Towards a Quality Aware Web Processing Service

Andreas Donaubauer, Tatjana Kutzner, Florian Straub

Fachgebiet Geoinformationssysteme, Technische Universität München, Germany
{andreas.donaubauer,tatjana.kutzner,florian.straub}@bv.tum.de

Abstract. Quality information is essential to determine the fitness for use of spatial data and furthermore to correctly interpret the results of spatial analyses. Up to now this aspect is still not sufficiently included in geospatial applications and geospatial web services. This paper discusses a concept for a Web Service with the ability to dynamically generate quality information for the results of spatial analyses by taking into account the quality of the input data. The concept is based on the current ISO quality and metadata standards and OGC web service standards. An implemented prototype together with a use case from forestry shows the practicability and relevance of this concept.

1 INTRODUCTION

In computer science web services support the interoperability between heterogeneous software components and hence the ability to exchange information between these components. This concept has been adopted in geoinformatics by employing so-called geospatial web services. These are web services specially designed for providing access to and processing heterogeneous spatial data stored at remote sources.

The Open Geospatial Consortium (OGC) aims at developing standards for accessing these geospatial web services. In the beginning the main focus was set on developing standards for merely accessing spatial data, as for example with the OGC Web Map Service (WMS) or Web Feature Service (WFS), whereas by now there has been a shift towards providing web services for the processing of spatial data. The Web Processing Service (WPS) is such a standard developed by the OGC.

When using spatial data or generating new data with data processing, e.g. by applying spatial analysis functionality provided by a Web Service, it is important to know about the quality of the data. The quality not only informs the user about the fitness for use of the data in certain fields of application, but also enables the user to interpret the results from an analysis.

This paper introduces a concept for a Web Service with the ability to dynamically generate quality information for the results of spatial analyses by taking into account the quality of the input data. Quality information can be added to spatial data by means of metadata. The concept includes a discussion about at which data level to store the metadata and how to integrate the latter into the Geography Markup Language (GML), which will be used as transfer format for the spatial data.

2 RELATED WORK

In the field of geoinformatics ISO has developed the ISO 191xx Geographic information series of standards, including the standard ISO 19113 (2003) which defines quality elements for describing the quality of geographic information and the standard ISO 19115 (2003) for metadata, which amongst others contains these quality elements. In addition, the standard ISO 19139 (2007) was developed to provide a technical specification for encoding the metadata in XML.

For exchanging geospatial data over the World Wide Web several standards have been developed by the OGC, most notably the Geography Markup Language (GML) (2004) which is based on XML to facilitate the modeling and transfer of spatial data. Spatial analysis functionality for processing spatial data can be provided using the Web Processing Service (WPS) (2005) standard, whereas the Web Feature Service (WFS) (2005) standard simply provides for accessing spatial data encoded in GML.

As an example of quality information which can be generated dynamically, this paper focuses on the positional accuracy of spatial data. Based on former work by Caspary, Scheuring and Frank, several models have been developed by Glemser (2000) for integrating information on the accuracy of geometry into spatial data and he describes how to treat this information in spatial analysis.

3 A CONCEPT FOR DYNAMICALLY GENERATING METADATA

3.1 Requirements and Constraints

To be able to determine the quality of spatial data several requirements have been identified resulting in some constraints.

First, we need to have quality information for the input data. This requirement is quite trivial in theory. In practice however, quality information recorded at feature or sub-feature level – although important – can hardly be found.

Second, we need to have a consistent and commonly understood quality description (semantics). A conceptual model for recording quality information is provided by the metadata standard ISO 19115 which is sufficient for our purpose.

Third, a syntax for encoding the quality information is required both in the input and in the output data of the spatial analysis to be carried out. Our concept relies on GML and the ISO 19139 standard in order to include the quality information in the spatial data. However, it should be made clear in this context that GML does not mandate to use ISO 19139.

Fourth, a Web Service interface is required for performing spatial analyses and generating quality information for the output data based on the quality of the input data and the spatial analysis method(s) applied. All spatial analysis methods differ from each other which as a consequence makes it necessary to determine for each spatial analysis method separately the way of generating quality information. This paper however concentrates on generating quality information, in particular spatial accuracy, for the spatial analysis methods polygon overlay and area calculation. The OGC Web Processing Service (WPS) can be used as web service interface, it provides an interface specification which is appropriate for our concept and subsequent prototyping.

3.2 Integration of Metadata and Geospatial Data

This section describes how quality information can be included conceptually into spatial data by using ISO 19115 and logically and physically by using GML and ISO 19139.

