Towards a Knowledge Base to Support Geoprocessing Workflow Development

Barbara Hofer, Stephan Mäs, Johannes Brauner & Lars Bernard

Abstract: The spatial analysis functionalities of geographic information systems (GIS) are increasingly used across the web. Interface specifications of geoprocessing web services define the syntactic properties of the services (number and type of parameters) and provide textual descriptions of the operations. The discovery and reuse of web services based on these syntactic properties is restricted and has led to the quest for extended operation descriptions. A number of extended operation descriptions have been proposed and are reviewed in this paper. The reviewed descriptions focus on particular steps of the workflow development process. In this paper we analyse all phases of the development process of a spatial analysis workflow regarding the requirements of operation descriptions. These requirements are translated into a knowledge base that contains information about spatial analysis operations for the operations' discovery, selection and composition. The knowledge base also foresees the automated discovery of operations in well-defined application contexts such as the transformation of data types or coordinate reference systems. The knowledge base combines the geooperator approach with elements of ontology-based approaches. The geooperator browser is the implementation of the discovery and selection of geoprocessing operations based on the knowledge base. A draft formalization of the knowledge base demonstrates its use and the support it provides during the composition of operations.

Keywords: spatial analysis workflows, operation descriptions, geooperators, web services, geoprocessing

This is an Accepted Manuscript of an article published by Taylor & Francis in International Journal of Geographical Information Science on 11th Sept 2016, available at: https://doi.org/10.1080/13658816.2016.1227441.

To cite this article:

Hofer, B., Mäs, S., Brauner, J., & Bernard, L. (2016). Towards a knowledge base to support geoprocessing workflow development. International Journal of Geographical Information Science, 31(4), 694–716. https://doi.org/10.1080/13658816.2016.1227441

1. Introduction

Geoprocessing is a core application of geographic information systems (GIS) and refers to the spatial analysis of data to derive information. The construction of a spatial analysis workflow requires considerable expertise on the user side. The user has to translate the problem at hand into the necessary GIS operations and then prepare the required input data. Support provided by GIS tools during the composition of geoprocessing operations to workflows requires information about the data and operations on the system side. Current developments related to data and processing web services lead to an increased availability of tools for online analyses of spatial data. Leaving the bounds of established GIS upraises questions about the documentation required for the appropriate application of geoprocessing functionalities and for the support for users during workflow development.

Kiehle et al. (2007) identified the semantic descriptions of operations for automation of workflow composition as a frontier in GIScience. Also, a research agenda related to online geoprocessing (Brauner et al. 2009) highlighted the importance of the semantic description of geoprocessing functionalities and the orchestration of operations. Web processing services (WPS), which are a standard for providing geoprocessing functionalities, are described by syntactic elements like the input, output and parameters of an operation, together with the title and description of the operation. These elements have been found to be too limited for the automated discovery of operations (Lutz 2007, OGC 2012). In addition, knowledge about valid combinations of data and operations is not available in a formalized manner, which hinders support during geoprocessing workflow development using web services (Stasch et al. 2014, Cruz et al. 2012, Hong and Huang 2012, Laniak et al. 2013, Qi et al. 2015).

Several approaches of semantic or extended descriptions of geoprocessing operations have been brought forward (Lutz 2007, Fitzner et al. 2011, Stasch et al. 2014, Lemmens 2006) (section 2). The objectives of these approaches vary from improved discovery, over comparison and composition to meaningful application of geoprocessing operations. However, an analysis of the complete geoprocessing workflow development process is not yet available. It is also not clear which elements need to be contained in an operation description to address all the phases of the workflow creation process. These phases comprise user interactions beginning with the search for and comparison of operations, over the chaining of operations and ending with the workflow execution. An analysis of the workflow development process is provided in this paper together with a conceptualization of user and machine interaction during workflow development (section 3). Workflow execution is not analysed in detail herein, but left for future investigations. The questions guiding this work are: *what needs to be known about an operation to use it appropriately in workflow development?* And, *how can the computer support users in their tasks given a knowledge base of geoprocessing operations?*

The identified requirements of operation descriptions determine the conceptualization of a knowledge base. The knowledge base is based on the geooperator approach (Brauner 2015) and extended following the analysis of the workflow development process. Geooperators provide a formalism of operation descriptions. They comprise categories specifying characteristics and possible applications of operations, links between operations, and a description to support the discovery of operations and operator comparability. The overall contribution of this work is a conceptual framework of a knowledge base listing elements of operation descriptions that are required during the workflow composition process (section 4). More specifically, the knowledge base supports discovery of operations across bounds of specific tools, provides feedback during workflow creation and supports the automated discovery of operations for transformations. Examples of how this knowledge base is used during workflow development illustrate the added value of the extended descriptions (section 5) and lead to a discussion and conclusions (section 6).

2. Context and Related Work

The overview of context and related work is separated into two sections: section 2.1 reviews existing model builders and the support they provide to users. Section 2.2 reviews approaches to extended operation descriptions and their core elements.

2.1. Model Builders and User Support

Established workflow development tools such as the ArcGIS ModelBuilder or the GRASS GIS Graphical Modeler provide various functionalities to their users: search for operations by keywords, syntactic checks and validation of appropriate input and parameters for operations, error handling during the execution of the workflow, export of the model etc. Such support for users during workflow development requires information about operations and their input and output parameters. Syntactic characteristics of operations, i.e. the respective input and output data type, are required, and rules that influence the correct functioning of the operations need to be known (e.g., the relationships between input datasets or specific properties of input and output). Some of the support is provided during the composition of tools, but many issues are not detected until runtime. The semantic correctness of the workflow is not evaluated, nor do ModelBuilder or Graphical Modeler include a recommender functionality that would point the user to operations required for closing gaps between operations that the user wants to connect.

In a closed environment like a desktop GIS tool, the variety of possible data formats is well known and the developers can control the rules checked upon execution of the spatial analysis operations. In recent years, workflow development has been widely discussed in the context of loosely-coupled web services (Lemmens et al. 2006, Kiehle 2006, Schäffer et al. 2010, Granell et al. 2013, Stollberg and Zipf 2007). Service-oriented architectures are open systems in regard to present data formats and the level of detail of operation specifications. So far workflow development may be better supported by legacy GIS tools, yet, objectives of scientific geodata infrastructures (Bernard et al. 2013) like workflow sharing (de Jesus et al. 2012) and provenance (Yue et al. 2011) are supported by web-service approaches. In a recent review of intelligent GIServices Yue et al. (2015) highlighted automated service chaining as a research challenge that has the potential of increasing the benefit of web processing services (WPS).

Workflow development tools like GeoBrain (Di 2004, Qiu et al. 2012), GeoPW (Yue et al. 2010) or the CyberGIS tool (Wang 2010, Wang et al. 2012) have been developed in the geoprocessing domain. These existing workflow tools primarily build on syntactic operation descriptions. In WPS an actual geoprocessing function is described as text. The *DescribeProcess* request for a particular operation returns the name, schema, encoding and format of input, output and parameters. In the WPS 2.0 specification the elements related to data can be extended by more detailed descriptions (OGC 2015).

The specification of input and output parameters of an operation provided by syntactic operation descriptions is not sufficient for automated discovery, service chaining, and identification of appropriate services for available data (Yue et al. 2008, Brauner et al. 2009). For example, the operations *union* and *intersect* can have the same input and output elements, namely two vector features as input and one vector feature as output (Fitzner et al. 2011). In case a search operation focuses on syntactic operation signatures only, the two operations cannot be differentiated.

As confirmed by the recent RichWPS project, which developed a framework for web processing services orchestration (Bensmann et al. 2014), the exploitation of extended operation descriptions has not been implemented in a workflow development tool for a general set of spatial operations. The intention of the project was a facilitation of workflow development based on WPSs similar to what has been stated in Dadi and Di (2009). The RichWPS ModelBuilder includes a semantic proxy, which is a registry of available services. However, semantic properties of services are not considered in the workflow development process.

The gap we address in this work is an analysis of the requirements of operation or service descriptions to support workflow development. Although this contribution indicates a possible integration of automation in workflow development it yet remains a vision. We orient our work towards the principles of service-oriented infrastructures, but aim at a specification of descriptions of geoprocessing operations that is independent of the mechanism for providing the functionality. The actual translation of the analyzed conceptual workflow into an executable workflow remains a future challenge. The question of workflow execution has been researched before, so far mostly under the use of the business process execution language (BPEL) (Zhang et al. 2006, Stollberg and Zipf 2007, Kiehle 2006, Schäffer and Foerster 2008, Chen et al. 2010).

