

# Spatial data fusion in Spatial Data Infrastructures using Linked Data

*Stefan Wiemann, Lars Bernard*

*Chair of Geoinformatics, Technische Universität Dresden, 01062 Dresden, Germany*

Contact: [stefan.wiemann@tu-dresden.de](mailto:stefan.wiemann@tu-dresden.de)

*This is an Accepted Manuscript of an article published by Taylor & Francis in International Journal of Geoinformation Science on 12 August 2015, available online: <http://dx.doi.org/10.1080/13658816.2015.1084420>*

## Abstract

The synthesis of spatial data from the various sources available on the Web is a major challenge for current applications based on Web-based information retrieval and spatial decision-making. This paper addresses spatial data fusion, with particular emphasis on its application in Spatial Data Infrastructures (SDIs). Possibilities for the integration of SDI and Semantic Web developments in the context of spatial data fusion are reviewed with a focus on the harmonized description and usage of feature relations. Specifically, potential applications of Linked Data principles are discussed in detail. On this basis, a classification and a decomposition of fusion processes in a service-oriented environment are proposed. A prototype implementation demonstrates the feasibility and usability of the approach using Open Geospatial Consortium (OGC) and Semantic Web standards.

Keywords: data fusion, Spatial Data Infrastructure, Linked Data, Semantic Web

## 1. Introduction

The rapid development of the Web, with its evolution from data-oriented to service-oriented structures, together with the widespread use of location-enabled mobile devices has considerably influenced the public awareness, availability and use of spatial data. Driven by legislation, commercial interest, community engagement and open data movements, the amount of spatial data accessible via the Web in both governmental Spatial Data Infrastructures (SDIs) and from volunteered, scientific and corporate

initiatives is continuously increasing. To prevent data isolation and enable cross-dataset analysis, methods of relating and combining spatial data from a variety of sources are indispensable. Here, it is envisioned that once future web services enable the derivation of spatial information from an arbitrary number of underlying data sources at little cost, these services will offer much greater potential than today's SDIs, which still primarily act as web-based (spatial) data provision platforms. Spatial data fusion techniques play an important role in fostering an integrated view of distributed spatial data sources on the Web. Because flexibility and interoperability are key factors in such data integration, the use of standards is a crucial requirement. Therefore, in addition to the geospatial standards established by the OGC (Open Geospatial Consortium), the Semantic Web standards published by the W3C (World Wide Web Consortium), particularly those concerning Linked Data, are considered good candidates for the formalization and management of feature relations as part of the fusion process. The application of Semantic Web technologies in the geospatial community has developed considerably in recent years. Although SDIs and the Semantic Web are quite distinct with respect to their architectures, workflows and applied standards, they can be regarded as complementary (Auer and Lehmann 2010, Janowicz et al. 2010, Padrinaci and Dominique 2010, Schade and Smits 2012). Moreover, Semantic Web technologies offer a promising approach to link SDIs to mainstream IT and to enable further areas of SDI application (Schade and Smits 2012). Still, the derivation of meaningful information from distributed spatial data sources on the Web remains a major challenge. Consequently, discussions revolve around two main questions: (1) how spatial data on the Web can be related in an efficient manner and (2) how those relations can be utilized in a meaningful and productive manner.

This paper investigates approaches, requirements and limiting factors for service-based spatial data fusion, with particular focus on the interaction of SDI and Semantic Web standards. The main contributions are a service-oriented decomposition of fusion processes and strategies for feature relation management using Linked Data in combination with SDIs. For demonstration purposes, a prototype implementation illustrates the feasibility of the presented approach. It builds upon open-source software, using OGC Web Services (OWS) for spatial data provision and processing as well as W3C standards for relation encoding and semantic reasoning. As use cases, the fusion of road networks, the versioning of datasets with change detection and the linking of different feature representations are presented.

A classification and a service-oriented decomposition of spatial data fusion processes to facilitate the implementation of interoperable and flexible fusion workflows are outlined in section 2. A concept for the integration of SDI and Semantic Web developments, with particular focus on the formalization, management and application of spatial data relations, is developed in section 3. A prototype implementation is described, evaluated and discussed in Section 4. The paper concludes with a summary and outlook in section 5.

## 2. Spatial Data Fusion – Classification and Decomposition

The term data fusion is widely used in electronic data processing and comprises a number of different definitions and classifications. In this paper, spatial data fusion refers to the synthesis of spatial data from multiple sources to extract meaningful information with respect to a specific application context. Therefore, the understanding of data fusion proposed here includes what others describe as conflation (e.g., Saalfeld 1988, Ruiz et al. 2011), data integration (e.g., Devogele et al. 1998, Walter and Fritsch 1999, Schwinn and Schelp 2005) or data concatenation (e.g., Kiehle et al. 2007, Longley et al. 2010).

Existing publications offer a variety of classifications for spatial data fusion. These approaches can be discriminated (Figure 1) based on the application field, the level of automation, the operation frequency, the matching level, the underlying data model and the spatio-temporal orientation of the input data (Yuan and Tao 1999, Schwinn and Schelp 2005, Ruiz et al. 2011). The corresponding process implementations can be further distinguished based on the input and output data structures they support, the applied matching strategies, their computational performance or other common quality measures.

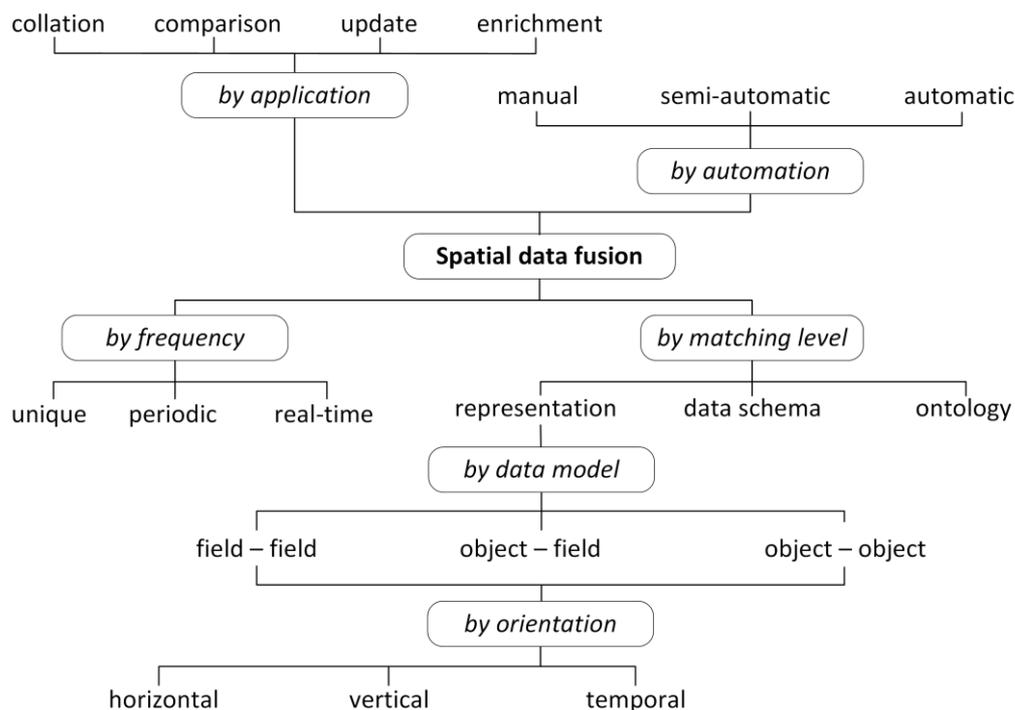


Figure 1: Classification schema for spatial data fusion processes.

Providing spatial data fusion processes in a distributed and (web-based) Service-Oriented Architecture (SOA), such as an SDI, requires the design and definition of appropriate service interfaces. These interfaces are meant to decompose and encapsulate appropriate sets of interoperable fusion function granules (also called services), which can then be orchestrated for different fusion workflows. Thus, such a design will ideally offer well-

defined interfaces that support the loose coupling, reusability and composability of the services (Erl 2008). Three levels of granularity can be defined (Erl 2008):

- (1) Atomic operations provide application-agnostic functionalities to the system, such as simple arithmetic operations and geographical distance measurements.
- (2) Low-level operations are already application driven but can still be characterized as generic and therefore usable in a number of applications. These include, for example, geometric transformations, feature similarity measurements and feature transfer operations.
- (3) High-level operations are application specific, with limited flexibility and reusability, such as recurring workflows tied to specific inputs and application targets.

Every operation can reuse or be a composite of lower-level operations. However, in practice, these levels cannot always be strictly separated. To enable service-based spatial data fusion in a flexible, reusable and practical manner, a toolbox of low-level operations appears to be the most promising approach. A typical fusion workflow and its logical decomposition are depicted in Figure 2. Although not bound to a SOA implementation, this example serves as an abstract framework for low-level fusion processes, comprising the following steps:



Figure 2: Low-level sub-processes for spatial data fusion.

