From Spatial Data Infrastructures To Spatial Knowledge Infrastructures – Evaluating the current state of concepts and cases

MSc. GIMA Master's Thesis
Thesis Report

Raggy Minten (1702629)
r.o.y.minten@students.uu.nl
MSc. Geographical Information Management and Applications (GIMA) Utrecht University

October, 2024
Wageningen University & Research (WUR)

Day-to-day Supervisor: dr. Jaap-Willem Sjoukema (WUR)
Responsible Professor: prof. dr. ir. Lammert Kooistra (WUR)

Abstract

This study provides a first-time systematic review into the concept of a Spatial Knowledge Infrastructure (SKI) in comparison with the well-established Spatial Data Infrastructure (SDI) concept. Generalized, an SDI is a data infrastructure where spatial data can be accessed by simple queries whereas the SKI should provide knowledge in the form of ready to go answers on questions regarding the spatial data available in the infrastructure. The systematic literature review on SKI results in four topics that were used to compare the concept with SDI: the definition, objectives, components, and architecture. Per topic, the identified differences between SKI and SDI could be classified into two perspectives: a data perspective and a user perspective. This resulted in a conceptual framework consisting of a set of parameters – data parameters and user parameters – that can be used to evaluate whether an infrastructure can be defined as SDI, SKI or something in-between. In total, 14 parameters were identified of which seven were considered to be a data parameter and seven to be a user parameter. The data parameters included the standards, producer, storage, update method, update frequency, spatial dimension, and temporal dimension. The user parameters were the level of expertise, query input, output suitability, output readiness, device readiness, analytics, and modeling.

In the second part of the study, two real-world case studies, the Kadaster Knowledge Graph and 3D Amsterdam were evaluated with the use of the set of parameters. For most of the parameters, at least one of the cases was classified as in-between SDI or SKI, or SKI-ready. With respect to the parameters update frequency and device readiness, both cases were still considered to be an SDI. Based on the findings, it can be concluded that they partially meet the criteria for being an SKI in order to fulfill present needs in the spatial domain, although it must be stressed that both cases were still under development and not fully operational. Besides, some bottlenecks already existing in SDI are not addressed. For instance, the dependency on the data quality and external data sources. The novelty of the SKI concept is reflected by the available scientific literature as the concept is predominantly discussed from a bird's eye view. Therefore, further research is advised to parse the concept in-depth.

In conclusion, SDIs are moving towards the envisioned SKI and the proposed set of parameters can serve as guidance for SDIs to improve into a knowledge-based system that is meeting current trends in automation and modeling.

Keywords: Spatial Data Infrastructure; SDI; Spatial Knowledge Infrastructure; SKI; spatial data; spatial knowledge; geospatial information

Preface

I present to you my thesis entitled '*From Spatial Data Infrastructures To Spatial Knowledge Infrastructures – Evaluating the current state of concepts and cases'*. This study was conducted between September 2023 and September 2024 at Wageningen University and Research (WUR) as part of the joint-degree master's program Geographical Information Management and Applications (GIMA).

Doing research besides a full-time job and having a move in the middle of the research cannot be recommended if you want to stay away from stress and sleep deprivation, but completing this research gives me quite the satisfaction.

Let me express my gratitude to Jaap-Willem Sjoukema, my day-to-day supervisor from WUR, for his dedication as he guided me through the sometimes challenging process and provided excellent insights during the research.

With a bit of luck, future GIMA students will find this thesis helpful when they first come across the concept of a Spatial Data Infrastructure (or SDI) and have no idea what kind of sorcery that is.

Sincerely,
Raggy Minten

Table of Contents

List of figures

List of tables

List of abbreviations

2D	Two-dimensional
3D	Three-dimensional
4D	Four-dimensional
AI	Artificial Intelligence
AR	Augmented Reality
DIKW pyramid	Data–Information–Knowledge–Wisdom pyramid
EU	European Union
GI	Geo-Information
GIS	Geographical Information System
GMES	Global Monitoring Environmental and Security
IoT	Internet of Things
ISO	International Organization for Standardization
KBS	Knowledge-Based System
KG	Knowledge Graph
KKG	Kadaster Knowledge Graph
LD	Linked Data
OGC	Open Geospatial Consortium
PO	Product Owner
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
SDI	Spatial Data Infrastructure
SDM	Spatial Data Mining
SoS	System of Systems
SWS	Spatial Web Services
SKI	Spatial Knowledge Infrastructure
UEM	Usability Evaluation Method
UI	User Interface
VGI	Volunteered Geographic Information
W3C	World Wide Web Consortium

1. Introduction

The need for digitalization plays a key role in today's society. The ever growing availability and flow of data offer various possibilities. In the domain of geo-information (GI), the use of spatial data fulfills the growing interest in automation and predictive modeling (Jobst Markus & Gartner, 2019; Olsson et al., 2018) and the demand for near-real-time monitoring of processes on earth (Zudilin & Iralieva, 2021). Therefore, the availability and use of spatial data could lead to new insights and developing knowledge (Omidipoor et al., 2021).

1.1 Problem statement

The way – spatial – data potentially lead to knowledge is demonstrated by the data–information–knowledge–wisdom (DIKW) pyramid (Figure 1). It shows that information can be retrieved from data, information can be used to increase knowledge and knowledge leads to the pinnacle of the pyramid: wisdom (Jobst Markus & Gartner, 2019; Rowley, 2007; Sjoukema, 2021).

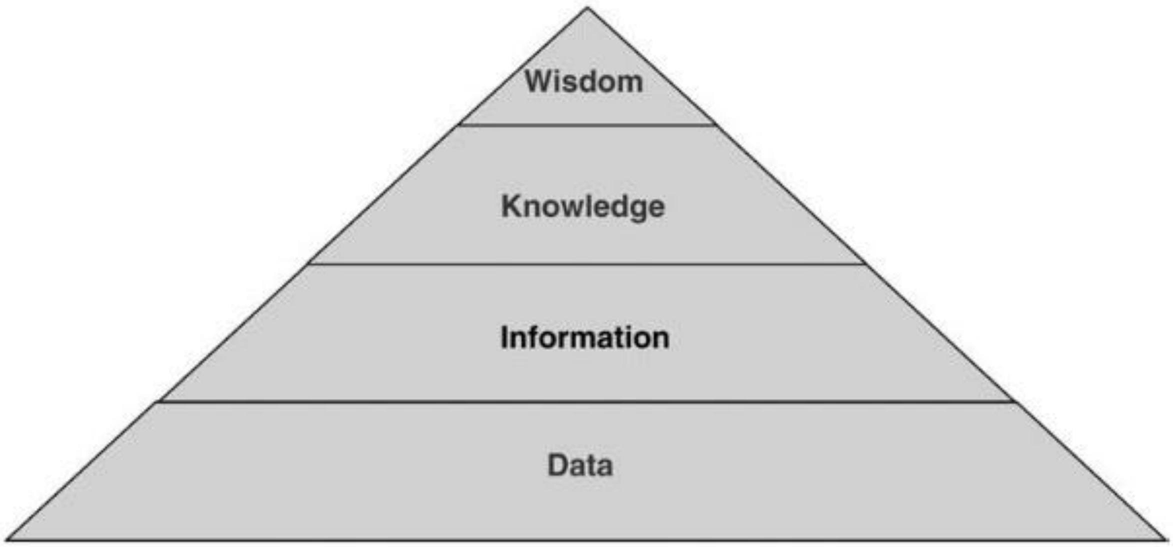

Figure 1 - Data-Information-Knowledge-Wisdom (DIKW) pyramid. Source: Rowley, 2007

In this ordering could knowledge be seen as aggregated information that gives answers to questions. If knowledge is available, it would not be necessary to dive into different information sources or the underlying data to find an answer to a question. Theoretically, knowledge is defined as meaningful information whereas information is contextualized data (Guigoz, 2016; Nativi et al., 2019).

Technological developments make it easier to derive such knowledge. In recent years, the introduction of Artificial Intelligence (AI) chatbots as mainstream tools provided people increasingly with these ready to use answers. However, a similar evolution has not been taking place in the geospatial domain yet. At present, for example, a policy advisor with a spatial question at hand need to go to a GI-expert who uses spatial data for retrieving information which will be formed into an answer for the advisor. Technological advancements could potentially result in the policy advisor using a system where the question is the input and the output will be a ready to use (correct) answer to that question. This meets the criterion for being knowledge because the user directly has its answer without the need for further processing information. In the geospatial domain, the Spatial Knowledge Infrastructure (SKI) is introduced to provide this spatial knowledge. The SKI is the successor to the well-established Spatial Data Infrastructure (SDI).

The SDI as a concept originated in the mid-1980s, but gained momentum in the early 1990s and is still relevant in present times (Coetzee et al., 2021; Rajabifard, Feeney & Williamson, 2002). Over the course of 30 years, the concept has evolved hugely. This evolution is characterized by different generations and objectives, although one of the latest generations, the user-centric SDI, does not seem to be implemented in practice (Sjoukema, 2021). In general, user-issues are still identified in SDIs

(Duckham et al., 2017; UN-GGIM, 2022). Open SDIs aim, among other objectives, to overcome these issues (Vancauwenberghe & van Loenen, 2018). However, despite these evolutions, the SDI is usually still considered to be an GI-expert system (Granell, Díaz & Gould, 2010; McEachen & Lewis, 2023) having the objectives to enhance the availability, integration, access, and sharing of spatial data (Arnold et al., 2021; Hendriks, Dessers & Van Hootegem, 2012). For getting knowledge, data still needs to be processed to provide answers on questions people have.

Therefore, as successor of the SDI a new concept has been introduced, called the Spatial Knowledge Infrastructure (SKI) (Arnold et al., 2021; Duckham et al., 2017; Woodgate et al., 2019). Although the SKI is introduced in recent years, concepts being called 'knowledge' systems have been in practice for over more than 30 years and started with the concept of a Knowledge-Based System (KBS) (Arinze & Banerjee, 1992).

The first KBS consisted of a Geographical Information System (GIS) on the one hand where the 'knowledge base' was in fact an expert system driven by manual input from a qualified user (Figure 2) (Fischer, 1994; Lee, Wu & Wei, 2008; Zhu, Healey & Aspinall, 1998). The manual input that serves as the 'knowledge base' is considered to be rule-based and derives on the expert user as they have the knowledge about desired outcomes (Lee, Wu & Wei, 2008; Sikder, 2009). After the GIS processed these rules on the data, the output still have to be evaluated and validated by the user (Sikder, 2009).

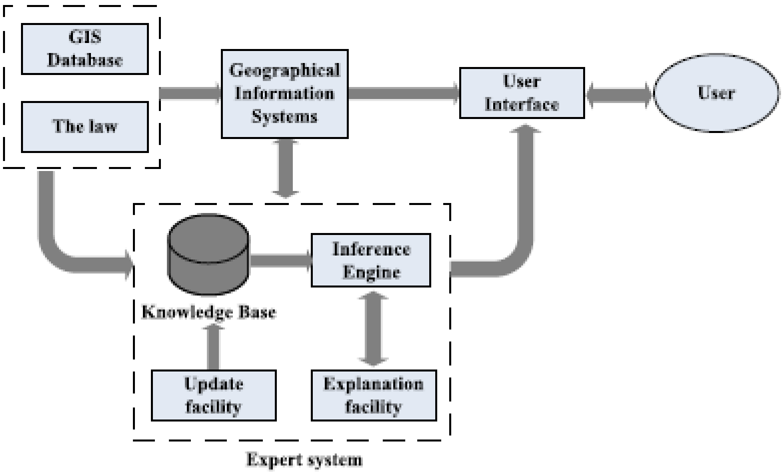

Figure 2 - Framework of a Knowledge-Based System (KBS). Source: Lee, Wu & Wei, 2008

About 25 years after the KBS, the SKI is the new spatial 'knowledge' concept. As the successor of SDI, the SKI aims for "*automatically creating, sharing, curating, delivering, and using knowledge (not just data or information)*" (Duckham et al., 2017, p.4). The SKI should be a response to the growing digitalization and latest developments in computer science by integrating GI approaches and technologies (Arnold et al., 2021; Coetzee et al., 2021; Duckham et al., 2017). Keeping the definition of knowledge in mind as used in this research, the SKI should provide ready to use answers on spatial data questions.

The novelty of the concept can be seen when looking to the SKI from a practical perspective. Figure 3 and Figure 4 represent two different views on the concept SKI and how the SKI differs from the SDI. The figures show the views of Duckham et al. (2017) and Geospatial World (2022) regarding the desired capabilities of an SKI. In general, it could be said that both articles address the same developments like making the infrastructure usable for non-geospatial users, emphasizing on looking forward (predictive) instead of backwards, having real-time or dynamic data, and they both state it should be in a 4D representation. On the other hand, differences can be seen such as whether mobile devices should be supported. This is one of the key points in Figure 3, but is not mentioned in Figure 4.

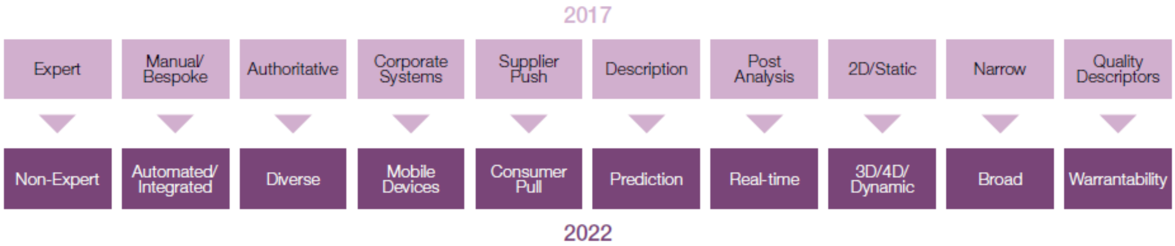

Figure 3 - Differences in capabilities when moving towards a spatial knowledge infrastructure. Source: Duckham et al., 2017

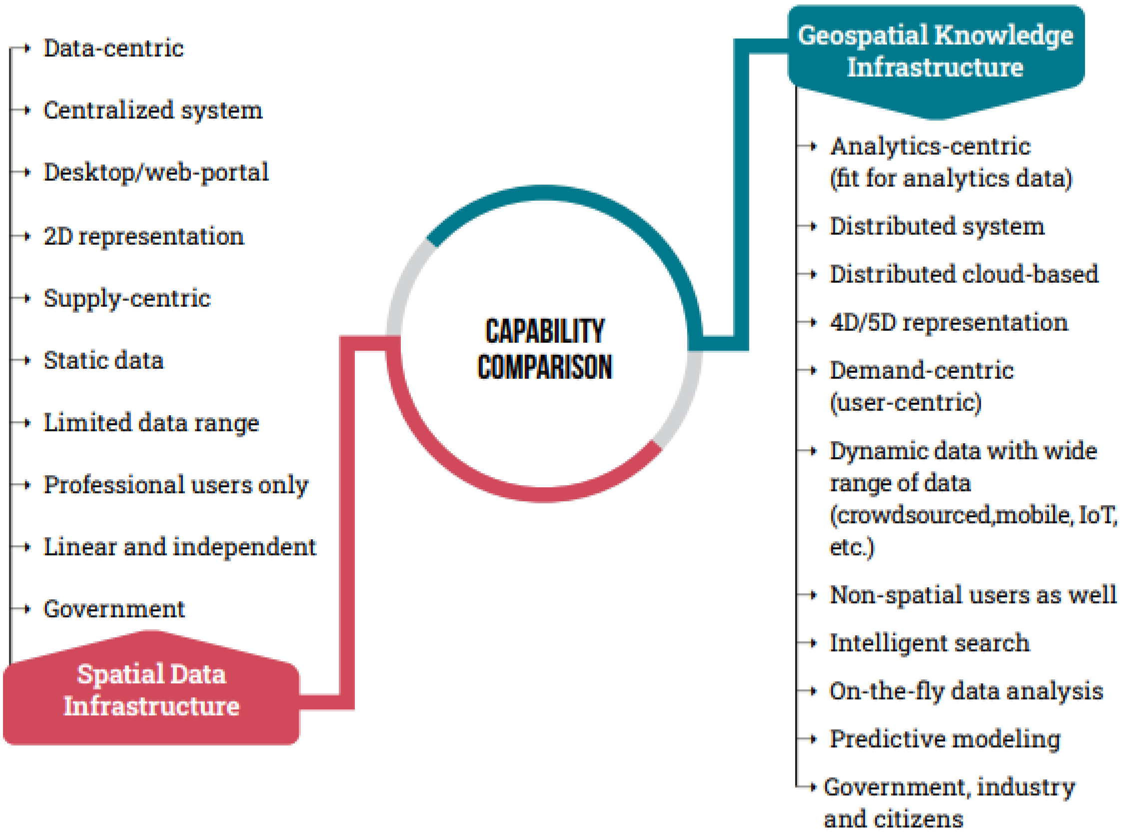

Figure 4 - From spatial data infrastructure (SDI) to spatial knowledge infrastructure (SKI) – comparison of its principles. Source: Geospatial World, 2022

All these differences must have an impact on the technicalities. Figure 5 shows the differences between an SDI and SKI based on a high-level architecture. As can be seen, the architecture will have to change for the greater part as well. Data has to be published as Linked Data (LD) and there are different applications for the different user groups, which are all classified as 'Knowledge On-demand applications'. Besides, the infrastructure is connected with the Web of Data. The data of the infrastructure is also published on the Web of Data.

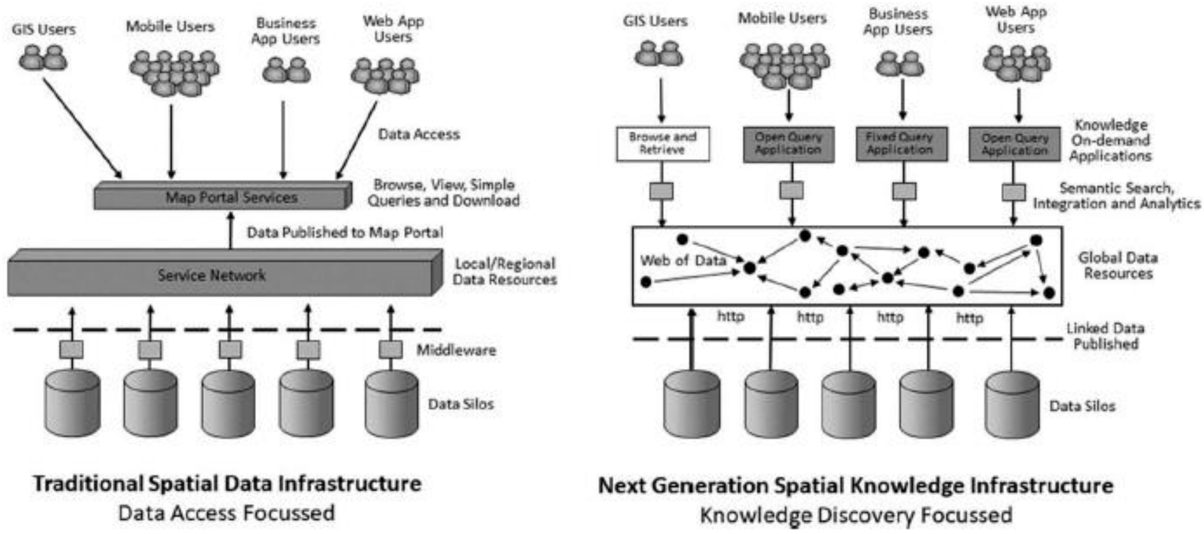

Figure 5 - Schematic representation of the differences between a spatial data infrastructure and a spatial knowledge infrastructure. Source: Arnold et al., 2021

Both the practical perspective as well as the technical perspective on what an SKI is highlight the differences with an SDI, at least according to the authors (Arnold et al., 2021; Duckham et al., 2017; Geospatial World, 2022). Besides, there is to some extent consensus of the concept and what should be included in both its goals (knowledge creation) and characteristics (e.g., semantic technologies, predictive). However, Figures 3, 4 and 5 also highlight existing differences within the literature. In addition, some of the characteristics are presented as being SKI, while SDI literature also recognize these developments. This makes it useful to study the concept of SKI to get an overall view of the concept and to identify to what extent the SDI is already evolving into an SKI.

1.2 Research Objectives

The previous paragraph emphasizes there is no overall consensus on the SKI concept as well as it is viewed from different angles. Therefore, the aim of this research is to provide a comprehensive and concise overview of the current literature on the SKI concept. Furthermore, the aim is to identify the differences between an SKI and SDI by providing a comparison between the concepts.

Reflecting to Duckham et al. (2017), they assumed the SKI would be in place in 2022 (Figure 3), which was the year Geospatial World (2022) published their proposal of the SKI (Figure 4) that is to some extent comparable to the proposed SKI by Duckham et al. in 2017. This raises the question whether the SKI is just a concept or development that exists in theory but not in practice. This makes another objective to evaluate to what extent contemporary spatial infrastructures are including characteristics assumed to be SKI instead of SDI. As can be seen, the term ‘spatial infrastructure’ has been used. This term will be used throughout this thesis for analytical purposes. Where this term is being used, it is meant to describe the infrastructure without trying to give a value judgement on whether it is considered to an SDI or SKI. To evaluate developments that are taking place in the GI-domain in practice, two case studies will be conducted on current projects that are focusing on spatial data.

1.2.1 Case studies

The first initiative to be studied is the Kadaster Knowledge Graph (KKG) by the Kadaster. The Kadaster is the Dutch cadastral organization that provides and manages various official governmental spatial datasets, like the cadastral registration and national reference coordinate. The aim of the KKG is to ‘*support the user with the use of integrated data*’ by linking different data upfront so that the user does not have to integrate, or link, the data themselves (Kadaster, n.d.b.). The KKG also presents different use cases such as a chatbot, ‘Self-Service GIS’ and an augmented reality (AR) app. These different use cases all have aspects linked to an SKI. The chatbot, Loki, should be able to answer questions from non-experts using layman terms (Ronzhin et al., 2019). ‘Self-Service GIS’ is still under development but is supposed to be a tool that provide ‘*nonexpert users with the opportunity to perform simple geospatial analysis and querying tasks on the web and, where possible, real-time data*’ (Bucher et al., 2021, p10). The AR-app should give you real-time information of the buildings by pointing your camera towards the building (Rowland, Folmer & Baving, 2022). When comparing the objectives of the KKG with the first insights in what an SKI is, it becomes clear that there are some resemblances. For instance the characteristics of (1) non-spatial users, (2) an intelligent search, (3) on-the-fly-data analysis and (4) dynamic data with wide range of data objectives of an SKI as well.

The second initiative that will be evaluated is 3D Amsterdam, which is a project of the municipality of Amsterdam. 3D Amsterdam aims to simulate what effects possible interventions have in the digital world before carrying them out in the real world (3D Amsterdam, n.d.b). Even though this is the aim of the case, it describes the current as ongoing and that it '*consists of a 3D-model of the city, various functionalities and an interactive web viewer*' (3D Amsterdam, n.d.b). Again, various SKI-aspects can be identified based on the objectives of the project: (1) 3D/4D representation, (2) predictive modeling, (3) non-spatial users and (4) on-the-fly data analysis.

When looking at the objectives and presumed SKI-characteristics, it can be seen that the projects do have similarities (like the non-spatial users), but also seem to have other focus points. Where the KKG focuses on Linked Data and a chatbot that provides ready to use answers, 3D Amsterdam intents to have a geoportal including 3D data and some form of analytics. The different objectives of the two case studies make them complementary to each other as it will be interesting to see how both relate to the SKI-concept as a whole.

1.2.2 Research questions

For realizing the research objectives, three research question are formulated. One main research question and two sub-questions. The main research question is as follows:

"To what extent does the concept of a spatial knowledge infrastructure differ from the concept of a spatial data infrastructure and to what extent are the Kadaster Knowledge Graph and 3D Amsterdam moving towards being a spatial knowledge infrastructure?"

The first sub-question is expressed as:

"To what extent does a spatial knowledge infrastructure differ from a spatial data infrastructure in terms of its characteristics?"

This question will be broken down into different components, which are (1) describing what an SDI is, (2) describing what an SKI is and (3) making a comparison between the two concepts. Answering this question will be part of the theoretical framework as this part serves as the scientific base of the study and elaborates on the theories and concepts dealt with. This leads to the second sub-question of this study:

"How could the Kadaster Knowledge Graph and 3D Amsterdam be classified as type of spatial infrastructure considering the characteristics of SDI and SKI?"

The second sub-question will assess some real-world spatial infrastructures based on the differences identified by answering sub-question 1. Two case studies will be analyzed: the Kadaster Knowledge Graph and 3D Amsterdam. The analysis has as main goal to describe what the characteristics of the cases are and put this into the concepts of SDI and SKI to evaluate to what extent the spatial infrastructure is an SDI, SKI or something in-between.

2. Methods

The research can be divided in two parts: (1) a (systematic) literature study on the two leading concepts SKI and SDI and (2) applying two case studies to these concepts. The objective of the literature study is to find all relevant literature that discusses the concept of SKI in order to form a complete overview of the concept's characteristics. While SDI is a much debated and evolving concept, the outcomes of the systematic review on SKI will form the basis for the literature search on SDI. Based on both the characteristics of SKI and SDI, it will be possible to evaluate whether a spatial infrastructure is either an SDI, SKI or something in-between. The case studies have, as the name says, the research design of a case study. While a case study typically involves only one case (Bryman, 2012), it could involve a few cases (Odoh & Chinedum, 2014) as is being done in this research. This section will discuss the research methods that will be applied when analyzing the cases.

2.1 Literature study

To get a complete overview of the research on the novel SKI concept, this part of the literature study will follow a systematic approach. A systematic approach has two main advantages of which one is that the explicit procedures enhance the possibility of reproducibility limiting bias and the other is that it is seen as an evidence-based approach (Bryman, 2012).

A well-known method for conducting a systematic review is the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) protocol (Page et al., 2021). PRISMA provides a set of guidelines in order to structure the search for and analyzing of existing research (Alkhaleel, 2024; Page et al., 2021; Rahman & Szabó; 2021). The guidelines of the PRISMA can be divided into three stages: article identification, screening and inclusion (Page et al., 2021). In the identification stage the search strategy will be described and duplicate articles are removed. The screening stage firstly requires that the title and abstract will be read which could lead to exclusion of the article. Secondly, articles that are still included are sought for retrieval and when accessible fully read to assess for eligibility (Page et al., 2021). After the assessment, the list with included articles is definite which brings the inclusion stage to an end.

Advantages of using the PRISMA protocol are to promote the transparency of the study and to efficiently and effectively search and report on the research done on the topic (Alkhaleel, 2024; Page et al., 2021; Rahman & Szabó, 2021). After the completion of the PRISMA protocol, a synthesis will be done to conclude the systematic review. The synthesis can be either author-centric or concept-centric (Klopper, Lubbe & Rugbeer, 2007; Schryen, 2015). An concept-centric synthesis is recommended instead of the author-centric synthesis and is commonly used when research on the topic is at an early stage (Cavallaro & Nocera, 2022; Schryen, 2015). While research into SKIs is assumed to be at an early stage and the synthesis is meant to identify clusters of information that will be the basis for the SKI framework, the synthesis will be concept-centric. In a concept-centric synthesis, a concept matrix will be presented. The concept matrix gives the researcher a clear overview of the topics discussed and which article discusses which of the topics (Klopper, Lubbe & Rugbeer, 2007).

The outcomes of the systematic review on SKI will serve as the basis for the literature review on SDI, meaning that the topics identified in the concept-matrix will be used as input in the search strategy. The main reason for including the topics of the concept-matrix is to refine the results of the search and find articles that are supposed to be suitable for a one to one comparison of the SDI and SKI concepts.

A first search in the Scopus database on "Spatial Data Infrastructure" resulted in 2,077 documents found. When adding one of the topics of the SKI concept matrix (Table 2) as refinement to the search, numbers were still varying between 81 and 289 documents per topic. Because of the limited time of this study and the amount of literature, it is not considered feasible to follow a systematic approach like the PRISMA protocol for the literature review on SDI. This means that this part of the literature study will follow a narrative approach. In a narrative review there are no standardized guidelines to be followed, making the scope of the literature study more broad (Bryman, 2012). A narrative approach has limitations in respect with a systematic review in terms of the reproducibility and relevant studies could be excluded or missed, which could lead to bias (Rahman & Szabó, 2021). However, these limitations are accepted given the outcomes of the preliminary search to SDI literature.

In order to enhance reproducibility and transparency of the literature study on SDI, guidelines of a systematic approach are partly included in the SDI literature review. Both the search strategy as well as the concept matrix adhere to the guidelines, but the screening and including or excluding of documents do not as this is not considered feasible.

2.1.1 Systematic literature review of SKI

The first step when executing the PRISMA protocol is to define the search strategy after which the screening takes place. The screening consists of three steps: excluding articles after reading the title and abstract, retrieving all articles that are still considered useful and assessing the retrieved articles for eligibility. After the assessment, the total number of included articles is established.

2.1.1.1 Search strategy SKI literature

In the first step, the search strategy, it is defined which databases are included in the search for literature, which search terms are being used and which search criteria are being used. As a start, a regular Google search has been performed with the search term 'Spatial knowledge infrastructure' as it will provide the author with first insights about the concept. Besides, this search could identify grey literature that would not be found otherwise (Atkison & Cipriani, 2018; Bramer et al., 2017; Haddaway et al., 2015). As this search is not part of the PRISMA protocol it is not included in the flow diagram, but it did provide one new source, which is Geospatial World (2022).

As for the databases, Scopus, Web of Science, ScienceDirect and Google Scholar are identified in the literature as useful sources for finding scientific literature (Alkhaleel, 2024; Coetzee & Wolff-Piggott, 2015; Korkou, Tarigan & Hanslin, 2023; Rahman & Szabó; 2021). Although it is advisable to use multiple databases, it is a time-consuming process making researchers limit the number of databases included in a systematic review (Bramer et al., 2017). In general, studies include two databases. Reasons to exclude certain databases are a limited numbers of articles or overlap between databases (Alkhaleel, 2024). Scopus is included in all studies, as it is considered to be the most extensive database (Coetzee & Wolff-Piggott, 2015). Google Scholar is advised to only use in addition to other databases as the disadvantage is that the search results are not necessarily reproducible and therefore go against the nature of a systematic review (Atkinson & Cipriani, 2018; Bramer et al., 2017). The advantage of Google Scholar is to find grey literature (Atkinson & Cipriani, 2018; Haddaway et al., 2015). In this case, all aforementioned databases are included as the assumption is to find a limited number of articles and therefore including more databases should lead to a better coverage of available literature on SKI.

