

Konformität von Geodaten (Qualitätssicherung)

Beschreibung

Amtliche Geodaten und Webdienste werden heutzutage unter Beachtung internationaler Standards und über selbst definierte Produktspezifikationen (incl. Erfassungskriterien) erstellt. Während der Produkterzeugung und nach Erstellung eines Produkts (z.B. eines Datensatzes, einer Karte) wird die Qualität des Ergebnisses auf Genauigkeit (räumlich, zeitlich und thematisch), Vollständigkeit und Konsistenz geprüft. Für die Beschreibung von Qualitätsmerkmalen von Datensätzen und die Festlegung von Erfassungskriterien gibt es im Interoperabilitätskonzept eigene Elemente.

Mit der Einführung moderner Informationstechnologien und umfassenden, sich beständig weiterentwickelnden Datenmodellen werden Qualitätskriterien zunehmend komplexer und deren Einhaltung schwieriger zu kontrollieren. Außerdem sind diese häufig in Textform und nicht standardisiert beschrieben, was Interpretationsspielräume offen lässt und keine automatisierte Prüfung unterstützt. Die Erfahrung zeigt, dass qualitätsgesicherte amtliche Daten mit Fehlern behaftet sein können. Datenfehler können u.a. auftreten durch:

- Fehlerhafte XML-Schemadateien
- Falsche Interpretation der Inhalte des Datenmodells
- Missachtung von Regeln zur Bildung und Änderung von Objekten
- Ungenauigkeiten bei der Datenakquisition und durch Datentransformationen
- Missachtung von Konsistenzbedingungen
- Technische Unzulänglichkeiten und Fehler in der GIS-Software
- Unstimmigkeiten bei aneinandergrenzenden oder sich überlappenden Datensätzen die zusammengeführt werden.

In der Konsequenz sind die Geodaten potentiell nicht mehr interoperabel verwendbar. Betroffen sind davon die Datennutzer, aber auch die Datenanbieter, wenn beispielsweise Datenbestände aus den Ländern zur zentralen Bereitstellung zusammengeführt werden müssen.

Bei der Qualitätssicherung kommen in der Regel automatisierte Prüfwerkzeuge zum Einsatz. Dabei erfolgt dieser Prüfprozess meist durch die verwendete GIS-Software mehr oder weniger in einer Blackbox, wobei oft nicht klar ist, was genau getestet wird und welche Fehler möglicherweise von einer Produktionssoftware toleriert bzw. nicht geprüft werden. Zu einem Interoperabilitätsproblem wird dies dann, wenn aufgrund fehlerhafter Daten die Zusammenführung von Datenbeständen oder die Analyse von Daten unmöglich werden. Ein transparenter Qualitätssicherungsprozess ist somit eine zentrale Anforderung für einen interoperablen Datentransfer. Zum Validieren der Konformität Geodaten zu den Vorgaben des Datenmodells werden unabhängige Test Frameworks oder Testsuites verwendet.

Ein Datenmodell enthält in verschiedenen Bereichen Qualitätsanforderungen (z.B. Kardinalitäten, textliche Konsistenzbedingungen, Erfassungsregeln, Einschränkung der erlaubten Geometrietypen etc.). Solche Regeln sind in maschinenlesbaren Code zu überführen und mit einer Testumgebung (Testsuite) zu prüfen. Qualitätsanforderungen, die schon jetzt relativ einfach mit Standardwerkzeugen überprüft werden können, wie z.B. die Validität einer XML-Datei, werden hier nicht betrachtet.

In der GDI-DE sind grundsätzlich sämtliche Vorgaben aus Spezifikationen, die den Datentransfer regeln (Daten, Metadaten und Dienste sowie Konventionen dazu), auf Einhaltung zu prüfen.