- (1) Data search and retrieval processes are aimed at the identification and gathering of input data for spatial data fusion. Whereas data search within an SDI primarily addresses metadata, data retrieval comprises means of accessing identified data, including the functionalities of spatial, temporal and thematic filtering.
- (2) Data enhancement addresses the quality characteristics of individual input sources relevant to the fusion process. It comprises tasks such as topology building and repair (e.g., Walter and Fritsch 1999, Butenuth et al. 2007), string preprocessing (Samal et al. 2004, Al-Bakri & Fairbairn 2012) and classification processes (Koukoletsos et al. 2012) and can significantly enhance the performance of the overall fusion process.
- (3) Data harmonization aims to minimize inequalities between inputs to ensure a common syntactic, structural and semantic base to support successful and meaningful data fusion. Harmonization is typically performed at the dataset level and includes processing such as positional re-alignment (Lupien and Moreland 1987, Doytsher et al. 2001, Song et al. 2006, López-Vázquez and Callejo 2013),

time synchronization (Uitermark et al. 1999), format and coordinate conversion (Stankute and Asche 2012) and model generalization (Sester et al. 1998, Uitermark et al. 1999).

- (4) Relation measurement involves the investigation and quantification of relations between input sources. It can be performed at the feature representation, schema or semantic level and typically expresses a certain kind of similarity. Whereas measurements at the feature level exploit spatio-temporal and thematic feature properties, measurements at the schema and semantic levels are based on terminological, structural and conceptual characteristics. Both can be combined to increase the probability and accuracy of a relation (Volz 2005). Several overviews of common approaches have been provided by Veltkamp and Hagedoorn (2001) for shape similarity, Shvaiko and Euzenat (2005) for schema similarity and Schwering (2008) for semantic similarity measurements.
- (5) Feature mapping exploits the relation measurements to determine the types of relation between features at the representation, schema or semantic level. In addition to the relation measurements, it can also be influenced by the reliability of the input sources and existing knowledge (Cholvy 2007). The probability of a relation can be expressed in terms of hard and soft confidence measures (Cobb et al. 1998, Bloch et al. 2001, Waltz and Waltz 2008).
- (6) Resolving steps are determined by the previous results and the specific application target. Conflict resolution strategies can be applied to eliminate any inconsistencies identified during the mapping process (Bloch et al. 2001). Finally, the actual objective of the fusion process can be achieved by utilizing the identified relations for data collation, comparison, update or enrichment. This includes the transfer of features and attributes that are not present in all datasets (e.g., mutual updates) as well as the merging of related features to obtain an enhanced or enriched feature representation.
- (7) Data provision addresses the encoding, storage and registration of results to facilitate access to and use of the results for visualization or further analysis. Ideally, the fusion results will include information on the provenance, uncertainty and processing lineage of the data.

The steps described above should not be regarded as forming a strict sequence. Individual processes may be skipped, reiterated or differently combined. In particular, steps 4 (relation measurement), 5 (feature mapping) and 6 (resolving) are often closely tied and are generally referred to as the process of identifying feature mappings (Rahm & Bernstein 2001).

### **3. Spatial data fusion using SDI and Linked Data concepts**

Established SDIs provide a means of publishing, searching, accessing and processing spatial data on the Web in an open and standardized manner (Bernard et al. 2005). The

Semantic Web, as introduced by Berners-Lee et al. (2001), allows for ubiquitous access to interlinked data on the Web. It is a manifestation of the Linked Data paradigm; it (1) uses Uniform Resource Identifiers to uniquely identify objects or concepts that can (2) be resolved to HTTP addresses, providing (3) data in standardized formats that are (4) linked to other data sources on the Web<sup>i</sup>.

When targeting interoperable, service-based spatial data fusion, all data fusion processing steps described in the previous section should be aligned with current SDI and Semantic Web standards. This can be achieved through the use of OGC<sup>ii</sup> standards for the registration, encoding, provision, visualization and processing of spatial data and the use of W3C standards<sup>iii</sup> for Linked Data, particularly RDF (Resource Description Framework) and SPARQL (SPARQL Protocol and RDF Query Language), respectively. Figure 3 shows an example of a possible means of communication between SDI and Semantic Web components for spatial data fusion on the Web. Whereas SDI components can be accessed using OWS interfaces, Semantic Web components offer functionalities based on RDF.

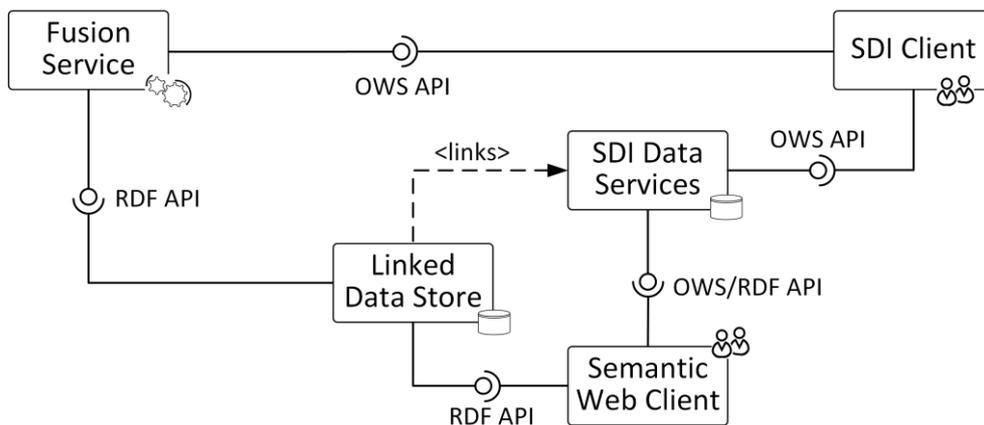


Figure 3: Interaction between SDI (OWS interface) and Semantic Web (RDF interface) components.

A method of unique identification management within SDIs is a prerequisite for the application of Linked Data concepts. On a small scale and with certain restrictions on reusability, flexibility and persistence, such a system of feature identification can be implemented using service and local feature identifiers (Schade et al. 2010). However, the governance of persistent unique identifiers and the realization of an effective resource identification and management strategy on a larger scale require comprehensive harmonization efforts beyond the technical level. The common identifier management strategy for environmental data in Europe (INSPIRE 2014) serves as an example of such an effort.

A suitable RDF encoding for spatial data is essential to ensure compatibility with

prevailing spatial data formats, particularly GML. Proposed solutions range from the use of plain GML literals (OGC 2012) to the implementation of a comprehensive RDF spatial data model (Cox 2013). Data providers can enable the exchange of spatial data between SDIs and the Semantic Web either by providing preprocessed data sets (Goodwin et al. 2008, Stadler et al. 2012) or by offering ad hoc transformations (Janowicz et al. 2013). However, reaching agreement on a common ontology set for the representation of spatial data in RDF remains a challenge.

### ***Representation of feature relations***

To facilitate access to and utilization of feature relations from various distributed data sources, these data sources must be well defined and well managed. Therefore, we propose a relation schema that describes feature relations compliant with SDI and Linked Data principles. The main components are as follows (Figure 4):

- *Resource* represents the fundamental concept of the Semantic Web. All feature relation components extend *Resource* and can be identified and linked accordingly by means of an internationalized resource identifier (IRI).
- *FeatureRelation* is the main class that contains all features participating in a relation as well as the associated relation types and measurements. A timestamp supports the versioning of the participating features and the maintenance of the relations between them.
- *RelationType* qualifies the types of relationships between input features. Where appropriate, it should adhere to existing vocabularies for object relations, such as those provided by OGC (2012) for topological relations and by SKOS<sup>iv</sup> (Simple Knowledge Organization System) for conceptual relations. Because features can be related in various ways, a single relation can have multiple relation types.
- *FeatureResource* represents a distinct feature instance that is linked by a relation. It points to the SDI data service (e.g., a Web Feature Service) that provides the feature instance (defined following the OGC notion of a *feature*) and must be uniquely identifiable and resolvable. In the case of a non-symmetric relation type, the role of each resource (reference or target) must also be defined.
- *RelationMeasure* provides information regarding the underlying measurements that support a relation type and constitute a relation, including information on similarity and confidence. If possible, lineage information for the conducted processes should be included for transparency and reproducibility. If the input data provide information regarding uncertainty, fuzzy measures can be applied.

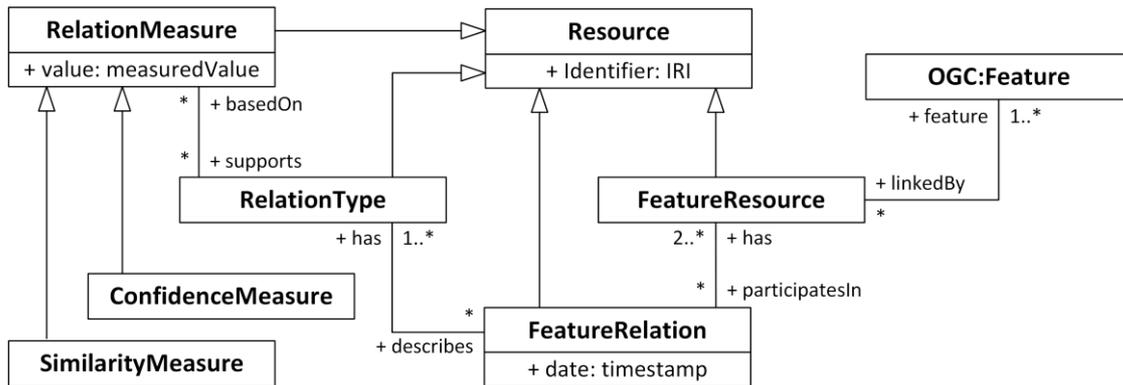


Figure 4: Proposed components for the formalization of feature relations.