The terms searched for were as following: "Geospatial Knowledge Infrastructure" OR "Spatial knowledge infrastructure" OR "Geospatial Knowledge Infrastructures" OR "Spatial knowledge infrastructures". The choice for including both 'Spatial' and 'Geospatial' is a result of the Google search, as it showed both writings are apparent in the literature (Duckham et al., 2017; Geospatial World, 2022). Coetzee & Wolff-Piggott (2015) include related terms like 'Geoportal' in their systematic review into SDIs, but in this case the choice was made to only search for the fully written concept. A search into the abbreviation "SKI" lead to articles about skiing and was therefore inadequate. Doing a more extensive search including key words found at articles on SKI in the first search, like "Semantic Web" lead to either articles in only that domain (e.g. Long, Shelhamer & Darrell, 2015) or, in combination with "Spatial Data Infrastructure" to papers that propose additions to SDI (e.g. Ulutaş Karakol & Cömert, 2022). As the main objective is to find literature specifically on the concept of SKI, the aforementioned search terms are deemed the most efficient.

As the first step of the screening is to read the title and abstract of the article, the search criteria were to search in the title, abstract and keywords. This kind of search was possible for Scopus, Web of Science (although it was called a search for 'Topic' in that database) and ScienceDirect. Due to its nature, a search in Google Scholar varies from the regular databases. A regular search without adjustments would lead to high numbers of unsuitable information (Haddaway et al., 2015), including articles that only cite the search terms (Atkison & Cipriani, 2018). To confirm this, a first search was done like this. This resulted in 176 hits, where a first scan confirmed that already on page 2 unsuitable articles or citations without reference to an article could be seen, Therefore, the box for 'Citations included' was switched off. Also, as recommended by Haddaway et al. (2015), the search was limited to only look in the title of the article. This can be done by using the option ‘advanced search’. This resulted in a search consisting of ‘Find articles with at least one of the words’ AND ‘where my words occur in the title of the article’. Google Scholar does not provide an option to search within the title, abstract and keywords. It is either anywhere in the article, or just in the title.

The search is carried out on February 26, 2024 and provided a total of 29 articles across the four databases. Of the 29 articles, 11 were removed due for being duplicates, making a total of 18 articles fit for screening. An overview of the strategy can be found in Table 1. The table follows the order of search, starting with Scopus and ending with Google Scholar. A complete overview of the PRISMA flow diagram can be seen in Figure 6.

Database	Search term	Search criteria	Number of articles found	Duplicates
Scopus	"GeoSpatial Knowledge Infrastructure" OR "spatial knowledge infrastructure" OR "GeoSpatial Knowledge Infrastructures" OR "spatial knowledge infrastructures"	Search within Article title, Abstract Keywords	11	n.a.
Web of Science	"GeoSpatial Knowledge Infrastructure" OR "spatial knowledge infrastructure" OR "GeoSpatial Knowledge Infrastructures" OR "spatial knowledge infrastructures"	Topic	4	4
ScienceDirect	"GeoSpatial Knowledge Infrastructure" OR "spatial knowledge infrastructure" OR "GeoSpatial Knowledge Infrastructures" OR "spatial knowledge infrastructures"	Title, abstract or author-specified keywords	1	1
Google Scholar	"GeoSpatial Knowledge Infrastructure" OR "spatial knowledge infrastructure" OR "GeoSpatial Knowledge Infrastructures" OR "spatial knowledge infrastructures"	Advanced search with at least one of the words AND where my words occur in the title of the article. Citations not included	13	7

Table 1 - Search strategy for the PRISMA protocol on literature of spatial knowledge infrastructure

2.1.1.2 Screening of articles SKI literature

In the first stage of the screening, articles were included when one of the search terms was mentioned in either the title or the abstract of the article. In case the search term was only mentioned in the key words, the abstract should either make a reference to SDI or otherwise give direction that the article is about a spatial infrastructure. In this stage, two articles were excluded. Thus, 16 articles were sought for retrieval. Of the 16 articles, another two were not retrieved due to inaccessibility. Therefore, 14 articles were assessed for eligibility.

2.1.1.3 Selection of articles SKI literature

The second stage of the screening requires to fully read the remaining articles for assessment. To be included in this study, articles should contribute to the concept of SKI, be it in terms of its origin, a definition, objectives, principles or technical requirements. Articles referring to other research without providing additional insights are excluded, as well as articles that only link to SKI's but not discuss it. In this step, 3 articles were excluded. Thus, a total of 11 articles were included in this study as fit for purpose.

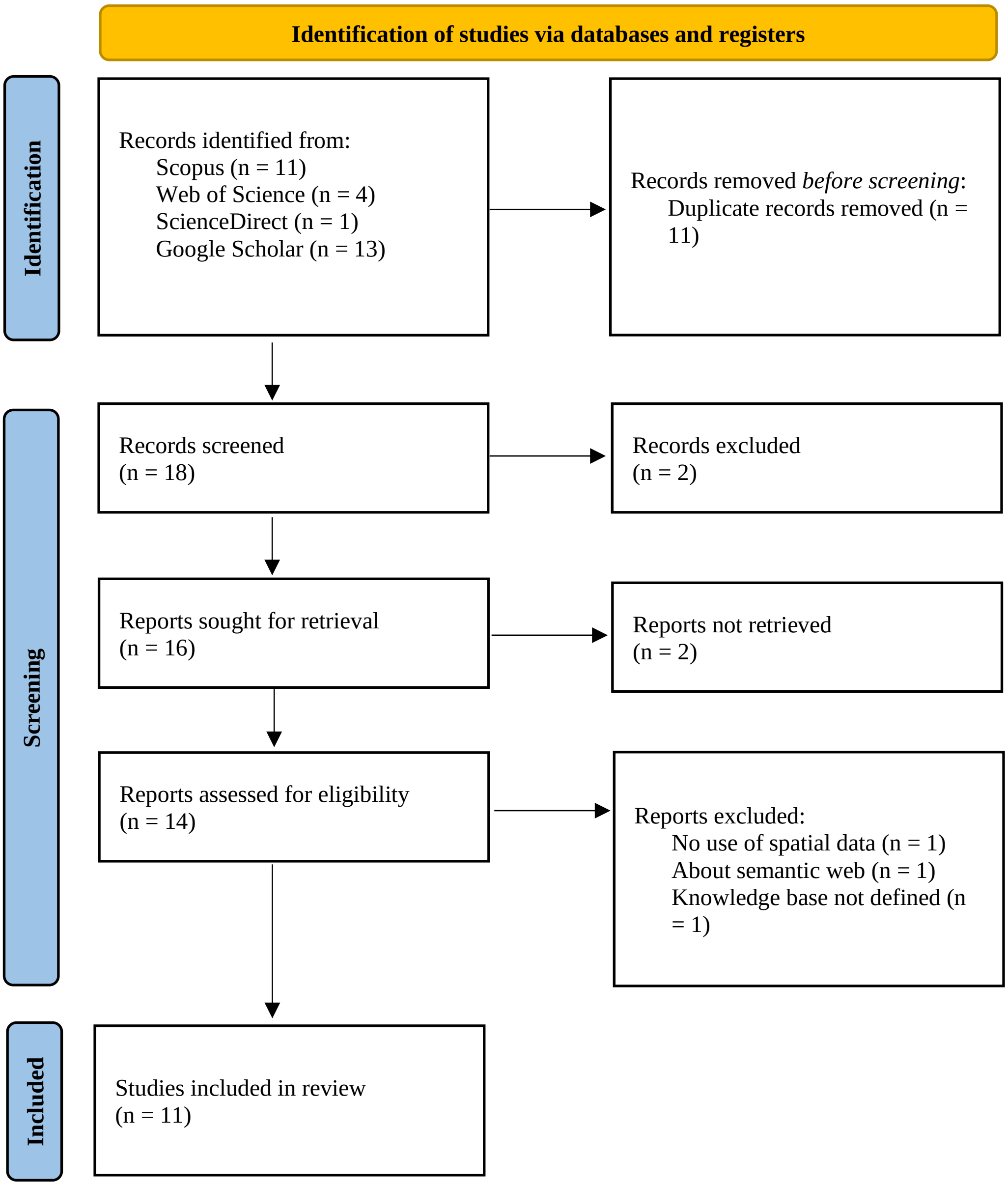

Figure 6 - PRISMA flow diagram. Source: Adapted from Page et al. (2021)

2.1.1.4 Concept matrix SKI literature

As stated, the synthesis follows the concept-centric approach leading to a concept matrix. The concept matrix includes which topic(s) the included articles discuss. The identified topics regarding SKI are (1) a definition, (2) the objectives, (3) the capabilities and (4) the architecture. These topics and the number of articles discussing the topic are shown in Table 2. The full concept matrix can be found in Appendix B – Concept matrix SKI literature review.

	Definition	Objectives	Components	Architecture
Number of articles	4	9	4	6

Table 2 - Summarized concept matrix for used spatial knowledge infrastructure literature

2.1.2 Narrative literature review of SDI

The SDI literature review was conducted after the systematic literature review on SKIs had been done. The starting point for the literature study on SDI are the identified topics of Table 2, which resulted in a search strategy that is following a systematic approach.

2.1.2.1 Search strategy SDI literature

Similar to the PRISMA protocol on SKI, the search strategy for the narrative literature study defines the databases searched for and the specifications of the search. For the SDI literature search, ScienceDirect and Google Scholar were used. ScienceDirect served as main database whereas Google Scholar was used in addition to identify other relevant literature. Including two databases meets the criterion for a systematic approach (Bramer et al., 2017).

The terms searched for were based on the outcomes of the systematic literature review on SKI. In this case, the concept matrix was used to refine the search terms. Therefore, the terms were constructed like “SDI <*topic*>” or “Spatial Data Infrastructure <*topic*>” where <*topic*> is the variable that entails one of the keywords ‘definition’, ‘objectives’, ‘components’ or ‘architecture’. Based on this logic, one of the searches resulted in the search “Spatial Data Infrastructure components”. The reasoning behind these search terms is two-sided. Firstly, as stated before, a preliminary search for “Spatial Data Infrastructure” resulted in 2,077 articles on Scopus. Screening and reading this amount of documents was not feasible within the time frame of the research meaning a refinement in the search was necessary. Secondly, the SKI literature study provided the topics to be studied in the SDI literature review for comparing the two concepts with each other. Thus, to both reduce the number of results as well as to find specific articles this search strategy is considered to be most practical for this research. Regarding the search criteria, no further specifications were applied. On ScienceDirect, the field ‘Find articles with these terms’ was used for the search whereas on Google Scholar the search terms were entered as input to search for all articles. Possible limitations of using these search terms and criteria are that articles using related terms (Coetzee & Wolff-Piggott, 2015) instead of “SDI” or “Spatial Data Infrastructure” are out of scope, as well articles that do not explicitly include one of the terms ‘definition’, ‘objectives’, ‘components’ or ‘architecture’.

The next stage of the PRISMA guidelines is the screening stage. Due to the narrative approach of the SDI literature study, these steps are not as well-documented as in the systematic approach on SKI literature. Therefore, this part deviates from the PRISMA protocol..

The search is carried out during March, 2024. An overview of the search strategy is provided in Table 3.

	Definition	Objectives	Components	Architecture
Search term	“Spatial Data Infrastructure definition” OR “SDI definition”	“Spatial Data Infrastructure objectives” OR “SDI objectives”	“Spatial Data Infrastructure components” OR “SDI components”	“Spatial Data Infrastructure architecture” OR “SDI architecture”

Table 3 - Search strategy for the literature on spatial data infrastructures

2.1.2.2 Concept matrix SDI literature

After performing the search strategy, the SDI literature study did not adhere to the PRISMA guidelines of screening and selecting the articles. Therefore, documentation, like Figure 6, is missing for this part. However, in order to provide insights in the number of included SDI articles that discuss one or more of the identified topics, Table 4 is constructed. Furthermore, the full concept matrix of this literature study can be seen in Appendix C – Concept matrix SDI literature review.

	Definition	Objectives	Components	Architecture
Number of articles	9	15	12	12

Table 4 - Summarized concept matrix for used spatial data infrastructure literature

2.1.3 Synthesis

In conclusion, the synthesis of the systematic literature review on the concept of SKI has led to the identification of four topics that all together make it possible to analyze the concept. The four topics are the definition of the concept, the objectives, the components and the architecture of the concept (Figure 7). These four topics are also used as search strategy in the literature study on SDI. A second use of the concept matrix of identifying different ways to study both concepts is that it forms the basis of the conceptual model, which will follow in chapter 4. The differences between SDI and SKI can be distinguished based on this analysis. These differences make it possible to make an assessment form for a spatial infrastructure. This assessment will make it possible to tell to what extent a proposed infrastructure is an SDI, SKI or something in-between. Thus, the case studies will be analyzed based on the conceptualized assessment form. The fact that the analysis will be done with the help of this form instead of only empirical research and interpretation makes the case studies straightforward and concise.

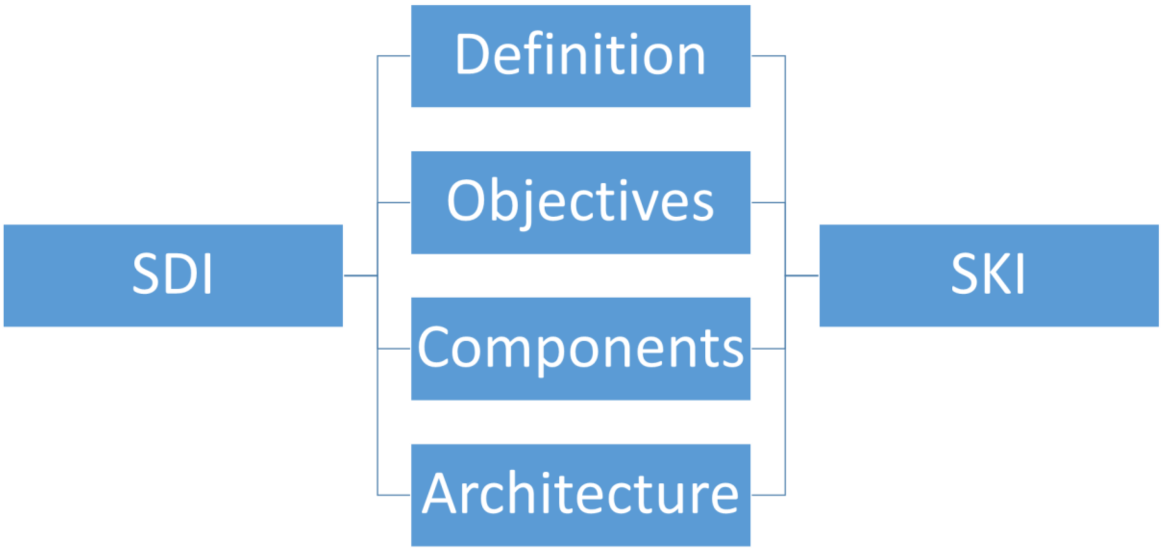

Figure 7 - The four topics identified to describe the two spatial infrastructures

2.2 Case studies

As stated before, two case studies will be conducted, which are the Kadaster Knowledge Graph (KKG) and 3D Amsterdam. For selecting the cases, it was important that a case should have a clear spatial component and it should have characteristics of a data infrastructure. Both cases are focused on spatial data and can be considered having a data infrastructure. The definition of a data infrastructure is that it consists of '*data assets (such as datasets, identifiers, and registers), standards and technologies used to curate and provide access to data assets, guidance and policies that inform the use and management of data assets and the data infrastructure itself, organizations that govern the data infrastructure and the communities involved in contributing to or maintaining it, and those who are impacted by decisions that are made using it*' (Dodds & Wells, 2019, p.262). Looking at the case studies, both include multiple data sources which can be accessed and downloaded via the user interface (UI). Furthermore, the municipality of Amsterdam for 3D Amsterdam and Kadaster for the KKG are the organizations that govern the infrastructure. This implies that both projects have (or maybe are) a spatial infrastructure. Furthermore, to make the cases an interesting subject for this study, they should have objectives and features that seem to move beyond usual SDI objectives and features that might be considered to be SKI characteristics. Section 1.2.1 identified that both the KKG as well as Amsterdam 3D have certain of these characteristics. Although the focus of the cases clearly differ from each other (Section 1.2.1), the different characteristics all are assumed to be characteristics of an SKI. Dutch cases with a focus on the surface area of the Netherlands are included in this study for practical reasons. Having the Netherlands as study area increases the correctness of the evaluation as this area is known by the author. There are also potential risks by choosing these two cases, as they both are experimental cases and only published in recent years (2019) having limited publicity. The infrastructures are therefore not used extensively by users yet and they are still under development. The risk for both infrastructures is that they, currently, seem to be depending on the continuity of the project. The question is whether the spatial infrastructure will be preserved in case, for example, the funding of the project stops. Especially because both cases have open data, funding of open data project could be stopped when the organization considers the business model is insufficient (Welle Donker, 2018).

To assess the cases, multiple assessment methods are included. This is advised due to the complex nature of SDIs, and thus SKIs (Grus, Crompvoets & Bregt, 2007). For this research, three methods are included which are a content analysis, a usability study, and semi-structured interviews. These three methods are chosen as they complement each other and should provide more concise answers. The case studies and subsequent methods are conducted mid-2024.

2.2.1 Content analysis

The methodological research technique of a content analysis is an approach to the analysis of a wide variety of visual and verbal data, like documents, text, video and audio in order to objectively quantify content in a systematic and replicable way (Bryman, 2012; Harwood & Garry, 2003; Stemler, 2015). The systematic and objective approach should suppress bias and enhance reproducibility (Bryman, 2012; Harwood & Garry, 2003). In a content analysis it is frequently desired to code the information in terms of subjects and themes (Bryman, 2012). Predefined categories serve as a base for the coding scheme. Different approaches can be used to conduct a content analysis depending on the aim of the study. Data could be sampled when the content analysis has to go through large amounts of data (e.g., mass media) (Bryman, 2012). In an ethnographic content analysis, the themes, subjects, and coding scheme can be iterative adjusted to the information distinguished (Sjoukema, 2021; Stemler, 2015). In latent content analysis, no coding scheme is applied as it focuses on interpreting the underlying meaning and themes without systematically categorizing themes and subjects (Graneheim & Lundman, 2004; Drahmann & Huijts, 2021). However, this approach affects the objectiveness and replicability of the study. The different approaches emphasize the need to discuss the approach and desired outcomes of the content analysis. For this research, a predefined set of subjects will be used to conduct the content analysis. These subjects will be based on the outcomes of the upcoming chapters. Therefore, Section 4.3.1 will specify the approach and subsequent subjects to be studied.

2.2.2 Usability study

The usability of any software product is deemed to be essential for its quality. Studying the usability of the product can be conducted by means of a usability evaluation method (UEM). This evaluation method focuses on the UI of a software application. UEMs are designed to establish whether the software application is usable for an end-user (Gupta, 2015; Paz & Pow-Sang, 2014). UEMs can be classified into three different methods: inspection, testing, and inquired methods (Gupta, 2015). EUM inspection methods include specialists inspecting the UI and identifying where aspects of the UI are deviating from the set guidelines of the application (Gupta, 2015; Paz & Pow-Sang, 2014). In the test methods, end-users will try and use the system while being recorded and issues are noted (Gupta, 2015; Paz & Pow-Sang, 2014). Inquiry methods are usually done during usability tests with the aim of collecting subjective impressions of users on the UI (Gupta, 2015).

In GI and SDI research, commonly used UEMs are the testing methods (Blake et al., 2017; Charalampos Gkonos & Hurni, 2019; Kalantari et al., 2020). Inquiry methods include semi-structured interviews (Kalantari et al., 2020), surveys (Blake et al., 2017) and questionnaires (Charalampos Gkonos & Hurni, 2019). One employed testing method is the think-aloud protocol, which is when testers ask participants about the UI to understand how the user thinks about the system (Gupta, 2015; Kalantari et al., 2020). Another included testing method is the shadowing method, where the expert user sits next to, but is prohibited from talking to, the participant (Charalampos Gkonos & Hurni, 2019; Gupta, 2015). During the test, the tester evaluates the participant's behavior (Gupta, 2015).

In this study, the aim is predominantly to evaluate to what extent the infrastructure is either an SDI, an SKI or something in-between. This assessment will be done based on the differences distinguished between the concepts, which is also part of this study. Therefore, the testing will be done only by the author of this article. The approach will be discussed in Section 4.3.2. There is no user-group or other tester involved, making the testing method different from the testing methods of UEMs. As for the inquiry method, the semi-structured interview will be used.

2.2.3 Semi-structured interviews

The first two methods, the content analysis and the usability study, will be conducted by the author alone. Due to this reason, it is beneficial to evaluate and substantiate the findings with someone who is directly involved in the case study in question. This could also lead to new information and new insights. To decide on the method, it is important to define whether the answers should provide quantitative or qualitative data. For this study, qualitative data will be used for the evaluation of the infrastructure. Besides, it is necessary to know how many and which people could be reached out to for information. As a spatial infrastructure has a technical system as its core foundation, the target audience includes staff involved in the product management team of the system. This could be someone who has the role of product lead, product owner or developer. The combination of qualitative data and a limited amount of possible interviewees exclude methods like a questionnaire (usually quantitative data) and surveys and focus groups (usually larger populations needed) (Bryman, 2012; Gupta, 2015). Therefore, an interview will be conducted. Qualitative interviews describe two main types of interviews: an unstructured interview and a semi-structured interview (Bryman, 2012). An unstructured interview is used when a range of topics is discussed whereas a semi-structured interview makes use of an interview guide to discuss a predefined and specific topic list (Bryman, 2012). Due to the topic list, a semi-structured allows for more in-depth discussion on the topics (Burroughs et al., 2019). While the interview should only provide answers to certain topics that are established in the upcoming chapters, there will be an interview guide and the interview will be semi-structured. The detailed approach and interview guide will be discussed in Section 4.3.3.

3. Differences between Spatial Data Infrastructure and Spatial Knowledge Infrastructure

Section 1.1 briefly addressed that the SDI evolved over time which is generally characterized by different generations. In fact, three generations have been discussed, with a potential new, fourth, generation proposed. However, other developments like the open SDI also impacts the SDI as a concept. In order to get a better understanding of the evolution of SDIs, it is important to elaborate on the history of the SDI.

The first generation SDI originated in the mid-1980s and is considered a product-based approach focused on producing and integrating spatial data (Rajabifard, Feeney & Williamson, 2002; Masó, Pons & Zabala, 2010; Nedovic-Budic, Pinto & Budhathoki, 2008). The first SDIs were authoritative driven by mapping agencies designed for national requirements (Rajabifard & Williamson, 2001; McDougall, 2010) and had as goal promoting economic development, stimulating better government and encouraging environmental sustainability (Masser, 1991; Vilches-Blázquez & Ballari, 2020). The approach focuses on developing a single, shared database to gather all available datasets (Masó, Pons & Zabala, 2010; Vilches-Blázquez & Ballari, 2020). The second generation SDI originated around 2000 and is assumed to be a process-based approach, focused on data processing and analyzing (Vilches-Blázquez & Ballari, 2020; Sjoukema, 2021). Although the second generation SDI had more attention to the user, both the first and the second generation are considered data-centric (Sadeghi-Niaraki et al., 2010). The third generation originated around 2008 and were supposed to switch towards a user-centric approach (Hennig & Belgiu, 2011; Nedovic-Budic, Pinto & Budhathoki, 2008; Sjoukema, 2021). Volunteered geographic information (VGI) and the appearance of sub-national SDIs contributed to the origin of this generation (Budhathoki, Bruce, & Nedovic-Budic, 2008; Vilches-Blázquez & Ballari, 2020). To what extent these generations are implemented in the real world can be questioned. At least for the third generation, the user-driven SDI, the implementation seems not to have taken place in practice (Hendriks, Dessers & Van Hootegem (2012; Sjoukema, 2021; Vancauwenberghe & van Loenen, 2018). This is emphasized by the fact that SDIs are usually still not considered to be user friendly (UN-GGIM, 2022).

Another movement in SDI evolution that, amongst other objectives, aims to overcome the user experienced issues is the ‘open SDI’. Open SDI originated from open data initiatives by governments which started around 2009 (Welle Donker & van Loenen, 2017). Regarding the GI-domain, the first initiative within the European Union (EU) for sharing spatial data with the public sector was INSPIRE (Izdebski, Zwirowicz-Rutkowska & Nowak da Costa, 2021; Vancauwenberghe & van Loenen, 2018). Enhancing the openness of spatial data by adhering to open data principles is the primary outcome of open SDIs, but even important is that the different stakeholders are included in the implementation and governance of the SDI as well as it serves the needs of all stakeholders (Izdebski, Zwirowicz-Rutkowska & Nowak da Costa, 2021; Vancauwenberghe & van Loenen, 2018). In this case, also non-governmental actors, such as citizens and business are identified as stakeholders (Vancauwenberghe et al., 2018; Vancauwenberghe & van Loenen, 2018).

In 2020, a fourth generation was mentioned by Vilches-Blázquez & Ballari (2020) as response to the latest societal needs and developments in computer science. Examples of these IT trends mentioned are cloud computing, AI and the semantic web (Vilches-Blázquez & Ballari, 2020). Similar examples of technological SDI developments being discussed are 3D/4D models within the SDI (Stoter et al., 2011) and making the SDI cloud-based (Tripathi, Agrawal & Gupta, 2020). Interestingly, the first insights into the presumed successor of the SDI, the Spatial Knowledge Infrastructure (SKI) also addressed these characteristics as being part of an SKI rather than part of an SDI (Figures 3 and 4 in Section 1.1). Nevertheless, also the SKI has a short history as term being used.

The first use of the term Spatial Knowledge Infrastructure (SKI) was by Markus (2005) to describe ‘*a framework for making data discoverable*’ (Woodgate et al., 2019 p.118). This objective does not significantly deviate from the objective of SDIs as stated in Section 1.1 as well as it does not make clear what the ‘knowledge’ component entails. In 2012, Fiedukowicz et al. (2012) and Stock et al. (2012) discussed the SKI from different viewpoints. Where Fiedukowicz et al. (2012) focus on the development of geoportals, Stock et al. (n.d.; 2012) focuses on the architecture of the system. The architecture should be based on ontologies in order to become a knowledge infrastructure. However, both articles do not explicitly discuss what the knowledge base is and how knowledge is extracted. Duckham et al. (2017) is the first article to fully conceptualize and define what an SKI is, by describing the differences in capabilities between an SDI and SKI, of which Section 1.1 provides some insights. After this article, more articles followed that explicitly study the concept of an SKI. These articles together form the foundation of this chapter, which describes both the concepts of SDI and SKI based on the identified topics in the systematic literature review. The upcoming sections are the (1) definition, (2) objectives, (3) components and (4) the architecture. The four topics combined will provide a comprehensive and overall overview of both concepts and the extent to which they differ from each other.

3.1 Definition

SDI, and thus SKI, are both complex concepts. Including a definition of the concepts is both useful for understanding the concepts as well as it is necessary as a basis for evaluating the concepts (Vandenbroucke et al., 2009 ; Warnest, 2005). The definition will provide a first look at what the concept entails.

3.1.1 SDI

The definition of what an SDI is, is debated ever since the early stages of the concept (Rajabifard & Williamson, 2001) and still continues to present (Arnold et al., 2021; Hendriks, Dessers & Van Hootegem, 2012; Vandenbroucke et al., 2009; Warnest, 2005). The study of Chan (2001) into SDI definitions is often used as starting point of follow-up research (Hendriks, Dessers & Van Hootegem, 2012; Warnest, 2005). Hendriks, Dessers & Van Hootegem (2012) already identified 28 definitions in their study and the number of definitions increased even more until present day (UN-GGIM, 2022; Van Westen, 2013). The binding factor between all identified definitions is that they refer to the components of an SDI, the objectives of an SDI or both (Vandenbroucke et al., 2009; Hendriks, Dessers & Van Hootegem, 2012). The components are both of technical and non-technical nature whereas the objectives vary from specific (access, exchange and use of spatial data) to broad (contributing to the performance of the business processes) (Hendriks, Dessers & Van Hootegem, 2012; Vandenbroucke et al., 2009; Warnest, 2005). Despite the large number of existing definitions, some authors (Chan, 2001; Ferreira et al., 2015; Vandenbroucke et al., 2009) advocate none of the given definitions can be considered as being holistic enough to describe the holistic framework an SDI is assumed to be. Nonetheless, it is difficult to agree there is no suitable wording available to describe what an SDI is. Accordingly, this study does choose a description that will be used as definition for SDIs. Though acknowledging the definition of SDIs is debatable, Arnold et al. (2021, p.369) state the term is generally used to describe '*a framework of technologies, policies and institutional arrangements to enable access to spatial data, and facilitate its discovery and sharing*'. This definition will be used as it approaches SDI as a holistic framework and it encompasses both technical and non-technical components as well as it mentions the objectives.

3.1.2 SKI

Three definitions are given in the SKI literature. These definitions can be divided into two categories: a technical definition and a holistic approach. The technical definition is stated by Omidipoor (2018, p.2) as '*an integrated set of technological components (typically hard components include computer and networking hardware and facilities, and soft components also include various software, services, procedures, protocols or standards) that are the foundation of a knowledge service*'. While an important part of SDI, and thus SKI, are also the non-technological components and the objectives of the infrastructure, this definition is not deemed holistic and inclusive enough to serve as definition for this study.

