INSTITUT NATIONAL DES SCIENCES APPLIQUÉES DE STRASBOURG

MÉMOIRE DE SOUTENANCE DE DIPLÔME D'INGÉNIEUR INSA

SPÉCIALITÉ : TOPOGRAPHIE

-

ENGINEERING DIPLOMA DEFENSE THESIS

SPECIALITY: SURVEYING ENGINEERING

Creation of quality dashboard for geospatial data and services

---

Written by Camille C. COSSEC

Defended September 22nd, 2022

Host organization :

National Land Survey of Finland
Maanmittauslaitos
P.O. Box 84
00521 HELSINKI
FINLAND

Company tutor :

Mr. Antti JAKOBSSON
Leading specialist & chief engineer

Academic tutor (INSA) :

M. Mathieu KOEHL
Responsable de la spécialité topographie

ABSTRACT

Nowadays, a colossal amount of data is continually being produced across the world. In this context and through three compelling use cases, the GeoE3 project tackles the challenge of taking advantage of this abundance of data, working on producing a European platform from which geospatial and statistical data with geospatial features can be accessible to the general public. However, when data from a wide variety of sources are gathered, the question of data quality assessment rises.

In the present thesis, a solution to this issue is presented as a Proof of Concept Quality Dashboard focused on the first use case of GeoE3, associated with solar energy potential and energy efficiency of buildings in changing climate and smart cities. This Proof of Concept, created using the ready-made software Microsoft Power BI, focuses on the identification and visualization of data quality indicators in an intelligible way, so that users from different backgrounds in geospatial data and data quality analysis can all benefit from it. Data quality indicators and ways to measure them have been identified, organized in tiers, and presented in a quality dashboard in three points of view in order to cater to every stakeholder.

Keywords: quality dashboard, data quality, metadata, geospatial data, interoperability, GeoE3

RÉSUMÉ

De nos jours, une quantité importante de données est produite en continu à travers le monde. Dans ce contexte, un projet Européen a vu le jour. À travers trois cas d’utilisation, GeoE3 a pour ambition de présenter une solution européenne sous la forme d’une plateforme depuis laquelle des données géospatiales ainsi que des données statistiques ayant des caractéristiques géospatiales deviennent accessibles au plus grand nombre. Cependant, lorsque des données de sources multiples sont réunies en un même endroit, il est nécessaire de se poser la question de l’évaluation de la qualité des données ainsi assemblées.

Le présent mémoire présente une solution à cette problématique sous la forme d’un de tableau de bord (Quality Dashboard) présentant l’analyse de la qualité de jeux de données et services considérés dans le cas d’utilisation concernant le potentiel d’énergie solaire et l’efficacité énergétique des bâtiments face au changement climatique et aux villes intelligentes. Cette Preuve de Concept, créée à l'aide du logiciel prêt à l'emploi Microsoft Power BI, se concentre sur l'identification et la visualisation d'indicateurs de qualité de données de manière la plus intelligible possible, afin que les utilisateurs novices comme experts en données géospatiales et/ou analyse de qualité de données puissent en bénéficier. Des indicateurs de qualité des données et leur mesures associées ont été identifiés, organisés en différents niveaux et présentés sous la forme de trois points de vue afin de répondre au mieux aux attentes de tous les utilisateurs.

Mots-clés : tableau de bord, qualité de données, métadonnées, données géospatiales, interopérabilité, GeoE3

FOREWORD

This memoire was written for my Engineering Diploma in Surveying Engineering at the Institut National des Sciences Appliquées de Strasbourg (INSA Strasbourg), France. It is the result of a 26-week internship at the National Land Survey of Finland.

The subject of this thesis is dedicated to the creation of a user-oriented dashboard focused on displaying quality analysis of geospatial data and other data with geospatial features and their metadata. The goal of this dashboard is to improve decision-making for users by helping them identifying possible issues in the quality of given datasets.

The topic of data and metadata quality analysis is a vast topic that has been of interest ever since the topic of collecting and using data was studied – this is reflected in the extremely expansive – seemingly never-ending – nature of the literature covering this area.

The complexity of this study resided mostly in the discovery and handling of a significant number of data quality notions over a short period of time.

In hindsight, as I studied data and metadata quality analysis, I uncovered many rabbit holes of very specific analysis outside the scope of this thesis's problematic. The extensive knowledge I learned from this research enlightened my vision of data and metadata quality analysis in the context of geospatial data and topography. In particular, it emphasized how important it is in our field, as surveying engineers, to include accurate metadata to the datasets we produce, especially if we aim to share it as open data through national portals.

ACKNOWLEDGMENTS

I would like to express my gratitude to Antti Jakobsson for being my tutor and thoroughly guiding me through this project.

I would also like to thank Lena Hallin-Pihlatie for being a great partner to work alongside with. In addition, I thank my team, the Paikkatietoteknologia group, led by Panu Muhli at the FGI, for welcoming me this year and advising me throughout my work.

More broadly, I would like to thank everyone at the National Land Survey of Finland and the FGI for warmly welcoming me during this internship.

Enfin, je souhaite remercier particulièrement ma famille qui m’a toujours soutenue sans faille.

ABBREVIATIONS

API	:	Application Programming Interface
DCAT-AP	:	DCAT Application profile for data portals in Europe
DSI	:	Data Science Institute
GeoE3	:	Geospatially Enabled Ecosystem for Europe GEOE3
FGI	:	Finnish Geospatial Research Institute
MML	:	Maanmittauslaitos
NLS	:	National Land Survey of Finland
OGC	:	Open Geospatial Consortium
UC	:	Use Case

TABLE OF CONTENTS

INTRODUCTION

1.1 Context

Geospatial data present a growing interest and an increasing number of modern applications. Companies and individuals from a wide variety of fields have been working and collecting large amounts of spatial data all over the world for years, usually with a very specific use in mind.

By taking a global approach to these challenges and processing data of various origins and scales, new efficient and durable solutions can emerge.

This study's focus resides within an ongoing project coordinated by the National Land Survey of Finland: the Geospatially Enabled Ecosystem for Europe project – or GeoE3.

The National Land Survey of Finland (NLS), or *Maanmittauslaitos* (MML) in Finnish, is a Helsinki-based public authority that operates under the Ministry of Agriculture and Forestry in Finland. Only they can oversee the land surveying in Finland by taking care of land survey deliveries, real estate information, map materials, and legal claims and mortgages. Additionally, they produce information, research, and services about the Earth.

In 2015, the NLS merged with the 1918-established Finnish Geodetic Institute (FGI), or *Geodeettinen laitos* in Finnish. Upon the merging, the FGI became the research branch of the NLS as the Finnish Geospatial Research Institute. Historically, the FGI worked towards creating and maintaining the national coordinate, height, and gravity systems. The Institute is also interested in projects regarding the spatial data infrastructure and carries out research work in the fields of geodesy, geoinformatics and remote sensing.

In particular, GeoE3 is a project housed by the Geoinformatics and Cartography department – *Geoinformatiikka ja kartografia* – of the Finnish Geospatial Research Institute. This department focuses on spatial data management and processing methods as well as on the large-scale utilization of spatial data. It is divided in several research groups, including the Geospatial technologies group – *Paikkatietoteknologiat* or *PaiTek* – within which this study's work has been conducted.

The GeoE3 project is co-financed by the Connecting Europe Facility of the European Union and is built upon previously EU-funded projects. GeoE3 was launched in 2020 and its development and implementation should span over 3 years, from October 2020 to September 2023.

This European project is part of the D.S.I.'s Open Data Initiative (Data Science Institute). In the long run, this ecosystem will allow the dynamic integration of various high-value geospatial data and services that originate from different national platforms, which will facilitate the access, analysis, understanding and visualization of information for people in both private and public sectors.

GeoE3's main objective is to take advantage of existing or emerging national, regional and cross-border geospatial data, platforms, and services in Europe in order to develop a Cloud-based ecosystem of services that dynamically integrate datasets and services with geospatial data. According to Reini (2021) this project's objectives encompass the improvement of data access and interoperability – geospatial or otherwise – and the harmonization of geospatial data based in use cases and APIs.

In particular, GeoE3's goals defined by the European Commission are detailed in section 2.1 below.

To fulfil its objectives, the GeoE3 project is developed through three relevant use cases related to climate change and urbanization (c.f. 2.3 Use cases).

1.2 Creation of a quality dashboard

Within the GeoE3 project, it was deemed valuable to order the creation of a quality dashboard that would be used to highlight the quality of datasets involved in the study of use cases and to make metadata information more intelligible for users.

This quality dashboard will serve as a visual representation of the data and metadata quality evaluation's results in such a way as to be accessible to novice users, while remaining sufficiently technical to be useful for experts in geospatial data.

Within the GeoE3 project structure, the quality dashboard is part of Task 5 in Activity 2.

As can be seen in Figure 0-1 Flow of data overview diagram, this dashboard shall be based on information obtained out of the data, metadata, and services as well as other available data sources including automated quality validation results, quality reviews and quality audits based on maturity models.

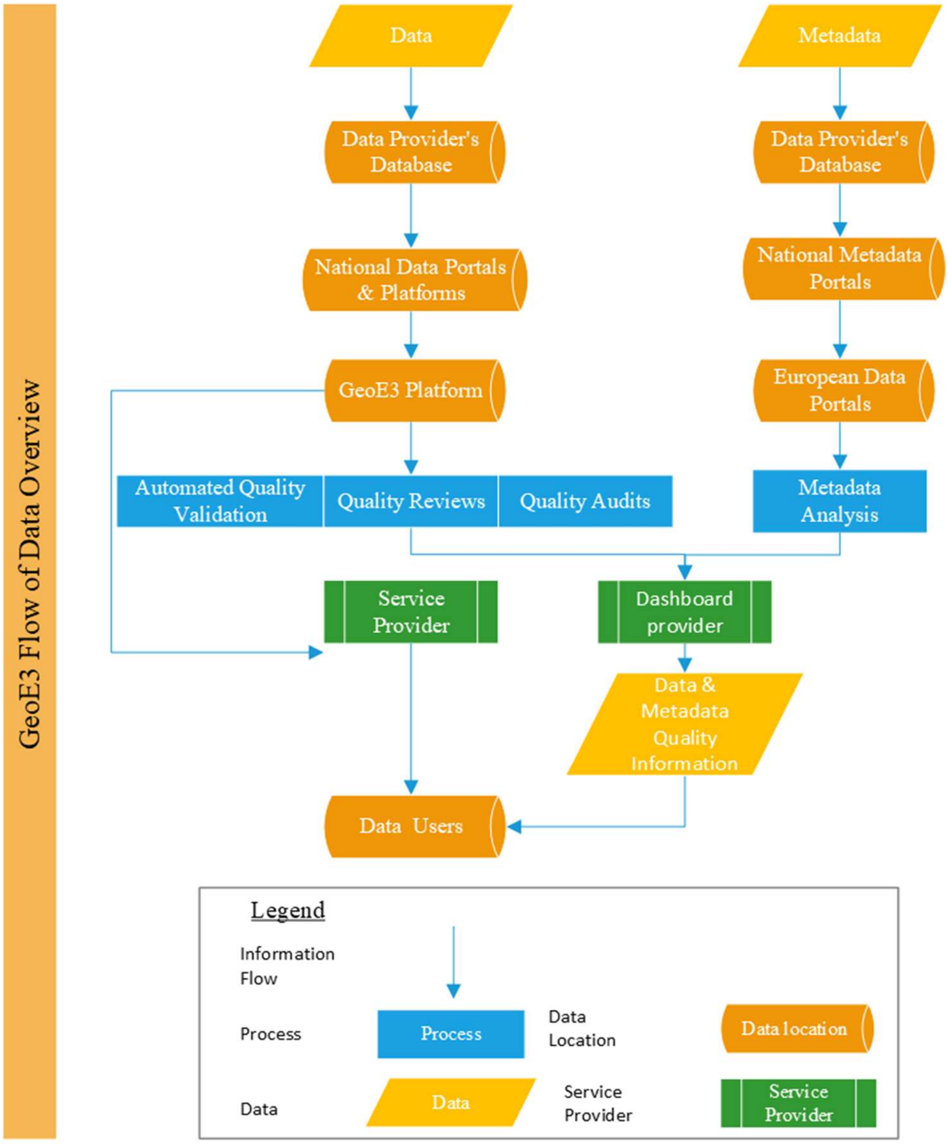


Figure 0-1 Flow of data overview diagram

Our goal here is to produce a proof of concept based on selected data and theme considering one of the use cases' user requirements, which shall be identified and confirmed through interviews with user groups at different stages of the dashboard's development.

Ultimately, this proof of concept shall be implemented in the continuation of this study, by the end of the year 2022.

This memoir is structured around three main axes. The first part will be focusing on reviewing the existing literature related to subject matters such as the GeoE3 project and its use cases, existing dashboards and data quality analysis methods and standards. Secondly, we will go through the methodology that led to the creation of our proof of concept, including the selection of platform for the quality dashboard pilot development, the assessment of user requirements, the design of mock-ups and first feedback. Finally, before discussing the future of this project, we will appraise our results consisting in the quality dashboard's final version, chosen data quality metrics and user feedback.

2 REVIEW OF LITERATURE

2.1 The GeoE3 project

First of all, it is essential to understand the environment in which the quality dashboard will be integrated. This dashboard is part of a much bigger project, and its creation cannot be viewed in isolation.

2.1.1 Introduction to the project

The Geospatially Enabled Ecosystem for Europe (GeoE3) project is a three-year project that aims to provide dynamic integration of high-value data sets and services with geospatial features from existing national geospatial data platforms.

On one hand, the GeoE3 project involves five European countries through their national cadastre and mapping organisations: Spain, the Netherlands, Norway, Estonia, and Finland. The National Land Survey of Finland – *Maanmittauslaitos* – is the coordinator and the driver of technical implementation.

On the other hand, an ample number of geospatial data operators are also associated with the project (cf. Appendix A). In Finland, those include Spatineo, the Finnish Meteorological Institute and Statistics Finland.

The technical implementation based on the Open Geospatial Consortium's interfaces focuses on enabling cross-boundary services through the use of new OGC API Features and OGC API Coverages interfaces.

The first services implemented in the project have already seen the light of day. The 2D and 3D building data of four countries (Finland, Norway, the Netherlands, and Spain) is available in the form of datasets through a single OGC API Features interface. Furthermore, the surface models for Finland, Norway and the Netherlands have been made available as datasets through the OGC API Coverages interface.

As a result of diverse data – including building data, surface models and climate data – becoming is readily accessible and available for integration, GeoE3 was able to create a web interface to present visualisations of these datasets. On Figure 2-1 below, buildings have been retrieved from the OGC API Features interface and displayed on a background map. On the right, the selected building is presented in 3D.

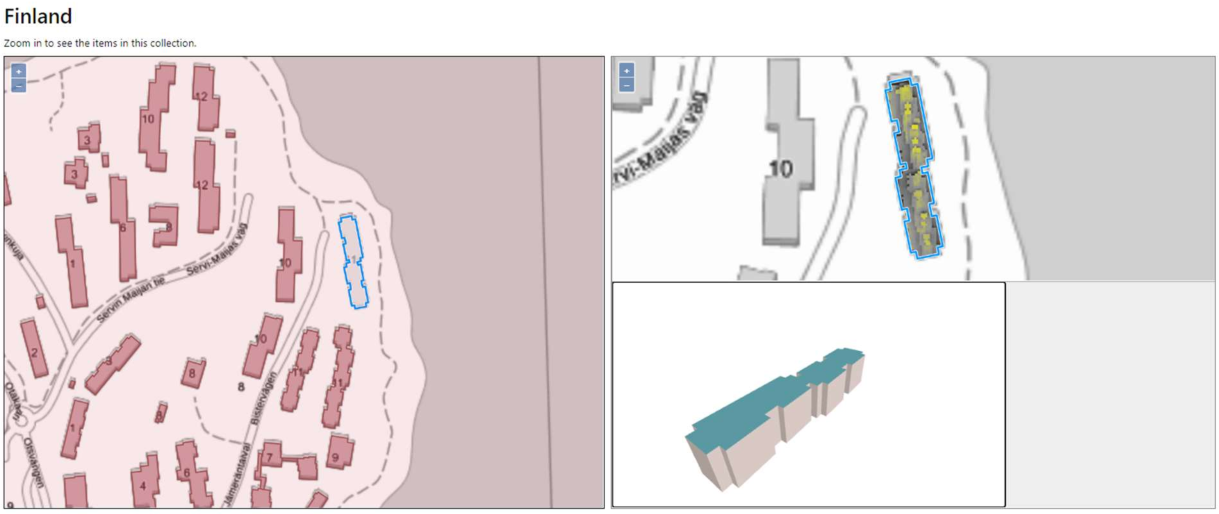


Figure 2-1 Example of buildings retrieved from the OGC API Features interface on a background map (geo3platform.eu@2022)

Several other presentations of building data are available, including a visualisation of the area's surface model and information about the average wind speed in the area, that could be used by stakeholders to assess solar energy potential and wind power in the area.

2.1.2 GeoE3's goals

The project's goals have been defined as follow by the European Commission (2019) as the following:

- **Simplify the discovery and evaluation processes for data and APIs by deploying new quality tools**;
- Enable integration of high value tabular data with geospatial content;
- Ensure data is DCAT-AP compliant and provide suggestions for improving the core service platform;
- Create interoperable datasets and modern, web friendly geospatial APIs supporting three highly relevant climate change and urbanization related use cases;
- Develop proven generic methods and tools for creating cross-border, cross domain applications for the defined use cases;
- Establish a permanent geospatially enabled ecosystem based upon national platforms supported by a network of API developers supporting an expanding user community;
- Prove the geospatially enabled ecosystems can be based on national geospatial platforms by implementing it with all national partners;
- Provide an infrastructure for sustaining GeoE3 based innovations.

Our task of creating a quality dashboard plays an integral part to the first goal of this list.

The European Commission has expressed the project's desirable outcomes. The first one, in which our quality dashboard will again play a part, is to enable high performance computing to improve public services and grow the ecosystem of data, and services, and the second one is to boost the data capability of the DSI with more readily discoverable and interoperable content.

According to Lehto et al. (2021), various technologies are to be tested for providing access to geospatial resources in the context of the new Open Geospatial Consortium's family of service interface standards. They additionally state that the main goal of GeoE3 is to develop use case-oriented, cross-border and cross-domain geospatial services and to integrate content by applying dynamic, service-level methods for joining other data with geospatial features, such as meteorological and statistical data.

The project's leading idea is that interoperability can be achieved using a data ecosystem based on national platforms. Therefore, if there is for instance a need from one entity for data covering several countries, the cross-border implementation would make it possible to fetch the data from several different data services and combine the results.

In conclusion, the GeoE3 project aims to improve the accessibility and interoperability of geospatial data, advance the harmonisation of geospatial data, and build an ecosystem based on national platforms.

As a concrete result, the project is to build and provide a test platform – whose demo version is at this time available on https://geoe3platform.eu/geoe3 – for demonstrating the benefits of interoperability. The project will also produce instructions, online courses and other support material related to the theme of the project.

Through the study of use cases associated with renewable energy, smart transport, and urbanization (see section 2.3 Use cases in the GeoE3 project), the project aims at supporting renewable energy- and smart city-related applications.

Access to a given dataset is provided by OGC API services, which therefore play a central role in the GeoE3 project.

2.2 Application Programming Interfaces

2.2.1 General principles of Application Programming Interfaces

The Web is made of a large network of connected servers. When a request is made on a given Web page displayed on a browser, it goes out to the website's remote server. Upon receiving the response, the browser interprets the code and displays the page. The part of the server that receives requests and send responses is an API.

According to Santoro et al. (2019), the acronym "API", standing for Application Programming Interface, probably first appeared in 1968 and was defined by Cotton & Greatorex (1968) as a collection of code routines that provide external users with data and data functionality. However, we will adopt a more modern definition of APIs coined by Shnier (as cited by Santoro et al., 2019) who describes them as “calls, subroutines, or software interrupts that comprise a documented interface so that an application program can use the services and functions of another application, operating system, network operating system, driver, or other lower-level software program”.

An API, in its most basic form, is an interface that allows a product and exterior queries to communicate. Essentially, they consist of contracts that define how developers communicate with a service, and the kind of output they should expect to receive back from the service.

APIs have multiple identified uses. The first one is to help developers perform their task. Applications can request from a device’s API to perform a function without being involved in how the request is resolved. APIs therefore allow developers to save time by taking advantage of a platform’s implementation and concomitantly allow existing system to extend their functionalities. This helps reduce the amount of code developers need to create and provides more consistent content across apps for the same platform.

Additionally, given that APIs are used to control access to hardware devices and software functions – that an application otherwise may not have permission to use – APIs can be considered gatekeepers of our personal data. For instance, when a browser displays a notification to request access to a device’s location, it means that the website is requesting geolocation data through the browser’s API. The only way to access such hardware resources – in the given example, the GPS sensor – is through the API. On that account, the browser can control access to the hardware and limit what apps can do.

2.2.2 Application Programming Interfaces and geospatial data

According to Steiniger & Hunter (as cited in Titov, 2021), given the increasingly heterogeneous nature of geospatial datasets as well as the tremendous volume of environmental datasets, the use of APIs when dealing with spatial data and Earth science data repositories has become essential.

Furthermore, in their presentation of Spatial Data Infrastructures, Phillips et al. (1999) emphasize the fact that data collected for a particular project are or could be, in most cases, useful for other projects. Since the costs involved with data collection are always considered in project planning, along with attempting to maximize the use of the data from a project, there is a great benefit of databases becoming a shared accessible resource.

In the context of this study, GeoE3’s principal focus is APIs’ ability to give users access to data from third party data providers.

In particular, the project uses the OGC API family of standards – a family based on the OGC Web Service standards destined to be an open standard for querying geospatial information on the Web. The Open Geospatial Consortium (OGC@2022) states that this family of standards is being developed specifically with geospatial data users and providers in mind in order to help influence and refine the process of standardization in related fields. OGC APIs therefore aim to ease web access to geospatial data and tools.

