GIMA
Geographical Information Management and Applications

Introducing knowledge footprints:
Enhancing knowledge graph data retrieval with visualisations for improved geoscience collaboration

MSc Thesis

Mark van Vliet
22 May 2024

Universiteit Utrecht
TU Delft
University of Twente.
Wageningen UR

Author	Mark van Vliet
Student number	1961845
Date	22 May 2024
Programme	Geographical Information Management and Applications (GIMA) Faculty of Geosciences Utrecht University (Delft University of Technology, University of Twente, Wageningen University & Research)
Supervision	Rob Lemmens, University of Twente Stanislav Ronzhin, University of Twente Sven Casteleyn, Universitat Jaume I, Castellón, Spain
Responsible professor	Prof. Menno-Jan Kraak, University of Twente

Acknowledgements

The thesis is a product I really enjoyed working on and combines the three technologies, graph data, ontologies, and JavaScript (D3.js in particular), I wanted to learn more of in the last semester of the Geographical Information Management and Applications (GIMA) master's programme. Now some acknowledgements.

First and foremost, I would like to thank Rob Lemmens for guiding me through the process of writing an MSc thesis and introducing me to various knowledgeable people.

I would also like to thank Stan Ronzhin for his knowledge on linked data and knowledge graphs. Besides that, I want to thank him for setting up a GraphDB server that I could use for this research.

In addition, I would like to thank Yuri von Engelhardt for his valuable knowledge about visualisations. I was surprised how much thought can go in creating a good visualisation.

I would also like to thank Sven Casteleyn and Carlos Granell for their valuable insights and letting me stay at the Universitat Jaume I in Castellón, Spain. A stay I really enjoyed and letting me get a taste of studying outside of The Netherlands.

In addition, I would like to extend my thanks to Upeksha Vidanelage for letting me use results from her research on using NLP methods for annotating geo-related scientific publications with EO4GEO BoK concepts. Without those annotations I would not be able to create knowledge footprints.

At last, I want to thank all the participants in the evaluation of knowledge footprints. I appreciate the time and effort these people put in. They provided me with valuable insights and helped me in a crucial part of good research.

Abstract
The EO4GEO Body of Knowledge (BoK) serves as a hierarchical structure of describing concepts in the domain of earth observation and geoinformation. Due to this nature the EO4GEO BoK provides an interesting data source for tagging and visualising organisational and personal knowledge. The hypothesis suggests that by publicly disclosing and simplifying access to information about an entity's specific expertise or knowledge, it becomes easier to identify experts, thereby improving collaboration within a particular field.

This study presents how the already existing relational nature of describing BoK concepts is transformed to a RDF graph dataset following an ontology for describing bodies of knowledge. In addition, this graph dataset is then semi-automatically enriched with expertise annotations, which are created by extracting EO4GEO concepts from research papers by natural language processing tools. These expertise annotations and the hierarchical and relational structure of the graph dataset creates the EO4GEO knowledge graph. With this knowledge graph, it becomes possible to ask questions that suggest which individual or organisation has specific expertise on a topic. These topics match the concepts in the EO4GEO body of knowledge.

This study further shows how visualisations can be leveraged and how they provide extra context in knowledge graph data retrieval through question and answering through visualisations. These visualisations are called knowledge footprints. They are created for the purpose of representing, promoting and retrieving someone's expertise. This study introduces a user-evaluated website that combines the EO4GEO knowledge graph, SPARQL, JavaScript and the D3.js library to interactively create these knowledge footprints. In addition, this website introduces potential applications for knowledge footprints.

The evaluation suggests that knowledge footprints do offer an interesting approach to visualise knowledge in our geospatial domain. While the evaluators see value in using knowledge footprints to identify collaborators with specific expertise, knowledge footprints likely won't entirely replace traditional search methods. Collaboration often involves certain human factors like reputation and familiarity that go beyond domain knowledge.

Keywords: Knowledge graphs, Ontology, Body of Knowledge, Knowledge footprints, Knowledge visualisation, Geoscience collaboration

Table of contents

List of tables

List of figures

List of listings

List of abbreviations

Abbreviation	Full form
KG	Knowledge graph
KA	Knowledge area
BoK	Body of Knowledge
RDF	Resource Description Framework
RDFS	RDF Schema
OWL	Web Ontology Language

1. Introduction

In January 2018 the EO4GEO alliance, existing of 25 partners, started the EO4GEO project. EO4GEO had the vision to create the European EO/GI workforce of the future with the right skills and knowledge to tackle the challenges the workforce faces today. To bridge the skill gap between the supply and demand of geospatial education and training in the EO/GI sector, EO4GEO created a Body of Knowledge (BoK) for the earth observation and geospatial domain. This BoK describes inner-related concepts within the EO/GI domain. These concepts are a combination of knowledge, skills, and competencies which are used to match occupational profiles, job offers and curricula by annotating them with these BoK concepts. The project ended in June 2022 and delivered the EO4GEO BoK and a variety of applications build upon the EO4GEO BoK.

The EO4GEO BoK provides an interesting data source for further applications. This thesis investigates how the EO4GEO BoK can be further utilised and create a new application that effectively links experts within the EO/GI domain to EO4GEO BoK concepts and make these connections visual. By creating these connections, expertise can be mapped which offers a new data source that can potentially be leveraged to improve geoscience collaboration. By providing people and organisations with the knowledge where specific expertise in the EO/GI domain is present, collaboration can be more targeted, which makes it more efficient and effective.

This thesis aims to achieve that by transforming the already existing relational nature and hierarchical structure of describing concepts in the EO4GEO BoK (Lemmens, Albrecht, et al., 2022), to a fully ontology-based knowledge graph for the geospatial information (GI) domain by using RDF triples. The usage of the RDF data model enhances the utility and integration capabilities of the EO4GEO BoK and helps build new applications. How this is done will be shown in this thesis.

1.1 Problem and its context

Incorporating semantic web principles, open standards and making use of ontologies increase the interoperability of the EO4GEO BoK. Improved interoperability allows for linking other data sources and potentially other bodies of knowledge related to the GI domain in the future. Besides that, it also solves a concern raised by Toppen & Reinhardt in 2009, on the completeness of the GIS&T BoK, a predecessor of the EO4GEO BoK (Toppen & Reinhardt, 2009). This concern is likely true for the EO4GEO BoK as well. This concern is understandable in a domain where technology is rapidly evolving (Hofer et al., 2020; Stelmaszczuk-Górska et al., 2020) and where different concepts gain or lose relevance based on technological development. Improved interoperability and the ability to link to other BoKs could reduce the workload of keeping BoKs up to date, thereby reducing and sharing the responsibility of ensuring a BoKs completeness, actuality, and relevance.

History of the EO4GEO BoK

The EO4GEO BoK origins from the original Geographic Information Science and Technology (GIS&T) BoK developed by the University Consortium for Geographic Information Science (UCGIS) in 2006 (Hofer et al., 2020), its main purpose was to inform GIS curricula and contribute to professional development in the US. However, by 2009, Toppen & Reinhardt recognized that due to differences in policies and regulations between the EU and US it was necessary to make an EU specific GIS&T BoK (Toppen & Reinhardt, 2009). From these ideas the GIS&T BoK got forked, serving as a starting point to develop an EU specific BoK known as the GI-N2K BoK. Worth mentioning is that the architecture of the GI-N2K BoK platform was based on linked data principles, made use of triple stores and ontologies (Vandenbroucke & Vancauwenberghe, 2016). Later during the EO4GEO project the GI-N2K BoK underwent further elaboration, incorporating the domain of earth observation, to become what is now known as the EO4GEO BoK (Hofer et al., 2020). During this transition, due to using a different platform (Living textbook) then GI-N2K, the use of linked data formats was discontinued. In the years following its creation, the EO4GEO BoK was expanded to include a more business-oriented approach (Hofer et al., 2020) and later integrated artificial intelligence concepts (Lemmens, Lang, et al., 2022).

EO4GEO BoK applications
The EO4GEO platform is all about knowledge management and knowledge transfer in a standardised way, it helps geospatial organisations and professionals answer questions like, what do I need to learn to do x, which tools do I need to do y, where do I find the necessarily resources and which people do I need to partner with or should be part of my team? All is done through standardised concept descriptions. The latter two are currently not implemented in the EO4GEO platform but will be explored in this thesis. Creating these new insights aims to improve geoscience collaboration.

Knowledge graphs, based on semantic web principles and ontologies, have the potential to integrate with other knowledge graphs. Part of this thesis is to explore whether individual knowledge graphs that capture the expertise of researchers can be linked to or annotated with the EO4GEO knowledge graph. When these knowledge graphs can be successfully linked, it becomes possible to visualise individual and organisational knowledge in so called "knowledge footprints" based on concepts in the EO4GEO BoK. The last step in this research is exploring what further applications these visualised knowledge footprints have. An example is footprint matching, wherein knowledge gaps are made visual or areas for collaboration through performing overlays are identified.

The AGILE (Association of Geographic Information Laboratories in Europe) full paper publications (Volume 2 - 4) (AGILE, 2021, 2022, 2023) will be the source of research papers to determine experts in this thesis. The classification of these research papers to EO4GEO BoK concepts will not be part of this thesis but is undertaken by a student of Universitat Jaume I (UJI) in Castellón, Spain. Her results will be used for further processing in this research.

1.2 Research objectives

The objective of this MSc research is to develop and evaluate a knowledge mapping and visualisation framework based on semantic web principles to support geoscience collaboration by effectively connecting experts to domain knowledge within the EO4GEO body of knowledge. To this main objective belongs the following main research question:

"Can a framework that uses knowledge graphs for knowledge mapping and visualization effectively represent and compare knowledge footprints of geoscience experts based on the EO4GEO Body of Knowledge?"

To answer this question, the main objective is broken down in the following three subobjectives with accompanying sub-research questions.

1. **Develop a standard format for describing the semantics of a BoK and apply this standard to the EO4GEO BoK.**
    a. How is the EO4GEO BoK semantically defined and how is it annotated?
    b. What relationships and properties should be included in a general ontology for bodies of knowledge?
    c. How can the existing EO4GEO body of knowledge be converted to graph data following this standard format?
2. **Develop a method that allows connecting individual expertise to the EO4GEO body of knowledge**
    a. How to make use of ontology relationships to connect knowledge graphs?
    b. How can expert's knowledge be connected to domain knowledge within the EO4GEO BoK?
    c. How can the EO4GEO knowledge graphs be used to query someone's expertise?

3. **Develop a user evaluated framework that visualises graph data-based knowledge footprints to support geoscience collaboration.**
    a. How can creating visual knowledge footprints based on the EO4GEO BoK be automated?
    b. Which use cases become possible when individual and organisational knowledge graphs can be visualised?
    c. What are potential further applications for knowledge footprints?

1.3 Research scope

- This research will involve creating an upper ontology for bodies of knowledge, which potentially enables BoK linking in the future. This research however will not go into linking the EO4GEO BoK to other BoKs.
- Only the latest version of the EO4GEO BoK will be used in this thesis and converted to the EO4GEO Knowledge graph.
- The research involves automatic JavaScript based visualisation creation, based on the to be created EO4GEO knowledge graph. But does not involve designing a good-looking website to show these visualisations.
- Nicely visualised and interactive visualisations is the ambition, but deemed a nice to have when developing this is too difficult. The creative process and explaining which visualisation in which case is best, how to deal with various parameters that influence visualisations, which data goes where and what data is used is the scope alongside the evaluation process with AGILE users.

1.4 Expected significance of the research

Upon the successful completion of this research, EO4GEO will have the ability to effectively annotate the EO4GEO knowledge graph with individual expertise, leading to the ability to create visual knowledge footprints. This ability is expected to significantly contribute to collaboration within the geoscience community. With these annotations, people can identify domain experts in a particular GEO related field. These insights help inform individuals or organisations with whom they need to collaborate with, in order to reach one's personal or organisational goals. Furthermore, by linking people to their respective organisations, organisational knowledge footprints can be created, providing a holistic view of the expertise of an organisation.

Another opportunity lays in footprint matching, a potential application of knowledge footprints. With footprint matching, gaps or overlaps in knowledge between entities can be identified. This is not restricted to matching knowledge footprints between individuals, but also envision the knowledge you need to fulfil a specific GI related role visualised in a knowledge footprint. Matching these two footprints can help individuals identify knowledge areas wherein someone needs to improve. Another opportunity is creating visual knowledge footprints from educational and research organisations. These can show in which GEO related field the organisation has expertise and could assist students help choose which organisation they want to apply too when they know which GEO field they want to study.

Besides this main objective, this research enables the first steps in achieving the ambition to establish connections between the EO4GEO BoK and other related BoKs in the future. This ambition potentially reduces the workload of keeping the EO4GEO BoK up to date and relevant, by outsourcing topics that are more comprehensively covered in another domain's BoK. By allowing linking of BoKs, more knowledge can be covered and potentially, with higher quality. An ontology that describes bodies of knowledge in general can enable this goal since it provides a common and machine-readable framework to structure and organize knowledge in a consistent manner. When BoKs adhere to this common framework, BoKs become interoperable with each other.

1.5 Overview of chapters

The second chapter "The foundations of knowledge graphs" describes the key elements and technologies the reader should have a basic understanding of, to understand the results of this thesis. The third chapter "A methodological framework for constructing the EO4GEO knowledge graph and its applications" describes the methods in detail how the different research objectives will be met. The fourth chapter "Towards the EO4GEO knowledge graph" are the results of subobjective one and two and explains how the EO4GEO knowledge graph will be constructed, and which decisions are made in this process. The fifth chapter "Knowledge footprints" goes into detail what becomes possible once the EO4GEO knowledge graph is constructed, in this chapter the created individual and organisational knowledge footprints and footprint matching will be introduced. Knowledge footprints are the result of subobjective three. The sixth chapter "Evaluating knowledge footprints" goes into detail how knowledge footprints are evaluated and what the results of these evaluations are. Chapter 7 contains the discussion and Chapter 8 the conclusion.

2. The foundations of knowledge graphs

This research makes use of fundamental technologies related to the semantic web. These technologies are further explained below to get an adequate understanding of these technologies. These technologies form the basis for the applications created in this research.

2.1 The technologies that define the semantic web

The semantic web started with a vision of the inventor of the World Wide Web Tim Berners-Lee. He envisioned an extension to the world wide web wherein resources are semantically described, given more meaning and contain links connecting resources to each other, making data on the world wide web machine readable (Matthews, 2005). Over the years the semantic web community has proposed several standards and best practices to realise this vision (Hogan, 2020). Some of these standards are relevant to this research, these include RDF, RDFS, OWL and SPARQL and will be explained below.

The Resource Description Framework (RDF) can be seen as the data model for describing resources on the semantic web (Hogan, 2020; Ryen et al., 2022). RDF is a standard for data interchange. This data model follows a graph like structure of nodes and edges, representing resources on the web and their relation. This way of describing resources is represented in triples (Bizer, 2011). Which makes it quite different to the common tabular structure of relational databases which requires a predefined data schema. Triples consist of three parts, a subject, predicate and object. Figure 1 gives an example of a triple (also known as a statement or RDF graph).

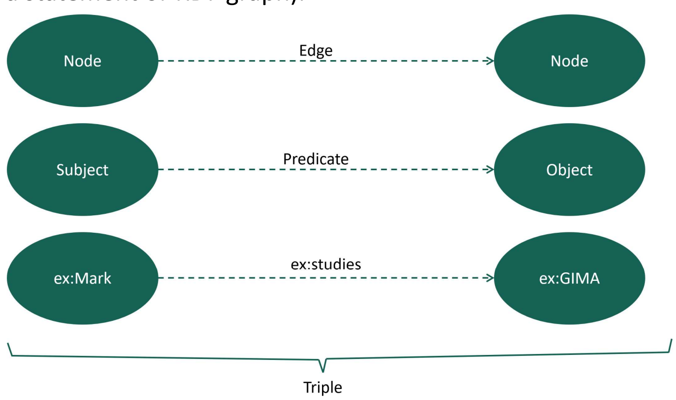


Figure 1: Example of a triple following the RDF language (Source: Author)

RDF has three kinds of nodes 1) resource nodes, which represent resources on the web through a URI (Unique Resource Identifier), 2) blank nodes, these represent resources for which a URI is not given and 3) literal nodes, which describe specific (attribute) values (Dong, 2023; Hogan, 2020). RDF statements can be serialized into different formats to improve readability (Po et al., 2020). Listing 1 shows the examples, Turtle (W3C, 2014) and JSON-LD (W3C, 2020), of these serializations. The displayed namespaces will be explained in Section 4.1.2.

```
eo4geo:IP3-5-3 a obok:Concept ;
    rdfs:label "[IP3-5-3] Local variance" ;
    obok:conceptStatus "New" ;
    obok:hasRecommendedMaterial <https://doi.org/10.1007/
    978-3-540-37258-5_66> ;
    obok:hasSkill eo4geo:bcdb158e-dc4e-45f6-8b97-0379607f8ad8 ;
    obok:isSubconceptOf eo4geo:IP3-5 ;
    dcterms:description "Local variance can be calculated as the value of
    standard deviation in a small neighborhood (e.g. 3x 3 moving window),
    then computing the mean of these values over the entire image. The
    obtained value is an indicator of the local variability in the image." ;
    rdfs:isDefinedBy eo4geo: ;
    skos:notation "IP3-5-3" .
```

```
{
  "@id": "eo4geo:IP3-5-3",
  "@type": "obok:Concept",
  "dcterms:description": "Local variance can be calculated as the value
  of standard deviation in a small neighborhood (e.g. 3x 3 moving
  window), then computing the mean of these values over the entire image.
  The obtained value is an indicator of the local variability in the
  image.",
  "obok:conceptStatus": "New",
  "obok:hasRecommendedMaterial": {
    "@id": "https://doi.org/10.1007/978-3-540-37258-5_66"
  },
  "obok:hasSkill": {
    "@id": "eo4geo:bcdb158e-dc4e-45f6-8b97-0379607f8ad8"
  },
  "obok:isSubconceptOf": {
    "@id": "eo4geo:IP3-5"
  },
  "rdfs:isDefinedBy": {
    "@id": "eo4geo:"
  },
  "rdfs:label": "[IP3-5-3] Local variance",
  "skos:notation": "IP3-5-3"
},
```

Listing 1: Examples of Turtle (left) and JSON-LD (right) serializations (Source: Author)

In the above sections RDF is introduced and how data can be stored in an RDF data model. However, the ability to store metadata or metainformation about RDF graphs is not well supported in RDF (Carroll et al., 2005). To solve this, named graphs are introduced. Named graphs extend RDF graphs by providing a name in the form of a URI to a set of RDF triples. In other words, it labels RDF graphs with metadata (Watkins & Nicole, 2006). Named graphs have benefits in keeping RDF statements with different purposes or different provenance separate from each other (Carroll et al., 2005; Watkins & Nicole, 2006). This approach enables separating big RDF graphs into smaller collections, or subgraphs. Which furthermore have benefits for access control, version management and query performance - by being able to query subsets of a bigger RDF graph (Shinavier, 2009). Named graphs can be serialized in a couple of formats, one of these is the TriG format which is used in this research. TriG (W3C, 2024) is an extension to the Turtle RDF format. Listing 2 shows samples of the two named graphs serialized in the TriG format, eo4geo:concepts and eo4geo:applications.

```
eo4geo:concepts {
    eo4geo:IP3-5-3 a obok:Concept ;
        rdfs:label "[IP3-5-3] Local variance" ;
        obok:conceptStatus "New" ;
        obok:hasRecommendedMaterial <https://doi.org/10.1007/978-3-540-37258-5_66> ;
        obok:hasSkill eo4geo:2a0ba19a-c8d1-4b81-ae44-9bc213848bdb ;
        obok:isSubconceptOf eo4geo:IP3-5 ;
        dcterms:description "Local variance can be calculated as the value of standard deviation in a small neighborhood (e.g. 3x 3 moving window),
        then computing the mean of these values over the entire image. The obtained value is an indicator of the local variability in the image." ;
        rdfs:isDefinedBy eo4geo: ;
        skos:notation "IP3-5-3" .
}
eo4geo:applications {
    eo4geo:ff6c6e18-983c-43e0-9587-e900df31cb02 a boka:Expert ;
        rdfs:label "Martin Tomko" ;
        boka:authorOf <https://doi.org/10.5194/agile-giss-4-2-2023> ;
        boka:hasKnowledgeOf eo4geo:WB3-6,
            eo4geo:WB4-3 ;
        org:memberOf eo4geo:fd67b88a-97b8-4f8f-947a-c235aa648936 ;
        foaf:name "Martin Tomko" .
}
```

Listing 2: Example of the TriG serialization. (Source: Author)

Naturally humans and machines want to be able to discover, access and query datasets to receive answers from specific questions. This is a foundational requirement for any application (Buil-Aranda et al., 2013). Relational databases know the Structured Query Language (SQL). However, SQL is not well suited for graph data or RDF datasets, since it requires a predefined schema to query over. RDF or graph datasets have a flexible data schema making SQL not usable. To solve this issue and being able to query RDF datasets the SPARQL Protocol and RDF Query Language (SPARQL) was developed. SPARQL works by giving conditions that match triple patterns (DuCharme, 2013).

RDF Schema (RDFS) is an extension to the RDF language, it provides RDF data with extra understanding and context to make it machine readable. For example, the triple ex:Mark ex:studies ex:GIMA from Figure 1 is fully understandable for humans, but not for machines, it has no context and it won't know how to place ex:Mark and ex:GIMA in context. Figure 2 provides the above triple extra semantics by defining the node's class through RDFS.

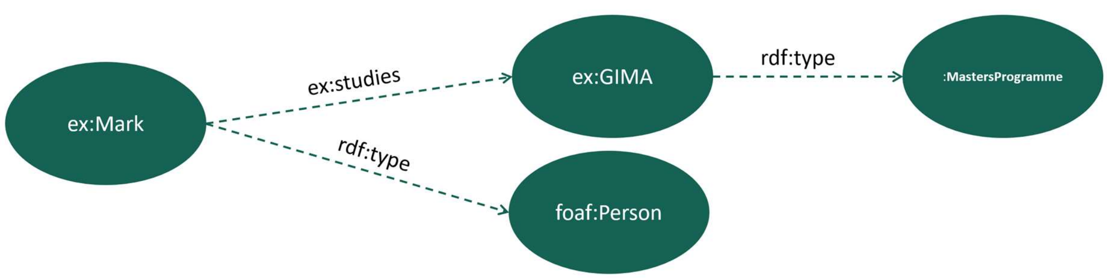


Figure 2: Example of how RDFS provide meaning to RDF data (Source: Author)

Besides that RDFS provide properties to define basic constraints, like the domain and range of a class or property, how classes relate to each other and provide a minimal set of basic terms for annotations (Hogan, 2020; W3C, 2023). While RDFS is relatively lightweight, the OWL language goes a step further to provide richer semantics. OWL is introduced in the next section.

2.2 Ontologies and ontology development

Feilmayr & Wöß (2016) defines an ontology as a formal, explicit specification of a shared conceptualization that is characterized by high semantic expressiveness required for increased complexity (Feilmayr & Wöß, 2016). Ontologies provide the semantics of a specific domain for both humans and machines and allow semantic modelling of knowledge (Ehrlinger & Wöß, 2016). They can be seen as the blueprint for a knowledge graph. Ontologies are often written in a combination of RDF, RDFS and the Web Ontology Language (OWL), which are all specifications for expressive knowledge representation. RDF and RDFS form the building blocks of an ontology, OWL specific constructions can be added to make the ontology more expressive.

Through providing semantic descriptions to classes and properties in an ontology, machines can reason over data in the graph and create new knowledge. This is called reasoning. Reasoning algorithms only work when an ontology explains the terms and relations between terms in a dataset in machine readable format, such as RDF, RDFS and OWL. Once these relations are understood a machine starts to process and draw conclusions based on these relations (Hogan, 2020). However, there is a trade-off to consider, the more expressive an ontology is, the more time reasoning algorithms take to derive new conclusions and finish reasoning (Hogan, 2020; Kang et al., 2020). The expressive OWL language further knows three sublanguages to help regulate the expressiveness of an ontology: OWL-Lite, OWL-DL and OWL-Full (Saha, 2007).

Researchers often view the EO4GEO BoK as the ontology for the EO/GI domain (Dubois et al., 2021; Hofer et al., 2020; Stelmaszczuk-Górska et al., 2020). This is indeed true as the EO4GEO BoK represents concepts, relationships and terminology specific to the GI domain. However, in this thesis an upper ontology (or foundational -/top level ontology) will be created that represents bodies of knowledge in general. Upper ontologies characterize themselves as domain-independent, focuses on high-level concepts and general information that are the same across all domains (Elmhadhbi et al., 2019; Schneider, 2003). This shows that there are different abstraction levels for ontology design. Figure 3 visualizes these abstraction levels well. The idea is that the EO4GEO BoK, as a domain ontology, falls under and adheres to this upper-level structure defined in the upper ontology created in this thesis. The key benefit of using this structure is that it enables easier integration with other bodies of knowledge in the future (Elmhadhbi et al., 2019; Mascardi et al., 2007) since it follows the same structure for describing bodies of knowledge. The created ontology in this thesis is expected to use RDF, RDFS and OWL.

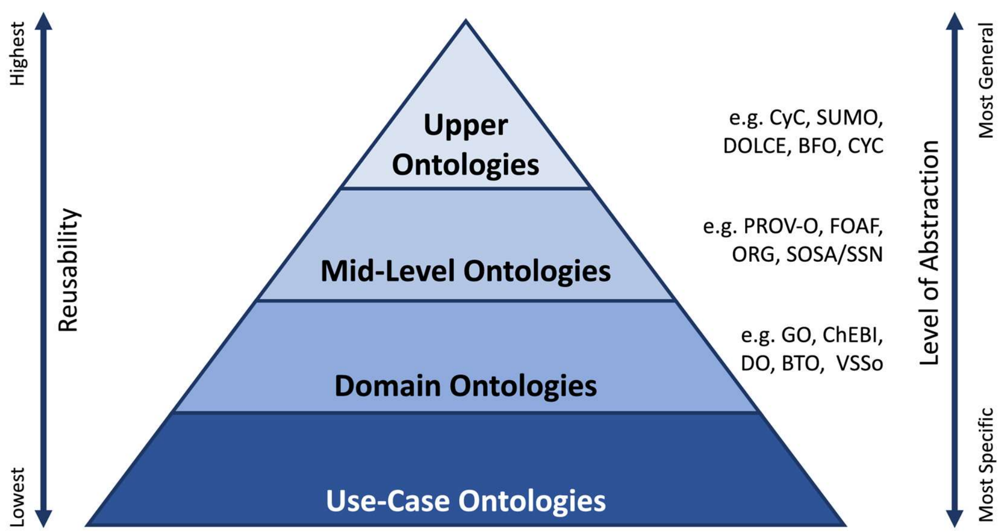


Figure 3: Abstraction levels for ontology design, the EO4GEO BoK positions itself as a domain ontology (Copied from: (Haller & Polleres, 2020))

2.3 Understanding knowledge graphs

One of the applications of the semantic web are knowledge graphs. Knowledge graphs (KGs) are structured representations of facts, consisting of entities, relationships, and semantic descriptions (Ji et al., 2022) and are represented in a machine-readable format (Abu-Salih, 2021). Nowadays, knowledge graphs are used for many possibilities, like supporting web search and question and answering systems like voice assistants, but also for product recommendations, biomedical research, or enterprise data integration and management solutions (Chaudhri et al., 2022; Dong, 2023). A common knowledge graph is the Google knowledge graph, which enriches Google's search capabilities through reasoning and inference (Kejriwal, 2019). Ehrlinger & Wöß (2016) further uses this reasoning aspect and integrates it in their definition for a knowledge graph: "A knowledge graph acquires and integrates information according to an ontology, which can be referred as the schema of the knowledge graph, and utilizes inference and reasoning to derive new knowledge" (Ehrlinger & Wöß, 2016; The Alan Turing Institute, 2020). Figure 4 shows their idea of an architecture of a knowledge graph. Enforcing that data instances follow ontologies as the schema, ensures clean semantics and structured data. By describing the meaning of entity classes in ontologies, this approach ensures that knowledge in the knowledge graph is understandable for machines, and understandable for humans through its visual graph structure (Dong, 2023). Knowledge graphs can be enriched with data through a combination of human-driven, semiautomated and/or fully automated methods (Chaudhri et al., 2022).

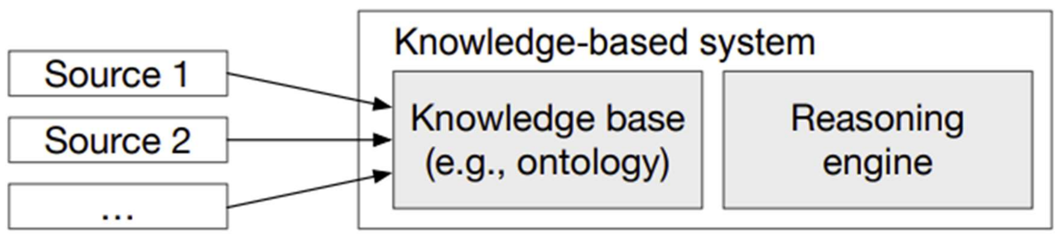


Figure 4: Architecture of a knowledge graph (Copied from: (Ehrlinger & Wöß, 2016))

3. A methodological framework for constructing the EO4GEO knowledge graph and its applications

This chapter covers the research methodology for each sub objective. Together the different subobjectives contribute to reaching the main research objective, which is to develop a framework that uses graph data to visualise knowledge footprints of geoscience experts based on the EO4GEO BoK (Section 1.2). Figure 5 shows a general overview of where each subobjective falls within reaching the main research objective. The first subobjective creates an ontology for bodies of knowledge and turns the EO4GEO BoK into the EO4GEO named graph using the concept of ontology-based data integration. This ontology provides a standard data model for content in the EO4GEO BoK. The second subobjective focusses on how individual expertise can be linked to the EO4GEO BoK, to do this an ontology for BoK applications will be created. Combining individual expertise and this ontology creates the EO4GEO applications named graph. In this step the EO4GEO concepts get enriched with expertise annotations which together form the EO4GEO knowledge graph. Combining subobjective one and two shows that the EO4GEO knowledge graph is enriched with data through human driven (Creating the EO4GEO BoK) and semi-automated methods (NLP processing of research papers). The last subobjective delves into how information in the EO4GEO knowledge graph can be used to generate knowledge footprints and later how these can be compared. Figure 5 graphically shows the general overview of the research methodology. The next sections go into more detail about the methodology for each subobjective.