2.2. Extended Operation Descriptions

Extended, functional or semantic operation descriptions have been tackled with ontology-based approaches (Lutz 2007, Lutz and Kolas 2007, Fitzner et al. 2011, Lemmens et al. 2006, Cruz et al. 2012), formalizations of data requirements of operations (Stasch et al. 2014) and WPS profiles (Müller 2015). We review this related work with a focus on ontology-based approaches as they contribute to the presented knowledge base.

Ontology-based approaches pose a set of requirements: they generally foresee a geooperation ontology and a data type ontology. The specification of operations in ontologies goes back to ontologies for web services like the web ontology language for services (OWL-S) (W3C 2004) and the Web Service Modeling Ontology (WSMO) (W3C 2005). Core elements of these web service ontologies are pre- and postconditions – i.e., constraints – for the execution of operations. Fitzner et al. (2011) list the following requirements for functional descriptions of geoprocessing operations:

- Input and output types of operations,
- constraints on operations,
- the operation itself,
- dependencies between input and output.

The ontology-based approaches by Lutz (2007) and Fitzner et al. (2011) have focused on the formalization of operation ontologies for operation discovery, focussing on specific operations like distance operations and overlay operations. A transfer of the formalizations to a larger set of GIS operations is missing. Lemmens et al. (2006) developed an ontology-based framework for operation discovery, composition and execution; however, their formalization based on description logics was deemed insufficiently expressive by subsequent work (Lutz 2007, Fitzner et al. 2011).

Cruz et al. (2012) presented an approach to web service composition based on data quality requirements. In their remote sensing-oriented use case, data quality refers to randomness of the spatial distribution, outliers, coverage completeness and pixel reliability. The data quality requirements are expressed with the semantic web rule language (SWRL) and evaluated in test expressions. Their work differs from our approach as they present a solution for the improved composition of their specific application case, while our work aims at identifying requirements throughout the workflow composition process.

The advantage of formalized and ontology-based approaches over textual or semi-formal descriptions as in WPS profiles (Müller 2015) is that the former avoids ambiguity of natural language (Lutz and Kolas 2007). Criticisms of ontologies are that they do not exist in an elaborated form and that they are slow in being processed (Müller et al. 2010). In addition, queries of an ontology need to be formulated in the formalism of the ontology (Lutz 2007, Fitzner et al. 2011). This is a hurdle concerning the use of service discovery by users who are not experienced in formal logics.

Lutz et al. (2007) propose a rule-based approach for service discovery and composition at the conceptual level. A rule thereby describes the link between the input and output of an operation; literals of the required elements are defined in a top-level and a domain ontology. The composition of services follows the principle of overcoming a lack of knowledge through backward chaining (Küster et al. 2005). The service composition focusses on the service outputs. The user specifies the required objective of the workflow and then the inputs of the required service are linked to services that provide them until the chain is complete. A weakness of this approach is that the formal description of workflow objectives is limited. Further, the rules that describe the link between input and output of an operation may become complex for more extensive operations like for example the *cost distance* operation of ArcGIS. Moreover, the rules must be expressive to describe the differences between operations that can have the same input and output.

Table 1 summarizes the discussed related work on ontology-based operation descriptions. The approaches are compared with regard to:

- the intended application purpose: discovery, comparison, composition or execution of operations,
- the data type ontologies: element definitions, level of abstraction and formalization,
- operation ontologies: syntactic elements, pre- and postconditions, categorization of operations by functionality, formalization,
- additional features like attributes of thematic data and the definition of conditional elements and loops.

The aspect of comparing discovered operations is not considered in the reviewed approaches. The comparison of analysis operations across software tools is one of the objectives of the approach by (Brauner 2015) on which our work is based.

Table 1: Overview of ontology-based operation descriptions.

		Lemmens 2006	Lutz 2007	Lutz, Lucchi, et al. (2007)	Fitzner, Hoffmann et al. 2011	Cruz, Monteiro 2012	Stasch, Scheider et al. 2014
objectives	**discovery**	X	X	X	X	_	_
	comparison	_	_	_	_	_	_
	composition	X	(X)	X	_	X	X
	execution	X	_	_	_	X	_
data type ontology / domain ontology	**element definition**	abstract elements in an object/field model based on ISO 19109	abstract elements based on ISO 19100 standards	top level ontology, domain ontology	ISO 19107/2003 Spatial Schema	geodata type, attribute type, data quality ontologies	observation procedures, prediction procedures, aggregation procedures, semantic reference systems.
	abstraction level (conceptual, application, data level)	conceptual, application, data level	conceptual level	conceptual, application level	conceptual level	conceptual, application level	conceptual level
	formalization	description logics	description logics	description logics; domain rules	description logics	description logics	higher order logics
operation ontology	**syntactic elements**	elements from domain ontology	elements from domain ontology	domain rules	elements from domain ontology	elements from geodata type ontology	prediction and aggregation procedures
	pre- and postconditions	following OWL-S	generic specifications of distance operations	_	overlay operations	data quality parameters defined for expressing constraints	operations are functions and data sets are predicates over function domains
	categorization of operations by functionality	15 feature processing operation classes	_	geographic data services and geoprocessing services	_	data type transformation and data property transformation operations	aggregation and prediction procedures
	formalization	description logics	first order logics for pre- and postconditions of operations	schema mapping rules, domain rules specifying the relationship between input and output of operations	conjunctive datalog queries	rule-based description using the semantic web rule language (SWRL)	higher order logics
additional elements	**thematic data attributes**	yes	_	yes	_	yes	yes
	control flow elements	yes	_	_	_	_	_

Recent work investigated the meaningful application of operations to spatial data (Stasch et al. 2014, Härtwig et al. 2014). Härtwig et al. (2014) propose including measurement scales of data to assure the meaningful application of spatial aggregation functions depending on the scale level of the input data. Stasch et al. (2014) provide a formalism of data and operations based on higher order logics for evaluating the meaningful application of spatial aggregation and interpolation operations. The scales of measurement are considered in the proposed knowledge base and it is assumed that the available metadata include this information about the datasets. Despite the difficulty in automatically assigning datasets to the categories of discrete or continuous data, this differentiation is included in the knowledge base. An intermediate solution is a manual tagging of datasets to close the gap in information about data. Improving the automatic tagging and metadata derivation remains a mid- and long-term objective. Work on extended operation descriptions has also been presented in the context of WPS profiles (Müller 2015, OGC 2015). WPS profiles specify the concepts behind implementations of geoprocessing functions with the objective of reducing implementation uncertainty in services from different sources (Müller 2015). The idea is that a user can be sure that a function provided from two different systems calculates the same results when it adheres to the same concept profile. Search would start from a concept and then proceed to specific implementations of this profile. To date no catalogue of WPS profiles exists, which would provide a starting point for such a search. The idea behind WPS profiles is not included in the conceptualization of a knowledge base here; it can provide an extension at a later stage of development.

3. User and System Interaction During Workflow Development

Creating spatial analysis projects based on web services generally requires different sets of user expertise, which frequently leads to user collaborations: domain expertise, knowledge about GIS functionality and technical skills. The development of a spatial analysis workflow and its application can be hindered by a lack of expertise in one of the areas mentioned. Domain specialists do not have their core expertise in spatial analysis, but are interested in deriving information about phenomena taking place in space. Recent work by Kuhn and Ballatore (2015) addresses the difficulty of translating questions about spatial phenomena and data into a GIS vocabulary, which is a major challenge for inexperienced GIS users. GIS analysts have knowledge about the development of spatial analysis workflows in GIS software, and develop individual solutions for tasks at hand. We consider GIS analysts as users, who have knowledge about spatial analysis operations and who have expertise in translating spatially related questions into GIS workflows. This assumption is necessary, as GIS are currently hardly usable without any understanding of their concepts.

The GIS analysts translate their analysis goals into an *abstract workflow concept.* That means that they are aware of the required *core geoprocessing subtasks* (not necessarily the precise operations implemented in the GIS) as well as the requirements on the input data. This abstract workflow concept is refined during the workflow construction process. This refinement means that the user identifies additional subtasks that need to be included in the workflow as the workflow develops and improves. In particular this refinement includes:

- pre-processing or preparing the input data for the core operations,
- operations that connect the core operations (e.g., format conversion or coordinate reference system conversion),
- postprocessing the outputs of the last core operation to prepare the final output data of the workflow.

