D4.1 Specification and Creation of the Database

Synergetics | Synergies for Green Transformation of Inland and Coastal Shipping

GRANT AGREEMENT NO.	101096809
DURATION OF THE PROJECT	42 months
DELIVERABLE NUMBER	D4.1
DISSEMINATION LEVEL	Public
DELIVERABLE LEADER	DST
STATUS	FINAL
SUBMISSION DATE	31-03-2024
AUTHOR	Marko Josipović josipovic@dst-org.de

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

Table of Content

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

List of Figures

List of Abbreviations

Abbreviation	Definition
DBMS	Database-management systems
DDL	Data Definition Language
DML	Data Manipulation Language
GUI	Graphical User Interface
MFA	Multi-Factor Authentication
PoLP	Principle of Least Privilege
RBAC	Role-Based Access Control
SQL	Structured Query Language
WP	Work Package

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

Modification Control

Version #	Date	Author	Organisation
V0.1	11-03-2024	M. Josipović	DST
V0.2	28-03-2024	M. Josipović	DST
V1.0	31-03-2024	B. Friedhoff	DST

Release Approval

Name	Role	REMARKS
B. Friedhoff – DST	WP-Leader/Coordinator	31-03-2024
A. Grasman – MARIN	Reviewer	27-03-2024
N. Kreukniet – SPB	Reviewer	20-03-2024
M. Quispel – SPB	Reviewer	20-03-2024

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

Executive Summary

This report of Task 4.1 outlines the specification and creation of the SYNERGETICS Database, focusing on developing of a scalable data management infrastructure for green technologies in inland and coastal shipping. The database design utilizes PostgreSQL for its robust, open-source architecture, considering relational data management. The list of relevant potential elements and parameters is given, prioritizing technological specifications, energy efficiency metrics, environmental impacts, and economic considerations, which will foster data-driven decisions in advancing green technologies. An important aspect of the report is the integration of the new/existing databases with a cloud server. The methodology described in the report is based on security, data integrity, and continuous improvement, which is evident in the implementation of access control, authentication measures, and the proactive maintenance and updating regimen.

As the project progresses, the groundwork laid in Task 4.1 will become pivotal for ongoing and future initiatives, serving as a crucial step for Task 4.2 by facilitating streamlined data collection and visualization. Simultaneously, it will act as a hub within Work Package 4, interlinking various databases and tools, thereby enhancing the project's overall efficacy and integration.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

1. Introduction

To elevate the data infrastructure of the SYNERGETICS project and establish a foundation for a scalable data management solution, Task 4.1, reported herein forms a crucial building block. This segment of the project elaborates on the plans and methodologies for accomplishing the task, providing a comprehensive overview of the strategy and implementation of the database. The Grant Agreement (GA) describes the task as follows:

"Task 4.1 Specification and creation of the database [M9-M15]

In the first task of WP4 the database, which will serve as a basis for the Catalogue of greening technologies, the SYNERGETICS tools and dissemination material, is specified. This includes the selection of relevant elements without overloading. The ESSF-SAPS database and studies available in the consortium will be used as a starting point.

The SYNERGETICS database shall not be limited to parameters like energy contents, densities, packaging factors, efficiencies or cost-predictions. References to external sources, worked examples (pilots), short text passages and illustrations will be included as well. The specification is also used to prepare an input collector for WPs 1 to 3.

The platform for the implementation of the database will be selected with a focus on interfaces to the SYNERGETICS tools and futureproof ways for updating and maintenance. At the end of Task 4.1 the database is set up and tested with exemplary data. The structure, maintenance concept, and interfaces are described in a report (D4.1).

Involved Partners: DST (lead), SPB, MARIN, OST"

Based on the description of the task, the initial implementation of the SYNERGETICS database has been executed on a local server managed by DST, named SDB (SYNERGETICS Database). The vision extends beyond current configurations, aiming for the integration of SDB with a cloud server solution. This strategy is conceived to maintain the integrity and confidentiality of existing data on private servers, mitigating risks linked to direct server access. Such a cloud-based approach is also facilitating a shared database environment that enhances accessibility while safeguarding private sensitive information.

This report delves into the methodologies adopted to realize this integration, outlining maintenance and update protocols. PostgreSQL has been selected for its robust database architecture, primarily due to its adherence to the relational model, which is ideal for managing table-based data. Additionally, PostgreSQL's open-source nature eliminates the need for any licensing, making it an attractive option for the database structure. As the SYNERGETICS project evolves, the insights obtained from the initial phases of database development will be key in honing the data management strategies in the future.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

2. Database Systems

2.1 Overview

In this chapter, a concise overview of the database system is presented, laying the groundwork for a better understanding of the subsequent chapters.