The OGC API standards, destined to be used to produce novel APIs, are constructed as "building blocks" and provide a way to connect diverse software program and automate operations. They are defined by both the requirements of the specific standards as well as through interoperability prototyping and testing in a forum led by the Open Geospatial Consortium – the OGC's Innovation Program.

According to the work of Jakobsson & Lehto (2021), using those APIs, cross-border services are being implemented. One of the goals of the GeoE3 project is to develop new service interfaces in accordance with the OGC API standard family across the boundaries of countries and domains.

2.3 Use cases in the GeoE3 project

GeoE3 addresses two main modern European challenges: climate change and urbanization. Even though the future GeoE3 services and tools are intended to be generic, it has been decided to approach these challenges through three use cases that will serve as a Proof of Concept for the deliverables.

Each use case requires the use and analysis of different types of data. A visual representation of the use cases and the main data involved in them can be seen below in Figure 2-2 : Visual representation of the three use cases involved in the GeoE3 project and the main types of data that have been identified for each use case.

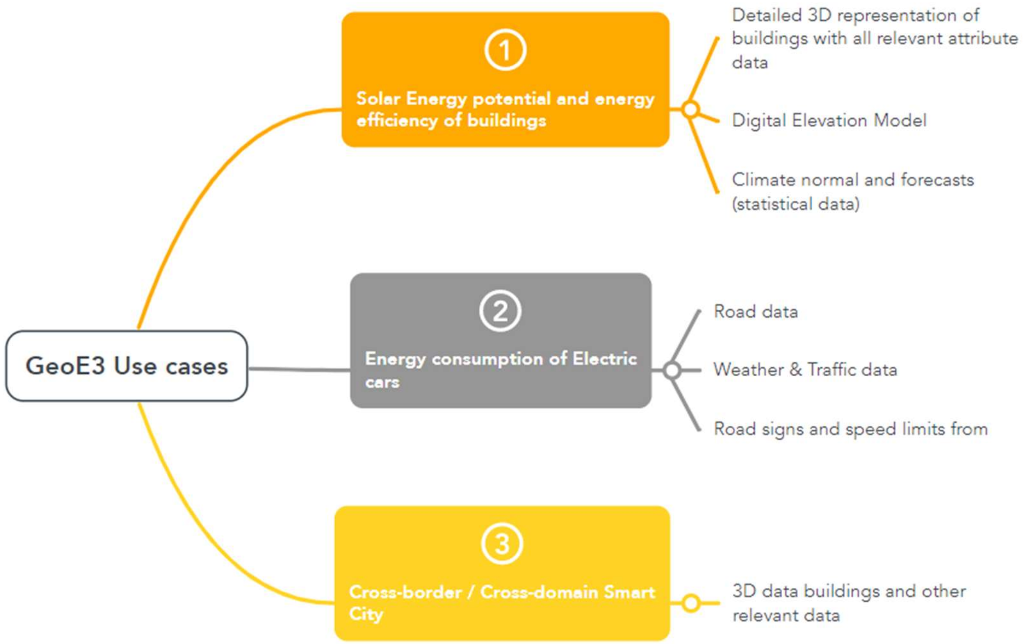


Figure 2-2 : Visual representation of the three use cases involved in the GeoE3 project and the main types of data that have been identified for each use case

Given that the present paper is dedicated to a task focused on the Use Case 1, we will go into more details regarding that specific use case than the two others.

The information related to the use cases has been obtained through Jakobsson (2020)'s Description of Work (internal document to the GeoE3 project).

2.3.1 Use case 2: Cooperative Intelligent Transport Systems (C-ITS) and Advancing map enhanced driver assistance systems leading to automated driving (ADASIS)

The second use case focuses on urbanization through the study of cooperative intelligent transport systems (C-ITS) and advancing map-enhanced driver assistance systems that will lead to automated driving.

This use case provides beneficial information in the context of climate change by studying electric cars and automated driving, as exploiting geographic information can reduce energy consumption.

Key datasets used for this use case and provided by the GeoE3 platform include fresh vector tiles for the visualisation of roads in both 2D and 3D, weather data and traffic data, traffic signs and speed limits from the national road databases in a selection of countries.

2.3.2 Use case 3: Cross border & Cross domain Smart City Finland Estonia

The objective of the third use case is to provide data and APIs to help create solutions for smart cities in collaboration with the FINEST project.

Innovative concepts and solutions are to be tested and implemented in Tallinn and Helsinki before being scaled globally.

Key resources used from the GeoE3 service platform include 3D vector tiles for visualisation of building and other data, tabular data on the smart cities for meeting the renewable energy solutions targets, and support for the companies developing the services.

2.3.3 Use case 1: Solar energy potential and energy efficiency of buildings in changing climate and smart cities

According to Jakobsson (2020)'s analysis, the first use case revolves around the study of buildings' solar energy potential and energy efficiency in the context of climate change and smart cities.

The goal of this use case is to investigate solar energy potential at building scale and ensuring their energy efficiency by providing data that helps in optimising their heating and cooling procedures. The GeoE3 service ecosystem will provide input data for the analysis via APIs.

The use case of renewable energy focuses especially on the generation of solar energy and the optimisation of heating and cooling systems in buildings. From the perspective of this use case, key datasets include building and elevation data.

Special focus has been placed on the processing of 3D building data and surface models. The positioning of solar panels can be optimised using roof shapes obtained from the LoD2 geometries of buildings. The local environment's surface model can be used to assess the amount of solar energy available in the area. The assessment also addresses climate conditions and variation based on local climate change scenarios.

A typical user could be a company providing intelligent analytics for real estate owners. To create preliminary analysis, the company might want to use resources from the GeoE3 service platform such as:

- Detailed 3D representation of buildings (LoD2) with all relevant attribute data;
- High resolution Digital Elevation Model of the surrounding terrain;
- Climate normals for the exact location of the property;
- Different scenarios for the future climate at the specific location.

To analyse solar energy potential, roof models are analysed together with the computational sun inclination models and predictions of the future climate. To optimize their installations, the company might use the same property information, but instead of employing climate statistics, use current weather observations and forecasts.

All the required information should be available from the GeoE3 service platform via easy-to-use interfaces and standardised data formats.

Data for the use case will be either provided by the GeoE3 partners or sourced from open data national portals.

2.3.4 Conclusion

These three use cases have the opportunity to involve many users with different backgrounds and levels of expertise interested in a variety of data all over Europe.

One of the identified goals of GeoE3 is to make metadata more visible and understandable for non-geospatial users and improve discoverability of data. To reach that goal, one task of GeoE3 has been dedicated to the development of a visualization tool: a dashboard. The present study's goal is to produce a Proof of Concept dashboard that allows visualization of data and metadata quality for datasets related to the Use Case 1 related to solar potential energy.

We will now check in the literature for possible existing solutions to this problematic. Dashboards as tools of data visualization are not new and have been broadly used in business analytics; therefore, the literature could present interesting suggestions in the application of this tool to geospatial data and data quality.

2.4 Existing Dashboards

In order to apprehend the opportunities and possibilities of geospatial dashboards, this paper's segment will go through some of the geospatial dashboard solutions – as well as their ambit – that have been developed around the world over the years.

2.4.1 A brief overview of geospatial dashboards

Dashboards are originally instruments that display the operation of a system in real time – the most obvious example being a vehicle dashboard that displays the necessary information to the user, the driver. According to Batty (2015), the first version of analytical dashboard appeared in the late 1960's and early 1970's and were essentially portals to hand-collected information. According to Changfeng et al. (2019), geospatial dashboard first appeared in the literature in 1990, with geospatial analysis features only being considered only recently in the 2010's. A timeline created by Changfeng et al. (2019) displaying significant geospatial dashboard development phases and events from 1990 to 2018 can be found in Figure B-1 in Appendix B.

According to Lwin et al. (2019) and Lwin et al. (2015), the term "dashboard" is nowadays commonly used by many developers and government agencies. They refer to them as tools whose primary function can be customarily defined as the collection and monitoring of remote objects. This monitoring is done by providing a rich graphical user interface that contains metrics of business performance to facilitate and improve decision-making.

From a decision-making point of view, there is a significant gap between the datasets and their end users. This gap limits the performance of the end users' analytical functions and, therefore, the extraction of information from the datasets. Geospatial dashboards currently present a strong attraction for academic and government institutions who hope to fill this gap using smart monitoring, visual analysis, better geovisualization and user-friendly, interactive, spatial analysis of urban big data.

2.4.2 City dashboards

Literature regarding geospatial dashboards, as of today, revolves mostly around a category of dashboard referred as "city dashboards", which are used to measure, interpret, and display in a synthetic way smart city performance. For this purpose, they generally aim to collect, visualize, and analyze big data. According to Damari (2017), this push for research in cities to be more data-driven is partly explained by the fact that geospatial information is becoming more common and available.

The review of city dashboards is interesting from the point of view of our study for a number of reasons. Indeed, the majority of literature covering geospatial dashboards is related to city dashboards. Therefore, a lot of analysis has been done on the different aspects that need to be taken into account when creating a dashboard – those will be more thoroughly detailed in section 3 Methodology of this paper. Moreover, city dashboards are designed to be used and understood by a wide range of users that have different backgrounds. For that reason, it is interesting to study how city dashboards have managed – or sometimes, have not, according to Young et al. (2020) – to accommodates to this challenge.

These dashboards are web-based interfaces with a scope usually limited to one city. Their focus varies from city to city, but we have compiled the main category of data that are displayed on city dashboards : social data, security, economics, population data, environmental data, social media information, mobility, energy, and health (see examples of information related to these categories in Appendix C, Table 9). Our review of city dashboards shows as well that they usually support static and / or real-time information, interactive maps, videos, sensor data, official reports in CSV or HTML format, indicator gauge data, KPI data generated from external services providers, data generated from citizens including citizen feedback data (Toronto.ca@2022, Changfenf et al. (2019), dashboard.edmonton.ca@2022, Young et al. (2020), and snap4city.org@2022).

According to Young et al. (2020), issues regarding existing dashboards have been raised by Kitchin and McArdle in several related papers (Kitchin and McArdle, 2017, Kitchin et al., 2016, McArdle and Kitchin, 2016a and 2016b). Those that relate the most to this paper's problematic include veracity and validity of data, usability, use and utility of the dashboard, and access.

2.4.3 Visualization and design considerations for geospatial data analysis

According to Badard & Dubé (2009), geospatial dashboards are a combination of Business Intelligence (BI) and GIS application on one single architecture. In that way, geospatial dashboards are communication tools between data and knowledge that better support the process of data analysis and help geodata users make more informed decisions.

The Cork Dashboard, launched in 2017 and decommissioned in January 2022 (dashboards.maynoothuniversity.ie@2022), was developed for web browsers. The landing page – which can be seen in Figure D- – presents the user with a number of general information and indicators and a possibility to dive deeper using clickable buttons.

A second layout can be seen in the Hong Kong Smart City dashboard in the appendices**,** Figure D-, which displays six different theme tabs that categorize the information displayed.

Finally, the London city dashboard (Figure 2-3 below) is one example among seven other similar city dashboards developed for cities in the United Kingdom. Young et al. (2020) describe it as an "at-a-glance" city dashboard. It summarizes and aggregates data and displays it using a modularized interface and an interactive map.

Users can view on a single web page real-time information about the weather, air pollution, public transport, public bike availability, river levels, electricity demand, the stock market, twitter trends, live traffic camera feeds, and the "happiness" level of the city. The data provided in the display are sourced from a diverse set of data suppliers using APIs.

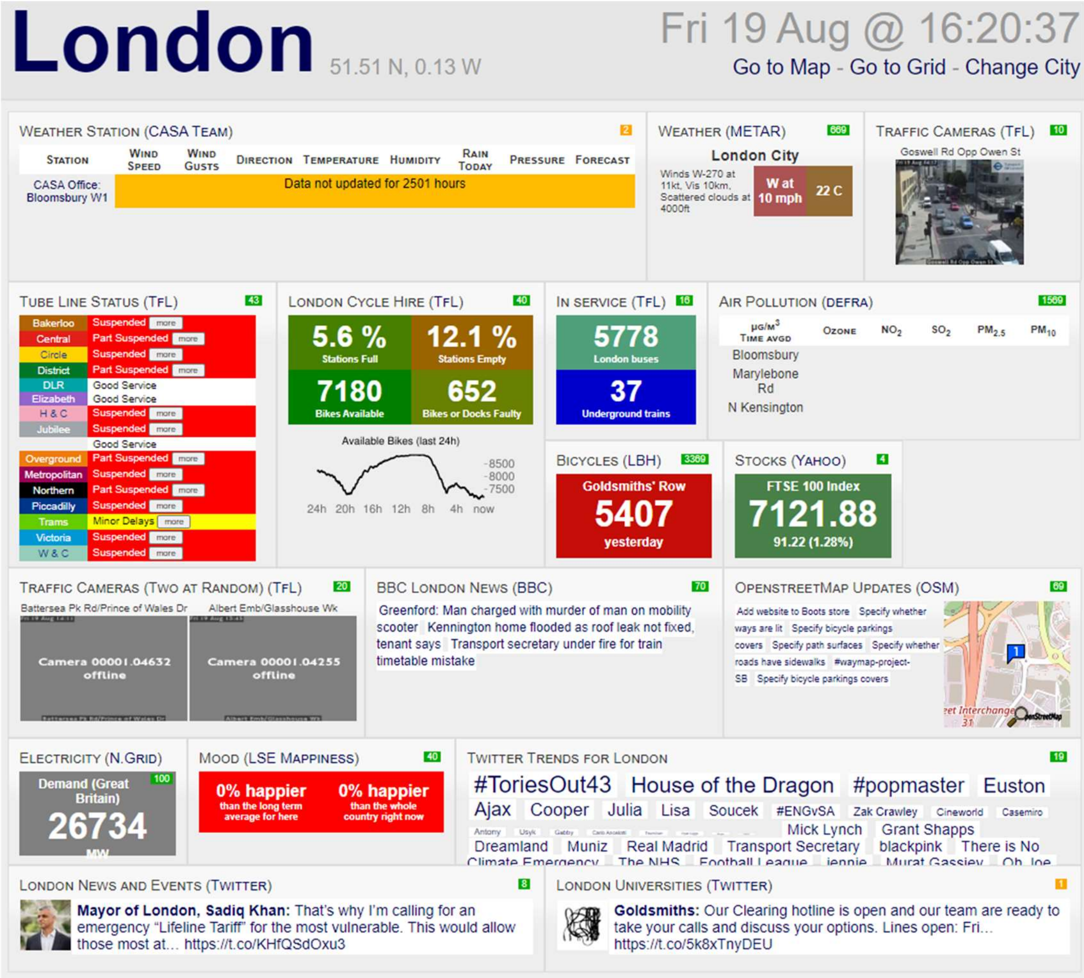


Figure 2-3: London city dashboard (citydashboard.org/london@2022)

According to Young et al. (2020), a number of authors have found that that many city dashboards are too specialized and complex, requiring higher-level data literacy, to be easily used by the general public ; this indicates that there is room for improvements to better the openness and accessibility of such open-data portals.

2.4.4 Dashboards focused on metadata and data quality assessment

There are a few dashboards that have been designed over the years to focus on metadata and data quality assessment – and even fewer are focusing on geospatial data.

Huisman-Van Zijp & Sjoukema (2020) describe their quality dashboard created for the Dutch Kadaster as a tool for communication dedicated to producers and quality managers ; one of the outcome of Huisman-Van Zijp & Sjoukem's paper is that the dashboard in question was too specific and needed to be more generic.

They however introduced promising functions into the development of this dashboard.

The history function allows progress of the measured indicators to be traced, while the crowdsourcing feedback function, thanks to which users can flag errors in the data when they spot it. Additionally, this function can be used to measure indicators that cannot be quantified automatically.

Vaidyambath et al. (2019) have created an intelligent dashboard (Figure 2-4 below) dedicated to help the user identify quality issues in order to prioritize and repair them appropriately.

They specifically investigated the use of two widgets to achieve these goals and found that it is still hard for novice users to understand the results displayed on their dashboard via these simple widgets. Therefore, there is still work to be done regarding the understandability of quality analysis on dashboards.

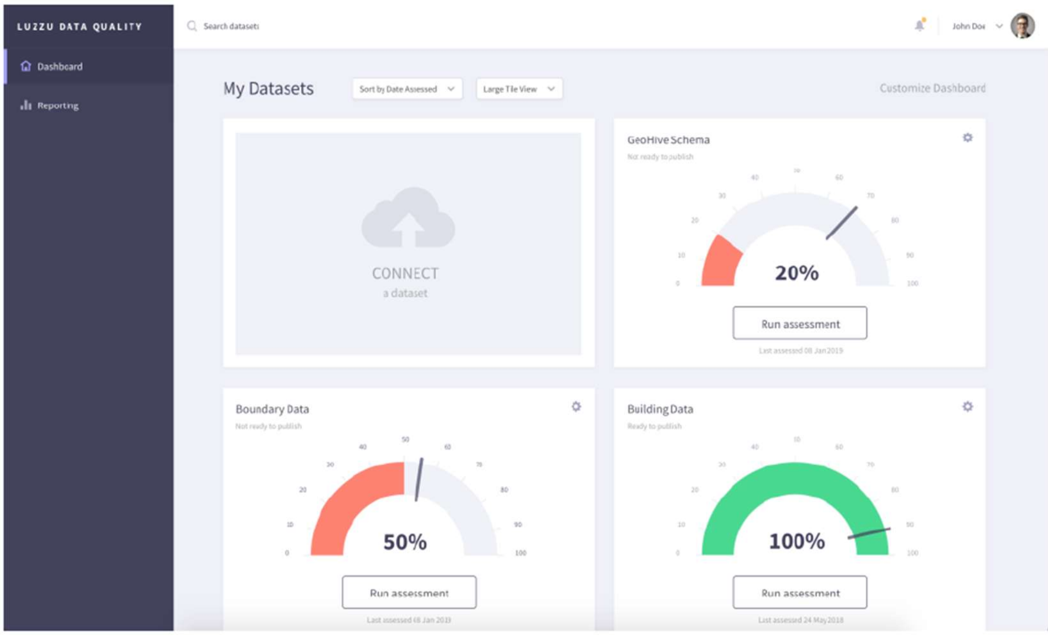


Figure 2-4: Vaidyambath et al. (2019)'s intelligent data-driven dashboard

Towards the end of July 2022, a metadata quality assessment tool was added to the data.europa.eu website, which is the official portal for European data. Given that the development of this dashboard came to our notice towards the end of the present thesis, we will take advantage of the situation by comparing our findings and method to theirs in Chapter 4 : Results, of this paper.

2.4.5 Conclusion

In addition, we will briefly evoke a couple of the more prominent uses found in our review of geospatial dashboards.

First, this tool can also be used to react to emergencies and to assemble data rapidly on a given event. Lwin et al. (2019) have discussed the use of the City Geospatial Dashboard as a geospatial solution provider for disaster management and the importance of this dashboard as a tool in the disaster response stage. The City Geospatial Dashboard has played important roles in collecting, sharing, and visualizing big data analytics results so that this data is available to planners, decision-makers, as well as any geospatial information users in a timely manner.

Moreover, in the last couple of years, many Covid-19 related dashboards have seen the light of day all over the world, whose scale vary from the world to country, city, neighborhood or even smaller.

Along with the example discussed above, a few geospatial dashboards are also being developed at the moment of the writing of this paper. Indeed, institutions such as Statistics Finland, the Dutch Kadastre, the Technische Universität Dresden – through the GeoKur project – for instance are in the process of creating their own geodashboard.

We have seen that dashboards are of great interest to a lot of institutions as they can be very efficient tools to visualize information that can be used to select the right dataset and interpret the data in the right way. They have been abundantly used in the study of smart cities and the way to visualize data has been central to the creation of dashboards from the beginning. However, geospatial data and metadata quality analysis are much rarer in the literature, and this added geospatial nature attaches different expectations and challenges to the selection and visualization of such information.

According to Batty (2015) and Batty et al. (2012), integrating data that is diverse is extremely difficult, often impossible due to the absence of common keys, the inconsistent format of the data, the presence of noise in dataset or missing data. The paths we have followed to challenge these issues will be addressed in chapter 3 of this paper. In the meantime, literature has been very proficient in identifying data quality standards. We will review those in the following section.

2.5 Data quality evaluation and interoperability

Data quality is multi-dimensional. Depending on the author or authors attempting to describe data quality, this concept can be split and labelled in different ways. In this section, with the help of different sources and point of views, we will try and gather the most appropriate ways to quantify geospatial data quality in a quality dashboard.

2.5.1 ISO standards

Ways to measure the quality of geographic data have been identified by several sources. One of the most high-quality sources is the International Organization for Standardization (ISO), who describe themselves as a worldwide federation of national standards bodies.

In our research, we have reviewed three ISO documents related to data quality:

- ISO 19157:2021, Geographic information — Data quality — Part 1: General requirements;
- ISO 19157:2021, Geographic information — Data quality – Part 3: Data quality measures register;
- ISO 19115:2011, Geographic information – Metadata – Part 1: Fundamentals.

The ISO 19157:2021 is a very thorough norm that aims to provide a framework for defining the quality of geographical data. It establishes in its Part 1 the principles for describing the quality of geographic data.

In particular, this norm presented several conceptual models of quality for geographical data that inspired us. For copyright reasons, we ask the reader to refer to the original document(s) to consult these models and schemas.

As per ISO 19157:2021, Part 1 (page 22), a data quality element is defined as "a component describing a certain aspect of the quality of geographic data". As illustrated in Figure 2-5 below, data quality elements cab themselves be described by measures, evaluations, results and by a metaquality evaluation.

Figure 2-5: Conceptual schema representing the descriptors of data quality elements – inspired by Figures 1 & 8 of the ISO 19157:2021, Part 1 document.

Metaquality is defined by the "information describing the quality of data quality". It is used to evaluate the suitability of the evaluation method – and therefore of its results.

An overview of the data quality elements can be seen in Figure 4 (page 21) of this norm. They are also listed in Table C.2 – Data quality elements in Annex C of the ISO 19157:2021 document. We will later refer and use these elements in to describe quality in our quality dashboard (cf. Chapter 3

Measures of data quality elements are also introduced as ways to make different datasets and data quality reports comparable. The norm suggests a list of standardized measures that can be completed by more use case-specific quality measures defined by the user in order to coincide with the nature of the dataset in question.