The operations involved in the refinement of the workflow are referred to as *bridging operations,* as opposed to core operations. Obviously, the categorization of an operation as a core or bridging operation depends on the workflow developer and also on the specific analysis goal of the workflow. In general, bridging operations provide data selections, data type conversion and reference system or other transformations.

To indicate requirements for geoprocessing operation descriptions we envision a workflow development process that supports the users during the following tasks:

- Search for operations, taking into consideration different characteristics of operations and also the different search approaches of users;
- Comparison of operations offering similar functionality;
- Composition of operations through checking consistency between data and operations and assuring the feasibility of the workflow;
- Automated discovery in the context of rather simple and bridging operations like coordinate transformation, data type conversion and similar.

The successful application of a geoprocessing workflow requires spatial data as input. The finding and binding of spatial data has been discussed elsewhere (Janowicz et al. 2010, Jones et al. 2004) and is not covered in this work. We assume here that the spatial data needed for the intended analysis are available and accessible, for example, through spatial data infrastructure. To successfully apply spatial operations to data, the data need to be sufficiently documented with metadata. Metadata are an important element of spatial data infrastructures. Here we assume that metadata are available and cover all elements that are required for checking the validity of certain inputs for operations. However, checking the consistency between data and operations requires corresponding descriptions on both sides.

The herein conceptualized process of workflow construction repeatedly follows a sequence of tasks as shown in Figure 1. We focus on workflow construction as a conceptual composition of operations. The execution of a processing chain, the execution monitoring and the low level interaction with operations (like accessing input data, managing intermediate results, service calls) are not in the foreground of this work. In the following subsections, we outline the main components in the workflow construction process.

3.1 Search and Discovery

As described in the previous section, we assume that a user has an abstract workflow concept in mind that refers to the overall analysis goal. The user is aware of the main workflow steps and core geoprocessing subtasks required, as well as of the requirements on the input data. This abstract workflow concept is incrementally refined during the workflow construction process.

To initiate workflow construction, the user selects one of the core geoprocessing subtasks (for example the *interpolate* or *intersect* operation) from the abstract workflow concept and defines corresponding search criteria. Ideally, these search criteria represent different characteristics of operations and the characteristics of the required input data and of the expected outputs. The search interface should allow the definition of criteria for different operator characteristics to support the differing search approaches of users. Resulting from the discovery step the user receives a list of operations or no result if no operation corresponds to the search criteria. In case the result set is empty, the search criteria have to be adapted.

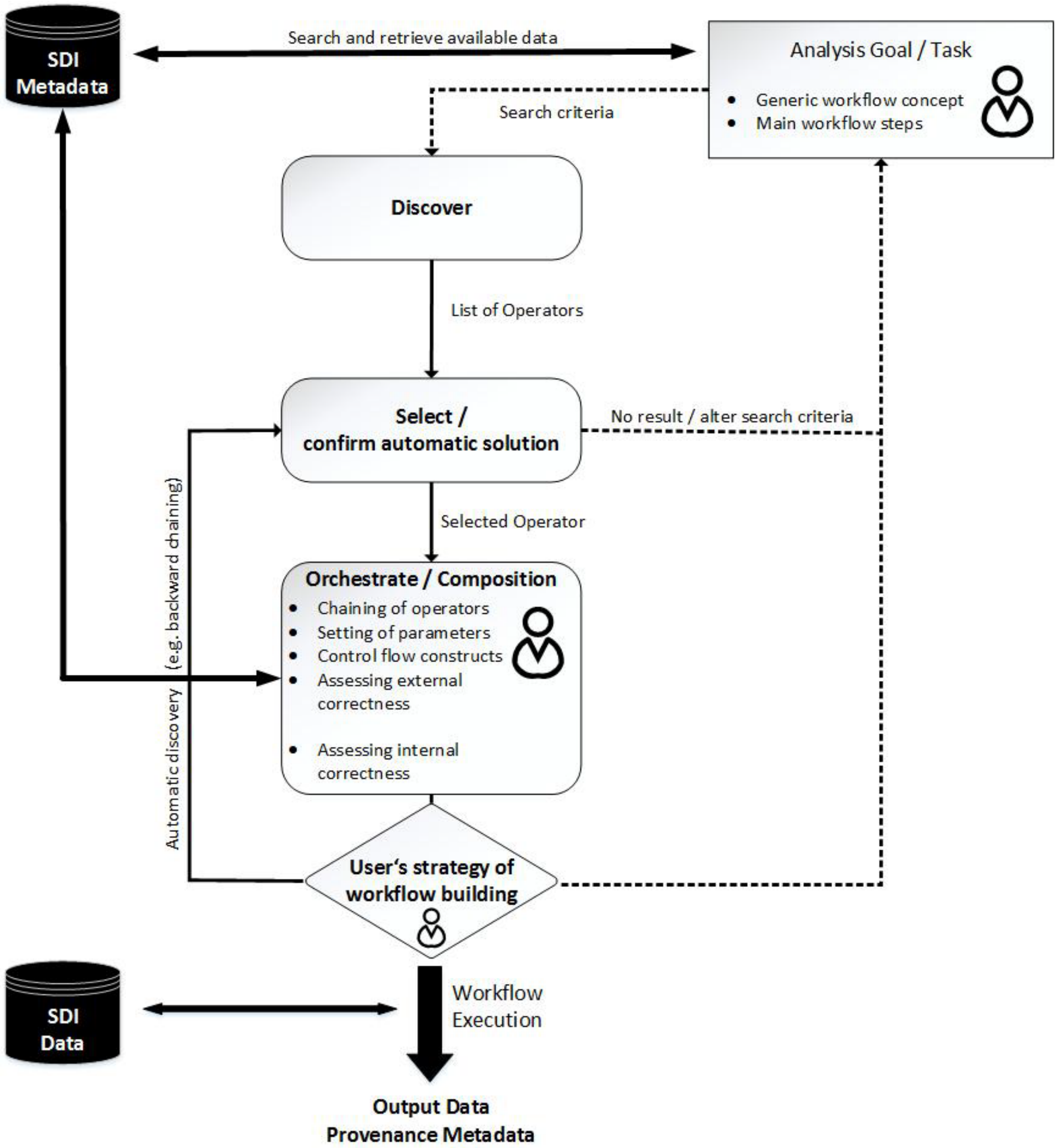


Figure 1: Process of workflow chaining with emphasis of user tasks and system support. Dashed lines indicate user actions outside of the system; the full line is automatic discovery by the system.

3.2 Selection and Composition

Having a list of operations as result requires the selection of a specific operation to be added to the workflow. These lists usually contain operations offering similar functionalities, e.g. implementations from different GIS software. To support the selection process by the user the operator descriptions should link to similar operators, specify differences, and describe use cases.

The composition of operations is carried out by the user. Parameters are set by the user as required by the operations. The workflow composition tool should support the validation of the logical consistency of the workflow to assess the internal correctness. The internal correctness assures that all requirements and consistency constraints of the operations and their in- and outputs are fulfilled and that the workflow is executable according to all specified rules. This means that the tool prevents the user from connecting operations in case prerequisites of the connected operation defined by operator descriptions are not fulfilled. We consider this internal validation of the workflow as part of the error handling mechanisms of the workflow composition tool before the workflow is executed. Here we assume that the user evaluates the external correctness, such that the user judges whether the workflow produces meaningful results for the anticipated objective. The user is free to test parts of the workflow while approaching the final goal, and can add and remove operations along the way.

After adding an operation, the process of discovery, selection and composition starts again unless the workflow is complete. Each time the user starts the next iteration, they are at a decision point. For tasks with a well-defined scope and in- and outputs such as the transformation of coordinate reference systems or data types (bridging operations) the workflow composition tool should offer an *automated discovery* function. This function can be initiated to automatically discover operations that could provide the missing link between two operations already present in the workflow. Reasoning methods such as backward chaining (Küster et al. 2005) have already been proposed by Lutz et al. (2007) for the automated chaining of web services. To overcome the limitations of their approach, as discussed in section 2.2, we suggest using the automated discovery only for the *bridging operations*, such as data type conversions or transformations, for which the differences between input and output of an operation can be described, and where reasoning processes can automatically derive matching operations.

