Master Thesis Geoinformatics

Cristina Gallego Torres

2015

Data Standardization and Data Modelling in compliance with the European Water Framework Directive.

Approaches and best practices to build and maintain a Data Warehouse for groundwater related information at GEUS

AALBORG UNIVERSITET
M. Sc. Geoinformatics
Aalborg University - Copenhagen

Table of Contents

Abstract

The new European Water Framework Directive (WFD) has established a legal framework to protect water across Europe so that long-term sustainable water use is assured. This community action regarding water quality requires Member States to report a wide range of water related data like, inland surface waters, transitional waters, coastal waters and groundwater. Hence, the need of data reporting to the European Environment Agency (EEA) has clearly originated a problem in the national organizations in charge of this purpose.

For the Danish case, GEUS (Geological Survey of Denmark and Greenland) is the organization that produces groundwater related data. As a research institution, GEUS carries out many researches and studies in relation to different fields of knowledge resulting in a large amount of data, which urges to be maintained, managed and controlled since it is valuable data not only at a national level but also to accomplish with European reporting obligations. This extensive data does not necessarily meet the WFD standards, in fact, some European requirements have not been even considered due to the apparently lack of relevance for the organization. Therefore, there seems to be a clear need of standardizing groundwater data while meeting with the European requirements.

This Master Thesis tries to solve this problem by modelling a database which could help in sorting internal groundwater related data in accordance with European legal framework. Along with the conceptual and logical data model, which was designed based on both, data modelling theoretical concepts and GEUS features; the thesis also provides a data standardization guide based on EEA resources, and a set of recommendations and best practices for a more efficient data management. All in all, this report attempts to offer GEUS resources so that they can deal with groundwater data with a particular regard to the European legislation. Last but not least, from a European level point of view, this report ultimately could contribute to the protection of water in Europe.

1 Introduction

First of all, it is important to remark that this thesis is the continuation of a previous project which took place during an internship of the author at GEUS. In fact it is not strictly the continuation of it but this thesis tries to solve the same problem by using a different approach. The semester project involved a deep research of WFD requirements from Member States as well as getting to know what the status of GEUS about the European requirements was. In this thesis author draws on this previous research to present the problem and the European regulation.

The report is structured as follow; the first chapter presents the problem statement and the organization in which the thesis took place, GEUS and the hydrology department. Chapter 2 is called Theory, and it is comprised of two important subsections named European Union Framework and database theory, in these two sections it can be found the theory on which the report has been based on. Notice that European Union Framework section is based on a former project (Gallego, 2015)developed by the author of this thesis regarding the same issue in the same organization. However, database theory provides the background information needed to understand the data model designed in the implementation section.

Chapter 3 explains the methodology of workflow taken for the development of the data model.

The implementation section, within chapter 4, mainly explains the data model here designed, its entities and their relationships. Finally chapter 5 and 6, represent the discussion of the report and wrap up conclusions respectively.

Also notices that the appendixes included at the end of the report comprise the actual data standardization guide and a complementary code list respectively

1.1 Problem statement

"We must improve the way we use and manage our water resources if we are to continue to benefit from the vital services our water ecosystems provide" (EU, 2014)

GEUS, as a national research institution, disposes of several databases where Denmark and Greenland information on many fields of knowledge is saved. The amount of data generated in order to perform different researches is huge. Regarding Geological and Hydrogeological information, data is mostly gathered into Jupiter database which stores information from over 240,000 wells, geological descriptions, graphical views, and groundwater chemistry data. (GEUS, 2010).

Due to the importance of water for human beings development and in order to protect water and ensure its good condition, a new WFD was adopted by the EU. Its official name is Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000, this regulation establishes a framework for Community action in the field of water policy. Consequently the WFD establishes data reporting obligations from Member States (WISE, 2008).

GEUS hydrogeological information regarding groundwater bodies is found and extracted from different sources like Jupiter database and National Water Resources Model (DK-MODEL). This information is structured in three different groups: Chemistry characterisation workgroup; Water balance workgroup; and Water Bodies workgroup. From these workgroups, data is constantly generated as output data ready to use in GEUS researches (GEUS, 2010). However, all this amount of produced data does not necessarily match EEA standards.

WFD reporting obligations regarding groundwater bodies are specified at both Annex II 2.1 and 2.2. GEUS has previously attempted to originate data, for the Annex II 2.1, developing polygons and raster grid information of approximately 400 groundwater bodies, which cover all mayor aquifers in Denmark (roughly 2700). Polygon information is based on horizontal extent measures, while raster grid gives vertical extent information including thickness and depth. Hence there seems to be a clear need for standardizing groundwater data meanwhile reporting obligations are considered, this is why it is necessary to develop a database which could potentially gather groundwater bodies' data as well as, containing the next specific information required within the Annex II 2.2:

- Geological characteristics of the groundwater bodies, including the extent and type of geological units.
- Hydrogeological characteristics of the groundwater body including hydraulic conductivity, porosity and confinement.
- Characteristics of the superficial deposits and soils in the catchment from which the groundwater body receives its recharge, including the thickness, porosity, hydraulic conductivity, and absorptive properties of the deposits and soils.
- Stratification characteristics of the groundwater within the groundwater body.
- An inventory of associated surface systems, including terrestrial ecosystems and bodies of surface water, with which the groundwater body is dynamically linked.
- Estimates of the directions and rates of exchange of water between the groundwater body and associated surface systems.
- Sufficient data to calculate the long term annual average rate of overall recharge.
- Characterization of the chemical composition of the groundwater, including specification of the contributions from human activity. Member States may use typologies for groundwater characterization when establishing natural background levels for these bodies of groundwater.

1.2 Approach

This Thesis intents to offer some tools to solve the WFD problem by modelling a database, which could be helpful to manage internal groundwater related data in accordance with European legal framework. In addition to the data model, the thesis also provides a data standardization guide based on EEA sources, and some recommendations and best practices for the development of the database. Overall, this report attempts to supply GEUS with resources so that they can deal with groundwater data, with a particular regard to the European legislation. Last but not least, from a European level point of view, this report ultimately could contribute to protect water in Europe.

It is important to remark that the Thesis has not been thought as a computer science project. Rather the purpose is to provide GEUS an overall document that presents a potential solution for the WFD issue. This is why the solution does not only focus on the data model, but also on the data standardization guide and data warehousing recommendations. It seems to be more logical to write a document involving these three aspects than focusing on just one, by acting this way GEUS could use this document to actually start implementing a solution for the WFD problem.

1.3 Presenting GEUS

GEUS, Geological Survey of Denmark and Greenland, is a national institution of Denmark that carries out many different geoscientific studies, research consultancy, and geological mapping, GEUS is not only a research institution but also it is advisory in the Ministry of Climate and Energy and Building, It is partner in Geocenter Denmark, which is a national centre of geoscientific research, education and consulting. GEUS is formed by different departments depending on their field of research (See Figure 1.) The report here presented has been developed in the Hydrology department (GEUS, 2014)

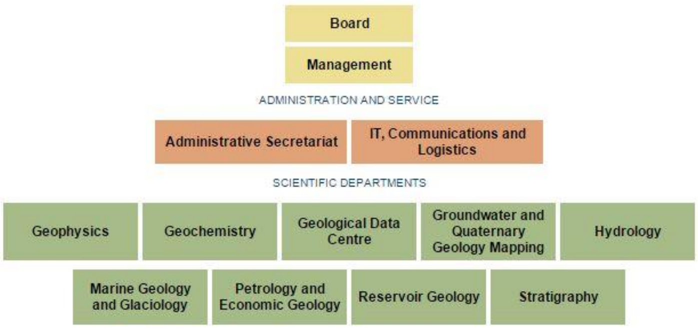


Figure 1 (GEUS, 2014)

1.3.1 Hydrology department

This section aims at providing some background information about the work at the Hydrology department; which focuses on characterisation, modelling, and management of water and solute transport in the hydrological cycle not just in Denmark but also internationally. It is well known that climate, land use, and water circulation highly influence the evolution of water quantity and quality in what concerns the hydrological cycle (GEUS, 2014),therefore the main areas in this department involve the interaction between land surface, root zone, groundwater bodies, watercourses, lakes, wetlands, and the sea. The researches that are taking place nowadays at the Hydrology department diverge into the following primary fields:

- Hydrogeological characterisation
- Water (flow) and solute transport
- Climate change effects and adaption
- Water management
- Capacity building

The compliance of the WFD could be placed in the water management field, which takes care of the integrated water resources, adaptive water management, groundwater protection and ground water management (GEUS, 2014). However, meeting WFD requirements involves more departments and people from the organization; this is explained in later sections.

1.3.2 Jupiter database

GEUS' Jupiter Database is a database that maintains, and allocates information regarding groundwater, drinking water, raw materials, environmental and geotechnical data, in general hydrogeological and geological information.

The main purpose of the Jupiter well database is to facilitate the access to information in order to be used for the creation of new well records, well construction diagrams, geological profiles, geological maps, and groundwater level maps, being this data new input to software like Geographic Information Systems (GEUS, 2014)

Well data archive currently lead information about wells into Jupiter database when a new well is under construction or is reported to GEUS; wells have a unique number of identification (Well File Number), and is recorded on the well file report, along with its sample description, water analysis and test pumping form, etc. (GEUS, 2014)

Currently, reports and maps related to wells are headed to the Well Data Archive from where information over 270,000 wells are saved into Jupiter database and distributed as follow:

- Water supply wells: 50%
- Geotechnical wells: 29%
- Raw material wells: 6%
- Others: 15%

(GEUS, 2014)

Why is Jupiter Database relevant for this report?. It's one of the main sources of water related data, hence, it should be taken into consideration for the solution that this thesis tries to provide. The link between the database and Jupiter will be explained in the implementation section.

2 Theory

2.1 European Union framework

Since the approach of this master thesis is providing GEUS with the necessary information to solve the issue induced by the WFD, it is necessary to briefly explain the main institutions and regulations behind the water policy in EU. European policy regarding water environmental institutions might seem confusing at some point, that's why this section is necessary. Table 1 Source: Author

EU regulations	EU institutions	EU infrastructure
Water Framework directive	Europe Environment Agency	Reportnet
Water Framework directive for Groundwater bodies		Eionet
River basin management plans and programmes of measures		WISE
Article 4 and Article 11 of WFD		

Table 1 Source: Author

2.1.1 EU institutions

2.1.1.1 European Environment Agency

EEA is part of the EU, and it's in charge of offering environmental information for those actors in relation to developing, adopting, implementing and evaluating environmental policy, and it also takes into account general public. The EEA, overall, supposes one of the principal information sources for this master thesis. (EEA, 2003)

EEA is extremely crucial because it can be considered highly relevant for being providers of *"timely, targeted, important and reliable information to policy-making agents and the public"* (EEA, 2003). The EEA serves to the European Institutions (the European Commission, European Parliament, and Council) and the member countries. Moreover, the Economic and Social Committees and the Committee of the Regions are also served, as well as business community, academia, non-governmental organizations and other parts of the civil society which are constantly users of the EEA's information. To sum up, *"the EEA aims to support sustainable development and to help achieve significant and measurable improvement in Europe's environment."* (EEA, 2003)

2.1.2 EU regulations

2.1.2.1 The European Union Water Framework Directive (WFD)

It's the main regulation behind this Master Thesis; in other words, this database aims to fulfil the WFD while providing cleansing in groundwater body related data. The main idea behind explaining this regulation is that GEUS comprehends the importance of meeting these requirements. Also, an additional reason is that, understanding this regulation when someone approaches for the first time might induce to mislead.

It all began seeking for palliating the fact that *"Europe's water is under pressure."* (EU, 2014) Economic activities, population growth and urbanization increase day by day the pressures on fresh water across Europe (EU, 2014)

According to EU (2014), citizens and environmental organizations concerns on the status of rivers and lakes, groundwater and coastal beaches, have been of discussion for long. Thus, to meet these demands the EU Commission has focused its policies on water protection, by establishing a new European Water Policy aiming at cleaning and restoring polluted water, for reaching this purpose the implication of citizens and, particularly organizations like GEUS is crucial too.

"The Water Framework Directive establishes a legal framework to protect and restore clean water across Europe and ensure its long-term and sustainable use. (Its official title is Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy.)" (WISE, 2008)

This directive attempts to create an innovative way for water management based on river basins, which are the natural geographical and hydrological units, establishing limited deadline to Member States for the protection of the aquatic ecosystems. In the Danish case, GEUS holds and generates a large amount of water related data. Inland surface water, transitional waters, coastal water and groundwater are involved.

More specifically, the key aims of the WFD are:

- *"Expanding the scope of water protection to all water, surface waters and groundwater*
- *Achieving "good status" for all waters by a set deadline*
- *Water management based on river basins*

- *Combined approach of emission limit values and quality standards*
- *Getting the prices right*
- *Getting the citizen involved more closely*
- *Streamlining legislation"*

(EU, 2014)

If we take into consideration that the WFD aims at the protection of the water of Europe, and that this report deals with providing GEUS with the necessary tools to meet the WFD. The resolution of this report, ultimately aims to protect groundwater in Denmark and consequently water across Europe.

2.1.2.2 The European Union Water Framework Directive for groundwater bodies

Groundwater is considered a main source of drinking water; this is why it is of vital importance to protect it. Thus, the good status of groundwater in terms of quantity and quality (chemical status), were established as one of the targets to be achieved by 2015 according to WFD for groundwater bodies, (WISE, 2008). The WFD for groundwater bodies is the part of the regulation that specifically deals with the protection of this kind of water.

Quantity, refers to the extracted water from a groundwater body so that, the quantity of water needed to replenish it, is not affected, thus ensuring the long-term sustainable use of, and therefore avoid its depletion. (WISE, 2008)

As far as concerns to the chemical status of groundwater bodies, in order to reach a good quality status, the Nitrates Directive is also established, which states measures to protect groundwater as well as surface water from pollution caused by nitrogen-based fertilizers used in agriculture. (EU, 2014; WISE, 2008)

"The Water Framework Directive recognizes the importance of the cycle linking groundwater and surface waters and it specifies that good status – in both quantity and chemical terms – of a groundwater body also means protecting the surface water bodies and terrestrial ecosystems that depend on its water." (WISE, 2008)

At last but not least, for a good status of groundwater bodies, delineation is extremely important: Groundwater bodies comprise a specific volume of water within an aquifer in which it could be found relevant water flows and extractions. Detailed monitoring of these bodies along with scientific knowledge about underground geology, are keys in the delineation of individual groundwater bodies. Having separate bodies is an excellent way of managing and protecting groundwater. Moreover, being focused on this delineation makes easy to identify bodies under pressure, thus making possible to find which ones are at risk or not (WISE, 2008). This Master Thesis, by providing a database model, will help GEUS in managing this kind of data, which will be valuable to assess quantity and quality status of groundwater bodies at European level

2.1.2.3 River basin management plan and programme of measures

The management activities designed in order to achieve the goals set by the WFD are considered within geographical boundaries, river basin districts (RBDs) instead of political boundaries. These river basins districts are based on surface water catchments along with, boundaries of associated groundwater bodies and coastal waters. This means that, the goals set by the WFD are specifically determined on the river basin management plan for each single European river basin district and not for countries (EEA). These plans for each individual river basin district, consists on, analysis, monitoring, objective-setting and consideration of measures to maintain or improve the water status within each hydrological unit (RBDs). In a more specific way the targets are:

- *"Record the current status of water bodies within the river basin district;*
- *Set out the measures planned to meet the objectives;*
- *Act as the main reporting mechanism to the Commission and the public."* (EEA)

The river basin management plan is made of three phases: planning, implementation, and evaluation of the program of measures, which aims at achieving the environmental objectives of the WFD. It is an iterative process since it is reviewed, established and updated every six years.

At last but not least, *"these RBMPs should be made available for information and consultation by the public."* (EEA)

2.1.2.4 Delineation and further characterisation (Article 4 and Article 11 of WFD)

The actual requirements of the WFD (Article 4 & 11) concerning groundwater bodies are developed within two annexes of the WFD. These two articles are specified in the Annex II, 2.1 and 2.2 respectively. The Annex II 2.1 defines the guides and the first steps to meet the WFD and it's mainly concerned about groundwater bodies delineations among other information, whereas, the 2.2 is more related to information for risk assessment of groundwater bodies (WISE Annex II)

The Article 4 enforces Member States to identify data related to:

- Locations and groundwater boundaries.
- Pressure to which the groundwater bodies are more likely to be subject to.
- General characteristics of the overlaying strata of the groundwater body catchment area.

(WISE, 2013)

Annex II 2.2, in reference to the Article 11 of the WFD requires information about other topics such as geological data, water balance, stratification etc.

2.1.3 EU infrastructures

2.1.3.1 Reportnet

The Repornet is the principal infrastructure that fosters and provides flows of information from Member States to the EEA. Its main objective is to support environmental policy developing in Europe. Descriptions on how trends and models about how environmental concerns modify and evolve are extremely important when it comes to help countries and European institutions to formulate policies. Thus, the relevance of good flows of information among European countries at an international level in order to easily understand environmental progress is crucial for the EU. Therefore, Reportnet's goal is to allow the access of multiple users to reported data (EEA, 2003)

Repornet's tool helps in enhacing this good flow of information. Aiming to increase transparency and reduce to a minimum the need of resources when sharing information, repornet's tool provides:

- A guide of reporting obligations to each country plan
- A smart software which help countries on data distribution tasks, making sure that, data is comparible , quality assured, and fully audited.
- One place to easily store information, referencing, and archiving it.
- It offers means for countries whenever they need other countries' information.

More specifically these tools are: Reporting Obligations Database (ROD); Network Directory (ND); Data Dictionary (DD); Data Exchange Modules (DEM); Data Repositories (CDR); Content Registry (CR); European Data Warehouses; and Indicator management.

The Data Dictionary (DD), keeps technical specifications about the data requested in reporting obligations. It provides detailed specification about how data should be stored and reported to make it comparable; such as, data file structures, definitions of each data set, tables, their data elements, permitted values and other issues which help in reporting good quality data.(EEA, 2003). This data dictionary has had a very important role for this report since the data elements of the database will have to meet data dictionary requirements. Additionally, the section of this report where the data elements are explained (Annex I: Data standardization guide) has been based on the data dictionary

"Reportenet is a contribution to what is a shared challenge for European countries, their networks of institutions and EU institutions to develop technical and organisational systems which facilitate streamlined, transparent and shared use of information." (EEA, 2003)

2.1.3.2 European Environment Information and Observation Network; EIONET

Eionet's main aim is *"to provide timely and quality-assured data, information and expertise…"* (EEA, 2012) this is done by gathering each country's individual environmental information so that can be shared on expected time, being validated at national level and ensuring quality. This information is made accessible through the EEA website becoming the main environmental assessment in terms of environmental management processes and environmental policymaking.

Overall, Eionet forms a mature and flexible network, it offers: cooperation in between national, regional, European and international level; a validation of high quality data by providing environmental assessment and knowledge; and share a common infrastructure, standards and tools that facilitate the information flows. (EEA, 2012)

"Data which countries are obliged to report to the European level are collected and analyzed in a transparent way by the EEA (...) to give an overview of Europe's environment. In this way, it also becomes possible to benchmark the environmental performance of countries. Furthermore, Eionet countries are encouraged to develop their national networks in ways which assure compatibility with EU and international reporting requirements. " (EEA, 2012)

In relation to water issues, Eionet mainly gathers information of:

- The status of the water resources across Europe in terms of quality and quantity
- Considering this status the study of its reaction to the effect of different pressures on the environment.

Data gathered by the country members is obtained on existing monitoring activities so that EIONET is able to provides a *"shared pool of common and timely data and information on the state of, and pressures on, Europe's water (SOE_WISE)* (EEA, 2012)*"* Eionet has been adapted in order to achieve the monitoring requirements established by the WFD. Eionet is important for this report because basically is the place where internal data generated at GEUS will eventually end up, so it is crucial to take into consideration that data will be shared and observed at European level trough Eionet in the future.

2.1.3.3 Water Information System of Europe; WISE

The Water Information System for Europe, WISE, is the main portal of information related to water. It involves a large collection of data gathered by the EU institutions to be used by several stakeholders (WISE). WISE is an entity shaped by collaboration between the European Commission (DG Environment, Joint Research Centre and Eurostat) and the EEA. WISE is a web-based service that can be used by the public, it is important because it has been designed to be a web portal access to water information covering from inland waters to marine. Currently the web-portal appearance is organised in the following sections:

- *"EU water policies (directive, implementation reports and supporting activities…)*
- *Data and themes (reported datasets, interactive maps, statistics, indicators…)*
- *Modelling (now – and forecasting services across Europe…)*
- *Projects and research (inventory for links to recently completed and ongoing water related projects and research activities)"* (WISE)

WISE users are:

- *"EU institutions as well as Member States national, regional and local administrations working in water policy development or implementation"*
- *Professionals working at water field from both sectors private and public, which shared a technical interest on water.*
- *Scientists whose work is based on the water field*
- *General public, within this group is considered those working in private or public entities which are not directly related to water directive but present an interest in water's issues."* (WISE)

WISE enables easy access to reference documents and thematic data, which can be downloaded for analyses by scientists and water professionals, between others. Additionally, WISE serves as main tool for the visualization of information related to water used by the public, by interactive maps, graphs and indicators. (WISE) Since one of the purpose of these Thesis is providing GEUS with resources to deal with the WFD compliance, it is important to remark WISE is a reference for any water related issue. It is particular useful for people to getting to know about european water topics. It played a vital role in this report when researching about european water legistlation.

2.2 Database theory

There are two main reasons of writing a section about database theory in this Master Thesis. First reason is to provide the reader, particularly those ones from GEUS, with background information in database design, so that employees can understand the data model and its components. The second reason is to present the theoretical concepts on which the decisions for the data model are based. Overall this section offers a good complement to the database model

What is a Database? A database (DB) can be intuitively defined as an organized collection of data, more specifically *"A database is a collection of interrelated data items that are managed as a single unit"* (Oppel, 2009)

In this report with the objective of designing a data warehouse in GEUS, the concept of a relational database has been undertaken because is the one suggested for future physical implementation, which is defined as:

"In simplest terms, a relational database is one that presents information in tables with rows and columns. A table is referred to as a relation in the sense that it is a collection of objects of the same type (rows). Data in a table can be related according to common keys or concepts, and the ability to retrieve related data from a table is the basis for the term relational database. A Database Management System (DBMS) handles the way data is stored, maintained, and retrieved. In the case of a relational database, a Relational Database Management System (RDBMS) performs these tasks" (ORACLE)

2.2.1 Date Warehouse and Data Warehouse Architecture

This section, aims at providing information about what's a Data Warehouse and Data Warehouse architecture, this will help the reader to understand how a data warehouse should operate.

In essence, a data warehouse is a database which is mostly used within organizations to lead information from many different departments into data marts or repositories of data, contained in the particular database or data warehouse. (1KeyData, 2015) (Figure 2 Source: Author).

There seems to be many definitions, it could be said, that it's a storage facility that collects information, manages it, and delivers it to different audiences. (Bowman, 2009). However a more comprehensive definition is the one provided by Bill Immon (Immom, 2005) who suggested that a data warehouse is a subject oriented, integrated, time variant and non-volatile collection of data in a support of management's decision making process. It's subject-oriented because it can be used to analyse a particular subject or area. It's integrated because it integrates data from multiple data sources. It's time variant because historical data is kept. Finally, it's non-volatile because data should not be changed once has entered the data warehouse. The other important definition is offered by Kimball, who states that a data warehouse is a copy of transaction data specifically structured for query and analysis (Kimball & Ross, 2011).