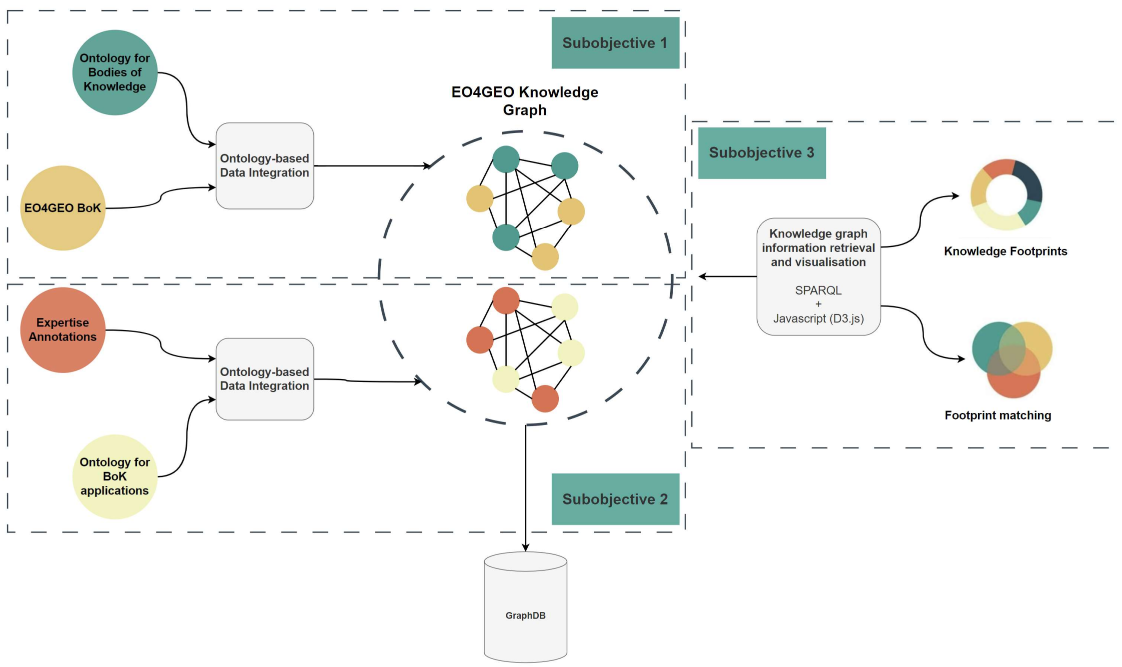


Figure 5: General overview of research methodology (Source: Author)

3.1 Creating a standard format for the EO4GEO BoK named graph

The first subobjective focuses on developing an ontology for bodies of knowledge in general. The outcome of this objective is a key part in the creation of the ontology based EO4GEO knowledge graph. This approach is accompanied by qualitative research design. Figure 6 shows the steps that need to be taken to reach the following desired results, an ontology that describes the semantics of a body of knowledge and next to that this step enriches the EO4GEO named graph with BoK content.

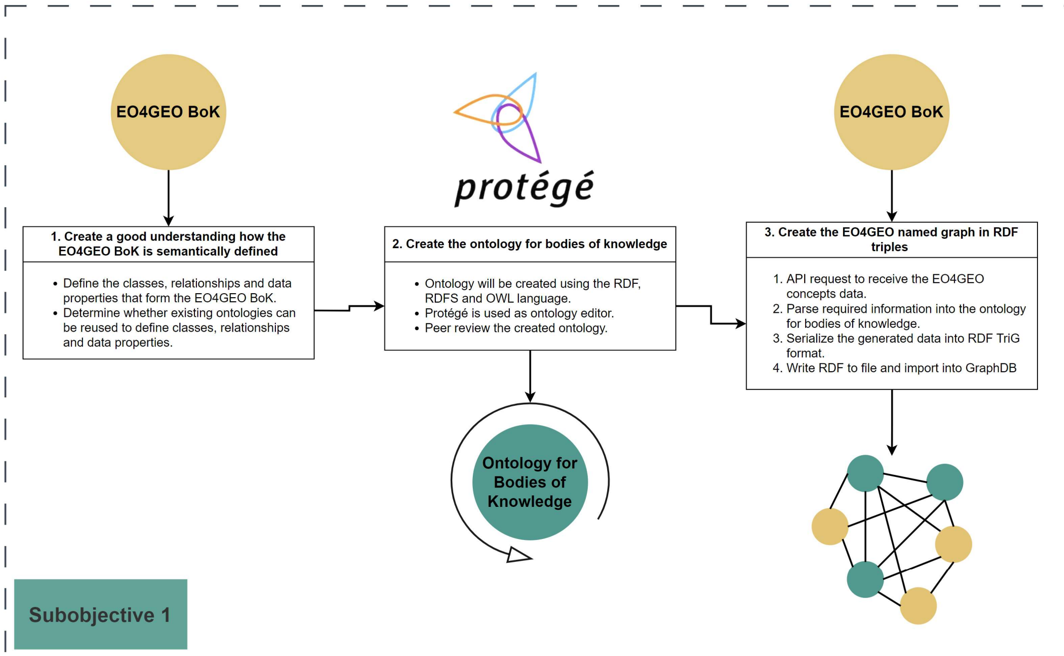


Figure 6: Detailed overview of the methodology for subobjective 1 (Source: Author)

The first step is about generating a good understanding on how the EO4GEO BoK is currently semantically defined. The semantics of the EO4GEO BoK will form the basis for the ontology for bodies of knowledge. During this process the main ontology classes, relationships and data properties are defined. Part of ontology development is to reuse existing ontologies as much as possible. This has benefits in interoperability and ensures that things are defined following a commonly agreed-upon understanding (Simperl, 2009). Classes or relationships that can not be defined using existing ontologies will become part of the ontology for bodies of knowledge (OBOK).

The second step is to use the knowledge from step one to create the ontology for bodies of knowledge. This ontology is created using the protégé application and outputs ontologies in the RDF, RDFS and OWL language.

The last step transforms the data within the EO4GEO BoK and maps it into the data format defined in the ontology for bodies of knowledge. This process is called ontology-based data integration. This step is done via python, and uses the JSON, requests and RDFLib library. RDFLib is a key library in this process, since it allows to parse data into RDF triples and gives options to further serialise the RDF data in the TriG format. The requests and JSON libraries are used to communicate with and correctly use data from the EO4GEO API. Once the triples are generated the data will be imported into GraphDB.

GraphDB is chosen since it allows for storing RDF data and allows for querying RDF data through SPARQL over the http protocol (RDF4J API). Another strong factor was that I am familiar with GraphDB, and it has a free license. Other triple stores like Virtuoso or Stardog are also suitable. A common other graph database, Neo4j, is not suitable since it makes use of property graphs instead of RDF graphs.

3.2 Incorporating individual expertise in the EO4GEO knowledge graph

The second subobjective focusses on developing a method that allows individual knowledge graphs to be linked to the EO4GEO knowledge graph. The outcome of this part is crucial since it enables the creation of individual and organisation footprints based on the EO4GEO knowledge graph. This approach is accompanied by qualitative research design. Figure 7 shows a detailed overview of the steps that need to be taken to incorporate individual expertise in the EO4GEO knowledge graph and how to retrieve the necessarily information for creating knowledge footprints.

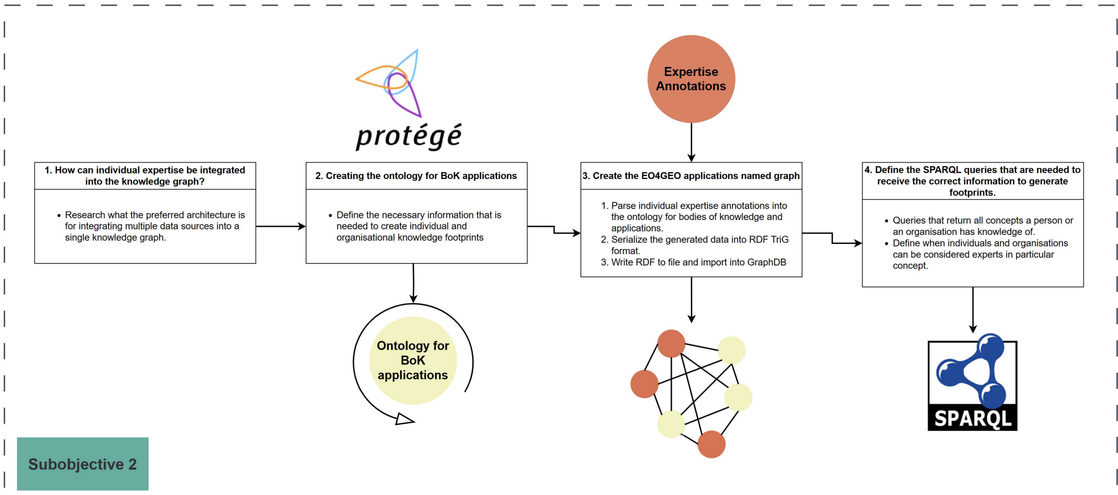


Figure 7: Detailed overview of the methodology for subobjective 2 (Source: Author)

The first step is researching what the preferred architecture is for integrating multiple different data sources into a single knowledge graph. This is important since the knowledge graph consists of multiple data sources with each a different objective.

The second step is creating an ontology for BoK applications. In this case all the semantics that are needed to create individual and organisational footprints. This ontology is created using the protégé application and outputs ontologies in the RDF, RDFS and OWL language.

The third step is to generate the EO4GEO applications named graph. This step is similar to how the EO4GEO named graph from the first subobjective is generated, and thereby follows the same procedure.

The last step is to develop queries that request the required data, which are needed for the creation of individual and organisational footprints. These queries are important since they need to return all the information to generate knowledge footprints in the next subobjective.

3.3 Generation and applications of knowledge footprints from the EO4GEO knowledge graph

The third objective focusses on receiving user feedback and creating the knowledge graph driven knowledge footprints. This is accompanied by a mixed-methods research design. Developing and creating visualisations is qualitative and evaluating whether these footprints are effective will be done quantitatively. Figure 8 shows a detailed overview of the steps that need to be taken to automatically generate knowledge footprints.

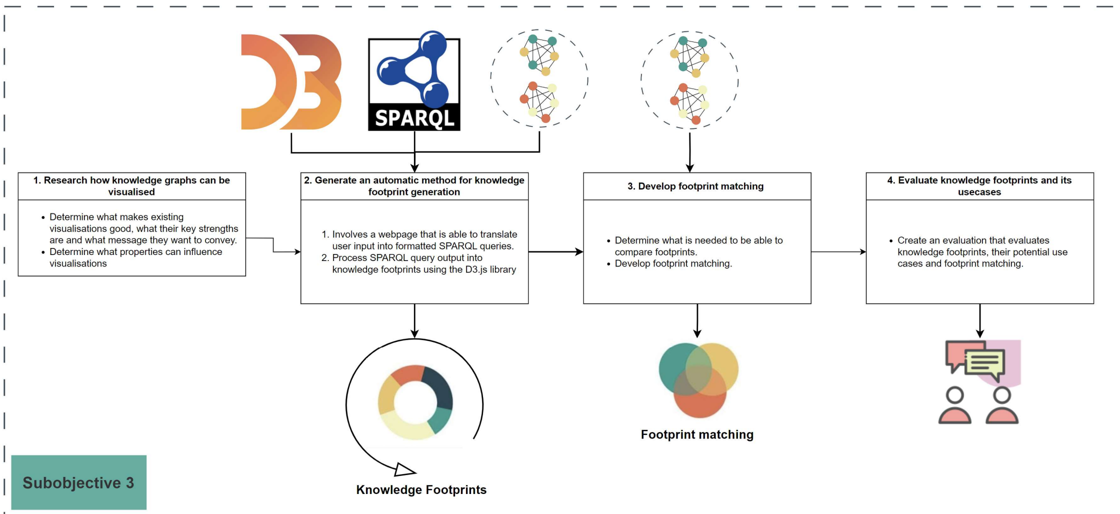


Figure 8: Detailed overview of the methodology for subobjective 3 (Source: Author)

The first step is researching how existing knowledge graphs are visualised, this is done to determine what makes existing visualisations good, what their strengths are and what message each visualisation wants to convey. Likely every visualisation answers a different question. Part of this step is also defining properties that influence knowledge footprints, ex. size, colour etc. to correctly translate those in good visualisations.

The second step is developing a JavaScript driven website that can send SPARQL queries to the EO4GEO knowledge graph and further parses the SPARQL query results into knowledge footprints using the D3.js library. D3.js is chosen due to its popularity and extensive visualisation methods and interactive possibilities. This website gives the user a form to fill in, this form determines what kind of knowledge footprint should be generated and generates this footprint after receiving the necessary information. With this approach the website offers a question and answer-based system, ensuring the end user will not have to write SPARQL queries themselves.

The third step investigates whether knowledge footprints can be visually compared and what potential use cases this application has. This process will be called footprint matching.

The last step is to determine the value of these knowledge footprints. This evaluation will be done interactively by creating a couple of tasks with accompanying questions. Each task will create a different personalised knowledge footprint. People that have published papers to AGILE in the past will be contacted and asked to assist in this evaluation.

4. Towards the EO4GEO knowledge graph

This chapter represents the results related to how the EO4GEO knowledge graph is constructed from the content in the EO4GEO BoK and represent the processes described for subobjective one and two in Section 3.1 and 3.2.

4.1 The EO4GEO BoK named graph

The first step in constructing the EO4GEO knowledge graph is to create an RDF graph from the content in the EO4GEO BoK. This named graph describes all content original to the EO4GEO BoK. The following sections describe the thought process towards creating the EO4GEO named graph. The reason why there is chosen for a named graph architecture will be described in Section 4.3.1.

4.1.1 How is the EO4GEO BoK semantically defined?

The EO4GEO BoK includes fourteen top level knowledge areas about the EO/GI domain described in a machine and human readable format. Each knowledge area holds theories, methods, technologies and applications described in various concepts. Concepts under each knowledge area follow a hierarchical structure to describe concepts in multiple granular levels. Wherein subconcepts of a concept describe the superconcept on a narrower level. Besides these sub and super concept relations, the EO4GEO BoK also holds the “pre-requisite of” and “is similar to” constructs to describe incoming and outgoing relationships between concepts (Dubois et al., 2021; Lemmens, Albrecht, et al., 2022). Figure 9 shows the fourteen knowledge areas differentiated by colour in a so-called zoomable circle packing chart.

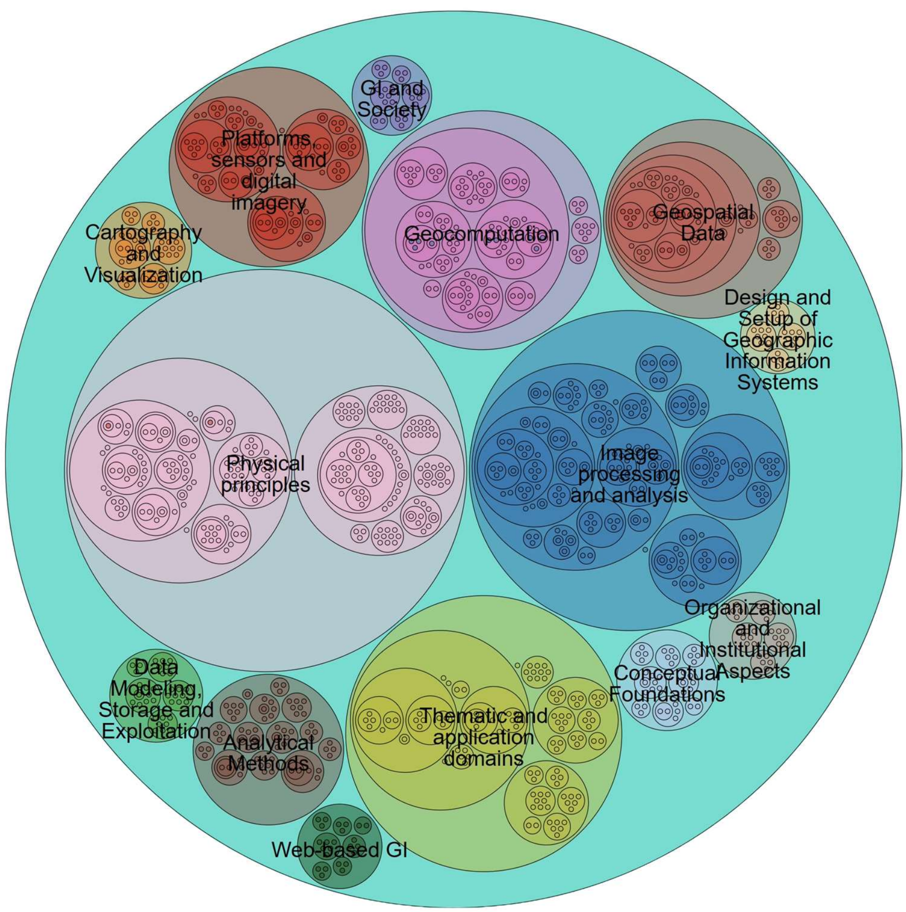


Figure 9: Overview of the fourteen knowledge areas in the EO4GEO BoK. (Adopted from: https://bok.eo4geo.eu/)

The EO4GEO BoK further annotates each concept with a unique identifier, the name of the concept, a focused description, some key literature references, the person who contributed to this concept, associated skills and a concept status. Listing 3 shows a JSON sample from data that can be extracted from the EO4GEO BoK using the EO4GEO API. This figure shows what data is stored with each concept in a structured format. What this JSON sample also shows is that some fields, e.g. the description value of a contributor to describe the organization this person works for and using the name field of a reference to store the ISBN and publisher information is semantically questionable, which could be better defined by using either a more extensive data schema or adopt a flexible schema like RDF.


```json
"AM10-1": {
  "contributors": [
    {
      "description": "University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC)",
      "name": "Rob Lemmens",
      "url": "https://research.utwente.nl/en/persons/rob-lemmens"
    }
  ],
  "id": "AM10-1",
  "name": "Problems of large spatial databases",
  "references": [
    {
      "description": " ",
      "name": "Karimi, H.A. (ed.) (2014) Big Data Techniques and Technologies in Geoinformatics. ISBN 9781138073197 Published April 20, 2017 by CRC Press 312 Pages.",
      "url": "https://www.routledge.com/Big-Data-Techniques-and-Technologies-in-Geoinformatics/Karimi/p/book/9781138073197"
    }
  ],
  "relations": [
    { "name": "is subconcept of", "source": "AM10-1", "target": "AM10" }
  ],
  "skills": [
    "Describe difficulties in dealing with large spatial databases, especially those arising from spatial heterogeneity",
    "Describe emerging geographical analysis techniques in geocomputation derived from artificial intelligence e.g., expert systems, artificial neural networks, genetic algorithms, and software agents",
    "Explain how to recognize contaminated data in large datasets",
    "Explain what is meant by the term contaminated data, suggesting how it can arise",
    "Outline the implications of complexity for the application of statistical ideas in geography"
  ],
  "uri": "http://bok.eo4geo.eu/AM10-1",
  "code": "AM10-1",
  "description": "Difficulties in dealing with large spatial databases, especially those arising from spatial heterogeneity and data quality issues.",
  "selfAssesment": "<p>In progress (GI-N2K)</p>"
},
```


Listing 3: Generated JSON structure from the AM10-1 concept from various sources of the EO4GEO API. (Source: Author)

4.1.2 An ontology for bodies of knowledge (obok)

Like mentioned in Section 2.3 knowledge graphs benefit from ontologies as they provide knowledge graphs with a clear semantic framework, giving machines context to the data that is present in a knowledge graph. Ontologies further enable reasoning and inference, examples will be explained in Section 4.3.1. Because of these benefits an upper ontology for bodies of knowledge (obok) in general is created. This ontology is designed in a way that it gives an RDF based structure for describing the various elements in the EO4GEO BoK. Figure 10 shows a simplified visual representation of this ontology. Appendix A shows this ontology in the OWL language and serialised in the Turtle format.

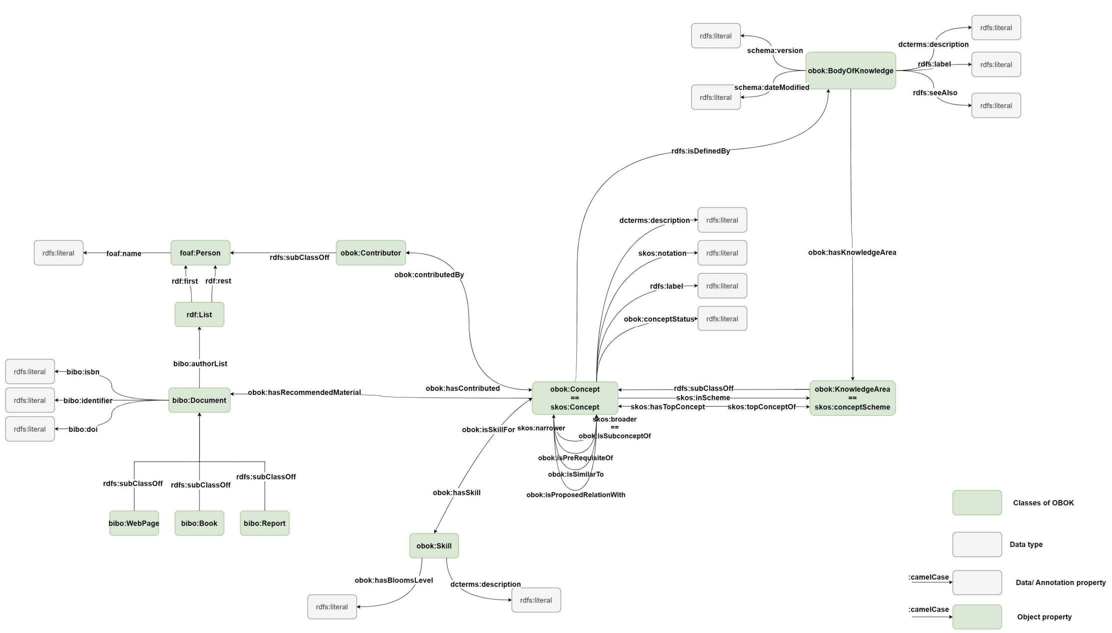


Figure 10: A simplified visual representation of the ontology for bodies of knowledge (obok) (Source: Author, draw.io)

A strong ideology in ontology design is to make use of existing ontologies as much as possible to leverage commonly accepted standards, ensure interoperability with other ontology-based systems and reduce redundancy. Because of this the following existing ontologies are used to define classes, relationships and properties in the obok ontology. Specific constructs used by these ontologies can be seen in Figure 10, but will not be explained in detail in this research as they are broadly explained on the web.

- **skos:** SKOS is used for its common data model for knowledge organisation systems. SKOS sees knowledge organisation systems as concept schemes containing multiple concepts (W3C, 2009b). Knowledge areas in the EO4GEO BoK will be classified as concept schemes and concepts in the BoK as SKOS concepts. Categorising each knowledge area as a concept scheme is chosen due to seeing them as standalone areas within the GIS domain. Modelling them this way allows for a more specific structure and offers the ability to make domain specific SPARQL queries.
- **bibo:** The bibliographic ontology is used to describe documents in RDF. The EO4GEO BoK makes references to interesting material accompanying a specific concept. The BIBO ontology is used to make this part of the BoK, RDF.
- **foaf:** The Friend of A Friend ontology is used to semantically describe persons who contributed to the BoK or are related to concepts through other relationships. The OBOK ontology extends the foaf:Person class with the obok:Contributor subclass.
- **dcterms:** The Dublin Core Metadata Initiative terms ontology (dcterms) is in the OBOK ontology often used to write descriptions related to multiple classes in the ontology.
- **schema:** The Schema.org ontology is used to describe administrative properties in the OBOK ontology. Mainly for version management purposes.

Not every construct in the EO4GEO BoK could be mapped using existing ontologies, to solve that constructs unique to the obok ontology are created. These are the following:

- **obok:BodyOfKnowledge:** An owl class that is used to represent the provenance of constructs in a BoK.
- **obok:KnowledgeArea:** An owl class that represents knowledge areas in a BoK. This class is equivalent to skos:ConceptScheme.
- **obok:Concept:** An owl class that represents various concepts in a BoK. This class is equivalent to skos:Concept.
- **obok:Contributor:** An owl class that represents the person who contributed to a specific concept in a BoK. This class is a subclass of foaf:Person, and thereby inherits all constructs associated with the foaf:Person class.
- **obok:Skill:** An owl class that represents skills or learning outcomes associated with a concept in a BoK.
- Semantic relations between obok classes:
    - **obok:hasKnowledgeArea:** indicates the knowledge areas a BoK holds.
    - **obok:isSubconceptOf:** Indicates that a concept has a lower granularity level then the related concept. This class is equivalent to skos:broader. Skos:broader should be read as “has broader concept” (W3C, 2009a).
    - **obok:isPreRequisiteOf:** Indicates that a concept needs to be known to understand the other.
    - **obok:isSimilarTo:** Indicates that a concept is similar to the other.
    - **obok:isProposedRelationWith:** A temporarily more administrative relationship between concepts.
    - **obok:contritibutedBy**: Links the person who contributed to the concept that person has contributed to.
    - **obok:hasContributed:** Inverse of obok:contributedBy.

- **obok:hasRecommendedMaterial:** Links an obok:Concept to bibo:Document. This relation represents documents that can be used for further reading material about a specific concept.

- Relations that describe data properties:
  - **obok:hasBloomsLevel:** Indicates the bloom's level[1] that accompanies a specific skill.
  - **obok:conceptStatus:** Indicates the status of a concept. This is an administrative property for BoK management.

4.1.3 The EO4GEO BoK RDF graph

Now having an ontology that describes a clear semantic framework for the EO4GEO BoK, the EO4GEO BoK can be transformed to RDF. This process is done via python, the accompanying script can be accessed in Appendix C. Figure 11 shows an RDF graph representation from a set of triples made in this process, multiple instances from most classes are removed for clarity. But what can be seen is that the constructs present in the ontology (Figure 10) can also be seen in this figure. Appendix E shows the same triples visualised in the below RDF graph in TriG format. The below figure does not show data properties in this visualisation method but are present in the raw RDF triples in Appendix E.

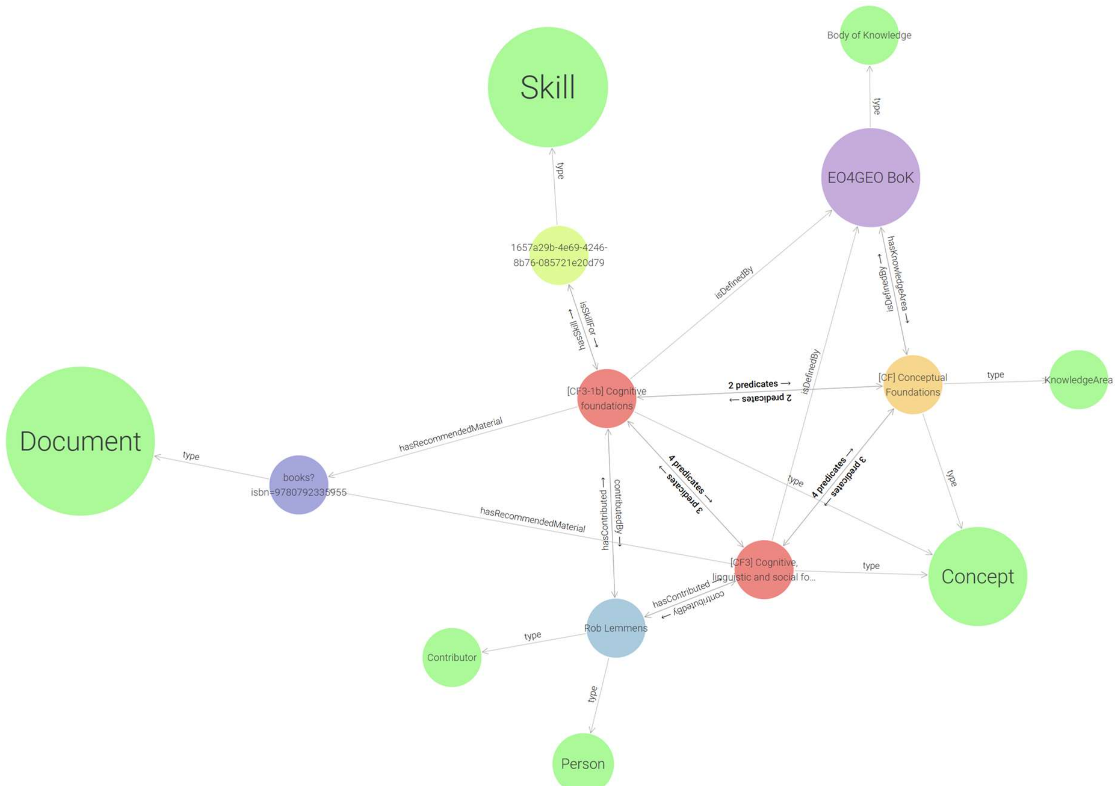


Figure 11: Visual RDF graph representation of RDF triples following the obok ontology applied to the EO4GEO BoK. (Source: Author, visualisation from GraphDB)

[1] The Bloom's taxonomy provides six levels of increasing cognitive skill described through specific verbs. (Bloom's Taxonomy, 2020)

4.2 The EO4GEO BoK applications named graph

The second step in constructing the EO4GEO knowledge graph is to create an RDF graph for data that is used by applications build upon the data in EO4GEO BoK. This named graph describes all content that currently does not exist in the EO4GEO BoK. The following subsections describe the thought process towards creating the EO4GEO applications named graph.