3.2.1 Conceptually

The international standard ISO 19115 defines more than 400 metadata elements for describing spatial information and services. Among them are several elements for describing the quality of spatial data and also the positional accuracy, which this concept focuses on, thus making the standard suitable for use here.

Quality information can refer to different levels of the spatial data, i.e. it can either describe the whole dataset (dataset level), individual features (feature level), feature properties, geometry or geometry sections (sub feature level). It is important that quality information can be added to the spatial data at the most appropriate level and not just to the complete dataset. The standard ISO 19115 defines the hierarchy levels dataset, feature and attribute. They correspond to the above mentioned levels and therefore allow for adding the metadata at the desired levels.

To avoid data redundancy, which appears when the complete metadata containing identical information is added to several data levels, a recommendation given by ISO 19115 is followed and all metadata describing general non-changing information, as for example the date of creating the metadata, will be added at dataset level only. Quality information however can vary, thus it is always added directly at the appropriate level.

3.2.2 Logically and Physically

The Geography Markup Language (GML, our concept relies on Version 3.1.1) is an XML-based language which facilitates modeling, transfer and storage of spatial information and also supports the integration of metadata. With GML metadata can be added to the levels dataset, feature and geometry, however, it is not possible to add metadata to a feature property or a geometry section. Here individual solutions have to be found. Metadata of a feature property for instance can be added to the feature level as well, hereby including some information for which property the metadata is meant to be. Quality information intended for geometry sections can be included in the spatial data by using the GML constructs MultiCurve aggregate or topological complex. Both methods are described in detail in Kutzner (2008).

ISO 19115 represents an abstract definition of metadata elements. To include the quality information in GML-encoded spatial data, we use the standard ISO 19139. It contains a consistent implementation of the abstract metadata elements by providing XML schemata for all metadata elements of ISO 19115.

Listing 1 shows how a surface polygon can be represented by a GML MultiCurve aggregate. The border of the polygon is split into GML LineStringSegments (lines 23 and 35) and each LineStringSegment is then provided with quality information on its positional accuracy (lines 6 and 31).

```
1 <app:area gml:id="idfafe29f9-8992-44c8-9718-c100a7bffdca">
2  <gml:multiCurveProperty>
3   <gml:MultiCurve>
4    <gml:curveMember>
5     <gml:Curve>
6      <gml:metaDataProperty>
7       <mdl:GeometryLevelMetadata>
8        <mdl:absoluteExternalPositionalAccuracy>
9         <gmd:DQ_AbsoluteExternalPositionalAccuracy>
10          <gmd:result>
11           <gmd:DQ_QuantitativeResult>
12            <gmd:valueUnit xlink:href="#m"/>
13            <gmd:value>
14             <gco:Record xsi:type="gml:LengthType" uom="#m">2.0</gco:Record>
15            </gmd:value>
16           </gmd:DQ_QuantitativeResult>
17          </gmd:result>
```

```
          </gmd:DQ_AbsoluteExternalPositionalAccuracy>
        </mdl:absoluteExternalPositionalAccuracy>
      </mdl:GeometryLevelMetadata>
    </gml:metaDataProperty>
    <gml:segments>
      <gml:LineStringSegment>
        <gml:posList>0 0 10 0 5 10</gml:posList>
      </gml:LineStringSegment>
    </gml:segments>
  </gml:Curve>
 </gml:curveMember>
 <gml:curveMember>
  <gml:Curve>
    <gml:metaDataProperty>
     ...
    </gml:metaDataProperty>
    <gml:segments>
      <gml:LineStringSegment>
        <gml:posList>5 10 0 0</gml:posList>
      </gml:LineStringSegment>
    </gml:segments>
  </gml:Curve>
 </gml:curveMember>
</gml:MultiCurve>
</gml:multiCurveProperty>
</app:area>
```

Listing 1: GML MultiCurve containing quality information

3.3 A Web Processing Service with the Capability of Generating Metadata at Runtime

The Web Processing Service (WPS) standard was developed by the OGC for providing web-service-based access to any kind of functionality for processing spatial data, may it be simple or complex calculations such as area calculation or polygon overlay.

In accordance with the forth requirement (see section 3.1 above) a WPS is used to provide the functionalities 1 and 2, which can be subdivided respectively into:

- *Functionality 1a* for carrying out spatial analyses, the focus being on polygon overlay in this concept.
- *Functionality 1b* for determining the quality for the analysis result based on the quality of the input data and the spatial analysis method to be performed, here polygon overlay. In this concept the quality information will be constrained to positional accuracy.

