

Freiheitsgrade in der UML/GML-basierten Geodatenmodellierung und deren Auswirkungen auf die Harmonisierung von INSPIRE-Datensätzen

Roland Grillmayer¹, Katharina Schleidt¹, Jan Schulze-Althoff¹

¹Umweltbundesamt GmbH · roland.grillmayer@umweltbundesamt.at

In den letzten Jahren ist eine Trendwende im Bereich der Geodatenmodellierung in Gange. Wurden früher die meisten konzeptionellen Geodatenmodelle anhand von ER-Diagrammen semantisch modelliert und im Anschluss im Rahmen des Implementierungsprozesses in ein logisches Datenbankschema übergeführt, so setzt sich in den letzten Jahren immer mehr der Trend zur modellgetriebenen Softwareentwicklung (MDS) auch im Bereich der Geodatenmodellierung durch. Verstärkt wird dieser Trend einerseits durch den vermehrten Einsatz der Geography Markup Language (GML – PORTELLE 2007) und andererseits durch die Integration dieses Entwicklungsparadigmas in verschiedener proprietärer (ENTERPRISE ARCHITECT 2016) und freier verfügbare (SHAPECHANGE 2016) Softwareprodukte. Für die konzeptionelle Modellierung von domänenspezifischen Geodatenmodellen in GML kommt meist die Unified Modelling Language (OMG UML 2015) zum Einsatz. Im Rahmen des ISO Standard ON EN ISO 19103: 2015 – Conceptual schema language wird ein UML-Profil für die Modellierung von Geodaten definiert. Zusammen mit dem ISO Standard ON EN ISO 19109:2006 – Rules for application schema, der die wesentlichen Komponenten einer UML-basierte Geodatenmodellierung zur Verfügung stellt, stecken diese beiden Standards den Rahmen für die UML-basierte Geodatenmodellierung ab. Das Encoding des konzeptionellen UML-Diagramms in ein physisches GML-Schema wird durch folgende Standards geregelt. Im Standard ON EN ISO 19118: 2012 – Encoding finden sich die allgemeinen Konzepte zur Übersetzung von konzeptionellen Geodatenmodellen in XML Applikationsschemas, welche im Annex E des frei verfügbaren OGC Standards GML 3.2 (PORTELLE 2007) konkrete Anwendung finden. Des Weiteren findet man im INSPIRE-Dokument Generic Conceptual Model (D2.5: Generic Conceptual Model) im Anhang F „Example for an extension to an INSPIRE application schema“ weiterführende Regeln, wie man existierende INSPIRE-Datenspezifikationen erweitern kann. Diese Standards stellen die wesentliche Grundlage für die Softwareimplementierung und die heute erreichte Praxistauglichkeit der MDS für die Entwicklung von GML-Applikationsschema dar.

1 Freiheitsgrade im Rahmen der Geodatenmodellierung

Trotz der zahlreichen verfügbaren internationalen Standards, welche das Ziel haben, Geodatenmodellierung so gut wie möglich zu präzisieren, ist man beim praktischen Einsatz von UML im Rahmen der Geodatenmodellierung immer wieder mit Modellierungsvarianten konfrontiert. Die Freiheitsgrade können zwar durch domänen- bzw. unternehmensspezifische Profile und Regelwerke eingeschränkt werden [5], eine völlige Unterbindung der Variabilität ist aber nicht möglich. Eine zu restriktive Einschränkung der Freiheitsgrade sollte aber aus Sicht der Entwickler von GML nicht angestrebt werden, um die universelle Nutzung von GML zu gewährleisten. Vor diesem Hintergrund wurde beim ISO TC211 die AdHoc-Gruppe „Best practices for UML modelling“ etabliert. Ziel der Gruppe ist es Erfahrungen und nutzbare

Beispiele für die UML-basierte Geodatenmodellierung zu Verfügung zu stellen und dazu beizutragen, dass diese Modelle sowohl für den Menschen als auch für die Maschine leicht und eindeutig interpretierbar sind.