Additionally, this norms (ISO 19157:2021, Part 1, pages 32-36) recommends that data quality measures be described by twelve data quality components:

- A unique measure identifier;
- A measure name;
- One or more aliases;
- The name of the data quality element or elements that the measure refers to;
- The basic measure from which the data quality measure is derived;
- The definition of the measure;
- The description of the measure – that includes the expression of its calculation if applicable
- Measure parameter or parameters, which are subsidiary variables used in the original measure;
- The value type of the data that reports the result of the measure;
- The value structure if the measure's result is expressed with more than one value;
- The source reference or references;
- One or more example of applying the measure and / or its result.

In Part 3 of the ISO 19157:2021 document, we can find the detailed definitions of the data quality elements and measures mentioned in Part 1 – they are incidentally summarized in Table 1 below, which is inspired by Figures 2 and 4 of ISO 19157:2021.

Quality Elements:	**Completeness**	**Thematic quality**	**Logical consistency**	**Temporal quality**	**Positional accuracy**	**Metaquality**
Subsidiary Quality Elements:	Commission	Thematic classification correctness	Conceptual consistency	Accuracy of a time measurement	Absolute positional accuracy	Confidence
	Omission	Non quantitative attributes correctness	Domain consistency	Temporal validity	Relative positional accuracy	Homogeneity
		Quantitative attribute accuracy	Format consistency	Temporal consistency	Gridded data positional accuracy	Representativity
			Topological consistency			

Table 1 : Overview of the data quality elements and their measures as identified by the ISO 19157:2021 document

The annexes of this document provide a list of standardized quality measures and their detailed definition. It is organized by element (completeness, thematic quality, logical consistency, temporal quality, positional accuracy and metaquality) that are described by numerous subsidiary elements, which themselves are described by measures made of several components. Within those quality components, we can find the actual parameters recommended to be used to quantify the quality of each measure.

On the other hand, the international standards required to describe geographic information and services with metadata are stated in Part 1 of the ISO 19115:2011 document. This norm focuses on covering the fundamentals principles of geographic metadata.

In particular, this document lists the minimum set of mandatory metadata required to “serve the full range of metadata applications, including data discovery, determining data fitness for use, data access, data transfer, and use of digital data and services”, as well as other conditional and optional metadata elements. A minutely detailed list can be accessed in Annex B of ISO 19115-1 and could be tested against metadata of datasets submitted to the GeoE3 platform.

As a side note, in comparison, the CSC – Science Information Technology Centre, a Finnish information technology competence centre owned by the state and universities involved in Fairdata, has developed a research dataset description tool, Qvain, in order to help harmonize metadata (Fairdata.fi@2022). It allows the user to fill-in a form and save the metadata before publishing it. Additionally, the tool is able to create identifiers automatically.

The information requested by Qvain includes the following categories and fields:

- Data origin: License and Access type;
- Description: Title, Description, Issued date, Keywords, Language, Field, Identifiers;
- Actors;
- Related publications and other material;
- Geographical area;
- Time period;
- History and events (provenance);
- Project and fundings.

In summation, the study of these norms will be very useful in our work. Indeed, not only will we be able to apply the quality measures presented – in length and great details – in those documents, but we also will be able to ask the data providers to refer to these international standards when they are producing dataset so that harmonization may be improved.

As previously stated, these documents are copyrighted. This is why we ask the reader to refer to the original documents if they are interested in the definitions of elements mentioned in this and the following chapters. Moreover, the measures drawn from the norms for our dashboard will be referenced as so in Chapter 3 Methodology.

2.5.2 FAIR guiding principles

Wilkinson et al. (2016) discuss data quality standards in a different way in their FAIR guiding principles, in which hey state that datasets should be Findable, Accessible, Interoperable and Reusable. Among those principles, they are pushing comprehensive description of data through extensive metadata.

The FAIR guiding principles are listed within four categories which are summarized in the diagram below (Figure 2-6).

F Findable		A Accessible		I Interoperable		R Reusable	
F1	Data & metadata are assigned a **globally unique and persistent identifier**.	A1	Data & metadata are **retrievable** by their identifier using a **standardized** open, free, and universally implementable (A1.1) **communications protocol** that allows for an authentication and authorization procedure, where necessary (A1.2).	I1	Data & metadata use a formal, accessible, shared, and broadly applicable **language** for knowledge representation.	R1	Data & metadata are **richly described** with a plurality of accurate and relevant **attributes:** Data & metadata are released with a clear and accessible data usage **license** (R1.1), associated with **detailed provenance** (R1.2) and meet domain-relevant **community standards** (R1.3).
F2	Data are described with **rich metadata** (defined by R1).						
F3	Metadata clearly and explicitly **include the identifier** of the data it describes.			I2	Data & metadata use **vocabularies** that follow FAIR principles.		
F4	Data & metadata are **registered or indexed in a searchable resource.**	A2	Metadata are **always accessible**, even when the data are no longer available.	I3	Data & metadata include **qualified references to other** data & metadata.		

Figure 2-6 : The FAIR Guiding Principles - diagram inspired by Wilkinson et al. (2016)'s classification[1]

Wilkinson et al. (2016)'s analysis on data management and harmonization not only provides us with leads to determine and classify data quality standards in our dashboard, but it also gives us straightforward items to be checked when analysing the quality of data and metadata. Therefore, we will be able to use some of those elements as measurements of data quality in our further work.

Additionally, Bahim et al. (2019) have published a FAIR data maturity model dedicated to developing an ensemble of assessment criteria. Out of their work came a list of indicators (cf. Appendix E FAIR data maturity model indicators, Bahim & al. (2019)) linked to the FAIR guiding principles and rated by three degrees of priority: useful, important, and essential (cf. Appendix F FAIR maturity model indicator priorities, Bahim et al. (2019)). This maturity model is a tool that can be used to assess the "FAIRness" of the data in question and to point out aspects of the data that need improvement.

2.5.3 A focus on interoperability

According to Buehler & McKee (1996) as mentioned in Philips et al. (1999), interoperability in the context of geospatial information can be defined as the capacity to exchange and manipulate geospatial information freely and cooperatively. Therefore, a user should not have to worry about the format of the data they are interested in nor about the system they are using.

[1] The numbering of these principles will used as a reference in later chapters of this paper.

As stated in section 2.1 of this paper, interoperability is at the heart of the GeoE3 project, and a lot of progress has been achieved in the last twenty years.

In their working paper, unpublished at this time, Jakobsson et al. (2022) discuss the evaluation and visualization of interoperability. For this purpose – and in the context of the GeoE3 project –, they have developed an Interoperability Map based on the assessment of a comprehensive four-level maturity model. Those levels are described in Table 9 below.

Level of Interoperability	**Description**
Level 0	Not interoperable and cannot be integrated
Level 1	Minimal interoperability and can be integrated with extra effort
Level 2	Intermediate interoperability and can be integrated mostly automatically
Level 3	Advanced / Optimal interoperability and can be integrated automatically

Table 2: Maturity model for Interoperability map as defined by Jakobsson et al. (2022)

The level of interoperability is determined for a number of different categories of aspects via specific criteria. These are described in Jakobsson et al. (2022) submitted – but unpublished at this date – work and can be seen in Appendix F.

The following table showcases an interoperability map for a couple of datasets in Finland. The different levels are a good way to inform any user as to the global degree of interoperability for one or more datasets.

This Interoperability map could be integrated in a quality dashboard as a highly informative, intelligible, and accessible overview of interoperability for a given dataset.

Country/Dataset	Legal aspects/Organizational aspects	Technical aspects/Data access		Semantic aspects		
	National data accessibility and integration arrangements	metadata discoverability	data accessibility	Vocabulary and data specifications	Data content and data quality	Quality assessment (QA)
Finland/Buildings, 2D	Level 1 (Open data, no national platform)	Level 1 (no DCAT AP)	Level 3 (OGC API)	Level 1 (definitions available but not MR)	Level 1 (national schema)	Level 1 (QA available but not published)
Finland/Buildings, 3D (test dataset)	Level 1 (Open data, no national platform)	Level 0 (no metadata)	Level 1 (no API)	Level 1 (definitions available but not MR)	Level 1 (data content limited)	Level 0 (no QA)

Table 3: Interoperability Map on building data, extracted form Jakobsson et al. (2022), Table 5, p13

2.5.4 Conclusion

We have herewith found several sources that describe ways to quantify data quality in a way that agrees with the challenges of the GeoE3 project. These measures need however to be structured in a comprehensive manner so that the targeted users can make the most of this heavily technical information.

In the next section, we will touch upon the process suggested by authors in order to achieve an efficient and thorough data quality analysis.

2.6 Data Quality Evaluation in context of use case

Rula & Zaveri (2014) have stated that there are numerous existing methodologies dedicated to the assessment of the quality of a dataset. However, those usually do not address a particular use case and demand that users be involved in the process, which means that they should possess a certain level of expertise as well.

2.6.1 Rula & Zaveri's six-step process (2014)

Rula & Zaveri (2014) describe a three-phased, six-step process to quality assessment of a dataset. This process is summarized in Figure 2-7 below.

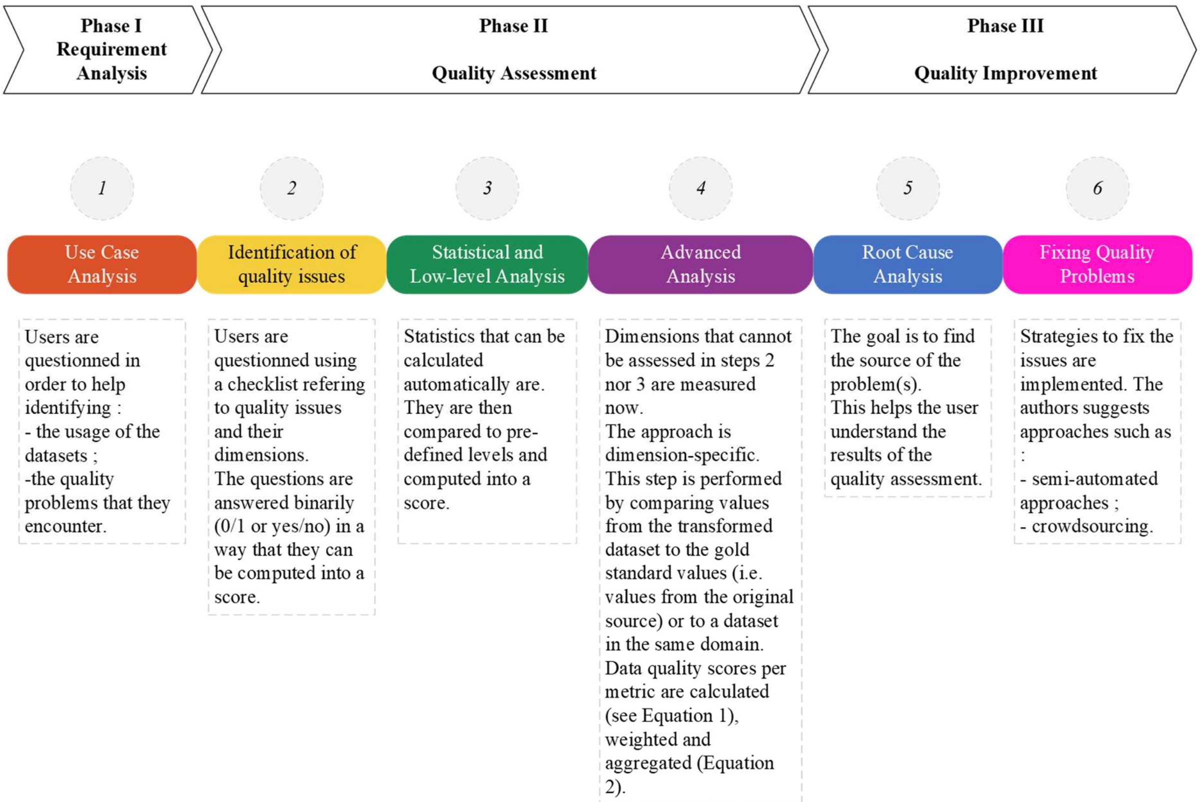


Figure 2-7 : Process steps of Rula & Zaveri (2014)'s quality assessment methodology – schema inspired by Rula & Zaveri (2014)'s schema (figure 1 in their paper).

In step 4 (see Figure 2-7 above) of their process, the authors refer to two equations whose objective is to calculate data quality scores. They are using values obtained by comparing values from the dataset in question to their "true" values – i.e., from either the original source or a dataset in the same domain.

The first equation (Equation 1) is a ratio calculation using the ratio between the total number of instances violating a data quality rule (V) and the total number of relevant instances (T). This ratio is then subtracted from 1 to obtain the data quality score (DQ_{score}).

$$DQ_{score} = 1 - \frac{V}{T}$$

Equation 1 :Data quality score for one property of the dataset (Rula & Zaveri, 2014)

To obtain an overall data quality score ($DQ_{weightedscore}$) for the dataset, a second equation is used (Equation 2); the individual data quality scores are weighted (w) relatively to the importance of the dimension measured, summed, and divided by the sum of all weighting factors (W).

$$DQ_{weightedscore} = \sum_{i=1}^{n} \frac{DQ_{score*w_i}}{W}$$

Equation 2 : Data quality weighted score for the overall dataset (Rula & Zaveri, 2014)

The weight given to each data quality score is heavily context-dependent – and therefore use case-dependent – and shall be determined manually.

This methodology involves automated, semi-automated as well as manual assessment tools but, in addition to regular quality assessment, the authors add a quality improvement phase to their process.

Furthermore, in the context of our study, it could be advantageous to add a step to this process to enable the evaluation of the tool used to communicate such quality analysis to the user.

2.6.2 Vaidyambath et al. (2019)'s evaluation method

In their development of an intelligent data quality dashboard, Vaidyambath et al. have developed an evaluation methodology that aims to allow users to give detailed feedback on their product so that they can understand how helpful their dashboard is in practice.

Two types of users, four experts and four novices in the domain of Linked data quality concepts – on which the dashboard focuses –, are invited to perform a list of tasks related to the identification and understanding of data quality problems.

The output data includes a response score based on the correct or incorrect answers given by the users to a questionnaire, as well as a grade assigned by the users for effort, confidence and usefulness in each task. The user data collected by Vaidyambath et al.'s experiment is shown in Table 4 below ; the "Analytics" and "Problem Report" lines refer to the two widgets analyzed during this experiment.

	User		*Response Score by Task*				*Effort by Task*				*Confidence by Task*				*Usefulness by Task*			
	Type	*No.*	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*
Problem Report	N	U1	10	10	10	0	4	4	3	7	8	6	8	8	8	8	8	8
		U3	10	10	10	0	4	4	4	7	7	7	7	6	5	6	6	6
	E	U2	10	10	10	10	2	2	2	4	9	7	9	7	8	8	9	8
		U7	10	10	10	10	3	3	4	5	8	7	8	7	8	8	8	8
Analytics	N	U5	10	10	0	3.3	5	4	5	7	9	10	5	5	8	8	5	8
		U6	10	10	0	5	6	3	5	9	9	9	7	3	3	5	5	4
	E	U4	10	10	10	10	1	1	1	9	1	10	9	10	9	10	6	7
		U8	10	10	10	10	2	3	2	7	9	9	9	9	8	8	8	8

Table 4: User data collected during the evaluation of Vaidyambath et al. (2019)'s intelligent dashboard.

Scores are finally calculated and compared for each user and each task in order to draw results (these results have been mentioned in 2.4.4 Dashboards focused on metadata and data quality assessment).

This step is an important one in the creation of any product destined to be used by a wider audience ; it will be a part of our process as well and will be discussed in chapter 4 Results.

2.7 Conclusion on the outcomes of the literature review

The present literature review has allowed us to understand the background on which our dashboard shall be built as well as the analysis expected in the context of the GeoE3 project.

We have seen that the concept of dashboard to display information has been used broadly over the last twenty years. In particular, dashboards focused on smart cities have brought a lot of new materials about the way to present data – that sometime possess a geospatial feature – to a wide range of users.

On the other hand, dashboards focused on data quality analysis have been designed, although they presented a few issues that prevents us to apply their model as is to our work. Their audiences often have to possess a certain level of expertise in data quality, their design could have been deemed unappealing or too complex, and they do not present the right data quality components that we aim to show.

Furthermore, we have not uncovered any existing dashboard specifically designed for data and metadata quality with a focus on geospatial data.

As emphasized by Nuradiansyah & Budi (2015), to allow efficient decision-making, the user needs to understand the data into the form of information, and this can be achieved by translating data in a visual form.

Therefore, using previous research on visualization of data quality, data quality standards and dimensions identified by authors and ISO norms, as well as suggested processes to the creation of data quality tools, we will be able to build a dashboard that fully fits in the context of the GeoE3 projects and its challenges.

In summary, our dashboard should make data more available to society in effective and engaging ways and appeal to a broad range of users with different backgrounds and levels of expertise. To handle the challenges raised to our attention by our review, we will tackle a set of fundamental challenges that we sorted in three main categories – user input, data, and visualization and interaction – that will be addressed through a process explained in the next chapter of this paper.

3 METHODOLOGY

This chapter and the next are dedicated to the creation of the proof of concept for a quality dashboard as part of the GeoE3 project.

Here, we will focus on the work that has led to the current final version of our quality dashboard, which will be presented in chapter 4

In the first of this chapter, we will go through the creation process we have constructed. This process will introduce the reader to the steps established in our proposed methodology that make up the subsequent sections of this chapter.

3.1 Process

This process is partly inspired by the work of several authors described in section 2.6, Data Quality Evaluation in context of use case, in this paper review of literature.

In view of the facts that this dashboard was to be created from scratch with multiple actors to consider along the way, it was necessary to establish a clear path to follow in our study. It is important to present in a clear way the steps that need to be followed for several reasons. The first is that, by identifying phases, it is easier to divide the work amongst different people and over time. The second is that this process could be reapplied for other applications than the one we are working on here.

This process is split into four main phases, one of which is to be the focus of future work within this project.

3.1.1 Phase I: Requirement analysis

During this phase, the requirements for the GeoE3 dashboard are identified and analysed. Even though the users are the one that should provide a majority of the requirements, they do not intervene yet (see 3.6 User interviews for further information).

The requirements and data generated from the GeoE3 project for the dashboard are reviewed and data quality standards are researched and transformed to fit the requirements.

3.1.2 Phase II: Data quality indicators assessment

Ways to assess data quality found in the previous phase are selected to fit our use case and classified in a comprehensive and logical manner that potential user may understand easily.

Additionally, the data quality measures part of our selection are defined in relation with our use case. For quantitative or statistical measures that are easily quantified, units are specified. For other or quantitative measures, ways to quantify them are described.

3.1.3 Phase III: Proof of Concept dashboard construction

This phase focuses on the production of a quality dashboard as a product and starts with the selection of the dashboard platform with a preference for readymade solutions. The functionalities necessary for the dashboard are estimated.

A simple test database with the previously selected indicators needs to be created to be linked to the chosen dashboard software.

This first version of the dashboard is evaluated by selected users through interviews and follow-up questionnaires and their suggestions are carefully considered before modifying the dashboard accordingly. A report on the evolution of user requirements is submitted to tasks leader within the GeoE3 project.

The user interviews in this project are done in collaboration with Lena Hallin-Pihlatie of the National Land Survey of Finland, who is in charge of collected feedback from users about the GeoE3 platform and their proposed services in general.

After a series of evaluation, the dashboard can be considered a proof of concept. The final version should be evaluated by users before the implementation.

3.1.4 Phase IV: Implementation

This is the next logical phase in the creation of our quality dashboard. However, it is not the focus of the present study.

3.1.5 Process overview

In Figure 3-1 below, the process drafted for the creation of a GeoE3 quality dashboard is summarized.

As can be gathered from that figure, some phases may be considered to overlap (phases represented in dotted lines) due to the necessary reassessment of features according to user feedback (which is represented in orange in the figure).

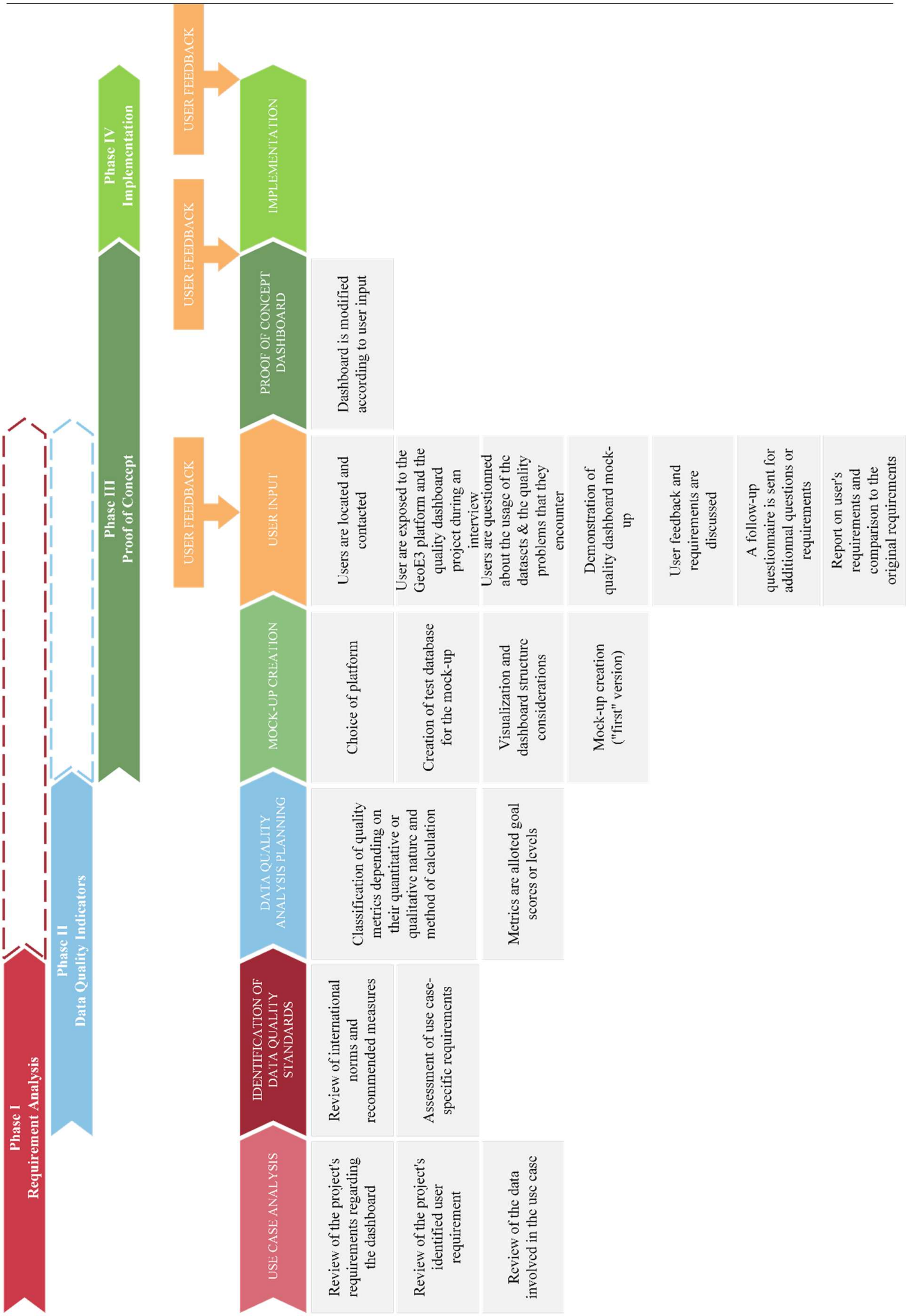


Figure 3-1: Overview of the process leading to the creation of a Proof of Concept quality dashboard for the GeoE3 project

3.2 Assessment of requirements

The first thing that needs to be considered in the creation of a quality dashboard is the users' requirement.