4.2.1 Integrating individual expertise into the EO4GEO knowledge graph

To be able to integrate individual expertise into the EO4GEO knowledge graph it needs to be able to be linked to concepts or content from the EO4GEO BoK. Listing 4 shows the JSON output from Upeksha's thesis (Vidanelage, [Forthcoming]) for every research paper or PDF that is processed through natural language processing algorithms. This process scans each research paper, extracts the title, author and organisation and annotates each research paper with related EO4GEO BoK concepts.

```json
{
  "DOI": "https://doi.org/10.5194/agile-giss-4-18-2023",
  "Title": "Predicting Pedestrian Counts using Machine Learning Molly Asher , Yannick Oswald , and Nick Malleson School of Geography , University of Leeds , UK Correspondence : Nick Malleson ( n.s.malleson @ leeds.ac.uk )",
  "Concepts": [
    "Discovery over linked open data",
    "Open data",
    "Machine learning",
    "Approaches to point, line, and area generalization",
    "Publishing linked open data",
    "Decision trees",
    "Time",
    "Information-as-data-interpretation"
  ]
}
```

Listing 4: Natural language processing output per research paper. (Source: (Vidanelage, [Forthcoming]))

This JSON is then further processed into the following JSON data structure (Listing 5), so it can be better transformed into RDF and thereby become part of the EO4GEO knowledge graph. This process further extracts the authors and organisation the author is a member of. This processing is done via python that connects with the ChatGPT API. This was deemed efficient but will be briefly reflected upon in the discussion. Appendix D shows this script.

```json
[
  {
    "doi": "https://doi.org/10.5194/agile-giss-4-18-2023",
    "authors": ["Molly Asher", "Yannick Oswald", "Nick Malleson"],
    "organisations": ["School of Geography, University of Leeds, UK"],
    "concepts": [
      "Discovery over linked open data",
      "Open data",
      "Machine learning",
      "Approaches to point, line, and area generalization",
      "Publishing linked open data",
      "Decision trees",
      "Time",
      "Information-as-data-interpretation"
    ]
  }
]
```

Listing 5: Further segmented JSON structure. (Source: Author)

4.2.2 An ontology for BoK applications

To realise a successful integration from individual expertise data with content from the EO4GEO BoK, an ontology is developed that extends the OBOK ontology with constructs specifically made to model persons with expertise and organisations with expertise. This ontology is called the ontology for BoK applications (BOKA). Figure 12 shows a simplified visual representation of this ontology. Appendix B shows this ontology in the OWL language, serialised in the Turtle format.

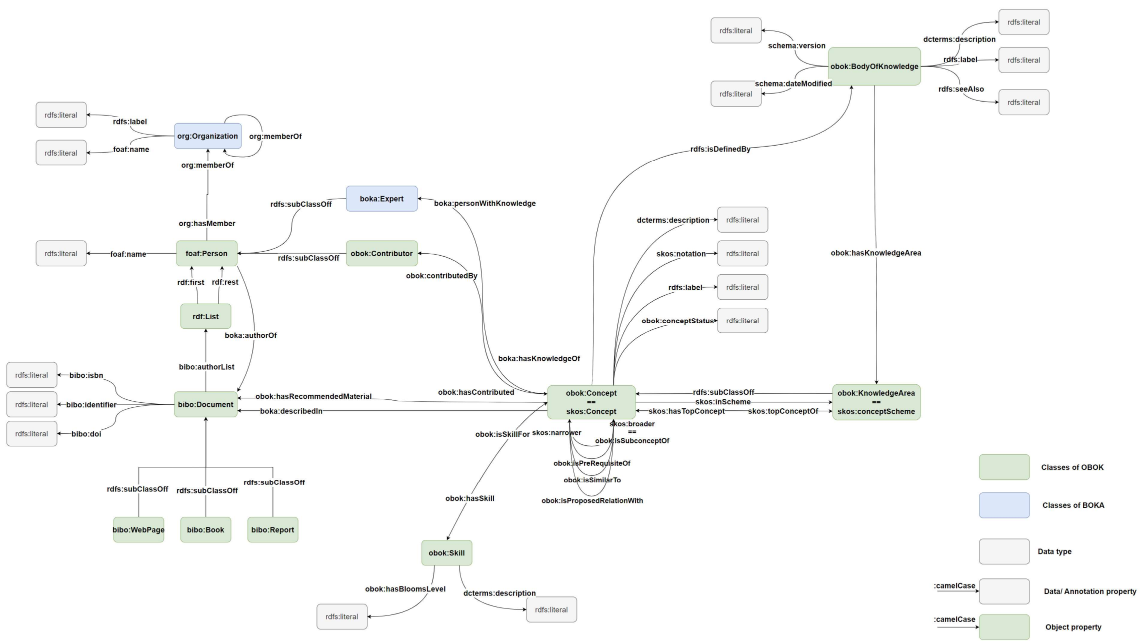


Figure 12: A visual representation of where the ontology for BoK applications (boka) falls within the ontology for bodies of knowledge (obok) (Source: Author, draw.io)

Looking at Figure 12 a couple of new constructs can be seen; these are explained below:

- **boka:Expert:** An owl class that is used to represent persons who have shown to hold expertise of have knowledge of a particular concept in the EO4GEO BoK. boka:Expert is a subclass of the foaf:Person owl class and thereby inherits all data properties and object properties from this class.
- **org:Organisation:** An owl class that is used to represent organisations. This owl class follows the class defined in the organisation ontology.
- **org:hasMember:** A construct that indicates that a foaf:Person is a member off a org:Organisation.
- **org:memberOf:** A construct that is the inverse of org:hasMember, but can also indicate that an organisation is a member of another organisation.
- **boka:authorOf:** A construct linking a bibo:Document to the author in this research this is scoped to the boka:Expert class
- **boka:describedIn:** A construct that links an bibo:Document to the obok:Concept. In these documents concept content is described.

4.2.3 The EO4GEO applications RDF graph

Figure 13 shows a similar visualisation of what has been shown in Section 4.1.3. The process to create this RDF graph is the same as creating the EO4GEO BoK RDF graph, however what can be seen is how the ontology for BoK applications integrates into the ontology and data from the EO4GEO BoK. In this figure a couple of newly created nodes are bounded in a red box, which makes visible what the BOKA ontology adds to the existing data structure defined through the OBOK ontology. An organisation class is added and a person who is considered an expert is defined. It also shows the new relations boka:hasKnowledgeOf and boka:personWithKnowledge. Besides that, documents that are used for determining which concept an author has knowledge of are linked to the obok:Concept class but through the boka:describedIn construct. This construct differs with the obok:hasRecommendedMaterial construct, as the preliminary indicates that this document is used for determining experts and indicates that content from the linked BoK concept is described in this document. The obok:hasRecommendedMaterial is particularly chosen by the concept contributor and indicates good further reading material. Appendix F shows the RDF graph in the figure below in TriG format, note that the TriG format clearly separates the triples specific to the EO4GEO BoK named graph and the EO4GEO BoK applications named graph and that the latter extends the triples in the EO4GEO BoK named graph.

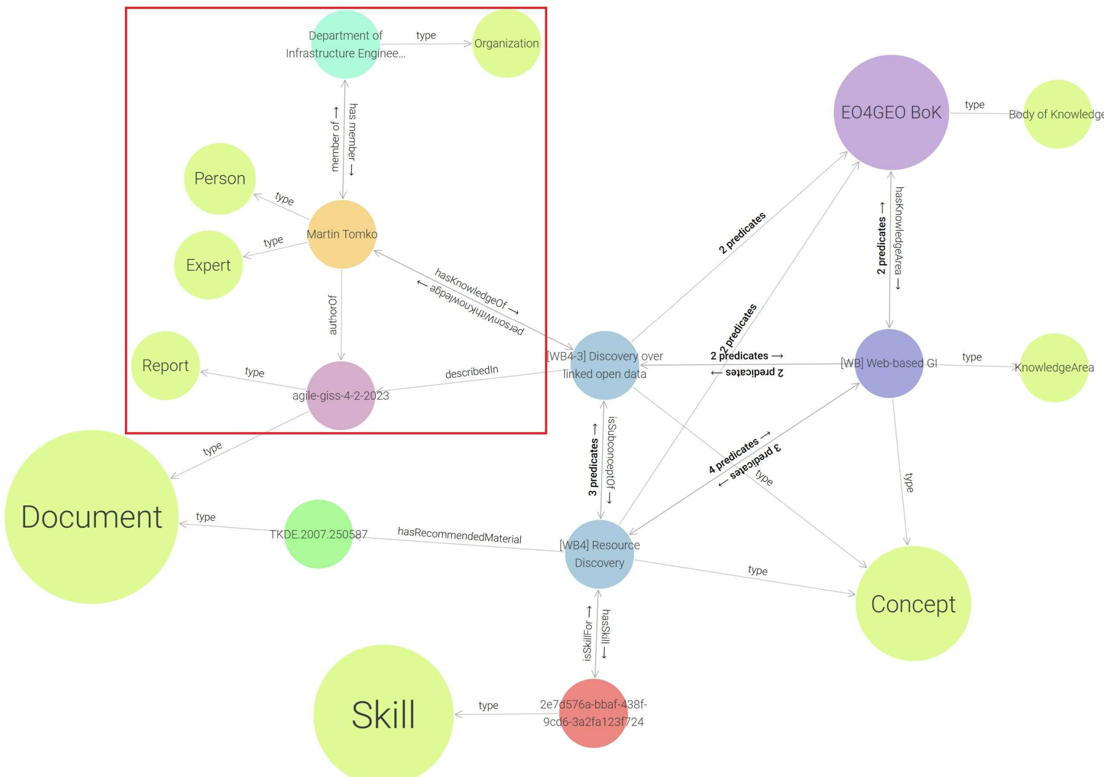


Figure 13: Visual RDF graph representation of RDF triples following the obok + boka ontologies applied to the EO4GEO BoK. (Source: Author, visualisation from GraphDB)

4.3 The EO4GEO knowledge graph

Together the EO4GEO BoK and EO4GEO BoK applications named graphs create the EO4GEO knowledge graph. The following subsections go into detail about the used architecture, explain the benefits of using ontologies and explain what drives knowledge footprint generation for the next subobjective.

4.3.1 The architecture of the EO4GEO knowledge graph

In this research there is chosen for a knowledge graph architecture that holds multiple named graphs. This allows for keeping multiple datasets separate based on provenance and the purpose of the data. This approach further has benefits for version management and allows for storing all the seven versions of the EO4GEO BoK over the years in a single knowledge graph. By the use of named graphs you can specifically request data from a specific BoK version. In this implementation the knowledge graph holds three graphs, note that graphs two and three are URI's and thereby adhere to named graphs. These graphs are:

1. **The default graph:** which is used to store the two ontologies in.
2. **https://bok.eo4geo.eu/concepts:** which is used to store the EO4GEO BoK data.
3. **https://bok.eo4geo.eu/applications:** which is used to store data created for the EO4GEO BoK applications.

Making use of ontologies in the knowledge graph allows the graph database to make uses of reasoning and inference over data in the knowledge graph. Figure 14 shows statistics about the EO4GEO knowledge graph in GraphDB. It also indicates that 23346 triples are created via the python script and 9335 triples are created through reasoning and inference. This is a great benefit of levering a knowledge graph to create links that did not exist yet.

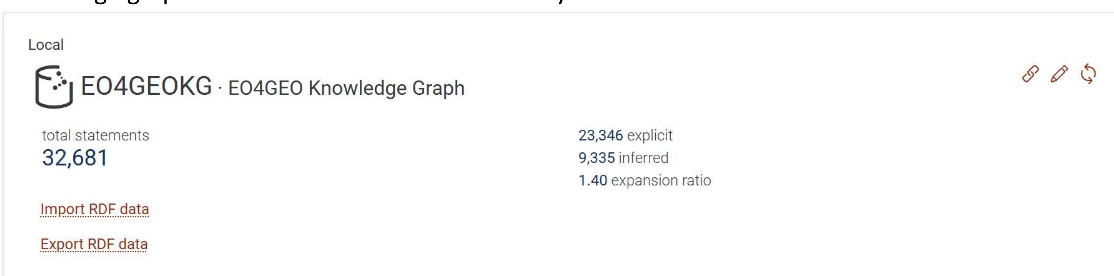


Figure 14: Statistics of the EO4GEO knowledge graph. (Source: Author, visualisation from GraphDB)

Listing 6 shows an example of which triples are inferred. For example, it now sees eo4geo:WB4 also as a skos:Concept and adds the skos:narrower construct with all concepts that are narrower then WB4. This inference happens due to specifying that obok:Concept is equivalent to skos:Concept and that skos:narrower is the inverse of skos:broader which is equivalent to obok:isSubconceptOf in the ontology. Listing 7 shows this in the OWL language.

```
eo4geo:WB4 a obok:Concept,
        skos:Concept ;
    rdfs:label "[WB4] Resource Discovery" ;
    obok:conceptStatus "In progress (GI-N2K)" ;
    obok:hasRecommendedMaterial <https://doi.org/10.1080/13658816.2012.739692>,
        <https://doi.org/10.1109/TKDE.2007.250587> ;
    obok:hasSkill eo4geo:d4c22c8f-449e-4d46-9d7f-606dc5efeb56 ;
    dcterms:description "Resource discovery means the discovery of resources including data and services
    needed for an application. Syntactic discovery refers to the discovery on the basis of syntactic
    comparison operations. It is classified as \"keyword-based\" and \"full-text-based\" discovery.
    Semantic discovery on the other hand, refers to the discovery of resources on he basis of some
    semantic definition. Therefore, semantic discovery requires that a resource be published by a semantic
    definition as defined in the topic WB3-5." ;
    rdfs:isDefinedBy eo4geo: ;
    rdfs:seeAlso eo4geo: ;
    skos:inScheme eo4geo:WB ;
    skos:narrower eo4geo:WB4-1,
        eo4geo:WB4-2,
        eo4geo:WB4-3 ;
    skos:notation "WB4" ;
    skos:topConceptOf eo4geo:WB .
```

```
eo4geo:WB4 a obok:Concept ;
    rdfs:label "[WB4] Resource Discovery" ;
    obok:conceptStatus "In progress (GI-N2K)" ;
    obok:hasRecommendedMaterial <https://doi.org/10.1080/13658816.2012.739692>,
        <https://doi.org/10.1109/TKDE.2007.250587> ;
    obok:hasSkill eo4geo:d4c22c8f-449e-4d46-9d7f-606dc5efeb56 ;
    dcterms:description "Resource discovery means the discovery of resources including data and
    services needed for an application. Syntactic discovery refers to the discovery on the basis of
    syntactic comparison operations. It is classified as \"keyword-based\" and \"full-text-based\"
    discovery. Semantic discovery on the other hand, refers to the discovery of resources on he basis
    of some semantic definition. Therefore, semantic discovery requires that a resource be published
    by a semantic definition as defined in the topic WB3-5." ;
    rdfs:isDefinedBy eo4geo: ;
    skos:inScheme eo4geo:WB ;
    skos:notation "WB4" ;
    skos:topConceptOf eo4geo:WB .
```

Listing 6: Left all triples including inferred constructs belonging to concept WB4 and right all generated triples without inference. (Source: Author)

```
###  http://example.org/OBOK/Concept
obok:Concept rdf:type owl:Class ;
             owl:equivalentClass skos:Concept ;
             dce:description "A concept is a theoretical construct that has been formed by combining
             particular instances into a general idea. These include theories, methods and technologies. All
             concepts together define a specific field."@en ;
###  http://example.org/OBOK/isSubconceptOf
obok:isSubconceptOf rdf:type owl:ObjectProperty ;
                    owl:equivalentProperty skos:broader ;
                    rdfs:domain obok:Concept ;
                    rdfs:range obok:Concept .
###  http://www.w3.org/2004/02/skos/core#broader
skos:broader rdf:type owl:ObjectProperty ;
             owl:inverseOf skos:narrower ;
             rdfs:comment "Broader concepts are typically rendered as parents in a concept hierarchy (tree).
             "@en ;
             rdfs:isDefinedBy <http://www.w3.org/2004/02/skos/core> ;
             rdfs:label "has broader"@en ;
             skos:definition "Relates a concept to a concept that is more general in meaning."@en ;
             skos:scopeNote "By convention, skos:broader is only used to assert an immediate (i.e. direct)
             hierarchical link between two conceptual resources."@en .
```

Listing 7: OWL constructs explaining why inference happens. (Source: Author)

4.3.2 Retrieving individual and organisational knowledge from the EO4GEO knowledge graph

Currently the EO4GEO knowledge graph consists of two named graphs, one with BoK content and one with data for the generation of knowledge footprints. Through SPARQL queries can be determined which EO4GEO concepts a specific organisation or a specific person has knowledge of. Two queries are explained below.

Figure 15 shows the first SPARQL query and its results to determine all the EO4GEO concepts a specific person has knowledge of. What can be seen is that SPARQL is able to query both the EO4GEO BoK concepts named graph and the EO4GEO applications named graph.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX obok: <http://example.org/OBOK/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX org: <http://www.w3.org/ns/org#>
PREFIX boka: <http://example.org/BOKA/>

select ?expertName ?organisationName ?conceptName where {
    ?expertURI rdf:type boka:Expert .
    ?expertURI foaf:name ?expertName .
    ?expertURI boka:hasKnowledgeOf ?conceptURI .
    ?conceptURI rdfs:label ?conceptName .
    ?expertURI org:memberOf ?organisationURI .
    ?organisationURI rdfs:label ?organisationName .
    filter(?expertName = "Martin Tomko")
} limit 100
```

Table | Raw Response | Pivot Table | Google Chart | Download as

Filter query results

Showing results from 1 to 2 of 2. Query took 0.1s, today at 10:18.

	expertName	organisationName	conceptName
1	"Martin Tomko"	"Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia"	"[WB3-6] Publishing linked open data"
2	"Martin Tomko"	"Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia"	"[WB4-3] Discovery over linked open data"

Figure 15: SPARQL query that returns every EO4GEO BoK concept that a specific person has knowledge of. (Source: Author, GraphDB)

Figure 16 shows the second SPARQL query and its results to determine all the members of a specific organisation and then subsequently all the concepts each person from that particular organisation has knowledge of. This result indicates which EO4GEO concepts an organisation has knowledge of. These kind of SPARQL queries will be used to generate organisational knowledge footprints in Chapter 5.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX obok: <http://example.org/OBOK/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX org: <http://www.w3.org/ns/org#>
PREFIX boka: <http://example.org/BOKA/>

select ?organisationName ?OrgMembersName  ?conceptName where {
    ?organisationURI rdf:type org:Organization.
    ?organisationURI rdfs:label ?organisationName .
    ?organisationURI org:hasMember ?membersOfOrganisationURI .
    ?membersOfOrganisationURI rdfs:label ?OrgMembersName .
    ?membersOfOrganisationURI boka:hasKnowledgeOf ?ExpertiseConceptURI.
    ?ExpertiseConceptURI rdfs:label ?conceptName .
    FILTER(CONTAINS(str(?organisationName), "University of Melbourne"))
} limit 100
```

Table | Raw Response | Pivot Table | Google Chart | Download as

Filter query results

Showing results from 1 to 8 of 8. Query took 0.1s, today at 10:25.

	organisationName	OrgMembersName	conceptName
1	"Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia"	"Kimia Amoozandeh"	"[WB3-6] Publishing linked open data"
2	"Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia"	"Kimia Amoozandeh"	"[WB4-3] Discovery over linked open data"
3	"Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia"	"Reza Arabsheibani"	"[WB3-6] Publishing linked open data"
4	"Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia"	"Reza Arabsheibani"	"[WB4-3] Discovery over linked open data"
5	"Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia"	"Ehsan Hamzei"	"[WB3-6] Publishing linked open data"
6	"Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia"	"Ehsan Hamzei"	"[WB4-3] Discovery over linked open data"
7	"Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia"	"Martin Tomko"	"[WB3-6] Publishing linked open data"
8	"Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia"	"Martin Tomko"	"[WB4-3] Discovery over linked open data"

Figure 16: SPARQL query that returns every EO4GEO BoK concept that a specific organisation has knowledge of. (Source: Author, GraphDB)

5. Knowledge footprints

The previous chapter introduced the foundation and how the EO4GEO knowledge graph is constructed. This chapter represents the findings related to the development of knowledge footprints and the footprint matching applications build upon the EO4GEO knowledge graph. These results represent the process described for subobjective three in Section 3.3. Besides showing the results, key parts in automating the creation of knowledge footprints will be explained.

5.1 Introducing knowledge footprints

The previous chapter has shown that concepts in the EO4GEO BoK can be transformed to graph data and later enriched with natural language processing annotations giving insight into whom holds specific expertise of an EO4GEO BoK concept. Section 4.3.2 has further shown that the EO4GEO KG can be queried through SPARQL which allows information to be returned in textual and/or tabular form. This section shows how query results from the EO4GEO KG can be made more understandable, visible and more easily placed into context through utilising visualisation techniques and making use of the hierarchical structure of the EO4GEO BoK. The results are knowledge footprints. A knowledge footprint is defined as a visual representation of the breadth of knowledge accumulated by a person or organisation and is in this thesis based on information in the EO4GEO knowledge graph. Figure 17 shows an example of a generated knowledge footprint, visualising knowledge of Wageningen University & Research. The footprint represents expertise shown in Agile published papers between 2021 – 2023 of whom members of Wageningen University & Research have contributed to. Appendix I further shows alternative visualisations, being the predecessors of the final knowledge footprint created on top of the EO4GEO knowledge graph and gives a brief overview what else is possible for hierarchical based data.

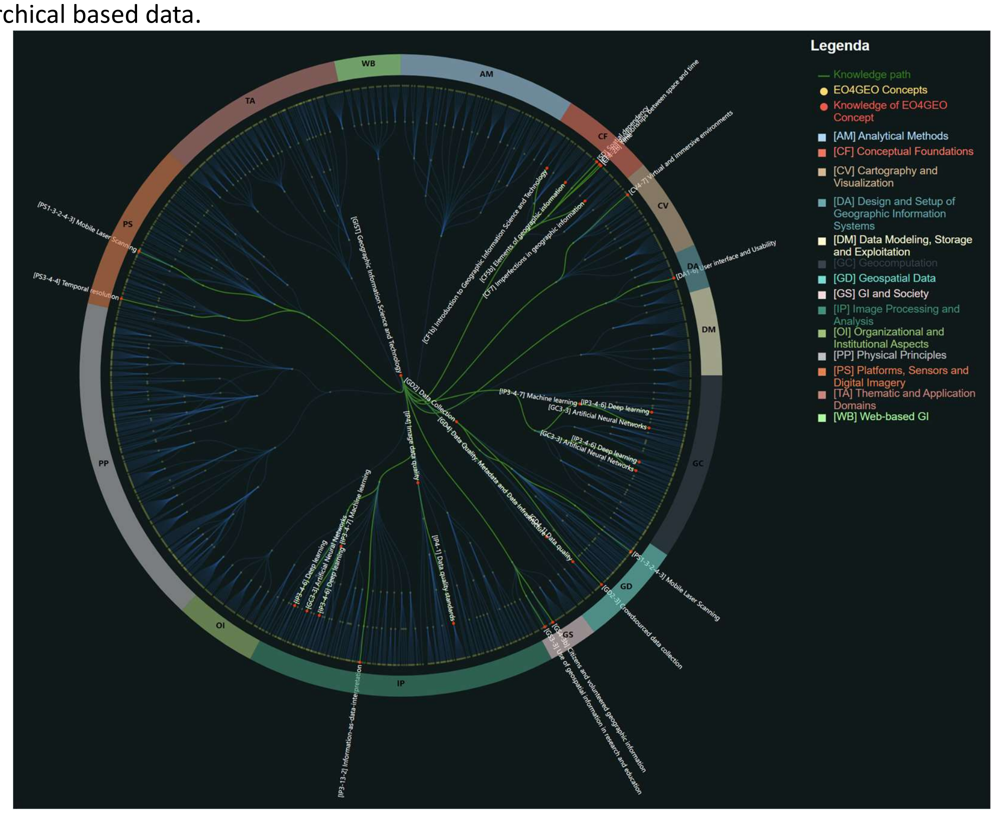


Figure 17: The organisational footprint of Wageningen University & Research. (Source: Author, mpvliet.github.io)

5.1.1 Knowledge footprint design decisions

Knowledge footprints are created by using a combination of a doughnut chart and D3's radial cluster tree visualisation. The outer doughnut chart is very much inspired by Elsevier's Wheel of Science (Elsevier, 2017), which can be seen in Appendix H. Radial cluster trees are suitable for hierarchical data and clustered data, which suites the EO4GEO BoK. For its design it was chosen to create a full radial cluster tree of all the EO4GEO concepts and let this visualisation serve as the basemap of a knowledge footprint. To effectively do that the visualisation is made a bit transparent which aims to remove focus and let the actual information, the EO4GEO concepts that an entity has knowledge of, be the main focus of the visualisation (J. von Engelhardt, personal communication, December 20, 2023). The yellow nodes are all the EO4GEO concepts, and the blue lines connect concepts to one another making the hierarchical structure of the EO4GEO BoK visible through parent child relations, or through the skos:broader and skos:narrower constructs in the ontology. The doughnut like chart around the knowledge footprint aims to make visual in which knowledge area, or skos:conceptScheme, the EO4GEO concepts are a part of. It aims to tell the viewer in which field an entity has knowledge without having to look at node labels. The yellow nodes, blue lines and the outer doughnut chart together form this basemap that is mentioned prior. Red coloured nodes indicate the EO4GEO concepts an entity has knowledge of and the green lines, which are called "knowledge paths", aim to make the hierarchical structure and thereby all the parent concepts of matched concepts visible through traversing this path to the root node. Offering the viewer more context about the expertise an entity displays.

5.1.2 The different types of knowledge footprints

Knowledge footprints can be created for different entities, the underlying ontology and the enriched information allow for creating knowledge footprints about the following entities:

- **A paper**: Which displays all the EO4GEO concepts that are matched to a paper. This is possible due to the boka:describedIn predicate between bibo:Document and skos:Concept. These prefixes are explained in Section 4.1.2 and 4.2.2.
- **An individual**: Which displays all the EO4GEO concepts that an individual has knowledge of. An individual knowledge footprint is an aggregation off all the knowledge displayed in papers the individual is an author of. This is possible due to the boka:hasKnowledgeOf predicate between boka:Expert/foaf:Person and skos:Concept. These prefixes are explained in Section 4.1.2 and 4.2.2.
- **An organisation**: Which displays all the EO4GEO concepts that an organisation has knowledge of. An organisational knowledge footprint is the aggregation off all the individuals and their knowledge that are a member off this organisation. Creating this footprint becomes possible due to the org:hasMember predicate between org:Organization and boka:Expert/foaf:Person. These prefixes are explained in Section 4.1.2 and 4.2.2.

5.2 Key components in automatic knowledge footprints creation

This section describes the key components that enable automatic knowledge footprint creation using the D3 library, JavaScript, a graph database and a front-end for interaction.

5.2.1 The SPARQL query that enables knowledge footprints

Every data retrieval or interaction with data on the website is built upon SPARQL queries that query the underlying graph database and request the needed information. Listing 8 shows the most crucial one which is used to retrieve all the information to create individual knowledge footprints.

```
1  PREFIX eo4geo: <https://bok.eo4geo.eu/>
2  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3  PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
4  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
5  PREFIX boka: <http://example.org/BOKA/>
6  PREFIX foaf: <http://xmlns.com/foaf/0.1/>
7
8  SELECT ?conceptName ?childName ?conceptID ?childID ?nodeColour ?showLabel ?labelSize ?nodeValue
9  FROM eo4geo:applications
10 FROM eo4geo:concepts
11 WHERE {
12   {
13     SELECT ?concept ?conceptName ?childName ?conceptID ?childID (IF(?knownByFirstEntity, 1 , 0 ) AS ?nodeValue) WHERE {
14       ?concept rdf:type skos:Concept;
15         rdfs:label ?conceptName;
16         skos:notation ?conceptID.
17       OPTIONAL {
18         ?concept skos:narrower ?child.
19         ?child rdfs:label ?childName;
20           skos:notation ?childID.
21       }
22       BIND(EXISTS {
23         ?expertURI rdf:type boka:Expert;
24           foaf:name ?expertName;
25           boka:hasKnowledgeOf ?concept.
26         FILTER(CONTAINS(LCASE(STR(?expertName)), LCASE("Sven Casteleyn")))
27       } AS ?knownByFirstEntity)
28     }
29   }
30   BIND(IF(?nodeValue = 1 , "#FF0000", "#FFFF00") AS ?nodeColour)
31   BIND(IF(?nodeValue = 1 , 16 , 0 ) AS ?labelSize)
32   BIND(IF(?nodeValue = 1 , 1 , 0 ) AS ?showLabel)
33 }
34 ORDER BY (?conceptName)
```

Listing 8: A SPARQL query to create individual knowledge footprints (source: Author, GraphDB)

Let's decompose the key parts in this query:

- Line 13-21 queries all the EO4GEO concepts that are of type skos:Concept, from those concepts return the name of the concept and the ID of the concept. Then through an optional statement, the query returns all the children of each concept, through the skos:narrower construct. If there is a child, return the name of the concept and its ID, if there is no child it does nothing.
- Line 22-27 returns a true or false value based on whether the subquery (23-26) matches triples in the query between line 13-21. So, it returns all the EO4GEO concepts a specific person has knowledge of. If one of those concepts is also in the main query (13-21), return true.
- On line 13 with the if statement, I am transforming the true or false, which the BIND exists statement returns, to 1 if true or 0 if false.
- Let's go to the main query (line 8-12 and 30-34). This query does a few checks on the results of the subquery with BIND statements. Lines 30-32 determine the style of nodes in the D3 visualisation. ?nodeValue = 1 indicates that the individual has knowledge of that specific concept.
- Line 9-10 are noteworthy in that I am specifically telling the query to extract data from two specific named graphs.

Table 1 show a sample of the output after running the SPARQL query from Listing 8. It shows that concept AM10 has three children (AM10-1, AM10-2, AM10-3), but those children have no children of their own. It also shows that this individual has knowledge of "[OI4-1] Adoption and implementation of standards" and therefore the nodeColour, showLabel, labelSize and nodeValue values get a different value, as stated in the various bind statements in the SPARQL query.

