MINERAL INTELLIGENCE
FOR EUROPE

Authors and affiliation:
Jasna Šinigoj, Andrej Vihtelič, Špela Kumelj (GeoZS)

E-mail of lead author:
spela.kumelj@geo-zs.si

Deliverable D3.2

TECHNICAL GUIDELINES – status after 1st year

Version: 27-11-2019

This report is part of a project that has received funding by the European Union's Horizon 2020 research and innovation programme under grant agreement number 731166.

Deliverable Data		
Deliverable number	D3.2	
Dissemination level	Public	
Deliverable name	Technical Guidelines	
Work package	WP3, Task 3.2	
Lead WP/Deliverable beneficiary	GeoZS	
Deliverable status		
Submitted (Author(s))	11/11/2019	Jasna Šinigoj, Andrej Vihtelič, Špela Kumelj
Verified (WP leader)	27/11/2019	Špela Kumelj
Approved (Coordinator)	29/11/2019	Jørgen Tulstrup

GENERAL INTRODUCTION

About Mintell4EU:

The European Union has identified security of supply, improvement in environmental management and resource efficiency as key challenges for the raw materials sector. Data regarding the location and spatial distribution of primary and secondary raw materials with respect to exploration, exploitation, production and trade activities, underpin decision making in government and industry. Given the dynamic character of such data regular updates of comprehensive, reliable and harmonized information across borders are required. The overall aim of this project is to improve the European Knowledge Base on raw materials by updating the electronic Minerals Yearbook (e-MYB) produced in the Minerals4EU project and to extend the spatial coverage and quality of data currently in the Minerals Inventory. The project will, furthermore, aim to increase the degree of harmonization, communication and interaction between existing data platforms, with the ambition of reaching a fully operational and reliable data knowledge management system, fulfilling the European needs and taking into account the Raw Materials Information System (RMIS) of the European Union. Importantly, the project will also integrate the electronic Minerals Yearbook into the Minerals4EU database, ensuring future sustainability as part of the EuroGeoSurveys-governed European Geological Data Infrastructure (EGDI, www.europe-geology.eu). All results will be integrated in the EGDI that will, by end of the project, disseminate European raw materials intelligence in a uniform way to end users through a common web portal interface. Finally, the applicability of the UNFC classification system for obtaining more accurate Pan-European mineral inventories will be tested.

Mintell4EU liaises and builds on results from the work of ORAMA project, in particular in case studies on UNFC, usage of technical guidelines in data insert into M4EU DB and compliance with INSPIRE.

EXECUTIVE REPORT SUMMARY

This report is a summary of existing technical guidelines or documentation, that is necessary for Minerals Inventory and e-MYB improvements to be carried out within Mintell4EU.
The report shows the status of technical documentation after the 1st year and will be updated at the end of the project to be in line with all the modifications yet to come.

TABLE OF CONTENTS

Table of Figures and Tables

1 INTRODUCTION

The overall aim of the Mintell4EU project is to improve the European Union Raw Materials Knowledge Base (EURMKB) by extending the spatial coverage and quality of data from past and ongoing European projects on raw materials (Minerals Inventory or MI) and by updating the electronic Minerals Yearbook (e-MYB) produced in the Minerals4EU (M4EU) project. The e-MYB will be developed in a separate schema in the M4EU database through the Mintell4EU project.

The e-MYB and MI in particular are associated with harvesting assurance systems and are highly dependent on already existing technical documentation.

To avoid duplication and/or repetition of deliverables between the ORAMA and Mintell4EU projects, the following chapters present only the summary of the most important technical information from ORAMA deliverables, which are crucial for successful implementation of MI and e-MYB improvements in the Mintell4EU. At the same time the report provides the links to the ORAMA Technical guidelines and webinars.

Finally, an overview of the currently running technical routines for the M4EU harvesting system and the most commonly user terminology are added to the report.
The main target group for this report are national data providers, who have to keep in mind that their data should be easily accessible and understandable as well as reliable and of a proven quality. They need to set up a practice for developing the national level data service with sufficient and appropriate datasets. To be able to achieve this, they need to understand the overall technical process behind it and this report tries to explain this process in a simple manner. Accordingly, it should be used by mineral resources researchers as well as IT personnel as their support to provide adequate web services.

2 DATA FLOW

In the ORAMA project a new M4EU/ORAMA/e-MYB data model has been jointly developed by the ORAMA project partners GEUS, BGS, BRGM, and GeoZS. The data model was created to serve several purposes:

- to harvest National Providers databases through a Web service,
- to set up Excel portrayals and an electronic input form for countries/institutions that do not have working Provider databases and Web services setup, and
- to build the Harvesting database.

2.1 Data flow of Minerals Inventory data

The European Minerals Knowledge Data Platform (EU-MKDP) adopted a professional architecture with a harvesting system and a diffusion system (*Figure 1*). The system is divided into three parts (Cassard et al., 2014).:

- The national level from which INSPIRE v.3 compliant WFS' (Web Feature Service) deliver data to the harvesting system either from (already) EuroGeoSource (www.eurogeosource.eu) structured databases or directly from other data sources in each country;
- The central harvesting system which regularly reads data from the national WFS' and stores it in the Harvesting Database. Data is subsequently being delivered to the central diffusion system through a database synchronization mechanism. The Harvesting Database is optimized towards reading data from the national level and delivering this data to the diffusion system;
- The central diffusion system which is updated with data from the harvesting system at regular intervals and makes this data available to users through the EU-MKDP web portal and the EGDI. The Diffusion Database has specific optimization to offer a better experience to the user for the delivery of the data and the computation of on-demand services.

The Information Factory is a software component able to process data (indexation, filtering, descriptive/exploratory statistics computation, automatic report generation…).

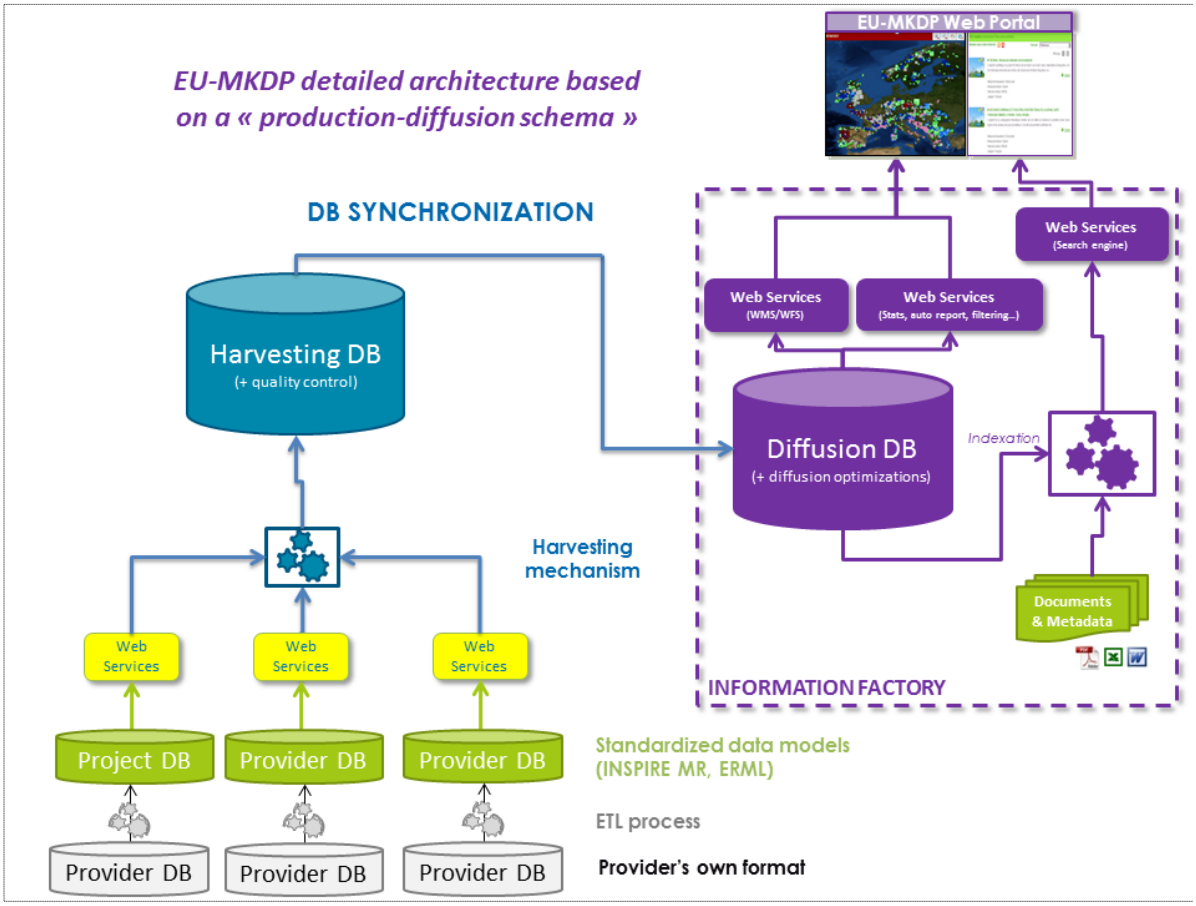


Figure 1: The EU-MKDP detailed architecture showing the Harvesting Database, the synchronization process with the Diffusion Database, the Information factory and the indexation process of both structured and semi- and non-structured data. The same information which is accessible on the EU-MKDP portal is also accessible on the EGDI portal.

2.2 Data flow of aggregated mineral reserve, resource, exploration and trade data

The e-MYB data flow includes country-wide aggregated mineral resources, reserves, production and trade data (described in Chapter 1 of the ORAMA D3.2 Serving aggregated data). It is also possible to incorporate the data available through the BGS website download tools for production and trade (import and export data).

