GIMA
Geographical Information Management and Applications

Workflow Sharing Based on Semantics

A Case Study in Marine Spatial Planning

Magali D. do Patrocínio Gonçalves

Thesis for the Degree of Master of Science

Supervisors

Dr. Ir. Rob Lemmens, Dr. Mahdi Farnaghi

Responsible Professor

Prof. Dr. Menno-Jan Kraak

April, 2024

Workflow Sharing Based on Semantics - A Case Study in Marine Spatial Planning

Thesis for the Degree of Master of Science

Author:	Magali D. do Patrocínio Gonçalves
Student number:	4473655
Date:	March 2024
Programme:	Geographical Information Management and Applications (GIMA)
	Faculty of Geosciences
	Utrecht University (Delft University of Technology, University of Twente, Wageningen University & Research)
Supervision	Dr. Ir. Rob Lemmens, Dr. M. Mahdi Farnaghi, University of Twente
Responsible Professor	Prof. Dr. Menno-Jan Kraak, University of Twente

Acknowledgements

This thesis has been submitted in partial satisfaction of the requirements for the master's degree in Geographical Information Management and Applications (GIMA).

Firstly, I would like to thank my supervisors, Dr. Ir. Rob Lemmens and Dr. Mahdi Farnaghi, for believing in me and all the time dedicated to meetings, reading my work and providing feedback and tips. I extend these words of gratitude to the Responsible Professor, Prof. Dr. Menno-Jan Kraak. Without their assistance, I would not have been able to complete this research.

To the volunteers who were kind enough to spend an hour testing the workflow repository prototype and providing feedback, know that you were also an essential contributor to this thesis. Your kindness and feedback were really appreciated.

I dedicate this work to my co-workers and managers who challenged me to do a master's and my friends who have kept me sane. It was a lot of work, but the knowledge gained was well worth all the sleepless nights. But most importantly, to my family. To my parents, who always inspired me, and to my sister, who could always calm me in times of stress. To my partner for accepting and supporting me in all he could during my journey. Finally, I dedicate this work to my daughters, who were able to adjust at such a young age to a working-studying mom. The time I could not spend with all of you will never return, but be sure that every minute I have spent with you and all the time I hope to get with you is precious.

Summary

Marine Spatial Planning aims to strategically allocate human activities in marine areas, ensuring natural resources, ecological systems, and cultural heritage preservation. In order to achieve this objective, Marine Spatial Planning relies on an extensive collection of geographical data and simulation models. Preparing and analysing this myriad of data for the Marine Spatial Planning process could benefit from automated processes such as workflows. Nevertheless, issues with findability and reusability, such as incomplete metadata, lack of or incomplete provenance tracking of data products, and interoperability (not only synthetical but also semantical), have hindered this development. Therefore, this study uses web semantics and linked data principles to define abstract workflows to enhance findability, reusability, and track provenance. The workflow description includes the workflow metadata and describes the input and output data from each workflow step.

The abstract workflows are described using the Resource Description Framework (RDF) in Turtle format (TTL - Terse RDF Triple Language) since it is easily read by humans and machines, is supported by modern RDF libraries, is faster to read and avoids redundancy, being able to handle more complex graphs.

The proposed workflow description is used to define several workflows used in Marine Spatial Planning and stored in a triplestore; the choice for this database was GraphDB. Triplestore databases need to be queried using SPARQL; few users would know how to use this query language. Therefore, Sparnatural enabled users to create SPARQL queries through a graphic interface, and a webpage with the workflow diagram and metadata was created for each workflow.

In the end, a semantic search engine for a Marine Spatial Planning workflow repository was designed, and a functional prototype was developed with three components: a triplestore containing the workflows described, the SPARQL graphical query builder and the workflow webpages with the workflow diagram and metadata.

The thesis contributes to Marine Spatial Planning (MSP) by introducing a novel semantic web workflow definition that uses linked data principles. The new semantic web workflow definition improves the discoverability of workflows by integrating rich metadata and standardized vocabularies, enabling users to easily locate specific workflows relevant to their needs via semantic searches.

The workflow, workflow purpose and domain ontologies, helped achieve the research goals. It would be great to be able to support the user community in expanding the vocabulary to fit their needs as it evolves.

Future research is needed with a bigger sample of workflows and testers, and a few suggestions are given to improve the user experience. Furthermore, implementing systems that can translate concrete workflows from popular GIS into abstract workflows and vice-versa is in dire need.

Keywords: workflow, interoperability, semantic web, linked data, marine spatial planning

Contents

1. Introduction

Marine Spatial Planning (MSP) aims to strategically allocate human activities in marine areas, ensuring natural resources, ecological systems, and cultural heritage preservation. To achieve this objective, MSP relies on an extensive collection of geographical data and simulation models. These resources are important for engaging with diverse stakeholders and comprehending the intricate interplay of factors within the marine environment. MSP aims to facilitate the decision-making process, effectively balancing the needs of various stakeholders while managing the inherent complexities of marine systems.

The scope of data utilised to support MSP is extensive and diverse, necessitating a range of operations to ready the data for analysis or decision-making support. Nonetheless, most data formats and operations are widely accessible and supported across various Geographical Information Systems (GIS), including open-source alternatives. MSP also benefits from the FAIR (Findable, Accessible, Interoperable and Reusable) data principle. Over the last ten years, several geodata portals dedicated to planning at sea have been developed, often offering open-source data. (e.g. Basemaps[1], EMODnet[2], Copernicus Marine Services[3]).

A substantial volume of data is essential to gather, organise, process, and analyse to support the MSP process. Certain tasks within this process are repetitive and might need to be replicated for different study areas. Experts engaged in data preparation and analysis could significantly save time by employing workflows that facilitate the acquisition of necessary datasets, allowing them to focus on the scientific aspects of their tasks. Workflows originated from business practices and have been devised to enhance efficiency and diminish the likelihood of human errors. In a workflow, the input data, the sequence of operations required, and the data dependencies are clearly defined to achieve a specific output (Goble et al., 2020).

Besides automating step-by-step processes, workflows are also useful for adapting data analysis methods, reproducing and verifying scientific processes (Duffy et al., 2012), and sharing knowledge between experienced and younger professionals (Kechagioglou et al., 2019). Workflows have thus become more popular over the last few decades and are being used in several fields of science, such as life sciences, biodiversity, astronomy, social sciences and geosciences (Goble et al., 2020; Khan et al., 2019), to improve performance, standardisation and reusability (Goble et al., 2020) as well as enable automation, scaling, adaptation and provenance support (Khan et al., 2019).

Furthermore, reusing scalable workflows within the context of Marine Spatial Planning automates geodata processing and potentially contributes to stakeholder engagement and process transparency. In Marine Spatial Planning, effective stakeholder engagement holds significant importance. A well-constructed workflow diagram could serve as a solid foundation for discussing or communicating the rationale behind a decision that was made or needs to be made. Additionally, stakeholders are often interested in the origin of the data presented to them when a decision needs to be made. By tracking data provenance, workflows could support MSP professionals in these exchanges.

[1] https://basemaps.helcom.fi

[2] https://emodnet.ec.europa.eu

[3] https://marine.copernicus.eu

1.1. Problem Description

Leveraging workflows offers numerous advantages; however, they must adhere to specific criteria to fully realise them. Workflows should be easily discoverable, accessible, reusable, and ultimately executable.

Several GIS offer tools to create geodata workflows on the proprietary spectrum (e.g. ArcGIS workflow Manager from Esri[4], FME Workbench from Safe Software[5]) and open source (e.g. graphical modeller in QGIS[6], GeoJModelBuilder[7]). Unfortunately, most of these tools do not allow sharing or searching for existing workflows; in some cases, this is offered within the organisation's environment. A clear disadvantage of these tools is that they are not shareable or reusable outside those environments (Meek et al., 2016). This results in an additional problem: open-source repositories for geodata workflows are still to be developed, and hardly any geodata workflows are registered in general science repositories. Geodata workflows, specifically workflows related to MSP data preparation in workflow repositories, could accelerate capacity building in several non-European countries embracing the Marine Spatial process at a growing pace.

Due to the growing interest in workflows, several workflow management systems (WfMSs) were developed with different approaches to workflow specification and execution (Khan et al., 2019). Workflow repositories have also been established next to WfMSs. The workflows are registered in repositories with a description and respective metadata, which is a good step towards findability. However, as Goble et al. (2020) mentioned, most repositories are dedicated to specific WfMSs, such as KNIME Hub[8] for KNIME or nf-core[9] for Nextflow, and workflows are only executable in the corresponding WfMS. Therefore, searching for workflows considering the many existing repositories is not a straightforward task, and interoperability can still be an issue since a workflow created in one WfMS will not necessarily work in another system.

On the other hand, geosciences can leverage the existence of the Open Geospatial Consortium (OGC[10]) standardised web services for raster and vector geodata (workflowS, WCS and WMS) and Web Processing Services (WPS). Using standardised web services contributes to workflow interoperability but does not solve all the problems with interoperability, findability, and reusability.

Interoperability issues can be synthetic or semantic. Synthetic interoperability issues can be by-passed by using system-agnostic workflow (Garijo et al., 2017; Ohuru, 2019; Rosser et al., 2016; Scheider et al., 2019; Scheider & Ballatore, 2018), common intermediate schemas (Goble et al., 2020; Ohuru, 2019), or standards (Crusoe et al., 2022).

Thus, the issues with findability and reusability, such as incomplete metadata, lack of or incomplete provenance tracking for reuse of products and semantic interoperability, also seen in the MSP context, have gained our attention. Furthermore, the idea of using workflows for MSP has also been explored by Campagna et al. (2014).

[4] www.esri.com

[5] www.safe.com

[6] www.qgis.org

[7] https://github.com/geoprocessing/GeoJModelBuilder

[8] hub.knime.com

[9] https://nf-co.re

[10] www.ogc.org

Effective ways of sharing interoperable executable workflows that can be easily found and understood for reuse by others are still in demand. Web semantics and linked data seem to help with this undertaking. For example, the work of Yue et al. (2016) shows that linked data has been used to share workflow results, data products and processing steps and can help track provenance. Additionally, several studies regarding geodata workflows also mention the potential of web semantics and linked data to improve the findability of data, processes and workflows, but this has yet to be implemented (Ohuru, 2019; Roos et al., 2010; Rowland, 2020; Ubels, 2018).

Marine Spatial Planning has been expanding globally as its advantages become clear, with several countries implementing the process worldwide. It seems logical that the next step for MSP is to arrange ways to share methods of producing data products and preparing data for MSP so it can be easily implemented and automated. An open-source, interoperable geodata repository would greatly develop the practice worldwide and encourage even more countries to implement the process.

This research provides guidelines for designing a geodata workflow repository leveraging web semantics and linked data to enhance the discoverability and reuse of data and workflows. Not to forget that the decision to reuse an existing workflow or to choose one workflow over another implies that users understand and trust this workflow, which can only be achieved by good metadata and clear documentation of the use and definition of the workflow. Additionally, a clear and intuitive representation of workflows can help accomplish this. This study aims to accelerate spatial planning processes in the marine environment, but potentially also on land, by reusing workflows and adapting input datasets and parameters.

1.2. Research Objectives

This research investigates whether web semantics and linked data principles can improve workflows' findability and reusability and support provenance tracking of geodata workflows.

The focus is on abstract workflows and their evolution on the premise that works like the one from Garijo et al. (2017), Kasalica (2022), Ohuru (2019), and Ubels (2018) will enable the conversion of abstract workflows to executable ones. This way, the format is portable with a complete description of the workflow steps, inputs, outputs and metadata, thus contributing to findability and reusability and enabling interoperability.

Therefore, this study aims to define a method to share abstract workflows with complete metadata and clear provenance using web semantics and linked data. Several workflows to prepare data for Marine Spatial Planning are built and registered in a repository prototype to test the hypothesis: Semantic-web representation and search can aid the discoverability and reusability of Maritime Spatial Planning related workflows.

The sub-goals of the research are:

1. To determine the workflow metadata needs to support discoverability and reusability.
2. To define methods to represent workflows by describing the input and (intermediary) output(s) with linked data.
3. Define the requirements for a geodata workflow repository that supports the discoverability and reusability of workflow represented by the method defined in sub-goal two.

The first sub-goal brings us to the following research questions:

1.1. What information about the workflow and workflow data needs to be described and stored?
1.2. What standards support storing and sharing of (workflow) metadata?
1.3. Which linked data format best supports our needs?
1.4. What (if any) extensions to existing web semantics vocabularies are needed to support the workflow data description?

The second sub-goal brings us to the following research questions:

2.1. To what extent can linked data best support the representation of workflows?
2.2. How to store workflows described with our method to support discoverability and reusability?
2.3. Are the workflows described using the proposed method easy to understand?

The third sub-goal brings us to the following research questions:

3.1. What are the workflow repository requirements to store and share workflows described with our method?
3.2. What are the workflow repository requirements to support the findability and reusability of workflows described with our method?
3.3. What are the requirements and functionalities of a user-friendly geodata workflow?

1.3. Scope Limitations

This research aims to contribute to workflow interoperability, findability and reusability of geodata workflows. Nevertheless, only the semantic aspect of workflow interoperability will be tackled.

This study will use existing software and not focus on developing a new repository for external use with a complete web-based user interface. As such, this work focuses on developing a prototype to create clear and concrete guidelines for the future adaptation of such a system. Additionally, due to time limitations, the work will be restricted to a limited number of relevant workflows for the MSP domain.

1.4. Thesis Outline

This thesis comprises five chapters: introduction, literature review, methodology, results, discussion and conclusion.

In the introduction, the reader will get acquainted with the context of the work to be developed and the problem at hand. The research goals and questions are also introduced, along with the work's limitations and the thesis's outline.

The second chapter discusses previous work pertinent to sharing, finding and reusing workflows. An overview of linked data and web semantics is also presented in this section.

In chapter three, the methodology is described, including the proposal for appropriately describing a workflow, the ontologies and vocabulary used, and the approach to testing the prototype.

The results of the prototype tests will be explored in chapter four, and finally, chapter five will present the results discussion and conclusions.

2. Sharing and Reusing FAIR Workflows with Semantic Web

This section reports and discusses previous work developed in workflows, workflow repositories, Semantic Web, and Linked Data relevant to our research.

2.1. Workflows

Workflows, or pipelines, originated from business practices to save time and reduce the chances of human error. Over the decades, workflows have had numerous definitions from different perspectives. However, the Workflow Management Coalition[11] (WfMC) defined it as "The automation of a business process, 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" (Workflow Management Coalition, 1999 p. 8).

The WfMC is responsible for the workflow reference model, which is the basis of most business process management and workflow software management today.

Tick (2006) considers the workflows to have several components, borrowing the definitions from the Workflow Management Coalition (1999):

- Processes: business processes are a set of activities that are needed to accomplish a business goal. A process description includes the activities and resources needed and their logical connections to execute a certain workflow.
- Activities: activities are "a piece of work that forms one logical step within a process" and can be regarded as "the smallest unit of work which is scheduled by a workflow engine during process enactment" (Workflow Management Coalition, 1999, p. 13). Activities can be executed sequentially, in parallel, or conditionally (Workflow Management Coalition, 1999).
- Instances: Instances represent individual executions of the process or activity with associated data, meaning each activity or process can have multiple instances.

Although not defined explicitly by the WfMC, explaining what a resource is in the context of workflows seems relevant. A resource is the component that carries out an instance of a workflow activity or task; it can be a person or a computer program.

There are different perspectives on modelling a workflow. For instance, focusing on the control flow or process perspective, emphasizing which activities occur and in which order, taking dependencies into account (Tick, 2006). The resource or organisation perspective would emphasis documenting the resources needed to execute the several tasks of the workflow, while the task or function perspective would focus on the "elementary operations carried out by the resources while performing a task" (Tick, 2006 p. 330). Another level of detail would be to look at the tools or applications needed to perform the task, Tick (2006) called it the "operation or application perspective". Finally, Tick (2006) also mentioned the "data or information perspective", which focuses on the data that is needed for the workflow execution and its transformations within the workflow.

[11] https://wfmc.org

2.1.1. Workflow Modelling

Workflows have been modelled in many ways over the years, including UML (Unified Modelling Language), BPMN (Business Process Model and Notation), BPEL (Business Processing Execution Language), and CWL (Common Workflow Language).

As the name indicates, UML is a modelling language, it can be used to specify, document, visualise and build artefacts of complex systems, from business modelling to software systems (*Unified Modeling Language*, 2005). It is a flexible language that allows modelling distributed systems running on any middleware at any level of detail (*Unified Modeling Language*, 2005).

BPMN has been widely used in business and science since the release of the first BPMN flowchart in 2014, probably due to the development in agreement between several tool vendors towards a standard in notations for describing business processes (Burattin, 2015) – Figure 1. Nevertheless, BPMN is a conceptual modelling tool and workflows specified using BPMN are not executable, they need to be translated to BPEL to be executable. Business process modelling such as BPMN and BPEL have a wide range of applications in science (Ohuru, 2019).

The Common Workflow Language (CWL), is an open standard to describe workflows using command-line tools (Crusoe et al., 2022; Goble et al., 2020; Khan et al., 2019). CWL is built upon the Command Line Tool Standard, is based on YAML (Yet Another Modelling Language) or JSON-style syntax and has explicit workflow inputs, outputs and documentation. CWL allows to describe workflows that use different software tools which are executable in their own command-line interface by leveraging software containerisation. Because CWL describes the runtime environment and uses software containers, it's a portable format that can be easily reused (Crusoe et al., 2022). Figure 2 shows a CWL workflow diagram.

Furthermore, workflows can be abstract or executable. Abstract workflows are conceptual models representing the logic of the data flows and the order in which processes have to be performed in the workflow. Abstract workflows can be described textually or be represented graphically, for example, using Business Process Model and Notation (BPMN). Executable workflows specify which software will execute each operation in the workflow (Lemmens et al., 2016). This type of workflows is also often called concrete and can be described using Business Processing Execution Language (BPEL) or CWL.

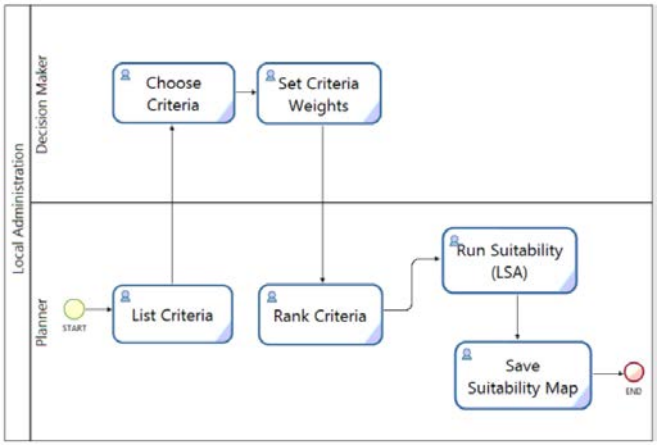


Figure 1 – BPMN workflow diagram for suitability analysis (Campagna et al., 2014)

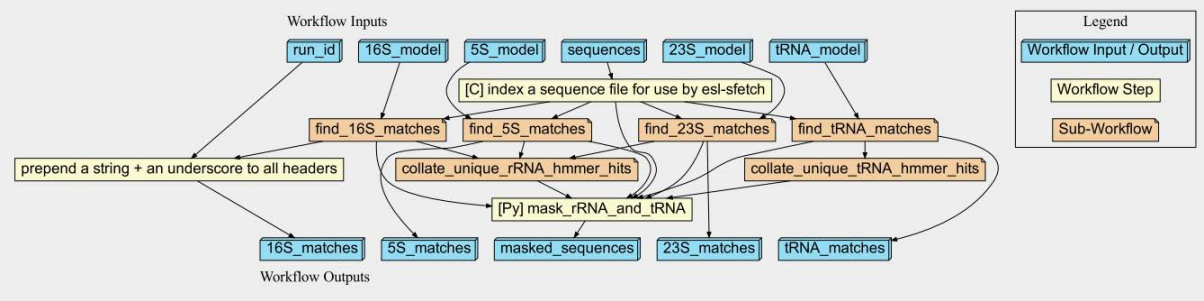


Figure 2 – Excerpt from a large microbiome bioinformatics CWL workflow (Mitchell et al., 2019 in Crusoe et al., 2022).

2.1.2. FAIR Scientific Workflows

workflows are used in multiple science fields to increase performance, standardisation and re-usability (Goble et al., 2020). In computational workflows, the input data, the sequence of operations required, and the data dependencies are defined to achieve a determinate output (Goble et al., 2020). The operations can comprise code or script execution, calling a service or tool, access to a database, the execution of a job on the cloud or the execution of another workflow (Goble et al., 2020). In practice, this results in many workflow systems with incompatibility issues, which deter from combining workflows or reusing them in different settings (Crusoe et al., 2022; Garijo et al., 2017).

Since workflows deal with and create new data, Goble et al. (2020) argue that workflows should follow the FAIR (Findable, Accessible, Interoperable and Reusable) data principles.

Findable

For data to be findable, it must have a rich metadata description with a unique and persistent identifier registered or indexed in a searchable resource and mention the identifier of the data it describes (Wilkinson et al., 2016). In the context of this thesis, we will use the terms findability as an item property and discovery as the action of finding such an item.

Accessible

For something to be accessible, (meta)data must be retrievable by their identifier using an open, free, universally implementable standardised communications protocol that allows for an authentication and authorization procedure, where necessary. In addition, the metadata needs to be accessible, even when the data are no longer available (Wilkinson et al., 2016).

Interoperable

Wilkinson et al. (2016) argue that data is interoperable when "(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation", "use vocabularies that follow FAIR principles", and "include qualified references to other (meta)data".

Khan et al. (2019) distinguish between syntactic and semantic interoperability. The exchange of information between two components using a common data format is referred to as syntactic interoperability, while semantic interoperability refers to the context of the information exchanged.

Reusable

Reusability is linked to the concepts of reproducibility and replicability. In science, reproducibility means that the same methods applied by independent scientists on the same data reproduce the same results (Ostermann & Granell, 2017). Replicability means similar results can be achieved by applying similar methods to the same type of data (Ostermann & Granell, 2017).

Additionally, a workflow can be repurposed, using the full workflow or just parts of it or using different input data to achieve a different result than the one from the original workflow (Garijo et al., 2017; Goble et al., 2020; Wroe et al., 2007). For the context of this thesis, we will use reusability as the ability to reproduce, replicate and repurpose a workflow.

For scientific experiences to be reproduced or replicated, scientists need detailed instructions regarding materials to use and procedures to perform. For a workflow, reusability means that all processes and inputs are fully described. Thus, reusability requires that the "meta(data) are richly described with a plurality of accurate and relevant attributes", "released with a clear and accessible data usage license", "associated with detailed provenance", and "meet domain-relevant community standards" (Wilkinson et al., 2016 p.4), for all inputs and operations.

FAIR Workflows

The FAIR principles of data can be translated for workflows if one also considers some particularities of workflows as process objects: structure, forms, versioning, executability, and reuse (Goble et al., 2020).

Structure refers to the fact that a workflow can be composed of nested workflows or sub-workflows, and thus, FAIR principles would need to be applied simultaneously on multiple levels. The data and tools used in the workflow must comply with FAIR principles for the workflow to be compliant as a whole (Goble et al., 2020). By forms, Goble et al. (2020) mean the workflow's specification level (abstract, ready to run with example data or executable), different requirements may apply to different forms of workflow.

Moreover, workflows can be changed or merged, need maintenance and updates, and eventually become deprecated, resulting in a new version. Tracking these changes with unique identifiers (versioning) is needed (Goble et al., 2020). Table 1 summarizes the FAIR criteria for workflows as a conclusion of this section.

Findability	Detailed metadata description regarding different components (data flowing in the workflow, operations, workflow instances, workflow versions). Unique and persistent identifier registered or indexed.
Accessibility	Components must be retrievable by their identifier using an open, free, universally implementable standardised communications protocol that allows for an authentication and authorization procedure, where necessary. Metadata must be accessible, even when the data or workflow are no longer available. Forms (Abstract, executable, ready to run or executed with input and output data).
Interoperability	Components use a formal, accessible, shared, and broadly applicable language for knowledge representation. Dependencies, reproducible state and associated persistent identifier. Can be operated by another system than the one in which the program was created without requiring major rework.
Reusability	Processes and inputs are fully described. Usage license. Version control and associated metadata.

Table 1 – Summary of FAIR criteria for workflows based on the literature review.

2.1.3. Metadata

The principles of FAIR data make the importance of metadata and provenance clear.

Metadata is generally defined as data about the data. Leipzig et al. (2021) mention three types of metadata standards that reveal what metadata entails: descriptive, administrative and structural. The descriptive metadata supports a resource's discovery and general assessment by including the format, content and creator. The administrative metadata supports technical and operational aspects of resource use; it includes preservation, technical and rights information. Finally, structural metadata supports "the linking among the components of a resource so that it can be fully understood" (Leipzig et al., 2021, p.4).

As findability is directly linked to metadata, efforts are being made to automate its production (Giuliani et al., 2016; Trilles et al., 2012), and several metadata standards have been defined (Brodeur et al., 2019).

According to the W3C (2013), provenance is "information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability, or trustworthiness." (p.1). Thus, information about the provenance of a dataset or workflow will be crucial in the decision to reuse such data or workflow.

More specifically, Jiang et al. (2017) defined provenance as the answers to the typical questions of “what”, “when”, “who”, “how”, “where”, and “why” of the generation of data. Workflows have been commonly referenced as a way to track the provenance of data since they document the input and the operations carried out to reach a certain output (Khan et al., 2019; Scheider & Ballatore, 2018; Zhang et al., 2020).

Khan et al. (2019) refer to three provenance types: retrospective provenance, prospective provenance and workflow evolution. Retrospective provenance supports reproducibility, detailing the tools used to execute each process. Prospective provenance corresponds to the abstract definition of workflow, in which the computational tasks and their order in the workflow are defined. Finally, the workflow evolution tracks any variant of workflows created by altering the original one (Khan et al., 2019).

Provenance can mean different things according to the context, which also holds for the workflow type: abstract or concrete. Abstract workflows will support only prospective provenance, while executable workflows can have prospective and retrospective provenance.

Closa et al. (2019) mention that provenance metadata is essential for reuse and reproducibility, while Leipzig et al. (2021) explain its importance for discovery and validation.

2.1.4. Standards

It has been argued above that metadata and standards are important in enabling FAIR data and workflows. Since metadata provides context and provenance (Leipzig et al., 2021), this section will focus on the different standards for metadata and provenance, both for geodata and workflows. Standards related to Linked Data will be covered under Section 2.2.

Metadata

Leipzig et al. (2021) mention the need for metadata at all component levels of a typical scientific computational analysis workflow: input data, tools used, reports produced, pipelines and publications. They state that "clearly no single metadata standard can support all aspects of the analytics stack" (Leipzig et al., 2021, p.17). Quite the opposite, several metadata standards have been defined over the years (Brodeur et al., 2019). The most relevant standards for geodata are ISO 19115-1:2014 and ISO 19139-1:2019, which were made mandatory in the European Community by the INSPIRE Directive[12]. The Data Catalog Vocabulary (DCAT[13]), a W3C recommendation to enhance data interoperability across data catalogues, is also relevant.

For this research, it is also relevant to mention RO-Crate[14]. Although RO-Crate is not a standard *per se*, its metadata specification, based on well-established standards (Schema.org and JSON-LD), offers a lightweight approach to packaging research data with their metadata (Soiland-Reyes et al., 2022). This approach applies to any resources, from single files to complex workflows, documenting each component. Research data can comprise anything from a simple data folder to a data-intensive computational research environment.

Provenance

W3C PROV[15] is a standard for provenance representation on the web; it includes three main concepts: entities, activities and agents. In PROV, the entities are the things we want to track the provenance of; they can be physical, digital, or conceptual, like a webpage, a chart, or a spellchecker (Figure 3). In the context of workflows, the activities refer to the actions and processes, and the agents correspond to the concept of resources from WfMS. W3C PROV standard targets general applications but is also used in geosciences and workflow applications (e.g. (Masó et al., 2015; Zhang, Jiang, Zhao, et al., 2020). PROV has also been extended to cater to workflow provenance through D-PROV (Missier et al., 2013) and ProvONE[16]. D-PROV and ProvONE both focus on retrospective and prospective provenance.

[12] https://inspire.ec.europa.eu/theme

[13] https://www.w3.org/TR/vocab-dcat-2

[14] https://www.researchobject.org/ro-crate/1.1

[15] https://www.w3.org/TR/prov-overview

[16] https://jenkins-1.dataone.org/jenkins/view/Documentation%20Projects/job/ProvONE-Documenta tion - trunk/ws/provenance/ProvONE/v1/provone.html

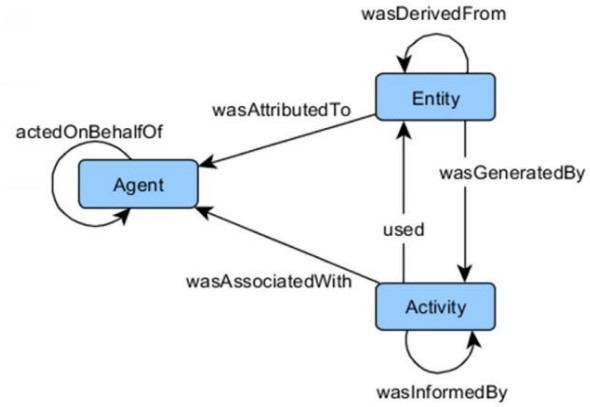


Figure 3 – PROV overview (Garijo et al., 2017).

The Open Provenance Model for Workflow (OPMW[17]) was developed specifically for scientific workflows (Garijo et al., 2017). OPMW was based on the Open Provenance Model (OPM[18]), which aimed to describe the generic provenance of any general thing, created by computer systems or not, in a system-agnostic and precise way (Garijo et al., 2017). OPM only accounted for one execution of the workflow through PROV-P. With the OPMW extension, it is possible to describe abstract workflows and link them to their execution while keeping track of the data flowing in the workflow and the parameters used in each operation (Garijo et al., 2017). Nevertheless, the parameters and values used are stored in the retrospective part of the model. In OPWM, the link between the data that comes in or out of a workflow step is only made by ordering the steps in the workflow, making the workflow description focused on its steps or operations.

CWL also has its provenance model CWLProv[19], which was developed to track the provenance of CWL workflow runs captured as a research object using Linked Data standards. A Research Object includes all resources involved in a given workflow execution, including the workflow inputs and outputs, as a bundle of documents that can be cited as a whole (Goble et al., 2020). CWL focuses on executable workflows, which is not the aim of this research study.

For the goals of this research, we are interested in a workflow model that can describe abstract workflows (prospective provenance) and can be linked to multiple executions of the same abstract workflow (retrospective provenance). Additionally, we are interested in tracking the parameters and their values even at the abstract level, which is not provided by any of the provenance models referred. Furthermore, we are interested in keeping track of the evolution of workflows.

Workflow Standards

Despite efforts (e.g. BPMN, CWL), no globalised standards have been implemented for scientific workflow representation (Garijo et al., 2017; Kanwal et al., 2017).

A workflow standard would have to be system agnostic and thus comprise the definition of an abstract workflow that could be translated to an executable workflow once a WfMS is chosen for its execution.

[17] https://www.opmw.org

[18] https://openprovenance.org/opm/model/opmo

[19] https://github.com/common-workflow-language/cwlprov

2.1.5. Sharing Workflows

There are several ways to share a workflow. One can save the workflow and share the file(s) via email or other communication means, make it available in the cloud protected by username and password, or load it in a workflow repository. In a workflow repository, there might be options to share the workflow with the general public or specific groups, and there might also be options to limit workflow reuse.

As mentioned in Section 2.1.1, workflows can be abstract or executable, and a choice needs to be made on which form to share the workflow so that other people can find, access, consult, understand and reuse, eventually executing it. Additionally, workflow creators need to be aware of the possibility of not sharing the data used as input for the workflow, depending on the license, sensibility and privacy issues. On top of that, they can share the details of the workflow execution(s), mentioning all the tools used for each step and each step's duration, and opt for sharing all the step's results or just the final output.

Some studies focused on workflow interoperability and translated their workflow into abstract ones using linked data and semantic web so anyone could reuse them regardless of the platform (Garijo et al., 2017; Scheider & Ballatore, 2018; Ubels, 2018). Nevertheless, not every platform has its operations and processed tagged with web semantics, so it is not guaranteed that this will work for all platforms. Ohuru (2019) used web services to chain processes defining a workflow based on a standard workflow interchange schema based on JSON data format. This approach requires the availability of an environment that can call all the web services necessary.

Workflow standards such as CWL could also contribute to more FAIR workflows; nevertheless, even if they are well accepted and have increasing usage, they are not universally used. Sharing a workflow in this format still does not mean anyone would be able to use it since not all WfMS enable the translation of CWL into their native language (Crusoe et al., 2022; Goble et al., 2020; Khan et al., 2019). CWL uses the so-called "configuration-based framework, " usually based on XML or JSON configuration files, to convey abstract workflows (Leipzig et al., 2021). Khan et al. (2019) used CWLProv to share an interoperable representation of CWL workflow instance that is executable using research objects.

Belhajjame et al. (2015) proposed preserving the workflow specification and its understandability, reusability and reproducibility using a workflow-centric approach to Research Objects.

The scientific community increasingly uses workflow repositories to share data and experiment results, which is essential to workflow reuse. The user's choice of repository to share a workflow may be influenced by the system used to build the workflow or by the sort of community supported by the workflow repository.

In some workflow repositories, DOIs (Digital Object Identifiers) can be assigned to workflows to reference workflows in scientific publications. DOIs can be important in sharing workflow resources as they ensure persistence and proper credit through citation (Garijo et al., 2017; Goble et al., 2020). However, in data repositories, every resource's description and the relationships between the rest of the workflow components would have to be manually added; thus, they are often absent (Garijo et al., 2017).

The amount of existing WfMS and workflow repositories is daunting; to have a glimpse of the existing systems, we can refer to "A curated list of awesome pipeline toolkits"[20] and the list of CWL workflow systems in Amstutz et al. (2023). For this research, the focus was on repositories that had "spatial" or "geo" in their name or URL. Additionally, systems mentioned in the reviewed literature or notably recognised were analysed to retrieve good practices and examples of general workflow tools.

Workflow repositories such as myExperiment[21] and Crowdlabs[22] allow workflows to be shared in their native language, no matter the system where they were created. Nevertheless, they often miss a link to their executions or instances, so they need to be manually documented to address this. Some repositories have started to adopt Research Objects (RO), which bundle together all the resources used in scientific experiments. However, these are currently in the early stages of adoption, relying heavily on user manual curation (Garijo et al., 2017; Goble et al., 2020).

The analysis of six workflow repositories and WfMSs in light of FAIR principles follows. In general, the tools that allow searching, editing and sharing, along with good metadata description, have more potential to be useful for users, as there is always a learning curve for each tool. When all functionalities are available in the same tool, it saves the user time, learning how to learn with different tools and the burden of file management and switching tools constantly. One of the observations is that there are not many geodata workflows registered in any of the platforms. Additionally, when a search retrieves many workflows, narrowing down the search is difficult due to the few filtering options presented in most repositories.

A table with the complete overview of the workflow repositories consulted in this study is presented in Appendix I. The overview comprises aspects regarding workflow FAIRness and points for improvement. A more detailed explanation follows per workflow repository.

KNIME

KNIME[23] is a general and versatile analytics platform that allows users to create and execute workflows (workflows). These workflows can be built within the platform or imported from the KNIME-Hub[24] repository. KNIME-Hub offers a collection of workflow examples suitable for beginners and workflows shared by the community for easy reuse. Furthermore, KNIME supports the creation of components, which are groups of nodes (operations) that can be used within other workflows or even chained together to form complete workflows.

The platform offers a graphical interface with drag-and-drop options to create workflows with a clear workflow diagram linking inputs, operations and outputs (Figure 4). Each workflow and node can be annotated, providing insights into each node's purpose. KNIME also offers real-time status monitoring during execution.

One notable feature of KNIME is its extensibility, with a range of extensions available, including those created by the community and in partnership with organizations like Harvard's Centre for Geographic Analysis. For instance, a geospatial analytics extension called "Geospatial Analytics for KNIME" has been released. However, the extension has limitations regarding supported input data formats and clarity in certain operations.

[20] https://github.com/pditommaso/awesome-pipeline

[21] https://www.myexperiment.org/home

[22] http://www.crowdlabs.org

[23] https://www.knime.com

[24] https://hub.knime.com

Searching for workflows, nodes, components, and extensions in the KNIME Community Hub is possible, but some limitations exist. Users can filter their searches based on types (workflow, components, nodes, extensions) and tags, but these tags are user-defined and may vary in spelling and synonyms. The search functionality could benefit from improvements, such as keyword combinations using operators like "+" or "AND" and the ability to exclude specific words with "-" or "\" signs. A search feature that helps find nodes or workflows based on specific input data or data types would enhance the platform's usability.

A crucial aspect of KNIME Hub is the description and provenance of nodes, components, and workflows, which help users make informed decisions. These are free-text fields with no mandatory requirements, allowing users to share information at their discretion. While this flexibility is valuable, enhancing descriptions with web semantics and linked data logic could significantly improve findability.

Regarding workflow reuse, users must save a copy in their local files. This prevents unauthorized changes to the original workflow but does not guarantee proper documentation or prevent users from sharing the workflow as if they were the original author. This raises ethical concerns and highlights the need for additional features to address these issues.

In summary, KNIME is a powerful analytics platform with several strengths and opportunities for improvement. Enhancements in search functionality, description standardization, and workflow suggestions can make it even more user-friendly and efficient.

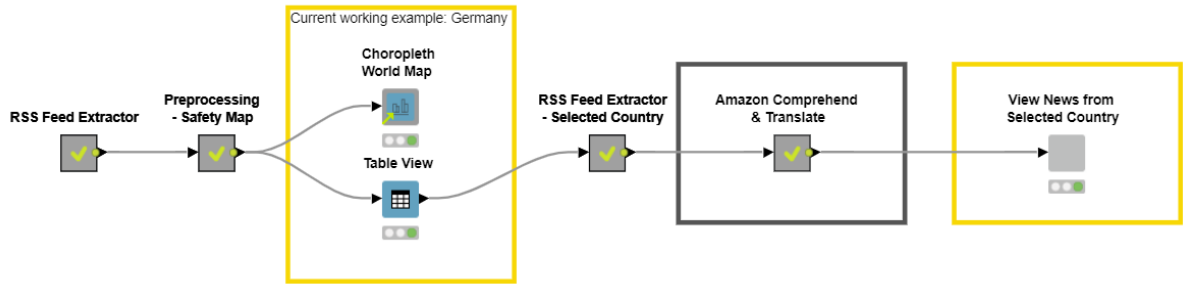


Figure 4 – Example workflow in Knime, retrieved from KnimeHub.

Pegasus

Pegasus[25] is a WfMS designed to optimize workflow execution by distributing tasks across multiple computational resources. This system converts abstract workflows in YAML format into executable workflows. However, users must provide the workflow in YAML format and a file listing the required execution environments and data inputs. Supported APIs include Python, Java, and R (*Pegasus Team*, 2019).

Abstract workflows offer portability across execution environments and enable optimizations at compile and runtime (Deelman et al., 2015). Pegasus can also transform CWL workflows into their format with minimal user intervention (Crusoe et al., 2022; *Pegasus Team*,2019). The translation from an abstract workflow to an executable one is performed by the Pegasus Mapper subsystem, which locates data, software, and computational resources for execution, along with optimizing performance and generating provenance information. Task logs and performance data are monitored by the Pegasus system (Deelman et al., 2015). It is possible to incorporate workflows into other workflows using them as nodes. Executable workflows in Pegasus use the DAG format (Directed Acyclic Graph); the nodes represent the jobs to be executed, and the edges represent the data and control flow dependencies (Deelman et al., 2015) (Figure 5).

