oder zumindest Kommandozeilenoptionen verfügen. Die interaktive Programmoberfläche kann verschiedene Ansichten (z.B. rollenbasierte Ansicht, Nur-Lese-Zugriff, Druckansicht, etc) bereitstellen. Thesauri sollten sich einbinden lassen, so dass Eingabefehler bei Ortsbezeichnungen vermieden werden. Man sollte den Editor an spezielle Metadatenprofile anpassen oder mit Plugins erweitern können (z.B. zum Parsen der Dateiheder von Inputdateien, zur Katalog-Nutzung, Dienste-Lokalisierung und Online-Suche in Diensten und Daten). Datenhaltung in einem proprietären Dateiformat allein ist unakzeptabel, also sollte eine Editorsoftware auch HTML- oder XML-basierte Ausgabeformate unterstützen.

Die folgenden drei geowissenschaftlichen Anwendungsszenarios sollen dies verdeutlichen.

Szenario 1 - Punktförmige Geodaten. Manuelle Eingabe eines simplen Datenbestandes, für den es noch keine Beschreibung durch Metadaten gibt. Für eine Bohrung sollen die Metadaten erfasst werden. Ein Wissenschaftler will dazu einen Editor verwenden, um die wichtigsten Projektdaten standardkonform einzugeben. Über die Lokation gibt es einen umfangreichen Bestand semistrukturierter Daten mit speziellen Detailinformationen und eigener, fachspezifischer Semantik: Es kann sich um ingenieurwissenschaftliche Daten handeln oder um provisorische Schichtenverzeichnisse mit deren Fachtermini. Ein standardkonformer Metadateneditor sollte hier die manuelle Dateneingabe unterstützen. Benutzerkomfort, Vorlagenverwaltung und Erweiterbarkeit stehen hier im Vordergrund.

Szenario 2 - Zweidimensionale geologische Karte. Als zweites Beispielszenario wurde eine digitale Geologische Karte aus einer externen Quelle mitsamt vorhandenem Metadatensatz genommen. Ziel des Szenarios war die interaktive Anpassung des mitgelieferten Metadatensatzes, und interne Weitergabe an andere Wissenschaftler. Hierfür eignen sich geologische Karten auf Grund ihrer relativ komplexen Sachdaten (und damit auch Metadaten) als Anwendungsbeispiel. Bei diesem Szenario liegt der Fokus auf den Editierfunktionen, den Import- und Exportmöglichkeiten der Editoren.

Szenario 3 – Simulationsrechnungen. In Szenario 3 soll ein Metadateneditor verwendet werden, um die Mastervorlagen für eine größere Anzahl sehr ähnlicher Dokumente (wie z.B. Metadatensätze für Simulationsrechnungen) zu erstellen. Es fallen Massendaten an, die sich inhaltlich nur in wenigen Attributen unterscheiden. Anfangs wird eine Mastervorlage mit einem interaktiven Metadateneditor erstellt, die dann als Input für ein Programm zur automatischen Metadatengenerierung genutzt wird. Die Metadaten setzen sich aus verschiedenen Teilen zusammen: einem veränderlichen Teil, der aus den Dateihedern (z.B. Inputparameter), den Daten selbst (z.B. Einbettungsrechteck) und aus dem Dateisystem (z.B. Erstellungsdatum oder Dateigröße) entnommen wird, und aus einem

konstanten Teil. Bei diesem Szenario sind die Gesichtspunkte Vorlagenverwaltung und Batchbetrieb essentiell, da eine manuelle Eingabe hoch redundanter und parametrisierbarer Information nicht praktikabel wäre. Das entsprechende Programm sollte eine Schnittstelle zum Parsen standardisierter Dateiheader-Formate (wie z.B. NetCDF-Attribute in der „Climate- and Forecast-Convention“), zum Lesen der Daten und der Filesystem-Datei-information haben, sollte möglichst eine Ausgabeschnittstelle für XML-Files haben, und muß per Script oder Kommandozeile aufzurufen sein.

2. Produktübersicht

In diesem Beitrag werden die in der Tabelle 1 aufgeführten Metadaten-Editoren mit den oben genannten Anwendungsszenarien herstellerneutral auf deren Brauchbarkeit in der Praxis geprüft.

Tab. 1: Überblick über die behandelten Metadaten-Editoren.

