International Journal of Spatial Data Infrastructures Research, 2012, Vol.7, 66-87.

Web Services for Spatial Data Exchange, Schema Transformation and Validation as a Prototypical Implementation for the LPIS Quality Assurance*

Stefan Wiemann[1], Lars Bernard[1]
Piotr Wojda[2], Pavel Milenov[2], Valentina Sagris[2], Wim Devos[2]

[1]Professorship of Geoinformation Systems, Technische Universität Dresden, Germany
{Stefan.Wiemann|Lars.Bernard}@tu-dresden.de
[2]Institute for Environment & Sustainability European Commission Joint Research Centre, Ispra, Italy
{Piotr.Wojda|Pavel.Milenov|Valentina.Sagris|Wim.Devos}@jrc.ec.europa.eu

Abstract

This article presents an SDI-based approach to implement selected web services within the Land Parcel Identification System (LPIS) Quality Assurance framework, according to the Commission Regulation 1122/2009. The Test Bed uses OGC conforming web services allowing for: (1) agricultural data transformation from national data schemas to the common LPIS Core Model, (2) transferring, validating and storing spatial and non-spatial observations of the quality inspections. The OGC Web Processing Service (WPS) interface specification is used as a basis to allow for interoperable accessing the schema transformation and content validation functionalities of the realised services. The implemented solutions demonstrate the feasibility of the proposed concepts and fit in with current INSPIRE activities.

Keywords: SDI, Agricultural Policy, Geoprocessing Services, Schema Transformation, Content Validation

DOI: 10.2902/1725-0463.2012.07.art4

1. INTRODUCTION

The Common Agricultural Policy (CAP) of the European Union (EU), since its reform in 2003, aims to provide farmers with a stable income, decoupled from production, within a framework of sustainable development of the rural areas while respecting environmental and societal needs. In order to distribute the related funds of the EU, each Member State had to establish its Paying Agency to collect and control the farmers' applications and to provide the respective funds. These procedures are managed through the Integrated Administration and Control System (IACS), which have therefore been established in the Member States. IACS also contains a module handling geographic information: the Land Parcel Identification System (LPIS). The LPIS's main functions are to provide unambiguously the geographical location and spatial extent of all agricultural parcels declared by the farmers and to quantify the area being eligible for funding. Moreover, the system shall support crosschecks during the administrative controls by the paying agency to prevent undue payments. Thus, if the Member State's LPIS fails to localise a reference parcel unambiguously, there is the risk that the parcel gets declared more than once or that inspections get less efficient. Furthermore, inadequate quantification of eligible area bares the risk that crosschecks for identifying and preventing over-declarations by farmers become impossible or at least less effective. To enable the Member States to address such weaknesses a Quality Assurance framework has been introduced by the Commission Regulation (EC) No 1122/2009, calling for an annual testing and reporting on seven prime quality elements.

To guarantee a correct interpretation of its results, the annual testing and reporting requires an intensive and standardized exchange of spatial and non-spatial data between the EU services and Member States. At first, the CAP legislative requirements on the LPIS were translated into the LPIS Core Model being a Geometry Markup Language (GML) application schema. However, on the Member State side many different LPIS implementations were developed. One of the proposed solutions to overcome those differences was to build a schema transformation web service to transform heterogeneous Member State data into the common LPIS Core Model. Additionally, another web service comprising content validation processes was proposed to support the testing and reporting procedure. In realising an SDI-based approach for LPIS the OGC Web Processing Service (WPS, OGC, 2007) standard was chosen to provide the corresponding functionality to the Member States. Concepts and prototypes of such an approach were developed in a study commonly conducted by the European Commission Joint Research Centre and the TU Dresden and are reported in this paper.

The remainder of the paper presents an overview on the LPIS Quality Assurance Framework, to derive the basic requirements on transformation and validation services which can support this framework. This is followed by the presentation on the concepts, service architecture and prototypical implementations of such transformation and validation services. A final discussion on the achieved solutions also links to the related INSPIRE Transformation Services.