Database-management systems, or DBMS, are used to store and retrieve large quantities of information. Data management involves not only defining structures for the systematic storage of information but also implementing efficiently mechanisms for manipulating this information, such as creating, reading, updating, or deleting. Moreover, these systems should ensure the integrity and security of the stored information, safeguarding against system crashes, unauthorized access, and data violations. The overall design/structure of the database is called the database schema.

The collection of data relevant to an analysis is commonly referred to as the database. This data can include but is not limited to, integers, characters, strings, floating-point numbers and arrays. To ensure the efficient usability of these data, it is essential that they can be retrieved easily.

Different levels of abstraction exist within a database system: physical/internal, logical/conceptual, and view/external. At the physical level, which is the lowest level of abstraction, the database represents how the data are actually stored employing various data structures. Moving up to the logical level, it can be noticed what data are stored in the database and the relationships that exist among them. Finally, at the view level, which is the highest level of abstraction, it describes only a part of the entire database, or more precisely this level is tailored to specific user requirements.

The fundamental framework for describing data on different levels is provided by the data model, which serves as the foundational structure underlying a database. Data models can be categorized into four main types:

1. **Relational Data Model**: This model organizes data into tables, where relationships between data entities are represented through keys. It is widely used in traditional database systems and offers a structured approach to data management.
2. **Entity-Relationship Data Model**: In this model, data is conceptualized in terms of entities and their relationships. Entities represent real-world objects or concepts, while relationships denote connections between these entities. It provides a graphical representation of the data schema, aiding in the visualization of complex data structures.
3. **Object-Based Data Model**: This model extends the concepts of the relational model by incorporating object-oriented programming principles. It allows for the representation of complex data types, inheritance, and encapsulation, making it suitable for applications with complex data structures and behaviors.
4. **Semi-structured Data Model**: Unlike the rigid structure of relational databases, the semi-structured data model allows for flexible and dynamic data schemas. It accommodates data with varying structures and supports hierarchical relationships, making it suitable for scenarios where data is evolving or unpredictable.

Before the relational data model, **network** and **hierarchical data models** were prevalent in the database field. However, the simplicity and flexibility of the relational model led to its widespread adoption, surpassing the network and hierarchical models that are nowadays used in legacy systems.

Overall, the selection of a data model depends on various factors, including the nature of the data, the specific requirements of the application, scalability, and the performance considerations of the database system. Each data model presents its own advantages and trade-offs, allowing database designers to choose the model best suited to their particular needs. Furthermore, it will be chosen and justified which data model is most suitable within the framework of the SYNERGETICS database in the following.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

2.2 Architectures

The architecture of a DBMS is a multi-layered structure that allows various user types to efficiently manage and manipulate data. The typical database system structure is represented in Figure 1.

At the highest level, different users interact with the database through customized interfaces. Naive users, such as web users, utilize application interfaces characterized by simple forms or web applications for specific tasks. Application programmers write application programs that communicate directly with the DBMS, in order to perform various operations. Sophisticated users, like analysts, use query tools to extract information from the database. In the end, database administrators employ administration tools to maintain, secure, and optimize the entire system's performance.

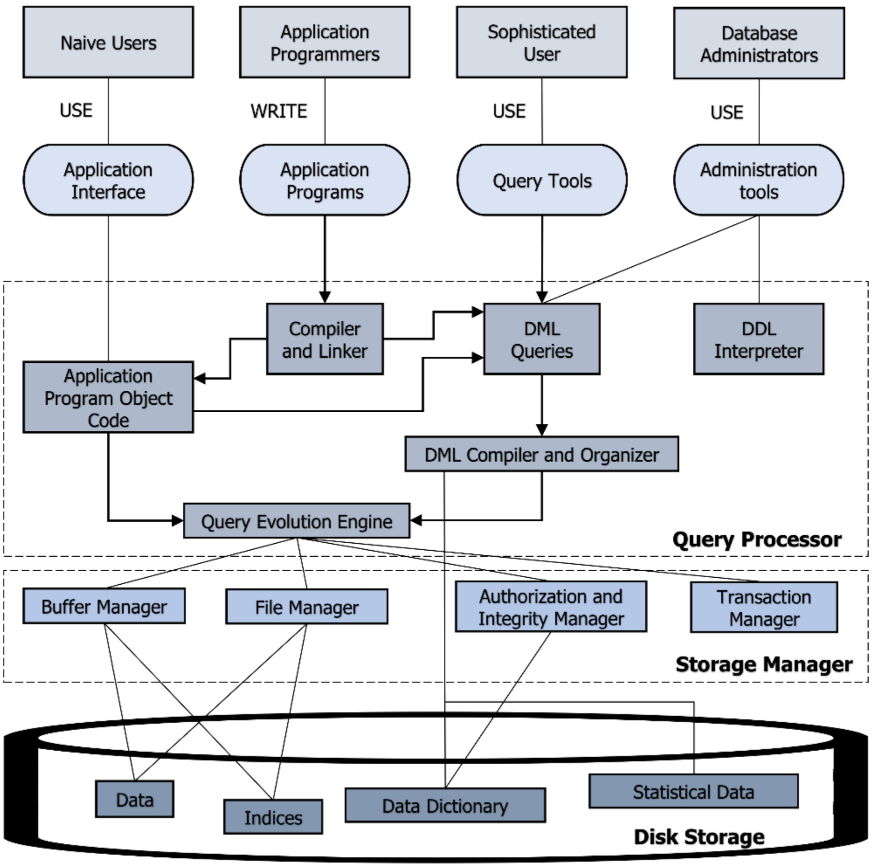

Figure 1: Database System Structure (Silberschatz, Korth, & Sudarshan, 2010).

