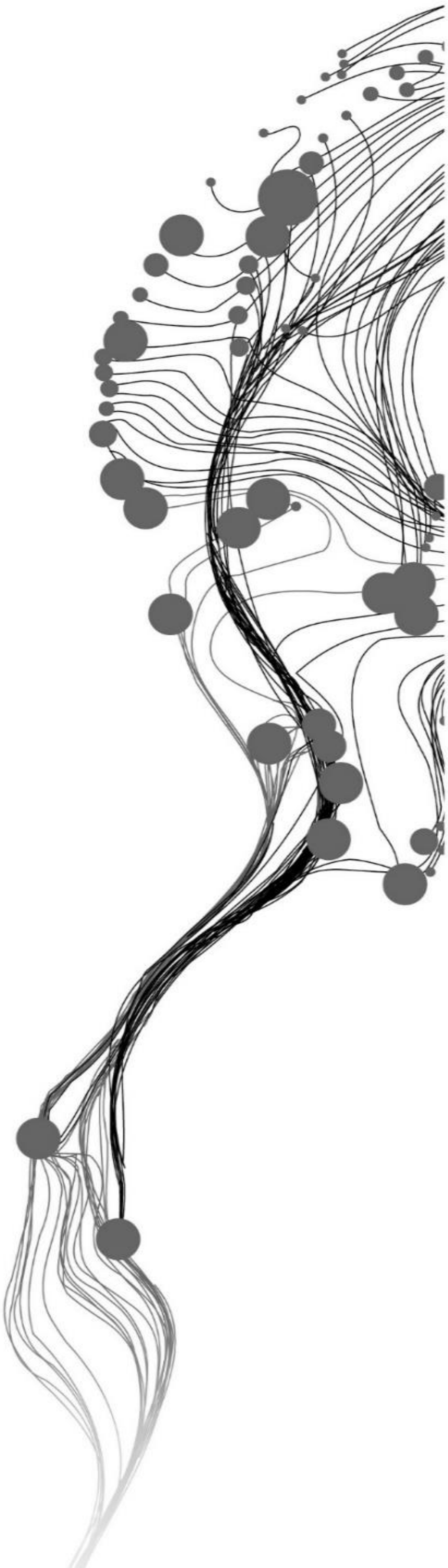
A METHOD FOR ENHANCING REPRODUCIBILITY OF GEOPROCESSING WORKFLOWS USING SEMANTIC ANNOTATION

DAWIT HAILE BEYENE

October 2021

SUPERVISORS:

Dr. Ir. R.L.G Lemmens
Dr. J.M. Morales



A METHOD FOR ENHANCING REPRODUCIBILITY OF GEOPROCESSING WORKFLOWS USING SEMANTIC ANNOTATION

DAWIT HAILE BEYENE

Enschede, The Netherlands, October, 2021

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

SUPERVISORS:
Dr. Ir. R.L.G Lemmens
Dr. J.M. Morales

THESIS ASSESSMENT BOARD:
Dr, F.O, Ostermann (Chair)
Dr. Ir, L.G.J, Boerboom (External Examiner, ITC)

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

In their everyday lives, people deal with spatial problems, whether or not they are conscious. These spatial problems could be specific, simple, straightforward, or complicated, which necessitate integrating various processes and data. In recent years, web technologies and cloud computing advancements have highlighted the importance of chaining geospatial operations and data to solve complex problems. Well-orchestrated sequential methods encapsulated as workflows can be used to allow the integration of datasets to solve complex scientific problems. The reproducibility of scientific applications has become increasingly important to advance computational science because it enables the original developer and other users or scientists to replicate, validate, and further expand original methods. However, most workflows are not reproduced for a variety of reasons, which is referred to as workflow decay.

There are several factors for the irreproducibility of geoprocessing workflows. Among these, one of the key problems for the irreproducibility of workflows is the lack of sufficient metadata description of the workflow. As a result, in this research, we developed a way to improve the semantics of workflows to enhance workflow reproducibility.

This research's main goal is to develop a method to enhance the reproducibility of geoprocessing workflows using semantic annotation. The implementation of the proof of concept required the development and integration of several components. An ontology has been developed in the Living Textbook (LTB), which helps model and store the semantic description used to embed the workflows. This research also studied the added values of concept mapping tools like LTB for ontology development and semantic enrichment. To automate the semantic annotation process, the JavaScript application has been developed. The prototype application has been designed for two user groups – workflow creator and workflow consumer.

This study has proven that, embedding additional semantic descriptions to the workflow increases the users understandability of a given workflow. This is attested by usability testing.

Keywords

Reproducibility, Geoprocessing Workflow, Living Textbook (LTB), Semantic Annotation.

ACKNOWLEDGMENTS

First and for most praise and thanks to almighty God, who gave me the knowledge and the ability to complete this work.

I wold like to thank the Orange Knowledge Program (OKP) for sponsoring my studies.

I would like to express my gratitude to my supervisors dr.ir. R.L.G. Lemmens and dr. J.M. Morales Guarin for their valuable guidance and support throughout the research process.

Finally, I want to thank my family and friends (Special thanks to Merhawi Fissehaye) for their love, prayers, advice and hope they had given to me.

Contents

ABSTRACT 4

ACKNOWLEDGMENTS 5

LIST OF FIGURES 8

LIST OF TABLES 9

LIST OF ABBREVIATIONS 10

1. INTRODUCTION 11

1.1. Background Information 11

1.2. Problem Statement 12

1.3. Research Objectives 13

1.4. Research Questions 13

1.5. Use Case 14

1.6. Thesis Outline 14

2. WORKFLOWS 16

2.1. Scientific Workflow. 16

2.2. Workflow Management Systems 17

2.3. Factors that Affect Reproducing Workflows. 18

2.3.1. Third-party resources 19

2.3.2. Nature of input data 20

2.3.3. Execution Environment 20

2.3.4. Metadata of the Workflow 21

2.4. Ways for Reproducing Workflows. 21

3. SEMANTIC ANNOTATION 23

3.1. Overview of Semantic Annotation 23

3.2. Ontologies 24

3.3. Criteria to develop an ontology 24

3.3.1. Decide the scope and domain of the ontology 24

3.3.2. Reuse existing ontology. 25

3.3.3. Define the class terms, the class hierarchy, and properties of the class 25

3.3.4. Languages and Tools to develop ontologies 26

3.4. Concept Maps for Ontology Development 26

3.4.1. The Living Textbook (LTB) 27

3.5. The Ontology 29

4. PROOF OF CONCEPT 32
4.1. System Overview Diagram 32
4.2. Prototype Implementation 35
4.2.1. Design 36
4.2.2. Development 38
4.2.3. Libraries used 38
4.2.4. Architecture 39
4.2.5. Deployment 39
4.3. Usability Testing and result discussion 39
4.3.1. Usability testing results 42
4.4. Code versioning and Deployment 46
5. CONCLUSION 47
5.1. Conclusions 47
5.1.1. Answers to research questions 48
LIST OF REFERENCES: 52
APPENDIX A: 56
APPENDIX B: 61
APPENDIX C: 66

LIST OF FIGURES

Figure 1: Composition of Workflow Management Systems (WFMS). Adapted from (Ohuru, 2019)............ 18
Figure 2: Factors contributing to workflow decay. Source (Zhao et al., 2012). 19
Figure 3: Workflow decay due to third-party resources. Source(Zhao et al., 2012)..................................... 20
Figure 4: The spectrum of reproducibility. Adapted from (Peng, 2011). 22
Figure 5: Generic Annotation Model (Andrews et al., 2012). 23
Figure 6: Hierarchy of Geo-processing ontology 25
Figure 7: Schema of Simple linked node export type. 28
Figure 8: Partial Schematic overview of the ontology. 29
Figure 9: Ontology shown in the LTB (See Appendix C for the list of concepts). 30
Figure 10: System overview diagram 32
Figure 11: A modal that asks the user to select a task. 33
Figure 12: The JSON structure of workflow created in the Workflowapp. 33
Figure 13: The result of parsed and semantically enhanced workflow... 34
Figure 14: The connection of the selected workflow element with other concepts in the LTB. 35
Figure 15: Mockup of the prototype application (A,B, & C). 37
Figure 16: workflow developed for lodging detection using DTM and CSM. Adapted from (Acorsi et al., 2019). 42

LIST OF TABLES

Table 1:Specification of LTB export types.. 27
Table 2: User profile of test participants.. 42
Table 3: Workflow consumers related questions response. 43
Table 4: workflow elements related participant response 44
Table 5: Participant response on the enhancement of the semantics of workflow. 44
Table 6: Summary of reproducibility questions participant 45

LIST OF ABBREVIATIONS

API Application Program Interface

BPEL Business Process Execution Language

BPM Business Process Management

BPMN Business Process Modelling Notation

CSS Cascading Style Sheets

ESRI Environmental Systems Research Institute

HCI Human Computer Interaction

ILWIS Integrated Land and Water Information System

ISO International Standards Organization

JSON JavaScript Object Notation

LTB Living Textbook

OGC Open Geospatial Consortium

SASM Semantic Annotation Structure Model

SDI Spatial Data Infrastructures

SWMS Scientific Workflow Management System

UI User Interface

UML Unified Modelling Language

URI Uniform Resource Identifier

VPN Virtual Private Network

WfMC Workflow Management Coalition

WfMS Workflow Management System

WWW World Wide Web

XML Extensible Markup Language

1. INTRODUCTION

1.1. Background Information

In their everyday lives, people deal with spatial problems, whether or not they are conscious. These spatial problems could be specific, simple, straightforward, or complicated, which necessitate integrating various processes and data. Data and operations are required to solve those problems. Fortunately, With the advancements in technology, those datasets are available. It is also not uncommon to find operations as services. Web technologies make the production, sharing, and reusing of data and services as fast as possible and process in distributed machines with high storage and processing capabilities. In order to achieve a result, there should be an appropriate combination of data and operations in spite of the availability.

In recent years, web technologies and cloud computing advancements have highlighted the importance of chaining geospatial operations and data to solve complex problems. Well-orchestrated sequential methods encapsulated as workflows can be used to allow the integration of datasets to solve complex scientific problems. The idea of workflows has been used for a long time to execute business processes automatically. Nowadays, workflows play an essential role in using data and processing operations and fueling scientific discoveries (Yolanda et al., 2007).

The workflow concept has been there for a relatively long period of time in facilitating and automating business processes. There has been a standardization of business processes since 1993 by the Workflow Management Coalition (WFMC) (Diniz, 2016). After three years of their establishment, in 1996, WFMC defined workflows as "*the automation of business processes, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules*" (Taylor et al., 2007). The above definition of workflow embraces only the business process domain, and its meaning is based on business process management (BPM). Whereas scientific workflows are "*workflows capture the individual data transformations and analysis steps as well as the mechanisms to carry them out in a distributed environment*" (Yolanda et al., 2008). Business processes are based on a control flow-driven approach, and they cannot be fully automated because they need the involvement of humans during their execution.

On the other hand, in scientific workflows, human beings are only involved during the creation of workflows. The computer can do all the remaining processing without the need for human intervention. Most scientific workflows are operated in a distributed environment because processing and storage on a single machine is not sufficient.

A workflow that incorporates geographic data and geoprocessing tools to accomplish a particular task is called a geoprocessing workflow. Geoprocessing workflows can be created using a variety of these tools, including the ESRI suite model builder, QGIS processing modeler, ILWIS suite model builder, ERDAS Imagine spatial modeler and the Workflowapp[1]. The main issue with these software packages (excluding the Workflowapp) is that they are proprietary and installed on a desktop machine, making it difficult to produce and share the workflow with others (Kechagioglou et al., 2019). The sharing and reproducing of workflows have a significant advantage in the scientific process as it allows scientists to understand scientific processes and methods developed by others. It can also be used as a starting point for a new approach. Therefore, the shareability and reproducibility of workflows have significance in achieving interoperability.

We can model workflows using Unified Modelling Language (UML) and Business Process Modelling (Ohuru, 2019). Compared to using unified modeling language, there is a wide range of using business process modeling for developing workflows. Business Process and Model Notation (BPMN) is used as a standard language to create workflows. It allows users to model an end-to-end sequence of a given process. It applies to both business and scientific workflows. After specifying a process using business process modeling notation, it should be saved as a BPMN document. The saved BPMN document has an XML representation of the graphical workflow. Before converting to their XML representation, BPMN documents cannot be executed by themselves (Decker et al., 2008). Due to this, a BPMN document should be converted to be executed. Its executable format is written in Business Process Execution Language (BPEL). A visual workflow should be serialized to BPEL scripts before execution (Ohuru, 2019).

1.2. Problem Statement

To advance computational science, the reproducibility of science applications has become increasingly important because it enables the original developer and other users or scientists to replicate, validate and further expand original methods (Meng et al., 2015). Scientific workflows need to enhance the reproducibility of scientific works by allowing sharing among different users. Reproducibility also plays a tremendous role in achieving interoperability. However, as indicated by (Hettne et al., 2012), most workflows are not reproduced for various reasons, and it is called a '*workflow decay*'. As (Zhao et al., 2012) indicated, one of the key problems for the irreproducibility of workflows is the lack of sufficient metadata description of the workflow. The metadata of workflows as one type of provenance information is helpful to reproduce workflows. A well-documented workflow associated with enough provenance information makes workflows efficiently reproduced (Bánáti et al., 2015).

[1] - https://gisedu.itc.utwente.nl/exercise/apps/workflow/

1.3. Research Objectives

The main research objective is to establish a method that can enrich the semantics of geoprocessing workflows. The data used to annotate the workflow has been created in the Living Textbook (LTB). This research will also investigate the added values to concept mapping tools to semantically annotate the geoprocessing workflows. Enhanced workflow semantics will enable users to understand what workflows do and improve the reproducibility of workflows.

There are three sub-objectives for this research:

Sub-Objective 1: To identify relevant existing approaches for the reproducibility of a geoprocessing workflow.

An assessment of the existing reproducibility approaches must be addressed to frame the scope of the research, investigate, and identify the complexity of the concept. The goal of this objective is to learn the fundamental concepts of the research area.

Objective 2: To devise a method to enrich the semantics of geoprocessing workflow for enhancing reproducibility.

This objective aims at investigating the technologies that helps to create a method which can be used to enhance the reproducibility of geoprocessing workflows.

Objective 3: Develop a prototype system to facilitate the annotation of geoprocessing workflow elements.

A system that can be used as a proof of concept should be created, used, and needs to be verified. Here, in this objective, we will build a JavaScript application to test the method created in objective 2.

1.4. Research Questions

Related to the first objective

I. What does it take for a geoprocessing workflow, in terms of its semantics, to be reproduced?
II. How well a workflow needs to be semantically enriched to be reproduced?
III. How do the existing reproducibility methods work with embedding semantic information into geoprocessing workflows?

Related to the second objective

I. What existing techniques are there to support semantic annotation?
II. What are the requirements to develop an ontology?
III. What are the pros and cons of languages and tools that are used to develop ontology?
IV. How can OGC standards and formal protocols help to enhance semantic enrichment?

Related to the third objective

I. What are the available methods to develop an annotation system?
II. For which type of user does the proposed system be applicable?
III. At which stages are user-system interactions needed?
IV. What are the limitations of semantic annotation in practice?

1.5. Use Case

The following criteria were used to select a use case to test the proof of concept:

- Workflows that address spatial problems (geoprocessing workflows).
- Workflows of which the reproducing benefits are clear.
- Workflows consist of operations that are available in most GI software.
- Workflows consist of various types of operations like Raster and Vector.
- The workflow is appropriate to consist of enough operations to see how versatile the application should be.

1.6. Thesis Outline

This thesis has adopted the following structure:

Chapter 1: provides a general introduction to the research, the research motivation, research objectives, and the questions that should be answered to accomplish the objectives.

Chapter 2: provides an explanation of workflow, types of workflows, and factors that affect reproducing geoprocessing workflows will be discussed. Additionally, a literature review on workflows will be covered. This section explains reproducibility and the criteria that should be met to call a geoprocessing workflow reproducible.

Chapter 3: explains the general overview of Semantic Annotation. Furthermore, it discusses the technologies behind semantic annotation systems. It discusses how to develop an ontology, tools to develop it, and the ontology developed for this research to be used as a data source by the prototype application. In addition to that, it discusses the added values of concept maps to develop ontology.

Chapter 4: explains the development and the details of the prototype application.

Chapter 5: explains the conclusions that are derived from the results.

2. WORKFLOWS

The idea of workflows existed for a long time in facilitating how business and scientific processes are designed and implemented. Workflows have been observed in the business and scientific realms in recent decades. Workflows are becoming increasingly popular due to various factors, the most important of which is the rise of web technologies and the availability of big data. The availability of spatial data and spatial services in Spatial Data Infrastructures (SDI) contributes to the growth of the workflow concept (Schäffer & Foerster, 2008).

Despite the fast production of workflows in every corner of the world, the lack of standards for sharing, reusing, and reproducing those workflows makes it challenging to use them to their full potential. To alleviate that problem, there was a standardization by Workflow Management Coalition (WfMC) in 1993 (Diniz, 2016). After three years of time, WfMC comes up with defining workflows as '*the automation of business processes, in whole or a part, during which documents, information or tasks are passed from one participant to another for action, according to procedural rules*' (Taylor et al., 2007). This definition is more related to business workflows. Scientific workflows can be defined as '*Workflows capture the individual data transformations and analysis steps as well as the mechanisms to carry them out in a distributed environment.*' (Gil et al., 2007). In most cases, business workflows need human intervention in the loop, whereas scientific workflows are automatic (Sonntag et al., 2010). Since business workflows orchestrate a step in a business domain, they need to be robust, which is not the same as in scientific workflows.