2. LPIS QUALITY ASSURANCE

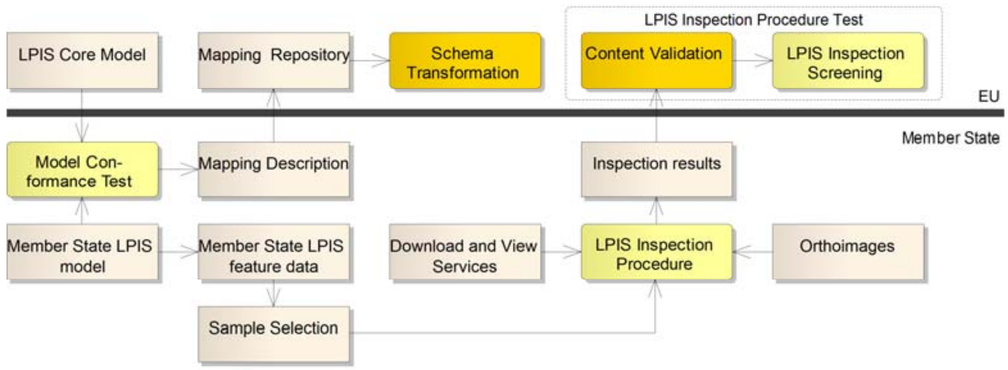
2.1. Overview

Since the CAP legislation sets up requirements but does not provide specific instructions on how the Member States should conceive and implement their LPIS systems, multiple solutions and designs have emerged and a need for harmonization appeared (Sagris et al, 2008). As the first step in the harmonization approach, a common LPIS Core Model has been established for translating directly legislative terms into the geospatial realm (Sagris and Devos, 2009). Thus, the LPIS Core Model defines the basic features including their feature types and properties, which should be present in the LPIS implementation of a Member State. Moreover, it specifies spatial overlays with additional datasets for cross-compliance (e.g. Natura2000 areas) as stipulated by European Commission (2004). In the given context, cross-compliance shall ensure the development of sustainable agriculture and making the CAP more compatible with the expectations of society by making the payments to the farmer dependent upon his respect of certain conditions. These conditions come from two legal sources: Statutory Management Requirements (SMR) in public, animal, plant health, environment and animal welfare that are based on the EU objectives as well as on the Good Agricultural and Environmental Conditions (GAEC) being defined in a national or regional context. For several SMR and GAEC conditions, appropriate geographic areas (e.g. protected sites) can be defined, which have to be respected by a farmer receiving direct payments.

Both individual Member States and the EU as a whole are interested in a high quality standard for LPIS to register reference parcels as geographically delimited areas retaining a unique identification (European Commission, 2004, article 2/26). Thus, a quality assurance system is required, to provide a well-structured process for the quality checks and corresponding tools to conduct these checks in a well-documented and reproducible manner. The corresponding framework relies on mutually agreed quality testing between the “consumer” (the European Commission) and the “suppliers” (the Member States). A simplified overview on the LPIS Quality Assurance process is depicted in Figure 1. The two main components are:

1. LPIS Model Conformance Test
2. LPIS Inspection Procedure Test

Figure 1: Simplified Overview on the LPIS Quality Assurance Process



The LPIS Model Conformance Test evaluates whether a Member State individual LPIS implementation has appropriate and corresponding feature types and properties in respect to the requirements embedded in the common LPIS Core Model. It deals with the database schema, its logical consistency and model completeness to assure that the database design is 'fit-for-purpose'. Here, logical consistency describes the degree of adherence to the rules as predefined in the LPIS Core Model concerning the data structure, the attribution and the relationships. In terms of model completeness, the Member State LPIS implementations are compared to the LPIS Core Model for the presence or absence of feature types, their attributes and relationships. An LPIS is considered to be conforming, if one unique reference parcel type and all mandatory attributes can be identified. Within the scope of this study, model conformance is a prerequisite for successful schema transformation and a critical condition for the LPIS Inspection Procedure Test that is defined and described in LCM terms. The mapping between corresponding entities and their attributes is used as input for the related schema transformation web service presented in this paper (chapter 3).

The objective of the LPIS Inspection Procedure Test is to collect the necessary and sufficient information, in order to assess the ability of the Member State to effectively support and control the farmers' applications, i.e. to locate unambiguously the reference parcels and to quantify the area of eligible land. It requires input data from different sources: vector data outlining the reference parcel, a randomly ordered list of reference parcels to be inspected, high resolution aerial or satellite imagery or ground-truth data collected during field surveys.

The Regulation specifies seven key quality elements for this procedure, and these are assessed through a specific set of measures, defined according to a quality framework set by ISO/TS 19138 and expressed in the EC No. 1122/2009:

- Correct quantification of the maximum eligible area;
- Proportion and distribution of reference parcels with ineligible area taken into account;
- Categorization of reference parcels with ineligible area taken into account;
- Occurrence of reference parcels with critical defects;
- Ratio of declared area in relation to the maximum eligible area inside the reference parcels;
- Percentage of reference parcels which have been subject to change, accumulated over the years;
- Rate of irregularities determined during on-the-spot checks.

Further explanations and a more detailed description of the inspection procedure including some examples can be found in Milenov et al, 2012.

The Member States send the inspection data to the European Commission, which proceeds with their validation through a screening process. Therefore, the Member State has to provide a complete inspection data package containing all items required for the screening. To assure a standardized geospatial data exchange, the European Commission Joint Research Centre defined a series of XML and GML schemas for such an inspection data package. The content validation web service presented here (Chapter 4) uses these schemas to check automatically the validity and completeness of the packages provided by the Member States. The subsequent test of the Member States LPIS Inspections is conducted as a semi-automated screening process.

2.2. A Service-based Approach for the LPIS Quality Assurance

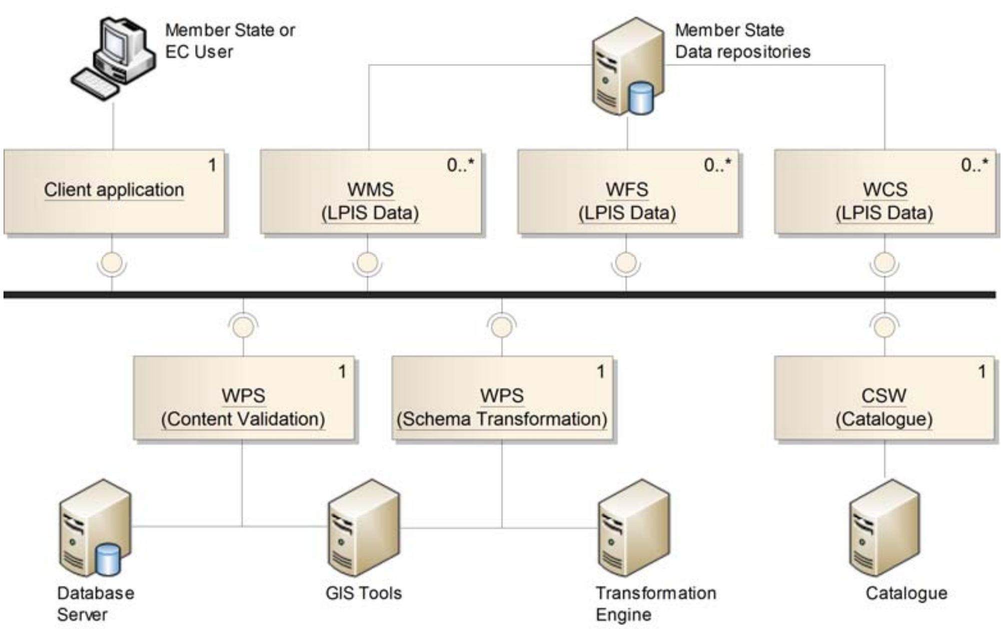
Several web services for the visualization, download and processing of spatial data can be used within the LPIS Quality Assurance framework. Furthermore, a catalogue service providing information on the service interfaces as well as the available spatial data is required. To facilitate interoperability, the implementation should rely on common web standards, in particular the ones specified by the Open Geospatial Consortium (OGC), and consider the technical guidance on INSPIRE Schema Transformation Services (INSPIRE, 2010b).

