Web-based Validation of Cadastral Data in Germany

Markus SEIFERT
DVW WG 2 Geoinformation,
Bavarian Agency for Digitization, High Speed Internet and Surveying, Germany

Key words: AdV, data modelling, quality check, test framework, official reference data

SUMMARY

Since the early beginning official data is collected following requirements of consolidated product specifications and their correctness (conformance to the requirements) is checked. Usually this quality check is done by the GIS implementation, which can be considered as a black box testing, where as it is often not clear, what is tested in detail and which errors are tolerated by the software. The result is, that problems occurs when central data providers try to merge data sets coming from different states, although they have passed a quality check by the responsible data collector.
The software-independent test of spatial data with an online test framework is a fairly new discipline compared to the test of web services. Additionally, it is very complex and technically demanding, with up to now unknown consequences for the performance of such test processes. Therefore, it should be figured out in a pilot project how feasible an implementation is by using representative test criteria.

With this pilot project, the surveying and cadastral authorities in Germany (AdV) wanted to address the following two issues:

- Evaluation of technical possibilities and limits of software independent data tests.
- Starting a common process on the definition of a comprehensive and consolidated set of test criteria in order to allow reliable conformity statements of the produced data.

The German data model for reference data (AAA application schema) comprises quality requirements in several areas. The main focus of this project is the automated test of requirements which are only textually formulated consistency requirements as well as data capture rules. Quality requirements, which can already be tested by commercial tools, such as a valid XML document, are not focus of this presented work. This article documents the results of the pilot project.

ZUSAMMENFASSUNG

Amtliche Geodaten und Webdienste werden heutzutage unter Beachtung von Richtlinien (z.B.

Produktspezifikationen) erstellt und auf Richtigkeit, also auf Konformität zu diesen Vorgaben geprüft. Kommen automatisierte Prüfwerkzeuge zum Einsatz, 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 werden. Wie auch immer, aufgrund fehlerhafter Daten und nicht konformer Webdienste kommt es bei der Zusammenführung von Datenbeständen immer wieder zu Problemen, obwohl sie eigentlich einen Qualitätssicherungsprozess beim Datenerfasser durchlaufen haben müssten.

Das Testen von Geodaten unabhängig von einer GIS-Software in einer Testumgebung im Internet (auch Test Framework oder Testsuite genannt) ist im Vergleich zum Testen von Webdiensten eine vergleichsweise neue, komplexe und technisch sehr aufwendige Disziplin, mit oft noch unbekannten Folgen für die Performance von Prüfläufen. Daher sollte zunächst im Rahmen eines Pilotprojektes die Realisierbarkeit anhand repräsentativer Prüfkriterien nachgewiesen werden. Vorranging sollte dies für AFIS-ALKIS-ATKIS-Bestandsdaten gemacht werden, da hier ein aufgrund der nunmehr flächendeckend verfügbaren Daten ein erheblicher Handlungsbedarf im Hinblick auf eine nachhaltige Qualitätssicherung besteht.

Mit der Pilotierung einer Testsuite für Datentests verfolgte die Arbeitsgemeinschaft der Vermessungsverwaltung der Länder der Bundesrepublik Deutschland (AdV) zwei wesentliche Ziele:

1. Die Ermittlung technischer Möglichkeiten und Grenzen von unabhängigen Datentests.
2. Die Initiierung der Erstellung eines umfassenden abgestimmten Katalogs mit zu prüfenden Testkriterien, damit ein Datensatz vollständig den Anforderungen der NAS und des AAA-Anwendungsschemas der GeoInfoDok entspricht.

Das AAA-Anwendungsschema enthält bereits in verschiedenen Bereichen Qualitätsanforderungen. Schwerpunkt dieses Projekts ist die automatische Überprüfung von Anforderungen, die oft nur textlich in Form von Konsistenzbedingungen und Bildungsregeln in der GeoInfoDok formuliert sind. Qualitätsanforderungen, die schon jetzt relativ einfach mit Standardwerkzeugen überprüft werden können, wie z.B. die Validität einer XML-Datei, wurden nur am Rande betrachtet. Dieser Bericht dokumentiert die bisherigen Ergebnisse dieser Pilotierung.