The BGS will receive a dump of Harvesting DB and will (i) create a copy of the GeoZS' Harvesting DB to store exploration, reserves and resources data, and (ii) extract production data from Harvesting DB, which will then be validated and incorporated in the BGS World Mineral Statistics DB together with production and trade data from other sources. This validated dataset will be provided by BGS to BRGM for the update of the previous version of

the interactive e-MYB (Minerals4EU Diffusion Platform, 2015 - http://minerals4eu.brgm-rec.fr/m4eu-yearbook/theme_selection.html) via Web service based on the M4EU/ORAMA/e-MYB data model. *Figure 2* gives a schematic overview of the data flow and which procedures will take place before the data will become visible on the new e-MYB portal.

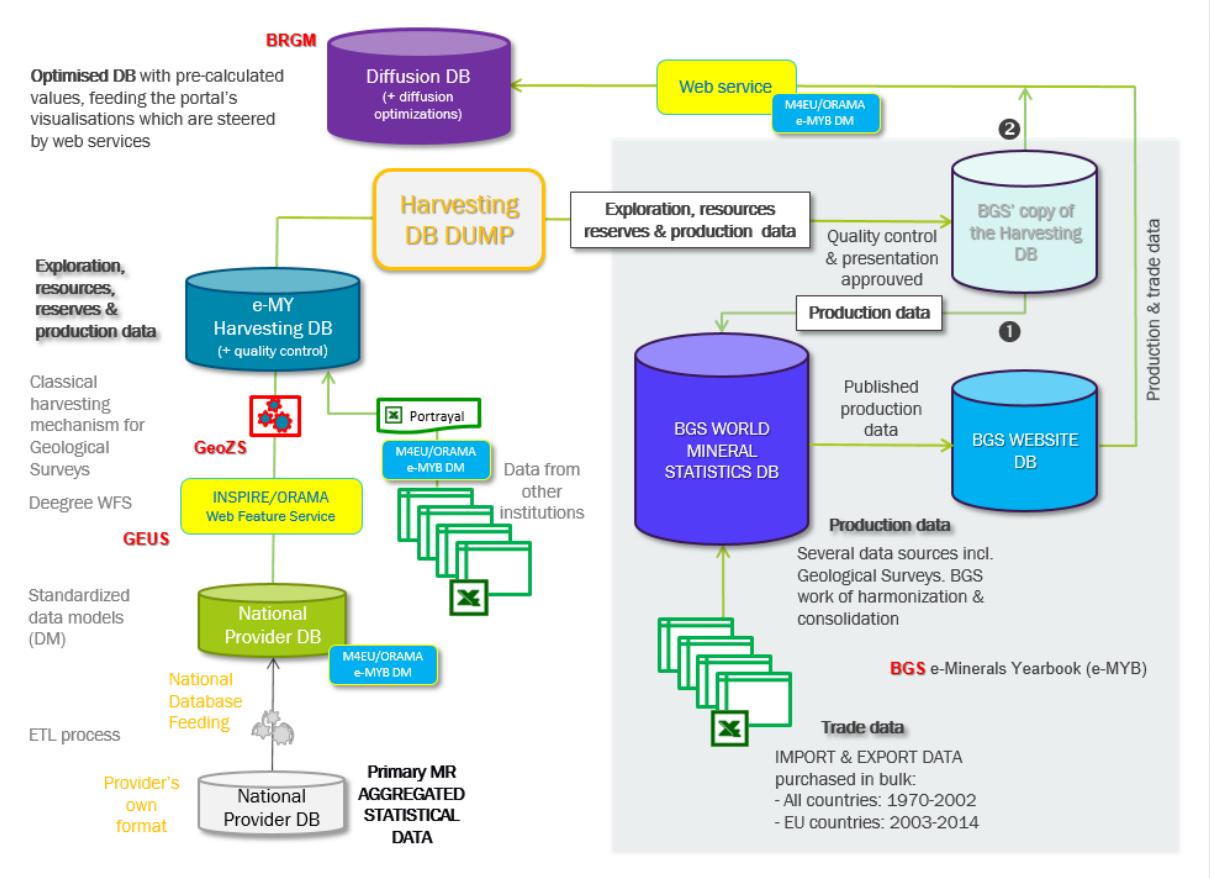


Figure 2: The e-MYB feeding data flow (© ORAMA project, D3.2 Serving aggregated data).

2.3 The Harvesting system

Data harvesting is a process which reads and automatically extracts large amounts of data from the Web (in form of Web pages or Web services). The Mintell4EU/e-MYB harvesting systems were programmed and set up by GeoZS, partly in the ORAMA project, partly in Mintell4EU. The harvesting systems periodically collect data from data providers and refreshes the harvesting Mintell4EU/e-MYB database using an INSPIRE compliant WFS.

During the harvesting phase the received data is checked if code lists conform to the INSPIRE registry code list values and if other data have the correct format (e.g., dates, numbers…). The Mintell4EU/e-MYB harvesting DB is hosted by GeoZS and connected to the Diffusion DB (hosted by BRGM) using SQL scripts (*Figure 3*).

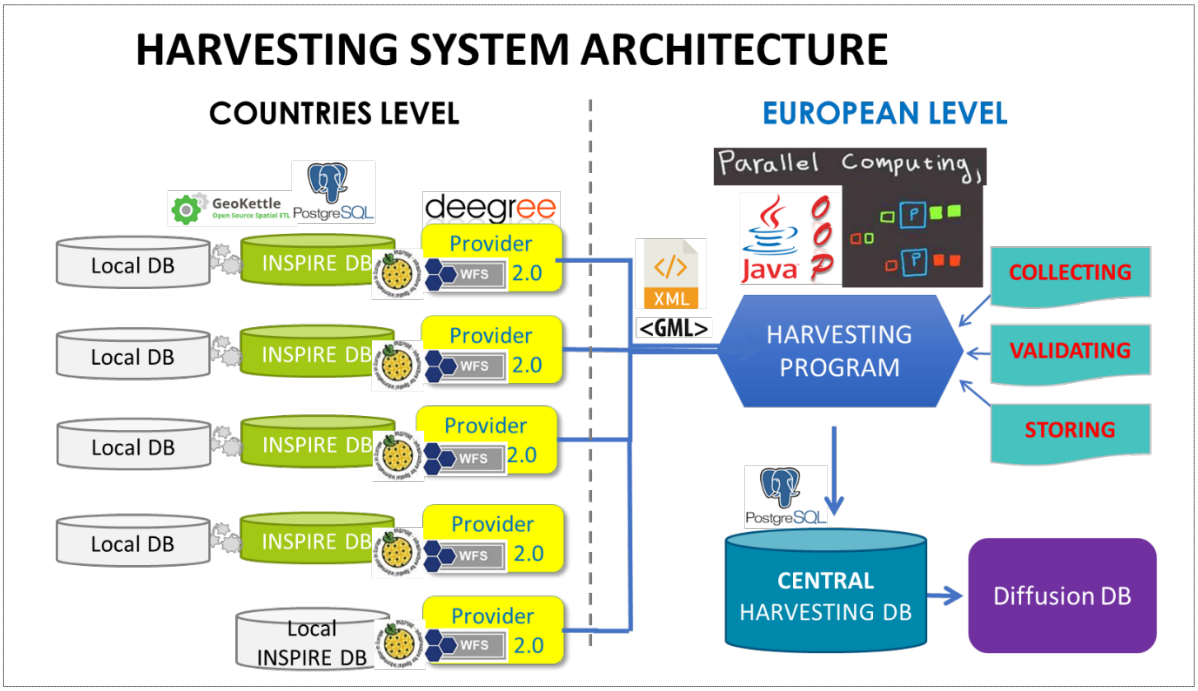


Figure 3: The harvesting system structure.

Some key outlines from the Mintell4EU Data Management Plan need to be highlighted:

1. Partners are encouraged to provide the latest version of the common database model and the latest data. The harvesting database model is always the latest version. The data from each partner for The Minerals Inventory is harvested each month. Data may change with each harvesting. Each record may have the attributes of origin (harvesting url) and date of latest harvesting (pr. country/survey) and database model version.
   The ORAMA project will produce a guideline for this and Mintell4EU will follow it.
   The production data in e-MYB will be recopied from the BGS web database each year in order to capture the eventual revisions to previous years as well as adding each additional year. The other data types will be added each year with no updates to prior years being expected. The mechanism for this data transfer is being developed under the ORAMA project and Mintell4EU will use this mechanism.
2. The quality of the data in the M4EU database depends on each partner quality assurance procedures. The Mintell4EU project is not aware of the procedures at each Geological Survey, but recommendations for QC procedures will be sent to the data providers in coordination with the ORAMA project. For the e-MYB the data are subject to stringent quality control procedures used by BGS for many decades. These procedures are not currently published.
3. The Minerals Inventory data is interoperable because the data from regional and national survey organizations are stored in a common database. The interoperability lies in the fact, that each survey is mapping its data into a scheme modelled in accordance with EarthResourcesML (see Appendix A of this report). The geospatial data is accessible in a standard format via OGC web services. The e-MYB data follows the ERML-Lite (see Appendix A of this report).

Overall data flow process provides compatibility of improved datasets with the INSPIRE Directive and existing data models and at the end shows the connection to the EU-MKDP portal, bridges the GeoERA Raw Materials projects via EGDI and provides the connection to the RMIS (*Figure 4*). Finally, Mintell4EU will also provide a detailed description of how you enter the system as a new data provider, how to harmonize your data with respect to INSPIRE and follow the rules of accessibility.