Currently, the most suitable solution to implement service based schema transformation and content validation in a standardized way is the utilization of the OGC WPS interface. According to OGC (2007), the WPS can be used to offer any sort of GIS processing functionality across the network. In addition, certain WPS implementations already include support to wrap established GIS software, thus combining the advantages of standardized interfaces and (mostly proprietary) functionality. Respective implementations are presented by Brauner (2008) for GRASS GIS or Müller et al (2009) for ArcGIS.

The proposed service infrastructure for the LPIS Quality Assurance framework is depicted in Figure 2 and includes the following components:

- OGC Web Map Service for the visualization of LPIS data and additional information for the LPIS Inspection Procedure,
- OGC Web Feature Service for the provision of LPIS feature data, e.g. for schema transformation
- OGC Web Coverage Service for the provision of orthoimages used during the LPIS Inspection Procedure,
- OGC Catalogue Service for the publishing and search for web services, spatial data or additional information,
- OGC Web Processing Service for service-based schema transformation and content validation,
- Web-based Client application as the main access point for the Member State or EC users.

Figure 2: Service Infrastructure for the LPIS Quality Assurance



3. A SCHEMA TRANSFORMATION SERVICE FOR LPIS

3.1. Background

The Model Conformance Test within the LPIS Quality Assurance framework yields conformance statements and it delivers schema transformation rules from a specific LPIS implementation towards the LPIS Core Model specification. Subsequently, LPIS data can be transformed using a schema transformation service and by this means be exchanged in a standardized way.

Since its origin in database engineering in the early 1980s, numerous research projects have dealt with the automation of schema mapping and corresponding data transformation processes. In the course of the development of Spatial Data Infrastructures, schema mapping and transformation allows the linkage and integration of information from different datasets and the use of common application schemas. Designing appropriate methods for efficient schema mapping and transformation of spatial data, as considered in this paper, is still one of the major research needs within the Geographic Information Science (Craglia et al, 2008).

A successful schema mapping should result in a set of transformation rules to modify data following a source application schema in order to match a target schema. In case of INSPIRE Schema Transformation Services, the target schema is the corresponding harmonised European schema. Following the classification of schema heterogeneities as described by Bishr et al (1999), three different levels of schema transformation are typically distinguished:

- Syntactic schema transformation – changes the encoding of the data itself or the basic data types respectively,
- Schematic schema transformation – changes the data structure and the data model’s schema vocabulary,
- Semantic schema transformation – requires a highly sophisticated in-depth-analysis of the involved datasets in cases where an exact mapping is not possible.

Whereas numerous approaches and software tools already cover syntactic and schematic schema mapping and transformation, there are only few approaches on the semantic level. Within the mdWFS project Donaubauer et al. (2007) achieve a semantic transformation of spatial data on the conceptual schema level using the OGC Web Feature Service (WFS) standard and the schema mapping formalization rules as described by Gnägi et al (2006). Klien (2007) describes an ontology-based schema mapping using semantic annotations with the prospect of using non-spatial data to improve the results. Based on a classification of semantic heterogeneity, Lutz et al (2009) discuss the integration of spatial data into SDI using ontologies and logical reasoning. However, the majority of approaches imply complex pre-conditions and rather complex user interactions for ontology mapping or semantic data enrichment. Further research on ontologies or the formalization of semantic reference systems as proposed by Kuhn (2003) is still necessary to advance semantic schema transformation processes.

Being the basis for later data transformations, the schema mapping defines relations between source and target schema elements and serves to derive the implicitly expressed transformation rules. This mapping must be precise, unambiguous and complete to avoid any information loss or inaccuracies during the later transformation processes. Schema mappings can be obtained manually, half-automated or fully automated. Whereas manual mapping is based on pure human interaction to match corresponding elements, computer-assisted attribute, instance or structure matching can facilitate automation of the schema mapping. The term matching hereby describes the automated detection of homologous elements in different datasets or schemas, leading to a mapping of those elements. Corresponding approaches are the data driven matching using the combination of feature geometry and thematic attributes proposed by Volz (2005) or the S-Match algorithm for matching semantically enriched graph structures proposed by Giunchiglia et al (2007).

A general approach to service-based schema transformation is described by Lehto (2007): It specifies schema transformation components and transformation processes as well as possible software architectures to implement the transformation services. Foerster et al (2010) propose a content transformation service, combining schema transformation and generalization processes, and provides a good overview on service based schema transformation processes.