To include the results in the workflow the user has to select an operation or confirm the automatic solution if only one operation has been found. If all required bridging operations are found, the user chooses the next of the core geoprocessing subtasks from their abstract workflow concept and reiterates the composition process.

4. Operation Descriptions for Workflow Development – The Knowledge Base

Our conceptualized workflow development process comprises a set of knowledge base requirements. The knowledge base needs to provide descriptions of operations that support their discovery and selection by the user, composition including feedback concerning the match between operations linked and automated discovery for bridging operations. This section presents the knowledge base that contains the required elements of operation descriptions. The knowledge base, founded on geooperators (Brauner 2015), considers elements of ontology-based approaches (Fitzner et al. 2011, Lutz 2007) to cover the entire workflow development process.

4.1. Geooperator Definitions

The main objective of the work on geooperators by Brauner (2015) is to support the discovery and comparison of geoprocessing operations by a human user. In the context of SDI the discovery and comparison of geoprocessing services is an open issue as no registries or catalogues for such services exist. Approaching such a catalogue requires a general classification of geoprocessing functionality. Various categorizations of geoprocessing functionality have been proposed, e.g. (Albrecht 1996, ISO 2005a). However, as this functionality is diverse, the categorization needs to be flexible enough to capture different perspectives on geoprocessing functionality and it needs to provide defined concepts with a common set of attributes of operations that can be compared (Brauner 2015).

In Brauner's work, geoprocessing operations are referred to as geooperators (Brauner 2015); a geooperator "*is a distinct, well defined and usually implemented piece of software serving a particular purpose for geospatial analysis or transformation*" (Brauner 2015, p.38). This focus on implemented functionality has the advantage that the users' knowledge of specific tools is considered and that user communities exist that can eventually contribute to the description of functionality.

The specification of geooperators comprises three main components: elements describing the operation together with its inputs, outputs and parameters; a categorization based on geooperator categories and matches with related geooperators from the same or other software products. Figure 2 provides an overview of this description.

Geooperator
id: String [1] label: String [1] implementedIn: String [0…1] definition: String [1] inputs: Operand [1…*] outputs: Operand [1…*] parameters: Parameter [0…*] categories: GeooperatorCategory [0…*] closeMatch: Geooperator [0…*] narrowMatch: Geooperator [0…*] broadMatch: Geooperator [0…*] contextRelated: Geooperator [0…*] references: String [0…1] comments: String [0…*]

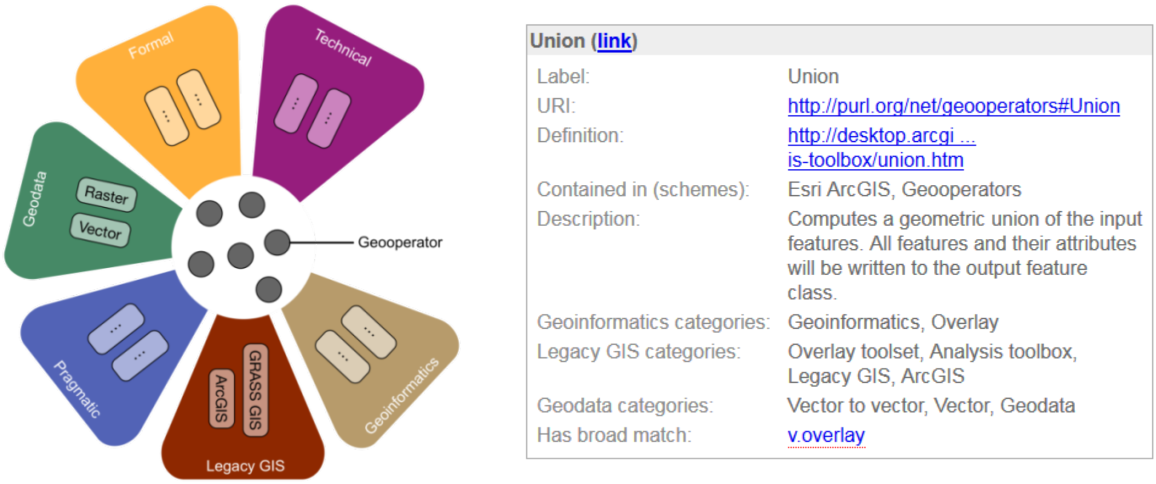
Figure 2: Elements of geooperator descriptions.

Geooperator *categories* specify the properties of geooperators and represent different views on geospatial operations as documented in literature. The categories proposed by Brauner are shown in Figure 3; the presented framework can be extended by additional categories, if they are considered useful for a further description of geooperators. The following categories are currently differentiated in geooperator descriptions:

- **Geodata Category:** The geodata category differentiates geodata models (raster, vector, networks etc.) used in geooperators. The geodata category is hierarchically structured into raster and vector operations, which are further divided into vector to vector, vector to raster, vector to attribute table operations etc.
- **Legacy GIS Category:** The legacy GIS category specifies in which GIS software the geooperator is implemented and to which toolbox it belongs.
- **Geoinformatics Category:** The geoinformatics category provides the option of relating a geooperator to a GIScience concept like Map Algebra, universal GIS operations (Albrecht 1996) or Egenhofer relations. In addition, this category can classify a geooperator as an analysis operation and into subcategories like overlay, interpolation, aggregation operations etc.
- **Pragmatic Category:** The pragmatic category consists of a variety of application fields and tasks for which a geooperator can be used.

- **Technical Category:** The technical category describes runtime environments and execution strategies on a conceptual level, and details runtime and implementation aspects.
- **Formal Category:** The formal category captures mathematical characteristics of geooperators such as arity, granularity, and reversibility.

(a) **(b)**



Label:	Union
URI:	http://purl.org/net/geooperators#Union
Definition:	http://desktop.arcgi ... is-toolbox/union.htm
Contained in (schemes):	Esri ArcGIS, Geooperators
Description:	Computes a geometric union of the input features. All features and their attributes will be written to the output feature class.
Geoinformatics categories:	Geoinformatics, Overlay
Legacy GIS categories:	Overlay toolset, Analysis toolbox, Legacy GIS, ArcGIS
Geodata categories:	Vector to vector, Vector, Geodata
Has broad match:	v.overlay

Figure 3: (a) Geooperator categories to provide keywords describing geoprocessing operations and (b) an example description of the ArcGIS geooperator *Union*.

4.2. Formalization of Geooperators

Brauner's approach provides a human and machine-usable formalism for representing descriptions of geooperators and links between functionality from different systems – the geooperator thesaurus (Brauner 2015). The Simple Knowledge Organisation System (SKOS) is used for capturing links between geooperators in a machine-readable form (Miles and Bechhofer 2009). SKOS is used in semantic web contexts for representing knowledge bases in RDF format (Resource Description Framework). The geooperator thesaurus represents the specification of geooperators through the defined categories and the relations among them[1]. Figure 4 shows a concept map of geooperators in black and categories describing geooperators in the colors brown (Legacy GIS category), green (Geodata category), and blue (Pragmatic category).

The geooperator perspectives are expressed as SKOS concept schemes, which are instantiated by SKOS concepts for representing geooperator categories (Brauner 2015). SKOS semantic relations link the concepts either hierarchically or associatively. For example, the category *Geodata* is instantiated by the concepts *raster*, *vector*, *raster to vector*, *vector to raster* and *vector to vector*. The semantic relation *broaderTransitive* represents the hierarchical structure between concepts; a *relatedMatch* links geooperators and categories; a *closeMatch* is used to link almost equivalent geooperators from different GIS; *narrowMatch* and *broadMatch* link geooperators from different GIS with less respectively more functionality. Figure 4 contains an example for a *narrowMatch*: the ArcGIS operation *Buffer* is in a *narrowMatch* relation with the GRASS operations *v.parallel* and *v.buffer*. This *narrowMatch* relation means that the GRASS operation *v.buffer* and *v.parallel* together cover the functionality of the ArcGIS *Buffer* operation.