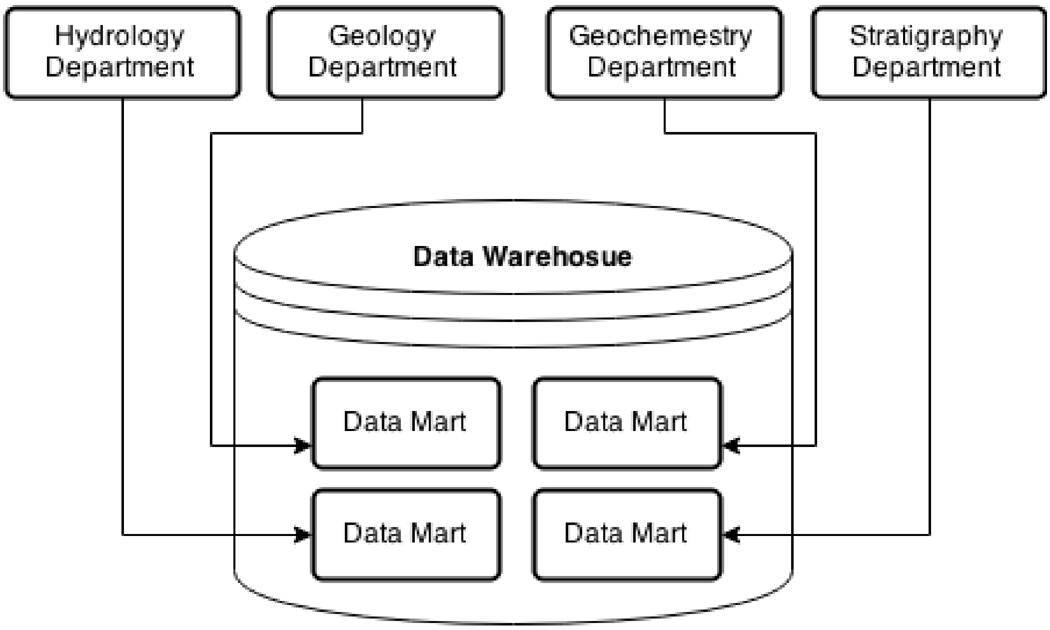


Figure 2 Source: Author

It's the foundation for information management. Among the potential benefits of a well-constructed data warehouse we can point out the improvement of the decision making process and the increase of the overall efficiency. (Oppel, 2010).

Data Warehouse architecture is basically the actual structure of the data warehouse. Obviously this differs from organization to organization, for instance while some may have few data sources, others involve a high number of them. In general and regardless sources and systems used, all data warehouses structures are composed of these layers. (1KeyData, 2015).(Figure 3 Source: Author adapted from

- Data Source Layer
- Data Extraction Layer
- Staging Area
- Extraction, Transformation, Loading Layer (ETL)
- Data Storage Layer

- Data Logic Layer
- Data Presentation Layer
- Metadata Layer
- System Operations Layer

(1KeyData, 2015)

Data source layer is the one that includes the data sources that would feed data into the data warehouse. Obviously data may be in different formats or come from different sources, in the case of GEUS, they can be monitoring networks, Jupiter database, National Water Resources Model etc. The data extraction layer is where data gets pulled from the data source layer into the data warehouse system without major transformation. Staging area layer is where data is storage prior to being transformed into the data warehouse. Within the Extraction, Transform, Loading (ETL) layer data gains value since logic is applied to transform it to an analytical nature. The data storage layer is where data sits. In the data logic layer business rules are stored; these ones will affect how report will look like. The presentation layer refers to the information that the users get, it could be tabular, graphical reports, etc. Furthermore, metadata layer is where information about the data is stored; it serves for data validation, analysis and reporting obligations. Due to the importance of metadata to achieve data quality and integrity at organizations; it will be described in a later section. Finally the system operations layer holds information about how the data warehouse system operates. (1KeyData, 2015)

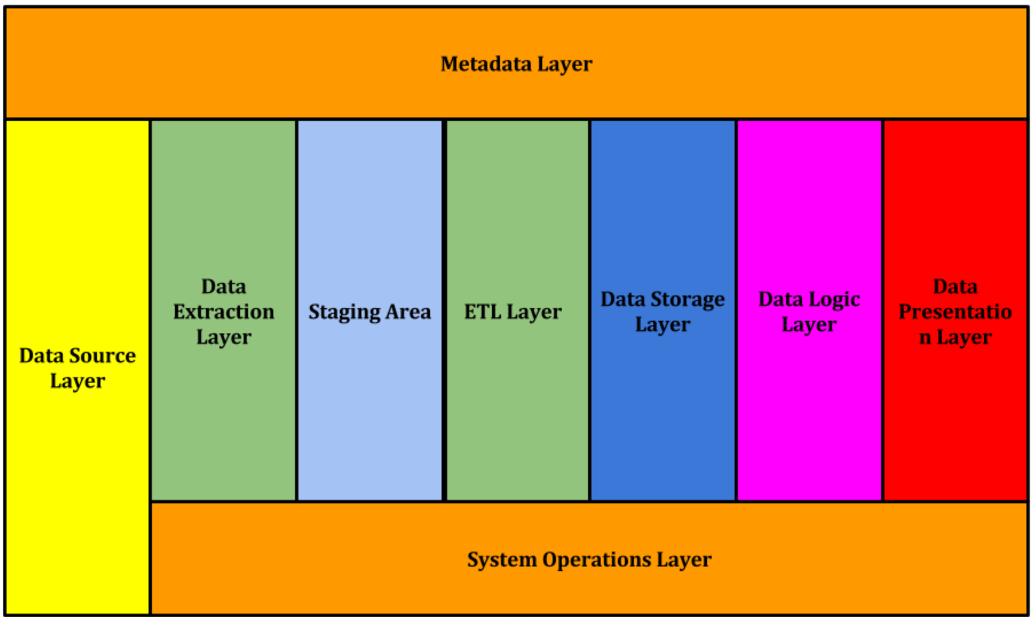


Figure 3 Source: Author adapted from (1KeyData, 2015)

There are several concepts relating to the design of a Data Warehouse that are of particular importance. For instance, dimensional data model, conceptual data model, logical data model, physical data model, metadata, facts and dimensions etc. Some of these are explained in the next sections.

2.2.2 Designing Data Warehouses: Data Models

A data model illustrates how the data within a specific information system is symbolized and accessed, in other words, it organizes data elements and indicates the way in which data elements are related to each other.

Data elements describes real life, people, places, and objects, along with the events occurring between them, therefore a data model represents reality. For instance a book has many pages or a car has just one owner. By using information systems, data and records regarding these real life events are stored, a data model offers then, the proper standard to facilitate and ensure communication among human beings. (Oppel, 2010)

Data models are the main assistance among organizations and the technical people, when defining the requirements for computer system and, deciding the design that meets those requirements. (Oppel, 2010) To sum up,

"A data model is a way finding tool for both business and IT professionals, which uses a set of symbols and text to precisely explain a subset of real information to improve communication within the organization and thereby lead to a more flexible and stable application environment." (Hoberman, 2009)

Finally, a data model particularly defines the structure of data, so they use a specifically data modelling notation, which is usually graphical in form. (McCaleb, 1999)

Role of Data Models

The main duties of the information systems such as databases are based on managing and organizing extensive amount of data, to enhance these duties, data models show the structure, control and integrity features of data stored within a data management system like relational databases. (West & Fowler, 2003)

In addition, the main goal and role of data model is to provide with the correct format and definition of data in other to support the development of information systems. To make this possible, the implementations of standards that will ensure both, meet business needs and data integrity, is essential. (Hoberman, 2009)

In consequence, the two key benefits which make data model relevant to applications when using and exchanging data are, communication and accuracy. A data model pretends to be the channel trough which project team members from different backgrounds and with distinct levels of experience can communicate with one another. Accuracy means that data elements and rules on a data model can be interpreted only one way and are not uncertain or ambiguous (Hoberman, 2009).

Types of Data Models

There are three commonly used data models which are: conceptual, logical and physical. Although there exists many ways to create and expose data models, it is better to focus on the most basic and simplest models, starting from the conceptual model and finishing at the physical model using the same data structure. (Oppel, 2010)A successful data warehouse design process begins with a series of data model: conceptual, logical, physical and dimensional. (Bowman, 2009).

The Conceptual Model

A conceptual data model is a high-level model; this means that it shows the highest level of information that takes part in the whole network. Conceptual models using a technology-independent manner covers main data and their relationships. Within conceptual models, the entities are drawn by rectangles which could be a person, place, thing, event or concept about which an organization save information. (Oppel, 2010).Figure 4

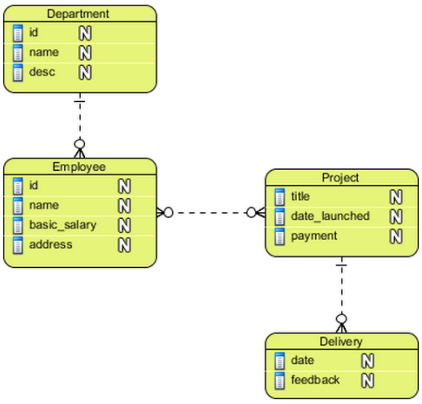


Figure 4 (Visual Paradigm, 2015)

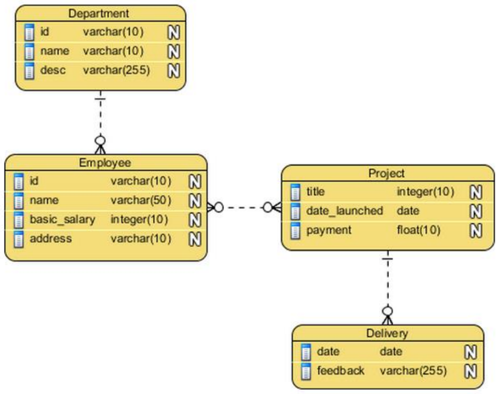


Figure 5 (Visual Paradigm, 2015)

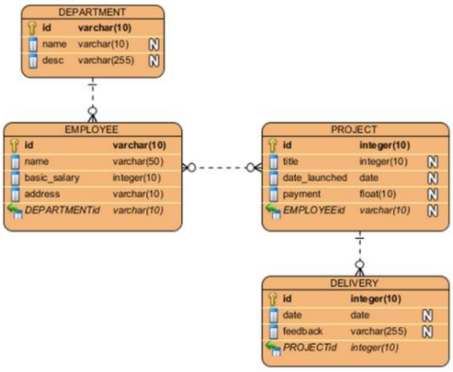


Figure 6 (Visual Paradigm, 2015)

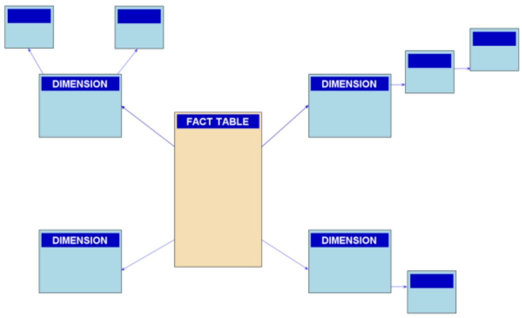


Figure 7 (1KeyData, 2015)

The Logical Model

Basically a logical model describes as much as possible details of data, regardless how it will be later physically implemented in the database. Therefore there exists different ways to draw this diagram. Commonly a logical model or diagram consists on different rectangles where each of them shows an entity, often the name of the entity will appear in the rectangle occupying the top place, and these entities shown have appeared previously in a conceptual model. The list of names shown within each rectangle, and separated from the entity's name by a line, comprises the list of the attributes which give detailed information of the entity they belong to. Each entity has an attribute that represents the unique identifier and it is placed the first in the list of attributes, whereas the ones below are the non-identifying attributes. Finally relationships between entities are drawn using lines which connect the rectangles (Oppel, 2010) Figure 5

The Physical Model

A physical model finally shows the way the diagram will be built in the database, thus the physical model is designed in order to be adapted to the features and constrains or rules, of the selected database management system (DBMS), such as MySQL, Oracle, or Microsoft SQL Server, among others. This last diagram pretends to show how the relational database should be implemented, for this purpose each entity contained previously on the logical model, will now represent at the physical model each of the tables that will be implemented in the database. Moreover, each attribute from the list of each entity corresponds to each column living inside of one specific table in the future database. Finally a unique identifier for each entity is identified and stablished, which as it will be explained in later sections will help to identify relationships between tables. (Oppel, 2010)Figure 6

The dimensional data model

It consists on detailed data tables, called fact tables, and dimension tables. Fact tables are surrounded by dimension tables creating a star like schema. These fact tables usually contain measures of interest that are generated out of data within the dimension tables. In other words, the data in the dimensions tables is necessary to compose a fact. In data warehousing design there are two commonly used approaches, snow flake and star flake. The main difference is that in snowflake schemas, dimensions can have their own dimensions that represent a lower level of granularity or information. On the other hand, star flake schemas, dimensions are not allowed to have a lower level of granularity, all the dimension tables will be at the same level. (Oppel, 2010)Figure 7

Relational model components: Conceptual and Logical model components

Before start reading this section, be aware that out there exists many ways of representing and drawing relational models, therefore do not be overwhelmed when discovering after this report other ways that, even though they might seem similar are not. In terms of data modelling, it is good to be open minded, notation and mapping can be slightly different, it is smarter to consider the pros and cons having in mind the main interests and the resources available when designing a database. Finally, in this report the data model designed was chosen considering the report and organization needs.

Entities

As mentioned, an entity (or entity class) can be a person, object or event on which data are collected. Entities, in essence, are real-world things on which there are a special interest to gather and store data in a database. It can be considered as an entity almost anything that can be named with a noun. Nevertheless, a model designer must limit the design to entities of interest to the people who will use the database which will be built from the model, in order to avoid designing a model that covers everything on the planet. Last but not least, consider that each entity within a conceptual model display the entire class for that entity (Oppel, 2010)

Entity notation, naming and definition

An entity within a conceptual model is mapped as a rectangle which contains the entity's name. Considering that it is expected organization users to understand data models, entity names should be defined from the organization names for the objects that entities represent. By acting this way, entity names will make sense to users. Bear in mind that entities are transformed into tables inside the final relational database, so it is at the conceptual model the best time to get them right. Finally it is good to consider adding to each entity a written definition of its use in business terms. This task will help to success when defining the relationships among entities. (Oppel, 2010)

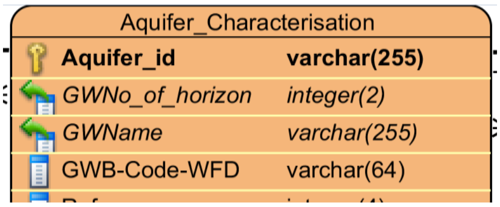


Figure 8 Source: Author

Attributes

An attribute is basically a unit fact that somehow describes an entity. For example: size, shape, colour, location, quantity, etc. Attributes are the smallest unit of data in databases, which means that it can't be divided in smaller parts. This rule should always be followed in the logical model because these attributes in the logical model will be the columns in the physical model (Oppel, 2010)

Attribute notations and Attribute notation for keys.

In both, conceptual and logical models attributes are represented as names inside the rectangles that represent the entity. The attributes that represent a unique characteristic of the entity are called Primary Key (PK) attribute. In other words, when an attribute is consider a primary key, is the one that provides an exclusive value for each record in the entity. It can be a normal attribute that is guarantee to be unique or can be composed of multiple attributes; in the latter case the records resulting from the combination of the multiple attributes must be unique. In logical models, see Figure 5, when the relationships between entities are established, the primary key attribute from the parent entity will appear as a foreign key (FK) within the child entity. (Oppel, 2010)(Figure 8 Source: Author)

Attribute Naming

In logical models attributes are important because the columns in the physical model will be made of the attributes. Names should effectively address the content of the attribute so that attributes can be easily understood by both, users and developers. If names aren't properly stated designers could use the columns for different purposes.

Relationships

Relationships is how we call to the association between entities, they are in charge of holding together the database. In data models relationships are represented by lines linking entities. The end of the line represents the maximum cardinality, which is *"the maximum number of instances of one entity that can be associated with the entity on the opposite end of the line. The maximum cardinality may be one, or many (the line has a crow's foot on the end). Just short of the end of the lines is another symbol that shown the minimum cardinality, which is the minimum number of instances of one entity that can be associated with the entity on the opposite end of the line. The minimum cardinality may be zero, denoted with a circle drawn on the line, or one, denoted with a short perpendicular line or tick mark drawn across the relationship line. Many data modelers use two perpendicular lines to mean one and only one"* (Oppel, 2010) It is easier to think about the associations in one direction and then, change the perspective to the opposite direction

Types of relationships and symbols

Considering the importance of relationships in order to design a successful data model, it proceeds now to explain briefly the most common of these ones. Figure 9

One-to-one Relationships In this kind of association an instance or record of one entity can at most be associated with one instance of the other entity and vice versa (Oppel, 2010)

One-to-Many Relationships: This is an association between two entities where any instance of the first entity may be associated with one or more instances of the second, and any instance of the second may be associated with at most one instance of the first. This association is commonly used and they are of vital importance in relational database models (Oppel, 2010)

Many-to-Many Relationships :Many to many relationship is an association among two entities where any instance of the first entity may be associated with zero, one, or more instances of the second and vice versa. (Oppel, 2010)

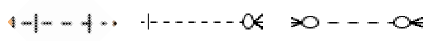


Figure 9 (Visual Paradigm, 2015)

Physical model components

The creation of the physical model serves to the DBMS to carry out the final implementation of the relational database. Physical components of a physical model mainly consist on physical storage and data types descriptions (file or tablespace name, size information and storage location) of each particular database entity as they are mapped from the logical model (Oppel, 2010).

In an ideal situation, physical implementation details should not appear on the logical model, however, some data model tools support a combined logical/physical data model which for time efficiency, components from the physical model can be shown on the final logical model (Oppel, 2010).

Although the main goal here is to present a conceptual and logical model (as they can be considered as a unique delivery) of data existing and required at GEUS, some physical components has been included during the process, as they were already known, leading to a better comprehension of data within the model and also reducing future workload. Hence, the physical component that have been used in the model in this report displayed are briefly explained below

Tables

Tables are considered as the primary unit of storage, and they are formed by rows and columns *"Each row corresponds to one occurrence of the entity that the table represents, and each column corresponds to one attribute for that entity."* (Oppel, 2010). These attributes are shown in the logical model as rows within the entities.

Data types

Each column, which represents an attribute, is assigned with a unique name and data type. This data type is defined as *"a category for the format of a particular column."* (Oppel, 2010). (Figure 6. The most commonly used are:

- Fixed length character: CHAR
- Variable length character: VARCHAR
- Integer: INTEGER
- Date: DATE

However different RDBMS like, MySQL, ORACLE, MicrosoftSQL Server, etc use different names for data types. (Oppel, 2009)

Constraints

Constraints can be defined as rules that are applied to a database object, and restrict the data values for the database object. Relationships and business rules from the logical model are implemented through constraints at the physical model.

Primary key constraints refer to the ones related to unique identifiers *"a unique identifier in the conceptual and logical model is implemented as primary key in the physical"* (Oppel, 2010).

Integrity Constraints

The purpose of an integrity constraint is to ensure the accuracy of data in the database; most of these constraints are based on the defined business rules. These kinds of constraints are invoked directly from the specific DBMS so there is no way to avoid them when a user connects to the database (Oppel, 2010).

The only integrity constraint used at GEUS' model here presented is: Not Null constraints

When building tables into a database, there is the option to specify whether a null value for a column is allowed. In this case, the "NOT NULL" constraint is not symbolized whereas a value that can be "NULL or NULLABLE" is represented by a white letter "N" located next to the attribute in the logical model, see Figure 5. This "N" indicates for the physical model that *"the value for that column in that row is unknown. A null value is not the same as a blank, an empty string, or a zero-it is indeed a special code (like "N") that has no other meaning in the database."* (Oppel, 2010).

Data Modelling: Diagramming techniques used

As explained in previous sections the process of modelling data mainly takes part during the development of the logical stage design. Nowadays, there exist several alternatives for drawing data models, the following two which are exposed below have been used to achieve the goal of this Master Thesis regarding resources and software provided (Oppel, 2010).

Snowflake dimensional model; this diagram was previously explain at (The dimensional data model), despite of not being crucial to reach the goal here pursuit, it serves to easily identify and better understand the hierarchy of data at GEUS. Basically this diagram will try to show at which level objects are found, and which ones are the main relations between them.

Entity Relationship Diagram (ERD); *"An ERD is a diagram that visually represents entities, attributes and relationships"* (Oppel, 2010). Data models are usually drawn using this diagram. The most important feature of an ERD is that non-technical people can understand it while it provides valuable information for technical staff (Oppel, 2010)

The importance of Data Modelling

This section aims at explaining why data modelling is actually important. The first reason, and probably most relevant is that, experience has shown that the more time spent in early stages gathering requirements and modelling, the better the implementation will occur, and therefore, the overall cost of the project will be lower. Here it is briefly explained some arguments supporting data modelling. (Oppel, 2010)

Documentation of business rules, organizations usually have constraints that define the way the organization is run, these rules are called business rules. Data models offers the option of illustrating business rules in such a way that, people in charge of the actual implementation of the database can dispose of the necessary information.

Visualization is another important argument, diagrams are often the best way for communicating data structures to a diverse audience with different background, particularly to those who are not familiar with databases but they have relation with the data generation.

Additionally, when models are depicted on diagrams, the data structure can be shared, and thus, it can be used in the future as a reference and as an excellent foundations for potential further expansions.

Also, it promotes common standards; data models are useful to identify common entities across the organization. Modellers can identify entities located in different departments and promote a common standard. Furthermore, data models can be a reference for new databases for naming entities and attributes

Last but not least, data integration is another benefit of data models, when data comes from different sources data models are a good way to integrate data into the common structures.

Overall, the best value that data models can provide is that they can be easily understood by a wide range of audience. Technical staff like database administrators can easily understand it, at the same time, the best models can be also understood by organization users.

(Oppel, 2010)

2.2.3 Metadata

First of all, it is important to get through the definition of metadata; the most intuitive definition is the one that defines it as *"Data about data"* (Beal, 2007). It might seem a simple statement but it captures the essence of metadata. More specifically, metadata is structured information that describes, explains and locates an information resource, making easier to retrieve, use or manage it. (NISO, 2004). For instance, metadata can include particular information about some monitoring network records such as, the date that new data was introduced into the database, the user that might have introduced this record, etc. (Layton, 2013). It seems to be obvious that the usefulness of the data increases by adding quality metadata to it.

There are three main types of metadata

- Descriptive metadata describes a resource for purpose like identification. It may include title, abstract, author etc.
- Structural Metadata refers how objects are put together. How pages are ordered could be an example.
- Administrative metadata offers information to help manage a resource, like when and how it was created, as well as technical information.
(NISO, 2004)

The usefulness of metadata will depend on its actual quality. Raw data will be less useful or valuable if the metadata isn't accurate, this would have consequences in the resulting analysis. The importance of metadata for an organization like GEUS is extreme. Metadata could contribute to identify the validity and reliability of data not just for internal purposes but to fulfil reporting obligations. From a database point of view, metadata is important in order to create more accurate database search and retrieval of information stored in a data warehouse. (Beal, 2007). From a European Level point of view, metadata can contribute to the interoperability of data among European countries.