The **Query Processor** is crucial to database functionality, enhancing the efficiency of data access and manipulation. It comprises a DDL (Data Definition Language) Interpreter, which interprets the DDL statements responsible for defining the database's structural schema and access methodologies—these statements are typically abstracted from the end-user's view. Concurrently, the DML (Data Manipulation Language) Compiler converts DML queries into a streamlined evaluation plan, a sequence of instructions that the Query Evaluation Engine can process and execute. In more simple terms, the DML language enables users to retrieve, insert, delete, or modify the data.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

In the structure of a database system, the **Storage Manager** handles the interaction between physical data and application queries. It incorporates several components:

- **Authorization and Integrity Manager** authenticated user requests and upholds data integrity, playing a key role in secure data operations.
- **Transaction Manager** ensures database consistency amidst system failures and oversees the execution of concurrent transactions, aiming to avoid operational conflicts.
- **File Manager** is in charge of allocating disk space and organizing data efficiently on the disk, contributing to the database's storage effectiveness.
- **Buffer Manager** focuses on improving database performance by managing the movement of data from disk storage to main memory and deciding on data caching strategies.

Positioned at the lowest layer of the architecture, disk storage serves as the repository where all data is stored. Together, these components support the database's operational needs, ensuring effective data management and system performance.

For further reading on database systems and their architecture, functionality, and management practices interested readers are encouraged to check Korth, Sudarshan, and Abraham Silberschatz (2010) or Sciore (2009).

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

3. SYNERGETICS Database Specification

3.1 Selection of Relevant Elements and Parameters

When determining the elements and parameters for the SYNERGETICS database, the focus should be on a balance between comprehensiveness and relevance. The criteria are centred on the project's aim to encourage advancements in greening technologies. Parameters should be carefully selected based on their potential to drive analysis and inform decision-making in the development and deployment of green technologies. In general, the SYNERGETICS database should focus on technological specifications, energy efficiency metrics, environmental impact assessments, and economic considerations. This includes following parameters, but is not limited to:

1. **Technology Description:** Detailing the operational principles and usage scenarios for alternative fuels.
2. **Alternative Energy Carriers:** Characteristics of energy carriers such as LNG, hydrogen, or biofuels in terms of their energy density both volume $[MJ/m^3]$ and weight $[MJ/kg]$ (Well-to-Tank).
3. **Characteristics of Energy Converters:** The amount of energy, in terms of both the caloric input value $[MJ]$ and the mechanical energy output $[kWh]$, along with emission factors, including but not limited to CO_2, NO_X, PM, and SO_X.
4. **Clean Technology Adoption Rate:** Evaluation of the uptake rate for various greening technologies within the fleet. This will involve leveraging insights from the T2.1 Pilot Database and ensuring the inclusion of a broad spectrum of use cases, such as LNG and emerging technologies like hydrogen, recognizing that some technologies may currently have limited application scenarios.
5. **Technological Advancement Timeline:** Estimations on how quickly new green technologies are progressing from concept to market-ready solutions (TRL).
6. **Cost of Green Tech Integration:** An assessment will be conducted on the capital expenditures (CapEx) associated with retrofitting vessels with green technologies. The analysis will consider variations across different vessel types and operational profiles, focusing on specific energy requirements i.e. kW capacity Additionally, the evaluation will factor in the retrofitting duration and the consequent opportunity costs incurred from the vessel's downtime.
7. **Operational Cost Impact**: This entails an assessment of the overall operational expenditures (OpEx) associated with the use of green technologies (based on CCNR studies), focusing on quantifiable metrics such as fuel costs and maintenance expenses. Additionally, the impact of energy storage systems on vessel payload, particularly regarding weight and volume, will also be considered.
8. **Bunkering Time Efficiency:** The time required for bunkering compared to traditional fuels and the associated logistical considerations.
9. **Regulatory Framework Tracking:** An analysis of the regulatory framework to identify how current and emerging regulations can support or impede the use of alternative energy sources for vessels. It also involves assessing the establishment of regulatory frameworks that could facilitate the development of bunkering infrastructure for these alternative fuels.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

3.2 Inclusion of Internal/External References

The database development significantly benefits from the integration of both internal insights gained from prior work within the project's work packages and external references.