- *Functionality 2a* for deriving values from the geometry of spatial data, here we focus on the area calculation of a given polygon. (This functionality is analogous to functionality 1a.)
- *Functionality 2b* for determining the quality of the values derived by functionality 2a, in this concept the accuracy of the area. (This functionality is analogous to functionality 1b.)

A WPS provides functionalities as so-called processes, therefore the defined functionalities are implemented as processes 1 and 2 with process 1 encapsulating functionalities 1a and 1b and process 2 encapsulating functionalities 2a and 2b. The processes are executed as described in the WPS specification by sending an Execute request via HTTP Post.

The data sent to the processes needs to be encoded as a GML FeatureCollection which consists of features representing surfaces. In the prototype each surface polygon of the input data is represented as MultiCurve aggregate (see section 3.2.2 above) and provided with quality information (ISO 19139 integrated in GML). The quality information is read from the input data, functionality 1a or 2a is executed and then the quality of the resulting data is determined according to the propagation model described in section 4.2. Finally the WPS returns a new FeatureCollection with the resulting data and its quality information, which is added at the same level as the quality information of the input data.

4 PROOF OF CONCEPT

Based on the above introduced concept for a quality aware WPS a prototype has been implemented and tested by applying a use case from forestry which is described below. The Java-based framework deegree 2.1 (2007) was used for the WPS. This framework offers implementations of several OGC web services, among them a WPS implementation. To provide the above defined functionalities by the deegree WPS they had to be implemented as processes in Java.

4.1 Use Case

For a given land parcel X the use case will identify the most suitable types of tree and the available surface area in hectares. Two kinds of data are needed:

1. Data about forest stands and the suitability of certain types of trees for these areas
2. A digital land register with information about individual land parcels

First a polygon overlay will be conducted on these data; the result of the overlay will be new surfaces which coincide with land parcel X. For these new surfaces the best types of tree will then be determined and the surface area will be calculated.

By integrating quality information into the result the use case can be expanded to determine the accuracy of the boundaries and of the surface area in addition. Let's assume that tree type Y is best suitable for a certain part of land parcel X, then the area gives the size of that part in ha and the quality gives the possible standard deviation. According to Lother (2003) this information is in forestry especially useful for economic reasons since the size indicates the value of a forest stand and also contributes to salary calculation.

4.2 System Architecture of the Prototype

Figure 1 depicts the components of the defined system architecture. Central component of the system is the so-called aggregate service which controls the whole workflow. A client informs the aggregate service about the data for which a user wants to carry out spatial analyses and for which the user wants the quality to be determined. The aggregate service then accesses the required data for example through WFS services, which return the data encoded in GML. The most important component is the WPS with processes 1 and 2 carrying out functionalities 1a, 1b and 2a, 2b respectively, i.e. here the polygon overlay is carried out, the positional accuracy is determined for the resulting data and the area and area accuracy are calculated.