2 INSPIRE – Freiheitsgrade in der Modellierung am Beispiel der Datenspezifikationen für Landbedeckung und Landnutzung

Die Möglichkeiten dieselben Sachverhalte sehr unterschiedlich zu modellieren, stellt für Initiativen wie INSPIRE, mit ihren hochkomplexen Projektprozessen und der großen Anzahl an Akteuren, einen kritischen Aspekt dar.

Bedingt durch die Freiheitsgrade in der Modellierung stößt man bei der Analyse verschiedener INSPIRE-Datenspezifikationen immer wieder auf unterschiedliche konzeptionelle Ansätze für ein und denselben Sachverhalt. Diese unterschiedlichen Modellierungskonzepte führen in weiterer Folge zu unterschiedlichen Implementierungen folgender für die Umsetzung von INSPIRE benötigter Prozesse:

- 1.) Unterschiedliches Encoding gleicher Sachverhalte in den GML-Applikationsschemata der INSPIRE-Datenspezifikationen.
- 2.) Unterschiedliche Namenskonventionen in den UML-Diagrammen und den daraus abgeleiteten GML-Applikationsschemata der INSPIRE-Datenspezifikationen (Beispiele: Dataset vs. DataSet; Unit vs. Object).
- 3.) Unterschiedliche Mapping Konzepte und Harmonisierungsstrategien im Rahmen der Datenharmonisierung.
- 4.) Teilweise wenig praxistaugliche Instanziierungen von INSPIRE-konformen “*Direct access download service(s)*” (Web Feature Service).
- 5.) Daraus folgend auch unterschiedliche GML-Instanzen mit verschiedenen Eigenschaften gegenüber Diensten und GI-Systemen.

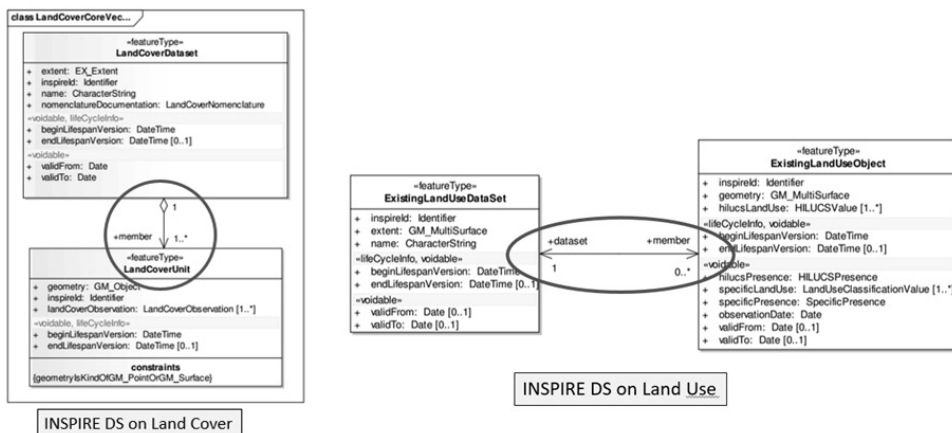


Abb. 1: Unterschiedliche Realisierung der Relation zwischen dem Feature Type LC/LU Dataset und dem Feature Type LC/LU Unit

3 Landbedeckung (LC) vs. Landnutzung (LU)

Die oben skizzierte Problematik wird anhand der unterschiedlichen Handhabung der Modellierung der Relation zwischen dem Feature Type *LC/LU Dataset* und dem Feature Type *LC/LU-Unit* erörtert.

Während in der Datenspezifikation für Landbedeckung die Relation zwischen dem Feature Type *lcv:LandCoverDataset* und Feature Type *lcv:LandCoverUnit* als Aggregation etabliert wird, wird in der Datenspezifikation für Landnutzung diese Relation als bidirektionale Assoziation etabliert (Abb. 1).