[25] https://pegasus.isi.edu

The PegasusHub[26], Pegasus' repository, provides access to workflows but has limited content. The search box appears ineffective, while the tags above the workflow list work as expected. Users share data about the workflows, and PegasusHub enforces certain requirements before a workflow is released, including proper documentation and a pegasushub.yml file containing metadata. Workflows typically include descriptions, release information, Pegasus version, dependencies, license, GitHub repository links, tags, prerequisites, execution instructions, and workflow diagrams (*Pegasus Team*, 2019).

Registering a workflow in PegasusHub is straightforward, involving forking the repository and adding a YAML file with the organisation's and repository's name (*Pegasus Team*, 2019).

Although Pegasus primarily targets the scientific community, projects like HubZero[27] and Wings[28] aim to make it more accessible. Wings is a workflow system that uses semantic reasoning to generate Pegasus workflows based on ontologies that describe workflow templates, data types and components (Deelman et al., 2015; Gil et al., 2007). It also supports execution provenance recording as linked data using OPMW (Garijo et al., 2017).

A graphical interface and an executable installation file for Pegasus would enhance user-friendliness. Improvements to PegasusHub could include enhancing the search functionality and enabling searches and filters in various fields, such as title, topics, license, description, dependencies, and job (activities) labels. Providing job descriptions, rather than just labels, would make workflows easier to understand for non-technical users.

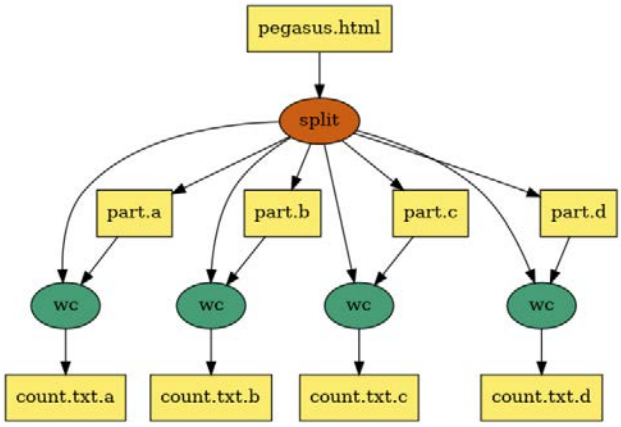


Figure 5 – Example workflow in PegasusHub, retrieved from PegasusHub (https://pegasushub.io).

[26] https://pegasushub.io

[27] https://hubzero.org

[28] https://www.wings-workflows.org

Taverna

Taverna[29] allows the design and execution of data-driven workflows, combining local tools or distributed web services. The core edition is suitable for scientific workflows in any domain, with REST or SOAP web services and command line tools. There are also domain-specific editions which support specific service types and provide examples of workflows and services in the specific domain (*Taverna*, n.d.). Taverna also allows tracking provenance by recording service calls, intermediate and final results from workflows in the OPM format and PROV model (Wolstencroft et al., 2013).

Taverna users can search existing workflows in myExperiment[30], a systems-agnostic workflow repository. myExperiment supports workflows in their native language from multiple systems, including Taverna, RapidMiner, Galaxy, and Knime. Users can download the workflows as their native formats or as Taverna or Galaxy workflows (Goble et al., 2010).

The tracking of prospective provenance and workflow evolution are supported in myExperiment, while Taverna registers the retrospective provenance using OPM and PROV.

Taverna was first developed for scientific research in bioinformatics; nowadays, support for geodata workflows can be achieved using REST or SOAP web services and command-line tools.

Queries can be made at workflow, files, users, groups or pack levels (combinations are possible), and the results are shown per level. The search history is kept, and users can tag workflows as favourites. The web browser version of myExperiment also offers filtering of results by file type, tags (domains or others), users and licence.

In myExperiment, it is possible to see attributions in other workflows or files, featured in packs, users that marked components as favourites and statistics of viewings and downloads, which usually indicates credibility. An example of a workflow diagram is presented in Figure 6.

Using controlled tags from a semantic vocabulary could improve the workflow findability in this platform.

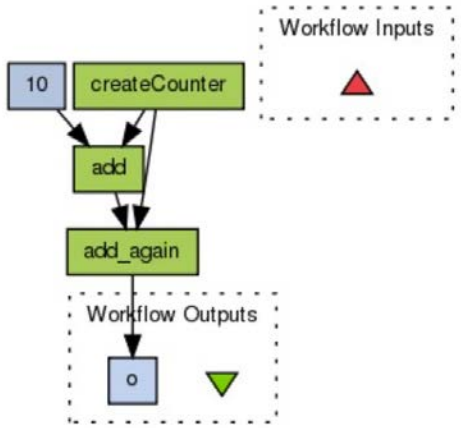


Figure 6 – Example of a Taverna workflow diagram in myExperiment, retrieved from myExperiment (https://www.myexperiment.org).

[29] http://www.taverna.org.uk

[30] https://www.myexperiment.org

Geoweaver

Geoweaver[31] is an open-source, in-browser Python-based software that enables less savvy users to compose and execute data processing workflows using online spatial data services, high-performance computation platforms and open-source deep learning libraries (Sun et al., 2020). Additionally, being a code repository and having the ability to manage servers also enables workflow composition and history recording. All the runs from workflows are saved along the code files supporting provenance. The fact that Geoweaver is an easy-to-use tool, a community effort and that the research's results can be later incorporated into the software is an incentive to use the tool.

Geoweaver allows saving the workflow and its execution with all its files, loading and sharing workflows saved on local files, and searching workflows from others. Nevertheless, users need access to the servers where the workflow is located, and its access may not be public. Recently, Geoweaver started to support conversion to CWL (Amruta et al. 2023).

The metadata documentation is poor; it only shows the workflow ID, the name and the description of the workflow. There is no workflow diagram, making it difficult to understand the workflow steps and evaluate its fit for purpose. There is also no workflow version tracking.

Galaxy

Galaxy[32] was developed to provide open access for research scientists without computer programming experience. It supports prospective and retrospective provenance, although it is unclear whether it follows any standard for provenance or metadata. Galaxy has a particular feature: users can run a set of tools and then decide to turn their analysis history into a workflow. Figure 7 depicts an example of a workflow diagram in Galaxy. While setting up a workflow, Galaxy will not allow connecting nodes if it does not make sense, for instance, connecting an output format incompatible with the input required in the next node.

Users looking to reuse a workflow can search using text and use the advanced options to filter names or special tags. If they find a potentially suitable workflow, users can view, edit, or run it. Galaxy will keep version control of the workflow in the environment each time changes are saved; it is up to the user to decide if and when to share it, with the option of adding itself as a contributor, updating the metadata and the license under which the workflow is shared (provided it respects the initial license).

A big downside of Galaxy is that the metadata available while browsing workflows is very restricted; there is no workflow description or diagram. These details are only visible when clicking on an eye icon, meaning users must click on all workflow results to analyse them.

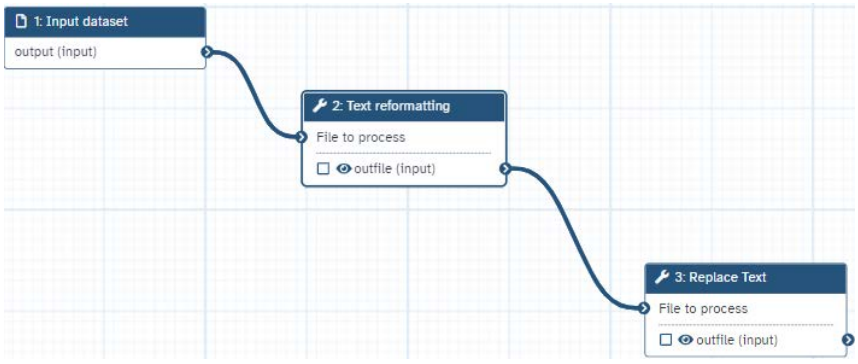


Figure 7 – Example of a Galaxy workflow diagram, retrieved from the Galaxy repository (public example workflow by a user registered as "mep").

[31] https://esipfed.github.io/Geoweaver

[32] https://galaxyproject.org

WorkflowHub project

The WorkflowHub[33] is a FAIR workflow registry which is system agnostic; users can register their workflows independently of the workflow system they used to create it, automatically creating an RO-Crate register. Workflow authors can choose the type of licence of their workflows (metadata only, full workflow, and edit privileges). It allows for workflows to be cited and keep versioning control, supporting all three types of provenance.

The fact that each component is assigned an identifier makes the workflow a FAIR digital object at every level. Additionally, workflow authors can assign DOIs to a workflow when registering it.

Users can browse existing workflows using keywords, filters, and tags, and they only need to register if they want to download or edit workflows.

Although there are specific rules for tags, for someone just browsing through so many workflows, a hierarchy of themes would make it easier for users to find relevant workflow categories.

2.2. Semantic Web

Web semantics, or the Semantic Web, is a broader vision and principles for organizing and structuring web content to make it more meaningful to humans and machines. The Semantic Web aims to add a layer of meaning to web resources so that software applications can understand and process them. This is achieved through the use of standardized data representation and metadata.

Linked Data is the practical implementation of Semantic Web principles that involves publishing data in RDF format, using standardized URIs, and interlinking datasets to create a web of connected, machine-understandable data.

RDF (Resource Description Framework) is a standard and key technology within the Semantic Web framework. RDF describes resources as 3-tuples: an ordered list of three elements called a triple. An RDF triple consists of the subject, the object, and the relationship between them - the predicate. These triples can express any information and form the foundation for representing and linking data on the Semantic Web. Data described as triples can be visualised in RDF graphs such as Figure 8.

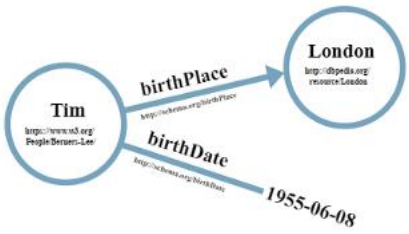


Figure 8 – An RDF graph example (retrieved from https://ontola.io/blog/what-is-linked-data)

[33] https://workflowhub.eu

RDF is also considered a standard since it provides a consistent way to represent and describe data in a machine-readable and human-understandable format. RDF files are commonly stored in 4 formats (Addlesee, 2018):

- N-Triples (.nt) - easy to parse and supported by most modern linked data databases and RDF libraries with high compression ratios;
- JSON-LD (.json) - can be manipulated by JSON libraries, but is not supported by newer linked data databases;
- RDF/XML (.rdf) - good to work with XML but not supported by newer linked data databases;
- Turtle (TTL – Terse RDF Triple Language) is similar to RDF but more human readable,while being supported by modern RDF libraries; also, data can be streamed in blocks as there are no start and end lines (unlike RDF/XML), good for data streaming or API use.

Linked Data is a set of best practices for publishing and interlinking data on the web using RDF and standardized URIs (Uniform Resource Identifiers) (Roos et al., 2010). The principles of Linked Data encourage the use of RDF to create structured, interlinked datasets, making it easier to connect and navigate between different data sources.

The linked data principles were defined by Tim Berners-Lee, the inventor of the internet and the initiator of the Linked Data project as such (Berners-Lee, 2006):

- Use URIs to name things;
- Use HTTP URIs so that things can be referred to and looked up ("dereferenced") by people and user agents;
- Provide useful information in the URIs using open web standards (e.g. RDF, SPARQL);
- Include links to related things using their URIs when publishing on the Web.

Besides the linked data principles, mentioning the Data on the Web Best Practices by W3C (Farias Lóscio et al., 2017) is important. There are too many principles to list and explain, so only a few are mentioned here due to their importance, which has already been stated before in the context of this thesis's goals. One of the most important principles is providing complete metadata and licence information for potential users to understand the data and access whether to use it. The use of persistent URIs for dataset identification and providing version information ensures that data is always citable through time, even if the data is no longer available. Reuse vocabularies, preferably standardised ones, to reduce redundancy, improve community consensus and understanding, and contribute to interoperability. Finally, another important practice is to cite the original publication to give credit and be transparent about the source of the data.

In 2010, Berners-Lee also devised a five-star ranking system to encourage good Linked Open Data (LOD), linked data shared under an open licence such as the Creative Commons CC_BY[34] as in Table 2.

No.of starts	Requirements
1	Available on the web (whatever format) but with an open licence, to be Open Data
2	As in no. 1 and also available as machine-readable structured data (e.g. excel instead of an image scan of a table)
3	As in no. 1 and 2 plus non-proprietary format (e.g. CSV instead of excel)
4	All the above plus using open standards from W3C (RDF and SPARQL) to identify things so that people can point at your objects
5	All the above, plus linking the data to other people's data to provide context

Table 2 - Good Linked Open Data Requirements (Berners-Lee, 2006).

[34] https://creativecommons.org/licenses/by/4.0

Linked Data uses the RDF family of standards for data interchange (e.g., RDF/XML, RDFa, TTL) and query (SPARQL) (Bernadette et al., 2013). The machine readability of linked data relies on OWL (Web Ontology Language[35]), a family of knowledge representation and vocabulary description languages for authoring ontologies based on RDF and standardised by W3C[36].

The data linkage relies on vocabularies to describe the data in different aspects and ontologies to express their relationship. SPARQL[37] is the RDF query language; its expansion – GeoSPARQL, allows the query of geographic data described with RDF (Open Geospatial Consortium, 2012). SPARQL and GeoSPARQL are also standards.

By adhering to Linked Data principles and using RDF and SPARQL, data becomes more accessible, discoverable, and useful in the context of the Semantic Web. This allows applications and services to make sense of the data, enabling smarter data integration, knowledge discovery, and automation on the web.

2.2.1. Linked Data, Geodata and Workflows

Linked Data principles can be used for many different purposes. In the context of geodata workflows several uses have been found in the literature, and tools to convert data (including geodata) in Linked Data have been developed over the years (Kyzirakos et al., 2018; Rowland, 2020; Scharffe et al., 2012). Additionally, ontologies were developed to describe data collections such as the U.S Geological Survey (Koubarakis et al., 2019; Usery & Varanka, 2012).

One of the obvious ways to use Linked Data in the context of geodata and workflows is to link raw data to data products and model outputs (Elliott et al., 2013) or link workflow components such as input data, output data, operations and workflow executions (Roos et al., 2010; Shaon et al., 2011; P. Yue et al., 2016).

Additionally, linked data can improve metadata (Nogueras-Iso et al., 2004) and semantic interoperability by annotating workflows using ontologies or controlled vocabularies (Mesbah et al., 2017; Salah et al., 2014). Tolovski (2020) also used semantic annotation, but his approach focused on the experiment's provenance, algorithm implementations, parameter settings and output metrics.

More recent studies express or publish workflow using linked data, improving workflow interoperability by converting abstract workflows into concrete workflows and vice-versa (Garijo et al., 2017; Lemmens et al., 2016), or converting workflows from one GIS tool to another (Kechagioglou et al., 2019; Ohuru, 2019). These conversions can be achieved by sharing geoprocessing methods as linked data (Scheider & Ballatore, 2018). Ubels' work (2018) is also relevant as it defines a method to convert abstract workflows into concrete ones using Semantic Web technologies. Scheider et al. (2019) went one step further; they described analysis methods using ontologies enabling users to find workflows based on their capacity to solve specific problems.

Linked data can also be used to bind abstract workflow to their executions and track provenance (Missier et al., 2010, 2013; Roos et al., 2010; P. Yue et al., 2016; Zhang et al., 2020)

The investigation by Skvortsov and Stupnikov (2022) proposes an ontological semantic approach to describing workflows at three levels: the control patterns, research workflow step patterns, and the meaning of the workflows in terms of domain knowledge. This approach enables semantic search for workflows for reuse and interoperability at all levels (Skvortsov & Stupnikov, 2022).

[35] https://www.w3.org/OWL

[36] https://www.w3.org

[37] https://www.w3.org/TR/sparql11-query

2.2.2. Workflow Ontologies

Several ontologies were developed over the years to describe workflows, with different degrees of generalisation, domain-specific and describing abstract or concrete workflows. A brief description of the most relevant workflow ontologies for our research is mentioned here.

On the domain-specific spectrum, it is relevant to refer to the Open Provenance Model for Workflow (OPMW[38]). The OPMW was developed specifically for scientific workflows. OPMW was based on the Open Provenance Model (OPM[39]), which aimed to describe the generic provenance of any general thing, created by computer systems or not, in a system-agnostic and precise way. OPMW extended OPM, making it workflow-specific, as well as P-PLAN and PROV, making it possible to describe abstract workflows and link them to their execution while keeping track of the data flowing in the workflow and parameters from operations. The relationship between data in an abstract workflow and the process that has generated it is described by the "IsGeneratedBy" property (Garijo et al., 2017).

In OPWM, the link between the data that comes in or out from a workflow step is only made by ordering the workflow steps, making the workflow description focused on its steps. Furthermore, OPMW uses the Dublin Core – DCMI Metadata Terms to document the workflow's metadata (Garijo et al., 2017).

Scheider and Ballatore (2018), proposed a vocabulary to express workflows as linked data and defined an ontology that enables the propagation of geodata types through the workflow. Their approach allows provenance tracking by connecting the output to the input via a node (operation) in a linked data pattern. Scheider and Ballatore, (2018) modelled the workflow, the operations and the data flowing in the workflow as resources, linking them via the properties "source", "input", "output", and "result". While the "source" and the "result" refer to what comes in and out of the workflow, respectively, the "input" and "output" refer to the data that comes in and out of operations (Figure 9). In Scheider and Ballatore (2018) linked data pattern, there is no mention of operations' parameters or links to the workflow instances.

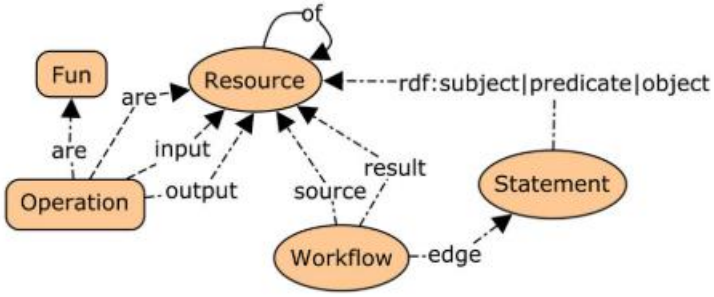


Figure 9 – Linked data pattern to describe workflows (Scheider & Ballatore, 2018).

[38] https://www.opmw.org

[39] https://openprovenance.org/opm/model/opmo

Billings (2019) also developed an ontology to describe scientific workflows that could be executed across distributed WfMS. This ontology focused on the control operation or application perspective and on the control flow or process perspective, as defined in Section 2.1.

On the semantic interoperability side, the work of Skvortsov and Stupnikov (2022) and The Workflow Motif Ontology[40] contribute to increasing workflow understanding (also for machines) through the semantic description of the motif of the workflow and its operations. This would also enable reusing of workflow fragments and automation of problem-solving using workflows.

In conclusion, despite all efforts and best practices to reuse vocabularies, there are still different workflow ontologies using different terms for the same concepts. Different studies were made in different fields of science but still focused on workflow description; there is a lack of uniformization of the terms and concepts used. There should be a common basic vocabulary for workflows that supports workflow prospective provenance and can be extended for users interested in tracking retrospective provenance or workflow evolution. OPMW makes a good candidate for tracking prospective and retrospective provenance. The missing link that also interests this research is the workflow evolution tracking. Although OPMW supports versions of a workflow it does not keep track of when or why did a change occur.

It would be interesting to track the provenance of any dataset within a workflow to its source(s) and see which transformations it underwent, including the parameters from each operation. The link between each operation in the abstract workflow and its execution would also be relevant for the next phase but is out of scope of this research. For this purpose, we propose the data perspective to describe abstract workflow.

[40] http://vocab.linkeddata.es/motifs

3. Towards FAIR Semantic Workflows

This chapter explores the research methodology employed to enhance the findability and reusability of workflows in Marine Spatial Planning. This chapter will be further developed and organised into sections focused on identifying personas and relevant use cases for the workflow repository, repository requirements, metadata needs, workflow description method and database querying capacity. Finally, the method to test and evaluate the solution is described.

3.1. Personas and Use Cases

Personas and use cases are identified to assist in defining the workflow repository requirements and the workflow metadata needs. In order to achieve this, four personas with different interests and roles around Marine Spatial Planning were devised, and a use case was defined.

3.1.1. Frederica, the Green Energy Stakeholder

Frederica is a 46-year-old strategist leading her company's transition from oil and gas to green energy. She wants an overview of how much area is suitable for offshore wind farms in the North Sea and where they are located. She knows the optimal conditions and constraints to develop wind parks but has limited GIS experience. Recently, a colleague has started to use workflows to automate his repetitive tasks; one of them is finding suitable areas for renewable energy. Frederica thinks she could adapt that workflow to help her find suitable areas for offshore wind farms.

Use case

Frederica searches for the workflow her colleague created, and she can retrieve it using the "creator" field. Since the workflow Frederica's colleague created is not specific for offshore green energy development, it does not include the environmental variables she needs for her analysis, such as the depth of the sea. Frederica decides to adapt the workflow she found to her specific needs.

3.1.2. Renata, the Marine Biologist

Renata is a 32-year-old marine biologist working in the research department of an environmental policy group. She won a PhD grant to investigate which areas in the North Sea need the most protection. Policymakers will then use her results to define more areas for environmental protection and contribute to the High Seas Treaty's goal to protect 30% of the oceans by 2030. Renata wants to protect vulnerable habitats that shelter important species and ensure that the new protected areas are part of a connected network to ensure that migratory species have a haven during their journeys.

Use case

Renata knows what kind of geodata she needs and wants to consider in her analysis. Additionally, she thinks strategically and considers a way of saving her analysis so that it can be run several times considering different species migration speeds and scenarios.

Furthermore, Renata can save different versions of her workflows, corresponding to different scenarios and use them to show policymakers how she reached her conclusions, making the decision-making process more transparent to stakeholders.

Before creating her own workflow, Renata searches for similar workflows in a workflow repository. Renata starts by searching workflows per type of analysis and retrieving several workflows, so she further filters per input data's theme. Renata finds a workflow that allows her to adapt to her needs and work more efficiently.

3.1.3. Franck, the GIS Specialist

Franck is a 54-year-old GIS expert at the state's department responsible for Marine Spatial Planning in Denmark. He has extensive experience in GIS projects in different contexts. Marine Spatial Planning has taken increasingly more of his time in the past ten years. Over this time, Franck has also gained insights into the process of Marine Spatial Planning and the different sectoral needs while working on several international projects related to MSP.

Use case

Due to high demand in recent years, Franck decided to organise his work differently. Instead of rebuilding similar processes for every new map or geodata analysis, he wants to define a parametric workflow that can be reused later. Franck can reuse the workflow for different study areas and re-run it when the data is updated or rules change.

Furthermore, over the years, Franck has seen growing interest in MSP from developing countries; nevertheless, technical skills and budgets are scarce in those countries compared with Europe. Franck would like to contribute to capacity building in developing countries and accelerate the Marine Spatial Planning processes. Therefore, Franck wants to share his workflows with the community. To ensure the portability of his workflows and increase findability, Franck shares his abstract workflows using linked data and the semantic web. He appropriately documents the input and output data types, formats, and GIS concepts behind the data at each step of the workflow to facilitate subsequent concretization of the workflows. He also provides a good description and context of the workflow, its purpose, authorship, and date of creation, and he keeps track of versioning, including the motivation behind changes.

3.1.4. Tom, the Student

Tom is a 20-year-old spatial planning student, and he is about to finish his second year of studies. Tom learned about planning processes and tools and gained some GIS skills, but he has little experience with projects. Tom recently heard about Marine Spatial Planning and wants to know more about it, so he joined a summer school on the topic.

Use case

During the summer school, Tom's working group is assigned the task of studying a region and coming up with suitable areas to develop offshore wave farms. Tom and his group have basic GIS knowledge, but they do not know enough about the requirements of wave farm implementation to know what data they need. Additionally, they are unsure what kind of operations they should run on the data to identify suitable areas for wave farms. Thus, the group searches for workflows related to offshore energy production, searching for that term in the workflow purpose description and for terms such as "suitable area" in the description of workflow's output data.

3.2. Repository and Metadata Requirements

The workflow repository requirements were determined based on the literature review, the personas, and the use cases defined in the previous section.

The workflows must be searchable by author, purpose, input, output data (intermediate or final), and Inspire data themes. Additionally, the results of workflow queries should include all the workflow versions and be sortable by date and filterable by maturity status. The collection of requirements is summarized in Table 3. Only the requirements specific to each persona are shown to avoid redundancy.

Personas and use cases	Requirement	Requirement Level	Metadata Needs
Frederica searches for a workflow created by a colleague.	Req. 1.1 Search and retrieve workflows by author or author's affiliation.	Workflow metadata	Workflow author, author affiliation and date of creation.
Renata searches for workflows by theme and input data.	Req. 2.1 Search & Retrieve workflows by type of analysis Req. 2.2 Search & Retrieve workflows by data theme.	Workflow metadata Input metadata	Input metadata, including theme. Workflow metadata, including the type of analysis.
Renata wants to use workflows as provenance data for similar outputs under different scenarios.	Req. 2.3 Save and retrieve parameter values at the operation level.	Operations' details	Operations' parameters and values.
Franck wants to enable colleagues around the world to reuse his workflows.	Req. 3.1 workflow needs to be understandable by others. Req. 3.2 Tracking of workflow evolution. Req. 3.3 Sorting by maturity status. Req. 3.4 Licence terms need to be clear.	Workflow metadata	Workflow description and purpose. Version control with modification date and motivation. Maturity status. Licence info.
Tom searches workflows that solve similar problems or yield similar types of results.	Req. 4.1 Retrieve workflows that produce similar outputs.	Workflow metadata Output metadata	Workflow theme, output metadata.
Tom might use the intermediary output of one workflow as input for his workflow.	Req. 4.2 Combine workflows as sub-workflows in a bigger workflow.	Output metadata	Intermediary outputs metadata.

Table 3 – workflow repository requirements summary

The workflow metadata needs were defined based on the repository's requirements to support the use cases. The workflow metadata list below was also created based on the literature review, OPMW, data on the web best practices (Farias Lóscio et al., 2017), and the tools evaluated during the research. At first, it seemed important to keep track of the tools used to create and modify the workflows since, in the case of many WfMS, the workflow is saved in their specific format, which may not be portable to other tools. Nevertheless, since the target of this research is abstract workflows in RDF, its format/language is more relevant than the software in which it was created.

The workflow metadata needs are presented below in five categories (Section 2.1.3): descriptive metadata, workflow provenance, workflow evolution, structural, and administrative metadata. The descriptive metadata includes data that identifies the workflow and relates to discovery. Workflow provenance data relates to the who, why and how of the workflow creation. Workflow evolution contains all the information necessary to keep track of different versions of a workflow, including the reasons behind changes. The workflow structural metadata refers to the objects related to the workflow and how they are combined. Finally, administrative metadata holds the license of the workflow and the language in which the information is available.

Descriptive Metadata

- Title: the title of the workflow (literal value).
- Identifier: a combination of tokens (letters, symbols and signs) identifying the workflow.
- Description: the description of the workflow (literal value).
- Vocabulary term(s): terms from the domain vocabulary that act like keywords. There can be more than one vocabulary term which can be related to other vocabulary terms employing “is part of”/”has part” from OWL.
- Purpose: the purpose of the workflow is set in the workflow vocabulary and relates to the type of data handling/analysis of the workflow. This can be more than one vocabulary term and be related to other vocabulary terms employing “is part of”/”has part” from OWL.
- Format: format of the workflow description file, in this case TTL.

Workflow Provenance

- Creation date: date on which the workflow was created.
- Author: person or organization that created the workflow.
- Contributor(s): person or organization that assisted in creating or modifying the workflow.
- Author/contributor contact: email address from the author/contributor. This term is associated with the author or contributor, respectively.
- References (if applicable): add the publication(s) that refer to the workflow specifically (preferably using DOI).
- Relevant Literature (if applicable): include publications pertinent to the workflow (preferably using DOI).

Workflow Evolution

- Version number: The version number or the workflow. Guidelines for workflow versioning are explained in Section 5.1.
- Is version of (if applicable): If the workflow is a version of another workflow, this field should indicate its identifier.
- Replaced by version (if applicable): If the workflow has become deprecated, this field should indicate the identifier of the workflow which replaces it.
- Replaces version (if applicable): If the workflow replaces another version, indicate the original workflow identifier in this field.
- Modification date (if applicable): Date of modification of the original workflow.
- Modified by (if applicable): Person or organization which changed the workflow.
- Description of the changes (if applicable): Description of the change made to the original workflow.
- Maturity: Use the terms work in progress, stable and deprecated, respectively, for workflows still under development, completely defined or replaced by other versions.

Structural Metadata

- Input data: URI of the input data, which has its own metadata file.
- Output data (intermediate and final): URI of the output data, which has its own metadata file.
- Data type: each dataset in the workflow should refer to its data type to enable automation of the workflow execution.
- Data vocabulary term: each dataset in the workflow should refer to at least one of the terms from the domain vocabulary that act like keywords.

Administrative Metadata

- License: The type of license of the workflow, so it is clear to users whether they can use the workflow and under which terms.

3.3. Workflow Description

As mentioned in Section 2.2.2, previous efforts have defined workflow ontologies; nevertheless, these approaches focused on documenting the steps and their order to produce an output based on the input data. In this study, the approach to workflow description focuses on the data or information perspective, that is, on the description of the data as it flows through the (abstract) workflows. The underlying premise is that this approach will significantly improve the discoverability of workflows and datasets it produces.

Therefore, this study defines workflows using web semantics by describing the input and output data at each operation within the workflow. The workflow metadata is included in the same file as well.

The RDF language chosen was TTL since it is easily read by humans and machines, is supported by modern RDF libraries and is fast to read. We focus on abstract workflows and their evolution on the premise that works like the one from Garijo et al. (2017), Ubels (2018), Ohuru (2019), and Kasalica (2022) will enable the conversion of abstract workflows to executable ones. This way, we can focus on a portable format with a complete description of the workflow's metadata, inputs, and outputs, thus contributing to findability and reusability and enabling interoperability.

Each dataset involved in a workflow has its corresponding metadata file, also described in TTL. The data description follows the ISO 19115-1:2014 Standard and includes all the mandatory fields from the INSPIRE directive and the data types defined by Scheider et al. (2020). The last was added to enable automatic concretization of workflows a posteriori.

Three ontologies are developed and combined to achieve the research goals: workflow ontology, workflow purpose ontology and domain ontology. The glue that holds these ontologies together is the workflow ontology, which includes the data flowing in the workflow, workflow provenances and the workflow evolution tracking. It links to the domain ontology through the tags assigned to the data that flows in the workflow, and to the workflow purpose via the property specifically created to document the workflow purpose.

The input data's metadata could be downloaded in RDF from the source but had to be converted to Turtle. EasyRdf[41] online converter was used to convert the metadata files to TTL language. After conversion, the files were checked for correctness, consistency, and completeness to ensure there were no mistakes or omissions, and the terms used to document each field were the same throughout all the files.

[41] https://www.easyrdf.org/converter

Workflows have a version number and can be part of other workflows. A person or institution creates them; the term creator can be associated with a person's or organization's name and email address. They have input and output data, but we know how it flows by looking at each description containing provenance data. If the data is not an input, the description includes an operation and "reference object" (can be more than one) - the data from which the dataset originates. The reference object can be a workflow input, intermediary or final output. Additionally, the operation will have a name and a parameter with name and value. Both workflow and operation have a purpose. Figure 10 illustrates the workflow pattern used in the workflow descriptions.

By backtracking from a workflow's final output to its reference object and doing that repeatedly until the workflow inputs, we obtain the full provenance of the output data and understand the flow of data within the workflow.

The URI definitions in this study follow the W3C recommendations (W3C, 2017); nevertheless, for this prototype, the URIs are not registered or online but local. The base URI is http://workflowRepo/MSP/workflow; the URI for data starts with http://workflowRepo/MSP/workflow/Data and is followed by the data identifier. In principle, we follow the camel case notation for namespaces and URIs but keep the original URIs from reused vocabularies. For new datasets, we add one or two words to the dataset identifiers related to their new meaning. The ontology URI is http://workflowRepo/MSP/ontology.

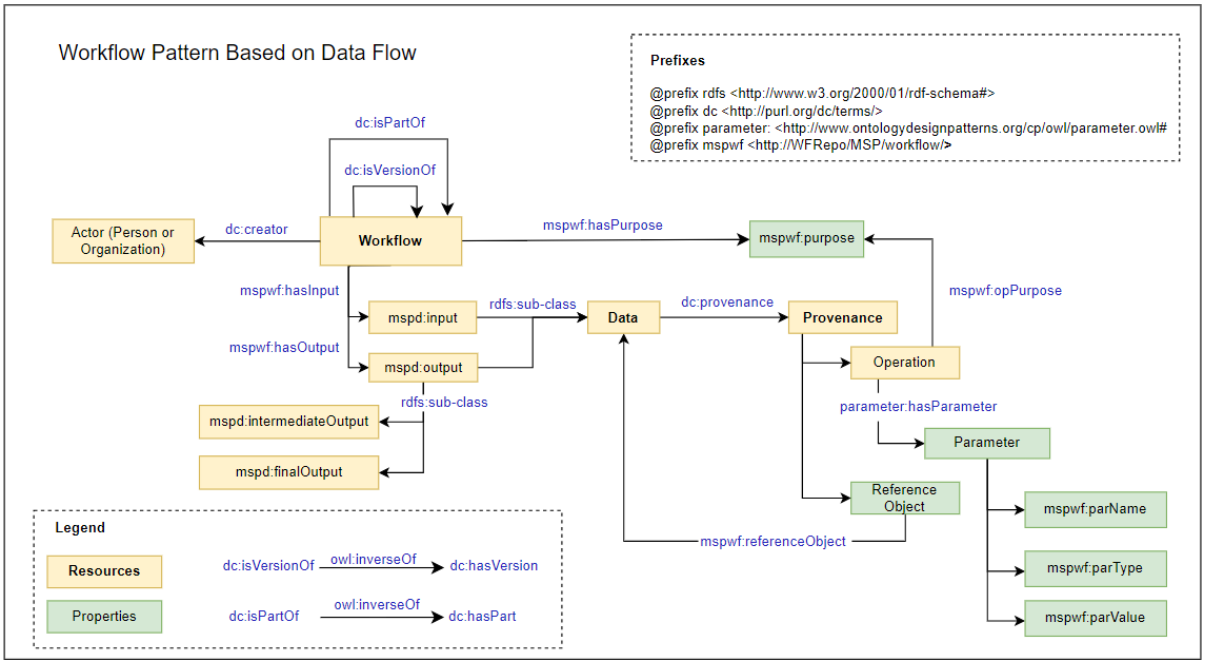


Figure 10 – Workflow ontology based on the data flow.

Inspired by Skvortsov and Stupnikov (2022) and hoping to support the automation of workflows in the future, as mentioned in Section 2.2.2, a workflow purpose pattern is suggested, as illustrated in Figure 11. This ontology can also be used to find workflows that solve similar problems.

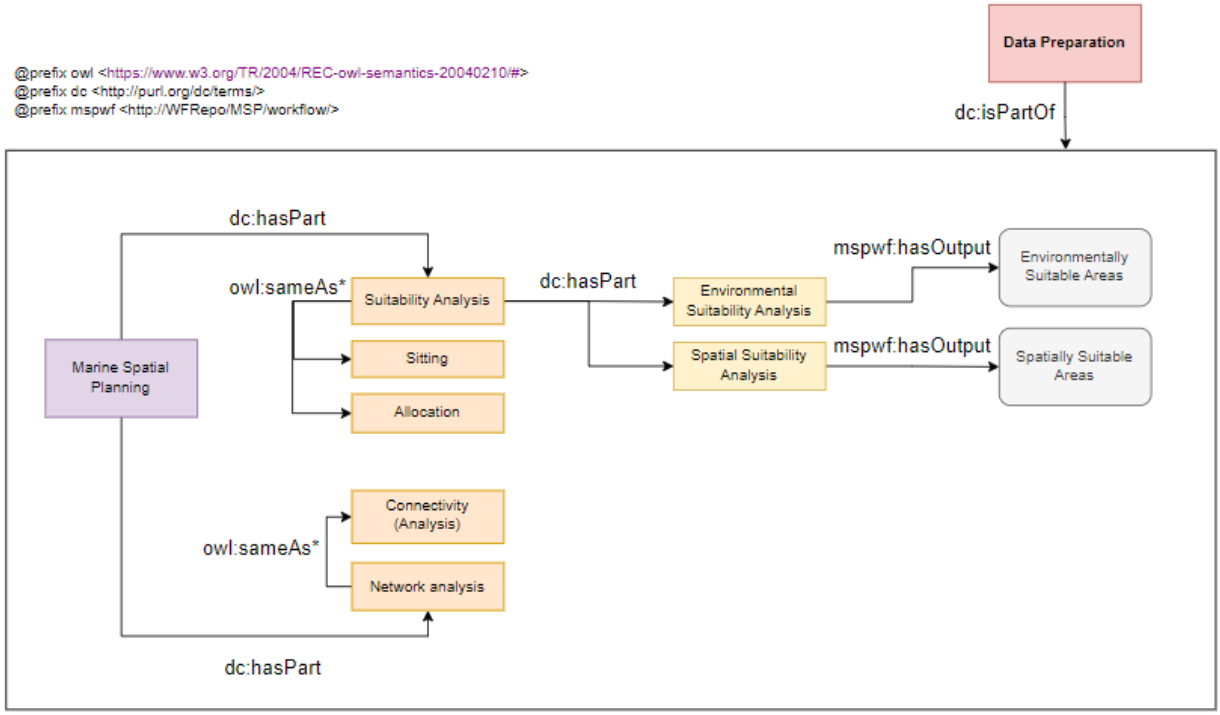


Figure 11 – Workflow Purpose Ontology. The main purpose of the workflows in the repository of this case study is to support the Marine Spatial Planning process. The data analysis can be of different sorts. In this example, data preparation is always needed as an over-arching process supporting other types of analysis, such as suitability and network analysis. The suitability analysis can be divided into spatial and environmental suitability.

3.3.1. Domain Ontology

Marine or Maritime Spatial Planning (MSP) is a process that aims to allocate areas for various human activities and nature conservation while preserving or enhancing the environmental status of the seas. One key aspect of MSP is recognising that different activities can interfere with each other. MSP can help identify potential conflicts and synergies among sectors and find solutions through stakeholder engagement.

Human activities in the marine environment may be "hard and fixed uses" or "fleeting and soft uses" (European Commission, 2019). We adopt the terms "hard uses" and "soft uses" for simplification. Hard uses are the ones that require the installation of (usually expensive) infrastructure and have a long period of permanence, such as an oil and gas platform. On the other hand, soft uses are temporary and often do not require infrastructure; think of shipping traffic, for instance. In addition, some human activities may be incompatible; for example, an oil and gas platform will stand in the way of shipping, but a pipeline will not interfere with it.

For the context of this research, we use the European Commission's (2019) concept of compatible maritime activities, assuming that any two uses that are not explicitly marked as compatible are incompatible. Nevertheless, incompatibility is not always black or white, especially with soft uses; many nuances exist. Therefore, the focus is on hard uses. In practice, multiuse is still limited to a few case studies such as Kyvelou and Ierapetritis (2021), Stancheva et al. (2022), and case studies conducted in the realm of the MUSES project (Multiuse in European Seas - https://muses-project.com). The conditions required for multiuse are too specific to implement in the scope of this research.

Due to their nature, some activities require safety area buffers around them. The buffer distance can vary per region and country for each activity (von Thenen, 2021).

There are two components to define suitable areas for human activities: spatial and environmental suitability. The first relates to areas not already occupied with incompatible activities and their safety area buffers; the latter refers to having the right environmental conditions, such as currents, bathymetry, and oxygen content (von Thenen, 2020).