As a number of authors have previously stated (including Vaidyambath et al., 2019 when referencing Tayi & Ballou, 1998), data quality is often defined as "fitness for use" and, according to many sources, this characterizes the ability of data to meet users' requirements (Lacagnia et al., 2021; Vaidyambath et al., 2019, referencing Tayi & Ballou, 1998; Rula & Zaveri, 2014, referencing Juran, 1974).

Therefore, the assessment of user requirements has become the common thread of this paper and should be the reason behind every development in the quality dashboard. First, we can consider the preliminary requirements that derive from the GeoE3 project and led to the creation of the task asking for the development of a quality dashboard.

3.2.1 Preliminary requirements

The new quality dashboard should collect data from various services to give an insight on the quality of data, metadata and services related to the project.

It was decided very early on to focus on only one of the three use cases, use case number 1, focusing on the solar energy potential and energy efficiency of buildings in changing climate and smart cities. This use case was introduced in our literature review, in section 2.3.3. Indeed, the user requirements to be defined are very different depending on the use case considered and it is more efficient to focus on one use case and thus one type of user only.

It was also decided that the dashboard could assess the quality of non-purely geospatial data; as emphasized by the users surveyed, who stated that other kind of data are essential and prominent in their work. Consequently, the dashboard should be able to display the analysis of data of varying types in addition to 2D and 3D data with attribute information.

The product should first and foremost be able to satisfy users in their inquiry of the quality of the data that they are interested in during their work, easing the decision-making process that can be time-consuming and discouraging. For that reason, the quality analysis results' visualization should be particularly user-friendly and reliable, and the feedback from users should be taken into account with great consideration.

Nonetheless, those requirements are not coming directly from the users, although they are simultaneously the ones that we are aiming to help – as they are the ones the dashboard is designed for – and the one who know the most about the use case. Therefore, it is essential to survey a selection of users and ask them about their hopes and expectations regarding a quality dashboard in GeoE3.

3.2.2 Further requirements assessment

To create a useful dashboard for the project, it is imperative to identify the user requirements for one use case. It is therefore necessary to identify and question groups of users, whose requirements must be compared to the already identified requirements presented in the previous section (cf. 3.2.1).

Chronologically, as shown in the process overview in Figure 3-1 above, we were able to schedule interviews with users only after a first draft of a dashboard (cf. 3.5). This came with positive repercussions as the users could give feedback on a mock-up instead of a concept that, as it appeared during the interviews, could have been too abstract to fully understand.

Therefore, for the design of a mock-up, firsthand expectations from users have not been considered.

Nevertheless, understanding the users' expectations has to come with understanding the data from which they require an analysis. In the following section, we will briefly assess the data that needs to be considered in our quality dashboard.

3.3 Data input

3.3.1 Data subject of analysis

The data that is to be analysed by the dashboard includes buildings, roads, Digital Terrain Model (DTM), Digital Surface Model (DSM), temperature, windspeed and sunshine data. It is summarized in Table 5 below).

Information	Data type	Countries in which the data is available via the GeoE3 platform
2D Buildings	Feature & Metadata	Finland, Norway, the Netherlands, Estonia, and Spain (National Cadastre)
3D Buildings	Feature	Parts of Finland, Norway, the Netherlands, Estonia, and Spain (National Cadastre)
2D Roads	Feature	Finland
3D Roads	Feature	Finland
DTM	Coverage	Finland, Norway, the Netherlands, Estonia, and Spain
DSM	Coverage	Finland, Norway, the Netherlands, and Spain
Temperature (yearly average from 30 years, from 1991 to 2020)	Coverage	Finland and Norway

Windspeed (yearly average from 30 years, from 1991 to 2020)	Coverage	Finland and Norway
Sunshine hours (yearly average from 30 years, from 1991 to 2020)	Coverage	Finland

Table 5: Summary of data involved in the GeoE3 project (geoe3platform@2022)

In particular, in the use case 1 related to solar energy, we are focusing on cadastral data useful to determine energy demands of buildings. That would be attribute data relative to buildings including building types, construction and renovation dates, and energy efficiency. The location of the building accessed through cadastral reference and / or its coordinates can provide information about its climatic zone and façade orientation for instance.

Finally, the shadows, sunshine and windspeed coverage at the location of the building are equally of interest, especially for solar panel installations in combination with 3D building models. It would also be profitable to be able to access energy performance certificates of the buildings in question. We have identified the following list of data that are significant in the use case 1:

- 3D buildings and attributes;
- 2D buildings and attributes;
- Solar energy potential data;
- Building footprints;
- Shadow index coverage;
- Number of sunshine hours at the nearest observation station;
- Average wind conditions;
- Normal air temperature at the nearest observation station;
- Monthly mean temperature based on climate scenarios;
- Digital Surface Model Normalized (DSMn);
- Weather Data;
- Building roof prints;
- Weather station's location;
- Lidar point cloud.

Most information about datasets is conveyed through their associated metadata, which therefore must be the subject of our attention as well.

3.3.2 Metadata

Metadata, which is commonly referred to as "data about data", according to Philips et al. (1999) (while referencing Anzlic, 1996; Kildow, 1996 and Anzlic, 1997), is an important part of data quality analysis. Its purpose is to describe the characteristics of the dataset. Some of the characteristics that are described by metadata may include the custodian of the data, the description of the data, the geographic extents of the data, the currency of the data, and the storage format.

In our case, metadata is provided from the European Data Portal catalogue service and the OGC API Records service.

3.3.3 Service information

Numerous supporting services, such as eTranslation, a European online translation system (cor.europa.eu@2022), automatic data quality validation services, and service quality monitoring are used to ensure fluent development and user experience of the GeoE3 platform and shall be part of the data quality dashboard's analysis when applicable.

Information about the quality of these service providers as well as the service provider allowing access to the data should also be included in the quality dashboard. This information will be provided through APIs by the company Spatineo, which is a partner in GeoE3.

In addition to these data, metadata and services information, other data that help with the assessment of data quality may be considered, including quality validation results or automated services, or quality audits based on maturity models.

3.3.4 Conclusion

The GeoE3 quality dashboard will display information about the quality of data and metadata involved in the GeoE3 platform.

The flow of data in the context of the GeoE3 project has been summarized in Figure 3-2 below, from the acquisition of data until the moment the data consumer finally is using the data. The data is accessed through the GeoE3 platform and goes through different evaluation processes, including automated quality validation processes, quality reviews and quality audits. These results, as well as the metadata analysis results and the service information, will then be fed to a database that will be accessed and visualized with a user-friendly dashboard to improve users' decision-making.

Now that we have a starting point for the requirements and an idea of the data that is involved in the use case 1, we can start investigating the actual building of a quality dashboard. The goal is to produce a first version of a dashboard that will be presented to users in order to collect feedback from users so that we can update, modify or specify the preliminary requirements.

The first step would be to choose a platform that meets our needs the best.

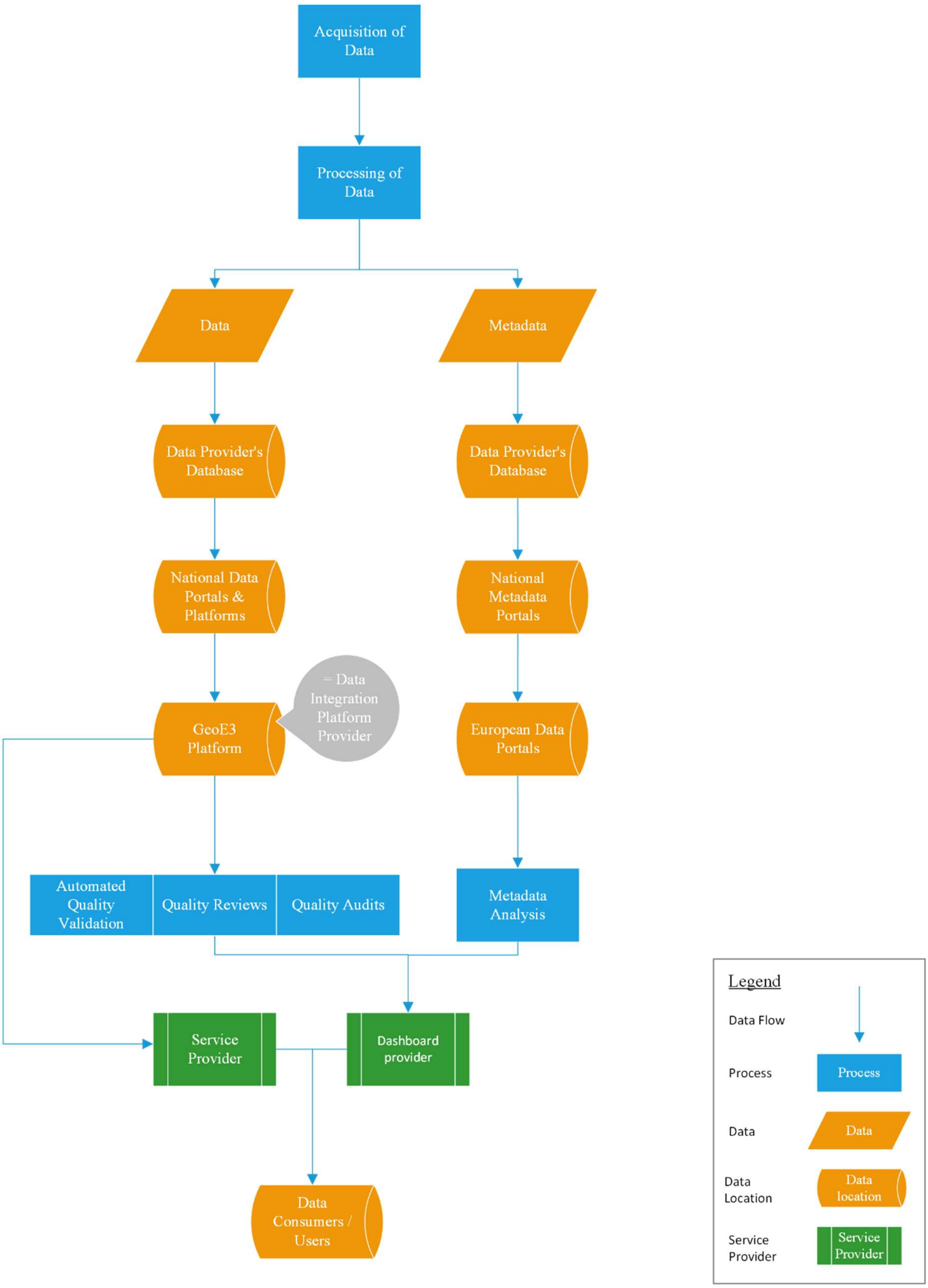


Figure 3-2: Flow of data diagram, complete from the acquisition of data to the data user

3.4 Choice of platform

The platform used to create the dashboard is now to be considered.

In the literature, the majority of dashboards have been made from scratch for their application. However, we will have here a preference for ready-made solutions. This presents several advantages. First, this allows us to focus more quicky on the design of a mock-up and the identification of necessary functionalities and quality indicators without delay. Moreover, we can take this opportunity to learn from software that specialize in the analysis and visualization of data.

In addition, the GeoE3 project, within which this task takes place, is to take place over the span of three years only and is dedicated to three use cases. Therefore, we can consider that the outcome of the GeoE3 project will be subject to modifications, improvements, and expansions.

In this section, we will compare a few licensed software in addition to free and open-source software.

3.4.1 Criteria used to compare platforms

In order to assess the strengths and weaknesses of candidate software, the comparison between software will be based on a list of criteria presented in Table 6 below.

In order to give each considered solution a score and rate them more easily, a weight between 0 and 1 has been given to each criterion based on the perceived importance of the criterion.

This rating makes it possible to go through a large list of candidates and select the most promising for further inquiry.

Criteria name	Description	Weight
Ready-made solution	The software should be ready to use. *This criterion is eliminatory.*	1
O/S	Operating System in which the software can be run. It should at least be available on the main current Operating Systems (Windows, MAC, Linux).	0.9
Data collection source	Describes whether the platform is able to collect data from classic data sources (CVS, Excel, SQL, P2P, API,…).	0.8
Ease of use for user of the dashboard	Perceived ease of use during our testing of the software from the point of view of the user of the dashboard.	0.8

Support	Characterization of the kind of support provided for the tool, if any. If the software is open source, an active community is valued. If the software is not open source, the company should offer quality support.	0.8
Type of license	Describes whether it is Open Source, Free or needs a subscription or a payment.	0.7
Perceived learning curve	Describe the perceived learning curve we have experienced.	0.6
Support for integration into web pages	Describes whether the platform pprovides support for embedding dashboards into third party web pages. This would be helpful for phase IV of the process.	0.6
Ease of use from the developer point of view	Perceived ease of use during our testing of the software from the point of view of the creator of the dashboard.	0.6
Price	If not free, describes the cost of the software. Preferred solutions are for free or open-source software or software whose license is already possessed by the organization.	0.5
Further development possibilities	Describes whether the software offers customizable possibilities for further development. Depends also on the coding languages further development can be done (at least Java, C++, or Python).	0.4
Geospatial features	Characterizes whether the software possesses geospatial features or features dedicated to the analysis of geospatial data in particular. The availability of geospatial features could be an asset and may be a requirement in further development. In addition, if the software has been developed with geospatial data in mind, it might be helpful in the future in ways that are not visible to our analysis now. However, our dashboard should focus on the *analysis* of geospatial data and not on geospatial data itself.	0.4
Software used in literature	Describes whether the software has been featured in previous literature. If it has been used in literature, it can help us determine its potential or its flaws as well.	0.2

Table 6: Criteria for the choice of platform for a quality dashboard

3.4.2 Comparison of different platforms

Platforms supporting the creation of dashboards are almost countless: some free, some paid, some Open Source. They also often have targets audience of different levels of knowledge and from different fields.

For the sake of brevity, this section will only display in Table 7 below the final score and remarks on each evaluated software.

The evaluation of each software was relatively brief due to the large number of options available in that department. The rating itself can be considered quite personal as it involves a personal assessment of the criteria and not a calculation based on facts or values. In addition, deeper or more complex functionalities could not possibly have been tested and therefore considered in our rating.

Software	Final weighted score	Open Source	Web-based or desktop	Comment
ArcGIS dashboards	0.81	X	Web-based	ArcGIS online license is necessary. BI features are similar to Power BI.
Microsoft Power BI	0.80	X	Desktop	This is Microsoft's dashboard designing software dedicated to Power BI. Its high score can be explained by the extensive support it provides to any user and the ability to link it to ArcGIS maps.
Tipboard	0.76	V	Web-based	Rates among the highest in all categories except for the support that, as in a lot of free software, might be lacking.
Smashing	0.74	V	Web-based	This platform cannot be used on Microsoft Edge. However, it offers a very satisfying design and is among the easiest to use.
Re: Dash	0.72	V	Web-based	Promising software that seems slightly harder to get used to and does not provide an extensive support. It also lacks slightly in BI tools.
Grafana	0.72	V	Web-based	A very big community of users might be able to provide support but help from the software itself is quite limited.

Freeboard	0.69	V	Web-based	We rated this software above average in most categories. However, tables were not suggested as data sources in the ready-made widgets and although it has a big community of users, tutorials are limited. Nonetheless, Plugins can be developed in Java, and it can display a limited Google map as a widget. The options for visualization seemed limited.
Mozaïk	0.62	V	Desktop	Promising software that however might lacked in BI tools.
query2report	0.60	V	Web-based	Free service that seemed to lack in the quantity of support and was not as easy to navigate.
Keen	0.52	X	Web-based	Paid web-based service. The service did not offer as much flexibility as others in our testing.
Metabase	0.44	V	Web-based	Metabase offered a solution that seemed hard to set up.
Metric fire	0.39	X	Web-based	Paid web-based service.
Metabase (licensed)	0.26	X	Web-based	
QGIS	0.00	V	Desktop	Plugins have been developed for the creation of dashboards within QGIS. However, it does not offer a ready-made solution nor BI tools. Could be of interest in phase IV or after.
Greppo	0.00	V	Desktop	Greppo is not a ready-made software. It could however be used in further development for building a dashboard in Python from scratch.
Stashboard	0.00	V	Desktop	Stashboard is not a ready-made software. It is a tool to create dashboards based on services and API information. Therefore, it does not really meet our requirements.

Table 7: Summary of the comparison of sixteen platforms

A short list of software that could have been part of our analysis – but were dropped because of the time-consuming nature of the testing of each platform – can be found in Appendix G.

This list allowed us to test further the five top-scored platforms. The results and final choice are discussed in the following section.

3.4.3 Results & Explanation of choice

3.4.3.1 QGIS

A solution to display information in the form of a dashboard is to use the "QGIS Dashboard" GeoPackage developed by Tim Sutton in 2020.

However, this does not meet one of our main requirements in that it does not offer a ready-made solution, allowing only the creation of labels that can be displayed on the QGIS interface (as can be seen in Figure 3-3 below). This criterion being eliminatory, this solution's rating came down to 0. Additionally, the package developed by Sutton does not incorporate Business Analysis tools.

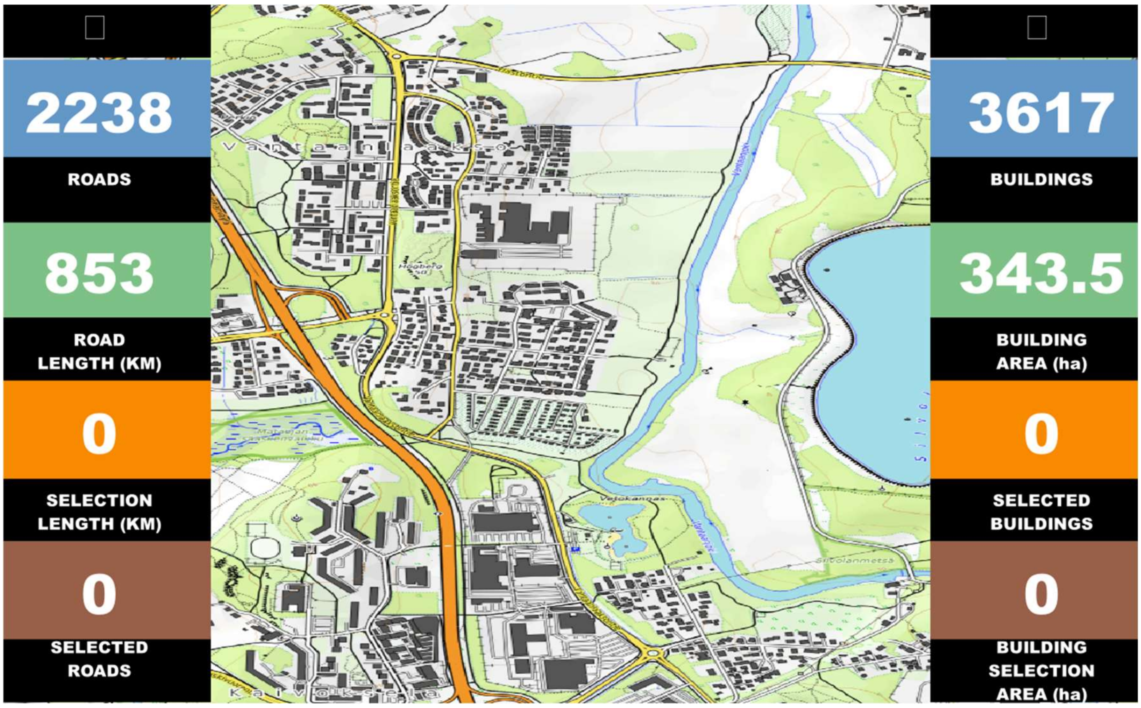


Figure 3-3: Example of a dashboard in QGIS using the QGIS GeoPackage developed by Tim Sutton (2020)

Having said that, this solution could be of interest in – or after – phase IV of the development of our quality dashboard. A QGIS plugin that allows the integration of datasets loaded from the GeoE3 platform in a QGIS environment has already been developed by Saskia Tuori within the Paikkatietoteknologiat group at the FGI. In that context, Tim Sutton's "QGIS Dashboard" GeoPackage could be a starting point to be used in future work to develop a GeoE3 quality data plugin to QGIS.

3.4.3.2 Open Source solutions

Tipboard, Smashing and Re:Dash are the Open Source solutions that rated the highest in the previously described rating system. All three can be found on their own GitHub webpages.