2.1. Scientific Workflow.

A scientific workflow is a process for achieving a scientific goal that is typically expressed in terms of tasks and their dependencies (Ludäscher et al., 2009). It is a well-orchestrated model of such scientific tasks (Deelman et al., 2018). Scientific workflow tasks are typically computational steps in scientific simulations or data analysis steps. Many of the current scientific works consist of several computational segments that are connected. Using scientific workflows to solve scientific calculations, conducting scientific experiments, and implementing many scientific analyses was inspired by the success shown in the realm of business workflows (Sonntag et al., 2010). Scientists of various disciplines are using scientific workflows for several benefits. Those benefits can range from getting the same result by reproducing the workflow to creating, documenting, and sharing scientific workflows among different user groups (Ubels, 2018).

Depending on the level of the details they provide, workflows can be classified as abstract workflows and concrete workflows (Kechagioglou et al., 2019). The abstract workflows provide a high-level insight into GI operations, the inputs, and the expected output of the execution. It doesn't explain the details of input types, parameter lists, etc. (Ohuru, 2019). They are not GI software consumables, and they are meant to be human-readable and used to establish workflow logic rather than being executed by the program. On the contrary, the concrete workflows encapsulate the detailed structure of the operation connections, parameter lists, execution environments, and the like. They can be directly consumed by WfMS and be executed to provide a result. Research conducted by Ubels, (2018), has proved the possibility of automatic conversion of abstract workflows to scientific workflows. Geoprocessing workflows are a type of scientific workflow that is used to develop a model that performs calculations and analytical experiments using geo-operations or tools in conjunction with spatial data.

2.2. Workflow Management Systems

Workflow Management System (WFMS) "*is a computer system used to automate, execute, and enact processes that are described in the workflows*" (Combi & Pozzi, 2008). Generally, WFMS has two components – the workflow client and the workflow engine. In most GI software, the workflow client is used to create the visual representation of the workflow by allowing users to drag and drop various workflow elements on the canvas. In addition to this, it enables users to make a logical connection to those workflow elements. Once the user finishes building the workflow, it should be translated to the code that can be understandable and executable by the workflow engine. When executing the workflow, the workflow engine follows the steps modeled by the user. The workflow engine runs at the back end, and when it finishes processing, it will send back the result to the workflow client so that the user can get the result.

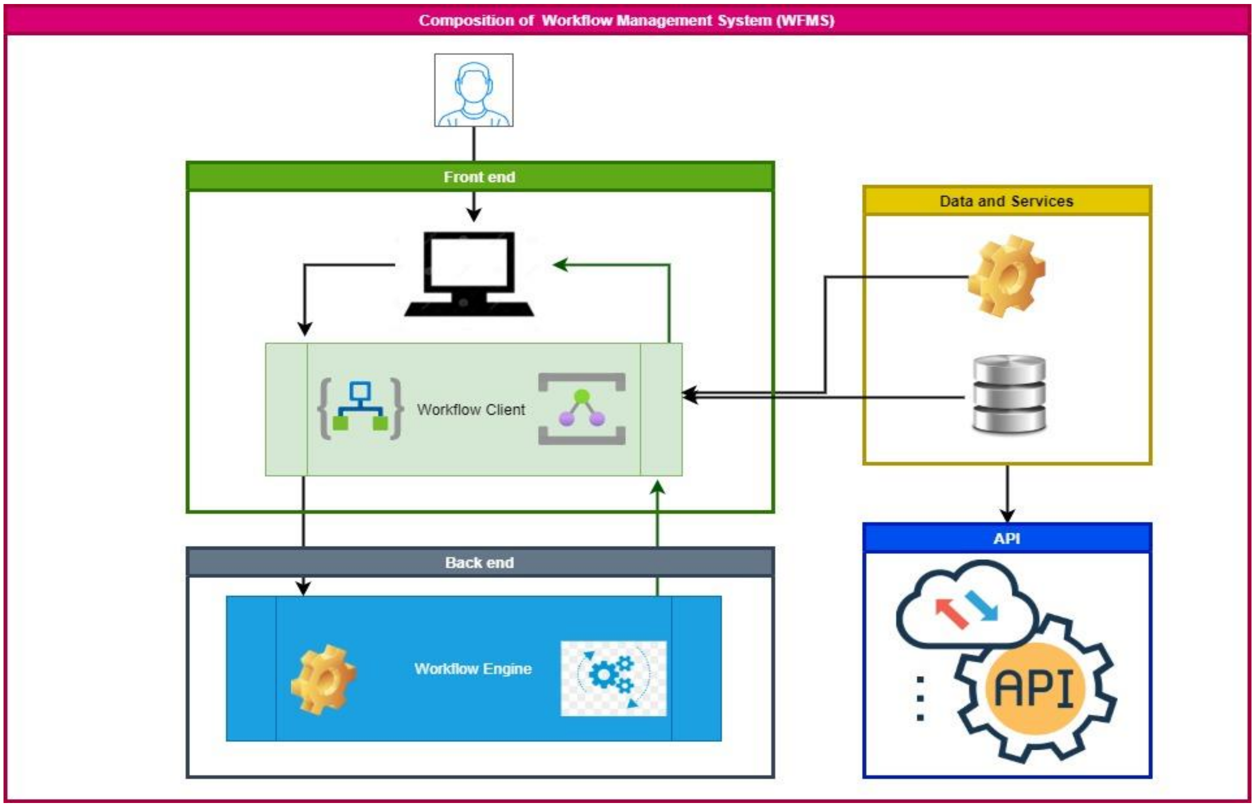


Figure 1: Composition of Workflow Management Systems (WFMS). Adapted from (Ohuru, 2019).

In the scientific community, the WFMS's are becoming progressively useful as a means to facilitate processing. To make standardized creation, sharing, and automation of workflows, several standardization organizations create standards. Object Management Group (OMG), Open Geospatial Consortium (OGC), and Workflow Management Coalition (WfMC), are organizations that create those standards. Through the use of these standards, as Schmidt, (1999) indicated, WFMS's would be able to automate and execute tasks that are developed by different vendors.

"*Reproducibility allows a workflow specified to address a particular scientific problem to be reused by different users under equivalent conditions without having to manipulate or change the original specification to produce scientifically similar results*".(Ohuru, 2019).

2.3. Factors that Affect Reproducing Workflows.

Reproducibility plays a crucial role in evaluating methods created by someone else by re-running the workflow somewhere else using various tools (Gil et al., 2007). Reproducibility enables a workflow designed to solve a specific issue to be reused by someone else without altering the original specification and using a similar dataset to get the same result. As Bechhofer et al., (2013) indicate, reproducibility helps to verify the accuracy of results by reusing workflows that someone already creates. It is necessary to share workflows to be able to reproduce them.

Different factors hinder the workflows from being reproducible. The term "*Workflow decay*", as defined by Zhao et al., (2012), is the context in which there is an inability to reproduce workflows. The terms *Workflow decay* and *workflow irreproducibility* can be used interchangeably. According to their findings, around 80% of workflows could not be reproduced due to several problems. The factors include volatile third-party resources, missing data, the execution environment, and insufficient metadata about the workflows. Each of these factors will be briefly discussed, even though this research only focuses on workflow irreproducibility caused by insufficient semantic metadata describing the workflow.

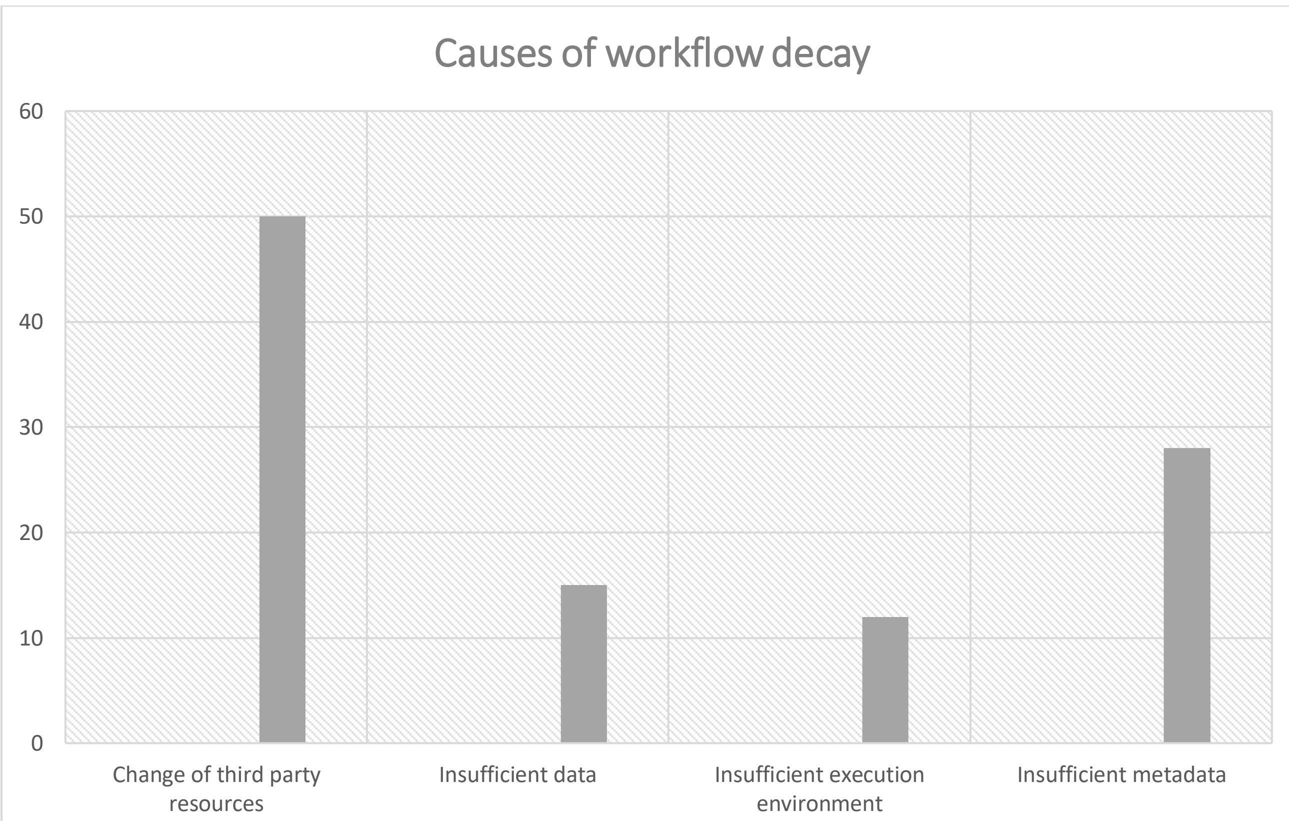


Figure 2: Factors contributing to workflow decay. Source (Zhao et al., 2012).

2.3.1. Third-party resources

Web services and databases are examples of third-party resources that can be used in the implementation of a workflow. It is obvious that, the workflow will not function properly if the web service that the workflow consumes are changed or modified. Web service providers may alter the design and execution of the services, resulting in a different outcome or making workflow execution difficult. As indicated in Figure 3, Third party resources could be unavailable, inaccessible or changed due to updates.

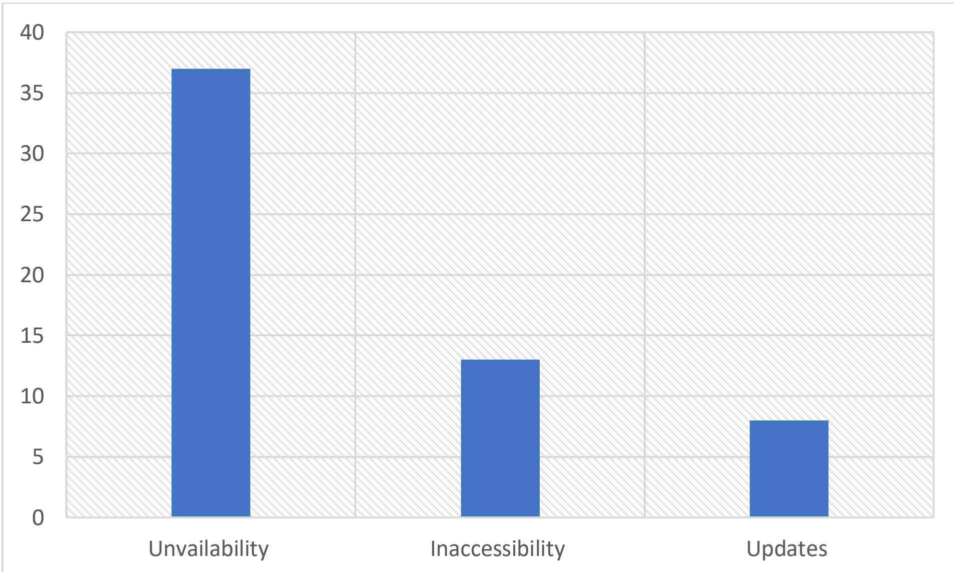


Figure 3: Workflow decay due to third-party resources. Source(Zhao et al., 2012).

Unavailability of web services can be caused due to a failure of servers where the services were running. If the server which hosts the services on the web can be down due to several reasons. If that is the case, the services can not be available and unable to use them in the workflow. Inaccessibility to web resources – *data* and *services* can happen due to change of naming of the services or it can happen due to the access right to those services. Updating the web services could happen there is enhancement of script or libraries. It will have direct impact on the change of quality of the result of workflow which uses those services.

2.3.2. Nature of input data

Unavailability or insufficiency of input data affects reproducibility of workflows. Unless the data is optional, if there is a missing data which the workflow needs to run, the workflow could not be executed, hence not reproduced. Not only the availability of data enables the workflow to run, but its compatibility to be consumed by the workflow.

2.3.3. Execution Environment

The workflow needs a workflow execution environment which consists of the software and libraries required to execute it. To reproduce the workflow, every library that the workflow uses should be available. In most cases, vendors of the software and libraries keep regularly updating them. This might cause an incompatibility issue with the original implementations of the workflow.

2.3.4. Metadata of the Workflow

It is important to embed workflows with metadata descriptions to enhance reproducibility of workflows. Having insufficient metadata information about the purpose of workflow, its inputs, the connection of workflow elements, and its expected result can make a workflow irreproducible. According to Zhao et al., (2012), about 28% of irreproducibility of workflows happened due to the unavailability of enough metadata information about the workflow.

Metadata is a collection of data that elaborates and provides an explanation of other data. Semantic annotation can be seen as formal metadata which is readable by machines and humans (Liao et al., 2011). Semantic annotation is explicitly and formaly defined as an ontology. Semantic annotation is a way of tagging, embedding, or attaching additional descriptive information or knowledge on workflows to improve their understandability. As a result, the term annotation can refer to both the process of annotating and the outcome of that process.

As discussed earlier, irreproducibility is not a back and white problem that can be solved by fixing one of the four factors. Creating a method which is able to fix all mentioned factors of irreproducibility is far from the scope of this research. This study aims to contribute to the reduction of workflow irreproducibility caused by a lack of metadata information. To do that, a semi-automated system that facilitates the semantic annotation process has been implemented. The details of the system will be covered in the coming chapters.

2.4. Ways for Reproducing Workflows.

In today's scientific world, an enormous amount of geo-spatial datasets and geo-computational operations are available. There are several ways of sharing those datasets and methods. Due to that, duplication of methods are increasing among different stakeholders. Enhancing the reproducibility of workflows has tremendous benefits in many scientific domains. However, because of the factors described in paragraph 2.3, reproductivity has not been achieved at the desired rate and level.

Reproducing workflows demands the fulfillment of several criteria. According to the area where the reproducibility of workflows is needed, there could be different criteria to be met to call a workflow reproduced or not. There are some approaches to be considered to enhance the reproducibility of workflows.

To successfully reproduce workflows, Peng, (2011) indicated conditions to be met to call workflows reproducible. The criteria range from least reproducible to fully reproducible workflows. Data and code are

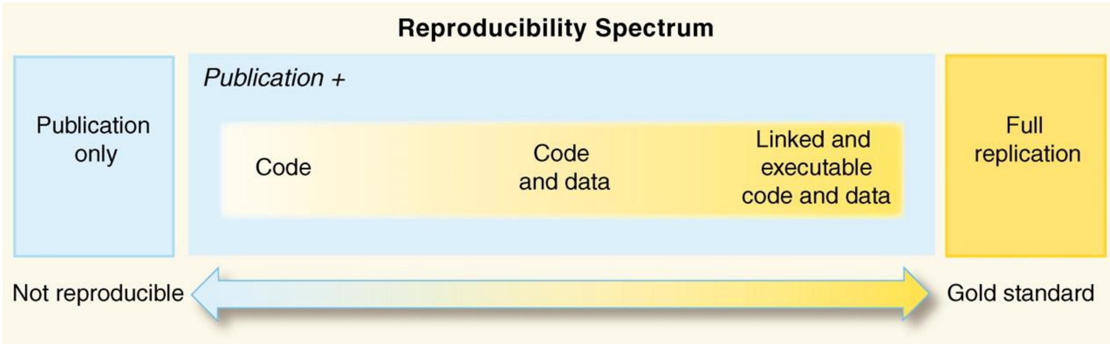


Figure 4: The spectrum of reproducibility. Adapted from (Peng, 2011).

mandatory to reproduce a method and to get the same result. A sufficient semantic description should also be given to describe the workflow logic and the data. So that one can better understand it and make the best out of it.

As Scheider et al., (2017) have outlined, Scientific Workflow Management Systems (SWMS) is one possible approach for making scientific workflows easier to reproduce. Most of the available SWMS has a graphical user interface that helps to create the workflow interactively. It uses nodes as a process and arrow that links each node to represent a data flow. This graphical way allows people to create, update, and execute the workflows.

Furthermore, there are several organizations like OGC that set standards to facilitate the sharing and reproducibility of workflows. Standards allow several resources to be linked together. Those resources can have different types of data formats. There are several standards like WPS, WCS, WFS, SWE , etc that OGC has defined to enable the creation of workflows using web services. WPS is one of these standards, and it may be included into workflows to run remote operations exposed by various GIS software. WFS provides vector datasets. The utilization of such a resource is confined to a narrow user group without sufficient descriptions. The OGC standard features like Service Capabilities provide descriptions and information to retrieve required data or services.