The two holistically approached definitions are:

- Duckham et al. (2017, p.4): '*A network of data, analytics, expertise and policies that assist people, whether individually or in collaboration, to integrate in real time spatial knowledge into everyday decision-making and problem solving*'

- Geospatial World (2022, p.17): '*An infrastructure to integrate geospatial approaches, data and technologies into the wider digital ecosystem. In so doing it delivers the location-based knowledge, services and automation expected by economies, societies and citizens in the 4IR* [fourth Industrial Revolution] *age*'

When comparing these two definitions, it could be said that there is an overall consensus. They both emphasize the objectives of providing spatial knowledge and include technological components (data) as well as non-technological components (people). Subtle differences could also be seen. Where Duckham et al. (2017) seems to highlight the importance of policies, where Geospatial World (2022) leave this out of its definition. Another difference could be identified from extracting the aim of the concept. Defined by Geospatial World (2022) the result of an SKI is to have '*location-based knowledge, services and automation expected by economies, societies and citizens in the 4IR age*', which seems rather abstract compared what Duckham et al. (2017) aimed for in '*to integrate in real time spatial knowledge into everyday decision-making and problem solving*'. According to the author, the definition of Duckham et al. (2017) is considered to provide the clearest definition in terms of the components and objectives of SKI. The definition of Duckham et al. (2017) is also used in other articles studying SKIs (Arnold et al., 2021; Ivánová et al., 2020; Kopsachilis, Vachtsavanis & Vaitis, 2023).

3.1.3 Differences

As an SKI emerged from SDI it is not surprising that the definition of both concepts has quite some resemblance to each other. The foundation of both concepts is based on a holistic framework regarding technological and non-technological components aiming to support users when working with spatial data. If only the two preferred definitions would be taken into consideration, it could be said that an SKI is highlighting the user, or people, in contrary to the SDI. However, some definitions of SDI also explicitly include the user (Hendriks, Dessers & Van Hootegem, 2012; Vandenbroucke et al., 2009; Warnest, 2005). Due to this reason, this will not be included as being a difference. Main differences regarding the definition of both concepts are that an SKI aims to provide spatial knowledge, real-time use of spatial data and that it is able to assist in some form of analytics. Spatial knowledge is in this study defined as providing ready to use answers on questions.

3.2 Objectives

In order to extract what the objectives of SDI or SKI as a concept are, it is necessary to predefine what is understood with the meaning of objective. A spatial infrastructure is not a stand-alone system, but part of a larger interrelated system striving for wider goals. The objectives specifically for the SDI or SKI could therefore be different from the overall objectives. For example, one project that included an SDI is the Global Monitoring Environmental and Security (GMES) having the main objective '*to monitor and better understand our environment .. and to contribute to the security of every citizen*' (Idrizi, 2018, p.59). From this program objective, it is difficult to derive in which way and to what extent the spatial infrastructure contributes to achieving the objective. Although objectives for a specific spatial infrastructure are context-dependent to some extent, they should all be based on a common set of objectives. The objectives can be studied in a twofold way. On the one hand an objective could be very narrow and simple, like facilitating easy use of spatial data. But it is also interesting to see what this means, for example whether the use of spatial data means visualized data on a map or querying a dataset. This also impacts on who is being seen as user, while are usually not familiar with querying whereas geospatial or IT experts are able to do this.

3.2.1 SDI

Already identified in the previous section on the definition of SDI, is that in some cases the objectives are referred to in the definition. In these cases, the objectives follow a same progression as the SDI generations do going from product based to process based and ultimately user oriented. This is also reflected in the objectives. Initially, SDI initiatives started in order to save money and improve efficiency and effectiveness after which it turned into the more holistic approach (Sjoukema, 2021; Zwirowicz-Rutkowska. 2017). Hence, the objectives also started as solely data-oriented and later included user-oriented terminology (Hendriks, Dessers & Van Hootegem, 2012; Vandenbroucke et al., 2009). A data-oriented objective could be identified when the aim is to integrate, share and exchange spatial data between stakeholders in the spatial data community (Hendriks, Dessers & Van Hootegem, 2012; Grus et al., 2011; Toomanian, 2021) whereas a user-oriented objective would be enabling an easy search for, access to and use of geoinformation (van Loenen, Crompvoets & Poplin, 2010). These objectives can be considered as being specific objectives (Vandenbroucke et al., 2009). Broader defined objectives combine both the data-related and user-related view or are broader in other terms, for instance by including business processes or business objectives (Hendriks, Dessers & Van Hootegem, 2012; Vandenbroucke et al., 2009). Though, only mentioning business objectives is limited looking at the various SDI projects place, making ‘program objectives’ a more suitable approach. Program objectives combine both data-oriented and user-oriented perspectives into a set of objectives that should be dealt with by constructing the SDI (Vandenbroucke et al., 2009; Zwirowicz-Rutkowska. 2017). A comprehensive aim, or objective, for SDIs could be formulated as integrating, sharing and exchanging spatial data between stakeholders enabling an easy search for, access to and use of this spatial data by the stakeholders in accordance with the program objectives. As there is much overlap with the definition of Arnold et al. (2021), it does substantiate this definition in terms of objectives included.

Although these objectives are in general the objectives of an SDI, every SDI is unique and has its own objectives. Especially the introduction of the open SDI had its implications on how these objectives are implemented. The data should be open instead of closed and also non-governmental stakeholders must be acknowledged (Vancauwenberghe et al., 2018). Open data is data being shared by the provider to everyone who wishes to access and (re)use it without having to legally or financially compensate the provider (Welle Donker & van Loenen, 2018). While the objective of open SDIs is to include users like citizens, SDIs are usually considered to support geospatial experts as user (McEachen & Lewis, 2023) as an SDI in general provides spatial data and information instead of knowledge. Citizens can access the SDI and download the data, but need to know what to look for in order to understand it. This implies the user has to be familiar with the GI-domain.

3.2.2 SKI

Stock et al. (2012) state the objectives of an SKI are the same as an SDI with the addition of semantic richness. The SDI objective referred to is access to spatial data, which is a data-oriented objective of SDI. Section 3.1.2 SKI identified the objectives of providing spatial knowledge, real-time data and assisting in analysis. Semantic richness can contribute to these objectives while integration with the semantic web, and adding new techniques like spatial data mining (SDM), is needed to support in knowledge creation (Fiedukowicz et al., 2012; Geospatial World, 2022; Kopsachilis, Vachtsavanis & Vaitis, 2023).

The availability of real-time data comes down to the data being open (available for everyone) and updated frequently. To integrate data within the semantic web does mean the standards have to comply with the standards applying for the semantic web, according to standards set by the W3C. This would support user-oriented objectives as on-the-fly real-time data analysis (Duckham et al., 2017; Geospatial World, 2022) and ready to use spatial data (Arnold et al., 2021). Another recurring objective is the need for advanced querying. An SKI should be able to deal with queries formulated by the user in terminology of the user and provide ready to use data, or answers (Arnold et al., 2021; Ivánová, Armstrong & McMeekin, 2017; Woodgate et al., 2017). As for the user, the aforementioned objectives should lead to non-geospatial expert use of the SKI (Duckham et al. 2017; Geospatial World, 2022; Woodgate et al., 2017). Another objective is to present the data in 3D or higher (Duckham et al. 2017; Woodgate et al., 2017). 3D will improve the visualization, use and interpretability of the data, whereas a 4D representation adds spatial-temporal information which can be used for predictive modeling (Geospatial World, 2022; Woodgate et al., 2017).

The overall objective of SKI could therefore be stated as semantic web integrated spatial data facilitating advanced querying and on-the fly real-time data analysis by the user providing ready to use (spatial) information resulting in spatial knowledge. The last paragraph indicates that the way to this objectives predominantly lie within the technical aspects. Although this might be true, these developments need other aspects, like the governance, to be in place as well (Coetzee et al., 2021; Sjoukema, 2021).

3.2.3 Differences

The main difference between SDI and SKI is that the objectives change from data sharing to knowledge creation. While the objective of an SDI is to access, share and exchange spatial data (Hendriks, Dessers & Van Hootegem, 2012; Grus et al., 2011; Toomanian, 2021), the SKI aims to provide real-time spatial knowledge (Duckham et al., 2017). This spatial knowledge is defined as providing ready to use answers, something that is also identified in the SKI literature (Arnold et al., 2021).

Interesting is that the SKI literature mainly focuses on technological developments that should lead to the spatial knowledge, which could therefore be seen as technical objectives of the evolution. For example, the data in an SKI should be integrated with the semantic web meaning it is automatically shared and accessible (Fiedukowicz et al., 2012; Geospatial World, 2022; Kopsachilis, Vachtsavanis & Vaitis, 2023). However, the development of including semantic web standards is also included in open SDIs as it enhances interoperability (Izdebski, Zwirowicz-Rutkowska & Nowak da Costa, 2021; Welle Donker & Van Loenen, 2018). However, an SDIs is usually considered to be using standards specifically designed for the GI-domain (Granell, Díaz & Gould, 2010; Omidipoor, 2018). Another technical objective is that the SKI should have a 3D or 4D representation, whereas the data in an SDI is usually presented in a two-dimensional representation (Duckham et al. 2017; Geospatial World, 2022; Woodgate et al., 2017). However, also regarding the 2D versus 3D, this is a development discussed in SDI literature (Stoter et al., 2011).

From a user perspective, the SDI is in general considered to be a system fit for geospatial experts and not as much for other users (Duckham et al., 2017; Granell, Díaz & Gould, 2010; McEachen & Lewis, 2023) despite developments in open SDIs that aim for incorporating other users as well (Vancauwenberghe et al., 2018). SDI support the access to and download of data for analytical purposes or to query it in a simple and hardcoded way which make it difficult for non-experts to use (Arnold et al., 2021; Woodgate et al., 2017). The SKI aims the querying should be easy to do by non-experts as well which should support real-time data analysis and predictive modeling (Arnold et al., 2021; Ivánová, Armstrong & McMeekin, 2017).

3.3 Components

Besides objectives, also components are referred to in the definitions of SDI and SKI (Hendriks, Dessers & Van Hootegem, 2012). The components, both non-technological and technological, together form the infrastructure. This next section discusses the included components for SDI and SKI.

3.3.1 SDI

As identified in the definition, it is widely varying what the components of an SDI are (Grus, Crompvoets & Bregt, 2007; Vandenbroucke, 2009). The core components and schematic representation proposed by Rajabifard & Williamson (2001) (Figure 8) is still used in most of the SDI literature as being the components of an SDI (Sjoukema, 2021; Toomanian, 2012; Welle Donker & Van Loenen, 2017).

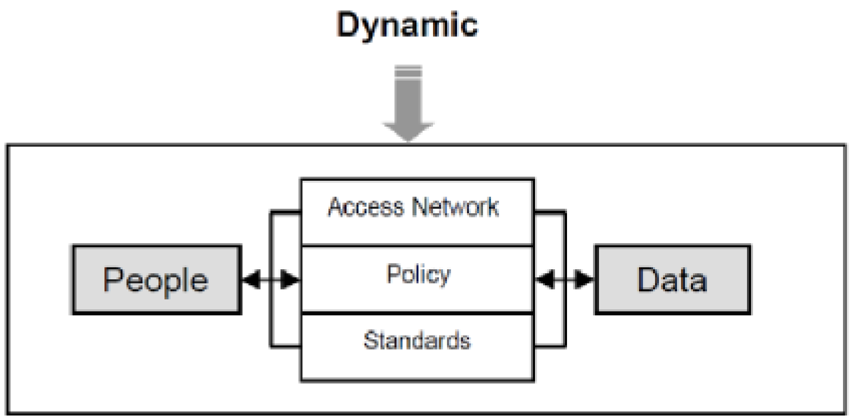

Figure 8 - Most used set of components of a spatial data infrastructure. Source: Rajabifard & Williamson, 2001

In this representation people and data are the two fundamental components that want to interact with each other. People want to access, share and/or use (spatial) data. To do so, the three technological components policy, standards and access network need to be in place. Access network could also be rephrased as technology. Initially, this representation was based on first generation SDIs (Rajabifard & Williamson, 2001).

The wide variety of components identified in SDI literature on the definitions indicate that the five proposed core components might not be the holy grail of SDI components. On itself this is not unusual as both new insights in and understanding of the concept evolve over time, as well as the technical possibilities and thus expectations do. As for technology, the desktop stand-alone GIS tools have made place for internet accessible geoportals (Steiniger & Hunter, 2012) and standards have been improved due to organizations as the Open Geospatial Consortium (OGC), Organization for Standardization (ISO) (Masó, Pons, & Zabala, 2012) and the The World Wide Web Consortium (W3C). The OGC designs the standards specifically for the spatial domain (Ferreira et al., 2015), whereas ISO is an overarching organization dealing with standards (Kopsachilis, Vachtsavanis & Vaitis, 2023). W3C designs the standards for the (semantic) web. Improved data standards make room to address new components like clearinghouses (geoportals) and metadata (Ferreira et al., 2012; Hennig & Belgiu, 2011; Tripathi, Agrawal & Gupta, 2020). Apart from technological evolutions, the soft components are also addressed. Stressed by Sjoukema (2021) is that policies alone are insufficient in developing and maintaining an SDI. It should have a governance structure, of which policies are deemed only a sub-component. The term ‘people’ in the framework refer to every stakeholder, from producer to user (Rajabifard & Williamson, 2001). Nonetheless, facilitating the user has been unachieved (Hennig & Belgiu, 2011). As users have their own unique goals, it is needed to both reconsider who the users are as well as addressing their needs (Welle Donker & Van Loenen, 2017).

A revised understanding of the SDI components is proposed by Arnold et al. (2021) (Figure 9). This SDI representation consists of four core components: data, users, technology and governance. In contrary to the components introduced by Rajabifard & Williamson (2001), each core component is built up out of four sub-components. In this case, the component ‘policy’ is one of the four sub-components of the ‘governance’ component and ‘standards’ is a sub-component of the core-component ‘data’. Furthermore, the component ‘access network’ has been rephrased to ‘technology’. The component ‘people’ needs more explanation. In Figure 9 a component ‘users’ can be seen, but ‘people’ are visible around the edges of the four components just as ‘capacity’. Though not discussed by Arnold et al. (2021), this could be explained as the ‘people’ component being rephrased to ‘users’ while these are the stakeholders. The ‘people’ and ‘capacity’ around the edges could mean that the core components still need people and resources to achieve what they aim for. Overall could be concluded that this revised view on the components address the issues of SDI over time and deliver a wider (sub-)set of components than only the core components.

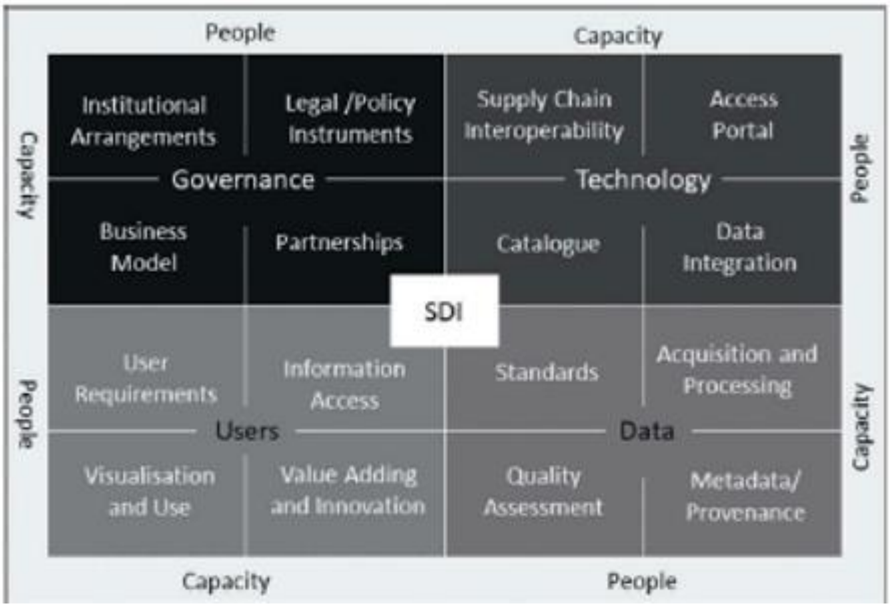

Figure 9 - Revised components of a spatial data infrastructure. Source: Adapted from Arnold et al., 2021

3.3.2 SKI

Just like the definition, Omidipoor (2018, p.4) is describing the components from a technical viewing point, stating the main components of an SKI are '*data, ETL* [Extraction-transformation-load], *spatial warehouse engine, Spatial On-line Analytical Processing (SOLAP) server and a collection of web knowledge services*'. Except for the data component, the other 'components', as will become clear in the next section, are architectural elements instead of what is usually referred to as components of a spatial infrastructure. Arnold et al. (2021) propose a set of two components that need to be added to the aforementioned SDI components (Figure 10) in order to move from an SDI towards SKI. The SKI part includes the two components of 'knowledge representation' and 'analytics'. These two additional components are explained as being (Arnold et al., 2021, p.370):

- Knowledge representation – '*the aspects that represent real world information in a form that a computer can utilise to solve complex tasks'*
- Analytics – '*the procedures required to execute complex queries*'

Noteworthy about these two components is that the sub-components include some aspects that are also included in an SDI, such as data integration, standards, metadata and acquisition. The two ontology sub-components and the sub-components vocabularies and custodianship and QA are new. What exactly is meant with these sub-components and how it contributes is not discussed, but it is data-oriented.

Figure 10 - Components of a spatial knowledge infrastructure. Source: Adapted from Arnold et al., 2021

Iványová et al. (2020) also propose two components for an SKI, named the 'SKI Broker' and 'SKI Metadata Catalogue'. Three components are part of this SKI Broker:

- Processing user queries
- Managing and storing information about SKI resources (e.g. ontologies)
- Providing answers to the user

When comparing these two SKI components with the SKI base as proposed by Arnold et al. (2021), it can be said that these components represent the elements 'Knowledge Representation' and 'Analytics'. The metadata catalogue is a sub-component of the analytics component of Arnold et al. (2021) whereas the processing and answering of queries by the user entail the aim of both the knowledge representation as well as the analytics.

In contrary to the aforementioned authors, Geospatial World (2022) discusses six, as they call it, elements that, at first sight, seem to have little resemblance to the predominantly used components. These six elements are: (1) integrated policy framework, (2) foundation data, (3) partnerships and collaboration, (4), industry leadership, (5) applications, analytics and modeling and (6) geospatial dimension to the wider digital ecosystem. The integrated policy framework element encompasses aims such as an integrated digital governance and open data which is in line with the governance component of an SDI. Foundation data has some confluences with the data component in SDI as proposed by Arnold et al. (2021) considering the acquisition and quality assessment, as Geospatial World (2022) states authoritative data collection ensures the quality. Partnerships and collaboration fit into the sub-component of partnerships as part of the main component governance as proposed by Arnold et al. (2021). Industry leadership, again, seems to coincide with the governance component in terms of the business model sub-component as one of the outcomes is to gain value for business processes. The element 'Applications, Analytics and Modelling' target multiple components. It is emphasizing the users, the need for knowledge on demand and answers on questions from the users to acquire knowledge. Therefore, the users component of SDI and both SKI components as proposed by Arnold et al. (2021) are included. However, the sub-components Arnold et al. (2021) are mentioning are quite technical and detailed whereas the description of Geospatial World (2022) is rather abstract and non-technical. The last element, geospatial dimension to the wider digital ecosystem, is also having some resemblances to multiple components identified by Arnold et al. (2021). Presumed outcomes are data integration into the wider web and standards. Other than the sub-components of data integration and standards, connecting into the web of data has its need for ontologies, assuming that sub-component is also included.

3.3.3 Differences

Proposed by Arnold et al. (2021) is that the complete SKI infrastructure is an integrated set of, an updated version of, the most commonly identified SDI components (Rajabifard & Williamson, 2001) with two new SKI components added. These two SKI components address the knowledge representation and analytics. Iványová et al. (2020) also propose two components which are called differently, but do have the same resemblances as the two components by Arnold et al. (2021). While the aims of these new SKI components are processing and answering complex queries and providing ready to use information or answers to the user, the new components are predominantly data-oriented. Geospatial World (2022) assume to propose a complete new set of components, but when analyzed they also partly readdress components part of an SDI with the addition of new components aiming for new capabilities. Complex querying, data integration, modeling and standards are recurring themes as these are also part of the components of Geospatial World (2022). Nonetheless, Geospatial World (2022) add a 'knowledge on demand' component to this. In short, regarding the components it could be said that an SKI is adding new components and accompanied objectives to the already existing set of components of SDI.

3.4 Architecture

Describing a software architecture can be done on different levels. A high-level description of an architecture discusses the included components in general terms without going into detail on the technicalities (Arnold et al., 2021). A detailed level architecture is describing the technical details of the components, or it emphasizes on one of the components (Iványová et al., 2020; Stock et al., 2012). Discussing an architecture could also be done by applying it to a case study (Béjar et al., 2009; Granell, Díaz & Gould, 2010). Though it must be clear what an architecture comprises. The architecture can be seen as the structure of the system describing the technical and software elements (Béjar et al., 2009; Cipolloni, 2018). This is important, as some literature also include the user as one of the components of an architecture (Ferreira et al., 2015). This study will stick to the technical and software elements as parts of an architecture.

3.4.1 SDI

In the most simplified way can the architecture of an SDI be categorized into three elements being data, middleware and user applications (Boguslawski, Borzachiello & Perego, 2020). In terms of layers, the layers are the data layer, service layer and application layer (Béjar et al., 2009; Cipolloni, 2018; Tripathi, Agrawal & Gupta, 2020) (Figure 11).

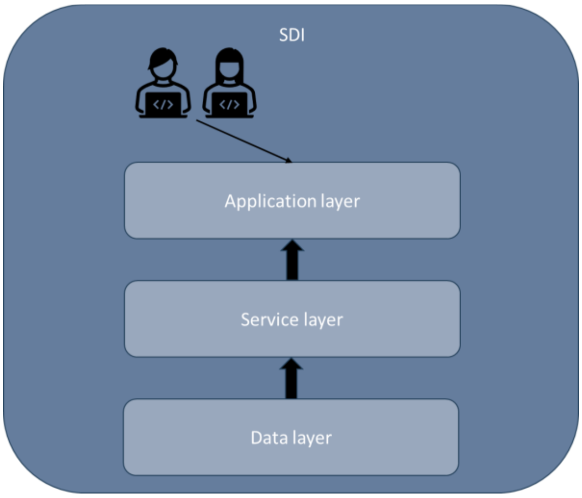

Figure 11 - Simplified high-level architecture of spatial data infrastructures

The data layer consists of elements like the databases, metadata and SDI documentation (Béjar et al., 2009; Cipolloni, 2018; Granell, Díaz & Gould, 2010). Thus, besides the storage of data itself, this layer contains the data management. Therefore, it can tell something about who produced the data in the datasets and in what way the data is stored. As most SDIs stem from governmental bodies, the databases are traditionally authoritative, both in produced data as in maintenance of the data (Duckham et al., 2017; Geospatial World, 2022). Although it must be noted that in open SDIs, the data could be produced by all users, which also includes non-authoritative users (Mulder et al., 2020; Vancauwenberghe & van Loenen, 2018), SDIs mainly provide authoritative data This also means the databases are generally being stored locally and have to be actively shared and approached by means of the service layer.

The service layer is the connecting part between the data layer and the application layer, thus can be considered the middleware of the architecture. It is the link that aid the objectives of discovering, accessing and using the data by the user. This layer usually consists of GI-domain services and standards (OGC services and standards like Web Service Feature) that facilitate the discovery, viewing, downloading and processing of the data (Cipolloni, 2018; Granell, Díaz & Gould, 2010; Omidipoor, 2018). These services processes the queries and map interactions of the user and translate them to OGC standards (Cipolloni, 2018; Granell, Díaz & Gould, 2010). As standards proposed by the OGC change over time to comply with the latest insights, it must be noted that SDIs usually make use of OGC standards like WMF, WFS, WMTS and WPS (Granell, Díaz & Gould, 2010; Omidipoor, 2018). However, developments in open SDIs aim to moving towards LD or at least semantic web standards (Izdebski, Zwirowicz-Rutkowska & Nowak da Costa, 2021).

The application layer could also be named the UI or, if present, the geoportal layer (Béjar et al., 2009; Cipolloni, 2018; Granell, Díaz & Gould, 2010). In this layer, the data can be queried and visualized. Though, the visualization is only possible in case the SDI has map services. SDI geoportals generally visualize the data in 2D (Béjar et al., 2009; Duckham et al., 2017; Woodgate et al., 2017) and are able to deal with simple queries only (Arnold et al., 2021; Duckham et al., 2017; Geospatial World, 2022). If there are no map services, queries are being used for finding, accessing and retrieving data. Downside of the limited querying possibilities is that the use of an SDI is, in most of the cases, reserved for geospatial experts instead of non-spatial acquainted users as well while it does not support ‘natural’ language but only structured query language (SQL) (Duckham et al., 2017; Geospatial World, 2022; Granell, Díaz & Gould, 2010). Besides, the output of simple queries means the output will be of more generic content that needs additional analysis by the user. Furthermore, SDI web-geoportals are predominantly designed for desktops (Duckham et al., 2017; Geospatial World, 2022; Granell, Díaz & Gould, 2010), although a shift towards mobile applications is proposed (Tripathi, Agrawal & Gupta, 2020).

3.4.2 SKI

Within the various descriptions of an SKI architecture, differences in approaches can be distinguished. Omidipoor (2018) is following the same layered structure as the SDI approach. He describes a data layer, a service layer and an application (portal) layer. Based on his perspective that an SKI is a combination of the concepts SDI, SWS (Spatial Web Services) and SDM (Spatial Data Mining) it is not surprising that the architecture is based on SDI on a high-level. The necessity of SDM in order to achieve an SKI is also mentioned in Fiedukowicz et al. (2012). According to Omidipoor (2018) knowledge extraction can be facilitated by combining concepts of SWS, SDM and the traditional SDI. Omidipoor (2018) emphasize on the fact that all three concepts on its own are not (yet) fit for knowledge discovery, but combining them could lead to this. SDM methods enhance usability and interoperability of the data, whereas the latest SWS can be used for implementing semantics and ontologies (Omidipoor, 2018). The main architecture follows the same layered approach as the SDI, consisting of the same three layers. As main components, Omidipoor (2018) name ‘data’, ‘Extract-Transform-Load (ETL)’, ‘spatial warehouse engine’, Spatial On-Line Analytical Processing (SOLAP) Server’ and ‘Web Knowledge Services (WKS)’. ETL is needed for getting the data prepared and stored in the spatial warehouse engine, the SOLAP server should enhance the querying and the WKS are performing SDM algorithms and providing summarized data (Omidipoor, 2018). However, Omidipoor (2018) also visualizes components ‘Mining’, ‘Knowledge Extraction’ and ‘Knowledge Reporting’ as being new. Unfortunately, what is meant by ‘Knowledge Extraction’ and ‘Knowledge Reporting’ is not substantiated by the author. Due to the differences between the visualized architecture and high-level description of the stated components it is difficult to fully grasp on the architecture Omidipoor (2018) is envisioning. It is now visualized by as if an SKI is in fact an SDI with the addition of new SWS and SDM techniques.

Ivánová et al. (2020) propose two new components, on top of the classic SDI architecture, which are an SKI metadata catalogue and an SKI broker. The SKI metadata catalogue is visualized as part of the SDI architecture where the SKI is built on. This is not surprising, as Ivánová et al. (2020) mention a metadata catalogue in itself is also part of an SDI. The difference in this case is the connection of the metadata catalogue with the semantic web, making the interaction with the catalogue easier and possible for various machines (Ivánová et al., 2020). As the basis of the metadata catalogue is no different from SDI, this will be part of the data layer. The SKI broker ‘*manages the interactions between the user application and spatial resources*’ (Ivánová et al., 2020, p.1534) which indicates the broker is part of the usual service layer. This results in a SDI-like architecture consisting a data layer, service layer and application layer.

Kopsachilis, Vachtsavanis & Vaitis (2023) also propose a high-level real-world SKI which is in line with the two aforementioned architectures by Omidipoor (2018) & Ivánová et al. (2020). Kopsachilis, Vachtsavanis & Vaitis (2023) identify the data layer in terms of the datasets, a service layer that transforms the data into semantic-web-ready data (RDF standard) and an application layer including an RDF explorer and a webGIS.

Arnold et al. (2021) proposes a different style architecture. Arnold et al. (2021) do identify five ‘layers’ being part of an SKI: storage layer, publishing layer, compliance layer and the application layer (Figure 12). Based on this visualization of the architecture, the ontology based SKI architecture as proposed by Stock et al. (n.d.) and Stock et al. (2012) fit in. Both the storage layer and the compliance layer include a sub-component regarding ontologies. Both ontology based architectures are an in-depth architectural view on these sub-components. The data standard for ontologies is the OWL, set by the W3C.

