DiMAT

# D4.1 DIMAT DATA AND ASSESSMENT SUITE

31/10/2024

Digital Modelling and Simulation for Design, Processing and Manufacturing of Advanced Materials

Grant Agreement No.: 101091496
Call: HORIZON-CL4-2022-RESILIENCE-01
Topic: HORIZON-CL4-2022-RESILIENCE-01-25
Type of action: HORIZON Innovation Actions

# D4.1 DIMAT DATA AND ASSESSMENT SUITE

| | | | |
|---|---|---|---|
| Grant agreement number | 101091496 | Acronym | DiMAT |
| Full title | Digital Modelling and Simulation for Design, Processing and Manufacturing of Advanced Materials | | |
| Start date | 01/01/2023 | Duration | 36 months |
| Project url | HTTPS://CORDIS.EUROPA.EU/PROJECT/ID/101091496 | | |
| Work package | WP4 - BUILD: DiMAT Data and Assessment | | |
| Deliverable | D4.1 – DiMAT Data and Assessment Suite | | |
| Task | T4.4 – Continuous Integration and Validation (D&A) | | |
| Due date | 30/06/2024 | | |
| Submission date | 31/10/2024 | | |
| Nature | OTHER | Dissemination level | Public |
| Deliverable lead | 3-Fraunhofer | | |
| Version | 1.1 | | |
| Authors | Lukas Morand, Yoav Nahshon (Fraunhofer) | | |

Co-funded by the European Union

| | |
|---|---|
| Contributions | UPV, CERTH, NTUA, DRAXIS, Imerys, Technical Partners, Toolkit Development Leaders |
| Reviewers | Simone Zuercher (IMERYS), Jorge Domenech (AITEX) |
| Abstract | This document presents the status of the implementation of the toolkits implemented within the DiMAT Data and Assessment Suite. |
| Keywords | Data management, Knowledge acquisition, Life-Cycle Assessment |


Co-funded by the European Union

Document Revision History

| Version | Date | Description of change | List of contributor(s) |
|---|---|---|---|
| 0.1 | 01-May-2024 | ToC | UPV, CERTH |
| 0.2 | 10-Jun-2024 | 1st Draft Available for Internal Review | Fraunhofer, NTUA, DRAXIS |
| 0.3 | 12-Jun-2024 | Internal Review | AITEX, IMERYS, UPV |
| 0.4-0.5 | 27-Jun-2024 | 2nd Draft addressing the comments from internal reviewers | Fraunhofer |
| 1.0 | 28-Jun-2024 | Quality check and issue of final document | CERTH |
| 1.1 | 31-Oct-2024 | Updated document based on PO remarks | CERTH |

## DISCLAIMER



## COPYRIGHT NOTICE

## EXECUTIVE SUMMARY

DiMAT Data and Assessment Suite provides a central set of digital technologies for data storage, data management, and data usage, driven by semantic technologies as well as environmental and cost life cycle assessment on the basis of the gathered data. The suite includes the following toolkits that interact with each other:

- DiMAT Cloud Materials Database – $\mathrm{Di}^{\mathrm{CMDB}}$
- DiMAT Knowledge Acquisition Framework – $\mathrm{Di}^{\mathrm{KAF}}$
- DiMAT Materials Environmental and Cost Life Cycle Assessment – $\mathrm{Di}^{\mathrm{MEC\text{-}LCA}}$

$\mathrm{Di}^{\mathrm{CMDB}}$ is a powerful materials and manufacturing data storage and management tool that uses semantic technologies for meta data annotation to generate a knowledge graph. The $\mathrm{Di}^{\mathrm{KAF}}$ toolkit aims to develop a knowledge graph that represents relevant concepts to describe materials and manufacturing processes. It furthermore hosts the ontology for the $\mathrm{Di}^{\mathrm{CMDB}}$. The $\mathrm{Di}^{\mathrm{MEC\text{-}LCA}}$ toolkit supports the high-level assessment of the environmental and economic impact of processes within the material value chain.

## TABLE OF CONTENTS

## ABBREVIATIONS

| | |
|---|---|
| API | Application Programming Interface |
| CPU | Central Processing Unit |
| CSV | Comma Separated Values |
| D | Deliverable |
| DSMS | Dataspace Management System |
| FAIR | Refers to FAIR data (Findable, Accessible, Interoperable, Reusable) |
| GUI | Graphical User Interface |
| HTTP | Hypertext Transfer Protocol. |
| IP | Internet Protocol |
| JSON | JavaScript Object Notation |
| KAF | Knowledge Acquisition Framework |
| KG | Knowledge Graph |
| KPI | Key Performance Indicator |
| LCA | Life Cycle Assessment |
| LCC | Life Cycle Costing |
| NER | Named Entity Recognition |
| NLP | Natural Language Processing |
| PHP | Hypertext Preprocessor |
| PNG | Portable Network Graphics |
| RAM | Random Access Memory |
| RDF | Resource Description Framework |
| SDK | Software Development Kit |
| SQL | Structured Query Language |
| T | Task |
| UI | User Interface |
| WP | Work Package |