In the simplest case, a relation consists of a number of similarity and confidence measurements and exhibits the default *relatedTo* relationship. Based on the underlying ontological description, each type of relation can be further characterized as a symmetric, reflexive or transitive relation to facilitate semantic reasoning. An inverse relation can also be defined. Similarly, the relation type can also express the role (reference, target, whole, part-of, etc.) of a feature resource participating in a relation.

The presented approach builds upon the OGC Abstract Specification for feature relations (OGC 2009), which defines lightweight and heavyweight relations. A lightweight relation consists of two feature resources connected by a specific type of relation. If additional data are stored, then relation entities with further attributes (timestamps, measurements, etc.), called heavyweight relations, are created. This paper extends this approach by introducing a Linked Data schema to enhance the usability and extensibility of existing feature relations.

Following the classification of links by Heath and Bizer (2009), the proposed schema can be used to formalize relationship links and identity links. Whereas identity links are used to link different representations of the same real-world object, relationship links establish connections between features to describe certain common types of connections.

These relations can either be embedded into the features of interest or be stored independently of the features. Therefore, different relation management approaches can be distinguished (Figure 5):

- (1) *Embedding relations as feature attributes*: The responsibility for the creation and maintenance of relations remains with the data provider, and users can directly access and explore the relations attached to particular features. This approach has been implemented for spatial data by Stadler et al. (2012), for example. However, two-way relations require mutual agreements among all participating data providers.

- (2) *Building a mediation layer between data sources to allow relations to be created and stored independently*: This separation of concerns allows third parties to relate spatial data and prevents feature attributes from being inflated with data regarding relations. Furthermore, it simplifies the encoding of multipart or multi-dataset relations because these relations are not bound to a specific feature representation. As an example, the Similarity Links implemented by Bröring et al. (2014) follow this approach. However, the lack of return links from the data stores to the relations hampers the traversal and browsing of relations, and therefore, this approach demands sophisticated registration strategies.
- (3) *Creating proxy representations by establishing relations between RDF representations of the original spatial data sources*: This approach is a hybrid of the two described above and enables the traversal of relations while remaining independent of the data providers. Nevertheless, the need for redundant data storage poses additional synchronization and security challenges and most likely prevents this approach from being scalable to extremely large datasets.

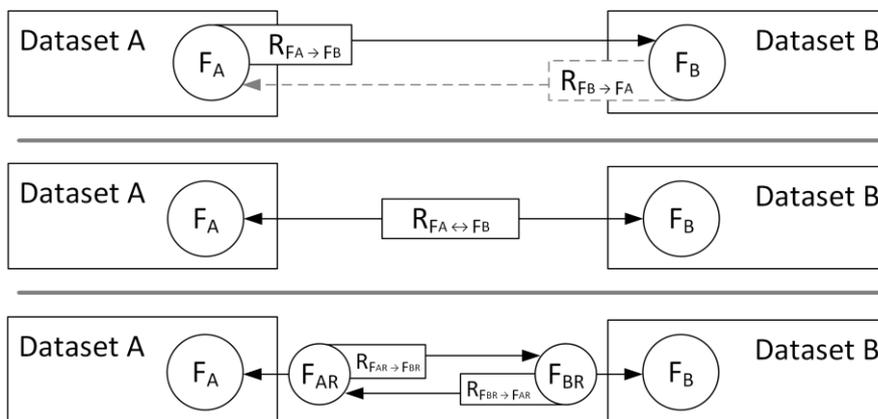


Figure 5: Approaches to the management of feature relations between two data sources.

In each case, a change in the underlying data sources must trigger an update of the affected relations to preserve their validity. This can be achieved either through the usage of recalculated change sets, as applied by Stadler et al. (2012), or through the implementation of a maintenance protocol, as proposed by Volz et al. (2009). Additionally, user feedback mechanisms can be of great benefit, especially for frequently used feature relations; such mechanisms have been investigated by Acosta et al. (2013), who compared contest-based and microtask-based approaches to enhance the quality of data links, and by Karam and Melchiori (2013), who focused on the use of a quality assurance framework leveraging the local knowledge of users.

### *Usage patterns*

For the application of spatial data fusion in SDIs in combination with Linked Data

concepts, we distinguish four distinct usage patterns (Figure 6):

- (4) *Linked Data crawling* is applied to construct new or maintain existing feature relations on the Web. It enables the regular creation, evaluation and updating of relations between data sources (Heath and Bizer 2011). All SDI data services that offer unique feature identification are supported. Crawling procedures can be either restrictive, operating on known data sources, or non-restrictive, operating on available registries. (Example: A number of SDI services offering road network data are published in a registry [1:publish]. A Crawler identifies the presence of new datasets [2:find] and triggers a fusion process. The data are accessed by fusion services [3:bind], and the relations are identified and stored as Linked Data [4:store].)
- (5) *Restrictive fusion* operates upon request on provided spatial data sources. The fusion process synthesizes those data sources and provides access to the identified relations as Linked Data. (Example: A client identifies two SDI services offering road network data [2:find] that have previously been registered [1:publish]. The client triggers the fusion process [3:bind] with a reference to the input sources. The fusion process accesses the data [4:bind], executes the fusion procedures and stores the identified relations as Linked Data [5:store]. Finally, the client can access the stored relations for further use [6:bind].)
- (6) *Non-restrictive fusion* considers a defined data source upon request and identifies suitable sources for spatial data fusion from available registries. In this context, the suitability of a source depends on a number of constraints, including data access restrictions, the application target and the available fusion processes; these last are also dynamically identified. Exit conditions and a means of reducing the search space are required to prevent infinite searching and processing. All identified relations are provided as Linked Data. (Example: A client identifies an SDI service offering road network data [2:find] that has previously been registered [1:publish]. The client triggers the fusion process [3:bind] to identify and relate complementary data sources. In the case of a successful search for suitable data sources [4:find], a fusion process is triggered based on the identified data sources [5:bind]. The identified relations are stored as Linked Data [6:store] and made available for further use by the client [7:bind].)
- (7) *Linked Data access* does not involve fusion processes directly but rather is aimed at extracting information from related spatial data. This can be achieved either by traversing an RDF graph or through the application of SPARQL queries to a triple store. The results of such a data search can be evaluated and resolved against SDI data services. (Example: A client obtains a road network from an SDI data service. Subsequently, the client attempts to identify data services that provide related networks [2:find], which have previously been registered [1:publish]. If any relations are found, they can be accessed [3:bind] and resolved against the underlying data service [4:resolve].)

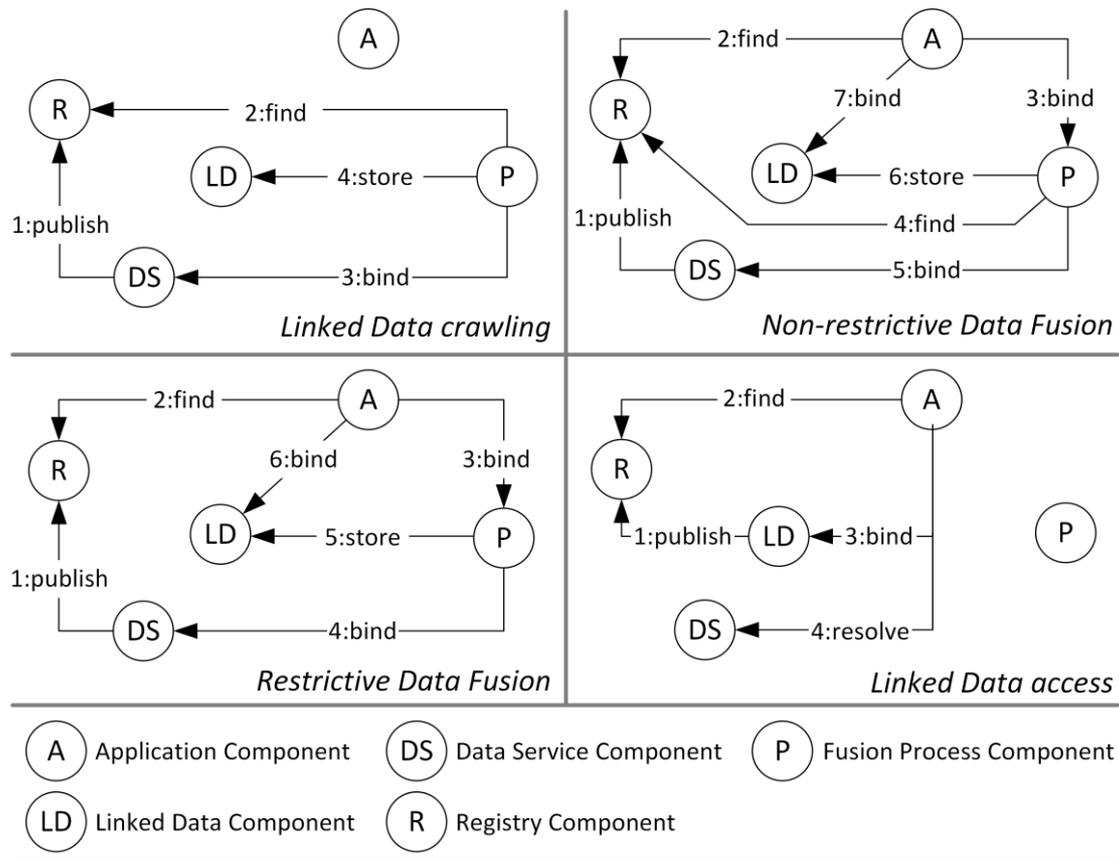


Figure 6: Usage patterns for spatial data fusion using Linked Data.