[1] https://github.com/GeoinformationSystems/GeooperatorBrowser

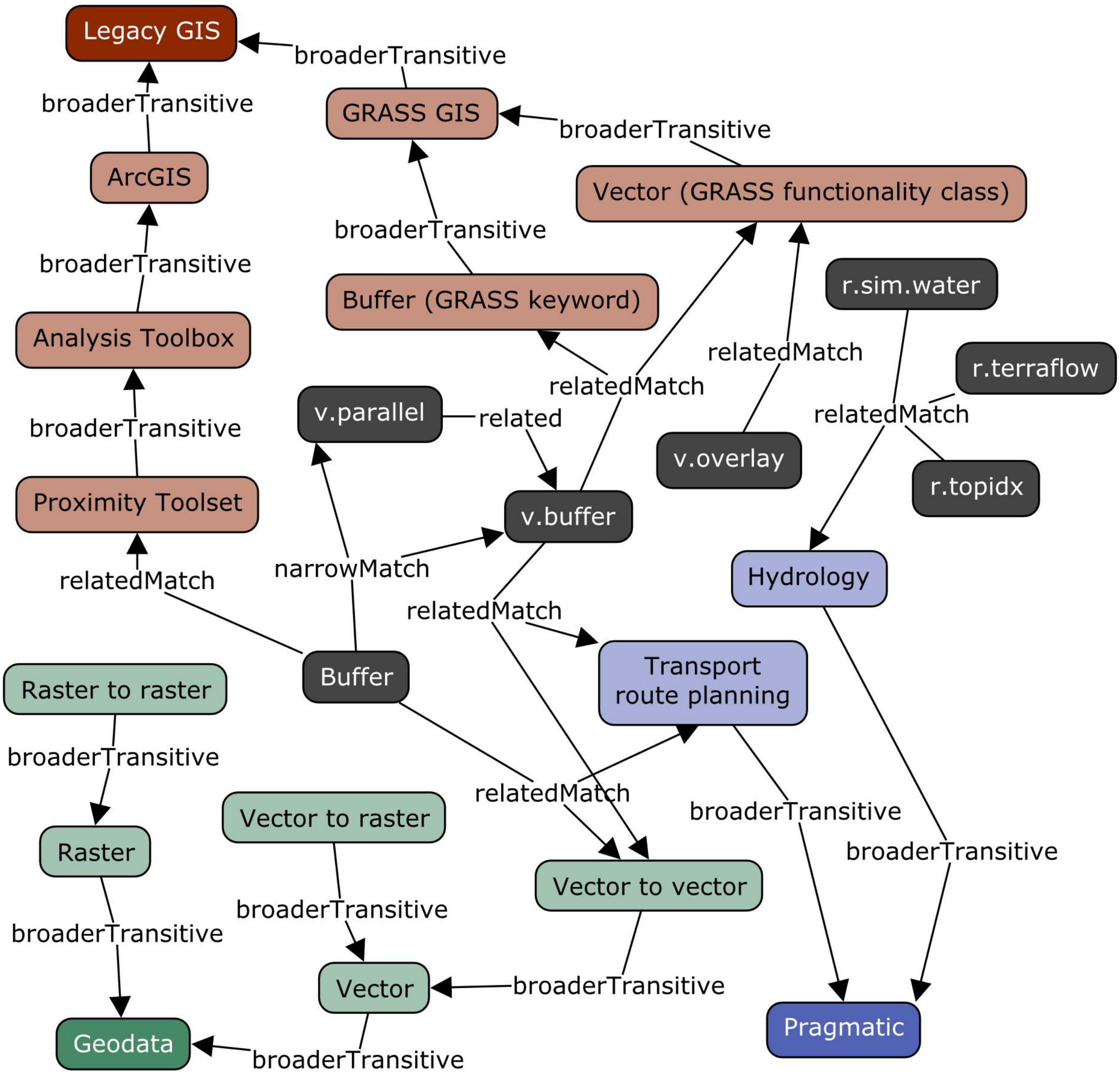


Figure 4: Geooperator concept map (source: (Brauner 2015)).

4.3. Proposed Extensions of the Geooperator Approach

The geooperator approach was developed with a focus on the tasks of discovery and comparison. In reference to the summary presented in Table 1, the tasks that remain to be addressed are composition and execution. As a discussion of execution exceeds the scope of this work, the extension of the geooperator approach concerns the task of composition. Composition of operations in the envisioned workflow requires an evaluation of the internal correctness of the workflow. This internal correctness refers to the correspondence between inputs provided to an operation and the inputs required by operations. In addition, the workflow development process foresees automated discovery of operations for bridging operations. The conceptualized knowledge base needs to provide sufficient information about operations to satisfy these two points.

The geooperator approach categorized operations based on geooperator categories. The approach currently does not include specifications of the input and output data types, and of the pre- and postconditions of the operations. Therefore, we introduce an extension of the geooperator approach with these elements of ontology-based operation descriptions.

The extension requires the description of operands, which are inputs and outputs of operations, non-spatial parameters and constraints on operations. In the following, we sketch a spatial data type specification that provides the basis for the named elements of functional operation descriptions. The specification mirrors the abstract ISO geographic feature model and is guided by previous work by Fitzner et al. (2011) and Lutz (2007). Assuming that the spatial data type description is available in the web ontology language (OWL), the OWL classes describing operands can be linked with the geooperator SKOS thesaurus (Figure 4).

This spatial data type specification is intended to specify operands of geooperators and includes information about constraints on operations. It specifies type signatures on a conceptual level. Compared to the ISO feature models or the core concepts of spatial information (Kuhn 2012) this specification is not meant to provide a general ontology for geographic features but is driven by available implementations of geooperators: We only differentiate between representations of spatial objects as vector data (ISO 2003) and raster data (ISO 2005b). We do not yet refer to data formats that implement these types (formats come into play for the execution of a geoprocessing workflow).

Constraints on spatial operations from previous work show that they concern geometric, topologic, thematic and temporal dimensions of data (Härtwig et al. 2014, Stasch et al. 2014, Brauner 2015). Constraints are required for detecting errors during workflow composition, which need an adaptation of the workflow, such as its extension for a data type conversion operation. In addition, the formal documentation of functional elements is required to provide recommendations in the context of automated discovery. The following pieces of information are initially included for specifying constraints on operations:

- geometric properties of a spatial object represented as raster or vector data,
- coordinate reference system (CRS),
- thematic properties represented as attributes with a description (which could link to a reference system, classification or attribute catalogue),
- scales of measurements (ordinal, nominal, interval, ratio), unit of measurements and data type (integer, floating point etc.) of a thematic attribute,
- differentiation of raster data about grid point based or cell based representation.

Figure 5 shows the current draft of the spatial data type specification. A detailed analysis of the constraints of geooperators is foreseen in future work, and may lead to the inclusion of additional elements in this specification. Further, 3D and 4D data are currently not considered in the model.

The geooperator descriptions are extended for elements describing input, output, parameters and constraints based on the spatial data type specification. Constraints on operations can be described on different levels of detail. A clip operation, for example, requires two vector features as input; yet there are rules concerning the possible combinations of input features: points can be clipped with points, lines, or polygons. A polygon, however, can only be clipped with a polygon. This can be expressed in the ontology with the object property *composes* that is transitive and reflexive. Points compose lines, lines compose polygons and because of the transitive and reflexive nature of the object property, the geometry types are interlinked.

If the user is to receive feedback during composition of operations, such rules have to be documented in the geooperator description. The constraints of operations that are represented in the ontology provide details about the required check routines, which can be implemented in a workflow composition tool.