3. SEMANTIC ANNOTATION

3.1. Overview of Semantic Annotation

The Merriam Webster online dictionary and thesaurus defines the word "Annotation" as '*a note added by way of comment or explanation*'. The online Oxford dictionary defined it as '*a note by way of explanation or comment added to a text or diagram*' (Liao et al., 2011). We can derive from these two definitions that annotation is the addition of information to a target resource to improve its semantics (Figure 5).

Under different scenarios, annotations can be defined and used differently. In computer programming, annotations can be used as comments or docstrings that are embedded inside the source code to make it more understandable by the reader. Those annotations in computer programming intend to elaborate the details. Due to that, the compiler doesn't consider them as executable lines of code. In other cases, in addition to text, annotations can also be images aiming at enriching the target's information.

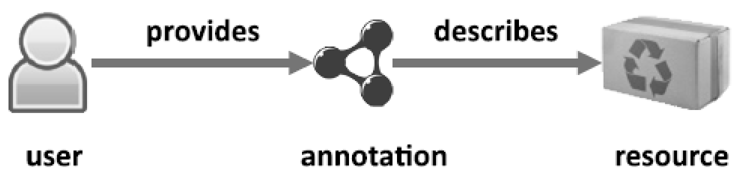


Figure 5: Generic Annotation Model (Andrews et al., 2012).

Depending on the type of information used to enhance the target, there are various types of annotations. Mosses, (2002) classified annotation as Semantic annotation, Textual annotation, and Linking annotation. Liao et al., (2011) defined semantic annotation as '*the action and results of describing (part of) an electronic resource by means of metadata whose meaning is formally specified in an ontology*'. Y. Lin, (2008), in his research, describes semantic annotation as a way of connecting knowledge that is organized as ontologies to the target information source. Semantic annotation is crucial to enrich models, systems, and other target information (Liao et al., 2011). Nowadays, they are used in many fields to integrate resources with their domain ontologies. Textual annotation is tagging the target with notes and comments. The Linking annotation enriches the semantics of the target by connecting the object to the human-readable information.

There are three main components of semantic annotation: Ontology, Semantic Annotation Structure Model (SASM), and Application (Liao et al., 2011). Semantic Annotation Structure Model (SASM) which is component of semantic annotation is the connection between electronic resources, applications, and ontology concepts

(Liao et al., 2011). The structure or schema of an annotation can be organized by SASM and the mappings between electronic resources and one or more ontologies can be described by SASM (Liao et al., 2011). By using SASM, an application can be designed to achieve user purposes like sharing and reuse, composition, integration, etc.

3.2. Ontologies

Gruber, (1993), defined an ontology as '*An ontology is an explicit specification of conceptualization*'. Recently, ontologies have been entering in every sphere of scientific research and experiments. They are also becoming common in the World Wide Web (WWW). According to their intended usage, ontologies could range from large taxonomies to small ones. Ontologies have the potential to establish a common perspective for various people under a similar domain to share information. It can provide definitions that are understandable and processable using machines and map the connection of concepts in the field.

Ontologies can be developed for a variety of reasons, depending on the type of users and the scenario in which they are needed. For example, ontologies can be designed to share information among people in the same domain and software agents (Gruber, 1993). Ontologies are also intended to enable reusing a specific domain knowledge. To name a few applications, ontologies can be used to define domain-general concepts in a top-level ontology, knowledge exchange and reuse, communication in multi-agent systems, natural language interpretation, and document search (Ohgren, 2004). Their potential to make an explicit assumption on certain domain alleviates the problems of hardcoding. The development of an ontology for a certain domain is not useful in and of itself, but it allows systems to be developed on it to get the most out of it.

3.3. Criteria to develop an ontology

The size and applicability of ontologies can be different in several scenarios. So, all of them are useful for which they are going to be used. Regardless of the size of the ontology, its intended use, which user groups will use it, or the tool that will be used to create it, the following points should be followed when developing an ontology.

3.3.1. Decide the scope and domain of the ontology

Before starting the process of developing an ontology, certain factors should be considered in order to answer questions such as: *What is the range that the ontology will cover at the end? For what purpose is the ontology will be developed? What type of questions is going to be answered by the data inside the ontology? Who is responsible for creating and maintaining the ontology?* etc. This research is not aiming at creating an all-rounded ontology that holds the semantic descriptions for all GI operations and for their various implementation environments. The ontology developed for this research contains data that is used to enhance the semantics of the geoprocessing workflow of some GI operations. The data is collected from online ESRI documentation. Most of the GI operations that are described in their documentation can be found in several GI software packages. Since the process of developing an ontology is iterative, allowing its contents to be updated as needed. According to their intended use, it can be scaled up or down.

3.3.2. Reuse existing ontology.

Developing an ontology is not a trivial task that can be done within a short period of time. Due to that, before starting to make a new one, it is worthy if there is existing ontology that might answer the problem, we would like to answer using the ontology.

3.3.3. Define the class terms, the class hierarchy, and properties of the class

Different approaches can be used to define the class hierarchy. Using the top-down approach, the ontology development process can begin by defining the highest classes, followed by the child classes. It can also be done starting by creating the child classes followed by grouping them to into their parent class. The combination of these two approaches can also be used when designing the class hierarchy. In addition to creating the classes, a proper connection of classes is needed. For this research, we followed the top-down approach, as shown in Figure 6.

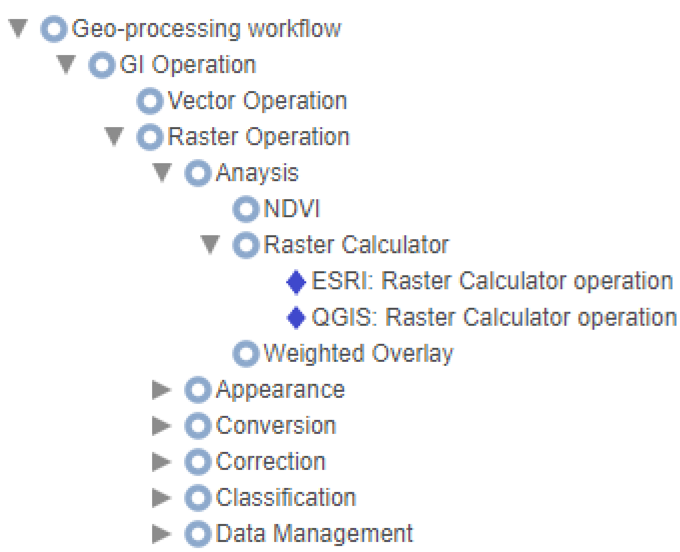


Figure 6: Hierarchy of Geo-processing ontology

3.3.4. Languages and Tools to develop ontologies

The ontology languages are a kind of formal language used to develop an ontology. Several languages for representing knowledge using ontologies have been created in recent years. Web Ontology Language (OWL), RDF Schema (RDFS), Simple Knowledge Organization System (SKOS), and more languages are available for creating ontologies. It's important to note that all the ontology languages mentioned above share the Resource Description Framework (RDF) concept as a common foundation (Ubels, 2018).

There are various ontological tools, and some of them can perform automated reasoning by using ontologies (Durán-Muñoz & Bautista-Zambrana, 2017). So, it will give advanced services for applications like conceptual search, semantic search, and retrieval. From the various tools, there are mainly two choices for building ontologies: standard ontology editing tools, like Protégé, and an ontology-based terminological resource editor, such as Ontoterm (Durán-Muñoz & Bautista-Zambrana, 2017). The drawback of using standard ontology editing tools is that it is not easy to adapt them to terminological purposes. Also, the work may be time-consuming; thus, this will discourage translators and terminologists from building the knowledge. Furthermore, standard ontology editing tools include several technical capabilities (for example, logical inferences) that aren't required for terminological projects and can slow the work process due to the quantity of knowledge required (Durán-Muñoz & Bautista-Zambrana, 2017).

3.4. Concept Maps for Ontology Development

Due to the advancements in technology, better interactive systems are becoming available to facilitate the conventional teaching and learning approach (Lemmens, R. L. G., Ronzhin, S., Augustijn, P. W. M., Verkroost, M. J., & Walsh, 2018). Concept mapping tools have been implemented to make that happen. Concept maps are used to organize specific knowledge using graphical representations (Novak & Cañas, 2008). In addition to using concept mapping tools for representing certain knowledge and use them to teach students in an interactive way, there is a possibility to use the data generated from them for further studies and analysis (Conceição et al., 2017). The proposed research will use the data that is generated from concept mapping tools to enhance the semantics of geoprocessing workflows. The process to embed textual description to a resource or contents aiming at making them more understandable and reusable is called Semantic Annotation or tagging (*Semantic Annotation*, n.d.). The semantic metadata can be embedded directly to the resource or can also be stored outside as ontology.

Ontologies are far more descriptive, but concept maps and ontologies have some similarities in terms of structure. Both ontologies and concept maps have concepts and define a relationship among them (Graudina

& Grundspenkis, 2008). Existing concept map tools are plentiful, and many more are on the way. In the next section, we will go over the concept map tool we used to create an ontology for this study.

3.4.1. The Living Textbook (LTB)

The Living Textbook (LTB) is one of the concept map tools developed in ITC to facilitate the conventional teaching-learning method more interactively. LTB, like other concept map tools, has a structure that resembles ontologies, as was discussed in the previous section. It is a tool to interactively construct course concepts along with their relationships. In this research, we used LTB to develop an ontology that contains descriptions of geo-operations.

To date, LTB does not provide an API that would allow third-party applications to make use of its real-time information. The ontology created in it can, however, be exported in a standard way. Because of the LTB's multiple export types, the ontology can be exported in a variety of formats. LTB's export types include *concept/instance (with id)*, *Learning path*, *Simple linked node*, *RDF (as JSON-LD)*, and *Relations*.

Table 1:Specification of LTB export types

NO	Name of export type	Data type	Holds description/definition of concepts
1	Concept/instance (with id)	CSV	No
2	Learning path	JSON	No
3	Simple linked node	JSON	Yes
4	RDF (as JSON-LD)	JSON	No
5	Relations	CSV	No

Of all mentioned export types, the Simple linked node is the one selected format to export the developed ontology from the LTB. There are several reasons for this, including that it provides a relatively rich full schema that contains the LTB concepts' definition. The LTB concepts' definition will be used as a semantic description to tag the workflow elements using the protoype application developed as a proof of concept.

```
{
    "nodes": [
        {
            "instance": "<concept-instance>",
            "label": "<concept-name>",
            "link": "<concept-url>",
            "numberOfLinks": <number-of-relations>,
            "definition": "<concept-definition>",
            "selfAssessment": "<concept-self-assessment>",
        }
    ],
    "links": [
        {
            "target": <target-id>,
            "source": <source-id>,
            "relationName": "<relation-name>"
        }
    ],
    "contributors": [
        {
            "nodes": [<node-ids>],
            "name": "<contributor-name>",
            "description": "<contributor-description>",
            "url": "<contributor-url>",
            "email": "<contributor-email>"
        }
    ],
    "external_resources": [
        {
            "nodes": [<node-ids>],
            "title": "<external-resource-title>",
            "description": "<external-resource-description>",
            "url": "<external-resource-url>",
        }
    ]
    "learningOutcomes": [
        {
            "nodes": [<node-ids>],
            "number": "<learning-outcome-number>",
            "name": "<learning-outcome-name>",
            "content": "<learning-outcome-content>",
        }
    ]
}
```

Figure 7: Schema of Simple linked node export type.

The schema of simple linked node format (Figure 7) contains five segments. These segments include *nodes*, *links*, *contributors*, *external_resources*, and *learningOutcomes*. The *node* contains the *label,* which is the name of the concept, *definition*, definition of the concept, and link that is the link to the concept's description. The details of how the developed JavaScript application uses this simple linked node schema is explained in chapter four.

3.5. The Ontology

Well-structured and standardized data is required to improve the semantics of geoprocessing workflows, to enable users better comprehend them, and to eventually reproduce them. To accomplish that, an ontology was developed as part of the research to hold a semantic description for workflow elements. This section describes the details of the developed ontology in the LTB. Creating an ontology that contains semantic descriptions for all geo-operations is out of the scope of this research. The ontology we developed in this research should be viewed as a proposed structure that could serve as the basis for the development of such comprehensive ontology.

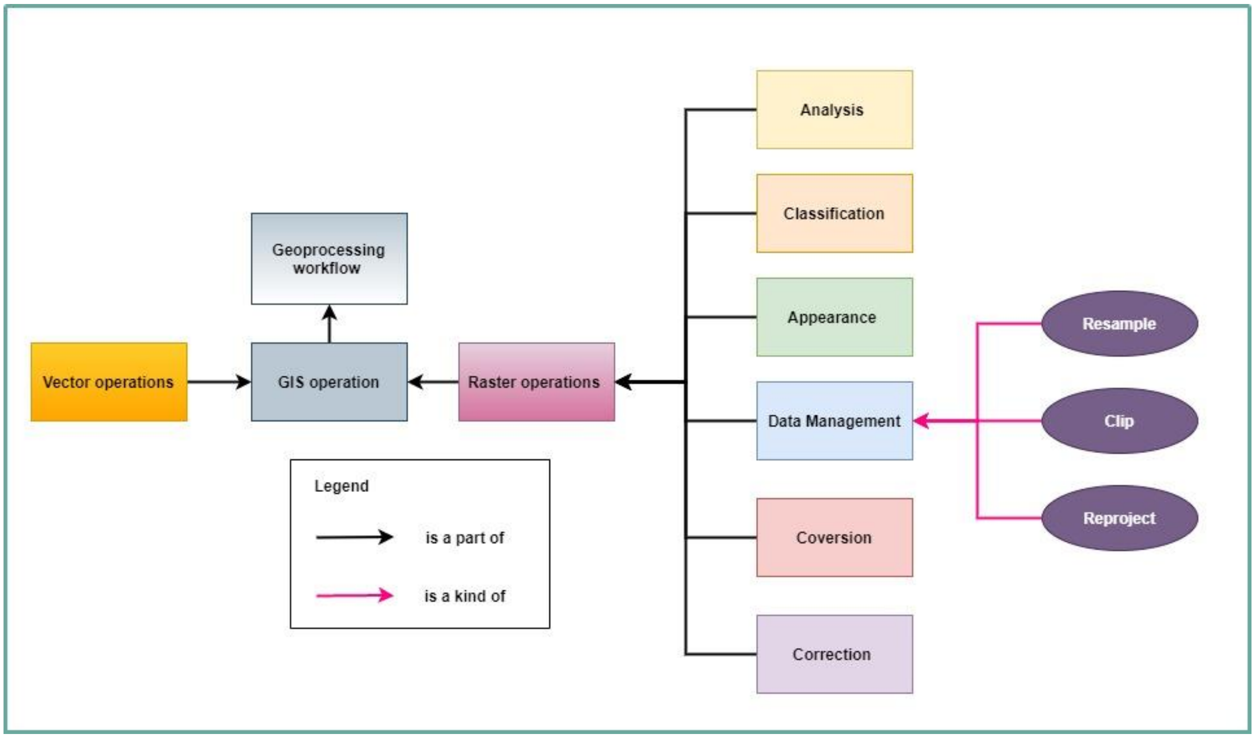


Figure 8: Partial Schematic overview of the ontology.

The ontology shown in Figure 9 has been developed in the LTB, to hold semantical descriptions to annotate the workflow elements. The ontology consists of 59 classes, of which 50 are geo-operations. As described earlier, creating an ontology that contains all geo operations is beyond the scope of this research. As a result these concepts are samples used as proof of concept. The remaining classes in the ontology are used to create the hierarchy of the ontology and to categorize geo-operations. For example, the operations *Resample*, *Clip*, and *Reproject* are grouped and put under the *Data Management* class which is not a actually a geo-operation.

The central class of the ontology is Geoprocessing workflow. The Geoprocessing workflow class has one immediate subclass, GIS Operation. The GIS operation class itself has two sub-classes: Vector operation and

Raster operation. These two sub-classes have further sub-classes describing categories of several types of geo-operations: Analysis, Classification, Appearance, Data Management, Conversion, and Correction. These six subclasses have other subclasses, which are geo-operations. The geo operations connected with their respective six classes with predicate *is a kind of*.