Who creates metadata? That is another question to be answered when managing a large amount of data. Then answer to this question varies by discipline, organization, data being described, the available tools and the expected outcome, but it is often a cooperative effort which involves different parts (NISO, 2004). Basic structural and administrative metadata is created by automated process or those in charge of generating the digital objects. For descriptive metadata the originator of the resource usually provides the information, this is often the case of scientific datasets where the originators have a particular understanding of the field of knowledge. In some cases a combination of researchers and information professionals, is the best solution, researchers can provide the scientific knowledge and IT can check for consistency (NISO, 2004) (Layton, 2013).

Last but not least, metadata needs to be stored somewhere, there seems to be two main options, first one is putting metadata in a central location for all data, this implies gathering metadata and store it, often in a database. The main disadvantage of this way is the interaction between the metadata and files. For instance, what happens if a files gest updated, how does that synchronize with the metadata? In this case mechanisms to update the centralized metadata server are required. The second option, is storing metadata with the data so that if data moves, metadata moves with it. However, there are some disadvantages; for example, searching metadata effectively is more difficult. To sum up, the importance of where metadata is stored seems to be crucial in trying to make data useful and reliable. (Layton, 2013) This is definitely something that GEUS should consciously consider seriously, not just for WFD purposes but for internal data consistency. It will require particular attention from the organization and expertise in the field of computer science.

2.2.4 Visual Paradigm

Visual Paradigm is a software that provides a set of tools for software development like software design, and programming. It also provides a wide range of diagrams, such as business process diagrams and entity relationship diagrams. For the

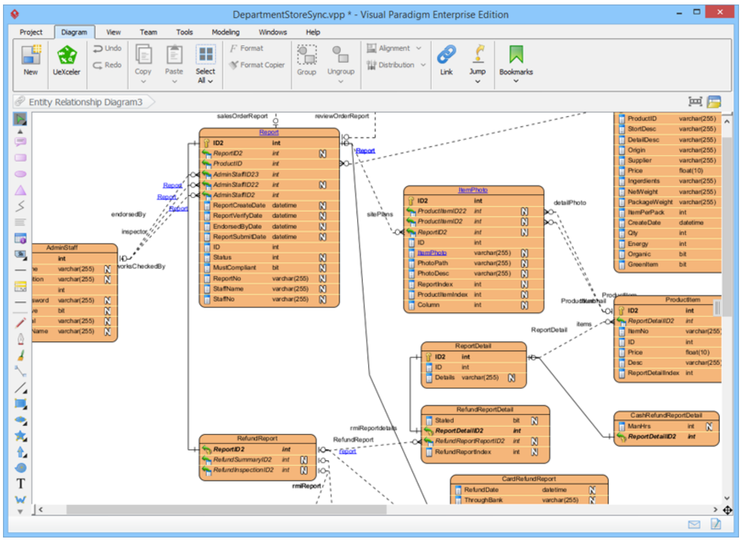


Figure 10 (Visual Paradigm, 2015)

development of this master thesis, Visual Paradigm 12.0 for enterprises has been used. Particularly, the database design, entity relationship diagram toolset was of extreme importance. This tool allowed representing graphically data elements, tables and relationships among them. These types of diagrams are helpful when it comes to modelling databases that involve several tables and data elements. By using this software to model the database, it is tried to present useful information for the future actual implementation of the database. At last, the toolset covers, conceptual, logical and physical models, which are the three levels of abstractions of database modelling. (Visual Paradigm, 2015).Figure 10

2.3 Hydrology concepts

In this subsection some hydrological concepts are defined so that the reader understands better the nature of the data. It is known that internal users from GEUS are generally familiar with hydrology concepts; however, external readers might find it useful. It is common to confuse the concepts of, groundwater body and aquifer, this mainly occurs due to the fact that, humans mostly interact with groundwater in the term of aquifers (Smith & Jhon, 2005)

Groundwater bodies comprise the water that seeps down through the soil or rocks and thus, into a layer of saturation. A layer of saturation is the area situated underground, which contains water under pressure of more than one atmosphere. It implies the main source of well water; this means that if a hole is dug through a layer of saturation, the water will fill it reaching the top of the layer. Nevertheless, there are areas considered of saturations where it cannot be found enough groundwater to support a well or spring, these ones are discarded as aquifers (Smith & Jhon, 2005; NCDENR).

Only if a useful quantity of water supplies the surface becoming, a well or a spring, the saturation layer can be considered an aquifer. Aquifers are often illustrated as big lakes or river situated under the ground because, aquifers maintain their water in the same manner that a river or a lake does, even though aquifers do not occupy an empty cavity in the rock. Water inside an aquifer plays a similar role that the water coming out from a hose watering a garden, flowing through the porous material (Smith & Jhon, 2005; NCDENR).Figure 11

Finally, since river basin districts play an important role in the EU regulation, it might be important to actually learn its definition. River basin is "*The area of land that is drained by a specific river and its tributaries*" (Smith & Jhon, 2005). Due to its transboundary limits, it is also considered by the WFD as "*the natural geographical and hydrological unit*" therefore comprising the hydrological boundaries for the development of the River Basin Management plans. (EU, 2014)

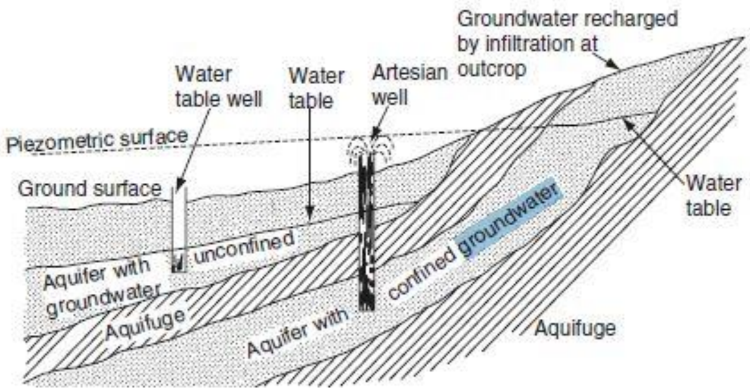


Figure 11 (Smith & Jhon, 2005)

3 Methodology

In general, the project management for this master thesis report could be defined as some type of Agile methodology. When this project began the main target was to apply the WFD to GEUS as organization, in such a way that the obligations depicted on this regulation were matched, while organizing groundwater related data. In other words, the main goal was to provide the organization with the necessary knowledge to cover European regulations regarding to groundwater taking in consideration internal data status at GEUS.

Additionally, it could also be said that GEUS fellow workers missed deep knowledge about the actual implications of the directive for the organization in terms of data, tasks, and previous attempts to meet WFD.

Besides, to the regulation framework understanding, and the technical nature of the data treated in this project, hydrological information about aquifers and groundwater bodies, it was crucial to understand the organizational process at GEUS. Getting to know how hydrological data is generated, where it is stored, how it is stored, and departments involved, etc. was of vital importance.

Therefore, the project began with a defined idea about the final desired output, a database model that would put in order groundwater related data and WFD needs, together with a data standardization guide which would facilitate the implementation of the database. Nevertheless, it was impossible to prevent in advance every necessary step and issues that might come up during the process. Hence, instead of working in a linear way where the steps are previously defined and planned, a more flexible methodology has been used. This permitted to review every step along the process in order to identify, errors and more efficient process, and if needed, re-design some of them, for instance relationships between entities. Thus, regular scheduled meetings with the person in charge at GEUS Lars Troldborg, provided a continuous exchange of insights about the process, these meetings were an important source of information regarding internal processes about data generation and storage. During the meetings that took place almost on daily basis, next steps in the process where set. Thesis supervisor from the university Thomas Balstrøm also developed such an important role, thanks to his experience, he provided important insights and suggestion, in finding the right balance between academic goals, GEUS needs and author's knowledge. There were regular meetings which were a source of relevant knowledge about the best way of reaching the desired output.

Regarding the actual workflow, this one could be described in Figure 12 Source: Author.

- **Data model design**: Based on EEA documents and database theory
- **Meetings**: In meetings with GEUS supervisor, feasibility based on GEUS current status was reviewed. Thesis supervisor, provided insights regarding the academic and learning goals so that the thesis was not too biased toward the organization.
- **Insights for modifications**: Meetings were source of modifications to be done to the data model design.

At last, this iterative workflow was repeated until a satisfactory data model design was achieved. Finally, is important to remark that, this Master Thesis report was written once the design of the data model was done and not along the process.

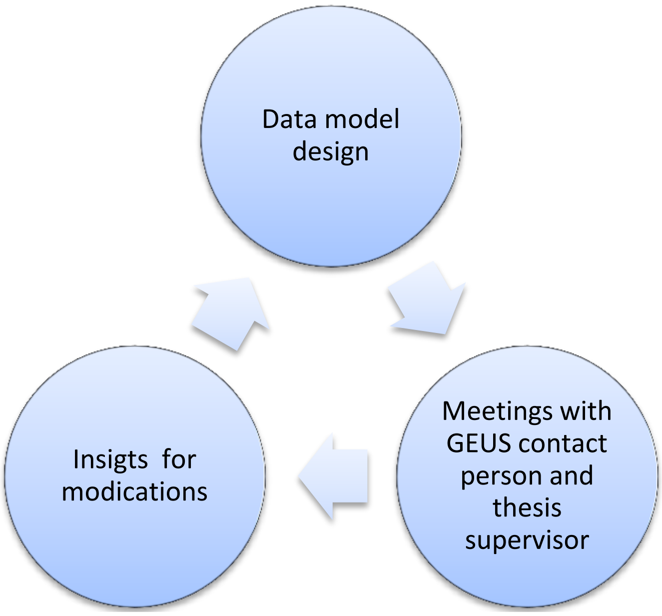


Figure 12 Source: Author

4 Implementation

This section explains the solutions taken to give responses to the main problem here presented according to WFD requirements and GEUS needs. Reading through, it is given a more specific view of data required by EEA at the WFD within some tables (Table 2). Looking into the annexes: Annex I: Data standardization guide, and Annex II: Code list , a data standardization guide is also offered; which contains data standardization in order to better comprehend WFD data, this is the reason why inside this implementation chapter is included a sub-section describing why a data standardization guide has been written as one of the solution to GEUS problem.

Last but not least, a dimensional data model and an ERD were drawn as the main part of the problem-solving process.

4.1 Actual WFD requirements

The importance of the objectives of the EEA and the EU commission within the WFD has been remarked, and also, the relevance of a data warehouse at GEUS, which will enhance reporting required deliveries concerning groundwater quality within deadline. However, before going through the explanation of the data model designed, in in this section it is presented the actual data requirements in the Groundwater Quality dataset of the WFD. In other words, this is the data that GEUS is required to report, and therefore the data that should be standardized and/or processed. It could also be considered as the target data of the data model here presented. The asterisk in front of some of the element name indicates that the element is mandatory in that particular table. “C” sing marks an element as common.

Tables	Data elements
GW-Body_Characterisation: Detailed information on the physical characteristics and proxy pressures relating to groundwater bodies are requested from EEA Member Countries on an annual basis.	Groundwater Body Code - EIONET (GWB-Code-EIONET) C * Groundwater Body Code - WFD (GWB-Code-WFD) C * Name (GWName) C * Reference year (Reference_year) C * Number of Horizon (GWNo_of_horizon) C * National Code (National_code) River Basin District Code (RBDcode) C River Basin District Name (RBDname) C Groundwater Body Area (GWArea) Minimum Depth to Groundwater Body (Depth_to_groundwater_min) Mean Depth to Groundwater Body (Depth_to_groundwater_mean) Maximum Depth to Groundwater Body (Depth_to_groundwater_max) Depth to Groundwater Body - Period (Depth_to_GW_period) Minimum Thickness (Thickness_min) Mean Thickness (Thickness_mean) Maximum Thickness (Thickness_max) Maximum Width (Maximum_width) Maximum Length (Maximum_length) Petrographic Description (Petrographic_descript) Stratigraphy (Stratigraphy) Main Aquifer Type (Main_aquifer_type) Overlying Strata (Overlying_Strata) Confined (Confined) Associated Aquatic Ecosystems (associated_aquatic_ecosystems) Description of the Associated Aquatic Ecosystems (Associated_Aquatic_Ecosystems_Purpose) Main Infrastructures (main_infrastructures) Description of the Main Infrastructures (Main_Infrastructures_Purpose) Minimum Hydraulic Conductivity (Hydraulic_conductivity_min) Mean Hydraulic Conductivity (Hydraulic_conductivity_mean) Maximum Hydraulic Conductivity (Hydraulic_conductivity_max) Minimum Annual Groundwater Level Amplitude (Annual_gw_level_amplitude_min) Mean Annual Groundwater Level Amplitude (Annual_gw_level_amplitude_mean) Maximum Annual Groundwater Level Amplitude (Annual_gw_level_amplitude_max) Annual Groundwater Level Amplitude - Period (Annual_GW_Level_Amplitude_period) Minimum Annual Precipitation (Annual_precipitation_min) Mean Annual Precipitation

	(Annual_precipitation_mean) Maximum Annual Precipitation (Annual_precipitation_max) Annual Precipitation - Period (Annual_Precipitation_period) Water Abstractions (water_abstraction) Purpose for the Water Abstractions (water_abstraction_purpose) Artificial Recharge (artificial_recharge) Purpose for Artificial Recharge (artificial_recharge_purpose) Main Recharge Source (Main_recharge_source) Remarks (Remarks) C
StationsGroundwater Detailed information on the physical characteristics of the sampling sites of the groundwater bodies are requested from EEA Member Countries on an annual basis.	* Station ID (GWStation_ID) C (PK) Groundwater Body Code - EIONET (GWB-Code-EIONET) C * Groundwater Body Code - WFD (GWB-Code-WFD) C National Station Code (National_station_code) National Station Name (NationalStationName) C WFD station (WFDstation) C WFD Station Code (WFD_EU_CD) C * Longitude (Longitude) C * Latitude (Latitude) C Type of Use (Type_of_use) Well or Spring (well_or_spring) Remarks (Remarks) C
NutrientsGW_Agg Data on Ammonium, Dissolved Oxiygen, Nitrate and Nitrite in groundwater bodies, including information on the number and type of sampling sites, the monitoring frequency, the number of sampling sites for each range of concentration values and statistics, are requested from EEA Member Countries on an annual basis.	Groundwater Body Code - EIONET (GWB-Code-EIONET) C * Groundwater Body Code - WFD (GWB-Code-WFD) C * Determinand Code - agg (DeterminandCode-agg) (PK) * Year (Year) C (PK) * Number of Sampling Sites (GWNumberOfSites) C Drinking Water Sites (GWDrinking_Water_Sites) C Industrial Sites (GWIndustrial_Sites) C Surveillance Sites (GWSurveillance_Sites) C Other Sites (GWOther_Sites) C Average Sampling Frequency (GWSamplingFrequency) C * Number of Samples (NumberOfSamples) C Limit of Quantification (LimitOfQuantification) C Minimum (Minimum) C * Mean (Mean) C Maximum (Maximum) C Median (Median) C Class 1 (Class_1) Class 2 (Class_2) Class 3 (Class_3)

	Class 4 (Class_4) Class 5 (Class_5) Remarks (Remarks) C
NutrientsGW_Disagg Chemical quality data on the concentration of Nutrients and general physic-chemical determinands (especially Nitrate, Nitrite, Ammonium and Dissolved Oxygen) in groundwater are requested from EEA Member Countries on an annual basis.	* Station ID (GWStation_ID) C (PK) * Determinand Code (DeterminandCode) (PK) * Value (Value) * Date (Date) (PK) Remarks (Remarks) C
HazSubstGW_Disagg Chemical quality data on the concentrations of hazardous substances and other chemical determinands in groundwater are requested from EEA Member Countries on an annual basis.	Element name * Station ID (GWStation_ID) C (PK) * Determinand Code (DeterminandCode) (PK) * Value (Value) * Date (Date) (PK) Remarks (Remarks) C
Saltwater-Intrusion Any occurrences of saltwater intrusion caused by groundwater over-exploitation are requested from EEA Member Countries on an annual basis.	* SALT-Code (SALT-Code) C (PK) Groundwater Body Code - EIONET (GWB-Code-EIONET) C * Groundwater Body Code - WFD (GWB-Code-WFD) C Reference year (Reference_year) C * Longitude (Longitude) C * Latitude (Latitude) C Name of Area Affected by Saltwater Intrusion (SaltName) Area of Saltwater Intrusion (SaltArea) Cause (Cause) Since (Since) Remarks (Remarks) C
GW-Body_GIS Attribute data referring to geographical data sets of groundwater bodies are requested from EEA Member Countries on an annual basis.	Groundwater Body Code - EIONET (GWB-Code-EIONET) C * Groundwater Body Code - WFD (GWB-Code-WFD) C Number of Horizon (GWNo_of_horizon) C Scale (Scale) C
Saltwater-Intrusion_GIS Attribute data referring to digital maps of occurrences of saltwater intrusion affecting groundwater bodies are requested from EEA Member Countries on an annual basis.	* SALT-Code (SALT-Code) C (PK) Scale (Scale) C

Table 2 (EEA, 2013a)

4.2 Why an organization data standardization guide?

In this section instead of describing the actual standardization guide, the reasons behind the decision of including the so called guide in this report are presented. See Annex I: Data standardization guide.

The first reason is that during the internship at GEUS, where the previous project was developed (Gallego, 2015), it was noticed that there was a clear need of communication. In other words, even though the organization knows that it is required to report data to the EEA, when it comes to know what data and in which format there seems to be a lack of knowledge. Obviously, this data standardization guide is also offered by the EEA trough the Reportnet. However, previous experience has evidenced that looking for specific information within EEA resources could become complex and doubts might come up, due to the different reporting obligations in many fields and different versions of the Reportnet tools. The standardization guide would palliate these doubts so that GEUS can easily find the format.

The second reason is that a data model which ultimately aims at fulfilling WFD requirements would not make sense if it is not accompanied by the actual specification of the European standards. For instance, (see Table 2) take as example the data element Class 1 (Class_1), if this is not accompanied with extra information it becomes useless, thus if you read it together "The number of sampling sites with average concentration of disaggregated values of:Ammonium ≤ 0.1 mg/l; Dissolved Oxygen ≤ 2 mg/l, Nitrate ≤ 10 mg/l, Nitrite ≤ 0.01 mg/l within the aggregation period" it acquires much more sense for the reader.

From a theoretical point of view, there are also arguments supporting this decision. A data dictionary is defined as a collection of descriptions of the data objects for the benefit of the users of the data. (Harris, 2011), This data dictionary or so called in this report, data standardization guide, can be consulted to understand the nature of the data element, what values it may contain and its meaning.

It helps in the process of improving data quality by sharing agreed-upon definitions for the data elements and their descriptions. It contributes to make data more trustful since developers can easy access to approved definitions. Also, describing each data elements reduces the creation of redundant elements because it allows controlling the addition of new data elements that might already exist. Additionally a data standardization guide promotes sharing information, in this case WFD data dictionary helps to spread information across Europe. Moreover, it foster easier data analysis, users can use the data dictionary for report generation, in the data dictionary users can easily see data elements and descriptions so that they see what can they dispose of for generating reports. (Harris, 2011).

4.3 Database modelling

In compliance with the definition given about data modelling, this next sub-section will illustrates the process of drawing a data model, defining roles of the main entities involved, their attributes or properties, and which are the key items that enable one relationship or another.

As described before, this process is mainly divided in three basic models, conceptual, logical and physical, which are designed prior to the beginning of the physical construction of the data warehouse.

This Master Thesis is based on modelling the conceptual and logical models, for this purpose both are used, a snowflake dimensional model and an ERD; aiming to give the best perspective or outlook of how the data warehouse can be implemented in the near future.

4.3.1 Snowflake: Dimensional Data Model

Aiming to easily identify and recognize the level of data and information gather into the organization (GEUS) a dimensional model was first edited.

A dimensional model based on snowflake concepts consists on central fact tables, containing some relevant measurements which in this case mainly include physical characteristics of the aquifers and so, groundwater bodies; proxy pressures submitted on groundwater area, as well as chemical quality data and hazardous substances in groundwater; also primary information stored in Jupiter database at GEUS. These measurements are computed by data led from their respective dimension tables. Figure 14 Source Author

Fact tables found at this model in a green colour are:

Aquifer_Characterisation
GW-Body_Characterisation
JupiterDB_link

The fact tables named above, in order to hold a set of real events like the measurements shown, depend on the information gathered in their respective dimension and sub-dimension tables, establishing a hierarchy which means that, data go from the lowest to higher level of information making possible the situation within the fact tables.

The dimension and sub-dimension tables, which correspond to the remaining blue tables from the model, are further normalized; this implies that data is reorganized meeting some essential requirements: all data is stored in only one place so that there is not redundancy; and all related data items are stored together, giving a logical sense to data dependencies.

Noticed that in this snowflake (Figure 14 Source Author) the intention is not to go into details of the information, and the focus is on the visualization of the hierarchy of data, making it easier to understand by any user, that is why only primary and foreign keys that will enable future relationships of the ERD are shown.