Werden GML-Instanzen entsprechend dem INSPIRE-Applikationsschema für Landbedeckung codiert, so wird oft der Feature Type *lcv:LandCoverUnit* als „Nested Property“ *lcv:member* innerhalb des Feature Type *lcv:LandCoverDataset* realisiert (Abb. 2). Grundsätzlich gestattet das Schema auch die getrennte Bereitstellung von *lcv:LandCoverUnit*, und dessen Referenzierung aus dem *lcv:LandCoverDataset* mittels xLink in der Feature Property *lcv:member*. Ein wesentlicher Nachteil dieser Form der Realisierung ist, dass es keine Möglichkeit gibt, aus der einzelnen *lcv:LandCoverUnit* den dazugehörigen *lcv:LandCoverDataset* zu ermitteln, da es sich hierbei um eine unidirektionale Assoziation handelt. Durch diese Variante des Encodings würden somit essenzielle Informationen, wie zum Beispiel die für den Datensatz verwendete Landbedeckungsnomenklatur verloren gehen.

```

▼<gml:boundedBy>
  ▼<gml:Envelope srsName="EPSG:3042" srsDimension="2">
    <gml:lowerCorner>860648.69 5426007.469</gml:lowerCorner>
    <gml:upperCorner>878212.243 5429200.77</gml:upperCorner>
  </gml:Envelope>
</gml:boundedBy>
▼<gml:featureMembers>
  ▼<lcv:LandCoverDataset gml:id="FR0011">
    <gml:metaDataProperty xlink:href="http://144.76.207.166:8080/geonetwork/srv/ger/xml/metadata.gml">
    </gml:metaDataProperty>
    <lcv:inspireId>...</lcv:inspireId>
    <lcv:beginLifespanVersion>2015-05-15T11:22:45+01:00</lcv:beginLifespanVersion>
    <lcv:endLifespanVersion xsi:nil="true" nilReason="unpopulated"/>
    <lcv:extent>...</lcv:extent>
    <lcv:name>Urban Atlas 2012 - Paris</lcv:name>
    <lcv:nomenclatureDocumentation>...</lcv:nomenclatureDocumentation>
    <lcv:validFrom>2012-01-01</lcv:validFrom>
    <lcv:validTo>2017-12-31</lcv:validTo>
    ▼<lcv:member>
      ▼<lcv:LandCoverUnit gml:id="FR0011-17192">
        ▼<lcv:inspireId>
          <base:Identifier>...</base:Identifier>
        </lcv:inspireId>
        <lcv:beginLifespanVersion>2015-05-15T11:22:43</lcv:beginLifespanVersion>
        <lcv:geometry>...</lcv:geometry>
        ▼<lcv:landCoverObservation>
          ▼<lcv:LandCoverObservation>
            <lcv:class xlink:href="http://dd.eionet.europa.eu/vocabulary/landcover/UA2012/11100"/>
            <lcv:mosaic>...</lcv:mosaic>
            <lcv:observationDate>2012-01-01T00:00:00</lcv:observationDate>
          </lcv:LandCoverObservation>
        </lcv:landCoverObservation>
        </lcv:LandCoverUnit>
      </lcv:member>
    </lcv:member>...</lcv:member>
  </lcv:LandCoverDataset>
</gml:featureMembers>
</gml:FeatureCollection>

```

Abb. 2: Resultierendes Encoding der Aggregation zwischen *LandCoverDataset* und *LandCoverUnit* bei der Ableitung des GML-Schema aus dem UML-Diagramm.

