MINERAL INTELLIGENCE FOR EUROPE

Deliverable D3.3

Quality control system for harvesting report

Authors and affiliation:
Blaž Bahar, Špela Kumelj, Andrej Vihtelič (GeoZS)

E-mail of lead author:
blaz.bahar@geo-zs.si

Version:13-10-2021

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	D3.3	
Dissemination level	Public	
Deliverable name	Quality control system for harvesting report	
Work package	WP3, Minerals Inventory	
Lead WP/Deliverable beneficiary	GeoZS	
Deliverable status		
Submitted (Authors)	13/10/2021	Blaž Bahar, Špela Kumelj, Andrej Vihtelič
Verified (WP leader)	13/10/2021	Špela Kumelj
Approved (Coordinator)	27/10/2021	Lisbeth Flindt Jørgensen

GENERAL INTRODUCTION

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 MINTELL4EU is to improve the European Knowledge Base on raw materials as there are several sources of non-harmonized data with different coverages developed for different purposes during national and international projects over recent decades. All data are shared at the European Geological Data Infrastructure, EGDI.

Tasks include updating the electronic Minerals Yearbook produced in the Minerals4EU project as well as extending the spatial coverage and quality of data currently in the Minerals Inventory. Furthermore, MINTELL4EU aims 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. Finally, the applicability of the UNFC classification system for obtaining more accurate Pan-European mineral inventories is tested through a large number of case studies on different commodities across Europe.

MINTELL4EU has 27 partners, each representing a national or regional geological survey organisation, from 25 European countries.

EXECUTIVE REPORT SUMMARY

This report is an overview of the existing harvesting quality control system, and how it is developed to check the mapping of national data on mineral resources to the MIN4EU DB (v2020.08.02). We describe different interactive tools which were developed to support data providers in performing data quality control of their data after each harvesting.

Detailed technical description of the harvesting system is described within the GIP-P project in the deliverable "*D7.3 Final version of Central database / harvesting*".

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1 OVERVIEW

One important 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), stored in, what used to be M4EU DB and renamed into the MIN4EU DB.

MIN4EU DB is associated with technical documentation addressing procedures and quality, which can be summarized into:

a) Technical guidelines and webinars set up within the ORAMA project, modified and supplemented partly within MINTELL4EU and partly within GIP-P
b) Technical routines of the harvesting system and MINTELL4EU Quality Control Application, both developed within MINTELL4EU to ease the data quality control, mapped through harvesting into MIN4EU DB.

MIN4EU harvesting is built on the MIN4EU data model as well as on WFS services. The existing technical routines that deal with the correctness of the harvesting process and its quality are interdependent. For example, if there are errors in the data model and/or WFS services, there will also be errors in the harvesting process, and if the errors do occur this will be visible on the error detection tools.

MINTELL4EU Quality Control Application (QCA) was developed to examine the latest data reported by data providers and can be assess at https://mintell4eu-qca.geo-zs.si/. Accordingly, a harvesting system for the collection and validation of mineral resources data has been established by the Geological Survey of Slovenia (GeoZS), which periodically collects data from different sources (data providers) and refreshes the MIN4EU database using an INSPIRE compliant WFS (Figure 1).

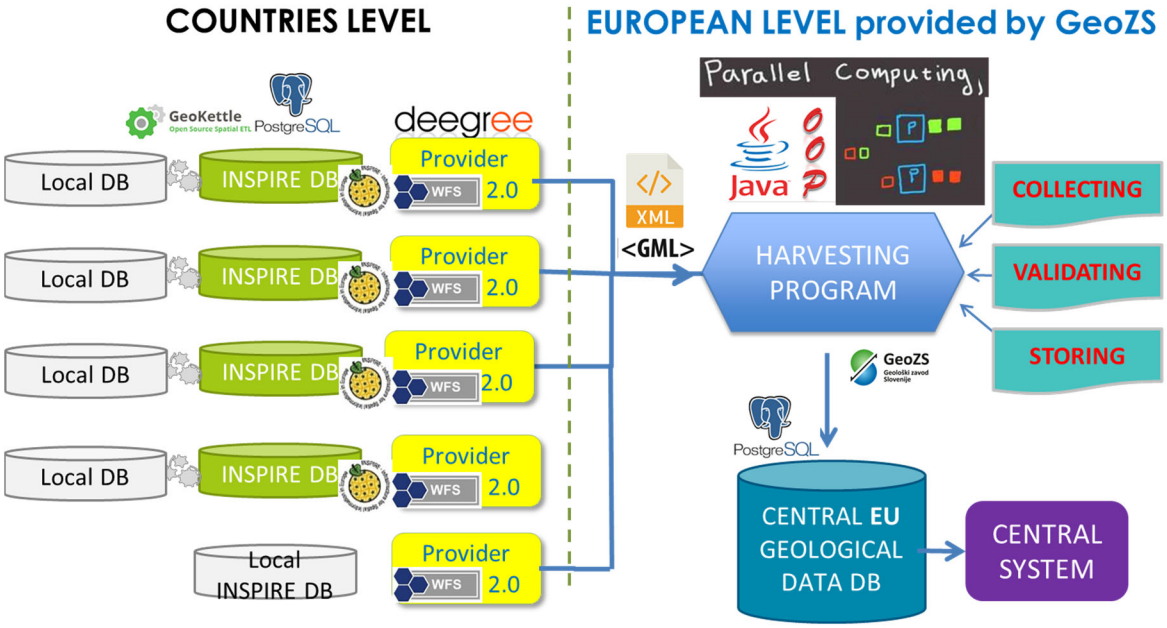

Figure 1. *Harvesting system architecture*

Currently, the harvesting is performed at least once a month for 36 data providers from 31 countries across Europe – Figure 2. The process utilizes INSPIRE as a common standard for data model, data specifications, and data services.