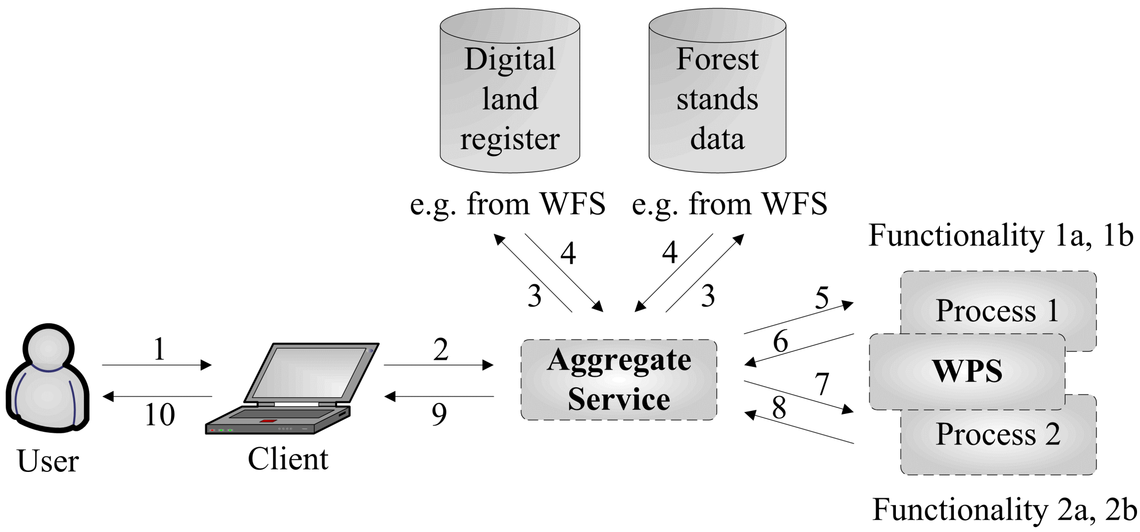

Figure 1: Components of the system

4.3 Accuracy Model and Propagation Model Applied

Figure 2 shows, that it is essential to be able to store the quality information and especially the positional accuracy directly with the geometry of the data. The forest stands have natural and therefore uncertain boundaries which derive from vegetation. These boundaries cause difficulty in exactly locating the border points, so that each forest stand can exhibit varying values of positional accuracy within its border. To know all accuracy values one border can have, it is necessary to store each value directly with the corresponding geometry section of that border. The land parcels in contrast have artificial boundaries with exact border points. For them it is sufficient to store the quality information once for the whole dataset.

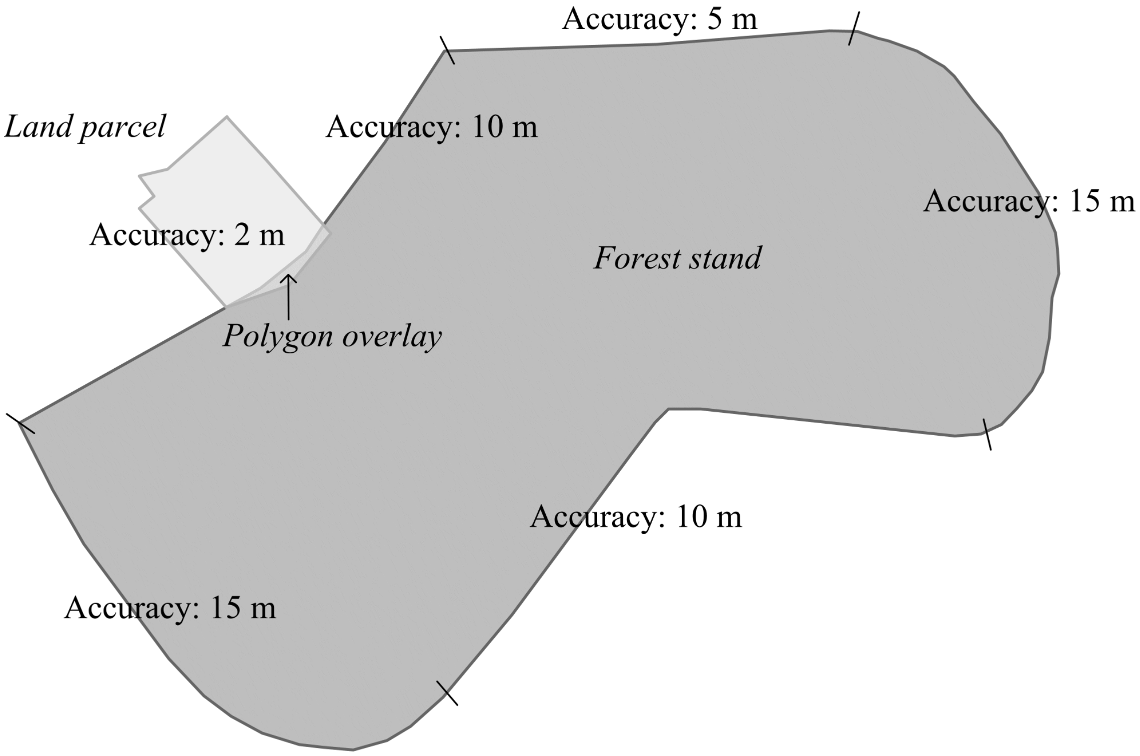

Figure 2: Land parcel and forest stand data with positional accuracies

Storing the quality at geometry level is done using the line-by-line approach described in Glemser (2000). With this approach each polygon can be separated into line segments and the quality is added directly to the corresponding line segment. When applying a polygon overlay on such surfaces, the propagation model for quality information used by functionality 1b (compare section 3.3) is quite simple: qualities of the line segments are simply transferred to the line segments of the newly created polygons.

Figure 3 exemplifies this: the polygon which results from the polygon overlay conducted on the land parcel and forest stand data from figure 2 is shown here. It contains information about the most suitable types of tree and the surface area. By using the quality information from the input data, the resulting polygon can additionally be provided with line segment accuracies describing the border of the polygon.