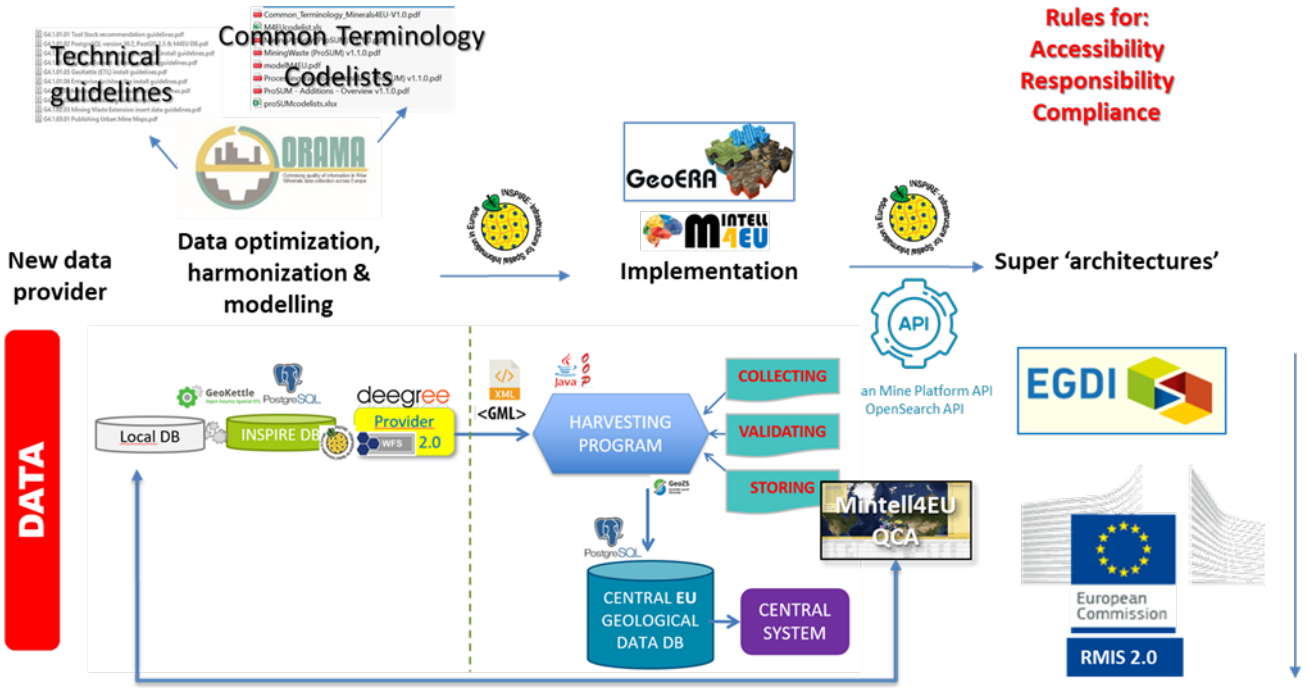


Figure 4: The compatibility of overall data flow process.

2.3.1 Codelists

The harvesting process is being done using the INSPIRE MR code list of commodities. As the e-MYB uses different code lists, mapping has been performed in order to transfer harvested data to BGS as recommended by ORAMA (Bide et al., 2018b). The mapping done by BGS for Minerals4EU has been reused and modified.

Data providers must ensure that identifiers (INSPIRE ID's: inspireid, inspirens) always stay the same for one given record (even if they delete all the data and replace it). **An identifier should always correspond to the same occurrence and should not be changed.** When a modification is made in a national database, the provider should normally change just the versionID and the begin/end lifeSpan.

For e-MYB a unique and stable triple key which identifies country, commodity and year is also needed. The fact is, not all organizations are able to use WFS for providing aggregated data at the national level. That is why during the Mintell4EU project a dedicated Web data entry form will be available where each data provider will be able to enter data for their country.

The final versions of codelists will be validated by ORAMA project and when final available to download from their website.

2.4 National providers of Web services

Each national provider should execute the necessary scripts for generating a database management system and working WFS setup (*Figure 5*). These scripts allow quick implementation of a standardized database model (DM) based on a standardized conceptual standard (UML). Once database and services are implemented and each countries data are stored by the data providers, the data can be harvested by a central harvesting database (which implements the same database model).

National providers should use ORAMA guidelines for common installation of M4EU/Mintell4EU/e-MYB database and WFS:

- G4.1.01.01 Tool Stack recommendation guidelines.pdf
- G4.1.01.02 PostgreSQL version 10.7, PostGIS 2.5 & M4EU DB.pdf
- G4.1.01.03 Java SE Development Kit openJDK install guidelines.pdf
- G4.1.01.04 Apache Tomcat & Deegree install guidelines.pdf
- G4.1.01.05 GeoKettle (ETL) install guidelines.pdf
- G4.1.01.06 Enterprise Architect lite install guidelines.pdf

All the guidelines can be found at the ORAMA project website https://orama-h2020.eu/#tab-id-4, under the ORAMA Deliverables, WP4 Sharing data, D4.1 Technical guidance for the data harmonization of raw materials.

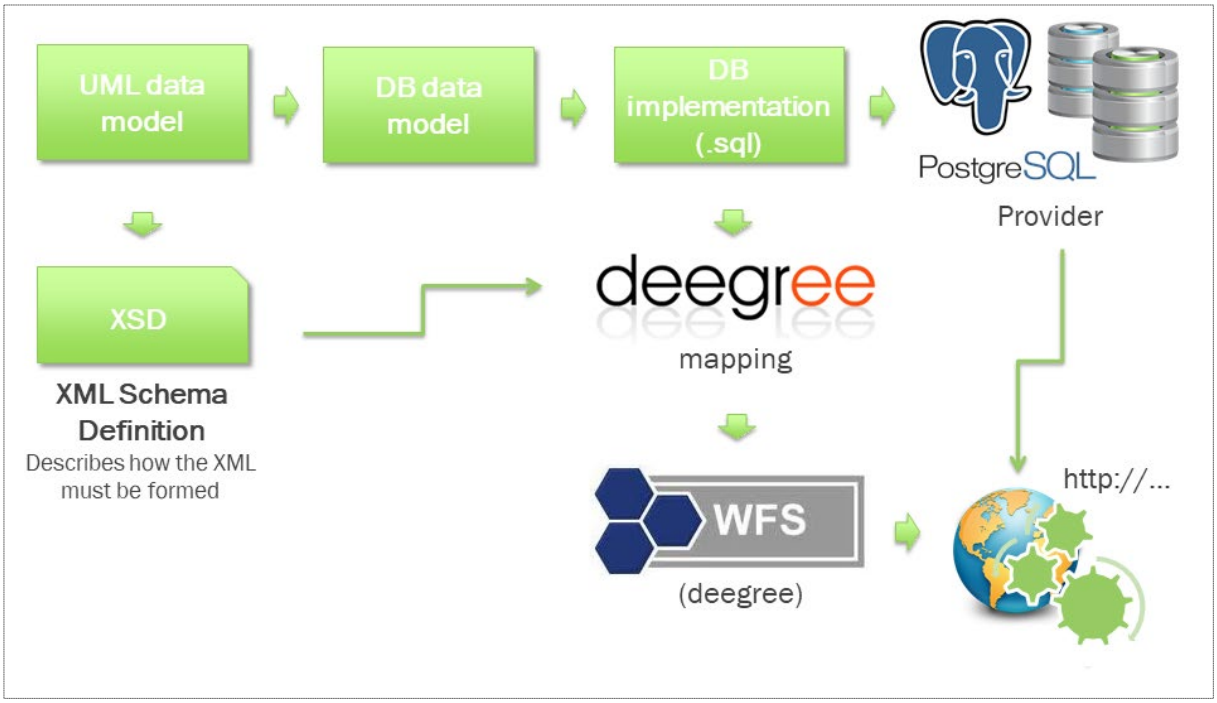


Figure 5: The stages from the UML data model to the setup of a Web feature service (WFS)

All required software is open source and to ensure common installation to manage the M4EU/ORAMA/e-MYB database it is recommended to follow suggested versions as described in the *Table 1*(Schjøth et al., 2019).

Table 1: e-MYB/Mintell4EU recommended Open Source Tool Stack

Software logo	Name and link	Recomm. Ver.
	PostgreSQL PostgreSQL (also referred to as Postgres) is an open-source relational database management system (RDBMS) https://www.postgresql.org/ https://geusgitlab.geus.dk/m4eu/2019-ORAMA-v2.xx/tree/master/03-postgresql	10.7 Or later
	PostGIS (bundle within PostgreSQL) PostGIS provides spatial objects for the PostgreSQL database, allowing storage and query of information about location and mapping. http://postgis.net/ https://geusgitlab.geus.dk/m4eu/2019-ORAMA-v2.xx/tree/master/04-postgis	2.5 Or later
	Java Development Kit 8 SE The JDK is a development environment for building applications, applets, and components using Java programming language. https://adoptopenjdk.net/ https://geusgitlab.geus.dk/m4eu/2019-ORAMA-v2.xx/tree/master/02-java-se-development-kit	Jdk8u202-b08 Or later
	Apache-Tomcat Apache Tomcat implements several Java EE specifications including Java Servlet, JavaServer Pages, Java EL, and WebSocket, and provides a "pure Java" HTTP web server environment in which Java code can run. https://tomcat.apache.org/download-80.cgi https://geusgitlab.geus.dk/m4eu/2019-ORAMA-v2.xx/tree/master/05-tomcat	8.5.38 Or later
	Deegree3 deegree is an open source software for spatial data infrastructures and the geospatial web. It offers components for geospatial data management, including data access, visualization, discovery and security. https://www.deegree.org/download https://geusgitlab.geus.dk/m4eu/2019-ORAMA-v2.xx/tree/master/06-deegree3/deegree-webservices-3.4.3	3.4.3 Or later
	GeoKettle (ETL) GeoKettle is a powerful, metadata-driven spatial ETL (Extract, Transform and Load) tool dedicated to the integration of different data sources for building and updating geospatial databases, data warehouses and web services. https://sourceforge.net/projects/geokettle/files/geokettle-2.x/2.5/ https://geusgitlab.geus.dk/m4eu/2019-ORAMA-v2.xx/tree/master/07-geokettle	2.6-r192

Software logo	Name and link	Recomm. Ver.
	Enterprise Architect viewer The EA Viewer is a free, read-only version of EA intended for distribution with UML models. The read-only version supports all viewing functions, however documentation generation and all 'update' ability is disabled. https://www.sparxsystems.com/products/ea/downloads.html https://geusgitlab.geus.dk/m4eu/2019-ORAMA-v2.xx/tree/master/08%20Enterprise%20Architect%20model/EA-lite-Reader-software	14.1.1427 Or later

2.5 Technical Guildelines for inserting and harmonization of data in the M4EU database

To optimize the quality of information, national providers should also use ORAMA Technical guidelines for harmonization of data, described in ORAMA project Deliverable 4.1: “Technical guidance for the data harmonization of raw materials” and available from ORAMA project website https://orama-h2020.eu/#tab-id-4 , under the ORAMA Deliverables, WP4 Sharing data, D4.1 Technical guidance for the data harmonization of raw materials.