Table 1: Sample of the results of the SPARQL query in Listing 8. (Source: Author, GraphDB)

	conceptName	childName	conceptID	childID	nodeColour	showLabel	labelSize	nodeValue
1	"[OI4-1] Adoption and implementation of standards"		"OI4-1"		"#FF0000"	"1"^^xsd:integer	"16"^^xsd:integer	"1"^^xsd:integer
2	"[AM1-2] Analytical approaches"		"AM1-2"		"#FFFF00"	"0"^^xsd:integer	"0"^^xsd:integer	"0"^^xsd:integer
3	"[AM10-1] Problems of large spatial databases"		"AM10-1"		"#FFFF00"	"0"^^xsd:integer	"0"^^xsd:integer	"0"^^xsd:integer
4	"[AM10-2] Data mining approaches"		"AM10-2"		"#FFFF00"	"0"^^xsd:integer	"0"^^xsd:integer	"0"^^xsd:integer
5	"[AM10-3] Knowledge discovery"		"AM10-3"		"#FFFF00"	"0"^^xsd:integer	"0"^^xsd:integer	"0"^^xsd:integer
6	"[AM10] Data mining"	"[AM10-2] Data mining approaches"	"AM10"	"AM10-2"	"#FFFF00"	"0"^^xsd:integer	"0"^^xsd:integer	"0"^^xsd:integer
7	"[AM10] Data mining"	"[AM10-1] Problems of large spatial databases"	"AM10"	"AM10-1"	"#FFFF00"	"0"^^xsd:integer	"0"^^xsd:integer	"0"^^xsd:integer
8	"[AM10] Data mining"	"[AM10-3] Knowledge discovery"	"AM10"	"AM10-3"	"#FFFF00"	"0"^^xsd:integer	"0"^^xsd:integer	"0"^^xsd:integer

Appendix J further shows SPARQL queries that provide the data for generating knowledge footprints for organisations and research papers.

5.2.2 Transforming SPARQL results into the D3 hierarchy data format

The next step is to transform the SPARQL output into a suitable format that D3 can process and create D3 hierarchy-based visualisations from. To do that D3 requires a root node which is the top node of a hierarchy. And under each node D3 wants to know what the children of that node are and progressively what the children of that node are until a leaf node, a node with no children is reached. With this data structure D3 is then able to create those tree-like visualisations. Listing 9 shows the data structure D3 requires.

```
const data = {
  name: "Eve",
  children: [
    {name: "Cain"},
    {name: "Seth", children: [{name: "Enos"}, {name: "Noam"}]},
    {name: "Abel"},
    {name: "Awan", children: [{name: "Enoch"}]},
    {name: "Azura"}
  ]
};
```

Listing 9: JavaScript object showing an example of the basis of a D3 hierarchy structure. (Source: (D3, 2023))

The skos:narrower construct is a key factor in creating these parent child relationships through SPARQL queries. Listing 10 shows the output of transforming SPARQL JSON output into this D3 hierarchy structure. It shows that "[GIST] Geographic Information Science and Technology" is the root node, that this node has fourteen children nodes, which are also the fourteen knowledge areas in the EO4GEO BoK. As a further example it shows that node "[AM] Analytical Methods" has fourteen children nodes, which potentially also have children of their own. Besides showing parent-child relations the data structure also shows various other properties, like "labelSize", "nodeColour" etc. these properties are not required by D3 but are added by my own to influence how D3 draws nodes in the created knowledge footprints.

```
▼ Object
  ▼ children: Array(14)
    ▼ 0:
      ▶ children: (14) [{…}, {…}, {…}, {…}, {…}, {…}, {…}, {…}, {…}, {…}, {…}, {…}, {…}, {…}]
        id: "AM"
        labelSize: "0"
        matched: null
        name: "[AM] Analytical Methods"
        nodeColour: "#FFFF00"
        nodeValue: "0"
        nodeValueFirstEntity: null
        nodeValueSecondEntity: null
        showLabel: "0"
        value: 1
      ▶ [[Prototype]]: Object
    ▶ 1: {name: '[CF] Conceptual Foundations', id: 'CF', nodeColour: '#FFFF00', showLabel: '0', labelSize: '0', …}
    ▶ 2: {name: '[CV] Cartography and Visualization', id: 'CV', nodeColour: '#FFFF00', showLabel: '0', labelSize: '0', …}
    ▶ 3: {name: '[DA] Design and Setup of Geographic Information Systems', id: 'DA', nodeColour: '#FF0000', showLabel: '1', labelSize: '16', …}
    ▶ 4: {name: '[DM] Data Modeling, Storage and Exploitation', id: 'DM', nodeColour: '#FFFF00', showLabel: '0', labelSize: '0', …}
    ▶ 5: {name: '[GC] Geocomputation', id: 'GC', nodeColour: '#FFFF00', showLabel: '0', labelSize: '0', …}
    ▶ 6: {name: '[GD] Geospatial Data', id: 'GD', nodeColour: '#FFFF00', showLabel: '0', labelSize: '0', …}
    ▶ 7: {name: '[GS] GI and Society', id: 'GS', nodeColour: '#FFFF00', showLabel: '0', labelSize: '0', …}
    ▶ 8: {name: '[IP] Image processing and analysis', id: 'IP', nodeColour: '#FF0000', showLabel: '1', labelSize: '16', …}
    ▶ 9: {name: '[OI] Organizational and Institutional Aspects', id: 'OI', nodeColour: '#FFFF00', showLabel: '0', labelSize: '0', …}
    ▶ 10: {name: '[PP] Physical principles', id: 'PP', nodeColour: '#FFFF00', showLabel: '0', labelSize: '0', …}
    ▶ 11: {name: '[PS] Platforms, sensors and digital imagery', id: 'PS', nodeColour: '#FFFF00', showLabel: '0', labelSize: '0', …}
    ▶ 12: {name: '[TA] Thematic and application domains', id: 'TA', nodeColour: '#FFFF00', showLabel: '0', labelSize: '0', …}
    ▶ 13: {name: '[WB] Web-based GI', id: 'WB', nodeColour: '#FFFF00', showLabel: '0', labelSize: '0', …}
      length: 14
    ▶ [[Prototype]]: Array(0)
    id: "GIST"
    labelSize: "16"
    matched: null
    name: "[GIST] Geographic Information Science and Technology"
    nodeColour: "#FF0000"
    nodeValue: "1"
    nodeValueFirstEntity: null
    nodeValueSecondEntity: null
    showLabel: "1"
    value: 1
  ▶ [[Prototype]]: Object
```

Children of the first child of the root

The first child of the root node

Children of the root node

Root node

Listing 10: The created hierarchy data structure from the EO4GEO knowledge graph. (Source: Author)

To transform the SPARQL JSON output into the D3 hierarchy visualised in Listing 10, a JavaScript script is developed[2] that picks up the JSON response from the SPARQL query. Listing 12 on the next page shows this script. The script is a JavaScript function that processes a JSON object. This JSON object is the output of a SPARQL request and contains a "head" and "results" member. The "results" member contains the results of the SPARQL query and shows these as an array of bindings, which are the variables from the SPARQL query and its value. Listing 11 shows a sample of a SPARQL JSON output.

```
▼ {head: {,…},…}
  ▼ head: {,…}
    ▶ vars: ["conceptName", "childName", "conceptID", "childID", "nodeColour", "showLabel", "labelSize",…]
  ▼ results: {bindings: [{conceptName: {type: "literal", value: "[AM1-2] Analytical approaches"},…},…]}
    ▼ bindings: [{conceptName: {type: "literal", value: "[AM1-2] Analytical approaches"},…},…]
      ▼ [0 … 99]
        ▼ 0: {conceptName: {type: "literal", value: "[AM1-2] Analytical approaches"},…}
          ▶ conceptID: {type: "literal", value: "AM1-2"}
          ▶ conceptName: {type: "literal", value: "[AM1-2] Analytical approaches"}
          ▶ labelSize: {datatype: "http://www.w3.org/2001/XMLSchema#integer", type: "literal", value: "0"}
          ▶ nodeColour: {type: "literal", value: "#FFFF00"}
          ▶ nodeValue: {datatype: "http://www.w3.org/2001/XMLSchema#integer", type: "literal", value: "0"}
          ▶ showLabel: {datatype: "http://www.w3.org/2001/XMLSchema#integer", type: "literal", value: "0"}
```

Listing 11: Sample of a SPARQL JSON output, showing the results member and the array of bindings. (Source: Author)

Let's decompose the key parts of this script:

- The function on line 2 creates a "Map" object, which holds key-value pairs. The key can be seen as an EO4GEO concept, and the value will store information about the children and other information that influence how this node appears in the D3 visualisation.
- Line 5 – 35 loops through each "binding" or EO4GEO concept in the SPARQL JSON output and creates a key value pair in the created "Map" object.
- Line 38 – 76 again loops through each EO4GEO concept in the SPARQL JSON output and extracts information about a possible child. Then creates this child object and pushes that in the children array from the parent concept object in the created map.
- Line 79 – 84 determines the root node for the D3 hierarchy through looping through the SPARQL JSON response and finding the EO4GEO concept that is not a child of any other concept. If that concept is found return the value from the key value pair and use that as the D3 hierarchy.

[2] https://github.com/MPvliet/Thesis-GIMA-2023-2024/blob/main/Footprint-Website/src/js/sparql/sparqlToD3Hierarchie.js

```javascript
function transformSPARQLtoD3Hierarchie(json) {
  const nodes = new Map();

  // Fill the map with all conceptnodes, since each concept can be a parent and can have childs.
  json.results.bindings.forEach(item => {
    const parent = item.conceptName.value;
    const parentId = item.conceptID.value;
    const nodeColour = item.nodeColour.value;
    const showLabel = item.showLabel.value;
    const labelSize = item.labelSize.value;
    const nodeValue = item.nodeValue.value;
    const nodeValueFirstEntity = item.nodeValueFirstEntity
      ? item.nodeValueFirstEntity.value
      : null;
    const nodeValueSecondEntity = item.nodeValueSecondEntity
      ? item.nodeValueSecondEntity.value
      : null;
    const matched = item.matched ? item.matched.value : null;

    if (!nodes.has(parent)) {
      nodes.set(parent, {
        name: parent,
        id: parentId,
        nodeColour: nodeColour,
        showLabel: showLabel,
        labelSize: labelSize,
        nodeValue: nodeValue,
        nodeValueFirstEntity: nodeValueFirstEntity,
        nodeValueSecondEntity: nodeValueSecondEntity,
        matched: matched,
        value: 1,
        children: [],
      });
    }
  });

  // Then check if parent nodes have childs and push these nodes into the childrenArray.
  json.results.bindings.forEach(item => {
    const parent = item.conceptName.value;
    const nodeColour = item.nodeColour.value;
    const showLabel = item.showLabel.value;
    const labelSize = item.labelSize.value;
    const child = item.childName ? item.childName.value : null; // not all concepts have childs. so without child equals null
    const childId = item.childID ? item.childID.value : null;
    const nodeValue = item.nodeValue.value;
    const nodeValueFirstEntity = item.nodeValueFirstEntity
      ? item.nodeValueFirstEntity.value
      : null;
    const nodeValueSecondEntity = item.nodeValueSecondEntity
      ? item.nodeValueSecondEntity.value
      : null;
    const matched = item.matched ? item.matched.value : null;

    if (child !== null && !nodes.has(child)) {
      nodes.set(child, {
        name: child,
        id: childId,
        nodeColour: nodeColour,
        showLabel: showLabel,
        labelSize: labelSize,
        nodeValue: nodeValue,
        nodeValueFirstEntity: nodeValueFirstEntity,
        nodeValueSecondEntity: nodeValueSecondEntity,
        matched: matched,
        value: 1,
      });
    }

    if (child !== null) {
      const parentNode = nodes.get(parent);
      if (!parentNode.children) {
        parentNode.children = [];
      }
      parentNode.children.push(nodes.get(child));
    }
  });

  // Creates an array from all nodes.values and then find the root node / most top concept, through looping through this list. The rootnode is the node that has no parent. and thus is no child for any node in this list. // For the EO4GEO BoK this is always 'GIST' // so it is basically looking for the object with the name GIST and using that object as the full and complete datastrucutre, since GIST has all children nodes.
  const d3DataStructure = Array.from(nodes.values()).find(
    node =>
      !json.results.bindings.some(
        binding => binding.childName && binding.childName.value === node.name
      )
  );

  d3DataStructure.children.sort((a, b) => a.name.localeCompare(b.name)); //Sorting the children so I can predict how the knowledge footprint draws, I'm doing this so the outher dougnut chart is in the right order.
  return d3DataStructure;
}

export { transformSPARQLtoD3Hierarchie };
```

Listing 12: JavaScript script that transforms SPARQL JSON output into the D3 hierarchy data structure. (Source: Author)

5.2.3 A graph database

The graph database GraphDB is used for storing RDF triples and allows for performing SPARQL queries over the web through their RDF4J based SPARQL endpoint. While simple, using the rest API for data retrieval is a crucial part in communicating with the knowledge graph and creating knowledge footprints. Listing 13 shows a JavaScript function which makes use of the fetch method to interact with the SPARQL endpoint.

```javascript
async function genericSPARQLQuery(query) {
  try {
    const response = await fetch(
      `https://graphdb.gch.utwente.nl/repositories/EO4GEOKG?query=${encodeURIComponent(
        query
      )}`,
      {
        method: 'GET',
        headers: { Accept: 'application/sparql-results+json' },
      }
    );
    const data = await response.json();
    return data;
  } catch (error) {
    console.error('Fetch error:', error);
  }
}

export { genericSPARQLQuery };
```

Listing 13: JavaScript code that is used to interact with the RDF4J based SPARQL endpoint. (Source: Author)

5.2.4 HTML + CSS + vanilla JavaScript front-end

A combination of HTML, CSS and JavaScript is used to create a webpage that allows users to easily create knowledge footprints. Figure 18 shows a screenshot of the webpage. Users can influence knowledge footprint creation through configuring various properties in a form element, which can be seen at the left side of the screen. The right side of the screen is used to display the knowledge footprints and provide various elements to interact with the footprint. Appendix G provides a comprehensive explanation of the various possibilities and functionalities of the website, this explanation is written as part of the task description in the user evaluation (see also Chapter 6). Besides that the full code can be found at https://github.com/MPvliet/Thesis-GIMA-2023-2024/tree/main/Footprint-Website and the website is available on https://mpvliet.github.io/

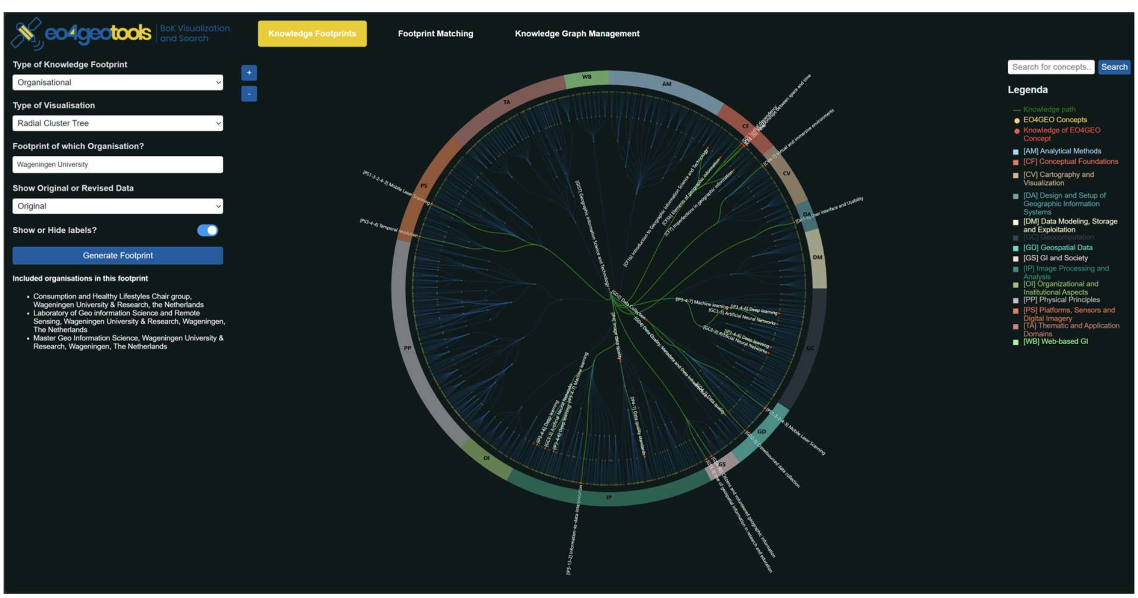


Figure 18: A snapshot of the website that allows for knowledge footprint generation. (Source: Author, mpvliet.github.io)

5.3 Benefits and interpretations of visualising expertise through knowledge footprints

Section 5.1, 5.2.1 and 5.2.4 have shown figures showing someone’s expertise through tabular format (e.g. Table 1) or through visualisations (e.g. Figure 17 and Figure 18). While a SPARQL query might be perfectly suitable in conveying who has specific knowledge of an EO4GEO concept, knowledge footprints focus on trying to create the bigger picture and places knowledge into context. It makes visible how their knowledge relates to other related EO4GEO concepts. The hierarchical nature of the EO4GEO BoK is herein extensively used.

This additional context also raises a new problem. Visualising the current implementation of annotating knowledge in knowledge footprints might cause the wrong interpretation of someone’s expertise. Figure 19 tries to highlight this problem.

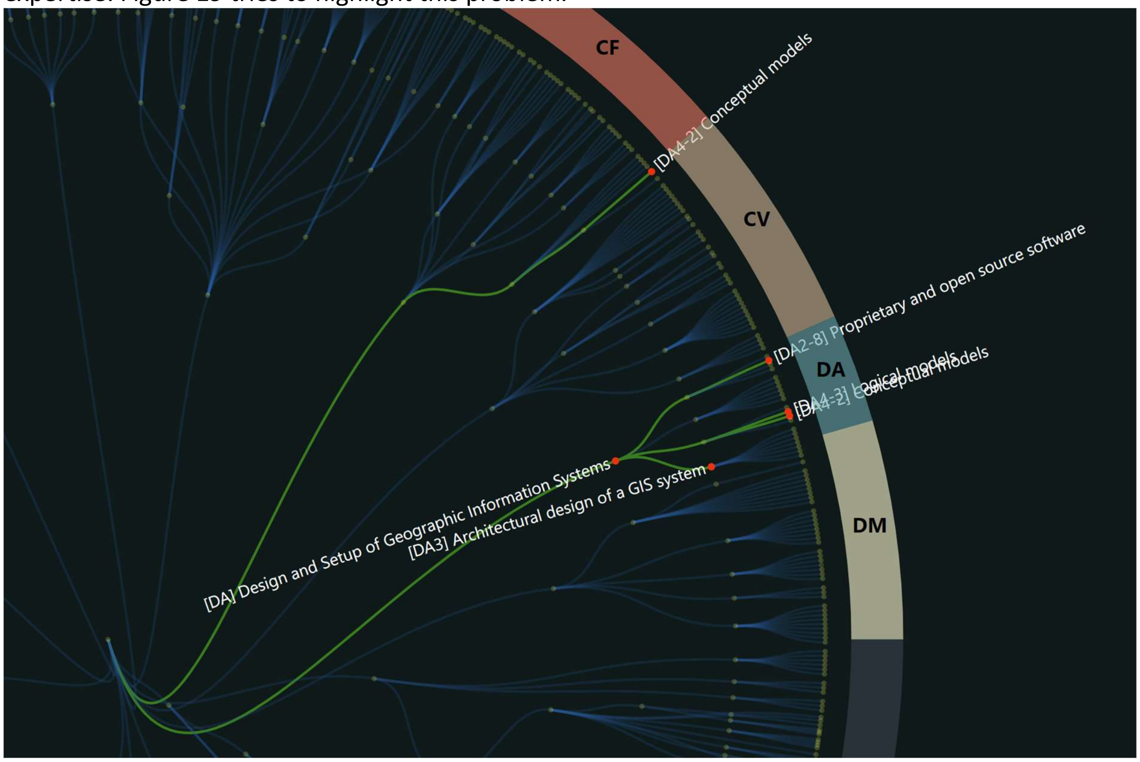


Figure 19: Part of a knowledge footprint of a fictional person. (Source: Author, mpvliet.github.io)

This footprint shows a footprint of a fictional person. One can interpret this footprint by thinking that this person has knowledge of design and setup of geographic information systems, knowledge of architectural designs of a GIS system, knowledge of proprietary and open-source software and is knowledgeable about logical and conceptual models. However, since not all subconcepts of “design and setup of geographic information systems” are coloured red, you could interpret this person as a person who is specialised in logical and conceptual models and has some general knowledge on (architectural) design and setup of a GIS. While this could potentially be true, this does not reflect the design and hierarchical structure of the EO4GEO BoK. Table 2 provides a look at the semantic meaning of relationships between concepts in the EO4GEO BoK and knowledge graph.

Table 2: Overview of the semantic meaning of relationships in the EO4GEO BoK and knowledge graph.

Definitions of semantic relationships in the EO4GEO BoK:	The EO4GEO BoK defines a sub concept as “A concept on a lower granularity level” (Lemmens, Albrecht, et al., 2022). Indicating that a sub concept is a specification of the parent concept.

Definitions of semantic relationships in the EO4GEO knowledge graph:	In the knowledge graph, the relation isSubconceptOf is equivalent to skos:broader. SKOS uses skos:broader and skos:narrower to enable the representation of hierarchical links. skos:broader is used to assert that a concept is broader or more general in meaning than another (W3C, 2009a). However, the W3C (2009) also mentions that these properties are flexible and can be interpreted in multiple ways. For example they mention the relation between one genre and its more specific species, and another interpretation is the relationship between one whole and its parts (W3C, 2009a).

Translating these semantic relationships between EO4GEO concepts to representing someone's knowledge, indicate that a person has to show knowledge of all subconcepts to infer that someone has adequate knowledge of a concept on a higher or broader granular level. However, the reality is probably more nuanced, and it might not be necessarily to know every detail of each sub concept to be knowledgeable of a broader concept, it does however indicate someone's level of expertise. The recommendation section suggests ideas to remove this degree of interpretation (see Section 7.5).

5.4 Use cases for visualising expertise through knowledge footprints

The previous sections introduced knowledge footprints, described how they are built upon the EO4GEO knowledge graph, that they are driven by SPARQL queries and that visualising knowledge through knowledge footprints provide extra context. This section goes into a potential more concrete use case.

The main use case for knowledge footprints is for promotional purposes, knowledge footprints can be used to represent and share someone's expertise, or someone's expertise shown in a recently written research paper. These knowledge footprints can be used for promoting research and proving potential interested parties in a quick overview which fields or what kind of knowledge is being discussed before reading the abstract of a research paper.

To extent that, knowledge footprints can be created to visually describe datasets, conference topics, workshops or any other activity to indicate what the activity is about.

Another use case could be annotating study programme courses with EO4GEO concepts and using these knowledge footprints to provide a quick overview of what students are expecting to learn during the specific study programme. Besides that, these footprints could provide an interesting starting point for further research, they could potentially do that by traversing the "knowledge path" and looking up semantically related EO4GEO concepts.

5.5 Footprint matching

Now that knowledge footprints can be created, it becomes possible to compare knowledge of different entities. The hypothesis is that comparing the expertise of different entities has benefits in finding suitable partners for collaboration and doing that through knowledge footprints offers context that textual comparisons are not able to deliver. The next section outlines the choices made in creating knowledge footprints through footprint matching.

5.5.1 Footprint matching design decisions

The first idea for footprint matching was to generate two knowledge footprints of two different entities next to each other and let the user visually compare the differences between these footprints. That however would not offer an efficient user experience. It was then decided to create a single knowledge footprint that incorporates the footprints of both entities. Figure 20 shows a screenshot of the footprint matching webpage showing a knowledge footprint that combines knowledge of two entities, University of Twente and Utrecht University. The reference corpus for both organisations are three years of AGILE papers (Volume 2 – 4, 2021 - 2023).

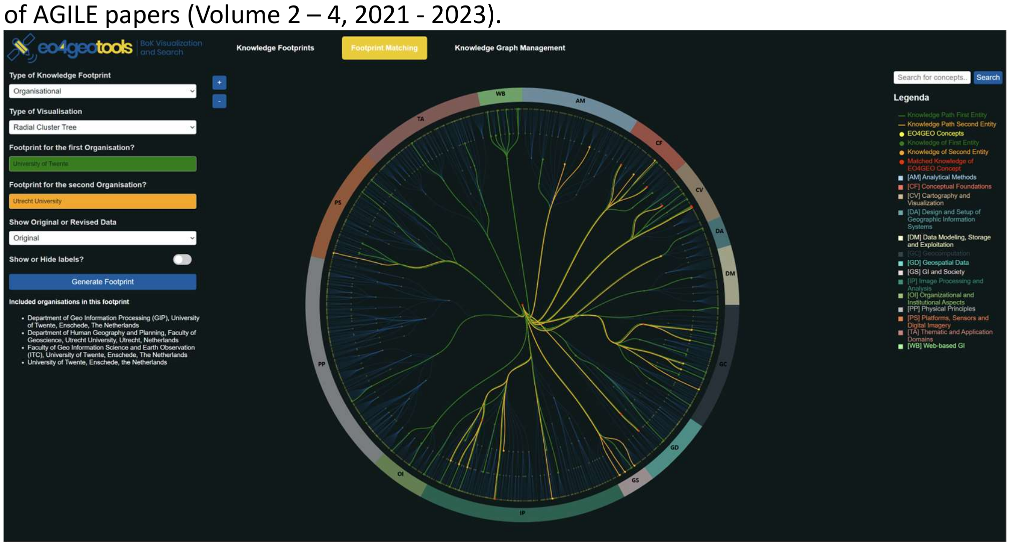


Figure 20: Overview of the footprint matching webpage, comparing University of Twente with Utrecht University. (Source: Author, mpvliet.github.io)

In the basis the created knowledge footprint is like a regular knowledge footprint, it shows the same basemap of EO4GEO concepts, highlights knowledge paths and highlights concepts that an entity has knowledge of. But uses colours to distinguish the two compared entities from each other. Knowledge paths and matched EO4GEO concept nodes get coloured green for the first entity and coloured orange for the second entity. Besides that, EO4GEO concept nodes that both entities have knowledge of get coloured red. Knowledge paths from EO4GEO concept nodes to the root node that both entities have do not overlap, but are drawn parallel to each other. Adopting a visualisation style often used to draw metro lines. Figure 21 shows this.

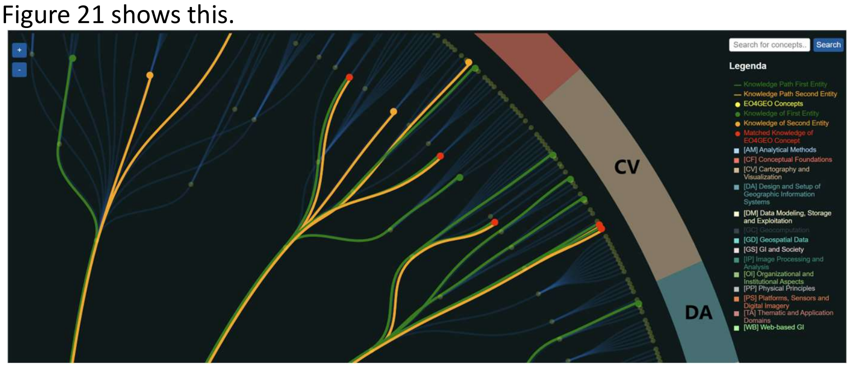


Figure 21: Figure 20 zoomed in, showing the similar visualisation style of drawing metro lines. (Source: Author, mpvliet.github.io)

5.5.2 The SPARQL query that enables comparing knowledge footprints.

To compare multiple entities in a knowledge footprint through footprint matching, data from two entities is needed. Requesting this data also goes via SPARQL queries. Listing 14 shows this query. In the basis the query is like the SPARQL query used to generate knowledge footprints, which is explained in Section 5.2.1. The main difference is that it contains a second BIND EXISTS statement on line 29 – 35 which is used to check whether this entity has knowledge of an EO4GEO concept.

```
1  PREFIX eo4geo: <https://bok.eo4geo.eu/>
2  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3  PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
4  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
5  PREFIX org: <http://www.w3.org/ns/org#>
6  PREFIX boka: <http://example.org/BOKA/>
7
8  SELECT ?conceptName ?childName ?conceptID ?childID ?nodeColour ?showLabel ?labelSize ?matched ?nodeValue ?nodeValueFirstEntity ?nodeValueSecondEntity
9  FROM eo4geo:applications
10 FROM eo4geo:concepts
11 WHERE {
12   {
13     SELECT ?concept ?conceptName ?childName ?conceptID ?childID (IF(?knownByFirstEntity, 1 , 0 ) AS ?nodeValueFirstEntity) (IF(?knownBySecondEntity, 1 , 0 ) AS ?
   nodeValueSecondEntity) WHERE {
14       ?concept rdf:type skos:Concept;
15         rdfs:label ?conceptName;
16         skos:notation ?conceptID.
17       OPTIONAL {
18         ?concept skos:narrower ?child.
19         ?child rdfs:label ?childName;
20           skos:notation ?childID.
21       }
22       BIND(EXISTS {
23         ?organisationURI rdf:type org:Organization;
24           rdfs:label ?organisationName;
25           org:hasMember ?membersOfOrganisationURI.
26         FILTER(CONTAINS(LCASE(STR(?organisationName)), LCASE("University of Twente")))
27         ?membersOfOrganisationURI boka:hasKnowledgeOf ?concept.
28       } AS ?knownByFirstEntity)
29       BIND(EXISTS {
30         ?organisationURI rdf:type org:Organization;
31           rdfs:label ?organisationName;
32           org:hasMember ?membersOfOrganisationURI.
33         FILTER(CONTAINS(LCASE(STR(?organisationName)), LCASE("Utrecht University")))
34         ?membersOfOrganisationURI boka:hasKnowledgeOf ?concept.
35       } AS ?knownBySecondEntity)
36     }
37   }
38   BIND(IF((?nodeValueFirstEntity + ?nodeValueSecondEntity) = 2 , "Match", "noMatch") AS ?matched)
39   BIND(IF((?nodeValueFirstEntity + ?nodeValueSecondEntity) = 2 , "#FF0000", IF(?nodeValueFirstEntity = 1 , "#008000", IF(?nodeValueSecondEntity = 1 , "#FFA500",
   "#FFFF00"))) AS ?nodeColour)
40   BIND(IF((?nodeValueFirstEntity = 1 ) || (?nodeValueSecondEntity = 1 ), 1 , 0 ) AS ?nodeValue)
41   BIND(IF((?nodeValueFirstEntity = 1 ) || (?nodeValueSecondEntity = 1 ), 16 , 0 ) AS ?labelSize)
42   BIND(IF((?nodeValueFirstEntity = 1 ) || (?nodeValueSecondEntity = 1 ), 1 , 0 ) AS ?showLabel)
43 }
44 ORDER BY (?conceptName)
```

Requests knowledge for the first entity.

Requests knowledge for the second entity.

Listing 14: The SPARQL query that requests all the data to enable footprint matching. (Source: Author, GraphDB)

Table 3 shows the first four results of the output of this SPARQL query.