The patterns of Linked Data access and Linked Data crawling are already well supported in the Semantic Web domain. However, on-demand information retrieval from distributed sources within SDIs requires additional capabilities of restrictive and non-restrictive data fusion.

#### 4. Prototype Implementation

To validate the applicability of the described concepts and to demonstrate the benefits of performing spatial data fusion in SDIs in combination with Linked Data, a prototype was implemented and evaluated for three different use cases.

The implemented infrastructure for service-based spatial data fusion comprises four components (Figure 7):

- Fusion input data are provided via the OGC Web Feature Service (WFS) interface. A local WFS is set up using the GeoServer<sup>v</sup> framework. However, externally accessible WFS instances can also be used.
- The Web client is built on JavaServer Faces (JSF) in combination with OpenLayers<sup>vi</sup> for map rendering and jsPlumb<sup>vii</sup> to offer a user interface for the

orchestration of the fusion process. It supports the retrieval of data via WFS interfaces, the customization of fusion processes, the encoding and invocation of the workflow using BPMN (Business Process Modeling Language) and the interactive visual analysis of relations using pre-defined SPARQL queries.

- Fusion processes are provided via the OGC Web Processing Service (WPS) interface using the 52°North WPS framework<sup>viii</sup>. The underlying software library allows additional processes to be implemented and directly exposed through the WPS interface.
- A Linked Data store for feature relations based on the Apache Jena Fuseki triple store<sup>ix</sup> provides a SPARQL endpoint for requesting and reasoning over identified relations. Server-side access, SPARQL queries and updating are realized using the provided Java API. SPARQL queries issued by the client are executed using the HTTP interface of the triple store.

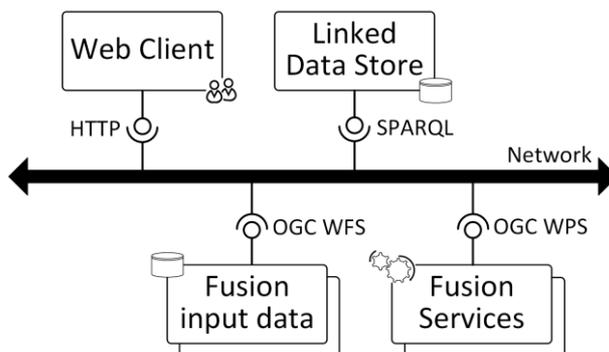


Figure 7: Implemented service infrastructure for spatial data fusion.

All feature relations identified by the fusion processes are represented in accordance with the previously described formalization schema (Figure 4). The generalized RDF encoding for a feature relation is shown in Listing 1 and contains information regarding the features participating in the relation, the assigned relation types and the relation measurements conducted. To comply with current implementations of the WFS interface, which do not support RDF, and to avoid the custom conversion of spatial data into RDF, the implementation applies a mediation approach to relation management. Thus, the relations are stored independently of the input sources. The feature identifiers are represented by local identifiers in the form `'[WFS server URL]?[WFS GetFeature request]#[GML id]'`, which can be resolved to a GML representation of a feature or set of features. The relation measurements provide lineage information by linking to the originating fusion process(es) and specific measurement values. The relation types are typically based on an interpretation of the relation measurements. Those measurements are explicitly linked in the relation type description, which also contains information concerning the roles of the features participating in a particular relation. Examples of relation types include the feature set relations, comprising disjoint, same as, subset of, superset of and multiple relations (see Figure 11), and the conceptual relation types implemented by SKOS,

describing related, broad, narrow, close and exact semantic relations.

Listing 1: Simplified RDF Turtle representation of a feature relation.

```
@[prefix definitions]
-----
foreach relation
_:relation_[uid]
  a [namespace]/featureRelation ;
  fusion:hasReference [reference URI] ;
  fusion:hasTarget [target URI] ;
  xsd:dateTime [timestamp]
-----
foreach relation type
fusion:hasRelationType [
  a [namespace]/[relation type] ;
  fusion:hasProcessURI [process URI] ;
  fusion:hasDescription [description URI]
] ;
-----
foreach relation measurement
fusion:hasRelationMeasurement [
  a [namespace]/[measurement type] ;
  fusion:hasProcessURI [process URI] ;
  rdf:value [measurement value]^^[value type] ;
  fusion:hasDescription [description URI]
] .
-----
```

The web client supports the creation and definition of fusion workflows. A workflow begins with the selection of the input sources, or more precisely, two WFS service endpoints. After the reference dataset and target dataset have been selected, the software provides information about the potential comparability of the datasets by analyzing the spatial reference systems used and their spatial extents. If a user wishes to compare and combine the datasets using fusion processes, a suitable WPS instance must be provided. The WPS used for demonstration here provides a set of low-level fusion functions, as defined according to the decomposition of processes described earlier in this paper. All processes provided by the WPS are listed with a name and description and are filtered according to pre-defined restrictions on possible inputs and ordered in accordance with their decomposition (c.f. Figure 2). As an example, angle, length and sinuosity difference measurements will not be displayed for input point data. The desired processes can be selected and manually orchestrated from within the client (Figure 8). To improve usability, the process descriptions and the supported input and output formats are read from the WPS DescribeProcess document and provided to the user. Any required literal

inputs can be added to the workflow using a corresponding input field and connected to the processes as appropriate.

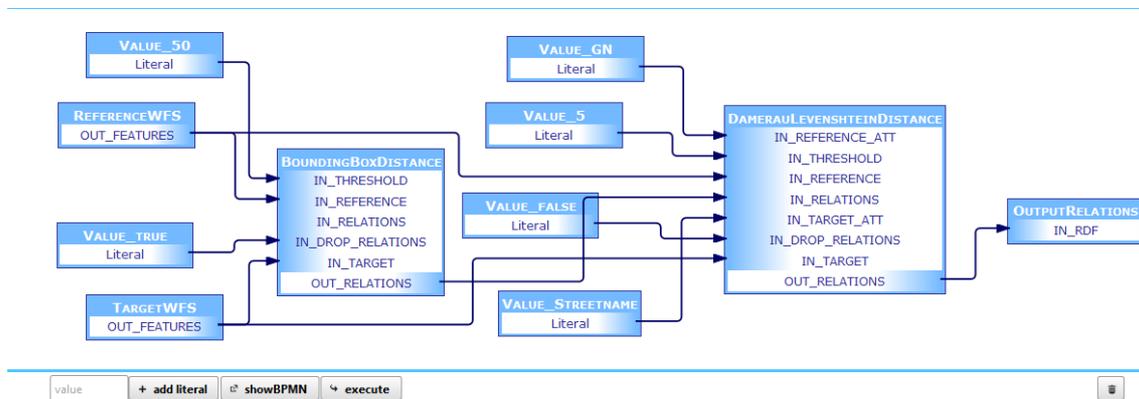


Figure 8: User-defined orchestration of WPS processes (screenshot).

The WPS process interfaces share the same design principles. The mandatory inputs are the reference features and target features. For the subsequent relation measurements, existing relations can optionally be provided to limit the number of required comparisons. In parallel with the graphically user-performed process orchestration, a BPMN file is created that contains information on the workflow. Once the fusion process is invoked by clicking the execute button, this BPMN file is sent to a separate WPS process for interpretation and execution. After the successful completion of the workflow, all identified feature relations are written to the triple store via SPARQL update queries. If requested, the WPS can also directly return relations encoded in RDF following the structure shown in Listing 1. The entire workflow sequence for client-based spatial data fusion is depicted in Figure 9.