Environmental suitability is also an important factor when defining areas to protect. There is no point in restricting marine uses in areas where important habitats or species are absent. Additionally, it is important to consider the connectivity between Marine Protected Areas (MPAs). There are different aspects of MPA Connectivity, but for simplification purposes, the distance between MPAs (Carr et al., 2017) is used as a proxy in this study.

The concepts used in the context of Marine Spatial Planning are described in Table 4. Their relationships are mapped using owl concepts of "same as", "inverse of", "part of", and "has part", as seen in Figure 12. The "same as" relationship links names with the same meaning, such as renewable and green energy. The "part of" concept conveys relationships with a larger class, e.g. bathymetry is part of the elevation theme. Finally, the "has part" concept links an object to its different parts. The objective of mapping all these concepts and their parts or super-classes is to support the discovery of similar datasets and workflows when looking for the super-class or a part of the object of search. Due to the system's complexity, only the concepts used in the workflow descriptions are included. Appendix II contains the domain ontology in TTL format.

Concept	Function in the ontology	Definition
Hard Uses	Super-class	Hard uses require fixed infrastructure that is typically expensive to install, is placed for a long period and is not easily relocated (European Commission, 2019).
Soft Uses	Super-class	Uses that are mobile, such as fishing and shipping, and not permanent, such as dredging (European Commission, 2019).
Environmental Conditions	Super-class	Conditions such as temperature, salinity, and wind speed.
Suitable area	Super-class	Area where spatial suitability and environmental suitability are favourable to a specific purpose.
Safety Buffer Area	Class	Area around a certain structure or activity which has special rules to ensure safety.
Environmentally Suitable Areas	Class	Areas where the environmental conditions are suitable for a determined purpose or species
Spatially Suitable Areas	Class	Areas that are free from uses that hinder a determined activity.
Require	Object property	Denotes a relationship between 2 infrastructures, where it does not make sense for one to exist without the other. For instance, having a wind farm without the respective energy cables that bring the energy to shore would not make sense.
Incompatible uses	Object property	Relationship between two activities that cannot coexist in time and space.
Influence	Object property	Certain parameters (depth of the sea), infrastructures (ports) or presence of species (fish) influence each other. The influence can be a positive reinforcement – the more the fish density, the more fishing vessels density or negative – the more protected areas, the less fishing density.
Distance relationship	Object property	Denotes the fact that some infrastructures, activities, or areas benefit from being close to others. For example, some species may benefit from marine protected areas that can be reached under a certain amount of time.

Table 4 – Domain ontology

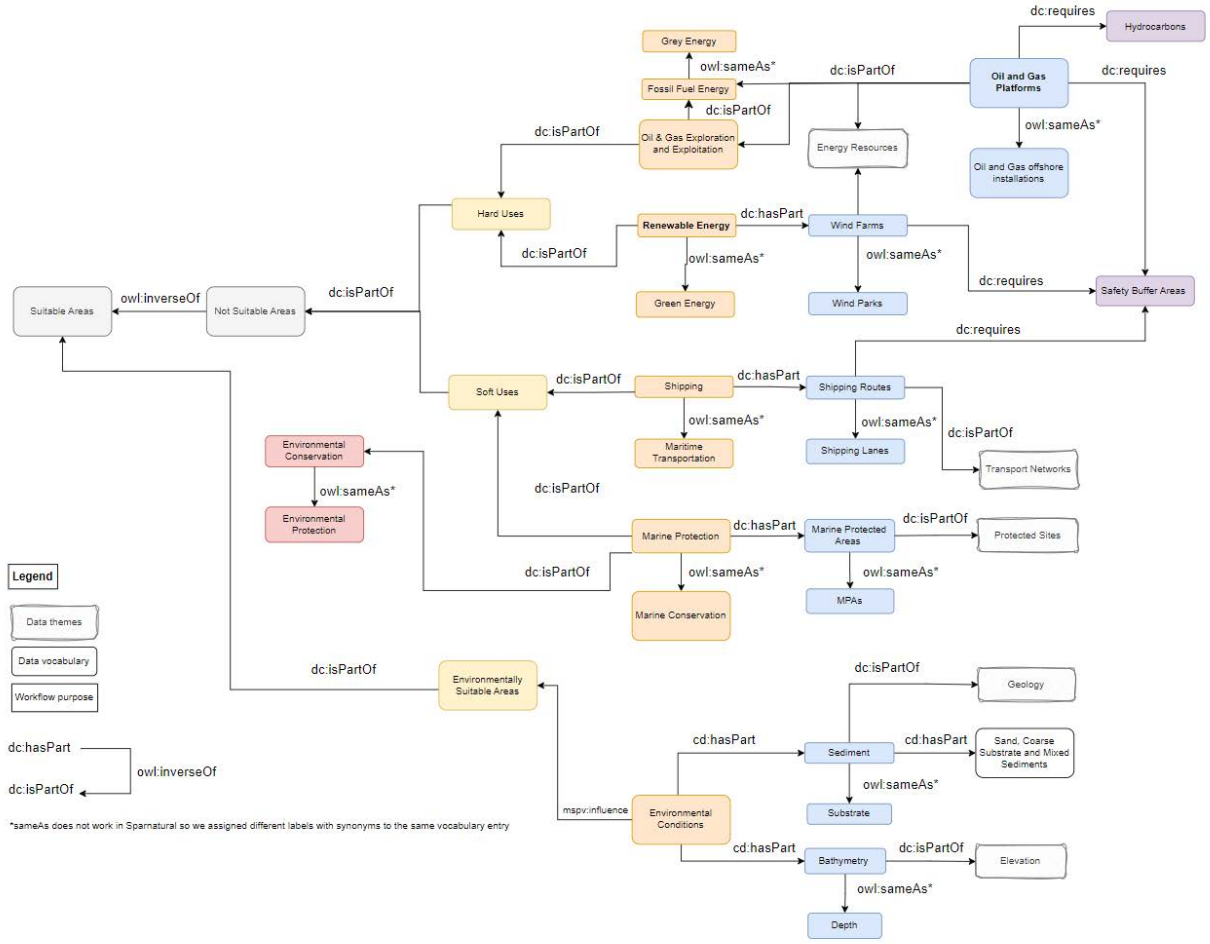


Figure 12 – Domain ontology with relationships between the different human activities, environmental conditions and biological components

3.3.2. Vocabularies

Section 3.3 and Section 3.3.1 define the links between concepts that support the description of workflows and the data flowing within them. In this section, the vocabularies used to support these descriptions are mentioned. Whenever possible and suitable, the W3C Data on the Web Best Practices (W3C, 2017) are followed and existing vocabulary is used.

Based on the relevant metadata standards explored in Section 2.1.2, the significant vocabularies for metadata are the Dublin Core DCMI Metadata Terms and GeoSPARQL. The Dublin Core uses generic terms describing data details such as title, authors and creation dates. GeoSPARQL is specific for geographic data descriptions and complements Dublin Core in describing metadata regarding geodata.

The themes from the INSPIRE directive are a good vocabulary to classify the datasets used in the workflows into thematic categories, and the domain ontology developed in the context of this research complements it, as pictured in Figure 12. Additionally, the same Directive provides terms to describe geodata formats. However, it lacks terms to describe the geometry type in a file. The data types Scheider et al. (2020) defined are used for that.

A few vocabularies pertinent to the marine environment and maritime activities are registered in the Linked Open Vocabularies (LOV[42]). One example is the SeaLiT ontology[43], which describes maritime history. Another example is the "Marine Top Level Ontology"[44], designed to describe spatial-temporal phenomena (with natural or human causes), biotic and abiotic things and their relationships and thus for Marine Spatial Plans description. Nevertheless, this is a high-level ontology that does not support the goals of this research. The domain ontology developed in this research could be fitted under the MarineTLO. However, it is not in the scope of this research.

3.4. Workflow Repository

3.4.1. Geodata Workflows

Several workflows used in Marine Spatial Planning were described using the proposed workflow description and stored in a triplestore database. The workflows were defined by the author of the thesis not only because they are used in her line of work but also because they deal with different data types and involve several common spatial operations. These workflows are intended to be reused (to reproduce the same results, replicate an analysis with similar data, or be repurposed) by the author or by others.

In this section, we present the workflows, which are dedicated to data preparation (Figure 13), spatial suitability analysis (Figure 14 and Figure 15), environmental suitability analysis (Figure 16 and Figure 17) and network analysis (Figure 18). The suitability analysis workflows have two versions each, with slight modifications to provide examples of variation. Also, the data preparation workflow is used in the environmental suitability analysis as a sub-workflow. The TTL files describing the workflow and containing their metadata are also stored in a public GIT repository, accessible at https://github.com/PGMagali/SemanticWorkflowSharing. A sample of those files is provided in Appendix III.

[42] https://lov.linkeddata.es/dataset/lov

[43] http://www.sealitproject.eu/ontology

[44] http://www.ics.forth.gr/isl/MarineTLO

Since developing a workflow repository is out of the scope of this research, the prototype used a triplestore database and a graphical SPARQL query builder to support the users in querying the database. A webpage was also developed to effectively represent each workflow, showing the workflow diagram and metadata (Figures 13 to 18).

The workflow RDF graphs did not offer a satisfactory way to visualize the workflows, so the workflow diagrams were created using draw.io online[45]. The workflow diagram represents the workflow operations as an edge, and the data flowing in the workflow as a node. This approach is the opposite of the typical DAG diagram used, for instance, in Pegasus.

[45] https://app.diagrams.net

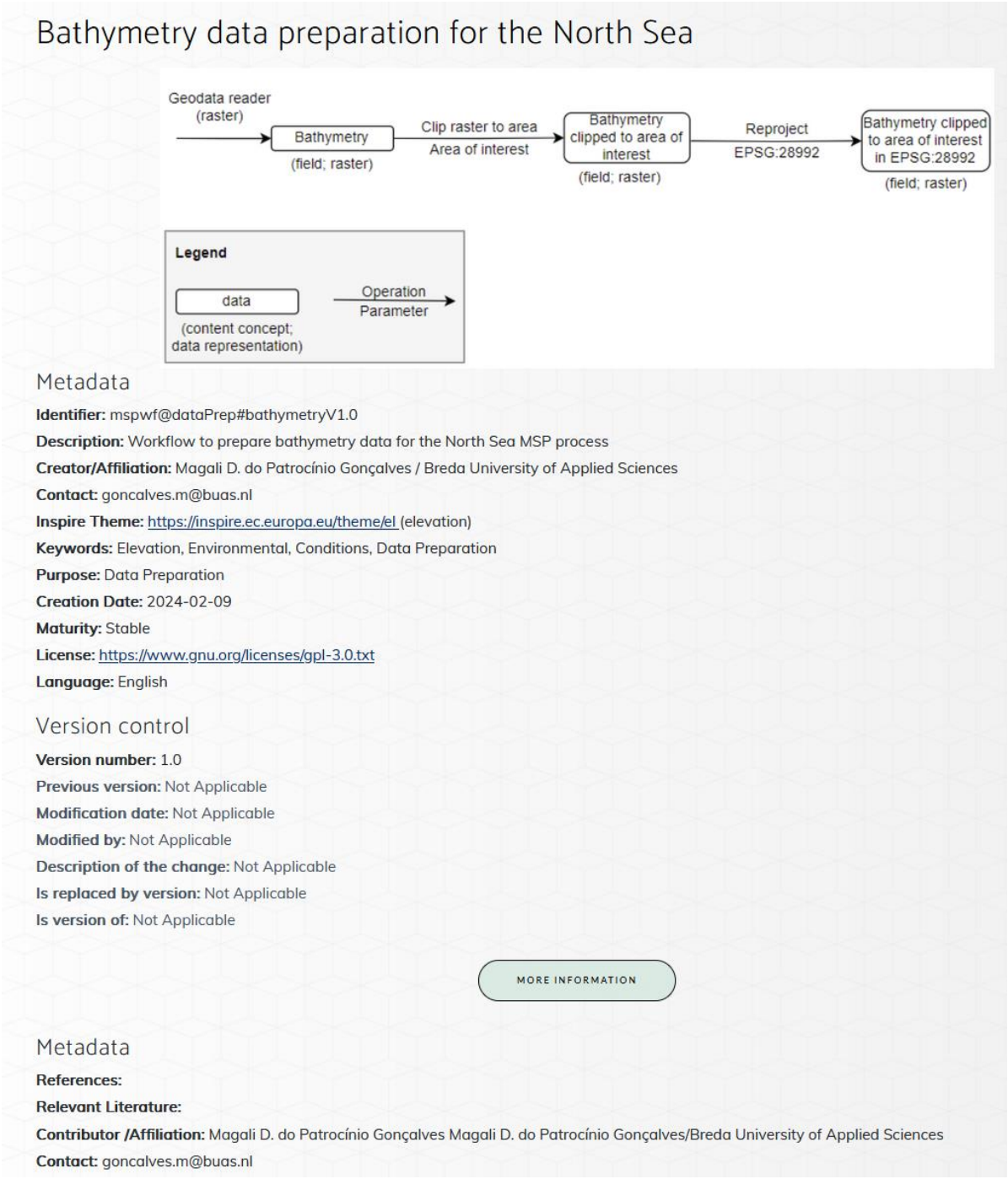
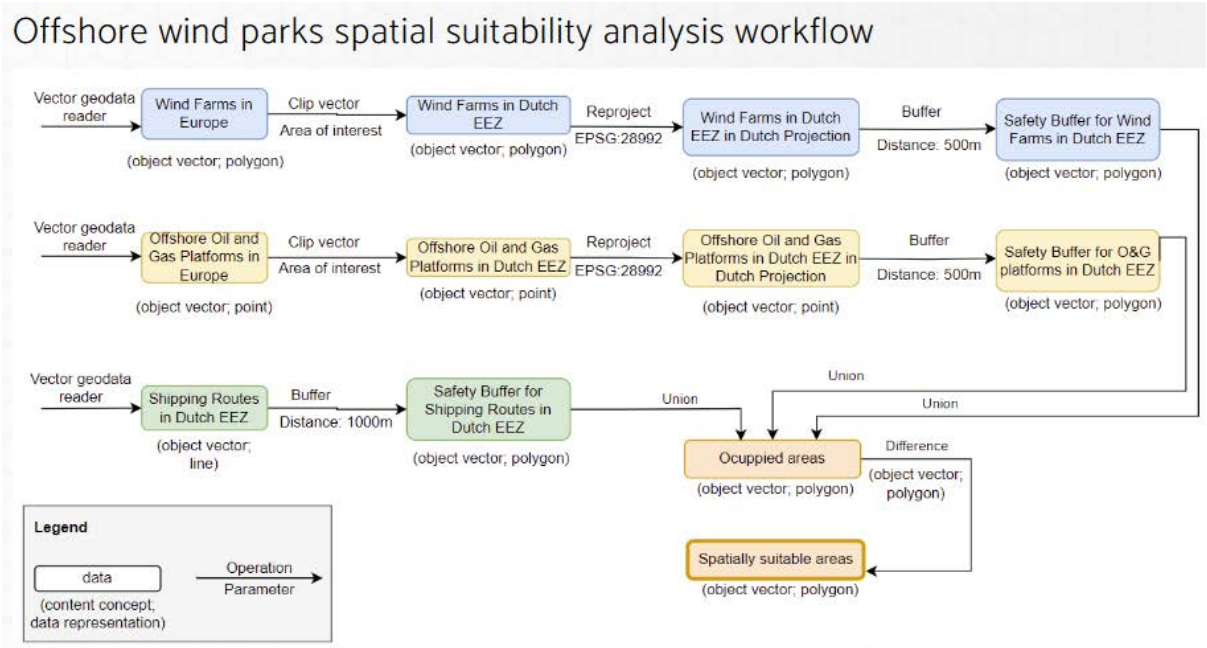


Figure 13 – Webpage with the workflow diagram prepared to illustrate the data preparation workflow and its metadata. The workflow takes the bathymetry as an input (field raster type), clips it to the area of interest and reprojects it to the projection used in The Netherlands.



Metadata

Identifier: mspwf@suitability#windfarmsV1.0
Description: Workflow to find spatially suitable areas for offshore wind parks by exclusion of non-suitable areas.
Creator/Affiliation: Magali D. do Patrocínio Gonçalves / Breda University of Applied Sciences
Contact: goncalves.m@buas.nl
Inspire Theme: https://inspire.ec.europa.eu/theme/er (energy)
Keywords: Offshore energy, Spatial Suitability, Energy
Purpose: mspv:spatialSuitability, mspv:offshore, mspv:marineSpatialPlanning, mspv:greenEnergy, mspv:renewableEnergy
Creation Date: 2024-01-14
Maturity: Stable
License: https://www.gnu.org/licenses/gpl-3.0.txt
Language: English

Version control

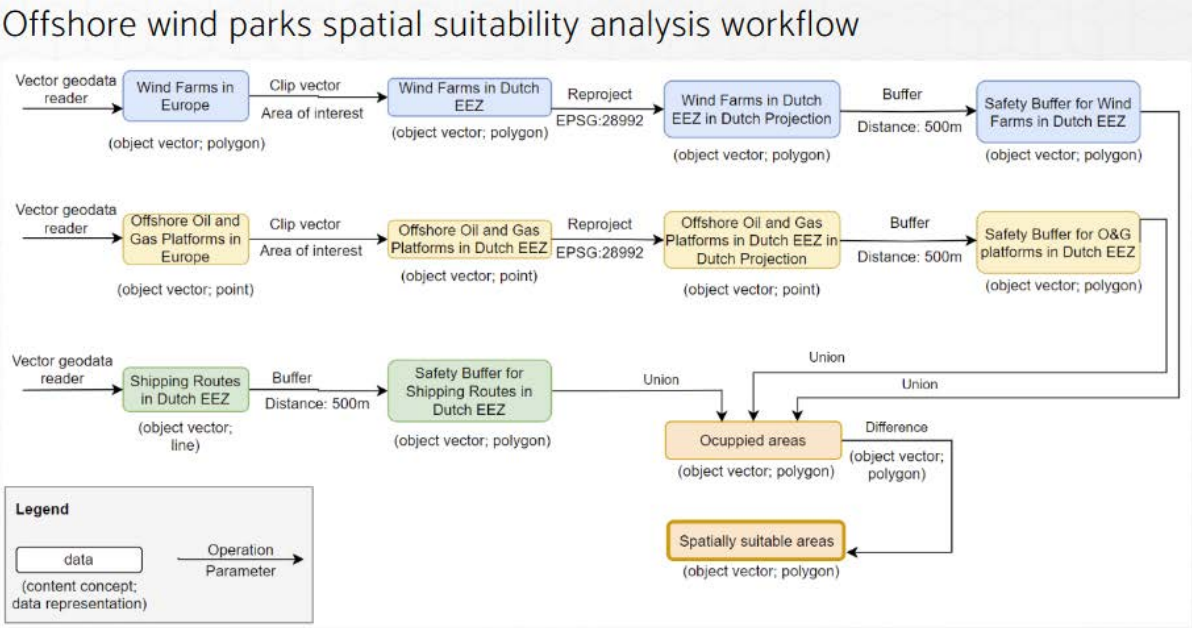
Version number: 1.0
Previous version: Not Applicable
Modification date: Not Applicable
Modified by: Not Applicable
Description of the change: Not Applicable
Is replaced by version: Not Applicable
Is version of: Not Applicable

MORE INFORMATION

Metadata

References: Ooms, E., Onwona Ansong, J., Lukic, I., et al. (2019). Addressing conflicting spatial demands in MSP -- Considerations for MSP planners. https://doi.org/https://data.europa.eu/doi/10.2826/151447
Relevant Literature:
Contributor /Affiliation: Magali D. do Patrocínio Gonçalves Magali D. do Patrocínio Gonçalves/Breda University of Applied Sciences
Contact: goncalves.m@buas.nl

Figure 14 – Webpage with the workflow diagram prepared to illustrate the spatial suitability analysis workflow. It takes several human activities as input (vector files of point, line and polygon types), clips the datasets to the area of interest when needed, reprojects them to the projection used in The Netherlands and adds safety buffers around each geometry, transforming all datasets into polygon-based types. The area of interest not covered by the union of the layers identifies the area suitable for wind park development. This is version 1.0 of the workflow.



Metadata

Identifier: mspwf@suitability#windfarmsV1.1
Description: Workflow to find spatially suitable areas for offshore wind parks by exclusion of non-suitable areas
Creator/Affiliation: Magali D. do Patrocínio Gonçalves / Breda University of Applied Sciences
Contact: goncalves.m@buas.nl
Inspire Theme: https://inspire.ec.europa.eu/theme/er (energy)
Keywords: Offshore energy, Spatial Suitability, Energy
Purpose: mspv:spatialSuitability, mspv:offshore, mspv:marineSpatialPlanning, mspv:greenEnergy, mspv:renewableEnergy
Creation Date: 2024-01-14
Maturity: Stable
License: https://www.gnu.org/licenses/gpl-3.0.txt
Language: English

Version control

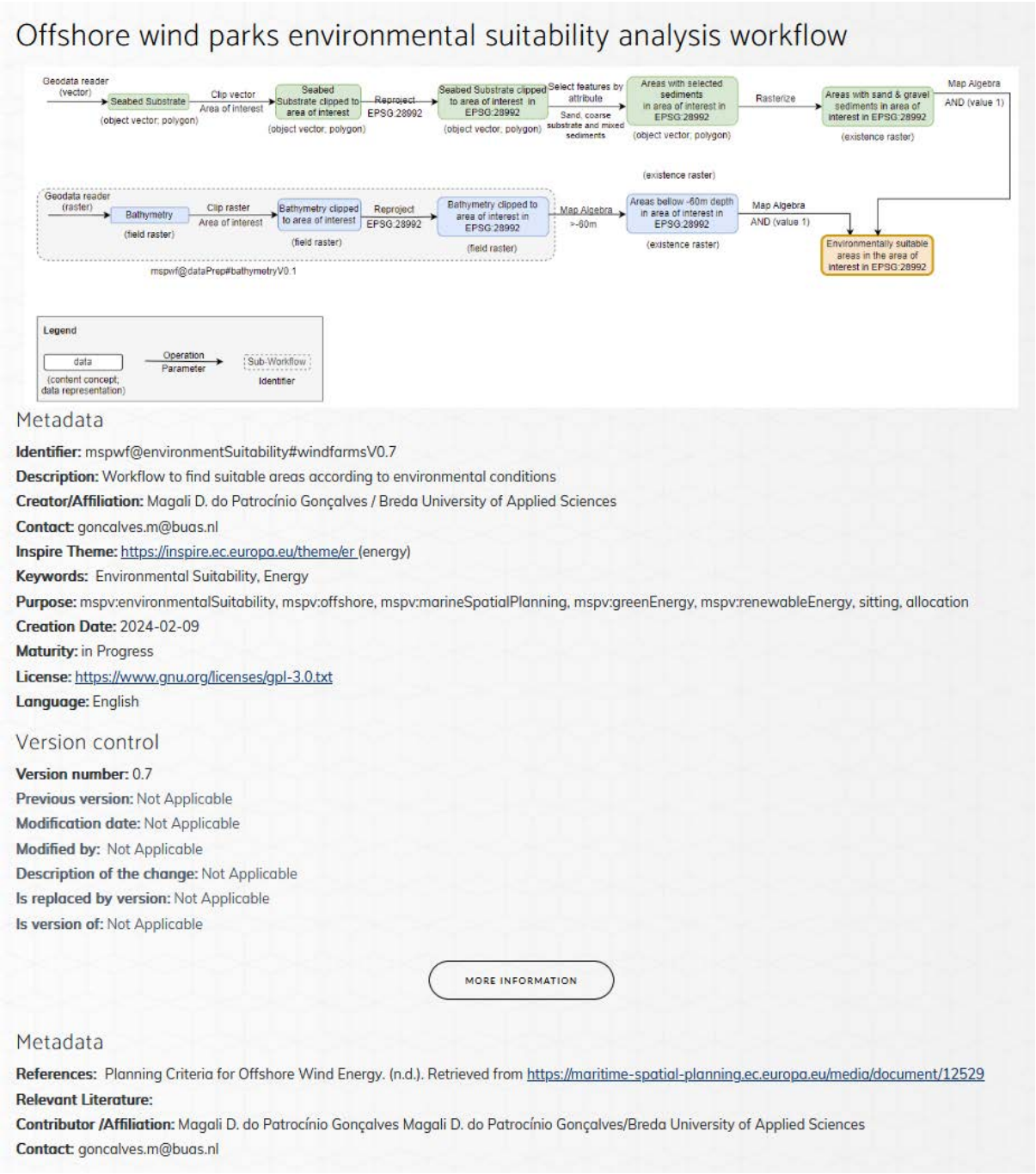
Version number: 1.1
Previous version: 1.0
Modification date: 2024-02-01
Modified by: Magali D. do Patrocínio Gonçalves
Description of the change: Changed safety buffer distance around shipping lanes to 500m instead of 1000m.
Is replaced by version: Not Applicable
Is version of: Not Applicable

MORE INFORMATION

Metadata

References: Ooms, E., Onwona Ansong, J., Lukic, I., et al. (2019). Addressing conflicting spatial demands in MSP -– Considerations for MSP planners. https://doi.org/https://data.europa.eu/doi/10.2826/151447
Relevant Literature:
Contributor /Affiliation: Magali D. do Patrocínio Gonçalves Magali D. do Patrocínio Gonçalves/Breda University of Applied Sciences
Contact: goncalves.m@buas.nl

Figure 15 – Webpage with the workflow diagram prepared to illustrate the spatial suitability analysis workflow. It takes several human activities as input (vector files of point, line and polygon types), clips the datasets to the area of interest when needed, reprojects them to the projection used in The Netherlands and adds safety buffers around each geometry, transforming all datasets into polygon-based types. The area of interest not covered by the union of the layers identifies the area suitable for wind park development. This is version 1.1 of the workflow, in this version the buffer distance considered for the shipping routes was 500m instead of 1000m as described in the version control field “Description of the change”.



Metadata

Identifier: mspwf@environmentSuitability#windfarmsV0.7
Description: Workflow to find suitable areas according to environmental conditions
Creator/Affiliation: Magali D. do Patrocínio Gonçalves / Breda University of Applied Sciences
Contact: goncalves.m@buas.nl
Inspire Theme: https://inspire.ec.europa.eu/theme/er (energy)
Keywords: Environmental Suitability, Energy
Purpose: mspv:environmentalSuitability, mspv:offshore, mspv:marineSpatialPlanning, mspv:greenEnergy, mspv:renewableEnergy, sitting, allocation
Creation Date: 2024-02-09
Maturity: in Progress
License: https://www.gnu.org/licenses/gpl-3.0.txt
Language: English

Version control

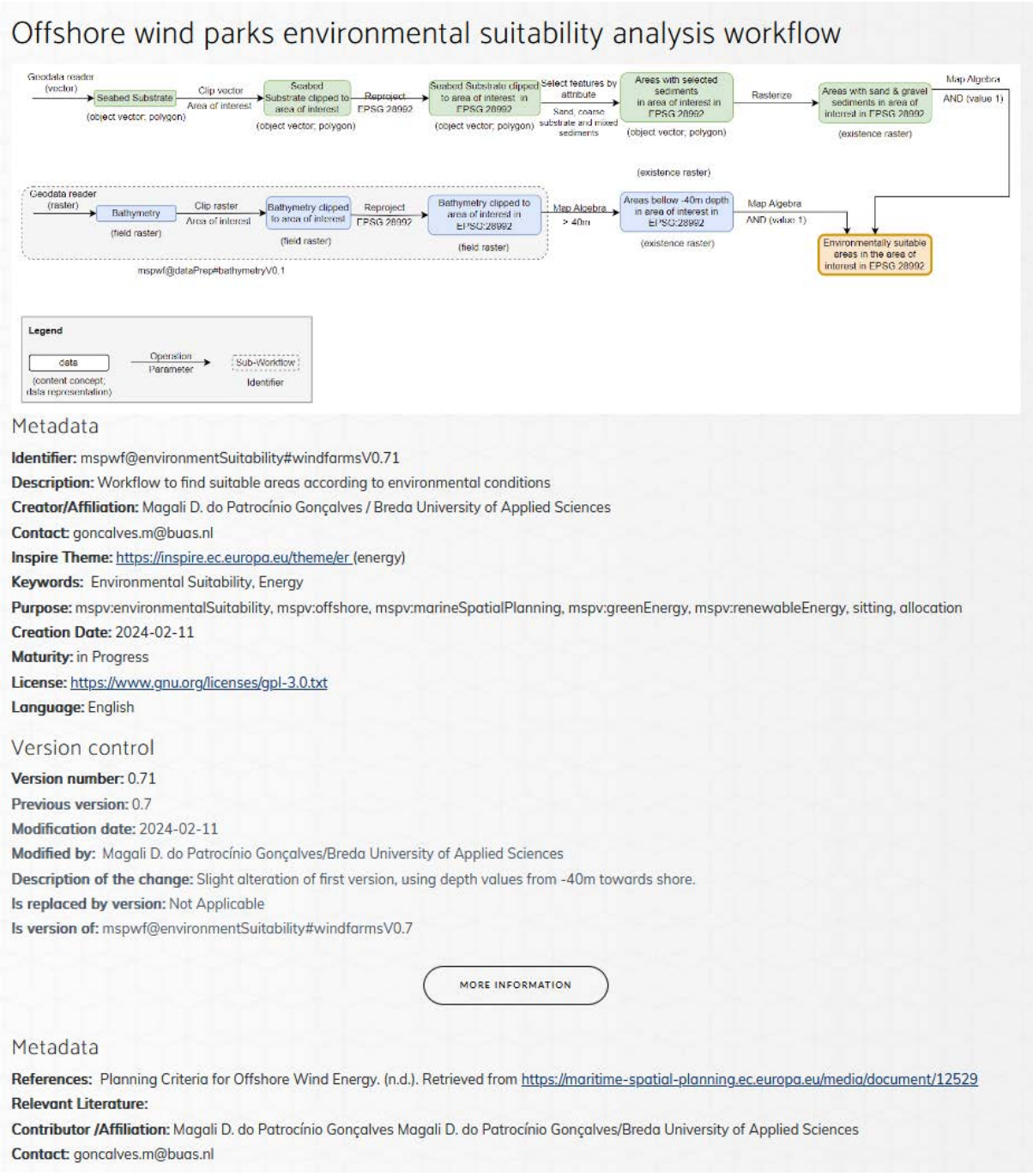
Version number: 0.7
Previous version: Not Applicable
Modification date: Not Applicable
Modified by: Not Applicable
Description of the change: Not Applicable
Is replaced by version: Not Applicable
Is version of: Not Applicable

MORE INFORMATION

Metadata

References: Planning Criteria for Offshore Wind Energy. (n.d.). Retrieved from https://maritime-spatial-planning.ec.europa.eu/media/document/12529
Relevant Literature:
Contributor /Affiliation: Magali D. do Patrocínio Gonçalves Magali D. do Patrocínio Gonçalves/Breda University of Applied Sciences
Contact: goncalves.m@buas.nl

Figure 16 – Webpage with the workflow diagram prepared to illustrate the environmental suitability analysis workflow. It uses the data preparation sub-workflow to prepare the bathymetry dataset, prepares the seabed substrate as well. Then, it filters the suitable sediment by attribute and associates it with the suitable depth to define areas where it is possible to build wind parks based on the environmental conditions. This is version 0.7 of the workflow which is still in progress.



Metadata

Identifier: mspwf@environmentSuitability#windfarmsV0.71
Description: Workflow to find suitable areas according to environmental conditions
Creator/Affiliation: Magali D. do Patrocínio Gonçalves / Breda University of Applied Sciences
Contact: goncalves.m@buas.nl
Inspire Theme: https://inspire.ec.europa.eu/theme/er (energy)
Keywords: Environmental Suitability, Energy
Purpose: mspv:environmentalSuitability, mspv:offshore, mspv:marineSpatialPlanning, mspv:greenEnergy, mspv:renewableEnergy, sitting, allocation
Creation Date: 2024-02-11
Maturity: in Progress
License: https://www.gnu.org/licenses/gpl-3.0.txt
Language: English

Version control

Version number: 0.71
Previous version: 0.7
Modification date: 2024-02-11
Modified by: Magali D. do Patrocínio Gonçalves/Breda University of Applied Sciences
Description of the change: Slight alteration of first version, using depth values from -40m towards shore.
Is replaced by version: Not Applicable
Is version of: mspwf@environmentSuitability#windfarmsV0.7

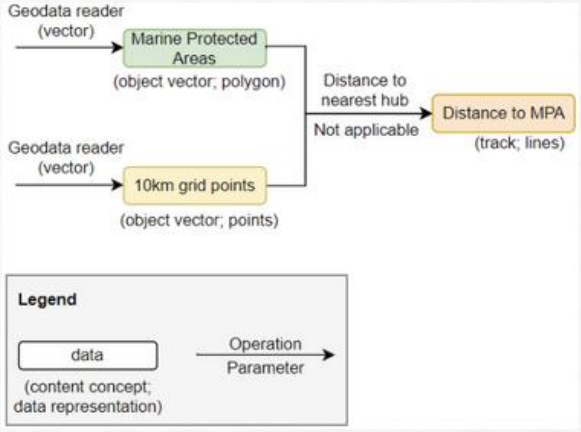
MORE INFORMATION

Metadata

References: Planning Criteria for Offshore Wind Energy. (n.d.). Retrieved from https://maritime-spatial-planning.ec.europa.eu/media/document/12529
Relevant Literature:
Contributor /Affiliation: Magali D. do Patrocínio Gonçalves Magali D. do Patrocínio Gonçalves/Breda University of Applied Sciences
Contact: goncalves.m@buas.nl

Figure 17 – Webpage with the workflow diagram prepared to illustrate the environmental suitability analysis workflow. It uses the data preparation sub-workflow to prepare the bathymetry dataset, prepares the seabed substrate as well. Then, it filters the suitable sediment by attribute and associates it with the suitable depth to define areas where it is possible to build wind parks based on the environmental conditions. This is version 0.71 of the workflow, in this version the depth considered suitable is above -40m instead of -60m as described in the version control field “Description of the change”.

MPAs Connectivity evaluation in the Baltic Sea



Metadata

Identifier: mspwf@MPAConnectivityBSV0.2
Description: Workflow to evaluate MPAs Connectivity in the Baltic Sea
Creator/Affiliation: Magali D. do Patrocínio Gonçalves/Breda University of Applied Sciences
Contact: goncalves.m@buas.nl
Inspire Theme: http://inspire.ec.europa.eu/theme/ps (Protected Sites)
Keywords: , Marine Protection, Marine Conservation, Marine Protected Areas, MPA
Purpose: Connectivity, Network Analysis, Marine Conservation, Marine Spatial Planning
Creation Date: 2024-02-09
Maturity: In Progress
License: https://www.gnu.org/licenses/gpl-3.0.txt
Language: English

Version control

Version number: 0.7
Previous version: Not Applicable
Modification date: Not Applicable
Modified by: Not Applicable
Description of the change: Not Applicable
Is replaced by version: Not Applicable
Is version of: Not Applicable

MORE INFORMATION

Metadata

References:
Relevant Literature:
Contributor /Affiliation: Magali D. do Patrocínio Gonçalves Magali D. do Patrocínio Gonçalves/Breda University of Applied Sciences
Contact: goncalves.m@buas.nl

Figure 18 – Webpage with the workflow diagram prepared to illustrate the network analysis workflow and its metadata. The inputs are a dotted grid layer of 10km and the Marine Protected Areas in the Baltic Sea. The operation will define the distance from each point to its closest Marine Protected Area. A point layer and a polygon layer originate, thus, a line-based layer.

3.4.2. Triplestore

Several triplestore databases are suitable for storing the workflow description, proprietary (e.g. RDF studio) or open source (e.g. Virtuoso). GraphDB was chosen because of its free access, support of TTL files, ease of use, and good graphical interface. Additionally, GraphDB is compatible with Sparnatural, a SPARQL query builder, allowing users unfamiliar with SPARQL to explore the RDF graphs.

Each TTL file describing a workflow was loaded into the database, including the metadata files for the data flowing in the workflow. A TTL file is added with the domain and workflow purpose ontology, as seen in Figure 11 and Figure 12, respectively. The ontology TTL file ensures that the relationships between terms are known to the database without stating them in each of the separate files that use the terms. Figure 19 shows the overview of the several components of this research and how they connect.

The system's core is the triplestore (on the left-hand side of the figure), which contains the workflow description files in TTL and the ontologies and vocabularies used. The query end-point grants users access to the database. A conversion system would be needed *a posteriori* to convert the abstract workflows into executable ones, but that is out of the scope of this research. The workflows would transform geodata into new datasets or data products, all with metadata descriptions, which are translated into TTL files as well, as to be used in the workflow description.

GraphDB offers the option to visualise the content of the database as a graph where users can click on a node to see more details regarding that node, or double-click and see other nodes connected to it. Additionally, the relationships between nodes are shown in the connecting arrows, and the nodes are colour-coded. Nevertheless, the node colours are assigned randomly each time a graph is loaded, which makes it harder to interpret the graph. Understanding a workflow by looking at the RDF graph is not straightforward as the workflow overview is not visible at first, and users can get lost in the myriad of nodes and connectors. Figure 20 shows one of the workflow diagrams in GraphDB.

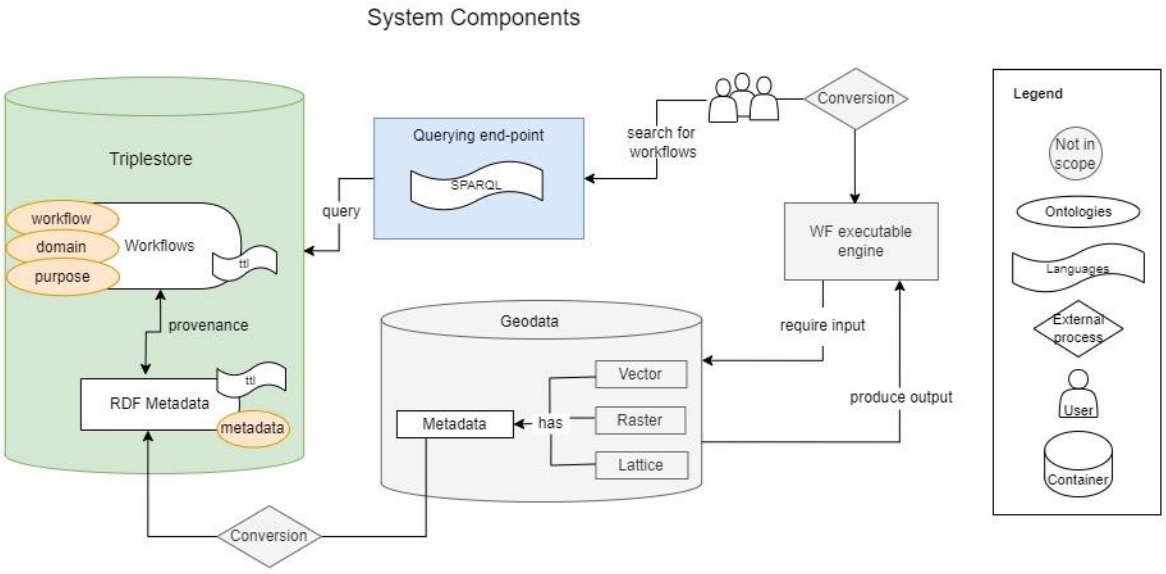


Figure 19 - overview of the several components of this research and how they connect. Grey components are not part of the scope of this research.