*DiMAT Toolkits*

| | |
|---|---|
| CMDB | Cloud Materials Database |
| KAF | Knowledge Acquisition Framework |
| MEC-LCA | Materials Environmental and Cost Life Cycle Assessment |
| MDF | Materials Design Framework |
| MM | Materials Modeler |
| MD | Materials Designer |
| MMS | Materials Mechanical Properties Simulator |
| MPS | Materials Processing Simulator |
| DTPC | Digital Twin for Process Control |

Co-funded by the European Union

# 1 DIMAT CLOUD MATERIALS DATABASE – DI$^{CMDB}$

This section provides a comprehensive examination of the latest version of the DI$^{CMDB}$, beginning with an overview that offers an understanding of the toolkit scope and objectives. The features delve into the key functionalities and unique aspects that distinguish this toolkit. Following to this, the technical specifications segment details the technical requirements and standards adhered to during development. The implementation status subsection provides a snapshot of the toolkit current stage, including a breakdown of the current implementation and a look ahead at the next developments, outlining planned advancements and improvements.

## 1.1 OVERVIEW

The Cloud Materials Database (Di$^{CMDB}$) is a system for storing, sharing, and exploration of material data for materials design, processing, and manufacturing processes. The system follows the FAIR-principles and supports traceability of data along different kind of operations, data sovereignty, and data safety. The system is based on semantic technologies to handle the manifold of data sourced and to provide possibilities for data exploration and generation of new knowledge. All linked and visualized in a knowledge graph.

## 1.2 FEATURES

Features of Di$^{CMDB}$ are the following:

- Semantic data integration: On the basis of a given ontology, data can be integrated in a semantic manner into the Di$^{CMDB}$ (i.e. using data2rdf tool [1]) and annotated and linked to each other. Meta data is stored as RDF. Time series data, which is for example typically gathered during experiments (including units) is stored as HDF5 container. Attachments in the form of complete files can also be stored. Data upload, annotation and linkage can be done via a web-frontend but also using the Python package DSMS-SDK [2].
- Furthermore, already existing data from other toolkits, like KAF can be accessed. Therefore, Di$^{CMDB}$ can be understood as a more general data providing platform.
- Data download: Data contained in the database can be accessed and downloaded in the desired unit via the web-frontend as well as the Python package DSMS-SDK.
- Data management (create, read, update, delete): Datasets, but also organizations, experts, testing machines, simulation models etc., can be stored in the CMDB as so-called

Knowledge-Items. The user creating the item has permissions to read, update and delete the item permanently.

- Data search/exploration: The web-frontend and also the Python package DSMS-SDK provides a SPARQL console with which data can be searched and retrieved from the database. In order to make the tool usable to domain experts, an additional search function is given, that allows for (i) searching for annotations given to data and (ii) a semantic search based on a large language model looking for similarities between a given word and descriptions of data.

## 1.3 TECHNICAL SPECIFICATIONS

- Provides a friendly graphical user interface (GUI) based on Angular [3].
- Employes a micro-structured architecture founded on the client-server model, integrating REST principles to facilitate efficient communication and enhance scalability.
- Runs in Docker containers, ensuring compatibility across different computing environments.
- Requires an internet connection for updates and authentication via Keycloak.
- Minimum hardware Requirements: 2-core CPU, 16GB RAM.

## 1.4 IMPLEMENTATION STATUS

In the following sections, the current implementation status is detailed and also the next developments to be released for the DiMAT Cloud Materials Database.

### 1.4.1 Current Implementation

The current implementation of Di$^{CMDB}$ includes the following:

- Semantic data integration.
- Data upload.
- Data management.
- Data search/ exploration.

Link to the GITLAB repository of the solution:


Co-funded by the European Union

https://gitlab.cc-asp.fraunhofer.de/dimat/data-and-assessment-suite/cloud-materials-database

We note that this repository only contains modules developed in the DiMAT project. The main code of Di$^{CMDB}$ is based on background IP (see Dataspace Management System [4]) and is located elsewhere.

### 1.4.2 Next developments

The future development and implementation of Di$^{CMDB}$ will include the following:

- Advanced settings for access rights, data sovereignty.
- Fill Di$^{CMDB}$ with data from DiMAT pilots.
- Connect Di$^{CMDB}$ with other toolkits.