Snowflake dimensional model works with the normalization process of data, relevant to occupy the minor space in the system and to increase performance (Jansen)

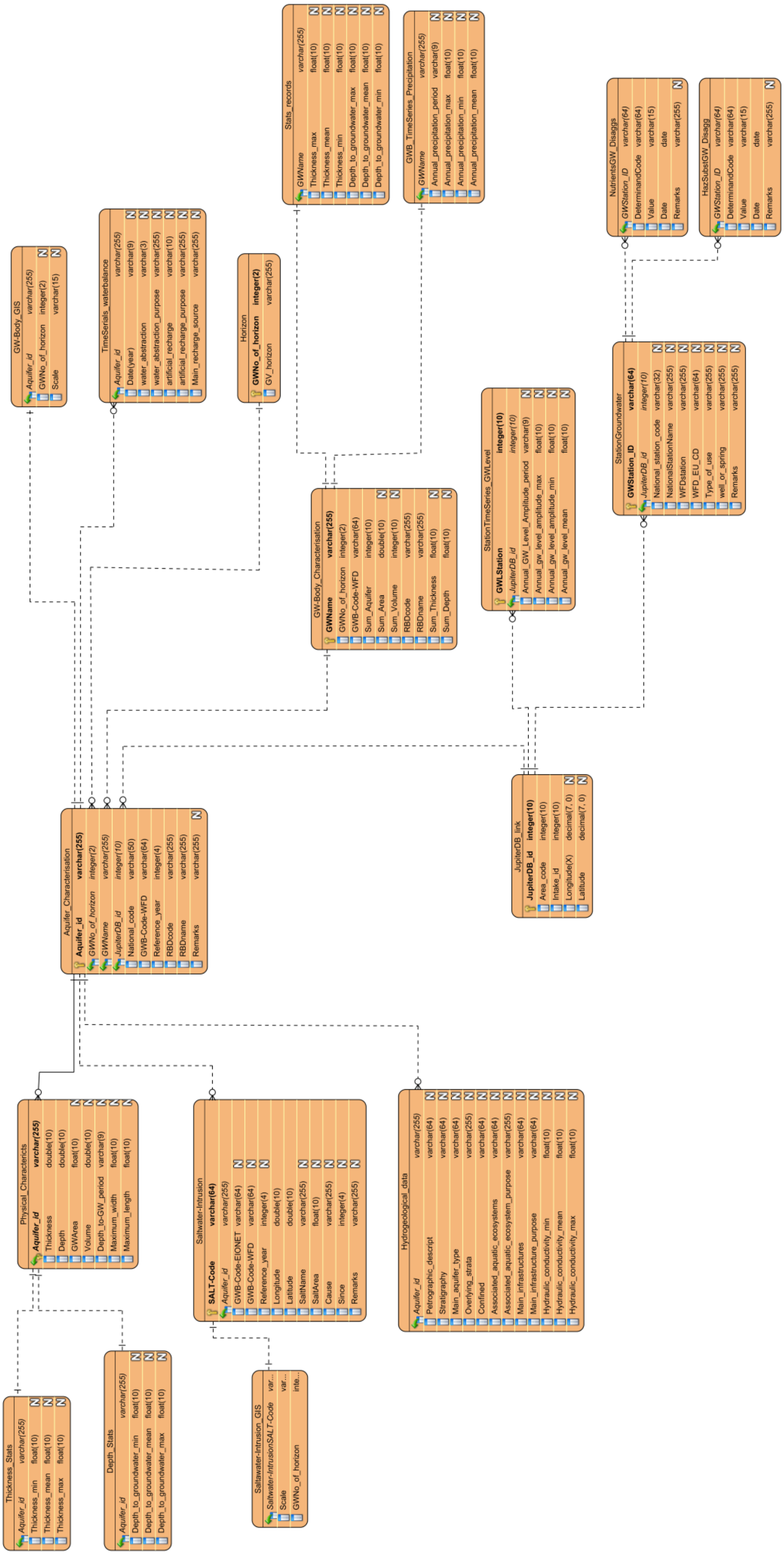


Figure 13 Source Author

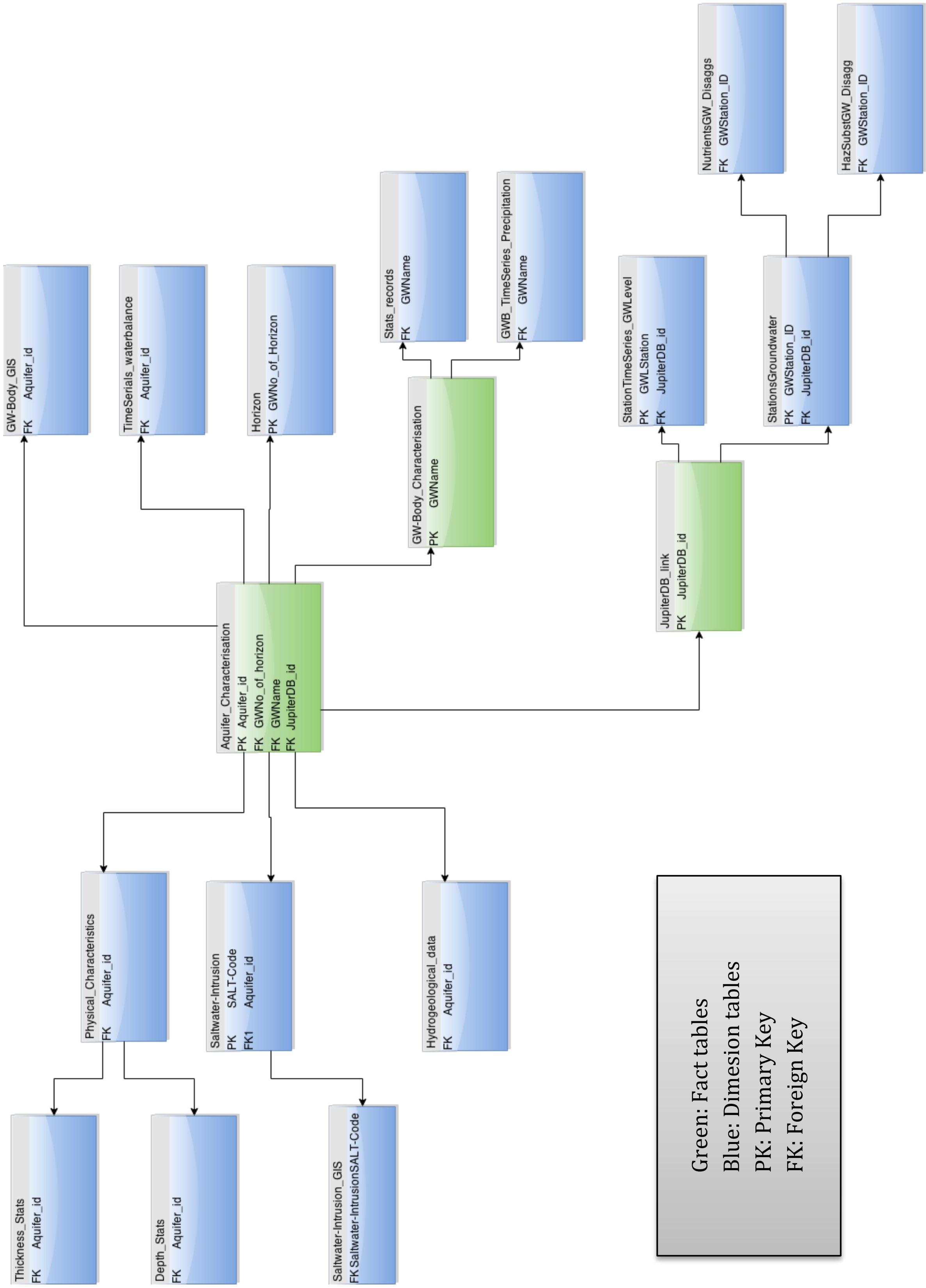


Figure 14 Source Author

4.3.2 ERD: Entities, relationships and cardinality

To promote a well understanding of the implementation undertaken in this Master Thesis, the description regarding process of drawing the data model will be based on the use of the granularity concept (as shown in the snowflake, Figure 14 Source Author) this means that explanation starts from the higher level of information to drill down into detail .

With the intention to recall the main concepts of an ERD, before starting the following information might be helpful in understanding the process taken during the drawing of the model. See Figure 13 Source Author

Keep in mind that the common elements of an ERD format are:

Entities which are drawn as rectangles or boxes
Relationships are depicted as lines
Symbols at the end or closer to the end of lines indicates cardinality (Figure 9)

Furthermore, if you take a quick look to the entire diagram or model, you will easily identify some of the physical components that were included during the drawing process. These ones, are placed next to each attribute in each raw in every entity, and as described in earlier sections (Physical model componentsThe Physical Model), they refer to the data type, in other words, they define how values or instances of attributes should be written or introduced in the tables of the future data warehouse. These are rules and constrains that must be considered by the person in charge of the implementation and thus build the database or data warehouse.

Before start reading through the following figures' interpretation, be mindful that connections made by lines between entities allow going more deeply into data, going from the higher level of information to the lowest. This depends on needs and interests, and it's possible not only always when entities are connected through relationships, but also when detailed data for attributes exist or is available to users within the specific entity.

To begin with, two main datasets can be distinguished:

Firstly, Groundwater Quality dataset, which collect data, required by WFD, and which comprises the following entities displayed in the ERD (Figure 13 Source Author)

- GW_Body_Characterization
- StationGroundwater
- NutrientsGW_Dissag
- HazsubstGW_Disagg
- Saltwater-Intrusion
- Saltwater-Intrusion_GIS
- GW-Body_GIS

Notice that, according to theory prior provided there should be eight tables forming the entities of this ERD, the absence of one of them (NutrientsGW_Agg) on the list above is a result of the statement stablished by WFD which says that Disaggregated data of Nutrients is preferable in any case if available, because GEUS disposes of this disaggregated data then the aggregated data is not needed, as well as its entity in this model.

The second relevant dataset for the performance of this ERD or data model is the one which makes possible the link with Jupiter database, which gather data generated and stored at GEUS and needed to reach WFD purposes. This dataset is composed by the following entities:

JupiterDB_link
StationTimeSeries_GWLevel
GWB_TimeSeries_Precipitation

Be aware that, the manner in which the latter dataset covering data coming from GEUS is presented; depends on how this data has been stored and provided.

The remaining entities comprise a lower level of information, which derive from the main entities. The information contained within this remaining entities are more likely to suffer changes and updates through time, thus, to ensure data integrity and avoid redundancy as much as possible they have been placed separated in several tables at a lower level of information and joined with the appropriate relationship to the entity they belong to.

Next, to facilitate comprehension, the ERD presented (Figure 13 Source Author) will be divided into eight figures. Each figure will represent the main relationships involved considering different levels of information.

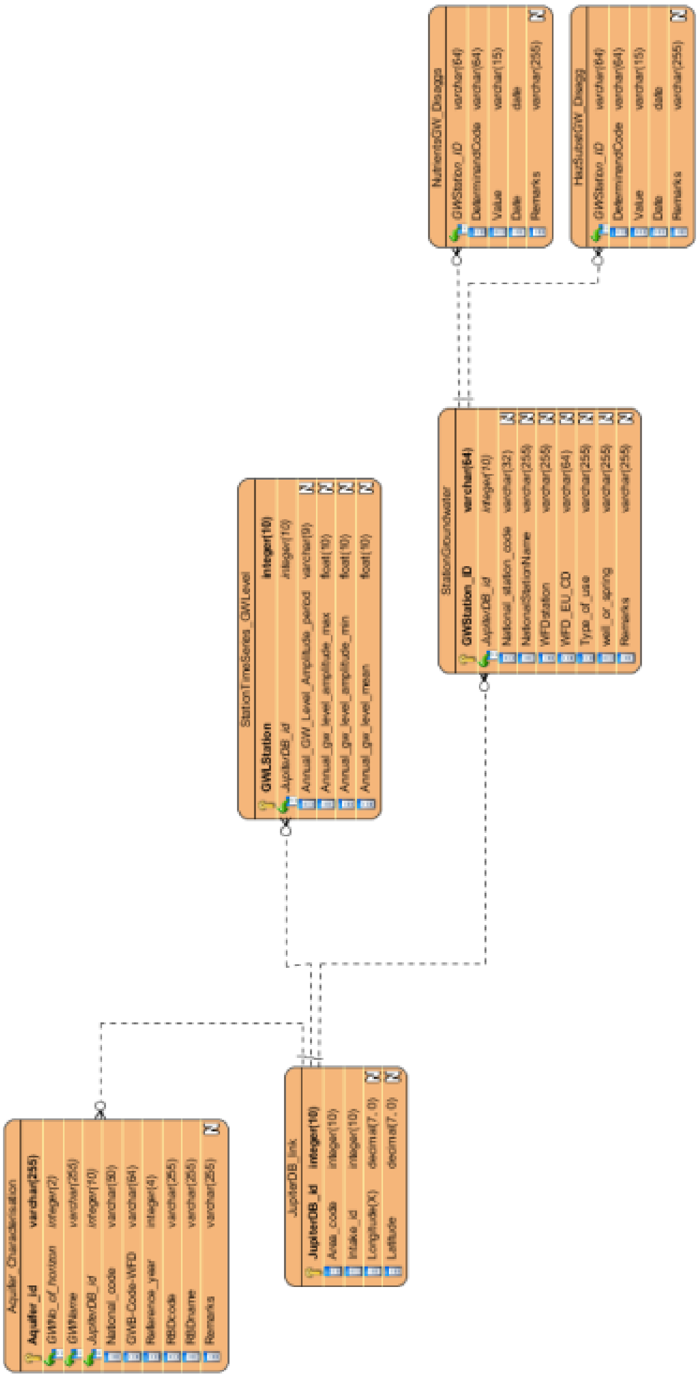


Figure 16 Source Author

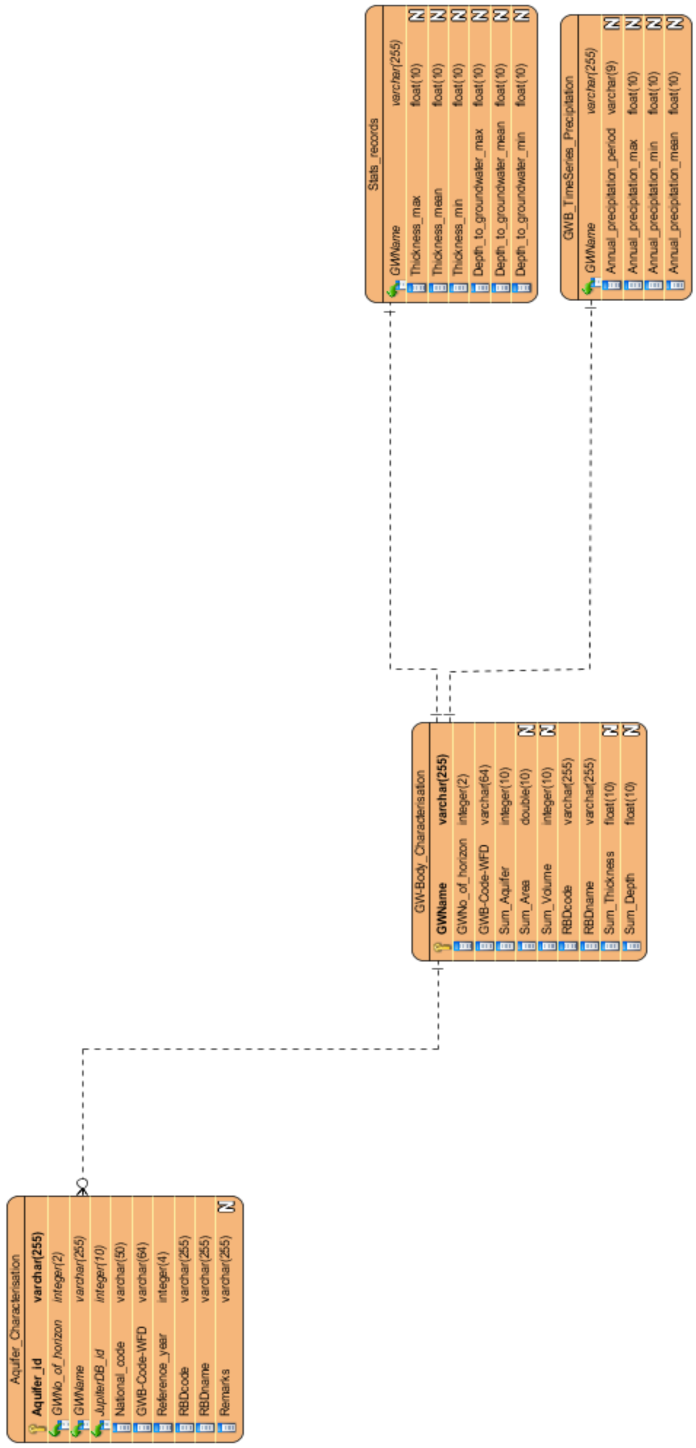


Figure 15 Source Author

Figure 16

The main entities shown in this first figure are:

Aquifer_Characterisation (WFD)
JupiterDB_link (GEUS)
StationTimeSeries_GWLevel (GEUS)
StationGroundwater (WFD)
NutrientsGW_Disagg (WFD)
HazSubstGW_Disagg (WFD)

As cited, the list of names placed below each name of entities represents the main attributes that tell some kind of information about the respective entity.

Considering as relevant the higher level of information, the attention should be focused on the entities which gather the major of data within the relationship taking part at this phase of the diagram: Aquifer_Characterisation and Jupiter DB_link; let's look at them individually.

Aquifer_Characterisation entity (WFD): Most of the information given by the attributes in this entity refers to main physical characteristics and proxy pressures related to the aquifers underlying Denmark. Data such as aquifer national code, names, river basin district name, etc; is stored. This information is required by the WFD.

The data stored and also derived from this entity, belongs to an aquifer area this is because the majority of the data coming from the Jupiter database is based on this extension area. The unique identifier and so primary key (PK) in this entity is the attribute: Aquifer_id. This PK determines the relationship established with the JupiterDB_link entity. The line connecting both entities means that every single Aquifer_id in the table can be associated with at most one attribute's value from the JupiterDB_link.

JupiterDB_link entity (GEUS): It gathers information from GEUS' Jupiter database, there can be found information regarding aquifer coordinates, areas, monitoring network, etc.

The PK in this case is: JupiterDB_id. This PK makes possible two relationships; with the entity StationTimeSeries_GWLevel and StationGroundwater respectively.

The relationship with StationTimeSeries_GWLevel displays that at any point in time, each value from JupiterDB_id can have zero to many values from the attributes in the entity StationTimeSeries_GWLevel; and each data from the entity StationTimeSeries_GWLevel must be related with one and only one value from the attributes in the JupiterDB_link entity.

StationTimeSeries_GWLevel entity, holds information about the groundwater level measured for a specified period of time.

The remaining relationship previously mentioned among JupiterDB_link and StationGroundwater entity, establishes the following association; each value of JupiterDB_id can be associated with zero, one or many values from the attributes stored in StationGroundwater.

StationGroundwater entity is an entity which gathers information required by the WFD, which is asked to store detailed data on the physical characteristics of the sampling sites of the groundwater bodies.

The PK for StationGroundwater entity for being a unique identifier value is: GWStation_ID

The last two entities in this figure shown, NutrientsGW_Disagg and HazSubstGW_Disagg, derived both from two different relationships drawn from the entity StationGroundwater.

Although both entities (NutrientsGW_Disagg and HazSubstGW_Disagg) represent a separate relationship, they have the same meaning which tells that, each GWStation_ID value from StationGroundwater, can be related to zero, one or many of the values of the attributes stored within the entities that StationGroundwater is connected to.

NutrientsGW_Disagg entity, involves data required by the WFD and due to its nature, its information might exist or not depending on labs works focus, it comprises chemical quality data on the concentrations of Nutrients and general physico-chemical determinands.

HazSubstGW_Disagg entity; this entity gathers information required by the WFD based on chemical quality data on the concentration of hazardous substances and other chemical determinands in groundwater

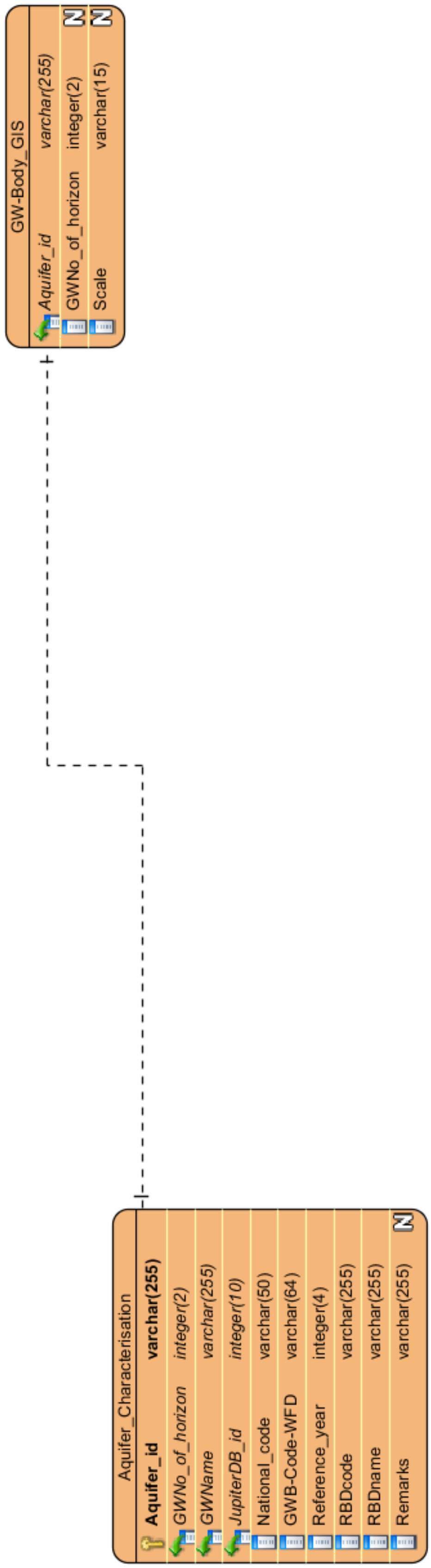


Figure 18 Source: Author

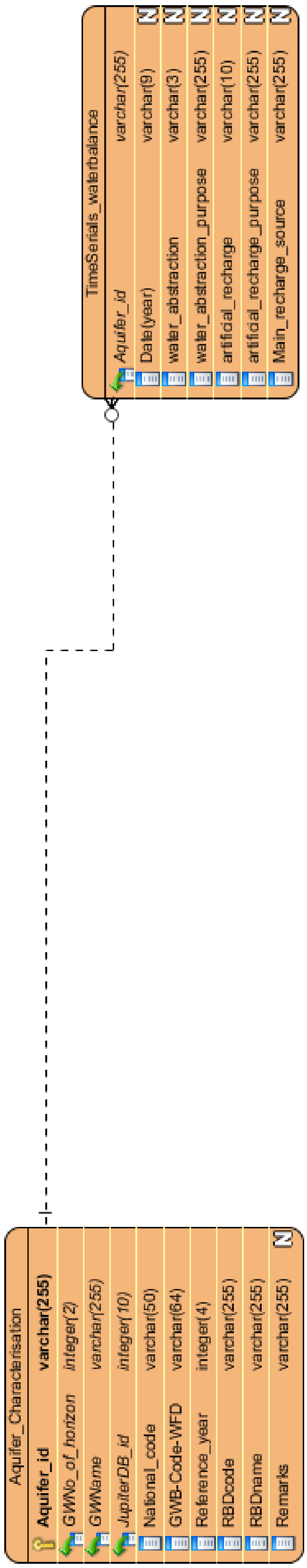


Figure 17 Source: Author

Figure 15

The first relationship to look up is the one between Aquifer_Characterisation and GW-Body_Chracterisation. Because it was already described the information held by the entity Aquifer_Characterisation, let's introduce some relevant details from GW-Body_Characterisation.

GW-Body_Characterisation entity: collects a set of information which refers to the physic characteristics and proxy pressures relating to groundwater bodies located under Denmark. Such information is demanded by EEA to fulfil purposes regarding WFD. Anyway, this entity mainly gathers the same data as Aquifer_Characterisation, but the purpose here has been to summarise those values in order to cover a bigger surface, groundwater bodies instead of aquifers.

The PK for the latter entity is: GWName

This connection allows the logical relation that says that a groundwater body can be associated, to zero, one or many of the data available in Aquifer_Characterization entity; conversely the data held in Aquifer_Characterization can be related to one and only one groundwater body.

At last, this part of the diagram represents two more relationship where two new entities take part: GWB_TimeSeries_Precipitation and Stats_Records.

GWB_TimeSeries_Precipitation entity: contains information on precipitation occurring over the groundwater body area in mm within defined time series.

This information is found at GEUS databases and it must be part of the tables of the future database to accomplish WFD requirements.

Stats_Records entity: Even though this entity can be easily seen as part of GW-Body_Characterisation because in appearance it just stores more physic data of groundwater bodies, consciously has been separated into another entity, the reason behind this movement is that most of the data here found is based on statistic calculations which are subjected to changes and updates through time.

The idea of having this specific data split into a new entity is to avoid errors or not desired changes that might occur, when combining large amount of data in only one entity or future table. As Stats_Records, is more likely to be modified at any time, by separating this particular data into a smaller entity, changes won't influence the remaining data at GW-Body_Characterisation, which do not need to change at the same time. This is helpful when trying to maintain data integrity and quality within a database.