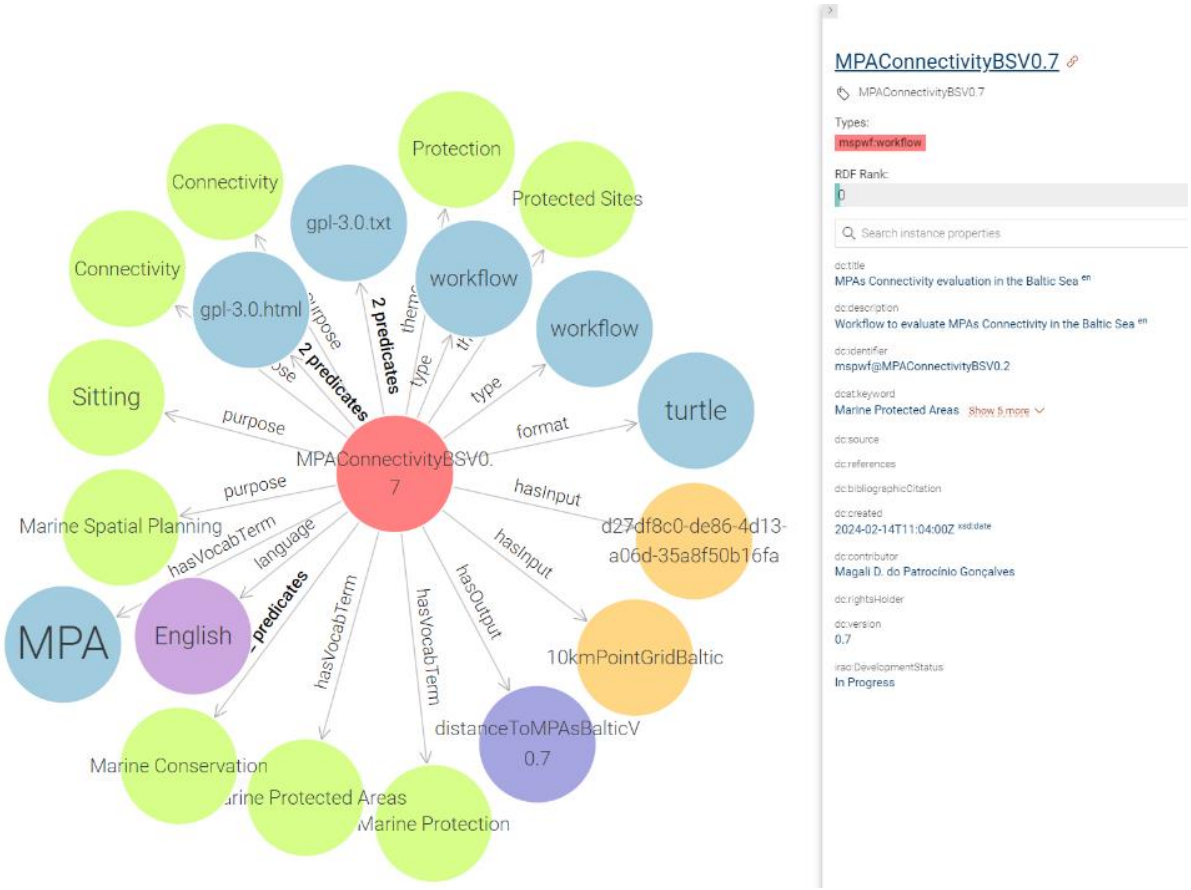


Figure 20 – RDF graph of one of the Workflows in the GraphDb database. In the centre of the diagram is the workflow (subject), the edges represent the properties (predicate), and around the centre, the nodes represent the properties' value (objects).

3.4.3. Querying the Triplestore

Several queries are expected to be supported by the workflow ontology and metadata, as identified in Section 3.2. SPARQL queries were defined and successfully tested for each persona and use case described in Section 3.1. One example of a query in GraphDB (Code snippet I) and the results retrieved (Figure 21) is illustrated below. All the queries made in GraphDB are in Appendix IV.

Unfortunately, only expert users will be able to know how to build SPARQL queries, which would hinder the adoption of a workflow repository that uses linked data. Therefore, a user-friendly way to query the triplestore database was found. Sparnatural[46] allows users to build SPARQL queries through a graphic interface, provided it is adapted to the database's ontology.

Sparnatural is free, open source and easy to configure. Only two files need editing to customize Sparnatural to a use case: a TTL file with the ontology and the webpage file. In the TTL file, besides the ontology, the search options can be customized for each class, e.g. search fields, lists of options, or autocomplete options from the database. There is also the option to customise object properties for more complex queries. For instance, changes between workflow versions were mapped under the Dublin Core term "Provenance" and then "Provenance Statement". This was customised to be named "Changes description" in Sparnatural; selecting that object property would immediately add both terms as one in the SPARQL query – Figure 22.

[46] https://sparnatural.eu

The TTL file created to customize Sparnatural to our database is in Appendix V.

Although Sparnatural is very useful in allowing any user to explore the RDF graph knowledge, it does not come without limitations.

Sparnatural does not recognize sub-classes of classes, so it does not cascade down their properties. This motivated the adaptation of the workflow ontology to fit the tools used, which is not the preferred solution. Initially, workflows had sub-workflows as a sub-category, but only the workflows remained. Similarly, the intermediate and final outputs were defined as a sub-class of outputs but are now defined as "part of" outputs (which in turn are "part of" data). Nevertheless, the workflow ontology still supports sub-workflows by mapping workflows that are part of another workflow; the OWL property "is part of" is used.

Additionally, if different classes share the same property, the properties need to be defined in the TTL file for each class that shares them, so that they can be differentiated when being queried. This is the case, for instance, of the "part of" property used for the domain vocabulary terms and the workflows. Thus, in the TTL file in Appendix V, readers can see the property "this:workflowIsPartOf" with domain:workflow and range workflow, and "this:vocabularyIsPartOf" with domain:vocabulary and range vocabulary.

Another constraint is that inverse relationships such as "is part of" and "has part" were not recognised unless they were both stated in the ontology. Simply having those properties defined as the inverse of each other did not mean users could query the data both ways; the system could not retrieve A as being part of B if, in the ontology file, the relationship stated that B has part A. Both relationships must be stated in the database's ontology file for each class/property.

Additionally, it was expected that having a term linked to another term with the "same as" property would mean that any query with one of the terms would also retrieve results if the other term was mentioned, but that was not the case. All tags need to be added to the data to force that behaviour.

Lastly, any vocabulary mentioned must be linked to its "literal value" (or label), which was an unnecessary step for the final users.

Although Sparnatural was a big help in enabling anyone to query the triplestore, implementing the vocabulary terms' relationships was unsatisfactory. Some features were left unexplored due to time limitations.

```
PREFIX msp: <http://WFRepo/MSP>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX mspwf: <http://WFRepo/MSP/workflow/>
SELECT *
WHERE {
   ?workflow a msp:workflow.
   ?workflow dc:creator ?creator.
   ?creator foaf:name ?creatorname.
   FILTER (REGEX (STR (?creatorname), ".*Magali.*")).
}
```

Code snippet I – SPARQL query for Frederica's use case, searching for a workflow created by a specific individual.

	workflow	creatorName
1	msp:workflow/MPAConnectivityBSV0.7	"Magali D. do Patrocínio Gonçalves"
2	msp:workflow/BathymetryDataPrepNSV1.0	"Magali D. do Patrocínio Gonçalves"
3	msp:workflow/WindFarmEnvironmentalSuitabilityV0.7	"Magali D. do Patrocínio Gonçalves"
4	msp:workflow/windFarmSpatialSuitabilityV1.0	"Magali D. do Patrocínio Gonçalves"
5	msp:workflow/WindFarmSpatialSuitabilityV1.1	"Magali D. do Patrocínio Gonçalves"
6	msp:workflow/WindFarmEnvironmentalSuitabilityV0.71	"Magali D. do Patrocínio Gonçalves"

Figure 21 – Results of SPARQL query from Frederica's use case, searching for a workflow created by her co-worker (from GraphDB)

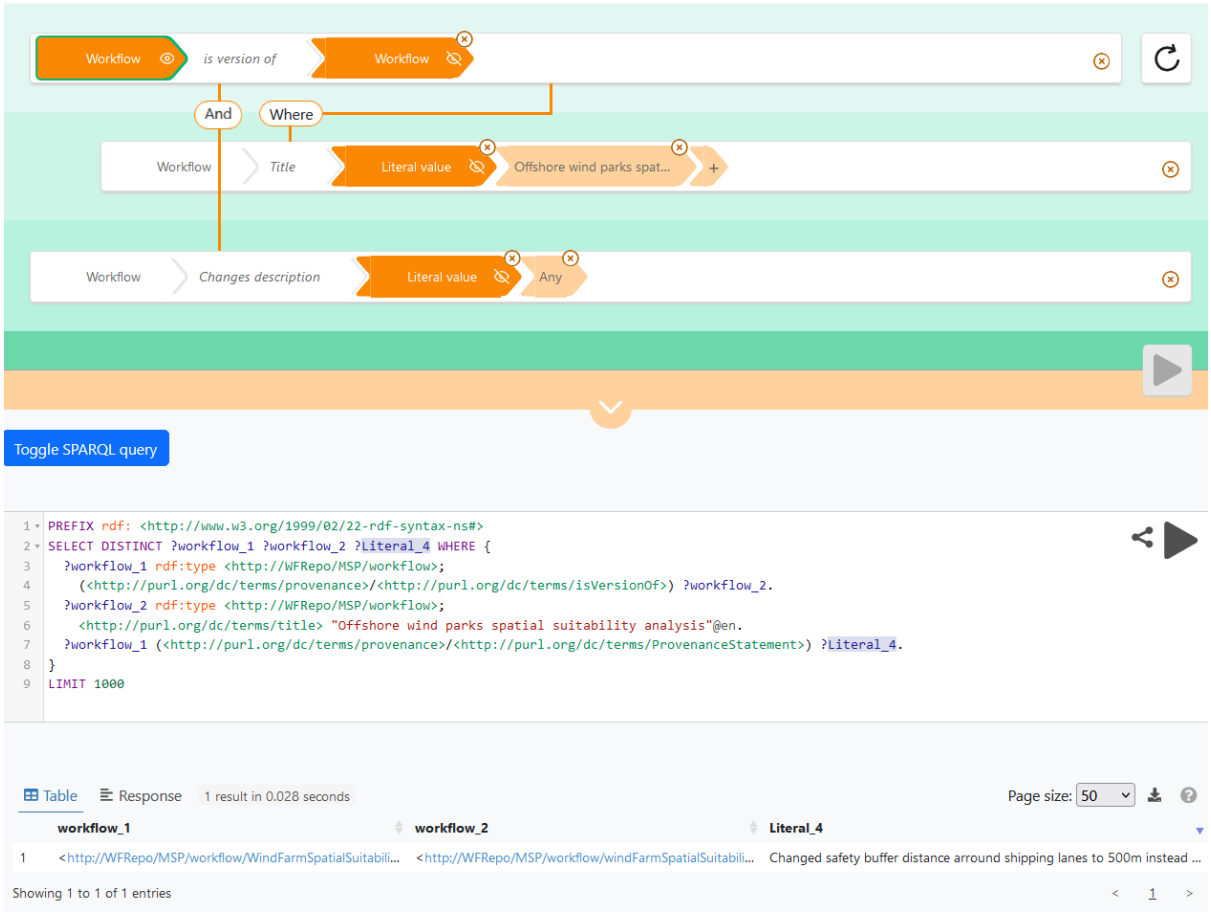


Figure 22 – Example of a query which needs to retrieve more than the first subject of the query, in this case, the two related workflows and the difference between them.

3.5. Approach's Evaluation

The success of the proposed method in describing workflows in facilitating workflow findability, accessibility, tracking of workflow evolution, and reusability was evaluated qualitatively.

Volunteers were asked to undertake a relevant predetermined set of tasks based on the use cases and the research goals, such as "find a workflow that provides a suitability map" or "find a workflow suitable for a particular type of data input". Participants were given access to the prototype and a (reasonable) time limit but no instructions on how to perform the task. After each task, volunteers were asked what the difficulty level was in accomplishing the tasks given. Next, they were asked if they could understand a small workflow's goal and workflow steps and if they found the information provided regarding the workflow was sufficient to decide whether to use it. The tasks and questions asked to the participants with the intent behind them are in Appendix VI – Questionnaire. The participants were informed of the research goals, and a consent form was provided. The questionnaire was implemented in Qualtrics[47].

The target audience was people acquainted with geodata and with a planning or technical background. Thus, volunteers were recruited from the 3rd year of the Built Environment program at Breda University of Applied Sciences; researchers and interns from the same education institute also volunteered. In total, eight persons participated. Table 5 shows the participant's characteristics.

The testing/surveying was anticipated by a short introduction of the research goals, workflow, linked data, and web semantics, as well as showing the workflow and theme ontology. Next, a brief demo of Sparnatural was made, demonstrating the relationship between workflow and data, showing the vocabulary implementation, and that the "literal value" needs to be used for text-based searches. After that, the survey page was shown, with the first page referring again to the research goals and requiring consent from participants.

The setting during the experiment was a quiet room with a 34-inch screen connected to the computer hosting the database, the Sparnatural customized website, and the webpages with the workflow diagram and metadata. On the left side of the screen was the Qualtrics survey with the tasks to perform and the questions; on the right side of the screen was the Sparnatural site to build the queries (Figure 23). In total, there were ten tasks, and for each task, testers had to find and, or interpret a workflow (see Questionnaire in Appendix VI. Testers were given access to the prototype and a (reasonable) time limit but no instructions on how to perform the task. However, sometimes, minor tips were given regarding how to use Sparnatural. During the testing, users filled in their findings and the difficulty level in accomplishing the tasks given.

When opening the customised Sparnatural webpage, users can choose to search for workflows or Data. Independently of the subject chosen, one can search for related workflows or data, search per creation date, vocabulary term (defined by the different ontologies; workflow, purpose or domain – see Section 3.3) or literal value. Naturally, the vocabulary terms and the literal values associated with the workflows differ from those associated with the data (Figure 23). All literal values are self-explanatory, except maybe the data "contact point" – the person to contact regarding a dataset, and the "requires" term, which is meant to be used to select datasets regarding marine or maritime uses that require a buffer zone. By turn, the vocabulary terms are linked to a label (literal value), which is easier to read and to other vocabulary terms as defined in the domain ontology defined in Section 3.3.1. The vocabulary terms were shown or selectable by users; instead, they would have to specify the vocabulary's label or connect it to another vocabulary term employing one of the terms: "same as", "part of" or "has part".

Using the query builder, users can search for all workflows with input or output data tagged with a

[47] www.qualtrics.com

certain vocabulary term. The user will use the graphic interface, and Sparnatural will build a SPARQL query (editable for more advanced users) and retrieve the results from the triplestore. An example is provided in Figure 24. The results are shown under the query; by default, only the first selected subject (data or workflow) appears in the result; users can use the "advanced mode" to edit the query to show also the objects and predicates or show more than one subject if that applies. One example of the need to show several subjects and objects on the query's results would be to see the data associated with a workflow, or a workflow related to another workflow, as shown in Figure 25. The queries' results consist of URIs, which users can click on to open a new page with more information, provided the webpage exists. As mentioned in Section 3.4.1, to provide users with the workflow details, a webpage containing the workflow diagram and metadata was created for each workflow.

Conclusions were drawn regarding the effectiveness and usability of our workflow description solution based on whether the participants could find (all) the workflows they were asked to find and whether they understood the workflow diagram and found the metadata enough. Relevant comments and suggestions during the participation were also noted, and the interactions were recorded (voice only) to ensure they could be replayed to take further notes.

Participant no.	Occupation	Background	Familiarity with…		
			Geodata	workflows	Web Semantics
1	Student	International media and entertainment	Slightly familiar	Extremely familiar	Not familiar at all
2	Post-doc researcher	Spatial planning, msp, coastal governance, GIS	Very familiar	Moderately familiar	Slightly familiar
3	Researcher	International Spatial Development	Moderately familiar	Not familiar at all	Not familiar at all
4	Student	Built Environment (Urban Design)	Slightly familiar	Very familiar	Not familiar at all
5	Student	Built Environment	Slightly familiar	Slightly familiar	Not familiar at all
6	Student	Built Environment	Moderately familiar	Not familiar at all	Not familiar at all
7	R & D Project Leader	Serious/simulation game R&D	Very familiar	Very familiar	Slightly familiar
8	Interactive Media Designer	Interactive Media Design	Very familiar	Very familiar	Very familiar

Table 5 – Characteristics of questionnaire respondents

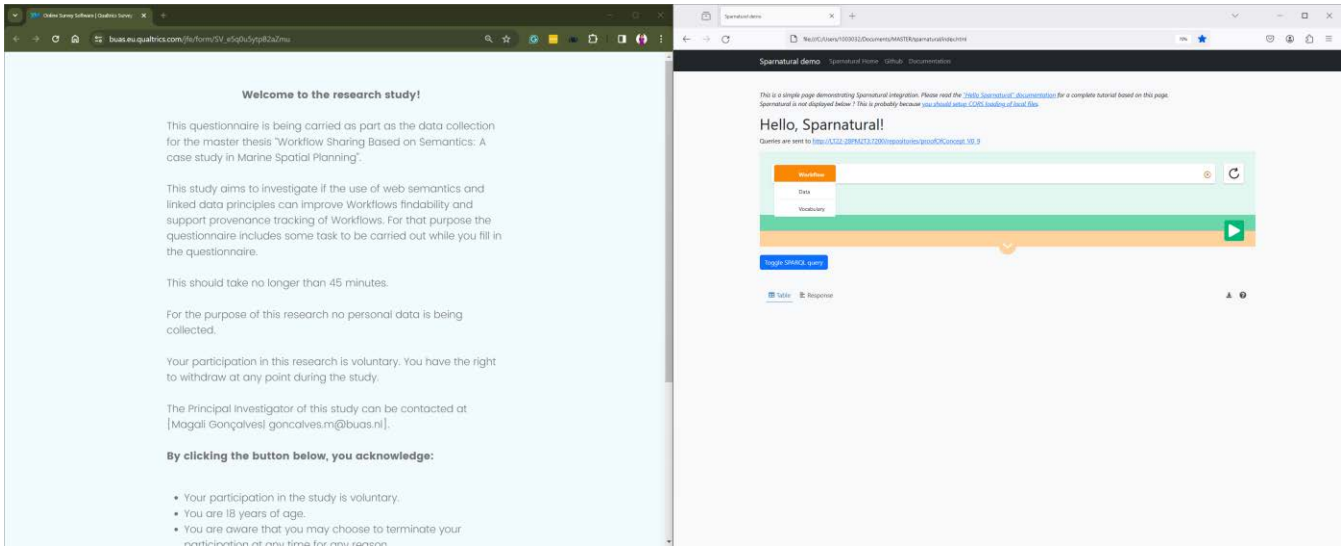


Figure 23 – Set-up of the testing

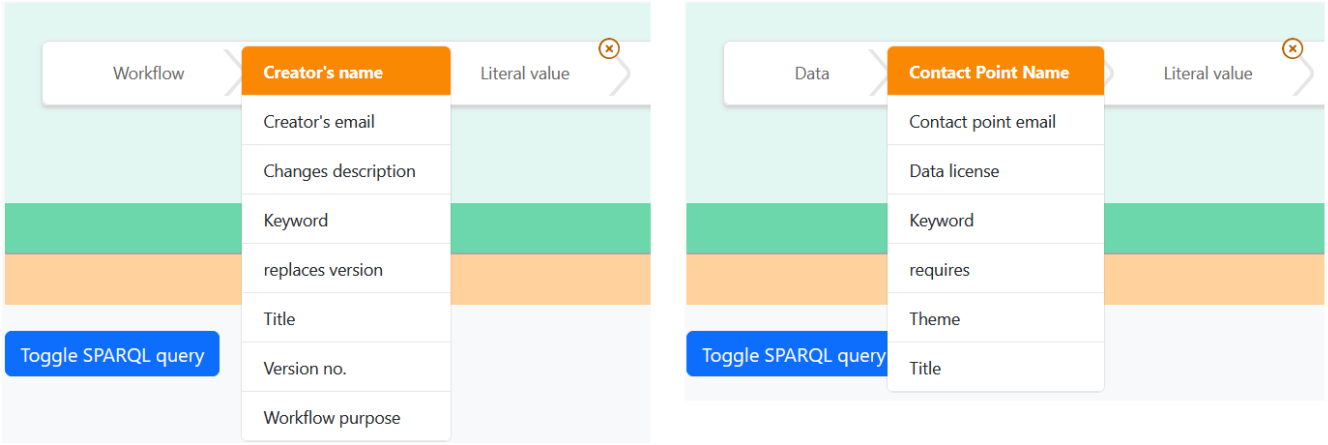


Figure 24 – Search options in Sparnatural when selecting the literal value relationship. Options for workflows on the left-hand side, and for data on the right-hand side.

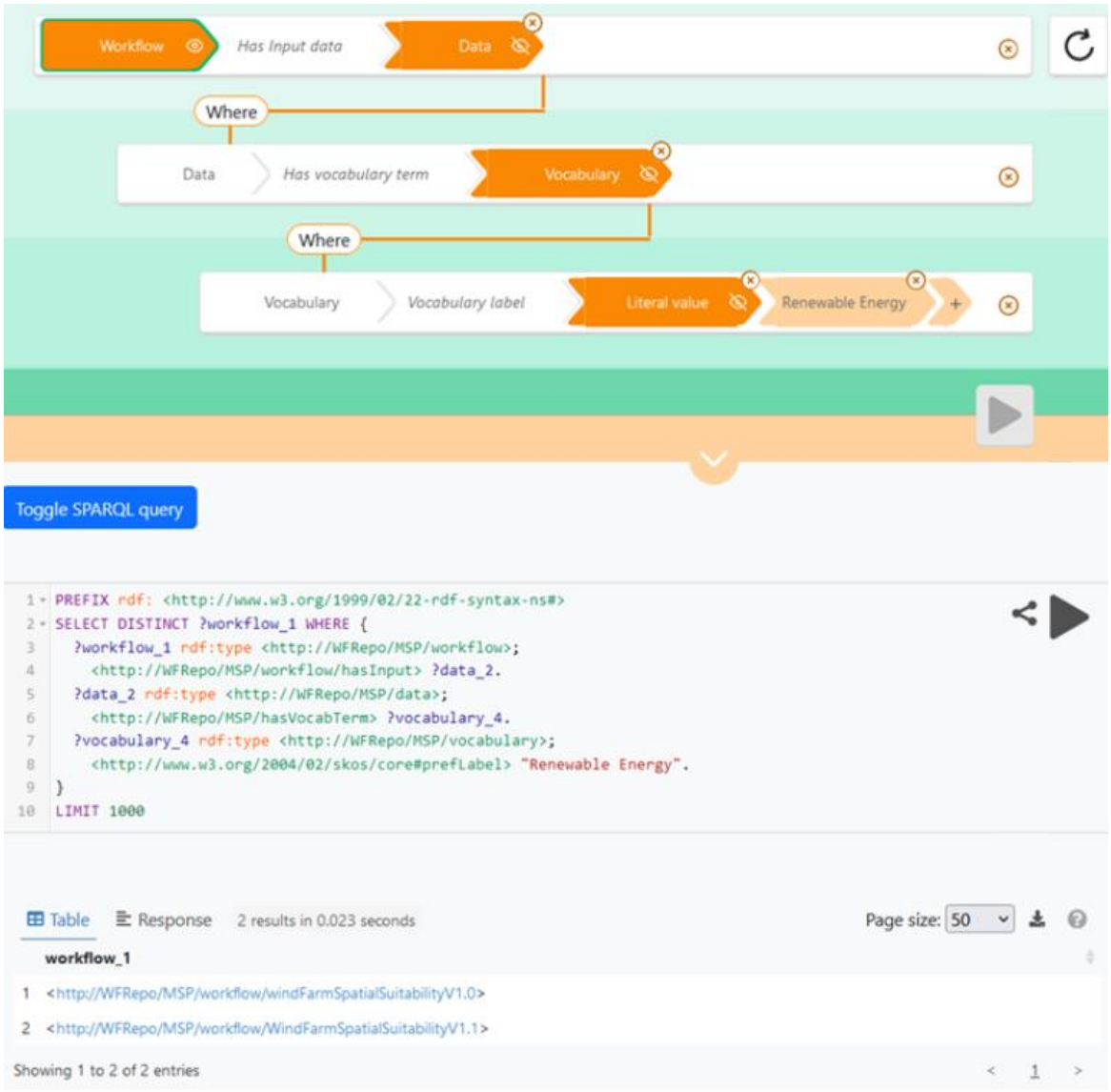


Figure 25 – An example of a query in Sparnatural using the workflow and domain ontologies.

4. Results

As mentioned in Section 3.5, after the volunteers experimented with the workflow repository prototype and filled in the questionnaire (Appendix VI), the results were gathered and stored in an Excel file for analysis and to create visualisations to support their interpretation. The raw results from the questionnaire are in Appendix VII and were analysed against the repository requirements of findability, accessibility, tracking workflow evolution, and reusability defined in Section 3.2. Additionally, observations are discussed regarding the participants' current use of workflows, workflow repositories and potential uptake of a geodata workflow repository that uses linked data.

4.1. Findability

The findability criteria defined in Section 3.2 are discussed individually in this Section.

Req. 1.1 Retrieve workflows by author or author's affiliation.

This requirement was analysed mainly through the first two questions below. However, the last four questions have also evaluated similar tasks:

- p2a - How many workflows created by someone at Breda University of Applied Sciences (AKA BUas) can you find?
- p2a.d - Please indicate the level of difficulty in replying to this question.
- p2h How many workflows have been created since February 2024?
- p2h.d Please indicate the level of difficulty in replying to this question:
- How many Workflows use input data from the HELCOM Secretariat?
- Please indicate the level of difficulty in replying to this question

All participants could find the six workflows in the repository by searching by BUas as a creator. Half of them found the task to be of medium difficulty. In contrast, the other half was divided between difficult and easy (2 each) – Figure 26.

The main difficulty of this task was linking the workflow and the literal value and only then being able to select what kind of literal value was being searched for (creator). Interestingly, there was a similar question, asking how many workflows had been created since February 2024; all participants found this question easier – Figure 26. For this query, users need to select a date and then the creation date, similar to the previous query, where they need to choose the literal value and then which literal value. As some participants mentioned, "literal value" does not mean much to most users. Additionally, several properties fall under that category, making it more difficult to find the correct one.

Nevertheless, it is visible that there is a learning curve since participants were also asked to find workflows that used input data provided by the Helcom Secretariat. This query had an additional step: linking the workflow to the input data and applying a literal value. Despite the bigger complexity, users did not find it much more difficult than retrieving the workflow per author – Figure 26.

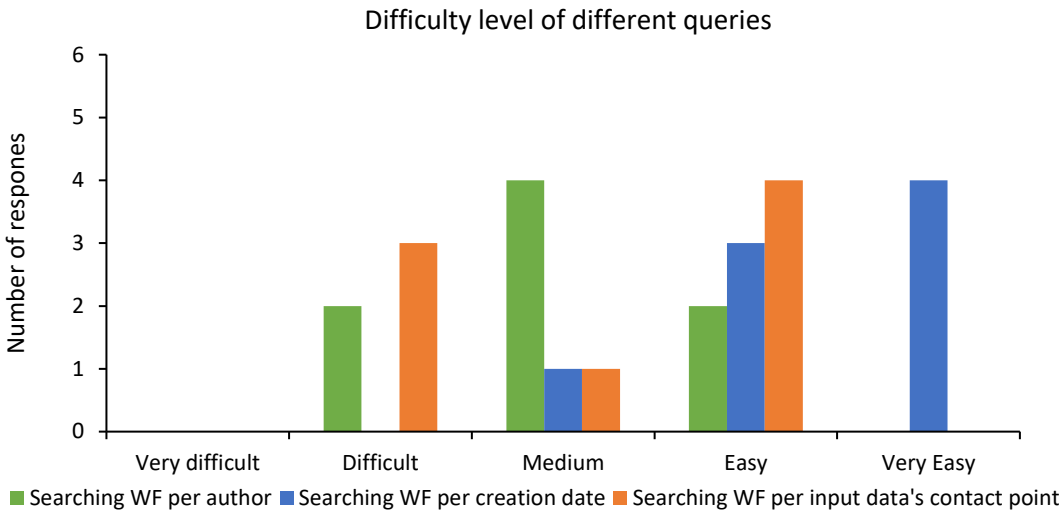


Figure 26 – Difficulty finding workflows by author, filtering by creation date, and by the contact point of the input data. The association with the date field is easier than with the literal value, and although the complexity of the last task is bigger than the first two tasks, it is not much more difficult than the first one.

Req. 2.1 Retrieve workflows by type of analysis.

This requirement was analysed through the questions:

- p3a - There is a Workflow to support decision-making over the allocation of Marine Protected areas (MPAs) so they can act as a protection network. Can you find it?
- p3a.d - Please indicate the level of difficulty in replying to this question.

This question was thought to trigger participants to search for a workflow using the type of analysis; nevertheless, not all participants searched by the expected field (workflow purpose). The mention of the Marine Protected Areas can be why some participants searched the workflow by input data. Nevertheless, at least two participants successfully used the workflow purpose for another task (finding the environmental suitability workflows), indicating that this field is useful for users.

All but one participant found the network analysis workflow; half considered it easy to find, two medium and one very easy - Figure 27. The participant who did not find the workflow classified the task as very difficult. This participant eventually searched per workflow purpose using the term “allocation”, retrieving the spatial and environmental suitability workflows together with the desired workflow. The tester tried adding the term “marine conservation” to filter out the undesired workflows, but that acted as an “or” query instead of an “and" query, which was not useful (Figure 28). For Sparnatural to build the query properly, one would have to add an “and” proposition to the query by hovering under “Workflow” and adding another workflow purpose (Figure 29). This is not an obvious solution for everyone, so these options must be clear to users beforehand.

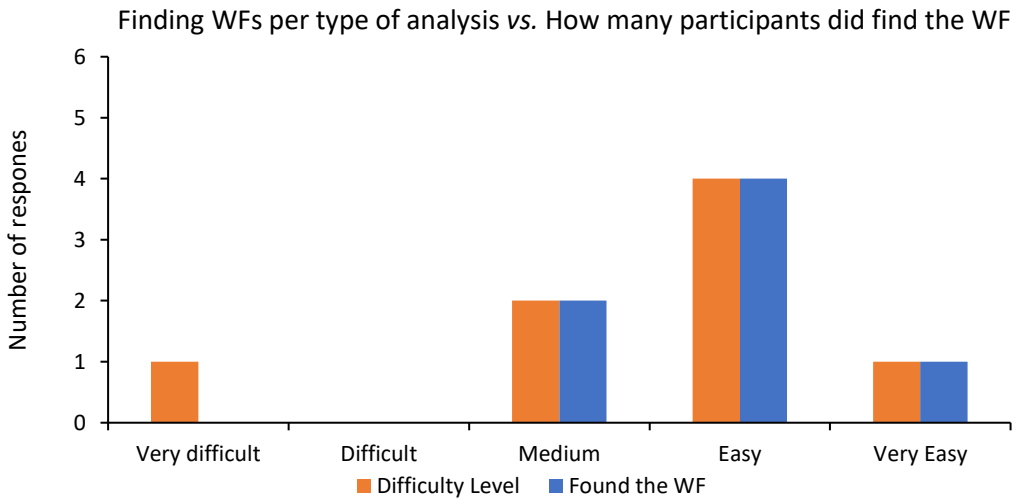


Figure 27 – Finding the workflow per network analysis was not deemed difficult; nevertheless, one user did not find it.

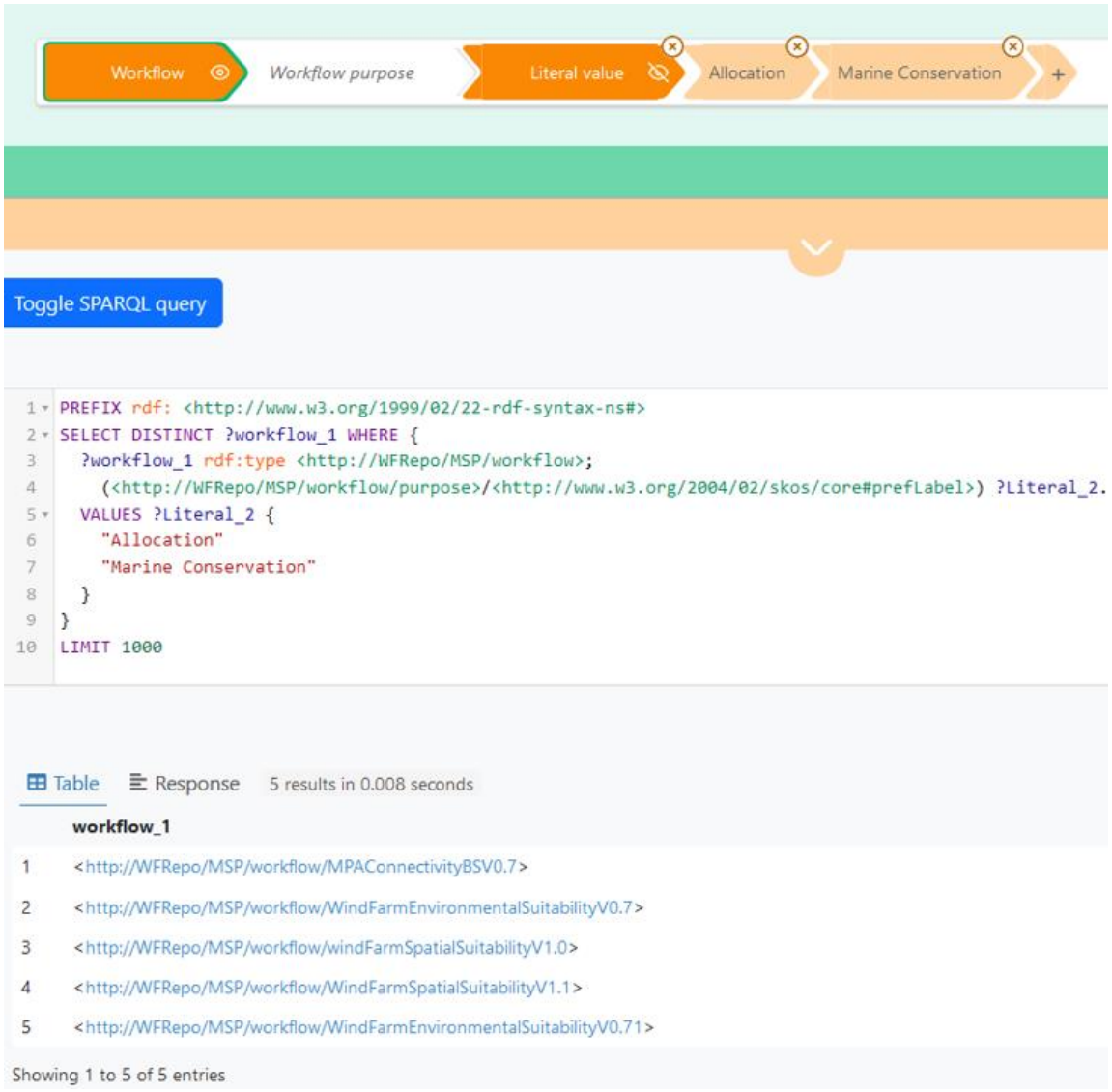


Figure 28 – Searching for the network analysis/allocation workflow by querying the workflow purpose using two different workflow purposes. It acts as an “and” query.

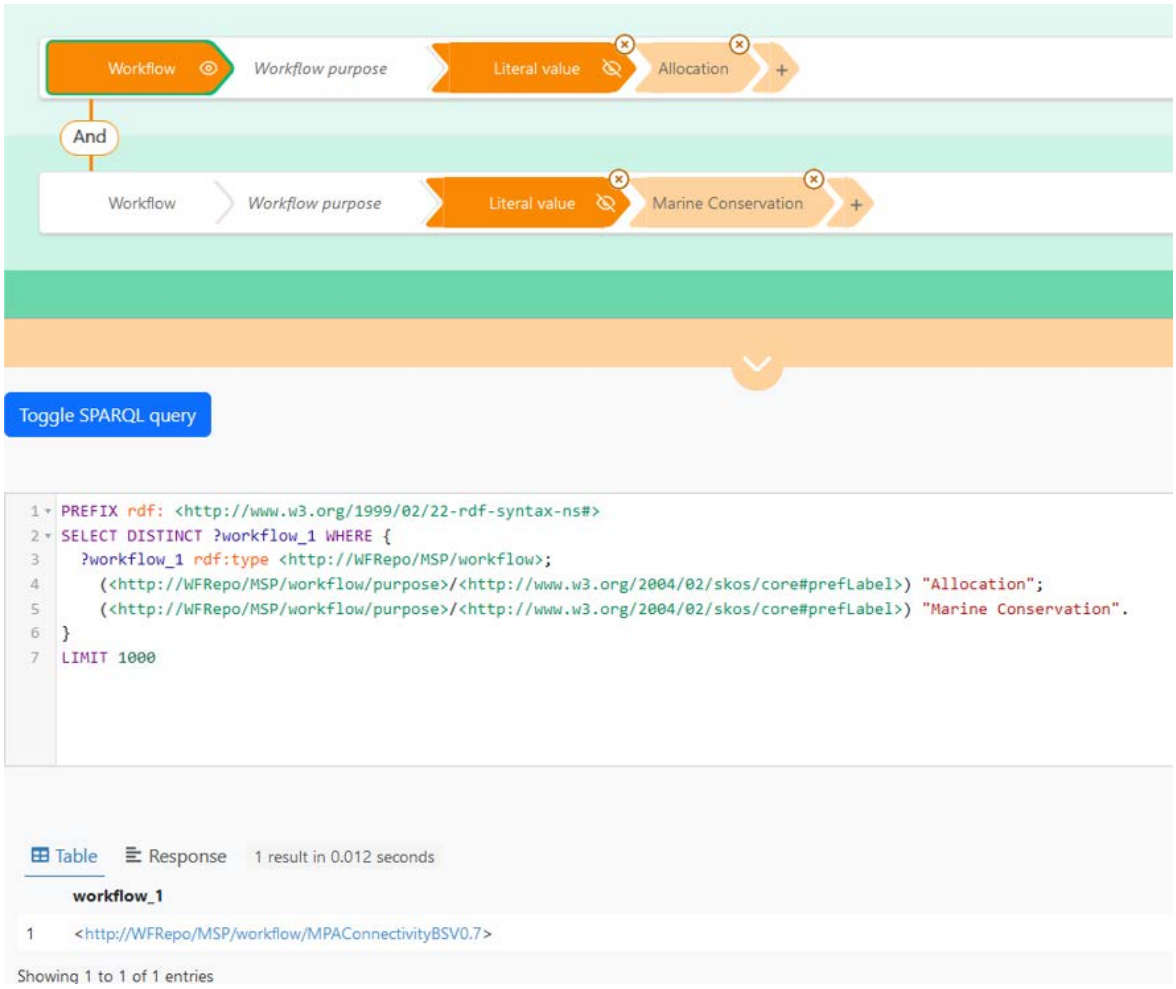


Figure 29 – Searching for the network analysis/allocation workflow by querying the workflow purpose using two different workflow purposes. It acts as an "or" query.

Req. 2.2 Retrieve workflows by data theme

This requirement was analysed through the questions:

- p2e - How many Workflows can you find that have input data regarding environmental conditions?
- p2e.d - Please indicate the level of difficulty in replying to this question.
- p2g - How many Workflows use input data related to renewable energy with data from 2022/01/01 onwards?
- p2g.d - Please indicate the level of difficulty in replying to this question.

Although these tasks were not considered easy by the participants, they all could find the three workflows that corresponded to each criterion, except the first participant, who only found two workflows defining Spatially Suitable Areas instead of three. This seems to have been due to a glitch in the database, not how the query was built.

Half of the participants replied that the first task was of medium difficulty, three found it very difficult, and one found it easy. Regarding the second task, two participants reported it as medium difficulty, two as easy, and two as very difficult, with one as very easy and one as difficult - Figure 30. The second task required an extra step in the query (the date of the input data), and revealed to be more difficult for the participants than the first task.

Besides domain vocabulary terms, the data's metadata had keywords; both options were kept as searchable. The vocabulary terms from the domain ontology were more useful to the users than the keywords; those should have been removed from the search options so as not to overwhelm the users.