In der Landnutzung hingegen wird die Relation mittels eines einfachen xLink Mechanismus etabliert. Features des Typen `eluc:ExistingLandUseDataset` und Features des Typen `eluc:ExistingLandUseUnit` werden unmittelbar als Kind-Element in einer Feature Collection encoded, wobei von der jeweiligen `eluc:ExistingLandUseUnit` mittels `xlink:href` eine Referenzen auf den `eluc:ExistingLandUseDataset` etabliert wird (Abb. 3). Zugleich ist es aufgrund der bidirektionalen Assoziation möglich, eine symmetrische Referenz mittels `xlink:href` von dem `eluc:ExistingLandUseDataset` Objekt auf die `eluc:ExistingLandUseUnit` Objekte bereitzustellen.

```

<gml:featureMember>
  <eluc:ExistingLandUseDataset gml:id="idS2005">...</eluc:ExistingLandUseDataset>
</gml:featureMember>
  <gml:featureMember>
    <eluc:ExistingLandUseObject gml:id="id118b6d92-8083-4a61-b0f6-7b89069d16d2">
      <eluc:inspireId...</eluc:inspireId>
      <eluc:beginLifespanVersion>2015-05-21T15:19:08</eluc:beginLifespanVersion>
      <eluc:geometry...</eluc:geometry>
      <eluc:hilucsLandUse xlink:href="http://inspire.ec.europa.eu/codelist/HILUCSValue/4_1_TransportNetworks"/>
      <eluc:hilucsPresence xsi:nil="true"/>
      <eluc:specificLandUse xsi:nil="true"/>
      <eluc:specificPresence xsi:nil="true"/>
      <eluc:observationDate>2011-08-01</eluc:observationDate>
      <eluc:validFrom>2005-01-01</eluc:validFrom>
      <eluc:validTo>2008-12-31</eluc:validTo>
      <eluc:dataset xlink:href="idS2005"/>
    </eluc:ExistingLandUseObject>
  </gml:featureMember>

```

Annotations in the image: "Etablierung der Assoziation durch Referenzieren des Datensatzes" points to the `<eluc:dataset xlink:href="idS2005"/>` tag. "Etablierung der Assoziation durch Referenzieren des Datensatzes" also points to the `<eluc:dataset xlink:href="idS2005"/>` tag. "Etablierung der Assoziation durch Referenzieren des Datensatzes" also points to the `<eluc:dataset xlink:href="idS2005"/>` tag.

Abb. 3: Resultierendes Encoding der Assoziation zwischen LandUseDataset und LandUse-Unit bei der Ableitung des GML-Schema aus dem UML-Diagramm.

Die oben skizzierten unterschiedlichen Realisierungen in den beiden GML-Schemata erfordern in weiterer Folge im Rahmen der Datenharmonisierung unterschiedliche Harmonisierungsstrategien sowie Mapping Konzepte, die sich in ihrer Komplexität deutlich unterscheiden. So lässt sich die Harmonisierung von Datensätzen zum Thema Landnutzung mit den im Rahmen der Studie untersuchten Softwarelösungen (HUMBOLDT ALIGNMENT EDITOR 2016 + FME: Online-Reference 2016) relativ einfach umsetzen. Hingegen sind für die Harmonisierung von Datensätzen zur Landbedeckung wesentlich komplexere Mapping-Regeln sowie vorbereitende Arbeiten für die Ausgangsdatsätze erforderlich. Um den Harmonisierungsprozess möglichst einfach zu gestalten empfiehlt es sich, die Ausgangsdaten in ein eigenes Datenbanksystem zu migrieren und das physische Datenbankschema möglichst nahe dem Datenschema der jeweiligen INSPIRE-Datenspezifikation anzunähern. Speziell für komplexe Datenspezifikationen wie die Landbedeckung sollte jeder Feature Type in Form einer Tabelle und die existierenden Relationen zwischen den Feature Typen in dem Datenbankschema realisiert werden.

Die unterschiedliche Modellierung dieser Relation wirkt sich in weiter Folge sowohl auf die Praxistauglichkeit vor INSPIRE-konformen „Direct access download service(s)“ (Web Feature Service) als auch der INSPIRE-konformen Predefined Datasets (Atom Feeds) aus. So ist zum Beispiel die Darstellung der `eluc:ExistingLandUseUnits` Geometrien durch das Encoding als direkte Element einer Feature Collection in allen gängigen proprietären und Open-Source-GIS-Programmen möglich. Hingegen scheitern gängige GIS-Clients beim Rendern der Geometrien der einzelnen `lcv:LandCoverUnits`. Das erklärt sich durch die Codierung dieser in Form einer „Nested Property“ `lcv:member`.

