

Standards für den Datenaustausch in der Geologie:

pro und contra GML

Marcus Apel

TU Bergakademie Freiberg, Lehrstuhl für Mathematische Geologie und Geoinformatik, Bernhard-von-Cotta-Straße 2, 09596 Freiberg/Sachsen, Email: marcus.apel@geo.tu-freiberg.de

GML ist eine XML- basierte Sprache zur Repräsentation von georeferenzierten Daten. Sie wurde vom OpenGIS Konsortiums (OGC) entwickelt. GML- kompatible Datenmodelle können zur Datenspeicherung in Datenbanken genutzt werden. Dieser Standard erlaubt darüber hinaus den effizienten Austausch von Geodaten zwischen Datenbanken, Mapservern, und Nutzeranwendungen wie z.B. GIS und Internet.

Dieser Artikel zeigt, wie GML für geologische Datenmodelle genutzt werden kann. Dabei werden Vor- und Nachteile im Hinblick auf geologische Aufschluss- und Modelldaten beleuchtet und eine GML-kompatible Lösung vorgeschlagen.

1 Einführung

1.1 Bedeutung von Standards für Geodaten

Mit der rasch zunehmenden Verbreitung und Vernetzung von GIS gewinnen Standards rasch an Bedeutung. Wichtige Ursachen sind:

- Produktivitätssteigerung von Unternehmen und Behörden durch die Nutzung standardisierter Programmschnittstellen, Austauschformate, und wieder verwendbarer Softwarekomponenten
- Standardisierung wird auf einer systemneutralen Daten- und Modellierungsebene geführt; damit Investitionsschutz durch Herstellerunabhängigkeit
- alle Beteiligten sprechen die gleiche "Sprache", damit bessere Konsistenz der Daten und Vermeidung von Informationsverlusten

1.2 Stellung des OpenGIS Consortium (OGC) und GML

OGC ist ein 1994 gegründetes internationales Konsortium mit Mitgliedern aus Industrie, Verwaltung und Hochschulen, welches sich die Schaffung von offenen Standards für den Umgang mit Geodaten zum Ziel gesetzt hat. GML ist Bestandteil einer ganzen Serie von Geonomen des OGC. Eine detaillierte Aufstellung aller OGC Standards ist unter www.opengis.org zu

finden. GML basiert auf weiteren wichtigen internationalen Standards:

- Object Management Group (OMG) Standards rund um UML (UML, MOF, XMI, MDA) zur objektorientierten Modellierung von Daten und Prozessen
- World Wide Web Consortium (W3C) Standards rund um XML (XML, XML Schema, XLink, XSLT etc.)

1.3 XML - Merkmale

In einer Mark-up Sprache wird der Inhalt eines Dokuments durch besondere Markup Elemente strukturiert. XML ist eine Metasprache zur Definition von Markup Sprachen, wie z.B. HTML, SVG und GML. Vorteile von XML-basierter Datenspeicherung sind:

- klare Trennung von Inhalt (Daten) und Darstellung (z.B graphisch oder Text)
- einfache Integration von geographischen und nicht-geographischen Daten
- ein Framework zur Entwicklung untereinander kompatibler Anwendungen.

XML als Datenformat bietet im Zusammenspiel mit XSLT und SVG viele Vorteile für die graphische Repräsentation und für Endnutzer editierbare Karten. Deren Definition erfolgt in XML Schema Dokumenten. XML Schema ist – wie XML selbst – zwar grundsätzlich durch Menschen lesbar, aber primär für eine möglichst einfache Verarbeitung durch Computer entwickelt.

Das Lesen und Interpretieren von XML Schema wird durch eine Reihe von XML Parsern in Programmiersprachen oder in vollständigen Softwaretools unterstützt. XML Schema ist mächtig genug, um die grundlegenden Konzepte von objektorientierten Sprachen abzubilden:

- Vererbung kann über die Elemente <extension> und <restriction> modelliert werden
- Abstrakte Klassen werden durch Substitution Groups unterstützt
- primitive Datentypen (String, Integer, Float, etc.) sind Teil der XML Schema Spezifikation XML Namespaces
- Verschiedene XML Schemas können gleichzeitig genutzt werden (durch import/ include)

Mit XLink kann man innerhalb eines XML Dokuments Beziehungen zwischen resources und locators definieren. Eine resource bezeichnet ein XML- Element im gleichen Dokument wie die Link-Definition. Ein locator identifiziert eine externe Quelle bei der es sich nicht um ein XML- Dokument handeln muss. Mit XLink kann man gerichtete Graphen beschreiben.