To build a foundation for the SYNERGETICS database and to include potential mentioned parameters and more, in the next step, when the database starts to fill, the starting point will be the ESSF-SAPS (Sustainable Power @ MARIN) database and relevant studies within the consortium, using them as an initial benchmark. Alongside, integration of the CCNR studies (RQ C) and incorporate insights from the NEED model (NEEDS (marin.nl)) developed by MARIN and SPB will be considered. This strategy ensures that we build upon a robust dataset, enhancing it with additional relevant information.

It is imperative that the SYNERGETICS database adds unique value. By cross-referencing with the ESSF-SAPS database and others, we will avoid data duplication in the future, ensuring that our database serves as a unique and valuable resource. This will not limit the database exclusively to ESSF-SAPS sources; rather, it will enable us to have a starting point with a validated data set that is well-recognized within the consortium. Data collection between an existing database and additional data ensures the SYNERGETICS database provides comprehensive coverage of the sector.

Moreover, the database will be enhanced with insights from Task 2.1: Relevant Pilot Database, which currently considers 185 pilots. These pilots, showcasing innovative greening technologies, are separated based on their learning potential, high, possible, or low, with consideration given to those conducted internally under the SYNERGETICS project. Each pilot has undergone thorough validation to ensure relevance and potential for synergy. This knowledge base is pivotal for WP4, as it directs a focus toward real technological solutions with the highest impact, advancing the project's scope on alternative propulsion and/or energy efficiency in Inland Waterway Transport and Coastal shipping. The pilot database is accessible at: https://www.synergetics-project.eu/wp-content/uploads/2023/12/T2.1_Pilot_Database_for_external_sharing_V2.0.xlsx.

The latest academic and industry research findings will be incorporated and evaluated to ensure the SYNERGETICS database remains current, informative, and grounded in reality. Data from companies developing green technologies will be balanced with independent research to provide a practical, yet critical perspective.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

4. Database Implementation

4.1 Database Creation

The initial phase of establishing a foundational structure for the Catalogue of Greening Technologies, the SYNERGETICS tools, and dissemination material involved creating a database on the server hosted by DST. The choice of PostgreSQL as the database management system was driven by its open-source nature, extensive support for advanced data types, robustness, and compatibility with a wide array of standards and programming languages.

Looking ahead, the database is designed with the intention of integration with a cloud server, leveraging the database dump method. This strategy will be applicable not only to the DST database but also to the ESSF-SAPS database and others. It is aimed at facilitating easier sharing and collaboration with other institutions or companies within the SYNERGETICS project, thereby ensuring the integrity and confidentiality of data while mitigating the risks associated with direct access to private servers.

Regardless of the specific database under consideration, whether it resides on a private or a cloud server, it is possible to implement on both types of platforms the methodologies of maintenance and updates explained in the following.

4.2 Selection of Database Management System

PostgreSQL was selected to serve as the backbone of the database architecture. One of the reasons for choosing PostgreSQL is that it could make use of the relational model, which is particularly adept at managing structured, table-based data. This attribute makes PostgreSQL exceptionally suitable for the needs of the project, as the data being stored are predominantly tabular and relational.

One significant advantage of selecting PostgreSQL for the database management system is its open-source licensing. PostgreSQL is distributed under the PostgreSQL License, a liberal open-source license, similar to the MIT License. This type of licensing offers the freedom to use, modify, and distribute the software in any form without license fees or proprietary restrictions. The absence of licensing fees makes PostgreSQL an economically viable option for projects of all sizes, enhancing its suitability for the database's foundational architecture. Additionally, it is important to note that although PostgreSQL is open-source licensed, it comes with no warranty. This means the software is provided 'as is', without any guarantees of performance, stability, or freedom from defects. Users are responsible for ensuring the software meets their requirements and mitigating any risks associated with its deployment and use.

PostgreSQL employs Structured Query Language (SQL), a powerful and versatile language that includes DDL for specifying database schemas, and DML for querying and updating data, offering a comprehensive tool for database management.

PostgreSQL offers advanced security measures including authentication mechanisms, access controls, and encryption of data both in transit and at rest. PostgreSQL's support for various authentication methods such as password and certificate-based authentication allows for versatile security configurations. Its role-based access control system ensures that access is tightly regulated, granting users permissions specific to their roles. These capabilities make PostgreSQL a secure choice for managing the SYNERGETICS database, aligning with adequate data protection and cybersecurity.

For a source that further elucidates PostgreSQL capabilities, features, and advantages consider referencing Obe & Hsu (2015).

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

4.3 Database Structure and Design

For database management and manipulation, DBeaver, a user-friendly database tool, was utilized. This decision was predicated on DBeaver's compatibility with PostgreSQL and its ability to offer a graphical interface for database interactions, significantly streamlining the process of database design, querying, and maintenance. Its features include:

- A comprehensive graphical user interface (GUI) that simplifies database administration and data manipulation.
- Support for various DBMS, making it a flexible tool for diverse database management needs.
- Advanced data editing capabilities that facilitate the management of large datasets and complex data structures.