Metadateneditor	Betriebs-System	Verfügbar-keit	Unterstützte Standards	Datenhaltung	Gruppe
ArcCatalog	Windows	kommerziell (Fa ESRI)	FGDC	Proprietär-XML	Spezial
ArcView Metadata Extension	Unix, Windows	kommerziell (ArcView, Fa ESRI)	FGDC	Proprietär-XML	Spezial
Authentic 2005	Windows	kommerziell (Fa. Altova)	W3C-XML-Schema	DB oder XML-(proprietär)	Generisch
CatMDEdit	Unix, Windows	frei	Dublin Core, FGDC, ISO		Spezial
Corpsmet95	Windows	eingestellt	FGDC		Spezial-EOL
Enraemed	Windows	frei	Dublin Core, ISO	SQL-Datenbank	Spezial-EOL
Infopath	Windows	kommerziell (Microsoft)	W3C-XML-Schema	DB oder XML-(proprietär)	Generisch
Mediator	Windows	frei	ISO-19115		Spezial-EOL
Preludio	Windows	gratis, komm. (Fa. Disy)	ISO-19115	Datenbank	Spezial
PubliStar	Linux, Windows	Kommz. (Fa. Delphi IMM)	ISO-19115	Datenbank (DB)	Spezial
USGS FGDC Metadata Tools	Unix, Windows	frei	FGDC	DB, Files	Spezial

EOL: End Of Life (Entwicklung und Vertrieb eingestellt)

Bei den evaluierten Tools handelt es sich um ausführbare Desktop-Anwendungen. Webbasierte "Editoren" oder freie Online-Validierungsdienste wie zum Beispiel <http://www.ukoln.ac.uk/metadata/dcdot/> wurden nicht in die Auswahl aufgenommen, weil dieser Mechanismus für vertrauliche Daten nicht geeignet ist.

3. Ergebnisse

Die bewerteten Merkmalkategorien der untersuchten Editoren sind in Tabelle 2 dargestellt. Eine Webseite mit den Ergebnissen der Vergleichstests wird noch für einige Zeit unter diesem URL zu finden sein: <http://dc.gfz-potsdam.de/~knb/agit/>.

Tab. 2: Bewertung der Hauptanforderungs-Kategorien der Editoren

	Installtn / Konfigtn / Administr.	Leistungsumfg. / Komfortfkt. / Textbearbeitg	Güte der Unterstützung / MD-Stds.	Integration in Untern. / IT / Infrastrukt.	Datenhaltung / Daten-Export	FAZIT
ArcCatalog	+	+	+	+	+	direkter Bezug zu den Daten
ArcView Metadata	+	-	0	0	0	GUI wirkt etwas minimal; Link zu GIS Daten
Authentic	+	+	0	+	+	Standalone-V:ers. gratis; in XMLSpy
CatMDEdit	+	+	+	0	+	mehrere Thesauri vorhanden
Corpsmet95	-	-	-	-	-	Entwicklung eingestellt
Enraemed	-	-	-	0	+	Thesaurus vorh., MS-SQL-DB-basiert
Infopath	0	+	0	+	+	Teil v MS-Office, Konfigurationsaufwand rel. hoch
Mediator	+	-	0	0	0	primitives Erfassungstool
Preludio	+	+	+	+	+	Hohe Startupzeit da java -basiert
PubliStar	+	+	+	+	+	Client-Server-Anwendung
USGS FGDC MD Tools	+	+	+	+	+	Suite aus 7 kleinen Tools z.T. für Kommandozeile

Bewertungsschlüssel: + geeignet/gut, 0 neutral, - weniger gut

Man kann diese Editoren in zwei Gruppen einteilen: Spezielle Metadateneditoren und Generische Editoren.

Spezial-Editoren sind vom Softwarehersteller bereits auf einen Metadatenstandard oder Metadatenprofile zugeschnitten worden. Damit sollte ein normaler User sofort in der Lage sein, standardkonforme Metadaten einzugeben.

Generische Editoren sind Softwarewerkzeuge, die nicht speziell auf Geodaten zugeschnitten sind, sondern allgemein der formularbasierten Bearbeitung von XML Daten dienen. Mit ihnen können auch (Meta-)Daten aus völlig anderen Fachgebieten verwaltet werden. Bei dieser Gruppe von Editoren sind teils aufwendige Anpassungen und Programmierarbeiten zu leisten, bevor ein Geowissenschaftler damit Geodaten-Metadaten bearbeiten kann. Tools aus dieser Kategorie haben dafür z.B. flexible Import- und Präsentationsfunktionen, umfangreiche XML-Unterstützung; Schnittstellen und Integrationsmöglichkeiten in die vorhandene IT-Infrastruktur (z.B. Verzeichnisdienste, Groupware- oder GI-Systeme).