1.4 GML - Merkmale

GML ermöglicht es geographische Objekte entsprechend den OGC Abstract Specifications zu modellieren. In GML finden die folgenden 2 Datenmodell Darstellungen Verwendung:

- GML-Applikationsschema (normativ)
- UML-Klassendiagramm (informativ)

Die Modellbeschreibung erfolgt in GML primär in der Form von XML Schema. Allerdings sind sämtliche GML-Basis-Schemas, wie zum Beispiel feature.xsd oder topology.xsd auch in der Form von UML-Klassendiagrammen dokumentiert. Zurzeit laufen Bestrebungen zur Aufwertung von UML als eigentlichem Modellierungsmechanismus mit der Möglichkeit der Ableitung der entsprechenden GML-Applikationsschemas.

Als GML **Datentypen** können die ca. 42 Standard-Typen von XML-Schema, die in den GML-Basis-Schemas vordefinierten Datentypen und selbst definierte Typen verwendet werden. Eine Typendefinition kann dabei im ganzen Namensraum verwendet werden.

Das **Geometriemodell** von GML (XML Schema: geometry.xsd) folgt demjenigen vom Topic 1 der OGC Abstract Specification (entspricht

ISO 19107). Unter anderem können Punkte (Point, 0d), Polylinien und Polygone (Curve, 1d), triangulierte Flächen (Surface, 2d) und Volumenkörper (Solid, 3d) modelliert werden.

Abstrahierte Phänomene der Realwelt werden in GML als **Feature** (XML Schema: feature.xsd) bezeichnet und sind gleichbedeutend mit Geo-Objekt. Ein Feature kann als Typ (Abstraktion) oder auch als Instanz (Repräsentation) vorkommen. Eine Feature Collection, kann selbst auch als Feature betrachtet werden.

Der Zustand eines Features ist durch ein Set von **Properties** (Name, Typ und Wert) definiert. Typ-Definitionen legen die Anzahl, Namen und Wertebereiche von Eigenschaften fest, die ein Feature haben darf. Features können geometrische, aber auch andere Eigenschaften haben. Die geographischen Eigenschaften werden mittels den aus geometry.xsd importierten Geometrie-Typen gespeichert. Gewisse Eigenschaften wie etwa Position oder Ausdehnung werden sicher von den meisten Features benötigt. Um für solche gewöhnlichen Eigenschaften eine einheitliche Namensgebung zu erreichen definiert feature.xsd eine Reihe von Namen. Über die Semantik werden hier allerdings keine Aussagen gemacht.

Beziehungen zwischen GML-Features werden entweder durch das direkte Einsetzen des untergeordneten Elementes in das übergeordnete Element ("by value"), oder mit XLink ("by reference") beschrieben.

1.5 GML Application Schema

Geometry.xsd und feature.xsd bilden ein Framework für die Erstellung von Datenmodellen für Anwendungen (Application Schema), die GML Elemente erweitern oder einschränken. Für die Nutzung des Frameworks gelten die folgenden Richtlinien:

- die grundlegenden GML-Elemente dürfen nicht verändert werden, neue Typen/Elemente werden durch Erweiterung der GML Basistypen erzeugt
- neue Feature Elemente erweitern AbstractFeatureType
- neue Geometrie-Typen erweitern AbstractGeometryType
- neue geometrische Eigenschaften erweitern GeometryPropertyType

2 Bewertung des GML Daten-Modells für die Geologie

Das GML Datenmodell ist *feature*-orientiert. Die abgebildete Welt besteht aus Objekten, die beschreibende Eigenschaften (Attribute) und Verhalten (Funktionen) besitzen. Einige Attribute sind räumlich. Deren Werte definieren die Lage und Form geometrischer Objekte. Features sind durch Beziehungen verbunden und erhalten über Metadaten einen Zusammenhang, Semantik (*Feature Type*), und Semiotik (*Cartography*, vgl. Abbildung 1). Diese Sichtweise repräsentiert die geologischen Verhältnisse als eindeutig und unfragwürdig. Demgegenüber sind in den Geowissenschaften grundsätzlich zwei verschiedene Arten von Daten vorhanden:

- Fakten (z.B. Beobachtungen, Messwerte, Gesetzmäßigkeiten) und
- Modelle (z.B. geologische Karten und Profile, 3d Modelle).