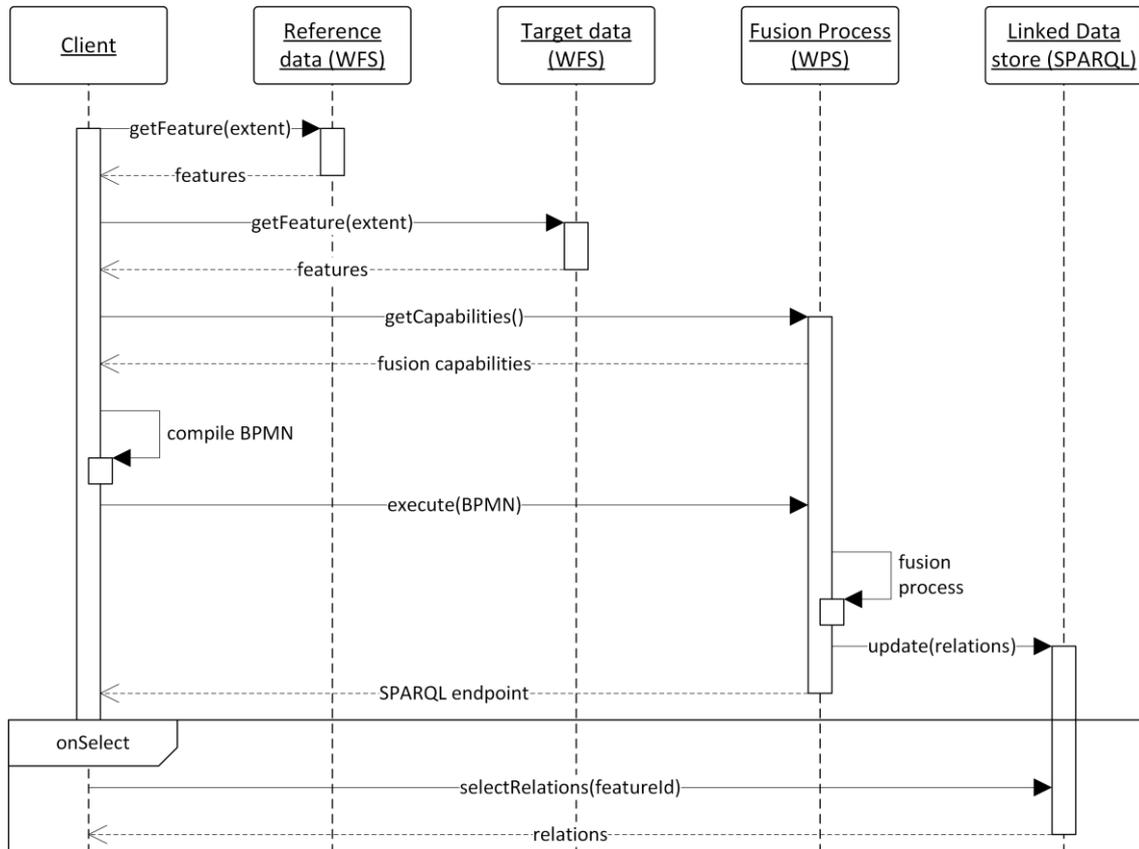


Figure 9: Sequence diagram for client-based compilation and execution of a data fusion process for defined spatial data sources.

Once the process has completed, the triple store can be accessed using SPARQL queries. Building on this functionality, the client supports the interactive visualization of the relations and feature attributes. When a feature is selected, a SPARQL query is compiled and executed to highlight related features on the corresponding map (Figure 11). The relation types and measurements associated with the related features are listed in the attribute table.

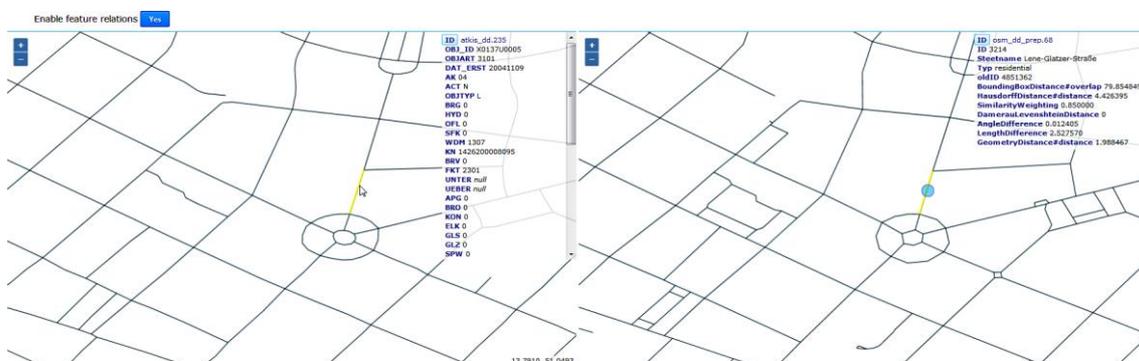


Figure 10: Comparison of input datasets based on previously identified relationships (screenshot).

Because standardized interfaces are used, other applications can also use the provided processes to create spatial data relations in an automated fashion without client interaction. This can be achieved using a third-party WPS orchestration engine and allows for access to and use of relations encoded in RDF for a multitude of purposes, including information transfer, data integration and cross-data analysis.

### *Evaluation*

Three use cases were chosen for evaluation to demonstrate the applicability and effectiveness of the prototype implementation for data inspection, integration and cross-data analysis. In addition, a performance evaluation is presented to demonstrate the scalability of the approach.

#### *Matching of road networks*

The first use case addresses the synthesis of overlapping road networks to detect identity relations and mutual updates. For this purpose, authoritative data following the German ATKIS model<sup>x</sup> (Authoritative Topographic Cartographic Information System) was compared to street network data provided by the OpenStreetMap project<sup>xi</sup>. Both data sources are set up as WFS layers.

For relation building, a number of geometric, topological and confidence criteria were selected. Subsequent to data retrieval, the bounding box relation was computed for each of the input features, thereby constructing a spatial index on the reference dataset and creating a relation for all target features with a bounding box distance of less than 50 m. This threshold was selected as a decisive criterion and significantly reduced the number of possible relations for further processing. Subsequently, calculations were performed to determine the minimum spatial distance (between closest points), Hausdorff distance, angle difference, length difference and topological relation between feature geometries that were identified as being possibly related in the previous bounding box selection. Moreover, a string distance metric, namely, the Damerau-Levenshtein distance, was calculated between road names. Finally, a simple confidence measure was added to weight the similarity measurements associated with each relation and assigns a confidence value from 0 (very unlikely) to 1 (very likely). The confidence values were calculated as follows:

- +0.1, if the bounding boxes intersect or the distance is less than 50 m;
- +0.2, if the angle difference is less than  $\pi/10$ ;
- +0.1, if the minimal geographic distance is less than 50 m;
- +0.1, if the length difference is less than 20 m;
- +0.15, if the Hausdorff distance is less than 50 m;

- +0.15, if the topology relation is not disjoint;
- +0.2, if the Damerau-Levenshtein distance between the road names is less than 4.

Once the relation measurements were performed and the results inserted into the triple store, the set relation types could be determined via SPARQL queries following the identification schema depicted in Figure 11. The determined relation type could be attached to each existing feature relation and used for further analysis. As shown in Table 1, more restrictive criteria for the relations (in this case, a higher confidence threshold) led to the identification of fewer feature relations but increased the number of *sameAs* relations. As the precision increased (from ~0.4 to 0.9), the recall rate of existing feature matches declined (from >0.9 to ~0.6). More complex feature matching procedures would certainly increase both precision and recall. Nevertheless, the simple specification of a confidence threshold significantly influences the quality and the quantity of the requested relations.

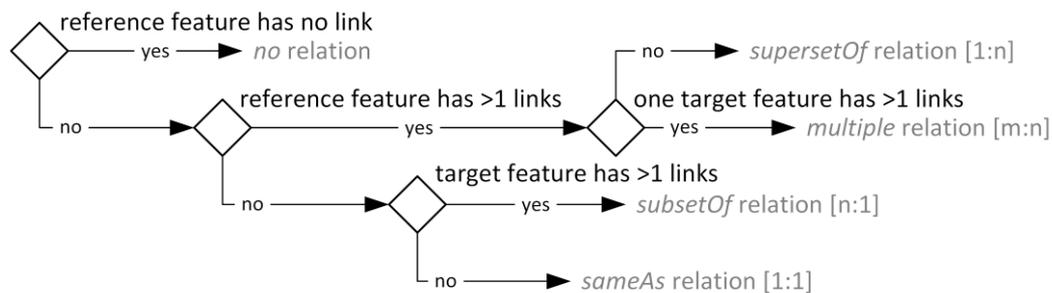


Figure 11: Identification schema for types of relations between road segments.

Table 1: Cardinality of relations with regard to the selected confidence threshold.

ATKIS → OSM	confidence $\geq 0.4$	confidence $\geq 0.8$
Total number of features	687	687
Feature with relation	670	360
sameAs [1:1] relation	11	296
subsetOf [n:1] relation	53	28
supersetOf [1:n] relation	10	11
Multiple [m:n] relation	470	25

The ATKIS licensing and the administrative requirements for quality assurance prevent the transfer of data to or from OSM. However, hints regarding missing, incomplete, outdated or erroneous data can be derived to support update cycles and data maintenance (Wiemann and Bernard 2014). SPARQL queries on the relations assist in selecting and resolving features in each dataset that have no counterparts or have significantly differing representations in the other dataset.