The domain vocabulary was developed to assist findability by using related terms, or equivalent terms. Since the relationships between terms were assigned to the terms' URI, it meant that after defining the relationship of the vocabulary terms, users had to add the label of the term (the literal value) to be able to search for it. Another option would be to focus on the last part of the vocabulary term, which would not be user-friendly. An example of such a query is shown in Figure 31, using Renata's use case, searching for a workflow related to Marine Conservation.

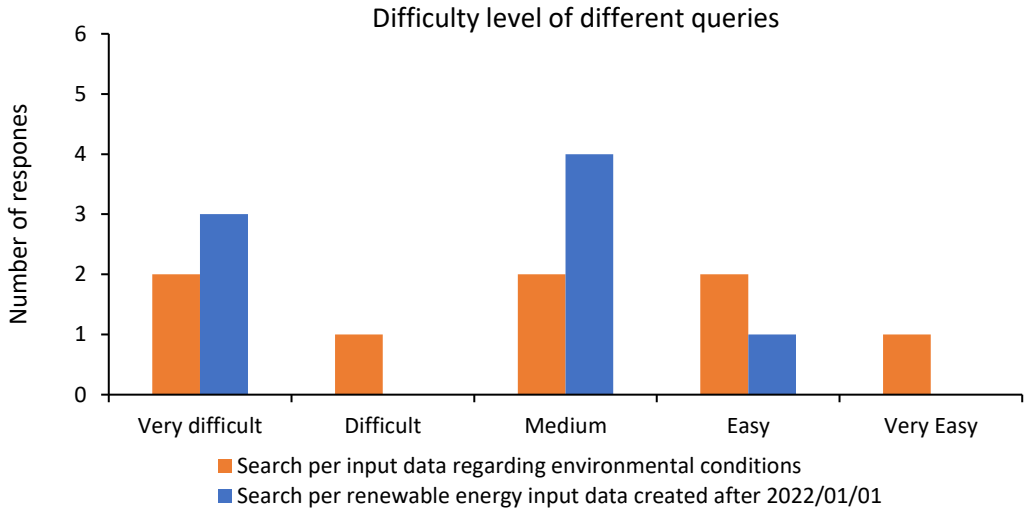


Figure 30 – The second task was more complex than the first one and was considered more difficult by the participants.

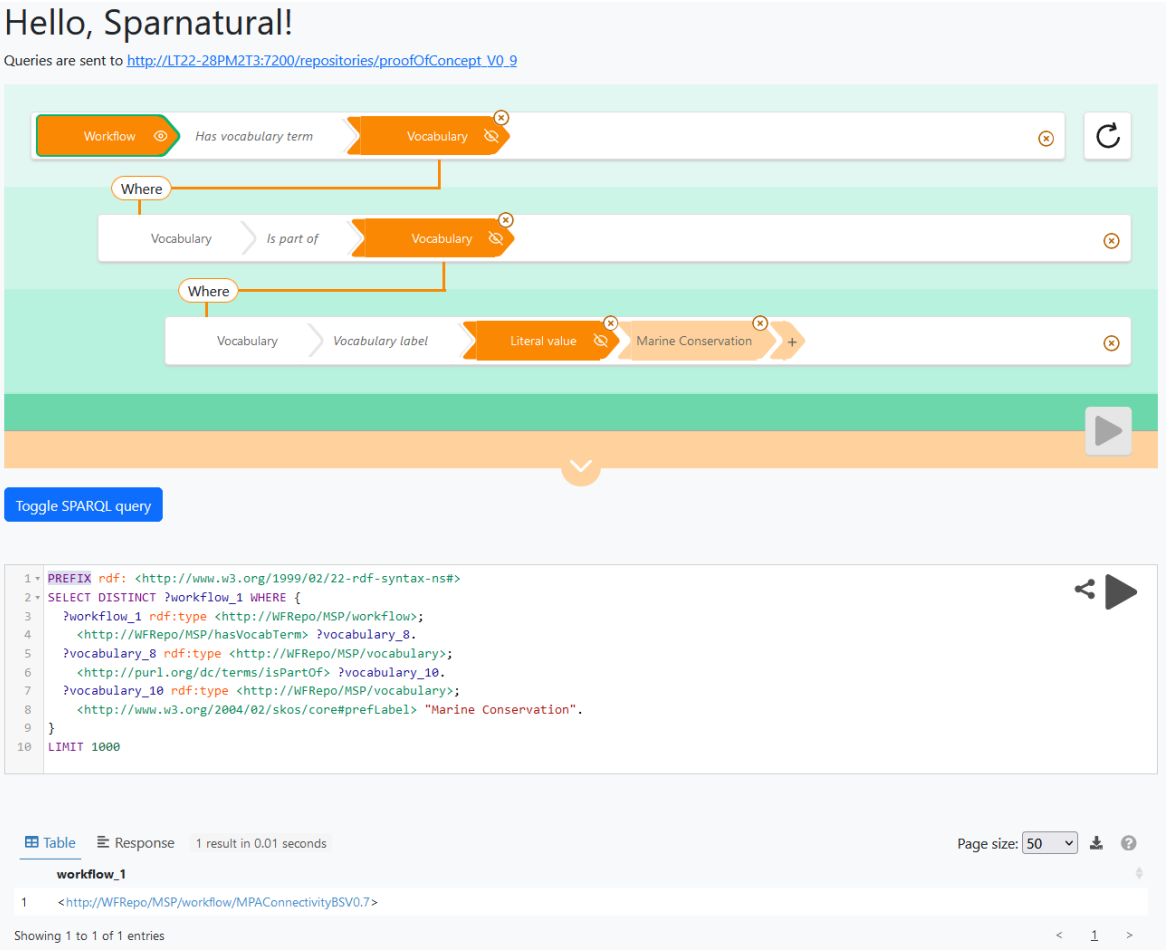


Figure 31 – Using Sparnatural to build an SPARQL query for Renata's use case: searching for a workflow related to Marine Conservation

Req. 4.1 Retrieve workflows that produce similar outputs.

This requirement was analysed through the questions:

- p2b - From those workflows, how many define Spatially Suitable Areas? (from the workflows found in question p2g – with input related to renewable energy).
- p2b.d- Please indicate the level of difficulty in replying to this question.

Needs further testing; most participants did not try searching by output type; they used the workflow purpose or looked at the workflow page to see the workflow description. The difficulty level varied from very easy to very difficult, depending on the method chosen to undertake the task. Nevertheless, users who looked at the workflow description would not have found it easy if the previous query had retrieved many results; this could indicate the need for more filtering options after the results.

4.2. Accessibility

The workflow metadata requirements in Section 3.2 were defined to support full workflow metadata description. Question p3b.3 intended to assess if the users deemed this information enough; if participants indicated it was insufficient, they could specify what they observed as missing.

- p3b.3 - Is the metadata provided enough for you to decide on whether to (re)use one of the workflows in the repository?
- p3b.3.no - If not, what are you missing?

Only one participant found the metadata insufficient; even after all the quality control was done on the input data, one input dataset did not have the license information. Therefore, this comment can be considered a good example of the importance of all the metadata, not be given too much importance in evaluating the method of describing workflows or deeming the metadata insufficient. However, it does bring up the importance of good, thorough quality control.

4.3. Tracking workflow Evolution

The workflow evolution criteria defined in Section 3.2 are discussed one by one in this Section.

Req. 2.3 Save and retrieve parameter values at the operation level.

This requirement was analysed through the questions:

- p2c - Can you understand the differences between the workflows found in the question above? (Finding two workflow versions of Spatially Suitable Areas)
- p2c.d - Please indicate the level of difficulty in replying to this question

The difference between the workflows was only the value of one parameter; in one version, the safety buffer distance around shipping lanes was 500m instead of 1000m.

Most participants tried to see the difference between workflows by looking at the diagram at first, but with such a small change, participants were not finding the difference and were unsure where to look. At some point, the tip that there was a version control in the metadata was given, so most users found the difference in the webpage of the workflow. Some users were curious whether they could query this through Sparnatural; the option was there (Figure 32). Nevertheless, this query was more demanding for users who were getting acquainted with SPARQL. The “Literal_6” corresponding to the description of changes made in the workflow must be added manually in the SPARQL query box to be included in the results.

The answers from the difficulty level ranged from very easy (2) to difficult (1), with three participants classifying it as medium difficulty and two as easy. Nevertheless, one of the respondents who classified the task as very easy did not find the difference between the workflow versions – Figure 33.

The exact parameter from each operation in the workflow is stored in the workflow description and can be queried in GraphDB - see Appendix IV, query 7. There was no way to build this query in Sparnatural as we would only retrieve the first workflow, and there would be no easy way to interpret the results either way.

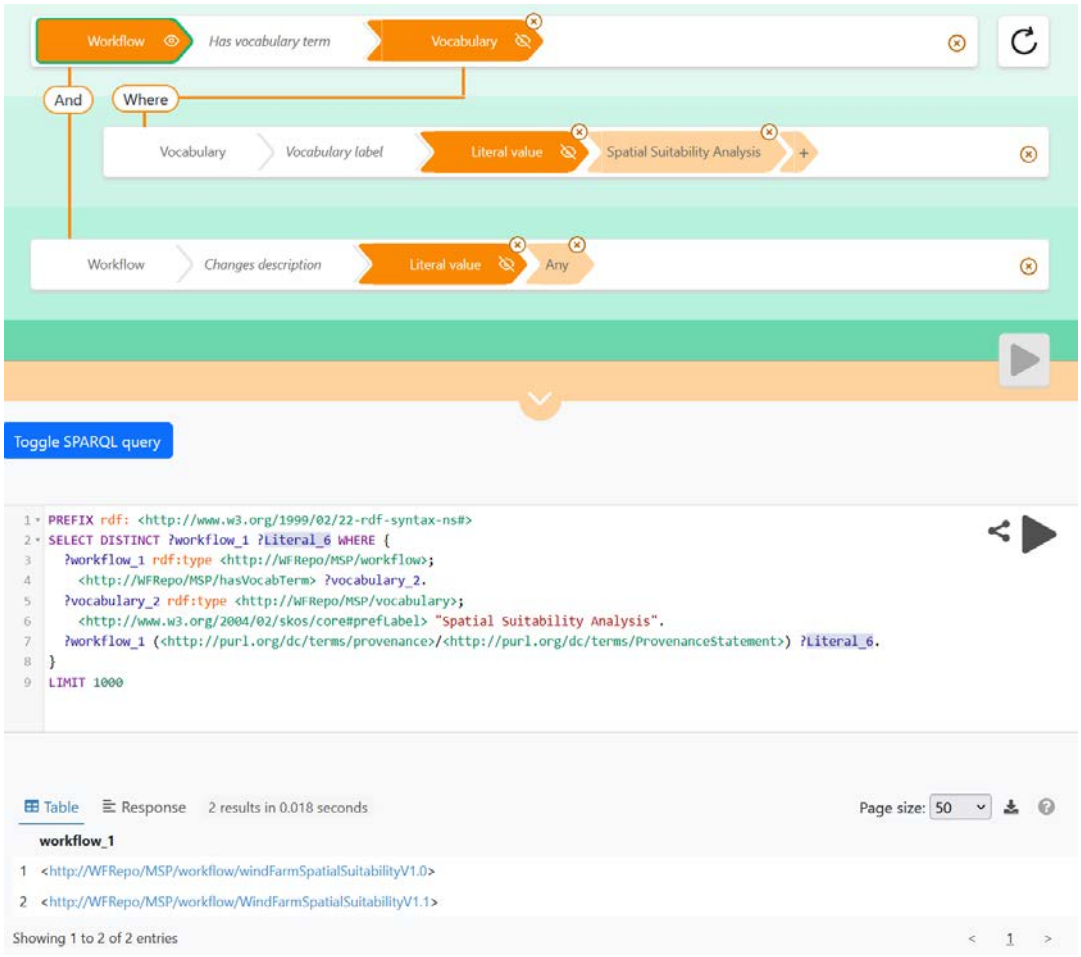


Figure 32 – Querying changes description from different workflow versions in Sparnatural.

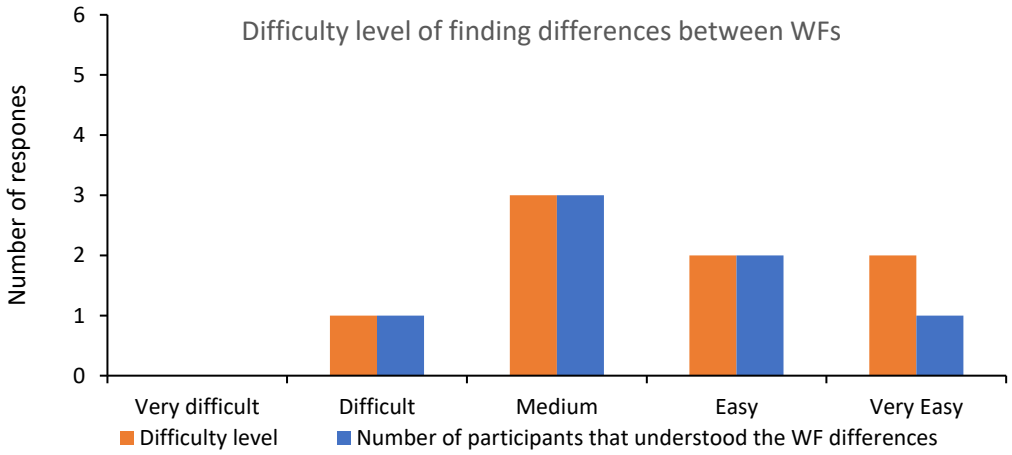


Figure 33 – Understand the difference between workflows was only found difficult by one responder but one did not find it.

Req. 3 . 2 Tracking of workflow evolution.

This requirement was analysed through the questions:

- p3b.1 - There is more than one workflow to define environmentally suitable areas. Are they somehow related?
- p3b.1/d - Please indicate the level of difficulty in replying to this question.

All participants found the two workflows and could infer, using different approaches, that they were versions of each other. Possibly as a consequence of the various methods used, some users found the task very difficult (2), others very easy (2), or easy (2), with the two other users finding it medium (1) and difficult (1). Overall, the task was not considered too difficult; that could have changed if the number of workflows defining environmentally suitable areas was big. Then, users would need to query the relationship using SPARQL, with the inconvenience of having many workflows - workflows relationship options: “has part”, “is replaced by”, “is version of”, “replaces”, and “is part of”. It might be too much for users to handle all the options offered, and a sensible choice would have been to maintain only the two “is version of” and “is part of”.

Req. 4 . 2 Combine workflows as sub-workflows in a bigger workflow.

This requirement was analysed through the question:

- p3b.5 - Can you see if there are any workflows that were already used to create other Workflows?

All respondents could ascertain that workflows in the database were used to create other workflows, probably because the question just before this one reminded them that there were workflow – workflow relationships. One participant inferred this from the workflow diagram legend – Figure 34.

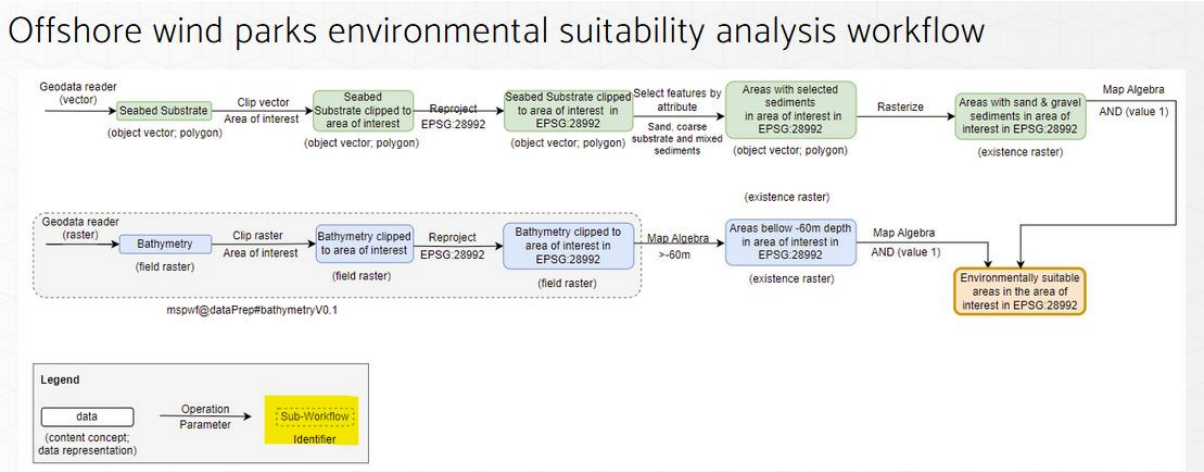


Figure 34 – workflow diagram with a sub-workflow reference

4.4. Reusability

The workflow reusability criteria defined in Section 3.2 are discussed individually in this Section.

Req. 3.1 workflow needs to be understandable by others.

This requirement was analysed through the questions:

- Q45 - If you look at the workflow diagram, can you clearly understand the steps of the workflow and the data transformations? (at this point, participants were looking at the network analysis workflow)
- Q51 - What would you do with this Workflow?
- p3b.2 - Is there an environmental suitability analysis workflow you would choose over the other? Why?
- p3b.2.d - Please indicate the level of difficulty in replying to this question.

Five users mentioned it took some effort to understand the workflow diagram, and the other three considered it clear. Admittedly, the workflow chosen to interpret the diagram was not the easiest one, and it included the "distance to nearest hub" operation, which people less familiar with GIS would not immediately recognize. Additionally, when finding the spatial suitability workflows, three participants tried to interpret the workflow diagrams and succeeded, indicating that the workflow diagrams are clear.

Three participants mentioned they would take ideas from the network analysis workflow and implement it in their GIS of choice; two would improve it, and the other two responders said they would run it several times with different scenarios. Understandably, the participant with a media background could not find a use for this particular workflow.

When asked if it was easy to choose a version of an environmental suitability analysis workflow, the answers ranged from very difficult (2) to very easy (1), with three participants finding it easy and two medium. The main difficulty of choosing between these two workflows seems to be related to their similarity. With a bigger sample size and maturity status sorting, this subject could have been better evaluated.

Req. 3.3 Sorting by maturity status.

Although the metadata was in most cases deemed enough to decide to reuse a workflow, deciding which workflow to reuse in case of different versions with the same maturity and slight differences was difficult. The repository/database would need more workflows, with different maturity states and differences between them, to test this better. Additionally, having a ranking in terms of the number of downloads, recommendations from users or "likes" could help.

Req. 3.4 Licence terms need to be clear

This requirement was analysed through the questions:

- p3b.4.1 - What is the license type of the input data?
- p3b.4.2 - Is the license type of the Workflow clear?

The workflow license was stated on the webpage, along with the workflow diagram and its metadata. Not every participant was aware of what it meant. Nevertheless, all of them found the type of license. For the input data, one of the participants stumbled upon an input dataset without the license information.

4.5. Participants' Experience and Suggestions

None of the participants had ever used WfMSs; only one mentioned using a workflow repository: GitHub. Three participants were very familiar with Geodata; most had acquaintance with workflows, with only two respondents noting they had no workflow knowledge. Only one participant was very familiar with web semantics; two were slightly familiar, and the rest were unfamiliar. So, the participants' experience on the topic was limited.

When asked if they could imagine themselves using a repository like the prototyped one, most replied yes (question no. p4d - Do you think you would use a repository like this for your workflows?). Only the International Media and Entertainment student replied "no"; although they are familiar with workflows, they meant conceptual workflows of how a production is run, for instance.

If volunteers responded yes to the previous question, they were asked, "For which purpose would you use such a workflow repository?" (Question p4d.2_1). This question had multiple options and allowed selecting more than one option: "To search for workflows", "To share my workflows", "To keep track of my workflow versions". All participants choose "To search for workflows", with only three participants also willing to share their workflows. Three other participants also found it useful to keep track of their workflow versions. Two participants selected all the options. This allows us to conclude that workflow repositories are valuable for repetitive and complex tasks, as well as for searching, consulting, referencing, and sharing information.

Finally, volunteers were asked if they missed anything in the prototype and if they had any suggestions (open questions):

- (p4e) Do you have any suggestions for a geodata workflow repository?
- (p4f) Did you miss something in the prototype?
- (p4g) Do you have ideas on improving the workflow description or workflow repository?

Since most volunteers had little experience with workflows or workflow repositories, it was difficult for some to provide suggestions. Nevertheless, a few of them could provide suggestions as mentioned below. They are organised into three categories: user interface, vocabulary, and workflow repository.

User interface

SPARQL is not a very intuitive way to query a database. Sparnatural helped bridge that gap, but users needed to select a subject and an object, and only then the predicate. That was very confusing for testers, especially when they had to search for subjects through literal objects, for instance, searching for a workflow with the title "x". First the subject workflow needs to be selected, next the "literal value", and only then can users select the title as a predicate and specify what they expect to be the title.

The colours, gradients and highlights while using Sparnatural could be improved; they were sometimes very confusing. For instance, after selecting the subject to search for, users had to select the object; when clicking on the selection menu, the first option was marked in bold and highlighted as if it was selected. This would lead to either testers thinking they had already made a selection or that the highlighted term was the menu's header and was not selectable.

The option to add another statement to the query is not obvious; the "and" button is only visible when hovering below the subject. It is also unclear if adding another term in the search box will act as an "or" or "and" query - Figures 28 and 29.

The fact that only the first object of the query would appear in the results automatically was limiting. For more details regarding the results, one would have to add manually to the SPARQL query.

There could have been more tooltips when users hovered over a term; those would have been easy to define.

Additionally, it was not obvious to testers that when querying for all objects related to a subject with a specified relationship (without specifying the object or any associated properties), they had to select “any”. It would have been more obvious to users to click on “all”.

Vocabulary

Using more familiar terms for predicate and object names could provide less savvy users with a clearer and more straightforward experience. For instance, using “properties” instead of “literal value” could have made it more clear that the metadata regarding the workflow was “hiding” under that term. Of course, this was done, as explained in Section 3.4.3, but it needs further refinement.

Mention if a certain term in the vocabulary is a parent or a child of any other term.

Workflow Repository

To ensure that all users know how to use the repository, it should provide demo videos, manuals, and user support or a community forum.

It would be nice to have the possibility to click on an operation in the workflow diagram and access its description and, eventually, different ways to run it, already relating to the workflow’s concretisation. This became clear regarding the “distance to nearest hub” from Question Q45.

Including user feedback options, such as the possibility of rating a workflow or providing feedback to the author, could be useful in assisting other users in deciding whether to use a workflow.

There could be an RDF comparison tool to analyse differences between workflows to support comparing different workflow versions. Another simpler feature mentioned to support the same task was to automatically highlight anything that is changed in a workflow when saving it as a new version.

5. Discussion and Conclusion

This chapter discusses the results of the research methods and their limitations. Concrete recommendations for metadata, workflow description, and features for a geodata workflow repository are defined, replying to the research questions and general conclusions are drawn. The suggestions for future research are also included.

5.1. Replying to the Research Questions

This section revisits the research goals and replies to the research questions outlined in Section 1.2. Each sub-question of each sub-goal is addressed separately.

Question 1.1 - What information about the workflow and workflow data needs to be described and stored?

Sections 3.2 and Section 3.3 describe the metadata and the workflow description requirements suggested to achieve the research goals. The workflows were described taking those requirements into account, and volunteers tested the approach's effectiveness in supporting the research goals.

Regarding the workflow metadata, the literature (Section 2.1.3 and Section 2.1.4) and the workflow repositories consulted (Section 2.1.5) were crucial to deciding the metadata fields that would help us support our goals. Section 3.2 defines the workflow metadata that supports our needs. Furthermore, the author would suggest the following fields to be mandatory to load a workflow in the repository:

Descriptive Metadata

- Title: the title of the workflow (literal value).
- Identifier: a combination of tokens (letters, symbols and signs) identifying the workflow.
- Vocabulary term(s): terms from the domain vocabulary that act like keywords
- Purpose: The purpose of the workflow is set in the workflow vocabulary and relates to the type of data handling/analysis of the workflow. This can be more than one vocabulary term and be related to other vocabulary terms employing "is part of"/" has part" from OWL.

Workflow Provenance

- Creation date: the date on which the workflow was created.
- Author: person or organisation that created the workflow.
- Author's contact: email address from the author/contributor. This term is associated with the author or contributor, respectively.

Workflow Evolution

- Version number: The version number or the workflow.
- Maturity: Use the terms work in progress, stable and deprecated, respectively, for workflows still under development, completely defined or replaced by other versions.

Structural Metadata

- Input data: URI of the input data, which has its own metadata file.
- Output data (intermediate and final): URI of the output data, which has its own metadata file.
- Data type: each dataset in the workflow should refer to its data type to enable automation of the workflow execution.
- Data vocabulary term: each dataset in the workflow should refer to the domain vocabulary that acts like keywords

Administrative Metadata

- License: The type of license of the workflow, so it is clear to users whether they can use the workflow and under which terms.
- Language: The language used in the workflow description is preferably English.

The metadata regarding the data flowing in the workflow should follow the ISO 19115-1:2014 Standard and Inspire Directive as mentioned in Section 2.1.4. The geodata data types defined by Scheider et al. (2020) should also be added to the metadata to facilitate workflow concretisation, as mentioned in Section 3.3.

Looking at the survey results, in general, the participants did not miss any metadata regarding the workflows, but one participant found one input dataset without license information. The participant missed that information when deciding whether to reuse the workflow, confirming that that information is crucial for that decision. This problem could have been anticipated and solved by contacting the data provider or using an alternative dataset with complete metadata.

Question 1.2 - What standards support storing and sharing of (workflow) metadata?

Section 2.1.3 discussed what metadata means and the different categories and functions. Most of the literature generally refers to data for which standards already exist. However, some also consider workflow-specific metadata focusing on provenance. Additionally, workflow repositories were consulted (Section 2.1.5).

Although there are no standards for workflow metadata, it is clear that due to its composite nature, workflow metadata is needed at different levels: input data, workflow, workflow tools, and publications (Section 2.1.2 and Section 2.1.4).

Ultimately, the decision was to use descriptive metadata, workflow provenance, workflow evolution, Structural Metadata and Administrative Metadata, as mentioned in Section 3.2.

Some workflow metadata properties were described using Dublin Core, just like for the input data (e.g. creator, identifier); others were specific to workflows, such as the input and output data used and sub-workflow; these used specific vocabulary.

Ultimately, the only standards used in the workflow description were related to geodata (ISO 19115-1:2014 and GeoSPARQL) and linked data (RDF).

Question 1.3 - Which linked data format best supports our needs?

Despite the JSON_LD format being popular in the literature and web applications, technically speaking, it seemed that the best format for our needs was TTL see Section 2.2 and Section 3.3. TTL has worked as intended, enabling the implementation of several vocabularies, avoiding redundancies, supporting the system's complexity, and allowing its expansion. The fact that humans can easily read the format has proven to be important since describing the workflows in the context of this research has been a manual task. TTL is a format designed specifically for RDF data and aligns directly with RDF Standards. Ideally, workflows would be created using a graphical tool, and the corresponding TTL file would be created automatically.

Question 1.4 - What (if any) extensions to existing web semantics vocabularies are needed to support the workflow data description?

As mentioned in Section 3.3.2, the ISO 19115-1:2014 could be implemented using mainly the Dublin Core vocabulary for the data metadata and GeoSPARQL. The geodata data types defined by Scheider et al. (2020) were also deemed important to describe the data flowing in the workflow, especially for more advanced users seeking specific types of data or looking to combine workflows since the output of one needs to be compatible with the input of the other. Additionally, the data types could be useful for workflow automation.

The Inspire Directive themes were combined with Marine Spatial Planning terms to enhance the discovery of workflows and data. After scouting the internet for vocabulary used in Marine Spatial Planning, it was clear that none conveyed the concepts and relationships at the level necessary for this research activity. Therefore, the domain ontology was developed based on the literature – see Section 3.3.1.

Terms from the data catalogue vocabulary (dcat) could describe formats and ways of sharing the data. Dcat and the aforementioned vocabularies make the data description quite complete. Nevertheless, adding the attributes from the attributes table or the measurements (in the case of rasters) of each dataset would also be desirable. For different formats and nature of data, different things might be relevant; for instance, for a NetCDF file, the periodicity and accuracy of measurements would also be relevant. Vocabulary to convey uncertainty of modelled data or measurement accuracy also seems not to exist. To further improve this research area, we advise developing the vocabulary in this domain in future research as it can greatly improve it.

Question 2.1 - To what extent can linked data best support the representation of workflows?

As seen in Section 2.1.2, workflows are composed objects; they include data, operations, execution software, and agents. Linked data seems like the perfect way to link all those components and clearly define the relations between them, including details about operations, or reusing parts of a workflow in another workflow. With every component having its identifier and metadata, it becomes clear how each component relates to each other and their provenance.

Theoretically, the advantages of using linked data to represent workflows are querying the workflow and all its components, linking to similar concepts, enhancing the discovery of similar products, and supporting versioning and composed workflows. Furthermore, it has the potential to enable the automated concretization of abstract workflows in different WfMS, making the workflows interoperable.

A workflow ontology was proposed to achieve all the potential advantages. Simultaneously, the workflow ontology aimed to track the data provenance of any dataset within the workflow.

Question 2.2 - How to store workflows described with our method to support discoverability and reusability?

Since the workflows are described using RDF, a triple datastore is needed. However, for ease of access, since not everyone knows SPARQL, an easy, user-friendly way is required to query the database, as mentioned in Section 3.4.3. The workflows are described by including a link to the data used as input and created as output of each workflow step, as well as workflows related to each other, whether they are versions of each other or part of each other. Furthermore, with the domain and purpose ontologies, workflows, and data can also be found due to their relationship to other items in the database.

A triplestore is needed to store the workflow files using the approach developed in this research. The triplestore needs to be configurable on a server that is publicly accessible through an API. For ease of use, the repository should also be accessible through an appropriate web interface.

Additionally, having a user-friendly querying interface or enabling the connection to a platform that offers an easy way to query the database is very important.

Enabling user authentication with different roles and permissions is desirable. Three important roles are administrators, authors and general users. Administrators ensure all registered workflows have a valid description and comply with the mandatory metadata requirements. Authors of workflows share their workflows with the community and submit at least the mandatory metadata fields. All users can search and reuse workflows from the repository, provided the workflow license allows it. Ideally, the system will not allow infringements and will track the provenance of workflow, changes, and evolution. That is why the repository must be connected to a WfMS.

The workflow description method helps discover workflows per any data type used or produced in the workflow and is enhanced by the domain ontology. The workflow description includes the data created by intermediate steps, so any intermediate output can be reused in another workflow instead of being limited to the input and output data.

The fact that the parameter values used are stored in the prospective provenance part instead of the retrospective part of the workflow description (unlike OPMW) makes the planning requirements clearer for users and stakeholders and supports the workflow repository requirement 2.3 - see Section 3.2.

Question 2.3 - Are the workflows described with our method easy to understand?

Although TTL is an easier-to-read format for humans than other RDF formats, it would not be the most straightforward way to understand a workflow; a graphical visualisation tends to give a better overview than text descriptions. Nevertheless, the graphic representation of workflows using RDF graphs in GraphDB was unclear (see Figure 20). Thus, although the workflows were described with RDF, their graphical representation was done using a workflow diagram, similar to a flowchart. The workflow diagram represents the operations of the workflow as an edge and the data flowing in the workflow as a node, which is the opposite of the typical DAG diagram used, for instance, in Pegasus and Knime. Although this was not initially done on purpose, it did make sense to the author as the workflow description is data-centred.

From the survey results, the participants could interpret and understand the workflow diagram, but it was more difficult for some participants than others. As mentioned in Chapter 4, there were some limitations in evaluating this aspect. The workflow diagram chosen to be interpreted was the smallest diagram to save participants time. However, this workflow had a geodata operation that might not be commonly known. The participants were unfamiliar with network analysis, making it difficult to understand the workflow from the workflow diagram. A link to the operations' description in the diagram could help solve this problem; users could get more information about the operations if desired.

Question 3.1 - What are the workflow repository requirements to store and share workflows described with our method?

The workflow repository must be linked to a triple datastore, and options must be provided to load new workflows and search for existing ones. Ideally, the workflow repository would also be connected to a WfMS that enables creating, editing, and saving new workflows with the option to create concrete workflows for the main GIS used in the geodata community.

Clear instructions for uploading new workflows must be provided to the users. During the survey, it became clear that several participants were making conclusions about the content and version of the workflow by looking at its title and URI. Therefore, one of the recommendations is to give meaningful names to the workflows shared and follow the version numbering as follows:

- Versions in progress should have a number under one and use a dot (.) as a separator, e.g. 0.6.
- Stable versions should have numbers from one up, the first stable version being 1.0.
- Versions of a workflow that are alternatives to the same workflow should use the number of the version and a decimal separator with the number of the sub-version. For instance, a workflow version 1.0 would become 1.1 if slightly altered. A slight alteration would be a change in the input data or parameter values. This is useful to save workflows that simulate different scenarios.
- Versions that update or replace a workflow version should use one digit above the version they replace and the number after the decimal separator is restarted. For instance, if a workflow on version 1.5 is replaced, the version should now be 2.0.

Another recommendation is to provide as much workflow and data metadata information as possible (see Section 3.2); the suggested mandatory fields were already stated when replying to Research Question 1.1.

To support discovery, the ontologies used in the repository must be available for users to use in their queries and workflow descriptions. Furthermore, it is important to enable users to expand these vocabularies, subject to an administrator or steering committee revision to maintain consistency.

Question 3.2 - What are the workflow repository requirements to support findability and reusability of workflows described with our method?

The geodata workflow repository must be connected to a triplestore database such as GraphDB, which must be installed in a publicly available server and facilitate a user-friendly way to query the database. Additionally, the repository needs to offer a way to visualise the workflow metadata and diagram with links to more information about entities (e.g. data providers), operations used in workflow steps and geodata involved in the workflow, as suggested in Section 4.5. This information is crucial to the users' decision to reuse a workflow for their specific goals. In principle, the discovery of workflows and data is enhanced by using the domain vocabulary and the possibility of searching for workflows per author, per data flowing in the workflow, vocabulary terms, or workflow purpose.

Search options need to meet the users' expectations to support findability. Through the survey implementation, it was observed that not all search options provided were used. It is debatable, for instance, if having both the workflow title and purpose is useful, as they provide similar information. It might be preferable to provide only the workflow purpose as a search field given that the workflow title can be subjective while the workflow purpose uses well-defined vocabulary terms. On the other hand, users would benefit from having a flexible search; sometimes, they can expect to see a keyword in one field when, in fact, it was stored in a different field. Searching for one keyword in different fields simultaneously could be more efficient than specifying the search field. Nevertheless, this does not hold for all keywords and fields; for instance, it might be useful to query for a workflow created by a certain entity for a certain purpose. In this case, it would be useful to specify the workflow purpose in the workflow purpose field and the author's name in the author's field to narrow down the results.

Remaining on the topic of findability, it is important to avoid unclear or unfamiliar terms to users in the search options, such as the "literal value" for the vocabulary terms. This could mean renaming it to "properties" or using a more familiar term to be clearer and so that users do not depend so much on tooltips or manuals for trivial tasks. This was discussed during the survey as well – Section 4.1 and Section 4.5. Furthermore, anything that is not useful for users should not be visible in the search options, such as any URIs from data or workflows.

Furthermore, as mentioned in Section 3.4.3, a good solution must be found to enable the use of the vocabulary term's label directly in the query to bypass the not-so-straightforward approach of linking a subject to a vocabulary term with the label "x". For one, using the vocabulary term's label is more user-friendly; on the other hand, the relationships between vocabulary terms are set at the URI level. A solution might be to use SKOS (Simple Knowledge Organization System). In Sparnatural, this could be implemented and shown as a tree type of search where the related concepts are shown in a tree structure. Additionally, the domain vocabulary could provide workflow or data suggestions to users next to the results.

For a cleaner look, the recommendation would be to limit the search options for workflows by workflow purpose, authoring entity, creation date, vocabulary terms used in the workflow, and related data and workflows as entry points. Once the first criteria are chosen, the search options differ per criteria; those are specified in Table 6, where each column represents the search options in the next step of the search.

First search criteria for workflows	Search option/value	Searched value
Workflow purpose	Enter label from vocabulary term – having auto-fill options helps in this case (*see vocabulary term)	
Author	Enter name of organisation or person	
Creation date	Date or time interval (select in calendar)	
Vocabulary term	Is part of Has part Same as	Enter label from vocabulary term
Data (specify whether it is input, intermediate output or final output data)	Identifier	Enter identifier
	Title	Enter title or words expected to be present in title
	Creation date	Date or time interval (select in calendar)
	Contact point	Enter name of organisation or person
	Vocabulary term (*see vocabulary term)	
workflow (with all the above options mentioned plus)	Is part of Has part Is version of Is replaced by Replaces	workflow (all the above options mentioned can be added to the related workflow as well)

Table 6 – Workflow search options depending on the first criteria chosen. Users can also search for data with the same options, using the relationship with workflow (is input/intermediate output/final output of workflow).

The workflow results can be sorted by the maturity status, creation date and version number of the workflows. The version control aspect of the workflow description is useful for tracking workflow evolution and can be activated in "advanced mode". These refer to the workflow–workflow relationships of "has part" (inverse of " is part of"), "is replaced by" (inverse of "replaces"), and "is version of", which might not be interesting for the most common users. This is also the case for queries to find all workflows that use data provided by a certain data provider; it is a nice feature, but it will not be interesting for most users.

Regarding reusability, results from the survey indicate that users could understand the workflows' purpose and steps and have the information necessary to decide whether to reuse workflows – Section 4.4. Nevertheless, to reuse any workflows, users would need a system to translate the abstract workflow to an executable format. Otherwise, they will need good guidelines on implementing the workflow themselves in their software of choice, but without the option to track workflow evolution. Having options to edit and save the workflow as a new version is needed to support reusability and workflow evolution. The system should be able to automatically generate metadata regarding the new workflow version.

Also, regarding metadata, strategic choices should be made to provide users with the most relevant information and reveal more information at users' request with a click of a button.

Additionally, as mentioned in Section 4.5, there should be a clear, easy way to highlight workflow changes to support users in choosing between different versions. The differences could be highlighted in the diagram, or an RDF comparison tool could be available in the workflow repository. Since the operations and parameter values are stored for each operation in the workflow description, they can be queried. Sparnatural (or another SPARQL query builder) should be customisable enough to enable querying each operation of two workflows, including the parameter names and values to assist in workflow comparison. More time would be needed to explore this possibility since Sparnatural only shows results for the first subject selected by default.

Another tool to assist users when choosing between several versions of a workflow is a rating system. A simple rating system can be achieved by allowing users to "like" a workflow, mentioning how many times the workflow was downloaded or allowing users to provide reviews or comments. Some of these examples are implemented in the reviewed workflow repositories - Section 2.1.5.

Question 3.3 - What are the requirements and functionalities of a user-friendly geodata workflow?

Ideally, a Geodata workflow repository would be web-based so users do not need to download and install software on their local machines. The web interface should encompass a range of functionalities to enhance user experience.

Firstly, the platform should facilitate the construction of SPARQL queries even for less savvy users; tooltips should aid users in navigating search options effectively, and advanced users should be allowed to customise their search options (Section 3.4.3, and Section 3.5). Something like the implementation done in Sparnatural is useful. However, further research and customisation are needed to make it more user-friendly, as mentioned in Section 3.4.3 and Section 5.3. Furthermore, the results of the queries need to be listed in a sensible way, including all the entities in the query, not just the first one selected.

Additionally, clear visualisation of the workflow (workflow) metadata and diagrams with embedded links to additional information about entities, operations, and geodata is essential for enhanced comprehension of the workflows. A good balance between the information provided by default and via links to other resources should be achieved. Ideally, information about operations and metadata regarding data flowing throughout the workflow should be provided through links describing those resources, while the workflow diagram and metadata should be available on the workflow page. Some workflow metadata can be hidden behind a “more information” button, as shown on the proof of concept webpage (see Figures 13 to 18).