Um Daten entsprechend zu qualifizieren, kann man in GML Metadaten - so genannte Feature Catalogs - nutzen. Damit ist jedoch die Semantik von Geomodellen unzureichend beschrieben. Vor allem in der Wissenschaft sind Modelle oft zahlreich, widersprüchlich, und Veränderungen unterworfen und daher nicht Fakten gleichzusetzen. Da geowissenschaftliches Arbeiten bestrebt ist räumlich-zeitliche Geomodelle zu erstellen, muss das Datenmodell sowohl Fakten als auch die Erstellung räumlich-zeitlicher Geomodelle unterstützen.

Das logische ISO/OpenGIS-Datenmodell umfasst vier Abstraktionsniveaus: *Meta-Metamodel*, *Metamodel*, *Model* und *Instance*. Das Meta-Metamodel spezifiziert Datenmodellierungssprachen wie UML, das Metamodel nutzt diese Sprache um Feature-Typen zu spezifizieren, die Modell-Ebene definiert darauf basierende nutzerguppen- und anwendungsspezifische Features, und Instanzen sind die eigentlichen Geodaten. Um geowissenschaftliches Modelldenken zu unterstützen, ist es notwendig Feature Catalogs (d.h. Modelle) als Daten zu betrachten und nicht als Meta-Daten. Damit können Geomodelle einen Zustand, Verhalten, und Beziehungen besitzen. Geologische Situationen können auf somit, auch in einer Menge, unterschiedlichen Modellen abgebildet werden.

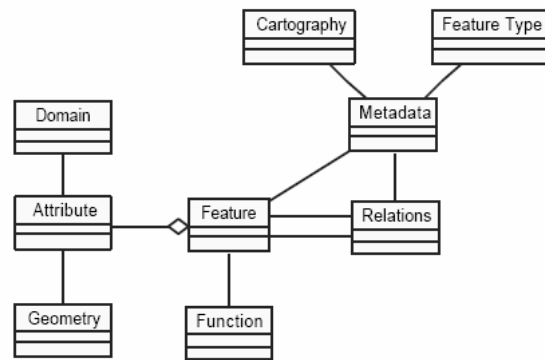


Abb. 1: Vereinfachtes GML Datenmodell.

3 Ein Geomodel-orientiertes Datenmodell

Im OpenGIS Datenmodell wird zwischen Feature Types und Attribute Domains unterschieden. Feature Types bezeichnen reale Objekte (z.B. "Bohrung"), während Attribute Domain den Wertebereich von Eigenschaften von Objekten (z.B. "Verrohrung") festlegt. Jedoch kann es vorkommen dass sich Feature Type und Attribute Domain überlappen: beispielsweise kann "Rhyolith" sowohl ein Feature (z.B. ein Polygon auf einer Karte) als auch einen petrographischen Domain-Wert eines Features repräsentieren. Eine Lösung wäre, Domain Value Definitions und Feature Definitions zusammenzufassen. Ein solcher Catalog entspräche einem konzeptionellen Klassifikationsschema für Geomodelle. Damit ihr Kontext umfassend festgelegt werden kann, sollten geowissenschaftliche Modelle Funktionen, Attribute und Beziehungen zu anderen Modellen, Feature Types und Instances besitzen.

Ein modellorientiertes Datenmodell erschwert die Identifikation von Feature Types und Instances. Das beruht auf Erkenntnis dass in einem Geomodel die Identität eine Features nicht eindeutig durch eine Menge von persistenten Eigenschaften bestimmt ist: Attribute, Funktionen, Beziehungen müssen transient definiert werden, da sie sich im Verlauf des Geomodellierungsprozesses ändern können. Demzufolge ist es notwendig, die Konzepte und Geomodellierungsparameter im Datenmodell zu erfassen. Es sind zwei Fälle zu unterscheiden:

- 1) Mit unterschiedlichen Konzepten erstellte Geomodelle, einschließlich Simulationen. Die einzelnen Geomodelle müssen durch Dokumentation der Konzepte und Modellierungsparameter im Datenmodell reproduzierbar gespeichert werden.

- 2) Mit unterschiedlicher Datengrundlage erstellte Geomodelle. Diese können durch Versionierung der konzeptionell identischen Geomodelle eines Gebietes gespeichert werden. Inkrementelle (d.h. lokale) Versionierung ist bei Geomodellierung mit lokal und global optimierenden Interpolationsalgorithmen wie z.B. DSI nicht sinnvoll.