The database, named SDB (SYNERGETICS Database), is created on a server hosed by DST, leveraging the relational model provided by PostgreSQL. This model is suitable for managing structured, table-based data, as mentioned before, aligning with the project data requirements, primarily consisting of tabular and relational data. To illustrate the practical implementation of the SDB, Figure 2 presents a snapshot of the database as visualized in the DBeaver interface.

It is intended to create a series of schemas that include the essential elements of the project, such as "Energy Storage Systems," outcomes of Work Packages with names 'WP1', 'WP2', and 'WP3' and more. The figure below also displays the mentioned schemas. Each schema will serve a specific purpose and contain data types tailored to the needs of the stored information. Additionally, each schema will have a short description of the data that is planned to be stored there. With PostgreSQL, each schema within the database encapsulates a variety of database objects each serving a distinct purpose:

1. **Tables:** Serving as the primary storage entities, tables are where all our data is systematically organized in rows and columns.
2. **Foreign Tables:** These special tables act as gateways to data stored in external databases, allowing seamless integration without the need to duplicate information.
3. **Views:** Views are essentially predefined SQL queries that are stored for later use. It enables to streamlining of complex data retrieval processes and presents customized perspectives of the data to different user groups.
4. **Materialized Views:** Going a step beyond standard views, materialized views store the query result as a physical table that can be refreshed. This mechanism is beneficial as it provides quick data access points where real-time information is not paramount.
5. **Indexes:** To expedite data retrieval operations, indexes provide an optimized pathway to quickly locate data within tables.
6. **Functions:** Objects that perform various operations, such as data validation and complex calculations.
7. **Sequences:** Sequences are utilized to generate unique identifiers for database records automatically, ensuring consistency and integrity across datasets.
8. **Data Types:** The PostgreSQL system offers a rich set of data types, allowing us to accurately define the nature of data.
9. **Aggregate Functions:** These functions are employed to perform calculations on a set of values, thereby providing summarized data insights.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