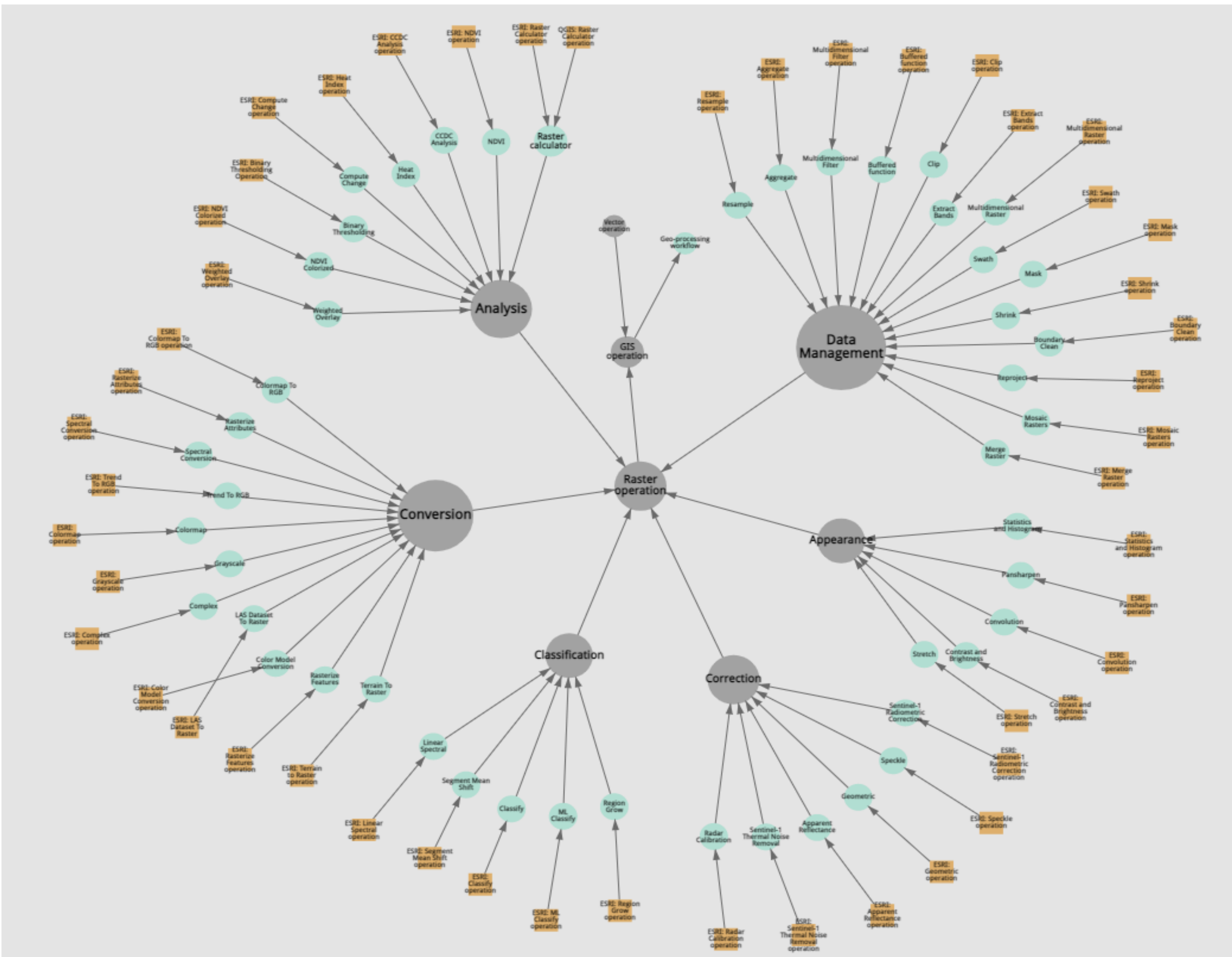


Figure 9: Ontology shown in the LTB (See Appendix C for the list of concepts).

Ontologies can be used for better data management. It can improve the data quality by improving the metadata of certain resource. The semantic descriptions that can be used to enrich the semantics of geoprocessing workflows can be modeled using ontologies. The ontology shown in Figure 9, has been modeled in the LTB and holds semantic description that can be used to improve the semantics of geoprocessing workflows.

As previously discussed in Section 2.3.4, one of the contributing factors for the irreproducibility of geoprocessing workflows is lack of sufficient metadata information. To enhance the semantics of geoprocessing workflows a well structured model of semantical dscriptions as ontology and a system which uses the ontology knowledge can be used. So by developing a system, which uses ontology data to annotate the workflows would potentially enhances the understandability of workflows and improves reproducibility.

4. PROOF OF CONCEPT

In this chapter, the tasks performed to establish a method that allows geoprocessing workflows to be semantically enhanced and the development of a prototype application will be explained. We explained here the different components of the system and how they are organized. We elaborate on the technologies used to build the system in terms of its design, development, and deployment. Finally, to verify the use of the developed application, we will discuss the result of the conducted usability testing.

4.1. System Overview Diagram

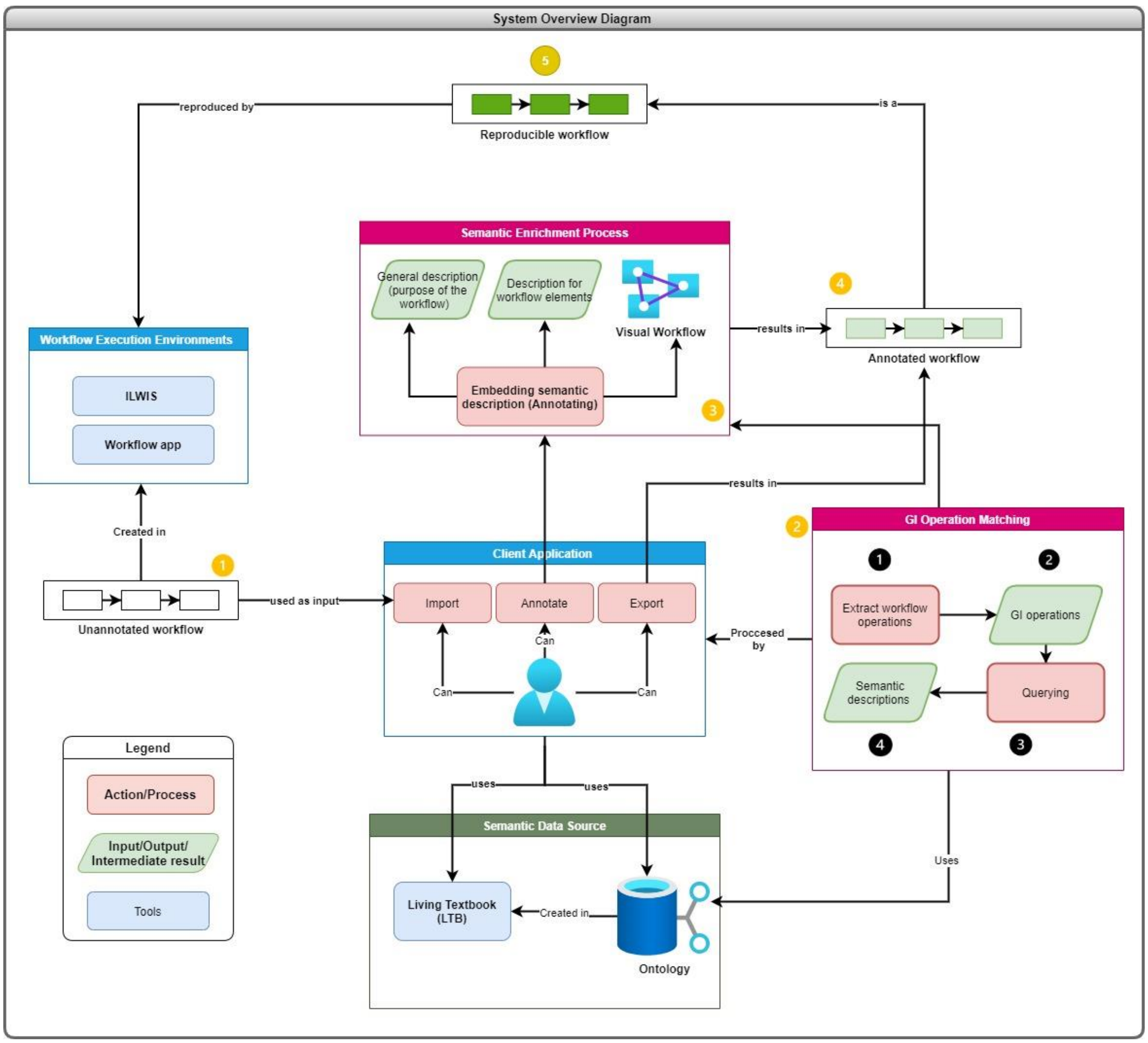


Figure 10: System overview diagram

A system overview diagram that shows how the application and all its parts work is shown in Figure 10. It shows the interaction of several components of the system. The main components of the system are the tools where the unannotated workflow is created, the ontology, the Living Textbook, and the prototype application. There are also automatic or semi-automatic tasks in some of the components.

Firstly, the user is served with an interface that asks what the user wants to do. The available options are:

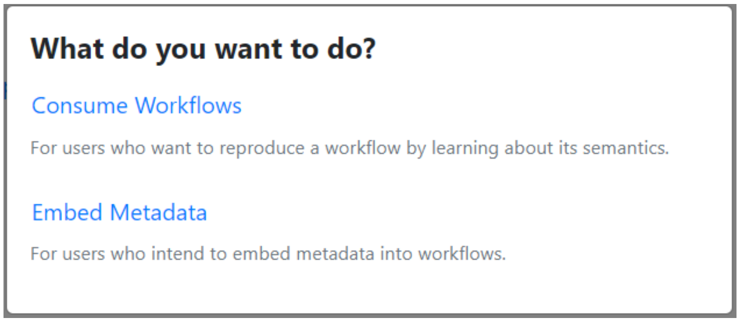


Figure 11: A modal that asks the user to select a task.

A. **To** **Consume workflow** - there are some workflows that comes embedded with the prototype application. If the user wants to reproduce them, he/she can go to the application, observe what is the workflow for, what are the operations used to build the workflow, what data is needed, what is the logical connection of the workflow elements, etc. Finally, the user can export the workflow and execute it.

B. **To** **Embed Metadata** **into workflows** – this option can be used by the user who already have a workflow created in the Workflowapp and wants to enhance its semantics before sharing it to some one else.

Depending on which option is selected by the users, the next step is to ask the user select/import the workflow. This is the first part where human – system interaction is needed. Then the system parses the workflow to

```
{
  "workflows": [
    {
      "id": 1,
      "metadata": {
        "longname": "Subworkflow"
      },
      "operations": [],
      "connections": []
    }
  ]
}
```

Figure 12: The JSON structure of workflow created in the Workflowapp.

extract the geo-operations. The operations are the elements which are connected each other to build the workflow. The JSON structure shown in Figure 12, is not a generic structure that all workflow development environments used. The internal data structure of the GI tools is different and needs to be handled accordingly. The prototype application is designed to handle workflow created using the Workflowapp.

Then after, the selected/imported workflow is going to be parsed. The system does two separate things here. The first is to filter out the name of the GI operation from the workflow and put them in the array and secondly querying the ontology for the matching semantic description for those GI operations. This specific process is shown in Figure 10, label 2, *GI Operation Matching*.

Once the ontology for existing semantic descriptions is searched, the system automatically annotate those workflow elements for which the system founds a matching semantic descriptions. The system uses string matching technique to find matches between the parsed workflow GI operations and the ontological GI operations.

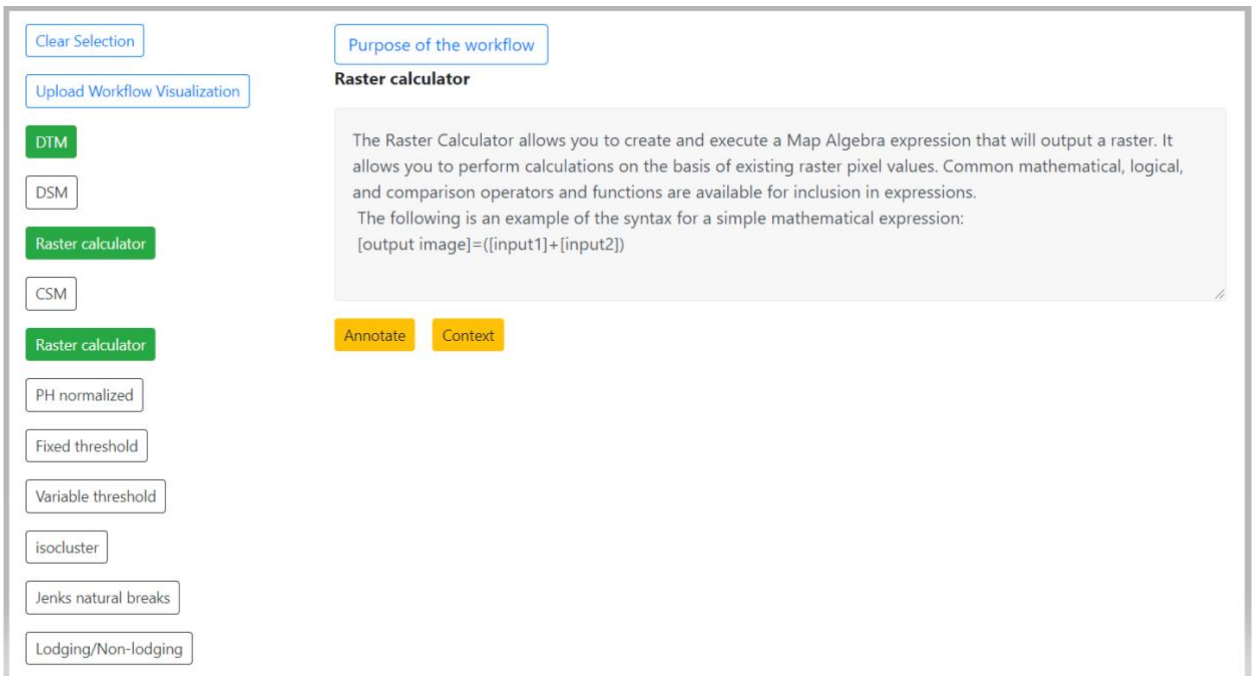


Figure 13: The result of parsed and semantically enhanced workflow.

From the listed workflow elements, those appeared in green are automatically annotated with semantic descriptions from the ontology. The system could not found semantic descriptions for the workflow elements shown in white. The *Semantic Enrichment Process* which is labeld in phase 3, in the overview diagram offers both automatic and manual annotation features. Due to that, the user can manually tag the elements with his/her descriptions. Furthermore, the user can also update the semantic descriptions of the workflow elements that the system automatically tagged. In addition to tagging each element of the workflow with semantic description, the general description of the purpose of the workflow can be found/tagged using the Purpose of the Workflow button.

In order also to better comprehend the semantics of the workflow element, the system allows users to view the context selected workflow element with other concepts in ontology. This can be done by clicking the Context button (Figure 14).

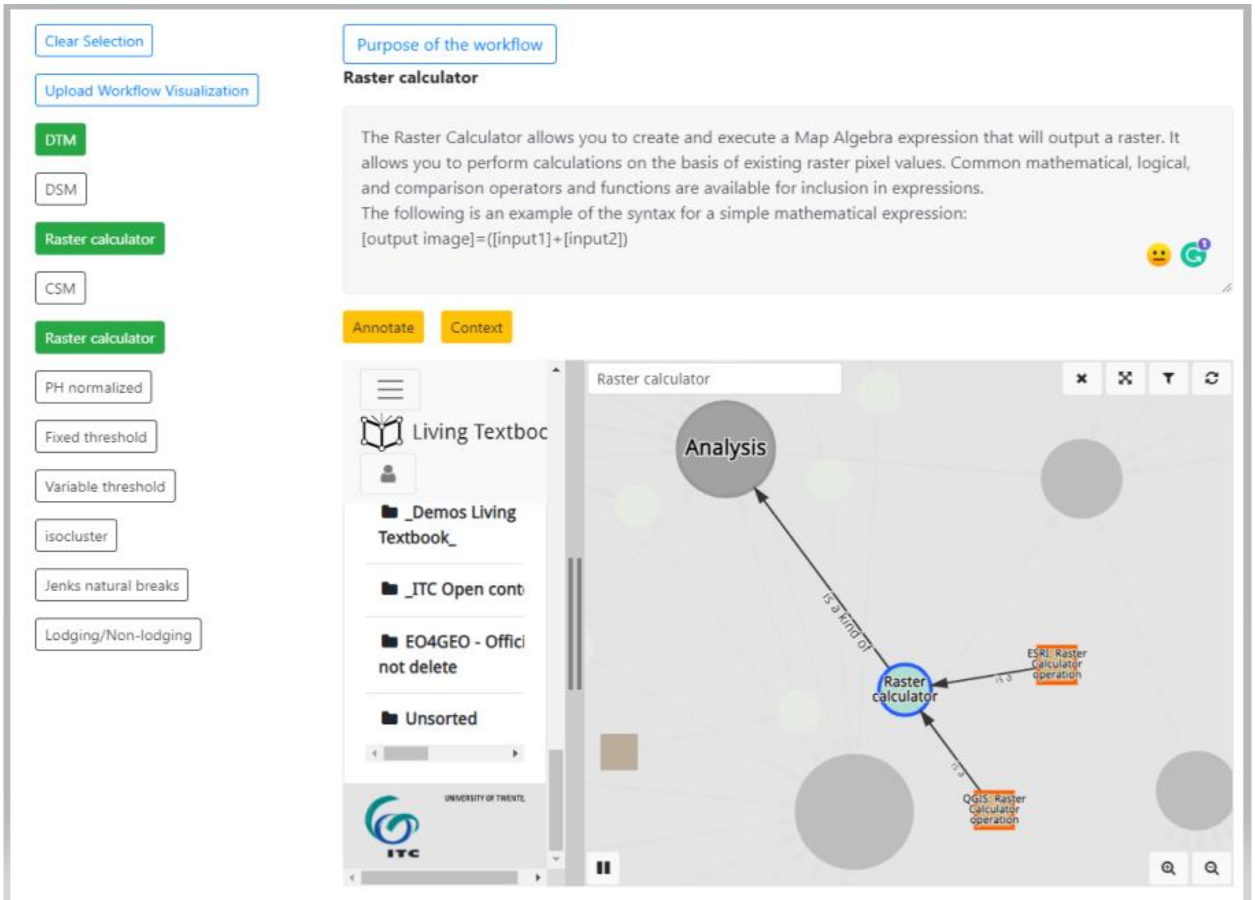


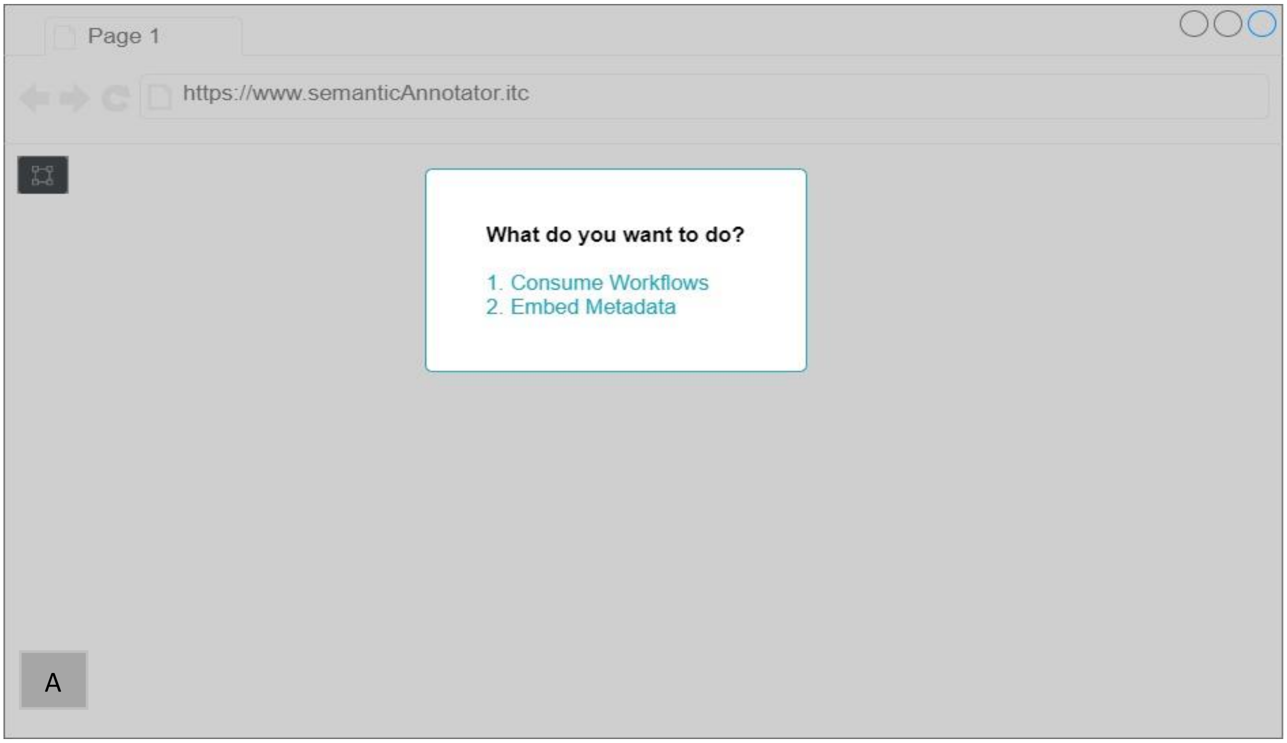
Figure 14: The connection of the selected workflow element with other concepts in the LTB.