### *Provenance and change detection*

The same matching strategy described above can be used to describe and store different versions of a feature, thus supporting versioning and history management. For this purpose, the relations must model the origin and evolution of the features over time. Accordingly, the reference and target features can be regarded as the ancestor and successor of a particular feature. If stored with the features, the Dublin Core vocabulary<sup>xii</sup> can be used to assign timestamps, versions or replacements. When the implemented mediation approach is used, changes can be explicitly stored in change sets, similar to the DBPedia-based approach described by Orlandi and Passant (2011) using the RDF ChangeSet vocabulary<sup>xiii</sup>.

As an example, two OSM datasets were chosen, and the changes in these dataset over a given period of time (01.01.2014 to 19.05.2015) were identified and stored. Because OSM features do not provide persistent unique identifiers or an identifier scheme, a simple join is insufficient to link different versions of the database. Instead, it is recommended that homologous features are identified by a certain combination of tags, including geometric and thematic attributes, which requires a number of relation measurements.

The older and newer OSM datasets contained 7875 and 9674 features, respectively, and were made accessible via WFS. Because primarily road features were selected, the same relation measurements and identity relation types that were used for the previous use case were chosen. Subsequently, the relations were requested from the triple store, the reference and target features were resolved against the WFS, the changes in geometry and attributes were computed following the workflow outlined in

Listing 2, and finally, the relations were complemented with those changes. With reference to the proposed feature relation components (Figure 4), the change sets were treated as relation measurements representing the measured changes between input features. A sample relation expressed using the ChangeSet vocabulary is shown in

Listing 3.

Listing 2: Pseudo-code of the workflow for the detection and creation of change sets between feature relations.

```
foreach reference attribute RA do:
  TA = target attribute corresponding to RA
  if TA == none do:
    create change set(
      subjectOfChange = RA
      removal = RA.value
      addition = none)
  else:
    TA.visited = true
    if RA.value != TA.value do:
      create change set(
        subjectOfChange = RA
        removal = RA.value
        addition = TA.value)
    else do:
      nothing
foreach target attribute TA where TA.visited != true do:
  create change set(
    subjectOfChange = TA
    removal = none
    addition = TA.value)
```

Listing 3: RDF-encoded relation between two OSM features expressed in terms of change

sets.

```
@[prefix definitions]
_:FeatureRelation_1b8ca48f-6391-4ab5-a1bf-e3c2771f43de
  a fusion:featureRelation ;
  fusion:hasReference <[WFS request]#roads_old.6922> ;
  fusion:hasTarget <[WFS request]#roads_new.6705> ;
  ...
  fusion:hasRelationMeasurement [
    a cs:changeSet ;
    cs:subjectOfChange osm:geometry ;
    cs:removal [
      rdf:subject osm:geometry ;
      red:predicate rdf:value ;
      rdf:object "MULTILINESTRING ((...))"
    ] ;
    cs:addition [
      rdf:subject osm:geometry ;
      red:predicate rdf:value ;
      rdf:object "MULTILINESTRING ((...))"
    ]
  ] ;
  fusion:hasRelationMeasurement [
    a cs:changeSet ;
    cs:subjectOfChange osm:type ;
    cs:removal [
      rdf:subject osm:type;
      red:predicate rdf:value ;
      rdf:object "footway"
    ] ;
    cs:addition [
      rdf:subject osm:type ;
      rdf:predicate rdf:value ;
      rdf:object "path"
    ]
  ] .
```

In the example shown, both the feature geometry, encoded as well-known text, and the attribute ‘type’ exhibited changes; both are represented as literal objects in RDF. In total, 3623 geometry changes and 2733 attribute changes were identified between the two OSM versions. Among the latter, id changes (1444) and type changes (592) were the most common.

### *Linking different feature representations*

As long as features can be uniquely identified, they can be referenced by feature relations, as defined in this paper. For vector data, this is typically achieved using feature or feature

collection identifiers. For raster data, these identifiers can be equated with cells or collections of cells, respectively. Within the SDI framework, the OGC Web Coverage Service (WCS) offers the capability of separating raster data into subsets to restrict their spatial, temporal and thematic extent. Thus, similar to the WFS example discussed earlier, we assume that a feature expressed as raster data can be identified and referenced via a WCS request.

When linking vector data to raster data, the spatial extent of a feature within a coverage may be implicitly indicated by the vector geometry. Accordingly, relations can describe zonal statistics for polygons, linear profiles for polylines and single coverage values for points. The first scenario was implemented as an example to relate municipality data provided via WFS to an elevation coverage provided via WCS. An example of a corresponding RDF encoding of a relation is shown in Listing 4. Although both reference and target are referenced by their corresponding OGC service requests, the extent of the target raster feature is implicitly specified by the geometry of the reference feature. The relation type defaults to *relatedTo* because no other specific relation type is defined.

Listing 4: Sample RDF relation between a polygon feature provided via WFS and an elevation raster provided via WCS.

```
@[prefix definitions]
_:relation_[uid]
  a [namespace]/featureRelation;
  fusion:hasReference <http:// ... #municipalities.317> ;
  fusion:hasTarget <http:// ... coverageId=fusion__dem> ;
  fusion:hasRelationMeasurement [
    a [namespace]/relationMeasurement ;
    fusion:hasProcessURI process:ZonalStatistics ;
    rdf:value "320.34"^^xsd:decimal ;
    fusion:hasDescription measurement:ZonalStatistics#min ;
  ] .
...
```

### *Performance evaluation*

An evaluation of the performance with respect to process runtime and memory consumption was conducted based for the first use case (synthesizing road networks). For this purpose, all measurements were applied to three feature samples — small, medium and large — for the city of Dresden (Figure 12). All tests were performed within a local test environment and therefore were unaffected by overheads incurred as a result of network transfer. The test computer contained an Intel Core i5-4300M CPU at 2.6 GHz, 8 GB of memory and a solid-state drive (SSD), running a 64-bit Windows 7 operating system with Java version 7. The results of the comparison are shown in Table 2.

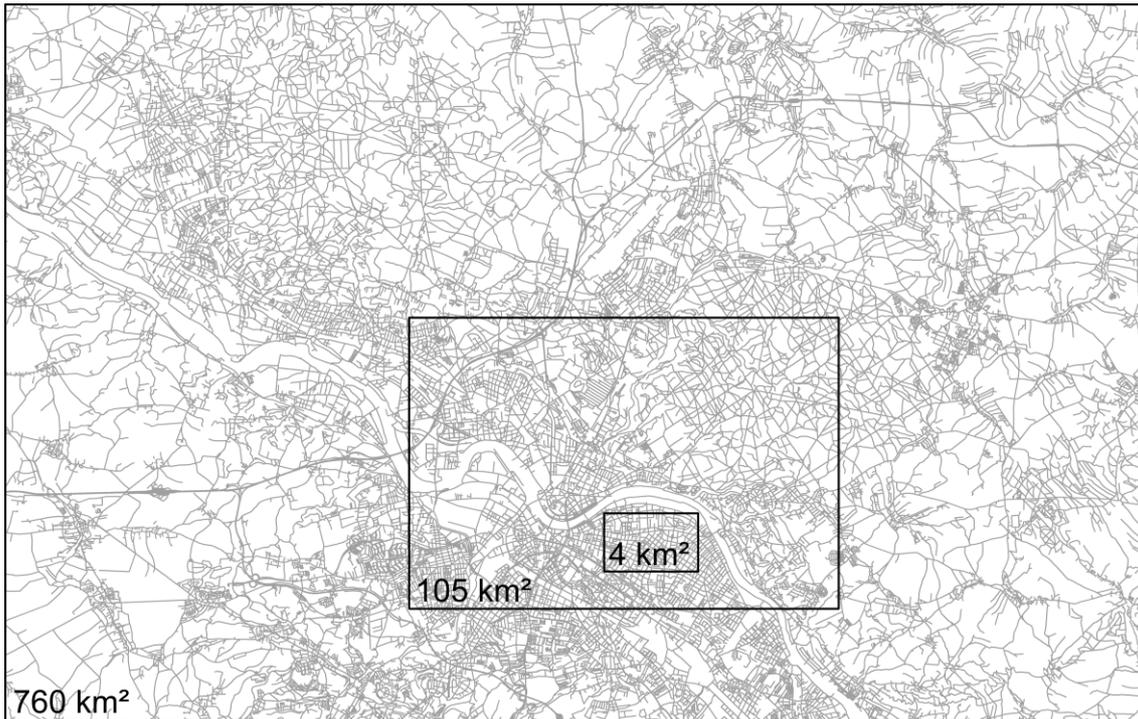


Figure 12: Feature samples for the city of Dresden considered for the performance evaluation.

Table 2: Results of the performance evaluation.

Size of dataset	small	medium	large
Reference features (ATKIS)	687	11,996	54,859
Reference size (uncompressed GML)	1MB	18MB	86MB
Target features (OSM)	1,257	23,383	67,568
Target size (uncompressed GML)	<1MB	16MB	49MB
Identified relations	1383	23,384	87,808
Number of triples	34,085	658,944	2,186,240
Total runtime	~15s	~120s	~340s
Data retrieval	~8s	~27s	~95s
Writing relations	~6s	~78s	~230s
JVM internal memory footprint	47MB	245MB	724MB

As seen from the comparison table, the performance of the application scales well with an increasing number of features and feature relations. The most time-consuming components are the data retrieval via WFS GetFeature requests and the writing of feature relations to the triple store via SPARQL update queries. Consequently, efforts concerning performance optimization should concentrate first on the reduction of the amount of data to be transferred, e.g., by setting a certain confidence threshold for relations to be recorded, where reasonable. The same applies for non-restrictive data fusion, which will introduce additional overhead due to the identification and retrieval of suitable datasets and fusion processes on the Web.