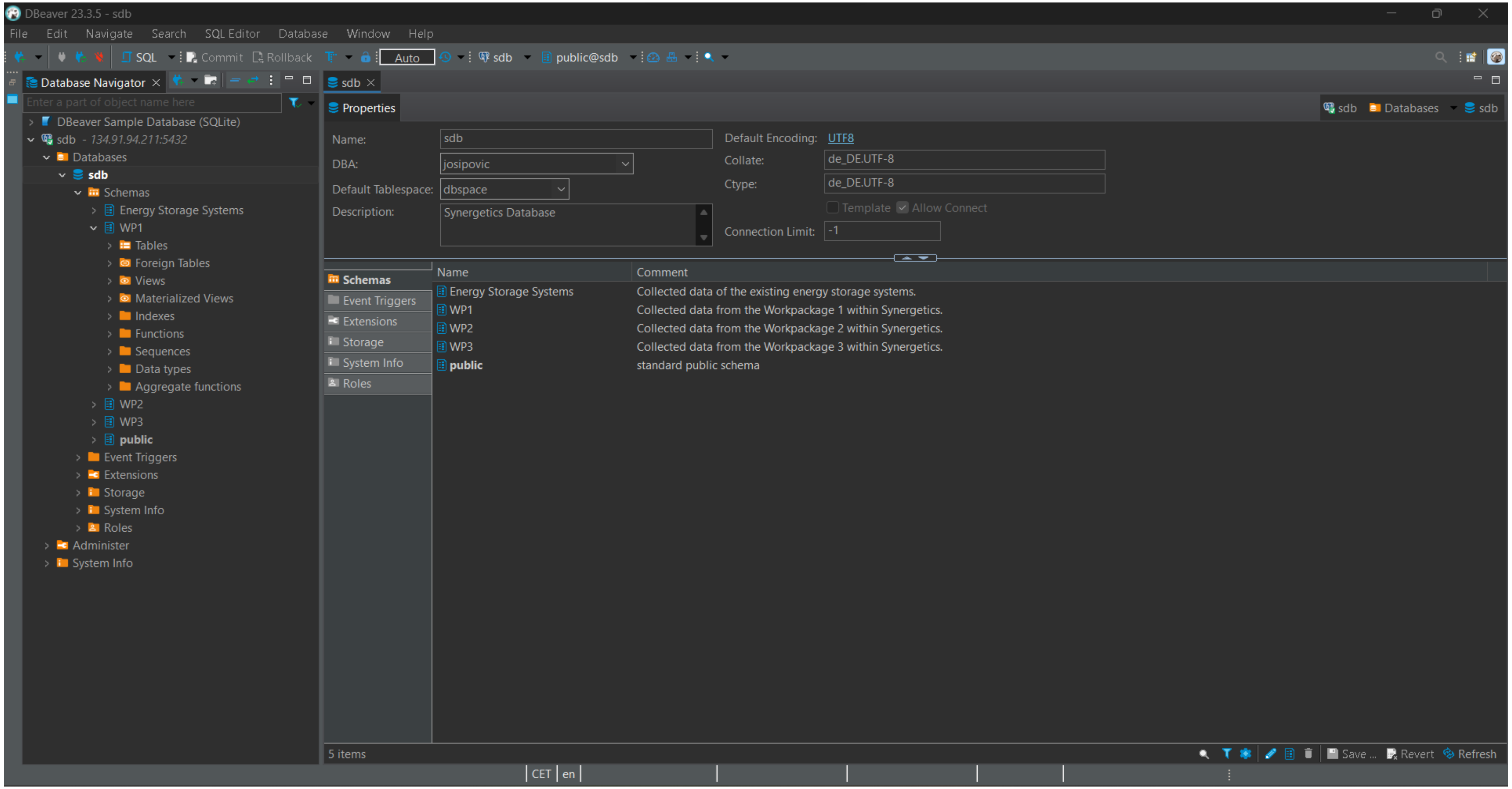

Figure 2: Interface of SDB (SYNERGETICS Database) within DBeaver.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

4.4 Access Control and Authentication

Maintaining the security and integrity of the database is of paramount importance. The access control and authentication framework should be designed to ensure that only authorized users can access and interact with the database. This framework is a crucial component of our database maintenance and updating regimen, safeguarding against unauthorized access and potential data breaches. The suggested implementation is explained in the following:

- **Access Control**
  The approach to access control should be multi-faceted, incorporating both role-based access control (RBAC) and the principle of least privilege (PoLP). Assigning users to roles based on their job functions and granting them the minimum levels of access necessary for their roles minimizes the risk of unauthorized data exposure or manipulation.
- **Authentication Mechanisms**
  The database should employ strong authentication mechanisms to verify the identity of users before granting access to the database. This includes the use of multi-factor authentication (MFA), which requires users to provide two or more verification factors, significantly enhancing security by adding an extra layer of protection beyond just passwords.
- **Continuous Monitoring and Auditing**
  To complement access control and authentication measures, the implementation of continuous monitoring and auditing of access logs and authentication attempts is necessary. This allows us to detect and respond to unauthorized access attempts in real time, as well as identify and remediate any potential security vulnerabilities promptly.
- **User Awareness**
  Users should be aware of the importance of data security, the proper handling of sensitive information, and how to recognize and respond to security threats. This element ensures that all team members, regardless of privilege, are equipped to contribute to the database's security.

4.4.1 Access Rights for Different User Groups

The accession control strategy will be designed to provide precise control over who can view, modify, or manage data within the database, based on their roles and responsibilities within the SYNERGETICS project. Below, is an outline for categorizing user groups and assigning appropriate access rights:

1. **Administrators:** Administrators have the highest level of access, including the ability to create, modify, and delete user accounts, assign roles, and configure database settings. Their access also covers the full range of database operations, such as backup and recovery tasks, as well as performing security audits. This group typically includes IT staff responsible for database maintenance and security.
2. **Data Managers:** Data managers are responsible for the management of data. They have the rights to add, update, and delete data records, as well as to run reports and analyze data trends. Access is restricted to data manipulation functions, excluding administrative tasks such as user management or system configuration changes.
3. **Analysts:** This user group includes project team members whose primary role involves querying the database for information, conducting analyses, and generating reports. Their access is generally read-only, preventing any modification or deletion of data. This ensures that researchers and analysts can perform their roles without risking the integrity of the database.
4. **External Viewers:** For projects involving collaboration with external partners, limited access may be granted to specific sections of the database (scheme/table) relevant to the partnership. This access is tightly controlled and monitored, with permissions set to ensure that external partners can only view or interact with data necessary for their contribution to the project.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

4.5 Maintenance of the Database

To sustain the database's efficiency, security, and alignment with the evolving needs of the SYNERGETICS project, we advocate a proactive maintenance and updating regimen. This section elaborates on a strategy for the continuous management of the sharing server as well as other, ensuring it remains a cornerstone for our project's data management framework.