The following instructions available to insert data into M4EU DB are:

- G4.1.02.01 Mineral Occurrence insert data guidelines.pdf
- G4.1.02.02 Mine insert data guidelines.pdf
- G4.1.02.03 Mining Waste Extension insert data guidelines.pdf

The additional technical guidelines regarding data harmonization for raw material are available for SRM:

- Technical Guideline Tools for harmonization of data collection on ELV.pdf
- Technical Guideline Tools for harmonization of data collection on Mining Waste.pdf
- Technical Guideline Tools for harmonization of data collection on WEEE/PV Panels.pdf
- Technical Guideline Tools for harmonization of data collection on Batteries.pdf
- ORAMA_D1.4_Guidance To Harmonisation of Resource and Reserve data.pdf

The ORAMA project also organized webinars on primary and secondary raw material, where all the above-mentioned contents were presented. On the ORAMA website https://orama-h2020.eu/ all webinars recordings and presentations are available for download.

Under the Mining waste webinar, performed on 30 September 2019, there are training materials on ProSUM data (MiningActivity (ProSUM) v1.1.0.pdf, MiningWaste (ProSUM) v1.1.0.pdf, ProcessingTransformationPlant (ProSUM) v1.1.0.pdf and ProSUM - Additions - Overview v1.1.0.pdf), but maybe even more importantly link to:

- Common_Terminology_Minerals4EU-V1.0.pdf
- modelM4EU.pdf
- M4EUcodelist.xls
- proSUMcodelists.xlsx

3 TECHNICAL ROUTINES WITHIN THE HARVESTING SYSTEM

Harvesting quality assurance was described in D3.1. Minerals inventory Improvements – status after 1st year (Chapter 3). Below only the main findings of harvesting quality assurance are underlined.

All activities, carried out to verify and ensure quality of the data, can be summarized in a few points:

- more frequent run of harvesting procedure;
- setting up a procedure for building referential databases (see Appendix A of this report) for each country/provider;
- individually communication with providers;
- analysis of harvested data;
- preparing new reports and harvesting procedure log for each country.

Main findings and recommendations concern both data providers and the harvesting system itself. The main requirements are:

1. Data providers must fill the data correctly.
2. Special care from providers must be taken towards (XOR) table connections (the mandatory connection to or from only one table).
3. Data providers must ensure that identifiers (INSPIRE ID's) for the same data stays the same all the time.
4. Data providers must ensure to change the versionID and the begin/end lifespan when a modification is made in a national database.
5. Data providers must always use the actual (the latest from svn repository) version of the database
6. Data providers must ensure that hardware & software services work permanently.

We are constantly solving the technical problems of the harvesting system, which we will continue to do in the future as well. The communication process between data providers and harvesting operators must be optimized and speed up so that any updates will be performed in time.

The proposed workflow is:
1. A data provider identifies errors or missing information;
2. A data provider updates a local (national) database and sends information on updated service to the e-mail address harvesting@geo-zs.si;
3. The harvesting is updated accordingly in the shortest possible time (the process of updating might be delayed due to human or technical factor);
4. The data provider gets feedback on updated data via e-mail and checks the application again to verify that the new information has been transferred to the harvesting database successfully.

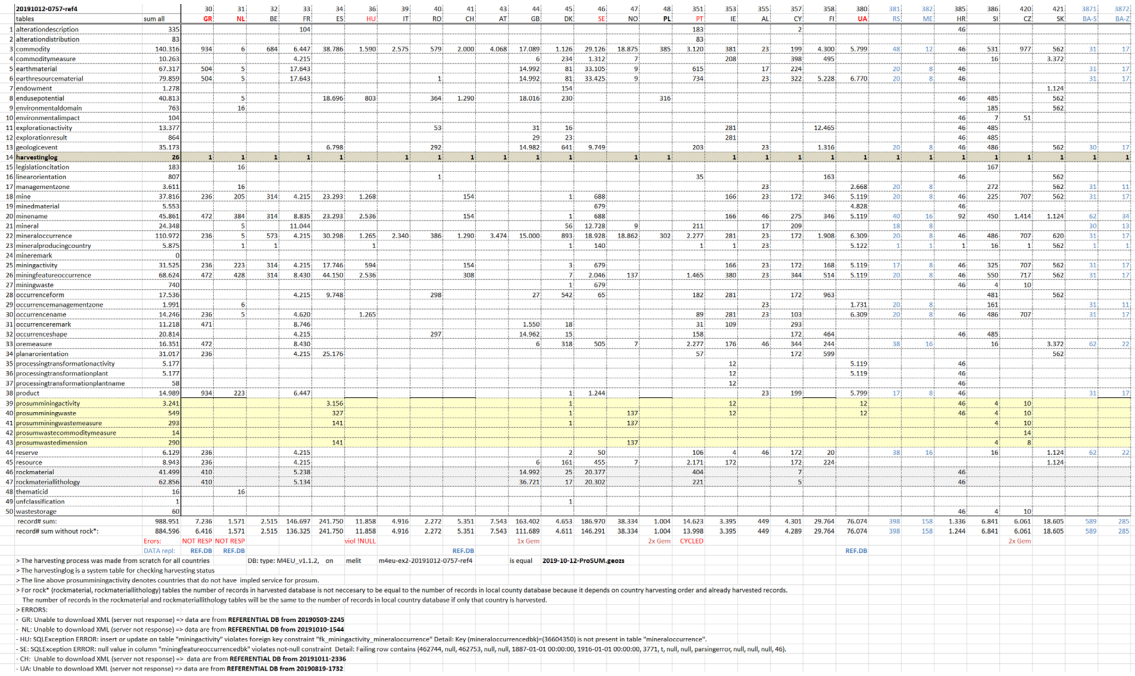
At the moment we are using different technical solutions to check that the harvesting system was performed correctly. For each data provider (country) various technical reports are produced to verify that data is being transmitted correctly:

- Count report,
- Geometry check report,
- Log files,
- Warning/ErrorLog Files,

- CSV files,
- SQL command files,
- M4EU v1.1.2 Database Analysis report.

3.1 Count report

Count report (*Figure 6*) represents a number of harvested records for specific provider/country (in columns) and harvested table (in rows with table name in first column). The second column denotes the number of records harvested from all providers/country.