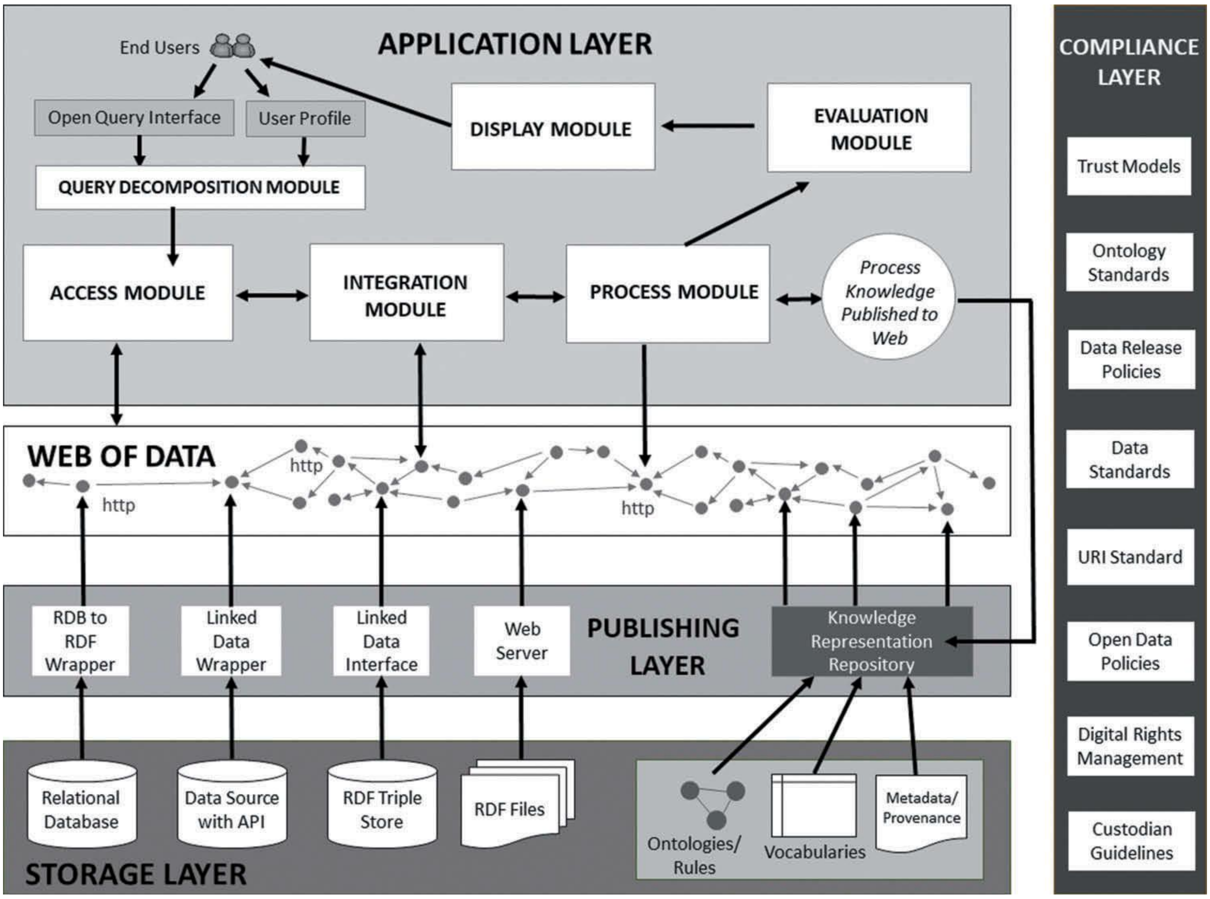

Figure 12 - High-level spatial knowledge infrastructure architecture. Source: Arnold et al., 2021

Arnold et al. (2021) also state that the middleware, or service layer, will be redundant in the proposed architecture. In the simplified architecture (Figure 13), the five layers are narrowed back into two layers, with the web of data replacing the level of the service layer from classic SDI architecture. The two-layered architecture could be explained by the fact that the storage and publishing layer are being seen as complementary to each other and might be seen as one layer, the data layer. In this point of view, the compliance layer would also be part of the data layer. In a simplified way, the result would be a two-layer architecture including a data layer and an application layer that are both directly connected to the web of data(Figure 13) Therefore, the terminology of what a 'layer' is could be depending on the level of detail of the architecture.

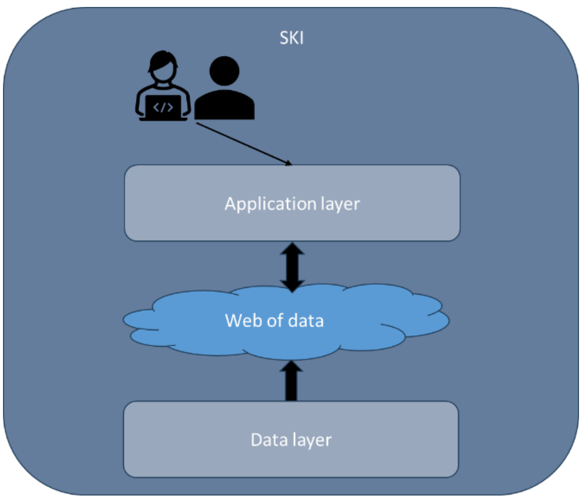

Figure 13 - Simplified high-level architecture of spatial knowledge infrastructure based on Arnold et al., 2021

This means, on a high-level, there is a difference between the proposed architectures of on the one hand Ivánová et al. (2020), Kopsachilis, Vachtsavanis & Vaitis (2023) and Omidipoor (2018) and on the other hand Arnold et al. (2021) regarding the service layer. As Arnold et al. (2021) state data stored could be transformed into Linked Data (LD) to be published and one can make use of a D2R server to deploy related databases as LD, this could be seen as the classic service layer. Thus, in terms of layers the SKI still makes use of a data layer, a service layer and an application layer (Figure 14). Both the service layer and the application layer interact with the web of data. The data layer assumably will still be a data silo.

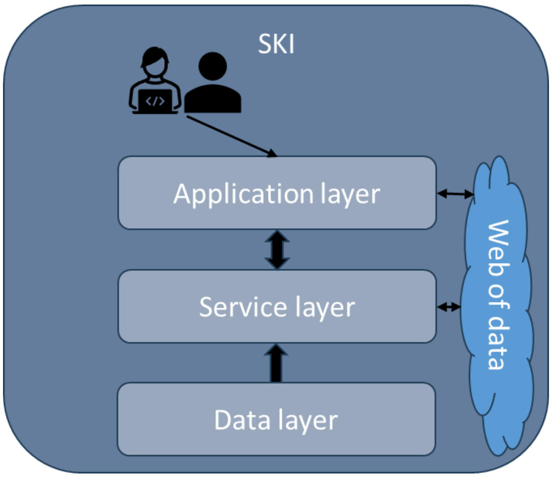

Figure 14 - Simplified high-level architecture of a spatial knowledge infrastructure as proposed by the author

The data stored and managed in the data layer of an SKI is supposed to be directly published on the semantic web and the data should be to real-time (Duckham et al., 2017; Fiedukowicz et al., 2012; Geospatial World, 2022). The storage of the data is therefore cloud-based (Geospatial World, 2022). In order to do so, the data standards should comply with standard set by the W3C, like RDF, HTTP and URI (Woodgate, 2017; Arnold et al., 2021; Kopsachilis, Vachtsavanis & Vaitis, 2023). Though this is only possible when the data is linked, for which more domain ontologies must be created as it prevents duplicates and irregularities (Stock et al., 2012). As for the production of data, the SKI should harvest data from diverse producers like the Internet of Things (IoT) or crowdsourcing, instead of only authoritative agencies (Duckham et al., 2017; Geospatial World, 2022). Regarding the maintenance of the data, Geospatial World (2022) advocates to keep it authoritative to ensure data quality.

The SKI application layer should be designed to support non-geospatial experts as well (Arnold et al., 2021; Duckham et al., 2017; Geospatial World, 2022). This should be supported by the querying possibilities. Yet, Arnold et al. (2021) and Kopsachilis, Vachtsavanis & Vaitis (2023) state SPARQL is already an enhanced way of querying. As SPARQL is still a form of SQL, the non-geospatial expert group changes into ‘regular’ IT, or web, experts. Other authors (Duckham et al., 2017; Geospatial World, 2022; Ivánová, Armstrong & McMeekin, 2017; Woodgate et al., 2017) state the querying should substantiate everyday language. In this case, every citizen is able to use an SKI. The output of the querying should deliver ready to use answers or information enhancing the possibilities of on-the-fly data analysis for the user (Geospatial World, 2022; Ivánová, Armstrong & McMeekin, 2017; Woodgate et al., 2017). The spatial dimension of the data, this should be at least 3D (Section 3.2.2). Furthermore, the application layer should support mobile devices as well instead of only desktop (Arnold et al., 2021; Duckham et al., 2017; Geospatial World, 2022).

3.4.3 Differences

On a high-level, the architecture of both SDI and SKI are similar where they both have a data(management) layer, a service layer and an application layer. The main difference between both architectures is the interconnectedness of the SKI with the web of data (Duckham et al., 2017; Fiedukowicz et al., 2012; Geospatial World, 2022) in contrary to the SDI. An SDI is perceived to be an isolated system. The difference is visualized in Figure 15. This figure could also clarify what is meant with an isolated or open system. This should not be confused with open SDI. The SKI is connected to the web of data and can also use and include data sources within its infrastructure. Open SDIs aim for having diverse producers of data and making its data available for everyone via web dowloads (e.g., via APIs) (Mulder et al., 2020; Welle Donker & Van Loenen, 2017), but the SDI as concept is usually not considered to be connected with external data sources (Tripathi, Agrawal & Gupta, 2020; Arnold et al., 2021).

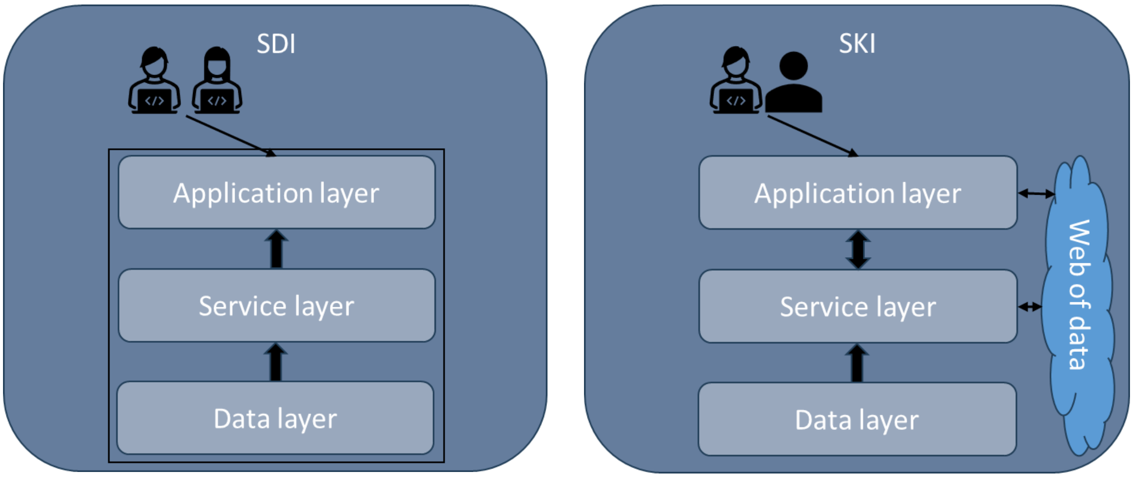

Figure 15 - Simplified architectural differences between the spatial data infrastructure (left) and spatial knowledge infrastructure (right)

Although the differences seem modest when looking at the simplified architecture of the SDI and SKI, a closer look is required. Especially the data layer and the application layer are affected by the differences between the SDI and SKI. The data of an SKI should be real-time (Duckham et al., 2017; Fiedukowicz et al., 2012; Geospatial World, 2022) whereas using SDI data is usually seen as looking backwards (Coetzee et al., 2021). This could be possibly to who the presumed data producers are. Developments in open SDIs aim to incorporate non-authoritative data producers (Mulder et al., 2020), but the producers are in general considered to be authoritative (Duckham et al., 2017; Geospatial World, 2022). The data has to be published directly, which will be done by the service layer. However, it requires that the standards of the data have to be semantical web ready (e.g., RDF) (Arnold et al., 2021; Kopsachilis, Vachtsavanis & Vaitis, 2023; Woodgate, 2017). Despite developments in open SDIs (Izdebski, Zwirowicz-Rutkowska & Nowak da Costa, 2021), data in SDIs are usually GI-domain specific (Cipolloni, 2018; Granell, Díaz & Gould, 2010; Omidipoor, 2018). Furthermore, SKIs should be cloud-based whereas are generally perceived to be local stored (Duckham et al., 2017; Fiedukowicz et al., 2012; Geospatial World, 2022). Though it must be noted that this shift is also a topic of interest in SDI literature (Tripathi, Agrawal & Gupta, 2020). Furthermore, the data of an SKI should be 4D in order to support modeling (Arnold et al., 2021; Geospatial World, 2022; Ivánová, Armstrong & McMeekin, 2017) whereas an SDI usually is 2D (Béjar et al., 2009; Duckham et al., 2017; Woodgate et al., 2017). 4D means that the data should have both a Z-value (height) as well as a timestamp (temporal). According to the SKI literature, assuming SDIs usually have 2D representation (Arnold et al., 2021; Geospatial World, 2022; Ivánová, Armstrong & McMeekin, 2017; Woodgate et al., 2017), either the data misses this information or the application layer is not able to process this.

The application layer is most affected by the user as the user interacts with the application layer. Although open SDI aims for including non-geospatial experts as users (Mulder et al., 2020; Vancauwenberghe et al., 2018), the SDI is, in general, still considered to be an expert system (Duckham et al., 2017; Geospatial World, 2022; Granell, Díaz & Gould, 2010). The SKI also aims for including non-geospatial experts. However, the literature is undecisive whether the users should be IT experts or citizens. This distinction can be made by the way the data can be queried. Querying an SDI is usually simple and hardcoded (Duckham et al., 2017; Geospatial World, 2022; Granell, Díaz & Gould, 2010). The output of such queries will result in either too much information, or data that have to be downloaded and processed by the user. For SKI, some of the literature state SPARQL should be used, which is not a computer language everyone understands (Arnold et al., 2021; Kopsachilis, Vachtsavanis & Vaitis, 2023). Other literature state speaking language should be supported which makes it possible for everyone to use the interface (Duckham et al., 2017; Geospatial World, 2022; Ivánová, Armstrong & McMeekin, 2017; Woodgate et al., 2017). The result of these queries should provide the answer the user is looking for as well as it should support the possibility of conducting on-the-fly data analysis (Geospatial World, 2022; Ivánová, Armstrong & McMeekin, 2017; Woodgate et al., 2017). Furthermore, an SKI should be working on mobile devices while an SDI is mainly desktop focused (Arnold et al., 2021; Duckham et al., 2017; Geospatial World, 2022; Granell, Díaz & Gould, 2010).

4. Conceptualizing the evaluation of going from Spatial Data Infrastructure to Spatial Knowledge Infrastructure

The differences between the concepts identified in the sub-sections of Chapter 3 lead to a conceptual framework that makes it able to evaluate the state of a spatial infrastructure and to what extent it is an SDI or an SKI.

4.1 Conceptual model

The differences are used for identifying a set of parameters constituting the conceptual model. These parameters make it possible to evaluate a spatial infrastructure. When reviewing all the differences, two perspectives can be distinguished that categorize the parameters and serve as basis of the conceptual model. The differences can be viewed on the one hand from a data perspective and on the other hand from a user perspective.

As for the data perspective, the first difference identified is regarding the standards. SDIs commonly use GI-domain (non-semantic web) standards (Granell, Díaz & Gould, 2010; Omidipoor, 2018) whereas an SKI needs to use semantic web standards (Arnold et al., 2021; Kopsachilis, Vachtsavanis & Vaitis, 2023; Woodgate, 2017). A second difference is the producer. Despite developments in open SDIs, the data producers in SDIs are generally considered to be authoritative producers while Data in SKIs should be produced by diverse producers (Duckham et al., 2017; Geospatial World, 2022). The third difference is the storage of the infrastructure. The entire SKI infrastructure should be cloud-based while the assumption is made by the SKI literature (Duckham et al., 2017; Fiedukowicz et al., 2012; Geospatial World, 2022) is not entirely cloud-based. The fourth difference between an SDI and SKI is whether the data is real-time or looking backwards. An SDI normally has no real-time data but the user is always looking backwards while an SKI should have real-time data (Arnold et al., 2021; Coetzee et al., 2021; Duckham et al., 2017). However, conceptualizing real-time data is two-sided. Firstly, the frequency the data is updated is important because if the data is only updated periodically instead of (near) real-time, it cannot be defined as real-time. Secondly, it is assumed that the data should be updated automatically instead of manually, while real-time data requires ongoing updates which will not be possible when a person has to do this manually. The last difference from a data perspective is regarding the dimensions of the data. An SDI is assumed to mainly have 2D data whereas the SKI requires to have 3D/4D data (Arnold et al., 2021; Béjar et al., 2009; Geospatial World, 2022; Ivánová, Armstrong & McMeekin, 2017; Woodgate et al., 2017). This difference also needs some extra. 3D data needs x,y,z-coordinates whereas 2D only needs x,y-coordinates and is therefore included in the conceptual model as the 'spatial dimension'. The fourth dimension (4D) requires that the data should have time stamps for making travelling through time possible being depicted as the 'temporal dimension'. These five differences result in seven data parameters to be included in the conceptual model.

From a user perspective, the level of expertise is identified as difference where an SDI usually requires to be an geospatial expert (Duckham et al., 2017; Geospatial World, 2022; Granell, Díaz & Gould, 2010) and an SKI also includes non-geospatial expert. For SKI, the literature is to some extent contradictory in who is assumed to count as non-geospatial expert as it could be either regular IT-experts or citizen. This distinction is based on the query input. The SPARQL language as way of querying is mentioned in some SKI literature (Arnold et al., 2021; Kopsachilis, Vachtsavanis & Vaitis, 2023), while other literature (Duckham et al., 2017; Geospatial World, 2022; Ivánová, Armstrong & McMeekin, 2017; Woodgate et al., 2017) state the input should support everyday language. However, both ways of querying are considered to be more advanced than the usually simple querying that can be done in SDIs (Duckham et al., 2017; Geospatial World, 2022; Granell, Díaz & Gould, 2010). This, on its turn, impacts the query output. The query output is being divided into two parameters. Firstly, the output suitability which indicates whether the user only gets the information needed and required for (advanced querying presumed to be SKI) or whether too much information is presented (simple querying assumed to be SDI). Secondly, the output readiness is defined to determine if the output needs processing (SDI) or can be used directly (SKI). Furthermore, device readiness is mentioned. SDIs usually focus on desktop whereas SKIs should support mobile devices (Arnold et al., 2021; Duckham et al., 2017; Geospatial World, 2022; Granell, Díaz & Gould, 2010). Analysis possibilities (analytics) is also a difference identified. The SKI literature (Arnold et al., 2021; Duckham et al., 2017) states an SKI should support in some form of analytics (on-the-fly analysis), implying an SDI usually does not have analysis possibilities and the analysis should be done afterwards (post-analysis). The same rhetoric goes for modeling. It is assumed that SDIs generally do not have modeling possibilities, but this should be possible in an SKI to conduct spatial-temporal modeling (Geospatial World, 2022; Woodgate et al., 2017). So, seven user parameters can be distinguished.

In all, a total of 14 parameters that have been established to evaluate to state of the spatial infrastructure. Seven data oriented parameters and seven user oriented parameters. The developed conceptual model can be seen in Figure 16.

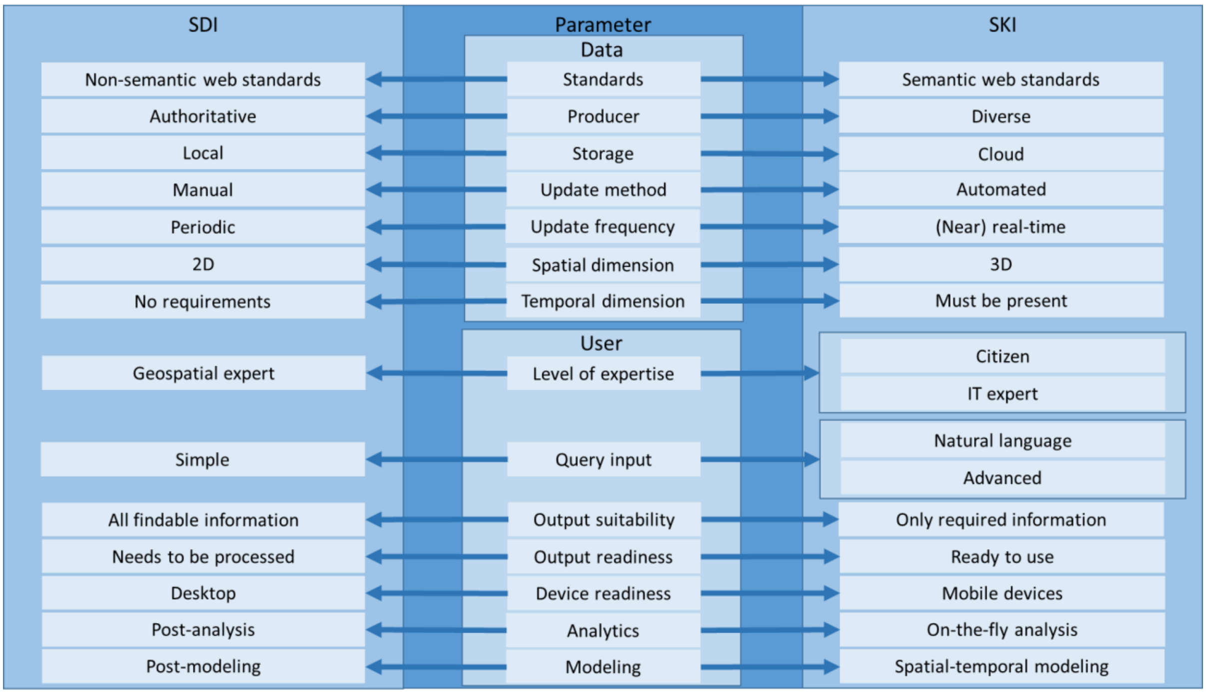

Figure 16 - Conceptual model with a set of data and user parameters to evaluate a spatial infrastructure

Despite that the parameters have been set, this is not sufficient to make a justified evaluation of the state of the spatial infrastructure. For one, SDIs evolve over time incorporating new technologies and insights, like applying a 3D spatial dimension (Stoter et al., 2011), enabling storage in the cloud (Tripathi, Agrawal & Gupta, 2020). Though, these examples only focus on one of the identified parameters. Additionally, spatial infrastructures consist of a wide arrange of data sources presumably produced and managed by different data producers. When the different data producers manage their datasets in different ways, the infrastructure where the data is coming together could have partly SDI-based data and partly SKI-ready data. So, the infrastructure as a whole will gradually evolve as well as the individual parameters that identify the infrastructure to be an SKI. This also makes the state of the individual parameters expected to be not as binary as depicted by the conceptual model. By operationalizing the parameters, it is possible to define an intermediate (or in-between) state where the parameter is neither considered to be completely SDI nor completely SKI.

4.2 Operationalization

Due to the fact that the state of the parameter in general is not black or white, it is needed to further operationalize them. This operationalization is dependent on the parameter in question.

4.2.1 Data parameters

The data parameters say something about the origin of the data, whether the data can be integrated into the semantic web and if the data updated real-time. Out of the data oriented differences, six parameters can be derived.

- <u>Standards</u>

Explicitly mentioned are standards. The standards needed to integrate the spatial data within the semantic web have to comply with the standards set by the W3C. Examples of these standards mentioned in the SKI literature are RDF and OWL. This parameter has as aim to identify which standards are being used and whether the data in current form is already integrated or ready to be integrated within the semantic web. Using these standards will also increase the readability for both man and machine. In case the standards need to be adapted to comply with the semantic web standards, it is not considered to be an SKI. An intermediate form between the traditional SDI and SKI could be web standards (e.g., REST API). Web standards are also set by the W3C and enhance interoperability and web accessibility, but are not designed specifically for the semantic web and thus need some sort of processing to be integrated into the semantic web.

- SDI: No semantic standards
- In-between: Web standards or using semantic standards for some datasets
- SKI: Using semantic standards for all datasets

- Producer

The producers of spatial data in SDIs are considered to be authoritative agencies like governments or spatial authoritative organizations. As for SKIs, the producers of data can be diverse. Diverse producers mean the SKI also uses data produced by new sources like crowdsourcing, the IoT, or wearables (e.g., smartphones and smartwatches). To evaluate the spatial infrastructure, there is no real intermediate state when it comes to the producer of the various data sources included. Either all the data is authoritatively produced and thus considered SDI or at least part of the data sources are stemming from diverse producers. It is important to distinguish the producer of the data and the agency maintaining the data. If the data is produced in a diverse way (e.g., crowdsourcing), but the agency that maintains the subsequent dataset of this produced data is an authoritative agency, the producer of the data itself is considered diverse and thus SKI.

- SDI: Only authoritative datasets
- SKI: Also non-authoritative datasets included

- Storage

The storage parameter does refer to the system as a whole. In an SDI, data is stored local (or regional) in data warehouses SKIs have the system operating in the cloud, from data to applications. This should enhance cloud computing and analytics, but also has better possibilities for scaling up (or down).

- SDI: Stored local or in a data center
- SKI: Cloud-based solution

- Update method

The data accessible within the infrastructure needs to be updated when new data becomes available. This updating can be done either manually or automatically. For this parameter, it is perceived that the update scheme of an SDI is done manually and updates within an SKI are done automatically. Automated updates increase the possibility of having real-time data. As a spatial infrastructure contains multiple data sources, possibly from multiple producers, it is possible that the different data sources have different update schemas. If all the data in the infrastructure need manual action to be updated, the data will never be as real-time as possible and therefore considered to be SDI. If all data is updated automatically, e.g., by having an API connection with the publisher of the data, the infrastructure is as updated as possible and therefore an SKI. When part of the data has to be updated manually and some data is updated automatically, the infrastructure is in an intermediate form between SDI and SKI. However, only automating updates instead of doing it manually does not fully ensure real-time data. Complementary to the update schema, is the update frequency.

- SDI: Data is updated manually
- In-between: Some of the data is updated automatically, some data is updated manually
- SKI: All data is updated automatically

- Update frequency

The update frequency refers to how often the data sources are updated. In order to provide real-time data, the data sources should be up to date. The frequency of updates is both dependent on the data source as well as the administrator or update method of the infrastructure. For example, borders of municipalities do not change monthly or even yearly and therefore could be updated irregularly. Other data, like addresses and functions of buildings, could change daily. If the source data publishes new versions regularly it is still needed for a spatial infrastructure to get and deploy this new data as well. Overall, the aim of an SKI is to have real-world changes processed as soon as possible, but no later than near real-time. In an SDI, the data is not updated frequently enough and real-time analyses will not be possible. Just as with the update schema, it is dependent on the various data sources to what extent the infrastructure is an SDI, SKI or somewhere in-between.

- SDI: The data within the infrastructure is (near) real-time
- In-between: Some of the data is (near) real-time, some data is outdated
- SKI: All data is updated periodically resulting in an information gap

- <u>Spatial dimension</u>

One important characteristic of spatial data is that it has a spatial dimension. This is also reflected by using the terms of a 2D or 3D dimension. In SDI, the data will be 2D (X and Y axes). Also 2.5D (including height) is still considered a flat representation and therefore considered to be SDI. In SKI, the minimum requirement is to have 3D representations (X, Y and Z axes) of the data. A 4D representation in SKIs is also mentioned in the literature, meaning changes over time can be seen. This temporal dimension will be discussed as separate parameter. For the spatial dimension, the difference between SDI and SKI will be whether the data is 2D or 3D. When evaluating the infrastructure, also for this parameter it is necessary to think about which dimension is suitable for the given data source. Not all spatial data has, or requires, three dimensions (e.g., borders of a municipality). If data has three dimensions (e.g., buildings), the data should be in 3D to comply with being an SKI. Like the former two parameters, it could therefore be that the system is in an intermediate state where part of the data is in still in an 2D (SDI) state and some data is already 3D (SKI).

- SDI: All data is 2D
- In-between: Some potentially 3D data is still in 2D
- SKI: All potential data is 3D

- <u>Temporal dimension</u>

This parameter, regarding the temporal dimension, is linked with the modeling parameter. Objectives for an SKI are 4D and spatial-temporal modeling. As the spatial component is evident in a spatial infrastructure, this is not the case for the temporal component. In order to make changes over time visible or facilitating modeling over time, the data needs to have some sort of time stamp. This could be either by including multiple versions of the same dataset or to combine multiple versions of the same dataset into a new dataset. In the first case, the versions over time stay separate datasets with its own time stamp. If a new dataset is created by joining the different versions, the time stamp needs to be a value within the dataset for distinguishing the differences over time. Therefore, an SKI does require all data to have different versions in time that can be accessed to model differences over time whereas this is not necessary for SDIs.

- SDI: No possibility to access different data versions in time
- In-between: Some data has different versions in time that can be accessed
- SKI: For all data different versions in time can be accessed

4.2.2 User parameters

For the user perspective, seven parameters are derived from the differences between SDI and SKI.

- <u>Level of expertise</u>

SDIs are developed for users active in the geospatial domain and SDIs are still considered to support only geospatial experts as users. The SKI aims to support non-geospatial users as well. In the SKI literature, the definition of non-geospatial users is varying. Both 'regular' IT experts are mentioned as SKI-user group as well as every citizen who use the SKI to get answers on questions they have. Based on current SKI literature, the infrastructure is considered to be SKI in both cases. This would make the aim of the system leading in to whether it fulfills the criterion regarding the level of expertise. The result for evaluating an infrastructure based on the level of expertise of the user can be twofold. It could be possible that the aimed new user group is the general IT expert, all citizens or both.

- IT experts:
    - SDI: Only usable for geospatial experts
    - SKI: Also usable for general IT experts
- Citizens:
    - SDI: Only usable for geospatial experts
    - SKI: Also usable for citizens

- <u>Query input</u>

Querying of data in SDIs can only be done in a hardcoded way and supporting simple queries. In SKIs, the data is linked and part of the integrated web, making it possible to perform complex queries. Following up on the different interpretations of the previous parameter, the form which this querying should support is depending on the aimed user groups. In case the identified non-geospatial users are IT experts, the proposed querying is based on SPARQL. SPARQL differs from SQL in terms of the data(sets) it is looking for, as SPARQL is a semantic web standard for querying. In case the user groups are everyone interested in information from the SKI, the querying should support everyday language as input. Comparable with the level of expertise, it depends on the aim of the infrastructure what the state of the system is. One possibility is that the everyday language is supported to a certain extent, but not fully. For example, it might be possible to use certain keywords but no full sentences.

- IT experts:
  - SDI: Only simple queries in SQL
  - SKI: Complex queries in SPARQL
- Citizens:
  - SDI: Not usable for citizens
  - In-between: Certain words can be used, but not full written sentences
  - SKI: Asking questions in natural language

- <u>Output suitability</u>

The output of a query performed in an SDI will give all information that meets the query criteria. While the queries are somewhat simple and hardcoded, the user often receives an abundance of information and needs to process and sort the information to find the information needed. In an SKI, it is envisioned that the user only retrieves the information the user is looking for. This implicates the suitability of the output. Notable is that the suitability of the output is somewhat intertwined with the input. For instance, when the querying can be more specific than using quite generalized queries, the number of irrelevant output could decrease.