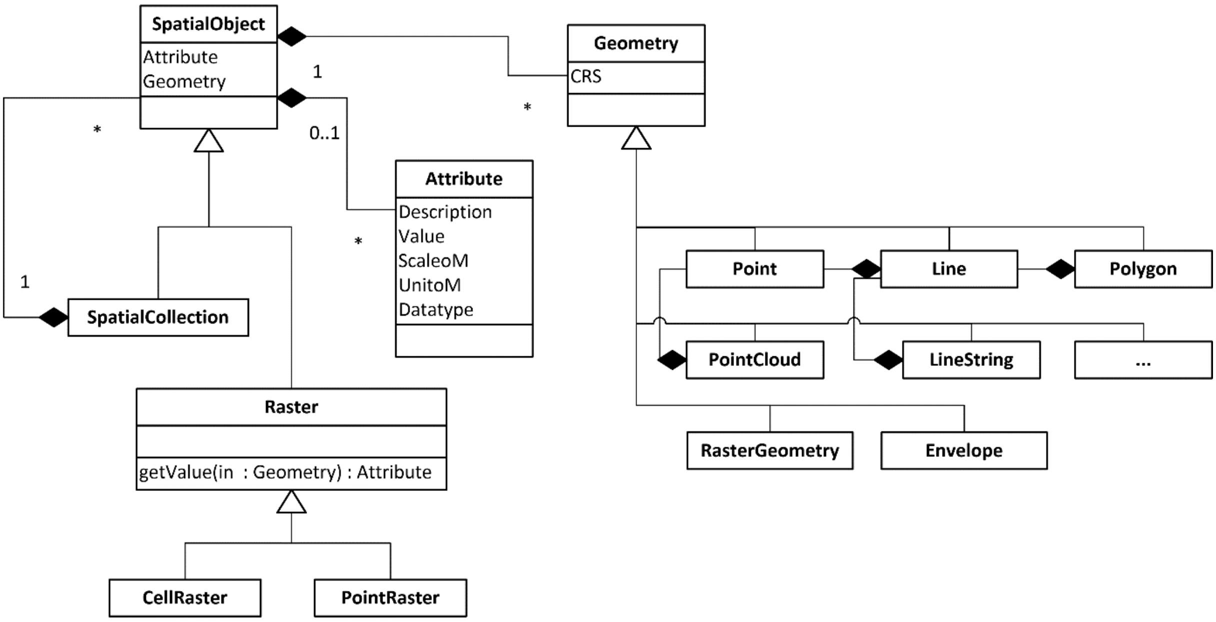


Figure 5: Feature model of the proposed data type ontology.

Functional operation descriptions in previous work did not consider parameters of operations. Nevertheless, these parameters are important as these settings can influence the result. For example, it is possible to request the output of an operation in a specific coordinate reference system, which triggers a transformation to this CRS during the execution of the operation.

We propose explicitly marking geooperators that provide bridging operations: data type conversion and coordinate reference system transformations. This description element can be included in the geoinformatics perspective of the geooperator descriptions and can separate bridging operations from analysis operations. The choice of the required bridging operation depends on the results of a check routine: if the CRS of the input does not match the one of the operation, this is the criterion used in the discovery of operations; if the data types do not match, they are considered when querying the ontology. The pieces of information required for differentiating the changes in data type or data parameter are provided by the data type specification.

5. Demonstration of Support Through the Knowledge Base

This section illustrates how the conceptualized knowledge base can support the user in discovery, selection and composition of geoprocessing operations once it is implemented. The support for the user includes a validation of the internal correctness and an automated discovery of bridging operations.

5.1. Discovery and Selection of Geooperators

The users in our conceptualized workflow development process have an idea about the required tool. They can specify characteristics of geooperators or provide keywords for the search. Therefore, the discovery is not automated but in the hands of the user. The discovery of operations is implemented in the geooperator browser[2] (Brauner 2015), which is part of the Geoprocessing Appstore[3] (Henzen et al. 2015). The geooperator categories support faceted browsing, which allows

[2] http://purl.org/net/geooperators
[3] http://apps1.glues.geo.tu-dresden.de:8080/appstore/catalog/main/home.page

the search results to be narrowed down by selecting characteristics of the operations. Matches with related geooperators broaden search results. The links between geooperators also increase the potential result-set in keyword-based searches when keywords are matched to functionality descriptions of different systems.

Upon retrieval of the search results, the user can compare and subsequently select operations. The definition of the geooperator and additional elements of its description support the selection. An example of a search result is displayed in Figure 6. Here, the user specified that the result should contain vector operations that belong to the ArcGIS proximity toolset. For each of the returned operations, the user can browse related geooperators.

The geooperator browser is a new approach to the discovery and selection of spatial processing functionality across software tools and technical realizations. Currently, the geooperator browser contains 40 geooperator descriptions. The geooperator approach and possibilities of extending the geooperator descriptions through community involvement have been discussed at a workshop (Hofer et al. 2015).

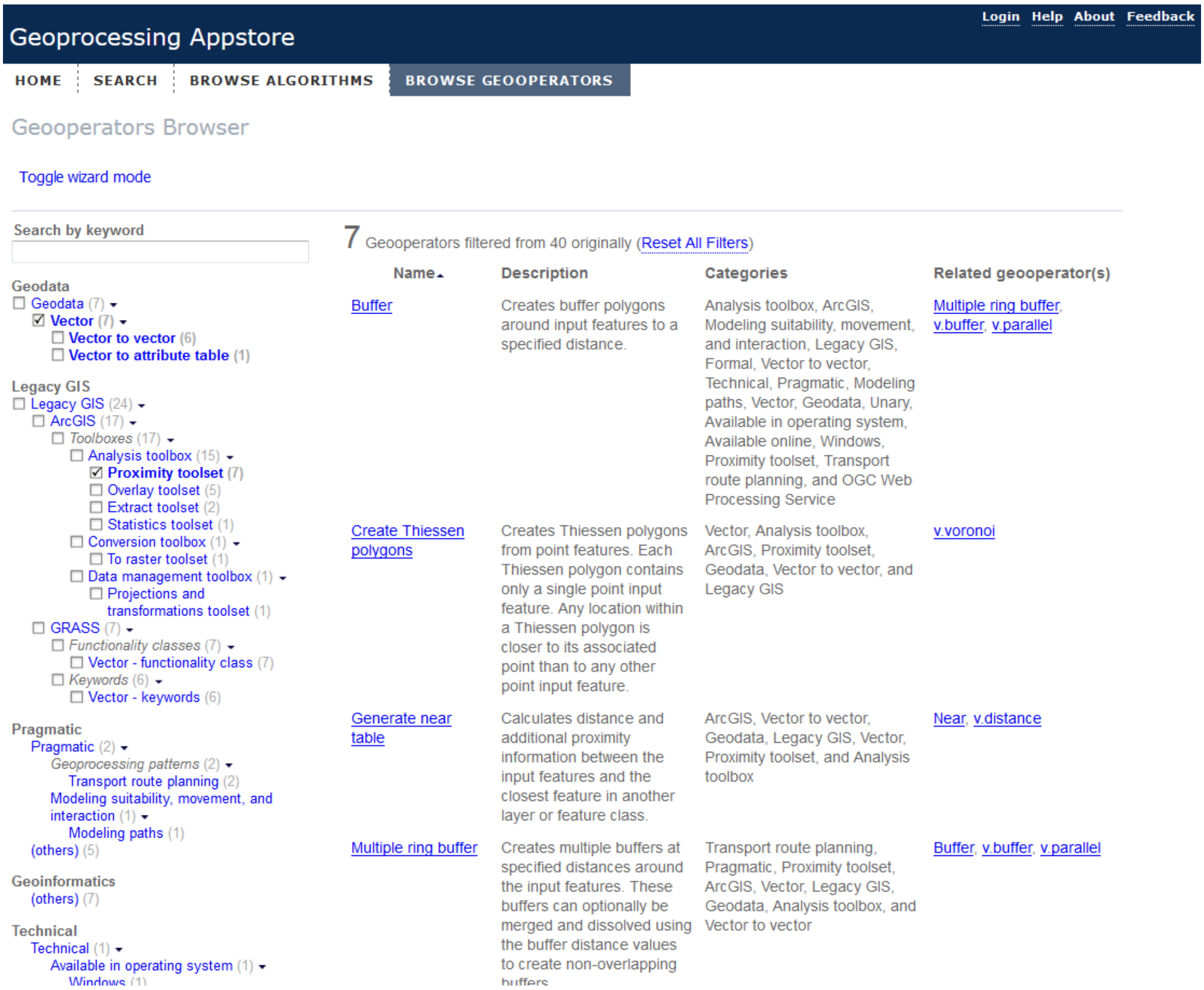


Figure 6: Results for a search of vector operations that belong to the proximity toolset.

5.2. Composition of Operations

Users select the required operations according to their abstract workflow concept and compose them into a sequence. The errors that occur upon composition generally refer to the specification of the input of operations: mismatching data type, wrong CRS, wrong number of inputs, violation of additional constraints etc. The geooperator description includes a specification of these constraints, which provides information about what has to be implemented in check routines for the specific operations. These check routines can provide feedback during the composition of operations and not only during execution as in the case of direct interaction with web services or specific GIS. Feedback provided during composition allows the users to adapt the workflow before executing it and is preferable to sending erroneous requests to web services (Cruz et al. 2012).