20191012-0757-ref4		30	31	32	33	34	36	39	40	41	43	44	45	46	47	48	351	353	355	357	358	380	381	382	385	386	420	421	3871	3872
tables	sum all	GR	NL	BE	FR	ES	HU	IT	RO	CH	AT	GB	DK	SE	NO	PL	PT	IE	AL	CY	FI	UA	RS	ME	HR	SI	CZ	SK	BA-S	BA-Z
1 alterationdescription	335				104												183			2					46					
2 alterationdistribution	83																83													
3 commodity	140.316	934	6	684	6.447	38.786	1.590	2.575	579	2.000	4.068	17.089	1.126	29.126	18.875	385	3.120	381	23	199	4.300	5.799	48	12	46	531	977	562	31	17
4 commoditymeasure	10.263				4.215							6	234	1.312	7			208		398	495					16		3.372		
5 earthmaterial	67.317	504	5		17.643							14.992	81	33.105	9		615		17	224			20	8	46				31	17
6 earthresourcematerial	79.859	504	5		17.643				1			14.992	81	33.425	9		734		23	322	5.228	6.770	20	8	46				31	17
7 endowment	1.278												154															1.124		
8 endusepotential	40.813		5			18.696	803		364	1.290		18.016	230			316									46	485		562		
9 environmentaldomain	763		16																							185		562		
10 environmentalimpact	104																								46	7	51			
11 explorationactivity	13.377								53			31	16					281			12.465				46	485				
12 explorationresult	864											29	23					281							46	485				
13 geologicevent	35.173					6.798			292			14.982	641	9.749			203		23		1.316		20	8	46	486		562	30	17
14 **harvestinglog**	**26**	**1**	**1**	**1**	**1**	**1**		**1**	**1**	**1**	**1**	**1**	**1**		**1**	**1**		**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
15 legislationcitation	183		16																							167				
16 linearorientation	807								1								35				163				46			562		
17 managementzone	3.611		16																23			2.668	20	8		272		562	31	11
18 mine	37.816	236	205	314	4.215	23.293	1.268			154			1	688				166	23	172	346	5.119	20	8	46	225	707	562	31	17
19 minedmaterial	5.553													679								4.828			46					
20 minename	45.861	472	384	314	8.835	23.293	2.536			154			1	688				166	46	275	346	5.119	40	16	92	450	1.414	1.124	62	34
21 mineral	24.348		5		11.044								56	12.728	9		211		17	209			18	8					30	13
22 mineraloccurrence	110.972	236	5	573	4.215	30.298	1.265	2.340	386	1.290	3.474	15.000	893	18.928	18.862	302	2.277	281	23	172	1.908	6.309	20	8	46	486	707	620	31	17
23 mineralproducingcountry	5.875		1	1			1						1	140			1	1	23			5.122	1	1	1	16	1	562	1	1
24 mineremark	0																													
25 miningactivity	31.525	236	223	314	4.215	17.746	594			154			3	679				166	23	172	168	5.119	17	8	46	325	707	562	31	17
26 miningfeatureoccurrence	68.624	472	428	314	8.430	44.150	2.536			308			7	2.046	137		1.465	380	23	344	514	5.119	20	8	46	550	717	562	31	17
27 miningwaste	748												1	679											46	4	10			
28 occurrenceform	17.536				4.215	9.748			298			27	542	65			182	281		172	963					481		562		
29 occurrencemanagementzone	1.991		6																23			1.731	20	8		161			31	11
30 occurrencename	14.246	236	5		4.620		1.265										89	281	23	103		6.309	20	8	46	486	707		31	17
31 occurrenceremark	11.218	471			8.746							1.550	18				31	109		293										
32 occurrenceshape	20.814				4.215				297			14.962	15				158			172	464				46	485				
33 oremeasure	16.351	472			8.430							6	318	505	7		2.277	176	46	344	244		38	16		16		3.372	62	22
34 planarorientation	31.017	236			4.215	25.176											57			172	599							562		
35 procesingtransformationactivity	5.177																	12				5.119			46					
36 processingtransformationplant	5.177																	12				5.119			46					
37 processingtransformationplantname	58																	12							46					
38 product	14.989	934	223		6.447								1	1.244					23	199		5.799	17	8	46				31	17
39 prosumminingactivity	3.241					3.156							1					12				12			46	4	10			
40 prosumminingwaste	549					327							1		137			12				12			46	4	10			
41 prosumminingwastemeasure	293					141							1		137											4	10			
42 prosumwastecommoditymeasure	14																										14			
43 prosumwastedimension	290					141									137											4	8			
44 reserve	6.129	236			4.215								2	50			106	4	46	172	20		38	16		16		1.124	62	22
45 resource	8.943	236			4.215							6	161	455	7		2.171	172		172	224							1.124		
46 rockmaterial	41.499	410			5.238							14.992	25	20.377			404			7					46					
47 rockmaterialllithology	62.856	410			5.134							36.721	17	20.302			221			5					46					
48 thematicid	16		16																											
49 unfclassification	1												1																	
50 wastestorage	60																								46	4	10			
record# sum:	988.951	7.236	1.571	2.515	146.697	241.750	11.858	4.916	2.272	5.351	7.543	163.402	4.653	186.970	38.334	1.004	14.623	3.395	449	4.301	29.764	76.074	398	158	1.336	6.841	6.061	18.605	589	285
record# sum without rock*:	884.596	6.416	1.571	2.515	136.325	241.750	11.858	4.916	2.272	5.351	7.543	111.689	4.611	146.291	38.334	1.004	13.998	3.395	449	4.289	29.764	76.074	398	158	1.244	6.841	6.061	18.605	589	285
Erors:		NOT RESP	NOT RESP				viol !NULL					1x Germ				2x Germ	CYCLED										2x Germ			
DATA real:		REF.DB	REF.DB										REF.DB									REF.DB								

> The harvesting process was made from scratch for all countries DB: type: M4EU_v1.1.2, on melit m4eu-ex2-20191012-0757-ref4 is equal **2019-10-12-ProSUM.geozs**
> The harvestinglog is a system table for checking harvesting status
> The line above prosumminingactivity denotes countries that do not have implied service for prosum.
> For rock* (rockmaterial, rockmaterialllithology) tables the number of records in harvested database is not necessary to be equal to the number of records in local county database because it depends on country harvesting order and already harvested records.
The number of records in the rockmaterial and rockmaterialllithology tables will be the same to the number of records in local country database if only that country is harvested.
> ERRORS:
- GR: Unable to download XML (server not response) => data are from **REFERENTIAL DB from 20190503-2245**
- NL: Unable to download XML (server not response) => data are from **REFERENTIAL DB from 20191010-1544**
- HU: SQLException ERROR: insert or update on table "miningactivity" violates foreign key constraint "fk_miningactivity_mineraloccurrence" Detail: Key (mineraloccurrencedbk)=(36604350) is not present in table "mineraloccurrence".
- SE: SQLException ERROR: null value in column "miningfeatureoccurrencedbk" violates not-null constraint Detail: Failing row contains (462744, null, 462753, null, null, 1887-01-01 00:00:00, 1916-01-01 00:00:00, 3771, t, null, null, parsingerror, null, null, null, 46).
- CH: Unable to download XML (server not response) => data are from **REFERENTIAL DB from 20191011-2336**
- UA: Unable to download XML (server not response) => data are from **REFERENTIAL DB from 20190819-1732**

Figure 6: The example of the Count report table.

After each harvesting the '<date_time>_count_report.csv' file is generated automatically, so the report shown in the upper picture is semiautomatic. Please note that the harvesting log is a system table for checking harvesting status, and that the line above *prosumminingactivity* denotes providers/countries that do not have implied service for prosum, the lines marked with yellow are ProSUM tables, and that for rock* (*rockmaterial, rockmaterialllithology*) tables the number of records in the harvested database is not necessary to be equal to the number of records in local county database because it depends on country harvesting order and already harvested records. The number of records in the *rockmaterial* and *rockmaterialllithology* tables will be the same to the number of records in local country database if only that country is harvested.

If the number of harvested records is significantly lower than the number of previously provided records by data provider, it is already an indicator of an error in the harvesting process.

3.2 Geometry Check Report

The Geometry check creates a harvested '<date_time>_geometry_errors.csv' file where the records with logically incorrect geometric data are written for all countries/providers. Please note that the *miningfeatureoccurrencedbk* field is a provider *mineraloccurrencedbk* with country code prefix, that is why from *miningfeatureoccurrencedbk* the country and provider *miningfeatureoccurrencedbk* can be determined. In fact, the geometry check report is the output if the sql sentence

select mineraloccurrencedbk, ST_GeometryType(geometry), ST_ISValid(geometry), ST_ISValidReason(geometry), ST_ASTEXT (geometry) FROM mineraloccurrence_vw WHERE ST_ISValid(geometry) = false ORDER BY mineraloccurrencedbk;

and as results we get information on (*Figure 7*):

1. **mineraloccurrencedbk -** a provider *mineraloccurrencedbk* with country code prefix;
2. **st_geometrytype -** a PostGIS geometry type;
3. **st_isvalid** - a Boolean value false if the geometry is not valid;
4. **st_isvalidreason** - a detailed report of invalid geometries and its reasons
5. **st_astext -** the well-known text representation of the geometry/geography.

mineraloc	st_geometrytyp	st_isvalid	st_isvalidreason	st_astext
4864	ST_MultiPolygo	f	Ring Self-intersection[21.85	MULTIPOL
48109	ST_MultiPolygo	f	Ring Self-intersection[21.82	MULTIPOL
4206124	ST_Polygon	f	Self-intersection[15.133851	POLYGON(
4206773	ST_MultiPolygo	f	Ring Self-intersection[16.58	MULTIPOL
44283345	ST_MultiPolygo	f	Self-intersection[-2.527754	MULTIPOL

Figure 7: The example of the Count report table.

3.3 Log Files

During harvesting process the log files are created for all countries. In file for each country/provider all activities of the harvesting procedure for that provider/country is recorded. The log files can be very big. For example, the Spain log file for harvesting from October 12th 2019 *'20191012-0757_M4EU_ProSUM_harvesting_log_ES-34.log*' has 724.569 lines and its size is 197.035 KB.

Creating those files is not the activity of Mintell4EU directly, but part of the GeoERA Information Platform Project (GIP-P), which main objective is to establish a common platform for organizing, disseminating and sustaining the digital results of all GeoERA projects. Nevertheless, log files serve as background information for the preparation of Warning/ErrorLog Files in Mintell4EU.

3.4 Warning/ErrorLog Files

After each harvesting the **Warning/Error files** are generated for all countries with Warning/error messages in log files. Below are examples of the files generated for harvesting performed on November 10th 2019 at 07:57:

- 955 20191012-0757_WarningErrorLog_AT-43.txt
- 2.462 20191012-0757_WarningErrorLog_CH-41.txt
- 1.755.531 20191012-0757_WarningErrorLog_ES-34.txt
- 928 20191012-0757_WarningErrorLog_FI-358.txt
- 2.484 20191012-0757_WarningErrorLog_GR-30.txt
    - 20191012-0757_WarningErrorLog_HU-36.txt
- 2.341 20191012-0757_WarningErrorLog_NL-31.txt
- 951 20191012-0757_WarningErrorLog_PL-48.txt
- 967 20191012-0757_WarningErrorLog_RO-40.txt
- 394.531 20191012-0757_WarningErrorLog_SE-46.txt
- 952 20191012-0757_WarningErrorLog_SK-421.txt
- 2.474 20191012-0757_WarningErrorLog_UA-380.txt

WARNING(s)/ERROR(s) list (example):