4.2. Prototype Implementation

The prototype implementation was initially carried out using Python. But at the later stages, the JavaScript application was found better to implement the prototype system in terms of workflow processing (JSON workflow), designing and implementing user interfaces, and deployment. Explanation of the prototype system, starting from designing to deployment, will be discussed in the coming section.

4.2.1. Design

The ultimate aim of the prototype application is to facilitate the semantic annotation of geoprocessing workflows. As the initial design concept, the low fidelity of the prototype application was designed on paper. After designing on paper, we created a mockup (Figure 15) using https://app.diagrams.net/.



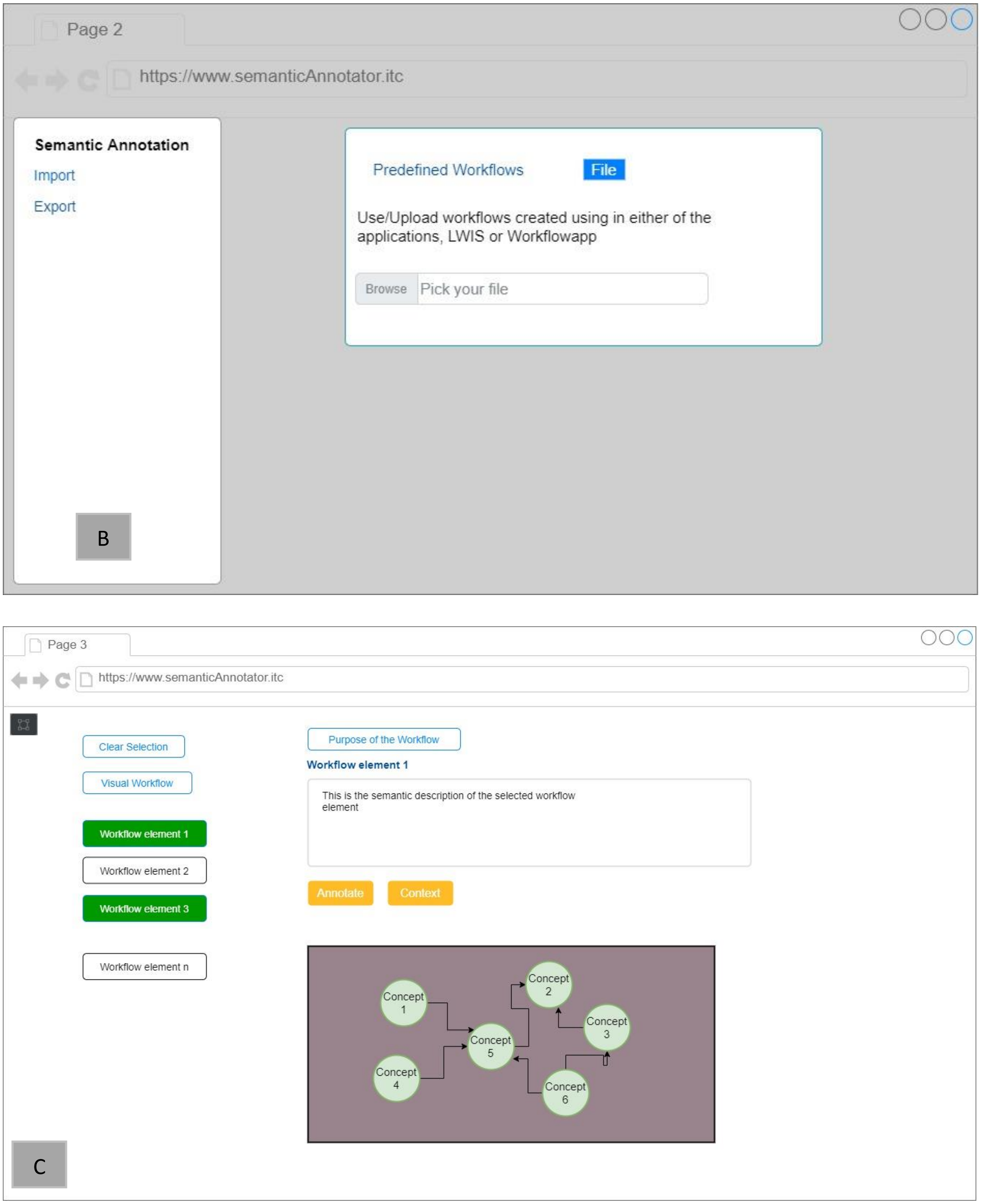


Figure 15: Mockup of the prototype application (A,B, & C).

4.2.2. Development

A web based platform was selected to make it easier for deployment, because a javascript project can be deployed on a server and accessed via the browser. A kind of desktop application was proposed using python at first but a web application is found preferable because desktop applications are platform dependent and require certain desktop environments to be installed in the system to be run including other constraints.

There are several web application development frameworks. I had filtered down the different options Angular, React and Vue. These frameworks have a very good support for two way data binding that is essential to control data flow between logic and presentation. As the main development framework, we used VueJs to implement the prototype system. VueJs is a modern JavaScript framework with strong support for advanced tasks like two way data binding.

4.2.3. Libraries used

Bootstrap Vue - a bootstrap CSS framework integration that works well with Vuejs application (Philippe Hong, 2018). Bootstrap is a User Interface (UI) framework with many important components like input boxes, dropdowns, button groups navs, paginations etc. In the implementation of the application some of those components were used. Bootstrap Vue adds modules like sidebar, rating and the like.

Vuex – a state management library that provides an API into the general state of the application (Nelson, 2018). It represents a global state that can be forwarded or reversed at anytime and provides a single point of entire state of the application.

Axios - is a lightweight Hyper Text Transfer Protocol (HTTP) based on XMLHttpRequests (XHR) service . It can be used to perform HTTP requests. The XHR object is used to interact with the server. You can retrieve data from the URL without refreshing the whole page. This allows the web page to update only part of the page without interrupting what the user is doing. It can be used to perform HTTP requests. The predefined workflows are stored in GitHub. To fetch data from GitHub API, axios enables communication with the external world via network calls.

Vue Router – is used to define different pages which can be navigated using different URI's (Lim & LaFranchi, 2019).

Firebase – it stores the user responses and image annotations provided by the users.

4.2.4. Architecture

I have used a state management pattern to manage the different states of the application as we continue to define workflow elements and annotate them.

4.2.5. Deployment

A minified distribution version of the prototype application has been generated and deployed on the ITC server. It can be accessed within that network. The application can be accessed using (https://gisedu.itc.utwente.nl/student/s2257335/annotator/index.html).

4.3. Usability Testing and result discussion

This section explains the selected usability testing methods, the criteria to choose the participants, and what elements of the proof-of-concept application need to be tested.

Several researchers defined *usability* in different ways. This is mostly determined by the researcher's goal and objective, as well as the study area. For instance, usability is defined as the capability of human function that can be utilized easily as well as effectively, given training and user support (Shackel, 1991). The author also illustrated that usability could be used to fulfill a specific range of tasks within the specified framework. Also, (Preece, J., Benyon, D., 1993) defined useability as an approach that the users execute tasks; safely, effectively, efficiently to improve the system. Furthermore, International Standards Organization (ISO) explained usability could be the quality of use with respect to effectiveness, efficiency, and satisfaction. The users achieve the intended goals in a particular task.

There are different reasons to conduct usability testing.

- **To validate prototype**: refers to assessing whether a given product meets its expected goal from the perspective of its end users.
- **To identify problems**: this is mostly concerned with finding defects in a product's design.
- **To uncover opportunities:** means discovering and opening space for product improvement based on user feedback and suggestions.
- **To learn the behavior of target user:** related to learning users' preferences from the product while observing their interaction to the product or service.

According to Murage, (2020), usability reports should incorporate a method that explains the approaches, the participants of usability testing, the tasks to be performed, the rationale behind preparing the evaluation method, and finally, processing the feedback. The purpose of this usability testing is to evaluate how users can achieve a specific objective.

Based on the intended goals of the study, usability testing methods are either qualitative or quantitative. Qualitative usability testing mostly incorporates collecting findings, insights, and others on how users utilize the product or service (Murage, 2020). In contrast, quantitative useability test relies on statistical data description of the user's experience.

In most cases, the qualitative usability test is more frequently used than the quantitative usability test. For this research, a qualitative usability test is considered. According to Roth et al., (2015), one might think that usability testing can be conducted in some design and implementation stages, including user requirements, prototyping the system, or final test and evaluation steps. This research focuses on usability evaluation made on the final stage of the application. Roth et al., (2015) stresses on the selection and categorizing the participants of usability testing. We have selected different participants from different domains with diverse backgrounds and experiences. This research adopted the criteria used by (Bowman et al., 2002). Those criteria include: What are the objectives of the evaluation method, at what point the method should be used in the assessment, and What results are expected and how they can be used to check the prototype application's usability.

In order to better understand the system and to understand the limitations that can emerge in the system, open and closed questions are developed to collect users' views, observations, and ideas. Appendix B presents all the questionnaires and profiles of the study participants. Numerous studies used questionnaires to gather participant's or user's ideas (Delikostidis, 2011). Moreover, to ensure the consistency of results and guide participants to give envisaged feedback, this research used a combination of open and closed-ended questions (van Elzakker & Wealands, 2007).

The tasks analysis method was selected according to the concept of the quality of use of a product. ISO/IEC 25010 defined quality of use as the result of an interaction between a user and a product measured as effectiveness, efficiency, and satisfaction (Bevan et al., 2016). Effectiveness is linked to the functioning of the components of a software system and their performance (H. X. Lin et al., 1997). Efficiency measures relate to the effectiveness of resource expenditure. Satisfaction measures the overall ease of the product or system.

The participants in the Usability testing are expected to complete a set of tasks using different techniques. The aim of conducting such kind of testing is to identify usability problems. The overall reason for conducting usability testing can be summarized as:

- To assess whether the participant finishes certain tasks successfully or not.

- Evaluate how satisfied the participants when using the product.
- Get feedback from the participant on the parts of the product to be changed, enhanced, or modified.
- To analyze the performance of the product if it fulfills the usability objectives.

We have developed a prototype application to allow users to embed semantic metadata into geoprocessing workflows in order to enhance their understandability and reproduce them. To test the application's capability, two different use cases were chosen. The first is to allow the creator of the workflow to embed semantic metadata on the workflows. And the second is to see how better the workflow consumer can understand the semantically enhanced workflow.

Use Case 1: Workflow Creator

A workflow creator is a person who developed an original workflow using available tools. In this use case, we will assess how well the prototype application helps the workflow creator enhance their workflow with semantic descriptions.

Use Case 2: Workflow Consumer

An individual who takes advantage of already established workflows is a workflow consumer. In this case, the application aims to allow consumers to reproduce the *predefined* workflows that are embedded in the prototype application. People from several disciplines can consume workflows, and they can have various experiences as well.

Scenario:

Peter is a junior GIS analyst hired a month ago by Glob-GIS. He conducts an analysis that the company receives from its customers. There are many customers who need to conduct a certain study for themselves. The study results must be obtained as soon a possible by clients. The results of the study were therefore used as an input for their needs. Recently, the company has been tasked with studying how much rice is yielded from the coming year by the local government and how much product they can lose from lodging. As a result, Peter wishes to reuse methods in order to make use of the model, which has already been tested and is time efficient. He searches for an existing method that can make such an analysis shown in Figure 16. His search revealed a method used by the company, but the analyst wants to understand what the workflow does with regard to data use, operations, the input data, and the outcome of the method.

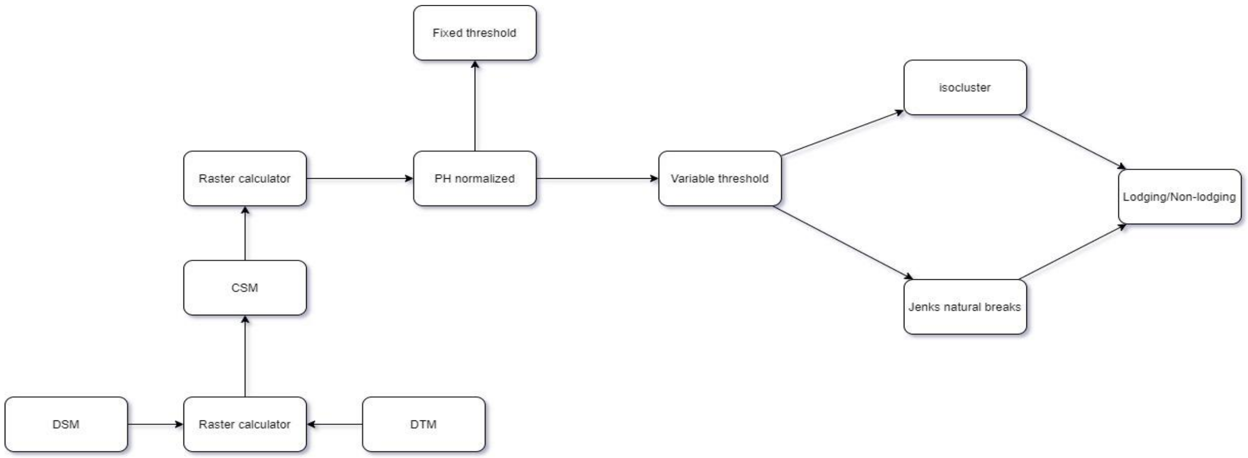


Figure 16: workflow developed for lodging detection using DTM and CSM. Adapted from (Acorsi et al., 2019).

4.3.1. Usability testing results

A) User profile

The usability testing was conducted with 7 participants. Table 2 presents the participant's academic background and their experience in using workflows. Of the participants, 71% have a master's degree, whereas 29% are bachelor holders. All participants have some sort of experience using workflows; however, the level of experience they have differs. Furthermore, they have varying experiences in creating, reusing, or reproducing geoprocessing workflows.

Table 2: User profile of test participants.

Participants	Education level	Background	Status	Workflow experience
P-1	Masters	Geography	Student	To some extent
P-2	Masters	Computer Science, Geoinformatics	Recent graduate	To great extent
P-3	Bachelor	Computer Science	Student	To some extent
P-4	Masters	Geoinformatics	Student	To some extent
P-5	Masters	Geoinformatics	GIS consultant	To some extent
P-6	Bachelor	Geography	Student	To some extent
P-7	Masters	Geoinformatics	Student	To some extent

B) Task-based questions

There were both open-ended and closed-ended questions prepared. The user experience was evaluated based on the user interface, affordances, and other factors in line with the question. The participant response on the prototype application for each task is presented in this section.

The usability questions are organized under two use cases:

1. Workflow consumer
2. Workflow creator

To determine how well the prototype application can assist users in understanding the semantics and reusing the predefined workflows. Also, through the prototype application, to see how well the user can integrate semantic descriptions into their workflows.

Scenario-1 (Workflow consumer)

Table 3 illustrates the summary of the participant's feedback for each question (1-4), shown in Appendix B. All of the respondents perceived the intent of the components of the prototype application. Of the participant, 85.7% easily navigated and founded the Lodgedetection workflow, while 14.3 % gets difficulties. About 14.3 % did not easily find Lodgedetection workflow, possibly because the level of interactions with apps differs from person to person.

Observing the visual representation of the Lodgedetection workflow helps all the participants understand the workflow's logic much better. The visual representation of the workflow, as explained by participants, allowed them to understand the logic, interaction, and sequential steps of the workflow.

With respect to the color used when displaying the workflow element, most participants (57.1%) respond that it gives them a clue. The participants explained that the color used to display the workflow elements provides users to look through. For instance, green implies certain things done in it, whereas white hints that nothing is done in it. This implies that the color representation of the workflow elements is found essential to enhance the understanding of the users.