4. Diskussion

Vielfalt der Spezialdisziplinen und Verwendungszwecke erfordert aufwendige Anpassungen. Bei der Breite der potentiellen Anwendungsmöglichkeiten wird man feststellen, dass die Szenarios und Produkte wahrscheinlich nicht auf die eigenen Anforderungen übertragbar sind. Ein gewähltes Softwaretool ist oft nur für einen Verwendungszweck wirklich gut geeignet, und eine Integration in die Prozesse im Unternehmen bzw. Projekt erfordert erheblich mehr Einarbeitung, Schulung und Programmieraufwand als der Einzelnutzer-Betrieb.

Unterschiedliche Entwurfsprinzipien. Die Standards DC und ISO19115 erscheinen auf der Ebene der zwingend vorgeschriebenen Elemente weitgehend miteinander kompatibel. Die zugrunde liegenden Entwurfsansätze entscheiden sich jedoch fundamental. Der Standard Dublin Core ist vom Bibliothekswesen geprägt, und ein zentrales Entwurfsziel ist, ein möglichst flexibles und dennoch allgemeingültiges Beschreibungssystem von Online-Ressourcen aller Art zu entwickeln. Im Gegensatz dazu geht ISO 19115 vom Geoinformationswesen aus, mit der Zielstellung auf vielen Ebenen (Semantik, Applikation, Datenübertragung) maximale Interoperabilität von verteilten Systemen zu gewährleisten. Die Diversität der Metamodelle ist eine mögliche Ursache für das Fehlen eines Editors, der alle Standards gleich gut unterstützt.

Extension Methodologies nicht fertig spezifiziert. Um zu ermöglichen, den Metadatenstandard an Fachtermini und Regelwerke einzelner Spezialgebiete anzupassen und weiterzuentwickeln, verfügen DC und ISO 19115 über dedizierte Spezifikationen von Erweiterungsmöglichkeiten zum Erstellen von "Community Profiles" (ISO) oder "Element Refinements" (DC). Dies wird von keinem Editor hinreichend berücksichtigt. Das könnte daran liegen dass diese Erweiterungsmöglichkeiten neu und noch nicht zu Ende spezifiziert sind.

Schneller technischer und normativer Wandel. Bei den Spezialeditoren hat die technische Umsetzung nur zeitlich verzögert mit den Standardisierungsbemühungen und der Normenvereinheitlichung Schritt halten können. Einige Editoren wurden schon veröffentlicht als ISO 19115 noch nicht endgültig verabschiedet war. ISO 19115 ist Teil einer Familie von ca 40 Standards, die z. T. noch im Entwurfsstadium sind, aber von ISO 19115 schon referenziert werden. DC scheint sich in Richtung des ebenfalls unfertigen W3C-Standards RDF weiterzuentwickeln, das bedeutet, der künftige Schwerpunkt wird auf semantische Interoperabilität gelegt.

Trend zur Bündelung und Integration. Leistungsfähigere Editoren werden oft gebündelt an Dienstleistungen oder sind Komponente größerer Softwarelösungen. Editor-Applikationen werden auch zunehmend mit Onlinediensten und Webangeboten gekoppelt. Dies scheint praktikabel zu sein, wegen der Komplexität der Sachverhalte, der anfallenden Datenmengen oder den Gegebenheiten bei der Projektarbeit vor Ort.

5. Literatur:

- o.A. (2003): *Final Draft International Standard ISO/FDIS 19115 (ISO/TC211) Geographical Information - Metadata (Version 2003-03-23)*. ISO, Genève.
- o.A. (currently under review) *ISO/TR 19120 Functional Standard - an Informative Report that identifies and reviews a number of geographic standards currently in use*. ISO, Genève.
- Moellering, H, H. J. G. L. Aalders & A. Crane [Hrsg] (2005): *World Spatial Metadata Standards: Scientific and Technical Descriptions, and Full Descriptions with Crosstable*. Elsevier, Amsterdam. ISBN: 0080439497.
- Nogueras-Iso, J, F. J. Zarazaga-Soria & P. Muro-Medrano (2005): *Geographic Information Metadata for Spatial Data Infrastructures*. Springer, Berlin/Heidelberg. ISBN: 3540244646.
- Senkler, K, A. Remke, U. Voges (2003): *Metadaten und Metainformationssysteme in einer GI*. In: Bernard, L., Sliwinski, A. und K. Senkler (Hrsg.): *Geodaten- und Geodienste-Infrastrukturen - von der Forschung zur praktischen Anwendung*. Beiträge zu den Münsteraner GI-Tagen 26./27. Juni 2003, Münster 2003. IfGIprints Band 18.
- Sietmann, R. (2005): *Wissenschaft + Web-Service = e-Science*. Heise Verlag: c't 2005, Heft 24, S. 46-49.