- SDI: Information that needs to be analyzed for suitability
- SKI: Information that is meeting the users' requirements

- <u>Output readiness</u>

The following step is the readiness of the output. In case of an SDI, the retrieved information still needs processing. The processing could be filtering redundant information or transforming the data to make it suitable for analysis. The output in an SKI should be ready to use, the information can be used as-is.

- SDI: Data needs to be processed
- SKI: Data can be used as-is

- <u>Device readiness</u>

The SDI is predominantly supporting desktop. In first instance only programs on desktops that were locally connected. Later, also web browsers were supported. The SKI is supposed to support mobile devices (e.g., tablet or smartphone) as well by either having applications for the device or via browsers on the mobile device. When evaluating the infrastructure, when only desktops are supported the infrastructure is an SDI. In a mobile-first society, an SKI is assumed to support smartphones.

- SDI: Only desktop browsers are supported
- SKI: Mobile devices are supported

- <u>Analytics</u>

Within an SDI, it is not possible to perform an analysis. This is partly related to the output suitability and readiness of the data. Data still needs to be processed by the user in order to make analysis possible. Processed predominantly means that the data have to be downloaded to be used on a local (GIS) system by the user. In an SKI, it should be possible to perform analyses on-the-fly within the infrastructure. Regarding the state of the infrastructure, it is either possible to perform analysis or it is not possible. Nonetheless, it might be possible that only certain data within the infrastructure can be used for analysis while other data cannot.

- SDI: No analysis possibilities
- SKI: Analysis can be done within the infrastructure

- <u>Modeling</u>

As stated in the spatial dimension, predictive modeling is one of the characteristics of an SKI. The aim of predictive modeling is to visualize future outcomes of certain changes. Modeling changes over time and space are called spatial-temporal modeling. Therefore, the SKI should be able to facilitate spatial-temporal modeling. In SDI, just as with analytics, it is necessary for the user to download the static data to be able to do some kind of modeling. While no modeling versus spatial-temporal modeling does describe whether the infrastructure is either SDI or SKI, some modeling possibilities indicate that the infrastructure is in-between concepts. One possibility of this intermediate state would be the possibility to spatial modeling. In spatial modeling, the temporal, or change over time, is still absent.

- SDI: No modeling possibilities
- In-between: Only spatial modeling possibilities
- SKI: Spatial-temporal modeling can be done within the infrastructure

4.3 Case study methods

The evaluation of the case studies will be based on the operationalization of the 14 parameters that have been developed for assessing the state of a spatial infrastructure. Three methods are included.

4.3.1 Content analysis

The content analysis will be based on the available documentation of the case in question, such as descriptions on the website or within the interface of the case and other findable written documentation about the case. The aim of the analysis is to get a general description of the case study and will be used for evaluating the parameters as much as possible.

4.3.2 Usability study

The usability study predominantly focuses on the testing method of a UEM, while the inspection method is part of the content analysis and the inquiry method will be the semi-structured interview. The testing will be done on the UI of the case study in question. The testing will look at which features and functionalities are present and to what extent these features and functionalities are usable and operational. Therefore, the aim is to evaluate some of the parameters. For example, whether there are query possibilities and if so, what the output will be. Other examples of parameters that will be evaluated are the capabilities in terms of data analysis and modeling and if the data are in 2D or 3D.

4.3.3 Semi-structured interviews

For 3D Amsterdam, the interview will be held with the product owner (PO). The interviewee from the Kadaster Knowledge Graph is the product lead. The case specific topic lists are depicted in Appendix E – Interview scheme 3D Amsterdam - Dutch (3D Amsterdam) and Appendix F – Interview scheme Kadaster Knowledge Graph - Dutch (Kadaster Knowledge Graph). As the interviewees are Dutch, the topic lists are set up in Dutch, a generalized English version of the topic list is shown in Appendix D – Generalized interview scheme - English. The interviews start with an introduction followed by some general information about the objectives and the current state of both the case study. Thereafter, the questions based on the 14 parameters will be discussed.

5. Results

The identified set consisting of seven data parameters and seven user parameters will now be applied to evaluate two real-world cases in the field of spatial infrastructures. Both cases are situated in The Netherlands and developed by authoritative bodies. The first case will be the Kadaster Knowledge Graph (KKG) and the second case will be 3D Amsterdam. The analysis will be based on the three research methods as described: a content analysis, a usability study and one semi-structured interview per case. The sections start with some general information about the case studies and its objectives, followed by a general overview of the architecture of the spatial infrastructure after which the data parameters and user parameters will be evaluated for the cases in question. The sections ends with a synthesis of the two cases.

5.1 Case study – Kadaster Knowledge Graph

The Kadaster Knowledge Graph (KKG) is developed by the Dutch cadastral agency Kadaster, which has the task to maintain various (national) spatial datasets. The KKG is part of Kadaster Labs (Figure 17). The 'Labs' environment is invented to explore the newest technologies aiming to overcome spatial challenges (Kadaster, n.d.a). In the Labs environment, the results of these explorative tests or studies are published.

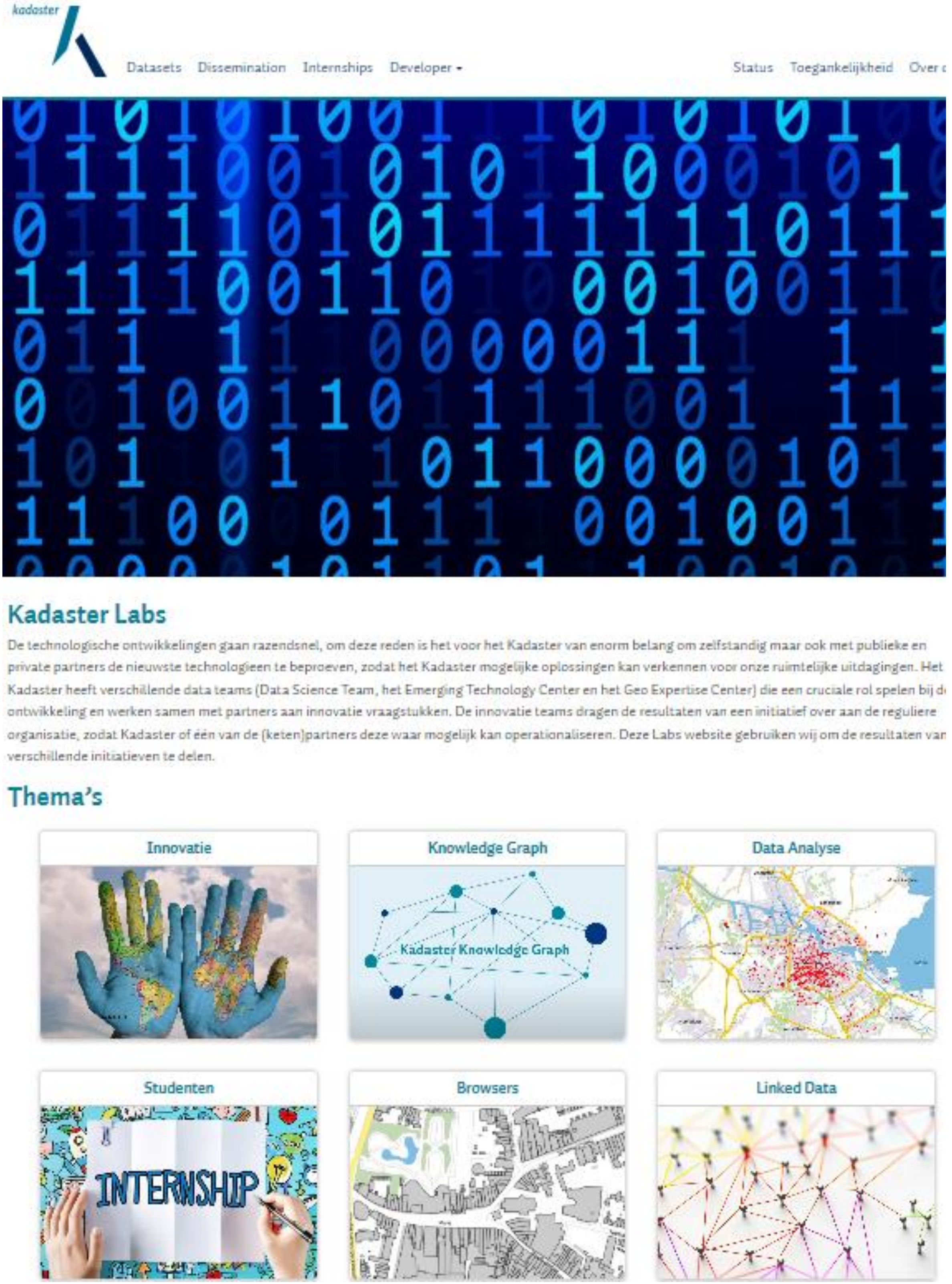

Figure 17 – Snapshot of the Kadaster Labs environment with some general information about what Labs is and the different themes. Source: Kadaster Labs[1]

The aim of the KGG itself is to integrate multiple governmental open spatial datasets and using the semantic web to make integration of additional dataset s possible leading to an extensive network of spatial data (Kadaster, n.d.b). The available open datasets are maintained by Kadaster (Kadaster,

[1] https://labs.kadaster.nl/

n.d.c). Kadaster tested this by multiple demonstrators and applications like an AR-app, self-service GIS and the already operational chatbot. During the interview with the product lead of the KKG, it was stated that at present time (mid 2024) the Labs environment, where the KKG is published and presented to the world, is actually the acceptance environment. They are working on getting the KKG fully operational on a production environment and to do so, the most important thing was to get the governance of the infrastructure in place. The evaluation of the case study will be based on the current version of the spatial infrastructure. Notable is the consequence of the focus, while this means that currently the upkeep and development of the KKG is on hold until the KKG has gone formally live. Another important outcome of the interview is the state of the product (KKG) that will be brought into the production environment. As said, the current KKG includes multiple demonstrators of applications that can be developed based on the KKG (e.g., AR-app, chatbot). However, most of them will not be part of what Kadaster envisages to facilitate with the KKG. When brought to production, it includes a SPAQRL endpoint making it possible to query the data and the architecture in an as-is state.

5.1.1 Architecture

Kadaster published both the envisioned architecture of infrastructure as well as the current status (Figure 18). Although it is updated in October 2021 (Kadaster, n.d.d) for the last time, it is still the state of the architecture as of mid-2024. The architecture follows the same layered structure as in case of both SDI and SKI, consisting of a data layer, service layer and application layer.

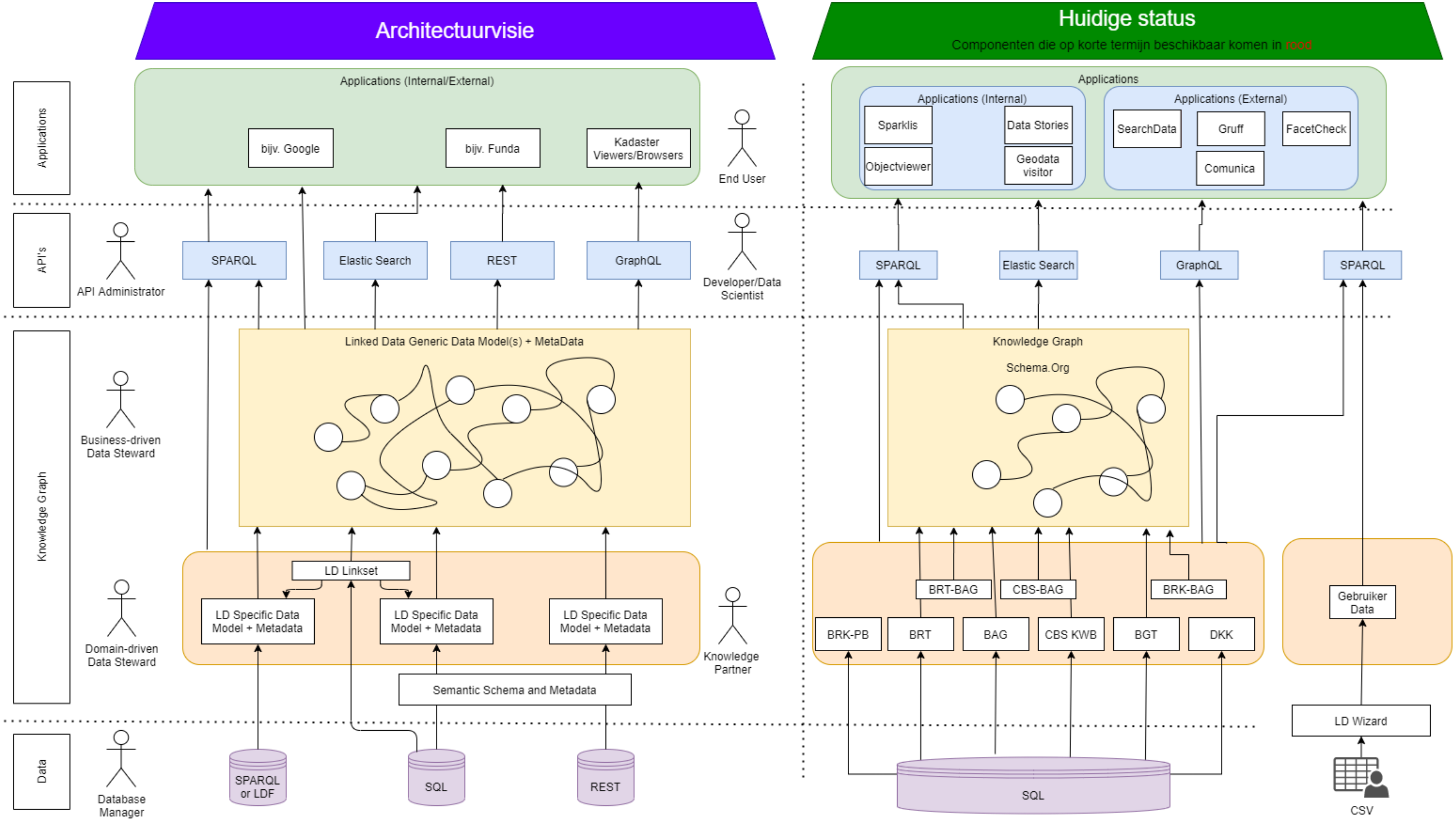

Figure 18 - Envisioned architecture (left) and current architecture (right). Source: (Kadaster, n.d.e)

In the envisioned architecture, the data sources in the data layer (purple) can be approached by the service layer via techniques like SPARQL, SQL and REST (Kadaster, n.d.e). The service layer is visualized as being the 'Knowledge Graph' (KG). In the first step (orange), the data is transformed into a domain-specific LD-model. This model is not yet fit for users who are not familiar with the source data (Kadaster, n.d.e). Therefore, a second step (yellow) is needed to make the data available in a more commonly LD-model which can be understood by common IT experts (Kadaster, n.d.e). After this step, the data can be extracted by API or applications. The API service is predominantly fit for IT experts whereas the applications are fit for every end-user (Kadaster, n.d.e). However, as Figure 18 is showing, this is not the case in the current situation. In the current state, the data sources can only be queried by using SQL. As can be seen in the KG-layer, at the moment only a limited number of datasets are present in the KKG. In the application layer can be seen that only a couple of internal and external applications are connected. For evaluation of the infrastructure, the current architecture will be used.

Currently, the KKG includes six open spatial databases of which one dataset is provided by the Statistics Netherlands (CBS) while the other five are administered by Kadaster (see the orange part of figure 18). Within the KG, some databases are connected to form a new data source. As for the CBS dataset, the product lead of the KKG indicated that pulling this dataset in current form, transforming it and publishing it as part of the KKG is not the most desirable solution. He would prefer to have the CBS itself make their datasets part of a LD infrastructure which can be approached and linked directly to the KKG.

Regarding the applications, the KKG has various demonstrators in the Labs environment. Most of the demonstrators can predominantly be seen as showcases for the possibilities of the KKG as those demonstrators only have videos and some documentation about how it should work, but no possibility of using it. This is in line with what is envisioned for the production environment of the KKG. In basis, the product lead mentioned the envisioned KKG is the infrastructure of the six LD-datasets that can be queried with the SPARQL endpoint. Compared to Figure 18, this means the KG does not necessarily include the application layer. However, the chatbot, which is still operable in the acceptance environment, might be incorporated at a later stage. For evaluating the KKG, the SPARQL endpoint and Loki chatbot will be used in the usability study as these two applications are up and running.

5.1.2 Data parameters

- Standards

Already mentioned is that the KKG transforms the source data into LD, ready to be integrated in the semantic web. GraphQL is being used to generate this LD (Kadaster, n.d.d). The standards used are therefore based on the standards complying for the semantic web. Mentioned are the W3C standards of JSON-LD, SHACL, OWL, PROV, RDF, RDFS and SKOS (4,5). Another standard set by the W3C are URIs, however, the KKG follows the Dutch URI strategy resulting in supporting HTTPS URIs instead of HTTP URIs (Kadaster, n.d.d). This is also substantiated by the product lead of the KKG, who said that most standards used for the KKG are developed by the W3C instead of the previously used GEO-domain standards. The Dutch URI strategy is set by Geonovum. Geonovum is a Dutch governmental organization aiming to enhance the exchange of spatial data by setting national standards. NEN 3610 is another standard of Geonovum that is used by the KKG. The GeoSPARQL standard set by the OGC and is an extension of the SPARQL standard to support spatial queries. However, the GeoSPARQL standard is considered to be complex and therefore not working efficiently as stated by the product lead of the KKG. This means that the published data by the KKG complies with the international standards to facilitate semantics. As for the standards, it can be said that they are SKI-ready.

   - Status: SKI

- Producer

The six connected databases are all authoritative databases. One of the data sources is produced by the CBS. The other five are administered by Kadaster, but for some data sources the data is produced and delivered by other authoritative governments (e.g., municipalities and water boards).

Hence, the KKG only has authoritative produced data making the state of the system an SDI.

   - Status: SDI

- Storage

For developing the KKG, a cloud-based solution has been chosen. Though it must be said that the product lead of the KKG stated that storing closed data in a cloud sometimes makes people anxious about privacy leaks making it possible that such data will not be stored in a cloud but in a data warehouse. He also did not see a direct need for storing the data in a cloud. Nonetheless, the cloud-based architecture means the infrastructure is SKI-ready on this parameter.

- Status: SKI

- Update method

The data sources within the infrastructure are updated manually. Despite it is possible to automate the process and automatically publish the update, it is a deliberate choice to do it manually. The reason for this are possible issues within new versions of the data which could cause issues when updating the data automatically without any checks. Despite the uploads are manually, the data is retrieved automatically by the administrator of the infrastructure. In order to be classified SKI, the data has to be updated automatically. Hence, the current status of the KKG is SDI in terms of the update method.

- Status: SDI

- Update frequency

According to the website (2), the datasets will be updated either quarterly (Kadaster datasets) or yearly (CBS dataset). As the current state of the acceptance environment means there is no regular maintenance or upkeep of the data, the update frequency is currently not met. Nevertheless, it is interesting to see whether the supposedly update frequency meets the SDI or SKI criteria. For the CBS dataset this is as frequent as possible while the CBS publishes this dataset on a yearly base. For the Kadaster administered datasets, three are updated daily, one is updated five times per year and one infrequently. The update frequency has a dependence on the source data, but in this case three of the six connected databases are updated daily at the source, but only every three months in the KKG. Thus, this not meeting SKI standards and considered SDI.

- Status: SDI

- Spatial dimension

In the current situation, all data within the KKG is 2D. During the interview it was mentioned that in the future some data could be 3D, but as of now it is 2D. Therefore, in the current state the KKG is an SDI.

- Status: SDI

- Temporal dimension

As for the temporal dimension, it depends on the data source in question whether there is a temporal dimension present. As said, the main goal of this parameter is to make it possible to see changes over time. One data source (BAG) has historic information available in the KKG due to the history model. For another data source (BRT) all historic is deleted after a new version is published. While some data has a temporal dimension whereas other data has not, the infrastructure as a whole is in-between SDI and SKI on this parameter.

- Status: In-between SDI and SKI

5.1.3 User parameters

- Level of expertise

The current available documentation as well as the use cases indicate the KKG do facilitate non-geospatial experts. In fact, depending on the use case the target group is either a general IT expert or a citizen. For example the chatbot, Loki, is developed for citizens as questions can be asked in natural language and the accompanied documentation is also written in a simple way suitable for non-IT experts. On the other hand, SPARQL does require IT-knowledge but not necessarily Geospatial-knowledge. Important to notice is that this is all within the acceptance environment. When the KKG is going to production, only a SPARQL endpoint will be available. Nonetheless, the product lead of the KKG did state it is envisioned that over time also applications for citizens will be realized, like the Loki chatbot. While both the acceptance environment and the production environment are supporting non-geospatial IT experts, the infrastructure is already in the state of an SKI.

- IT experts:
  - SKI
- Citizens:
  - SKI

- Query input

Already mentioned in the previous parameter is the possibility of using everyday language as input for the chatbot (figure 19). This is a requirement for being an SKI in case citizens are within scope as potential user group. Also mentioned is SPARQL as way of querying and retrieving the data, which is suitable for general IT experts. With the possibility of using SPARQL in order to support general IT experts as users, the KKG is in both the acceptance and production environment SKI-ready.

  - IT experts:
    - SKI
  - Citizens:
    - SKI

- Output suitability

Querying the KKG with the SPARQL end-point is working and providing the right answers to some extent. The SPARQL endpoint is 'equipped' with GeoSPARQL for geographic questions. For this case, it is important to note the difficulties experienced with GeoSPARQL by the product lead of the KKG. Even in benchmark studies GeoSPARQL only scores in a range between 40% and 66%, making geographical queries almost impossible to do even if the infrastructure itself complies with the standard. As can be seen in Figure 19, it is possible to give input in a natural language. It is also clear that the response is not always right. In this example, the input 'show me all schools in the neighborhood' gives no results, but afterwards when the input is just 'schools' there is a result. In the latter case, the result is specific and clear. During other tests, similar bugs were encountered where questions only resulted in errors or a repeating answer. Due to these findings, the current state of the suitability of the query output is partly beyond the state of an SDI, but cannot be considered SKI-ready.

    - Status: In-between SDI and SKI

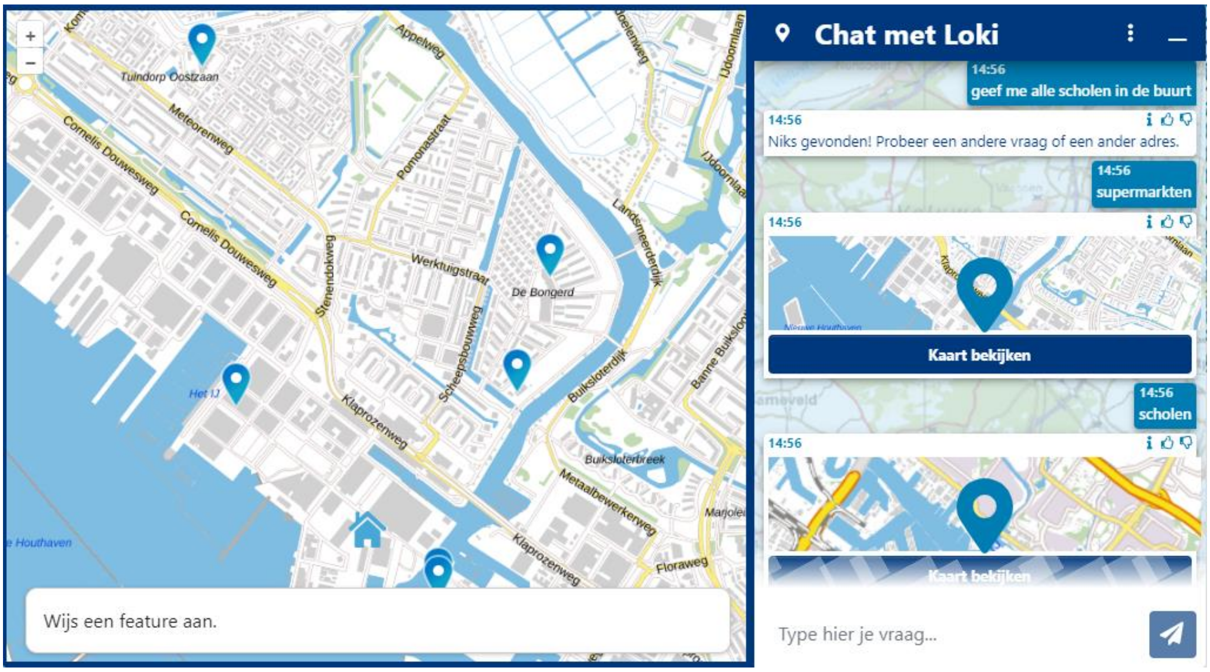

Figure 19 - Input and output of a question for the chatbot, Loki. The input (right) requests all schools in the neighborhood. After no response, a prefilled request for supermarkets popped up after which schools was asked for again. The result (left) of the latest request shows six school locations. Source: Loki chatbot Kadaster Labs[2]

- Output readiness

The readiness of the output is to some extent correlated with the suitability of the output. At least for the chatbot, when no answers are given there is no readiness of the output. Though, the answers that were given were ready to use (Figure 19). In some cases the chatbot referred to another part of the Kadaster website, for example to purchase the information. To some extent, this can be seen as getting information as good as it gets, while incorporating a web shop within the chatbot would feel somewhat strange and therefore not preferable. In conclusion, the output readiness of the chatbot would fit the SKI narrative when working properly. As for the SPAQRL endpoint, the readiness of the data should be

[2] https://labs.kadaster.nl/cases/lokiv3

ready to use. However, the difficulties encountered with GeoSPARQL could impact this parameter. When wrong data or too much information is given, the readiness is still considered SDI. The output in Figure 20 is specific and fit for being classified SKI. Altogether does the KKG does move towards an SKI, but due to the bugs or errors encountered it is not yet considered SKI.

➢ Status: In-between SDI and SKI

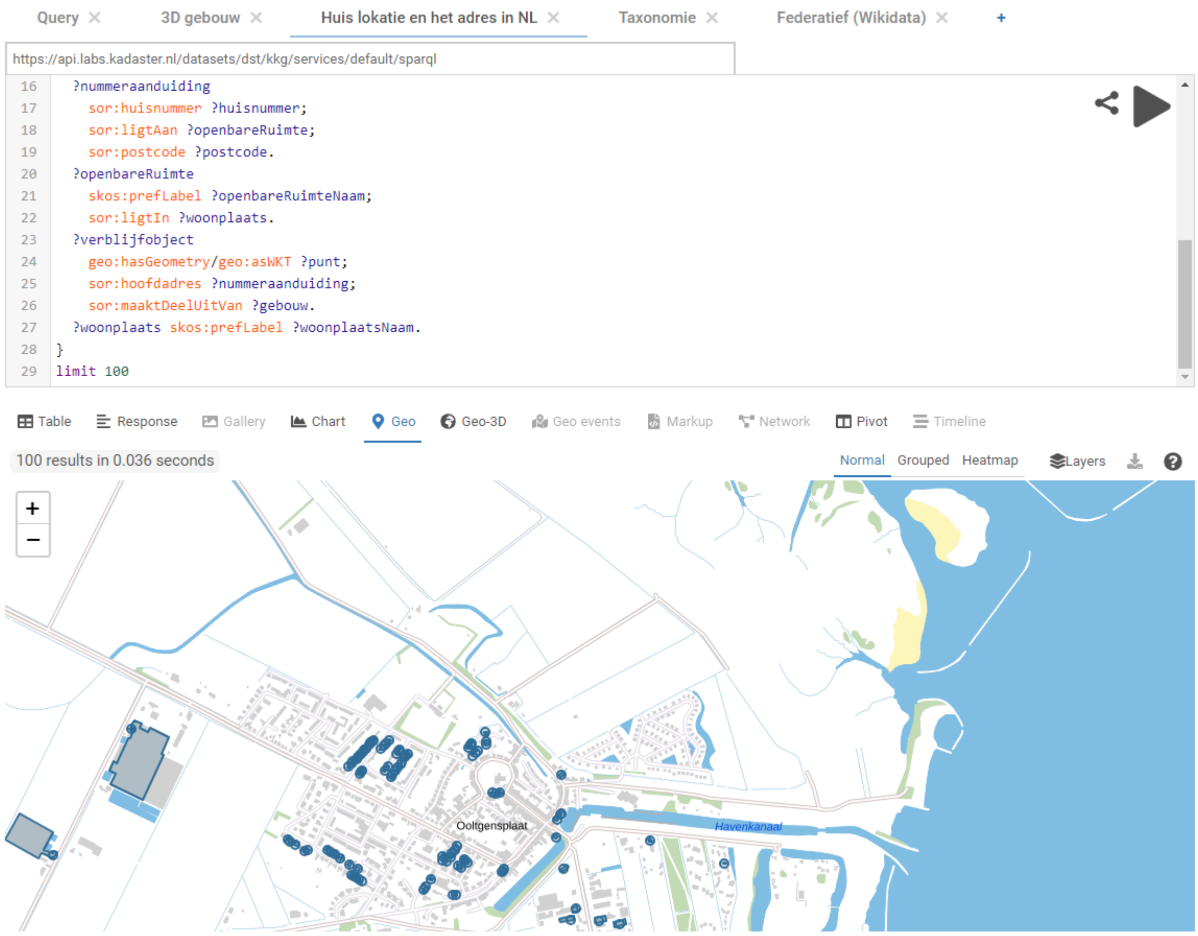

Figure 20 - SPARQL input and output. The query (above) some locations and addresses of houses within the Netherlands. The polygons (below) are visualizing the result of the query. Source: SPARQL Endpoint Kadaster[3]

- <u>Device readiness</u>

The KKG is predominantly supporting desktop browsers. Only the use case of the AR-app is specifically developed for mobile use, but this app is not (yet) available in the app stores and there is no other way to download the app. The product lead also indicated that the KKG is, at least for now, only focused on desktop browsers. This means that the infrastructure will be considered an SDI in both current and envisioned state.