Given that the data type ontology sketched in section 4 is fully integrated into the geooperator descriptions, the system can provide feedback to the user during the composition of operations in the following exemplary situations. The rules stated are formulated with the semantic web rule language (SWRL) and implemented in Protégé OWL.

- Coordinate reference system of the data does not correspond with the requirements of the operation: e.g. it can be evaluated for the geooperator *Near*, which works either with geodesic or planar distances, whether the given input fulfils the requirements of the chosen method or not. The expression of the rule tests, whether the spatial reference system (*srs*) of the inputfeatures (*nearfeatures*, *vector3035*) is EP*SG:4326*, which is used as a representative of a geographic coordinate system in the example. The *nearfeatures* fulfil the condition whereas the *vector3035,* having the coordinate references system with EPSG:3035, does not.

  Geooperator(Near), hasValue(NearMethod, "geodesic"), srs(nearfeatures, "EPSG:4326") -> Precondition(nearfeatures)

  Geooperator(Near), hasValue(NearMethod, "geodesic"), srs(vector3035, "EPSG:4326") -> Precondition(vector3035)

- Unsuitable geometry type of input: as mentioned before, polygons can only be clipped with polygons. If the clip feature for a polygon dataset is a point dataset, this mismatch can be detected by the system. The following rule says that the input features *clippoint* only fulfil the precondition of the *Clip* geooperator, if *clippoly* (the feature to be clipped) composes *clippoint*. As this is not the case, *clippoint* does not fulfil the precondition.

  Geooperator(Clip), composes(clippoly, clippoint), input(Clip, clippoint), input(Clip, clippoly) -> Precondition(clippoint)

- Scale of measurement of data does not comply with the operations: given the scale of measurement of input data and comparing it to constraints on the operation, would prevent the user from applying an average operation to ordinal data (the data need to be on an interval- or ratio-scale). For the operation *Mean* the input values are *numericValues* with a level of measurement (*levelMeasurement*) of *interval* and therefore the input supplied fulfils the precondition.

  Geooperator(Mean), input(Mean, numericValues), levelMeasurement(numericValues, "interval") -> Precondition(numericValues)

- Types of raster: the raster to polyline operation from ArcGIS only works when the raster field assigned to the polyline is an integer field. Expressed in a rule this input condition of the *Raster_to_polyline* geooperator needs to have an *integer* value in the specified *Rasterfield*.

Geooperator(Raster_to_polyline), input(Raster_to_polyline, Rasterfield), hasValue(Rasterfield, "integer") -> Precondition(Rasterfield)

5.3. Automated Discovery of Operations

During workflow composition the need to transform the coordinate reference system or the data type of the input dataset can be identified. This need can be detected by the user or by a check routine. As described in section 3.2, we envision an automated discovery of these bridging operations that utilizes the backward chaining approach presented in Lutz and Kolas (2007).

In the backward chaining approach, the operation that shall finally be executed is set and operations that prepare the required input are iteratively added to the workflow. In the case of bridging operations, the goal of the operation is to have, e.g., a dataset in a specific CRS, which requires an operation that transforms the available dataset into the correct CRS. The transformation of the CRS is the goal of the operation that shall be discovered. In the geooperator description we include a tag of operations that provide bridging functionalities. The specific operation can be selected from this set of operations, by querying the CRS parameter of input and output in the ontology. In case of a data type conversion, the data type of the input and output are different as well. In addition, the input needs to be subsumed by the data type of the available dataset and the output needs to be subsumed by the data type of the input of the targeted operation. Following the geooperator approach, the automated discovery can identify a *feature to raster* operation, but also related operations like a *polyline to raster* operation. In case a CRS and a data type conversion are required, these two operations would be discovered one after the other and put in a sequence in the workflow. The parametrization of the discovered operations lies in the hands of the users.

The automated discovery of operations is a step towards the recommendation of suitable operations for a specific task, which is a function that existing tools do not provide. An extension of this recommendation function requires the representation of additional knowledge in operation descriptions. For example, in case a CRS and data type transformation are required, it would be sensible to transform the coordinate reference system of a vector dataset rather than a raster dataset as the projection of raster data requires resampling. The recommended order of operations could be considered upon composition if it is represented in the knowledge base or heuristics are applied accordingly.

6. Conclusions and Future Work

The vision in the context of web processing services is the automation of spatial analyses. Yue et al. (2015) state that automation bears the added value of online geoprocessing. Automation is hindered by a lack of registries of geoprocessing services and a lack of formalized GIS expert knowledge to date. The knowledge base that we present in this paper is based upon a requirement analysis of the geospatial workflow development process. The knowledge base provides descriptions of spatial operations that support discovery, selection, and composition. Following the geooperator approach, discovery and selection of operations is done by the users and supported by the geooperator browser. Concerning composition, a specific achievement of the conceptualized knowledge base is that the users receive feedback during this phase in case an adaptation of the workflow is necessary. In well-defined situations the knowledge base can provide recommendations of required operations, which is a step towards automated discovery of operations.

The provision of feedback during the composition of operations and the conceptualization of automated discovery required an extension of the geooperator approach. The geooperator descriptions are extended for details on input, output, and parameters of operations together with their constraints; these elements support the evaluation of the internal correctness of combinations of operations. The external verification of the workflow, i.e. assessing whether the workflow generates meaningful results and conforms to analysis goals, remains in the hands of the users.

The work shows that the supportive feedback to users during workflow composition requires more than the available syntactic elements of operation descriptions. The vision of a workflow development tool that can evaluate the correct use of functionality from various tools and services requires the description of operations and their constraints.

A problem of all approaches that use extended operation description is the above mentioned lack of registries of services that include the required elements. The extension of the set of descriptions that are currently available could be achieved with community involvement. In addition, we assumed that the metadata of data specify the level of measurement of attributes and whether the observed phenomenon is continuous or discrete. These elements would be extensions of existing metadata descriptions as has been requested before (Bernard and Krüger 2000, Härtwig et al. 2014, Stasch et al. 2014).

An extensive analysis of constraints on operations, the development of a workflow composition tool that includes the conceptualized knowledge base and the transfer of the conceptual workflow into an executable workflow are necessary future work. A detailed evaluation of preconditions and postconditions of operations is currently performed. This analysis looks at a large set of spatial analysis operation with the objective to denote a set of recurring elements in constraints. The results of the analysis will complement the sketched data type ontology. It is a trivial observation that more detail in operation descriptions allows more specific feedback to users. However, the objective is not to replicate error handling that is implemented in the tool itself but rather to provide feedback in case an adaptation of the workflow is necessary and to provide recommendations for operations.

The implementation of a workflow development tool will follow. The RichWPS framework (Bensmann et al. 2014) could serve as a basis for the implementation and could be extended with mechanisms to evaluate the operation descriptions proposed here. The RichWPS framework works with web processing services, which is not a contradiction to the geooperator approach. Brauner (2015) demonstrated that geooperator descriptions can be injected as semantic annotations in metadata of web processing service descriptions. In the long run, the knowledge base we conceptualized could be part of SDIs as metadata of data are today.

Acknowledgements

We thank the anonymous reviewers and the editor for their suggestions and comments that helped to improve the quality of the paper. Barbara Hofer was partly funded by a DRESDEN Junior Fellowship by Technische Universität Dresden.

References

Albrecht, J., 1996. Universal GIS-Operations - A Task-Oriented Systematization of Data Structure-Independent GIS Functionality Leading Towards a Geographic Modelling Language. *ISPA - Mitteilungen.* 1-94.

Bensmann, F.*, et al.* 2014. The RichWPS Environment for Orchestration. *ISPRS International Journal of Geo-Information,* 3(4), 1334.

Bernard, L. and Krüger, T. 2000. Integration of GIS and Spatio-temporal Simulation Models: Interoperable Components for Different Simulation Strategies. *Transactions in GIS,* 4(3), 197-215.

Bernard, L., *et al.* 2013. Scientific geodata infrastructures: challenges, approaches and directions. *International Journal of Digital Earth*.

Brauner, J., 2015. *Formalizations for geooperators - geoprocessing in Spatial Data Infrastructures.* PhD Thesis (Dr.). Technische Universität Dresden.