- **Performance Optimization**
  SDB will benefit from indexing strategies, data partitioning, and query optimization techniques. These measures are aimed at improving query performance, facilitating quick access to large datasets, and maintaining the database's responsiveness.
- **Data Normalization**
  By logically organizing data into distinct tables, each dedicated to a specific type of information, the structure of the SDB will be significantly streamlined and reduce redundancy. This not only enhances the efficiency but also the performance of the database system. The implementation of higher-order data normalization techniques should be used in the future.
- **Data Quality Assurance**
  To uphold the highest standards of data quality within the SDB, robust validation checks, data cleansing routines, and error detection mechanisms will be implemented. These procedures will ensure the database's accuracy and reliability by enforcing automated validation rules during data entry and engaging in regular cleansing to correct any discrepancies. In specific instances where the full rigor of these data quality assurance processes cannot be applied, a distinct schema may be established in the database.
- **Change Management Plan**
  The evolution of SDB will be managed through a detailed change management strategy, encompassing procedures for schema updates, including the addition or removal of tables and columns. A structured documentation process will support each change, involving review, testing, and approval phases to minimize disruptions and uphold data integrity.
- **System Integration**
  SDB will be designed for seamless integration with external systems, utilizing well-defined APIs, data exchange mechanisms, or communication protocols. This integration capability enhances the database's utility and scalability, supporting effective data communication and exchange with other systems.
- **Backup and Recovery**
  To prevent data loss or catastrophic events, the strategy includes scheduled backups and recovery solutions. Regular backups, coupled with efficient storage techniques, and quick recovery procedures, will safeguard data integrity and ensure minimal operational downtime.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

5. Governance and Intellectual Property

5.1 Data Governance

Central to the governance of the SDB are the pivotal partners responsible for its inception and ongoing evolution—all partners within WP4. As the leader in WP4, DST takes on the main government role, ensuring the SDB's alignment with the objectives of WP.

The DST has specific roles and responsibilities for scheduling regular meetings to review progress, set goals, and address challenges, ensuring that all WP4 partners are aligned and engaged according to their tasks. All WP4 partners have responsibilities to oversee the development and implementation of the management database as described in the report to ensure the database's relevance and responsiveness to the project's needs.

It suggests that access levels be allocated with consideration to the specific roles and contributions of the consortium partners within the project. The following distribution is proposed:

1. **Administrators:** It is comprised of members directly involved in Task 4.1, who oversee the technical and strategic management of the database. These will be technical personnel from DST, SPB, MARIN, and OST given their central role in the database's setup and maintenance.
2. **Data Managers:** All partners involved in WP4 could take on the roles of Data Managers, due to their integral involvement and expertise in the relevant tasks. This would ensure comprehensive oversight and up-to-date data curation within the SDB. These partners include DST, SPB, MARIN, OST, SNAOS, KOE, TTS, ANZ, and ANLEG.
3. **Analysts:** Encompassing all partners within the consortium, with access focused on data analysis to facilitate understanding and leveraging the project's collective knowledge.
4. **External Viewers:** To grant access to select external stakeholders, allowing them to view data pertinent to their involvement without compromising the database's integrity.

5.2 Intellectual Property Rights

The management of Intellectual Property Rights (IPR) within the SYNERGETICS Database is guided by a stringent policy framework to protect data integrity and the encompassing innovations. In accordance with the SYNERGETICS Grant Agreement and the Consortium Agreement, the following guidelines have been established:

1. **Ownership of Results (Article 16.2 of the Grant Agreement and Article 8 of the Consortium Agreement):**
   - The consortium owns results produced under SYNERGETICS, aligning with the grant authority's stipulation of non-ownership of these results. Detailed joint ownership arrangements are specified in Annex 5 of the Grant Agreement.
2. **Access and Use Rights (Articles 16.1, 16.3, and 17 of the Grant Agreement and Article 9 of the Consortium Agreement):**
   - Consortium partners are granted access to the SYNERGETICS Database (SDB) for project tasks, with usage rights extending as detailed in the agreements.
   - The grant authority retains rights for non-sensitive information use for communication and dissemination, in compliance with Articles 16.3 and 17.
3. **Third-Party Content (Article 16.3 of the Grant Agreement and Article 8.3 of the Consortium Agreement):**
   - Necessary licenses and authorizations will be obtained for the incorporation of third-party content, ensuring compliance with the agreement's obligations.
   - An informed consent form and a draft non-disclosure agreement will be prepared to address the IPR of knowledge providers.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

4. **Communication and Visibility (Article 17 of the Grant Agreement and Article 8.4 of the Consortium Agreement):**
   - All communication and dissemination activities will appropriately highlight EU support, feature the European flag and funding statement, and adhere to the quality and disclaimer requirements.
5. **Non-Compliance and Consequences (Articles 16.5, 17.5 of the Grant Agreement and Article 8.3 of the Consortium Agreement):**
   - Breaches of IPR articles may lead to a reduction in the grant and other measures, as described in the respective articles of the agreements.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

6. Integration of the Database with a Cloud Server

In the pursuit of advancing the data infrastructure for the SYNERGETICS project and securing a scalable data management solution, this chapter outlines a proposal for integrating the private database with a cloud-based server. The selection of the cloud hosting provider will be determined in subsequent steps within the project, tailored to align with the specific requirements and architecture of the private databases. This transition is planned to guarantee minimal operational disruption and to ensure the security of the private server, from which specific datasets will be migrated to the cloud server.