Both of these entities, as seen, are related to the same entity: GW-Body_Characterisation, using the same relationship, one-to-one, this comes to tell, in both cases, that each groundwater body is related with at most one of the values of these two entities respectively; And any value of the two entities can be associated to just one groundwater body at a time.

Figure 17

GW-Body_GIS entity: The new entity appearing in the above relationship, hold spatial information, meaning that, the attribute data refers to geographical data sets of groundwater bodies, these ones have to be provided as polygons in ETRS89 reference system. Although these polygons need to correspond to groundwater bodies surface; as it was beforehand told most of the data at GEUS' databases has been originally computed and/or stored based on aquifers surface.

The only relationship appearing here reveals that each aquifer can be related with one and only one polygon shape, which will allows in future time, to visualize information of aquifers on a map.

Figure 18

TimeSeries_waterbalance entity: Mainly stores information regarding flow of water within the aquifer system, data such as condition on how water is maintained in the hydrogeological unit, and many other hydrological details; during a period of time

The relationship existing between this entity and Aquifer_Characterisation is, one-to-many. It states that each aquifer from Aquifer_Characterisation, can have zero, one or many values regarding data in TimeSeries_waterbalance.

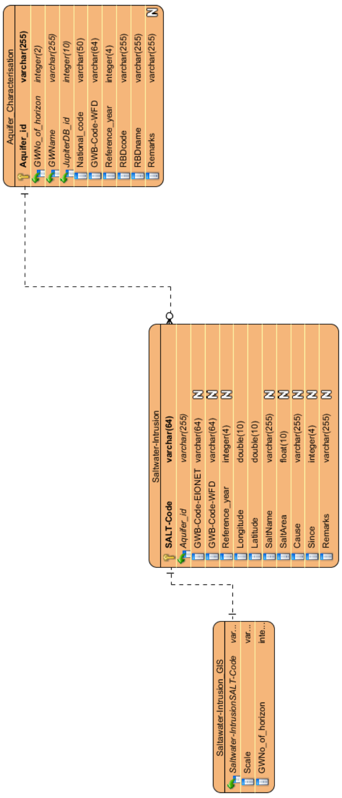


Figure 22 Source: Author

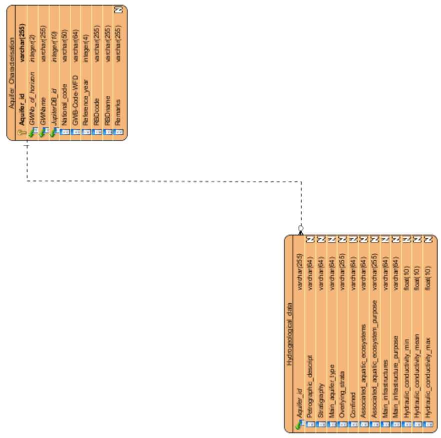


Figure 21 Source: Author

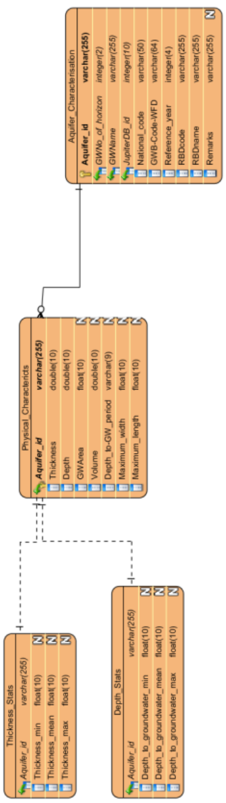


Figure 19 Source: Author

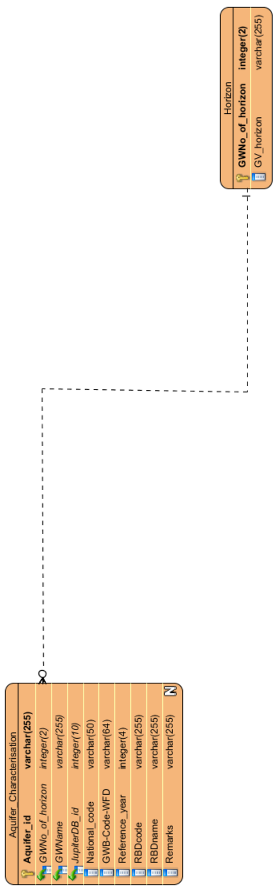


Figure 20 Source: Author

Figure 19

Physical_Characteristics entity: This entity's data is also considered as the same kind of information given by the entity Aquifer_Characterisation data, required by the WFD: which as cited, involves detailed information on the physical characteristics and proxy pressures related, in this case, to aquifers.

In this specific situation, data from Aquifer_Characterisation is based more on fixed values (names, codes, dates…) which are not supposed to change in a long term whereas data inside Physical_Characteristics is dynamic data, figures that due to measurements probably will need to be updated. Thereby, this is the reason for this group of information to appear spread between two entities.

The relation here represented between the last two described entities is, one-to-many, establishing that; every single aquifer from Aquifer_Characterisation is associated with zero, one or many values of each attributes belonging to Physical_Characteristics.

The entities **Thickness_Stats** and **Depth_Stats,** are a subset of data related also to physical characteristics of the surfaces of aquifers. Both entities hold statistics measurements about thickness and depth respectively.

As shown in the figure (Figure 19 Source: Author), the attributes: Aquifer_id; was declared as PK in Physical_Characteristics, by doing this, it is easy to associate each aquifer with a potential existing value of Thickness or/and Depth; and through the relationship: one-to-one, it is possible to connect to both statistics entities. Again, computing these statistics values within smaller entities will avoid future data redundancy.

As observed, within this part of the diagram there is a new element which have not appeared in other figures, the continuous black line drawn which represents the relationship between Aquifer_Characterisation and Physical_Characteristics, this line defined the existence of an identifying relationship. An identifying relationship means that the child entities shown (Physical_Characteristics, Thickness_Stats and Depth_Stats), depends all on the parent entity (Aquifer_Characterisation) and they cannot be uniquely identified without the parent entity. So, these child entities shared the same primary key.

Figure 22

Saltwater-Intrusion entity: As its name suggests, the data stored is referred to any occurrences of saltwater intrusion caused by aquifer overexploitation.

The one-to-many relationship seen, tells that an aquifer can be related to zero, one or many values regarding salt water intrusion data, for instance a particular location where a salt intrusion might have occurred; date, cause, etc.

Saltwater-Intrusion_GIS entity: It contains data referring to digital maps of occurrences of saltwater intrusion affecting aquifers in Denmark. This data, as in a prior situation explained, represents spatial information as polygon shapefiles that have been developed by using a specific software as ArcGIS, this dataset must be provided in the ETRS89 reference system.

The relationship one-to-one, assign one and only one polygon area corresponding to a specific saltwater intrusion with a single aquifer where this situation might occur.

For the entities in this figure involved, data is not yet found at GEUS, despite, as I was informed, that salt intrusion events are likely to occur, affecting groundwater bodies in Denmark. Nevertheless, is required information in order to reach objectives regarding WFD, so, this entity must be maintained when modelling the diagram for future purposes.

Figure 20

Horizon entity: Using the meaning of the horizon concept as a data element makes easier the description of this entity. Horizon is defined as the vertical position beforehand defined by a particular country (Denmark in this case), in which a groundwater body is located.

At GEUS the horizon attribute: GV_horizon, was prior defined as a name (formed by letters and numbers) whereas the horizon attribute or data element required from the WFD: GWNo_of_horizon; must be a numeric value or integer, therefore this simple entity was created to give solution to this particular issue, inside it, each name of horizon given at first by GEUS, is correlated to a number as indicated by requirements from the WFD.

The PK found at Horizon is: GWNo_of_horizon

The relationship here addressed from the new entity to the, now well-known, Aquifer_Characterisation is: one-to-many, this establishes the next statement:

One specific horizon can be related to zero, one or many aquifers, so within a horizon zero, one or many aquifers can be found along with the respective characteristics stored

In contrast, one aquifer can be associated to one and only one horizon, meaning that each aquifer is located within a unique horizon

Figure 21

Hydrogeological data entity: Mainly gathers information that deals with the distribution and movement of groundwater in the aquifers, as well as, main geological characteristics of the same ones and descriptions of the aquatic ecosystems involved.

The one-to-many relationship made from Aquifer_Characterisation tells that; one aquifer is related to zero, one or many data available within the Hydrogeological_data entity.

This Entity Relationship Diagram (ERD) (Figure 13 Source Author) has been modelled according to data requirements by the EEA and considering data availability at GEUS. This latter aspect is important to have in mind, because an organization such as GEUS comprises many departments which generate and load to the system huge amount of data. Might be a tough task to get to know and load all data (generated in-house and also coming from external sources). Therefore, be aware that at the start of the design of this data model, GEUS' data used, was data available at that time. Issues regarding to organizing and dealing with data at GEUS will be discussed in later sections.

5 Recommendations: Multidisciplinary team & Agile Best practices for Data Warehousing

One of the main recommendations presented in this thesis is the creation of a multidisciplinary team in charge of the development of the data warehouse. Aiming to suggest the composition of this team, the arguments written by (Oppel, 2010) about team building have been used; in his book the author mentions the roles that should be included in such a development team.

Firstly, it is recommended to include the executive sponsor, this is the individual who is in charge of allocating funds to the project. Thus he can be aware at any time of the budgetary needs of the project.

It is also important that the users that will use the developed data warehouse are in the team. Users are the ones that actually will use the database once developed. In this case GEUS users don't just use data for analysis, but also they are the ones generating data, therefore, the organizational structure of GEUS plays an important role because the required data is generated in different departments (Hydrology, Geology etc.). Hence people from the departments that generate data should be included. Basically, users will be the ones who will use the data warehouse on their operations.

Subject matter experts are also recommended to be in the team because they have deep knowledge about the nature of the data, thus, they can advise about the sense of including and excluding data. Furthermore, data modelers should be in the team because they have an overall picture of the current data status and how it should be implemented.

The Database Administrator is the person who will build the actual physical database; hence, he must be incorporated. In fact, database administrator should take over this Master Thesis in order to complete the physical model.

Operation specialists comprise the IT department that will be responsible of keeping the data warehouse running, therefore, their involvement is crucial because they can offer the knowledge for planning the necessary computer resources that will serve the database.

Last but not least, considering that in this data warehouse WFD play such an important role, in the team there should be a person who is familiar not just with the data status at GEUS but also with the European requirements.

Agile Best practices

Agile best practices for data warehouse development are thought to reduce the overall risk of the process while increasing the likelihood of matching end-user's needs. Some of these best practices (Ambler & Lines, 2012) are expected to be useful for further physical development of the database.

The first data warehousing best practice is doing an initial architecture envisioning. This basically consists on doing a high level vision of the architecture at the beginning of the project. By doing this we can obtain an overall picture of, for example, the relationships among departments and the data generation and flow created in the organization. In the GEUS case, seems to be important in order to visualize what data is generated by departments and where is stored.

The second principle is to develop the data warehouse in a just in time model, it means that it should be assumed that requirements and unforeseen features will come up along the development. By working just in time it is possible to keep on incorporating and modifying on continuous basis. In the GEUS case, due to the lack of specific knowledge of the status of the data it is likely that new events come up during the development.

Proving the data warehouse earlier, it is also recommended as best practice. It refers to actually testing the code of the data warehouse continuously during the work process so that errors can be easily identify and solve

To develop a system efficiently, it is very important to understand how people will potentially use it. This implies that the development should be performed under a combined focus, not only in data but usage. This will avoid the risk of designing something not aligned and fitted to the organization.

Additionally, being aware of what it's called the "one truth" philosophy and not strictly stick to it is another best practice. Obviously it is desirable to have a single definition for each data element. But the reality is that organizations are made of different department with its own terminology, constraints, way of working etc.

This principle suggests that we should consider the “one truth” philosophy but also the organizations features and not let prevent the data definitions over the value of the organization.

Furthermore, the work should be organized by requirements. Agile projects usually base work on prioritized requirements, so that each iteration of the process aims at meeting the highest requirement of the stakeholder.

Finally, the involvement of the team through the process is critical; actually it’s even better that stakeholder’s take part in the modelling process directly. In this report this principle has been applied as much as possible during the modelling stage.

(Ambler & Lines, 2012)

6 Discussion

This section comes to expose some reflections about this report and the year spent working with data at GEUS.

First of all it is essential to mention that this Master Thesis was built partly upon the 3rd Semester project taken at GEUS (Gallego, 2015). At the beginning of that project, GEUS focus was to find out which kind of data was required for reporting purposes in order to accomplish with deadlines established by the WFD. Understanding main regulations, specific data requirements and European Union Institutions, involved a hard task which supposed a time consuming process. Thereafter, the primary object was to compute and process as much data as possible to be finally reported. However getting to this point revealed an immediate necessity of reorganizing all the data gathered at GEUS; the first decision undertaken to solve this matter was the construction of a data warehouse, where the data will be stored maintaining its integrity. To reach this goal the prior step to proceed with, is the design of a data model which provides a first view of the architecture of a data warehouse, and will help in its future implementation.

At start, the main identified issue was; a lack of communication among departments "in-house" resulting on a complex access to data. There doesn't seem to be a work group or at least one person officially in charge and skilled in the field regarding the new directives, or knowing to a certain extent the highlights concerning particular requirements of the WFD.

Hence, there was no sense to keep calculating new data. Thus, this Master Thesis evolved from the idea of building a consistent structure for the organization where data can be stored, organized, and guided, keeping its reliability at any time, thereby preserving data quality. There was also a very important goal behind this Master Thesis, though is a common aspect often forgotten in many organizations; duly inform users involved, both those implied in the implementation and maintenance of software and those who access it, in order to somehow work with data, ensuring a long lifetime to the system.

With regards to the WFD, as it has been cited several times, it aims to protect water and ensure its good status across Europe, seemingly, by 2015. This is an ambitious challenge that supposes a change of mind-set, considering not only, rivers as water channels leading and aquifers as geological storage pools, but as an integrated and connected system in which its quality depends on each other. For this to be real there is a need to promote a great update and movement from the old water legislations prevailing in each European country into the new WFD. This is being a narrow road to drive, incurred by the big resistance to change from the National administrations who still believe in some "proven" practices related to water management.

At this point of the report, it has been hinted several times the fact of non-existence awareness about the way of storing data at GEUS, but it has been discovered a relevant issue that might happen in most of the European countries involved regarding WFD reporting requirements. Even though data achieved at GEUS to accomplish deliveries established at a National level were referred to the same topic (Groundwater quality); some needed data such as names, codes, chemistry values, etc. differ from requirements found at the data dictionary provides by the WFD . It seems that the focus is set on the National reporting obligations rather than the European ones. Adapting routine practices of, data generation and storage, has been also a drawback for a long time not only at GEUS and in Denmark, as I was able to check while working with GEUS data, but also for some other Member States which have not managed to fit their old National water directives into the new WFD (WWF)

Aiming to get insights and valuable information about this matter, I was encouraged to contact with the EEA, nonetheless, it did not succeed. Hence, it was impossible to know whether Member States have done significant improvements concerning directives adaptation and data requirements, or not.

Additionally, the develop of the 3rd Semester project enabled, specially myself, to be familiar with WFD, as well as all the institutions that take part on it, and even more important, I was able to learn data status at GEUS and ways of dealing with it. Thanks to the role played by this previous project, this Master Thesis was possible.

It is also relevant to notice that this Master Thesis has not been performed from a "strictly" computer-science approach. Instead, it has been raised to give solution to some matters encountered at GEUS related to data management, which were directly affecting data report requirements set by WFD. Moreover, within this report it is found a hidden purpose that is trying to convert its reading comprehension into a comfortable line to follow, easy not only to catch up but also to imagine and visualize by an outside reader of the topic. Likewise computer scientist or IT professionals can add valuable insights upon this report.

Concerning the author's learning process this report has represented a great and thrilling challenge. Besides all the knowledge acquired from the 3rd Semester approach, this Master Thesis gave the facility to gain and develop skills on databases and data warehouses; data models concepts and data modelling design process; and above all to know and understand better the relevance of data structure.

To wrap up, the opportunity to work at GEUS this last course, dealing with data referring to WFD and groundwater quality concerns, has provided the author with a wider knowledge regarding water related data management, considering this a productive aspect due to author's educational and professional background, civil engineer specialized in hydrology.

Furthermore, this last academic year can be considered as a valuable addition upon the skills gained at the first year of the Master. The 1st year completed of the MSc programme was highly significant at obtaining knowledge mostly related to geodesign (geovisualization and geocomputation tools among others) due to projects developed at that time. Whereas this last year, internship taken at GEUS along with the chosen Thesis topic, have supposed an improvement, regarding data management capabilities, and GI-technology and information systems.

All in all, these two years have been reached from two different approaches which, in terms of learning and experiences, complement each other.

Moreover, is interesting to mention software used for data modelling process, Visual Paradigm; as it was described previously in the section (2.2.4) offers many tools especially for the design of different data models, between other relevant design purposes and tasks, related to databases. However, since it was the first time working with this software, somehow it was needed a period of learning in order to obtain the best outcome possible of Visual Paradigm.

It is noteworthy that intentionally the writing style on this report is seeking to be an engaging content. Creating theory and implementation sections where, explanations try to be as accurate as possible while keeping a certain degree of abstraction from computer-science view, is a completely challenge which is expected to be achieved.

7 Conclusion

As an overall conclusion, it can be said that the need of compliance with WFD has revealed the necessity of data organization at GEUS. This Master Thesis has been proposed as a solution to this issue by proposing a data model, a data standardization guide, and some recommendations about data warehousing. However in order to fully meet WFD requirements, a clear commitment by GEUS will be crucial not just for reporting obligations but also to improve the management of data related to water resources, and this is of vital importance for the water quality in Denmark and so in Europe. Therefore this report should be considered as the foundation upon which GEUS could develop a clear data management policy and data warehouse.

Acknowledgements

I would like to express my gratitude and special recognition for their collaboration on this report, to GEUS and everyone that comprises the team from the Hydrology Department, with special mention to Lars Troldborg, who without his experience, knowledge, and help this would have not been possible. To the organization itself, for giving me the opportunity to live this experience during my last academic year, serving me at any time with the needed resources.

My special gratitude to my supervisor in this Master Thesis, Thomas Balstrøm, whose expertise and patience together with his vast knowledge regarding every field here discussed, supposed an important value and source of insights, which helped not only to develop this report but also to fulfil my learning goals.

Last but not least, to those special people that gave their great support and positive energy, closer and from distance, in order to be confidence enough to take and face this challenge.

List of references

1KeyData. (2015, May 15th). *Data Warehousing*. Retrieved from 1Keydata.com offers free online programming tutorials and information on data warehousing: http://www.1keydata.com/

Ambler, S., & Lines, M. (2012). *Disciplined agile delivery: A practitioner's guide to agile software delivery in the enterprise.* IBM Press.

Beal, V. (2007, May 18th). *Metadata: Data About Data.* Retrieved from Webopedia Computer Science: http://www.webopedia.com/DidYouKnow/Computer_Science/metadata.asp

Bowman, D. (2009, May 15th). *Information management check list.* Retrieved from http://www.information-management-architect.com/: http://www.information-management-architect.com/

EEA. (2003). *European Environment Agency, ReportnetElectronic infrastructure and tools for streamlining flows of environmental information in Europe.* Retrieved September 23, 2014, from https://www.eionet.europa.eu/reportnet/reportnet2003.pdf

EEA. (2012). Eionet connects sharing environmental information in Europe. *EEA Publications*. Copenhagen, Denmark: European Environment Agency.

EEA. (2013a, July). *Data dictionary- Definition of: WISE_Soe Reporting: Groundwater body Quality dataset_version2013.* Retrieved September 15th September, 2014, from EEA - http://www.eea.europe.eu/

EEA. (July de 2013b). *Data Dictionary docmentation: Concepts & terms - datasets, tables and data elements*. Obtenido de http://dd.eionet.europa.eu/documentattion/doc2

EEA. (n.d.). *River basin management plans and programmes of measures*. Retrieved May 15th, 2015, from European Environment agencty: European Environment Agency, Reportnet

EU. (2014). *Introduction to the new EU WFD European Commission*. Retrieved September 10th September, 2014, from htttp://ea.europa.eu

Gallego, C. (2015, January 6th). Modelling a spatial database for Danish groundwater bodies: First approach to the design of a spatial database at GEUS under the European Water Framework Directive. Copenhagen: Aalborg University Copenhagen Student Project Library.

GEUS. (2010). *National water resources model DK-MODEL*. Retrieved from National Vandressource Model: http://vandmodel.dk/fs_national_water_model..pdf

GEUS. (2014). *Official website*. Retrieved November 1st, 2014, from http://www.geus.dk/geuspage-uk.htm

Harris, J. (2011, May 15th). *Implemengint an enterprise data dictionary.* Retrieved from Erwin: http://erwin.com/content/tips_and_techniques/Implementing_an_Enterprise_Data_Dictionary_Standard.pdf

Hoberman, S. (2009). *Data Modeling Made Simple: A Practical Guide for Business it Professionals.* US: Technics Publications LLC.

Immom, W. (2005). *Building the data warehouse.* John wiley & sons.

Jansen, C. (n.d.). *Techopedia.com*. Retrieved from Normalization: http://www.techopedia.com/definition/1221/normalization

Kimball, R., & Ross, M. (2011). *The data warehouse toolkit: the complete guide to dimensional modeling.* John Wiley & Sons.

Layton, J. (2013, 18th May). *Where should you store your metadata?* Retrieved from http://www.enterprisestorageforum.com/: http://www.enterprisestorageforum.com/storage-management/where-should-you-store-your-metadata.html

McCaleb, M. R. (1999). A Conceptual Data Model of Datum Systems. *Journal of Research of the National Institute of Standards and Technology, 104*(4-6).

NCDENR. (n.d.). *North Carolina Department of Environment and Natural Resources*. Retrieved from Basic Hydrology: http://www.ncwater.org/?page=560

NISO. (2004). *Understanding Metadata.* Betheseda: NISO, National Information Standards Organization.

Oppel, A. (2009). *Databases A Beginner's Guide* (1st ed.). McGraw-Hill Osborne Media.

Oppel, A. (2010). *Data Modeling: A Beginner's Guide.* McGraw-Hill.

ORACLE. (n.d.). *A Relational Database Overview*. Retrieved from ORACLE Tutorials: https://docs.oracle.com/javase/tutorial/jdbc/overview/database.html

Smith, P., & Jhon, G. (2005). *Dictionary of Water and Waste Management* (2nd ed.). London: IWA Publishing.

Visual Paradigm. (2015, May 4). *Database Design (ERD) Toolset*. Retrieved from Visual Paradigm: http://www.visual-paradigm.com/features/database-design/

West, M., & Fowler, J. (2003, September 2nd). Developing High Quality Data. *EPISTLE*. London, United Kingdom: European Process Industry STEPTechnical Liasion Executive.

WISE. (2008). Water Notes on the implemetatio of the Water Framework Directive, Groundwater at risk: Managing the water under us. *Water Notes*. Copenhagen, Denmark: EEA.