Web-based Validation of Cadastral Data in Germany

Markus SEIFERT
DVW WG 2 Geoinformation,
Bavarian Agency for Digitization, High Speed Internet and Surveying, Germany

1. CONTEXT

Within the German mapping agencies (AdV) the creation of official geospatial reference data has to be conducted under consideration and control of prescribed and agreed quality criteria, which are formulated in different specifications. With the introduction of modern information technologies and comprehensive, constantly evolving data models, such as: As the AFIS-ALKIS-ATKIS application scheme, quality criteria are becoming increasingly complex and difficult to control. In addition, these quality requirements are not standardized, but often described only in text form, which potentially cause errors due to interpretation gaps. Experience shows that even quality-assured official data can also contain mistakes. Data errors can occur due to

- invalide XML schema files
- incorrect interpretation of the contents of the data model
- disregard of rules for creating and modifying objects
- disregard of consistency requirements
- technical shortcomings and errors in the GIS software.

Whatever the reasons for data errors are, they do occur. The consequence is that the geodata are not interoperable. This is felt directly by the data user, but also by the AdV itself, when data sets from the German Laender have to be integrated for central provision, for example at the Central Office for House Coordinates and House Boundaries (ZSHH) or at the Central Office for Geotopography (ZSGT).

Although many GIS software solutions used by the data providers have generally implemented extensive data tests, it is often unknown what exactly is being tested and how to deal with errors, e.g. whether they may even be tolerated to some extent. Despite the officiality of the data and data providers' declarations of conformity issued to the best of their knowledge, it is therefore urgently necessary to define test criteria based on the requirements contained in the AdV specifications and to implement them in a test environment independent of the GIS software.

The GIS-software-independent testing of geodata in a test framework is a relatively new and technically complex discipline compared to the testing of web services, with still unknown consequences for the performance of test runs. Accordingly, feasibility should first be demonstrated in the context of a pilot project on the basis of representative test criteria.

2. THE QUALITY MEASURES FOR THE GERMAN REFERENCE DATA

The AdV has agreed the following key points of the quality assurance system for the geodata of official surveying and mapping:

"Through national regulation, designation and descriptive, quantitative quality features, the AdV identifies and guarantees the quality of the geotopograpical and real-estate descriptive products of official surveying and mapping. National topicality, uniformity, completeness and availability of the products are essential characteristics in this regard. The surveying authorities guarantee compliance with AdV product quality by standardised test procedures and declare conformance with the AdV standards.

The objective is a comprehensive quality assurance for the geodata of official surveying and mapping as a result of the conception and production process. The conception (AAA-basic schema, AAA-technical schema) is task of the state communities represented by the AdV, during which production of the data inventories in harmony with the AAA-application schema is the task of the surveying authority of each individual state.

2.1 Quality Assurance Model

The relationship structure of the aspects to be quality tested is shown in the following quality assurance model for the AAA-application schema:

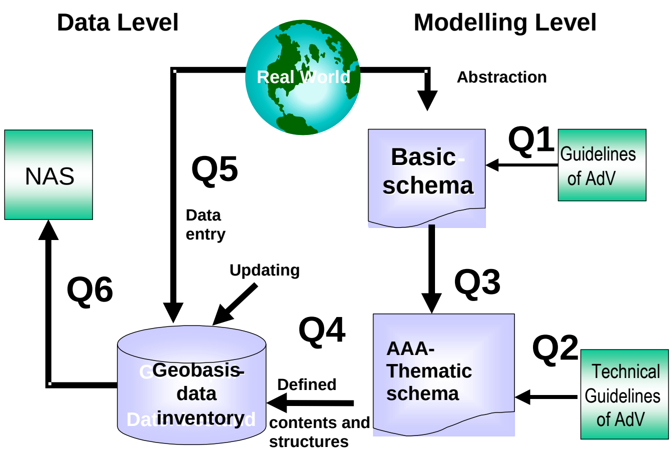

Figure 1: The quality assurance model of the AFIS-ALKIS-ATKIS project

Q1 measures the AAA basic schema against the strategic and technical stipulations of the AdV, Q2 measures the AAA technical schema against the technical stipulations of the AdV. Q3 determines whether the AAA technical schema corresponds to the regulations of the AAA basic schema. Q1, Q2 and Q3 verify the conceptual, internal quality.