Table 3: Workflow consumers related questions response.

Questions	Response(s)		Percentage(%)	
	Yes	No	Yes	No
1	7	-	100	0
2	6	1	85.7	14.3
3	7	-	100	0
4	4	3	57.1	42.9

Table 4, presents the participant response summary interacting with the workflow elements. All the participants showed that the UI components are displayed when the users interact with the workflow elements. More than half(57.1%) were very satisfied, and 42.9 were satisfied with the capabilities of the application to facilitate the reproduction of existing workflows. In summary, the participant stated that the used prototype is self-explanatory, interactive, and presents tasks in an understandable way. In line with this, they suggested that it will become more usable if the window used to display the LTB is slightly bigger.

Table 4: workflow elements related participant response

Questions	Response(s)		Percentage(%)	
	Yes	No	Yes	No
a	7	-	100	0
b	7	-	100	0
C	7		100	0
Level of satisfaction among participants				
	Unsatisfied	Poorly satisfied	Satisfied	Very satisfied
Percentage	0	0	42.9	57.1

Scenario 2: (Workflow Producer)

In order to analyze by the users (participants) with respect to the enhancement of the semantics of created, some questions are used:

1) In this prototype application, was it possible to find a way where you could embed semantic descriptions about the purpose of your workflow?
2) Using the prototype, could you able to tag each element of your workflow with semantic descriptions?
3) Related to question 2, Did you get a predefined semantic description for the elements of your workflow in the ontology (LTB)?
4) In the prototype, did you find a way to save all changes you made into your workflow and be able to find an outlet to export it for later use?

83.3 % of the participants indicated that annotating a workflow with general descriptions is possible. Similarly, it is possible to tag each element of the workflow with semantic descriptions. When the user tries to enrich their workflows, most of them (83.3%) reveal that they found a predefined semantic description for the workflow elements in the ontology (LTB), whereas 16.7% did not. This might arise from the experience of the participants and educational backgrounds. For the fourth question (4), all the participants find a way to save and export the semantically enhanced workflow. This is summarized in Table 5.

Table 5: Participant response on the enhancement of the semantics of workflow.

Questions	Response(s)		Percentage(%)	
	Yes	No	Yes	No
1	5	1	83.3	16.7
2	5	1	83.3	16.7
3	5	1	83.3	16.7
4	6	-	100	0

Additionally, all participants indicates that the prototype application helped them add their own semantical descriptions to workflow.

Section 3: Reproducibility questions

Table 6 presents the levels of participants' understanding of AfriAlliance workflow embedded in the prototype application as compared to Workflowapp. Of the participant 14.9 % indicated level 4 and 5, where as 14.3% gives ranks as level 3. Of the participant, 85.7% are executed the AfriAlliance workflow that shown in the prototype application using the Workflowapp, whereas 14.3% are not. From this result, we can conclude that users got a much better understanding of the same workflow after semantic enhancement.

Table 6: Summary of reproducibility questions.

Questions	Response (Level 1-5) and number of participants				
	1	2	3	4	5
1	-	-	1	3	3
Percentage (%)	0	0	14.3	42.9	42.9
Questions	Response (s)		Percentage (%)		
	Yes	No	Yes	No	
2	6	1	85.7	14.3	

Lack of metadata information contained in workflows is one of the causes of workflow decay, accounting for 28% of the total (Zhao et al., 2012). In this study we assess the ways to enrich the semantics of workflows to improve reproducibility. Semantic enrichment of workflows improves the understandability of geoprocessing workflows, according to the results of usability testing.

The results of usability testing have verified that the procedure described in the previous chapters is effective. The usability testing result has revealed that, semantic enrichment potentially helps users to better understand the workflow. The results has also shown that the application that was developed as a proof of concept helps in facilitating the semantic annotation process in an interactive way. There were also some feedbacks that the participants suggested to improve like increasing the number of concepts in the ontology and minor modifications in the user interfaces.

4.4. Code versioning and Deployment

To keep track of changes to the application's source code and data, a version control system is used. Version control systems come in a variety of shapes and sizes. The version control system used in this study is Github. We can gain a lot of advantages by employing version control systems. Github can also be utilized as a repository for all source code, data, and implementation. The predefined workflows that are embedded in the prototype application can be accessed (https://github.com/Dawit225/Geo_workflow/). A minified distribution version of the prototype application has been generated and deployed on the ITC server. The application can be accessed using (https://gisedu.itc.utwente.nl/student/s2257335/annotator/index.html). We have used firebase storage to deploy the workflow images that are uploaded by users who reproduce the workflow.

5. CONCLUSION

5.1. Conclusions

This research presents the use of a semantic annotation method to improve the understandability which contributes to the reproducibility of geoprocessing workflows. It established a methodological framework on how to make geoprocessing workflows more reproducible by improving their semantics. An ontology and a prototype system that uses the data from the ontology was established to improve the semantics of the geoprocessing workflow. This research also demonstrated that tagging semantic descriptions to the workflow would potentially enhance reproducibility.

Reproducibility is not a black-and-white problem that can be handled with a single solution. Several criteria must be examined while determining reproducibility (see Section 2.3). A workflow needs to be shared before reproduced. Depending on the export type used in the GI software, there are several methods for sharing workflows. BPMN (XML), JSON, and other formats can be used for sharing. Some workflows use third-party resources like data and services. So, the workflow will not function properly if the web service that the workflow consumes are changed or modified. Furthermore, the workflow should contain enough detail of descriptions so that it can be easily understandable.

The focus of this study is enhancing the reproducibility of workflows caused by lacking sufficient metadata descriptions (see Section 2.3.4). The developed ontology and a prototype system were integrated well and facilitated the semantic annotation process. As the usability testing on the implemented system prototype has shown, additional semantic descriptions enhance the understandability of workflows. As a result, it contributes to the improvement of the reproducibility of workflows.

This research also proves that, concept mapping tools like the Living Textbook (LTB) can be used to model and store knowledge as ontology development tools. In this study, we used the LTB to store data that can be used to annotate geoprocessing workflows.

5.1.1. Answers to research questions

Objective 1: To study relevant existing approaches for the reproducibility of geoprocessing workflows.

This objective aimed to look through existing methods to study what strategies are available to improve the reproducibility of geoprocessing workflows.

RQ 1.1. What does it take for a geoprocessing workflow, in terms of its semantics, to be reproduced?

Reproducibility is a process of re-running the workflow created by someone else using the same or various tools. To be able to reproduce a workflow, the following properties should be followed.

- ✓ A workflow must be shared before it can be reproduced.
- ✓ Third-party resources should be available and accessible for workflows that utilize them.
- ✓ The required data needs to be available to reproduce a workflow.

The above-mentioned criteria are the general criteria that one needs to think of when reproducibility comes to mind. In addition to that, workflows should have sufficient semantic information. The semantic information ranges from the overall description and purpose of the workflow to describing each element of workflow.

RQ 1.2. How well a workflow needs to be semantically enriched to be reproduced?

To improve the semantics of geoprocessing workflows, this study followed a semi-automated semantic annotation approach. We used the top-down approach to embed semantic description to the geoprocessing workflows. At the first level, an overview description of a workflow should be embedded with it. It includes the purpose of a workflow, on which environment it can be reproduced, which data is needed to execute the workflow, etc. The next level with which semantic description should be embedded is elements of the workflow. The prototype system uses a string matching technique to search and embed semantic descriptions to the workflows elements. The description associated with the ontology concept is embedded in the workflow element when it discovers a match between the workflow element's label and an ontology concept. The accuracy of a description can be determined by how much information it gives on a given concept. The usability testing showed that embedding semantic descriptions in a stage-based approach has helped users better understand the workflow.

RQ 1.3. How do the existing reproducibility methods work with embedding semantic information into geoprocessing workflows?

Most of the currently available GI tools that are used to develop workflows use the self-documenting technique to embed semantic information into their workflow. Most GI tools used to develop workflows provide a drag and drop feature for creating workflows. So, for each element of the workflow, there is a predefined description to explain it. The descriptions can be stored in the form of files, database tables, or in an ontology.

Objective 2: To devise a method to enrich the semantics of geoprocessing workflow for enhancing reproducibility.

RQ 2.1. What existing techniques are there to support semantic annotation?

Semantic annotation can be widely used in various fields and domains for several purposes. It can be used for composition, integration, sharing, or reusing resources. There are some scientific methods that help to develop semantic annotation systems. Semantic annotations systems should contain the following three components. The first one is ontology. It can be used to describe and express a body of knowledge as well as define concepts. Ontologies can be modeled in a way that can be reused. The second one is the Semantic Annotation Structure Model (SASM). It helps to arrange the structure of annotation and describe the mapping between resource and ontology. The last one is the application which is designed and implemented to make use of the ontology. For the implementation of this application, the required technologies are discussed in RQ 3.1.

RQ 2.2. What are the requirements to develop an ontology?

To develop an ontology the following criteria should be considered (see Section 3.2). These criteria can be used as a standard to follow for developing an ontology.

- ✓ Decide the scope and domain of the ontology
- ✓ Reuse existing ontology (if any).
- ✓ Define the class terms, the class hierarchy, and properties of the class
- ✓ Identify languages and tools used to build an ontology.

The size and applicability of ontologies can be different under several scenarios. Regardless of the size of the ontology, its intended use, which user groups will use it, or the tool that will be used to create it, the above points should be followed when developing an ontology.

RQ 2.3. What are the pros and cons of languages and tools that are used to develop ontology?

There are a variety of ontological languages such as RDF, RDF Schema, OIL, DAML+OIL, and OWL. The availability of these languages promotes data transfer across different ontology-related applications, but it makes data interchange or reuse between systems that do not share the same languages unfeasible. If all ontological editors use the same languages, such as OWL, the problem will be solved easily, however, this isn't always possible or hasn't been done yet. From various ontological tools, some of them can perform automated reasoning by using ontologies. On the other hand, the standard editing tools have a drawback which is not easy to adapt them to terminological purposes and can be time-consuming.

RQ 2.4. How can OGC standards and formal protocols help to enhance semantic enrichment?

Standards allow several resources to be developed and shared in a standard way. Those resources can have different types of data formats. There are several standards like WPS, WCS, WFS, and SWE that OGC has defined to enable the creation of workflows using web services. For example, WFS provides vector datasets. The utilization of such a resource is confined to a narrow user group without sufficient descriptions. The OGC standard features like Service Capabilities provide descriptions and information to retrieve required data or services.

Objective 3: Develop a prototype system to facilitate the annotation of geoprocessing workflow elements.

RQ 3.1. What are the available methods to develop an annotation system?

In this study, we developed an ontology to store semantic descriptions that can be used to enrich the semantics of elements of a workflow. Ontologies are used in a variety of domains to store knowledge that can be used to enrich the semantics of a target resource. Exporting the ontology in JSON, a lightweight and most popular data interchange format, can be parsed and queried with several high-level programming languages. JSON data-interchange format is well supported by languages such as C++, C#, Java, PHP, Python, and JavaScript. JavaScript has better support for parsing JSON files. As a result, the semantic annotation process was facilitated by the integration of ontology, the JSON data-interchange format, and the JavaScript programming language (see Section 4.2).

RQ 3.2. For which type of user does the proposed system be applicable?

Two user groups can use the prototype application that we developed as part of the research:

1. Workflow consumer

An individual who takes advantage of already established workflows is a workflow consumer. In this case, the developed system allows consumers to understand and reproduce the predefined workflows embedded in the prototype application. People from several disciplines can consume workflows, and they can have various experiences as well.

2. Workflow creator

A workflow creator is a person who developed an original workflow using available tools. The workflow creator can use the prototype application to document their workflow by tagging semantic information on it. The prototype application provides both automatic and manual annotating features.

RQ 3.3. At which stages are user-system interactions needed?

The prototype application that we developed as part of the research is semi automated system. Due to that, user-system interaction is needed in several stages of the annotation process. The first point where user needs to interact with the application is to import a workflow to be semantically enhanced. Once the system receives the workflow, it automatically parses the workflow, searches semantic descriptions for workflow elements from the ontology, and displays the finding to the user. Following this, as a second point where user-system interaction is needed, the user can check the semantic description given by the system, edit it, and embed it with the workflow.

RQ 3.4. What are the limitations of semantic annotation in practice?

Various reasons limit the full use of semantic annotation depending on the domain where they are required. One of the factors is the size and unstructured nature of the target resource to be annotated. The data structure containing the target resource has to be machine-readable. It should also be easily processable and can easily integrate into the underlying ontological structure.

LIST OF REFERENCES:

Acorsi, M. G., Martello, M., & Angnes, G. (2019). Identification of maize lodging: A case study using a remotely piloted aircraft system. *Engenharia Agricola*, *39*(specialissue), 66–73. https://doi.org/10.1590/1809-4430-ENG.AGRIC.V39NEP66-73/2019

Andrews, P., Zaihrayeu, I., & Pane, J. (2012). A classification of semantic annotation systems. *Semantic Web*, *3*(3), 223–248. https://doi.org/10.3233/SW-2011-0056

Bánáti, A., Kacsuk, P., & Kozlovszky, M. (2015). Four level provenance support to achieve portable reproducibility of scientific workflows. In *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. https://doi.org/10.1109/MIPRO.2015.7160272

Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., Couch, P., Cruickshank, D., Delderfield, M., Dunlop, I., Gamble, M., Michaelides, D., Owen, S., Newman, D., Sufi, S., & Goble, C. (2013). Why linked data is not enough for scientists. *Future Generation Computer Systems*, *29*(2), 599–611. https://doi.org/10.1016/J.FUTURE.2011.08.004

Bevan, N., Carter, J., Earthy, J., Geis, T., & Harker, S. (2016). Masaaki Kurosu (Ed.) Human-Computer Interaction Theory, Design, Development and Practice. In *Human-Computer Interaction. Theory, Design, Development and Practice*. http://www.springer.com/series/7409

Bowman, D. A., Gabbard, J. L., & Hix, D. (2002). A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence: Teleoperators and Virtual Environments*, *11*(4), 404–424. https://doi.org/10.1162/105474602760204309

Combi, C., & Pozzi, G. (2008). Workflow Management Systems for Healthcare Processes. *Https://Services.Igi-Global.Com/Resolvedoi/Resolve.Aspx?Doi=10.4018/978-1-59904-889-5.Ch177*, 1412–1416. https://doi.org/10.4018/978-1-59904-889-5.CH177

Conceição, S. C. O., Samuel, A., & Biniecki, S. M. Y. (2017). Using concept mapping as a tool for conducting research: An analysis of three approaches. *Http://Www.Editorialmanager.Com/Cogentsocsci*, *3*(1), 1404753. https://doi.org/10.1080/23311886.2017.1404753

Decker, G., Kopp, O., Leymann, F., Pfitzner, K., & Weske, M. (2008). Modeling service choreographies using BPMN and BPEL4Chor. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5074 LNCS*, 79–93. https://doi.org/10.1007/978-3-540-69534-9_6

Deelman, E., Peterka, T., Altintas, I., Carothers, C. D., van Dam, K. K., Moreland, K., Parashar, M., Ramakrishnan, L., Taufer, M., & Vetter, J. (2018). The future of scientific workflows. *International Journal of High Performance Computing Applications*, *32*(1), 159–175. https://doi.org/10.1177/1094342017704893

Delikostidis, I. (2011). *Improving the Usability of Pedestrian Navigation Systems Volunteered Geographic Information (VGI) for Disaster Management: A Case Study for Floods in Jakarta View project Vario-scale geo-information View project*. https://www.researchgate.net/publication/272025784

Diniz, F. D. E. C. (2016). *Composition of Semantically Enabled Geospatial Web Services*. University of Twente, Faculty of Geo-information Science and Earth Observation.

Durán-Muñoz, I., & Bautista-Zambrana, M. R. (2017). Applying Ontologies to Terminology: Advantages and Disadvantages. *HERMES - Journal of Language and Communication in Business*, *51*, 65–77. https://doi.org/10.7146/HJLCB.V26I51.97438

Gil, Y., Deelman, E., Ellisman, M., Fahringer, T., Fox, G., Gannon, D., Goble, C., Livny, M., Moreau, L., & Myers, J. (2007). Examining the challenges of scientific workflows. *Computer*, *40*(12), 24–32. https://doi.org/10.1109/MC.2007.421

Graudina, V., & Grundspenkis, J. (2008). *CONCEPT MAP GENERATION FROM OWL ONTOLOGIES*. www.google.com

Gruber, T. R. (1993). 1-s2.0-S1042814383710083-main.pdf. In *Knowledge Acquisition* (pp. 199–220). http://ac.els-cdn.com/S1042814383710083/1-s2.0-S1042814383710083-main.pdf?_tid=28c5c43c-df11-11e2-b19c-00000aacb35f&acdnat=1372327738_c3518b0aab22af436aef0c5e3e2dd48e

Hettne, K., Wolstencroft, K., Belhajjame, K., Goble, C., Mina, E., Dharuri, H., Verdes-Montenegro, L., Garrido, J., De Roure, D., & Roos, M. (2012). Best practices for workflow design: How to prevent workflow decay. *CEUR Workshop Proceedings*, *952*(January).

Kechagioglou, X., Lemmens, R., & Retsios, V. (2019). Sharing geoprocessing workflows with business process model and notation (BPMN). *ACM International Conference Proceeding Series*, *Part F1482*, 56–60. https://doi.org/10.1145/3318236.3318239

Lemmens, R. L. G., Ronzhin, S., Augustijn, P. W. M., Verkroost, M. J., & Walsh, N. (2018). Space Education with The Living Textbook, A web-based tool using a Concept Browser. In L. Bacsárd (Ed.), *Proceedings of the 2nd Proceedings Symposium on Space Educational Activities, Budapest, Hungary.* (Issue November).