3.2. Towards a Schema Transformation Service for LPIS

According to the Technical Guidance for INSPIRE Schema Transformation Services (INSPIRE, 2010b), SOAP web services are recommended for implementing schema transformation services for INSPIRE. Furthermore, the Rule Interchange Format (RIF) is proposed as the most suitable standard for schema mapping descriptions. However, both solutions are not yet fully supported by the geospatial community. Thus, to our knowledge, the more widespread OGC Web Processing Service (WPS) in combination with established geoprocessing functionality is currently the best approach for offering service-based schema transformation in a standardized way.

The setup of the LPIS Quality Assurance web services follows the general idea of the INSPIRE network service architecture (INSPIRE, 2008). Since the different LPIS database implementations were independently built by each Member State, a distributed structure of heterogeneous spatial data and services is already existent. In order to deal with those structures, INSPIRE Network Services recommendations can be directly adopted, even though agricultural reference parcels and the LPIS Core Model are not included in the existing INSPIRE Annex Themes.

Whenever a Member State LPIS implementation conforms to the LPIS Core Model specifications, a schema mapping process and subsequently data transformation can be performed. As a result, every Member State should be capable of using the transformation service to transform and exchange their data in a standardized way and to ease communication with the responsible EU authority.

Although not yet on an operational basis, it is presumed that any Member State can provide LPIS feature data via the standardized OGC WFS interface following an arbitrary LPIS GML application schema. Thus, the schema transformation rules between two GML feature collections described by Lehto (2007) can be applied. The following components can serve as a basis for an implementation:

- A number of software systems offer a wide range of schema mapping and transformation functionality. Widely-used examples are the Feature Manipulation Engine (FME) (Safe Software), GoPublisher (Snowflake Software) and HALE (HUMBOLDT project),
- Extensible Stylesheet Language Transformation (XSLT) allows transforming XML documents by applying previously defined template rules. In addition GeoXSLT introduced by Klausen (2006) is capable of performing simple spatial operations,
- Numerous spatial data processing and XML libraries offer basic schema and data transformation functionality.

This list reflects only the approaches taken into consideration within the presented study. Since these were identified as the most widespread and matured approaches, alternatives like the candidate model mapping languages identified by INSPIRE (2010a) were not further investigated.

An LPIS to LPIS Core Model schema mapping can be created either by the Member State, the responsible EU authority or by third-party service providers. The choice typically depends on aspects like the chosen software system, available budget and required user expertise. The mapping description contains a set of schema element assignments for national LPIS implementations towards corresponding elements in the LPIS Core Model schema as shown in Listing 1.

Listing 1: Python Snippet of a Mapping Description for Parcel Type and Id

```
#----------------------------------------------------
# Reference parcel type
# valid entries: 'PHY_BLOCK', 'AGR_PARCEL', 'FAR_BLOCK', 'TOPO_BLOCK'
#
feature.setStringAttribute(globals['RP_DEF'], globals['PHY_BLOCK'])


#----------------------------------------------------
# rpID
#
rpID = feature.getStringAttribute('MSID') #LPIS feature attribute for rpID
feature.setStringAttribute(globals['RP_ID'], rpID) #set LCM attribute value
```

As the different LPIS implementations generally lack formalised semantic descriptions, a full automation of the matching and subsequent mapping process seems not feasible. To enable the translation of schema mapping descriptions into transformation rules, the mapping must be consistent and formalized. To ensure both, there will be a need for a best practise for creating the schema mapping as well as possibilities for its validation. The latter might be performed in parallel or subsequent to the schema mapping process. In this regard, we currently suggest a manual definition of the mapping using predefined templates as well as the application of simulated datasets for prototypical transformation. However, the design of a comprehensive validation procedure can only be envisaged for future developments.

As a result, the Model Conformance Test provides a statement on the degree of conformance of Member States' LPIS implementations regarding the LPIS Core Model. If the report states full conformance, the derived schema mapping descriptions can be uploaded and registered within a central mapping repository, to allow for further usage by the transformation service.

One of the following two-implementation approaches can be taken to realise the LPIS transformation service:

1. A processing service that is tightly coupled to the source schema and the target schema; the schema mapping procedure is hard wired behind the service interface,

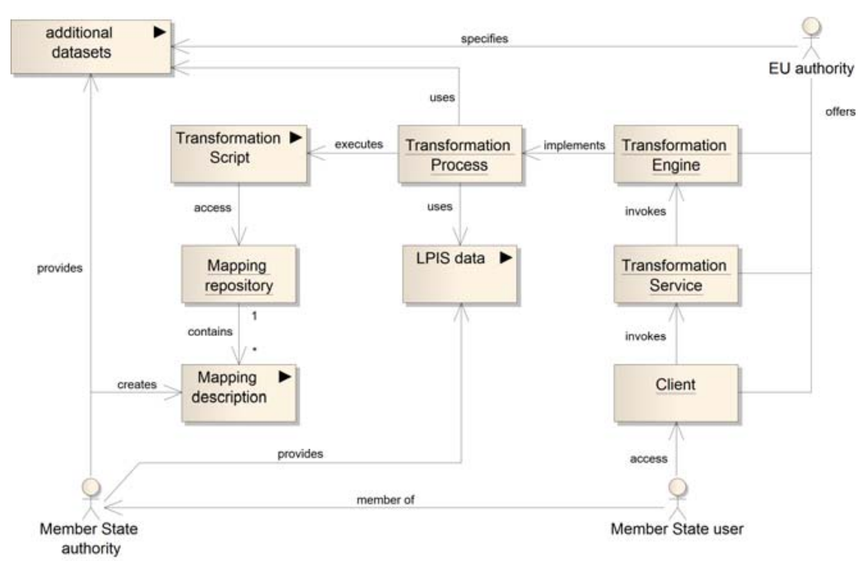
2. A generic processing service that offers a common schema transformation procedure; the source schema and the target schema are given as process parameters.