Support and tutorials are not as extensive as in paid software and the learning curve was slower. In addition, they required the use of an already set up and ready database to load data into a dashboard. At the time of this platform comparison, we mostly worked with spreadsheets due to the low number of entries required to create a mock-up and therefore did not have a personal database to try on these solutions.

3.4.3.3 Microsoft Power BI

Microsoft Power BI allowed us to have quick results to display to users and does not require an expertise in web design or programming, allowing us to focus on the elements that need to be displayed themselves. It is a powerful tool that, if the need arises in the future, allows us to link the dashboard to another powerful software prominent in geospatial data, ArcGIS online by ESRI.

Microsoft Power BI presents however a few drawbacks that have been taken into account. Firstly, it is not an Open Source software. That in itself limits possibilities of further eventual development. Secondly, like most of the other software we have tested, it is a business analysis software that does not specialize in geospatial data.

3.4.3.4 Final choice: Microsoft Power BI

As a result of our testing and comparison, we have chosen not to focus on a software specialized in geospatial data. Indeed, although we will be manipulating geospatial data in our analysis, the data that the dashboard needs to display does not have geospatial features in most cases.

Open Source software would be a fine choice and an excellent place to start if we consider future more complex developments – with, for instance, the direct integration of quality analysis functions into the quality dashboard. We can consider that the creation of this dashboard arrives relatively soon in the development of the GeoE3 project, seeing as the GeoE3 platform is not yet widely used by the population nor displays all the data such a platform should ultimately aim to display. Additionally, it is at the moment focusing only on three use cases (although they do extend over a large range of applications). Therefore, choosing an Open Source software would have presented more possible freedom for future developments.

However, in the context of our project, they at the moment do not offer as many advantages as our top paid software do from a guidance point of view. Indeed, these paid software benefit from having an extensive support available for their users and present many alternative visualizations of data analysis results. Given that we started from scratch with few ideas of smart visualizations or layout, using an easier platform to navigate allowed for more trials and error with less lost time.

Finally, between ESRI's ArcGIS Dashboards and Microsoft Power BI, we have chosen to focus on the latter, which edge over ArcGIS Dashboards mostly resides in the fact that the

National Land Survey of Finland already had acquired the license, and we would therefore jump into the creation of a mock-up – and its presentation to stakeholders for feedback – faster. Power BI also allowed us to work fluently in tandem with data sourced from spreadsheet formats via Excel.

Additionally, Power BI possesses the option of integrating a powerful spatial analysis tool via ArcGIS for Power BI, therefore allowing us to keep the door open to access and use parts of ArcGIS's functionalities if the need arises in the future.

We have now established some of the expectations of a quality dashboard, the data involved in the use case and the platform on which we are building the quality dashboard.

Thereafter, we will present the metrics that should be used to assess the quality of data and the way we have classified and organized them, before presenting the first mock-up version of our quality dashboard.

3.5 Building a first version of a Quality Dashboard: the mock-up

3.5.1 Data quality metrics

In section 2.5 Data quality of chapter 2, we have listed a variety of quantifiers suggested by authors to measure data quality. Here, we will classify them in multi-level categories.

The first identified labels, inspired by Rula & Zaveri (2014)'s vocabulary, will be called *dimensions*. Sub-levels are considered quality *elements* – this term is inspired by the ISO norm' vocabulary. These elements are quantified by *measurements* that are characterized by quality *components*, as defined by the ISO 19157:2021 and ISO 19115:2011 international standards. One of those *components* is defined as a measure parameter by the ISO 19115:2011 international standard; here, it referred to as a *metric.*

This classification is summarized in Figure 3-4: Summary of data quality composition and vocabulary below.

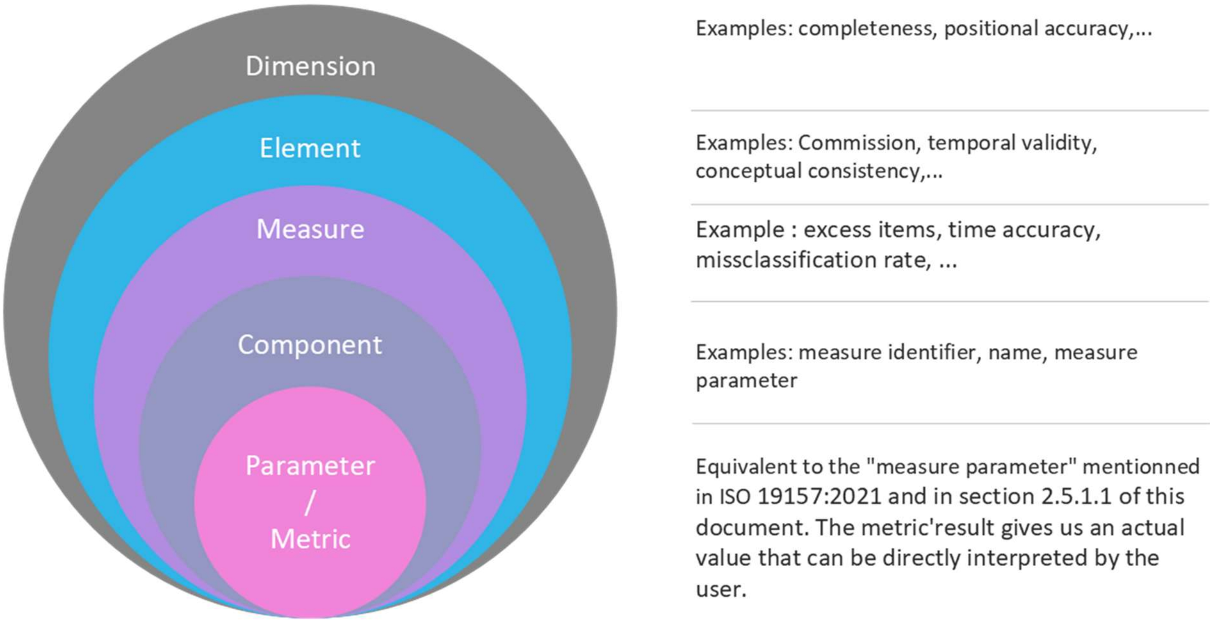


Figure 3-4: Summary of data quality composition and vocabulary

Data quality is a multi-dimensional concept with each dimension focused on one aspect of quality. The dimensions can be measured with one or more quality metrics that report a series of score or values. Those can then be compared to predetermined – and often use-case specific – thresholds for quality evaluation.

The dimensions we have identified are the following:

- Description clarity: This dimension focuses on measuring the quality of the description available through the metadata. This dimension would be evaluated through the consideration of quality elements such as key-words and the actual textual description.

- Reliability: This dimension provides information to the user regarding the level of trust they can put into the dataset considered. It takes into consideration two quality elements. The first is the credibility; measurements considered for this will be the source reliability, the authenticity of data and the actual credibility, which is defined by the ISO 25000 standard as a measure that establishes the degree to which data has attributes that are regarded as true and believable by users in a specific context of use.

- Relevance: This dimension takes five elements into account and focuses on providing the user with information relative to how much the dataset analysed is pertinent to its research. This dimension would require the user to add input into either the dashboard or the GeoE3 platform so that the metadata attached to the dataset in question can be compared to the user's specific requirements.

  For instance, a user interested in a given area will specify it to the platform so that this area can be compared to the surface the dataset covers. Similarly, the user may request a specific file format or type of geometry.

- Portability: This dimension is designed to help the user understand whether a dataset can easily be used internationally, without the barrier of language or the mistakes that can be induced by the use of units that are not in the International System of Units (SI).

- Information security: This dimension is two-fold and focuses on the assessment of the traceability and confidentiality elements. The first element should inform the user if the dataset cannot be traced all the way to its original source. The second element should give information related to the accessibility of datasets by assessing whether it is Open Source or protected in some way against public use.

- Service quality: This is the only dimension that evaluates the quality of a service in the context of GeoE3. The robustness of the service will be assessed through the measure of its availability. For this purpose, four metrics will be considered: the availability of the service over a pre-determined period of time (previous 7 days, 30 days, or previous year), average and longest service interruption, and response time.

  However, the needs of a user relative to a service will vary a lot depending on the use case or the user's field of work, and a service displaying quality scores on the low end might be still perfectly useable in certain circumstances. For instance, a high response time is not necessary

  Therefore, in accordance with Spatineo, the Finnish company providing the analysis on these services for GeoE3, we have decided that those metrics will not be used to generate a score.

These next dimensions and their definitions are derived from the ISO 19157:2021 standard.

- Completeness: The completeness dimensions should identify the degree to which a given dataset possesses all features, attributes, and relationships it is supposed to. For that reason, the omission and commission quality elements are measured through the identification of missing, excess, or duplicate items in the collection.

- Thematic quality: This dimension measures the correctness of classifications of qualitative attributes and the accuracy of quantitative attributes in a dataset through three data quality elements, the classification correctness, the non-quantitative attribute correctness, and the quantitative attribute accuracy. Those will ultimately be assessed through counts of incorrectly classified items and misclassification rates.

- Logical consistency: This dimension measures the degree of adherence to logical rules of data structure through the following quality element: conceptual consistency, domain consistency, format consistency and topological consistency.

- Temporal quality: This dimension is particularly important for a lot of users. It assesses the accuracy of a time measurement, the temporal consistency and, most importantly, the temporal validity which is evaluated through the maintenance frequency and the "currentness" – or "up-to-dateness" – of the dataset. Those two measures can heavily affect the way a user will consider data as useable and should be emphasized in our dashboard.

- Positional accuracy: This dimension relates to measurement accuracy and, depending on the reference system used by the dataset, it can be measured through absolute accuracy, relative accuracy, or gridded data positional accuracy.

Based on their interoperability map, we have identified three other dimensions from the work of Jakobsson & al. (2022): semantic interoperability, legal and organizational interoperability, and technical interoperability. Those dimensions are assessed thanks to the interoperability map produced for the GeoE3 project and referenced earlier in this paper (cf. Table 3: Interoperability Map on building data, extracted form Jakobsson et al. (2022), Table 5, p13).

We now have a list of quality dimensions and the quality elements, measures, and metrics that we would like to evaluate in order to assess the data quality of a dataset. A detailed classification is available in section 3.5.3 below.

In order to make our analysis more intelligible for users, we have identified three categories in which to organize these dimensions.

3.5.2 Three points of view

As was emphasized by Nuradiansyah & Budi (2015), what is currently needed for the non-technical user to understand the data into the form of information, decision, and action is the translation of data into a form of visualization.

In order to adapt to every type of user in spite of their different expectations and level of data quality expertise, we have adopted a navigational structure that categorizes the presentation of information in three windows.

The first point of view identified is the Use Case view, or Useability view. This approach should focus on answering questions related to how well this dataset or service is suited to the use case and is mostly aimed to satisfy the needs of a novice user not familiar with data quality or geospatial data. This point of view will gather the reliability, description clarity and relevance dimensions (cf. Figure 3-5, page 47).

The second point of view is the Data & Service view and is mostly aimed at the data provider or data integrator. In addition to the service quality dimension, this view gathers the data quality dimensions that we have identified through the ISO norms: completeness, thematic quality, logical consistency, temporal quality, and positional accuracy (cf. Figure 3-6, page 47).

Finally, the third point of view focuses on Interoperability and features data from the interoperability map designed using an interoperability maturity matric approach. Depending on the selected dataset or datasets, this view would basically show the status of each dataset, service, or GeoE3 partner country. In practice, the interoperability view gathers the technical interoperability, semantic interoperability, legal and organization interoperability, portability, and information security dimensions (cf. Figure 3-7, page 49).

3.5.3 Summary diagrams

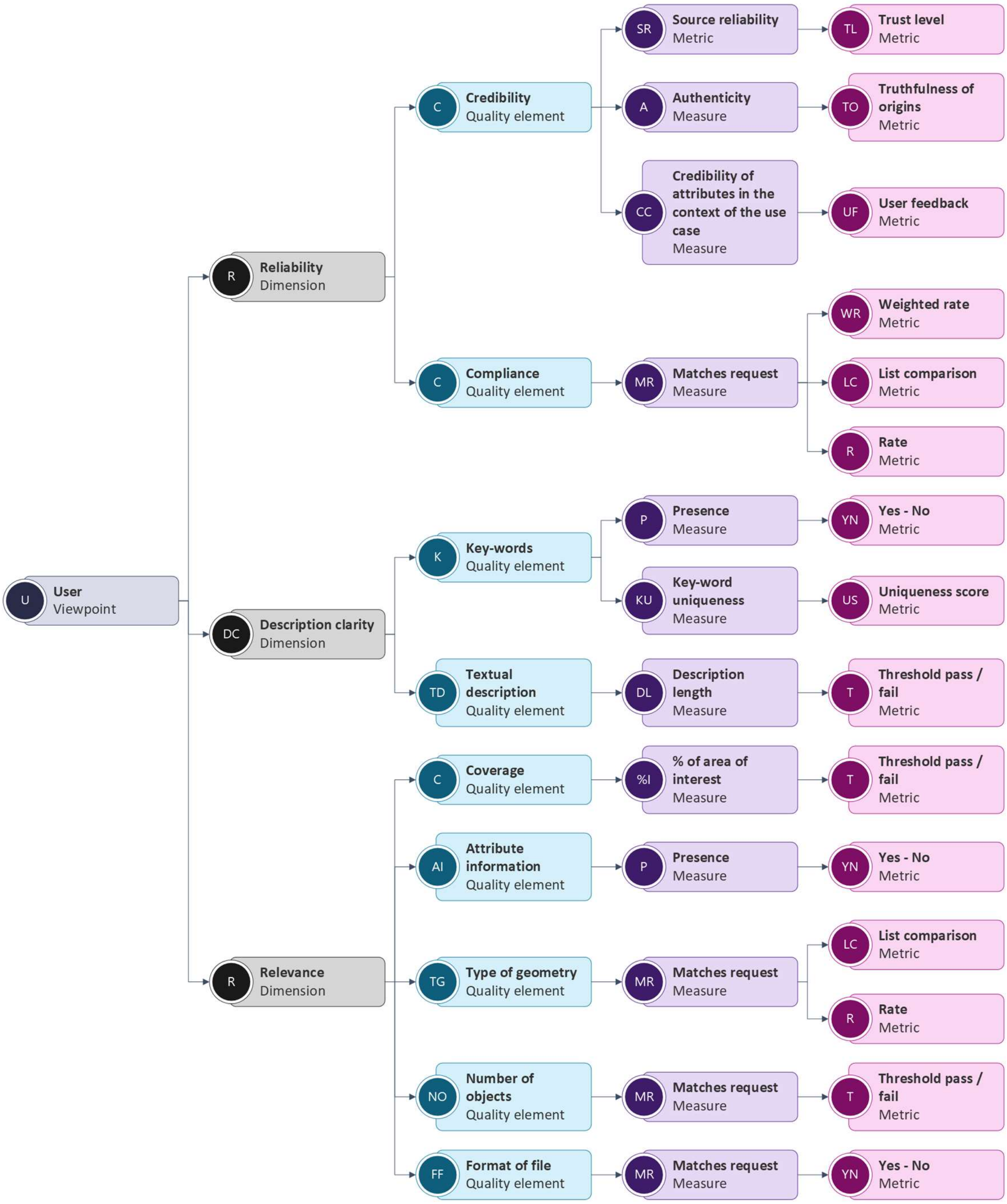
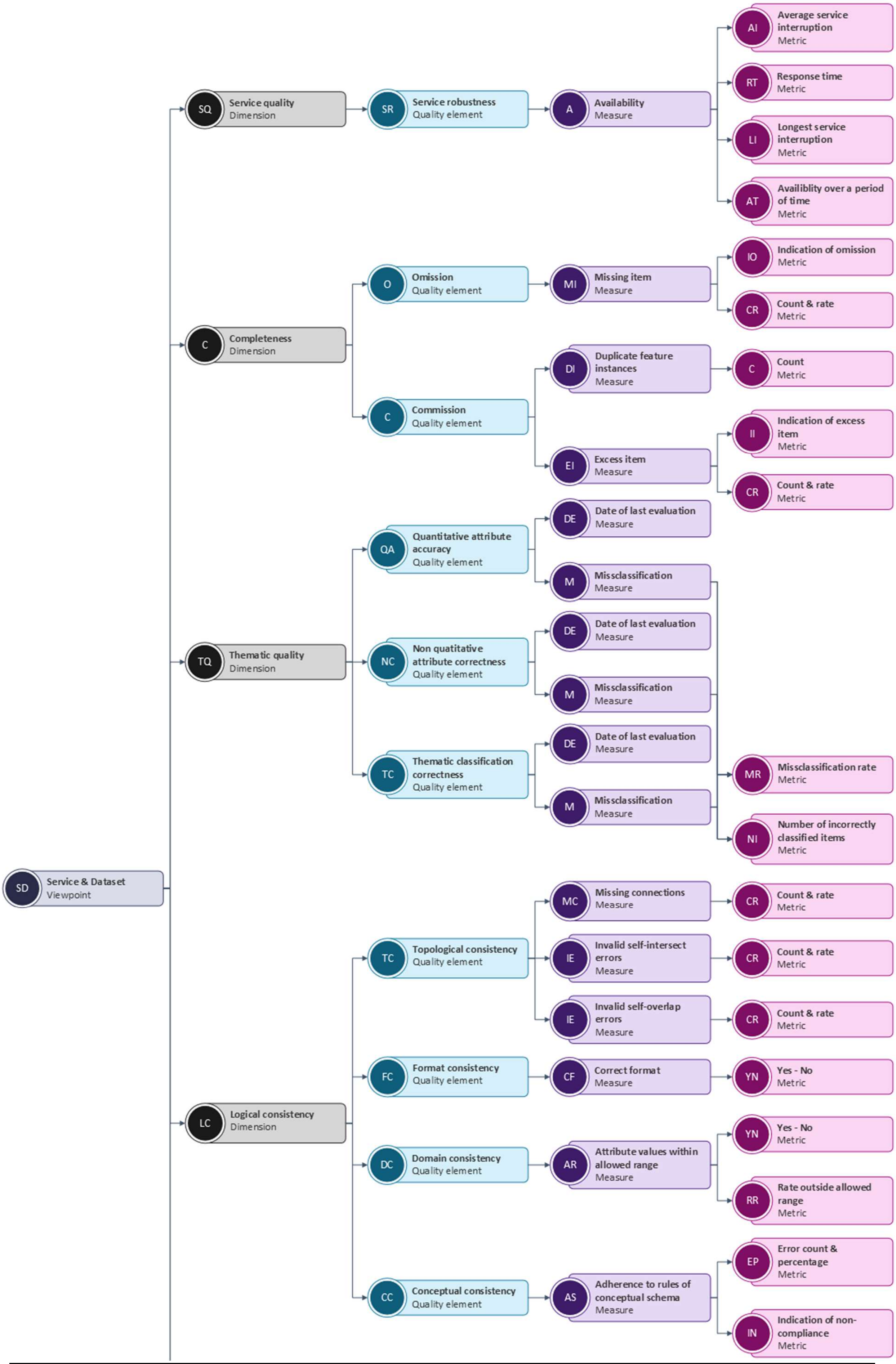


Figure 3-5: Summary diagram of classification of quality dimensions, elements, measures, and metrics for the quality dashboard

Figure 3-6 (below): Summary diagram of classification of quality dimensions, elements, measures, and metrics for the quality dashboard, Service & Dataset point of view (part 1)


SD
Service & Dataset
Viewpoint
SQ
Service quality
Dimension
SR
Service robustness
Quality element
A
Availability
Measure
AI
Average service interruption
Metric
RT
Response time
Metric
LI
Longest service interruption
Metric
AT
Availiblity over a period of time
Metric
C
Completeness
Dimension
O
Omission
Quality element
MI
Missing item
Measure
IO
Indication of omission
Metric
CR
Count & rate
Metric
C
Commission
Quality element
DI
Duplicate feature instances
Measure
C
Count
Metric
EI
Excess item
Measure
II
Indication of excess item
Metric
CR
Count & rate
Metric
TQ
Thematic quality
Dimension
QA
Quantitative attribute accuracy
Quality element
DE
Date of last evaluation
Measure
M
Missclassification
Measure
NC
Non quatitative attribute correctness
Quality element
DE
Date of last evaluation
Measure
M
Missclassification
Measure
TC
Thematic classification correctness
Quality element
DE
Date of last evaluation
Measure
M
Missclassification
Measure
MR
Missclassification rate
Metric
NI
Number of incorrectly classified items
Metric
LC
Logical consistency
Dimension
TC
Topological consistency
Quality element
MC
Missing connections
Measure
CR
Count & rate
Metric
IE
Invalid self-intersect errors
Measure
CR
Count & rate
Metric
IE
Invalid self-overlap errors
Measure
CR
Count & rate
Metric
FC
Format consistency
Quality element
CF
Correct format
Measure
YN
Yes - No
Metric
DC
Domain consistency
Quality element
AR
Attribute values within allowed range
Measure
YN
Yes - No
Metric
RR
Rate outside allowed range
Metric
CC
Conceptual consistency
Quality element
AS
Adherence to rules of conceptual schema
Measure
EP
Error count & percentage
Metric
IN
Indication of non-compliance
Metric