Co-funded by the European Union

# 2 DIMAT KNOWLEDGE ACQUISITION FRAMEWORK – DI^KAF

This section provides a comprehensive examination of the latest version of the DI^KAF, beginning with an overview that offers an understanding of the toolkit scope and objectives. The features delve into the key functionalities and unique aspects that distinguish this toolkit. Following this, the technical specifications segment details the technical requirements and standards adhered to during development. The implementation status subsection provides a snapshot of the toolkit current stage, including a breakdown of the current implementation and a look ahead at the next developments, outlining planned advancements and improvements.

## 2.1 OVERVIEW

The Knowledge Acquisition Framework (Di^KAF) provides access to advanced analytics, information retrieval and visualizations related to manufacturing materials. The core component of the toolkit is the developed Knowledge Graph (KG). The entities (nodes) of the KG represent graphically materials and relevant concepts for manufacturing such as their properties, acceptable material states, manufacturing devices, business processes and their structure while relationships (edges) join relevant concepts. The schema of this KG draws inspiration from popular ontologies such as EMMO [5] and the Industrial Data Ontology [6].

## 2.2 FEATURES

The features of the KAF toolkit are the following:

- Data Exploration: Users are able to write their queries and retrieve relevant information through the front-end. For the users experienced with Neo4J Cypher there is the possibility to write freely their own queries. For less experienced users, pre-selected filters which can be used to retrieve the most basic information are provided.
- Data Visualization: Results of the queries posed to the KG are visualized, offering to the user an overview of a subset of the KG, revealing how the information is structured.
- Data download: The users can select to download the results of their queries in a CSV file, allowing thus further analysis and inspection. Data can also be retrieved from an HTTP API.

Co-funded by the European Union

- Data Upload/modification: A user with administrative capabilities can modify the data in the KG or include new datasets. To include new datasets, a template should be filled that ensures that the graph's schema is respected.
- Data analytics: Capitalizing on the structured graph format a number of algorithms can be applied in order to extract useful information, revealing correlations and similarities between materials.

## 2.3 TECHNICAL SPECIFICATIONS

Technologies employed:

- The Knowledge Graph of the KAF toolkit is developed with Neo4j.
- The front-end is developed with Vue.Js.
- The services that run in the back-end are developed with Python (Flask API) and Javascript (Node.Js).
- The secure login mechanism is implemented with Keycloak.
- The graph embedding algorithms are provided by Ampligraph.
- Graph visualization is performed with NeoVis.
- Data visualization is also performed with Tensorboard.
- The solution is provided as a docker image.

Recommended computing resources 8GB RAM and 4-core CPU.

## 2.4 IMPLEMENTATION STATUS

In the following sections, the current implementation status is detailed and the next developments to be released for the DiMAT Knowledge Acquisition Framework.

### 2.4.1 Current Implementation

The current implementation of Di$^{KAF}$ includes the following:

- Data exploration through queries and filters.
- Data uploading by uploading a filled-in template CSV file provided by the toolkit for downloading.

- Data modification through cypher queries.
- Query results downloading as CSV files.
- Query results visualization with NeoVis.
- Explanation of the KG structure through a dedicated interactive webpage.
- Similarity search for Material nodes.
- Visualizations of information (currently only Material nodes) with Tensorboard.
- Modifications of the KG embedding model used through the front-end.

Link to the GITLAB repository of the solution:

https://gitlab.cc-asp.fraunhofer.de/dimat/data-and-assessment-suite/knowledge-acquisition-framework/kaf/-/tree/kaf-v1?ref_type=heads

### 2.4.2 Next developments

The future development and implementation of Di$^{KAF}$ will include the following:

- Recommendations about selection of materials and business processes.
- Interactions with the KG using free text also besides filters and cypher queries. The text should be handled with NLP techniques (NER, Intent classification etc.)
- Connection of KAF with other toolkits
- Population with more data from Pilot partners


Co-funded by
the European Union

# 3 DIMAT MATERIALS ENVIRONMENTAL AND COST LIFE CYCLE ASSESSMENT – DIMEC-LCA

This section provides an update and documentation of the latest version of the MEC-LCA toolkit, beginning with an overview of the toolkit scope and objectives. It explains the key functionalities and unique aspects that distinguish this toolkit, along with technical specifications. The implementation status subsection provides a snapshot of the toolkit current stage, including a breakdown of the current implementation and a look ahead at the next developments for the following versions.

## 3.1 OVERVIEW

The Materials Environmental and Cost Life Cycle Assessment (Di$^{MEC-LCA}$) provides a high-level assessment of the environmental and economic impact of processes within the material value chain through the graphical representation of results. The users, such as product engineers or decision-makers, can review indicators of the environmental and cost LCAs (WP7), as well as information on the pilots' KPIs (WP2). Information on core life cycle assessment issues, such as identification of environmental and financial hotspots, evaluation of design options performance, and benchmarking will be visualized through the platform's UI.