Table 3: Sample of the results of the query in Listing 14. (Source: Author, GraphDB)

	conceptName	childName	conceptID	childID	nodeColour	showLabel	labelSize	matched	nodeValue	nodeValueFirstEntity	nodeValueSecondEntity
1	"[CF1b] Introduction to Geographic Information Science and Technology"	"[CF1-1b] What is Geographic Information Science and Technology"	"CF1b"	"CF1-1b"	"#FF0000"	"1"^^xsd:integer	"16"^^xsd:integer	"Match"	"1"^^xsd:integer	"1"^^xsd:integer	"1"^^xsd:integer
2	"[CF1b] Introduction to Geographic Information Science and Technology"	"[CF1-2b] Contributions to GIS and T by key allied fields"	"CF1b"	"CF1-2b"	"#FF0000"	"1"^^xsd:integer	"16"^^xsd:integer	"Match"	"1"^^xsd:integer	"1"^^xsd:integer	"1"^^xsd:integer
3	"[CF7] Imperfections in geographic information"	"[CF7-1] Vagueness"	"CF7"	"CF7-1"	"#FF0000"	"1"^^xsd:integer	"16"^^xsd:integer	"Match"	"1"^^xsd:integer	"1"^^xsd:integer	"1"^^xsd:integer
4	"[CF7] Imperfections in geographic information"	"[CF7-2] Error-based uncertainty"	"CF7"	"CF7-2"	"#FF0000"	"1"^^xsd:integer	"16"^^xsd:integer	"Match"	"1"^^xsd:integer	"1"^^xsd:integer	"1"^^xsd:integer

5.5.3 Footprint matching use cases

The main use case for footprint matching lays in finding potential suitable partners for collaboration. It could be a valuable tool in the research phase, while you compare potential suitable partners it could help decision making. This works through providing a quick overview of all the knowledge an entity holds beyond your own familiar domain.

Another potential use case is to make use of the hover functionality on the webpage to receive a list of all individuals and organisations who hold knowledge of the concept, you hover over. So, for example, if someone wants to find candidates with knowledge of machine learning, the footprint matching site is able to provide that. This again can prove a starting point for finding a potential partner. A next step could be comparing an organisation's knowledge with the knowledge of one of the entities of the generated list and compare both entities' profile. Footprint matching leverages the extra context provided by knowledge footprints and assists in finding domain specialists for potential collaborations.

These two scenarios are explained and validated in the knowledge footprint evaluation. The next chapter describes how this evaluation is set up and gives insight into the results of the evaluation.

6. Evaluating knowledge footprints

This chapter discusses the systematic assessment used to evaluate knowledge footprints. The purpose of this assessment is to gain insight into whether knowledge footprints are valuable, and evaluate the effectiveness of the created website. From the evaluation results, areas for improvements and strengths of knowledge footprints will be identified. These findings help to answer the main research question of this study and provide valuable guidance for further recommendations. The evaluation design is presented first, followed by showing the evaluation results.

6.1 Evaluation design

For the evaluation twelve participants are chosen. These twelve participants have in common that they are all active in the EO/GI domain and have all submitted research papers to AGILE in the last three years. The last part is crucial, while the evaluation structure and questions are the same, the tasks in the evaluation are personalised and are related to their published AGILE papers. Appendix G shows the full evaluation document that these twelve participants received. The evaluation consists of five tasks and nine questions. These tasks and their purpose are described below.

Task 1: Generate a knowledge footprint of one of your papers.

Brief description:	Task one asks the evaluator to generate a knowledge footprint of one of their AGILE papers they have published. Besides that, it walks through the main functionalities of the website.
Accompanying questions:	<u>Q1</u>: Performing this task generates a knowledge footprint of one of your AGILE papers. Looking at this visualisation and the matched concepts, would you say this paper is correctly annotated with the right EO4GEO concepts?
Goal:	Introduces and gives a thorough explanation of the website and its functionalities, providing the evaluator with the knowledge to properly answer the next questions. In addition, it introduces the evaluator to the first entity, papers, of which knowledge footprints can be generated. The goal of question one is to gather insight into the accuracy of the NLP tool, used to annotate research papers with EO4GEO concepts.

Task 2: Make your personal AGILE footprint

Brief description:	Task two asks the evaluator to generate their own personal knowledge footprint, showing their expertise.
Accompanying questions:	<u>Q2</u>: What are your first thoughts about the way knowledge is visualised in these knowledge footprints. Does it offer you new insights?
Goal:	Introduce personal/individual AGILE knowledge footprints and explains that individual footprints are an aggregation of all knowledge, extracted from papers, the individual was an author of. The goal of question two is to gather insight into the effectiveness of visualising knowledge through knowledge footprints. It introduces the hierarchical structure of the EO4GEO BoK in a different visualised way. This task aims to evaluate this.

Task 3: Improve your personal AGILE footprint

Brief description:	Task three asks the evaluator to improve their personal AGILE knowledge footprint. Through the functionalities on the “Knowledge graph management” webpage.

Accompanying questions:	No extra questions.
Goal:	The goal of this step is partly to evaluate whether the website is easy in use, meaning that people can easily refine their personal knowledge footprints by adding and/or removing matched EO4GEO concepts through the “Knowledge graph management” page. It party accesses the accessibility of the framework. If people barely added/removed concepts, and say the annotations are likely incorrect. Then this shows that the current method is to much of a burden and not an ease to use.

Task 4: Make your organisational AGILE profile

Brief description:	Task four asks the evaluator to create an organisational knowledge footprint of the organisation they are a member of.
Accompanying questions:	<u>Q3</u>: How do you interpret this organisational footprint, can you explain what you see (e.g. is the organisation very specialised or has knowledge about a lot of general concepts?) <u>Q4</u>: Could you say these organisational footprints effectively indicate in which field or fields an organisation is contributing in? Please explain briefly why or why not.
Goal:	Introduces the evaluator to organisational knowledge footprints, and that organisational footprints are an aggregation of all personal knowledge footprints from organisation members. The goal of question three is to create an understanding how people perceive knowledge footprints and what conclusions they create. The goal of question four is quite directly asking people whether organisational knowledge footprints are effective in displaying in which fields a specific organisation is active in. Which validates whether this use case is valuable.

Task 5: Discover Footprint Matching

Brief description:	Task five introduces the evaluator to footprint matching. And a potential workflow that describes a use case for this application.
Accompanying questions:	<u>Q5</u>: Could you see yourself using these matched footprints to effectively try and find out whether a certain organisation is a good fit to collaborate with? Please explain briefly why or why not. <u>Q6</u>: Could you see these footprint offer you insights in what kind of knowledge an organisation might also have to offer, besides looking for the specific concepts. Besides that, could this help you in this process of finding potential collaborations?
Goal:	Evaluates the use case “Is footprint matching a helpful tool in finding potential partners for collaboration?”. The goal of question five is evaluating whether someone would use this approach. The goal of question six is evaluating whether the extra benefit and context knowledge footprints give, is helpful for decision making.

Concluding questions

Accompanying questions:	<u>Q7</u>: What do you think of the way expertise is displayed in these knowledge footprints? <u>Q8</u>: Would you say these knowledge footprints provide possibilities to improve geoscience collaboration? Please explain briefly why or why not. <u>Q9</u>: What other potential use cases for these visualisations have come to mind?
Goal:	Gather a general opinion and offering the reviewer some questions to conclude their final thoughts on knowledge footprints. Question seven validates the design principles of the visualisation. Question eight partly answers the main research question of this research from the perspective of a potential knowledge footprint user. Question nine is an open question and provides potential unexplored use cases.

6.2 Evaluation results

From the 12 participants 4 participants provided an answer. This section summarises and evaluates their responses. Appendix K provides the complete anonymised responses.

6.2.1 Evaluation question 1

Question 1: Performing this task generates a knowledge footprint of one of your AGILE papers. Looking at this visualisation and the matched concepts, would you say this paper is correctly annotated with the right EO4GEO concepts?

Key points:

- Participant A mentions that the publication keywords are similar than the matched EO4GEO concepts, yet the detected concepts are broader than what the paper covers.
- Participant C mentions that some fully relevant concepts where not found. Some concept where relevant but not in the right context as shown in the framework. Some concepts where matched but not related to the paper.

Interpretation from author

The results from this question indicate that the script is able to pattern match keywords in EO4GEO concepts and words used in the paper. However, the NLP method has no real clue about the context a EO4GEO concept operates in, it does not keep parent and sub concepts and their relation into account. What indicated this the most is the following quote from the evaluation: "The concept data quality is relevant but it is associated to image processing in the framework and our paper did not consider image data." (Participant C, 2024). For reference, Figure 22 shows part of the footprint participant C used to form the quoted conclusion. The participant talks about concept "[IP4-1] Data quality standards", this concept has the following parent concept "[IP4] Image data quality". However, to make a counterpoint it seems like the EO4GEO BoK has not placed these concepts well in their hierarchical structure. Subconcepts of the IP4-1 concept are related to data quality standards, and not directly applied to the image processing and analysis field.

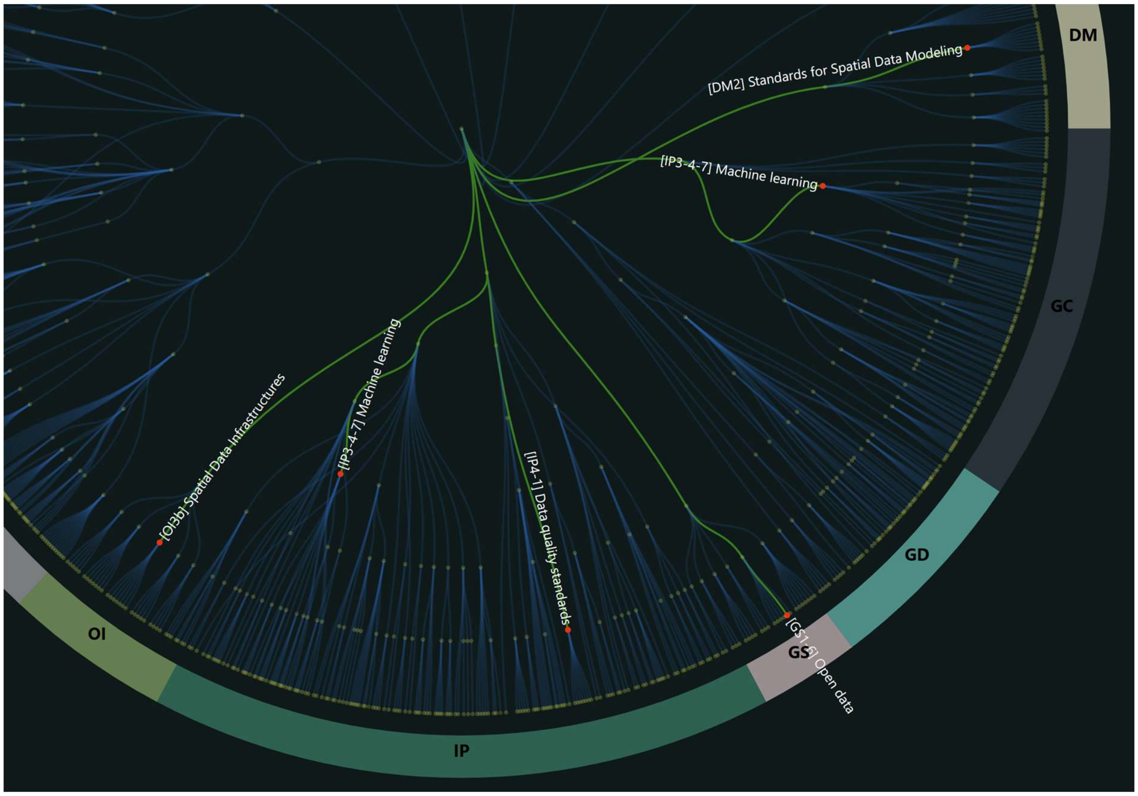


Figure 22: Footprint relevant to participant C, that matches for example IP4-1 Data quality standards. (Source: Author)

6.2.2 Evaluation question 2

Question 2: What are your first thoughts about the way knowledge is visualised in these knowledge footprints. Does it offer you new insights?

Key points:

- Participant A mentions that creating your own personal AGILE footprint should generally not offer you new insights about yourself, however sees purposes for collaboration and networking, by checking the nodes that match themselves, one can find others with similar knowledge.
- Participant B questions why some concepts are listed multiple times and questions why concepts this person has the most knowledge of are not made more visible. Participant B also mentions that there are concepts this person does not recognize themselves in.
- Participant C does not mention new insights, but made more general comments about the application itself, the participant liked browsing the concepts, but wishes some concepts are grouped, as the interface requires many clicks.
- Participant D likes the visualisation as it showcases knowledge in GIScience. Participant D, like participant B, was thrown off by why concepts where present in several categories (e.g. Machine learning), but understands the reasoning behind it.

Interpretation from author

This question aimed to evaluate whether visualising knowledge in this type of visualisation, using the EO4GEO BoK hierarchy was effective. The received responses do not specifically give insight into whether visualising the BoK hierarchy was effective. The answers from participant B and D however do suggest that people noticed parent sub relationships better, when they were present. For example, they both noticed that “Machine learning” is present multiple times in the graph and have multiple parents.

6.2.3 Evaluation question 3

Question 3: How do you interpret this organisational footprint, can you explain what you see (e.g. is the organisation very specialised or has knowledge about a lot of general concepts?)

Key points:

- Participant A mentions that the generated organisational footprint has a balance between general and specialised concepts. However, as this person is new to the organisation, the person can not effectively evaluate whether the footprint accurately reflects the expertise within the organisation.
- Participant B is unable to answer.
- Participant C liked looking at the organisational footprints, but mentioned that he/she was more interested in comparing organisations.
- Participant D suggests some standardisation in concept depth or concept levels within the graph/hierarchy. Participant made the example for EO4GEO concept “Deep learning”

Interpretation from author

Question three aims to start people thinking about knowledge footprints and their effectiveness. Participant D made an interesting point, the text below dives a bit deeper into her suggestion. The concept “Deep learning” appears three times in parent child relations within the EO4GEO BoK. Below these three paths are shown, traversing from concept “Deep learning” to the root concept.

1. GIST **->** [**GC**] Geocomputation **->** [**GC3**] Artificial intelligence (AI) in EO and GI **->** [**GC3-12**] AI algorithms **->** [**IP3-4-6**] Deep learning.
2. GIST **->** [**IP**] Image processing and analysis **->** [**IP3**] Image understanding **->** [**IP3-4**] Image classification **->** [**IP3-4-6**] Deep learning.
3. GIST **->** [**GC**] Geocomputation **->** [**GC3**] Artificial intelligence (AI) in EO and GI **->** [**IP3-4-7**] Machine learning **->** [**IP3-4-6**] Deep learning.

Path one and two seem to follow a logical depth sequence where deep learning is the fourth child in the tree, path three is probably the one that made participant D question the usage of depth in the graph. In this path concept “[IP3-4-7] Machine learning” is the fourth child in the tree and “[IP3-4-6] Deep learning” the fifth child.

Due to only matching someone’s expertise based on the concept title and not taking into account the parent classes, someone’s expertise gets matched with all these three knowledge paths. This indicates a person has knowledge about deep learning in the field of geocomputation, and in the field of image processing and analyses. However, this assumption cannot theoretically be made without determining if this person has knowledge about deep learning in both fields.

Furthermore, the use of depth as a measure of specialisation has consequences on the interpretation of the footprint. If concepts appear on multiple depths, expertise can not be interpreted well.

6.2.4 Evaluation question 4

Question 4: Could you say these organisational footprints effectively indicate in which field or fields an organisation is contributing in? Please explain briefly why or why not.

Key points:

- In general, this question was difficult to judge, not everybody knows the domains exactly wherein their organisation is working.

- Participant C made the observation that the organisational footprint does not reflect the scope of their university but probably only the AGILE perspective on it.
- Participant D mentions that when you use the main categories of concepts, the main knowledge areas, the footprint reflects in which fields they are working on. However, mentioned that the footprint does not indicate the amount of contribution to topics within the graph.

Interpretation from author
Question 4 aimed to evaluate whether organization footprints effectively visualize in which fields an organization is contributing in. The answers were a bit mixed, this was a hard question, as you would need to have a clear picture of the domains in which their organization are contributing to. The answers from participant C and D indicate that organisational footprints are able to draw a picture in which fields an organisation is working in, but that these organisational footprints lack the clarity regarding the extent of their organisations contributions to each matched concept in the graph. Participant D suggested doing something with the width or size of elements in the graph which is an interesting thought and will be taken into account for improvements.

6.2.5 Evaluation question 5

Question 5: Could you see yourself using these matched footprints to effectively try and find out whether a certain organisation is a good fit to collaborate with? Please explain briefly why or why not.

Key points:
- Participant A mentions that using the footprint matching workflow to compare organisations is a good starting point, but that decisions would not solely be based on the tool.
- Participant B mentions that the participant rather preferred a tool that shows other organisations working on a specific topic.
- Participant C mentions that the participant could use this tool indeed, but rather to look for papers to read first.
- Participant D mentions that he/she would use the tool to explore potential collaboration partners, but also mentions that collaboration is more than fitting research interest. So only sees use cases for exploration.

Interpretation from author
This question aimed to evaluate whether footprint matching could be a potential tool in finding collaborations. In general, the responses all indicate that the tool is valuable in the exploration phase of finding potential partners, but underline that deciding to collaborate with a potential organisation, can not be done purely on a statistical match and thus is not sufficient on its own. Participant B and C provide improvements or new ideas that can be made by leveraging the data in the EO4GEO knowledge graph, for further elaboration see Section 6.3.1.

6.2.6 Evaluation question 6

Question 6: Could you see these footprint offer you insights in what kind of knowledge an organisation might also have to offer, besides looking for the specific concepts. Besides that, could this help you in this process of finding potential collaborations?

Key points:
- Participant B does not think he/she would often look at what other organisations are contributing to.
- Participant C mentions that knowing the source, or what determines that an organisation has knowledge of a specific concept, would be important. He/she also suggests introducing a

chatbot, a question and answering interface, which suggest interesting organisations to look at. Which can later be used to match with your own organisation.
- Participant D believes these visualisations have benefits, but again suggests that something should be introduced to give an idea how much knowledge or how much papers contributed to a specific concept.

Interpretation from author
Question 6 aims to evaluate whether showing the full knowledge footprint, containing all the EO4GEO concepts is beneficial in understanding what kind of expertise this organisation holds. Evaluating the answers provided by the participants, conclude that some might deem it as beneficial and others do not. Participant C and D provide interesting improvements to improve the efficiency and trustworthiness of the tool, for further elaboration see Section 6.3.1.

6.2.7 Evaluation question 7
Question 7: What do you think of the way expertise is displayed in these knowledge footprints?

Key points:
- Participant A mentions that the visualisation is easy to understand, read and follow.
- Participant B mentions that the visualisation is not easy to read, remarking that all topics seem to be equally important, which they are not from the participants perspective.
- Participant C likes the general idea and thinks looking at our domain in that way is promising. But would like more context on how the full classification was achieved.
- Participant D discusses the semantics of “expertise” and “knowledge” but does not comment on the visualisation itself.

Interpretation from author
Question seven aims to evaluate whether the current visualisation, using the radial cluster tree from D3 as basis, helps show an entities expertise. Answers for this question are quite mixed, the number of responses is too low to form a conclusion.

6.2.8 Evaluation question 8
Question 8: Would you say these knowledge footprints provide possibilities to improve geoscience collaboration? Please explain briefly why or why not.

Key points:
- Participant A underlines that a tool can not replace the human element in collaboration. In the participants experience collaboration is often sought in people they already worked with, know or are recommended. The participant also mentions that specific domain knowledge is not always the dealbreaker as reputation is often considered as well.
- Participant B does not think they will provide possibilities for improved collaboration.
- Participant C believes the tool could be helpful to identify similarities and complementarities between papers, and believes the tool could be helpful in the bidding process.
- Participant D believes opportunities for collaboration may certainly be improved, but that this mainly depends on the nature of the collaboration. The participant provides the example that the visualisation could be helpful when he/she is writing a paper and wants to find someone knowledgeable in the same field. But that the tool is less helpful when to find a complementary partner, i.e. a group that does research in areas where the participant does not themselves.

Interpretation from author
Question eight helps to gauge whether the participants believe knowledge footprints could be a tool that could improve geoscience collaboration. The evaluation suggest that knowledge footprints won't replace decision making in the search for collaboration, as there are to much other factors in play, like reputation and familiarity. However, the evaluation underlines that there lays an opportunity in the exploration phase in the search for collaboration.

6.2.9 Evaluation question 9

Question 9: What other potential use cases for these visualisations have come to mind?

Key points:

- Participant A suggests the following functionalities on top of the current tool; show a list of related papers when hovering over a node/EO4GEO concept. Some statistics, most popular topic, most cited paper, most active people/organisation.
- Participant B would like to find people that work on the same topic.
- Participant C thinks the tool can be used to help in exchange of master students.
- Participant D suggests that once all base data (not only AGILE papers) is included, it becomes possible to use the tool to find reviewers for paper reviews or finding workshop organisers. Participant D also sees more personal use cases; creating a personal profile to showcase, or use the tool as a method for critically reflecting on your own research contributions that have been published in contrast to research done, but not published. This gives you an idea about how other might view you. He/she also suggests it could be interesting to create time series of contributions over time.

6.3 Evaluation takeaways

This section summarises areas for improvements in the process of creating knowledge footprints and identifies the strengths of knowledge footprints.

6.3.1 Areas for improvement

The evaluation reveals the following areas for improvement, these are grouped by the area they are relevant for.

EO4GEO BoK and concept extraction through NLP

1. The used NLP processing method can be improved by taking into account the context of EO4GEO concepts, what are the parent and child concepts of a concept and beyond only taking the title of each concept, also use the description that is available with each concept in the BoK. By making this improvement, papers can be more accurately matched with EO4GEO concepts and potentially reduces the number of matched concepts someone does not recognise themselves in. In addition to this, using the parent and child relationships between concepts it becomes possible to accurately determine whether an individual has knowledge of a concept that appears multiple times in the EO4GEO BoK under different knowledge areas. Emphasizing different fields that leverage a specific concept. (e.g. Deep learning appears in the knowledge area "Geocomputation" and in knowledge area "Image processing and analysis")
2. EO4GEO suggests a depth/hierarchy in each EO4GEO concept title (e.g. "[IP3-6-4-1] Gauss filter"), however this hierarchy is not always enforced when EO4GEO concepts can have multiple parents. Additionally, different knowledge area's within the EO4GEO BoK show varying depths. For example, the knowledge area "Image Processing and Analysis" contains leaf nodes that are 5 levels deep, whereas the knowledge area "Web-based GI" only goes 3 levels deep. This difference does not imply that the leaf concepts of Image Processing and

Analysis are more specialised than the leaf concepts of Web-based GI. Rather, it reflects the differences in interpretation of the creators of the EO4GEO BoK.

Knowledge footprint website and interface

1. Pursue improvements to reduce the number of clicks (e.g. many mouse actions to check EO4GEO concept nodes).
2. Pursue the usage of visual indicators, e.g. size or width to emphasize the amount of knowledge in each EO4GEO concept (personal footprint), the number of papers/contributions to a EO4GEO concept (organisational footprint).
3. Pursue better tooling or a visualisation that shows all the organisations/individuals that are contributing to a specific topic. The word better is chosen here as the current website does support listing organisations and individuals that contribute to a specific topic, but the evaluation suggests that this is not visible or good enough.
4. Pursue tooling or a visualisation that recommends related papers to a specific EO4GEO concept. This can be used to find similarities and complementarities between papers.
5. Pursue a question and answering system, in the form of a chatbot, that can analyse the EO4GEO knowledge graph and suggests potential collaborators.
6. Incorporate the source of knowledge, the specific AGILE paper, in each matched node that indicate an entity has knowledge about a concept.
7. Pursue a better way to highlight organisations that complement your own organisation. This could be beneficial in finding groups that do research in areas where your own organisation does not.
8. Pursue showing statistics, most popular topics, most active person and/or organisation.
9. Pursue creating time series, to show contribution of a specific entity over time and how their research interests change.

General improvements

1. Explain why concepts can be present in multiple categories.
2. Explain how the full classification of creating knowledge footprints is achieved.

6.3.2 Strengths

The evaluation reveals the following strengths, these are grouped by the area they are relevant for.

Knowledge footprint website and interface

The evaluation shows that knowledge footprints are able to visualise knowledge in the GIScience domain and that knowledge footprints offer a promising and unique way to look at the GIScience domain.

Use cases of knowledge footprints

The evaluation shows that knowledge footprints are a good starting point in the exploration phase of finding collaboration but that they will not completely replace the process of finding partners. At last knowledge footprints are seen as a building block for many other use cases as summarised in Section 6.3.1. A new use case for personal footprints have been identified, using the footprint to critically reflect on their own research contributions and thus providing personal insights.

7. Discussion and conclusion

This chapter discusses and concludes the main findings of this research and the relevance of the underlying technologies of knowledge footprints in detail.

7.1 Summary of key findings

This thesis started with the following research objective: "Develop and evaluate a knowledge mapping and visualisation framework based on semantic web principles to support geoscience collaboration by effectively connecting experts to domain knowledge within the EO4GEO body of knowledge". This objective comes with the following main research question: "Can a framework that uses knowledge graphs for knowledge mapping and visualization effectively represent and compare knowledge footprints of geoscience experts based on the EO4GEO Body of Knowledge?" To answer this question, the sub questions are answered first. This is done for each subobjective in the next section.

7.1.1 Develop a standard format for describing the semantics of a BoK and apply this standard to the EO4GEO BoK.

To achieve this subobjective, three sub questions are formed. These are answered below.

How is the EO4GEO BoK semantically defined and how is it annotated?

The EO4GEO BoK is semantically defined through various relationships between concepts. The relationships that were crucial to creating knowledge footprints were relationships that describe the hierarchical structure between EO4GEO concepts (ex. the isSubclassOf relation). Besides these relationships, concepts in the EO4GEO BoK are enriched with information like a description, the person who contributed to this concept, an optional literature reference and associated skills. Whilst these properties are included in the created ontology for bodies of knowledge, they were not extensively used in generating knowledge footprints. Besides that, the EO4GEO BoK is annotated through a human-driven procedure. Section 4.1.1 provides a detailed answer to this question.

What relationships and properties should be included in a general ontology for bodies of knowledge?

For the purpose of creating knowledge footprints, the semantic relation between EO4GEO concepts to make the hierarchical relation needs to be defined. This is done by using the skos:broader and skos:narrower constructs. Besides that, constructs for concept descriptions and concept labels are added. For representing concepts and knowledge areas the skos:Concept class is used for the former and skos:ConceptScheme for the latter. In addition, and enabling combining BoKs in the future the obok:BodyOfKnowledge class is added to the ontology for bodies of knowledge. This class has administrative properties and stores the provenance of skos:Concept and thereby BoK content. For other data in the EO4GEO BoK the FOAF ontology is used to represent people in the BoK and the BIBO ontology is used for representing resources in the BoK. Section 4.1.2 provides a detailed answer to this question.

How can the existing EO4GEO body of knowledge be converted to graph data following this standard format?

The EO4GEO BoK has an API where all the information in the BoK can be accessed in JSON format. With the help of python packages like requests and JSON, communication was made with the EO4GEO API and EO4GEO BoK content could be extracted. The python package RDFLib, in combination with the created ontology acting as a guide for the data schema, was used to transform BoK data in JSON format to RDF triples.

These three steps create an ontology for bodies of knowledge and create the first RDF representation of the EO4GEO BoK, thereby achieving this subobjective.

7.1.2 Develop a method that allows connecting individual expertise to the EO4GEO body of knowledge.

To help achieve the second objective, three sub questions are formed. These are answered below.

How can expert's knowledge be connected to domain knowledge within the EO4GEO BoK? & How to make use of ontology relationships to connect knowledge graphs?

These two questions are closely related to each other. The received NLP expertise annotations (Vidanelage, [Forthcoming]) were leveraged to create an RDF dataset that connects the matched EO4GEO concepts to each author, and each author to their organisation. These NLP annotations were already semantically rich and contained the DOI of the research paper, a list of authors, a list of organisations and the list of related EO4GEO concepts. Due to leveraging the RDF data model and using the same EO4GEO concept URI's that were used in the creation of the EO4GEO BoK in RDF, it became possible to automatically integrate and link these datasets together. However, it was important to keep the provenance of these two datasets separate from each other. To achieve this separation, named graphs are utilised. Utilizing named graphs has benefits in keeping RDF statements for different purposes and/or different provenance separate from each other. The two created named graphs together form the EO4GEO knowledge graph.

The evaluation showed that using these NLP expertise annotations is a good starting point, however there is room for improvement. The evaluation indicates that the NLP process can be improved by also looking at the context a specific EO4GEO concept operates in, using the concept description and the parent and child relations of the concept, this context becomes better known.

How can the EO4GEO knowledge graphs be used to query someone's expertise?

The query language for RDF data, SPARQL, was used to query the EO4GEO knowledge graph and receive someone's expertise. By specifying in the query from which named graph someone wants to retrieve statements, it becomes possible to receive query results that combine the EO4GEO BoK concepts and the NLP expertise annotations. Section 4.3.2 and 5.2.1 describe the details what these queries look like.

This objective is achieved by leveraging the RDF dataset, named graphs and use of SPARQL.

7.1.3 Develop a user evaluated framework that visualises graph data-based knowledge footprints to support geoscience collaboration.

To achieve this subobjective, three sub questions are formed. These are answered below. The created visualisations are the created knowledge footprints and the user evaluated framework is the created website.

How can creating visual knowledge footprints based on the EO4GEO BoK be automated?

The D3.js library was used to create knowledge footprints. This library was chosen due to offering extensive methods to create highly customisable visualisations. Alternatives like Charts.js exist, but were not explored as D3.js was sufficient. A knowledge footprints is thereby a combination of D3's radial cluster tree visualisation and a D3 doughnut chart. This knowledge footprint is based on data from the EO4GEO knowledge graph. By hosting this knowledge graph in a cloud instance of GraphDB, the knowledge graph became accessible on the web. Through a combination of SPARQL queries and JavaScript scripts that transforms sparql-json to the required data format, that D3 likes to receive to create radial cluster trees, it becomes possible to automate knowledge footprint creation.

Which use cases become possible when individual and organisational knowledge graphs can be visualised?