- **Product_Production_Q is null**. The uomproduction is set to null. ProductID: M4EU.PSPR_2 url: http://lnegmineral4eu.lneg.pt:8080/prosum/services?request=GetFeature&service=WFS&version=2.0.0&typeName=prosum:Product_Extension&srsName=EPSG:4258&outputFormat=text/xml;%20subtype=gml/3.2.1
- **NULL value for mineraloccurrence.name** for mineraloccurrencedbk = M4EU.MO_103979
- **value too long** for type character varying(50) for mineraloccurrencedbk = M4EU.MO_400170. Field mineraloccurrence.name $$Halimba IV. (Halimba II. és Halimba III. bü.) - bauxit$$will be shorten on 50 characters.
- SQLException ERROR: insert or update on table "miningactivity" **violates foreign key constraint** "fk_miningactivity_mineraloccurrence"
- SQLException ERROR: null value in column "miningfeatureoccurrencedbk" **violates not-null constraint**
- **Product_Production_Q is null**. The uomproduction is set to null. ProductID: M4EU.PSPR_2 url: http://lnegmineral4eu.lneg.pt:8080/prosum/services?request=GetFeature&service=WFS&version=2.0.0&typeName=prosum:Product_Extension&srsName=EPSG:4258&outputFormat=text/xml;%20subtype=gml/3.2.1
- **NULL value for mineraloccurrence.name** for mineraloccurrencedbk = M4EU.MO_103979
- **value too long** for type character varying(50) for mineraloccurrencedbk = M4EU.MO_400170. Field mineraloccurrence.name $$Halimba IV. (Halimba II. és Halimba III. bü.) - bauxit$$will be shorten on 50 characters.
- SQLException ERROR: insert or update on table "miningactivity" **violates foreign key constraint** "fk_miningactivity_mineraloccurrence"
- SQLException ERROR: null value in column "miningfeatureoccurrencedbk" **violates not-null constraint**
- Unable to download XML from <link for **MappedFeature** service>

- Unable to download XML from < link for **MineralProducingCountry** service>
- Unable to download XML from <link for **MiningFeatureOccurrence** service>
- Unable to download XML from <link for **ManagementRestrictionOrRegulationZone_Extension**>
- Unable to download XML from <link for **Specimen** service>
- Unable to download XML from <link for **Product_Extension** service>
- Unable to download XML from <link for **prosum:MiningActivity** service>
- Unable to download XML from <link for **prosum:ProcessingTransformationPlant** service>
- Unable to download XML from <link for **prosum:ProcessingTransformationActivity** service>
- Unable to download XML from <link for **prosum:MiningWaste** service>
- Unable to download XML from <link for **prosum:Product_Extension** service>

3.5 CSV files

For each country/provider some csv (comma separated value) files are created for purpose of the Mintell4EU Quality Control Application program which is available at https://mintell4eu-qca.geo-zs.si/.

- <countryCode_countryName> _m4eu_base_vw
- <countryCode_countryName>_m4eu_commodity_geologic_data_vw.csv
- <countryCode_countryName>_m4eu_commodity_managementzone_vw.csv
- <countryCode_countryName>_m4eu_commodity_mineraloccurrence_vw.csv
- <countryCode_countryName>_m4eu_commodity_mining_activity_vw.csv
- <countryCode_countryName>_m4eu_commodity_prosum_vw.csv
- <countryCode_countryName>_m4eu_commodity_transf_plant_vw.csv

3.6 SQL command files

Sql command files are generated for each country from log files. They can be used to restore the country database as it was harvested. Please note that harvested database has additionally the harvesting log table, a county code column added for each table, and dbk values with country code in front of the provider dbk value.

3.7 M4EU v1.1.2 Database Analysis report

A database Analysis report is an html routine where we can examine structure of the database but cannot see data. It is a tool that includes database relation diagrams to help user to become familiar with the database.

The documentation of the M4EU v1.1.2 database is stored on *geusgitlab* repository and can be downloaded from https://geusgitlab.geus.dk/m4eu/2017-ProSUM-v1.1.2/tree/master/09-db-m4eu/v1.1.2/DBPostgreSQL/DB_analysis .

The complete directory download is in zip or tar format and must be unpacked in any directory before use. To start using Analysis report the index.html file in unpacked directory must be open in web browser. The main overviews can be seen in *Figure 8* to *Figure 13*.

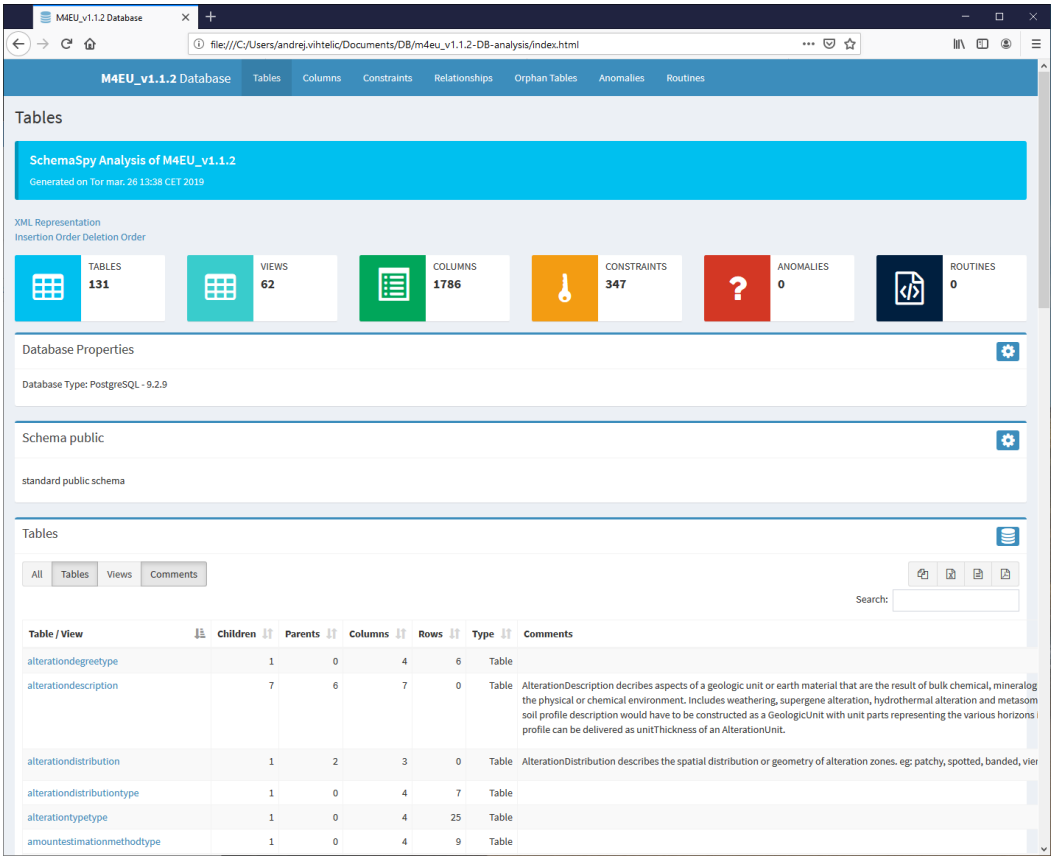


Figure 8: Database tables overview.

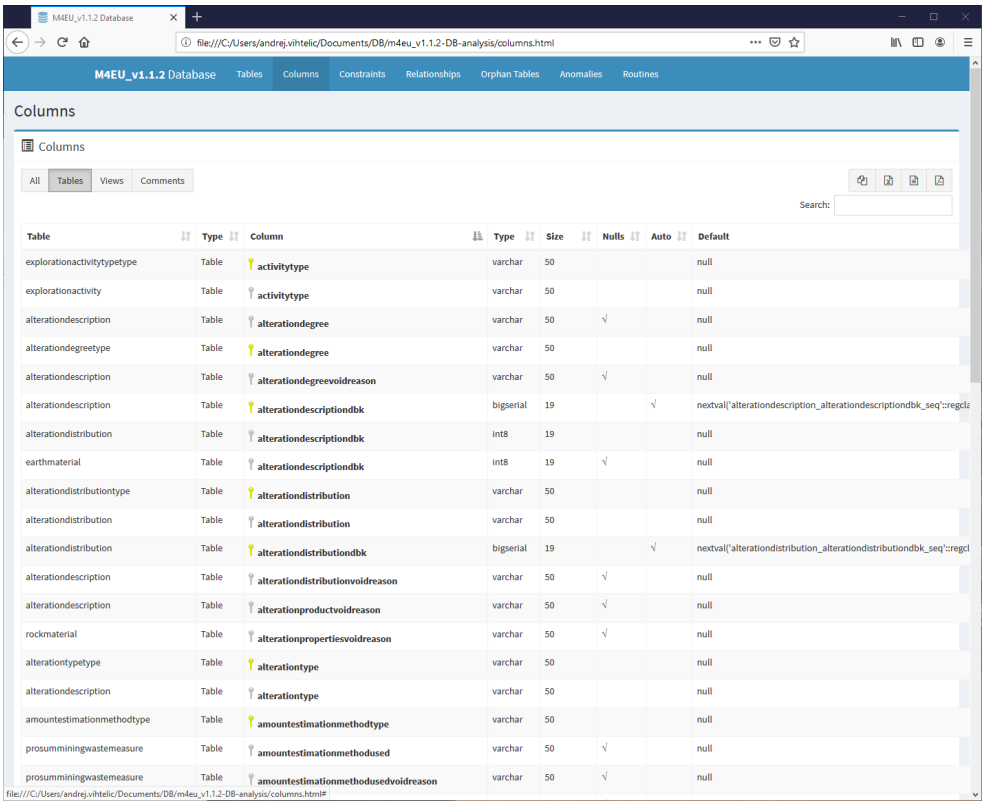


Figure 9: Columns overview.