Q4 verifies the database internally as a product for logical agreement with the AAA application schema and compliance with the defined quality specifications, while Q5 compares the geodatabase externally with the real world. Q6 relates the quality of the NAS to the user.

The following quality testing schema is derived:

Table 1 – Test schema and responsibilities

		AdV	Laender
1.	**AdV regulations and standards for the development of procedures and program systems**		
	Quality assurance of the AAA-basic schema against stipulations of the AdV (Q1)	X	
	Quality assurance of the common AAA-technical schema against the technical stipulations of the AdV (Q2)	X	
	Quality assurance of the common AAA-technical schema against the AAA-basic schema (Q3)	X	
	Quality assurance of data inventories (ALKIS/ATKIS/AFIS) against the common AAA-application schema (Q4)		X
	Quality assurance of the exchange data against NAS (Q6)	Fundamental prin-ciples	X
2.	**Stipulations for AdV product quality**		
	Stipulation of descriptive and evaluating quality features for unified products including topicality, uniformity, completeness and availability	X	
3.	**Stipulations for quality assurance of the primary database data**		
	Quality assurance of the primary database data against technical reality (Q5)		X
4.	**Quality assurance (as part of quality management)**		
	Conformity declaration by the surveying authority		X

The quality assurance principles for Q6 assume that when data is submitted from AFIS®/ALKIS®/ATKIS®, the created NAS files do not have to be checked against the model. The model-compliant implementation must guarantee this using the valid XML Schema files (XSD); interoperability must be guaranteed. Data acceptance is part of the qualification process. For this purpose, appropriate test tools must be provided which ensure the required quality of the accepted data by using the currently applicable XML Schema files (XSD). When testing exchange data against the NAS schema, a distinction is drawn between testing for a well-formed XML file and testing for validity of the XML file.

2.2 Systems and recording of quality assurance

On the basis of ISO 19105 "Geographic Information – Conformance and testing", abstract test suites (ATS) are to be formulated and used to examine conformity. Each AAA-quality criteria can then be analysed and recorded according to the following schema:

- Theories (conformance requirements).
- Examination solutions, formulated as questions. Each of the questions can result in separate test modules and test cases, which are structured as follows:
  a) Test purpose
  b) Test method
  c) Reference
  d) Test type
- Test for confirming or refuting these theories (executable test suite – ETS with executable test cases).

At the end of the quality assurance scheme of the AdV is the declaration of conformity of the Laender to comply with all specifications of the specifications. The AdV test suite presented here is intended to give Laender sustainable support in quality assurance steps Q4 and Q6 with a component that is independent in terms of GIS software manufacturer and includes all the necessary test criteria so that an AdV declaration of conformity ultimately meets its own requirements in terms of official data.

3. THE TEST SOFTWARE

3.1 Application of existing test frameworks

The interoperability of official spatial data and the services used for the data provision requires that geospatial reference data, geodata services and metadata which are uniformly distributed nationwide, can be used in web-based geographic information systems. This succeeds only under consideration of the numerous community standards, the AdV specifications. The requirements of the AdV specifications form the benchmark of the data quality checks of a test software.

For the quality check of the data and services, a suitable test platform (AdV test suite) is needed in order to operationalize the quality tests. This is not an official certification, but the technical process to verify compliance with requirements from AdV specifications as part of the overall quality assurance of official geospatial reference data described above. Until now, no appropriate standard software for data tests that support the requirements of AdV exists, so the pilot had also the goal of evaluating existing test software, if they are possible to be used and if there are additional requirements to be implemented.