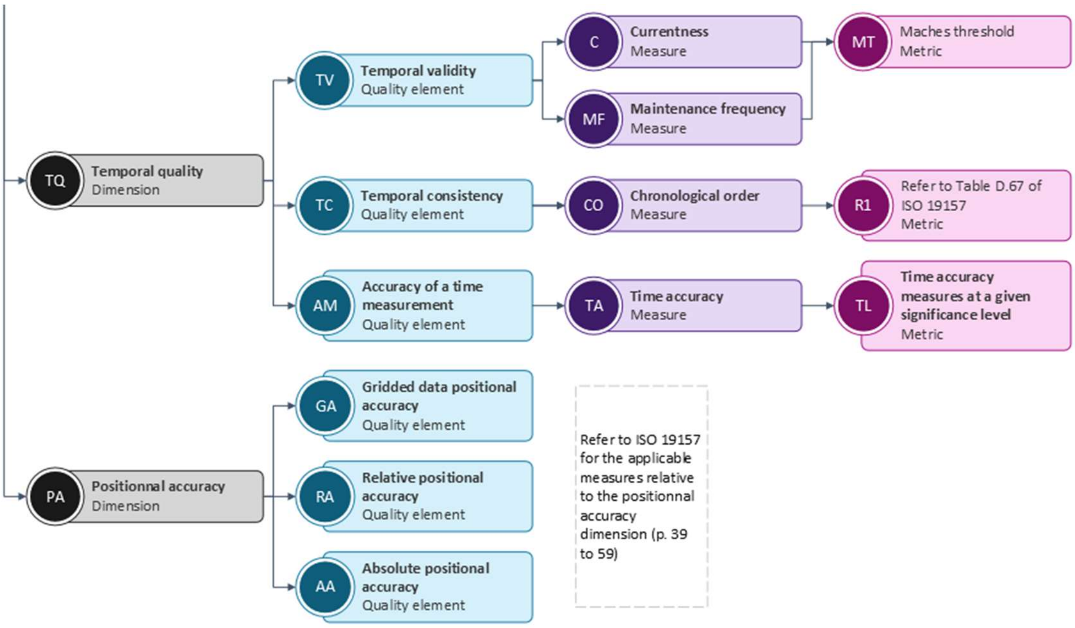
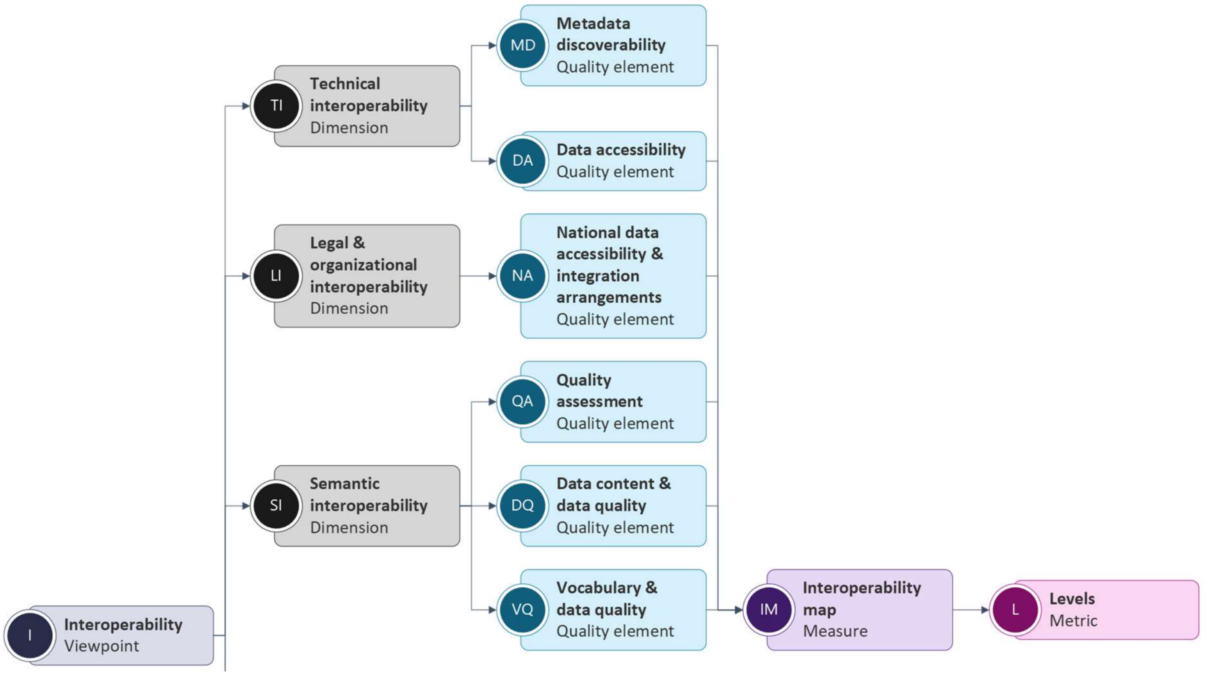


Figure 3-7 : Summary diagram of classification of quality dimensions, elements, measures, and metrics for the quality dashboard, Service & Dataset point of view (part 2)

On these diagrams, the categorization and hierarchy of data quality dimensions (in grey), elements (in blue), measures (in purple) and metrics (in pink) are highlighted.

Figure 3-8: (below) Summary diagram of classification of quality dimensions, elements, measures, and metrics for the quality dashboard, Interoperability point of view (part 1)



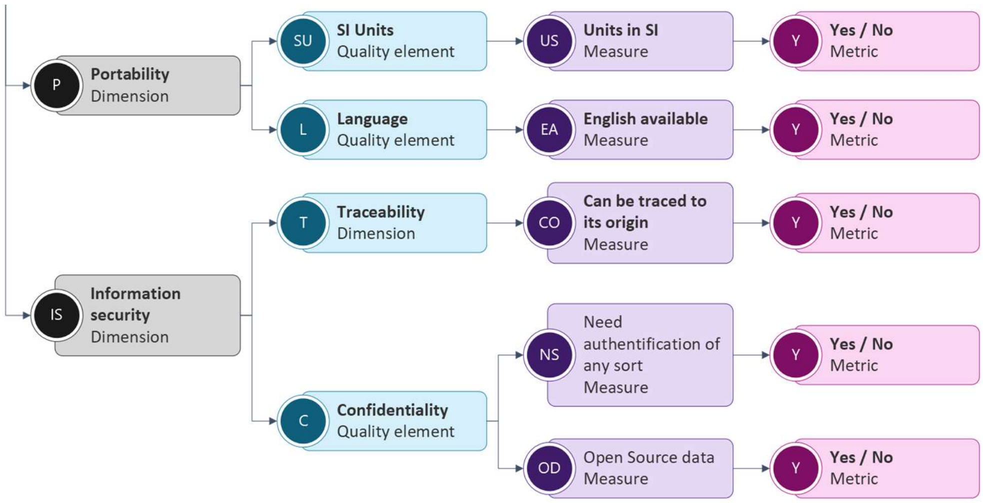


Figure 3-9 : Summary diagram of classification of quality dimensions, elements, measures, and metrics for the quality dashboard, Interoperability point of view (part 2)

These diagrams are a good reference point to have all throughout our analysis and made the creation of a mock-up easier.

3.5.4 Mock-up: a first version of a quality dashboard

This analysis led us to a mock-up dashboard build on Microsoft Power BI that was aimed to be presented to users during interviews in order to collect feedback and assess more precisely the users' needs and expectations.

Two of the three points of view can be seen in Figure 3-10 and Figure 3-11 below. In addition, the planned layout of the interoperability point of view can be seen in Appendix H.

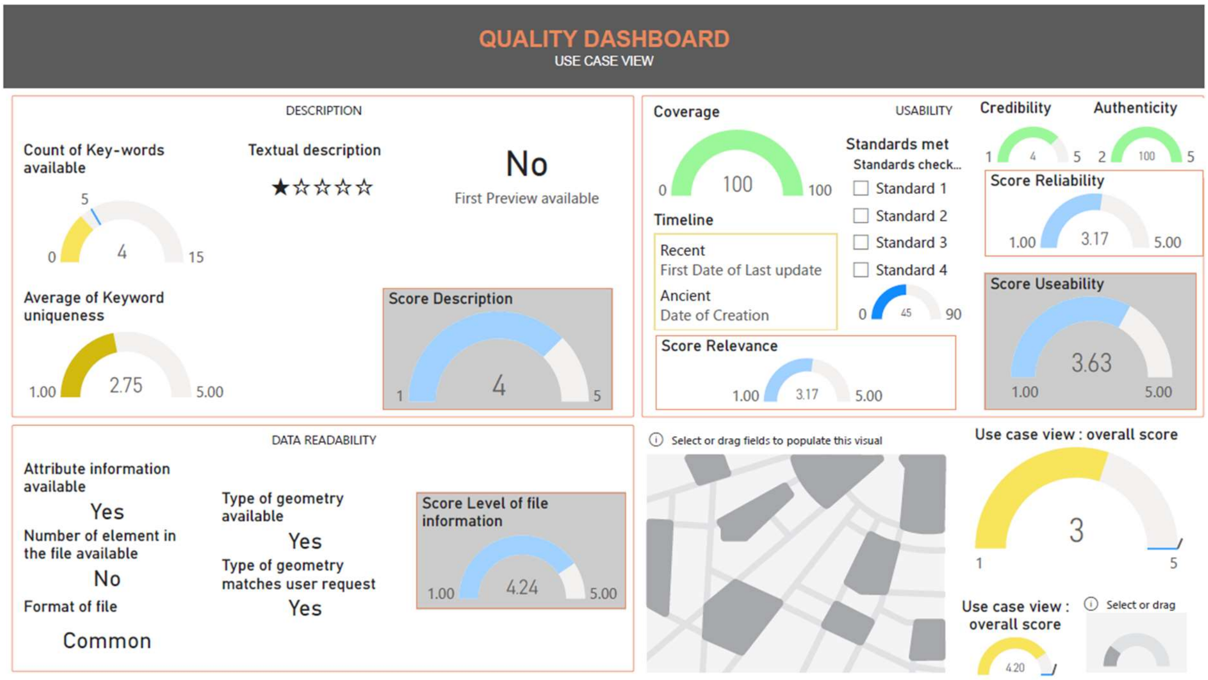


Figure 3-10: First quality dashboard version: the use case point of view

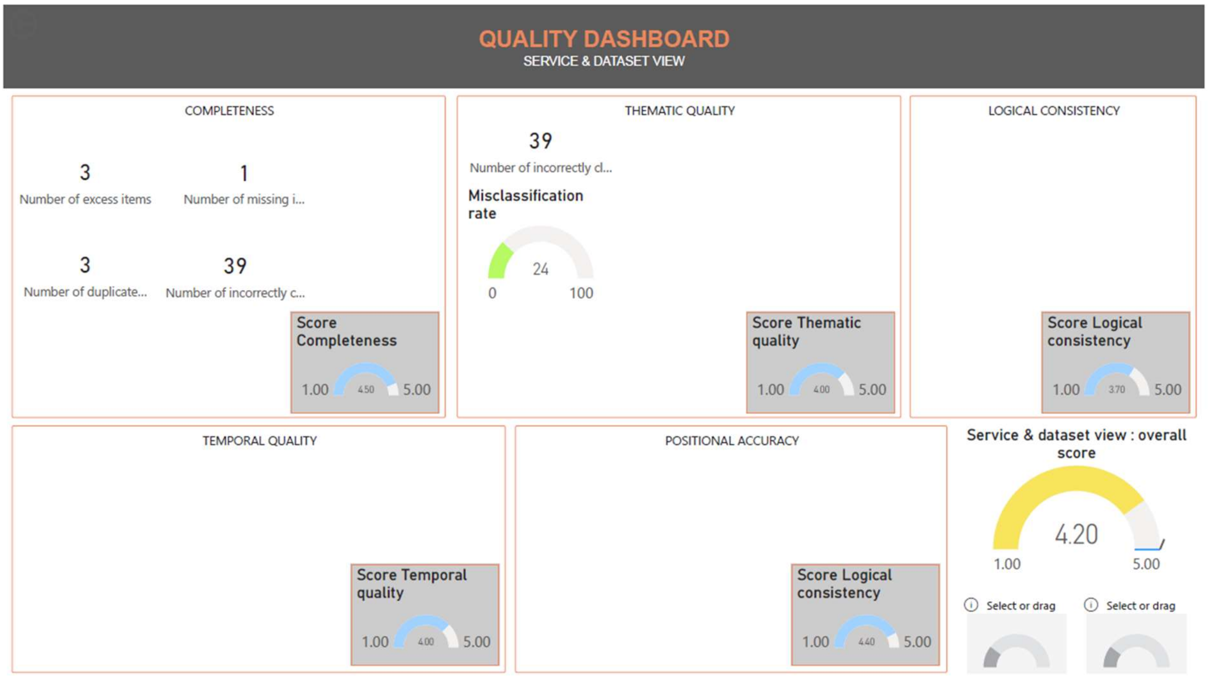


Figure 3-11: First quality dashboard version: the service and dataset point of view

As can be seen in the above figures, the first version of a dashboard we conceived was divided in three windows for each point of view, allowing the user to switch between the windows as needed. The layout of each point of view was divided by dimension – encompassed by an orange box –, within which appear metrics whose evaluation is represented by business intelligence widgets that allow the results of a quality analysis to be visualized in a very user-friendly way. Each dimension was assigned an overall score, represented in a grey box in the above figures.

Following the solution presented by Rula & Zaveri (2014) discussed in section 2.6.1, we aimed to assign scores to each metric. This mock-up has been developed with a mock database in a spreadsheet format that plays the part of collecting the results of previously identified quality metrics through quality analysis processes. In this database, each metric is given a factitious value that could otherwise believably have resulted from an actual quality analysis process.

For some metric, we obtain Boolean (pass / fail or 0 /1) scores using different techniques:

- Metrics that are measured by yes or no – for instance, the traceability dimension and the presence or absence of key-words – are already Boolean;
- Some values can be compared against a pre-determined threshold. For instance, the coverage percentage, count or rate of missed items or the number of characters in the dataset description can be evaluated this way.
  Although this threshold is for now arbitrarily set, in the future, thresholds could either be manipulated by the user to fit their personal requirements or set based on default user profiles.

For other metrics, we have decided to grade them on a 1 to 5 scale, 1 being the lowest and 5 the highest. This decision is based on the fact that a homogeneous rating system should be easier to understand and more comfortable to manoeuvre through for the users.

Such a score is to be calculated based on the Equation 1 presented by Rula & Zaveri (2014) in section 2.6.1 (page 25) but slightly modified to fit our chosen scale.

$$DQ_{score} = \left(1 - \frac{V}{T}\right) * 5$$

Equation 3: Data quality score for one metric, slightly modified from Rula & Zaveri (2014)

Inspired again by the same authors, to obtain an overall data quality score for the dataset, a second equation is used, based on Equation 2 (cf. page 25). The individual data quality scores are weighted relatively to our perceived importance of the metric.

$$DQ_{weightedsco} = \sum_{i=1}^{n} \frac{DQ_{score} * w_i * 5}{W}$$

Equation 4 : Data quality weighted score for an agglomerate of several metrics, a dimension, or a point of view, slightly modified from Rula & Zaveri (2014)

To cater to different user to the best of our abilities, it was decided to display scores not only for every metric, but for each dimension, and subsequently each point of view. That way, any user can choose for themselves how much they want to learn from the quality analysis process.

It was this version of dashboard that was presented to a first wave of users during interviews. Those will be addressed thenceforth.

3.6 User interviews

The assessment of user requirement in this project was done in association with Lena Hallin-Pihlatie, who is in charge of collecting feedback related to the GeoE3 project.

Given that the quality dashboard we aim to produce is a tool dedicated to help users in their experience of the GeoE3 platform, requiring feedback from potential users was a central point of our methodology.

The medium chosen to collect feedback was online interviews. This presents several advantages. First, we could individualize the questions depending on the background of the interviewee, therefore focusing on different aspects in each interview. This also allows us to evaluate more easily the feedback of users or ask more for precision if it an answer was unclear – likewise, the user was able to ask questions from us. Finally, users were not limited by their location to be part of this process and we had access to different perspectives.

3.6.1 Profile of interviewees

The first step of this process was to identify users who deal with data related to solar energy potential and energy efficiency of buildings to interview. The users could be researchers, scientists, representatives of environmental government agencies or electrical power producing companies and could come from any of the five partner countries of GeoE3.

Most importantly, the profiles of the people we were looking to interview had to fit in the context on GeoE3 in that they had to be potential users of the platform. Therefore, they had to deal with building or solar energy related data. We were not specifically looking for experts however and welcomed anyone that shown an interest in being presented the project and interviewed.

3.6.2 Program of the interview

The focus of these interviews was on checking whether the offered services and tools met users' needs and identifying possible improvements. Lena Hallin-Pihlatie's questions were centered on the GeoE3 platform and APIs while ours were on the quality dashboard itself.

Once the background of the interview has been established, the interviews were designed in two parts, with Lena Hallin-Pihlatie first focusing on the GeoE3 platform and APIs demonstration, explaining at the same time the context in which the dashboard is to function. A demonstration of the mock-up quality dashboard followed an introduction to the concept and goal of this task. This was followed by a free discussion with the interviewed user when they could give feedback as well as ask more targeted questions.

Questions asked regarding the quality dashboard are summarized in the following table.

Theme	Examples of questions
Usefulness in relation to the user's work	How useful and relevant for solar energy estimation in your company would a quality dashboard be in your opinion?
Usefulness in relation to the user's work	How often would you need our service / to check on the quality of data?
Metrics	What information would you like to see on the dashboard?
Visualization / Layout	What information should be prioritized in your opinion?
Metrics	What problems with the data do you usually encounter in your work? How would you like it to be analysed?
Development	On which scales should the analysis be able focus on (one building, one neighbourhood, one city, …)?
Metrics	Who would be the users of the dashboard in your company? Do they have prior knowledge in geospatial data? What is their level of competence?
Visualization / Layout	Would an interactive map associated with the dashboard, or preview of the data, be of use to the users?
Platform	Is there a need for a mobile app; is the web user interface sufficient?

Usefulness in relation to the user's work	How do you perform this task currently (perhaps a similar service)? What would make our dashboard better those other services? Under what conditions would you consider changing from those other services to ours?
Visualization / Layout	What functionalities could be further improved?
Content / Data	What contents could be further improved?
Usefulness in relation to the user's work	Would you work from a computer or mobile?

Figure 3-12: List of recurring questions asked in the interviews

Our modifications following the interviewed are demonstrated in chapter 4 Results, section 4.1 Interview results.

3.7 Conclusion on the methodology

In this chapter, we have established the methodology adopted in order to create an efficient, practical, and user-friendly quality dashboard for the GeoE3 platform in the context of the first use case of the project, Solar energy potential and energy efficiency of buildings.

As a result of a series of dashboard software solutions, we have created a first version of a quality dashboard on Microsoft Power BI, a software that possess the advantages of guiding us through business intelligence visualization. With the help of extensive diagrams, we have also established the hierarchy of data quality indicators that should be displayed on a quality dashboard.

In the next chapter, we will go through the results of interviews that we have performed, the issues raised by the users and the changes in the mock-up that their input led to.

The final version of the quality dashboard will then be presented before being compared to a metadata dashboard created by Data Europa and released during the summer 2022.

4 RESULTS & DISCUSSION

4.1 Interview results

4.1.1 Challenges

The process of finding users from partner countries was quite challenging. Indeed, we had to change strategy in the beginning of this project, when trying to identify user requirements to take into account to create a mock-up and refocus on trying to produce a dashboard that could suit needs of the users without having their input at the start.

GeoE3 partners provided contact information to companies and organizations, which are relevant stakeholders for the use case Solar energy potential and energy efficiency of buildings. More than a dozen requests and follow-up reminders were sent out. In our requests, it was stressed that the interviews will be confidential and that no personal information or business secrets would be included in the report. As it turns out, only a small part of contacted users answered our solicitations and requests for interviews.

After reaching out to contacts within the national cadastre and mapping organizations of partnered countries, we have managed to schedule three interviews at this time. As a result, we were able to hold interviews during a period from the months of June to August 2022.

4.1.2 Summary of the interviews

The duration of the interviews was of approximately two hours and were held through Teams and Zoom. We interviewed people from Norway and Spain working for small companies in the solar energy industry. They either had experience using GIS, APIs or both.

The first two interviews were held close together in June 2022 and the interview subjects were presented with the mock-up version of the quality dashboard discussed earlier in section 3.5.4 (cf. Figure 3-10 and Figure 3-11). The last person interviewed in August 2022 was presented with a version close to the final version of the quality dashboard (cf.

The first person interviewed was working for a small company, with under ten employees, installing bifacial photovoltaic systems for flat roofs and buildings and integrated systems for sloping roofs, walls, and facades. They tackled the roles of product manager, project manager, and sales representative among others. Although they did not have any GIS experience, they were familiar with APIs.

The second person interviewed was working for a company of similar size to the first one. His activities included photovoltaic installation calculations for buildings based on relevant information. They were also a civil servant of one of the national cadastres and were familiar with geospatial data, GIS, and APIs.

Finally, the third person interviewed worked in a company employing about thirty people focused on solar, thermal and wind projects. They were familiar with geospatial data and had GIS experience.

In summary, although the number of interviews we have managed to arrange was fairly low – and in any case lower than what we would have hoped at the start of this process –, the profile of the persons interviewed fit the use case perfectly and were, at the same time, different enough to offer diversified input. Moreover, although none were very knowledgeable in the data quality field, they all presented different level of expertise in geospatial data.

Acquiring input directly from potential users is very valuable in this stage of the creation of a product when it is early enough to mould it according to their suggestions. Interviewee are put in the position of being able to correct the direction of the project if necessary or encourage it. Moreover, inviting users to feel active in the process inspire more engagement towards the project, which could even lead to a possible longer collaboration.

4.1.3 Feedback and issues raised by users' assessment of the dashboard

The feedback from the interviews was overall very positive. The people interviewed were enthusiastic, understood the concept of a quality dashboard and were very interested in the value that new features could bring to their work in association with the GeoE3 platform. Quoting interview subject number one:"I think this is a very useful tool" and "wish[ed] everyone could have this".

We received suggestions to upgrade indicators to include a definition of its quality, the indicator source – the ISO standard for instance –, the score computation method, and finally the original values used for the score computation. This sort of input verbalizes the fact that users often require more information than what we assumed in the first place. In our case, we can take away from this that the quality indicators appear abstract and unintelligible to users, and that we should aim to popularize not only the visualization of the score of indicators, but the explanations relatively to the quality indicators themselves. Moreover, although the people interviewed were curious about the metrics we used and where they came from, it was not something they would have thought to require by themselves. On that account, we can draw the conclusion that the quality dashboard, even in a version that appeared chaotic and cluttered to some, succeeded in raising awareness and curiosity on data quality. This is a positive point for two reasons: first, this means that a quality dashboard might attract users to the GeoE3 platform with little knowledge in data quality, and secondly it might raise awareness about the importance of selecting the right dataset.