➢ Status: SDI

- <u>Analytics</u>

Both the output of the chatbot as well as the SPARQL output makes is possible to do some form of analysis within the infrastructure. In the future the product lead expects this level of limited analytics will be possible. Only limited, as he considers more complex and extensive analyses will be too heavy for desktop browsers. He does not expect web-based applications can compete with local or specialized tools like a GIS. Although cloud computing should enhance the possibilities of analytics of online applications, at this moment the product lead does not think this technology will compete with specialized tools either. Nonetheless, the complexity of or amount of data used in an analysis is not specified for SKI. Just being able to do 'some' analysis would suffice. Hence, the current state is already considered to be SKI.

➢ Status: SKI

[3] https://labs.kadaster.nl/sparql/

- Modeling

Although the strength of SPARQL is data analysis, the query language can also be used for certain aspects of spatial-temporal modeling like computing simple simulations. Complex spatial-temporal modeling requires other tools like a GIS or programming software (e.g., R or Python). Also, not all data sources have a temporal dimension, so the modeling cannot be performed on all data. While it is technical possible to do a spatial-temporal modeling of some of the data within the infrastructure, despite the possibilities are limited, the parameter is defined as being SKI-ready.

➢ Status: SKI

5.2 Case study – 3D Amsterdam

The second case to which the developed set of parameters will be applied is the case of 3D Amsterdam. While there is no documentation available on the origin of the project, this was first discussed in the interview with the PO of 3D Amsterdam. 3D Amsterdam is a project by the municipality of Amsterdam and started in response to the new Dutch Environment and Planning Act (in Dutch: Omgevingswet). The aims were both to enhance the communication and participation for the citizens of Amsterdam and to create an equal level playing field in information flows. People should be able to understand why a permit was denied or why construction work was taking place in their street. A 3D, or digital twin, of the city should help in answering those questions. A snapshot of the web interface can be seen in Figure 21. The local scene has been set at the Royal Palace of Amsterdam, situated at the Dam in the city center. For the technology of building the city in 3D, a gaming engine (Unity) was chosen. Using a game engine for real-world (governmental) purposes was a new application of such technology, but it was also deemed the only possibility for realizing the 3D twin of the city.

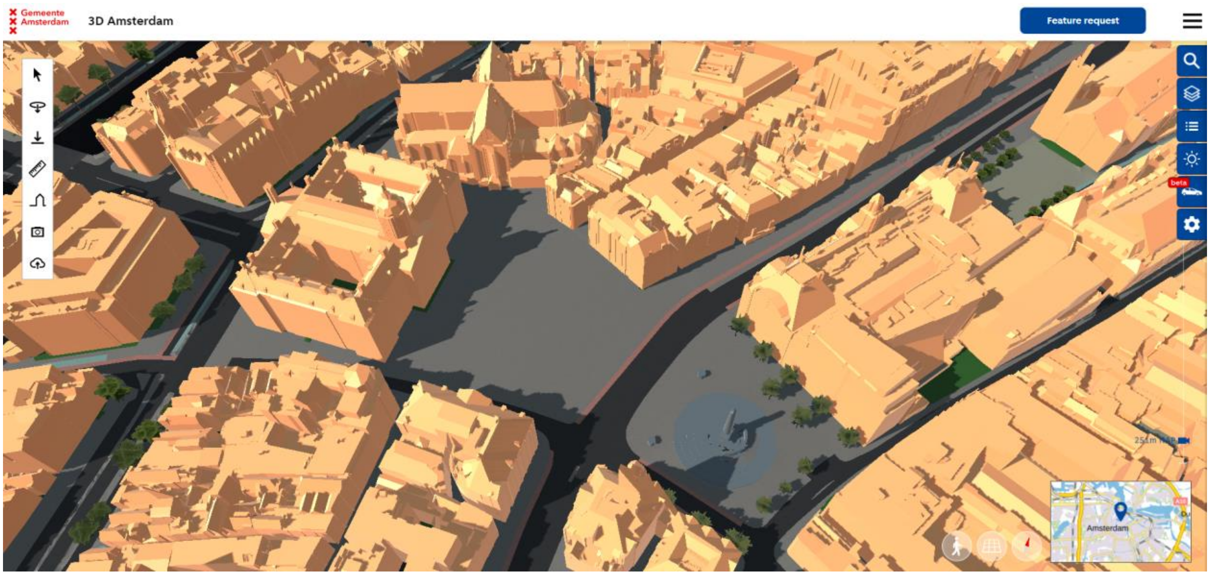

Figure 21 - Snapshot of the 3D Amsterdam interface. Location Royal Palace of Amsterdam. Source: 3D Amsterdam[4]

After a hackathon in 2019, the municipality of Utrecht got interested in the idea resulting in a cooperation of both cities. On its turn, this resulted in even more momentum and parties interested in the possibilities. For instance, the municipality of Rotterdam started 3D Rotterdam, which was based on the two existing projects. At this point in time, the project group developed the idea of scaling up and making a nationwide coverage possible. This idea was substantiated by the fact that the data source used for the 3D visualization of Amsterdam already was a nationwide database meaning all needed data was available. However, the architecture was not scalable or sharable causing to reevaluate the architecture.

[4] https://3d.amsterdam.nl/index.html#121360.74,487243.31,232.91,56.00,344.51,0.00

5.2.1 Architecture

The solution for overcoming scaling and sharing issues was to develop the system in terms of a hub and spoke model. In this solution, a new cloud environment was created that would host the 3D model of the whole of the Netherlands. This new environment or system, is called Netherlands 3D. Netherlands 3D will be the hub, and the local environments are the spokes (Figure 22). At this point, five 3D 'systems' are findable on the web: Netherlands 3D, Flevoland 3D (the province of Flevoland) and the municipalities of Amsterdam, Utrecht and Rotterdam.

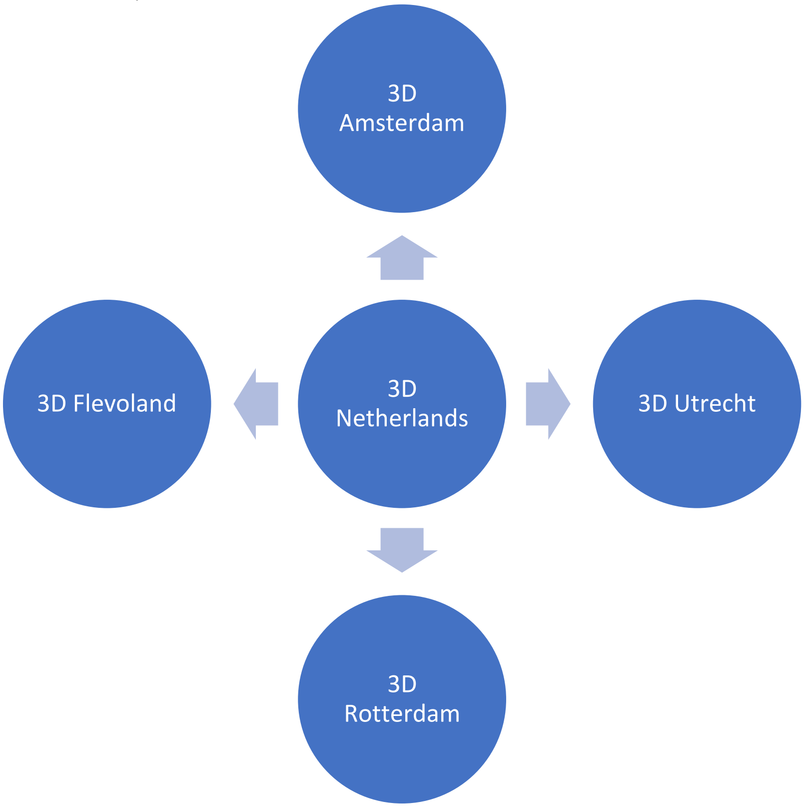

Figure 22 - Visualization of the hub and spoke model for the development of the various 3D projects as based on the interview with the product owner of 3D Amsterdam

In this model, every new function and feature will be developed for Netherlands 3D (hub) and can be requested by the local 3D project groups (spokes) to be integrated. Every separate function or feature developed for the hub can be seen as stand-alone components and be requested by the spokes to integrate into the local environment as a tool. This means, while the overall toolbox for the hub has all possibilities, (local) spokes only integrate the needed tools within its local (cloud-based) environment. Regarding the requests for new features and functions, the team of Netherlands 3D takes wishes and requests, coming from colleagues involved in the spoke projects, into consideration and development, foremost working as city engineer or spatial planning. Suggestions coming in via the browser by everyday citizen do not get prioritized unless it is notified multiple times. The development team predominantly works for the authoritative bodies.

Important to note is that the PO of 3D Amsterdam stated that the 3D Amsterdam team is also responsible for Netherlands 3D. The impact of this is that the team is currently fully working on Netherlands 3D meaning there is no upkeep for 3D Amsterdam and the project, and therefore the spatial infrastructure, is currently on hold. However, since the infrastructure is in a production environment and the fact that Netherlands 3D is based on 3D Amsterdam, this local project is still a promising case study. Because there is currently no monitoring on 3D Amsterdam some functionalities or features do not work properly. Since the other 3D case studies are, more or less, duplicates and consist of the same possibilities, during the case study it is possible that these functionalities or features are studied at one of the other 3D cases.

5.2.2 Data parameters

- Standards

Data of the different systems cannot be approached via web techniques like (REST) API and is not envisioned to be integrated within the semantic web. Limited surface area can be downloaded via the interface in two data formats: Autodesk DXF and Collada DAE (3D Amsterdam, n.d.a). Although these standards are not specifically designed for the Geo-domain, they are also not designed for regular IT experts. Both data formats are developed for professions like engineers, designers and architects used by graphic programs. Due to the fact that the data cannot be classified as either semantic web standards nor as 'general' web standards, the standards are still considered to be in an SDI state despite the fact that the standards might not be the classic SDI standards.

> ➢ Status: SDI

- Producer

Most of the layers in the infrastructure are produced by authoritative agencies. The 'Buildings' (gebouwen) layer has the dataset 3D BAG as its data source. 3D BAG is a collaboration between authoritative agencies and non-authoritative agencies, but is published and maintained by a Dutch university (3DBAG, n.d.b). Although a university can be considered to be an authoritative body, creating and maintaining spatial data is not a core business making the role of the university not authoritative in this regard. Thus, the state of the infrastructure in terms of the data producers is SKI.

> ➢ Status: SKI

- Storage

Although the hub and spoke model could suggest both the hub as the spokes are stored at the same location and are in fact one database, this is not the case. Both 3D Amsterdam as well as Netherlands 3D are both stored in the cloud, but in their own environment. The model works as such that the project group for Netherlands 3D facilitates in developing the desired software and data of the local scene, but this must be implemented by the local organization into their own product. Nonetheless, the cloud-based solution makes this parameter SKI-ready.

> ➢ Status: SKI

- Update method

Updates for new versions of the data sources are done semi-automatically. There is an API connection with the source and new versions will be downloaded automatically and a notification is given to the administrator that new data is available. However, as the PO shared deploying the update is done manually. The main reason given for this method is the fact that checks are performed on the new versions if it can be published without problem. During a previous update, the size of one tile was altered into new dimensions without any prior notice. This caused the whole infrastructure to fail as the dimensions between the data source and infrastructure deviated. Interesting in this case is the information on Utrecht 3D about the 'Sewer' (rioolnetwerk), stating that a change in the source data leads to a direct change within the application layer. This assumes already some data is updated automatically. According to the parameter, this makes the infrastructure in an intermediate state between SDI and SKI.

> ➢ Status: In-between SDI and SKI

- Update frequency

Within the dashboard, for some data layers the date of collection is known as well as the supposedly update frequency. For example, for the layer 'Trees' (Bomen) it is stated the data could lie three years behind. Nevertheless, the year of collection is stated to be 2019 making the lag at least four years. This emphasizes the fact that the upkeep of the infrastructure is on hold. For other data sources the date of collection or update frequency is not mentioned (e.g., the layer 'Other' (Overige). Whether data is up-to-date or what the update frequency should be is not known for those data sources. For the data sources that are produced by other organizations, there is also a dependency on the update frequency of those sources. Although 3D Amsterdam could always use the latest version, when the latest version is outdated, the criterion of supplying real-time data cannot be met. In this case, 3D Bag is such source. One of the input data for creating the 3D model is dated in 2022 (3DBAG, n.d.a). With the regular (2D) BAG being updated on a daily basis, this means 3D Amsterdam is not (near) real-time (Figure 23).
Therefore, the update frequency is still SDI.

> ➢ Status: SDI

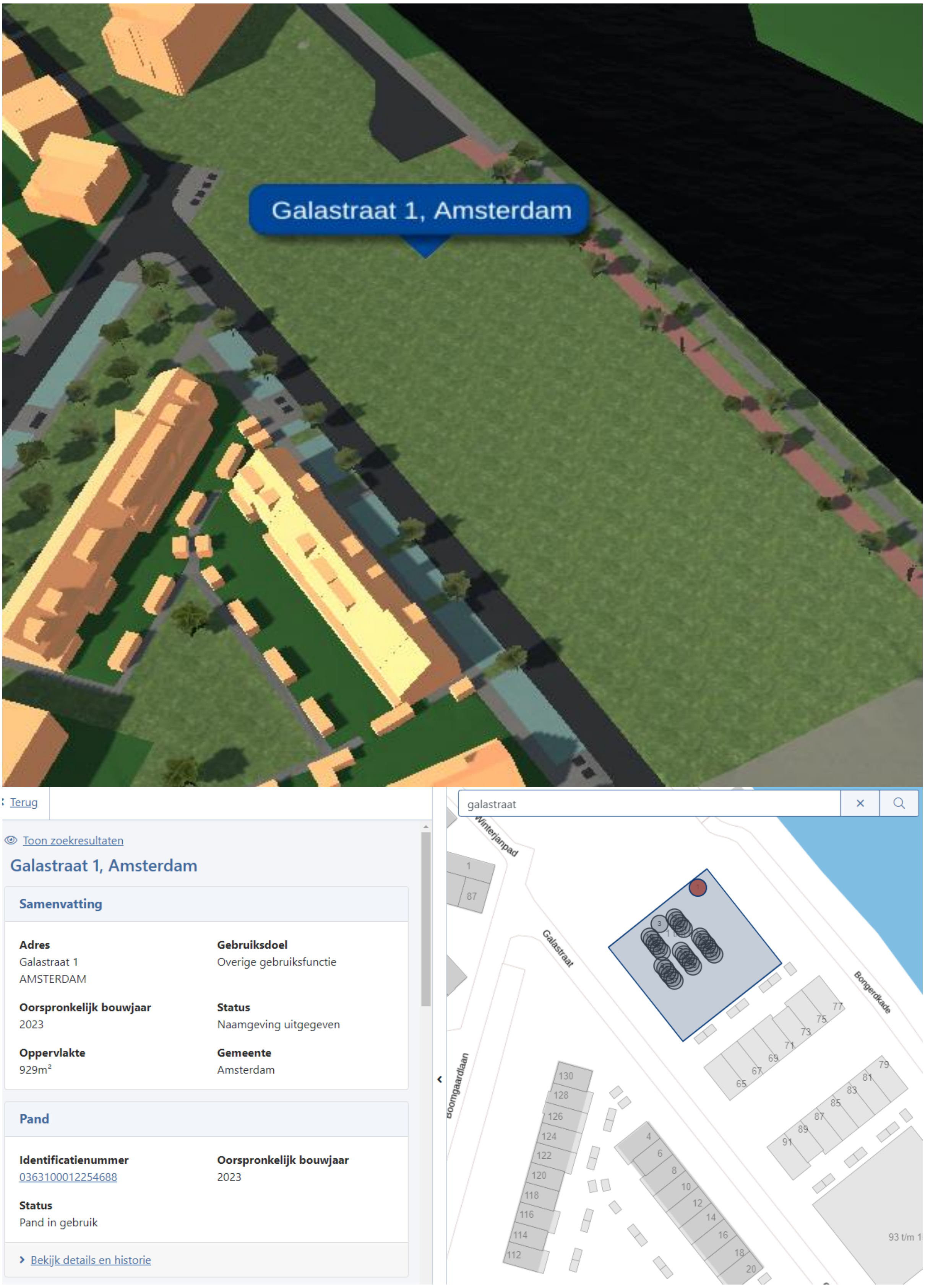

Figure 23 – Example of data difference between 3D Amsterdam and daily updated 2D data. Above: Snapshot of 3D Amsterdam. Below: snapshot of (2D) BAG data source. Sources: 3D Amsterdam[5] (top) and Kadaster BAG

---

[5] https://3d.amsterdam.nl/index.html#122228.23,491368.63,200.00,54.74,315.00,0.00

Viewer[6] (below)

- <u>Spatial dimension</u>

Not all data has 3D dimensions, but this is not necessary. One example of 2D data are the names of the neighborhood and their borders. There is no height dimension in this case. For the rest, even the sewer has a 3D dimension and is visualized (in 3D Utrecht) in that matter compared to the ground level (Figure 24). In its current state does the infrastructure definitely fit the SKI narrative.

    ➢ Status: SKI

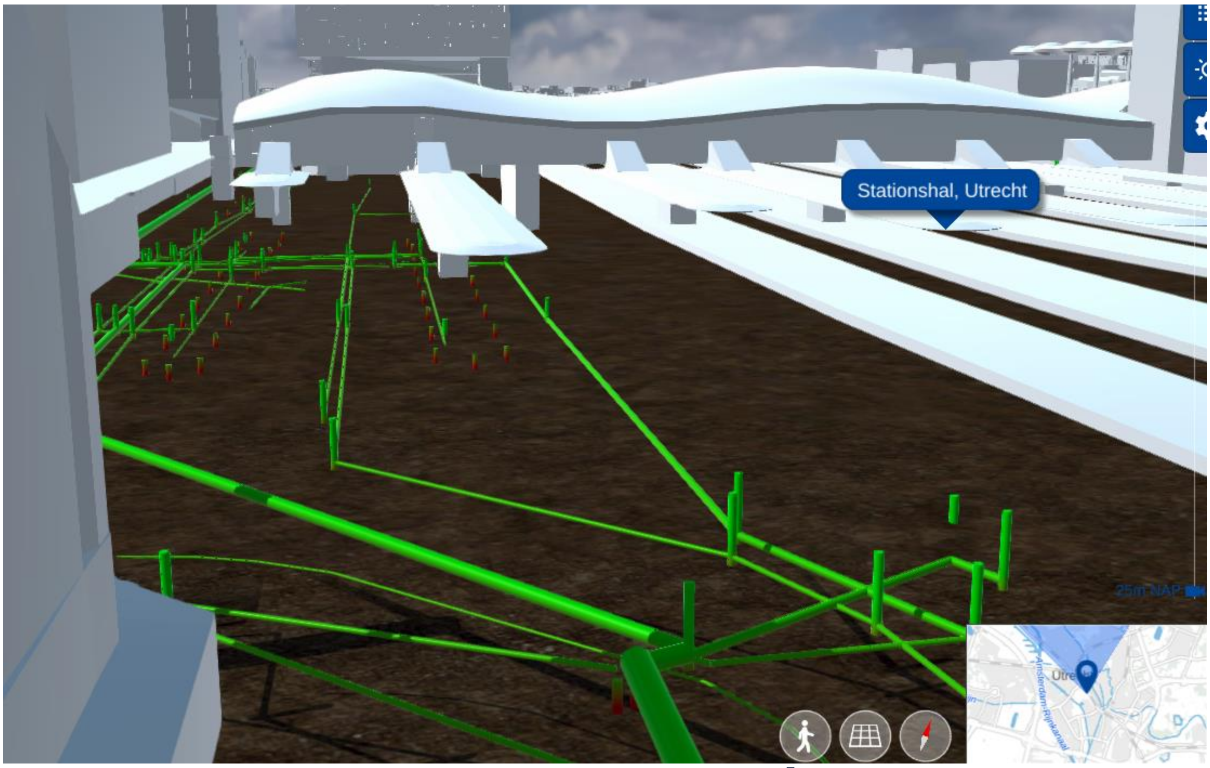

Figure 24 – Snapshot of 3D sewer network in Utrecht. Source: 3D Utrecht[7]

- <u>Temporal dimension</u>

This parameter is to some extent difficult to evaluate. The PO of 3D Amsterdam showed in a beta developer environment that for one part of Amsterdam (IJburg) it was possible to visualize the creation and development of the island which fits the SKI definition fully and therefore shows the possibilities. It was even possible to do the visualization in first person. However, this is not available for users. One available feature is the positioning of the sun during the whole day and for a specific date. With this feature the shade of buildings can be visualized (Figure 25). Notable about this feature is that it is not part of the data or data, but in the application itself. Still could be said that both features are not considered to be SDI, but at least moving towards SKI making the infrastructure in an intermediate state.

    ➢ Status: In-between SDI and SKI

5.2.3 User parameters

- <u>Level of expertise</u>

The PO mentioned development of the infrastructure is predominantly based on colleagues working as city engineer or spatial planner. It is a nice-to-have if everyday citizens can use the infrastructure for own purposes, but the development of features and functions will be based on authoritative colleagues working for governmental agencies. As engineers and architects are also part of the presumed users,

---

[6] https://bagviewer.kadaster.nl/lvbag/bag-viewer/?searchQuery=galastraat+1&objectId=0363200012169559&theme=BRT+Achtergrond&geometry.x=122127.4595&geometry.y=491450.873&zoomlevel=15

[7] https://3d.utrecht.nl/app#136027.52,455531.84,32.70,10.70,341.22,0.00

this parameter is to some extent deviating from the identified criteria as these professions are not necessarily geospatial experts nor are they IT experts or citizens. It is possible for citizens to use the system because addresses can be searched and the changing-sun-feature does not require any specific knowledge. However, citizens are not included as user group for the infrastructure and furthermore does the infrastructure provide little value to citizens.. This will place the level of expertise in-between SDI and SKI. As for IT experts, this profession is also not thought of as part of the target audience. Also, the data is not downloadable in a format where IT experts usually work with. Therefore, as for IT experts the current state is considered SDI.

- IT experts:
  - SDI
- Citizens:
  - In-between SDI and SKI

- Query input

The only query function is to search for an address. Other than this, the only possibility is to download specific views of the data based on certain tile dimensions. This is currently the only way of querying the data. It is not possible to perform more complex advanced query or using natural language. Therefore, the query input possibilities are still considered to be an SDI.

- IT experts
  - SDI
- Citizens:
  - SDI

- Output suitability

The limited possibilities of querying the data also results in limited output possibilities. If one wants to visualize the sunlight or rising water level on present buildings, the output is what is required. Other than that, the output almost certainly needs processing while data can only be downloaded by selecting predefined square vector tiles making the area not adjustable to the needs of the user. Also, the information of a building when selecting it is faulty though this is possibly the result of the current upkeep of 3D Amsterdam while this functionality does work in the case of 3D Utrecht. Overall, due to this combination of (limited) possibilities, the current state is in-between SDI and SKI while a traditional SDI would not provide such specified information regarding the sun positioning.

  - Status: In-between SDI and SKI

- Output readiness

As stated before, the user has the possibility to adjust sun positioning and water level to see the impact on existing buildings. It must be emphasized that these results can only be ‘used’ by visual notice (or taking a screenshot) while these results cannot be downloaded. The user can only download predefined extents of data and in two data formats. The downloaded need to be processed by the user within the local environment. Just as with the previous parameter, the readiness of the data is only SKI-ready when it comes to the functionalities regarding the water level and sun positioning.

  - Status: In-between SDI and SKI

- Device readiness

The website of 3D Amsterdam clearly states it only supports the most used web browsers (3D Amsterdam, n.d.b). Less popular browsers and mobile devices are not supported.

  - Status: SDI

- Analytics

Already mentioned are the possibilities regarding the water level and the position of the sun. These functions can shed light on when certain parts of the city are subject to flooding in case of a dam breach or rising water level, or to what extent you will have sunlight on your balcony (Figure 25). Whether the sunlight function is completely true can be questioned. This option had three options for cloudiness which are clouds, a clear sky and grey. In the ‘grey’ scenario the visualization still shows that buildings create shades. There is no shade when the sky is grey. Nonetheless, these two functionalities can be distinguished as facilitating analytics and thus the state of the infrastructure is already SKI ready.

  - Status: SKI

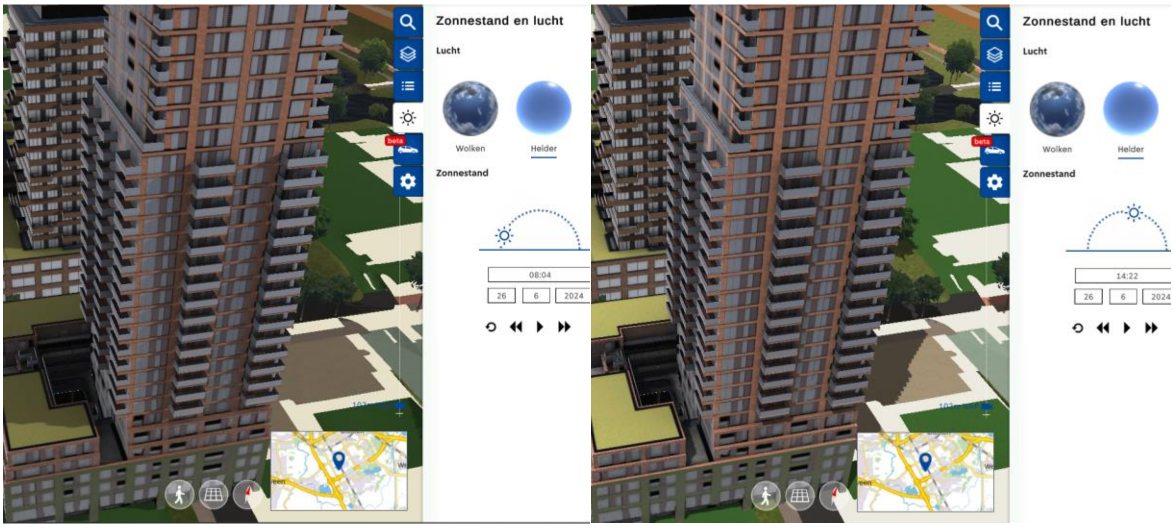

Figure 25 – Example of changing the time of day to see results of changing sun positioning. Source: 3D Amsterdam[8]

- Modeling

The functionalities regarding the positioning of the sun have all characteristics of spatial-temporal modeling as the user can see changes over space and time meaning the infrastructure meets the criteria set for an SKI.

  - Status: SKI

5.3 Synthesis

An overview on the state of the parameters of both case studies is depicted in Table 5.

Parameter		**State of the parameter**	
		3D Amsterdam	*Kadaster Knowledge graph*
Data	*Standards*	SDI	SKI
	Producer	SKI	SDI
	Storage	SKI	SKI
	Update method	In-between	SDI
	Update frequency	SDI	SDI
	Spatial dimension	SKI	SDI
	Temporal dimension	In-between	In-between
User	*Level of expertise*	IT experts : SDI Citizen : In-between	IT experts : SKI Citizen : SKI
	Query input	IT experts : SDI Citizen : SDI	IT experts : SKI Citizen : SKI
	Output suitability	In-between	In-between
	Output readiness	In-between	In-between
	Device readiness	SDI	SDI
	Analytics	SKI	SKI
	Modeling	SKI	SKI

Table 5 – Synthesis of case study results

One interesting outcome is that two parameters are still in an SDI state in both case studies, which are the update frequency and the device readiness. Both cases only focus and support desktop browsers and the data is updated periodically. Regarding the update frequency, 3D Amsterdam has a dependency on the source data which also is updated periodically. The KKG uses daily updated data sources, but chooses to implement updates less frequently. The update method is the only parameter that is still SDI in the KKG and in-between SDI and SKI in 3D Amsterdam. Both case studies mention

[8] https://3d.amsterdam.nl/#125173.71,479825.37,336.40,54.74,315.00,0.00

difficulties in automating updates due to needed quality checks and potential (and unaccounted for) changes in the source data that can impact the working of the infrastructure. Three parameters are considered to be in an intermediate state between SDI and SKI in both case studies, which are the temporal dimension, output suitability and output readiness. To some extent do these parameters have characteristics of an SKI, but not for all data or queries in the infrastructure. Three parameters are already fully SKI-ready in both cases, which are the storage of data, the possibility of performing on-the-fly data analysis and the possibility of modeling some data in the infrastructure. The case studies both run their infrastructure cloud-based and some possibilities of analytics is possible. For the other seven parameters it is dependent on the case study in question what the state of the infrastructure is (Table 5).

6. Discussion

This study provides a first comprehensive overview of the novel SKI concept. A comparison with the well-established SDI concept has been made which have led to a set of parameters that can be used to evaluate whether contemporary spatial infrastructures are moving towards being an SKI. In total 14 parameters are proposed and two cases are evaluated, the KKG and 3D Amsterdam. It was found that most of the parameters are beyond the SDI classification in at least one of both case studies.

6.1 Findings

The main difference between the SDI and SKI is that it is changing from data sharing to knowledge creation and sharing. Knowledge in this case means ready to use answers on questions from a user. The way an SDI can evolve into being an SKI is described by changes in the objectives, components and architecture. When reviewing these changes, the SKI predominantly focuses on the technical aspects by incorporating the latest technologies in order to meet current expectations by people who expect to get an answer. These expectations are raised by AI chatbots that are introduced in the last couple of years, but is not been discussed in SKI literature yet. Although the differences focuses on the user besides the data, all innovations seem to be technology driven. Soft aspects, which are considered to be just as important as technical aspects for having a good spatial infrastructure (Sjoukema, 2021), seem to be out of scope of current SKI literature. Therefore, the evolution looks mainly like a technology push of the SDI from the theoretical part. This is emphasized by the proposed set of components by Arnold et al., (2021) who discuss that an SKI is an SDI with two additional new components. This would mean that the basics of an SDI should already be excellent. Although every SDI is unique, SDI literature, for example, identify issues when it comes to data quality and data standards (Innerebner et al., 2017; Tripathi, Agrawal & Gupta, 2020). This implies there is still work to be done for SDIs.