Aktueller Stand in INSPIRE und der GDI-DE

Mit der GDI-DE-Testsuite steht bereits ein Werkzeug für den Konformitätstest von Metadaten und Diensten zur Verfügung. Die Testsuite bietet derzeit Tests für die Prüfung von:

- Metadaten,
- Katalog-/Suchdiensten (CSW),
- Karten-/Darstellungsdiensten (WMS) und
- Downloaddiensten (WFS, Atom)

Die Testsuite prüft noch nicht:

- Daten (uploads)
- Daten von Downloaddiensten (WFS, WCS)

Auf INSPIRE-Ebene existiert seit 2017 ebenfalls eine Testsuite für Daten (INSPIRE Validator, <http://inspire-sandbox.jrc.ec.europa.eu/validator/>), die jedoch nur „INSPIRE-Konformität“ prüft, aber keine national festgelegten Konventionen oder zulässige Ergänzungen und Einschränkungen zu den INSPIRE-Vorgaben. Eine Möglichkeit zur Erweiterung um solche nationale Festlegungen ist auf europäischer Ebene nicht vorgesehen. Eine Nutzung des INSPIRE-Validators für INSPIRE-Konformitätstest innerhalb der GDI-DE ist jedoch möglich und braucht in der GDI-DE-Testsuite nicht noch einmal implementiert werden.

Vom OGC wird mit dem OGC Web Validator (<http://cite.opengeospatial.org/teamengine/>) ein Werkzeug zur Validierung von Implementierungen von zahlreicher Dienststandards angeboten. Für Datenstandards wird derzeit die Validierung von GML (Version 3.2.1) und OGC KML (Version 2.2) sowie GeoPackage (Version 1.0, beta-Status) angeboten. Es können keine Daten gegen konkrete eigene Anwendungsschemata geprüft werden.

Bewertung und Handlungsbedarf

Konformitätstests bestehen aus zwei Komponenten:

1. Konformitätsklassen, von Modellieren nach fachlichen Vorgaben für die zu prüfende Datenspezifikationen festgelegt
2. Eine (unabhängige) Prüfsoftware, die einen Datensatz auf die Einhaltung der Konformitätsklassen überprüft und entsprechend die Testergebnisse ausgibt

Innerhalb einer Konformitätsklasse sind Testparameter festgelegt, welche jeweils zu erfüllen sind.

Testparameter können in Abhängigkeit vom Datenthema unter anderem sein:

- Die Einhaltung der im Datenschema vorgegebenen Datenstrukturen (z.B. GML), Datentypen (z.B. Integer, Geometrietypen), Regeln für Identifikatoren und Relationen zwischen Objekten (Kardinalitäten)
- inhaltlichen Anforderungen (z.B. Vollständigkeit von Pflichtattributen)
- Zusätzliche explizite Regeln oder Abhängigkeiten zwischen einzelnen Datenelementen welche nicht im Datenschema enthalten sind
- logische Konsistenz von einzelnen Geometrien (u.a. Topologie),
- Konsistenz der Geometrien in einem Datensatz (z.B. Prüfung disjunkter Polygone oder geschlossene Flächenabdeckung der Polygone im Datensatz)
- Topologische Integritätsbedingungen bei Netzwerkdaten (z.B. Knoten-Kanten-Bedingungen)

- Weitere semantische Integritätsbedingungen (z.B. konsistente topologische Relationen mit den Objekten anderer Datenthemen)

Die Testparameter sollten in der GDI-DE in UML und OCL formalisiert bzw. textuell beschrieben werden.

Wenn bei Prüfungen Fehler aufgedeckt werden, wird in den meisten Fällen eine Beseitigung dieser Fehler notwendig. In Ausnahmefällen können weniger gravierende Fehler auch toleriert werden. Dies ist bei der Definition der Konformitätsklassen anzugeben.

Handlungsbedarf:

Maßnahmen	<u>Federführer, Akteure</u>
Erweiterungsmöglichkeit der GDI-DE-Testsuite um Datentests vorsehen	<u>AK Architektur</u>
Bei Bedarf der Implementierung von Datentests in der GDI-DE-Testsuite (Klärung der Finanzierung, Zusammenarbeit etc.)	<u>GDI-DE Betriebsstelle</u> Datenmodellierungsexperten nach fachlichen Vorgaben, Fachanwender