Furthermore, an intuitive graphical interface should enable users to create or customise workflows. This interface should automatically generate the corresponding TTL file for storage and automatically create (or update in case of editing) the workflow metadata.

Basic metadata such as creation date and authorship can be automated. However, the platform should offer more options for metadata creation, necessitating further development.

Users should be able to convert abstract workflows into executable formats of their choice. Support for at least three popular formats should be provided, we would recommend Galaxy,Taverna or YAML.

Enabling user feedback mechanisms, such as ratings and the ability to provide feedback to authors, can assist other users in evaluating the suitability of workflows for their needs.

The platform should offer comprehensive user support, including demo videos, manuals, FAQs, and avenues for users to ask questions and receive assistance. Clear instructions on sharing workflows must be provided to ensure correct usage of ontologies and continued support for findability and reusability.

Users should have the option to easily extend vocabularies, subject to approval from developers, to accommodate evolving requirements and domain-specific terminology.

Additionally, the graphical user interface design options play a big role in the learning curve and ease of use of a tool; although this is not the focus of this study, it should not be discarded.

5.2. General Conclusions

Workflows are composed objects which can be described using web semantics by linking each component using defined vocabularies. Another consequence of being composed objects is that all its components must adhere to FAIR principles for a workflow to be FAIR (Section 2.1.2). During the extensive literature review, it was noted that workflow ontologies (Section 2.2.2) are not fully aligned with FAIR principles, and a multifaceted approach is necessary to improve the description of the Workflow metadata. Future research should address this gap; here, we point out the areas considered important to cover for comprehensive FAIR metadata:

- Descriptive metadata: Enhance workflow ontologies to include detailed descriptions that provide a clear overview of the workflow's purpose and scope. Include references to literature if applicable.
- Provenance and version tracking: Extend existing ontology to comprehensively track provenance, which includes version history, detailed records of changes, contributors, and the rationale for modifications.
- Structural metadata: This should detail the data type, format, source, quality, and constraints. Standardising this documentation enhances interoperability and facilitates integration with other data systems and workflows.
- Administrative metadata: Specify conditions under which workflows were created and can be reused, including operational environments and software dependencies. This transparency supports reusability, allowing others to adapt the workflow with confidence.
- Linking to External Standards: Integrate references to external metadata standards and data-sharing protocols within the domains. For instance, linking to standards like ISO/IEC or Dublin Core for metadata documentation can help align with broader data management practices.

Implementing these strategies would enhance workflow ontologies from procedural documentation to robust, standalone resources integral to the scientific data ecosystem. These enhancements make workflows more usable and reliable and ensure they contribute effectively to a stronger knowledge base.

This thesis substantially contributes to Marine Spatial Planning (MSP) by introducing a novel workflow definition using web semantic principles (Section 3.3). This workflow definition improves the discoverability of workflows by integrating rich metadata and standardised vocabularies, enabling users to easily locate specific workflows relevant to their needs by exploring RDF graphs through a third-party interface (Section 4.1). It also promotes the reusability of workflows across different projects and applications by standardising descriptions, encouraging efficient use of resources and reducing redundancy. Furthermore, the definition includes robust mechanisms for provenance tracking, which enhance transparency and traceability of data processes within workflows (Section 3.2).

To describe the workflows, three ontologies are developed and combined to achieve the research goals: workflow ontology, workflow purpose ontology and domain ontology (Section 3.3). The glue that holds the ontologies together is the workflow ontology, which includes the data and workflow provenances and the workflow evolution tracking. It links to the domain ontology through the tags assigned to the data that flows in the workflow, and to the workflow purpose via the property specifically created to document the workflow purpose. The workflow purpose and domain ontologies developed are not exhaustive but are a first step to enhancing the findability and reusability of workflow in MSP. It would be great to see these ontologies grow with contributions from the GIS and MSP communities, respectively.

In addition, several MSP-related workflows were developed using the proposed workflow description and added to a semantic web workflow repository with a searchable interface Section 3.4.1. This prototype was useful in providing very specific design insights and clear guidelines and specifications for developing a more mature repository linked to a WfMS, as well as the identification of gaps that can be considered for future research (Section 5.1 and Section 5.4).

Using a graphical SPARQL query builder was essential in enabling non-technical users to query the triplestore. During the survey, there was a noticeable learning curve while querying the database. The author thinks anyone could use the defined prototype with training or tutorials; nevertheless, as already mentioned, there is room for improvement (Section 4.5).

5.3. Limitations

Since the workflow description relies on RDF and is stored in a triplestore, the search for workflows must be done through SPARQL queries. One of the limitations of this approach is that the targeted users are not expected to know (or learn) SPARQL. One way to overcome this serious limitation was to use a graphical query builder. Sparnatural was successfully used; nevertheless, it was not the ideal solution described in Section 3.4.3. More time would be needed to research how to overcome the restrictions found, search for a better solution, or create a customised one.

Another limitation was due to the restricted number of workflows available in the database; the participants did not test advanced filtering and sorting options as they were irrelevant due to the limited number of results retrieved. Additionally, although the participants who tested the system and replied to the survey had a diverse background, the sample was small, as mentioned in Section 3.5.

This research was only a small contribution to an MSP ontology regarding data pertinent to the theme and the data analysis involved to test a hypothesis.

5.4. Future Research

The research presented in this thesis can be a stepping stone to designing a geodata workflow repository that can satisfy users' needs. However, it does not have all the answers yet. Several limitations are already mentioned in Section 5.3, and some elements are missing to fulfil the requirements defined for such a repository.

Previous work has been done to convert abstract workflows to executable ones using web semantics (Ubels, 2018), which is why the data types description was added to the workflow description. Having the input and output data type description at each step of the workflow should enhance the automation of workflow concretisation. Nevertheless, this depends on software or service providers describing or tagging their operations using web semantics. Additionally, the data type can be useful for WfMS, such as Wings, which suggests compatible operations based on the input data type and operations used in other workflows in the database.

Keeping track of the data types and core concepts within the workflow could be done automatically using the work of Scheider and Ballatore (2018). The relationships between GIS concepts and rules defined in their article could also apply to the workflow description approach of this study to validate any changes made to the workflow, assisting users in reusing the workflow for their purposes.

Furthermore, an automated way to convert existing workflow formats to the described TTL format would accelerate the repository uptake and enrich the database. The translation of workflows from GIS to TTL to be able to share and back for people to use it in their (proprietary) software of choice seems crucial for the uptake of a workflow repository like the one proposed in this research. An extensive dictionary of operations for different GIS to TTL and the other way around is crucial. Although some research has already been done in this context, more research is needed to accommodate more data transformations and GIS.

Incomplete metadata is often a problem, as also seen during the survey implementation; one of the participants confessed not expecting the metadata to be complete because, in their experience, that was often the case. Another participant found incomplete metadata in one of the input datasets even after this was checked. Research has been previously done on automating metadata creation (Giuliani et al., 2016; Trilles et al., 2012); exploring how that research could apply to a geodata workflow repository would be interesting. Artificial intelligence could probably assist in this task. Furthermore, it would be useful to automatically validate the workflow description to make sure it fits the repository requirements before sharing the workflow in the repository.

The domain vocabulary can be further developed with more marine or maritime uses, environmental conditions, marine habitats and species. Likewise, workflow and data format descriptions were not added to the workflow or data descriptions; they are important for executable workflow validation and automatisation. Furthermore, data attributes could also be included with their relationships when applicable.

To conclude, the graphical user interface design options play an important role in the user-friendliness and uptake of (online) tools and software. Although this is not the focus of this study, it should not be overlooked; more research into applying user-friendly options in a geodata workflow repository would be useful.

References

Addlesee, A. (2018). *Understanding Linked Data Formats*. https://medium.com/wallscope/understanding-linked-data-formats-rdf-xml-vs-turtle-vs-n-triples-eb931dbe9827

Amruta, K., Sun, Z., Fan, C., & Ma, X., (2023). Geoweaver_cwl: Transforming geoweaver AI workflows to common workflow language to extend interoperability. *Applied Computing and Geosciences*, *19*. https://doi.org/https://doi.org/10.1016/j.acags.2023.100126

Amstutz, P., Mikheev M., R. Crusoe M., Tijanić N., Lampa, S. et al. (2023): *Existing Workflow systems.* Common Workflow Language wiki, GitHub. https://s.apache.org/existing-workflow-systems

Beckett, D., Berners-Lee, T., Prud'hommeaux, E., & Carothers G., (2014). *RDF 1.1 Turtle - W3C Recommendation 25 February 2014*. https://www.w3.org/TR/turtle/#language-features

Belhajjame, K., Zhao, J., Garijo, D., Gamble, M., Hettne, K., Palma, R., Mina, E., Corcho, O., Gómez-Pérez, J. M., Bechhofer, S., Klyne, G., & Goble, C. (2015). Using a suite of ontologies for preserving workflow-centric research objects. *Journal of Web Semantics*, *32*, 16–42. https://doi.org/10.1016/j.websem.2015.01.003

Cagle, K. (2022, February 13). *Why JSON Users Should Learn Turtle*. Data Science Central. https://www.datasciencecentral.com/why-json-users-should-learn-turtle

Farias Lóscio, B., Burle, C., & Calegari, N. (2017). *Data on the Web Best Practices*. W3C Recommendation 31 January 2017. https://www.w3.org/TR/dwbp

Bernadette, H., Atemezing, G., Pendleton, M., & Srivastava, B. (2013). Linked Data Glossary. *W3C Recommendation*. https://www.w3.org/TR/ld-glossary

Berners-Lee, T. (2006). *Linked Data*. https://www.w3.org/DesignIssues/LinkedData.html

Billings, J. (2019). *Ontological Considerations for Interoperability in Scientific Workflows*. University of Tennessee, Knoxville.

Brodeur, J., Coetzee, S., Danko, D., Garcia, S., & Hjelmager, J. (2019). Geographic information metadata—an outlook from the international standardization perspective. *ISPRS International Journal of Geo-Information*, *8*(6). https://doi.org/10.3390/ijgi8060280

Burattin, A. (2015). Process mining techniques in business environments: Theoretical aspects, algorithms, techniques and open challenges in process mining. In *Lecture Notes in Business Information Processing* (Vol. 207). https://doi.org/10.1007/978-3-319-17482-2_1

Campagna, M., Ivanov, K., & Massa, P. (2014). *Orchestrating the spatial planning process: from Business Process Management to 2nd generation Planning Support Systems*. 3–6.

Carr, M. H., Robinson, S. P., Wahle, C., Davis, G., Kroll, S., Murray, S., Schumacker, E. J., & Williams, M. (2017). The central importance of ecological spatial connectivity to effective coastal marine protected areas and to meeting the challenges of climate change in the marine environment. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *27*(January), 6–29. https://doi.org/10.1002/aqc.2800

Crusoe, M. R., Abeln, S., Iosup, A., Amstutz, P., Chilton, J., Tijanić, N., Ménager, H., Soiland-Reyes, S., Gavrilović, B., & Goble, C. (2022). Methods included. *Communications of the ACM*, *65*(6), 54–63. https://doi.org/10.1145/3486897

Deelman, E., Vahi, K., Juve, G., Rynge, M., Callaghan, S., Maechling, P. J., Mayani, R., Chen, W., Ferreira Da Silva, R., Livny, M., & Wenger, K. (2015). Pegasus, a workflow management system for science automation. *Future Generation Computer Systems*, *46*, 17–35. https://doi.org/10.1016/j.future.2014.10.008

Duffy, C., Gil, Y., Deelman, E., Marru, S., Pierce, M., Demir, I., & Wiener, G. (2012). Designing a road map for geoscience workflows. *Eos*, *93*(24), 225–226. https://doi.org/10.1029/2012EO240002

Elliott, S., Turek, G., Snow, V., Rutledge, D., & Alistair, R. (2013). *Framework for Interoperable Freshwater Models: Testing and Recommendations. Prepared for Ministry of Business, Innovation and Employment* (Issue February).

European Commission, Executive Agency for Small and Medium-sized Enterprises, Ooms, E., Onwona Ansong, J., Lukic, I. (2019). *Addressing conflicting spatial demands in MSP : considerations for MSP planners : final technical study*, Publications Office. https://data.europa.eu/doi/10.2826/151447

International Organization for Standardization. (2006). *Geographic information Metadata.* (ISO 19115-1:2014). The International Organization for Standardization.

International Organization for Standardization. (2019). *Geographical Information - XML Implementation.* (ISO 19139-1 :2019). The International Organization for Standardization.

Garijo, D., Gil, Y., & Corcho, O. (2017). Abstract, link, publish, exploit: An end to end framework for workflow sharing. *Future Generation Computer Systems*, *75*, 271–283. https://doi.org/10.1016/j.future.2017.01.008

Gil, Y., Ratnakar, V., & Deelman, E. (2007). Wings for pegasus: Creating large-scale scientific applications using semantic representations of computational workflows. *Proceedings of the National Conference on Artificial Intelligence*, 1767–17774. https://www.aaai.org/Papers/AAAI/2007/AAAI07-284.pdf

Giuliani, G., Guigoz, Y., Lacroix, P., Ray, N., & Lehmann, A. (2016). Facilitating the production of ISO-compliant metadata of geospatial datasets. *International Journal of Applied Earth Observation and Geoinformation*, *44*, 239–243. https://doi.org/10.1016/j.jag.2015.08.010

Goble, C. A., Bhagat, J., Aleksejevs, S., Cruickshank, D., Michaelides, D., Newman, D., Borkum, M., Bechhofer, S., Roos, M., Li, P., & De Roure, D. (2010). myExperiment: a repository and social network for the sharing of bioinformatics workflows. *Nucleic Acids Research*, *38*(suppl_2), W677–W682. https://doi.org/10.1093/nar/gkq429

Goble, C., Cohen-Boulakia, S., Soiland-Reyes, S., Garijo, D., Gil, Y., Crusoe, M. R., Peters, K., & Schober, D. (2020). Fair computational workflows. *Data Intelligence*, *2*(1–2), 108–121. https://doi.org/10.1162/dint_a_00033

Jiang, L., Kuhn, W., & Yue, P. (2017). An interoperable approach for Sensor Web provenance. *6th International Conference on Agro-Geoinformatics*, 1–6. https://doi.org/10.1109/Agro-Geoinformatics.2017.8047046

Kanwal, S., Khan, F. Z., Lonie, A., & Sinnott, R. O. (2017). Investigating reproducibility and tracking provenance - A genomic workflow case study. *BMC Bioinformatics*, *18*(1), 1–14. https://doi.org/10.1186/s12859-017-1747-0

Kasalica, V. (2022). *Synthesis of Scientific Workflows : Theory and Practice of an Anstance-Aware Approach*. [Unpublished doctoral dissertation]. Utrecht University.

Kechagioglou, X., Lemmens, R., & Retsios, V. (2019). Sharing Geoprocessing Workflows with Business Process Model and Notation (BPMN). *Proceedings of the 2019 2nd International Conference on Geoinformatics and Data Analysis*, 56–60. https://doi.org/10.1145/3318236.3318239

Khan, F. Z., Soiland-Reyes, S., Sinnott, R. O., Lonie, A., Goble, C., & Crusoe, M. R. (2019). Sharing interoperable workflow provenance: A review of best practices and their practical application in CWLProv. *GigaScience*, *8*(11), 1–27. https://doi.org/10.1093/gigascience/giz095

Koubarakis, M., Bereta, K., Nikolaou, C., & Stamoulis, G. (2019). Linked Geospatial Data. *Encyclopedia of Big Data Technologies*, 1124–1131. https://doi.org/10.1007/978-3-319-77525-8_71

Kyvelou, S. S. I., & Ierapetritis, D. G. (2021). Fostering spatial efficiency in the marine space, in a socially sustainable way: Lessons learnt from a soft multi-use assessment in the mediterranean. *Frontiers in Marine Science*, *8*(March), 1–15. https://doi.org/10.3389/fmars.2021.613721

Kyzirakos, K., Savva, D., Vlachopoulos, I., Vasileiou, A., Karalis, N., Koubarakis, M., & Manegold, S. (2018). GeoTriples: Transforming geospatial data into RDF graphs using R2RML and RML mappings. *Journal of Web Semantics*, *52–53*, 16–32. https://doi.org/10.1016/j.websem.2018.08.003

Leipzig, J., Nüst, D., Hoyt, C. T., Ram, K., & Greenberg, J. (2021). The role of metadata in reproducible computational research. *Patterns*, *2*(9). https://doi.org/10.1016/j.patter.2021.100322

Lemmens, R., Farifteh, J., Piccinini, C., Retsios, B., Schouwenburg, M., & Bonina, J. (2016). Visual and coded geoprocessing workflows based on ILWIS and Python. *Journal of GeoPython*, *1*(1), 21–24. http://www.geopython.net/pub/Journal_of_GeoPython_1_2016.pdf

Masó, J., Closa, G., & Gil, Y. (2015). Applying W3C PROV to express geospatial provenance at feature and attribute level. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8628*, 271–274. https://doi.org/10.1007/978-3-319-16462-5_31

Meek, S., Jackson, M., & Leibovici, D. G. (2016). A BPMN solution for chaining OGC services to quality assure location-based crowdsourced data. *Computers and Geosciences*, *87*, 76–83. https://doi.org/10.1016/j.cageo.2015.12.003

Mesbah, S., Fragkeskos, K., Lofi, C., Bozzon, A., & Houben, G. J. (2017). Semantic annotation of data processing pipelines in scientific publications. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10249 LNCS*, 321–336. https://doi.org/10.1007/978-3-319-58068-5_20

Missier, P., Dey, S., Belhajjame, K., Cuevas-Vicenttín, V., & Ludäscher, B. (2013). D-PROV: Extending the PROV provenance model with workflow structure. *5th Workshop on the Theory and Practice of Provenance, TaPP 2013*, *April*.

Missier, P., Sahoo, S. S., Zhao, J., Goble, C., & Sheth, A. (2010). Janus: From Workflows to Semantic Provenance and Linked Open Data. In *Ipaw* (Vol. 6378, Issue November, pp. 129–141). https://doi.org/10.1007/978-3-642-17819-1_16

Nogueras-Iso, J., Zarazaga-Soria, F. J., Lacasta, J., Béjar, R., & Muro-Medrano, P. R. (2004). Metadata standard interoperability: application in the geographic information domain. *Computers, Environment and Urban Systems*, *28*(6), 611–634. https://doi.org/10.1016/j.compenvurbsys.2003.12.004

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* [Unpublished Master Thesis]. University of Twente.

Ostermann, F. O., & Granell, C. (2017). Advancing Science with VGI: Reproducibility and Replicability of Recent Studies using VGI. *Transactions in GIS*, *21*(2), 224–237. https://doi.org/10.1111/tgis.12195

Pegasus Team (2019). *Pegasus User Guide*. Pegasus WMS. https://pegasus.isi.edu/documentation

Roos, M., Bechhofer, S., Zhao, J., Missier, P., Newman, D. R., De Roure, D., & Marshall, M. S. (2010). A linked data approach to sharing workflows and workflow results. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6415 LNCS*(PART 1), 340–354. https://doi.org/10.1007/978-3-642-16558-0_29

Rosser, J., Pourabdollah, A., Brackin, R., Jackson, M., & Leibovici, D. (2016). Full Meta Objects for flexible geoprocessing workflows: profiling WPS or BPMN? *The 19th AGILE International Conference on Geographic Information Science*, *June*.

Rowland, A. (2020). *Using Semantic Technologies for Effective GIS Workflow Translation*. [Unpublished Master Thesis]. Utrecht University.

Salah, H., Nacer, B., & Mahmoud, B. (2014). Semantic annotations for workflow interoperability. *Informatica (Slovenia)*, *38*(4), 347–366.

Scharffe, F., Atemezing, G., Troncy, R., Gandon, F., Villata, S., Bucher, B., Hamdi, F., Bihanic, L., Képéklian, G., Cotton, F., Euzenat, J., Fan, Z., Vandenbussche, P. Y., & Vatant, B. (2012). Enabling linked data publication with the datalift platform. *AAAI Workshop - Technical Report*, *WS-12-13*(July), 25–30.

Scheider, S., & Ballatore, A. (2018). Semantic typing of linked geoprocessing workflows. *International Journal of Digital Earth*, *11*(1), 113–138. https://doi.org/10.1080/17538947.2017.1305457

Scheider, S., Ballatore, A., & Lemmens, R. (2019). Finding and sharing GIS methods based on the questions they answer. *International Journal of Digital Earth*, *12*(5), 594–613. https://doi.org/10.1080/17538947.2018.1470688

Shaon, A., Callaghan, S., Lawrence, B., Matthews, B., Woolf, A., Osborn, T., & Harpham, C. (2011). A linked data approach to publishing complex scientific workflows. *Proceedings - 2011 7th IEEE International Conference on EScience, EScience 2011*, 303–310. https://doi.org/10.1109/eScience.2011.49

Skvortsov, N. A., & Stupnikov, S. A. (2022). A Semantic Approach to Workflow Management and Reuse for Research Problem Solving. *Data Intelligence*, *4*(2), 439–454. https://doi.org/10.1162/dint_a_00142

Soiland-Reyes, S., Sefton, P., Crosas, M., Castro, L. J., Coppens, F., Fernández, J. M., Garijo, D., Grüning, B., La Rosa, M., Leo, S., Ó Carragáin, E., Portier, M., Trisovic, A., RO-Crate Community, Groth, P., & Goble, C. (2022). Packaging research artefacts with RO-Crate. *Data Science*, *5*(2), 97–138. https://doi.org/10.3233/DS-210053

Stancheva, M., Stanchev, H., Zaucha, J., Ramieri, E., & Roberts, T. (2022). Supporting multi-use of the sea with maritime spatial planning.The case of a multi-use opportunity development - Bulgaria, Black Sea. *Marine Policy*, *136*(October 2021), 104927. https://doi.org/10.1016/j.marpol.2021.104927

Sun, Z., Di, L., Burgess, A., Tullis, J. A., & Magill, A. B. (2020). Geoweaver: Advanced Cyberinfrastructure for Managing Hybrid Geoscientific AI Workflows. *ISPRS International Journal of Geo-Information*, *9*(2), 119. https://doi.org/10.3390/ijgi9020119

Taverna. (n.d.). Taverna Project Incubation Status. http://www.taverna.org.uk

Tolovski, I., Džeroski, S., & Panov, P. (2020). Semantic Annotation of Predictive Modelling Experiments. In: Appice, A., Tsoumakas, G., Manolopoulos, Y., Matwin, S. (eds) Discovery Science. DS 2020. Lecture Notes in Computer Science(), vol 12323. Springer, Cham. https://doi.org/10.1007/978-3-030-61527-7_9

Tick, J. (Eds.). (2006). Workflow Model Representation Concept. https://www.researchgate.net/publication/228936659_Workflow_Model_Representation_Concepts

Trilles, S., Díaz, L., Gil, J., & Huerta, J. (2012). Assisted Generation and Publication of Geospatial Data and Metadata. *Bridging the Geographic Information Sciences International AGILE2012 Conference Avignon France April 24- 27, 2012*, *9*(April), 105–110. https://agile-online.org/Conference_Paper/CDs/agile_2012/proceedings/papers/Paper_Trilles_Assisted_Generation_and_Publication_of_Geospatial_Metadata_2012.pdf

Ubels, S. (2018). *Understanding abstract geo-information workflows and converting them to executable workflows using Semantic Web technologies*. [Unpublished Master Thesis]. University of Twente.

Unified Modeling Language. (2005). What is UML. https://www.uml.org/what-is-uml.htm

Usery, E. L., & Varanka, D. (2012). Design and development of linked data from the National Map. *Semantic Web*, *3*(4), 371–384. https://doi.org/10.3233/SW-2011-0054

von Thenen, M., Hansen, H. S., & Schiele, K. S. (2021). A generalized marine planning framework for site selection based on ecosystem services. *Marine Policy*, *124 (February)*, 104326. https://doi.org/10.1016/j.marpol.2020.104326

von Thenen, M., Maar, M., Hansen, H. S., Friedland, R., & Schiele, K. S. (2020). Applying a combined geospatial and farm scale model to identify suitable locations for mussel farming. *Marine Pollution Bulletin*, *156*(May), 111254. https://doi.org/10.1016/j.marpolbul.2020.111254

W3C. (2013 April 30). *W3C PROV-Overview*. https://www.w3.org/TR/prov-overview

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 1–9. https://doi.org/10.1038/sdata.2016.18

Wolstencroft, K., Haines, R., Fellows, D., Williams, A., Withers, D., Owen, S., Soiland-Reyes, S., Dunlop, I., Nenadic, A., Fisher, P., Bhagat, J., Belhajjame, K., Bacall, F., Hardisty, A., Nieva de la Hidalga, A., Balcazar Vargas, M. P., Sufi, S., & Goble, C. (2013). The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. *Nucleic Acids Research*, *41*(Web Server issue), 557–561. https://doi.org/10.1093/nar/gkt328

Workflow Management Coalition. (1999). The Workflow Management Coalition Specification, Terminology & Glossary. In *Workflow Management Coalition* (workflowMC-TC-1011; Issue 3). https://doi.org/10.1016/s0019-0578(00)00014-8

Wroe, C., Goble, C., Goderis, A., Lord, P., Miles, S., Papay, J., Alper, P., & Moreau, L. (2007). Recycling workflows and services through discovery and reuse. *Concurrency and Computation: Practice and Experience*, *19*(2), 181–194. https://doi.org/10.1002/cpe.1050

Yue, P., Guo, X., Zhang, M., Jiang, L., & Zhai, X. (2016). Linked Data and SDI: The case on Web geoprocessing workflows. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 245–257. https://doi.org/10.1016/j.isprsjprs.2015.11.009

Yue, S., Chen, M., Wen, Y., & Lu, G. (2016). Service-oriented model-encapsulation strategy for sharing and integrating heterogeneous geo-analysis models in an open web environment. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 258–273. https://doi.org/10.1016/j.isprsjprs.2015.11.002

Zhang, M., Jiang, L., Yue, P., & Gong, J. (2020). Interoperable web sharing of environmental models using OGC web processing service and Open Modeling Interface (OpenMI). *Environmental Modelling and Software*, *133*(August), 104838. https://doi.org/10.1016/j.envsoft.2020.104838

Zhang, M., Jiang, L., Zhao, J., Yue, P., & Zhang, X. (2020). Coupling OGC WPS and W3C PROV for provenance-aware geoprocessing workflows. *Computers and Geosciences*, *138*(June 2019), 104419. https://doi.org/10.1016/j.cageo.2020.104419

Appendices

Appendix I – Assessment of the Workflow Repositories consulted in this study

Workflow Repositories Assessment

Platform's Name	Access Type	Metadata (Standards)	WF Portability (WF format)	Provenance (Standards)	WF Reusability	Type of WF (Supports Execution)	Supports Geodata Operations	Geodata Format(s) Supported	Input/output data shared	Annotations support	Search fields/options	Points to improve
KNIME Hub	Open access. No identifiers used for WFs or data.	User dependent. In the description users can make a more or less extensive description. (Not explicit)	KNIME format only (from the KNIME analytics platform). There are tools to convert to other formats (e.g. KNIME2Grid [1])	Retrospective and prospective provenance supported. (Not explicit)	License of WF is clearly stated. Shows diagram. Input metadata, description and scope of WF, and instructions on how to use the WF are user dependent. Adapting a WF to reuse in KNIME analytic tools is easy.	Executable. (Yes)	Yes - Via extension (Geospatial Analytics for KNIME) More extensions can be added.	Shapefile, single layer zipped Shapefiles and Geopackage,	User's choice	Yes (it becomes visible in the WF's graph and helps the WF to be found)	Text, filter by tags, WFs, Nodes, Components, Extensions, Collections	Search workflows by type and coverage of input data (when applicable). Download to a common format that can be run in any platform. Search workflows that include a certain operation. Enable search in chosen field such as: "description", "nodes", "creator" "input data", "output data". Create and use a tag ontology that users can expand upon if approved by development team.
Pegasus HUB	Open access via GitHub repository link. No identifiers used for WFs or data.	User dependent but there are clear metadata requirements for when uploading a WF. (Not explicit)	Yes, for tools that support CWL (YAML)	Retrospective and prospective provenance supported. (OPMW in Wings)	License of WF is clearly stated. Shows diagram. Description of the purpose, the latest release, Pegasus version, dependencies, topics (tags) prerequisites, and instructions on running the WF.	Abstract to executable (Yes in Pegasus)	Yes, wrapped in YAML.	NA	Sometimes. Mapped via Replica Catalogs (accepted formats: YAML (default), File, Regex, Directory, Database via JDBC, and MRC)	Not explicit	Text and tags	Adding functionality to the search box and to enable search in different fields, such as title, topics, license, description, dependencies and job labels (processes names). It would also be useful for non-technical users to have operations' description available, rather than just the label; this would make it easier to understand the steps and overall purpose of the workflow.
myExperiment	Open access. No identifiers used for WFs or data.	Metadata at the pack level (bundle of files and WFs) and item level. Metadata includes inputs, outputs, the operations it makes on data and information about credit, attribution and licensing. Version control. Shows diagram. (Not explicit)	System agnostic, including Taverna, Rap idMiner, Galaxy, Knime and Kepler. User can download WF as their native format or as Taverna or Galaxy Tool.	Prospective provenance and WF evolution supported. Taverna registers the retrospective provenance. (OPM and PROV)	License of WF is clearly stated. Shows diagram. Version control. Description on how to run the WF and dependencies are available. Usage of packs (bundle of files and WFs with all the info needed in one place).	Executable. (Yes, in Taverna or refer to native tool)	Possible via REST or SOAP web services and command line tools.	NA	Not sure	Not explicit but mentions of linked data support in several sources.	Text, filter (type/tool, tag, user, license, group, wsdl, curation) Sorting per rank, title, latest, last updated, user, community ranking, most viewed, most downloaded, type, licence, topic.	Implementation of identifiers to WFs and WF components. Create and use a tag ontology that users can expand upon if approved by development team.
Geoweaver	Based on server definitions No identifiers used for WF or data.	Poor metadata, only shows ID, name and description. No WF diagram, not clear what the steps of the WF are. (not mentioned)	Recently started to support conversion to CWL (CWL)	Retrospective provenance. (PROV mentioned in literature but not implemented)	Does not seem to provide enough info for fit-for-purpose evaluation. Easily adaptable.	Executable. (in Geoweaver or CWL supporting tools).	Yes	Not listed, it depends more on the services used than on the platform itself.	Hosted in servers, access controlled)	Not explicit but when using code it can be commented and the use of Jupyter notebooks make great annotation tools.	Text (search only on titles). No filters, no tags.	More documentation on how to search for WFs made by others. More metadata description options, readable WF file or diagram.
Galaxy	Open access. Name, sometimes includes version.	Users can opt to include a WF description. Users that opt to view the WF will be able to see the code, the diagram is only visible when editing.	Translation from Taverna to Galaxy[4]	Prospective and retrospective provenance. (Not explicit)	No licence details, users can view, download or run WF.	Supports execution	Yes	geojson, NetCDF	Can be	Not sure (genome annotations for sure)	Text, tags and filters (name, tag, is published, is importable, is shared with me, and is deleted)	Include link to previous or superseding versions. Create and use a tag ontology that users can expand upon if approved by development team. Provide clear license for WF reuse. Let users view the WF diagram (not only the code) when they choose view WF.
Workflow hub	Open access. Controlled by the author, there are different levels of access, metadata access, access to content and edit privileges. (Users need to register to download WFs.	Metadata includes, input, output and steps (when it's a known format as galaxy, taverna, CWL). Shows diagram. Version control. DOI Citations possible in different formats.	Any format is welcome as long as you fill in the metadata. No conversions. (Native and CWL)	Prospective, retrospective, and evolution. Identifiers for all components (Goble et al., 2021).	License of WF is clearly stated. WF diagram not available. Download and reuse in another compatible platform.	Abstract (CWL) and can include concrete (native).	Not clear but with CWL it should be possible to wrap any functions.	Not specified but should be possible.	Yes, all components packed using RO-Crate	Yes	Text, filter (created at, updated at, tool, WF type, tag, submitter, team, test), and sort (last update, creation data, title, downloads, views).	Create and use a tag ontology that users can expand upon if approved by development team. Implement more metadata fields as mandatory (at the moment only the team and the title are mandatory). There is no standard way to show the WF details, so it's not always easy to define fit for purpose.

(1) https://github.com/WorkflowConversion/KNIME2Grid

(2) https://trac.nbic.nl/elabfactory/wiki/eGalaxy & http://www.taverna.org.uk/documentation/taverna-galaxy/

Appendix II – WF Purpose and Domain Ontology

```
@prefix owl: <https://www.w3.org/TR/2004/REC-owl-semantics-20040210/#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix msp: <http://WFRepo/MSP/> .
@prefix mspv: <http://WFRepo/MSP/vocabulary/> .
@prefix dc: <http://purl.org/dc/terms/> .
@prefix mspwf: <http://WFRepo/MSP/workflow/> .

mspv:dataPreparation a msp:vocabulary;
       skos:prefLabel "Data Preparation";
       dc:partOf mspv:marineSpatialPlanning.

mspv:marineSpatialPlanning  a msp:vocabulary;
       skos:prefLabel "Marine Spatial Planning";
       dc:hasPart             mspv:dataPreparation,             mspv:suitabilityAnalysis,
mspv:environmentalSuitabilityAnalysis,                   mspv:spatialSuitabilityAnalysis,
mspv:networkAnalysis .

mspv:suitabilityAnalysis a msp:vocabulary;
       skos:prefLabel "Suitability Analysis";
       dc:hasPart mspv:environmentalSuitabilityAnalysis, mspv:spatialSuitabilityAnalysis.

mspv:environmentalSuitabilityAnalysis a msp:vocabulary;
       dc:isPartOf mspv:suitabilityAnalysis, mspv:marineSpatialPlanning;
       skos:prefLabel "Environmental Suitability Analysis".

mspv:spatialSuitabilityAnalysis a msp:vocabulary;
       dc:isPartOf mspv:suitabilityAnalysis;
       skos:prefLabel "Spatial Suitability Analysis";
       dc:partOf mspv:suitabilityAnalysis, mspv:marineSpatialPlanning.

mspv:networkAnalysis  a msp:vocabulary;
       skos:prefLabel "Network Analysis".

mspv:bathymetry  a msp:vocabulary;
       skos:prefLabel "Bathymetry";
       dc:isPartOf mspv:environmentalConditions, <http://inspire.ec.europa.eu/theme/el>;
       owl:sameAs mspv:depth.

mspv:depth  a msp:vocabulary;
       skos:prefLabel "Depth";
       dc:isPartOf mspv:environmentalConditions, <http://inspire.ec.europa.eu/theme/el>;
       owl:sameAs mspv:bathymetry.
#data theme
<http://inspire.ec.europa.eu/theme/el>  a msp:vocabulary;
       skos:prefLabel "Elevation".

#data theme
<http://inspire.ec.europa.eu/theme/ps>  a msp:vocabulary;
       skos:prefLabel "Protected Sites"@eng.
mspv:MarineProtectedAreas a msp:vocabulary;
       skos:prefLabel "Marine Protected Areas";
       dc:isPartOf mspv:MarineProtection, mspv:MarineConservation,
       mspv:softUses, <http://inspire.ec.europa.eu/theme/ps>;
       owl:sameAs mspv:MPA .
mspv:MPA a msp:vocabulary;
```

```
        skos:prefLabel "MPA";
        dc:isPartOf mspv:MarineProtection, mspv:MarineConservation,
        mspv:softUses, <http://inspire.ec.europa.eu/theme/ps> ;
        owl:sameAs     mspv:MarineProtectedAreas.

mspv:MarineProtection  a msp:vocabulary;
        skos:prefLabel "Marine Protection";
        dc:isPartOf      mspv:environmentalConservation,      mspv:environmentalProtection,
mspv:softUses;
        dc:hasPart mspv:MarineProtectedAreas ;
        owl:sameAs     mspv:MarineConservation.

mspv:MarineConservation       a msp:vocabulary;
        skos:prefLabel "Marine Conservation";
        dc:isPartOf mspv:environmentalConservation, mspv:environmentalProtection;
        owl:sameAs     mspv:MarineProtection.

mspv:environmentalConservation  a msp:vocabulary;
        skos:prefLabel "Environmental Conservation";
        owl:sameAs     mspv:environmentalProtection.

mspv:environmentalProtection  a msp:vocabulary;
        skos:prefLabel "Environmental Protection";
        owl:sameAs     mspv:environmentalConservation.
mspv:connectivity  a msp:vocabulary;
        skos:prefLabel "Connectivity".

mspv:environmentalConditions  a msp:vocabulary;
        mspv:influence mspv:environmentalSuitability;
        skos:preLabel "Environmental Conditions";
        dc:hasPart mspv:bathymetry,mspv:depth, mspv:sediment, mspv:seabed, mspv:substrate.

mspv:renewableEnergy a msp:vocabulary;
        skos:prefLabel "Renewable Energy";
        dc:isPartOf mspv:hardUses, <https://inspire.ec.europa.eu/theme/er>;
        dc:hasPart mspv:windFarms;
        owl:sameAs mspv:greenEnergy.

mspv:greenEnergy a msp:vocabulary;
        skos:prefLabel "Green Energy";
        dc:isPartOf mspv:hardUses, <https://inspire.ec.europa.eu/theme/er>;
        dc:hasPart mspv:windFarms;
        owl:sameAs mspv:renewableEnergy.

<https://inspire.ec.europa.eu/theme/er> a msp:vocabulary;
        skos:prefLabel "Energy Resources".

mspv:influence a msp:vocabulary;
        skos:prefLabel "Influence".
mspv:environmentalSuitability a msp:vocabulary;
        skos:prefLabel "Environmental Suitability".

mspv:offshore a msp:vocabulary;
        skos:prefLabel "Offshore".
```

```
mspv:sediment a msp:vocabulary;
       skos:prefLabel "Sediment";
       dc:isPartOf mspv:environmentalConditions;
       owl:sameAs mspv:substrate.

mspv:seabed a msp:vocabulary;
       skos:prefLabel "Seabed";
       dc:isPartOf mspv:environmentalConditions.

mspv:substrate a msp:vocabulary;
       skos:prefLabel "Substrate";
       dc:isPartOf mspv:environmentalConditions;
       owl:sameAs mspv:sediment.

mspv:environmentallySuitableAreas a msp:vocabulary;
       skos:prefLabel "Environmentally Suitable Areas";
       mspwf:isOutputof mspv:environmentalSuitabilityAnalysis.

mspv:suitable a msp:vocabulary;
       skos:prefLabel "Suitable";
       owl:sameAs mspv:appropriate.

mspv:appropriate a msp:vocabulary;
       skos:prefLabel "Appropriate";
       owl:sameAs mspv:suitable.

mspv:suitability a msp:vocabulary;
       skos:prefLabel "Suitability".
mspv:sitting a msp:vocabulary;
       skos:prefLabel "Sitting";
       owl:sameAs mspv:allocation.

mspv:allocation a msp:vocabulary;
       skos:prefLabel "Allocation", "Sitting";
       owl:sameAs mspv:sitting.

mspv:hardUses a msp:vocabulary;
       mspv:influence mspv:spatialSuitability;
       skos:prefLabel "Hard Uses";
       dc:hasPart       mspv:windFarms,       mspv:windParks,       mspv:oilGasPlatforms,
mspv:oilGasOffshoreInstallations.

mspv:softUses a msp:vocabulary;
       mspv:influence mspv:spatialSuitability;
       skos:prefLabel "Soft Uses";
       dc:hasPart   mspv:shippingRoutes,   mspv:shippingLanes,   mspv:MarineProtectedAreas,
mspv:MPA.

mspv:windFarms a msp:vocabulary;
       skos:prefLabel "Wind Farms";
       dc:requires mspv:safetyBufferAreas;
       dc:isPartOf  mspv:hardUses,  mspv:renewableEnergy,  <https://inspire.ec.europa.eu/
theme/er>;
       owl:sameAs mspv:windParks.
```

```
mspv:windParks a msp:vocabulary;
       skos:prefLabel "Wind Parks";
       dc:requires mspv:safetyBufferAreas;
       dc:isPartOf  mspv:hardUses,  mspv:renewableEnergy,  <https://inspire.ec.europa.eu/
theme/er>;
       owl:sameAs mspv:windFarms.

mspv:shippingRoutes a msp:vocabulary;
       skos:prefLabel "Shipping Routes";
       dc:isPartOf mspv:shipping, mspv:softUses, <https://inspire.ec.europa.eu/theme/tn>;
       dc:requires mspv:safetyBufferAreas;
       owl:sameAs mspv:shippingLanes.

mspv:shippingLanes a msp:vocabulary;
       skos:prefLabel "Shipping Lanes";
       dc:isPartOf mspv:shipping, mspv:softUses, <https://inspire.ec.europa.eu/theme/tn>;
       dc:requires mspv:safetyBufferAreas;
       owl:sameAs mspv:shippingRoutes.

mspv:shipping a msp:vocabulary;
       dc:isPartOf mspv:softUses, <https://inspire.ec.europa.eu/theme/tn>;
       skos:prefLabel "Shipping", "Maritime Transportation".

mspv:oilGasPlatforms a msp:vocabulary;
       skos:prefLabel "Oil and Gas Platforms", "Oil and Gas Offhore Installations";
       dc:isPartOf                                  <https://inspire.ec.europa.eu/theme/er>,
mspv:oilGasExplorationExploitation, mspv:fossilFuelEnergy, mspv:hardUses;
       dc:requires mspv:safetyBufferAreas, mspv:hydrocarbons;
       owl:sameAs mspv:oilGasOffshoreInstallations.

mspv:oilGasOffshoreInstallations a msp:vocabulary;
       skos:prefLabel "Oil and Gas Offhore Installations";
       dc:isPartOf                                  <https://inspire.ec.europa.eu/theme/er>,
mspv:oilGasExplorationExploitation, mspv:fossilFuelEnergy, mspv:hardUses;
       dc:requires mspv:safetyBufferAreas, mspv:hydrocarbons;
       owl:sameAs mspv:oilGasPlatforms.

mspv:oilGasExplorationExploitation  a msp:vocabulary;
       dc:isPartOf mspv:hardUses, mspv:greyEnergy, mspv:fossilFuelEnergy.

mspv:safetyBufferAreas a msp:vocabulary;
       skos:prefLabel  "Safety Buffer Areas".

mspv:fossilFuelEnergy a msp:vocabulary;
       skos:prefLabel "Fossil Fuel Energy";
       dc:isPartOf                                  <https://inspire.ec.europa.eu/theme/er>,
mspv:oilGasExplorationExploitation;
       owl:sameAs mspv:greyEnergy.

mspv:greyEnergy a msp:vocabulary;
       skos:prefLabel "Grey Energy";
       dc:isPartOf <https://inspire.ec.europa.eu/theme/er>;
       owl:sameAs mspv:fossilFuelEnergy.

mspv:hydrocarbons a msp:vocabulary;
```

```
       skos:prefLabel "Hydrocarbons";
       dc:isPartOf       <https://inspire.ec.europa.eu/theme/er>,       mspv:greyEnergy,
mspv:fossilFuelEnergy.

mspv:oilGasExplorationExploitation a msp:vocabulary;
       skos:prefLabel "Oil and Gas Exploration and Exploitation";
       dc:isPartOf    <https://inspire.ec.europa.eu/theme/er>,    mspv:fossilFuelEnergy,
mspv:hardUses.

mspv:spatiallyNotSuitableAreas a msp:vocabulary;
       skos:prefLabel "Not Suitable Areas";
       dc:hasPart mspv:hardUses, mspv:softUses, mspv:safetyBufferAreas;
       owl:inverseOf mspv:spatiallySuitable.

mspv:spatiallySuitable a msp:vocabulary;
       skos:prefLabel "Spatially Suitable".

<https://inspire.ec.europa.eu/theme/tn>    a msp:vocabulary;
       skos:prefLabel "Transport Networks".
<http://inspire.ec.europa.eu/theme/ge> a msp:vocabulary;
       skos:prefLabel "Geology".
```