Despite that the case studies were based on the conceptual model and therefore predominantly focused on the technology driven developments as well, the fact that that governance of the KKG was not yet in place for a fully operational infrastructure could be seen as an indicator that there is also technology push going on in practice. Nevertheless, the results of the case studies can shed light on which of the (technological) developments are currently taking place in practice.

Looking at the evaluated data parameters, only the storage parameter is SKI-ready in both case studies, meaning the entire infrastructure is cloud-based (Table 5). Although storing data in the cloud instead of using on-premise servers can be seen as the standard (Sunyaev, 2020), this does not have to apply for applications. The product lead from the KKG thinks these cloud-based infrastructures will not replace specialized 'local' software applications (Section 5.1.3). This directly relates to the user parameters of modeling and analytics. While both case studies do facilitate a form of analytics and modeling within the infrastructure, the possibilities are limited. Some of the literature assumes the future of cloud computing will overcome these limitations (Duckham et al., 2017; Geospatial World, 2022), but the KKG product lead thinks this is not going to be the case due to the existence of specialized tools. Other interesting outcomes are related to providing real-time data. In order to provide this, the data should be updated frequently and automatically. As for the update frequency, the most frequently updated data source included in one of the case studies, is updated on a daily basis (Section 5.1.2). Both case studies do not use 'real' real-time data sources (e.g., sensors or other VGI). Regarding the automated updates, both interviewees stated it is technically possible to implement new updates automatically, but this is not feasible due to checks that need to be done. In the case of 3D Amsterdam, it was even said that data was updated automatically at first, but a non-communicated change within the source data resulted in down-time of the webpage as the infrastructure could not process the changes resulting in an error (Section 5.2.2). This issue related to the data quality and standards emphasizes the need for excellent SDI basics in order to be able to realize an SKI. With the aim of including new data producers by incorporating crowdsourced or VGI data on the one hand and data transformation to make it linked data on the other hand, data issues keep existing (Medeiros & Holanda, 2019). To overcome these issues, Geospatial World (2022) addresses to have the maintenance of spatial data sources in hands of authoritative agencies to enhance interoperability by ensuring good data quality and standards.

It is useful to discuss the implications of the case study results on the conceptual framework and set of parameters. The literature reviews of both SKI and SDI have resulted in the 14 parameters that are established. The novelty of the SKI concept resulted in limited available literature on the concept, although also this stimulated the choice for a systematic approach. As only 11 articles explicitly used the terminology of a Spatial Knowledge Infrastructure, the term has not been used extensively. The novelty is also emphasized by the fact that the literature predominantly discusses the concept on a high-level basis without going into detail too much. The white papers of Duckham et al. (2017) and Geospatial World (2022) describe the concept on a holistic view while focusing on the definition and objectives whereas Arnold et al. (2021) and Omidipoor (2018) generally describe the components and architecture (Appendix B). This high-level approach was useful for establishing the set of parameters as the overall differences between the concepts are clearly described. Though, one downside of this approach is that the details of the differences are not described. This absence of this in-depth information in what exactly is meant can cause a faulty classification when evaluating a case. For example, the analytics parameter is established out of the need for performing on-the-fly analysis in an SKI. However, what kind of analysis and how complex these analyses should be is not described. This resulted in an SKI classification on this parameter for the case study of 3D Amsterdam as it is meeting the set criteria (Section 5.2.3), but to what extent the current available analytical possibilities (raising the water level and changing the positioning of the sun) are really sufficient to provide knowledge remains the question. There is also contradiction within the SKI literature on what the envisioned level of expertise is of people working with an SKI, because in the current situation this could be either IT experts or citizens (Section 4.2.1). Remarkable is that 3D Amsterdam does not focus on the geospatial expert nor on IT experts or citizens, but on other professionals like engineers and architects. Besides, advanced and complex querying should be possible in an SKI, but whether it is realistic that non-geospatial experts can produce complex spatial queries can be doubted. With respect to the querying, evaluating the output is challenging, both in terms of the two parameters (suitability and readiness) as well as for defining what 'knowledge' is. The SKI should provide specific and right information that can be used straight away, but to be able to assess whether the information is correct, it is still needed to either analyze the output or to already have the knowledge about the presumed output. For instance, when taking the usability test done in Figure 19 where it was asked to show all schools in a particular neighborhood, the specific output indicated the user got the information needed. Nonetheless, in case there is another school in the area which is not in the output (e.g., due to the data quality or outdated data), this will stay undetected.

The objectives of the cases also impact the parameters of the developed conceptual model and its parameters. On one hand there is the KKG, that focuses on the semantic web by providing an infrastructure of spatial Linked Data (LD) while encouraging others to develop their own application(s) on this infrastructure. On the other hand, 3D Amsterdam focuses on a digital twin of the city where users can analyze and model themselves with the given data and tools. The result is that the focus of the case defines which parameters (Figure 16) are deemed important for the spatial infrastructure at its core. This explains why certain parameters are SKI in one of the cases and SDI in the other, and vice versa. It is also the reason why 'device readiness' is considered to be an SDI in both cases: there is simply no need to develop an app or facilitate mobile browsing. This raises the question if it is necessary, or even desired, to aim for all-encompassing SKIs.

Taking this question into consideration while reflecting once more to the overall objective of an SKI, which is knowledge creation, it is interesting to contemplate if the various established parameters do contribute to this knowledge creation. For example, although web designs could make use of a 'mobile first' method because people prefer to use their smartphone instead of a desktop, it is questionable that which device is being used makes or breaks the knowledge creation of the infrastructure. This could, again, give the impression that the proposed SKI-characterization is predominantly a technology push instead of solely focused on knowledge creation. On the other hand, taking into consideration the user needs, this example would adhere to their needs and is user-centric. Comparable reflections are possible for most of the parameters, but it is indicative that defining the user groups and objectives of the SKI in question results in a sub-set of parameters that need to be included instead of all parameters.

6.2 Limitations

The findings highlighted the novelty of the SKI concept as a limited number of articles were identified in the systematic literature review. The search terms used in the protocol (Table 1) provide a concise and specified overview of literature, but the consequence of using such specific search terms is that the number of results could be limited. Literature using deviant terminology will be out of scope. In contrary to the systematic approach for finding SKI literature, a narrative approach was applied for including SDI literature. The search was based on the identified topics in the SKI review and although this enhances reproducibility and limited the results to a manageable number of articles to screen, it also means potentially useful articles or developments within the SDI-domain using deviant terminology were out of scope. This could affect the conceptualized parameters.

Whether the search strategies have affected the parameters cannot be said while articles are out of scope. However, it is identified that some of the characteristics assumed to be missing in SDI but occurring in SKI are discussed in SDI literature as developments in SDIs. Examples are semantic web standards (Izdebski, Zwirowicz-Rutkowska & Nowak da Costa, 2021), non-geospatial users (Mulder et al., 2020; Vancauwenberghe et al. 2018), diverse producers (Budhathoki, Bruce, & Nedovic-Budic, 2008; Vilches-Blázquez & Ballari, 2020), cloud-based system (Tripathi, Agrawal & Gupta, 2020) and 3D (Stoter et al., 2011. Although these developments are acknowledged, they are not considered to be usual practice. Nonetheless, this does give some points to ponder. For one, the so-called fourth generation SDI proposed by Vilches-Blázquez & Ballari (2020) and SKI can probably be seen as having comparable aims for the development of SDIs. Secondly, it shows that the established conceptual model and parameters are to some extent presented rather black-and-white. However, this is foremost a result of the aim to compare the SDI and SKI with each other and evaluate the case studies and particularly affects concepts that are continuously evolving, of which the SDI is a prime example.

The conceptual model and parameters are used, and tested, on two real-world case studies in the Netherlands. Based on two case studies, it is too early to draw decisive conclusions whether the conceptual model and parameters are fully applicable for evaluating all spatial infrastructures. In addition, it must be noted that, although both spatial infrastructures are accessible through the web, the KKG is still part of an acceptance environment and currently not operational whereas 3D Amsterdam is not fully operational anymore and development is on hold while the designated team is focusing on 3D Netherlands, which on its turn is not yet fully operational. Besides, both case studies are developed only recently, with 3D Amsterdam having been launched in 2019 (3D Amsterdam, n.d.c) and the KKG, although open to public, under development. The outcomes could provide useful insights into the latest developments when developing new SDIs, but might not be generalizable for SDIs that are implemented and operational for a longer period of time which could lead to bias.

6.3 Recommendations

This study provides several points to ponder for future research. The first recommendation would be to invert the methods of this study. Where this research developed a conceptual framework based on the literature and tested it against two case studies, new research could take a case study with the objective of providing knowledge and a spatial infrastructure at its core as starting point for the research. Based on the case study, a set of characteristics can be developed which subsequently can be compared with the characteristics (parameters) as described in this study.

A second recommendation that is somewhat in the same direction would be to compare other related concepts with the concept of SKI. Netherlands 3D explicitly mentions that they are building a digital twin, making a comparison between the concepts of digital twin and SKI an interesting subject with Netherlands 3D as case study. Meanwhile, the KKG fully focuses on an LD-infrastructure and integration within the semantic web. Comparing the semantic web with an SKI would be a different point of view. Both studies could have different views on which parameters are important and which are either less important or redundant.

As a third suggestion, it would be recommended to further study contemporary developments in the SDI field to evaluate which parameters are already incorporated in current SDIs without being directed into the SKI-narrative. In case some parameters are already business as usual in current SDIs, the development of SDI into SKI could be limited to less parameters. In particular the recent developments in the field of AI, like the emergence of AI-tools, could have its implications on the view on SKIs. It would not be unexpected that the focus of the SKI shifts to study the possibilities of implementing AI-technologies for reaching knowledge rather than the currently proposed changes.

A fourth recommendation would be to test the conceptual framework by studying many case studies instead of just two. Such research could provide a more general view into the possibilities and state of the parameters. If many case studies show that certain parameters seem to be staggering in the development, more attention should be drawn to these specific parameters because this could either mean the parameter is redundant or, if relevant, why this is happening.

7. Conclusion

The aim of this study was to explore to what extent the concept of a spatial knowledge infrastructure (SKI) differs from the well-established concept of a spatial data infrastructure (SDI) in the need for intelligent spatial infrastructures and to what extent current (mid-2024) case studies in the Netherlands are on its way to becoming an SKI. For evaluating the state of the spatial infrastructure, a set of 14 parameters has been developed which are divided into seven data parameters and seven user parameters. The data parameters include the standards, producer, storage, update method, update frequency, spatial dimension, and temporal dimension. The user parameters are the level of expertise, query input, output suitability, output readiness, device readiness, analytics, and modeling. This research is a first-time approach to develop a conceptual model that could evaluate whether spatial infrastructures are developing towards a knowledge infrastructure. The conceptual model was applied to two real-world Dutch case studies that have a spatial infrastructure at its core.

Both cases, the Kadaster Knowledge Graph (KKG) and 3D Amsterdam, are to some extent considered to be SKI-ready. In fact, it was found that, of the 14 parameters, only the data parameter "update frequency" and the user parameter "device readiness" are still considered SDI in both cases. The data are not updated (near) real-time and although society is 'mobile-first', both cases do not support mobile devices. One reason mentioned for not supporting mobile devices is the necessary computing power. The other 12 parameters are in at least one of both case studies moving beyond the SDI categorization. Which parameters are moving towards the SKI spectrum depends on the focus of the case study. For example, the KKG aims to be a semantic web integrated data infrastructure whereas 3D Amsterdam focuses on a web-based application where the user can interact with a 3D reproduction of the city. Of the parameters evaluated, the data parameter "update method" is unlikely to be SKI-ready in the near future. Both case studies are technically ready to automatically update the data, but due to dependencies on external data producers, a manual quality check is required before updating the data. On the other hand, some of the parameters could nowadays already be considered business as usual due to evolving standards within society. For example, a cloud-based solution is more the rule rather than the exception within the field of data storage.

The outcomes show that it is depending on the objectives of a case to what extent individual parameters are considered to be necessary SKI-ready. Overall, it does not seem likely that infrastructures aim to have all 14 parameters SKI-ready to be an all-encompassing SKI. As the main objective of an SKI is to provide knowledge, the question is also if all 14 parameters are necessary for realizing this knowledge creation. Currently, the evolution of the SKI seems predominantly a technology push of an SDI. This emphasizes that the basics of an SDI should be excellent in order to reach the state of an SKI, which is still work in progress. Furthermore, the conceptual model is a rather black-and-white display of an SDI and SKI. Although this is the result of the literature reviews on SKI and SDI, it is acknowledged that this has its impact on an ever evolving concept an SDI.

Whilst current SKI literature predominantly describes the requirements with a lack of detail, future research should go in-depth on the established parameters to specify the detailed requirements for the parameters. In this regard, the user parameter "level of expertise" is interesting as IT experts and citizens are identified as user groups in the literature, but 3D Amsterdam focusses on other user groups like architects and city engineers. Another suggestion is to invert the research methodology of this study and reevaluate the proposed conceptual model and accompanying set of parameters based on case studies instead of literature. Moreover, recent developments in the field of AI have not been part of the SKI literature yet and therefore could also impact the current view on the characteristics of an SKI. It should be studied what these developments can contribute to the need for knowledge in the geospatial domain. Nevertheless, this study could serve as a foundation for evaluating real-world spatial infrastructures and provides some guidance for assessing present SKI and SDI characteristics.

References

3D Amsterdam. (n.d.a). *Downloads*. Gemeente Amsterdam. Retrieved May 10, 2024, from https://3d.amsterdam.nl/index.html#downloads

3D Amsterdam. (n.d.b). *Over 3D Amsterdam*. Gemeente Amsterdam. Retrieved May 10, 2024, from https://3d.amsterdam.nl/index.html#over

3D Amsterdam. (n.d.c). *Wat is nieuw?*. Gemeente Amsterdam. Retrieved May 10, 2024, from https://3d.amsterdam.nl/#watisnieuw

3DBAG. (n.d.a). *Data sources*. 3D geoinformation research group (TU Delft) and 3DGI. Retrieved May 10, 2024, from https://docs.3dbag.nl/en/overview/sources/

3DBAG. (n.d.b). *About Us*. 3D geoinformation research group (TU Delft) and 3DGI. Retrieved May 10, 2024, from https://docs.3dbag.nl/en/overview/group/

Alkhaleel, B. A. (2024). Machine learning applications in the resilience of interdependent critical infrastructure systems—A systematic literature review. *International Journal of Critical Infrastructure Protection*, *44*, 100646. https://doi.org/https://doi.org/10.1016/j.ijcip.2023.100646

Atkinson, L. Z., & Cipriani, A. (2018). How to carry out a literature search for a systematic review: a practical guide. *BJPsych Advances*, *24*(2), 74–82. https://doi.org/10.1192/bja.2017.3

Arinze, B., & Banerjee, A. (1992). A knowledge-based approach for facilities location planning. *Expert Systems with Applications*, *5*(1), 131–139. https://doi.org/https://doi.org/10.1016/0957-4174(92)90102-X

Arnold, L., Mcmeekin, D., Ivánová, I., & Armstrong, K. (2021). Knowledge on-demand: a function of the future spatial knowledge infrastructure. *Journal of Spatial Science*, *66*, 1–18. https://doi.org/10.1080/14498596.2019.1654942

Béjar, R., Latre, M. Á., Nogueras-Iso, J., Muro-Medrano, P. R., & Zarazaga-Soria, F. J. (2009). An architectural style for spatial data infrastructures. *International Journal of Geographical Information Science, 23*(3), 271–294. https://doi.org/10.1080/13658810801905282

Blake, M., Majewicz, K., Tickner, A., & Lam, J. (2017). Usability analysis of the Big Ten Academic Alliance geoportal: Findings and recommendations for improvement of the user experience. *The Code4Lib Journal*, 38. https://journal.code4lib.org/articles/12932

Boguslawski, R., Borzachiello, M., Perego, A. (2020). *Architectures and standards for spatial data infrastructures and digital government : European Union location framework*, Publications Office. https://data.europa.eu/doi/10.2760/806776

Bramer, W. M., Rethlefsen, M. L., Kleijnen, J., & Franco, O. H. (2017). Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Systematic Reviews*, *6*(1), 245. https://doi.org/10.1186/s13643-017-0644-y

Bryman, A. (2012). *Social Research Methods*. Oxford: Oxford University Press; 4th edition.

Bucher, B., Folmer, E., Brennan, R., Beek, W., Hbeich, E., Würriehausen, F., Rowland, L., Maturana, R. A., Alvarado, E., Buyle, R., & Di Donato, P. (2021). *Spatial linked data in Europe: Report from spatial linked data sessions at knowledge graph in action*. (73 ed.) (Official Publication - EuroSDR; No. 73).

Budhathoki, N.R., Bruce, B., and Nedovic-Budic, Z. (2008). “Reconceptualizing the role of the user of spatial data infrastructure.” *GeoJournal 72* (3):149-60. https://doi.org/10.1007/s10708-008-9189-x

Burroughs, D. R., Storie, J., Storie, C. D., & Onstein, E. (2019). Evaluating the Arctic SDI: An Assessment of the Foundations needed for Success. *Int. J. Spatial Data Infrastructures Res.*, *14*, 1–34. https://ijsdir.sadl.kuleuven.be/index.php/ijsdir/article/view/490

Cavallaro, F., & Nocera, S. (2022). Integration of passenger and freight transport: A concept-centric literature review. *Research in Transportation Business & Management*, *43*, 100718. https://doi.org/https://doi.org/10.1016/j.rtbm.2021.100718

Chan, T. O. (2001). The dynamic nature of spatial data infrastructures: a method of descriptive classification. *GEOMATICA*, *55*(1), 65–72. https://doi.org/10.5623/geomat-2001-0008

Charalampos Gkonos, I. I. E., & Hurni, L. (2019). Spinning the wheel of design: evaluating geoportal Graphical User Interface adaptations in terms of human-centred design. *International Journal of Cartography*, *5*(1), 23–43. https://doi.org/10.1080/23729333.2018.1468726

Cipolloni, C. (2018). ARCHITECTURE OF SPATIAL DATA INFRASTRUCTURE (SDI) (DRAFT). retrieved from http://www.info-rac.org/en/infomap-system/infomapnode

Cocchia, A. (2014). Smart and Digital City: A Systematic Literature Review. In C. Dameri Renata Paola and Rosenthal-Sabroux (Ed.), *Smart City: How to Create Public and Economic Value with High Technology in Urban Space* (pp. 13–43). Springer International Publishing. https://doi.org/10.1007/978-3-319-06160-3_2

Coetzee, S., Gould, M., McCormack, B., Mohamed Ghouse, Z., Scott, G., Kmoch, A., Alameh, N., Strobl, J., Wytzisk, A., & Devarajan, T. (2021). *Towards a sustainable geospatial ecosystem beyond SDIs*. https://doi.org/10.13140/RG.2.2.22555.39203

Coetzee, S. & Wolff-Piggott, B. (2015). A Review of SDI Literature: Searching for Signs of Inverse Infrastructures. In C. B. and L. de M. P. M. Robbi Sluter Claudia and Madureira Cruz (Ed.), *Cartography - Maps Connecting the World: 27th International Cartographic Conference 2015 - ICC2015* (pp. 113–127). Springer International Publishing. https://doi.org/10.1007/978-3-319-17738-0_9

Coleman, D., & McLaughlin, J. (1998). *Defining global geospatial data infrastructure (GGDI): components, stakeholders and interfaces*. *52*, 129–143.

Dodds, L., & Wells, P. (2019). Data infrastructure. In Davies, T., Walker, S., Rubinstein, M., & Perini, F. (Eds.). (2019). *The State of Open Data: Histories and Horizons* (pp. 260-273). African Minds and International Development Research Centre

Drahmann, A., & Huijts, J. H. M. (2021). Het Digitaal Stelsel Omgevingswet: ICT-flop of belofte voor de toekomst?. *Computerrecht*, *2021*(3), 241-249. Retrieved from https://hdl.handle.net/1887/3278511

Duckham, M., Arnold, L., Armstrong, K., McMeekin, D., & Mottolini, D. (2017). Towards a spatial knowledge infrastructure. *White Paper*, 24. https://www.crcsi.com.au/assets/Program-3/CRCSI-Towards-Spatial-Knowledge-Whitepaper-web-May2017.pdf

Ferreira, K. R., de Queiroz, G. R., Vinhas, L., Câmara, G., Maurano, L. E., Souza, R. C. M., & Sanchez, A. (2015). Towards a Spatial Data Infrastructure for Big Spatialtemporal Data Sets. *In 17th Brazilian Symposium on Remote Sensing (SBSR), 2015. Proceedings* (pp. 7588-7594).

Fiedukowicz, A., Gasiorowski, J., Kowalski, P., Olszewski, R., & Pillich-Kolipińska, A. (2012). The statistical geoportal and the ``cartographic added value'' - creation of the spatial knowledge infrastructure. *Geodesy and Cartography*, *61*, 47–70. https://doi.org/10.2478/v10277-012-0021-x

Fischer, M. M. (1994). From conventional to knowledge-based geographic information systems. *Computers, Environment and Urban Systems*, *18*(4), 233–242. https://doi.org/https://doi.org/10.1016/0198-9715(94)90026-4

Geospatial World (2022). *A Geospatial Knowledge Infrastructure To Enhance The World Economy, Society & Environment*. https://www.geospatialworld.net/gw-assets/pdf/GKI-White-Paper.pdf

Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse education today*, *24*(2), 105–112. https://doi.org/10.1016/j.nedt.2003.10.001

Granell, C., Díaz, L., & Gould, M. (2010). Service-oriented applications for environmental models: Reusable geospatial services. *Environmental Modelling & Software*, *25*(2), 182–198. https://doi.org/https://doi.org/10.1016/j.envsoft.2009.08.005

Grus, Ł., Crompvoets, J., & Bregt, A. (2007). Multi-view SDI Assessment Framework. *International Journal of Spatial Data Infrastructures Research*.

Grus, L., Crompvoets, J., & Bregt, A. (2010). Spatial data infrastructures as complex adaptive systems. *International Journal of Geographical Information Science*, *24*, 439–463. https://doi.org/10.1080/13658810802687319

Grus, Ł., Castelein, W., Crompvoets, J., Overduin, T., van Loenen, B., van Groenestijn, A., Rajabifard, A., & Bregt, A. K. (2011). An assessment view to evaluate whether Spatial Data Infrastructures meet their goals. *Computers, Environment and Urban Systems*, *35*(3), 217–229. https://doi.org/10.1016/j.compenvurbsys.2010.09.004

Guigoz, Y. (2016). Spatial Data Infrastructures for addressing environmental challenges: stocktaking, capacity building, implementation and assessment. https://doi.org/10.13097/archive-ouverte/unige:88107

Gupta, S. (2015). A Comparative study of Usability Evaluation Methods. *International Journal of Computer Trends and Technology*, *22*, 103–106. https://doi.org/10.14445/22312803/IJCTT-V22P121

Gürdür Broo, D., Bravo-Haro, M., & Schooling, J. (2022). Design and implementation of a smart infrastructure digital twin. *Automation in Construction*, *136*, 104171. https://doi.org/https://doi.org/10.1016/j.autcon.2022.104171

Haddaway, N., Collins, A., Coughlin, D, Kirk, S. (2015) The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. PLoS ONE 10(9): e0138237. https://doi.org/10.1371/journal.pone.0138237

Harwood, T. G., & Garry, T. (2003). An overview of content analysis. *The marketing review*, *3*(4), 479-498. https://doi.org/10.1362/146934703771910080

He, X., Persson, H., & Östman, A. (2012). Geoportal Usability Evaluation *. *International Journal of Spatial Data Infrastructures Research*, *7*, 88–106. https://doi.org/10.2902/1725-0463.2012.07.art5

Hendriks, P.H.J., Dessers, E., & van Hootegem, G. (2012). Reconsidering the definition of a spatial data infrastructure. *International Journal of Geographical Information Science*, *26*(8), 1479–1494. https://doi.org/10.1080/13658816.2011.639301

Hennig, S., & Belgiu, M. (2012). *User-centric SDI: Addressing Users Requirements in Third-Generation SDI. The Example of Nature-SDIplus*. *10*. https://doi.org/10.5278/ojs.persk..v10i20.448

Idrizi, B. (2018). General Conditions of Spatial Data Infrastructures. *International Journal of Natural and Engineering Sciences*, *12*, 57–62.

Izdebski, W., Zwirowicz-Rutkowska, A., & Nowak da Costa, J. (2021). Open data in spatial data infrastructure: the practices and experiences of Poland. *International Journal of Digital Earth*, *14*(11), 1547–1560. https://doi.org/10.1080/17538947.2021.1952323

Innerebner, M., Costa, A., Chuprikova, E., Monsorno, R., & Ventura, B. (2017). Organizing earth observation data inside a spatial data infrastructure. *Earth Science Informatics*, *10*(1), 55–68. https://doi.org/10.1007/s12145-016-0276-0

Ivánová, I., Siao Him Fa, J., McMeekin, D. A., Arnold, L. M., Deakin, R., & Wilson, M. (2020). From spatial data to spatial knowledge infrastructure: A proposed architecture. *Transactions in GIS*, *24*(6), 1526–1558. https://doi.org/https://doi.org/10.1111/tgis.12656

Ivánová, I., Armstrong, K., & McMeekin, D. (2017). Provenance in the next-generation spatial knowledge infrastructure. Proceedings - 22nd International Congress on Modelling and Simulation, MODSIM 2017, 410–416

Jobst Markus & Gartner, G. (2019). Changing Paradigm in Map Production and Geoinformation Management—An Introduction. In M. and S. P. Döllner Jürgen and Jobst (Ed.), *Service-Oriented Mapping: Changing Paradigm in Map Production and Geoinformation Management* (pp. 3–42). Springer International Publishing. https://doi.org/10.1007/978-3-319-72434-8_1

Kadaster. (n.d.a) *Kadaster Labs*. Retrieved May 25, 2024, from https://labs.kadaster.nl/

Kadaster. (n.d.b) *Kadaster Knowledge Graph*. Retrieved May 10, 2024, from https://labs.kadaster.nl/thema/Knowledge_graph

Kadaster. (n.d.c) *Kadaster Knowledge Graph (KKG)*. Retrieved May 25, 2024, from https://data.kkg.kadaster.nl/

Kadaster. (n.d.d) *Welkom*. Kadaster. Retrieved May 25, 2024, from https://labs.kadaster.nl/demonstrators/architectuur-selfservice

Kadaster. (n.d.e) *Solution Architecture*. Kadaster. Retrieved May 25, 2024, from https://labs.kadaster.nl/demonstrators/architectuur-selfservice/Architecture/

Kalantari, M., Syahrudin, S., Rajabifard, A., Subagyo, H., & Hubbard, H. (2020). Spatial Metadata Usability Evaluation. *ISPRS International Journal of Geo-Information*, *9*(7). https://doi.org/10.3390/ijgi9070463

Klopper, R., Lubbe, S. & Rugbeer, H. (2007). The Matrix Method of Literature Review. *Alternation*, *14*(1), 262–276. https://doi.org/10.10520/AJA10231757_377

Kopsachilis, V., Vachtsavanis, N. & Vaitis, M. (2023). A Spatial Knowledge Infrastructure for the Aegean Archipelago. In M. L. and K. A. Klonari Aikaterini and De Lázaro y Torres (Ed.), Re-visioning Geography: Supporting the SDGs in the post-COVID era (pp. 35–55). Springer International Publishing. https://doi.org/10.1007/978-3-031-40747-5_3

Korkou, M., Tarigan, A. K. M., & Hanslin, H. M. (2023). The multifunctionality concept in urban green infrastructure planning: A systematic literature review. *Urban Forestry & Urban Greening*, *85*, 127975. https://doi.org/https://doi.org/10.1016/j.ufug.2023.127975

Lee, T.-Z., Wu, C.-H., & Wei, H.-H. (2008). KBSLUA: A knowledge-based system applied in river land use assessment. *Expert Systems with Applications*, *34*(2), 889–899. https://doi.org/https://doi.org/10.1016/j.eswa.2006.10.038

van Loenen, B., Crompvoets, J., & Poplin, A. (2010). Assessing geoportals from a user perspective. In *Workshop proceedings : 2nd workshop on Value of Geoinformation* (pp. 29–38). Hafencity University press; Hamburg.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June*, 431–440. https://doi.org/10.1109/CVPR.2015.7298965

Markus, B. (2005). Building spatial knowledge infrastructure. In *ISPRS Workshop on Services and Applications of Spatial Data Infrastructure* (pp. 65–70).

Masó, J., Pons, X., & Zabala, A. (2012). Tuning the second-generation SDI: theoretical aspects and real use cases. *International Journal of Geographical Information Science*, *26*(6), 983–1014. https://doi.org/10.1080/13658816.2011.620570

Masser, I. (1999). All shapes and sizes: the first generation of national spatial data infrastructures. *International Journal of Geographical Information Science*, *13*(1), 67–84. https://doi.org/10.1080/136588199241463

Mcdougall, K. (2010). From silos to networks-will users drive spatial data infrastructures in the future? *Proceedings of the 24th International Federation of Surveyors International Congress (FIG 2010)* (pp. 1-13). International Federation of Surveyors (FIG).