WISE. (2013). *WFD - Annex II - 2.1. Initial characterisation*. Retrieved September 15, 2014, from http://www.wise-rtd.info/en/info/annex-ii-21-initial-characterisation

WISE. (n.d.). *About WISE water information system of Europe*. Retrieved from WISE.com: http://water.europa.eu/info

WISE Annex II . (n.d.). *Annex II - 2.2. Furhter characterisation*. Retrieved November 1st, 2014, from WISE: http://www.wise-rtd.info/en/info/annex-ii-22-furhter-characterisation

WWF . (n.d.). *Water Frame Work directive [Directiva Marco del Agua]*. Retrieved from Wordl Wild Life Spain: http://www.wwf.es/que_hacemos/agua_y_agricultura/nuestras_soluciones/directiva_marco_de_agua/

List of figures

List of tables

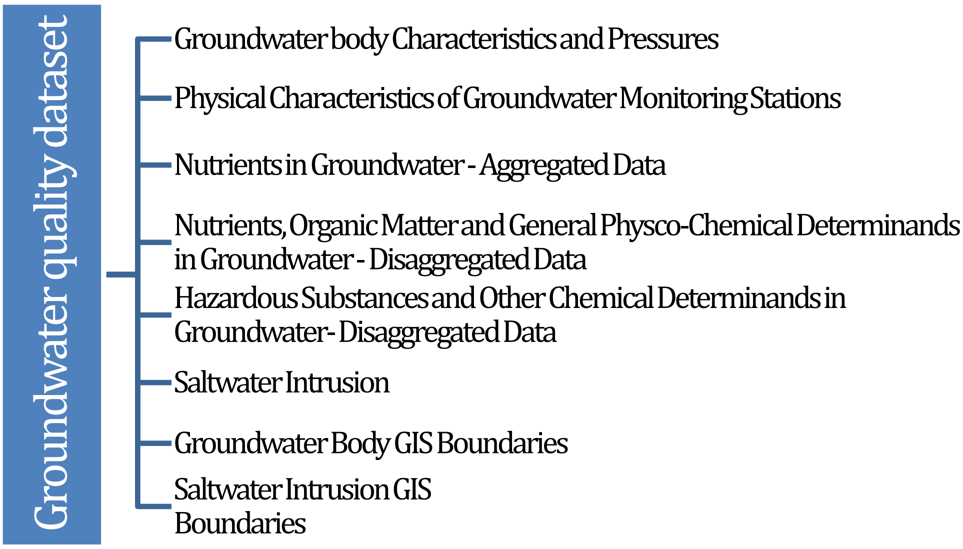
List of Acronyms

DD	Data Dictionary
DB	Database
DBA	Database Administrator
DBMS	Database Management System
EEA	European Environment Agency
Eionet	European information and observation network
ETRS89	European Terrestrial Reference System 1989
EU	European Union
FK	Foreign Key
GIS	Geographic Information System
GEUS	Geological Survey of Denmark and Greenland
GWB	Groundwater Bodies
IS	Information System
NWMR	National Water Model Resources
ORDBMS	Object-Relational Database Management System
PK	Primary Key
PoM	Programme of Measures
RBD	River Basin District
RBMP	River Basin Management Plan
SDBMS	Spatial Database Management System
SQL	Structure Query Language
WFD	Water Framework Directive
WISE	Water Information System for Europe

Annex I: Data standardization guide

This section could be considered as *"translation"* of the data requirements in such a way that GEUS can actually understand what and in what format they are to report. This section is actually based on EEA terminology, more specifically the data dictionary provided by the Reportnet, Dataset specification for WISE-SoE Reporting: Groundwater quality Version July 2013 (EEA, 2013a; EEA, 2013b).It aims to eases the understanding of European requirements. Therefore, details of data requested on groundwater are presented here by table, this includes, the physical characteristics of the groundwater bodies, proxy pressures on the groundwater area, as well as chemical quality data on nutrients and organic matter, and hazardous substances in groundwater.

The tables already presented together form the groundwater body quality dataset.



WISE-SoE: GW quality. Each table requests different information.

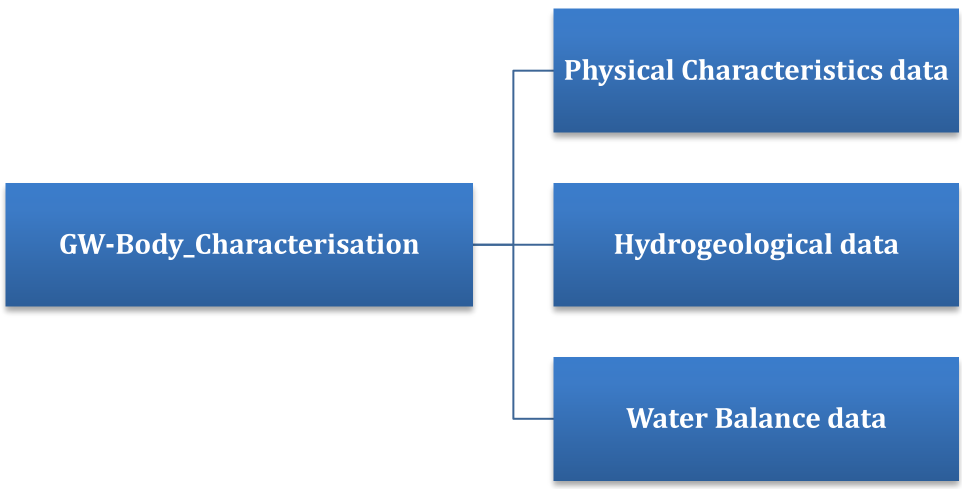
Groundwater Body Characteristics and Pressures: GW-Body_Characterisation

Table definition: Detailed information on the physical characteristics and proxy pressures relating to the groundwater bodies on annual basis.

Data requirements:

- Groundwater Body Code (**GWB-Code-EIONET** and **GWB-Code-WFD)** should be unique in this table and no duplicate records should exist
- Only "**ONE**" of the fields **GWB-Code –EIONET** or **GWB-Code-WFD** can be filled for each record of this table. Only if the delineation of certain groundwater body is fully identical within the EIONET and simultaneously according to the Water Framework Directive, Article 5, both of these fields should be filled in such record.
- Filling of **GWB-Code-WFD** field is **mandatory** for EU member countries.

This table gathers data of Physical characteristics, Hydrogeological data and water balance data. It is important to remark this, because this kind of data in GEUS is generated in different departments but the EEA requires holding data in one table. That's the reason why in this section we make this differentiation.



Physical Characteristics data	
Groundwater Body Code - Eionet **GWB-Code-EIONET**	•Internally produced, unique identifier of the GW body which was delineated for the use within EIONET •Based on ISO 3166-alpha-2 country code elements, concatenated with an identificantion number or alphanumeric code
Groundwater Body Code - WFD **GWB-Code-WFD**	•Internally produced, unique identifier of the GW body which was delineated according to WFD, Article 5 and under which it was reported the European Commission •Based on ISO 3166-alpha-2 country code elements, concatenated with an identification number or alphanumeric code •This field is **mandatory** for EU member countries
Name **GWName**	•Name of the groundwater body •Use UTF-8 codelist in case non-ascii character occur
Reference year **Reference_year**	•Last year of revision of update of characteristics •Min inclusive value: 1800 •Max inclusive value: 2014
Number of Horizon **GWNo_of_horizon**	•Vertical position of a groundwater horizon in which a groundwater body is situated •Value from 1 (top) to 10 (bottom)
National Code **National_Code**	•Nationally assigned identifier groundwater body •Not null field
River Basin District Code **RBDcode**	•River Basin District Code, as defined in the codelist •Not null field
River Basin District Name **RBDname**	•Name of the River Basin District •Not null field; keep blank if the field RBDcode is filled
Groundwater Body Area **GWArea**	•Area of the groundwater body in km^2
Minimum Depth to Groundwater Body **Depth_to _groundwater_min**	•Minimum depth from the surface tot he top of the groundwater body in m.
Mean Depth to Groundwater Body **Depth_to_groundwater_mean**	•Mean depth from the surface tot he top of the groundwater body in m.
Maximum Depth to Groundwawter Body **Depth_to_groundwater_max**	•Maximum depth from the surface to the top of the groundwater body in m.
Depth to Groundwater Body - Period **Depth_to_GW_period**	•Monitoring period (time span) for which the Minimum / Mean / Maximum Depth to Groundwater Body was determined •Use format YYYY or YYYY - YYYY, where YYYY is a numeric expresion
Minimum Thickness **Thickness_min**	•Minimum thickness of the groundwater body in m.
Mean Thickness **Thickness_mean**	•Mean thickness of the groundwater body in m.
Maximum Thickness **Thickness_max**	•Maximum thickness of the groundwater body in m.
Maximum Width **Maximum_width**	•Maximum width of the groundwater body perpendicular to groundwater flow direction in km.
Maximum Length **Maximum_length**	•Maximum length of flow path within the groundwater body in km.

Hydrogeological data	
Petrographic Description **Petrographic_descript**	• Short petrographic description of the dominant components of the stratigraphy of the groundwater body • Multiple choice from codelist is allowed. Use comma as a delimeter
Stratigraphy **Stratigraphy**	• Description of the stratigraphy (geological period) of the groundwater body, as defined in the codelist • Multiple choice from codelist is allowed. Use comma as a delimeter
Main Aquifer Type **Main_aquifer_type**	• Predominant aquifer type, as defined in codelist • Multiple choice from codelist is allowed. Use comma as a delimeter
Overlying Strata **Overlying_Strata**	• Description of the groundwater body's overlying strata • e.g. Quaternary deposits (mainly till and loess)etc.
Confined **Confined**	• A confined aquifer isa aquifer that is confined or overlain by a layer that does not transmit water y any appreciable amount or that is impermeable. An unconfined aquifer is a aquifer with water table open to the atmosphere through permeable overlying material. Enter "yes" or "no"
Associated Aquatic Ecosystem **asociated_aquatic_ecosystems**	• Information, wheter the aquatic ecosystems are associated with the groundwater body. Enter "yes" or "no"
Description of th Associated Aquatic Ecosystems **Associated_Aquatic_Ecosystems_Purpose**	• Description of the aquatic ecosystems associated with the groundwater body. Enter "yes" or "no"
Main Infrastructures **main_infrastructures**	• Information, whether the main infrastructures affect the dynamics of the groundwater body. Enter "yes" or "no"
Description of the Main Infrastructures **Main_Infrastructures_Purpose**	• Description of the main infrastructures affecting the dynamics of the groundwater body

Water Balance data

Minimum hydraulic conductivity **Hydraulic_conductivity_min**	• Minimum hydraulic conductivity of the groundwater as a kf-value. Free text description. Unit kf=m/s
Mean hydraulic conductivity **Hydraulic_conductivity_mean**	• Mean hydraulic conductivity of the groundwater as a kf-value. Unit kf=m/s
Maximum hydraulic conductivity **Hydraulic_conductivity_max**	• Maximum hydraulic conductivity of the groundwater as a kf-value. Unit kf=m/s
Minimum Annual Groundwater level amplitude **Annual_gw_level_amplitude_min**	• Minimum of the range between the lowest and the highest groundwater level within a year in m.
Mean Annual Groundwater Level Amplitude. **Annual_gw_level_amplitude_mean**	• Mean of the rage between the lowest and highest goundwater level within a year in m.
Maximum Annual Groundwater Level Amplitude **Annual_gw_level_amplitude_max**	• Maximum of the range between the lowest and the highest groundwaterlevel within a year in m.
Annual Groundwater Level Amplitude - Period **Annual_gw_level_amplitude_period**	• Monitoring period (time span) for which the Minimum / Mean / Maximum Annual Groundwater Level Amplitude was determined. • Use of format YYYY OR YYYY-YYYY where YYYY means the year in numeric expression
Minimum Annual Precipitation **Annual_precipitation_min**	• Minimum long term annual precipitation over the groundwater body area in mm.

Mean Annual Precipitation **Annual_precipitation_mean**	• Mean long term annual precipitation over the groundwater body area in mm.
Maximum Annual Precipitation **Annual_precipitation_max**	• Maximum long term annual precipitation over the groundwater body area in mm.
Annual Precipitation Period **Annual_precipitation_period**	• Monitoring period (time span) for which the Minimum / Mean / Maximum Annual precipitation was determined • Use of format YYYY OR YYYY-YYYY where YYYY means the year in numeric expressi
Water Abstractions **water_abstraction**	• Information, whether water is abstracted from the groundwater body. • Enter "yes" or "no"
Purpose of the water abstractions **water_abstraction_purpose**	• Purpose for which the water is abstracted from the groundwater body.
Artificial recharge **artificial_recharge**	• Information, whether the groundwater body is artificially recharged. • Enter "yes" or "no"
Purpose for Artificial Recharge **artificial_recharge_purpose**	• Purpose for the artificial Recharge of the groundwater body
Main Recharge Source **Man_recharge_source**	• Main source of recharging the groundwater body, as defined in the code list. • Multiple choice from the codelist is allowed
Remarks **Remarks**	• Remarks, comments or explanatory notes

Physical Characteristics of Groundwater Monitoring Stations: StationsGroundwater

Table definition: Detailed information on the physical characteristics of the sampling sites of the groundwater on an annual basis.

Data requirements:

- ✓ At least one of the fields GWB-Code-EIONET or GWB-Code.-WFD has to be filled for each record of this table or entity.
- ✓ Filling GWB-Code-WFD field is mandatory for EU member countries.
- ✓ The station-ID ode should be unique in this table and no duplicate records should exist

Station ID GWStation_ID	•Internally produced, unique identifier of the groundwater monitoring station •Based on ISO 3166-alpha-2 country code elements, concatenated with the nationally assigned unique identifier of the groundwater monitoring station.
Groundwater Body Code - EIONET **GWB-Code-EIONET**	•Internally produced, unique identifier of the groundwater body which was delineated for the use within the EIONET •Based on ISO 3166-alpha-2 country code elements, concateneted with an identification number or alphanumeric code.
Groundwater Body Code - WFD **GWB-Code-WFD**	•Internally produced, unique identifier of the groundwater body, which was delineated according to the Water Framework Directive, Article 5 and under which it was reported to the European Commission. •Based on ISO 3166-alpha-2 country code elements, cocatenated with an identification number or alphanumeric code.
National Station Code **National_station_code**	•Nationally assigned, unique identifier of the groundwater monitoring station
National Station Name **NationalStationName**	•National name of the monitoring station •Use UTF-8 codelist in case non-ascii characters occur
WFD station **WFDstation**	•Specification yes / no, whether the monitoring station was reported to European Commission as Water Framework Directive, Article 8 monitoring station.
WFD Station Code **WFD_EU_CD**	•Internally produced, unique international identifier of the monitoring station under which the station was reported as WFD Art.8 monitoring station •Keep blank if WFD station = no

Longitude
Longitude

- (X) International geographical co-ordinates in decimal degrees format
- Not null field
- Use the common geodetic datum ETRS89. WGS84 should be used for overseas areas and can be used for TCM data as well
- Use negative values for coordinates west of the Greenwich Meridian (0º)
- Round the coordinates to 4 - 5 decimal places (0.0001º = about 10 m)
- Unit: decimal degrees
- Min inclusive value: -180
  Max inclusive value: 180
- Decimal precision: 7
- This is a required not nulled field.

Latitude
Latitude

- (Y) International geographical co-ordinates in decimal degrees format
- Not null field
- Use negative values for coordinates south of the Equator (0°)
- Min inclusive value: -90
  Max inclusive value: 90
- Rest of requirements are equal to those defined for: Longitude
- This is a requried not nulled field

Type of use
Type_of_use

- Type of use of groundwater, as defined in codelist
- Multiple choice from codelist is allowed. Use comma as a delimiter

Well or Spring
well_or_spring

- Sampling site is a Well or Spring, as defined in codelist

Remarks
Remarks

- Remarks, comments or explanatory notes (free text
- Rivers, Lakes, Groundwater quality - concentration data tables: Enter the text "value confirmed" in the case you are sure the value exceeding the Potentially high value is correct.

Nutrients in Groundwater – Aggregated Data: NutrientsGW_Agg

Table Definition: Data on Ammonium, Dissolved Oxygen, Nitrate and Nitrite in groundwater bodies, including information on the number and type of sampling sites, the monitoring frequency, the number sampling sites for each range of concentration values and statistics, are requested from EEA Member Countries on an annual basis.

Data requirements:

- ✓ Data on Ammonium, Dissolved Oxygen, Nitrate and Nitrite are requested as disaggregated data but can be provided as annually aggregated values. On aggregation, sample concentration values recorded as below the limit of quantification, should be replaced with a value equivalent to half the limit of quantification. Where a calculated mean value of the measurement results is below the limit of quantification, the value shall be referred to as 'below the limit of quantification'.
- ✓ The following fields combine to create a unique record: groundwater body code (GWB-Code-EIONET or GWB-Code-WFD), DeterminandCode-agg and Year. No duplicate records should exist within this combination.
- ✓ It is encouraged to all countries to provide data on Ammonium, Dissolved Oxygen, Nitrate and Nitrite in groundwater, for as long time series as possible. Please use the unit mg/l for reporting Minimum, Mean, Maximum and Median in case of all these substances.
- ✓ The disaggregated data are always preferred. In case of disaggregated data reported for certain substances, their reference area and period, DO NOT report the aggregated data resulting from identical disaggregated data anymore. Such data redundancy is not taken into account.

Only the field GWB-Code-WFD can be filled and it's mandatory for EU member countries.

Groundwater Body Code - EIONET **GWB-Code-EIONET**	• Internally produced, unique identifier of the groundwater body which was delineated for the use within the EIONET. • Based on ISO 3166-alpha-2 country code elements, concatenated with an identiification number or alphanumeric code. • This is a mandatory field
Groundwater Body Code - WFD **GWB-Code-WFD**	• Internally produced, unique identifier of the groundwater body which was delineated according to the WFD, Article 5 and under which it was reported to the European Commission • Based on ISO 3166-alpha-2 country code elements, concatenated with an identification number or alphanumeric code
Determinand Code - agg **DeterminandCode-agg**	• Identification number of the determinand monitored, as defined in the codelist. • Codes 500 - 503 representin Ammonium, Dissolved oxygen, Nitrates and Nitrites can be used only • This is a required not nulled field
Year **Year**	• Year of aggregation period (aggregated data sets) or year in which sample taken (disaggregated data sets), in format YYYY • Year should be no later than requested in the latest data request • Min inclusive value: 1800 Max inclusive value: 2012 • This is a required not nulled field.
Number of Sampling Sites **GWNumberOfSites**	• Total number of sampling sites within aggregation period • The sum of the number of all sampling site types in all concentration classes • Min inclusive value: 0 • This is a requried, not nulled field
Drinking Water Sites **GWDrinking_Water_Sites**	• Number of sampling sites used for drinking water purposes • Min inclusive value: 0
Industrials Sites **GWIndustrial_Sites**	• Number of sampling sites used for industrial purposes • Min inclusive value: 0
Surveillance Sites **GWSurveillance_Sites**	• Number of sampling sites used for surveillance purposes • Min inclusive value: 0
Other Sites **GWOther_Sites**	• Number of sampling sites used for other purposes • Min inclusive value: 0

Average Sampling Frequency **GWSamplingFrequency**	•Average frequency that the sampling sites were monitored during the aggregation period •Min inclusive value: 0 •Decimal precision: 1
Number of Samples **NumberOfSamples**	•Number of samples included in aggregated data •Min inclusive value: 1
Limit of Quantification **LimitOfQuantification**	•The smallest concentration that can be distinguished from the analytical blank at a chosen level of statistical confidence (usually 95%) •Nutrients: Voluntary field •Aggregated data reporting: In case of using different limits within the aggregation period, please enter the highest value.
Minimum **Minimum**	•Minimum disaggregated sample concentration value of the input data used for aggregation •Min inclusive value: 0
Mean **Mean**	•Mean concentration value of aggregated data •Min inclusive value: 0
Maximum **Maximum**	•Maximum disaggregated sample concentration value of the input data used for aggregation •Min inclusive value: 0 •Not null field for reporting of SoE Rivers and Lakes water quality, table Hazardous substances - Aggregated data •Min inclusive value: 0
Median **Median**	•Median concentration value of aggregated data •Min inclusive value: 0

Class 1
Class_1

- The number of sampling sites with average concentration of disaggregated valures of:

  Ammonium ≤ 0.1 mg/l

  Dissolved Oxygen ≤ 2 mg/l

  Nitrate ≤ 10 mg/l

  Nitrite ≤ 0.01 mg/l

  within the aggregation period
- Min inclusive value: 0

Class 2
Class_2

- The number of sampling sites with average concentration of disaggregated values of:

  Ammonium (> 0.1 and ≤ 0.3) mg/l

  Dissolved Oxygen (>2 and ≤ 5) mg/l

  Nitrate (> 10 and ≤ 25) mg/l

  Ntrite (> 0.01 and ≤ 0.03) mg/l

  within the aggregation period
- Min inclusive value: 0

Class 3
Class_3

- The number of sampling sites with average concentration of disaggregated values of:

  Ammonium (>0.3 ≤ 0.5) mg/l

  Dissolved Oxygen > 5 mg/l

  Nitrate (> 25 and ≤ 50) mg/l

  Nitrite (> 0.03 and 0.06) mg/l

  within the aggregation period

- Min inclusive value: 0

Class 4
Class_4

- The number of sampling sites with average concentration of disaggregated values of:

  Ammonium (> 0.5) mg/l

  Dissolved Oxygen - not relevant, keep blank

  Nitrate > 50 mg/l

  Nitrite (> 0.06 and ≤ 0.1) mg/l

  within the aggregation period
- Min inclusive value: 0

Class 5
Class_5

- The number of sampling sites with average concentration of disaggregated values of:

  Ammonium - not relevant, keep blank

  Dissolved Oxygen - not relevant, keep blank

  Nitrate - not relevant, keep blank

  Nitrite > 0.1 mg/l

  within the aggregation period
- Min inclusive value: 0

Remarks
Remarks

- Remarks, comments or explanatory notes (free text)
- River, Lakes, Groundwater quality - concentration data tables: "value confirmed" in the case you are sure the value exceeding the Potentially high value is correct.
- Min size : 0

Nutrients, Organic Matter and General Physico-Chemical Determinands in Groundwater – Disaggregated Data: NutrientsGW_Disagg

Table Definition: Chemical quality data on the concentrations of Nutrients and general Physico-Chemical determinands (especially Nitrate, Nitrite, Ammonium and Dissolved Oxygen) in groundwater are requested from EEA Member Countries on an annual basis.

Data requirements:

- ✓ Comparable and harmonised data are requested on the concentrations and distribution of Nitrate, Nitrite, Ammonium, Dissolved Oxygen and other nutrients or general physic-chemical determinands in groundwater bodies.
- ✓ Countries are asked to provide the Name and CAS Number in case of reporting any additional determinands missing in the code list.
- ✓ Data on Nitrate, Nitrite, Ammonium and Dissolved Oxygen are primarily requested as disaggregated data. If the country cannot provide disaggregated values, data can be provided as annually aggregated per groundwater body in the format described in the Nutrients in Groundwater – Aggregated Data table instead.
- ✓ It is encouraged to all countries to provide data on Ammonium, Dissolved Oxygen, Nitrate and Nitrite and other nutrients or general physic-chemical determinands for as many groundwater bodies, for as many determinands, for as long time series as possible.