The main use case for knowledge footprints that were identified is for promotional purposes. Knowledge footprints can be used to represent and or promote someone's published paper or visualise someone's expertise. The main benefit the visualisation provides is giving users context about EO4GEO concepts, The hierarchical nature of the BoK enables this. Another use, not shown in this research, is to apply the methods to generate knowledge footprints to visualise which EO4GEO concepts are taught in study programme courses. The evaluation showed that people recognize the potential of visualising expertise. However, the evaluation also showed that people have different views on the created knowledge footprints, and consequently see different use cases for leveraging knowledge footprints. Section 6.3 has summarised those.

What are potential further applications for knowledge footprints?

The concept of footprint matching was identified as a potential further application for knowledge footprints. Footprint matching gives users the opportunity to compare the expertise of different entities. This thesis shows that this process can be leveraged during the research phase, while looking for potential partners to collaborate. The evaluation results underline this thought, but mention that the tool won't replace the search for collaboration as there is a certain human aspect to consider and factors like reputation are often prioritized higher than specific domain knowledge.

7.2 Evaluation of existing theories and models

To critically evaluate the created ontology for bodies of knowledge the ontology is compared with the ontology or schema Microsoft uses for their Microsoft Academic Knowledge Graph (MAKG) (Microsoft, 2021). Figure 23 shows a visual representation of their ontology.

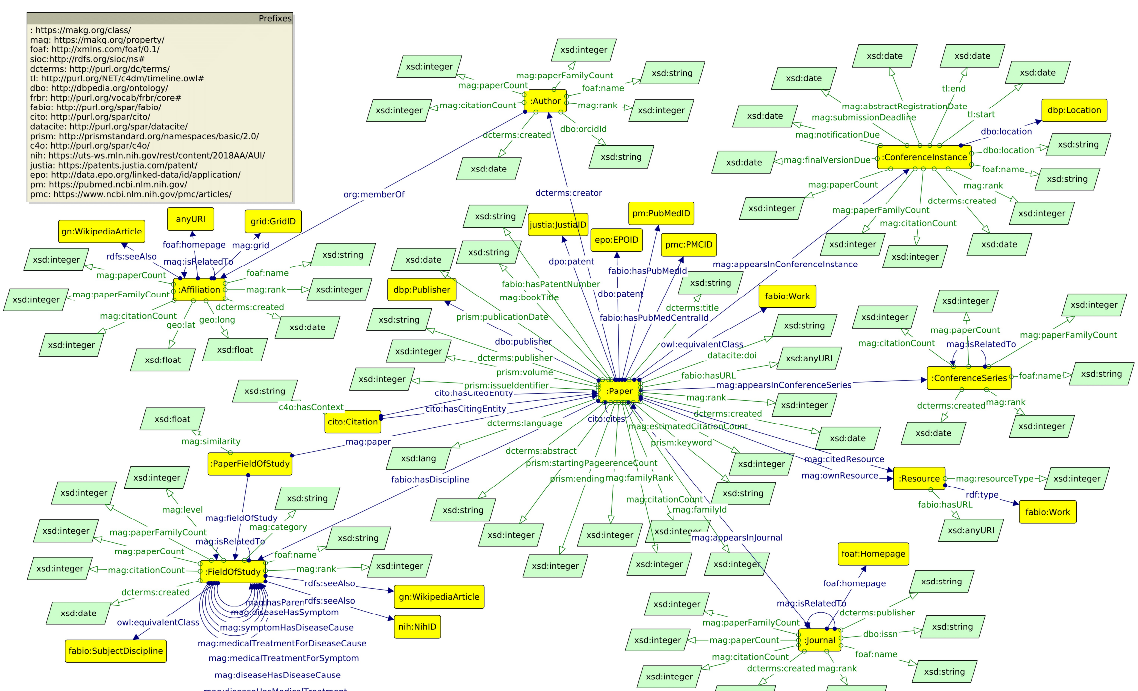


Figure 23: A visual representation of the Ontology Microsoft uses for their Microsoft Academic Knowledge Graph. (Copied from: (Microsoft, 2021))

Besides the difference that MAKG stores quite a bit more data properties to each class, there are definitely similarities. MAKG also models everything around a paper, linking papers to author, albeit via the dcterms:creator construct (OBOK uses :authorOf), and author links to a :Affiliation through the org:memberOf construct, which OBOK also uses. What is curious is that they did not adopt the org:Organisation class instead they used :Affiliation. Furthermore, they link papers to a :FieldOfStudy.

Adopting linking :Papers to the journal or conference they appear in could be a great addition to OBOK. This allows creating knowledge footprints from all papers in a journal or all papers part of a conference, potentially offering new promotional visualisations. At last, a nice touch is that MAKG defined :Affiliation and :ConferenceInstance spatially through dbp:Location or geo:lat and geo:long enabling potential GEOSPARQL spatial filtering techniques. OBOK does not currently support that.

Microsoft further mentions that :FieldOfStudy are also hierarchical in nature, grouping fields of study under more general fields of study. One of their products build upon this graph dataset is the topic graph explorer. Figure 24 shows a screenshot of their Topic Graph Explorer. Quite interesting is that Microsoft also opted for representing concepts, or in their case topics, through a hierarchy-based visualisation. Microsoft however chose for a force directed tree layout, which was deemed not suitable in this thesis, due to its unpredictable way of placing nodes. The stable background of the knowledge footprints serves as the basemap to create a visualisation upon.

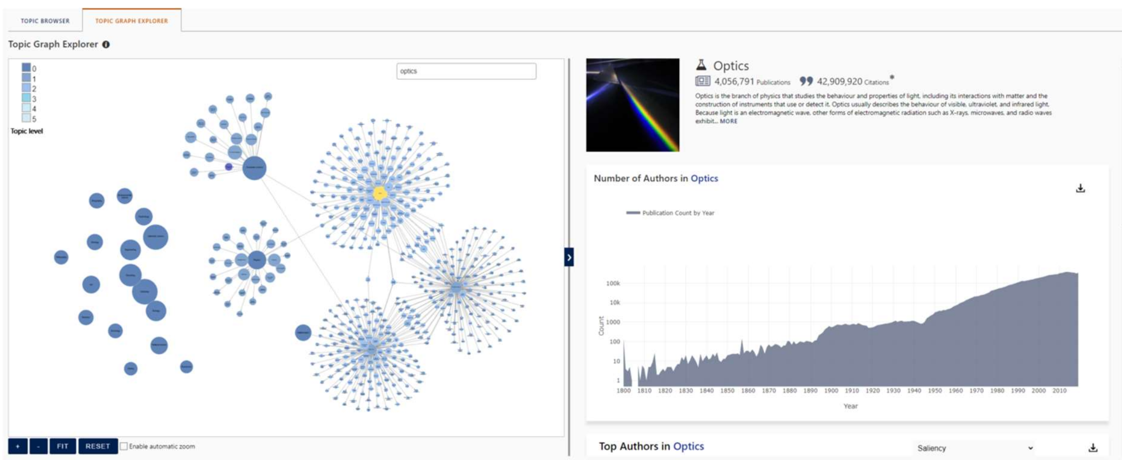


Figure 24: Microsoft's Topic Graph Explorer. (Copied from: (Microsoft, 2020))

Microsoft (2020) further mentions the following “We feel that a visual representation of our topic hierarchy can give our users better context. Given that topics in Microsoft Academic can have many parents and children, seeing these relationships in a directed graph brings perspective to their structure.” Microsoft’s thoughts on the benefits of visual representations are similar to the conclusions of this thesis on the main benefit of knowledge footprints.

Elsevier’s Wheel of Science can also be compared to knowledge footprints. An introduction to the Wheel of Science can be found in Appendix H. What stood out about their visualisation is that they did something with node sizes. Wherein the size of the node tells something about the momentum or visibility of a particular topic (Elsevier, 2021). While something like that could also be interesting to introduce to knowledge footprints and create a sort of heatmap indicating current popular topics. The evaluation suggested that a variable node size of an EO4GEO concept could also tell something about how many times someone has shown expertise on that concept, telling something about someone’s level of expertise.

At last, lets briefly reflect upon the usage of ontologies in knowledge graph architectures. Some sources (DuCharme, 2021; Polikoff, 2023) suggest an ontology is not needed in knowledge graphs, while other sources or other big knowledge graphs like Google’s Knowledge Graph, Microsoft’s

Academic Knowledge Graph (MAKG) or Amazon's product catalog show that ontologies are quite valuable and are used. This is a question that I asked myself while creating the EO4GEO knowledge graph and its applications. I do make use of the semantic relationships, classes and properties defined in the ontologies I use. Besides that, the ontology assists me in modelling RDF statements following the schema defined in the ontology. However, the applications I made still work fine when I remove the RDF statements from the ontologies in GraphDB/ knowledge graph. In this study I mentioned that I make use of the reasoning and inference capabilities of GraphDB that the OWL ontologies enable. Most inferred statements were constructs that were the inverse of another construct, so this meant for me that I had to do write less python code to create RDF statements from the EO4GEO BoK. However, since I moved to a multiple named graph architecture and wrote queries that specifically requests statements from specific named graph, meant that I could not retrieve and access inferred statements. This is due to the fact that GraphDB stores inferred statements in the default graph of your graph database and the default graph is a combination of inferred statements and the aggregation of all named graphs. Meaning if I wanted to include the default graph and retrieve inferred statements, I also retrieved statements from other named graphs I did not want in my SPARQL query results. So, to come back to this question, are ontologies needed in knowledge graph architectures? In the end I think it differs on the use case of the knowledge graph and how you enrich it with data, are you enriching the knowledge graph via semi-automated or automated processes, which the likes of Google, Microsoft and Amazon do, then I guess the extra context and meaning about constructs that an ontology provides is beneficial for various machine learning purposes. Is the knowledge graph more manually enriched and the meaning of classes is not necessarily needed for the end user, then the usage of ontologies might offer to be less beneficial.

7.3 Significance of this research

The significance of this research lies in demonstrating that applications can be developed on knowledge graphs. Although the RDF data format might seem unfamiliar or intimidating to some, its flexible schema and ability to easily link things together is beneficial. I was sceptical myself, but I realised that a SPARQL query results is just a structured JSON response. The main learning curve involves learning how to retrieve data from a knowledge graph through SPARQL.

In addition to the benefits of levering knowledge graph technologies, visualising knowledge graph content, and in particular content in the EO4GEO knowledge graph, has shown that knowledge footprints have potential and may have a place in the search for collaboration.

At last, the created ontology for bodies of knowledge could serve as a foundation for digitising and utilising other bodies of knowledge for similar purposes, such as knowledge footprint creation, thereby enhancing their accessibility, usability and their ability to integrate with other datasets. Which align with the main objective of the semantic web.

7.4 Limitations of this research

This study and the followed methodology have some limitations, these are either limitations that appeared during writing this thesis, or limitations that became apparent once the evaluation results were processed. These limitations are described in the next paragraphs.

Let's put some emphasis on the method used that links the matched EO4GEO concepts from the NLP extracting process to expertise from the authors. The reality is that often when papers contain multiple authors, is that each author has written a different piece of the paper. Meaning that in theory each author has shown their expertise by writing a different part of the whole. The assumption made in this research is that all authors of the processed paper have knowledge about all EO4GEO concepts displayed in the paper. This is assumption is not correct and shows a limitation in the generated knowledge footprints. This assumption is made since there is no automatic method to determine which author has written what, and thereby unable to determine their specific contribution.

Another thing that could have been done better is how organisations are modelled in the ontology for bodies of knowledge and how the organisation triples are created with the NLP extractions data. For example, an organisation name in the current knowledge graph could be the following: “Delft University of Technology, Department Architectural Engineering & Technology, Delft, The Netherlands” This perfectly matches the organisation one of the authors has submitted as their organisation while publishing their paper. However, this text string stores more information about this organisation, the department, the name of the university, the city where the university resides and in which country this organisation is located. These all could have been separate classes, but automatically extracting that data would have been harder. However, this limitation has also brought some benefits, as the underlying knowledge footprint SPARQL query uses contains filter statements to filter organisation names, it becomes possible to generate knowledge footprints on other granular levels. For example, create the knowledge footprint of all organisations in Spain or generate a knowledge footprint about all Agile papers from volume 4 (query contains “https://doi.org/10.5194/agile-giss-4”).

In addition, another limitation has been identified during extracting organisation names from research papers. In the above section an example of the Delft University is given, but when the knowledge graph for all organisation names that contain “Delft University” is queried. 12 different variations of naming conventions appear. It perfectly represents heterogeneity challenges on attribute level.

Finally, although the evaluation produced promising results, the described objectives were not met for a number of questions. This was probably due to a combination of the duration of the questionnaire and too open-ended questions that left too much room for interpretation.

7.5 Future research and recommendations

I want to disclose that I am not an expert on natural language processing methods, but what I would like to see in future research are advances in natural language processing and I am curious if machine learning can successfully extract and classify all the different parts of this text string: “Delft University of Technology, Department Architectural Engineering & Technology, Delft, The Netherlands”. Advancements in this area potentially solve this part of the limitation of this study.

To solve the heterogeneity problem displayed in the limitation section I would suggest looking into how Google solves heterogeneity challenges on entity levels. So, problems like; are these 12 organisations from different sources the same organisation, could be solved. Google calls their method to solve this “Entity Linkage” (Dong, 2023). While they operate on a whole different level of scale, I suspect we can learn something and apply parts of their methodology.

For footprint matching I would recommend adding more flexibility in which entities and or provenance of knowledge that can be compared with each other. Currently you can only compare AGILE based footprints, meaning the current footprints only show expertise displayed through AGILE papers. In addition to that, it could be beneficial to compare a paper knowledge footprint with a footprint from an individual.

Speaking of the visualisation of knowledge footprints, I want to suggest looking into the usage of visual indicators, e.g. size or width to emphasize the amount of knowledge in each EO4GEO concept (personal footprint) or the number of papers/contributions to a EO4GEO concept (organisational footprint).

To remove the degree of interpretation of knowledge footprints it is suggested to make the following adjustments in the process of creating knowledge footprints.

- Someone's expertise should only be able to be (automatically) annotated with leaf concepts. As leaf nodes specify knowledge on the lowest granular level in the BoK.
- The annotation process should take parent and child relations into account to determine in what field a concept operates in. This could result in more accurately determining if someone's expertise is actually applicable in the knowledge are the concept appears in.
- For parent concepts modify the styling of these nodes in the knowledge footprint, to indicate the level of expertise. For example, use different colours or partly colour the node (e.g. through a small pie chart), to indicate the level of expertise they hold. This could be done by calculating how many subconcepts of the max subconcepts they have expertise of.
- It could even be possible to let the knowledge graph handle this, through usage of ontology rules.

At last, besides the full list of areas of improvement in Section 6.3.1 from the evaluation, I want to briefly underline the idea to research whether the EO4GEO knowledge graph can be the basis for a question and answering system. A chatbot, as suggested in the evaluation, is likely more efficient and effective in suggesting potential partners for collaboration. The created knowledge footprints could then be used as extra context. I believe the defined semantics in both the domain ontology (the EO4GEO BoK) and the expertise annotations could be an interesting starting point to research whether a ChatGPT like system, build upon the EO4GEO knowledge graph could make interesting collaboration suggestions.

References

Abu-Salih, B. (2021). Domain-specific knowledge graphs: A survey. *Journal of Network and Computer Applications*, *185*, 103076. https://doi.org/10.1016/j.jnca.2021.103076

AGILE. (2021). *Volume 2, 2021 | 24th AGILE Conference on Geographic Information Science*. https://agile-giss.copernicus.org/articles/2/

AGILE. (2022). *Volume 3, 2022 | 25th AGILE Conference on Geographic Information Science "Artificial Intelligence in the service of Geospatial Technologies"*. *3*. https://agile-giss.copernicus.org/articles/3/

AGILE. (2023). *Volume 4, 2023 | 26th AGILE Conference on Geographic Information Science "Spatial data for design"*. *4*. https://agile-giss.copernicus.org/articles/4/

BigKnowledge LLC. (2021). *Geospatial Technology Explorer*. Geospatial Technology Explorer. Retrieved 23 Octobre 2023, from https://geospatial.bigknowledge.net/explorer/

BigKnowledge LLC. (2023). *Data Science & Analytics. Mapped.* ArcGIS StoryMaps. Retrieved 23 Octobre 2023, from https://storymaps.arcgis.com/stories/0e8ae7ab747042598da6545f6bb1c98f

Bizer, C. (2011). Evolving the Web into a Global Data Space. In A. A. A. Fernandes, A. J. G. Gray, & K. Belhajjame (Eds.), *Advances in Databases* (Vol. 7051, pp. 1–1). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24577-0_1

Bloom's Taxonomy. (2020). *Blooms Taxonomy: Resource for Educators*. What Is Bloom's Taxonomy? Retrieved 6 May 2024, from https://bloomstaxonomy.net/

Buil-Aranda, C., Hogan, A., Umbrich, J., & Vandenbussche, P.-Y. (2013). SPARQL Web-Querying Infrastructure: Ready for Action? In C. Salinesi, M. C. Norrie, & Ó. Pastor (Eds.), *Advanced Information Systems Engineering* (Vol. 7908, pp. 277–293). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41338-4_18

Carroll, J. J., Bizer, C., Hayes, P., & Stickler, P. (2005). Named graphs. *Journal of Web Semantics*, *3*(4), 247–267. https://doi.org/10.1016/j.websem.2005.09.001

Chaudhri, V. K., Baru, C., Chittar, N., Dong, X. L., Genesereth, M., Hendler, J., Kalyanpur, A., Lenat, D. B., Sequeda, J., Vrandečić, D., & Wang, K. (2022). Knowledge graphs: Introduction, history, and perspectives. *AI Magazine*, *43*(1), 17–29. https://doi.org/10.1002/aaai.12033

Corea, F. (2019). *An Introduction to Data: Everything You Need to Know About AI, Big Data and Data Science* (Vol. 50). Springer International Publishing. https://doi.org/10.1007/978-3-030-04468-8

D3. (2023). *Hierarchies | D3 by Observable*. Hierarchies. Retrieved 18 February 2024, from https://d3js.org/d3-hierarchy/hierarchy

Dong, X. L. (2023). *Generations of Knowledge Graphs: The Crazy Ideas and the Business Impact* (arXiv:2308.14217). arXiv. http://arxiv.org/abs/2308.14217

Dubois, C., Jutzi, B., Olijslagers, M., Pathe, C., Schmullius, C., Stelmaszczuk-Górska, M. A., Vandenbroucke, D., & Weinmann, M. (2021). Knowledge and skills related to active optical sensors in the body of knowledge for earth observation and geoinformation (EO4GEO BOK). *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *V-5–2021*, 9–16. https://doi.org/10.5194/isprs-annals-V-5-2021-9-2021

DuCharme, B. (2013). *Learning SPARQL: Querying and updating with SPARQL 1.1* (2. ed). O'Reilly. ISBN: 9781449371470

DuCharme, B. (2021). *You probably don't need OWL*. You Probably Don't Need OWL. Retrieved 6 May 2024, from https://www.bobdc.com/blog/dontneedowl/

Ehrlinger, L., & Wöß, W. (2016). Towards a Definition of Knowledge Graphs. *In Proceedings of the Posters and Demos Track of International Conference on Semantic Systems—Semantics 2016 and International Workshop on Semantic Change & Evolving Semantics, Leipzig, Germany*. https://www.researchgate.net/publication/323316736_Towards_a_Definition_of_Knowledge_Graphs

Elmhadhbi, L., Karray, M.-H., & Archimède, B. (2019). Toward the Use of Upper-Level Ontologies for Semantically Interoperable Systems: An Emergency Management Use Case. In K. Popplewell, K.-D. Thoben, T. Knothe, & R. Poler (Eds.), *Enterprise Interoperability VIII* (Vol. 9, pp. 131–140). Springer International Publishing. https://doi.org/10.1007/978-3-030-13693-2_11

Elsevier. (2017). *The dawn of predictive analytics to measure research performance: SciVal's Topic Prominence*. Elsevier Connect. Retrieved 23 Octobre 2023, from https://www.elsevier.com/connect/the-dawn-of-predictive-analytics-to-measure-research-performance-scivals-topic-prominence

Elsevier. (2021). *SciVal topic prominence | Elsevier*. Www.Elsevier.Com. Retrieved 23 February 2024, from https://www.elsevier.com/products/scival/overview/topic-prominence

Feilmayr, C., & Wöß, W. (2016). An analysis of ontologies and their success factors for application to business. *Data & Knowledge Engineering*, *101*, 1–23. https://doi.org/10.1016/j.datak.2015.11.003

Haller, A., & Polleres, A. (2020). Are we better off with just one ontology on the Web? *Semantic Web*, *11*(1), 87–99. https://doi.org/10.3233/SW-190379

Hofer, B., Casteleyn, S., Aguilar-Moreno, E., Missoni-Steinbacher, E., Albrecht, F., Lemmens, R., Lang, S., Albrecht, J., Stelmaszczuk-Górska, M., Vancauwenberghe, G., & Monfort-Muriach, A. (2020). Complementing the European earth observation and geographic information body of knowledge with a business-oriented perspective. *Transactions in GIS*, *24*(3), 587–601. https://doi.org/10.1111/tgis.12628

Hogan, A. (2020). *The Web of Data*. Springer International Publishing. https://doi.org/10.1007/978-3-030-51580-5

Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2022). A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(2), 494–514. https://doi.org/10.1109/TNNLS.2021.3070843

Kang, Y.-B., Krishnaswamy, S., Sawangphol, W., Gao, L., & Li, Y.-F. (2020). Understanding and improving ontology reasoning efficiency through learning and ranking. *Information Systems*, *87*, 101412. https://doi.org/10.1016/j.is.2019.07.002

Kejriwal, M. (2019). *Domain-Specific Knowledge Graph Construction*. Springer International Publishing. https://doi.org/10.1007/978-3-030-12375-8

Lemmens, R., Albrecht, F., Lang, S., Casteleyn, S., Stelmaszczuk-Górska, M., Olijslagers, M., Belgiu, M., Granell, C., Augustijn, E.-W., Pathe, C., Missoni-Steinbacher, E.-M., & Monfort Muriach, A. (2022).

Updating and using the EO4GEO Body of Knowledge for (AI) concept annotation. *AGILE: GIScience Series*, *3*, 1–6. https://doi.org/10.5194/agile-giss-3-44-2022

Lemmens, R., Lang, S., Albrecht, F., Augustijn, E., Granell, C., Olijslagers, M., Pathe, C., Dubois, C., & Stelmaszczuk-Górska, M. (2022). Integrating concepts of artificial intelligence in the EO4GEO body of knowledge. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIII-B4-2022*, 53–59. https://doi.org/10.5194/isprs-archives-XLIII-B4-2022-53-2022

Mascardi, V., Cordì, V., & Rosso, P. (2007). *A Comparison of Upper Ontologies*. 55–64. ISBN: 978-88-6122-061-4 http://woa07.disi.unige.it/papers/mascardi.pdf

Matthews, D. B. (2005). Semantic Web Technologies. *Semantic Web Technologies. E-Learning. 6.* https://www.researchgate.net/publication/30408878_Semantic_Web_Technologies

Microsoft. (2020). Visualizing the Topic hierarchy on Microsoft Academic. *Microsoft Research*. Retrieved 23 February 2024, from https://www.microsoft.com/en-us/research/project/academic/articles/visualizing-the-topic-hierarchy-on-microsoft-academic/

Microsoft. (2021). *Schema & Linked Dataset Descriptions – Microsoft Academic Knowledge Graph (MAKG)*. Retrieved 23 February 2024, from https://makg.org/schema-linked-dataset-descriptions/

Po, L., Bikakis, N., Desimoni, F., & Papastefanatos, G. (2020). *Linked Data Visualization: Techniques, Tools, and Big Data*. Springer International Publishing. https://doi.org/10.1007/978-3-031-79490-2

Polikoff, I. (2023). *Why I Don't Use OWL Anymore – TopQuadrant*. Why I Don't Use OWL Anymore. Retrieved 6 May 2024, from https://www.topquadrant.com/resources/why-i-dont-use-owl-anymore/

Ryen, V., Soylu, A., & Roman, D. (2022). Building Semantic Knowledge Graphs from (Semi-)Structured Data: A Review. *Future Internet*, *14*(5), 129. https://doi.org/10.3390/fi14050129

Saha, G. K. (2007). Web ontology language (OWL) and semantic web. *Ubiquity*, *2007*(September), 1–1. https://doi.org/10.1145/1295289.1295290

Schneider, L. (2003). How to Build a Foundational Ontology: The Object-Centered High-Level Reference Ontology OCHRE. In A. Günter, R. Kruse, & B. Neumann (Eds.), *KI 2003: Advances in Artificial Intelligence* (Vol. 2821, pp. 120–134). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-39451-8_10

Shinavier, J. (2009). *Position paper: Named Graphs in Linked Data*. https://www.w3.org/2009/12/rdf-ws/papers/ws25.pdf

Simperl, E. (2009). Reusing ontologies on the Semantic Web: A feasibility study. *Data & Knowledge Engineering*, *68*(10), 905–925. https://doi.org/10.1016/j.datak.2009.02.002

Stelmaszczuk-Górska, M. A., Aguilar-Moreno, E., Casteleyn, S., Vandenbroucke, D., Miguel-Lago, M., Dubois, C., Lemmens, R., Vancauwenberghe, G., Olijslagers, M., Lang, S., Albrecht, F., Belgiu, M., Krieger, V., Jagdhuber, T., Fluhrer, A., Soja, M. J., Mouratidis, A., Persson, H. J., Colombo, R., & Masiello, G. (2020). Body of knowledge for the earth observation and geoinformation sector—A basis for innovative skills development. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIII-B5-2020*, 15–22. https://doi.org/10.5194/isprs-archives-XLIII-B5-2020-15-2020

The Alan Turing Institute. (2020). *Knowledge graphs*. Knowledge Graphs: How Do We Encode Knowledge to Use at Scale in Open, Evolving, Decentralised Systems? Retrieved 12 February 2024, from https://www.turing.ac.uk/research/interest-groups/knowledge-graphs

Toppen, F., & Reinhardt, W. (2009). *A European approach towards the UCGIS Geographic Information Science and Technology Body of Knowledge a discussion paper*. https://www.researchgate.net/publication/268291133_A_European_approach_towards_the_UCGIS_Geographic_Information_Science_and_Technology_Body_of_Knowledge_a_discussion_paper

Vandenbroucke, D., & Vancauwenberghe, G. (2016). *Towards a new body of knowledge for geographic information science and technology.* ISSN: 1857-9000 EISSN: 1857-9019

Vidanelage, U. I. E. K. ([Forthcoming]). *EO4GEO BoK annotation of GI resources.*

W3C. (2009a). *SKOS Simple Knowledge Organization System Primer*. SKOS Simple Knowledge Organization System Primer. Retrieved 22 February 2024, from https://www.w3.org/TR/skos-primer/

W3C. (2009b). *SKOS Simple Knowledge Organization System Reference*. SKOS Simple Knowledge Organization System Reference. Retrieved 29 November 2023, from https://www.w3.org/TR/skos-reference/

W3C. (2014). *RDF 1.1 Turtle*. RDF 1.1 Turtle. Retrieved 6 May 2024, from https://www.w3.org/TR/turtle/

W3C. (2020). *JSON-LD 1.1*. A JSON-Based Serialization for Linked Data. Retrieved 6 May 2024, from https://www.w3.org/TR/json-ld11/

W3C. (2023). *RDF 1.2 Schema*. RDF 1.2 Schema. Retrieved 7 November 2023, from https://www.w3.org/TR/rdf12-schema/#acknowledgments-for-rdf-1-2

W3C. (2024). *RDF 1.2 TriG*. RDF 1.2 TriG. Retrieved 6 May 2024, from https://www.w3.org/TR/rdf12-trig/

Watkins, E. R., & Nicole, D. A. (2006). Named Graphs as a Mechanism for Reasoning About Provenance. In X. Zhou, J. Li, H. T. Shen, M. Kitsuregawa, & Y. Zhang (Eds.), *Frontiers of WWW Research and Development—APWeb 2006* (Vol. 3841, pp. 943–948). Springer Berlin Heidelberg. https://doi.org/10.1007/11610113_99

Appendices

Appendix A: Turtle serialization of the Ontology for Bodies of Knowledge defined in the OWL language.

The ontology is published to GitHub, due to size reasons, and can be seen by following this link: https://github.com/MPvliet/Thesis-GIMA-2023-2024/blob/main/Ontology-Development/Ontology-BoK.ttl

Appendix B: Turtle serialization of the Ontology for BoK applications defined in the OWL language.

The ontology is published to GitHub, due to size reasons, and can be seen by following this link: https://github.com/MPvliet/Thesis-GIMA-2023-2024/blob/main/Ontology-Development/Ontology-BoK-Applications.ttl

Appendix C: Python script that turns the EO4GEO BoK into RDF.

The script is published to GitHub, due to size reasons, and can be seen by following this link: https://github.com/MPvliet/Thesis-GIMA-2023-2024/blob/main/EO4GEO-BoK-Extraction/EO4GEO-BoK-to-KG.py

Appendix D: Python script that further processes the received JSON.