McEachen, N., & Lewis, J. (2023). ENABLING KNOWLEDGE SHARING BY MANAGING DEPENDENCIES AND INTEROPERABILITY BETWEEN INTERLINKED SPATIAL KNOWLEDGE GRAPHS. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 48(4/W7-2023), 117–124. https://doi.org/10.5194/isprs-archives-XLVIII-4-W7-2023-117-2023

Medeiros, G., Holanda, M. (2019). Solutions for Data Quality in GIS and VGI: A Systematic Literature Review. In: Rocha, Á., Adeli, H., Reis, L., Costanzo, S. (eds) New Knowledge in Information Systems and Technologies. WorldCIST'19 2019. Advances in Intelligent Systems and Computing, vol 930. Springer, Cham. https://doi.org/10.1007/978-3-030-16181-1_61

Mulder, A. E., Wiersma, M. G., & van Loenen, B. (2020). Status of National Open Spatial Data Infrastructures: a Comparison Across Continents. *International Journal of Spatial Data Infrastructures Research*, *15*, 56-87. https://doi.org/10.2902/1725-0463.2020.15.art3

Nativi, S., Santoro, M., Giuliani, G., & Mazzetti, P. (2019). Towards a knowledge base to support global change policy goals. *International Journal of Digital Earth*, *13*(2), 188–216. https://doi.org/10.1080/17538947.2018.1559367

Nayak, A., Božić, B., & Longo, L. (2022). Linked Data Quality Assessment: A Survey. In: Xu, C., Xia, Y., Zhang, Y., Zhang, LJ. (eds) Web Services – ICWS 2021. ICWS 2021. Lecture Notes in Computer Science(), vol 12994. Springer, Cham. https://doi.org/10.1007/978-3-030-96140-4_5

Nedovic-Budic, Z., Pinto, J., & Budhathoki, N. (2008). *SDI Effectiveness from the User Perspective*.

Odoh, M., & Chinedum, I. E. (2014). Research designs, survey and case study. *Journal of VLSI and signal processing*, *4*(6), 16-22.

Olsson, P.-O., Axelsson, J., Hooper, M., & Harrie, L. (2018). Automation of Building Permission by Integration of BIM and Geospatial Data. *ISPRS International Journal of Geo-Information*, *7*(8). https://doi.org/10.3390/ijgi7080307

Omidipoor, M. (2018). Towards Spatial Knowledge Infrastructure (SKI) : Technological Understanding. https://api.semanticscholar.org/CorpusID:53631028

Omidipoor, M., Toomanian, A., Neysani Samany, N., & Mansourian, A. (2021). Knowledge Discovery Web Service for Spatial Data Infrastructures. *ISPRS International Journal of Geo-Information*, *10*(1). https://doi.org/10.3390/ijgi10010012

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*. https://doi.org/10.1136/bmj.n71

Paz, F., & Pow-Sang, J. A. (2014). Current Trends in Usability Evaluation Methods: A Systematic Review. *2014 7th International Conference on Advanced Software Engineering and Its Applications*, 11–15. https://doi.org/10.1109/ASEA.2014.10

Rahman, Md. M., & Szabó, G. (2021). Multi-objective urban land use optimization using spatial data: A systematic review. *Sustainable Cities and Society*, *74*, 103214. https://doi.org/https://doi.org/10.1016/j.scs.2021.103214

Rajabifard, A., & Williamson, I. (2001). *Spatial data infrastructures: concept, SDI hierarchy and future directions*.

Rajabifard, A., Feeney, M. E. F., & Williamson, I. P. (2002). Future directions for SDI development. *International Journal of Applied Earth Observation and Geoinformation*, *4*(1), 11–22. https://doi.org/10.1016/S0303-2434(02)00002-8

Rajabifard, A., Binns, A., Masser, I., & Williamson, I. (2006). The role of sub-national government and the private sector in future spatial data infrastructures. *International Journal of Geographical Information Science*, *20*(7), 727–741. https://doi.org/10.1080/13658810500432224

Ronzhin, S., Folmer, E., Maria, P., Brattinga, M., Beek, W., Lemmens, R., & van't Veer, R. (2019). Kadaster Knowledge Graph: Beyond the Fifth Star of Open Data. *Information*, *10*(10). https://doi.org/10.3390/info10100310

Rowland, A., Folmer, E. J. A., & Baving, T. (2022). The Knowledge Graph As The Interoperability Foundation For An Augmented Reality Application: The Case At The Dutch Land Registry. *2022 ITU Kaleidoscope- Extended Reality – How to Boost Quality of Experience and Interoperability*, 1–8. https://doi.org/10.23919/ITUK56368.2022.10003053

Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, *33*(2), 163–180. https://doi.org/10.1177/0165551506070706

Rula, A., Maurino, A., & Batini, C. (2016). *Data Quality Issues in Linked Open Data*. 87–112. https://doi.org/10.1007/978-3-319-24106-7_4

Sadeghi-Niaraki, A., Rajabifard, A., Kim, K., & Seo, J. (2010). Ontology based SDI to facilitate spatially enabled society.

Schryen, G. (2015). Writing Qualitative IS Literature Reviews—Guidelines for Synthesis, Interpretation, and Guidance of Research. Communications of the Association for Information Systems, 37, pp-pp. https://doi.org/10.17705/1CAIS.03712

Sikder, I. U. (2009). Knowledge-based spatial decision support systems: An assessment of environmental adaptability of crops. *Expert Systems with Applications*, *36*(3, Part 1), 5341–5347. https://doi.org/https://doi.org/10.1016/j.eswa.2008.06.128

Sjoukema, J.-W. (2021). *Governance Dynamics of Spatial Data Infrastructures*. https://doi.org/10.18174/550452

Steiniger, S., & Hunter, A. (2012). Free and Open Source GIS Software for Building a Spatial Data Infrastructure. In *Geospatial Free and Open Source Software in the 21st Century* (pp. 247–261). https://doi.org/10.1007/978-3-642-10595-1_15

Stemler, S. E. (2015). Content Analysis. In *Emerging Trends in the Social and Behavioral Sciences* (pp. 1–14). John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.1002/9781118900772.etrds0053

Stock, K., Stojanovic, T., Reitsma, F., Ou, Y., Bishr, M., Ortmann, J., & Robertson, A. (2012). To ontologise or not to ontologise: An information model for a geospatial knowledge infrastructure. *Computers and Geosciences*, *45*, 98–108. https://doi.org/10.1016/j.cageo.2011.10.021

Stock, K., Small, M., Robertson, A., Reitsma, F., & Ou, Y. (n.d.) The Machinery of Knowledge: An Ontology-Registry for a Geospatial Knowledge Infrastructure. *Submitted to International Journal of Geographical Information Science*.

Stoter, J., van den Brink, L., Vosselman, G., Goos, J., Zlatanova, S., Verbree, E., Klooster, R., van Berlo, L., Vestjens, G., Reuvers, M., Thorn, S., & Delft, T. (2011). *A GENERIC APPROACH FOR 3D SDI IN THE NETHERLANDS.* In: Proceedings of the Joint ISPRS Workshop on 3D City Modelling & Applications and the 6th 3D GeoInfo Conference, Wuhan, pp. 22, 2011. https://gdmc.nl/publications/2011/Generic_approach_3D_SDI.pdf

Sunyaev, A. (2020). Cloud Computing. In *Internet Computing: Principles of Distributed Systems and Emerging Internet-Based Technologies* (pp. 195–236). Springer International Publishing. https://doi.org/10.1007/978-3-030-34957-8_7

Toomanian, A. (2012). *Methods to Improve and Evaluate Spatial Data Infrastructures*. [Doctoral Thesis (compilation), Centre for Geographical Information Systems (GIS Centre)]. Department of Physical Geography and Ecosystem Science, Lund University.

Tripathi, A. K., Agrawal, S., & Gupta, R. D. (2020). Cloud enabled SDI architecture: a review. *Earth Science Informatics*, *13*(2), 211–231. https://doi.org/10.1007/s12145-020-00446-9

Ulutaş Karakol, D., & Cömert, Ç. (2022). Architecture for semantic web service composition in spatial data infrastructures. *Survey Review*, *54*(382), 1–16. https://doi.org/10.1080/00396265.2020.1858255

UN-GGIM (2022). *Future Geospatial Information Ecosystem: From SDI to SoS and on to the Geoverse Making the Step Change Using the Integrated Geospatial Information Framework*. Discussion paper. https://ggim.un.org/meetings/GGIM-committee/12th-Session/documents/Future_Geospatial_Information_Ecosystem_Discussion_Paper_July2022.pdf

Vancauwenberghe, G., Valečkaitė, K., van Loenen, B., & Donker, F. W. (2018). Assessing the Openness of Spatial Data Infrastructures (SDI): Towards a Map of Open SDI. *International Journal of Spatial Data Infrastructures Research*, *13*, 88–100. https://doi.org/10.2902/1725-0463.2018.13.art9

Vancauwenberghe, G., & van Loenen, B. (2018). Exploring the Emergence of Open Spatial Data Infrastructures: Analysis of Recent Developments and Trends in Europe. In: Saeed, S., Ramayah, T., Mahmood, Z. (eds) User Centric E-Government. Integrated Series in Information Systems. Springer, Cham. https://doi.org/10.1007/978-3-319-59442-2_2

Vandenbroucke, D., Crompvoets, J., Vancauwenberghe, G., Dessers, E., & van Orshoven, J. (2009). A Network Perspective on Spatial Data Infrastructures: Application to the Sub-national SDI of Flanders (Belgium). *Transactions in GIS*, *13*(s1), 105–122. https://doi.org/10.1111/j.1467-9671.2009.01166.x

Vilches-Blázquez, L. M., & Ballari, D. (2020). Unveiling the diversity of spatial data infrastructures in Latin America: evidence from an exploratory inquiry. *Cartography and Geographic Information Science*, *47*(6), 508–523. https://doi.org/10.1080/15230406.2020.1772113

Warnest, M. (2005). A collaboration model for national spatial data infrastructure in federated countries. Philosophy. University of Melbourne, Australia

Welle Donker, F., & van Loenen, B. (2017). How to assess the success of the open data ecosystem? *International Journal of Digital Earth*, *10*(3), 284–306. https://doi.org/10.1080/17538947.2016.1224938

Welle Donker, F. (2018). Funding Open Data. In B. van Loenen, G. Vancauwenberghe, & J. Crompvoets (Eds.), *Open Data Exposed* (pp. 55-78). Information Technology and Law Series; Vol. 30. TMC Asser Press. https://doi.org/10.1007/978-94-6265-261-3_4

van Westen, C. J. (2013). 3.10 Remote Sensing and GIS for Natural Hazards Assessment and Disaster Risk Management. *Treatise on Geomorphology: Volume 1-14*, *1–14*, 259–298. https://doi.org/10.1016/B978-0-12-374739-6.00051-8

Woodgate, P., Coppa, I., Choy, S., Phinn, S., Arnold, L., & Duckham, M. (2017). The Australian approach to geospatial capabilities; positioning, earth observation, infrastructure and analytics: issues, trends and perspectives. *Geo-Spatial Information Science*, *20*(2), 109–125.

https://doi.org/10.1080/10095020.2017.1325612

Zhu, X., Healey, R. G., & Aspinall, R. J. (1998). A Knowledge-Based Systems Approach to Design of Spatial Decision Support Systems for Environmental Management. *Environmental Management*, *22*(1), 35–48. https://doi.org/10.1007/s002679900082

Zudilin, S. N., & Iralieva, Y. S. (2021). Automation of Land Use Planning Based on Geoinformation Modeling. *IOP Conference Series: Earth and Environmental Science*, *720*(1), 12039. https://doi.org/10.1088/1755-1315/720/1/012039

Zwirowicz-Rutkowska, A. (2017). A multi-criteria method for assessment of spatial data infrastructure effectiveness. *Earth Science Informatics*, *10*(3), 369–382. https://doi.org/10.1007/s12145-017-0292-8

Appendix A – PRISMA outcome SKI

1. Scopus – 11 hits

Arnold, L. M., McMeekin, D. A., Iványová, I., & Armstrong, K. (2021). Knowledge on-demand: a function of the future spatial knowledge infrastructure. *Journal of Spatial Science*, *66*(3), 365–382. https://doi.org/10.1080/14498596.2019.1654942

Ivánová, I., Siao Him Fa, J., McMeekin, D. A., Arnold, L. M., Deakin, R., & Wilson, M. (2020). From spatial data to spatial knowledge infrastructure: A proposed architecture. *Transactions in GIS*, *24*(6), 1526–1558. https://doi.org/https://doi.org/10.1111/tgis.12656

Ivánová, I., Armstrong, K., & McMeekin, D. (2017). Provenance in the next-generation spatial knowledge infrastructure. *Proceedings - 22nd International Congress on Modelling and Simulation, MODSIM 2017*, 410–416.

Kopsachilis Vasilis and Vachtsavanis, N. and V. M. (2023). A Spatial Knowledge Infrastructure for the Aegean Archipelago. In M. L. and K. A. Klonari Aikaterini and De Lázaro y Torres (Ed.), *Re-visioning Geography: Supporting the SDGs in the post-COVID era* (pp. 35–55). Springer International Publishing. https://doi.org/10.1007/978-3-031-40747-5_3

McEachen, N., & Lewis, J. (2023). ENABLING KNOWLEDGE SHARING BY MANAGING DEPENDENCIES AND INTEROPERABILITY BETWEEN INTERLINKED SPATIAL KNOWLEDGE GRAPHS. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *48*(4/W7-2023), 117–124. https://doi.org/10.5194/isprs-archives-XLVIII-4-W7-2023-117-2023

Stock, K., Stojanovic, T., Reitsma, F., Ou, Y., Bishr, M., Ortmann, J., & Robertson, A. (2012). To ontologise or not to ontologise: An information model for a geospatial knowledge infrastructure. *Computers and Geosciences*, *45*, 98–108. https://doi.org/10.1016/j.cageo.2011.10.021

Woodgate, P., Coppa, I., Choy, S., Phinn, S., Arnold, L., & Duckham, M. (2017). The Australian approach to geospatial capabilities; positioning, earth observation, infrastructure and analytics: issues, trends and perspectives. *Geo-Spatial Information Science*, *20*(2), 109–125. https://doi.org/10.1080/10095020.2017.1325612

1.1 Excluded at screening: 2

Fischer, T., & Jobst, M. (2020). Capturing the spatial relatedness of long-distance caregiving: A mixed-methods approach. *International Journal of Environmental Research and Public Health*, *17*(17), 1–23. https://doi.org/10.3390/ijerph17176406

Jobst, M., & Fischer, T. (2022). Approaching a Common Conscious Dataspace from a Data Provider Perspective – Requirements and Perspectives. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13376 LNCS*, 333–343. https://doi.org/10.1007/978-3-031-10450-3_28

1.2 Not retrieved: 0

1.3 Excluded after assessment: 2

Kopsachilis, V., Vachtsavanis, N., & Vaitis, M. (2022). Semi-automatic Semantification of Institutional Spatial Datasets. *CEUR Workshop Proceedings*, *3157*, 1–10.

Pilogallo, A., & Scorza, F. (2021). Regulation and Maintenance Ecosystem Services (ReMES): A Spatial Assessment in the Basilicata Region (Southern Italy). In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12955 LNCS*. https://doi.org/10.1007/978-3-030-87007-2_50

2. Web of Science – 4 hits – 4 duplicates

3. **ScienceDirect – 1 hit – 1 duplicate**

4. **Google Scholar – 13 hits – 7 duplicates**

Duckham, M., Arnold, L., Armstrong, K., McMeekin, D., & Mottolini, D. (2017). Towards a spatial knowledge infrastructure. *White Paper*, 24. https://www.crcsi.com.au/assets/Program-3/CRCSI-Towards-Spatial-Knowledge-Whitepaper-web-May2017.pdf

Fiedukowicz, A., Gasiorowski, J., Kowalski, P., Olszewski, R., & Pillich-Kolipińska, A. (2012). The statistical geoportal and the ``cartographic added value'' - creation of the spatial knowledge infrastructure. *Geodesy and Cartography*, *61*, 47–70. https://doi.org/10.2478/v10277-012-0021-x

Omidipoor, M. (2018). *Towards Spatial Knowledge Infrastructure (SKI) : Technological Understanding*. https://api.semanticscholar.org/CorpusID:53631028

Stock, K., Small, M., Robertson, A., Reitsma, F., & Ou, Y. (n.d.) The Machinery of Knowledge: An Ontology-Registry for a Geospatial Knowledge Infrastructure. *Submitted to International Journal of Geographical Information Science*.

4.1 Excluded after screening: 0

4.2 Not retrieved: 1

Arnold, L. ~M. (2016). Spatial Knowledge Infrastructures - Creating Value for Policy Makers and Benefits the Community. *AGU Fall Meeting Abstracts*, *2016*, PA11D-06.
Stock, K. (2009). COMPASS: A Geospatial Knowledge Infrastructure Managed with Ontologies. *EGU General Assembly Conference Abstracts*, 5719.

Excluded after assessment: 1

Markus, B. (2005). Building spatial knowledge infrastructure. In *ISPRS Workshop on Services and Applications of Spatial Data Infrastructure* (pp. 65–70).

Appendix B – Concept matrix SKI literature review

Article	**Definition**	**Objectives**	**Components**	**Architecture**
Arnold et al. (2021)		x	x	x
Duckham et al. (2017)	x	x		
Fiedukowicz et al. (2012)		x		
Ivánová et al. (2020			x	x
Ivánová, Armstrong & McMeekin (2017)		x		
Kopsachilis, Vachtsavanis & Vaitis (2023)		x		x
McEachen & Lewis (2023)		x		
Omidipoor (2018)	x		x	x
Stock et al. (n.d.)				x
Stock et al. (2012)	x	x		x
Woodgate et al. (2017)		x		
Geospatial World (2022)	x	x	x	

Appendix C – Concept matrix SDI literature review

Article	Definition	Objectives	Components	Architecture
Arnold et al. (2021)	x		x	
Béjar et al. (2009)				x
Boguslawski, Borzachiello & Perego (2020)				x
Chan (2001)	x			
Cipolloni (2018)				x
Duckham et al. (2017)				x
Ferreira et al. (2015)	x	x	x	
Geospatial World (2022)				x
Granell, Díaz & Gould (2010)		x		x
Grus, Crompvoets & Bregt (2007)			x	
Grus et al. (2011)		x		
Hendriks, Dessers & Van Hootegem (2012)	x	x		
Hennig & Belgiu (2011)			x	
Izdebski, Zwirowicz-Rutkowska & Nowak da Costa (2021)		x		x
Kopsachilis, Vachtsavanis & Vaitis (2023)		x		
van Loenen, Crompvoets & Poplin (2010)		x		
Masó, Pons, & Zabala (2012)			x	
McEachen & Lewis (2023)		x		
Mulder et al. (2020)				x
Omidipoor (2018)		x		x
Rajabifard & Williamson (2001)	x		x	

Sjoukema (2021)		x	x	
Steiniger & Hunter (2012)			x	
Toomanian (2021)		x	x	
Tripathi, Agrawal & Gupta (2020)			x	x
UN-GGIM (2022)	x			
Vandenbroucke et al. (2009)	x	x	x	
Vancauwenberghe et al. (2018)		x		
Vancauwenberghe & van Loenen (2018)				x
Warnest (2005)	x			
Welle Donker & Van Loenen (2017)			x	
Welle Donker & van Loenen (2018)		x		
Van Westen (2013)	x			
Woodgate et al. (2017)				x
Zwirowicz-Rutkowska. (2017)		x		

Appendix D – Generalized interview scheme - English

I. Introduction

1) Start of the interview
   - Possibility of recording the interview

2) Introduction of the interviewer

3) Introduction of the interviewee
   - What is your role in the project?
   - What amount of your working time are you dedicated to the project?

4) Background of the interview

Explain the study aim. Study aim is to evaluate current developments in spatial infrastructures. The case study in question has a spatial infrastructure at its core. Current development in the field of spatial infrastructures it to provide spatial knowledge and not only pushing data. Related concepts are the SDI and SKI. Study proposes a set of parameters that evaluate the transition from SDI to SKI and identifies to what extent the case study in question is on its way to be a knowledge-providing infrastructure.

II. General information about the project

1) Aim of the project
   - Can you provide some background information about the project, like when it started and what the projects tries to achieve?
   - To what extent is it defined when the end-form of the project has been reached?

2) Current status of the project
   - What is the current status of the project?

3) Future developments of the project
   - Which developments are predicted on the short term and on the long term?

III. Data-oriented questions

1) Current data sources
   - What is the general architecture of the infrastructure?
   - Which datasets are currently included in the project?
   - Is all the data open available or are there closed datasets?
   - Elaborate in order to answer following user parameters
     - ✓ Producers - Who are the producers of these datasets?
     - ✓ Standards – What are the standards? Are the standards uniform?
     - ✓ Spatial dimension – Is the data in 2D or 3D?
     - ✓ Update scheme – Are the datasets from other producers retrieved manually or automatically and published manually or automatically? And publishing own data?
     - ✓ Update frequency – How often is the data from author producers updated? And own data?
     - ✓ Storage – At what location is the data stored that is available in the system?

   - Do you need to transform certain data sources in order to connect them within the infrastructure?
   - To what extent can the data of the infrastructure be approached, or downloaded, by others?
     - ✓ Preliminary answering user parameters of query input and output

2) Future data sources
   - Are there other datasets envisioned to be connected to the project?
   - What do you want to achieve with incorporating those datasets?

IV. User-oriented questions

1) Users
   - ➢ Which user groups are projected to make use of the project?
     - ✓ Level of expertise of the user

2) Functionalities
   - ➢ Which functionalities are currently present and which functionalities are in the pipeline or envisioned?
     Elaborate in order to answer following user parameters
     - ✓ Query input
     - ✓ Query output
     - ✓ Analytics
     - ✓ Modeling
   - ➢ To what extent does the data support, or need working to support, new functionalities?

3) Devices
   - ➢ Which devices should the project support now and in the future?
     - ✓ Device readiness

V. Round up

Appendix E – Interview scheme 3D Amsterdam - Dutch

I. Introductie

1 start interview – vragen opnemen

2 introductie interviewer

3 introductie geïnterviewde

- Academische/professionele achtergrond
- Rol in het project
- In hoeverre voltijd met dit project bezig
- Bekend met ruimtelijke data infrastructuren

4 achtergrond interview

Ontwikkelingen binnen het geo-informatica veld met betrekking tot ruimtelijke data infrastructuren. Dit zijn ontwikkelingen die je meer in de huidige tijd ziet, waarbij we 'kennis' willen creëren en de gebruiker willen faciliteren in plaats van enkel data verzamelen en de data 'zenden'. Het doel van het onderzoek is een set van parameters ontwikkelen gebaseerd op de laatste inzichten waar een infrastructuur tegenaan gehouden kan worden om te evalueren in hoeverre het op weg is naar wat gezien wordt als een kennisinfrastructuur in plaats van een data-infrastructuur.

3D Amsterdam is een mooie case study omdat het een technisch/software project gebaseerd op ruimtelijke data is.

II. Algemene info 3D Amsterdam

1 huidige status en verwachting eindproduct

Ik heb gelezen dat Amsterdam 3D in 2019 is gelanceerd en dat de laatste wijziging/nieuwe functionaliteit begin 2023 is uitgerold. Wat is momenteel de status van het project en wat hopen jullie te bereiken?

III. User parameters

1 Doel: parameter device readiness

Momenteel worden enkel internetbrowsers ondersteund. Wat is de toekomstvisie op dit onderdeel?

2 Doel: parameter level of expertise

Het doel is om 'communicatie en participatie steeds makkelijker te maken'. Wat bedoelen jullie hiermee? Communicatie met of tussen wie? Participatie door/voor wie?

3 Doel: parameters analytics en modeling, query input, output suitability en output readiness

Een ander doel is simulaties uit te voeren om zo de uitkomsten van aanpassingen te testen. Hebben jullie hier al een concreet idee bij over hoe jullie dit voor je zien? Gebeurt dat binnen de interface van de website? Kan je bijvoorbeeld iets 'intekenen' of op andere wijze aangeven wat je wil wijzigen?

Wordt direct in binnen de interface duidelijk wat de uitkomsten zijn? Zo ja, op welke wijze?

Bijvoorbeeld de kaartlaag bomen betreft niet de daadwerkelijke hoogte maar een geschatte hoogte. Klopt dat? Heeft dit soort zaken nog impact op de mogelijkheden van het modeleren?

IV. Data parameters

1 Doel: parameter temporal dimension

Modeleren over tijd – gaat het dan alleen over toekomstige wijzigingen of ook om bijvoorbeeld historische wijzigingen situaties in te kunnen zien?

2 Doel: parameter spatial dimension

Doel is om zoveel mogelijk in 3D te visualiseren. Hoe zit dit met bijvoorbeeld het rioolstelsel? Is dat ook wenselijk en zo ja, mogelijk? Of hoe wordt bepaald wat wel/niet in 3D wordt weergegeven? (on)mogelijkheden data of op basis van eigen inzicht?

3 Doel: parameters producers

In de interface staan 5 kaartlagen met achterliggende informatie, maar daarnaast ook enkele andere lagen, zoals themakaarten met PDOK informatie. Hoeveel databronnen zijn in totaal aangesloten?

Planologische (geluid)zones – bron?

Wijk-buurt-straatnamen - bron CBS kerncijfers?

Dus de kaartlagen zijn afkomstig vanuit gemeente zelf, Kadaster (3D BAG), PDOK en CBS?

Wat valt onder de kaartlaag ‘overige’? Ik zie daar bijvoorbeeld geen informatie over.

4 Doel: architectuurplaat

Is er wellicht een hoog-over architectuurplaat beschikbaar? Zou ik die mogen ontvangen?

5 Doel: parameters update method + frequency

Hoe updaten jullie de databronnen? Automatisch (b.v. API) of handmatig? Welke bronnen handmatig?

Bijvoorbeeld de kaartlaag bomen heeft nog een bewerking nodig. Gebeurt dit automatisch na een update van het bronbestand of moet er binnen het project iemand zorgen voor deze translatie? Is het bij andere databronnen nog nodig na importeren wijzigingen toe te passen?

Kaartinformatie van de kaartlaag bomen geeft aan tot 3 jaar achter te lopen, maar ook dat laatste import 2019 was. Kan hier wat meer duiding aan gegeven worden?

6 Doel: parameter standards

Zijn de standaarden van de databronnen gelijk aan elkaar?

7 Doel: parameter storage

Draait het hele systeem in de cloud of nog (gedeeltelijk) in een datawarehouse? Is het makkelijk om op te schalen bij uitbreidingen?

Appendix F – Interview scheme Kadaster Knowledge Graph - Dutch

I. Introductie

1 start interview – vragen opnemen

2 introductie interviewer

3 introductie geïnterviewde

- Academische/professionele achtergrond
- Rol in het project
- In hoeverre voltijd met dit project bezig
- Bekend met ruimtelijke data infrastructuren

4 achtergrond interview

Ontwikkelingen binnen het geo-informatica veld met betrekking tot ruimtelijke data infrastructuren. Dit zijn ontwikkelingen die je meer in de huidige tijd ziet, waarbij we 'kennis' willen creëren en de gebruiker willen faciliteren in plaats van enkel data verzamelen en de data 'zenden'. Het doel van het onderzoek is een set van parameters ontwikkelen gebaseerd op de laatste inzichten waar een infrastructuur tegenaan gehouden kan worden om te evalueren in hoeverre het op weg is naar wat gezien wordt als een kennis-infrastructuur in plaats van een data-infrastructuur.

De Kadaster Knowledge Graph is een mooie case study omdat het een technisch/software project gebaseerd op ruimtelijke data is.

II. Algemene info Kadaster Knowledge Graph

1 Doel: ontstaansgeschiedenis KKG

Met welk idee en wanneer is de knowledge graph ontstaan?

2 Doel: Huidige status project

Wat is momenteel de status van het project? Is de (openbare) website nog operationeel? Wordt het nog bijgewerkt?

Bijvoorbeeld toegankelijkheidsverklaring is sinds 2021 niet bijgewerkt, sommige data ook niet up-to-date.

Verschilt dit per use case? Sommige use cases wel, anderen niet?

Bijvoorbeeld integrale gebruiksoplossing wel en losse use cases/testen niet.

3 Doel: Toekomst *(indien relevant na vraag 2)*

Wat hopen jullie nog te bereiken?

III. User parameters

1 Doel: parameter level of expertise

Welke gebruikersgroepen willen jullie ondersteunen?

Wie zien jullie als 'professionele gebruiker'? (Doel: Geospatial Expert of algemene IT-expert)

2 Doel: parameters query input, output suitability en output readiness

Wat is het algemene doel met betrekking tot het bevragen van de KKG te zijn?
De input voor het bevragen van de KKG lijkt afhankelijk van de gebruiker ofwel SPARQL ofwel alledaags taalgebruik te zijn.

Hoe zouden de antwoorden tot de gebruiker moeten komen?

Alleen benodigde antwoorden? Nog nodig om zelf te downloaden?

3 Doel: parameters analytics en modeling

In hoeverre moet de KKG analyses en modeleren binnen het systeem faciliteren?

Er lijken niet echt mogelijkheden te zijn om zaken te analyseren of te modeleren.

4 Doel: parameter device readiness

Welke apparaten moet de KKG ondersteunen denkend aan desktop en mobiele apparaten?

IV. Data parameters

1. Doel: parameter standards

Alle openbaar beschikbare bronnen in de KKG worden als linked data aangeboden. Voor alle datasets zetten jullie dit zelf om?

Bevat de KKG ook databronnen die niet openbaar zijn? Volgen deze databronnen hetzelfde proces en worden ze ook als linked data aangeboden?

Zijn er nog databronnen die jullie willen gaan aansluiten?

1. Doel: parameter temporal dimension

De datasets worden periodiek bijgewerkt. Is er via de KKG ook een versiebeheer om bijvoorbeeld datasets met verschillende tijdstempels te downloaden?

2 Doel: parameter spatial dimension

Zijn alle datasets in 2D of zijn er ook 3D datasets?

De BAG is in samenwerking met TU Delft ook in 3D ontwikkeld.

3 Doel: parameters producers

Wie zijn de producers van de databronnen? Beheren jullie alle databronnen of haal je sommige bronnen ook as-is binnen van andere producers?

Kadaster, CBS, PDOK, gemeentes?

4 Doel: architectuurplaat

Is de architectuurplaat nog up-to-date?

5 Doel: parameters update method + frequency

Hoe updaten jullie de databronnen? Automatisch (b.v. API) of handmatig? Welke bronnen handmatig? Waar is dit afhankelijk van?

Gebaseerd op de huidige updatestatus is vermoeden handmatig.

7 Doel: parameter storage

Draait het hele systeem in de cloud of nog (gedeeltelijk) in een datawarehouse? Is het makkelijk om op te schalen bij uitbreidingen?