Station ID **GWStation_ID**	• Internally produced, unique identifier of the groundwater monitoring station • Based on ISO 3166-alpha-2 country code elements, concatenated with the nationally assigned unique identifier of the groundwater monitoring station • Min size: 3
Determinand Code **DeterminandCode**	• Identification number of the determinand monitored, as defined in the codelist • Integer code specified in the codelist should be provided • Determinands monitored but not detailed in the codelist ca be included in the data file too. Please provide Name and CAS Number in such case. On this account, field is specified as string data type. • Determinands not detailed in the codelist should be reported in µg/l • Unit: mg/l, µS/cm or °C - see codelist for details
Value **Value**	• Concentration of determined sampled • Sample concentrations below the limit of quantification should be indicated as <Value_of_the_limit_of_quantification • Min size: 0
Date **Date**	• Date of sampling in format YYYY-MM-DD
Remarks **Remarks**	• Remarks, comments or explanatory notes (free text) • River, Lakes, Groundwater quality - concentration data tables: Enter the thext "value confirmed" in the case you are sure the value exceeding the Potentially high value is correct. • Min size: 0

Hazardous Substances and Other Chemical Determinands in Groundwater – Disaggregated Data: HazSubstGW_Disagg

Table definition: Chemical quality data on the concentrations of hazardous substances and other chemical determinands in groundwater are requested from EEA Member Countries on an annual basis.

Data requirements:

- ✓ Comparable and harmonised data are requested on the concentrations and distribution of hazardous substances in groundwater bodies. The focus is on pesticides and other Preferred SoE hazardous substances marked in the code list attached to the DeterminandCode field
- ✓ Reporting of any other nationally monitored hazardous substances is welcomed. Countries are asked to provide the Name and CAS Number in case of reporting any additional determinands missing in the code list
- ✓ Data on hazardous substances and other chemical determinannds are requested as disaggregated data
- ✓ When reporting the determinands which represent the sum of hazardous substances in one reported disaggregated data sample (e.g. the sum of DDT-like substances or Total PCBs), in cases of concentrations below the limit of quantification use "0" as an input of the given item for summarization to disaggregated data sample instead of the half of the limit of quantification (Article 5.3 of the QA/QC Directive)
- ✓ In the case that all summarized substances in one disaggregated data sample are below LOQ, enter the value of the limit of quantification with the prefix "<" into the "Value" field (LOQ should be identical for all summarized substances), do not divide the LOQ by 2
- ✓ It is encouraged to all countries to provide data on hazardous substances (especially pesticides and other Preferred SoE hazardous substances), for as many groundwater bodies, for as many determinands, and for as long a time series as possible.

Station ID **Station_ID**	• Internally produced, unique identifier of the groundwater monitoring station. • Based on ISO 3166-alpha-2 country code elements, concatenated with the nationally assigned unique identifier of the groundwater monitoring station
Determinand Code **DetermiandnCode**	• Identification number of the determinand monitored, as definded in the codelist • Integer code specified in the codelist should be provided • Selected substances in the codelist are marked as "Preferred SoE Hazardous Substances". Please focus on these substances in your reporting above all • Determinands not detailed in the codelist should be reported in µg/l • Min size: 0
Value **Value**	• Concentration of determinand sampled • Min size: 0
Date **Date**	• Date of sampling in format YYYY-MM-DD
Remarks **Remarks**	• Remarks, comments or explanatory notes (free text) • Rivers, Lakes, groundwater quality - concentration data tables: Enter the text "value confirmed" in the case you are sure the valu exceeding the Potentially high value is correct.

Saltwater Intrusion: Saltwater-Intrusion

Table Definition: Any occurrences of saltwater intrusion caused by groundwater over-exploitation are requested from EEA Member Countries on an annual basis.

Data requirements:

- ✓ Additional attribute data are requested on any occurrences of saltwater intrusion (from seawater or deep aquifers) caused by the over-exploitation of groundwater.
- ✓ The SALT-Code field should be unique in this table and no duplicate records should exist. This code also relate to the codes used in the Saltwater-Intrusion_GIS table
- ✓ Only GWB-Code-WFD field can be filled as it is mandatory for EU member countries

SALT-Code
SALT-Code

- Internally produced, unique idetifier of the area indicating the occurrence of saltwater intrusion
- Based on ISO-3166-alpha-2 country code elements, concatenated with 'SW' for Saltwater, and an identification number or alphanumeric code.
- Min size: 5

Groundwater Body Code - WFD
GWB-Code-WFD

- Internally produced, unique identifier of the groundwater body, which was delineated according to the Water Framework Directive, Article 5 and under which it was reported to the European Commission
- Based on ISO 3166-alpha-2 country code elements, concatenated with an identification number or alphanumeric code

Reference Year
Reference_year

- Last year of revision of update of characteristics
- Min inclusive value: 1800
  Max inclusive value: 2012

Longitude
Longitude

- (X) International geographical co-ordinates in decimal degrees format
- Use the common geodetic datum ETRS89. WGS84 should be used for overseas areas and can be used for TCM data as well
- Use negative values for coordinates west of the Greenwich Meridian (0°)
- Please round the coordinates to 4 - 5 decimal places, depending on your input data precision (0.0001° = about 10m)
- Min inclusive value: -180
  Max inclusive value: 180

Latitude
Latitude

- (Y) International geographical co-ordinates in decimal degrees format
- Use the common geodetic datum ETRS89. WGS84 should be used for overseas areas and can be used for TCM data as well
- Use negative values for coordinates south of the Equator (0°)
- Please round the coordinates to 4 - 5 decimal places, depending on your input data precision (0.0001° = about 10m)
- Min inclusive value: -90
  Max inclusive value: 90

Name of Area Affected by Saltwater Intrusion **SaltName**	• Name of the area, spot or groundwater body which indicates the occurence of saltwater intrusion • Min size: 0
Area of Saltwater Intrusion **SaltArea**	• Area affected by saltwater intrusion in km^2 • Min inclusive value: 0
Cause **Cause**	• The primary reason for the saltwater intrusion in the groundwater body (free text)
Since **Since**	• Year when saltwater intrusion started • Min inclusive value: 1800 Max inclusive value: 2012
Remarks **Remarks**	• Remarks, comments or explanatory notes (free text) • River, Lakes, Groundwater quality - concentration data tables: Enter the text "value confirmed" in the case you are sure the value exceeding the Potentially high value is correct.

Groundwater Body GIS Boundaries: GW-Body_GIS

Table Definition: Attribute data referring to geographical data sets of groundwater bodies are requested from EEA Member Countries on an annual basis.

Data requirements:

- ✓ It is assumed that your groundwater geographical data sets can be used by the ETC water for work under contract to the EEA and that the geographical data sets can be compiled with other groundwater boundaries and be published on the Waterbase web site. Please declare any restrictions or reservations regarding the publication of your geographical data sets on delivery.

Digital groundwater body geographical data sets have to be provided as polygons in ETRS89 (European Terrestrial Reference System 89) coordinate system in shape file vector data format.

Groundwater Body Code - **WFDGWB-Code-WFD**	•Internally produced, unique identifier of the groundwater body, which was delineated according to the Water Framework Directive, Article 5 and under which it was reported to the European Commission •Based on ISO 3166-alpha-2 country code elements, concatenated with an identification number or alphanumeric code
Number of Horizon **GWNo_of_horizon**	•Vertical position of a groundwater horizon in which a groundwater body is situated •Value from 1 (top) to 10 (bottom) •Min inclusive value:1 Max inclusive value: 10
Scale **Scale**	•Approximate scale at which map data captured (e.g. 1:100,000) •Min size:0

Annex II: Code list (EEA, 2013b)

4. Codelists

4.1 Non-common Elements Codelists

4.1.1 Codelists for **Groundwater Body Characteristics and Pressures** table

4.1.1.1 **Petrographic Description (Released at 24 Oct 2013)** codelist

Value	Definition	Short Description
Boulders		
Carbonate - conduit		
Carbonate - fractured		
Chalk		
Claystone		
Coarse gravel		
Coarse sand		
Conglomerate		
Fine gravel		
Fine sand		
Gravel		
Igneous rock		
Marlite		
Medium gravel		
Medium sand		
Metamorphic rock		
Other		
Sand		
Sandstone		
Shale		
Silt		
Silty sand		

4.1.1.2 **Stratigraphy (Released at 24 Oct 2013)** codelist

Value	Definition	Short Description
Cambrian		
Carboniferous		
Cretaceous		
Devonian		
Jurassic		
Mesozoic		
Neogene		
Ordovician		
Paleogene		
Paleozoic		
Permian		
Precambrian		
Quaternary		
Silurian		
Tertiary		
Triassic		

4.1.1.3 **Main Aquifer Type (Released at 24 Oct 2013)** codelist

Value	Definition	Short Description
Fractured media		
Fractured/Karstic media		
Fractured/Porous media		
Karstic media		
Porous media		

4.1.1.4 **Confined (Released at 24 Oct 2013)** codelist

Value	Definition	Short Description
no	not confined	
yes	confined	

4.1.1.5 **Associated Aquatic Ecosystems (Released at 24 Oct 2013)** codelist

Value	Definition	Short Description
no	No aquatic ecosystems are associated with the groundwater body.	
yes	Aquatic ecosystems are associated with the groundwater body.	

4.1.1.6 **Main Infrastructures (Released at 24 Oct 2013)** codelist

Value	Definition	Short Description
no	No main infrastructures affect the dynamics of the groundwater body.	
yes	Main infrastructures do affect the dynamics of the groundwater body.	

4.1.1.7 **Water Abstractions (Released at 24 Oct 2013)** codelist

Value	Definition	Short Description
no	Water is not abstracted from the groundwater body.	
yes	Water is abstracted from the groundwater body.	

4.1.1.8 **Artificial Recharge (Released at 24 Oct 2013)** codelist

Value	Definition	Short Description
no	The groundwater body is not artificially recharged.	
yes	The groundwater body is artificially recharged.	

4.1.1.9 **Main Recharge Source (Released at 24 Oct 2013)** codelist

Value	Definition	Short Description
Groundwaters		
Irrigation		
No info		
Precipitation		
Springs		
Surface waters		

4.1.2 Codelists for **Physical Characteristics of Groundwater Monitoring Stations** table

4.1.2.1 **Type of Use (Released at 24 Oct 2013)** codelist

Value	Definition	Short Description
DRW	Drinking Water	
IND	Industrial Supply	
OTH	Other Use	
SUR	Surveillance	

4.1.2.2 **Well or Spring (Released at 24 Oct 2013)** codelist

Value	Definition	Short Description
S	Spring	
T	Shaft	
W	Well	

4.1.3 Codelists for **Nutrients in Groundwater - Aggregated Data** table

4.1.3.1 **Determinand Code - agg (Released at 24 Oct 2013)** codelist

Value	Definition	Short Description
500	Preferred SoE Nutrients CAS: 14797-55-8	Nitrate NO3 [mg/l]
501	Preferred SoE Nutrients CAS: 14798-03-9	Ammonium NH4 [mg/l]
502	Preferred SoE Nutrients CAS: 14797-65-0	Nitrite NO2 [mg/l]
503	Preferred SoE Nutrients CAS: 7782-44-7	Dissolved Oxygen O2 [mg/l]

4.1.4 Codelists for **Nutrients, Organic Matter and General Physico-Chemical Determinands in Groundwater - Disaggregated Data** table

4.1.4.1 **Determinand Code (Released at 24 Oct 2013)** codelist

Value	Definition	Short Description
1279		pH [pH]
1289		Total organic carbon (TOC) [mg/l]
1300	CAS: 7723-14-0	Phosphorus [mg/l]
1301	CAS: 14265-44-2	Total phosphates [mg/l]
1302	CAS: 14265-44-2	Orthophosphates [mg/l]
1303		CODMn [mg/l]
1304		CODCr [mg/l]
1321		Total Nitrogen [mg/l]
1342		Temperature (water) [°C]
1343		Electrical conductivity [µS/cm]
1344		Hydrogen carbonate (Bicarbonate) HCO3 [mg/l]
1350		Total dissolved solids [mg/l]
500	Preferred SoE Nutrients CAS: 14797-55-8	Nitrate [mg/l]
501	Preferred SoE Nutrients CAS: 14798-03-9	Ammonium [mg/l]
502	Preferred SoE Nutrients CAS: 14797-65-0	Nitrite [mg/l]
503	Preferred SoE Nutrients CAS: 7782-44-7	Dissolved Oxygen [mg/l]

4.1.5 Codelists for **Hazardous Substances and Other Chemical Determinands in Groundwater - Disaggregated Data** table

4.1.5.1 **Determinand Code (Released at 24 Oct 2013)** codelist

Value	Definition	Short Description
1106	CAS: 79-00-5	1,1,2-trichloroethane [µg/l]
1107	CAS: 108-90-7	Chlorobenzene [µg/l]
1108	CAS: 95-50-1	1,2-dichlorobenzene [µg/l]
1109	CAS: 541-73-1	1,3-dichlorobenzene [µg/l]
1110	CAS: 75-01-4	Chloroethene (vinylchloride) [µg/l]
1111	Preferred SoE Haz.Subst. CAS: 79-01-6	1,1,2-trichloroethene [µg/l]
1112	Preferred SoE Haz.Subst. CAS: 127-18-4	1,1,2,2-tetrachloroethene [µg/l]
1113	Preferred SoE Haz.Subst. CAS: 56-23-5	Tetrachloromethane [µg/l]
1114	CAS: 108-88-3	Toluene [µg/l]
1115	CAS: 1330-20-7	Xylene [µg/l]
1116	CAS: NA	Dichlorobenzene [µg/l]
1117	CAS: 85-01-8	Phenanthrene [µg/l]
1118	CAS: 100-42-5	Styrene [µg/l]
1119	CAS: NA	Dichlorophenol [µg/l]
1120	CAS: 88-06-2	2,4,6-trichlorophenol [µg/l]
1121	CAS: 95-95-4	2,4,5-trichlorophenol [µg/l]

Value	Definition	Short Description
1122	CAS: 60145-21-3	PCB103 [µg/l]
1123	CAS: 70362-41-3	PCB106 [µg/l]
1130	CAS: 139-13-9	NTA [µg/l]
1131	CAS: 60-00-4	EDTA [µg/l]
1132	CAS: 156-59-2	Cis-1,2-dichloroethene [µg/l]
1133	CAS: 1634-04-4	MTBE [µg/l]
1134	CAS: 75-69-4	Trichlorofluoromethane [µg/l]
1135	CAS: 75-71-8	Dichlorodifluoromethane [µg/l]
1136	CAS: 75-27-4	Dichlorobromomethane [µg/l]
1143	CAS: 333-41-5	Diazinon [µg/l]
1205	CAS: 7440-42-8	Boron [µg/l]
1206	CAS: 57-12-5	Cyanides (as total CN) [µg/l]
1207	CAS: 7439-93-2	Lithium [µg/l]
1208	CAS: 7440-28-0	Thallium [µg/l]
1209	CAS: 3397-62-4	Deisopropyldeethylatrazine [µg/l]
1210	CAS: 30125-63-4	Desethylterbuthylazine [µg/l]
1211	CAS: 16655-82-6	3-hydroxycarbofuran [µg/l]
1212	CAS: 2599-11-3	Hydroxysimazine [µg/l]
1213	CAS: 1194-65-6	Dichlobenil [µg/l]
1214	CAS: 2008-58-4	2,6-dichlorobenzamide [µg/l]
1215	CAS: 100-02-7	Nitrophenol [µg/l]
1216	CAS: 75-99-0	Dalapon [µg/l]
1217	CAS: 2163-68-0	Hydroxyatrazine [µg/l]
1218	CAS: 1689-84-5	Bromoxynil [µg/l]
1219	CAS: 1698-60-8	Chloridazon [µg/l]
1220	CAS: 64902-72-3	Chlorsulfuron [µg/l]
1221	CAS: 26225-79-6	Ethofumesate [µg/l]
1222	CAS: 96-45-7	Ethylenethiourea (ETU) [µg/l]
1223	CAS: 67564-91-4	Fenpropimorph [µg/l]
1224	CAS: 1071-83-6	Glyphosate [µg/l]
1225	CAS: 51235-04-2	Hexazinone [µg/l]
1226	CAS: 1689-83-4	Ioxynil [µg/l]
1227	CAS: 2164-08-1	Lenacil [µg/l]
1228	CAS: 123-33-1	Maleinhydrazid [µg/l]
1229	CAS: 74223-64-6	Metsulfuronmethyl [µg/l]
1230	CAS: 40487-42-1	Pendimethalin [µg/l]
1231	CAS: 23103-98-2	Pirimicarb [µg/l]
1232	CAS: 60207-90-1	Propiconazole [µg/l]
1233	CAS: 66753-07-9	Hydroxyterbuthylazine [µg/l]
1234	CAS: 534-52-1	Dinitro-o-cresol (DNOC) [µg/l]
1235	CAS: 88-85-7	Dinoseb [µg/l]
1236	CAS: 76-03-9	Trichloroacetic acid [µg/l]
1237	CAS: 1563-66-2	Carbofuran [µg/l]
1238	CAS: 1066-51-9	Aminomethylphosphonic acid (AMPA) [µg/l]
1239	CAS: 106-93-4	1,2-dibromoethane [µg/l]
1240	CAS: 9016-45-9	Nonylphenolethoxylate [µg/l]
1242	CAS: 151-21-3	Sodium dodecyl sulfate [µg/l]
1246	Preferred SoE Haz.Subst. CAS: 71-55-6	1,1,1-trichloroethane [µg/l]

Value	Definition	Short Description
1247	CAS: 25167-83-3	Tetrachlorophenols [µg/l]
1248	CAS: 106-42-3	P-xylene [µg/l]
1249	CAS: 108-38-3	M-xylene [µg/l]
1250	CAS: 95-47-6	O-xylene [µg/l]
1251	CAS: 108-95-2	Phenol [µg/l]
1252	CAS: 59-50-7	3-methyl-4-chlorophenol [µg/l]
1253	CAS: 526-75-0	2,3-dimethyl-phenol [µg/l]
1254	CAS: 95-48-7	2-methyl-phenol [µg/l]
1255	CAS: 106-44-4	4-methyl-phenol [µg/l]
1256	CAS: 95-65-8	3,4-dimethyl-phenol [µg/l]
1257	CAS: 133-53-9	3,5-dimethyl-phenol [µg/l]
1258	CAS: 576-26-1	2,6-dimethyl-phenol [µg/l]
1259	CAS: 105-67-9	2,4-dimethyl-phenol [µg/l]
1260	CAS: 1570-64-5	4-chloro-2-methylphenol [µg/l]
1261	CAS: 85540-50-1	6-chloro-3-dimethylphenol [µg/l]
1262	CAS: 120-83-2	2,4-dichlorophenol [µg/l]
1263	CAS: 1570-65-6	4,6-dichloro-2-methylphenol [µg/l]
1264	CAS: 87-65-0	2,6-dichlorophenol [µg/l]
1265	CAS: 58-90-2	2,3,4,6-tetrachlorophenol [µg/l]
1266	CAS: 4901-51-3	2,3,4,5-tetrachlorophenol [µg/l]
1267	CAS: NA	octylphenols [µg/l]
1268	CAS: 95-87-4	2,5-dimethylphenol [µg/l]
1269	CAS: 100-41-4	Ethylbenzene [µg/l]
1270	CAS: 104-35-8	4-nonylphenol mono-ethoxylate (NP1EO) [µg/l]
1271	CAS: 20427-84-3	4-nonylphenol di-ethoxylate (NP2EO) [µg/l]
1272	CAS: NA	Adsorbable organic halogens (AOX) [µg/l]
1273	CAS: NA	Volatile organic halogens (VOX) [µg/l]
1275	CAS: 5598-13-0	Chlorpyrifos-methyl [µg/l]
1276	CAS: 17040-19-6	Demeton-S-methylsulfon [µg/l]
1277	CAS: 7439-89-6	Iron and its compounds [µg/l]
1278	CAS: 7439-96-5	Manganese and its compounds [µg/l]
1280	CAS: NA	Sulphate [mg/l]
1281	CAS: 24959-67-9	Bromide [µg/l]
1282	CAS: 20461-54-5	Iodide [µg/l]
1291	Preferred SoE Haz.Subst. CAS: 7440-50-8	Copper dissolved [µg/l]
1292	Preferred SoE Haz.Subst. CAS: 7440-66-6	Zinc dissolved [µg/l]
1293	Preferred SoE Haz.Subst. CAS: 7440-43-9	Cadmium dissolved [µg/l]
1294	Preferred SoE Haz.Subst. CAS: 7440-02-0	Nickel dissolved [µg/l]
1295	Preferred SoE Haz.Subst. CAS: 7439-92-1	Lead dissolved [µg/l]
1296	CAS: 50-00-0	Formaldehyde [µg/l]
1314	CAS: 7429-90-5	Aluminium dissolved [µg/l]
1315	Preferred SoE Haz.Subst. CAS: 7440-38-2	Arsenic dissolved [µg/l]
1316	CAS: 7439-96-5	Manganese dissolved [µg/l]
1317	Preferred SoE Haz.Subst. CAS: 7439-97-6	Mercury dissolved [µg/l]
1318	Preferred SoE Haz.Subst. CAS: 7440-47-3	Chromium dissolved [µg/l]
1319	CAS: 7439-89-6	Iron dissolved [µg/l]
1320	CAS: 1702-17-6	Clopyralid [µg/l]
1322	CAS: 13684-56-5	Desmedipham [µg/l]

Value	Definition	Short Description
1323	CAS: NA	Chloride [mg/l]
1324	CAS: 3307-39-9	2-(4-chlorophenoxy)propionic acid (4-CPP) [µg/l]
1325	CAS: 1113-02-6	Omethoate [µg/l]
1329	CAS: 133855-98-8	Epoxiconazole [µg/l]
1330	CAS: 26259-45-0	Secbumeton [µg/l]
1331	CAS: 834-12-8	Ametryn [µg/l]
1332	CAS: 1610-18-0	Prometon [µg/l]
1333	CAS: 57-74-9	Chlordane [µg/l]
1334	CAS: 2385-85-5	Mirex [µg/l]
1335	CAS: 2310-17-0	Phosalone [µg/l]
1336	CAS: 122-14-5	Fenitrothion [µg/l]
1337	CAS: 25140-90-3	2-(2,6-dichlorophenoxy)propionic acid (2,6-DCPP) [µg/l]
1338	CAS: 52236-30-3	Desamino-diketo-metribuzin [µg/l]
1339	CAS: 56507-37-0	Diketo-metribuzin [µg/l]
1340	CAS: 50-30-6	2,6-dichlorobenzoic acid [µg/l]
1345	CAS: NA	Chlorine Cl- [mg/l]
1346	CAS: 7440-70-2	Calcium [mg/l]
1347	CAS: 7439-95-4	Magnesium [mg/l]
1348	CAS: 7440-23-5	Sodium [mg/l]
1349	CAS: 7440-09-7	Potassium [mg/l]
1351	CAS: 7782-41-4	Fluorine [mg/l]
1353	CAS: 7440-32-6	Titanium [mg/l]
504	Preferred SoE Haz.Subst. CAS: 15972-60-8	Alachlor [µg/l]
505	Preferred SoE Haz.Subst. CAS: 1912-24-9	Atrazine [µg/l]
506	Preferred SoE Haz.Subst. CAS: 470-90-6	Chlorfenvinphos [µg/l]
507	Preferred SoE Haz.Subst. CAS: 2921-88-2	Chlorpyrifos [µg/l]
508	Preferred SoE Haz.Subst. CAS: 330-54-1	Diuron [µg/l]
509	CAS: 115-29-7	Endosulfan [µg/l]
510	Preferred SoE Haz.Subst. CAS: 58-89-9	Gamma-HCH (Lindane) [µg/l]
511	Preferred SoE Haz.Subst. CAS: 118-74-1	Hexachlorobenzene (HCB) [µg/l]
512	Preferred SoE Haz.Subst. CAS: 34123-59-6	Isoproturon [µg/l]
513	Preferred SoE Haz.Subst. CAS: 122-34-9	Simazine [µg/l]
514	Preferred SoE Haz.Subst. CAS: 1582-09-8	Trifluralin [µg/l]
800	CAS: NA	DDT total according Directive 2008/105/EC = (DDT, o,p' + DDT, p,p' + DDE, p,p' + DDD, p,p') [µg/l]
801	CAS: NA	DDT sum = (DDT, o,p' + DDT, p,p') [µg/l]
802	CAS: 7440-36-0	Antimony [µg/l]
803	Preferred SoE Haz.Subst. CAS: 7440-38-2	Arsenic [µg/l]
804	Preferred SoE Haz.Subst. CAS: 7440-50-8	Copper [µg/l]
805	CAS: 7440-39-3	Barium [µg/l]
806	CAS: 7440-41-7	Beryllium [µg/l]
807	Preferred SoE Haz.Subst. CAS: 7440-66-6	Zinc [µg/l]
808	Preferred SoE Haz.Subst. CAS: 7440-43-9	Cadmium [µg/l]
809	CAS: 7440-48-4	Cobalt and its compounds [µg/l]
810	CAS: 7440-31-5	Tin and its compounds [µg/l]
811	CAS: NA	Chromium 3+ [µg/l]
812	CAS: NA	Chromium 6+ [µg/l]