Liao, Y., Lezoche, M., Panetto, H., & Boudjlida, N. (2011). Why, Where and How to Use Semantic Annotation for Systems Interoperability. *Proceedings of the 1st UNITE Doctoral Symposium*, 71–78.

Lim, B., & LaFranchi, R. (2019). Model, Vue, and Controller. *Vue on Rails*, 41–79. https://doi.org/10.1007/978-1-4842-5116-4_3

Lin, H. X., Choong, Y. Y., & Salvendy, G. (1997). A proposed index of usability: A method for comparing the relative usability of different software systems. *Behaviour and Information Technology*, *16*(4–5), 267–277. https://doi.org/10.1080/014492997119833

Lin, Y. (2008). Semantic Annotation for Process Models: Facilitating Process Knowledge Management via Semantic Interoperability. *Norwegian University of Science and Technology (NTNU).*

Ludäscher, B., Bowers, S., & McPhillips, T. (2009). Scientific Workflows. *Encyclopedia of Database Systems*, 2507–2511. https://doi.org/10.1007/978-0-387-39940-9_1471

Meng, H., Kommineni, R., Pham, Q., Gardner, R., Malik, T., & Thain, D. (2015). An invariant framework for conducting reproducible computational science. *Journal of Computational Science*, *9*, 137–142. https://doi.org/10.1016/j.jocs.2015.04.012

Mosses, P. D. (2002). The semantics of semantic equations. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *28 LNCS*(May 2014), 409–422. https://doi.org/10.1007/3-540-07162-8_701

Murage, N. (2020). *Mapping flows of benefits from marine ecosystems across the globe AN INTERACTIVE WEB MAPPING APPLICATION* [Faculty of Geo-Information Science and Earth Observation of the University of Twente, Enschede, The Netherlands]. https://essay.utwente.nl/85213/1/murage.pdf

Nelson, B. (2018). State Management. *Getting to Know Vue.Js*, 127–147. https://doi.org/10.1007/978-1-4842-3781-6_7

Novak, J. D., & Cañas, a J. (2008). The Theory Underlying Concept Maps and How to Construct and Use Them. *IHMC CmapTools*, *January 2006*, 1–36. https://doi.org/Technical Report IHMC CmapTools 2006-01 Rev 2008-01

Ohgren, A. (2004). *Ontology Development and Evolution: Selected Approaches for Small-Scale Application Contexts*. https://www.diva-portal.org/smash/get/diva2:4295/FULLTEXT01.pdf

Ohuru, R. O. (2019). *A Method for Enhancing Shareability and Reproducibility of Geoprocessing Workflows. Case Study: Integration of Crowdsourced Geoinformation, Satellite and In-Situ Data for Water Resource Monitoring* [Faculty of Geo-Information Science and Earth Observation of the University of Twente, Enschede, The Netherlands]. https://essay.utwente.nl/83539/1/ohuru.pdf

Peng, R. D. (2011). Reproducible Research in Computational Science. *Science*, *334*(6060), 1226–1227. https://doi.org/10.1126/SCIENCE.1213847

Philippe Hong. (2018). *Practical Web Design: Learn the fundamentals of web design with HTML5, CSS3 ... - Philippe Hong - Google Books*. Packt Publishing Ltd. https://books.google.nl/books?hl=en&lr=&id=g_RZDwAAQBAJ&oi=fnd&pg=PP1&dq=bootstrap+vue&ots=o3DkAESu2X&sig=su9WuHKuiyLutrvxjmHc0IjppEc&redir_esc=y#v=onepage&q=bootstrap vue&f=false

Preece, J., Benyon, D., & O. U. (1993). *A Guide to usability : human factors in computing*. Wokingham, England: Addison-Wesley.

Roth, R. E., Ross, K. S., & MacEachren, A. M. (2015). User-centered design for interactive maps: A case study in crime analysis. *ISPRS International Journal of Geo-Information*, *4*(1), 262–301. https://doi.org/10.3390/ijgi4010262

Schäffer, B., & Foerster, T. (2008). A client for distributed geo-processing and workflow design. *Journal of Location Based Services*, *2*(3), 194–210. https://doi.org/10.1080/17489720802558491

Scheider, S., Ostermann, F. O., & Adams, B. (2017). Why good data analysts need to be critical synthesists. Determining the role of semantics in data analysis. *Future Generation Computer Systems*, *72*, 11–22. https://doi.org/10.1016/J.FUTURE.2017.02.046

Schmidt, M. T. (1999). Evolution of workflow standards. *IEEE Concurrency*, *7*(3), 44–52. https://doi.org/10.1109/4434.788778

Semantic Annotation. (n.d.). Retrieved December 2, 2020, from https://www.ontotext.com/knowledgehub/fundamentals/semantic-annotation/

Shackel, B. (1991). Usability - context, framework, design and evaluation. In Brian Shackel & Simon Richardson (Ed.), *Human factors for informatics usability* (pp. 21–38). Cambridge University Press.

Sonntag, M., Karastoyanova, D., & Deelman, E. (2010). Bridging the gap between business and scientific workflows: Humans in the loop of scientific workflows. *Proceedings - 2010 6th IEEE International Conference on e-Science, EScience 2010*, 206–213. https://doi.org/10.1109/eScience.2010.12

Taylor, I. J., Deelman, E., Gannon, D., & Shields, M. S. (2007). Workflows for e-Science: Scientific Workflows for Grids. *Workflows for E-Science: Scientific Workflows for Grids*, *January*, 1–523. https://doi.org/10.1007/978-1-84628-757-2

Ubels, S. (2018). *Understanding abstract geo-information workflows and converting them to executable workflows using Semantic Web technologies* [Utrecht University]. https://dspace.library.uu.nl/handle/1874/365897

van Elzakker, C. P. J. M., & Wealands, K. (2007). Use and Users of Multimedia Cartography. In W. Cartwright, M. P. Peterson, & G. Gartner (Eds.), *Multimedia Cartography* (pp. 487–504). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-36651-5_34

Yolanda, Ellisman, M., Fahringer, T., Fox, G., Gannon, D., Goble, C., & Myers, J. (2007). *"Examining the Challenges of Scientific Workflows," in Computer, vol. 40, no. 12,* (pp. 24–32).

https://doi.org/10.1109/MC.2007.421

Yolanda, G., Deelman, E., Ellisman, M., Fahringer, T., Fox, G., Gannon, D., Goble, C., Livny, M., Moreau, L., & Myers, J. (2008). Examining the Challenges of Scientific Workflows. *Computer*, *40*, 24-32. 10.1109/MC.2007.421.

Zhao, J., Gomez-Perez, J. M., Belhajjame, K., Klyne, G., Garcia-Cuesta, E., Garrido, A., Hettne, K., Roos, M., De Roure, D., & Goble, C. (2012). Why workflows break - Understanding and combating decay in Taverna workflows. *2012 IEEE 8th International Conference on E-Science, e-Science 2012*. https://doi.org/10.1109/eScience.2012.6404482

APPENDIX A:

Manual For Using The Prototype Application

To start interacting with the prototype application, the user needs to click (https://gisedu.itc.utwente.nl/student/s2257335/annotator/index.html). If the user is not in the ITC network, he/she might need to turn on Virtual Private Network (VPN) on their computer. Then after, a white modal window with self-explanatory text will be served for the user. It allows users to do two things here:

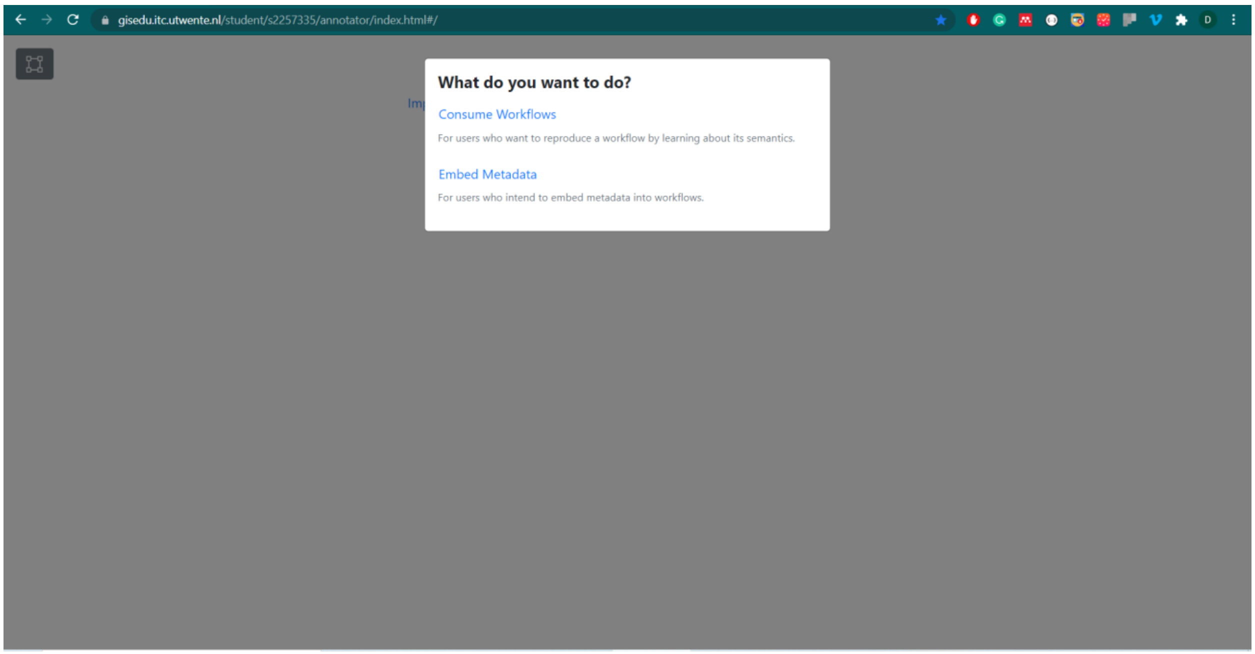


Figure A1: The first page of the prototype application served to the user.

1. To **Consume Workflows**

This functionality can be used by the users who want to reproduce the existing workflow by understanding its semantics. Once the user clicks on Consume Workflows, a new window shown in Figure 2A will be served. The user can continue interacting by clicking the Import link. After clicking the Import link, the white window modal will pop up to allow the user select Predefined Workflows or Import new one by clicking File from their device as shown in Figure 3A. For this specific use case, a user who is going to consume a workflow, needs to click on Predefined Workflows to select one of listed workflows.

Figure 2A: Consume Workflows page asking the user to select workflows

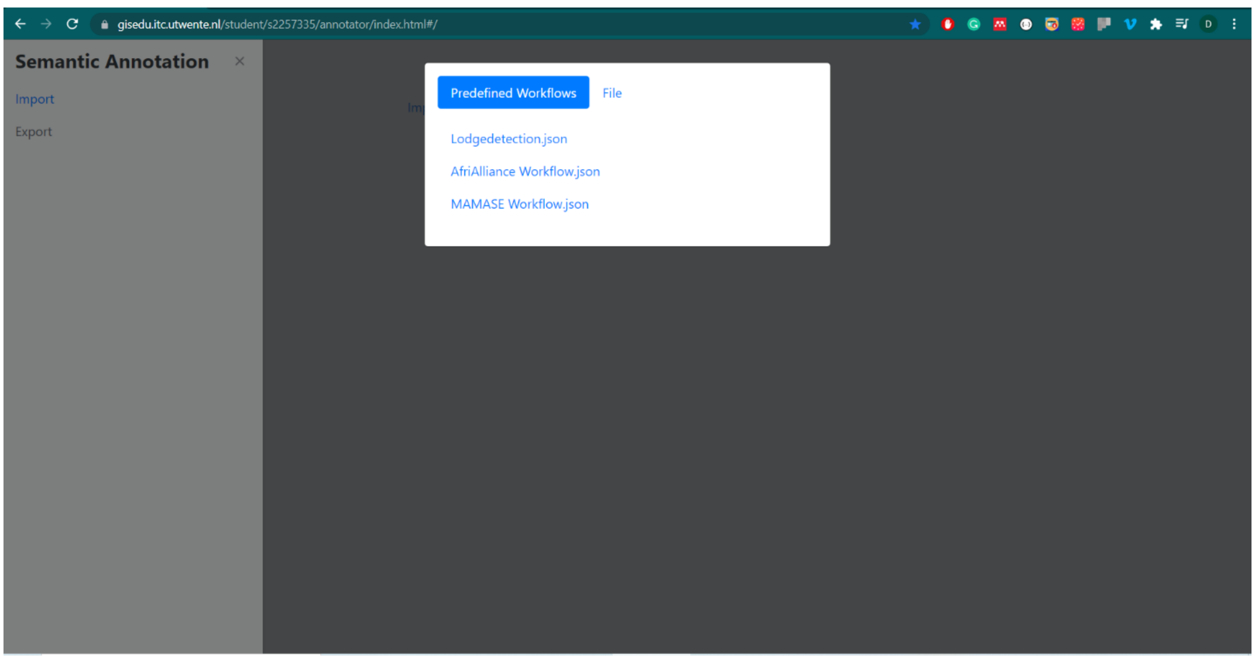


Figure 3A: Modal in the consume workflow page with list of predefined workflows.

After the user selects one of the three listed workflows, he/she will be served with an interface having some components on it as shown in Figure 4A. (Let's assume the user selects Lodgedetection.json workflow).

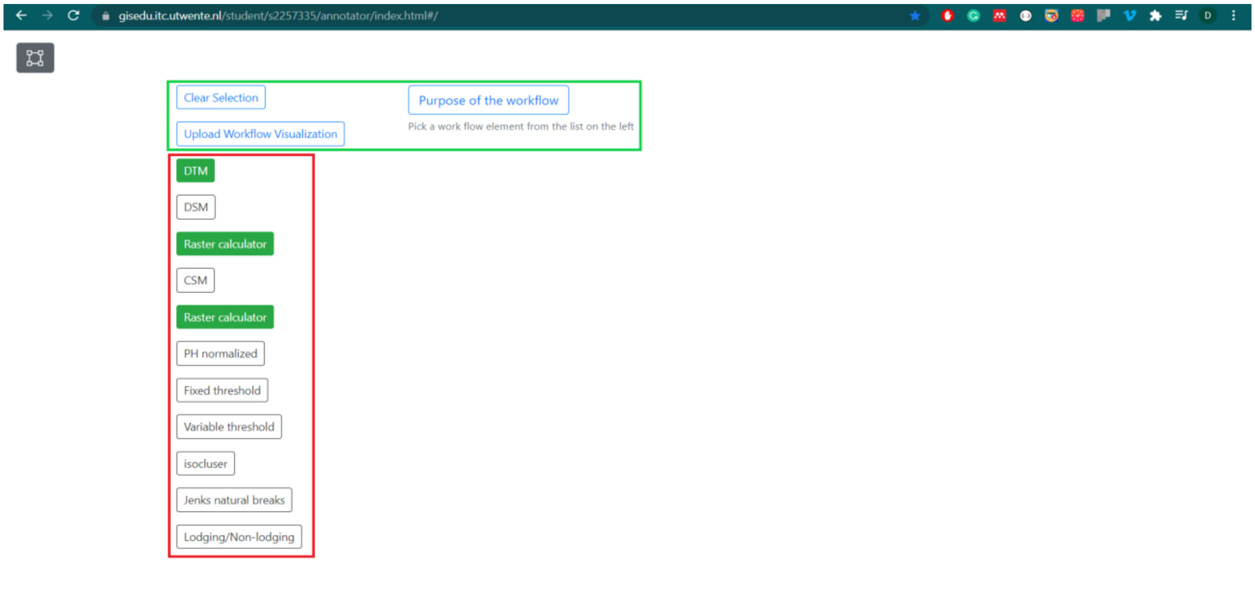


Figure 4A: Extracted workflow components (Red box) and other functionalities (Green box).

The components shown inside the red box are the extracted workflow elements. They are clickable and the user can click to Annotate them with semantic descriptions. The elements shown inside the green box provide different functionalities like seeing the visual representation of the selected workflow by clicking on Upload Workflow Visualization button, understanding the purpose of the workflow by clicking Purpose of the Workflow button, etc.

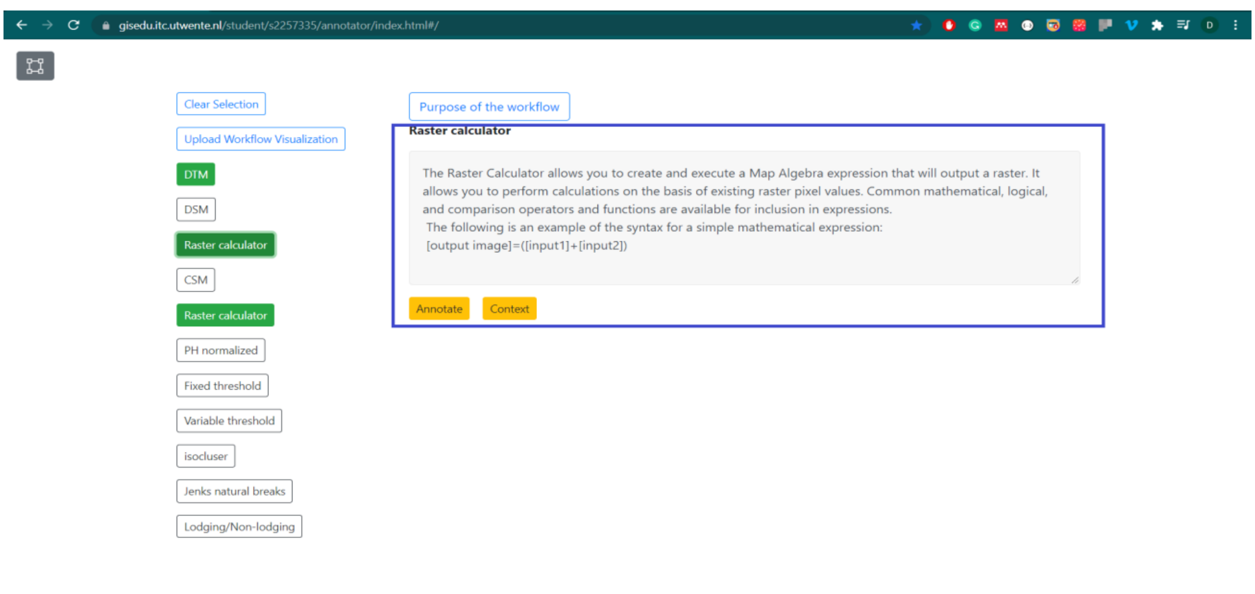


Figure 5A: Semantic description for the selected workflow element (Raster Calculator).