The recently conducted European Spatial Data Infrastructure Network (ESDIN) project has already laid the groundwork for testing extended INSPIRE data to build on. The results, including the ESDIN Test Framework (ETF) tool, are available at http://www.esdin.eu/project/summary-esdin-project-public-deliverables#test . The ETF tool has been used in the European Location Framework (ELF) project, which has been further developed with the goal of better reporting and use through a web application.

To evaluate this test software for AdV purposes, a pattern was developed by thematic experts of the AdV in order to formalize test criteria. Following the study, this pattern was discussed with the working groups responsible for ATKIS (topographical data) and ALKIS (cadastral data), further developed and initially filled by the working groups with prioritized test criteria.

3.2 Results of the pilot implementation

The test software should take into account the content and structural diversity of the existing databases in the AdV. Thus, test data from Berlin were used, which have a very complex inner-city structure. In addition, data volumes of varying sizes were used, all the way to a dataset covering a complete Land, which was possible with ATKIS data from Bavaria. This made it possible to draw conclusions about the performance requirements depending on the file sizes. Comprehensive tests were performed on the following model data provided by the respective surveying authorities for topographical reference data. Table 2 summarizes the most important results for the test runs:

Table 2 – Achieved performance of the test software

Data set	Number of objects	File size GB	Duration of test run
Area of Berlin	164.574	0,6	3 min
Area of Thuringa	1.385.605	5,4	53 min
Half area of Lower Saxony	2.031.780	8	3 h
Area Schleswig Holstein	1.472.860	5,9	1,5 h
Area Bavaria	3.542.920	13,95	6 h

The runtime is determined mainly by the access time of the hard disk: the demo server had 8 GB and 2 Xeon e5-2676 CPUs, but no SSD hard disk.

Based on the objectives of the pilot implementation described above, in particular two important results have been achieved which are preconditions for the subsequent implementation of an operable solution:

1. Although not yet complete, but representative test criteria of the AAA application scheme could be implemented with the used test environment so that meaningful test results were achieved. Last but not least, the complexity of some test criteria led in their various cases to iterative adaptations to their specification or their implementation. For example: An object of object type 53001 AX_ConstructionInTrafficAreas with the value type 1880 for the attribute "building function" always has to have the same geometry of one or more concatenated objects of the object type 42003 AX_StreetAxis, 42008 AX_RoadAxis, 42008 AX_Track or 53003 AX_Path or 42014 AX_RailroadTrack.

2. Some test criteria require expensive arithmetic operations (e.g., large-scale geometical intersections depending on additional requirements). The test framework had to be optimized to allow reasonable computing times. The need for optimization in geometry and topology tests was mainly related to two aspects:

    - Numerous tests deal with topological relationships with other objects. Without a geometric index, corresponding tests cannot be carried out sufficiently quickly with increasing amounts of data.
    - For geometry operations during tests, the GML geometries must be converted to the native representation of the used geometry library. If certain geometries are used more frequently, redundant transformations represent an avoidable loss factor for performance. Caching of the geometries can remedy this situation.

Thus, the geometry component of ETF was extended to support spatial indexing of the objects and caching of the geometries. With the exception of these geometry and topology tests, only little optimization for the piloting was necessary during development. To what extent further optimizations would be necessary for an operational AdV test suite will depend on the performance expectations to be determined.

The following performance measurements were made with the database of Bavaria on a standard notebook (MacBook Pro). The portions used for the test run were varied (the total number of data is 710 portions). Table 3 summarizes the most important test run numbers.

Table 3 - Overview of the test runs with the topographical data from Bavaria

	Test 1	Test 2	Test 3	Test 4	Test 5
Test range (number of tiles)	001-009	001-099	400-710	200-710	001-710
Number of objects	50913	1139856	2449406	5059277	7488518
Duration	74 s	29 min	2 h	4 h	6 h
Duration of one Object	1,45 ms	1,52 ms	3,02 ms	2,95 ms	2,91 ms

Particularly problematic was the test with regard to gapless and overlap-free area coverage of the objects of the object type Land Use, since particularly many topological and geometric operations (joining of all Land use areas) have to be carried out here.