Brauner, J., *et al.*, 2009. Towards a Research Agenda for Geoprocessing Services. *In:* Haunert, J., Kieler, B. and Milde, J. eds. *12th AGILE International Conference on Geographic Information Science.* Leibniz University Hannover, Germany.

Chen, N., *et al.* 2010. Geo-processing workflow driven wildfire hot pixel detection under sensor web environment. *Computers & Geosciences,* 36(3), 362-372.

Cruz, S. A. B., Monteiro, A. M. V. and Santos, R. 2012. Automated geospatial Web Services composition based on geodata quality requirements. *Computers & Geosciences,* 47(0), 60-74.

Dadi, U. and Di, L., 2009. Creating web service interfaces and scientific workflows using command line tools: A GRASS example.

de Jesus, J., *et al.* 2012. WPS orchestration using the Taverna workbench: The eScience approach. *Computers & Geosciences,* 47(0), 75 - 86.

Di, L., 2004. GeoBrain-A Web Services based Geospatial Knowledge Building System. *NASA Earth Science Technology Conference.* 22--24.

Fitzner, D., Hoffmann, J. and Klien, E. 2011. Functional description of geoprocessing services as conjunctive datalog queries. *Geoinformatica,* 15(1), 191-221.

Granell, C., Schade, S. and Ostländer, N. 2013. Seeing the forest through the trees: A review of integrated environmental modelling tools. *Computers, Environment and Urban Systems,* 41, 136-150.

Härtwig, M., Müller, M. and Bernard, L. 2014. A Generic Web Service for Ad-hoc Statistical Spatio-Temporal Aggregation. *Transactions in GIS,* 18(4), 527-543.

Henzen, C., *et al.*, Geoprocessing Appstore. ed. *The 18th AGILE International Conference on Geographic Information Science*, 2015 Lisbon, Portugal.

Hofer, B., *et al.* 2015. Descriptions of Spatial Operations – Recent Approaches and Community Feedback. *International Journal of Spatial Data Infrastructures Research,* 10, 124-137.

Hong, J.-H. and Huang, M.-L., Interoperable GIS Operations: A Quality-Aware Perspective. ed. *Multidisciplinary Research on Geographical Information in Europe and Beyond, AGILE'2012 International Conference on Geographic Information Science*, April, 24-27 2012 2012 Avignon.

ISO, 2003. ISO 19107: Geographic information - Spatial schema.

ISO, 2005a. ISO 19119: International Standard Geographic Information Services. Oslo, Norway: ISO/TC 211 Secretariat.

ISO, 2005b. ISO 19123: Geographic information - Schema for coverage geometry and functions.

Janowicz, K., *et al.* 2010. Semantic Enablement for Spatial Data Infrastructures. *Transactions in GIS,* 14(2), 111-129.

Jones, C. B., *et al.*, 2004. The SPIRIT Spatial Search Engine: Architecture, Ontologies and Spatial Indexing. *In:* Egenhofer, M. J., Freksa, C. and Miller, H. J. eds. *Geographic Information Science: Third International Conference, GIScience 2004, Adelphi, MD, USA, October 20-23, 2004. Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg, 125-139.

Kiehle, C. 2006. Business logic for geoprocessing of distributed geodata. *Computers and Geosciences,* 32(10), 1746-1757.

Kiehle, C., Greve, K. and Heier, C. 2007. Requirements for Next Generation Spatial Data Infrastructures-Standardized Web Based Geoprocessing and Web Service Orchestration. *Transactions in GIS,* 11(6), 819-834.

Kuhn, W. 2012. Core concepts of spatial information for transdisciplinary research. *International Journal of Geographical Information Science,* 26(12), 2267-2276.

Kuhn, W. and Ballatore, A., 2015. Designing a Language for Spatial Computing. *In:* Bacao, F., Santos, M. Y. and Painho, M. eds. *AGILE 2015.* Springer International Publishing, 309-326.

Küster, U., Stern, M. and König-Ries, B., A Classification of Issues and Approaches in Automatic Service Composition. ed. *First International Workshop on Engineering Service Compositions (WESC05) at the Third International Conference on Service Oriented Computing (ICSOC05)*, 2005.

Laniak, G. F.*, et al.* 2013. Integrated environmental modeling: A vision and roadmap for the future. *Environmental Modelling and Software,* 39, 3-23.

Lemmens, R., 2006. *Semantic interoperability of distributed geo-services.* (PhD). ITC.

Lemmens, R.*, et al.* 2006. Integrating Semantic and Syntactic Descriptions to Chain Geographic Services. *Internet Computing, IEEE,* 10(5), 42-52.

Lutz, M. 2007. Ontology-Based descriptions for semantic discovery and composition of geoprocessing services. *Geoinformatica,* 11(1), 1-36.

Lutz, M. and Kolas, D. 2007. Rule-based discovery in spatial data infrastructure. *Transactions in GIS,* 11(3), 317-336.

Lutz, M.*, et al.*, A rule-based description framework for the composition of geographic information services. ed. *Proceedings of the 2nd international conference on GeoSpatial semantics*, 2007 Mexico City, Mexico, 114-127.

Miles, A. and Bechhofer, S., 2009. *SKOS Simple Knowledge Organization System Reference. Recommendation.* W3C.

Müller, M. 2015. Hierarchical profiling of geoprocessing services. *Computers & Geosciences,* 82(0), 68-77.

Müller, M., Bernard, L. and Brauner, J. 2010. Moving code in spatial data infrastructures - web service based deployment of geoprocessing algorithms. *Transactions in GIS,* 14(SUPPL. 1), 101-118.

OGC, 2012. Semantic Annotations in OGC Standards. *OGC Best Practices.* Open Geospatial Consortium, 59.

OGC, 2015. OGC WPS 2.0 Interface Standard. *OGC document 14-065.*

Qi, K.*, et al.* 2015. An Extension Mechanism to Verify, Constrain and Enhance Geoprocessing Workflows Invocation. *Transactions in GIS*, n/a-n/a.

Qiu, F.*, et al.* 2012. GWASS: GRASS web application software system based on the GeoBrain web service. *Computers & Geosciences,* 47(0), 143-150.

Schäffer, B., Baranski, B. and Foerster, T., 2010. Towards Spatial Data Infrastructures in the Clouds. *In:* Painho, M., Santos, M. Y. and Pundt, H. eds. *Geospatial Thinking.* Springer Berlin Heidelberg, 399-418.

Schäffer, B. and Foerster, T. 2008. A client for distributed geo-processing and workflow design. *Journal of Location Based Services,* 2(3), 194-210.

Stasch, C.*, et al.* 2014. Meaningful spatial prediction and aggregation. *Enviornmental Modelling and Software,* 51, 149-165.

Stollberg, B. and Zipf, A. 2007. OGC Web processing service interface for Web service orchestration aggregating geo-processing services in a bomb threat scenario. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),* 4857 LNCS, 239-251.

W3C, 2004. OWL Web Ontology Language for Services (OWL-S) - Submission.

W3C, 2005. Web Service Modeling Ontology (WSMO) - Submission.

Wang, J.*, et al.* 2012. Using service-based GIS to support earthquake research and disaster response. *Computing in Science and Engineering,* 14(5), 21-30.

Wang, S. 2010. A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Annals of the Association of American Geographers,* 100(3), 535-557.

Yue, P.*, et al.* 2015. Towards intelligent GIServices. *Earth Science Informatics,* 8(3), 463-481.

Yue, P., Gong, J. and Di, L., 2008. Automatic Transformation from Semantic Description to Syntactic Specification for Geo-Processing Service Chains. *In:* Bertolotto, M., Ray, C. and Li, X. eds. *Web and Wireless Geographical Information Systems.* Springer Berlin Heidelberg, 50-62.

Yue, P.*, et al.* 2010. GeoPW: Laying Blocks for the Geospatial Processing Web. *Transactions in GIS,* 14(6), 755-772.

Yue, P., *et al.* 2011. Sharing geospatial provenance in a service-oriented environment. *Computers, Environment and Urban Systems,* 35(4), 333 - 343.

Zhang, J., Pennington, D. and Michener, W., 2006. Automatic Transformation from Geospatial Conceptual Workflow to Executable Workflow Using GRASS GIS Command Line Modules in Kepler. *In:* Alexandrov, V., *et al.* eds. *Computational Science – ICCS 2006.* Springer Berlin Heidelberg, 912-919.