### *Discussion of the implementation*

In terms of the proposed usage patterns, the current prototype implements restrictive data fusion and Linked Data access. Of the patterns that were not implemented, Linked Data crawling can be achieved by extending the capabilities for batch processing; however, non-restrictive data fusion remains subject to further research, mainly to address the identification of suitable datasets for data fusion and the search for and automated orchestration of fusion processes on the Web. The implemented input data restrictions and process classification, which will both be elaborated further, can serve as a starting point for this effort. The mediation approach was chosen for relation storage because the majority of SDI spatial data services on the Web do not support direct or proxy RDF encoding.

The main access point to the implemented prototype is the Web-based client, which allows for an ad hoc fusion of feature datasets provided via WFS and the visual inspection and interpretation of relations with no need for a desktop GIS and very low client-side computing requirements. Because it uses standardized interfaces and modular services, the implementation remains flexible and extensible. The SPARQL endpoint of the published triple store allows for the retrieval and analysis of feature relations in a standardized manner.

The implemented Web client enforces a separation between data processing and application and is therefore well suited for non-IT experts. However, direct access to the underlying standards-based service infrastructure enables additional functionality, such as batch processing, additional input sources (e.g., further OGC data services), and custom SPARQL queries for cross-data analysis.

Currently, the features are related based on their spatial and thematic attributes because these attributes are the best candidates for identifying homologous features. In an ideal scenario, semantics attached to the spatial data would improve the feature matchmaking process and support reasoning over relations. However, because the majority of spatial data sets available on the Web do not (yet) provide formalized semantic information, semantic matching methods are currently not implemented in the prototype. Nonetheless, the implementation allows for extension to semantic relation measurements.

For practical reasons, a number of technical compromises were made, and thus, the prototype would require further development to be considered an operational application. First, a number of the WFSs that are available on the Web are mirrored to facilitate ready access by the client. Furthermore, for security issues, the triple store and the WPS are tightly coupled and located on the same machine. Finally, the identified relations are currently stored only as blank nodes with no resolvable URI, thus preventing explicit referencing from outside the triple store.

## 5. Conclusion and challenges for further research

The suggested approach supports the fusion of spatial data from SDIs and makes the fusion results accessible to mainstream Web technologies. Furthermore, the approach is flexible and allows for extension. This is achieved through the use of open standards from both the geospatial and Semantic Web domains. The calculated feature relations are explicitly stored to enable their on-demand usage for various application targets. It should be noted, however, that in addition to technical solutions, a number of non-technical issues at the institutional, policy, legal and social levels must be solved before operational deployment (Mohammadi et al. 2010).

The presented implementations are meant as a framework and are not (yet) focused on achieving the most robust and reliable feature mappings; rather, they are focused on demonstrating use cases that prove the feasibility and applicability of the presented approach. Additional process implementations will certainly improve the fusion processes and increase the actual usability of the results. However, the flexibility of the framework allows for the extension of its functionality by means of atomic, low-level or high-level processes provided via the WPS interface.

For the future realization of non-restrictive data fusion and the ability to work dynamically with existing data sources and processes to fulfill given application targets, three major fields of research can be identified: First, further capabilities for service-based spatial data fusion on the Web in an interoperable and flexible manner should be built and maintained. This will require a review of existing implementations with respect to their suitability for service-based spatial data fusion. Because computational performance is crucial for Web-based processing, a balancing of process complexity with respect to particular use cases is required. Although complexity reduction lowers the precision and recall rates of matching, the matching performance might still be sufficient as long as the uncertainty is properly documented. On a general level, the definition of benchmarks for spatial data fusion in SDIs could facilitate the evaluation and validation of related implementations.

Second, the gap between SDIs and the Semantic Web should be further closed in a mutual learning process. This will require the elaboration of an appropriate demarcation between applications for which one or the other technology is more appropriate. Likewise, a distinction must be drawn among explicitly required, implicitly stated and non-relevant relations, as the significance of a relation decreases with increasing spatial, temporal or semantic distance. Furthermore, a common set of ontologies and vocabularies for describing the relationships between spatial features should be established, linked and harmonized to facilitate federated queries and the utilization of relations by third parties. However, to advance semantic reasoning capabilities, not only the data but also the concepts underlying the data must be formalized and interlinked.

Third, the target requirements for spatial data fusion should be formalized to enable the automated inference of reliable and meaningful information for decision-making. The selection of an appropriate fusion processes significantly depends on contextual information, such as minimum quality and computational needs, data access constraints and targeted information content, which therefore must be expressed in a formalized manner. Considering that Semantic Web applications are generally regarded as imperfect (Auer and Lehmann 2010, Hitzler and van Harmelen 2010), strategies for handling uncertain, inconsistent or missing data and relations should also be addressed. Finally, to further expand the possible uses of the Semantic Web for decision- and policymaking, additional abstraction layers are required to avoid conflict with specifics of the RDF or SPARQL language.

With respect to the idea of a spatially enabled Semantic Web, as expressed by Egenhofer (2002), the ability to combine spatial data from SDIs using Semantic Web technologies is considered an important step forward. Nevertheless, hope, hype and reality must be further separated to neither exaggerate expectations nor impair usability for real-world applications.

### **Acknowledgements**

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 308513, COBWEB.

We would like to thank the editor and the anonymous reviewers for their suggestions and comments that helped to improve the quality of the paper.

### **References**

- Acosta, M., et al., 2013. Crowdsourcing Linked Data quality assessment. In: 12th International Semantic Web Conference, Vol. 8219 of Lecture Notes in Computer Science, Sydney, Australia Springer-Verlag, 260–276.
- Al-Bakri, M. and Fairbairn, D., 2012. Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources. *International Journal of Geographical Information Science*, 26 (8), 1437–1456.
- Auer, S. and Lehmann, J., 2010. Creating knowledge out of interlinked data. *Semantic Web*, 1 (1-2), 97–104.
- Bernard, L., Kanellopoulos, I., Annoni, A., Smits, P., 2005. The European Geoportal - One step towards the Establishment of a European Spatial Data Infrastructure. *Computers, Environment and Urban Systems*, 29(1), 15–31.
- Berners-Lee, T., Hendler, J., and Lassila, O., 2001. The Semantic Web. *Scientific American*, 284 (5), 34–43.

- Bizer, C., Heath, T., and Berners-Lee, T., 2009. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5 (3), 1–22.
- Bloch, I., et al., 2001. Fusion: General Concepts and Characteristics. *International Journal of Intelligent Systems*, 16, 1107–1134.
- Bröring, A., et al., 2014. OGC Testbed-10 CCI VGI Engineering Report. Technical report, Open Geospatial Consortium.
- Butenuth, M., et al., 2007. Integration of heterogeneous geospatial data in a federated database. *ISPRS Journal of Photogrammetry & Remote Sensing*, 62 (2007), 328–346.
- Cholvy, L., 2007. Modelling Information Evaluation in Fusion. In: *Proceedings of the 10<sup>th</sup> International Conference on Information Fusion*, Quebec, Canada, 1–6.
- Cobb, M.A., et al., 1998. A Rule-based Approach for the Conflation of Attributed Vector Data. *GeoInformatica*, 2 (1), 7–35.
- Cox, S., 2013. An explicit OWL representation of ISO/OGC Observations and Measurements. In: O. Corcho, C. Henson and P. Barnaghi, eds. *Proceedings of the 6th International Workshop on Semantic Sensor Networks*, Vol. 1063 of *CEUR Workshop Proceedings*, Sydney, Australia, 1–18.
- Devoegele, T., Parent, C., and Spaccapietra, S., 1998. On spatial database integration. *International Journal of Geographical Information Science*, 12 (4), 335–352.
- Doytsher, Y., Filin, S., and Ezra, E., 2001. Transformation of Datasets in a Linear-based Map Conflation Framework. *Surveying and Land Information Systems*, 61 (3), 159–169.
- Egenhofer, M.J., 2002. Toward the semantic geospatial web. In: *Proceedings of the 10<sup>th</sup> ACM international symposium on Advances in geographic information systems*, McLean, Virginia, USA, 1–4.
- Erl, T., 2008. *SOA: Principles of Service Design*. Prentice Hall.
- Goodwin, J., Dolbear, C., and Hart, G., 2008. Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web. *Transactions in GIS*, 12 (Supplement 1), 19–30.
- Heath, T. and Bizer, C., 2011. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool.
- Hitzler, P. and van Harmelen, F., 2010. A reasonable Semantic Web. *Semantic Web*, 1 (1-2), 39–44.
- INSPIRE, 2014. D3.4 INSPIRE Generic Conceptual Model, INSPIRE Drafting Team "Data Specifications".
- Janowicz, K., et al., 2013. A RESTful Proxy and Data Model for Linked Sensor Data. *International Journal of Digital Earth*, 6 (3), 233–254.
- Janowicz, K., et al., 2010. Semantic Enablement for Spatial Data Infrastructures. *Transactions in GIS*, 14 (2), 111–129.
- Karam, R. and Melchiori, M., 2013. Improving Geo-spatial Linked Data with the Wisdom of the Crowds. In: *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, Genoa, Italy, 68–74.