Appendix III – Sample of an abstract WFs described in TTL

```
@prefix owl: <https://www.w3.org/TR/2004/REC-owl-semantics-20040210/#> . #protege already has this one by default
@prefix xml:<http://www.w3.org/XML/1998/namespace> . #protege already has this one by default
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns/> . #protege already has this one by default
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> . #protege already has this one by default
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> . #protege already has this one by default, used it also for date time
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dc: <http://purl.org/dc/terms/> .
@prefix gml: <http://www.opengis.net/gml/> .
@prefix ccd: <http://geographicknowledge.de/vocab/CoreConceptData.rdf#> .
@prefix schema: <https://schema.org/> .
@prefix irao: <http://ontology.ethereal.cz/irao> .
@prefix msp: <http://WFRepo/MSP/> .
@prefix mspwf: <http://WFRepo/MSP/workflow/> .
@prefix mspd: <http://WFRepo/MSP/data/> .
@prefix mspv: <http://WFRepo/MSP/vocabulary/> .

<http://WFRepo/MSP/workflow/MPAConnectivityBSV0.7> a mspwf:workflow;
       dc:title "MPAs Connectivity evaluation in the Baltic Sea"@en;
       dc:description "Workflow to evaluate MPAs Connectivity in the Baltic Sea"@en;
       dc:identifier "mspwf@MPAConnectivityBSV0.2"^^xsd:string;
       msp:hasVocabTerm        mspv:MarineProtectedAreas,        mspv:MarineConservation, mspv:MarineProtection, msp:MPA;

#it should be possible to infer this by the input/output data
       mspwf:purpose mspv:connectivity, mspv:networkAnalysis, mspv:marineSpatialPlanning, mspv:MarineConservation;
       dc:source ""^^xsd:string;
       dc:references ""^^xsd:string;#add relevant articles
       dc:bibliographicCitation ""^^xsd:string;
       #use the DOI here when relevant, it could also be a list of relevant literature
       dc:created "2024-02-14T11:04:00Z"^^xsd:date;
       dc:creator [
              vcard:fn "Magali D. do Patrocínio Gonçalves"^^xsd:string;
                     a vcard:Individual;
                     vcard:email "goncalves.m@buas.nl";
              vcard:Organization <https://www.buas.nl>;
                     a vcard:Organization;
                     vcard:fn "Breda University of Applied Sciences",
                            "BUas"];

       dc:contributor "Magali D. do Patrocínio Gonçalves"^^xsd:string;
       dc:rights <https://www.gnu.org/licenses/gpl-3.0.txt>,
       <https://www.gnu.org/licenses/gpl-3.0.html>;
       dc:publisher [
       foaf:name "mspwf";
       foaf:homepage <http://WFRrepo.info> ];
       #(This page does not exist)
```

```
        dc:accessRights <https://www.gnu.org/licenses/gpl-3.0.txt>,
        <https://www.gnu.org/licenses/gpl-3.0.html>;
        dc:rightsHolder ""^^xsd:string;
        dc:version "0.7"^^xsd:string;

dcat:theme <http://inspire.ec.europa.eu/theme/ps>,
        mspv:environmentalProtection;
        irao:DevelopmentStatus " In Progress";
        dc:provenance [
                dc:contributor "Magali D. do Patrocínio Gonçalves"^^xsd:string;
                dc:ProvenanceStatement       ""^^xsd:string ;
                dc:replaces ""^^xsd:string;
                dc:isReplacedBy ""^^xsd:string;
                dc:modified ""^^xsd:date;
                dc:isVersionOf ""^^xsd:string];
                dc:format <https://www.w3.org/TR/turtle/> ;
        dc:language <http://publications.europa.eu/resource/authority/language/ENG>;

        mspwf:hasInput   <http://localhost:8080/geonetwork/srv/resources/datasets/d27df8c0-
de86-4d13-a06d-35a8f50b16fa>,
        <http://WFRepo/MSP/data/10kmPointGridBaltic> ;

        mspwf:hasOutput <http://WFRepo/MSP/data/distanceToMPAsBalticV0.7>.

mspd:output rdfs:subClassOf msp:data.
mspd:intermediateOutput rdfs:subClassOf msp:data.
mspd:finalOutput rdfs:subClassOf msp:data.
mspd:input rdfs:subClassOf msp:data .
mspwf:workflow rdfs:subClassOf msp:workflow.

<http://localhost:8080/geonetwork/srv/resources/datasets/d27df8c0-de86-4d13-
a06d-35a8f50b16fa> a mspd:input,
        ccd:ObjectVector, ccd:region;
        msp:hasVocabTerm  mspv:MarineProtectedAreas, mspv:MPA.

<http://WFRepo/MSP/data/10kmPointGridBaltic> a mspd:input,
        ccd:ObjectVector, ccd:point.


<http://WFRepo/MSP/data/distanceToMPAsBalticV0.7> a mspd:FinalOutput,
        ccd:track, ccd:line;
        msp:hasVocabTerm   mspv:MarineProtectedAreas, mspv:MPA, mspv:connectivity.

# more details regarding metadata
dcat:theme  dc:VocabularyEncodingScheme  "controlled  vocabulary:  GEMET  -  INSPIRE  themes,
version 1.0 (reference date: 2008-06-01)".
dcat:keyword   dc:VocabularyEncodingScheme "controlled vocabulary: GEMET - INSPIRE themes,
version 1.0 (reference date: 2008-06-01)".

<http://publications.europa.eu/resource/authority/language/ENG>
  a skos:Concept, dc:LinguisticSystem ;
  skos:prefLabel "English"@en, "Englisch"@de, "anglais"@fr, "Engels"@nl ;
  skos:inScheme <http://publications.europa.eu/resource/authority/language> .

<http://inspire.ec.europa.eu/theme/ps> rdfs:label "Protected Sites"@eng.
```

```
mspv:MarineProtectedAreas skos:prefLabel "Marine Protected Areas", "MPA";
       dc:partOf mspv:MarineProtection, mspv:MarineConservation.

mspv:MarineProtection skos:prefLabel "Marine Protection";
       dc:isPartOf mspv:environmentalConservation, mspv:environmentalProtection.

mspv:MarineConservation      skos:prefLabel "Marine Conservation";
       dc:isPartOf mspv:environmentalConservation, mspv:environmentalProtection.

mspv:environmentalConservation      skos:prefLabel      "Environmental      Conservation",
"Conservation".

mspv:environmentalProtection skos:prefLabel "Environmental Protection", "Protection".

mspv:connectivity skos:prefLabel "Connectivity".

mspv:networkAnalysis skos:prefLabel "Network Analysis".
```

```
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dc: <http://purl.org/dc/terms/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix ns0: <http://www.w3.org/TR/void/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix ns1: <http://www.opengis.net/rdf#> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .


<http://localhost:8080/geonetwork/srv/resources/catalogs/649cc4b7-
cce8-4abe-817d-69cbe4678bdd> a dcat:Catalog ;
  dc:title "HELCOM Metadata catalogue"@en ;
  dc:description "" ;
  rdfs:label """HELCOM Metadata catalogue (Helcom)
      """@en ;
  foaf:homepage "https://metadata.helcom.fi:/geonetwork/srv/" ;
  ns0:openSearchDescription              "https://metadata.helcom.fi:/geonetwork/srv/eng/
portal.opensearch" ;
  ns0:uriLookupEndpoint       "https://metadata.helcom.fi:/geonetwork/srv/eng/rdf.search?
any=" ;
  dc:publisher <http://localhost:8080/geonetwork/srv/resources/organizations/Helcom> ;
  dcat:themes     <http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeumetadatacodelistConditionsApplyingToAccessAndUse-
ConditionsApplyingToAccessAndUse>,       <http://localhost:8080/geonetwork/srv/resources/
registries/vocabularies/external.theme.httpinspireeceuropaeumetadatacodelistSpatialScope-
SpatialScope>,    <http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeumetadatacodelistProtocolValue-ProtocolValue>,
<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeumetadatacodelistLimitationsOnPublicAccess-
LimitationsOnPublicAccess>,    <http://localhost:8080/geonetwork/srv/resources/registries/
vocabularies/
external.theme.httpinspireeceuropaeumetadatacodelistQualityOfServiceCriteria-
QualityOfServiceCriteria>,     <http://localhost:8080/geonetwork/srv/resources/registries/
vocabularies/external.theme.httpinspireeceuropaeumetadatacodelistPriorityDataset-
PriorityDataset>,              <http://localhost:8080/geonetwork/srv/resources/registries/
vocabularies/external.theme.httpinspireeceuropaeuapplicationschema-applicationschema>,
<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeumediatypes-media-types>,      <http://localhost:8080/
geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeutheme-theme>,   <http://localhost:8080/geonetwork/srv/
resources/registries/vocabularies/external.theme.gemet-en>,       <http://localhost:8080/
geonetwork/srv/resources/registries/vocabularies/external.place.regions>,        <http://
localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeumetadatacodelistOnLineDescriptionCode-
OnLineDescriptionCode> ;
  dc:language "eng" ;
  dcat:dataset         <http://localhost:8080/geonetwork/srv/resources/d27df8c0-de86-4d13-
a06d-35a8f50b16fa> ;
  dcat:record          <https://metadata.helcom.fi:/geonetwork/records/d27df8c0-de86-4d13-
a06d-35a8f50b16fa> .

<http://localhost:8080/geonetwork/srv/resources/organizations/Helcom>
  a foaf:Organization ;
  foaf:name "Helcom" .
```

```
<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeumetadatacodelistConditionsApplyingToAccessAndUse-
ConditionsApplyingToAccessAndUse>
  a skos:ConceptScheme ;
  dc:title "Conditions Applying To Access and Use" ;
  dc:uri              "https://metadata.helcom.fi:/geonetwork/srv/eng/thesaurus.download?
ref=external.theme.httpinspireeceuropaeumetadatacodelistConditionsApplyingToAccessAndUse-
ConditionsApplyingToAccessAndUse" .

<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeumetadatacodelistSpatialScope-SpatialScope>
  a skos:ConceptScheme ;
  dc:title "Spatial scope" ;
  dc:uri              "https://metadata.helcom.fi:/geonetwork/srv/eng/thesaurus.download?
ref=external.theme.httpinspireeceuropaeumetadatacodelistSpatialScope-SpatialScope" .

<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeumetadatacodelistProtocolValue-ProtocolValue>
  a skos:ConceptScheme ;
  dc:title "INSPIRE Protocol values" ;
  dc:uri              "https://metadata.helcom.fi:/geonetwork/srv/eng/thesaurus.download?
ref=external.theme.httpinspireeceuropaeumetadatacodelistProtocolValue-ProtocolValue" .

<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeumetadatacodelistLimitationsOnPublicAccess-
LimitationsOnPublicAccess>
  a skos:ConceptScheme ;
  dc:title "Limitations on public access" ;
  dc:uri              "https://metadata.helcom.fi:/geonetwork/srv/eng/thesaurus.download?
ref=external.theme.httpinspireeceuropaeumetadatacodelistLimitationsOnPublicAccess-
LimitationsOnPublicAccess" .

<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeumetadatacodelistQualityOfServiceCriteria-
QualityOfServiceCriteria>
  a skos:ConceptScheme ;
  dc:title "Quality of Service criteria code" ;
  dc:uri              "https://metadata.helcom.fi:/geonetwork/srv/eng/thesaurus.download?
ref=external.theme.httpinspireeceuropaeumetadatacodelistQualityOfServiceCriteria-
QualityOfServiceCriteria" .

<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeumetadatacodelistPriorityDataset-PriorityDataset>
  a skos:ConceptScheme ;
  dc:title "INSPIRE priority data set" ;
  dc:uri              "https://metadata.helcom.fi:/geonetwork/srv/eng/thesaurus.download?
ref=external.theme.httpinspireeceuropaeumetadatacodelistPriorityDataset-
PriorityDataset" .

<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeuapplicationschema-applicationschema>
  a skos:ConceptScheme ;
  dc:title "INSPIRE application schema register" ;
  dc:uri              "https://metadata.helcom.fi:/geonetwork/srv/eng/thesaurus.download?
```

```
ref=external.theme.httpinspireeceuropaeuapplicationschema-applicationschema" .

<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeumediatypes-media-types>
  a skos:ConceptScheme ;
  dc:title "INSPIRE media-types register" ;
  dc:uri               "https://metadata.helcom.fi:/geonetwork/srv/eng/thesaurus.download?
ref=external.theme.httpinspireeceuropaeumediatypes-media-types" .

<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeutheme-theme>
  a skos:ConceptScheme ;
  dc:title "GEMET - INSPIRE themes, version 1.0" ;
  dc:uri               "https://metadata.helcom.fi:/geonetwork/srv/eng/thesaurus.download?
ref=external.theme.httpinspireeceuropaeutheme-theme" .

<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.gemet-en>
  a skos:ConceptScheme ;
  dc:title "GEMET" ;
  dc:uri               "https://metadata.helcom.fi:/geonetwork/srv/eng/thesaurus.download?
ref=external.theme.gemet-en" .

<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.place.regions>
  a skos:ConceptScheme ;
  dc:title "Continents, countries, sea regions of the world." ;
  dc:uri               "https://metadata.helcom.fi:/geonetwork/srv/eng/thesaurus.download?
ref=external.place.regions" .

<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
external.theme.httpinspireeceuropaeumetadatacodelistOnLineDescriptionCode-
OnLineDescriptionCode>
  a skos:ConceptScheme ;
  dc:title "Online description code" ;
  dc:uri               "https://metadata.helcom.fi:/geonetwork/srv/eng/thesaurus.download?
ref=external.theme.httpinspireeceuropaeumetadatacodelistOnLineDescriptionCode-
OnLineDescriptionCode" .

<https://metadata.helcom.fi:/geonetwork/records/d27df8c0-de86-4d13-a06d-35a8f50b16fa>
  a dcat:CatalogRecord ;
  foaf:primaryTopic    <http://localhost:8080/geonetwork/srv/resources/d27df8c0-de86-4d13-
a06d-35a8f50b16fa> ;
  dc:conformsTo <https://www.iso.org/standard/32557.html> ;
  dc:issued "2023-05-03" ;
  dc:modified "2023-05-03" .

<http://localhost:8080/geonetwork/srv/resources/datasets/d27df8c0-de86-4d13-
a06d-35a8f50b16fa> a dcat:Dataset ;
  dc:identifier "d27df8c0-de86-4d13-a06d-35a8f50b16fa" ;
  dc:title "HELCOM MPAs" ;
  dc:abstract """This dataset contains borders of the HELCOM MPAs (former Baltic Sea
Protected Areas (BSPAs). The dataset has been compiled from data submitted by HELCOM
Contracting Parties. It includes the borders of designated HELCOM MPAs stored in the
http://mpas.helcom.fi. The designation is based on the HELCOM Recommendation 15/5 (1994).
```

```
The dataset displays all designated or managed MPAs as officially reported to HELCOM by
the respective Contracting Party. The latest related HELCOM publication based on MPA
related data is http://www.helcom.fi/Lists/Publications/BSEP148.pdf
  The dataset contains the following information:
MPA_ID: Unique ID of the MPA as used in HELCOM Marine Protected Areas database
Name: Name of the MPA
Country: Country where MPA is located
Site_link: Direct link to site's fact sheet in the http://mpas.helcom.fi where additional
information is available
MPA_status: Management status of the MPA
Date_est: Establishment date of the MPA
Year_est: Establishment year of the MPA""" ;
  dcat:keyword "MADS",
  "Marine Protected Areas",
  "MPA",
  "Marine Protection",
  "Marine Conservation";
  dcat:theme      <http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
GEMET%2520-%2520INSPIRE%2520themes%252C%2520version%25201.0/concepts/>,          <http://
localhost:8080/geonetwork/srv/resources/registries/vocabularies/GEMET/concepts/>,
<https://inspire.ec.europa.eu/metadata-codelist/TopicCategory/environment> ;
  dc:spatial [
    a <http://www.opengis.net/rdf#Polygon> ;
    ns1:asWKT """
            <http://www.opengis.net/def/crs/OGC/1.3/CRS84>
            Polygon((9.171427   52.525805,   9.171427   66.331433,   33.044318   66.331433,
33.044318 52.525805, 9.171427 52.525805))
          """^^ns1:WKTLiteral;
    rdfs:label "Baltic Sea"@en;
    dc:isPartOf "Europe"@en
  ] ;
  dc:temporal """2010-02-01
          /
          2022-12-31""" ;
  dc:updated "2022-12-05" ;
  dcat:contactPoint [
   a vcard:Organization ;
   vcard:fn "HELCOM Secretariat"@en ;
   vcard:hasEmail <mailto:data@helcom.fi> ;
   a vcard:Organization ;
    vcard:fn "MSP workflows"@en ;
    vcard:hasTelephone "" ;
    vcard:hasEmail "" ;
    vcard:hasURL <http://WFRepo/MSP/workflow>
    ];
    dc:publisher     <http://localhost:8080/geonetwork/srv/resources/organizations/HELCOM%
20Secretariat> ;
  dc:language "eng" ;
  dc:license "Use constraints: Data can be used freely given that the source (HELCOM) is
cited.",         <http://inspire.ec.europa.eu/metadata-codelist/LimitationsOnPublicAccess/
noLimitations> ;
  dcat:distribution    <https://metadata.helcom.fi:/geonetwork/records/d27df8c0-de86-4d13-
a06d-35a8f50b16fa#WWW%3ALINK-1.0-http--link-Download%20dataset>,                <https://
metadata.helcom.fi:/geonetwork/records/d27df8c0-de86-4d13-a06d-35a8f50b16fa#WWW%
3ALINK-1.0-http--link-Open%20in%20Map%20Viewer> ;
```

dcat:dataQuality """This dataset is compiled from data submitted by HELCOM Contracting States. It includes the borders of designated HELCOM MPAs.

The dataset was updated in May 2023 to update Status of German MPAs: "Fehmarnbelt", "Kadetrinne" and "Pommersche Bucht-Rönnebank"

The dataset was updated December 2022 to update new MPAs reported by Germany (MPA_IDs 593-600).

The dataset was updated June 2022 to update 3 new MPAs by Sweden and modifying extent of MPA_ID=118 (Fladen).

The dataset was updated October 2019 to accommodate updated status of management plans.

The dataset was updated April 2019 by HELCOM Secretariat to add a new MPA (MPA ID:522 - Södra Sandbäck) as designated by Finland, and to include new areas to 2 existing MPAs (MPA ID: 143 - Saaristomeri /Archipelago Sea and MPA ID: 157 - Tulliniemen linnustonsuojelualue/ Tulliniemi bird protection area) as designated by Finland. In addition the whole Finnish MPA network was re-projected to remove the shift in the features that occurred before. Also the spatial delineation of following Finnish MPAs were revised to remove islands and water areas that are not included in the MPA IDs: 161, 159 and 144.

The dataset was updated in January 2019 by HELCOM Secretariat in order to update MPA establishment dates for Germany based on corrected information.

The dataset was updated in February 2018 by HELCOM Secretariat in order to update attribute table content (management plan status) according to updates carried out by Contracting Parties to HELCOM MPA database.

In March 2017 two new MPAs were included (MPA_ID:533 - Sambija plynaukšte Biosphere Polygon and MPA_ID:532 - Klaipeda-Ventpilis plynaukšte Biosphere Polygon) as designated by Lithuania.

Following changes has been made until april 2016: Removed attributes "Location_S", "AreaTot_Ha", "AreaMar_Ha", "N2K_Site", changed following attributes names: "BSPA_ID" to "MPA_ID", "BSAPLink" to "Site_link" and added attribute "Date_est" according to new http://mpas.helcom.fi.

Following changes has been made until August 2015: Modified HELCOM MPA: Küstenlandschaft Bottsand - Marzkamp u. vorgelagerte Flachgründe, Östlichen Kieler Bucht

Following changes has been made until February 2015: Added HELCOM MPAs: Luodematalat (February 2015) Länsileton alue (February 2015) Sandkallanin eteläpuolinen merialue (February 2015) Hangon itäinen selkä (February 2015) Merikallat (February 2015) Björkör (February 2015) Boxö (February 2015) Långör - Östra Sundskär (February 2015) Signilskär - Märket (February 2015) Lågskär (February 2015) Bogskär (February 2015)

Following changes has been made until July 2013: Removed BSPAs: 1 Finnish BSPA, BSPA_ID: 146 (Åland Area: Signilskär/Märket) (June 2012) 2 Polish BSPAs, BSPA_ID: 87, 182 1 Russian BSPAs, BSPA_ID: 162 (Ingermanlandskiy) 4 Swedish BSPAs, BSPA_ID: 114-121 5 Latvian BSPAs, BSPA_ID: 96-100 (June 2012, April 2013) 1 Danish BSPA, BSPA_ID: 290 (July 2013) Added BSPAs: 4 Swedish (ID 297-300 (May 2011) 7 Latvian BSPAs, ID: 303-309 (April 2013) 2 Lithuanian BSPAs, ID: 310-311 (july 2013) Modied BSPA areas: 2 Danish BSPA:s, ID 244, 249 (July 2013) 2 Russian BSPA:s, ID 163-164, (July 2013).

```
During 2009-2010 the data was updated within a project to analyse the ecological coherence of the HELCOM protected areas network. Results are published in "HELCOM 2010. Towards an ecologically coherent network of well-managed Marine Protected Areas – Implementation report on the status and ecological coherence of the HELCOM BSPA network. Balt. Sea Environ. Proc. No. 124B." Available online at http://http://www.helcom.fi/Lists/Publications/BSEP124B.pdf""" .

<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/GEMET%2520-%2520INSPIRE%2520themes%252C%2520version%25201.0/concepts/>
  a skos:Concept ;
  skos:inScheme   <http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/GEMET%2520-%2520INSPIRE%2520themes%252C%2520version%25201.0> ;
  skos:prefLabel "" .

<https://metadata.helcom.fi:/geonetwork/records/d27df8c0-de86-4d13-a06d-35a8f50b16fa#WWW%3ALINK-1.0-http--link-Download%20dataset>
  a dcat:Distribution ;
  dcat:accessURL     "https://maps.helcom.fi/website/MADS/download/?id=d27df8c0-de86-4d13-a06d-35a8f50b16fa" ;
  dc:title "Download dataset" ;
  dc:format "WWW:LINK-1.0-http--link" .

<https://metadata.helcom.fi:/geonetwork/records/d27df8c0-de86-4d13-a06d-35a8f50b16fa#WWW%3ALINK-1.0-http--link-Open%20in%20Map%20Viewer>
  a dcat:Distribution ;
  dcat:accessURL           "https://maps.helcom.fi/website/mapservice/?datasetID=d27df8c0-de86-4d13-a06d-35a8f50b16fa" ;
  dc:title "Open in Map Viewer" ;
  dc:format "WWW:LINK-1.0-http--link" .
```

```
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dc: <http://purl.org/dc/terms/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix ns0: <http://www.w3.org/TR/void/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix ns1: <http://www.opengis.net/rdf#> .
@prefix parameter: <http://www.ontologydesignpatterns.org/cp/owl/parameter.owl#> .
@prefix msp: <http://WFRepo/MSP/> .
@prefix mspwf: <http://WFRepo/MSP/workflow/> .
@prefix mspd: <http://WFRepo/MSP/data/> .
@prefix mspv: <http://WFRepo/MSP/workflow/vocabulary/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .

<http://WFRepo/MSP/data/10kmPointGridBaltic> a dcat:Dataset,
  mspd:input;
  dc:identifier "10kmPointGridBaltic" ;
  dc:title "Ten kilometer grid points covering the Baltic Sea" ;
  dc:abstract "This layer contains a 10km points grid covering the Baltic Sea.";
  dcat:keyword "Baltic Sea", "grid", "10km)";
  dcat:theme "connectivity";
  dc:spatial [
    a <http://www.opengis.net/rdf#Point> ;
    ns1:asWKT """
            <http://www.opengis.net/def/crs/OGC/1.3/CRS84>
            Polygon((9.171427   52.525805,   9.171427   66.331433,   33.044318   66.331433,
33.044318 52.525805, 9.171427 52.525805))
          """^^ns1:WKTLiteral;
    rdfs:label "Baltic Sea"@en;
    dc:isPartOf "Europe"@en
    ] ;
       dcat:contactPoint [
    a vcard:Organization ;
    vcard:fn "MSP workflows"@en ;
    vcard:hasTelephone "" ;
    vcard:hasEmail "" ;
    vcard:hasURL <http://WFRepo/MSP/workflow>
    ];
  dc:temporal """2010-02-01/2022-12-31""" ;
  dc:updated "2024-02-05" ;
  dc:publisher "" ;
  dc:language "eng" ;
  dc:license "Use constraints: Data can be used freely given that the source (HELCOM) is
cited.",        <http://inspire.ec.europa.eu/metadata-codelist/LimitationsOnPublicAccess/
noLimitations> ;
  dcat:distribution "" ;
  dcat:dataQuality """""" .

<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/GEMET%2520-%
2520INSPIRE%2520themes%252C%2520version%25201.0/concepts/>
  a skos:Concept ;
skos:inScheme   <http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
GEMET%2520-%2520INSPIRE%2520themes%252C%2520version%25201.0> ;
  skos:prefLabel "" .
```

```
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dc: <http://purl.org/dc/terms/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix ns0: <http://www.w3.org/TR/void/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix ns1: <http://www.opengis.net/rdf#> .
@prefix parameter: <http://www.ontologydesignpatterns.org/cp/owl/parameter.owl#> .
@prefix msp: <http://WFRepo/MSP/> .
@prefix mspwf: <http://WFRepo/MSP/workflow/> .
@prefix mspd: <http://WFRepo/MSP/data/> .
@prefix mspv: <http://WFRepo/MSP/workflow/vocabulary/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .

<http://WFRepo/MSP/data/distanceToMPAsBalticV0.7> a
  mspd:finalOutput;
  dc:identifier "distanceToMPAsBaltic" ;
  dc:title "Distance to MPAs in the Baltic" ;
  dc:abstract "This layer contains lines connecting a 10km grid of points in the Baltic
Sea to a proposed network of MPAs to which the distance to an MPA will be lower than 200km
";
  dcat:keyword "MADS",
  "Marine Protected Areas",
  "MPA",
  "Marine Protection",
  "Marine Conservation";
  dcat:theme <http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/GEMET
%2520-%2520INSPIRE%2520themes%252C%2520version%25201.0/concepts/>, <http://localhost:8080/
geonetwork/srv/resources/registries/vocabularies/GEMET/concepts/>,              <https://
inspire.ec.europa.eu/metadata-codelist/TopicCategory/environment>,          "environment",
"conservation", "protection", "connectivity";
  dc:spatial [
    a <http://www.opengis.net/rdf#line> ;
    ns1:asWKT """
            <http://www.opengis.net/def/crs/OGC/1.3/CRS84>
            Polygon((9.171427   52.525805,   9.171427   66.331433,   33.044318   66.331433,
33.044318 52.525805, 9.171427 52.525805))
          """^^ns1:WKTLiteral;
    rdfs:label "Baltic Sea"@en;
    dc:isPartOf "Europe"@en
    ] ;
       dcat:contactPoint [
    a vcard:Organization ;
    vcard:fn "HELCOM Secretariat"@en ;
    vcard:hasEmail <mailto:data@helcom.fi> ;
    a vcard:Organization ;
    vcard:fn "MSP workflows"@en ;
    vcard:hasTelephone "" ;
    vcard:hasEmail "" ;
    vcard:hasURL <http://WFRepo/MSP/workflow>
    ];
  dc:temporal """2010-02-01
          /
          2022-12-31""" ;
  dc:updated "2024-02-05" ;
```

```
  dc:publisher "" ;
  dc:language "eng" ;
  dc:license "Use constraints: Data can be used freely given that the source (HELCOM) is
cited.",         <http://inspire.ec.europa.eu/metadata-codelist/LimitationsOnPublicAccess/
noLimitations> ;
  dcat:distribution "" ;
  dcat:dataQuality """""" ;
  dc:provenance [
        a dc:ProvenanceStatement;
        rdfs:label "The points layer was used to calculate the connectivity to the hubs,
the Marine Protected Areas being the hubs.";
        mspwf:referenceObject     <http://localhost:8080/geonetwork/srv/resources/datasets/
d27df8c0-de86-4d13-a06d-35a8f50b16fa>,
               <http://WFRepo/MSP/data/10kmPointGridBaltic> ;
        mspwf:distanceToHub [
                        parameter:hasParameter[
                        mspwf:parName "distance";
                        mspwf:parValue "units:km"]];
                        mspwf:opPurpose "Calculate the distance to the MPAs from equidistant
network to evaluate connectivity"].

<http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/GEMET%2520-%
2520INSPIRE%2520themes%252C%2520version%25201.0/concepts/>
  a skos:Concept ;
  skos:inScheme   <http://localhost:8080/geonetwork/srv/resources/registries/vocabularies/
GEMET%2520-%2520INSPIRE%2520themes%252C%2520version%25201.0> ;
  skos:prefLabel "" .
```

Appendix IV – SPARQL Queries

Query 1 (general use)

Describe workflow

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX gmd: <http://www.isotc211.org/2005/gmd#>
PREFIX mspwf:<http://WFRepo/MSP/workflow/>

DESCRIBE ?workflow WHERE {
  ?workflow a mspwf:workflow .
}
```

Query 2 (Use case I - Frederica)

Find Workflows by author

```
PREFIX msp: <http://WFRepo/MSP>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX mspwf: <http://WFRepo/MSP/workflow/>
PREFIX vcard: <http://www.w3.org/2006/vcard/ns#>
SELECT *
WHERE {
   ?workflow a mspwf:workflow.
   ?workflow dc:creator [
      vcard:fn ?creatorname] .
   FILTER (REGEX (STR (?creatorname), ".*Magali.*")).
}
```

Query 3 (Use case I - Frederica)

Find Workflows by author's affiliation

```
PREFIX msp: <http://WFRepo/MSP>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX mspwf: <http://WFRepo/MSP/workflow/>
PREFIX vcard: <http://www.w3.org/2006/vcard/ns#>
SELECT DISTINCT ?workflow ?creatorName
WHERE {
   ?workflow a mspwf:workflow.
   ?workflow dc:creator ?creator.
   ?creator vcard:fn ?creatorName.
   FILTER (REGEX (STR (?creatorName), ".*BUas.*")).
}
```

Query 4 (Use case II – Renata)

Find workflows by type of analysis

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX mspwf: <http://WFRepo/MSP/workflow/>

SELECT ?workflow ?analysis
WHERE {
   ?workflow a mspwf:workflow;
      mspwf:purpose ?analysis.
   FILTER  (REGEX (STR (?analysis), ".*network.*") )
} limit 100
```

Query 5 (Use case II – Renata)

Find workflows by theme (in workflow)

a) Using domain ontology:

```
PREFIX mspwf: <http://WFRepo/MSP/workflow/>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX irao: <http://ontology.ethereal.cz/irao>
PREFIX mspv: <http://WFRepo/MSP/vocabulary/>
PREFIX msp: <http://WFRepo/MSP/>
SELECT DISTINCT ? workflow
WHERE {
   ? workflow a mspwf:workflow.
   ? workflow msp:hasVocabTerm ?term.
   FILTER (?term =mspv:MarineConservation).
}
```

b) Using the inspire theme

```
PREFIX mspwf: <http://WFRepo/MSP/workflow/>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX irao: <http://ontology.ethereal.cz/irao>
PREFIX mspv: <http://WFRepo/MSP/vocabulary/>
PREFIX msp: <http://WFRepo/MSP/>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
SELECT DISTINCT ?workflow

WHERE {
   ?workflow a mspwf:workflow;
      dcat:theme ?theme.
FILTER (REGEX (STR (?theme), ".*Protection.*") ).
}
```

c) Using the WF description

```
PREFIX mspwf: <http://WFRepo/MSP/workflow/>
PREFIX mspv: <http://WFRepo/MSP/vocabulary/>
PREFIX msp: <http://WFRepo/MSP/>
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dc: <http://purl.org/dc/terms/>
SELECT DISTINCT ?workflow ?title
        WHERE {
   ?workflow a mspwf:workflow;
      dc:description ?title.
FILTER (REGEX (STR (?title), ".*MPAs Connectivity.*") ).
}
```

Query 6 (Use case II – Renata)

Find workflows by theme (in input data)

a) Looking at the input data abstract

```
PREFIX mspwf: <http://WFRepo/MSP/workflow/>
PREFIX dc: <http://purl.org/dc/terms/>

select ?wfTitle ?input ?abstract
where {
  ?workflow a mspwf:workflow;
        dc:title ?wfTitle;
     mspwf:hasInput ?input.
  ?input dc:abstract ?abstract.
  FILTER  (REGEX (STR (?abstract), ".*MPA.*") ).
}
```

b) Looking at the input data title

```
PREFIX mspwf: <http://WFRepo/MSP/workflow/>
PREFIX dc: <http://purl.org/dc/terms/>

select ?wfTitle ?input ?title
where {
  ?workflow a mspwf:workflow;
        dc:title ?wfTitle;
     mspwf:hasInput ?input.
  ?input dc:title ?title.
  FILTER  (REGEX (STR (?title), ".*MPA.*") ).
}
```

Query 7 (Use case II – Renata)

Search all versions with different parameters

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX mspwf: <http://WFRepo/MSP/workflow/>
PREFIX parameter: <http://www.ontologydesignpatterns.org/cp/owl/parameter.owl#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX mspd: <http://WFRepo/MSP/data/>
PREFIX msp: <http://WFRepo/MSP/>

SELECT DISTINCT ?workflow ?outputtitle ?parametername ?value ?purpose
WHERE {
 ?workflow a msp:workflow ;
   mspwf:hasOutput ?output.
 ?output a msp:data;
       dc:title ?outputtitle.