The script is published to GitHub, due to size reasons, and can be seen by following this link: https://github.com/MPvliet/Thesis-GIMA-2023-2024/blob/main/EO4GEO-BoK-Extraction/EO4GEO-BoK-to-KG.py

Appendix E: Figure 10 in TriG, a machine-readable format.

```
@prefix bibo: <http://purl.org/ontology/bibo/> .
@prefix boka: <http://example.org/BOKA/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix eo4geo: <https://bok.eo4geo.eu/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix obok: <http://example.org/OBOK/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix schema: <https://schema.org/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

eo4geo:concepts {
  eo4geo: a obok:BodyOfKnowledge ;
    rdfs:label "EO4GEO BoK" ;
    obok:hasKnowledgeArea eo4geo:AM,
      eo4geo:CF,
      eo4geo:CV,
      eo4geo:DA,
      eo4geo:DM,
      eo4geo:GC,
      eo4geo:GD,
      eo4geo:GS,
      eo4geo:IP,
      eo4geo:OI,
      eo4geo:PP,
      eo4geo:PS,
```

```
        eo4geo:TA,
        eo4geo:WB ;
      dcterms:description "A Body of Knowledge that describes the Geographic Information and Earth Observation domain." ;
      rdfs:seeAlso "https://bok.eo4geo.eu/" ;
      schema:dateModified "08/11/2023" ;
      schema:version "V7" .

   eo4geo:CF a obok:Concept,
        obok:KnowledgeArea ;
      rdfs:label "[CF] Conceptual Foundations" ;
      obok:conceptStatus "In progress (GI-N2K)" ;
      obok:hasRecommendedMaterial <http://books.google.com/books?isbn=9780198742845> ;
      obok:hasSkill eo4geo:4c58b2a9-4543-4399-aef1-e4e99d05f38c ;
      obok:isSubconceptOf eo4geo:GIST ;
      dcterms:description "The GIScience perspective is grounded in spatial thinking. The aim of this knowledge area is to recognize,
identify, and appreciate the explicit spatial, spatio-temporal and semantic components of the geographic environment at an ontological
and epistemological level in preparation for modeling the environment with geographic data and analysis. To do this, ... social constructs,
and the like." ;
      rdfs:isDefinedBy eo4geo: ;
      skos:notation "CF" .

   eo4geo:CF3 a obok:Concept ;
      rdfs:label "[CF3] Cognitive, linguistic and social foundations" ;
      obok:conceptStatus "In progress (GI-N2K)" ;
      obok:contributedBy <https://research.utwente.nl/en/persons/rob-lemmens> ;
      obok:hasRecommendedMaterial <http://books.google.com/books?isbn=9780226468044>,
        <http://books.google.com/books?isbn=9780792335955>,
        <http://dx.doi.org/10.1007/3-540-60392-1_1>,
        <http://dx.doi.org/10.1080/136588100415710> ;
      obok:hasSkill eo4geo:c2c25253-ab83-47c9-8b55-6688ee2ec04e ;
      dcterms:description "Geographic information is observed, comprehended, organized, used in human processes, with both personal
and social influences. Therefore, sound models of geographic information should be grounded on a sound understanding of human
perception, cognition, memory, and behavior, as well as human institutions." ;
      rdfs:isDefinedBy eo4geo: ;
      skos:inScheme eo4geo:CF ;
      skos:notation "CF3" ;
      skos:topConceptOf eo4geo:CF .

   eo4geo:CF3-1b a obok:Concept ;
      rdfs:label "[CF3-1b] Cognitive foundations" ;
      obok:conceptStatus "In progress (GI-N2K)" ;
      obok:contributedBy <https://research.utwente.nl/en/persons/rob-lemmens> ;
      obok:hasRecommendedMaterial <http://books.google.com/books?isbn=9780792335955>,
        <http://dx.doi.org/10.1080/136588100415710> ;
      obok:hasSkill eo4geo:4e621026-1043-4365-9d4d-b7c34fe2a2dc ;
      obok:isSubconceptOf eo4geo:CF3 ;
      dcterms:description "- Theories of human perception, cognition, and memory and their ability to model spatial knowledge acquisition
(e.g., Marr on vision, Piaget on cognitive development) - Types of mental representations (i.e., analogue, propositional, procedural) - ...
and GIS data representations thereof connections with cartography and maps" ;
      rdfs:isDefinedBy eo4geo: ;
      skos:inScheme eo4geo:CF ;
      skos:notation "CF3-1b" .

   eo4geo:4e621026-1043-4365-9d4d-b7c34fe2a2dc a obok:Skill ;
      obok:hasBloomsLevel 5 ;
      obok:isSkillFor eo4geo:CF3-1b ;
      dcterms:description "Explain the role of metaphors and image schemata in our understanding of geographic phenomena and
geographic tasks." .

   <http://books.google.com/books?isbn=9780226468044> a bibo:Document .

   <https://research.utwente.nl/en/persons/rob-lemmens> a obok:Contributor ;
      rdfs:label "Rob Lemmens" ;
      obok:hasContributed eo4geo:AM,
        eo4geo:AM1-2,
        eo4geo:AM10-1,
        eo4geo:AM10-2,
        eo4geo:AM10-3,
        eo4geo:AM11,
            ...
        eo4geo:CF6-3,
```

```
        eo4geo:CF6-4,
        eo4geo:CF7,
        eo4geo:CF7-2 ;
    dcterms:description "University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC)" ;
    foaf:name "Rob Lemmens" .

}
```

Appendix F: Figure 13 in TriG, a machine-readable format.

```
@prefix bibo: <http://purl.org/ontology/bibo/> .
@prefix boka: <http://example.org/BOKA/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix eo4geo: <https://bok.eo4geo.eu/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix obok: <http://example.org/OBOK/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix schema: <https://schema.org/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

eo4geo:concepts {
  eo4geo: a obok:BodyOfKnowledge ;
    rdfs:label "EO4GEO BoK" ;
    obok:hasKnowledgeArea eo4geo:AM,
      eo4geo:CF,
      eo4geo:CV,
      eo4geo:DA,
      eo4geo:DM,
      eo4geo:GC,
      eo4geo:GD,
      eo4geo:GS,
      eo4geo:IP,
      eo4geo:OI,
      eo4geo:PP,
      eo4geo:PS,
      eo4geo:TA,
      eo4geo:WB ;
    dcterms:description "A Body of Knowledge that describes the Geographic Information and Earth Observation domain." ;
    rdfs:seeAlso "https://bok.eo4geo.eu/" ;
    schema:dateModified "08/11/2023" ;
    schema:version "V7" .

  eo4geo:WB a obok:Concept,
      obok:KnowledgeArea ;
    rdfs:label "[WB] Web-based GI" ;
    obok:conceptStatus "In progress (GI-N2K)" ;
    obok:hasRecommendedMaterial
<https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/content/pdf/10.1007/s10109-004-0133-
4.pdf&casa_token=ahu2p72ykyIAAAAA:hxU9vtq7s3srHMLSANPfMha5CRqVBTGDwPwvEnpOIuGYpvi7kKSrtL4PO81mUUmY6AQtTPGnV_9A
ba8> ;
    obok:hasSkill eo4geo:5dbed0e8-a081-44ca-a589-119220cb648c,
      eo4geo:8994db32-310d-4afe-8f01-2300387dffdd ;
    obok:isSubconceptOf eo4geo:GIST ;
    dcterms:description "This knowledge area is about Web Based Geographic Information management aspects and therefore it was
given the name \"Web Based GI\" or \"WBG\" in short. ...  SA is covered by KA11 in as much as it should have been." ;
    rdfs:isDefinedBy eo4geo: ;
    skos:notation "WB" .

  eo4geo:WB4 a obok:Concept ;
    rdfs:label "[WB4] Resource Discovery" ;
    obok:conceptStatus "In progress (GI-N2K)" ;
    obok:hasRecommendedMaterial <https://doi.org/10.1080/13658816.2012.739692>,
      <https://doi.org/10.1109/TKDE.2007.250587> ;
    obok:hasSkill eo4geo:d4c22c8f-449e-4d46-9d7f-606dc5efeb56 ;
    dcterms:description "Resource discovery means the discovery of resources including data and services needed for an application.
Syntactic discovery refers to the discovery on the basis of syntactic comparison operations. It is classified as \"keyword-based\" and \"full-
text-based\" discovery. Semantic discovery on the other hand, refers to the discovery of resources on he basis of some semantic
definition. Therefore, semantic discovery requires that a resource be published by a semantic definition as defined in the topic WB3-5." ;
    rdfs:isDefinedBy eo4geo: ;
    skos:inScheme eo4geo:WB ;
    skos:notation "WB4" ;
    skos:topConceptOf eo4geo:WB .

  <https://doi.org/10.1109/TKDE.2007.250587> a bibo:Document .

  eo4geo:d4c22c8f-449e-4d46-9d7f-606dc5efeb56 a obok:Skill ;
    obok:hasBloomsLevel 5 ;
    obok:isSkillFor eo4geo:WB4 ;
```

```
    dcterms:description "Explain the differences between syntatic and semantic discovery of resources" .

  eo4geo:WB4-3 a obok:Concept ;
    rdfs:label "[WB4-3] Discovery over linked open data" ;
    obok:conceptStatus "In progress (GI-N2K)" ;
    obok:hasRecommendedMaterial <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.414.8933&rep=rep1&type=pdf> ;
    obok:hasSkill eo4geo:17a9f505-d823-44bd-90ba-9cb8aa24b2da,
      eo4geo:18692489-6e33-419f-9f45-d6b7754045c3,
      eo4geo:789fdecf-0c6b-404c-ada1-dcf8073c3848,
      eo4geo:b2d6b32c-4126-42bd-bb4c-f6fb739a7f93 ;
    obok:isSubconceptOf eo4geo:WB4 ;
    dcterms:description "Linked (open) data provides structured data which is interlinked in a machine readable way. This allows to
discover, access and combine data in an automatic way." ;
    rdfs:isDefinedBy eo4geo: ;
    skos:inScheme eo4geo:WB ;
    skos:notation "WB4-3" .
}

eo4geo:applications {
  eo4geo:8178cf63-461b-4165-b074-497e94dccb7f a boka:Expert ;
    rdfs:label "Martin Tomko" ;
    boka:authorOf <https://doi.org/10.5194/agile-giss-4-2-2023> ;
    boka:hasKnowledgeOf eo4geo:WB3-6,
      eo4geo:WB4-3 ;
    org:memberOf eo4geo:8eca2695-c64b-41c2-aa0f-b9542a0ba2a0 ;
    foaf:name "Martin Tomko" .

  eo4geo:8eca2695-c64b-41c2-aa0f-b9542a0ba2a0 a org:Organization ;
    rdfs:label "Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia" ;
    org:hasMember eo4geo:0b8ddb08-3635-4e59-9700-bb0183c808ee,
      eo4geo:6301988d-5b53-4d3e-a9a2-7ad6f5f9ad0b,
      eo4geo:7a5e98ce-1492-4750-ab1e-3cd7cbefd18e,
      eo4geo:8178cf63-461b-4165-b074-497e94dccb7f ;
    foaf:name "Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC 3010, Australia" .

  eo4geo:WB4-3 boka:personWithKnowledge eo4geo:0b8ddb08-3635-4e59-9700-bb0183c808ee,
      eo4geo:6301988d-5b53-4d3e-a9a2-7ad6f5f9ad0b,
      eo4geo:7a5e98ce-1492-4750-ab1e-3cd7cbefd18e,
      eo4geo:8178cf63-461b-4165-b074-497e94dccb7f,
      eo4geo:8c5ccb75-a340-4576-8b09-3dad194aebe0 ;
    boka:describedIn <https://doi.org/10.5194/agile-giss-4-2-2023> .

  <https://doi.org/10.5194/agile-giss-4-2-2023> a bibo:Report ;
    bibo:doi "https://doi.org/10.5194/agile-giss-4-2-2023" .
}
```

Appendix G: Evaluation of the knowledge footprint visualisations setup.

Linked data-based knowledge footprints a MSc thesis evaluation

Hi <name>,

I am mailing you to ask you to participate in the evaluation of my thesis. My thesis is part of the Geographical Information Management and Applications (GIMA) master's programme, a collaboration between Utrecht University, Delft University of Technology, University of Twente and Wageningen University. Rob Lemmens from the University of Twente is my primary supervisor.

The aim of my thesis is to answer the following question *"Can a framework that uses linked data for knowledge mapping and visualization effectively represent and compare knowledge footprints of geoscience experts based on the EO4GEO Body of Knowledge?"*

During the past months this framework is created in which the EO4GEO Body of Knowledge concepts (http://www.eo4geo.eu/bok/) are used to link expertise or knowledge to. This is done by processing the AGILE full paper publications from year 2021 – 2023 and leveraging NLP techniques to match papers to EO4GEO BoK concepts. In short you can then infer that the authors of a paper have knowledge of the EO4GEO concepts that are matched to that paper.

When this information is combined you can generate a knowledge profile or a so called "knowledge footprint" of a specific author or further aggregating to organisations and create organisational knowledge footprints. These knowledge footprints are created for the purpose of representing, sharing and promoting scientific and professional capabilities amongst organisations and people.

This evaluation comes with a couple of tasks. By following these tasks, you will be made aware of the various possibilities and use cases of these footprints. At the end of each tasks a couple of questions will be asked to evaluate the results. This evaluation is expected to take 45-60 minutes and includes nine questions, questions are coloured blue.

The following site offers interactive functionalities to generate these knowledge footprints and will be used in this evaluation: EO4GEO Knowledge Graph and Applications (mpvliet.github.io) The tasks and extra information can be found in the attachment.

Thank you for your time!

Best regards,
Mark van Vliet

Task 1: (Aim: getting familiar with the webpage.)
I would like you to pick a paper you contributed to that is published as one of the AGILE full paper publications (Volume 2-4 (2021-2023)) and keep the DOI available. You can choose from the following list, these are all the papers where you are one of the authors of:

-

After that I would like to show you how to create a knowledge footprint for this specific paper. To create this knowledge footprint, follow the following steps:

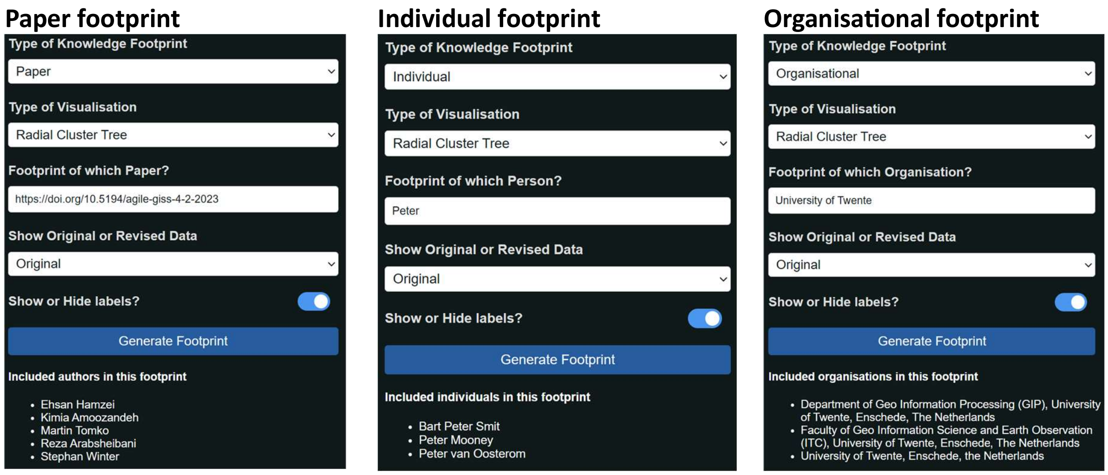
1. Go to [EO4GEO Knowledge Graph and Applications (mpvliet.github.io)]
2. In the form on the left side of the page enter the following parameters:
    - **Type of Knowledge Footprint:** Paper
    - **Type of Visualisation:** Radial Cluster Tree
    - **Footprint of which Paper?:** The DOI of the paper you choose.
    - **Show Original or Revised Data:** Original
    - **Show or Hide labels?**: Enabled
3. Click Generate Footprint.

Once the visualisation has loaded, you might have to zoom in or out, so the graph nicely fits the screen, and the circular graph becomes visible. After that lets go through the main interactive functionalities of the webpage.

The left side of the screen has two purposes, to configure and generate knowledge footprints by user input and to give information about the footprint or specific things in the graph. Directly under the "Generate Footprint" button, text will be shown indicating which entities are part of the knowledge footprint. The included entities differ based on the "Type of Knowledge Footprint" selection and are explained below.

- **Paper footprint:** All the authors that contributed to the paper that matches the DOI that the user inputted as parameter.
- **Individual footprints:** All the individuals/authors that matches the name you specified in the "footprint of which person" entry.
- **Organisational footprints:** All the organisations that are matched with the organisation entry.

Some examples below, be aware that your entry is used in a contains query. So, when you enter the name "Peter" multiple individuals are returned that contains the name "Peter". If you want to be more specific enter a more specific entry. Or use it to your advantage and generate footprints on multiple granular levels (e.g. create an organisational footprint of all organisations in "Spain").
You can use this detail section to validate that you are creating the correct footprint. The pictures below give some examples. Take a look at the text under the blue "generate footprint" button, to see which entities are part of the footprint.

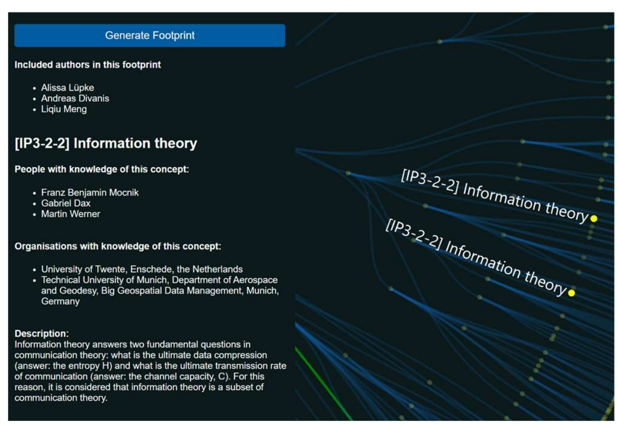


Then let's move to the created knowledge footprint. What you see is a radial or circular tree like structure, visualising hierarchical data. The yellow nodes depict all the EO4GEO concepts in the EO4GEO Body of Knowledge. By looking at the depth of the nodes and following the lines between the nodes the structure of parent- and subconcepts in the EO4GEO body of knowledge becomes visible. The centre or the root of the graph shows the highest concept level in the body of knowledge. Concepts at the outer bounds, or the leaf nodes of the graph indicate the highest depth in the body of knowledge.

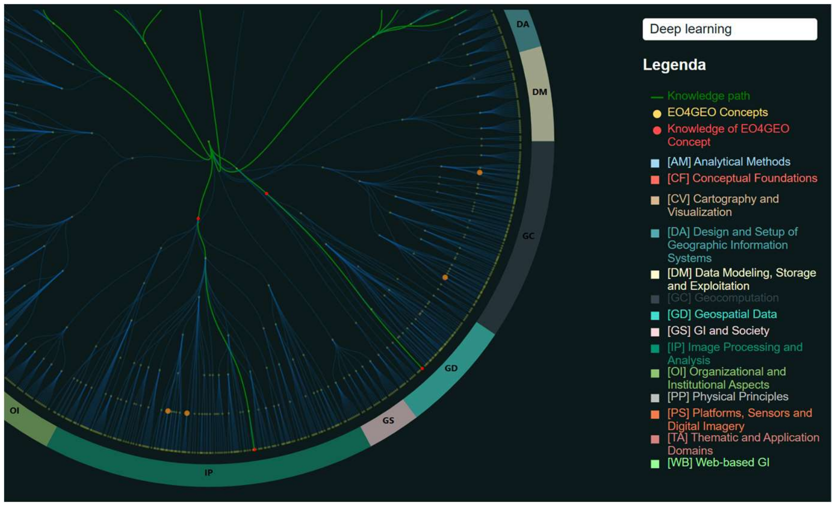
The doughnut chart around the footprint displays the fourteen knowledge areas in the EO4GEO BoK. It indicates or gives the viewer an idea in which knowledge area knowledge falls, without having to look at concept labels. The doughnut chart is labelled and further information can be seen in the legenda, on the right side of the screen.

With your mouse you can hover over all the nodes. Once your mouse is on a specific node it highlights and shows its concept name. On the left side of your screen it shows extra information, like the description of each concept.

Besides that, it also gives you information of all the individuals or organisations who are also knowledgeable about that specific concept.



In the top right corner of the page you find a search bar, which enables you to search for where certain EO4GEO concepts are placed in the visualisation. After typing the concept name and pressing enter the concepts that match the search string light up and receive an orange buffer. The highlight is visible for around 5 seconds. You might find duplicate nodes/concepts, this is due that concepts can have multiple parent concepts, and thus are also drawn multiple times in the graph. Note that the search bar is case sensitive. For example, I am searching for “Deep Learning”. The orange highlighted nodes indicate that these node names match the search string.



Nodes/concepts turn red once the paper/individual/organisation contains knowledge about that concept. The green lines indicate so called “knowledge paths” it makes all the parent nodes of a matched node visible and makes traversing to the root node possible. One can say that if you have knowledge about a specific node, you likely also have knowledge about the parent concepts.

At last, the “Knowledge Graph Management” tab, holds some functionalities to add new experts, new organisations, link persons to organisations and give you the possibility to add expertise to specific persons. This new information is written to a separate named graph/ data source and can be accessed in the footprints by configuring the ‘*Show Original or Revised Data*’ parameter to ‘*Revised’*. Feel free to play around and create your own footprints.

You have been involved with an earlier thesis evaluation from Upeksha, in which you were asked to annotate your AGILE paper with EO4GEO BoK concepts using the EO4GEO BoK annotation tool. I am using her results, the extracted EO4GEO concepts per AGILE paper, to create these knowledge footprints. The red nodes show these extracted EO4GEO concepts from your paper. But do not show the EO4GEO concepts you might have provided in your response.

Q1: Performing this task generates a knowledge footprint of one of your AGILE papers. Looking at this visualisation and the matched concepts, would you say this paper is correctly annotated with the right EO4GEO concepts.

Task 2: Make your personal AGILE footprint

Now that you have the gist of the footprints and its possibilities, the second task is to create your personal AGILE footprint. The personal footprint is an aggregation of all the EO4GEO concepts that are matched to AGILE papers you contributed too from year 2021 - 2023. This individual footprint can be created by following these steps:

1. Go to EO4GEO Knowledge Graph and Applications (mpvliet.github.io)
2. In the form on the left side of the page enter the following parameters:
   - **Type of Knowledge Footprint:** Individual
   - **Type of Visualisation:** Radial Cluster Tree
   - **Footprint of which Person?:** <your name>
   - **Show Original or Revised Data:** Original
   - **Show or Hide labels?** : Enabled
3. Click Generate Footprint.

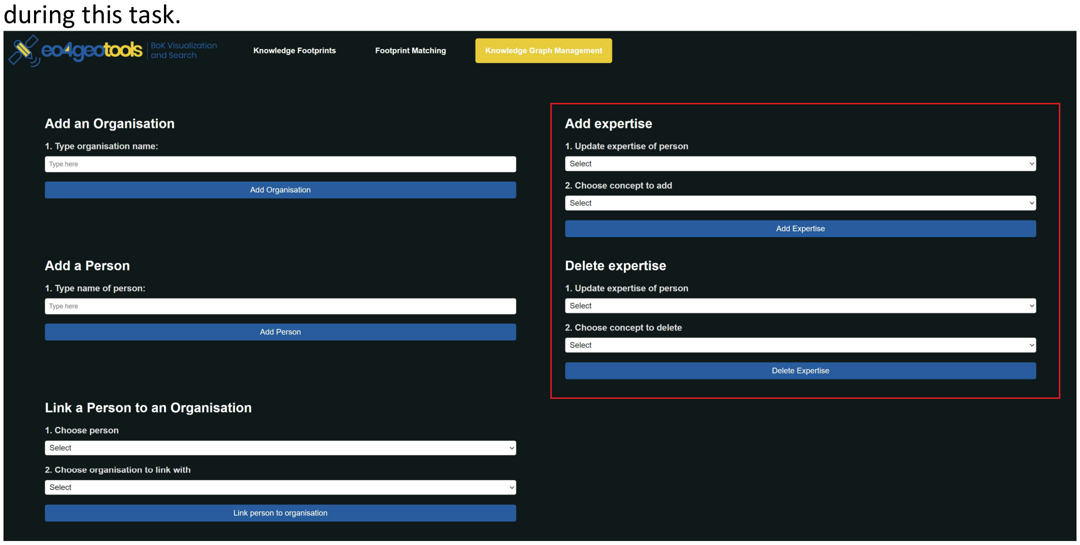
Q2: What are your first thoughts about the way knowledge is visualised in these knowledge footprints. Does it offer you new insights?

Task 3: Improve your personal AGILE footprint

Currently the footprint shows all the concepts that are extracted from your AGILE papers via NLP techniques. This process likely missed some concepts or might have matched your paper to concepts that are not related. I would like to ask you to make improvements to your personal AGILE footprint. Note: only make adjustments to your AGILE footprint that are related to work published to the AGILE association.

To do this follow the following steps:

1. Go to [EO4GEO Knowledge Graph and Applications (mpvliet.github.io)](mpvliet.github.io)
2. On the screen you see various forms, the "Add expertise" and "Delete expertise" are relevant during this task.



3. Select your own name under "Update expertise of person" and choose the concept you want to add or delete from your personal AGILE footprint under "Choose concept to delete"

You can view the improved footprint by making your personal footprint and setting the **Show Original or Revised Data** parameter to "revised".

Task 4: Make your organisational AGILE profile

In this task you will make an organisational footprint. An organisational footprint is the aggregation of all individuals that are a member of a specific organisation. (Paper -> Individual -> Organisational). This can be done by:

1. Go to [EO4GEO Knowledge Graph and Applications (mpvliet.github.io)](mpvliet.github.io)
2. In the form on the left side of the page enter the following parameters:
   - **Type of Knowledge Footprint:** Organisational
   - **Type of Visualisation:** Radial Cluster Tree
   - **Footprint of which Organisation?:** <Organisation name>
     - **Suggestion:** <suggestion>
   - **Show Original or Revised Data:** Original
   - **Show or Hide labels?** : Enabled
3. Click Generate Footprint.

Q3: How do you interpret this organisational footprint, can you explain what you see (e.g. is the organisation very specialised or has knowledge about a lot of general concepts?)

Q4: Could you say these organisational footprints effectively indicate in which field or fields an organisation is contributing in? Please explain briefly why or why not.

Task 5: Discover footprint matching

Task five is a little bit different and goes into a potential use case for these knowledge footprints. That is finding potential partners for collaboration in the GI field. I would like to take you along the following steps to create these insights:

1. Create an organisational footprint of your own organisation.
2. Find an EO4GEO concept that your organisation is knowledgeable about and which is a topic you would like to collaborate on with another organisation. (You can find organisations that also hold expertise about a specific concept upon hovering over the specific concept and checking the details section on the left side of the page)
3. Choose an organisation from the list and copy their organisation name.
4. Now go to the "Footprint Matching" tab or open this link, fill in the form and click generate footprint.
   - **Type of Knowledge Footprint:** Organisational
   - **Type of Visualisation:** Radial Cluster Tree
   - **Footprint for the first Organisation?:** <Your own organisation name>
   - **Footprint for the second Organisation?:** <The name of the organisation you would like to partner with>
   - **Show Original or Revised Data:** Original
   - **Show or Hide labels?** : Disabled

These four steps create a new type of knowledge footprint which combines the footprint of two entities, in this case two organisations. And tries to visualise where shared knowledge lays and where these organisations differ from each other. The visualisation does that by assigning a colour to each organisation and tries to visualise overlap/matches through this metro style like approach of visualising metro lines, but in our case, knowledge paths.

The green lines and green nodes belong to the first organisation and the yellow lines and nodes belong to the second organisation. Shared concepts (nodes) are coloured red and shared knowledge paths are drawn parallel to each other.

Q5: Could you see yourself using these matched footprints to effectively try and find out whether a certain organisation is a good fit to collaborate with? Please explain briefly why or why not.

Q6: Could you see these footprint offer you insights in what kind of knowledge an organisation might also have to offer, besides looking for the specific concepts. And could this help you in this process of finding potential collaborations?

Concluding questions:

Q7: What do you think of the way expertise is displayed in these knowledge footprints?

Q8: Would you say these knowledge footprints provide possibilities to improve geoscience collaboration? Please explain briefly why or why not.

Q9: What other potential use cases for these visualisations have come to mind?

Thank you for your time!
Mark van Vliet

Appendix H: An overview of other knowledge visualisations from other sources.

Elsevier's prominence map

These knowledge footprints can be compared to Elsevier's prominence maps or "Wheel of Science" in which they give an overview of topics an institution's researchers are active in (Elsevier, 2017). Figure H.1 shows an example of a prominence map of the Athena University. They will however differ in the level of detail, the origin of topics and how topics are represented. Wherein Elsevier uses their own topic classification and represent topics based on prominence. While the knowledge footprints in this thesis uses concepts from the EO4GEO BoK. How concepts will be visually represented in this thesis will be explored, for example how do we deal with weight of a concept?

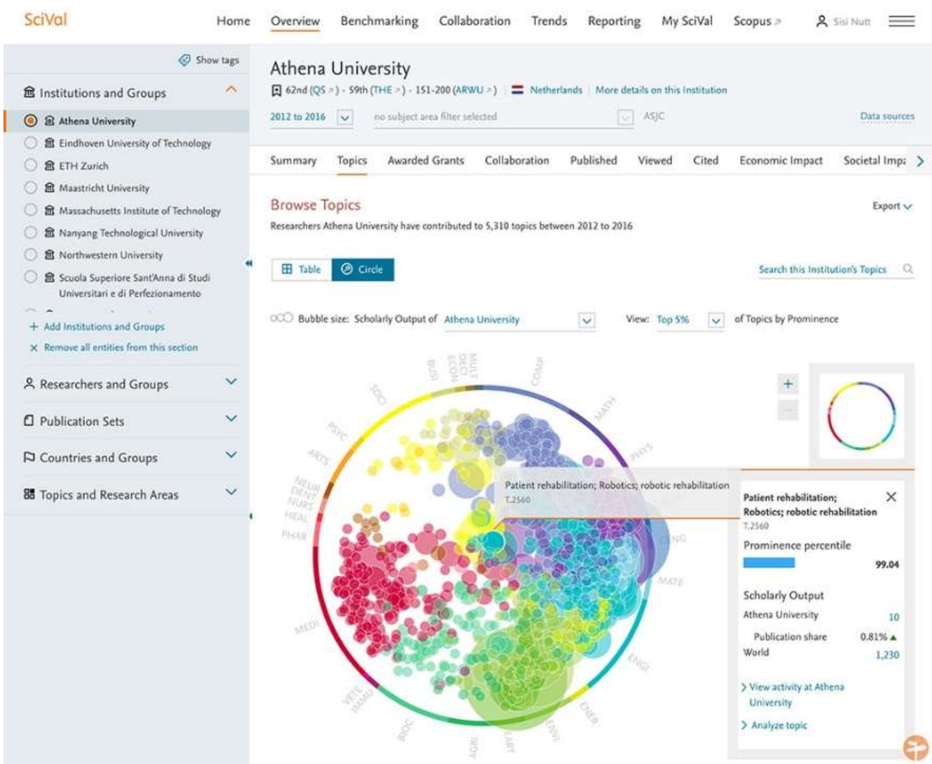


Figure H.1: Elsevier's Wheel of Science (Elsevier, 2017)

A limitation of this visualisation is the chosen granularity of the topics/categories. While it gives an abstract overview of multiple domains, it lacks depth once you want information of a specific domain.