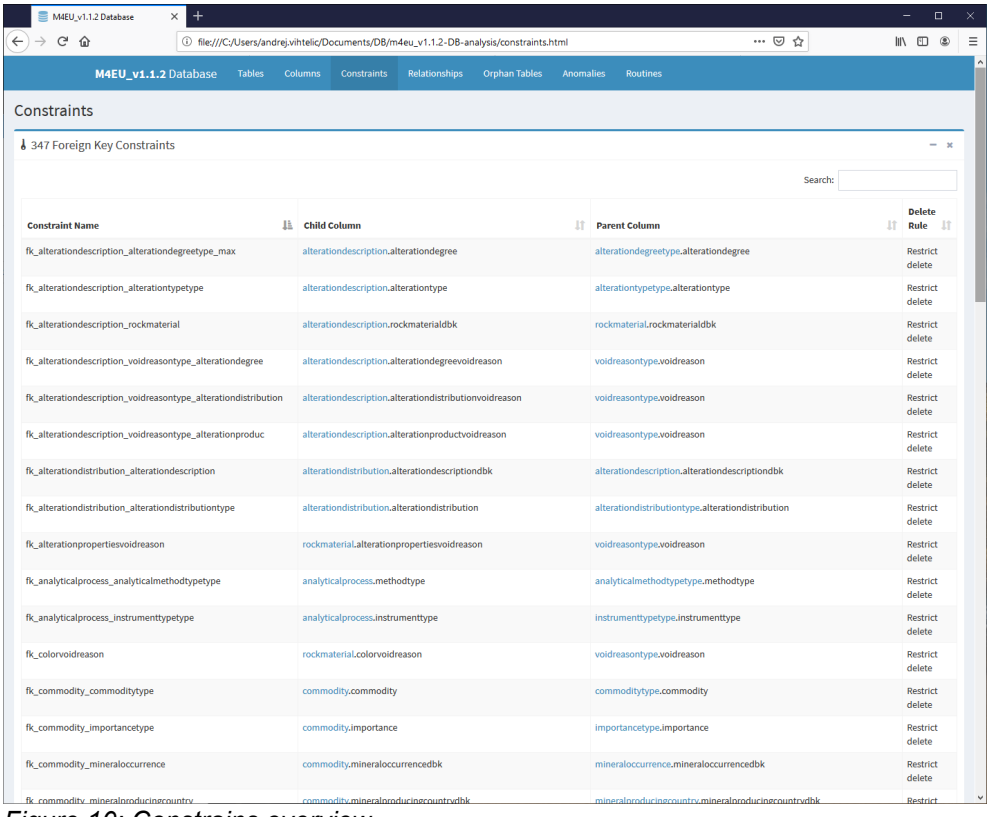


Figure 10: Constrains overview.

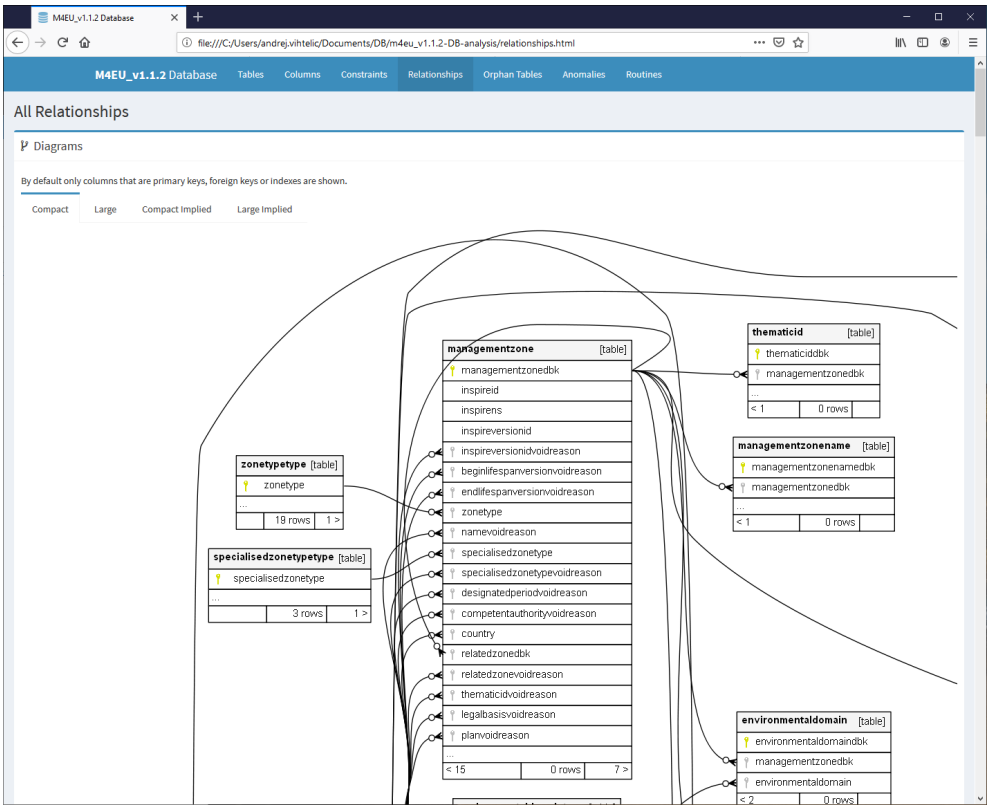


Figure 11: Relationship overview.

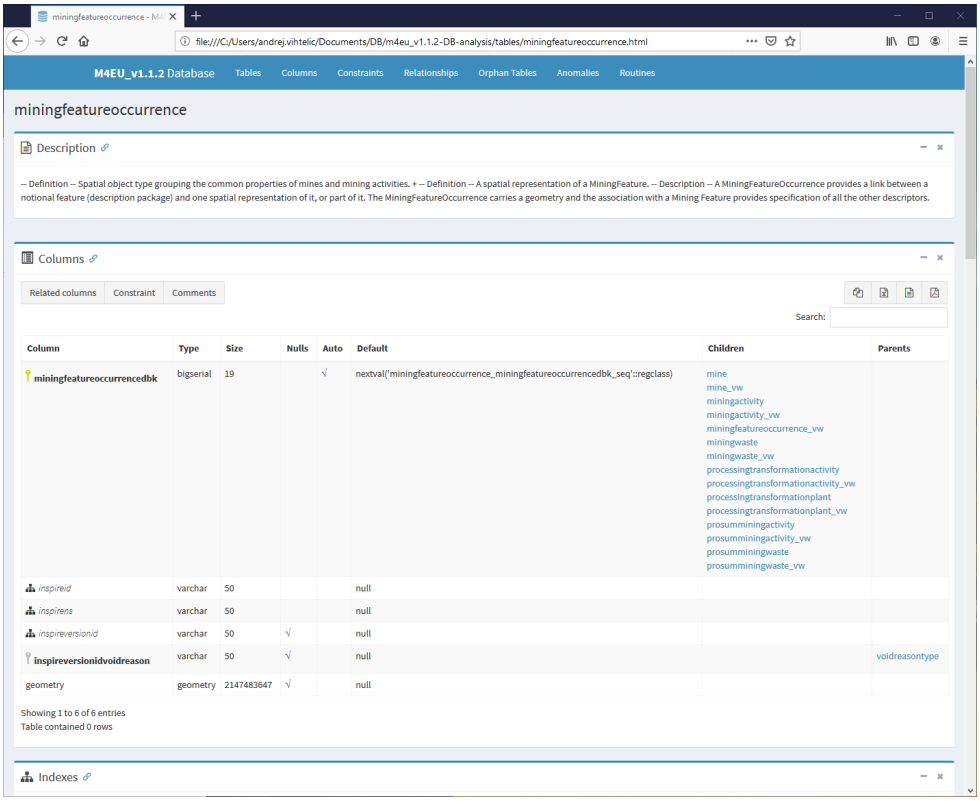


Figure 12: Table MiningFeatureOccurrence overview.

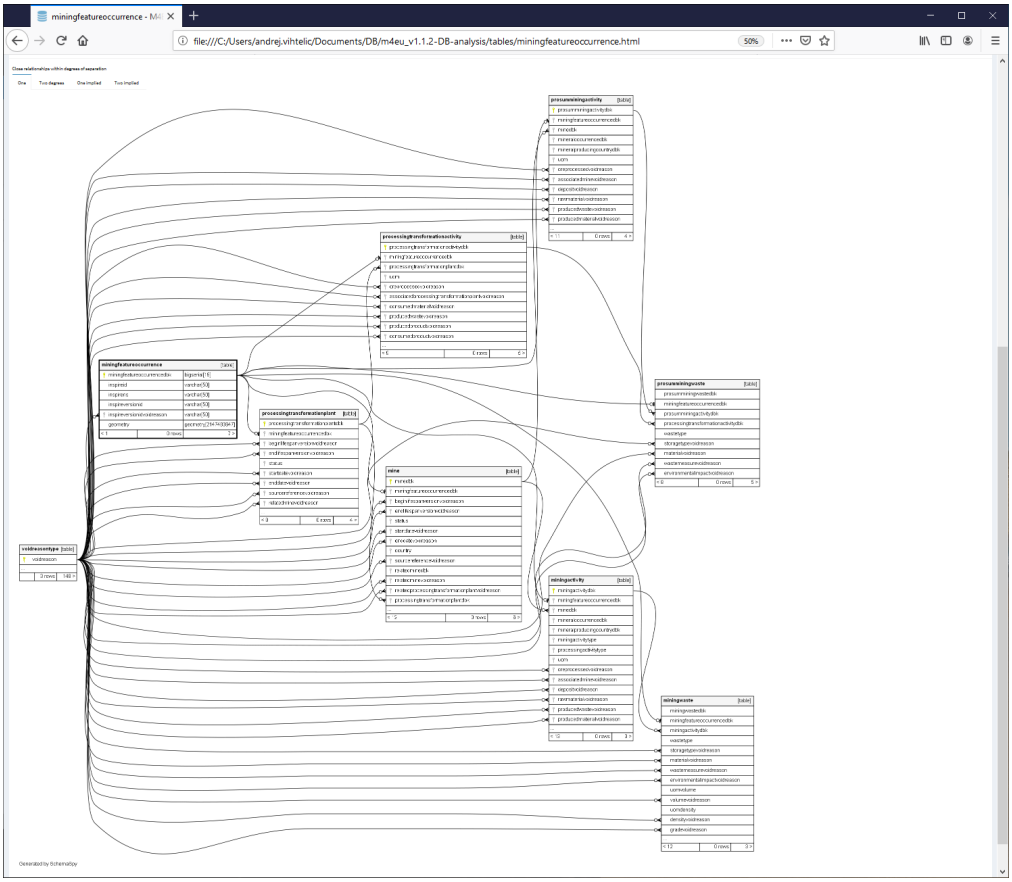


Figure 13: Table MiningFeatureOccurrence relationship – one degree separation.