## 3.2 FEATURES

Current features of Di$^{MEC-LCA}$ are the following:

- Data visualization and dashboards: Di$^{MEC-LCA}$ provides a user-friendly interface for creating, customizing, and managing dashboards. Through integration with Grafana, Di$^{MEC-LCA}$ allows users to interact with a variety of dashboards and plots.
- Data manipulation: Where applicable, users can modify existing data stored in the MySQL database, directly from the dashboard. Any changes made to the data are reflected in real-time within the dashboard.
- Data filtering: Users can dynamically adjust filter criteria, through the Variable plugin, and instantly see the impact on the displayed data.
- Data download: Di$^{MEC-LCA}$ facilitates data export in multiple formats, including PNG for graphical representations and raw data in CSV and JSON formats.

- Data calculations: Di$^{MEC-LCA}$ supports mathematical operations, documenting and parsing the various equations which generate the KPI and LCA/LCC results. Furthermore, for users with more complex analytical needs, additional math calculations can be conducted externally on the data through scripts before data is stored in the database or uploaded through the Infinity plugin.
- Panel Toolbar: Each panel in the dashboard has a toolbar, which contains options for editing the panel, such as changing the visualization type, modifying the query, and configuring display options.
- Time Range Selector: At the top of the dashboard view, there is a time range selector that allows user to specify the time range for the data displayed on the dashboard.
- Dashboard Toolbar: The dashboard toolbar appears at the top of the dashboard view and contains options for interacting with the entire dashboard, including saving the dashboard, adding annotations, sharing the dashboard, and accessing dashboard settings.
- Access management: Administrators have detailed control over user access, dictating which features and functionalities each user can access within the platform.

## 3.3 TECHNICAL SPECIFICATIONS

There are three primary Docker containers Laravel, MySQL, and Grafana, hosted on a dedicated server.

**Laravel**, a PHP framework, is utilized solely by the administrator and enables the creation of seeders to populate the MySQL database with data. By leveraging Laravel's seeders, the admin can easily insert data into the MySQL database, setting up the foundation for subsequent data querying and analysis.

**MySQL** database acts as a data storage solution within the stack. It provides a scalable relational database management system for storing structured data efficiently. The MySQL container ensures reliable data persistence and retrieval, facilitating seamless storage and retrieval of data.

Integrated with MySQL as a data source, **Grafana** serves as a visualization and dashboarding tool. Within Grafana:

- The **Infinity plugin** facilitates direct upload of CSV files into Grafana. Administrators use this plugin to create dynamic graphs and visualize data without having to query a database.


Co-funded by the European Union

- The **Variable plugin**, enabling users to select a variable for automatic calculations. This feature enhances user interaction and flexibility by allowing users to customize the displayed data directly through the interface.Users can choose the variable through dropdown lists, buttons or radio buttons.
- Data manipulation is possible through the **Data Manipulation plugin**. Users can update existing data in the MySQL database directly from dashboards, in an intuitive and straightforward manner.

**Keycloak** is used for identity and access management. This connection provides essential user authentication and authorization services, ensuring secure access to the platform's resources. Notably, Keycloak operates independently and is not part of the Docker stack.

## 3.4 IMPLEMENTATION STATUS

The following sections detail the current implementation status and the next developments to be released for the DiMAT Materials Environmental and Cost Life Cycle Assessment toolkit.

### 3.4.1 Current Implementation

The current implementation of Di$^{MEC-LCA}$ includes the following:

- User interface.
- Dashboarding and data visualization.
- Data manipulation.
- Data download.
- Data calculations.
- Keycloak integration.
- KPI data integration (WP2).
- Dummy LCA/LCC data integration.
- Docker image.

Link to the GITLAB repository of the solution:

https://gitlab.cc-asp.fraunhofer.de/dimat/data-and-assessment-suite/materials-environmental-and-cost-life-cycle-assessment


Co-funded by the European Union

### 3.4.2 Next developments

The future development and implementation of DiMEC-LCA will include the following:

- Integration of real LCA/LCC data (through the elaboration of Task 7.5).
- Detailed user testing.
- Update of user access roles and groups as needed once real LCA/LCC data is integrated.

Co-funded by the European Union

## REFERENCES

[1] https://github.com/MI-FraunhoferIWM/data2rdf

[2] https://github.com/MI-FraunhoferIWM/dsms-python-sdk

[3] https://v17.angular.io/guide/what-is-angular

[4] Yoav Nahshon et al. Semantic orchestration and exploitation of material data: A dataspace solution demonstrated on steel and cooper applications, submitted to Advanced Engineering Materials, 2024

[5] https://emmo-repo.github.io/

[6] https://www.iso.org/standard/87560.html


Co-funded by
the European Union