- Kiehle, C., Heier, C., and Greve, K., 2007. Requirements for Next Generation Spatial Data Infrastructures-Standardized Web Based Geoprocessing and Web Service Orchestration. *Transactions in GIS*, 11 (6), 819–834.
- Koubarakis, M. and Kyzirakos, K., 2010. Modeling and Querying Metadata in the Semantic Sensor Web: the Model stRDF and the Query Language stSPARQL. In: 7th Extended Semantic Web Conference, ESWC 2010, Vol. 6088 of Lecture Notes in Computer Science, Heraklion, Greece, 425–439.
- Koukoletsos, T., Haklay, M. and Ellul, C., 2012. Assessing Data Completeness of VGI through an Automated Matching Procedure for Linear Data. *Transactions in GIS*, 16 (4), 477–498.
- Longley, P.A., et al., 2010. *Geographic Information Systems & Science*. 3 John Wiley & Sons.
- López-Vázquez, C. and Callejo, M.A.M., 2013. Point- and curve-based geometric conflation. *International Journal of Geographical Information Science*, 27 (1), 192–207.
- Lupien, A.E. and Moreland, W.H., 1987. A General Approach to Map Conflation. In: *Proceedings of Auto-Carto 8*, Baltimore, Maryland, USA, 630–639.
- Mohammadi, H., Rajabifard, A. and Williamson, I.P., 2010. Development of an interoperable tool to facilitate spatial data integration in the context of SDI. *International Journal of Geographical Information Science*, 24 (4), 487–505.
- OGC, 2009. OGC Abstract Specification, Topic 8: Relationships Between Features. Version 4, Technical report, Open Geospatial Consortium Inc.
- OGC, 2012. OGC GeoSPARQL - A Geographic Query Language for RDF Data. OpenGIS Standard, Open Geospatial Consortium.
- Orlandi, F. and Passant, A., 2011. Modelling provenance of DBpedia resources using Wikipedia contributions. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9 (2), 149–164.
- Ruiz, J.J., et al., 2011. Digital map conflation: a review of the process and a proposal for classification. *International Journal of Geographical Information Science*, 25 (9), 1439–1466.
- Saalfeld, A., 1988. Conflation - Automated Map Compilation. *International Journal of Geographical Information Systems*, 2 (3), 217–228.
- Samal, A., Seth, S. and Cueto, K., 2005. A feature-based approach to conflation of geospatial sources. *International Journal of Geographical Information Science*, 18 (5), 459–489.
- Schade, S., Granell, C. and Días, L., 2010. Augmenting SDI with Linked Data. In: *Workshop On Linked Spatiotemporal Data*, in conjunction with the 6th International Conference on Geographic Information Science.
- Schade, S. and Smits, P., 2012. Why linked data should not lead to next generation SDI. In: *Geoscience and Remote Sensing Symposium (IGARSS)*, 2894–2897.
- Schwering, A., 2008. Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey. *Transactions in GIS*, 12 (1), 5–29.

- Schwinn, A. and Schelp, J., 2005. Design patterns for data integration. *Journal of Enterprise Information Management*, 18 (4), 471–482.
- Sester, M., Anders, K.H., and Walter, V., 1998. Linking Objects of Different Spatial Data Sets by Integration and Aggregation. *GeoInformatica*, 2 (4), 335–358.
- Shvaiko, P. and Euzenat, J., 2005. A Survey of Schema-Based Matching Approaches. In: *Journal on Data Semantics IV*, Vol. 3730 of *Lecture Notes in Computer Science*, Springer-Verlag, 146–171.
- Song, W., Haithcoat, T.L., and Keller, J.M., 2006. A Snake-based Approach for TIGER Road Data Conflation. *Cartography and Geographic Information Science*, 33 (4), 287–298.
- Stadler, C., et al., 2012. LinkedGeoData: A Core for aWeb of Spatial Open Data. *Semantic Web*, 3 (4), 333–354.
- Stankute, S. and Asche, H., 2012. A Data Fusion System for Spatial Data Mining, Analysis and Improvement. In: B. Murgante, O. Gervasi, S. Misra, N. Nedjah, A.A.C. Rocha, D. Taniar and B.O. Apduhan, eds. *Computational Science and Its Applications - ICCSA 2012*, Vol. 7334 of *Lecture Notes in Computer Science* Springer-Verlag, 439–449.
- Uitermark, H., Vogels, A., and van Oosterom, P., 1999. Semantic and Geometric Aspects of Integrating Road Networks. In: A. Včkovski, K.E. Brassel and H.J. Schek, eds. *Interoperating Geographic Information Systems, Proceedings of INTEROP'99*, Vol. 1580 of *Lecture Notes in Computer Science*, Zurich Springer-Verlag, 177–188.
- Veltkamp, R.C. and Hagedoorn, M., 2001. State-of-the-art in shape matching. In: M.S. Lew, ed. *Principles of Visual Information Retrieval*. Springer-Verlag, 87–119.
- Volz, J., et al., 2009. Discovering and maintaining links on the web of data. In: 8th *International Semantic Web Conference, ISWC 2009*, Vol. 5823 of *Semantic Web-ISWC*, Springer-Verlag, 650–665.
- Volz, S., 2005. Data-Driven Matching of Geospatial Schemas. In: A.G. Cohn and D.M. Mark, eds. *Spatial Information Theory*, Vol. 3693 of *Lecture Notes in Computer Science*, Springer-Verlag, 115–132.
- Walter, V. and Fritsch, D., 1999. Matching spatial data sets: a statistical approach. *International Journal of Geographical Information Science*, 13 (5), 445–473.
- Waltz, E. and Waltz, T., 2008. Principles and Practice of Image and Spatial Data Fusion. In: M.E. Liggins, D.L. Hall and J. Llinas, eds. *Handbook of Multisensor Data Fusion*. 2 ed. CRC Press, 89–114.
- Wiemann, S. and Bernard, L., 2014. Linking crowdsourced observations with INSPIRE. In: *Proceedings of the AGILE'2014 International Conference on Geographic Information Science*, Castellón, Spain.
- Yuan, S. and Tao, C., 1999. Development of Conflation Components. In: *The Proceedings of Geoinformatics'99 Conference*, Ann Arbor, Michigan, USA, 1–13.

## List of Figures

Figure 1: Classification schema for spatial data fusion processes. ....	3
Figure 2: Low-level sub-processes for spatial data fusion. ....	4
Figure 3: Interaction between SDI (OWS interface) and Semantic Web (RDF interface) components. ....	6
Figure 4: Proposed components for the formalization of feature relations. ....	8
Figure 5: Approaches to the management of feature relations between two data sources. ....	9
Figure 6: Usage patterns for spatial data fusion using Linked Data. ....	11
Figure 7: Implemented service infrastructure for spatial data fusion. ....	12
Figure 8: User-defined orchestration of WPS processes (screenshot). ....	14
Figure 9: Sequence diagram for client-based compilation and execution of a data fusion process for defined spatial data sources. ....	15
Figure 10: Comparison of input datasets based on previously identified relationships (screenshot). ....	15
Figure 11: Identification schema for types of relations between road segments. ....	17
Figure 12: Feature samples for the city of Dresden considered for the performance evaluation. ....	22

## List of Tables

Table 1: Cardinality of relations with regard to the selected confidence threshold. ....	17
Table 2: Results of the performance evaluation. ....	22

## List of Listings

Listing 1: Simplified RDF Turtle representation of a feature relation. ....	13
---	----

Listing 2: Pseudo-code of the workflow for the detection and creation of change sets between feature relations.....	19
Listing 3: RDF-encoded relation between two OSM features expressed in terms of change sets.....	20
Listing 4: Sample RDF relation between a polygon feature provided via WFS and an elevation raster provided via WCS.....	21

<sup>i</sup> <http://www.w3.org/DesignIssues/LinkedData.html>

<sup>ii</sup> <http://www.opengeospatial.org/standards>

<sup>iii</sup> <http://www.w3.org/standards>

<sup>iv</sup> <http://www.w3.org/TR/skos-reference>

<sup>v</sup> <http://geoserver.org>

<sup>vi</sup> <http://openlayers.org>

<sup>vii</sup> <http://www.jsplumb.org>

<sup>viii</sup> <http://52north.org/communities/geoprocessing/wps>

<sup>ix</sup> <https://jena.apache.org>

<sup>x</sup> <http://www.adv-online.de/Publications/AFIS-ALKIS-ATKIS-Projekt/GeoInfoDoc>

<sup>xi</sup> <http://www.openstreetmap.org>

<sup>xii</sup> <http://dublincore.org/documents/dcmi-terms>

<sup>xiii</sup> <http://purl.org/vocab/changeset>