4 CONCLUSIONS

The existing report provides technical guidance on how to implement MI improvements and e-MYB uploads, which will be updated and upgraded at all times, especially in terms of automation validation and control of harvesting systems.

It is a description of the status of the work carried out after one year after the project start, in a strong interdependence with ORAMA project.

A workshop in Ljubljana is planned for spring 2020, which will look at the needs of data providers, especially in order to present technical solutions and guidance to assist them in the process of mapping their national data into M4eu database. A transparent guidelines scheme will be provided for this purpose and technical report will be upgraded accordingly before the project is completed.

5 REFERENCES

Bide T., Horváth Z., Brown T., Idoine N., Laukó A., Sári K., Sőrés L., Petavratzi E., McGrath E., Bavec S., Rokavec D. (2018b). Deliverable D1.2: Final analysis and recommendations for the improvement of statistical data collection methods in Europe. ORAMA Public Report, 105 p.

Schjøth F., Whitehead D. and WP4 team (2019): G4.1.01.01 Tool Stack Recommendation Guidelines

Cassard D., Tertre F., Bertrand G., Tellez-Arenas A., Schjøth F., Tulstrup J., Heijboer T., Vuollo J., Čápová D., Šinigoj J., Gruijters S., Tomas R., Schubert C. (2014). Deliverable D5.2: Minerals4EU-WP5: The EU-MKDP: specifications of the system. Minerals4EU Public Report, 32 p.
https://geusgitlab.geus.dk/m4eu/2015-Minerals4EU-v0.7.0.2/blob/master/00-cookbook-m4eu/Deliverable/Minerals4EU_WP5_D5%202_07072014_FINAL.pdf
(Last accessed on 12 Nov 2019).

EarthResourceML (ERML) version 2.0, EarthResourceML-Lite version 2.0.1:
http://www.earthresourceml.org/
(*Last accessed on* 12 Nov 2019).

EarthResourceML (ERML) version 3.0:
https://www.seegrid.csiro.au/subversion/xmml/GGIC/trunk/doc/ERML_3_Doc/index.htm
(*Last accessed on* 12 Nov 2019).

GeoSciML: http://www.geosciml.net/
(*Last accessed on* 12 Nov 2019).

Minerals4EU Knowledge Data Platform: http://minerals4eu.brgm-rec.fr/
(Last accessed on 12 Nov 2019).

https://geusgitlab.geus.dk/m4eu/2019-ORAMA-v2.xx/blob/master/00-cookbook-m4eu/G4.1.01%20M4EU%20Tool%20Stack%20guidelines/G4.1.01.01%20Tool%20Stack%20recommendation%20guidelines.pdf
(Last accessed on 12 Nov 2019).

APPENDIX A:
Glossary of key-terms used in D3.2 Technical report

Key-terms descriptions are taken from Orama project, D3.2 Serving aggregated data, Glossary of key terms used in Work Package 3 (available on **https://orama-h2020.eu/wp-content/uploads/ORAMA_WP3_DEL3.2_20191029_v1.0.pdf**).

Data model: A **data model** organizes data elements and standardizes how the data elements relate to one another. The 'EURare – Minerals4EU' data model, as well as the **INSPIRE MR data model**, is an entity-relationship model (ERM). This is an abstract conceptual data model (or semantic data model) used in software engineering to represent **structured data**. This model is transformed into a relational model, which in turn generates a relational database. These conceptual entity/relationship models are developed using **UML** Class Diagram notation. Thus, **data models describe the structure, manipulation and integrity aspects of the data stored in data management systems such as relational databases. They typically do not describe unstructured data**, such as word processing documents, e-mail messages, pictures, digital audio, and video.

Diffusion database: The role of the (Central) Diffusion Database (DB) is to provide the portal with data. These data are sent to the portal using Web services (**WFS, JSON**). In order to speed the process, the structure of the Diffusion DB is optimized for diffusion. This means that its structure does not follow exactly the **data model** which has been 'flattened' or simplified without altering the data. For this project, the (Central) Diffusion DB is hosted by BRGM in France. At the origin – before the addition of specific tables used for optimization the **Diffusion DB** is an exact copy of the **Harvesting DB** made by using **SQL** scripts.

e-MYB: electronic version of the BGS' Minerals Yearbook.

ERML or **EarthResourceML** data model: This is the 'international – world-wide' fully compliant version of the **INSPIRE MR** data model (http://www.earthresourceml.org/). This data model is managed by the IUGS/CGI/ERMLWG and used in Europe, North America and Australia. **ETL: Extract, Transform and Load** process in database management that performs data **extraction** from homogeneous or heterogeneous data sources; data **transformation** for storing in the proper format or structure for the purpose of querying and analysis; and data **loading** into the final target.

ERML-Lite is a model and schema for simple map services (eg, WMS and WFS Simple Features). It is an abridged version of the full **EarthResourceML** model and can be used to deliver simplified views on mineral occurrences and their commodities, mines, mining activities and mine waste products.

EU-MKDP: the European Union Minerals Knowledge Data Platform developed in the frame
of the EU-FP7 Minerals4EU project. The **IKMS** (the EURare's Integrated Knowledge Management System) and the **EU-MKDP** are based on the same architecture and share

numerous components.

Harvesting system: The **Minerals4EU (Central) Harvesting System** including the database periodically refreshes the information available about mineral resources by requesting data from the data providers using INSPIRE compliant Web services (**WFS**). This DB is structured in such a way that a large part exactly reflects the **INSPIRE** Mineral Resources (MR) data model, but it also includes the ProSUM mining waste modifications. During the harvesting phase the data that is received is checked whether codelists conform to the INSPIRE registry code list values and other data have the correct format (e.g., dates, numbers…). The **Minerals4EU Harvesting DB** is hosted by the Geological Survey of Slovenia (GeoZS) and connected to the **Diffusion DB** using **SQL** scripts. This Harvesting DB delivers data related to primary mineral resources and mining wastes. **The ProSUM Harvesting DB** has been built using the **ProSUM Unified data model**, and is dedicated to the urban mine (WEEE, ELV & BATT). It is hosted by the Geological Survey of Denmark and Greenland (GEUS) and data is currently extracted from Excel sheets that have a standardized format (**portrayals**) provided by the different ProSUM work packages.

INSPIRE: The INSPIRE directive lays down a general framework for a Spatial Data Infrastructure (SDI) for the purposes of European Community environmental policies or activities which may have an impact on the environment. The INSPIRE Directive entered into force on 15 May 2007. INSPIRE is based on the infrastructures for spatial information established and operated by the Member States of the European Union. The directive addresses 34 spatial data themes needed for environmental applications, among which Mineral Resource and Geology. To ensure that the spatial data infrastructures of the Member States are compatible and usable in a community and transboundary context, the INSPIRE Directive requires that additional legislation or common Implementing Rules (IR) are adopted for a number of specific areas (metadata, interoperability of spatial data sets and services, network services, data and service sharing and monitoring and reporting). These are published either as Commission Regulations or as Decisions. See: http://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:32007L0002&from=EN

INSPIRE MR data model: This is the European approved data model for mineral resources (MR), including both primary and secondary (i.e., Mining wastes) resources. However, mining wastes do not belong to the core part of this data model, being only an extension. One objective of the ProSUM project is to improve and extend the mining wastes part of the INSPIRE MR data model (http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_MR_v3.0.pdf). The mineral resources data model used in ProSUM is the M4EU data model (Minerals4EU project) directly derived (with **GeoSciML** extensions for geology) from the INSPIRE MR data model.

JSON (**JavaScript Object Notation**) is a data-interchange format. Although not a strict subset, JSON closely resembles a subset of JavaScript syntax. Though many programming languages support JSON, JSON is especially useful for JavaScript-based apps, including websites and browser extensions.

M4EU: Abbreviation for Minerals4EU. Used only for the data model, the database model and the database, not for the project itself.

Referential database. The purpose of the referential database is to protect the related data from being accidentally modified and/or deleted, but at the same time to prevent that the harvesting of the entire database would be empty for a certain country/provider due to service failure or any other technical reasons.

SQL (script): **SQL** or **Structured Query Language** is a special-purpose programming language designed for managing data held in a relational database management system (RDBMS), or for stream processing in a relational data stream management system (RDSMS).

Structured data refers to any data that resides in a fixed field within a record or file. This includes data contained in relational databases and spreadsheets. Structured data first depends on creating a **data model**, i.e., a model of the types of business data that will be recorded and how they will be stored, processed and accessed. This includes defining what fields of data will be stored and how that data will be stored: data type (numeric, currency, alphabetic, name, date, address) and any restrictions on the data input (number of characters; restricted to certain terms...). Structured data has the advantage of being easily entered, stored, queried and analyzed.

ToolStack refers to a set of tools/softwares needed to perform a complex task such that no additional tools/softwares are needed to support this task. **UML,** the **Unified Modeling Language** is a standardized general-purpose modeling language in the field of software engineering. It is a graphical language for visualizing, specifying, constructing, and documenting the artifacts of a software-intensive system.

UNFC: United Nations Framework Classification for Resources. **Unstructured Data** (or **unstructured information**) refers to information that either does not have a pre-defined data model or is not organized in a pre-defined manner. Unstructured information is typically text-heavy, but may contain data such as dates, numbers, and facts as well. This results in irregularities and ambiguities that make it difficult to understand using traditional programs as compared to data stored in fielded form in databases or annotated (semantically tagged) in documents.

Web service: is defined by the World Wide Web Consortium (W3C - https://www.w3.org/) as 'a software system designed to support interoperable machine-to-machine interaction over a network'. Several types of Web services are used such as Web Feature Services (**WFS**) allowing the transfer of data, and Web Map Services (**WMS**) allowing the visualization of maps.