Abbildung 2 zeigt ein Diagramm, das die wesentlichen konzeptionellen Elemente und Assoziationen enthält: Aufschlusspunkte (ObservationPoints) umfassen eine Eigenschaftsbeschreibung, die auf geowissenschaftlichen Konzepten beruht. Entsprechend geowissenschaftlicher Konzepte, wie z.B. Stratigraphie, wird ein Geomodel bestehend aus einer Menge von Geoobjekten erstellt. Diese Geoobjekte bestehen aus interpolierten Modelldaten und Aufschlussdaten (ObservationPoints). Diese Unterscheidung zwischen Modelldaten und Aufschlussdaten ist auch in Abbildung 3 zu ersehen, die in vereinfachter Form das logische Datenmodell in UML darstellt.

Ein Geomodellierungsprojekt ("GISTriXP") enthält eine Menge Aufschlusspunkte (ObservationPointSet) und geophysikalische Modelle (vorläufig nicht enthalten) als Referenz (in UML: Aggregation) und eine Menge von Geomodellen (in UML: als Komposition). Geomodelle sind 3d topologische Grenzflächenmodelle (boundary representation) zur Modellierung von Strukturen und 3d Gittermodelle (z.B. Voxel) zur Modellierung von Eigenschaften. Der wichtigste Unterschied zum GML-Datenmodell ist: Jedes Modellobjekt enthält Referenzen auf die Ausgangsdaten sowie eine konzeptionelle geologische Beschreibung und ist damit reproduzierbar und dynamisch veränderbar. Von der UML-Repräsentation wurden XML Schemas erstellt, die weitgehend GML-kompatibel sind. Das betrifft unter anderem die Geometrie, Eigenschaften, Metadaten und einfache Datentypen.

4 Zusammenfassung

Das dem GML-Standard des OpenGIS Consortium zugrundeliegende Datenmodell wurde auf die

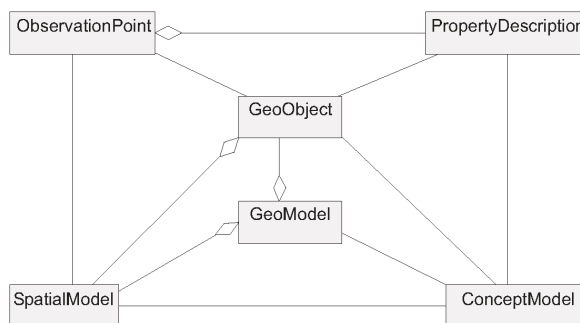


Abb. 2: UML-Entwurf eines konzeptionellen Datenmodells für Aufschlusspunkte und Geomodelle.

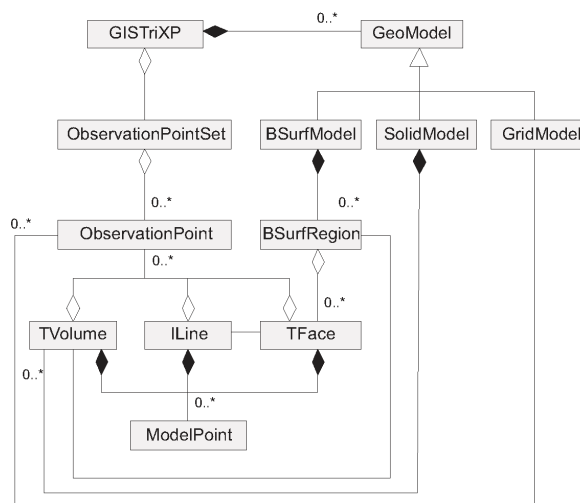


Abb. 3: Logisches Datenmodell eines Geomodellierungsprojektes.

Verwendbarkeit in den Geowissenschaften hin untersucht. Geomodelle besitzen im Gegensatz zu Features (wie z.B. ObservationPoints) eine intrinsische Semantik. Während unveränderliche geowissenschaftliche Primärdaten oft gut mit GML-Elementen abgebildet werden können, wird die GML-Speicherung von Geomodellen aufgrund der Trennung der Semantik von den Geoobjekten als Meta-Daten nicht empfohlen. Für Geomodelle wurde ein möglichst GML-konformes Geomodel-Applikationsschema entwickelt. Dieses umfasst eine Beschreibung der geologischen Konzepte und der Parameter der Modellerstellung.