 OPTIONAL {
  ?output dc:provenance [
   ?processType [
    parameter:hasParameter [
     mspwf:parName ?parametername ;
     mspwf:parValue ?value
    ] ] ;
    mspwf:opPurpose ?purpose
   ].
 }
}
ORDER by ?outputtitle
```

Query 8 (Use case III – Franck)

Retrieve all WF versions with WF evolution parameters.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX mspwf: <http://WFRepo/MSP/workflow/>
PREFIX parameter: <http://www.ontologydesignpatterns.org/cp/owl/parameter.owl#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX mspd: <http://WFRepo/MSP/data/>
PREFIX msp: <http://WFRepo/MSP/>

SELECT DISTINCT ?workflow ?version ?replaces ?replacedBy ?versionOf ?modification
WHERE {
 ?workflow a msp:workflow ;
    dc:version ?version;
  dc:provenance [
     dc:replaces ?replaces;
     dc:isReplacedBy ?replacedBy;
     dc:isVersionOf ?versionOf;
     dc:ProvenanceStatement ?modification
  ].
}
ORDER by ?workflow
```

Query 9 (Use case III – Franck)

Retrieve a certain WF's versions with version details and ordered by maturity status.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX mspwf: <http://WFRepo/MSP/workflow/>
PREFIX parameter: <http://www.ontologydesignpatterns.org/cp/owl/parameter.owl#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX mspd: <http://WFRepo/MSP/data/>
PREFIX msp: <http://WFRepo/MSP/>
PREFIX irao: <http://ontology.ethereal.cz/irao>

SELECT DISTINCT ?workflow ?description ?version ?replaces ?replacedBy ?versionOf ?modification ?maturity
WHERE {
 ?workflow a msp:workflow ;
   dc:description ?description;
    dc:version ?version;
     irao:DevelopmentStatus ?maturity;
  dc:provenance [
      dc:replaces ?replaces;
      dc:isReplacedBy ?replacedBy;
      dc:isVersionOf ?versionOf;
      dc:ProvenanceStatement ?modification
   ].
FILTER  (REGEX (STR (?description), ".*spatially suitable areas.*") ).
}
ORDER by ?maturity
```

Query 10 (Use case IV – Tom)

Search WFs by output description

```
PREFIX mspwf: <http://WFRepo/MSP/workflow/>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX msp: <http://WFRepo/MSP/>
PREFIX mspd: <http://WFRepo/MSP/data/>

SELECT DISTINCT ?workflow ?inputTitle ?wfdescription ?fOutputTitle
WHERE {
 ?workflow a mspwf:workflow ;
   dc:description ?wfdescription;
   mspwf:hasInput ?input;
   mspwf:hasOutput ?output.
 ?input a mspd:input;
   dc:title ?inputTitle.
 ?output a mspd:finalOutput;
   dc:title ?fOutputTitle.
FILTER (REGEX (STR (?wfdescription), ".suitable", "i"))
}
```

Query 11 (Use case IV – Tom)

Search WFs by intermediate output

```
PREFIX mspwf: <http://WFRepo/MSP/workflow/>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX msp: <http://WFRepo/MSP/>
PREFIX mspd: <http://WFRepo/MSP/data/>

SELECT DISTINCT ?WFTitle ?wfDescription ?intOutputTitle
WHERE {
  ?workflow a mspwf:workflow;
    dc:title ?WFTitle;
     dc:description  ?wfDescription;
    mspwf:hasOutput ?output.
  ?output a mspd:intermediateOutput;
    dc:title ?intOutputTitle.
FILTER (REGEX (STR (?intOutputTitle), ".safety", "i"))
}
```

Query 12 (Use case IV – Tom)

Search WFs that are part of other WFs

```
PREFIX mspwf: <http://WFRepo/MSP/workflow/>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX msp: <http://WFRepo/MSP/>
PREFIX mspd: <http://WFRepo/MSP/data/>

SELECT DISTINCT ?WFTitle1 ?workflow2
WHERE {
  ?workflow1 a mspwf:workflow;
    dc:title ?WFTitle1;
    dc:hasPart ?workflow2.
}
```

Appendix V – Sparnatural's Configuration File (TTL)

```
@prefix : <http://data.example.com/my-sparnatural-config#> .
@prefix dc: <http://purl.org/dc/terms/> .
@prefix qb: <http://purl.org/linked-data/cube#> .
@prefix sh: <http://www.w3.org/ns/shacl#> .
@prefix geo: <http://www.opengis.net/ont/geosparql#> .
@prefix msp: <http://WFRepo/MSP/> .
@prefix org: <http://www.w3.org/ns/org#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix adms: <http://www.w3.org/ns/adms#> .
@prefix core: <http://data.sparna.fr/ontologies/sparnatural-config-core#> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix doap: <http://usefulinc.com/ns/doap#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix geof: <http://www.opengis.net/def/function/geosparql/> .
@prefix mspd: <http://WFRepo/MSP/data/> .
@prefix mspv: <http://WFRepo/MSP/vocabulary/> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix this: <http://data.example.com/my-sparnatural-config/> .
@prefix euvoc: <http://publications.europa.eu/ontology/euvoc#> .
@prefix mspwf: <http://WFRepo/MSP/workflow/> .
@prefix vcard: <http://www.w3.org/2006/vcard/ns#> .
@prefix schema: <http://schema.org/> .
@prefix skosxl: <http://www.w3.org/2008/05/skos-xl#> .
@prefix xls2rdf: <https://xls2rdf.sparna.fr/vocabulary#> .
@prefix skosthes: <http://purl.org/iso25964/skos-thes#> .
@prefix parameter: <http://www.ontologydesignpatterns.org/cp/owl/parameter.owl#> .
@prefix datasources: <http://data.sparna.fr/ontologies/sparnatural-config-datasources#> .
@base <http://data.example.com/my-sparnatural-config#> .

<http://data.example.com/my-sparnatural-config> rdf:type owl:Ontology ;
       owl:imports <http://data.sparna.fr/ontologies/sparnatural-config-core> ,
       <http://data.sparna.fr/ontologies/sparnatural-config-datasources> ;
    dc:format   <https://xls2rdf.sparna.fr/rest/convert?url=https%3A%2F%2Fdocs.google.com%
2Fspreadsheets%2Fd%2F1DqvousSlnkSg4TB_5czBcuFeirp6FkofSlirtueXob0%2Fexport%3Fformat%
3Dxlsx&noPostProcessings=true> ;
    dc:source                                    <https://docs.google.com/spreadsheets/
d/1DqvousSlnkSg4TB_5czBcuFeirp6FkofSlirtueXob0> .

#################################################################
#    Annotation properties
#################################################################

###  http://purl.org/dc/terms/format
dc:format rdf:type owl:AnnotationProperty .


###  http://purl.org/dc/terms/source
dc:source rdf:type owl:AnnotationProperty .
```

```
#################################################################
#    Object Properties
#################################################################

###  http://WFRepo/MSP/hasVocabTerm
msp:hasVocabTerm rdf:type owl:ObjectProperty ;
        rdfs:subPropertyOf core:NonSelectableProperty ;
    rdfs:domain msp:data ,
        msp:workflow ;
         rdfs:range msp:vocabulary ;
         core:sparqlString "<http://WFRepo/MSP/hasVocabTerm>" ;
         rdfs:label "Has vocabulary term" .

###  http://WFRepo/MSP/workflow/hasInput
mspwf:hasInput rdf:type owl:ObjectProperty ;
        rdfs:subPropertyOf core:ListProperty ;
    owl:inverseOf mspwf:isInputOf ;
    rdfs:domain msp:workflow ;
    rdfs:range msp:data ;
    rdfs:label "Has Input data"@en .

###  http://WFRepo/MSP/workflow/hasOutput
mspwf:hasOutput rdf:type owl:ObjectProperty ;
        rdfs:subPropertyOf core:ListProperty ;
    owl:inverseOf mspwf:isOutputOf ;
    rdfs:domain msp:workflow ;
    rdfs:range msp:data ;
    rdfs:label "Has Output data"@en .


###  http://WFRepo/MSP/workflow/isInputOf
mspwf:isInputOf rdf:type owl:ObjectProperty ;
    rdfs:subPropertyOf core:ListProperty ;
    rdfs:domain msp:data ;
    rdfs:range msp:workflow ;
    core:sparqlString "^<http://WFRepo/MSP/workflow/hasInput>" ;
    rdfs:label "Is input of"@en .


###  http://WFRepo/MSP/workflow/isOutputOf
mspwf:isOutputOf rdf:type owl:ObjectProperty ;
        rdfs:subPropertyOf core:ListProperty ;
    rdfs:domain msp:data ;
    rdfs:range msp:workflow ;
    core:sparqlString "^<http://WFRepo/MSP/workflow/hasOutput>" ;
    rdfs:label "is output of"@en .


###  http://WFRepo/MSP/workflow/purpose
mspwf:purpose rdf:type owl:ObjectProperty ;
    rdfs:subPropertyOf core:AutocompleteProperty ;
    rdfs:domain msp:workflow ;
    rdfs:range this:Literal ;
    core:sparqlString    "<http://WFRepo/MSP/workflow/purpose>/<http://www.w3.org/2004/02/
skos/core#prefLabel>" ;
```

```
    rdfs:label "Workflow purpose" .


###  http://WFRepo/MSP/workflow/referenceObject
mspwf:referenceObject rdf:type owl:ObjectProperty ;
    rdfs:subPropertyOf core:ListProperty ;
    rdfs:domain msp:data ;
    rdfs:range msp:data ;
    core:sparqlString  "<http://purl.org/dc/terms/provenance>/<http://WFRepo/MSP/workflow/
referenceObject>" ;
    core:tooltip "the dataset that was modified to create the present dataset" ;
    rdfs:label "reference object" .

###  http://data.example.com/my-sparnatural-config/WFTtitle
this:WFTtitle rdf:type owl:ObjectProperty ;
       rdfs:subPropertyOf core:AutocompleteProperty ;
    rdfs:domain msp:workflow ;
    rdfs:range this:Literal ;
    core:sparqlString "<http://purl.org/dc/terms/title>" ;
    rdfs:label "Title" .


###  http://data.example.com/my-sparnatural-config/workflowHasPart
this:workflowHasPart rdf:type owl:ObjectProperty ;
                     rdfs:subPropertyOf core:SearchProperty ;
                     owl:inverseOf <this:workflowIsPartOf> ;
                     rdfs:domain msp:workflow ;
                     rdfs:range msp:workflow ;
                     core:sparqlString "<http://purl.org/dc/terms/hasPart>" ;
                     rdfs:label "Has part" .


###  http://purl.org/dc/elements/1.1/isVersionOf
<http://purl.org/dc/elements/1.1/isVersionOf> rdf:type owl:ObjectProperty ;
    rdfs:subPropertyOf core:AutocompleteProperty ;
    rdfs:domain msp:workflow ;
    rdfs:range msp:workflow ;
       core:sparqlString       "<http://purl.org/dc/terms/provenance>/<http://purl.org/dc/
terms/isVersionOf>" ,
       "^<http://purl.org/dc/terms/hasVersion>" ;
    core:tooltip "this workflow is a version of another WF (the input and parameter values
might differ, the operations and their order remain the same)" ;
       rdfs:label "is version of" .


###  http://purl.org/dc/elements/1.1/replaces
<http://purl.org/dc/elements/1.1/replaces> rdf:type owl:ObjectProperty ;
    rdfs:subPropertyOf core:AutocompleteProperty ;
    rdfs:domain msp:workflow ;
    rdfs:range this:Literal ;
    core:sparqlString    "<http://purl.org/dc/terms/provenance>/<http://purl.org/dc/terms/
replaces>" ;
    core:tooltip "This workflow replaces another workflow that is now deprecated" ;
    rdfs:label "replaces version" .
```

```
###  http://purl.org/dc/terms/ProvenanceStatement
dc:ProvenanceStatement rdf:type owl:ObjectProperty ;
       rdfs:subPropertyOf core:SearchProperty ;
    rdfs:domain msp:workflow ;
    rdfs:range this:Literal ;
    core:sparqlString    "<http://purl.org/dc/terms/provenance>/<http://purl.org/dc/terms/
ProvenanceStatement>" ;
    rdfs:comment "Changes made" ;
    rdfs:label "Changes description" .

###  http://purl.org/dc/terms/created
dc:created rdf:type owl:ObjectProperty ;
       rdfs:subPropertyOf core:TimeProperty-Date ;
    rdfs:domain msp:workflow ;
    rdfs:range <http://data.example.com/my-sparnatural-config/Literal/Date> ;
    core:exactDateProperty "" ;
    rdfs:label "created" .


###  http://purl.org/dc/terms/isReplacedBy
dc:isReplacedBy rdf:type owl:ObjectProperty ;
    rdfs:subPropertyOf core:AutocompleteProperty ,
       core:ListProperty ;
    rdfs:domain msp:workflow ;
    rdfs:range msp:workflow ;
    core:sparqlString    "<http://purl.org/dc/terms/provenance>/<http://purl.org/dc/terms/
isReplacedBy>" ;
    rdfs:label "Is replaced by"@en .

###  http://purl.org/dc/terms/license
dc:license rdf:type owl:ObjectProperty ;
       rdfs:subPropertyOf core:ListProperty ;
    rdfs:domain msp:data ;
    rdfs:range this:Literal ;
    core:sparqlString "<http://www.w3.org/ns/dcat#distribution>/<http://purl.org/dc/terms/
license>" ;
    rdfs:label "Data license" .

###  http://purl.org/dc/terms/modified
dc:modified rdf:type owl:ObjectProperty ;
       rdfs:subPropertyOf core:TimeProperty-Date ;
    rdfs:domain msp:data ;
    rdfs:range <http://data.example.com/my-sparnatural-config/Literal/Date> ;
    core:exactDateProperty "" ;
    rdfs:comment "Last modification date" ;
    rdfs:label "Last modification date" .

###  http://purl.org/dc/terms/replaces
dc:replaces rdf:type owl:ObjectProperty ;
       rdfs:subPropertyOf core:ListProperty ;
    rdfs:domain msp:workflow ;
    rdfs:range msp:workflow ;
    rdfs:label "Replaces"@en .

###  http://purl.org/dc/terms/requires
```

```
dc:requires rdf:type owl:ObjectProperty ;
        rdfs:subPropertyOf core:ListProperty ;
    rdfs:domain msp:data ;
    rdfs:range this:Literal ;
    rdfs:label "requires" .

###  http://purl.org/dc/terms/version
dc:version rdf:type owl:ObjectProperty ;
        rdfs:subPropertyOf core:ListProperty ;
    rdfs:domain msp:workflow ;
    rdfs:range rdfs:Literal ;
    rdfs:label "Version no."@en .

###  http://www.w3.org/2004/02/skos/core#prefLabel
skos:prefLabel rdf:type owl:ObjectProperty ;
        rdfs:subPropertyOf core:AutocompleteProperty ;
    rdfs:domain msp:vocabulary ;
    rdfs:range this:Literal ;
    rdfs:label "Vocabulary label" .

###  http://www.w3.org/2006/vcard/ns#email
vcard:email rdf:type owl:ObjectProperty ;
        rdfs:subPropertyOf core:AutocompleteProperty ;
    rdfs:domain msp:workflow ;
    rdfs:range this:Literal ;
    core:order 2 ;
    core:sparqlString   "<http://purl.org/dc/terms/creator>/<http://www.w3.org/2006/vcard/
ns#email>" ;
    rdfs:label "Creator's email" .

###  http://www.w3.org/2006/vcard/ns#hasEmail
vcard:hasEmail rdf:type owl:ObjectProperty ;
        rdfs:subPropertyOf core:AutocompleteProperty ;
    rdfs:domain msp:data ;
    rdfs:range this:Literal ;
    core:order 2 ;
    core:sparqlString   "<http://www.w3.org/ns/dcat#contactPoint>/<http://www.w3.org/2006/
vcard/ns#hasEmail>" ;
        rdfs:label "Contact point email"@en .

###  http://www.w3.org/ns/dcat#theme
dcat:theme rdf:type owl:ObjectProperty ;
        rdfs:subPropertyOf core:AutocompleteProperty ,
        core:ListProperty ;
    rdfs:domain msp:data ;
    rdfs:range this:Literal ;
    core:sparqlString  "<http://www.w3.org/ns/dcat#theme>/<http://www.w3.org/2004/02/skos/
core#prefLabel>" ;
    rdfs:label "Theme"@en .

###  https://www.w3.org/TR/2004/REC-owl-semantics-20040210/#sameAs
<https://www.w3.org/TR/2004/REC-owl-semantics-20040210/#sameAs>                    rdf:type
owl:ObjectProperty ;
               rdfs:subPropertyOf core:AutocompleteProperty ;
        rdfs:domain msp:vocabulary ;
```

```
        rdfs:range msp:vocabulary ;
        rdfs:label "Same as" .

###  this:WFKeyword
<this:WFKeyword> rdf:type owl:ObjectProperty ;
       rdfs:subPropertyOf core:AutocompleteProperty ;
    rdfs:domain msp:workflow ;
    rdfs:range this:Literal ;
    core:sparqlString "<http://www.w3.org/ns/dcat#keyword>" ;
    rdfs:label "Keyword"@en .

###  this:dataContactPointName
<this:dataContactPointName> rdf:type owl:ObjectProperty ;
       rdfs:subPropertyOf core:AutocompleteProperty ;
    rdfs:domain msp:data ;
    rdfs:range this:Literal ;
    core:order 1 ;
    core:sparqlString   "<http://www.w3.org/ns/dcat#contactPoint>/<http://www.w3.org/2006/
vcard/ns#fn>" ;
    rdfs:label "Contact Point Name" .

###  this:dataKeyword
<this:dataKeyword> rdf:type owl:ObjectProperty ;
       rdfs:subPropertyOf core:ListProperty ,
       core:SearchProperty ;
        rdfs:domain msp:data ;
        rdfs:range this:Literal ;
        core:sparqlString "<http://www.w3.org/ns/dcat#keyword>" ;
        rdfs:label "Keyword"@en .

###  this:dataTitle
<this:dataTitle> rdf:type owl:ObjectProperty ;
       rdfs:subPropertyOf core:SearchProperty ;
    rdfs:domain msp:data ;
    rdfs:range this:Literal ;
    core:sparqlString "<http://purl.org/dc/terms/title>" ;
    rdfs:label "Title"@en .

###  this:vocabularyHasPart
<this:vocabularyHasPart> rdf:type owl:ObjectProperty ;
       rdfs:subPropertyOf core:SearchProperty ;
    owl:inverseOf <this:vocabularyIsPartOf> ;
    rdfs:domain msp:vocabulary ;
    rdfs:range msp:vocabulary ;
    core:sparqlString "<http://purl.org/dc/terms/hasPart>" ;
    rdfs:label "Has part" .

###  this:vocabularyIsPartOf
<this:vocabularyIsPartOf> rdf:type owl:ObjectProperty ;
       rdfs:subPropertyOf core:AutocompleteProperty ,
              core:SearchProperty ;
       rdfs:domain msp:vocabulary ;
    rdfs:range msp:vocabulary ;
    core:sparqlString "<http://purl.org/dc/terms/isPartOf>" ;
    core:tooltip "Choose this property for data that might be part of other components" ;
```

```
    rdfs:label "Is part of"@en .

###  this:workflowIsPartOf
<this:workflowIsPartOf> rdf:type owl:ObjectProperty ;
       rdfs:subPropertyOf core:SearchProperty ;
    rdfs:domain msp:workflow ;
    rdfs:range msp:workflow ;
    core:sparqlString "<http://purl.org/dc/terms/isPartOf>" ;
    core:tooltip "Use this property when searching for a Workflow that is used in another
workflow" ;
    rdfs:label "Workflow is part of" .

###  this:worklowCreatorName
<this:worklowCreatorName> rdf:type owl:ObjectProperty ;
       rdfs:subPropertyOf core:AutocompleteProperty ;
    rdfs:domain msp:workflow ;
    rdfs:range this:Literal ;
    core:order 1 ;
    core:sparqlString   "<http://purl.org/dc/terms/creator>/<http://www.w3.org/2006/vcard/
ns#fn>" ;
                          rdfs:label "Creator's name"@en .

#################################################################
#    Classes
#################################################################

###  http://WFRepo/MSP/data
msp:data rdf:type owl:Class ;
         rdfs:subClassOf core:SparnaturalClass ;
         core:order 3 ;
         rdfs:label "Data"@en .


###  http://WFRepo/MSP/vocabulary
msp:vocabulary rdf:type owl:Class ;
               rdfs:subClassOf core:SparnaturalClass ;
               core:order 4 ;
               core:sparqlString "<http://WFRepo/MSP/vocabulary>" ;
               rdfs:label "Vocabulary" .

###  http://WFRepo/MSP/workflow
msp:workflow rdf:type owl:Class ;
             rdfs:subClassOf core:SparnaturalClass ;
             core:order 1 ;
             rdfs:label "Workflow"@en .

###  http://WFRepo/MSP/data/input
mspd:input rdf:type owl:Class ;
           rdfs:subClassOf core:SparnaturalClass ;
           core:order 7 ;
           rdfs:label "Input data"@en .

###  http://WFRepo/MSP/vocabulary/environmentalConditions
mspv:environmentalConditions rdf:type owl:Class ;
                             rdfs:subClassOf core:SparnaturalClass ;
```

```
                              rdfs:label "Environmental conditions"@en .

###  http://WFRepo/MSP/vocabulary/environmentalSuitability
mspv:environmentalSuitability rdf:type owl:Class ;
                              rdfs:subClassOf core:SparnaturalClass ;
                              rdfs:label "Environmental suitability"@en .

###  http://WFRepo/MSP/vocabulary/spatialSuitability
mspv:spatialSuitability rdf:type owl:Class ;
                        rdfs:subClassOf core:SparnaturalClass .

###  http://WFRepo/MSP/vocabulary/suitableAreas
mspv:suitableAreas rdf:type owl:Class ;
                   rdfs:subClassOf core:SparnaturalClass ;
                   rdfs:label "Suitable areas"@en .

###  http://WFRepo/MSP/vocabulary/unsuitableAreas
mspv:unsuitableAreas rdf:type owl:Class ;
                     rdfs:subClassOf core:SparnaturalClass ;
                     rdfs:label "Unsuitable areas"@en .

###  http://WFRepo/MSP/vocabulary/windFarms
mspv:windFarms rdf:type owl:Class ;
               rdfs:subClassOf core:SparnaturalClass .

###  http://WFRepo/MSP/vocabulary/windParks
mspv:windParks rdf:type owl:Class ;
               rdfs:subClassOf core:SparnaturalClass ;
               rdfs:label "Wind Parks"@en .

###  http://WFRepo/MSP/workflow/operations
mspwf:operations rdf:type owl:Class ;
                 rdfs:subClassOf core:SparnaturalClass ;
                 rdfs:label "operations" .

###  http://data.example.com/my-sparnatural-config/Literal
this:Literal rdf:type owl:Class ;
             rdfs:subClassOf rdfs:Literal ;
             core:faIcon "fa-solid fa-pen-to-square" ;
             core:order 2 ;
             core:tooltip "Choose this if the value you are searching for is a string,
such as a name, title, keyword, a (operation's or workflow's) purpose, version numbers,
etc."@en ;
             rdfs:label "Literal value"@en .

###  http://data.example.com/my-sparnatural-config/Literal/Date
<http://data.example.com/my-sparnatural-config/Literal/Date> rdf:type owl:Class ;
                                                             rdfs:subClassOf
rdfs:Literal ;
                                                             core:faIcon  "fa-regular  fa-
calendar-days" ;
                                                             core:order 5 ;
                                                             rdfs:label "Date" .

###  http://www.w3.org/2006/vcard/ns#Individual
```

```
vcard:Individual rdf:type owl:Class ;
                 rdfs:subClassOf core:SparnaturalClass ;
                 core:order 5 ;
                 rdfs:label "Individual"@en .


###  http://www.w3.org/2006/vcard/ns#Organization
vcard:Organization rdf:type owl:Class ;
                   rdfs:subClassOf core:SparnaturalClass ;
                   rdfs:label "organization" .

###  this:attribute
<this:attribute> rdf:type owl:Class ;
       rdfs:subClassOf core:SparnaturalClass ;
       rdfs:label "this attribute" .

###  Generated by the OWL API (version 4.5.25.2023-02-15T19:15:49Z) https://github.com/
owlcs/owlapi
```

Appendix VI – Questionnaire

Welcome to the research study!

This questionnaire is being carried as part as the data collection for the master thesis “Workflow Sharing Based on Semantics: A case study in Marine Spatial Planning”.

This study aims to investigate if the use of web semantics and linked data principles can improve Workflows findability and support provenance tracking of Workflows. For that purpose the questionnaire includes some task to be carried out while you fill in the questionnaire.

This should take no longer than 45 minutes.

For the purpose of this research no personal data is being collected.

Your participation in this research is voluntary. You have the right to withdraw at any point during the study.

The Principal Investigator of this study can be contacted at [Magali Gonçalves| goncalves.m@buas.nl].

By clicking the button below, you acknowledge:

- **Your participation in the study is voluntary.**
- **You are 18 years of age.**
- **You are aware that you may choose to terminate your participation at any time for any reason.**
- **The data collected, without any personal information that could identify you (not linked to you) may be shared with others.**

O I do consent, and I wish to participate

O I do not consent, I do not wish to participate

End of Block: Informed Consent

Start of Block: Demographics

p1a Demographic Information

What is your occupation? _____
What is your background/training? _____

How many years are you working in your field (if applicable)? _____

p1b Demographic Knowledge

	Not familiar at all	Slightly familiar	Moderately familiar	Very familiar	Extremely familiar
How familiar are you with Geodata in general?	O	O	O	O	O
How familiar are you with Workflows in general?	O	O	O	O	O
How familiar are you with web semantics in general?	O	O	O	O	O

End of Block: Demographics

Start of Block: Part II

p2intro

In this section, we would like you to use the website provided to try to answer to the questions below.

file:///C:/Users/1003032/Documents/MASTER/sparnatural/index.html

p2a How many workflows created by someone at Breda University of Applied Sciences (AKA BUas) can you find? _____

question's intention: see if we fulfil req. 1.1 - Retrieve WFs by author or author's affiliation. Also, warming up to how to build the queries.

p2a.d Please indicate the level of difficulty in replying to this question:

	Very difficult	Difficult	Medium	Easy	Very Easy
Difficulty Level	O	O	O	O	O

p2h How many workflows have been created since February 2024? _____

question's intention: **Gives an idea of the repo's update status and *warming up to how to build the queries***

p2h.d Please indicate the level of difficulty in replying to this question:

	Very difficult	Difficult	Medium	Easy	Very Easy
Difficulty Level	O	O	O	O	O

p2f How many Workflows use input data from the HELCOM Secretariat? _____

question's intention: **For data providers, it might be interesting to know what is being done with their data, still *warming up to how to build the queries***

p2f.d Please indicate the level of difficulty in replying to this question:

	Very difficult	Difficult	Medium	Easy	Very Easy
Difficulty Level	O	O	O	O	O

p2g How many Workflows use input data related to renewable energy with data from 2022/01/01 onwards? _____

question's intention: Req.2.2; to see if users can find workflows by data theme using the domain ontology

p2g.d Please indicate the level of difficulty in replying to this question:

	Very difficult	Difficult	Medium	Easy	Very Easy
Difficulty Level	O	O	O	O	O

p2b From those workflows, how many define Spatially Suitable Areas? _____

question's intention: Req. 4.1 see if users can retrieve WFs that originate similar output

p2b.d Please indicate the level of difficulty in replying to this question:<

	Very difficult	Difficult	Medium	Easy	Very Easy
Difficulty Level	O	O	O	O	O

p2c Can you understand the differences between the workflows found in the question above?

_____

(question's intention: Renata use case. Given different but similar results, how do users choose which one to use? Can the user understand the differences between the results, to then decide which one to use? Is the WF description clear enough for transparent decision making? Req. 2.3 and 3.1)

p2c.d Please indicate the level of difficulty in replying to this question:

	Very difficult	Difficult	Medium	Easy	Very Easy
Difficulty Level	O	O	O	O	O

p2e How many Workflows can you find that have input data regarding environmental conditions?

_____

question's intention: Req. 2.2,see if users can find workflows by input data theme by using the domain ontology

p2e.d Please indicate the level of difficulty in replying to this question:

	Very difficult	Difficult	Medium	Easy	Very Easy
Difficulty Level	O	O	O	O	O

End of Block: Part II

Start of Block: Part III

p3a There is a Workflow to support decision-making over the allocation of Marine Protected areas (MPAs) so they can act as a protection network, can you find it?

O Yes

O No

question's intention: requirement 2.1, searching per type of geodata analysis

p3a.d Please indicate the level of difficulty in finding the mentioned workflow:

	Very difficult	Difficult	Medium	Easy	Very Easy
Difficulty Level	O	O	O	O	O

Q45 If you look at the workflow diagram, can you clearly understand the steps of the workflow and the data transformations?

O Yes, it is clear

O With some effort

O No, I have doubts

question's intention: Req. 3.1 WF needs to be understandable by others

Q51 What would you do with this Workflow?

O Nothing

O Take ideas from it and make my own in the GIS I use

O Improve it/ make it more complete

O Run it several times with different scenarios for MPAs' location and evaluate the results

question's intention: Req. 3.1 WF needs to be understandable by others and see if users would adopt the technology behind the approach.

p3b.1 There is more than one workflow to define Environmental suitable areas. Are they somehow related? _____

question's intention: req 3.2 – tracking WF evolution

p3b.1.d Please indicate the level of difficulty in replying to this question:

	Very difficult	Difficult	Medium	Easy	Very Easy
Difficulty Level	O	O	O	O	O

p3b.2 Is there a environmental suitability analysis workflow you would choose over the other? Why?

_____

question's intention: evaluating if the WF is understandable, compare different WFs to see which one is more appropriate to reuse. Req. 3.1, Req. 2.3 understanding operation and parameter differences between WF

p3b.2.d Please indicate the level of difficulty in replying to this question:

	Very difficult	Difficult	Medium	Easy	Very Easy
Difficulty Level	O	O	O	O	O

p3b.4.2 Is the license type of the Workflow clear?

_____

question's intention: Req. 3.4 Licence terms need to be clear.

p3b.4.1 What is the license type of the input data?

_____

question's intention: Req. 3.4 Licence terms need to be clear.

p3b.3 Is the metadata provided enough for you to make a decision on whether to (re)use one of the workflows in the repository?

O Yes

O No

question's intention: Evaluate if the metadata provided is complete.

p3b.3.If not, what are you missing?

_____

question's intention: get information that might have been missing

p3b.5 Can you see if there are any Workflows that were already used to create other Workflows?

_____

question's intention: Req. 4 . 2 Combine WFs as sub-workflows in a bigger WF. Can users understand when a WF was reused in another WF or when they are looking at a WF that has a "sub-workflow"?

End of Block: Part III

Start of Block: Part IV

p4intro In this section, we would like you to reflect on what you saw.

p4a How difficult do you think this would be...

	Very difficult	Difficult	Medium	Easy	Very Easy
adapting the workflow to reuse it with different datasets	O	O	O	O	O
adapting the workflow to reuse it for a different purpose	O	O	O	O	O
combining two or more workflows into a bigger workflow?	O	O	O	O	O

question's intention: Infer reusability and technology acceptance

p4b Do you currently use any workflow management system?

O Yes

O No

question's intention: Get more insights on potential users

p4b.1 Which one?_____

question's intention: Get more insights on potential users

p4c Do you currently use any workflow repositories?

O Yes

O No

question's intention: Get more insights on potential users

p4c.1 Which one?_____

question's intention: Get more insights on potential users

p4d Do you think you would use a repository like this for your workflows?

- O Yes
- O Maybe
- O No

question's intention: Get more insights on technology acceptance of potential users

p4d.2 For which purpose would you use such a workflow repository:

- ☐ To search for workflows
- ☐ To share my workflows
- ☐ To keep track of my workflow versions

question's intention: Get more insights on potential users

p4e Do you have any suggestions for a geodata workflow repository?

_____

question's intention: Get more insights on what more could be useful for a geodata WF repository

p4f Did you miss something in the prototype?

_____

question's intention: Get more insights on what more could be useful for a geodata WF repository

p4g Do you have ideas on how to improve the workflow description or workflow repository?

_____

question's intention: Get more insights on what more could be useful for a geodata WF repository

End of Block: Part IV

Appendix VII - Questionnaire Results

Question number	Question	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Participant 6	Participant 7	Participant 8
p1a_1	Demographic Information - What is your occupation?	student	postdoc researcher	Researcher	Students	student	Student	R&D Project Leader	Interactive Media Designer
p1a_2	Demographic Information - What is your background/training?	International media and entertainment	spatial planning, msp, coastal governance, GIS	International Spatial Development	Built Environment (Urban Design)	built environment	Built Environment	serious/simulation game R&D	Interactive Media Design
p1a_3	Demographic Information - How many years are you working in your field (if applicable)?	-99	1	1	6 months	-99	5 month	17	8
p1b_1	Demographic Knowledge - How familiar are you with Geodata in general?	Slightly familiar	Very familiar	Moderately familiar	Slightly familiar	Slightly familiar	Moderately familiar	Very familiar	Very familiar
p1b_2	Demographic Knowledge - How familiar are you with Workflows in general?	Extremely familiar	Moderately familiar	Not familiar at all	Very familiar	Slightly familiar	Not familiar at all	Very familiar	Very familiar
p1b_3	Demographic Knowledge - How familiar are you with web semantics in general?	Not familiar at all	Slightly familiar	Not familiar at all	Not familiar at all	Not familiar at all	Not familiar at all	Slightly familiar	Very familiar
p2a	How many workflows created by someone at Breda University of Applied Sciences (AKA BUas) can you find?	6	6	6	6	6	6	6	6
p2a.d_1	Please indicate the level of difficulty in replying to this question	Easy	Difficult	Medium	Medium	Medium	Difficult	Easy	Medium

Question number	Question	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Participant 6	Participant 7	Participant 8
p2h	How many workflows have been created since February 2024?	4	4	4	4	4	4	4	4
p2h.d_1	Please indicate the level of difficulty in replying to this question	Very Easy	Very Easy	Easy	Easy	Very Easy	Medium	Very Easy	Easy
p2f	How many Workflows use input data from the HELCOM Secretariat?	1	1	1	1	1	1	1	1
p2f.d_1	Please indicate the level of difficulty in replying to this question	Medium	Easy	Difficult	Difficult	Easy	Difficult	Easy	Easy
p2g	How many Workflows use input data related to renewable energy with data from 2022/01/01 onwards?	2	2	2	2	2	2	2	2
p2g.d_1	Please indicate the level of difficulty in replying to this question	Medium	Easy	Very difficult	Very difficult	Medium	Very difficult	Medium	Medium
p2b	From those workflows, how many define Spatially Suitable Areas?	2	2	2	2	2	2	2	2
p2b.d_1	Please indicate the level of difficulty in replying to this questionl	Very Easy	Very Easy	Very difficult	Easy	Difficult	Medium	Medium	Difficult

Question number	Question	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Participant 6	Participant 7	Participant 8
p2c	Can you understand the differences between the workflows found in the question above?	No.	the version control	The version	Yes, different versions	Yes. Changed safety buffer distance around shipping lanes to 500m instead of 1000m.	with the help I managed to find the description of the changes between the versions	Changed safety buffer distance around shipping lanes to 500m instead of 1000m.	They are the First version, and the Updated version of the same workflow. Second one (V1.1) has the appropriate metadata in the version control.
p2c.d_1	Please indicate the level of difficulty in replying to this question	Very Easy	Difficult	Easy	Medium	Medium	Medium	Very Easy	Easy
p2e	How many Workflows can you find that have input data regarding environmental conditions?	2	3	3	3	3	3	3	3
p2e.d_1	Please indicate the level of difficulty in replying to this questionl	Very Easy	Difficult	Very difficult	Medium	Very difficult	Easy	Medium	Easy
p3a	There is a Workflow to support decision-making over the allocation of Marine Protected areas (MPAs) so they can act as a protection network ,can you find it?	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes

Question number	Question	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Participant 6	Participant 7	Participant 8
p3a.d_1	Please indicate the level of difficulty in finding the mentioned workflow	Very Easy	Easy	Easy	Very difficult	Easy	Easy	Medium	Medium
Q45	If you look at the workflow diagram, can you clearly understand the steps of the workflow and the data transformations?	Yes, it is clear	Yes, it is clear	With some effort	Yes, it is clear	With some effort	With some effort	With some effort	With some effort
Q51	What would you do with this Workflow?	Nothing	Take ideas from it and make my own in the GIS I use	Run it several times with different scenarios for MPAs' location and evaluate the results	Improve it/ make it more complete	Improve it/ make it more complete	Take ideas from it and make my own in the GIS I use	Run it several times with different scenarios for MPAs' location and evaluate the results	Take ideas from it and make my own in the GIS I use
p3b.1	There is more than one workflow to define Environmental suitable areas. Are they somehow related?	Now that I know where to check I can see that one is the newer version of the same workflow.	Yes. Versions of each other	Yes they are related	Yes	Yes. Slight alteration of first version, using depth values from -40m towards shore.	I could find them	yes, one is an improvement/alteration of the other	Yes, its the same workflow but the next version.
p3b.1.d_1	Please indicate the level of difficulty in replying to this questionl	Very Easy	Very difficult	Very difficult	Medium	Easy	Difficult	Easy	Very Easy
p3b.2	Is there a environmental suitability analysis workflow you would choose over the other? Why?	I would choose the newer version, because it is more up-to-date.	They are similar enough. I would probably choose the first version.	The newer version, becasue ther is slight alteration of first version, using depth values from -40m towards shore.	Yes, the latest version. Most accurate yet.	I do not know	I would use the newer version	it depends on the kind of wind turbine pilon you're choosing to work with. Difference is -60m or -40m.	Normally the latest version, but it depends on my usecase, do i need the -60m or the -40m.

Question number	Question	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Participant 6	Participant 7	Participant 8
p3b.2.d_1	Please indicate the level of difficulty in replying to this question	Very Easy	Medium	Very difficult	Easy	Very difficult	Easy	Easy	Medium
p3b.4.2	Is the license type of the Workflow clear?	Yes.	Yes.	Yes, its free to use if you are not charging others for it	Relatively. Once you click on the link it becomes clear, but not at first glance.	Yes	yes	yes, GPL3	Yes
p3b.4.1	What is the license type of the input data?	GPL 3	https://www.gnu.org/licenses/gpl-3.0.txt	GNU	gpl 3.0	GNU General Public License	gpl-3.0	also probably GPL3	Cant find from the available data sources
p3b.3	Is the metadata provided enough for you to make a decision on whether to (re)use one of the workflows in the repository?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
p3b.3.no	If not, what are you missing?	NA	NA	NA	NA	NA	NA	NA	License information, the info on the actual workflow is sufficient
p3b.5	Can you see if there are any Workflows that were already used to create other Workflows?	Yes, by looking at the legend.	Yes. I found one - data preparation w/f	Yes there were	Yes	Yes	yes	yes 1	Yes, WindFarmEnvironmentalSuitability
p4a_1	How difficult do you think this would be... - adapting the workflow to reuse it with different datasets	Very difficult	Easy	Medium	Medium	Easy	Very Easy	Easy	Easy

Question number	Question	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Participant 6	Participant 7	Participant 8
p4a_2	How difficult do you think this would be... - adapting the workflow to reuse it for a different purpose	Medium	Easy	Easy	Easy	Easy	Medium	Medium	Easy
p4a_3	How difficult do you think this would be... - combining two or more workflows into a bigger workflow?	Very Easy	Medium	Difficult	Very Easy	Difficult	Difficult	Medium	Medium
p4b	Do you currently use any workflow management system?	No	No	No	No	No	No	No	No
p4b.1	Which one?	NA	NA	NA	NA	NA	NA	NA	NA
p4c	Do you currently use any workflow repositories?	No	No	No	No	No	No	No	Yes
p4c.1	Which one?	NA	NA	NA	NA	NA	NA	NA	GitHub
p4d	Do you think you would use a repository like this for your workflows?	No	Yes	Maybe	Yes	Maybe	Yes	Maybe	Yes
p4d.2_1	For which purpose would you use such a workflow repository: - To search for workflows	To search for workflows	To search for workflows	To search for workflows	To search for workflows	To search for workflows	To search for workflows	To search for workflows	To search for workflows
p4d.2_2	For which purpose would you use such a workflow repository: - To share my workflows	0	To share my workflows	0	To share my workflows	0	0	0	To share my workflows

Question number	Question	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Participant 6	Participant 7	Participant 8
p4d.2_3	For which purpose would you use such a workflow repository: - To keep track of my workflow versions	0	0	To keep track of my workflow versions	To keep track of my workflow versions	0	0	0	To keep track of my workflow versions
p4e	Do you have any suggestions for a geodata workflow repository?	-99	No	-99	Compatibility with software, customizability.	-99	-99	User feedback functionalities: allowing workflow users to provide feedback, showing results of that. Tutorial material (videos? manual?). All other things I can think of are a bit too fundamental.	Mentioned in recording
p4f	Did you miss something in the prototype?	-99	No	-99	Tooltips on hover	-99	-99	License of the input data used in the workflow. Would be great if the system can feed back somehow if a certain term in the vocabulary is actually a 'parent' or a 'child' or both.	Mentioned in recording

p4g	Do you have ideas on how to improve the workflow description or workflow repository?	-99	No	Colours and gradients of the tags in Sparnatural	More user friendly interface/search options. Something more familiar?	-99	-99	Only the aforementioned, I think they all apply.	Mentioned in recording
Notes: 0 - option not selected; -99 no answer									

Appendix VIII - Tools Used in the Research

During the preparation of this work, the author(s) used Microsoft Word as a text processor and editor to write the thesis with the Grammarly plug-in and sometimes recurring to ChatGPT to edit text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the work.

Mendeley, with the Windows plug-in, was used to manage the references of previous relevant work. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the work.

All diagrams were built using draw.io online (https://app.diagrams.net/).

During the preparation of this work, the author(s) used Protégé to develop the ontology, rdfEditor to describe the WFs in TTL, GraphDB to store the TTL files and Sparnatural to let testers query the triple store.

Qualtrics was used for the questionnaire and data collection.