Diskurs

Abschließend stellt sich die Frage: Ist eine der beiden Modellierungen korrekt bzw. zu bevorzugen? Darauf lässt sich keine verbindliche Antwort finden; je nach Sichtweise sollte die eine bzw. andere Realisierung präferiert werden.

Aus Sicht des UML-Geodatenmodellierers ist die semantisch korrekte Etablierung der Relation zwischen einem Datensatz und den Dateneinheiten eine Aggregation. Diese signalisiert eine enge Bindung zwischen Datensatz und Dateneinheit, welche die konkreten Geoobjekte und deren Eigenschaften abbildet. In der GML-Codierung kann dies in Form einer Schachtelung der Features in den Datensatz dargestellt werden.

Aus Sicht des Praktikers kann eine einfache, bidirektionale Assoziation zwischen dem Datensatz und der Dateneinheit sinnvoll sein. Zwar wird die Enge der Bindung nicht abgebildet, aber die erzeugten GML-Applikationsschemas sind unkompliziert in der Nutzung mit Standardwerkzeugen.

Ohne Präferenzen für eine der beiden Varianten, stellt die Realisierung unterschiedlicher Modellierungskonzepte für ein und denselben Sachverhalt für den gesamten INSPIRE-Prozess eine unnötige Hürde dar, die im Rahmen der Konsolidierung der Datenspezifikationen behoben werden sollte.

4 Ausblick

Anhand weiterer Beispiele soll verdeutlicht werden, dass die unterschiedliche Handhabung von Modellierungskonzepten für ein und denselben Sachverhalt die ohnehin schon sehr komplexen Prozessketten, die für eine erfolgreiche Umsetzung von INSPIRE benötigt werden, unnötig erschweren.

Als Beispiel wird die unterschiedliche Handhabung und Darstellung von Einschränkungen (Constraints) in UML-Diagrammen zur Diskussion gestellt, wie auch deren Umsetzung in konkrete Validierungsroutinen. Mit einer kritischen Auseinandersetzung über die semantisch korrekte Verwendung von Datentypen des ISO/TC211 und INSPIRE-Basistypen endet der Beitrag.

Die im Rahmen der Studie gemachten Erfahrungen und oben skizzierten Diskussionspunkte werden zurzeit bei der Ad-hoc-Gruppe „Best practices for uml modelling“ des ISO/TC211 zur Diskussion gestellt. Erste Ergebnisse liegen voraussichtlich im Juni vor und fließen in den vorliegenden Beitrag ein.

Literatur

D2.5: Generic Conceptual Model, Version 3.4.

http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.5_v3.4.pdf.

ENTERPRISE ARCHITECT: <http://www.sparxsystems.de/uml/> (30.01.2016).

FME: <http://www.safe.com/> (30.01.2016).

HUMBOLD ALIGNMENT EDITOR: <http://www.esdi-community.eu/projects/hale> (30.01.2016).

- OMG UML – OMG UNIFIED MODELING LANGUAGE (2015): OMG Document Reference Nummer: formal/2015-03-01. Normativ Reference: formal/2015-03-01.
<http://www.omg.org/spec/UML/2.5>.
- PORTELLE, C. (2007): OpenGIS® Geography Markup Language (GML) Encoding Standard. Reference number of this OGC® document: OGC 07-036.
http://portal.opengeospatial.org/files/?artifact_id=20509 (30.01.2016).
- SEIFERT, M. (2007): Das AFIS-ALKIS-ATKIS-Anwendungsschema als Komponente einer Geodateninfrastruktur. In: Zeitschrift für Geodäsie, Geoinformation und Landmanagement (zfv), 130 (2).
http://geodaesie.info/system/files/privat/zfv_2005_2_Seifert.pdf (30.01.2016).
- SHAPECHANGE: <http://shapechange.net/> (30.01.2016).