EO4GEO Occupational Profiles

Figure H.2 shows another visualisation example, the output of the EO4GEO Occupational Profile Tool which is based on the concepts in the EO4GEO BoK. It shows which knowledge a test person has and in which knowledge area this falls.

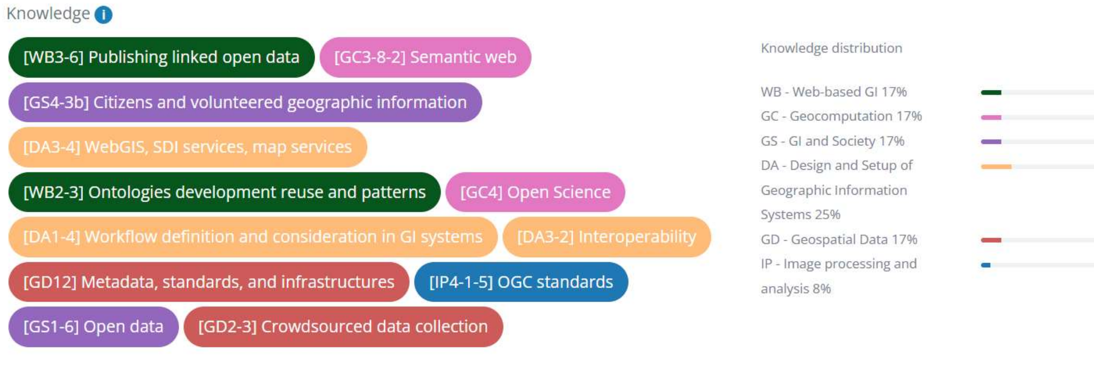


Figure H.2: Output of the EO4GEO Occupational Profile Tool (Source: https://eo4geo-opt.web.app/)

While it uses more in-depth topics, it is not directly clear how each topic relates to each other and how and how much they contribute to the knowledge distribution percentage. The use of colours is explored but it is not directly clear that they relate to the knowledge areas in the EO4GEO BoK. On the other hand, this test person shows knowledge about for example "[GC3-8-2] Semantic Web" couldn't you then infer that this person also has a bit of knowledge about "[GC3-8] Computational Linguistics" and even "[GC3] Artificial intelligence (AI) in EO and GI"?

Geospatial BoKMap explorer from BigKnowledge

Figure H.3 shows a base map about the Geospatial Technology domain. This base map is formed using text mining and machine learning on 100k domain artifacts (BigKnowledge LLC, 2023). It is again a different form of knowledge representation, and it looks like it uses a hierarchy of concepts to make concepts visual.

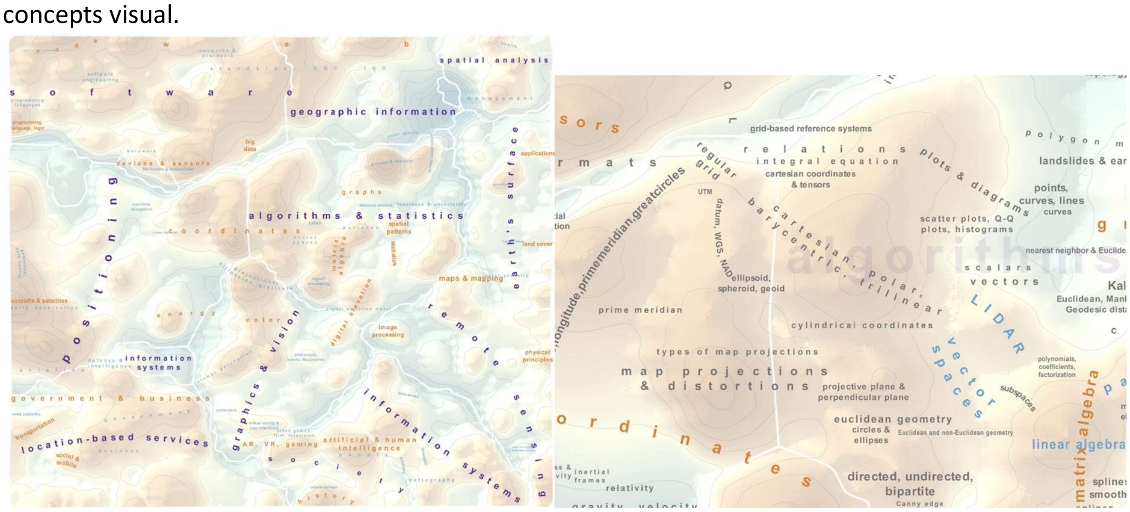


Figure H.3: BigKnowledge BoKMap Explorer (BigKnowledge LLC, 2021)

While it looks appealing, it is not directly clear how the use of height (visual representation) and contours are used to represent knowledge. I am uncertain what I can do with this information.

AI Knowledge Map (AIKM)

Figure H.4 illustrates the various AI paradigms and further categorizes them, showcasing how each form of AI is utilized today (Corea, 2019).

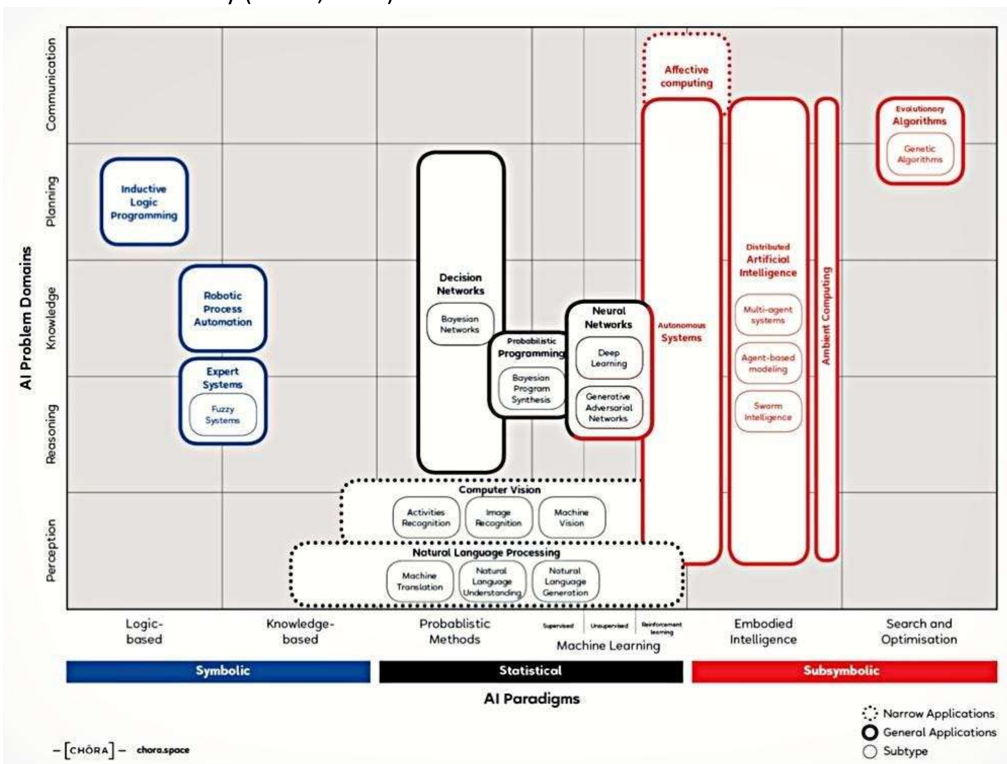


Figure H.4: The AI knowledge map (AIKM) (Copied from: (Corea, 2019))

While all the visualisations above are based upon knowledge areas, they all visually differ. It shows that each visualisation answers a different question and that there is not a single visualisation method that makes knowledge visual. Therefore, in this thesis multiple visualisations are explored and each answers a different question.

Appendix I: An overview of current knowledge footprint predecessors

This appendix shows a couple of other visualisations that are made via the D3 library and based on the EO4GEO knowledge graph. Note that they are all less developed.

Radial tidy tree

Quite like the current knowledge footprint, but how the depth of nodes is handled is different. The radial cluster tree has an equal depth for all leaf nodes, while Radial Tidy trees places nodes on their actual depth in the tree hierarchy.

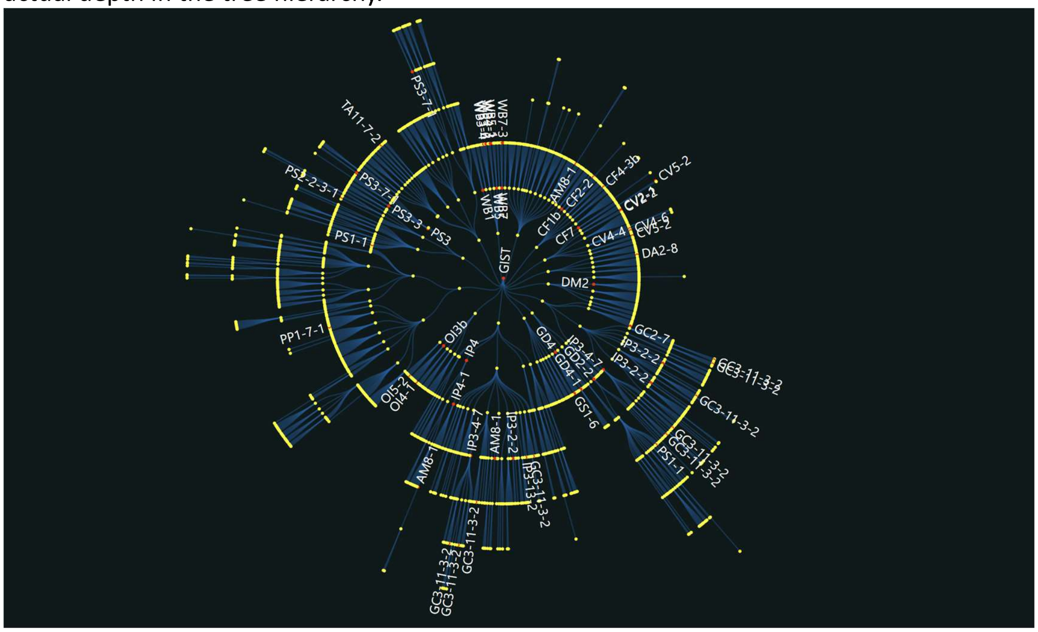


Figure !.1: Radial tidy tree (Source: Author)

Force directed trees

Force directed trees were also interesting and highly customisable. However further customisations were not pursued due to that each new creation of this visualisation generates different positionings for nodes. I was more looking at a stable basemap style background to build a visualisation on.

Figure !.2: Force directed tree (Source: Author)

Tree maps

Another method of visualising hierarchical data is using tree maps. During the thesis I investigated three variants. A zoomable tree map and a nested tree map. While interesting, I did not see much potential in showing personal or organisation expertise through these visualisations.

Figure I.3: Zoomable tree map. (Source: Author)

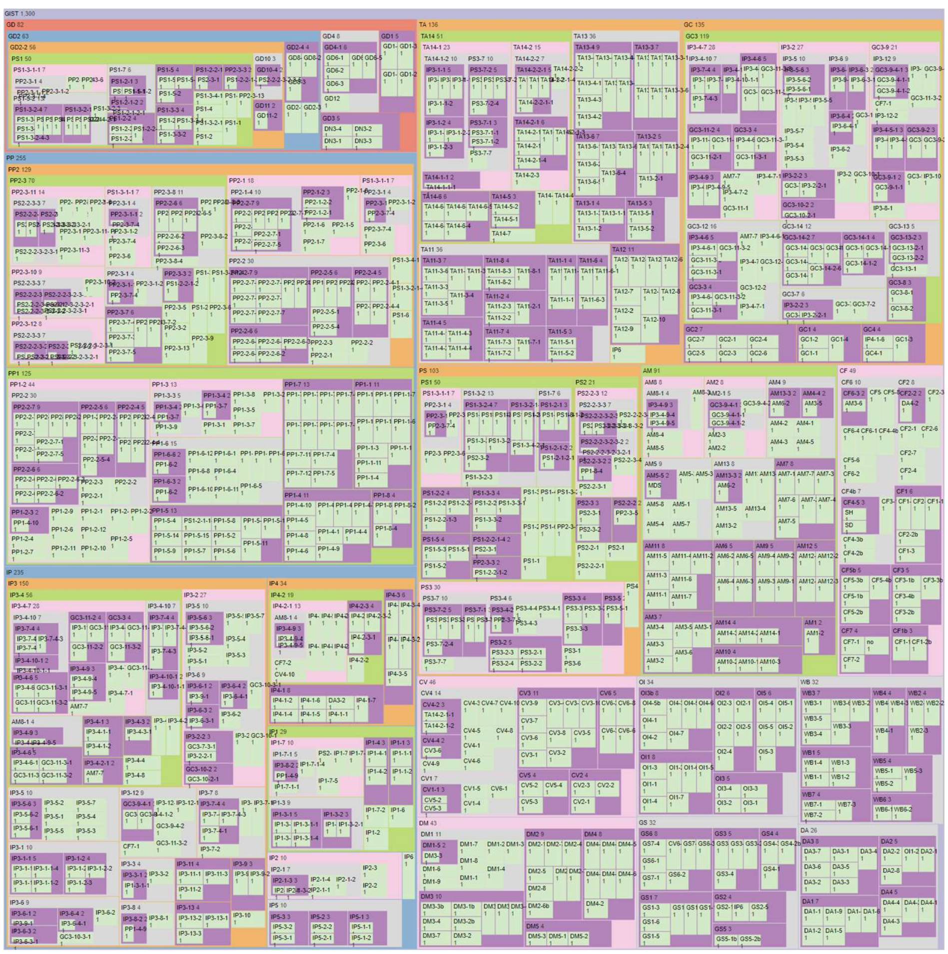


Figure I.4: Nested tree map. (Source: Author)

Circular packing

Another visualisation was the circular packing. EO4GEO BoK users are familiar with this visualisation as the original BoK visualisation webpage utilises this visualisation technique. The cool thing however is that this visualisation is created on RDF data, while the one on the BoK website uses a more tabular dataset, but still shows all the tree structures in the same way but a different position.

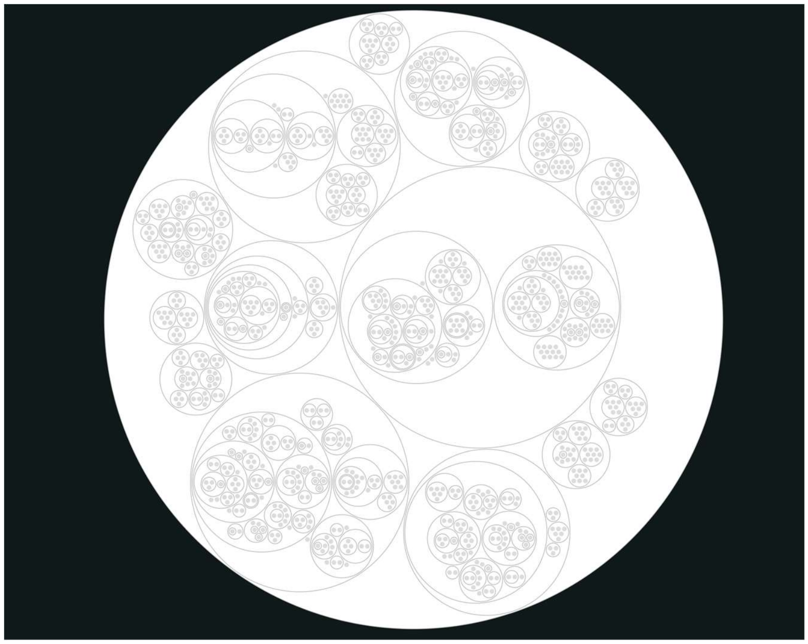


Figure I.5: Circular packing. (Source: Author)

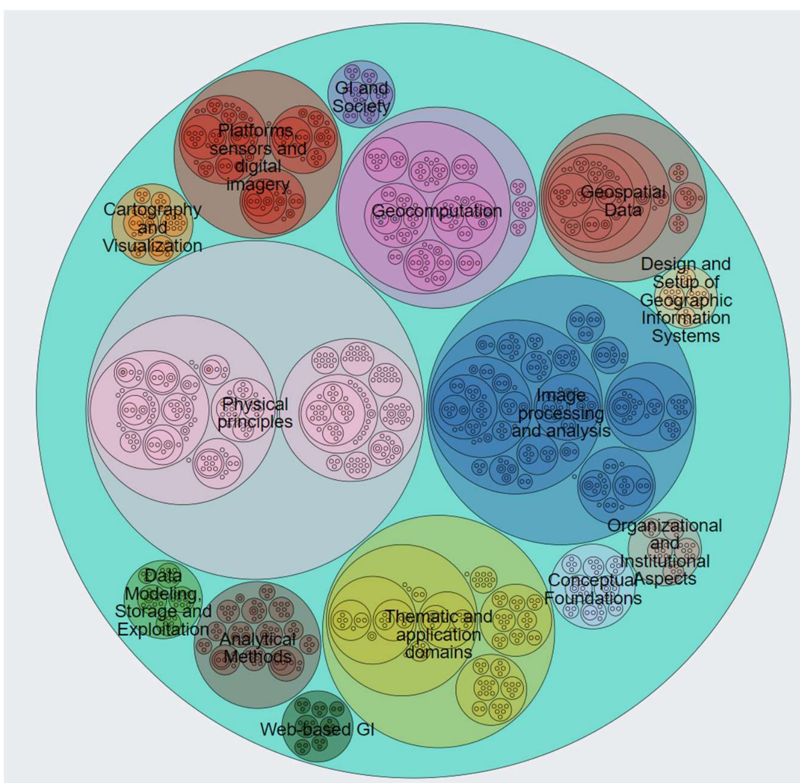


Figure I.6: Circular packing. (Source: https://bok.eo4geo.eu/)

Appendix J: An overview of SPARQL queries that query matched BoK concepts for organisations and papers.

Below are two different SPARQL queries that either create a knowledge footprint for a specific paper or for an organisation. Note that the BIND EXISTS statement is different and key to creating knowledge footprints for different entities.

```
PREFIX eo4geo: <https://bok.eo4geo.eu/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX bibo: <http://purl.org/ontology/bibo/>
PREFIX boka: <http://example.org/BOKA/>

SELECT ?conceptName ?childName ?conceptID ?childID ?nodeColour ?showLabel ?labelSize ?nodeValue FROM eo4geo:applications
FROM eo4geo:concepts
WHERE {
  {
    SELECT ?concept ?conceptName ?childName ?conceptID ?childID (IF(?knownByFirstEntity, 1 , 0 ) AS ?nodeValue) WHERE {
      ?concept rdf:type skos:Concept;
        rdfs:label ?conceptName;
        skos:notation ?conceptID.
      OPTIONAL {
        ?concept skos:narrower ?child.
        ?child rdfs:label ?childName;
          skos:notation ?childID.
      }
      BIND(EXISTS {
        ?paperURI rdf:type bibo:Report;
          bibo:doi ?DOIPaper.
        FILTER(CONTAINS(LCASE(STR(?DOIPaper)), LCASE("https://doi.org/10.5194/agile-giss-4-18-2023")))
        ?concept boka:describedIn ?paperURI.
      } AS ?knownByFirstEntity)
    }
  }
  BIND(IF(?nodeValue = 1 , "#FF0000", "#FFFF00") AS ?nodeColour)
  BIND(IF(?nodeValue = 1 , 16 , 0 ) AS ?labelSize)
  BIND(IF(?nodeValue = 1 , 1 , 0 ) AS ?showLabel)
}
ORDER BY (?conceptName)
```

Figure J.1: A SPARQL query to generate knowledge footprints for specific research papers.

```
PREFIX eo4geo: <https://bok.eo4geo.eu/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX org: <http://www.w3.org/ns/org#>
PREFIX boka: <http://example.org/BOKA/>

SELECT ?conceptName ?childName ?conceptID ?childID ?nodeColour ?showLabel ?labelSize ?nodeValue FROM eo4geo:applications
FROM eo4geo:concepts
WHERE {
  {
    SELECT ?concept ?conceptName ?childName ?conceptID ?childID (IF(?knownByFirstEntity, 1 , 0 ) AS ?nodeValue) WHERE {
      ?concept rdf:type skos:Concept;
        rdfs:label ?conceptName;
        skos:notation ?conceptID.
      OPTIONAL {
        ?concept skos:narrower ?child.
        ?child rdfs:label ?childName;
          skos:notation ?childID.
      }
      BIND(EXISTS {
        ?organisationURI rdf:type org:Organization;
          rdfs:label ?organisationName;
          org:hasMember ?membersOfOrganisationURI.
        FILTER(CONTAINS(LCASE(STR(?organisationName)), LCASE("University of Twente")))
        ?membersOfOrganisationURI boka:hasKnowledgeOf ?concept.
      } AS ?knownByFirstEntity)
    }
  }
  BIND(IF(?nodeValue = 1 , "#FF0000", "#FFFF00") AS ?nodeColour)
  BIND(IF(?nodeValue = 1 , 16 , 0 ) AS ?labelSize)
  BIND(IF(?nodeValue = 1 , 1 , 0 ) AS ?showLabel)
}
ORDER BY (?conceptName)
```

Figure J.2: A SPARQL query to generate knowledge footprints for specific organisations.

Appendix K: Evaluation responses

This appendix shows the raw responses of the evaluation participants. Note that all personal identifications are anonymised.

Participant A:

Q1: (I only compared the paper keywords with the concepts) It does match, yet the detected concepts are broader than what the paper covers

Q2: The question is unclear to me; for individuals searching themselves (you designed the task specifically for Prof. … , right?), there should be no surprise or new insight, but I can see how beneficial this could be for purposes like collaboration and networking. Maybe you could have mentioned in the task that people can find others with similar knowledge by hovering over the detected nodes.

Q3 and Q4: There is a balance between general and specialized concepts; however, since I am new to the faculty and haven't had the chance yet to get to know others and their work, I can not evaluate the effectiveness of the footprints. Maybe a comparison of the faculty footprint and generated footprint could be helpful:

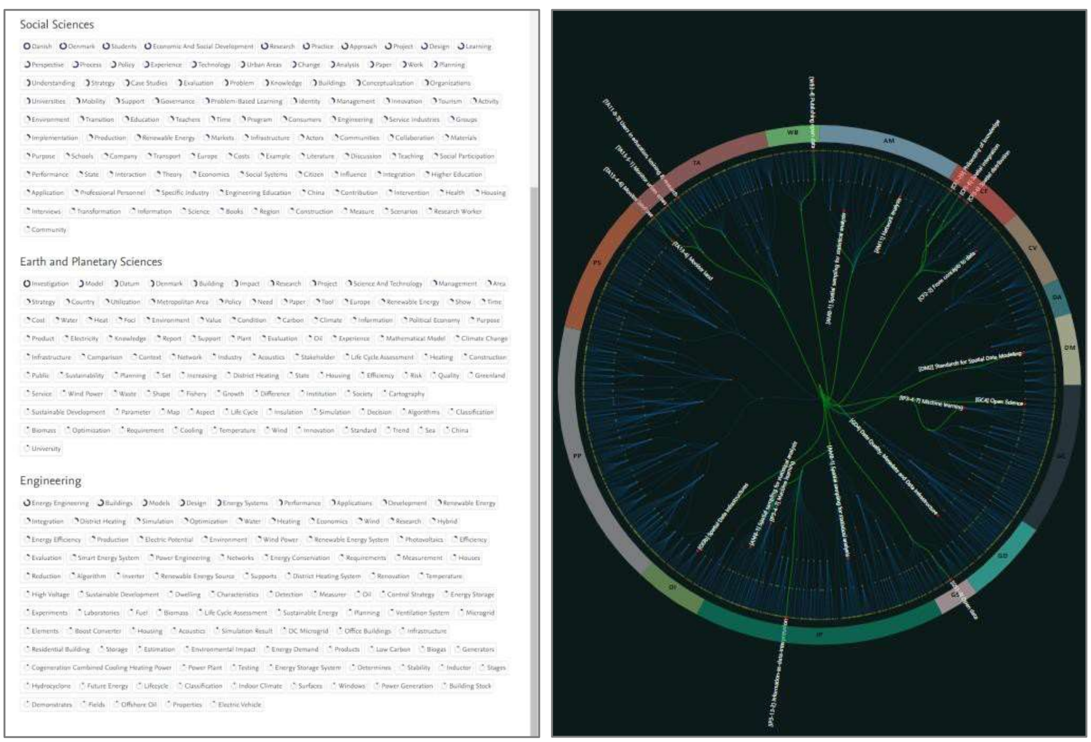


Figure K.1: left: https://vbn.aau.dk/en/organisations/institut-for-b%C3%A6redygtighed-og-planl%C3%A6gning/fingerprints/, right: https://mpvliet.github.io/index.html

The fields used to categorize knowledge in the left picture differ from the topics from EO4GEO, so a direct comparison is impossible. There could be more fields than shown in the right picture, but I can not say that with certainty.

Q5: Yes, but I would not base my decision solely on that (based on the importance of the task for which I need collaborators), but this is a good starting point.

Q6: I believe yes

Q7: It is easy to read, understand and follow

Q8: As explained in Q5, this could be a good start; however, first, there should be an investigation on motives and approaches for collaboration; in my experience, many seek collaboration with the ones they know in person or have already worked with or through recommendation, such tools although helpful can not completely replace the human aspect of what people seek in a collaboration. Also, some have other criteria and might, for example, choose reputation over specific domain expertise.

Q9: For the current version, nothing comes to mind; however, other functionalities can be built on top of this. It would be helpful if a list of related papers would also appear on the left side for each node. It could also be beneficial if some default options would show, e.g., most popular topics, most cited papers, most active people or organizations, etc.

Participant B:
Q1: note from author: question 1 was not asked to this person.

Q2: Machine learning seems to be listed double. For me, agent-based modelling is the most important aspect but it is not more visible. I do not recognise myself in the other topics.

Q3: I do not know how to answer this question.

Q4: This is difficult for me to judge. I do not know how well the actual domains the organisation is working on are represented in the graph.

Q5: I would have like another type of visualisation showing other institutes that work on ABMs.

Q6: I cannot say that I often look for other organisations working on other topics. I would however like to know organisations working on the same topic

Q7: it is not easy to read the diagrams. All topics seem equally important which they are not from my perspective.

Q8: I do not think they will provide possibilities for improved collaborations

Q9. See my earlier comment on finding people that work on the same topics

Participant C:
Hi Mark
Here is my evaluation: Congratulations on the hard work and best wishes.

Q1 : the concept « Semantic Discovery » was not found while fully relevant. The concept data quality is relevant but it is associated to image processing in the framework and our paper did not consider image data.
The concept SDI was good. Standards for Spatial Data Modeling : should rather be standards for metadata. There is no machine learning in the paper while the concept arose.

Q2 : done. I liked browsing the concepts despite it would be easier that some concepts are grouped because the interface requires many clicks

Q3 : It is nice to cope with the different labels for one organisation and maybe it could also be done at the stage where one need to select an organisation to relate to. I liked looking at the footprint. Yet i would be more interested at comparing footprints between different organisations at this stage.

Q4 the footprint does not reflect the scope of LASTIG lab nor Université Gustave Eiffel but only probably the AGILE perspective on it ?

Q5 : I could use such tool indeed, rather to look for papers to read at first

Q6 : In theory yes but it depends a lot on the source. And then also I imagine a chatbot interface could be more efficient than the one we have. It could be used to recommend organisations to look at.

Q7 : It was not clear the way the full classification was achieved but I liked the general idea and think it is promising to help us look at our domain that way

Q8 : To identify similarities and complementarities between papers, to improve the bidding process

Q9 : Help in exchange of master students

Participant D:
Hi Mark,
Q1: note from author: question 1 was not asked to this person.

Q2: What are your first thoughts about the way knowledge is visualised in these knowledge footprints. Does it offer you new insights?
I like the visualisation as presented. I think that this gives a good idea of the knowledge in GIScience. What threw me off at the beginning is that a concept may be present in several categories, e.g., machine learning may be found in several nodes. But I also understand the reasoning behind it.

Q3: How do you interpret this organisational footprint, can you explain what you see (e.g. is the organisation very specialised or has knowledge about a lot of general concepts?)
In general, yes. I am actually unsure if the concepts (e.g. deep learning) always occur at the same level within the graph - if not then this might be an indication for some standardisation that is needed.

Q4: Could you say these organisational footprints effectively indicate in which field or fields an organisation is contributing in? Please explain briefly why or why not.
If you accept the categories (e.g., data modelling or geospatial data) as given, then the organisational footprints indicate the contributions of the organisation. However, I cannot see if any given concept has received more than one publication. This could be indicated e.g., through the use of width of the lines - the wider the line the more publications in this area.

Q5: Could you see yourself using these matched footprints to effectively try and find out whether a certain organisation is a good fit to collaborate with? Please explain briefly why or why not.
Yes, I can see myself exploring the matched footprints for the purpose of finding potential collaboration partners. However, collaboration is more than fitting research interest's. So, I would use the footprints for exploration.

Q6: Could you see these footprint offer you insights in what kind of knowledge an organisation might also have to offer, besides looking for the specific concepts. And could this help you in this process of finding potential collaborations?

Yes, similar answer than to Q4. However, my critical comments concerning width or weight of a line apply here as well.

Q7: What do you think of the way expertise is displayed in these knowledge footprints?

Well, expertise to me is something I can apply and act on. Knowledge is something different from action. It might be the basis for the action, but by (philosophical) nature, it is rather passive than active or actionable.

Q8: Would you say these knowledge footprints provide possibilities to improve geoscience collaboration? Please explain briefly why or why not.

Opportunities for collaboration may certainly be improved, but this also depends on the nature of the collaboration. If I intend to write a paper together and I am looking for someone knowledgeable in the same field, then your visualisation will help. If, however, I am writing a grant for the European Union, then I will need a complementary partner, i.e. a group that does research in areas where I do not do so myself. In this case your tool would probably not be able to help.

Q9: What other potential use cases for these visualisations have come to mind?

Once it contains all base data (i.e., not only AGILE papers), it could be used for matching reviewers for paper reviews or finding workshop organisers…

I would certainly want to play around with it and produce a kind of personal profile to be put up on the webpage. It is also a good method for critically reflecting on your own research contributions that have been published in contrast to research done but not published and this gives you an inkling about how others might view you. It could also be interesting to produce a time series…

I hope that my short answers helped. I wish you all the best with your thesis.

Kind regards,