To demonstrate and prototype the functionality, the former option ought to be sufficient. For the future, the latter and more generic option should be explored. As one of the key principles of service-oriented architectures is service statelessness, interactive processes like the creation of schema mappings, have either to be automated or to be composed of multiple stateless processes. Consequently, within this study, the WPS could be used for:

1. Wrapping existing software solutions for transformation engines to access the respective functionality in a standardized way and to allow more freedom of choice for the underlying systems,
2. Automating schema matching approaches based on the syntactic or semantic information of the source and target schema,
3. Validating the derived schema mapping descriptions, like a check for completeness or conflict detection.

Figure 3 presents an overview on the proposed service architecture for schema transformation and the related processes. A central repository holds the schema mapping descriptions that are used to execute the transformation processes. The transformation service as a specific WPS offers interoperable access to the used transformation engine.

Figure 3: Architecture for Service-based Schema Transformation Using a Central Mapping Description Repository



More details on the intended schema mapping and transformation workflow for LPIS are depicted in Figure 4. The workflow consists of four consecutive sub-processes:

1. The mapping descriptions for a national LPIS implementation towards the LPIS Core Model as well as the corresponding Model Conformance Test report are compiled. If the national LPIS implementation conforms to the LPIS Core Model specification, the mapping description is uploaded to the mapping repository for later use by the schema transformation service.
2. The Member State prepares a specific reference parcel dataset within the national LPIS for schema transformation. This data is derived from a sample pre-selection procedure and provided for schema transformation.
3. A Member State user requests the schema transformation service to transform selected LPIS reference parcels. If the WPS successfully validates this request, it invokes the transformation engine. The transformation process is composed of the transformation of the provided LPIS data and, if requested, of defined overlays with additional datasets for cross-compliance. The result of the schema transformation is returned to the client either directly or by reference. It contains the transformed dataset, which now follows the LPIS Core Model and optionally holds information on the transformation process.

The proposed WPS interface for schema transformation from Member State LPIS data towards the LPIS Core Model is described in Table 1. As Inputs, a Member State identifier and the LPIS data to be transformed have to be provided. In addition, optional datasets to conduct spatial overlays for cross-compliance can be included. The result of the WPS is either a link to a web accessible resource or a direct return of the transformation results. In case of an error, a corresponding error report is returned.