When the user clicks the workflow elements (assume the user clicks ***Raster Calculator***), the components shown in the blue box will be displayed. The text displayed in the text area is the semantic description associated with the selected workflow element which comes from the ontology. It is editable so that the user can modify it with his/her own description. To get additional understanding of the selected workflow element, the user can click the Context button and observe the connection of the selected workflow element with other concepts inside the ontology (LTB) as shown in Figure 6A. When the user clicks the Context button, the embedded LTB window will be displayed. Then the user needs to click on the *Open map* button and write the label of the workflow element in the search bar of the LTB.

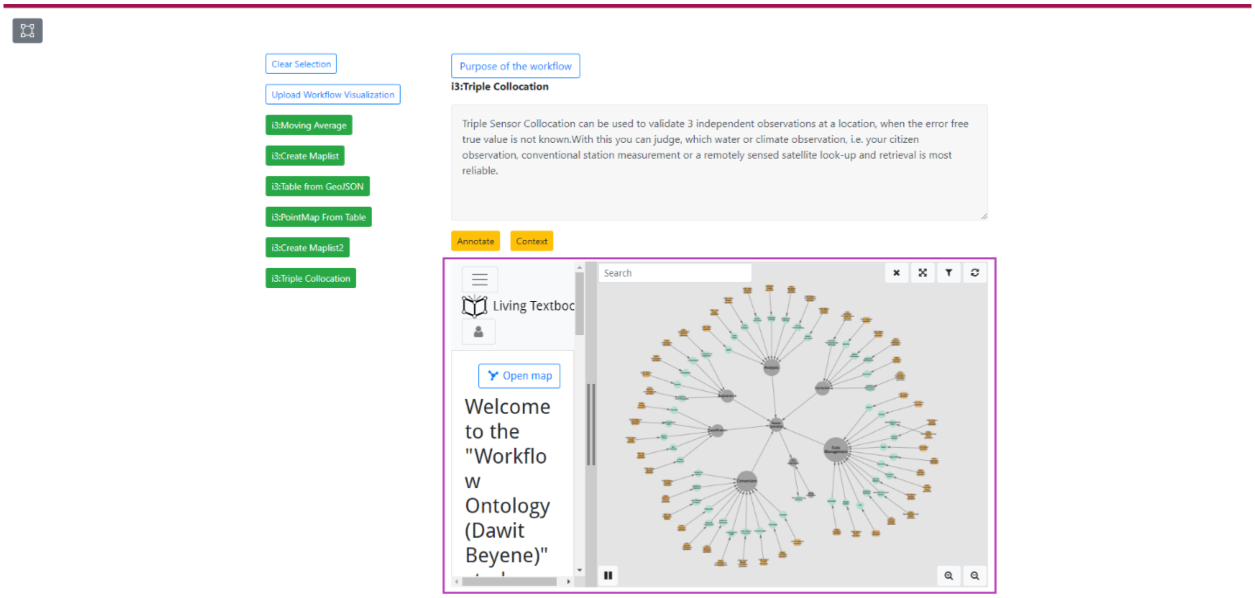


Figure 6A: The visual display of geoprocessing ontology in the Living Textbook (LTB).

Finally, the user can click the Annotate button to embed the semantic description on the element and export the workflow by clicking the Export link from the sidebar.

2. To **Embed Metadata into Workflows**

This functionality can be used by the users who want to embed semantic metadata into workflows for the sake of enhancing reproducibility. To accomplish this task the user needs to have a workflow developed in either Workflowapp or ILWIS and export it as a JSON file. Once the user clicks on Embed Metadata, a new window page with a window modal will be served. The user can continue interacting by clicking the Import link. After clicking the Import link, the white window modal will pop up to allow user Import workflow from the File as shown in Figure 8A.

Figure 7A: Embed Metadata page asking the user to import workflows for semantic enrichment.

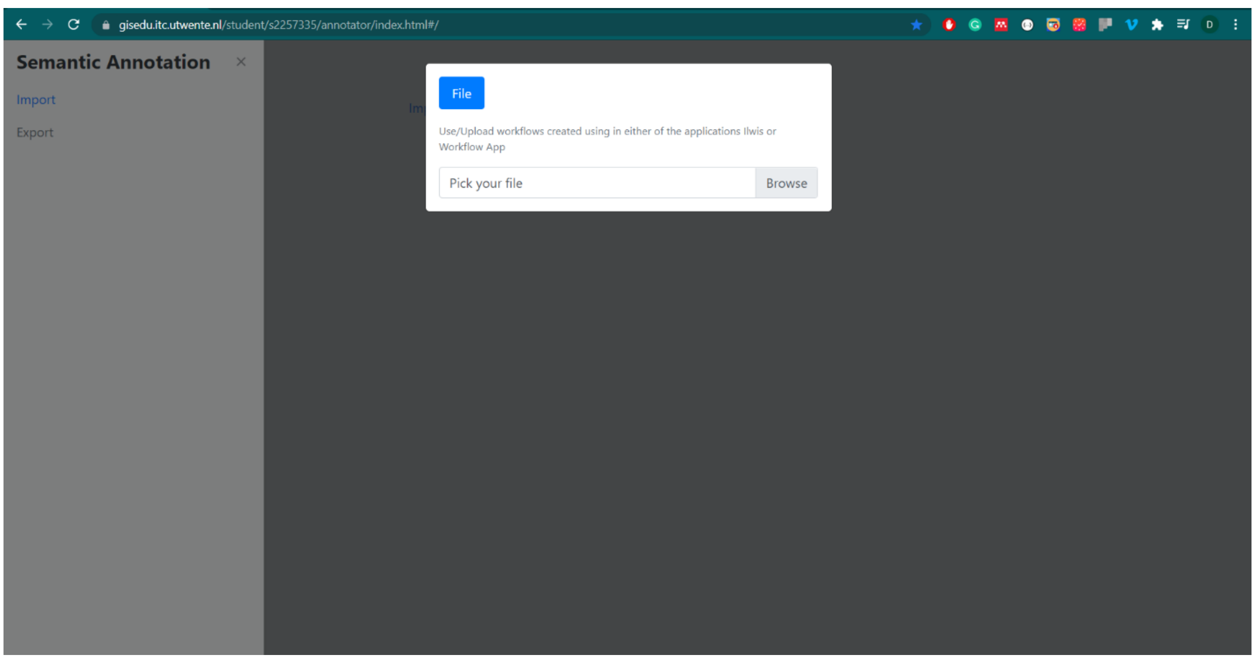


Figure 8A: Modal in the Embed Metadata page prompting users to import workflow.

After importing the workflow, the user can annotate his/her workflow by following the same steps described in Figure 4A, Figure 5A, and Figure 6A. Finally the user can export the semantically enhanced workflow by clicking the Export link from the sidebar.

APPENDIX B:

Usability Test Questionnaires

The profile of participants

Right questions must be asked first to have the most accurate and operating results from any usability test. The questionnaire includes both open and closed-ended questions. The questionnaire is structured in two different segments, the first one is about the profile of the participants shown in Table 1, and the second one is the task-based questions that can be answered while interacting with the prototype application. Knowledge about the profiles of the participants helps to understand who the product is for. It also has the value of further understanding of the potential user groups.

User profile criteria	
Highest level of completed education	The user is expected to have at least a bachelor's degree to understand the area.
Educational Background	It is expected to have Geo-spatial knowledge. Because the workflows that are going to be tested by the user are geoprocessing workflows.
Profession	Expected to be in geo-spatial field.
Experience related to creating, reusing, or reproducing geoprocessing workflows	It can be any. It helps to observe the understanding of experts and non-experts.
Experience with GI software	

This usability testing seeks to assess and verify the usage of the prototype application as a proof of concept. It also aims to test with the potential users. The ultimate goal of such an application is to help users reproduce workflows.

This usability testing has two main sections:

User profile:

This is the general description of the participant.

1. What is your highest completed education level?

   ☐Bachelor

   ☐Masters

   ☐Doctorate

☐Other (Please tell us what)

2. What is your educational background?

   ☐Computer Science

   ☐Geoinformatics

   ☐Geology

   ☐Urban Planning

   ☐Water Resources

   ☐Other (Please tell us what)

3. What is your current occupation?

   ____

4. Do you have experience related to creating, reusing, or reproducing geoprocessing workflows?

   ☐No

   ☐Yes, to a small extent

   ☐Yes, to a great extent

Task-based usability testing:

This section contains questions that will be answered by the user using the prototype application:

Scenario 1 (workflow consumer):

Peter is a junior GIS analyst hired a month ago by Glob-GIS. He conducts an analysis that the company receives from its customers. There are many customers who need to conduct a certain study for themselves. The study results must be obtained as soon a possible by clients. The results of the study were therefor used as an input for their needs. Recently, the company has been tasked with studying how much rice is yielded from the coming year by the local government and how much product they can lose from lodging. To this end, Peter wants to reuse methods to make use of the model which is already tested and time efficient. he searches for an existing method that can make such an analysis shown in Figure 3. His search revealed a method used by the company, but the analyst wants to understand what the workflow does with regard to data use, operations which process the input data, and also what result is expected at the end.

1. Did you get the components of the interface self-explanatory to interact with?

   ☐Yes

   ☐No (Please specify which elements were misleading)

2. Did you find it easy to navigate the page and get the Lodgedetection workflow from the Predefined one?

   ☐Yes

   ☐No (Please specify why)

3. The Upload Workflow Visualization button allows seeing the workflow in visual form. From the point of Peter, do you think it will help him to understand the logic of the workflow?

   ☐Yes

   ☐No (Please specify why)

4. Do the different colors used for *workflow elements* listed on the left side give a clue?

   ☐Yes (can you tell us what)

   ☐No

5. While interacting with the workflow elements listed on the left, when the elements get clicked, three different components are displayed: *Text Area* and two *Buttons* – Annotate and Context.

   a. If the workflow comes with a predefined semantic description, the description will be displayed on the text area for the user to read and understand its usage in the workflow. Do you think it would be helpful to know the semantics of each element to understand the workflow in general much better?

      ☐Yes

      ☐No (Please specify why)

   b. The Context button will provide a visual display of the connection of the selected workflow element with other *Concepts* inside the ontology (LTB). To do that, click the *Context* button then the embedded LTB will be displayed. Then click on Open map and write the exact name of the selected workflow element on the search bar of LTB. If the concept is available in the ontology, it will be displayed in connection with other concepts. If the concept is shown on LTB (check Raster Calculator for Lodgedetection workflow), do you think it can enhance the understanding of the user towards the selected workflow element?

      ☐Yes

      ☐No (Please specify why)

   c. The user can modify the semantic description displayed in the text area. When the user clicks the *Annotate* button, the description will be tagged into the selected workflow element. Export the semantically enhanced workflow by clicking the Export link from the sidebar. Does the exported workflow contain your added semantic description? (Check it by Importing the workflow again and observe the embedded semantic descriptions)

☐Yes

☐No

6. How satisfied are you with using the prototype application to reproduce the existed workflow by understanding its semantics?

   ☐Very satisfied

   ☐Satisfied

   ☐Poorly satisfied

   ☐Unsatisfied

7. Do you have any feedback while using the prototype application?

   ____

Scenario 2 (workflow creator):

A workflow creator is a person who developed an original workflow using available tools. After creating the workflow, to be able to share with other colleagues, the creator wants to enhance the semantics of the workflow for better understandability. Jack created a workflow using the ***Workflow app/ILWIS*** and intends to embed semantic descriptions using the prototype application.

Task: Enhance workflow semantics by embedding semantic descriptions into the workflow you have already created using Workflowapp. Use the prototype application to accomplish this task.

1. In the prototype application, was it possible to find a way where you could embed a semnatic description about the purpose of your workflow?

   ☐Yes (Please tell us how)

   ☐No

2. Using the prototype application, could you able to tag each element of your workflow with semantic descriptions?

   ☐Yes (Please tell us how can it be helpful to tag each element with semantic descriptions)

   ☐No

3. Related with question 2, did you get a predefined semantic description for the elements of your workflow in the ontology (LTB)?

   ☐Yes

   ☐No

4. In the prototype application, did you find a way to save all changes you made into your workflow and be able to find an outlet to export it for later use?

   ☐Yes (Can you please tell us how?)

   ☐No

5. On a scale of 1 to 4, what do you think the use of this prototype application towards enriching workflows with semantic descriptions?

	1	2	3	4	
Worst	○	○	○	○	Best

6. Do you have any feedback while using the application?

   ____

Section 3: Reproducibility questions

Task: In the workflowapp, there are two predefined workflows MAMASE and AfriaAllinace. The same workflows can also be found in the prorotype application. The first task is to test the participants if they can get much better understanding abut those workflows after semantic enrichment. To do this, first, you are asked to go to the workflowapp and see the AfriAlliance workflow. Then after, go to the prototype application and see the same workflow.

The second task s to reproduce the AfriAllinace workflow found in the prototype application using the workflowapp. To do that:

A. Export the AfriAlliance workflow from the prototype application.
B. Save the workflow on your computer.
C. Go to the workflowapp, import the workflow that you previously saved and execute it.

1. On a scale of 1 to 5, how did you rate your understanding level of the AfriAlliance workflow, in the prototype application comparing to the workflowapp?

	1	2	3	4	5	
Diminished	○	○	○	○	○	Enhanced

2. Are you able to execute the AfriAlliance workflow found in the prototype application using the workflowapp?

   ☐Yes (can you observe any difference in the result or in the process while executing?)

   ☐No

APPENDIX C:

List of concepts in Figure 9

1. Geo-processing workflow
    a. GIS operation
        i. Vector Operation
        ii. Raster Operation
            1. Analysis
                a. Binary thresholding
                    i. ESRI: Binary thresholding Operation
                b. CCDC Analysis
                    i. ESRI: CCDC Analysis Operation
                c. Compute Change
                    i. ESRI: Compute Change Operation
                d. Heat Index
                    i. ESRI: Heat Index Operation
                e. NDVI
                    i. ESRI: NDVI Operation
                f. NDVI Colorized
                    i. ESRI: NDVI Colorized
                g. Raster Calculator
                    i. ESRI: Raster Calculator Operation
                    ii. QGIS: Raster Calculator Operation
                h. Weighted Overlay
                    i. ESRI Weighted Overlay Operation
            2. Appearance
                a. Contrast and Brightness
                    i. ESRI: Contrast and Brightness Operation
                b. Convolution
                    i. ESRI: Convolution Operation
                c. Pansharpen
                    i. ESRI: Pansharpen Operation
                d. Statistics and Histogram
                    i. ESRI: Statistics and Histogram Operation
                e. Stretch
                    i. ESRI: Stretch Operation
            3. Classification
                a. Classify
                    i. ESRI: Classify Operation
                b. Linear Spectral
                    i. ESRI: Linear Spectral Operation
                c. ML Classify
                    i. ESRI: ML Classify Operation
                d. Region Grow
                    i. ESRI: Region Grow Operation

   e. Segment Mean Shift
      i. ESRI: Segmen Mean Shift Operation
4. Conversion
   a. Color Model conversion
      i. ESRI: Color Model conversion Operation
   b. Colormap
      i. ESRI: Colormap Operation
   c. Colormap To RGB
      i. ESRI: Colormap To RGB Operation
   d. Complex
      i. ESRI: Complex Operation
   e. Grayscale
      i. ESRI: Grayscale Operation
   f. LAS Dataset To Raster
      i. ESRI: LAS Dataset To Raster Operation
   g. Rasterize Attributes
      i. ESRI: Rasterize Attributes Opeartion
   h. Rasterize Features
      i. ESRI: Rasterize Features Operation
   i. Spectral Conversion
      i. ESRI Spectral Conversion Operation
   j. Terrain To Raster
      i. ESRI: Terrain To Raster Operation
   k. Trend To RGB
      i. ESRI: Trend To RGB Operation
5. Correction
   a. Apparent Reflectance
      i. ESRI: Apparent Reflectance Operation
   b. Geometric
      i. ESRI: Geometric Operation
   c. Radar Calibration
      i. ESRI: Radar Calibration Operation
   d. Sentinel-1 Radiometric Correction
      i. ESRI; Sentinel-1 Radiometric Correction Operation
   e. Sentinel-1 Thermal Noise Removal
      i. ESRI: Sentinel-1 Thermal Noise Removal Operation
   f. Speckle
      i. ESRI: Speckle Operation
6. Data Management
   a. Aggregate
      i. ESRI: Aggregate Operation
   b. Resample
      i. ESRI: Resample Operation
   c. Boundary Clean
      i. ESRI Roundary Clean Operation
   d. Buffered function
      i. ESRI Buffered function Operation
   e. Clip

   i. ESRI Clip Operation

f. Extract Bands
   i. ESRI: Extrat Bands Operation
g. Mask
   i. ESRI: Mask Operation
h. Merge Raster
   i. ESRI: Merge Raster Operation
i. Mosaic Rasters
   i. ESRI Mosaic Rasters Operation
j. Multidimensional Filter
   i. ESRI: Multidimensional Filter Operation
k. Multidimensional Raster
   i. ESRI: Multidimensional Raster Operation
l. Reproject
   i. ESRI: Reproject Operation
m. Shrink
   i. ESRI: Shrink Operation
n. Swath
   i. ESRI: Swath Operation