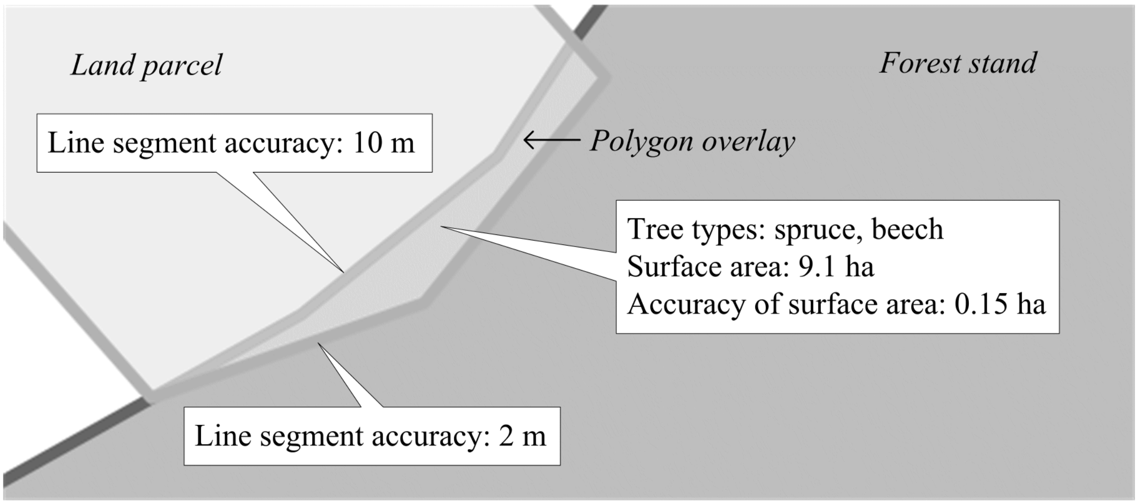

Figure 3: Polygon overlay and accuracy propagation

Based on the quality information associated with the line segments of the polygon, functionality 2b (compare section 3.3) can be carried out in order to calculate the standard deviation of the surface area. For our prototype, we adopted an error propagation model described by Magnussen (1996); compare Lother (2003). A strict approach according to the law of error propagation could be implemented alternatively.

5 Conclusions and Future Work

This paper introduced a concept for integrating quality information into the results of spatial analyses and showed the practicability and relevance of the concept by applying a use case from forestry on the developed prototype.

By using standards of ISO and OGC the feasibility of dynamically generating quality information was demonstrated. Furthermore it was shown that the WPS is suitable for generating quality information at runtime in a dynamic environment and that quality information can be generated at different data levels. The concept is generic and can therefore be applied to various areas of application.

As analysis function the polygon overlay was chosen in this concept to dynamically generate metadata for the overlay result. However, there are several other important analysis functions and it would make sense to find out if the concept can be applied to them as well. Since the concept is generic this should pose no problem.

Another important aspect is the visualization of metadata. GML is only suitable to a limited extent to present information in a user-friendly way and the structure of GML documents is difficult to be understood by users from other scientific fields. GML viewers are available but they must be supplemented with new functionality to be able to display quality information in a useful way.

Furthermore, the concept focuses on positional accuracy as quality information and how to determine this information for the analysis result. In reality spatial data will usually contain many more metadata elements. In future research should therefore be analyzed which other elements are suitable for being generated dynamically based on meta information associated with the input data.

6 REFERENCES

deegree 2.1 (2007). http://www.deegree.org/, last visited 11.05.2008.

Glemser, M. (2000). Zur Berücksichtigung der geometrischen Objektunsicherheit in der Geoinformatik. Dissertation, Universität Stuttgart.

GML (2004). OpenGIS Geography Markup Language (GML) Implementation Specification, Version 3.1.1.

ISO 19113 (2003). Geographic information – Quality principles.

ISO 19115 (2003). Geographic information – Metadata.

ISO 19139 (2007). Geographic information – Metadata – XML schema implementation.

Kutzner, T. (2008). Konzept zur dynamischen Generierung von Metainformation für service-orientierte Geoprocessing Workflows. Diploma thesis, TU München.

Lother, G. (2003). Konzeptionelle Aspekte eines landesweiten Fachgeoinformationssystems für die Bestandsdokumentation forstlicher Geodaten. Dissertation, TU München.

Magnussen, S. (1996). A Coordinate-Free Area Variance Estimator for Forest Stands with a Fuzzy Outline. In: Forest Science 42 1 (1996), S. 76-85

WFS (2005) Web Feature Service Implementation Specification. Version 1.1.0

WPS (2005). OpenGIS Web Processing Service. Version 0.4.0