To quote another interview: "if you can't evaluate what you're looking at in 5 seconds, there is too much information." This point came up in the first two interviews we held, during which the mock-up version of the dashboard was presented. This version displayed most of the quality indicator categorized by quality dimension (as identified in chapter 3) in three 16:9 format windows, one for each point of view (cf.

Figure 4-1, Figure 4-2, Figure 4-3 and Figure 4-4 pages 57 and 58). Therefore, we should be particularly careful not to put too much quality information on one page.

4.1.4 Interview perspectives

Two of the users have shown interest in the project and would be willing to meet again to collaborate further. Additionally, two work-arounds to the lack of users to interviews were found. In order to bypass the language barrier and to make the questions and answer process more adaptable, we decided to implement two 10 to 15 minutes questionnaires in relation to the GeoE3 platform and the quality dashboard.

The feedback from interviews led to rethink the whole architecture of the quality dashboard. It can be shortened into two main points: the dashboard should appear as simple as possible while offering the option to go deeper if the user shows interest, and the quality dimension, elements, metrics, and score should be detailed. The next section exposes the changes we implemented and the current layout.

4.2 Current version of the quality dashboard

In the final version of the dashboard, we took into consideration all aspects of feedback collected from the people interviewed. We have proceeded with a multi-level navigational interface that, instead of displaying all quality indicators relative to a point of view in one window, offered several small windows to navigate through.

Figure 4-1 shows the first pane that would appear to a user. The dashboard displays the three use cases view as well as their weighted data quality score calculated as described in section 3.5.4 (page 50).

From this “home” pane, we can access the dimension scores of each view point on separate windows (Figure 4-2, Figure 4-3 and Figure 4-4). On and on, we can go through the different levels of the data quality indicator hierarchy (as described in the diagrams in section 3.5.3 page 47).

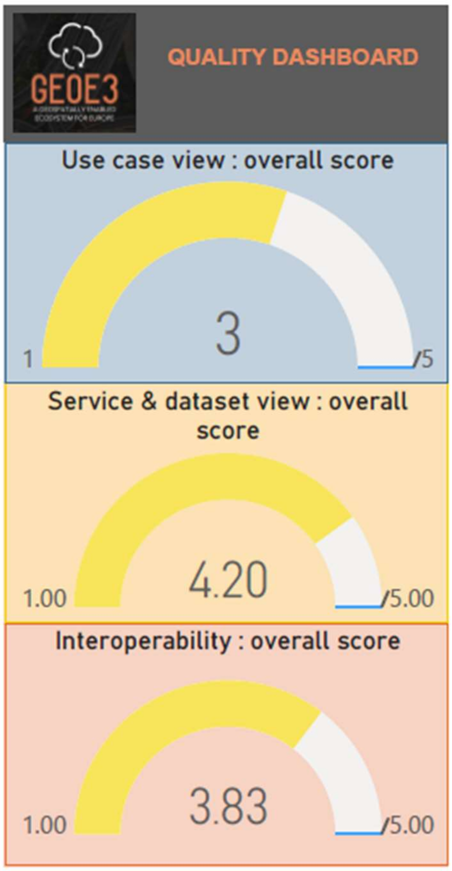


Figure 4-1:Current version of the quality dashboard: ”home” screen

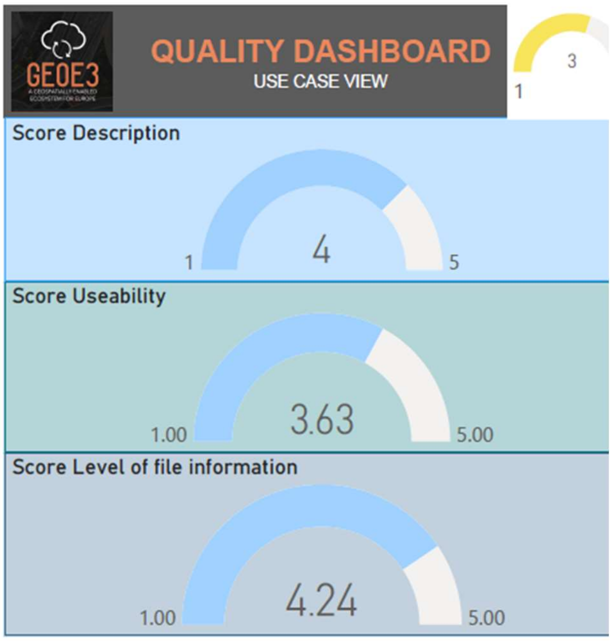


Figure 4-2: Current version of the quality dashboard: user point of view

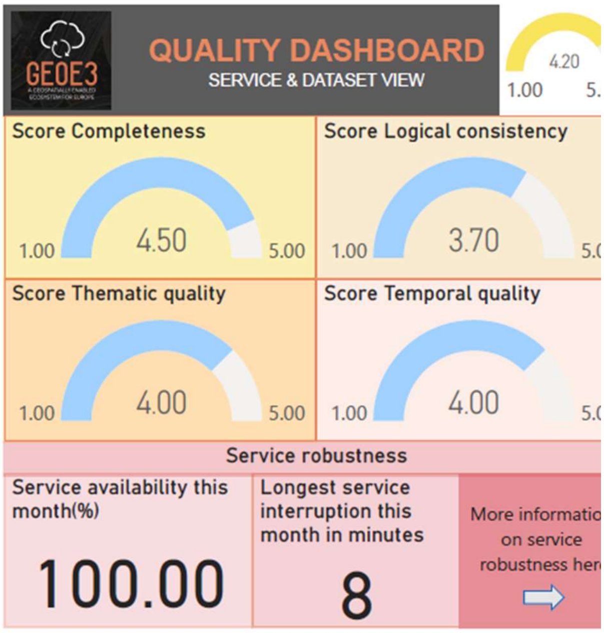


Figure 4-3: Current version of the quality dashboard: service & dataset point of view

Figure 4-4: Current version of the quality dashboard: Interoperability point of view (interoperability map is to be integrated in the future)

At the top right of each panel, the calculated score of the quality level that is the focus of the pane is displayed. The Interoperability view is to be completed by the interoperability map once it has been fully implemented.

This multi-level architecture allows the user to reach for just how much they want to know about a dataset. A user could deduce from Figure 4-2 above that the useability of the dataset considered is not optimal, with a score of 3 out of 5. But, seeing how it performed quite well in the two other viewpoints, if no other dataset were to offer better statistics, this dataset could be given a chance. In these circumstances, a more experienced user would want to gather more information about the scores, thus accessing quality dimensions or elements to pinpoint the exact shortcomings of the dataset.

This example raises a limitation of this solution to our attention. This analysis does not allow for an easy comparison of several datasets. The more immediate remedy would be to be able to put several panes side by side for each dataset compared, but maybe other approaches could be more user-friendly. For instance, a module dedicated to comparisons could be added.

Similarly, in the future, we have to consider ways to assess aggregate the quality analysis results of an ensemble of datasets so that several datasets can be considered together.

In the next section, we will discuss the metadata dashboard developed by Data Europa in the summer 2022. This is a interesting way to conclude this chapter as it allows us to compare findings with a solution that came out almost simultaneously to ours.

4.3 Data Europa metadata dashboard

Data Europa developed, in the summer of 2022, a dashboard dedicated to evaluating the metadata of datasets against indicators chosen by the website (data.europa.eu@2022). This dashboard is aimed at users with a different profile than ours: data providers, with the hope that they can check their metadata and improve it before publishing it.

The home page of this quality dashboard is presented in Figure 4-5.

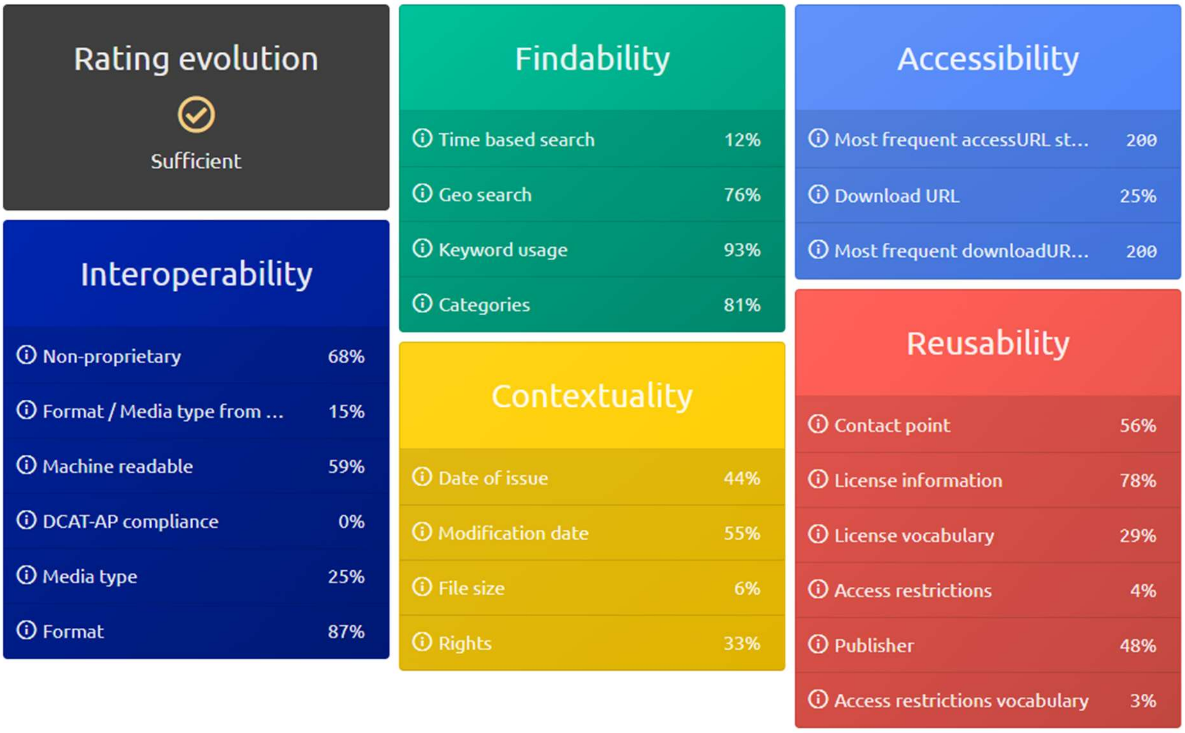


Figure 4-5 Metadata dashboard developed for the official portal for European data (data.europa.eu@2022)

According to their methodology, they defined dimensions based on the FAIR principles. This explains that some of the same ideas behind indicators are found in both dashboards.

Some indicators were not considered in our quality dashboard but could be added to our architecture. For instance, the indicator in their “accessibility” dimension are not considered in our quality dashboard, nor are categories, machine readability, the notion of using correct vocabulary (checked against data.europ.eu approved vocabulary) or contact point information.

Besides the quality indicator themselves, a there is a difference in our methodology in the way they weighted the dimensions. Where we scored the dimensions individually, they assigned a weight to each indicator which, summed, gives a weight to the dimension they refer to. This system allows them to calculate one final score for the whole dataset. The scores themselves also differ as they used a percentage grading.

In conclusion, this dashboard’s metrics seem to intersect ours. It assesses less quality measures than ours does, but it presents interesting metrics that would add value to our analysis. Moreover, the vocabulary used to describe license use and issues of rights is more precise than ours.

Therefore, in order to deal with homogeneous metrics, for the sake of comparison and to lessen any confusion than the already abstract quality analysis measures can instigate, we should consider matching some of our vocabulary to theirs.

5 CONCLUSION & PERSPECTIVES

The topic of the present study is integral to a project with great ambitions: the Geospatially Enabled Ecosystem for Europe project. GeoE3 is part of the D.S.I.'s Open Data Initiative and aims to facilitate the access, analysis, understanding and visualization of geospatial data for the general public through a unique and efficient ecosystem. Focal points of this project include the access, interoperability, and cross-border harmonization of data.

When data from a wide range of sources is gathered, the question of data quality assessment rises. But quality assessment might not be enough in itself to convey the information in a way that can be understood by any user. In the present thesis, a solution to this issue has been presented in the form of a Proof of Concept Quality Dashboard focused on the first use case of GeoE3, associated with solar energy potential and energy efficiency of buildings in changing climate and smart cities. This quality dashboard, dedicated to collect, summarize, and present information from multiple sources, was met with the challenge of having to cater to a wide range of users with different levels of knowledge regarding data quality. It must simultaneously popularize the scientific concepts of data quality and be specific enough to provide enough information for specialists with high expectations, therefore mending the perspectives of data producers, novices and expert data users.

We have herein established a detailed process that can be followed in the context of the development of a product designed to include non-expert users in its application, from the requirement analysis to the implementation. A method to acquire feedback that includes interviews and questionnaires has been proposed with focus points of questioning. We have also established a reusable and detailed data quality vocabulary hierarchy through terms such as quality dimension, element, measure, and metrics that take into consideration international standards as well as more trivial definitions. These terms have subsequently been categorized in three distinct points of view crucial to the project.

An analysis of sixteen dashboard software has led us to choose the ready-made software Microsoft Power BI to develop our Proof of Concept and identify several promising Open Source software whose code could inspire future developments in the GeoE3 project or its successor. In particular, a QGIS plugin could be developed following the work of Tim Sutton's QGIS dashboard plugin. Based on our dashboard, the tool he developed would need to be improved in terms of visualization of metrics.

The Proof of Concept we created focuses on the visualization of data quality metrics in an intelligible way, so that users from different backgrounds in geospatial data and data quality analysis can all benefit from it. From the identified data quality indicators, data quality scores have been calculated using Vaidyambath et al. (2019)'s equations.

We have produced a multi-level quality dashboard that has received positive feedback from stakeholders within and outside the scope of the GeoE3 project. The next step in this process would the actual implementation of the dashboard to the already established GeoE3 platform. This would be done in a number of steps, starting with collaborating with Spatineo to develop APIs in order to allow the dashboard access to actual data quality results.

In the future, Vaidyambath et al. (2019)'s evaluation method, discussed in section 2.6.1, could be applied to the dashboard. That would allow users to give detailed feedback on the quality dashboard through a series of tasks and would bypass the challenge we have been met with in our study of finding willing users to collect feedback from.

Further development could also open up the dashboard to the two other use cases to fully meet the demand of the GeoE3 project. Following the process we have established in this study, it would be very effective to adapt our quality dashboard to these two other use cases.

Additionally, the functionalities of the dashboard could be extended by implementing a way for the users to establish their own threshold or requirements regarding data quality, or by further visualizations of more complex information.

For instance, Devillers et al. (2022) suggested another very promising way to represent quality indicators. Referencing Rivest et al. (2001), Devilliers et al. discussed the possibility to display quality metrics directly on maps using Spatial OLAP systems. This category of system enables the analysis of great volume of geospatial data, allowing the management and the visualization of geometric entities at different levels of detail. It would then be possible to directly visualize quality indicators on the associated geometric entities. Appendix I provides an example of geospatial data quality visualization using a Spatial OLAP system. This solution would provide a way to visualize data quality on a bigger scale, becoming in the same time a visually attractive solution to the comparison of several datasets.

TABLE OF FIGURES

TABLE OF TABLES

TABLE OF EQUATIONS

REFERENCES

#1 Business Analytics Platform & KPI Dashboards [WWW Document], n.d. . Databox. URL https://databox.com/ (accessed 8.29.22).

5. Data quality [WWW Document], n.d. URL https://www150.statcan.gc.ca/n1/pub/92-500-g/2012001/qual-eng.htm (accessed 6.9.22).

About GeoKur [WWW Document], n.d. URL https://geokur.geo.tu-dresden.de/index.php (accessed 8.18.22).

ArcGIS Dashboards | Data Dashboards: Operational, Strategic, Tactical, Informational [WWW Document], n.d. URL https://www.esri.com/en-us/arcgis/products/arcgis-dashboards/overview (accessed 8.28.22).

Atzl, C., Vockner, B., Mittlboeck, M., n.d. Online Visualization of Streaming Data. GI_Forum 2, 57–66. https://doi.org/10.1553/giscience2016_02_s57

Bahim, C., Casorrán-Amilburu, C., Dekkers, M., Herczog, E., Loozen, N., Repanas, K., Russell, K., Stall, S., 2020. The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments. Data Science Journal 19, 41. https://doi.org/10.5334/dsj-2020-041

Batty, M., 2015. A perspective on city dashboards. Regional Studies, Regional Science 2, 29–32. https://doi.org/10.1080/21681376.2014.987540

Batty, M., Axhausen, K.W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., Portugali, Y., 2012. Smart cities of the future. Eur. Phys. J. Spec. Top. 214, 481–518. https://doi.org/10.1140/epjst/e2012-01703-3

BI Reporting & Dashboard Software | FineReport [WWW Document], n.d. URL https://www.finereport.com/en/ (accessed 8.29.22).

BigQuery : entrepôt de données cloud | BigQuery : entrepôt de données cloud [WWW Document], n.d. . Google Cloud. URL https://cloud.google.com/bigquery?hl=fr (accessed 8.29.22).

Building City Dashboards, 2016. URL https://dashboards.maynoothuniversity.ie/ (accessed 8.18.22).

Business Architecture · opengeospatial/GEOE3 Wiki [WWW Document], n.d. . GitHub. URL https://github.com/opengeospatial/GEOE3 (accessed 8.16.22).

CARTO, n.d. Introducing the new CARTO dashboard [WWW Document]. URL https://carto.com/blog/new-dashboard/ (accessed 8.29.22).

CityDashboard: London [WWW Document], n.d. URL https://citydashboard.org/london/ (accessed 8.19.22).

CityDashboard [WWW Document], n.d. URL https://citydashboard.org/about.php (accessed 8.19.22).

Cotton, I.W., Greatorex, F.S., 1968. Data structures and techniques for remote computer graphics, in: Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I on - AFIPS ’68 (Fall, Part I). Presented at the the December 9-11, 1968, fall joint computer conference, part I, ACM Press, San Francisco, California, p. 533. https://doi.org/10.1145/1476589.1476661

Cyfe | All-In-One Business Dashboard. Visualize your KPIs. [WWW Document], 2019. URL https://www.cyfe.com/ (accessed 8.29.22).

Dameri, R., 2017. Urban Smart Dashboard. Measuring Smart City Performance. pp. 67–84. https://doi.org/10.1007/978-3-319-45766-6_4

Dashboards and Dashboard Builder | Snap4City [WWW Document], n.d. URL https://www.snap4city.org/drupal/node/17 (accessed 8.19.22).

Dasheroo - Free Business Dashboards Done Right [WWW Document], n.d. URL https://dasheroo.com/ (accessed 8.29.22).

Dataset Metadata · Open Data - Metadata Guide [WWW Document], n.d. URL https://centerforgov.gitbooks.io/open-data-metadata-guide/content/dataset-metadata.html (accessed 2.22.22).

Devillers, R., Gervais, M., Bédard, Y., Jeansoulin, R., 2002. Spatial data quality: From metadata to quality indicators and contextual end-user manual 21–22.

Draft ISO/NP 19157-3, Geographic information – Data Quality – Part 3: Data quality measures register, 2021.

eTranslation [WWW Document], n.d. URL https://cor.europa.eu/en/engage/Pages/e-translation.aspx (accessed 8.30.22).

Eurooppalainen ponnistus paikkatietojen tehokäytön puolesta [WWW Document], n.d. . Maanmittauslaitos. URL http://www.maanmittauslaitos.fi/tietoa-maanmittauslaitoksesta/organisaatio/lehdet-ja-julkaisut/positio/eurooppalainen-ponnistus-paikkatietojen-tehokayton-puolesta (accessed 8.11.22).

freeboard - Dashboards For the Internet Of Things [WWW Document], n.d. URL https://freeboard.io/ (accessed 8.29.22).

GeoE3 – A geospatially enabled ecosystem for Europe, 2022. . geoforum.fi. URL https://geoforum.fi/geoe3-a-geospatially-enabled-ecosystem-for-europe/ (accessed 6.28.22).

Geospatially Enabled Ecosystem for Europe (GeoE3) | OGC [WWW Document], n.d. URL https://www.ogc.org/projects/initiatives/geoe3 (accessed 6.28.22).

Google Data Studio [WWW Document], n.d. URL https://datastudio.google.com/u/0/ (accessed 8.29.22).

HUISMAN-VAN ZIJP, A., SJOUKEMA, J.-W., 2020. Boosting Data Quality by a Participatory Approach: the Success of Quality Dashboards and Crowdsourced Feedback.

ISO 19157:2013 Geographic information - Data quality - ICA Wiki [WWW Document], n.d. URL https://wiki.icaci.org/index.php?title=ISO_19157:2013_Geographic_information_-_Data_quality (accessed 6.9.22).

ISO/CD 19157-1:2021(E) ; Geographic information — Data quality – Part 1: General requirements, 2021.

ISO_DIS_19115-1_(E) : Geographic information — Metadata — Part 1: Fundamentals, 2011.

Jakobsson, A., Lehto, L., 2021. Eurooppalainen ponnistus paikkatietojen tehokäytön puolesta. Positio 2, 26–27.

Jing, C., Du, M., Li, S., Liu, S., 2019. Geospatial Dashboards for Monitoring Smart City Performance. Sustainability 11, 5648. https://doi.org/10.3390/su11205648

Koshkarev, A.V., 2008. Geoportal as a tool to control spatial data and services Spatial data.

Lee, D., 2019. Big Data Quality Assurance Through Data Traceability: A Case Study of the National Standard Reference Data Program of Korea. IEEE Access PP, 1–1. https://doi.org/10.1109/ACCESS.2019.2904286

Lwin, K.K., Sekimoto, Y., Takeuchi, W., Zettsu, K., 2019. City Geospatial Dashboard: IoT and Big Data Analytics for Geospatial Solutions Provider in Disaster Management, in: 2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM). Presented at the 2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), pp. 1–4. https://doi.org/10.1109/ICT-DM47966.2019.9032921

Lwin, K.K., Zettsu, K., Sugiura, K., 2015. Geovisualization and correlation analysis between geotagged Twitter and JMA rainfall data: Case of heavy rain disaster in Hiroshima, in: 2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM). Presented at the 2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM), pp. 71–76. https://doi.org/10.1109/ICSDM.2015.7298028

Map Visuals for Power BI | ArcGIS for Power BI [WWW Document], n.d. URL https://www.esri.com/en-us/arcgis/products/arcgis-for-power-bi/overview (accessed 8.30.22).