The above described optimizations improved the performance behavior so that a test run on the complete dataset of Bavaria on a normal workstation was finished after only 6 hours. Thus, a significantly improved (linear increasing) calculation time could be achieved instead of a quadratic increasing runtime behavior depending on the number of objects (see Table 2). This may be even more important in the testing of cadastral data, as it requires more area-wide, gapless requirements to be fulfilled, especially the parcels.

Further improvements in the test software used were also made in the documentation of the test results.

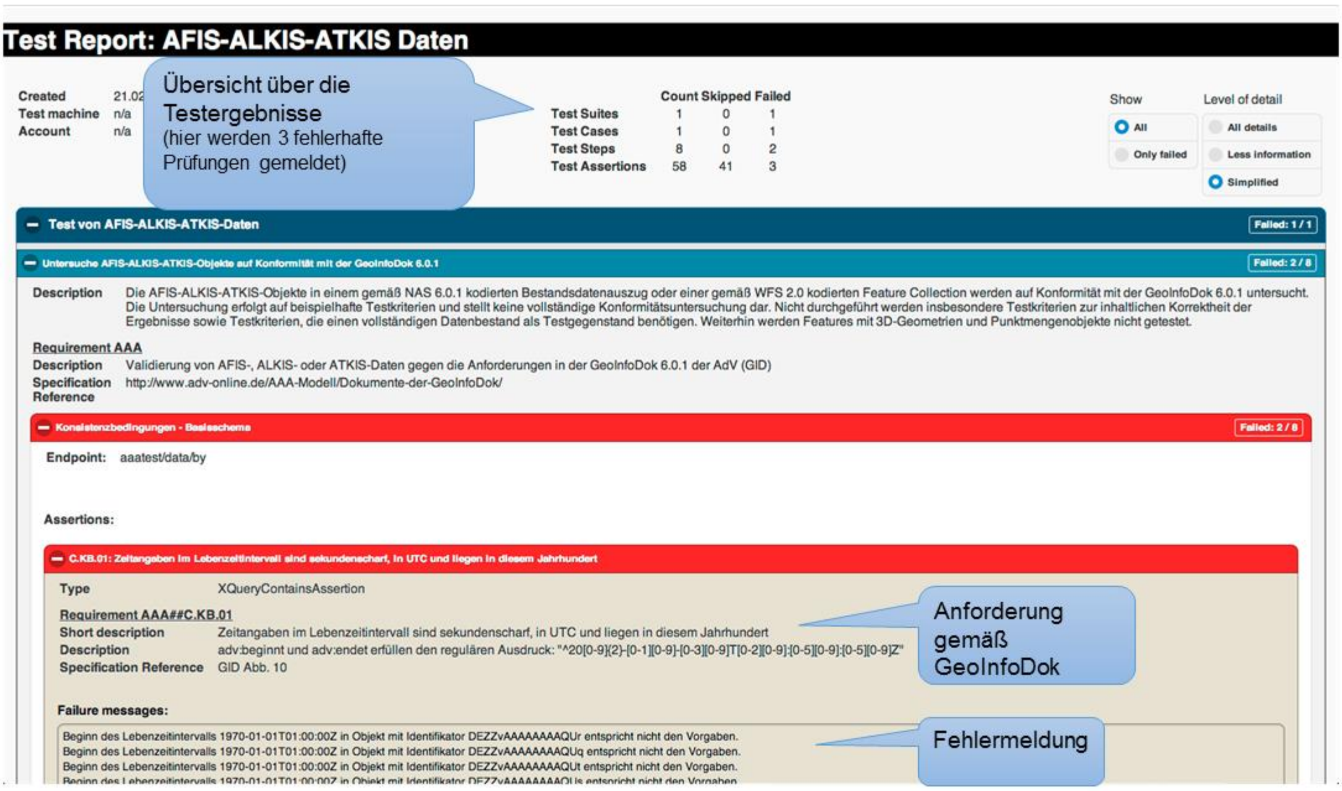

Figure 2: Example of a test report in case of a detected error

In addition to the calculation time, a description of the test including the reference to the corresponding specification is reported for the respective test criterion. This is followed by a list of detected objects with mistakes.