Figure 4: Schema Mapping and Transformation Workflow Sequence

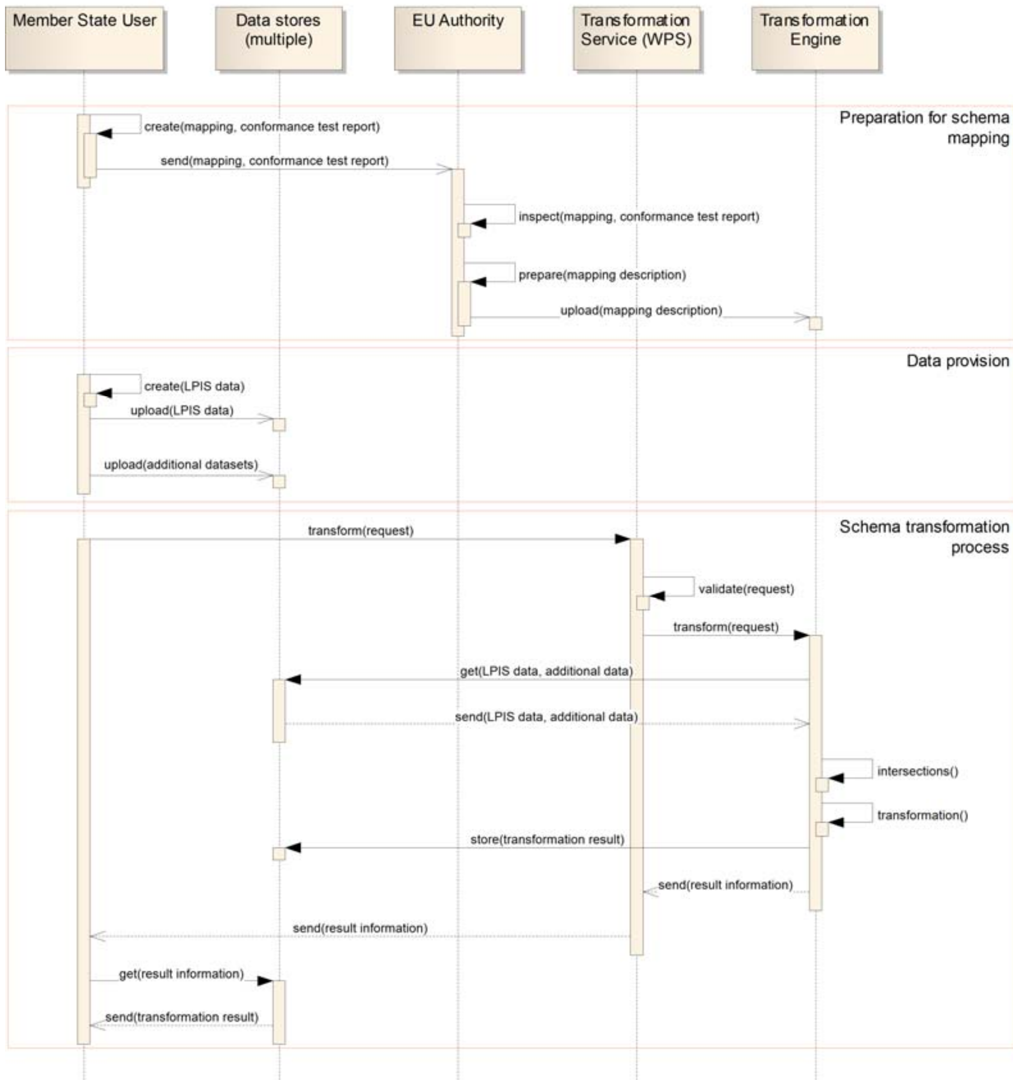


Table 1: WPS Process Parameter for LPIS Schema Transformation

Parameter	Definition	Data Type	Multiplicity
MEMBER_STATE	LPIS Member State identifier	UID	1 (mandatory)
LPIS_SOURCE	LPIS dataset for transformation (conform to the LPIS Core Model)	URL / Vector data	1 (mandatory)
LFA_SOURCE	Less Favoured Areas (overlay)	URL / Vector data	0..1 (optional)
NAT_SOURCE	Natura2000 Areas (overlay)	URL / Vector data	0..1 (optional)
NIT_SOURCE	Nitrate Vulnerable Zones (overlay)	URL / Vector data	0..1 (optional)
SBU_SOURCE	Stream Buffer areas (overlay)	URL / Vector data	0..1 (optional)
AEM_SOURCE	Agro-Environmental Measures (overlay)	URL / Vector data	0..1 (optional)
RESULT	Transformation Result (LPIS data following the LPIS Core Model)	URL / Vector data	1 (mandatory)

3.3. A prototype of a web service for schema transformation

A prototypical implementation should prove the feasibility of the proposed architecture. The following software components were utilized:

- FME Desktop – to conduct the schema mapping definition process with a graphical user interface (GUI) for interactive modelling of complex schema and format mapping processes.
- FME Server – offers the possibility to access and run previously created FME mapping scripts within a distributed environment and allows the realization of a high performance, scalable and reliable transformation service. Using the Java API of FME Server, the functionality to run the schema transformation is wrapped by the standardized OGC WPS interface.
- 52°North WPS – to implement the mediator WPS instance between the web client and the FME Server. Furthermore, it is used for the implementation of certain pre- and post-processing steps, like the validation of the input data, as well as for the correct parameterization of the transformation service.
- GeoServer WFS – to provide LPIS datasets for schema transformation and additional datasets for cross-compliance via the OGC WFS interface.

Following the proposed WPS interface, the implemented service requires a Member State identifier to select the appropriate mapping description from the repository and the WFS URL of the LPIS dataset to perform the schema transformation process. In addition, optional URLs pointing to datasets for performing spatial overlays with the reference parcels can be handed to the service.

To build a generic workflow, allowing changes and extensions to the mapping process, the mapping descriptions are coded in plain Python scripts. This solution has the advantages of software independence (no FME Desktop licence required for creating the mapping description on the Member State side) and easy editing with a standard text editor. Most of the GML transformation rules proposed by Lehto (2007) as well as their combinations can be applied, in particular attribute filtering, renaming, reclassification, merging/splitting, reordering, conversion and augmentation. The derivation of target features from multiple source features has not been a requirement for LPIS and is therefore not yet supported.

There are two scripts for each transformation, one for the mandatory transformation of reference parcels and one for the optional spatial overlays for cross-compliance. This allows for easier maintenance and traceability. However, a specific Python formatting needs to be respected within the FME Server environment. Thus, the scripting demands appropriate user skills.

To test the prototypical implementation of the transformation services, two local copies of Member State LPIS datasets holding reference parcels were prepared for schema transformation. These datasets were accessed via the OGC WFS interface using a GML schema derived from the underlying database implementations. The results of the previously performed Model Conformance Test showed that all required elements and mandatory attributes were present. On this basis a schema mapping was created in order to derive the transformation rules between the LPIS implementations and the LPIS Core Model. The mapping descriptions were formalized, mainly by an adaption to a previously created python-based mapping template, and stored in the mapping repository. These mapping scripts can be invoked by the FME Server Engine consecutive to a corresponding WPS request. In this way the selected national LPIS datasets were successfully transformed into the LPIS Core Model. Furthermore, the calculations of spatial overlays for cross-compliance, in particular with datasets for Natura2000 and Nitrate Vulnerable Zones, were performed.

3.4. Compatibility with INSPIRE Schema Transformation Services

Concerning the Transformation Network Services, INSPIRE defines three mandatory operations: *GetTransformationServiceMetadata, Transform* and *LinkTransformationService* (European Commission, 2010). The first two operations map to the *GetCapabilities*, *DescribeProcess* and *Execute* Operations defined for the OGC WPS interface. The *LinkTransformationService* operation is realised using a Discovery Service to register and publish the developed WPS instances. The input parameters defined for the *Transform* operation, have only partially been implemented: The source schema, the target schema and the mapping rules are not explicitly handled as input parameters of the transformation service, but given implicitly using Member State identifiers. However, most of the requirements for INSPIRE Transformation Network Services proposed by INSPIRE (2010b) are fulfilled by the implementation presented here:

- Source and target schema are provided as GML application schemas,
- Mapping descriptions are stored separately from the transformation service in a mapping repository, although not yet XML-based as proposed for INSPIRE Transformation Network Service,
- All INSPIRE transformation use cases (store configuration, transform dataset, gather technical information) map to the implemented LPIS use cases on schema transformation,
- The proposed implementations follow the architectural requirements of INSPIRE (open interface, statelessness, parameter by reference, schema agnostic interface, automated process, mapping flexibility)

The schema mapping on the feature level, in particular the derivation of a target feature from multiple source features is not supported, as it has not been a requirement for the presented study. Furthermore, only one target schema, the LPIS Core Model GML application schema, is available for schema transformation. Thus, only a subset (four out of six) of the transformation levels proposed by INSPIRE (2010b) is realised. However, the transformation functionality can be enhanced during the further development of the prototyped service, especially by taking advantage of the used WPS wrapping approach for utilizing advanced GIS functionality.