Value	Definition	Short Description
813	Preferred SoE Haz.Subst. CAS: 7440-47-3	Chromium [µg/l]
814	CAS: 7439-98-7	Molybdenum and its compounds [µg/l]
815	Preferred SoE Haz.Subst. CAS: 7440-02-0	Nickel [µg/l]
816	CAS: 7782-49-2	Selenium and its compounds [µg/l]
817	CAS: 7440-22-4	Silver [µg/l]
818	CAS: 7440-24-6	Strontium [µg/l]
819	Preferred SoE Haz.Subst. CAS: 7439-92-1	Lead [µg/l]
820	CAS: 7440-61-1	Uranium [µg/l]
821	CAS: 7440-62-2	Vanadium and its compounds [µg/l]
822	Preferred SoE Haz.Subst. CAS: 7439-97-6	Mercury [µg/l]
823	Preferred SoE Haz.Subst. CAS: 87-86-5	Pentachlorophenol [µg/l]
824	CAS: 51218-45-2	Metolachlor [µg/l]
825	Preferred SoE Haz.Subst. CAS: 309-00-2	Aldrin [µg/l]
826	CAS: NA	DDT + DDE + DDD total = (DDT, o,p' + DDT, p,p' + DDE, o,p' + DDE, p,p' + DDD, o,p' + DDD, p,p') [µg/l]
827	Preferred SoE Haz.Subst. CAS: 789-02-6	DDT, o,p' [µg/l]
828	Preferred SoE Haz.Subst. CAS: 50-29-3	DDT, p,p' [µg/l]
829	CAS: NA	DDE sum = (DDE, o,p' + DDE, p,p') [µg/l]
830	CAS: 3424-82-6	DDE, o,p' [µg/l]
831	Preferred SoE Haz.Subst. CAS: 72-55-9	DDE, p,p' [µg/l]
832	CAS: NA	DDD (TDE) sum = (DDD, o,p' + DDD, p,p') [µg/l]
833	CAS: 53-19-0	DDD, o,p' [µg/l]
834	Preferred SoE Haz.Subst. CAS: 72-54-8	DDD, p,p' [µg/l]
835	Preferred SoE Haz.Subst. CAS: 60-57-1	Dieldrin [µg/l]
836	Preferred SoE Haz.Subst. CAS: 72-20-8	Endrin [µg/l]
837	CAS: 76-44-8	Heptachlor [µg/l]
838	CAS: 1024-57-3	Heptachlorepoxide [µg/l]
839	Preferred SoE Haz.Subst. CAS: 25057-89-0	Bentazone [µg/l]
840	Preferred SoE Haz.Subst. CAS: 319-84-6	Alpha-HCH [µg/l]
841	Preferred SoE Haz.Subst. CAS: 319-85-7	Beta-HCH [µg/l]
842	CAS: 319-86-8	Delta-HCH [µg/l]
843	CAS: 6108-10-07	Epsilon-HCH [µg/l]
844	CAS: 608-73-1	Hexachlorocyclohexane (HCH) [µg/l]
845	Preferred SoE Haz.Subst. CAS: 959-98-8	Alpha-Endosulfan [µg/l]
846	CAS: 33213-65-9	Beta-Endosulfan [µg/l]
847	CAS: 56-38-2	Parathion [µg/l]
848	CAS: 298-00-0	Parathion-methyl [µg/l]
849	Preferred SoE Haz.Subst. CAS: 6190-65-4	Desethylatrazine [µg/l]
850	Preferred SoE Haz.Subst. CAS: 1007-28-9	Desisopropylatrazine [µg/l]
851	Preferred SoE Haz.Subst. CAS: 139-40-2	Propazine [µg/l]
852	Preferred SoE Haz.Subst. CAS: 7287-19-6	Prometryn [µg/l]
853	CAS: 21725-46-2	Cyanazine [µg/l]
854	Preferred SoE Haz.Subst. CAS: 5915-41-3	Terbuthylazine [µg/l]
855	Preferred SoE Haz.Subst. CAS: 886-50-0	Terbutryn [µg/l]
856	CAS: 33693-04-8	Terbumeton [µg/l]
857	CAS: 41394-05-2	Metamitron [µg/l]
858	CAS: 21087-64-9	Metribuzin [µg/l]

Value	Definition	Short Description
859	CAS: 314-40-9	Bromacil [µg/l]
860	CAS: 55512-33-9	Pyridate [µg/l]
861	Preferred SoE Haz.Subst. CAS: 94-75-7	2,4-D [µg/l]
862	CAS: 120-36-5	Dichlorprop (2,4-DP) [µg/l]
863	CAS: 93-76-5	2,4,5-T [µg/l]
864	Preferred SoE Haz.Subst. CAS: 94-74-6	MCPA [µg/l]
865	CAS: 94-81-5	MCPB [µg/l]
866	Preferred SoE Haz.Subst. CAS: 7085-19-0	Mecoprop [µg/l]
867	CAS: 7286-69-3	Sebuthylazine [µg/l]
868	CAS: 93-72-1	Fenoprop [µg/l]
869	CAS: 94-82-6	2,4-DB [µg/l]
870	CAS: 1918-00-9	Dicamba [µg/l]
871	CAS: 57837-19-1	Metalaxyl [µg/l]
872	CAS: 67129-08-2	Metazachlor [µg/l]
873	CAS: 34256-82-1	Acetochlor [µg/l]
874	CAS: NA	Pesticides (sum) [µg/l]
875	CAS: 72-43-5	Methoxychlor [µg/l]
876	CAS: 7429-90-5	Aluminium and its compounds [µg/l]
877	Preferred SoE Haz.Subst. CAS: 120-12-7	Anthracene [µg/l]
878	Preferred SoE Haz.Subst. CAS: 71-43-2	Benzene [µg/l]
879	CAS: NA	Brominated diphenylethers [µg/l]
880	CAS: 1163-19-5	Bis(pentabromophenyl) ether [µg/l]
881	CAS: 32536-52-0	Diphenyl ether, octabromo deriviate [µg/l]
882	CAS: 32534-81-9	Diphenyl ether, pentabromo derivative [µg/l]
883	Preferred SoE Haz.Subst. CAS: 85535-84-8	Chloroalkanes C10-13 [µg/l]
884	Preferred SoE Haz.Subst. CAS: 107-06-2	1,2-dichloroethane [µg/l]
885	Preferred SoE Haz.Subst. CAS: 75-09-2	Dichloromethane [µg/l]
886	Preferred SoE Haz.Subst. CAS: 117-81-7	Di (2-ethylhexyl) phthalate (DEHP) [µg/l]
887	CAS: 84-66-2	Di-ethyl phthalate [µg/l]
888	CAS: 84-69-5	Di-iso-butyl phthalate [µg/l]
889	CAS: 84-74-2	Di-(n-butyl) phthalate [µg/l]
890	CAS: 85-68-7	Butyl benzyl phthalate (BBP) [µg/l]
891	Preferred SoE Haz.Subst. CAS: 87-68-3	Hexachlorobutadiene (HCBD) [µg/l]
892	Preferred SoE Haz.Subst. CAS: 91-20-3	Naphthalene [µg/l]
893	CAS: 25154-52-3	Nonylphenol [µg/l]
894	Preferred SoE Haz.Subst. CAS: 104-40-5	4-nonylphenol [µg/l]
895	CAS: 84852-15-3	4-nonylphenol, branched [µg/l]
896	CAS: 1806-26-4	4-octylphenol [µg/l]
897	Preferred SoE Haz.Subst. CAS: 140-66-9	Para-tert-octylphenol [µg/l]
898	Preferred SoE Haz.Subst. CAS: 608-93-5	Pentachlorobenzene [µg/l]
899	CAS: NA	Polyaromatic hydro-carbons (PAH) (sum) [µg/l]
900	Preferred SoE Haz.Subst. CAS: 50-32-8	Benzo(a)pyrene [µg/l]
901	Preferred SoE Haz.Subst. CAS: 205-99-2	Benzo(b)fluoranthene [µg/l]
902	Preferred SoE Haz.Subst. CAS: 191-24-2	Benzo(g,h,i)perylene [µg/l]
903	Preferred SoE Haz.Subst. CAS: 207-08-9	Benzo(k)fluoranthene [µg/l]
904	Preferred SoE Haz.Subst. CAS: 206-44-0	Fluoranthene [µg/l]
905	Preferred SoE Haz.Subst. CAS: 193-39-5	Indeno(1,2,3-cd)pyrene [µg/l]
906	CAS: 688-73-3	Tributyltin compounds [µg/l]

Value	Definition	Short Description
907	CAS: 36643-28-4	Tributyltin cation [µg/l]
908	CAS: 12002-48-1	Trichlorobenzenes [µg/l]
909	CAS: 87-61-6	1,2,3-trichlorobenzene [µg/l]
910	CAS: 120-82-1	1,2,4-trichlorobenzene [µg/l]
911	CAS: 108-70-3	1,3,5-trichlorobenzene [µg/l]
912	Preferred SoE Haz.Subst. CAS: 67-66-3	Trichloromethane [µg/l]
913	CAS: 919-86-8	Demeton-S-methyl [µg/l]
914	CAS: 62-73-7	Dichlorvos [µg/l]
915	CAS: 60-51-5	Dimethoate [µg/l]
916	Preferred SoE Haz.Subst. CAS: 465-73-6	Isodrin [µg/l]
917	Preferred SoE Haz.Subst. CAS: 330-55-2	Linuron [µg/l]
918	CAS: 1336-36-3	Polychlorinated biphenyls [µg/l]
919	CAS: 31508-00-6	PCB118 [µg/l]
920	CAS: 32598-13-3	PCB77 [µg/l]
921	CAS: 32774-16-6	PCB169 [µg/l]
922	CAS: 80-05-7	Bisphenol A [µg/l]
923	CAS: 37680-73-2	PCB101 (2,2',4,5,5'-pentachlorobiphenyl) [µg/l]
924	CAS: 35065-28-2	PCB138 (2,2',3,4,4',5'-hexachlorobiphenyl) [µg/l]
925	CAS: 35065-27-1	PCB153 (2,2',4,4',5,5'-hexachlorobiphenyl) [µg/l]
926	CAS: 35065-29-3	PCB180 (2,2',3,4,4',5,5'-heptachlorobiphenyl) [µg/l]
927	CAS: 35694-08-7	PCB194 (1,2,3,4-tetrachloro-5-(2,3,4,5-tetrachlorophenyl)benzene) [µg/l]
928	CAS: 7012-37-5	PCB28 (2,4,4'-trichlorobiphenyl) [µg/l]
929	CAS: 35693-99-3	PCB52 (2,2',5,5'-tetrachlorobiphenyl) [µg/l]
930	CAS: 189084-64-8	PBDE100 (2,2',4,4',6-pentabromodiphenyl ether) [µg/l]
931	CAS: 68631-49-2	PBDE153 (2,2',4,4',5,5'-hexabromodiphenyl ether) [µg/l]
932	CAS: 207122-15-4	PBDE154 (2,2',4,4',5,6'-hexabromodiphenyl ether) [µg/l]
933	CAS: 41318-75-6	PBDE28 [µg/l]
934	CAS: 5436-43-1	PBDE47 (2,2',4,4'-tetrabromodiphenyl ether) [µg/l]
935	CAS: 60348-60-9	PBDE99 [µg/l]
936	CAS: 129-00-0	Pyrene [µg/l]
937	CAS: 53-70-3	Dibenzo(a,h)anthracene [µg/l]
938	CAS: 218-01-9	Chrysene [µg/l]
939	CAS: 56-55-3	Benzo(a)anthracene [µg/l]
940	CAS: NA	TCDD (dioxins and furans) [µg/l]

4.2 Common Elements Codelists

4.2.1 WFD station (Released at 10 Jul 2009)

4.2.1.1 Codelist

Value	Definition	Short Description
no	Station was not reported as WFD Art.8 monitoring station.	
yes	Station was reported as WFD Art.8 monitoring station.	

4.2.1.2 Referencing Tables

Table Name
Physical Characteristics of Groundwater Monitoring Stations

4.2.2 River Basin District Code (Released at 24 Jul 2012)

4.2.2.1 Codelist

Value	Definition	Short Description
AD1	Andorra	AD
AT1000	Danube	AT
AT2000	Rhine	AT
AT5000	Elbe	AT
BEEscaut_RW	Scheldt	BE
BEEscaut_Schelde_BR	Scheldt	BE
BEMaas_VL	Meuse	BE
BEMeuse_RW	Meuse	BE
BERhin_RW	Rhine	BE
BEScheldе_VL	Scheldt in Flanders	BE
BESeine_RW	Seine	BE
BG1000	Danube River Basin District	BG
BG2000	Black Sea River Basin District	BG
BG3000	East Aegean River Basin District	BG
BG4000	West Aegean River Basin District	BG
CH10	Rhine	CH
CH50	Rhone	CH
CH60	Po	CH
CH80	Danube	CH
CH90	Adige	CH
CY001	Cyprus	CY
CZ_1000	Danube	CZ
CZ_5000	Elbe	CZ
CZ_6000	Oder	CZ
DE1000	Danube River Basin District	DE
DE2000	Rhine River Basin District	DE
DE3000	Ems River Basin District	DE
DE4000	Weser River Basin District	DE
DE5000	German Elbe	DE
DE6000	Oder	DE

Value	Definition	Short Description
DE7000	Maas River Basin District (German Part)	DE
DE9500	Eider	DE
DE9610	Schlei/Trave	DE
DE9650	Warnow/Peene	DE
DK1	Jutland and Funen	DK
DK2	Zealand	DK
DK3	Bornholm	DK
DK4	Vidaa-Krusaa	DK
EE1	West-Estonian River Basin District	EE
EE2	East-Estonian River Basin District	EE
EE3	Koiva River Basin District	EE
ES010	Minho	ES
ES014	Galician Coast	ES
ES015	Basque County Internal Basins	ES
ES016	Cantabrian	ES
ES020	Duero	ES
ES030	Tagus	ES
ES040	Guadiana River Basin District	ES
ES050	Guadalquivir	ES
ES060	Andalusia Mediterranean Basins	ES
ES063	Guadalete and Barbate	ES
ES064	Tinto, Odiel and Piedras	ES
ES070	Segura	ES
ES080	Jucar	ES
ES091	Ebro	ES
ES100	Catalan River Basin District	ES
ES110	Balearic Islands	ES
ES120	Gran Canaria	ES
ES122	Fuerteventura	ES
ES123	Lanzarote	ES
ES124	Tenerife	ES
ES125	La Palma	ES
ES126	La Gomera	ES
ES127	El Hierro	ES
ES150	Ceuta	ES
ES160	Melilla	ES
FIVHA1	Vuoksi River Basin District	FI
FIVHA2	Kymijoki-Gulf of Finland River Basin District	FI
FIVHA3	Kokemäenjoki-Archipelago Sea-Bothnian Sea River Basin District	FI
FIVHA4	Oulujoki-Iijoki River Basin District	FI
FIVHA5	Kemijoki River Basin District	FI
FIVHA6	Tornionjoki IRBD	FI
FIVHA7	Teno, Näätämöjoki and Paatsjoki IRBD	FI
FIWDA	Åland River Basin District	FI
FRA	L'Escaut, la Somme et les cours d'eau côtiers de la Manche et de la mer du Nord	FR
FRB1	Meuse	FR
FRB2	La Sambre	FR

Value	Definition	Short Description
FRC	Rhine	FR
FRD	Le Rhône et les cours d'eau côtiers méditerranéens	FR
FRE	Les cours d'eau de la Corse	FR
FRF	L'Adour, la Garonne, la Dordogne, la Charente et les cours d'eau côtiers charentais et aquitains	FR
FRG	La Loire, les cours d'eau côtiers vendéens et bretons	FR
FRH	La Seine et les cours d'eau côtiers normands	FR
FRI	Les cours d'eau de la Guadeloupe	FR
FRJ	Les cours d'eau de la Martinique	FR
FRK	Les fleuves et cours d'eau côtiers de la Guyane	FR
FRL	Les cours d'eau de la Réunion	FR
GBNIIENB	Neagh Bann	IE; !!! RBDcode starts with abbreviation GB
GBNIIENW	North Western	IE; !!! RBDcode starts with abbreviation GB
GR01	Western Peloponnese	GR
GR02	Northern Peloponnese	GR
GR03	Eastern Peloponnese	GR
GR04	Western Sterea Ellada	GR
GR05	Epirus	GR
GR06	Attica	GR
GR07	Eastern Sterea Ellada	GR
GR08	Thessalia	GR
GR09	Western Macedonia	GR
GR10	Central Macedonia	GR
GR11	Eastern Macedonia	GR
GR12	Thrace	GR
GR13	Crete	GR
GR14	Aegean Islands	GR
HU1000	Hungarian part of the Danube River Basin District	HU
IEEA	Eastern	IE
IEGBNISH	Shannon	IE
IESE	South Eastern	IE
IESW	South Western	IE
IEWE	Western	IE
IS1	Iceland	IS; RBD not included in the RBD GIS reference layer v1.4
ITA	Eastern Alps	IT
ITB	Po Basin	IT
ITC	Northern Appenines	IT
ITD	Serchio	IT
ITE	Middle Appenines	IT
ITF	Southern Appenines	IT
ITG	Sardinia	IT
ITH	Sicily	IT
LI1	Liechtenstein	LI
LT1100	Nemunas River Basin District	LT
LT2300	Venta River Basin District	LT

Value	Definition	Short Description
LT3400	Lielupe River Basin District	LT
LT4500	Dauguva River Basin District	LT
LU RB_000	Mosel	LU; !!! RBDcode contains space
LU RB_001	Chiers	LU; !!! RBDcode contains space
LVDUBA	Daugava river basin district	LV
LVGUBA	Gauja river basin district	LV
LVLUBA	Lielupe river basin district	LV
LVVUBA	Venta river basin district	LV
MC1	Monaco	MC
MTMALTA	Malta	MT
NLEM	Ems	NL
NLMS	Meuse	NL
NLRN	Rhine	NL
NLSC	Scheldt	NL
NO1101	Moere og Romsdal	NO
NO1102	Troendelag	NO
NO1103	Nordland	NO
NO1104	Troms	NO
NO1105	Finnmark	NO
NO5101	Glomma	NO
NO5102	West Bay	NO
NO5103	Agder	NO
NO5104	Rogaland	NO
NO5105	Hordaland	NO
NO5106	Sogn and Fjordane	NO
NOFIVHA5	Kemijoki	NO
NOFIVHA6	Tornionjoki (Finnish part)	NO
NOSE1	Bothnian Bay	NO
NOSE1TO	Torne River	NO
NOSE2	Bothnian Sea	NO
NOSE5	Skagerrak and Kattegat	NO
PL1000	Danube River Basin District	PL
PL2000	Vistula River Basin District	PL
PL3000	Swieza River Basin District	PL
PL4000	Jarft River Basin District	PL
PL5000	Elbe River Basin District	PL
PL6000	Oder River Basin District	PL
PL6700	Ucker River Basin District	PL
PL7000	Pregolya River Basin District	PL
PL8000	Nemunas River Basin District	PL
PL9000	Dniester River Basin District	PL
PTRH1	Minho and Lima	PT
PTRH10	Madeira	PT
PTRH2	Cavado, Ave and Leca	PT
PTRH3	Douro	PT
PTRH4	Vouga, Mondego and Lis	PT
PTRH5	Tagus and Western Basins	PT
PTRH6	Sado and Mira	PT

Value	Definition	Short Description
PTRH7	Guadiana	PT
PTRH8	Algarve Basins	PT
PTRH9	Azores	PT
RO1000	Danube River Basin District	RO
SE1	1. Bothnian Bay (Sweden)	SE
SE1TO	1. Bothnian Bay (International district Torne river - Sweden)	SE
SE2	2. Bothnian Sea (Sweden)	SE
SE3	3. North Baltic Sea (Sweden)	SE
SE4	4. South Baltic Sea (Sweden)	SE
SE5	5. Skagerrak and Kattegat (Sweden)	SE
SENO1102	2. Bothnian Sea (International drainage basin Trondelagsfylkene - Sweden)	SE
SENO1103	1. Bothnian Bay (International drainage basin Nordland - Sweden)	SE
SENO1104	1. Bothnian Bay (International drainage basin Troms - Sweden)	SE
SENO5101	5. Skagerrak and Kattegat (International drainage basin Glomma - Sweden)	SE
SI_RBD_1	Danube	SI
SI_RBD_2	North Adriatic	SI
SK30000	Vistula	SK
SK40000	Danube	SK
TR01	Meric Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR02	Marmara Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR03	Susurluk Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR04	Kuzey Ege Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR05	Gediz Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR06	K. Menderes Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR07	B. Menderes Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR08	Bati Akdeniz Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR09	Antalya Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR10	Burdur Lakes Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR11	Akarcay Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR12	Sakarya Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR13	Bati Karadeniz Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR14	Yesilirmak Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR15	Kizilirmak Basin	TR; RBD not included in the RBD GIS reference layer v1.4

Value	Definition	Short Description
TR16	Konya Closed Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR17	Dogu Akdeniz Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR18	Seyhan Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR19	Asi Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR20	Ceyhan Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR21	Firat ve Dicle Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR22	Dogu Karadeniz Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR23	Coruh Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR24	Aras Basin	TR; RBD not included in the RBD GIS reference layer v1.4
TR25	Van Lake Basin	TR; RBD not included in the RBD GIS reference layer v1.4
UK01	Scotland	GB
UK02	Solway Tweed	GB
UK03	Northumbria	GB
UK04	Humber	GB
UK05	Anglian	GB
UK06	Thames	GB
UK07	South East	GB
UK08	South West	GB
UK09	Severn	GB
UK10	Western Wales	GB
UK11	Dee	GB
UK12	North West	GB
UKGBNIIENB	Neagh Bann	GB
UKGBNIIENW	North Western	GB
UKGBNINE	North Eastern	GB
UKGI17	Gibraltar	GB

4.2.2.2 Referencing Tables

Table Name
Groundwater Body Characteristics and Pressures