Even though the full compliance check of datasets was not the goal of this pilot implementation, even with the representative test criteria it quickly became clear that no tested dataset was error free. This not only justifies the further development of this pilot implementation, but also makes particularly clear the need for action.

4. CONCLUSIONS

With the pilot project for a test framework, important findings were obtained for the standardized testing of reference data from surveying agencies and thus the precondition for a forthcoming implementation was justified. Based on the experience, the test software used could be significantly optimized so that complex and time consuming test cases (such as topology) are possible with acceptable performance.

Based on the results of the piloting, the "Overall AdV Test Suite Concept" was approved by the AdV plenary meeting in autumn 2017 and is now to be brought into practical operation step by step. This will start with the test criteria for the AAA application scheme, followed by the test criteria for AdV metadata and AdV web services.

It has been shown that the use of an AdV test suite can improve the quality of the sometimes very complex databases sustainably through the systematic and permanent data check and the corresponding corrections. The centralized data providers of the AdV, which will bring together and provide the central databases of the federal states, as well as transnational users, will benefit directly from this. Quality assurance is an essential precondition for the central authorities of the federal states as well as the AdV for the further automation of work processes and for the reduction of updating cycles. Correct data sets significantly reduce the ongoing effort – considering the ongoing decreasing staff capacities.

The comprehensive catalog of test criteria for the AdV thematic data models can also be implemented in any GIS software. For the first time, there will be a validated and mintainedlist of test criteria agreed within the AdV, which must be successfully passed by any data set to ensure compliance.

The AdV test suite is modularly expandable, which could possibly also be used in the future to test the INSPIRE conformity of AdV reference data, transformed into the INSPIRE data models. Alternatively, a connection to other test machines is possible in principal, such as the INSPIRE Validator of the European Commission. The surveying and cadastral administrations of the federal states in Germany can also ensure conformity with the AdV specifications by using the test suite locally within the framework of the data production and thus meet the high quality standard of harmonized official data.

REFERENCES

Software ETF: https://hub.docker.com/r/iide/etf-webapp/ and https://github.com/interactive-instruments/etf-webapp, latest access 02/2018

Test framework oft he European project ESDIN: https://www.interactive-instruments.de/, latest access 02/2018

AdV product specificaktion for 3D data: http://www.adv-online.de/AdV-Produkte/Weitere-Produkte/3D-Gebaeudemodelle-LoD/ , latest access 02/2018

INSPIRE Reference Validator: https://ies-svn.jrc.ec.europa.eu/attachments/download/1593/SC_235_D4.2.1_-_INSPIRE_Test_Framework_-_design_report_-_v03.docx , latest access 02/2018

BIOGRAPHICAL NOTES

Dr. Markus Seifert is head of the project team “SDI Standards” of the Surveying Authorities of the States of the Federal Republic of Germany (AdV) and is working on the conceptual schema of the AAA data model. Furthermore he represents the Bavarian Organization for surveying and cadaster in several national working groups concerning the standardization of public geospatial data. On behalf of the AdV he is the head of the German delegation at ISO/TC 211 and CEN TC 287 and was chair of the INSPIRE Working Groups “Orthoimagery” und “Protected Sites”. He currently is also in charge with the implementation of the spatial data infrastructure in Bavaria and Germany as head of the SDI office in Bavaria. Markus Seifert is member of the DVW Working Group “Geoinformation” and is national delegate to FIG Commission 7 “Landmanagement”

CONTACTS

Dr. Markus Seifert
Head of the SDI unit
Agency for Digitisation, High Speed Internet and Surveying
Alexandrastr. 4, 80538 Munich
Germany
Telefone: +49 89 2129 1002
Mobile: +49 173 8633042
E-Mail: Markus.Seifert@ldbv.bayern.de
Internet: http://www.gdi.bayern.de