The brief procedure for database migration can be divided into the following steps:

1. **Initial Preparation**
   The process began with the backup of the database (such as the ESSF-SAPS databases) using PostgreSQL's pg_dump utility. This step ensured a complete snapshot of the database with all essential data, schemas, and structures, which are ready for transfer.
2. **Cloud Server Configuration**
   A new server is provisioned on the cloud, selecting an appropriate configuration and operating system that is well-supported by PostgreSQL. Following the server setup, PostgreSQL was installed and configured to accept connections, laying the groundwork for the new database environment.
3. **Secure Data Transfer**
   Utilizing secure copy protocols, the database snapshot is transferred from the private server to the Cloud server. This operation is conducted with attention to data integrity and security, ensuring that the transfer is both efficient and protected against potential vulnerabilities.
4. **Data Restoration and Verification**
   Upon successful transfer, the database snapshot is imported into the newly established PostgreSQL database on the cloud server. It should be ensured that the cloud-based database mirrors the structure, and functionality of the original system without any discrepancies. The verification process involves a series of checks, including but not limited to, data consistency validation, schema and data type alignment, and integrity constraint analysis.
5. **Data Updates**
   Following the successful verification, it becomes imperative to establish a systematic approach for future data updates and migrations. The frequency of these updates is determined based on the needs of the SYNERGETICS project, the data generation rate, and the criticality of ensuring up-to-date data availability. To streamline this step, an automation strategy will be developed and implemented. This strategy will include the scheduling of regular database snapshots, and their secure transfer to the cloud server, followed by the necessary restoration and verification procedures. Automation not only enhances efficiency but also minimizes the risk of human error and operational downtime, thereby ensuring continuous data availability and integrity.

The long-term hosting and maintenance plan of the SDB and the linked tools, including post-project continuation, will be outlined in the deliverable D6.2 - Update of the PEDR Strategy & Implementation plan.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

7. Conclusion

Based on the overview presented in the preceding chapters of the report, here are the key conclusions that can be drawn:

- **Parameter Selection**: The selection of data elements and parameters for the SYNERGETICS Database is strategically guided by the project's objectives. By prioritizing technological specifications, energy efficiency metrics, environmental impact, and economic considerations, our database will be set to become a key for data-driven decisions in the advancement of green technologies.
- **Database Implementation:** The initial creation of the SYNERGETICS Database (SDB) on the DST server demonstrates a step towards establishing a robust and scalable data management infrastructure.
- **Cloud Integration:** The strategy is to integrate SBD, as well as other databases with cloud services emphasizes the project's commitment to enhancing data sharing, collaboration, and security. This approach mitigates risks associated with direct private server access, leveraging cloud technology for efficient and secure data management.
- **Selection of Database Management System**: The selection of PostgreSQL, due to its open-source nature, compatibility with advanced data types, and strong security features, underscores the project's dedication to utilizing reliable and versatile technologies that align with the project's needs.
- **Commitment to Security and Data Integrity**: The implementation of sophisticated access control, authentication measures, and continuous security monitoring illustrates a comprehensive approach to safeguarding the database against unauthorized access and potential errors, reinforcing the importance of data security within the project.
- **Foundation for Continuous Improvement:** As the SYNERGETICS project advances, the database will transition from its current empty initial phase to a fully populated resource. The report details a maintenance approach that, while preliminary at this juncture, establishes the protocols to be adopted. Adherence to this outlined strategy will ensure the database's capability to adapt, providing a framework for its maintenance and enhancement as the project grows and evolves. The long-term plan of the hosting will be outlined in the deliverable 'D6.2 - Update of the PEDR Strategy & Implementation plan'.
- **Governance and IPR:** The SYNERGETICS Database is governed by a cooperative framework among WP4 partners, with DST leading to ensure alignment with project goals. The database operates under clear IPR guidelines to protect data integrity, ensuring proper use of policies of Grant Agreement.

In essence, the methodology of development and implementation of the SYNERGETICS database as outlined in this report marks a milestone in enhancing the data. As SYNERGETICS moves forward, the infrastructure established here will serve as an essential resource for ongoing and future initiatives aimed at promoting innovations in greening inland and coastal shipping.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation

8. References

Obe, R. O., & Hsu, L. S. (2015). *PostgreSQL: Up and Running* (3 Ausg.). O'Reilly Media.

Sciore, E. (2009). *Database Design and Implementation* (2 Ausg.). John Wiley & Sons.

Silberschatz, A., Korth, H. F., & Sudarshan, S. (2010). *Database System Concepts* (6 Ausg.). McGraw-Hill Education.

Funded by the Horizon Europe Programme of the European Union under grant agreement No 101096809
Funded by the Horizon Europe guarantee of the United Kingdom, under project No 10068310
Funded by the Swiss State Secretariat for Education, Research and Innovation