4. A CONTENT VALIDATION SERVICE FOR LPIS

In general, content validation is used to assess the suitability of data to fulfil an intended purpose. For LPIS Quality Assurance, content validation is applied to ensure completeness and validity of the inspection results provided by the Member States. The content validation service shall support the responsible EU authority in conducting screening procedures on the data, being currently exchanged using standardised XML and GML inspection files.

Content validation is understood here as the process of examining whether and to which extent datasets fulfil certain data quality requirements like completeness, consistency and correctness. For spatial datasets, this includes verification processes with respect to previously defined constraints on geometry, thematic attributes, data format restrictions, well-formed data structure, or naming conventions.

A number of examples for web services focusing on validation can be found, for instance:

- W3C validation services to check files against certain web standards like HTML, CSS or RDF,
- mobileOK service to evaluate web pages concerning their suitability for mobile devices,
- Contenthooks, a Google service to check repository commits for well-formed XML, forbidden strings and a valid utf-8 character encoding,
- INSPIRE Metadata Validator, to test conformance of uploaded metadata entries against the INSPIRE Metadata Regulation.

Most of the available services deal with set-actual comparisons of the data structure or in a specific way formalized thematic content. In a similar way, the content validation service for LPIS should also focus these topics. The web service for content validation would be the main access point for Member States to check pro-actively their inspection results against expressed constraints and, if successfully validated, to store their inspection test package in the database of the European Commission. As those constraints may change due to changes in the underlying requirements, the service should be kept as flexible as possible. In particular, the validation process for the LPIS Inspection Tests includes:

- Package validation – check if all requested files are uploaded and readable,
- File validation – check if files can be validated against the prescribed inspection schemas (e.g. well formed XML structure, mandatory elements and attributes, valid attribute values),
- Spatial content validation – check for topological consistency and spatial extent of contained spatial features,
- Thematic content validation – check if certain previously defined thematic constraints are fulfilled (e.g. complete list of reference parcels as requested for inspection)

A package is classified as valid, if it passes all of the above-mentioned validation processes. Upon successful validation, the package can be uploaded for further use within the subsequent screening process. Different ways for implementing such a web service are possible:

1. The validation service is a processing service, tightly coupled with the database (exclusive access),

2. The validation service is a processing service, loosely coupled with the database (reading and writing capabilities),
3. The validation service is a data service that stores the content and applies conformance checks internally.

As being identified as the most robust solution and being sufficient in the scope of this study, the first option was applied during the prototypical implementation, using an OGC WPS for wrapping the validation functionality. Thus, the validation service is tightly coupled to the database, both running on the same machine. As described in Table 2, a zipped package containing the inspection results and an optional reference to corresponding orthoimages are required as input for the validation service. As a response to the package validation request, a log file is returned to the client containing basic information on the uploaded datasets, the validation process and the validation result.

Table 2: WPS process parameter for LPIS inspection content validation

Parameter	**Definition**	**Data Type**	**Multiplicity**
PACKAGE	Zipped file containing the inspection results	URL / zipped file	1 (mandatory)
ORTHOIMAGE	Orthoimage used for inspection	URL / Raster image	0..1 (optional)
RESULT	Logfile containing information on the validation process	Text	1 (mandatory)

The prototypical validation process is implemented using Java and a PostGIS database. In order to test the performance of the implemented components, several artificial LPIS inspection packages were created, simulating different quantities of included reference parcels. Using those packages, the service proved efficient and scalable with constantly processing around 90 reference parcels per second. In addition, the OGC WPS interface encapsulates the validation functionality and offers communication in an as far as possible standardized way.

5. CONCLUSION AND FURTHER RESEARCH

The web services designed and developed within the presented study demonstrate the feasibility of the chosen SDI-approach for LPIS Quality Assurance. A web service for the transformation of heterogeneous LPIS database implementations towards the common LPIS Core Model as well as a web service for the validation of inspection results were designed, implemented and successfully tested in laboratory conditions. Both services offer their functionality via standardized OGC WPS interfaces including an approach to wrap FME Server technology acting as a sophisticated and customizable backend for schema transformation.

The implemented schema transformation allows for the exchange of LPIS datasets following the common LPIS Core Model specification and thus offers support for data harmonisation and validation. Furthermore, the package content checking facilitates the automation of various validity and conformance checks on the Member State LPIS inspections and reduces the costs for further screening. Those results reveal the great potential of geoprocessing services for the LPIS Quality Assurance improving the quality assessment process as well as the communication and collaboration between Member States and responsible EU authorities.

Additional aspects, such as security, robustness and usability, should be considered for the further development of both prototyped web services. This might include further investigation on service authentication, authorization issues, service failover configurations, continuous service monitoring and user-friendliness. Finally, the proposed services need to be integrated in a future geoportal implementation for LPIS.

Concerning the transferability of the proposed solutions, it must be noted, that both of the presented services were developed closely to the LPIS requirements. However, certain aspects could be transferred to similar applications. For the schema transformation, this especially applies to the generic service structure offering a highly customizable and flexible transformation towards one application schema. Moreover, the implemented service clearly shows the feasibility and advantages of wrapping FME Server technology for schema transformation behind a WPS interface. Content validation on the other hand can be considered as crucial for assuring data quality. The proposed service-based solution could easily be extended to serve corresponding functionality to a wide range of application areas.