MetricFire, n.d. MetricFire [WWW Document]. URL https://www.metricfire.com/ (accessed 8.28.22a).

MetricFire, n.d. Pricing | MetricFire [WWW Document]. URL https://www.metricfire.com/pricing/ (accessed 8.28.22b).

Mozaïk | Mozaïk [WWW Document], n.d. URL http://mozaik.rocks/ (accessed 8.29.22).

OGC API [WWW Document], n.d. URL https://ogcapi.ogc.org/ (accessed 8.11.22).

Phillips, A., Williamson, I., Ezigbalike, C., 1999. Spatial Data Infrastructure Concepts. Australian Surveyor 44, 20–28. https://doi.org/10.1080/00050351.1999.10558768

Portability [WWW Document], n.d. URL https://iso25000.com/index.php/en/iso-25000-standards/iso-25010/64-portability (accessed 8.31.22).

Premium Bootstrap Admin Dashboard Themes [WWW Document], n.d. . Keenthemes. URL https://keenthemes.com/ (accessed 8.29.22).

Qvain | Research Dataset Description Tool [WWW Document], n.d. URL https://qvain.fairdata.fi/dataset (accessed 8.25.22).

Qvain – Research Dataset Description Tool [WWW Document], n.d. . Fairdata. URL https://www.fairdata.fi/en/qvain/ (accessed 8.25.22).

Research Data Alliance FAIR Data Maturity Model Working Group, 2020. FAIR Data Maturity Model: specification and guidelines. https://doi.org/10.15497/RDA00050

Rula, A., Zaveri, A., 2014. Methodology for Assessment of Linked Data Quality.

Santoro, M., Vaccari, L., Mavridis, D., Smith, R., Posada, M., Gattwinkel, D., n.d. Web APIs: general-purpose standards, terms and European Commission initiatives. Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-13183-0. https://doi.org/10.2760/675, JRC118082

Setting up a Redash Instance [WWW Document], n.d. URL https://redash.io/help/open-source/setup (accessed 8.29.22).

Stodder, D., n.d. Data Visualization and Discovery for Better Business Decisions 35.

Titov, A.G., 2020. Geospatial meteorological and climatic data services application programming interface. IOP Conf. Ser.: Earth Environ. Sci. 611, 012059. https://doi.org/10.1088/1755-1315/611/1/012059

Toronto, C. of, 2017. Toronto's Dashboard [WWW Document]. City of Toronto. URL https://www.toronto.ca/city-government/data-research-maps/toronto-progress-portal/ (accessed 8.19.22).

Vaidyambath, R., Debattista, J., Srivatsa, N., Brennan, R., 2019. An Intelligent Linked Data Quality Dashboard 12.

Vladimir Zdraveski, V., Trajanov, D., Mishev, K., Kocarev, L., n.d. ISO-Standardized Smart City Platform Architecture and Dashboard. IEEE Pervasive Computing 35–43. https://doi.org/10.1109/MPRV.2017.31

What Is an API, and How Do Developers Use Them? [WWW Document], n.d. . How-To Geek. URL https://www.howtogeek.com/343877/what-is-an-api/ (accessed 8.11.22).

What is an API? In English, please. [WWW Document], 2019. . freeCodeCamp.org. URL https://www.freecodecamp.org/news/what-is-an-api-in-english-please-b880a3214a82/ (accessed 8.11.22).

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., ’t Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Yigitbasioglu, O.M., Velcu, O., 2012. A review of dashboards in performance management: Implications for design and research. International Journal of Accounting Information Systems 13, 41–59. https://doi.org/10.1016/j.accinf.2011.08.002

Young, G.W., Kitchin, R., Naji, J., 2021. Building City Dashboards for Different Types of Users. Journal of Urban Technology 28, 289–309. https://doi.org/10.1080/10630732.2020.1759994

TABLE OF APPENDICES

Appendix A List of partners in the GeoE3 project

Partners	*Geographical Area*
Finnish Meteorological Institute	Finland
Statistics Finland	Finland
Spatineo Inc.	Finland
Norwegian Mapping Authority	Norway
Cadastre, Land Registry and Mapping Agency	The Netherlands
Centro Nacíonal de Información Geográfica	Spain
Estonian Land Board	Estonia
Open Geospatial Consortium (OGC)	Europe
Information Technology Center of the Ministry of Environment (KEMIT)	Estonia
Aventi Intelligent Communication	Norway
Dirección General del Catastro	Spain

Table 8: List of partners in the GeoE3 project (maanmittauslaitos@2022)

Appendix B Timeline of some significant geospatial dashboard events

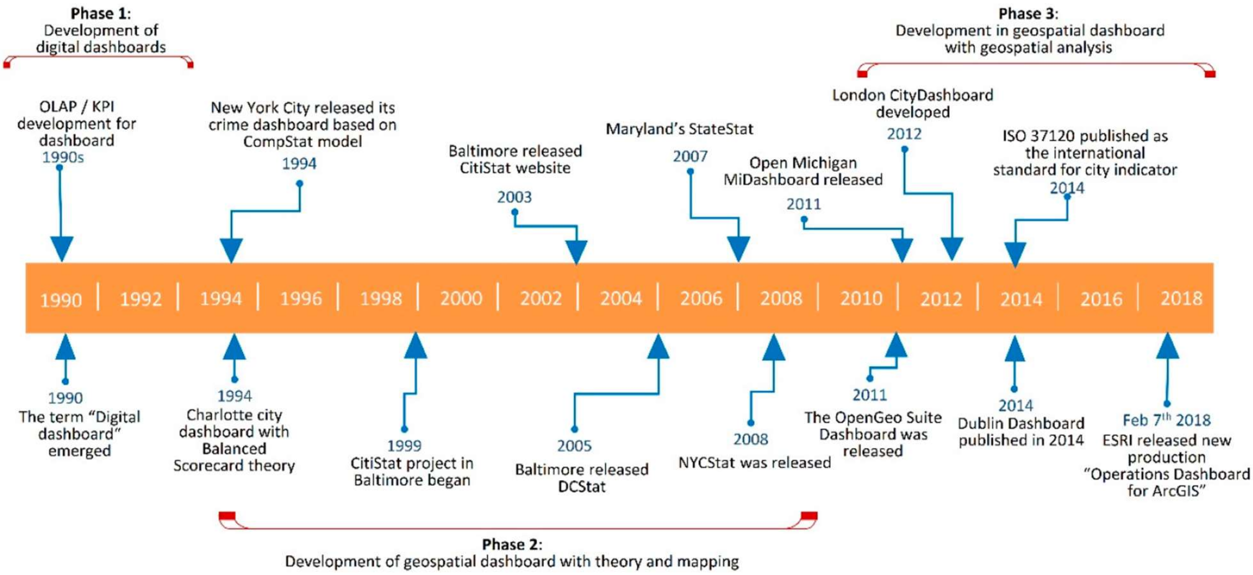


Figure B-1: Timeline of some significant geospatial dashboard events, created by Changfeng et al. (2019)

Appendix C Recap of the main category of data involved in city dashboard and examples

Category of data	**Examples of data analysed**
Social data & security	Social class, religion, life expectancy, population data, community vulnerability metrics, crime metrics,
Economic data	Development and construction, employment, real estate, business, revenues metrics
Population data	Density
Environmental data	Pollution, waste, garbage, water quality, weather, wind, temperature, air conditions
Social media information	Twitter feeds, Facebook feeds
Mobility	Public and private transportation: Availability, position in real time, average delay, number, and location of events in traffic, transportation safety metrics, traffic flow, cycling paths and flow
Energy	Energy consumed / saved
Health	Hospitals, risks, accident and emergency waiting tomes

Table 9 : Recap of the main category of data involved in city dashboard and examples;

Data in the above table was collected and compiled from Toronto.ca@2022, Changfenf et al. (2019), dashboard.edmonton.ca@2022 and snap4city.org@2022

Appendix D Examples of city dashboards

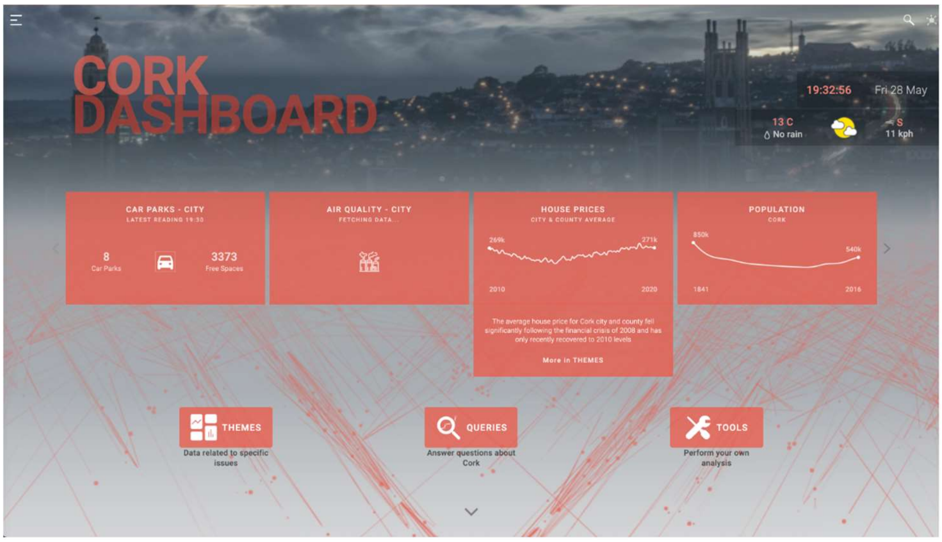


Figure D-1 Example of a smart city dashboard : the Cork Dashboard (maynoothuniversity.ie@2022)

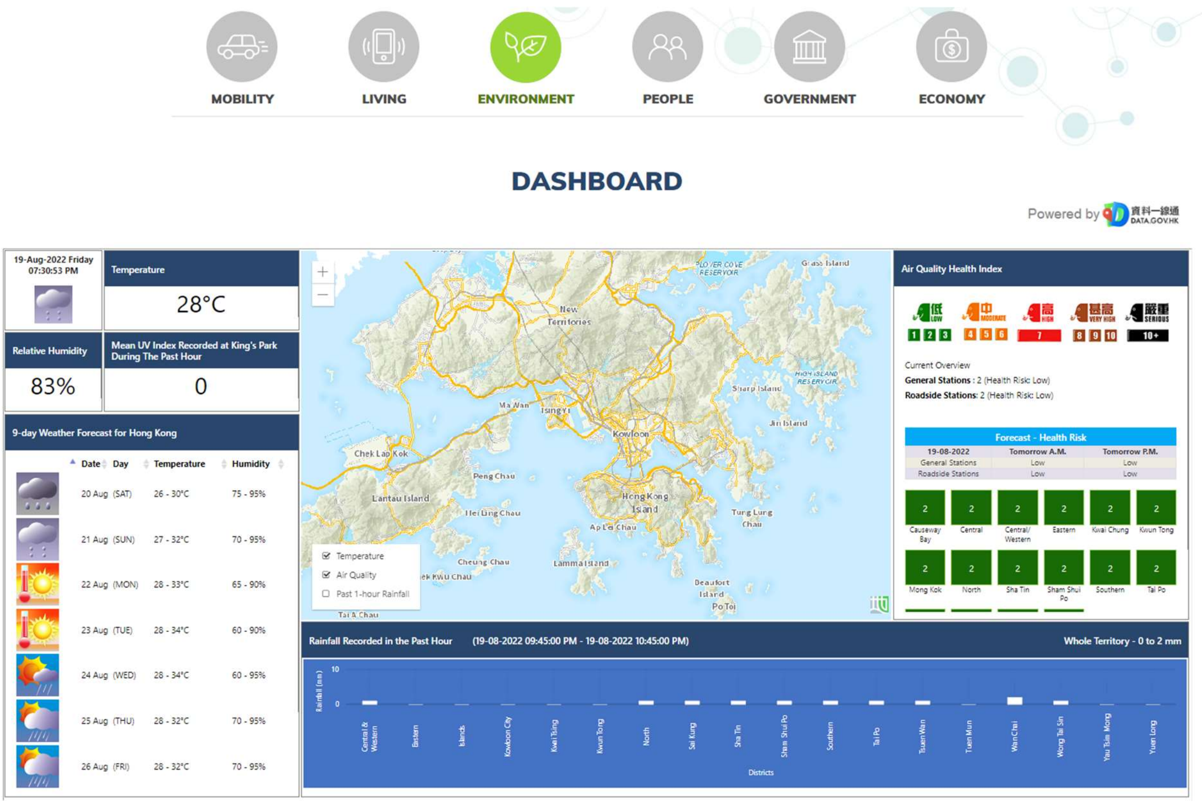


Figure D-2: Hong Kong Smart City Dashboard, Environment tab (smartcity.gov.hk@2022)

Appendix E FAIR data maturity model indicators, Bahim & al. (2019)

FAIR	ID	Indicator	Priority	
F1	RDA-F1-01M	Metadata is identified by a persistent identifier	□□□	Essential
F1	RDA-F1-01D	Data is identified by a persistent identifier	□□□	Essential
F1	RDA-F1-02M	Metadata is identified by a globally unique identifier	□□□	Essential
F1	RDA-F1-02D	Data is identified by a globally unique identifier	□□□	Essential
F2	RDA-F2-01M	Rich metadata is provided to allow discovery	□□□	Essential
F3	RDA-F3-01M	Metadata includes the identifier for the data	□□□	Essential
F4	RDA-F4-01M	Metadata is offered in such a way that it can be harvested and indexed	□□□	Essential
A1	RDA-A1-01M	Metadata contains information to enable the user to get access to the data	□□	Important
A1	RDA-A1-02M	Metadata can be accessed manually (i.e. with human intervention)	□□□	Essential
A1	RDA-A1-02D	Data can be accessed manually (i.e. with human intervention)	□□□	Essential
A1	RDA-A1-03M	Metadata identifier resolves to a metadata record	□□□	Essential
A1	RDA-A1-03D	Data identifier resolves to a digital object	□□□	Essential
A1	RDA-A1-04M	Metadata is accessed through standardised protocol	□□□	Essential
A1	RDA-A1-04D	Data is accessible through standardised protocol	□□□	Essential
A1	RDA-A1-05D	Data can be accessed automatically (i.e. by a computer program)	□□	Important
A1.1	RDA-A1.1-01M	Metadata is accessible through a free access protocol	□□□	Essential
A1.1	RDA-A1.1-01D	Data is accessible through a free access protocol	□□	Important
A1.2	RDA-A1.2-01D	Data is accessible through an access protocol that supports authentication and authorisation	□	Useful
A2	RDA-A2-01M	Metadata is guaranteed to remain available after data is no longer available	□□□	Essential
I1	RDA-I1-01M	Metadata uses knowledge representation expressed in standardised format	□□	Important
I1	RDA-I1-01D	Data uses knowledge representation expressed in standardised format	□□	Important
I1	RDA-I1-02M	Metadata uses machine-understandable knowledge representation	□□	Important
I1	RDA-I1-02D	Data uses machine-understandable knowledge representation	□□	Important
I2	RDA-I2-01M	Metadata uses FAIR-compliant vocabularies	□□	Important
I2	RDA-I2-01D	Data uses FAIR-compliant vocabularies	□	Useful
I3	RDA-I3-01M	Metadata includes references to other metadata	□□	Important
I3	RDA-I3-01D	Data includes references to other data	□	Useful
I3	RDA-I3-02M	Metadata includes references to other data	□	Useful
I3	RDA-I3-02D	Data includes qualified references to other data	□	Useful
I3	RDA-I3-03M	Metadata includes qualified references to other metadata	□□	Important
I3	RDA-I3-04M	Metadata include qualified references to other data	□	Useful
R1	RDA-R1-01M	Plurality of accurate and relevant attributes are provided to allow reuse	□□□	Essential
R1.1	RDA-R1.1-01M	Metadata includes information about the licence under which the data can be reused	□□□	Essential

R1.1	RDA-R1.1-02M	Metadata refers to a standard reuse licence	□□	Important
R1.1	RDA-R1.1-03M	Metadata refers to a machine-understandable reuse licence	□□	Important
R1.2	RDA-R1.2-01M	Metadata includes provenance information according to community-specific standards	□□	Important
R1.2	RDA-R1.2-02M	Metadata includes provenance information according to a cross-community language	□	Useful
R1.3	RDA-R1.3-01M	Metadata complies with a community standard	□□□	Essential
R1.3	RDA-R1.3-01D	Data complies with a community standard	□□□	Essential
R1.3	RDA-R1.3-02M	Metadata is expressed in compliance with a machine-understandable community standard	□□□	Essential
R1.3	RDA-R1.3-02D	Data is expressed in compliance with a machine-understandable community standard	□□	Important

Table 10: FAIR data maturity model indicators, Bahim & al. (2019)

Appendix F FAIR maturity model indicator priorities, Bahim et al. (2019)

Level of priority	Description
Essential	Such an indicator addresses an aspect that is of the utmost importance to achieve FAIRness under most circumstances, or, conversely, FAIRness would be practically impossible to achieve if the indicator were not satisfied.
Important	Such an indicator addresses an aspect that might not be of the utmost importance under specific circumstances, but its satisfaction, if at all possible, would substantially increase FAIRness.
Useful	Such an indicator addresses an aspect that is nice-to-have but is not necessarily indispensable.

Table 11: Description of the FAIR maturity model indicator priorities, Bahim et al. (2019)

Appendix G List of untested platforms that could have presented an interest in the project

Name of the platform	Source
Carto	https://carto.com/blog/new-dashboard/
Countly	https://support.count.ly/hc/en-us
Cyfe	https://www.cyfe.com/
Dasheroo	https://dasheroo.com/
Databox	https://databox.com/
Fine report	https://www.finereport.com/en/
Google data studio	https://datastudio.google.com/u/0/
Google bigquery	https://cloud.google.com/bigquery?hl=fr

Table 12: List of untested dashboard platforms

Appendix H First version of quality dashboard, Interoperability layout

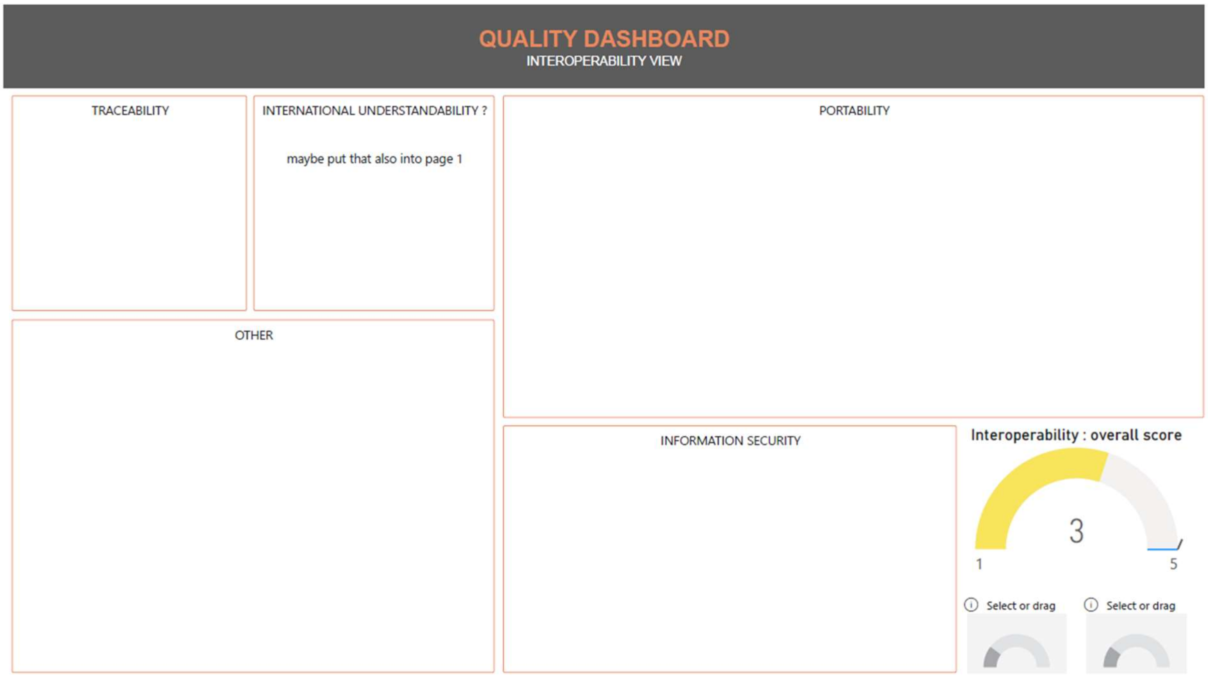


Figure H-1: First version of quality dashboard, Interoperability layout

Appendix I Example of visualization using a Spatial OLAP system (Devillers et al., 2022)

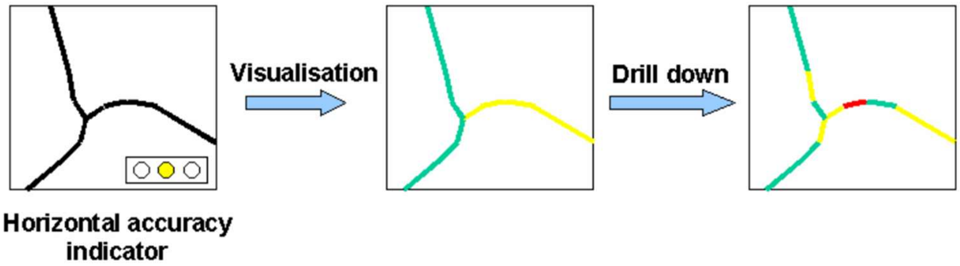


Figure I-1: Visualization of geospatial data quality using indicators and a Spatial OLAP system (Devilliers et al., 2022)

According to Devilliers et al. (2022), users can display quality indicators either with a streetlight representation, or by directly associating the quality indicators to the individual objects, at different levels of detail.