The development of RIF towards a common schema mapping language for spatial datasets as well as the SOAP/WSDL binding for geoprocessing services should be studied further in order to path the way for future interoperable transformation services. Another restraint is the lack of well-agreed service profiles for schema transformation and content validation in general. As it will facilitate interoperability and support ongoing standardization processes, more efforts should go into the corresponding development of generic schema transformation and content validation service profiles.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the valuable IT contributions of Vladimir Knezevic and Romuald Franielczyk, which helped in conducting the described software developments and tests. Moreover, we are grateful to the two anonymous reviewers and to Stephan Mäs (TU Dresden), who provided us with a number of constructive comments to enhance this paper.

REFERENCES

Bishr, Y.A., Pundt, H., Kuhn, W., Radwan, M. (1999). Probing the Concept of Information Communities – A first Step towards Semantic Interoperability, in Goodchild, M., Egenhofer, M.J., Fegeas, R., Kottman, C. (Eds). *Interoperating Geographic Information Systems*, Kluwer, Academic Publishers, pp. 55-69.

Brauner, J. (2008). Providing GRASS with a Web Processing Service Interface. *GI-Days 2008*. Münster, Germany.

Craglia, M., Goodchild, M.F., Annoni, A., Camara, G., Gould, M., Kuhn, W., Mark, D., Masser, I., Maguire, S., Liang, S., Parson, E. (2008). Next-Generation Digital Earth. *International Journal of Spatial Data Infrastructures Research*, 2008(3), pp. 146-167.

Donaubauer, A., Straub, F., Schilcher, M. (2007). mdWFS: A Concept of Web-enabling Semantic Transformation. *10th AGILE International Conference on Geographic Information Science*, 2007, Aalborg, Denmark.

European Commission (2004). Commission Regulations No 796/2004 laying down detailed rules for the implementation of cross-compliance, modulation and the integrated administration and control system provided for in of Council Regulation (EC) No 1782/2003 establishing common rules for direct support schemes under the common agricultural policy and establishing certain support schemes for farmers, April 2010.

European Commission (2010). Commission Regulations No 1088/2010 amending Regulation (EC) No 976/2009 as regards download services and transformation services, November 2010.

Foerster, T., Lehto, L., Sarjakoski, T., Sarjakoski, L.T., Stoter, J. (2010). Map generalization and schema transformation of geospatial data combined in a Web Service context. *Computers, Environment and Urban Systems*, 34(1), pp. 79-88.

Giunchiglia, F., Yatskevich, M., Shvaiko, P. (2007). Semantic Matching: Algorithms and Implementation, in Spaccapietra, S. (Ed). *Journal on Data Semantics IX*, Springer-Verlag, Berlin-Heidelberg, pp. 1-38.

Gnägi, H.R., Morf, A., Staub, P. (2006). Semantic Interoperability through the Definition of Conceptual Model Transformations. *9th AGILE International Conference on Geographic Information Science*, April 20-22 2006, Visegrád, Hungary.

INSPIRE 2008, INSPIRE Network Services Architecture, Version 3.0, July 2008.

INSPIRE 2010a, Development of Technical Guidance for the INSPIRE Transformation Network Service, State Of The Art Analysis, Version 2.0, December 2010.

INSPIRE 2010b, Technical Guidance for the INSPIRE Schema Transformation Network Service, Version 3.0, December 2010.

Klausen, F.M. (2006). GeoXSLT: GML processing with XSLT and spatial Extensions. Department of informatics. University of Oslo.

Klien, E. (2007). A Rule-Based Strategy for the Semantic Annotation of Geodata. *Transactions in GIS*, 11(3), pp. 437-452.

Kuhn, W. (2003). Semantic Reference Systems. *International Journal of Geographical Information Science*, Guest Editorial, 17(5), pp. 405-409.

Lehto, L. (2007). Schema Translations in a Web Service Based SDI. *10th AGILE International Conference on Geographic Information Science*, May 8-11 2007, Aalborg, Denmark.

Lutz, M., Sprado, J., Klien, E., Schubert, C., Christ, I. (2009). Overcoming semantic heterogeneity in spatial data infrastructures. *Computers & Geosciences*, 35(4), pp. 739-752.

Milenov, P., Devos, W., Wojda, P., Tarko, A. (2012). LPIS Quality Assurance Framework ANNEX II Executable Test Suite (ETS). Publications Office of the European Union, Luxembourg, 2012, JRC 68017.

Müller, .M, Bernard, L., Vogel, R. (2010). Multi-criteria evaluation for emergency management in Spatial Data Infrastructures, in Zlatanova, S., Konecny, M., Bandrova, T. (Eds). *Cartography and Geoinformatics for Disaster Management: Developments and Trends*, Springer-Verlag, Berlin-Heidelberg, pp. 273–86.

Open Geospatial Consortium Inc. (OGC) (2007). OpenGIS Web Processing Service, *OpenGIS Standard*, Version 1.0.0, June 2007.

Sagris, V., Devos, W., Milenov, P. and Kay, S. (2008). New evidence of land management in the frame of Common Agricultural Policy: needs for standardization. *FIG Working Week 2008*, June 14-19 2008, Stockholm, Sweden.

Sagris, V., Devos, W., (2009). Core Conceptual Model for Land Parcel Identification System (LCM). *GeoCap technical specification, v 1.1*. EUR-Scientific and Technical Research series.

Volz, S. (2005). Data-Driven Matching of Geospatial Schemas, in Cohn, A.G., Mark, D.M. (Eds). *Spatial Information Theory*, Springer-Verlag, Berlin-Heidelberg, pp. 115-132.