Each time harvesting is performed, several technical routines or error detection records are created. These allow online data quality control and verify that the harvesting was successful. The routines were first introduced and described in the D3.2. Technical Guidelines (Chapter 3) released in November 2019, but at that time they were organized as individual files in folders. Now, technical routines are grouped and available under a common website, called Harvesting info website (https://harvesting.geo-zs.si/main). More details are described in Chapter 2 of this report.

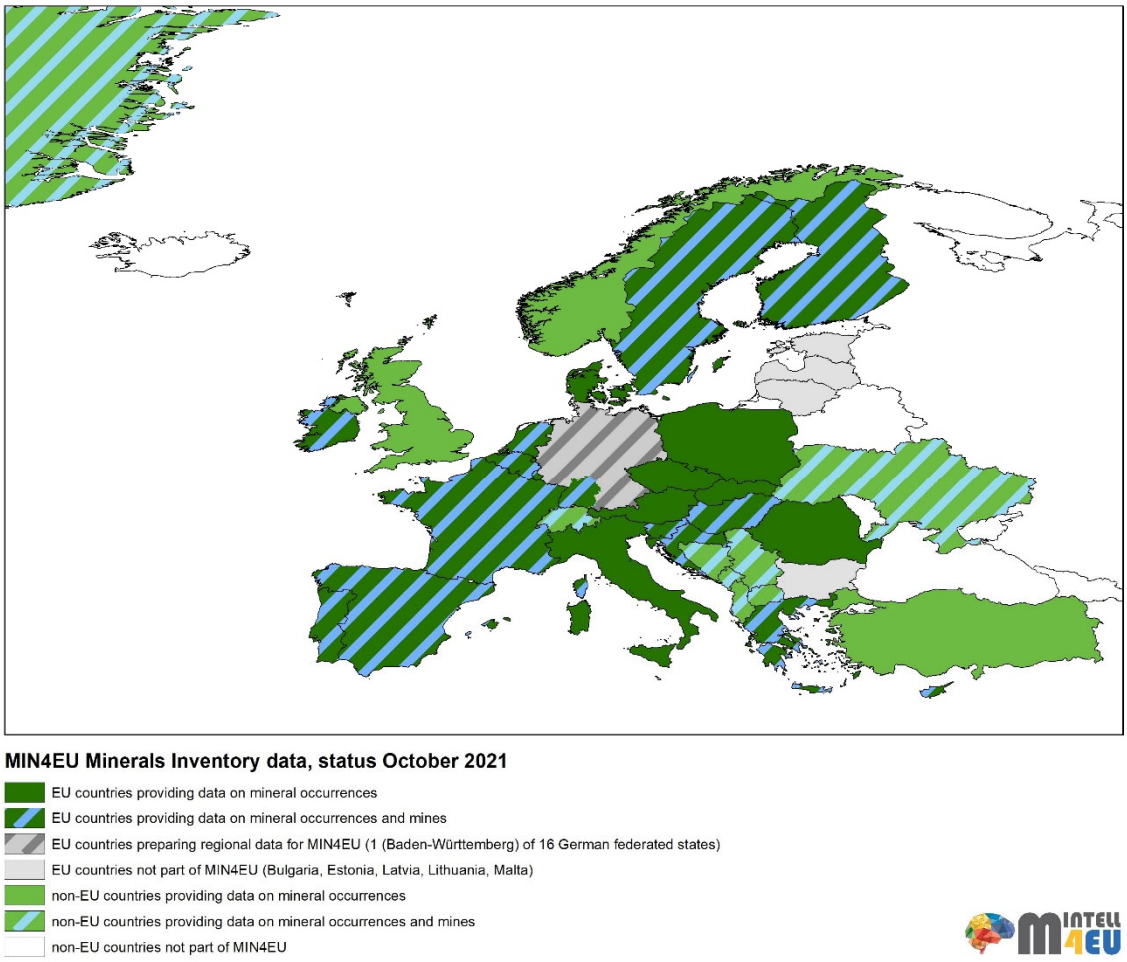

Figure 2. *Spatial coverage of MIN4EU DB Minerals Inventory data*

During the harvesting phase the received data is checked, e.g., if code lists conform to the INSPIRE registry code list values, and if data have the correct format (e.g., dates, numbers…), the correctness of entered void-reason fields is checked. The resulting MIN4EU harvesting DB is then transferred to The Geological Survey of Denmark and Greenland (GEUS) that publish the data at EGDI (Figure 1).

2 DESCRIPTION OF THE HARVESTING INFO WEBSITE

The harvesting info website (https://harvesting.geo-zs.si/main) contains information related to the harvesting of MIN4EU DB data (Figure 3).

On the top of the index page there are links to different harvesting quality control tools running on a GeoZS server. Data provider can access:

- MIN4EU service status report,
- MIN4EU count report,
- MIN4EU alert report,
- MIN4EU chart report,
- MIN4EU providers DB info and
- MIN4EU Quality Control Application.

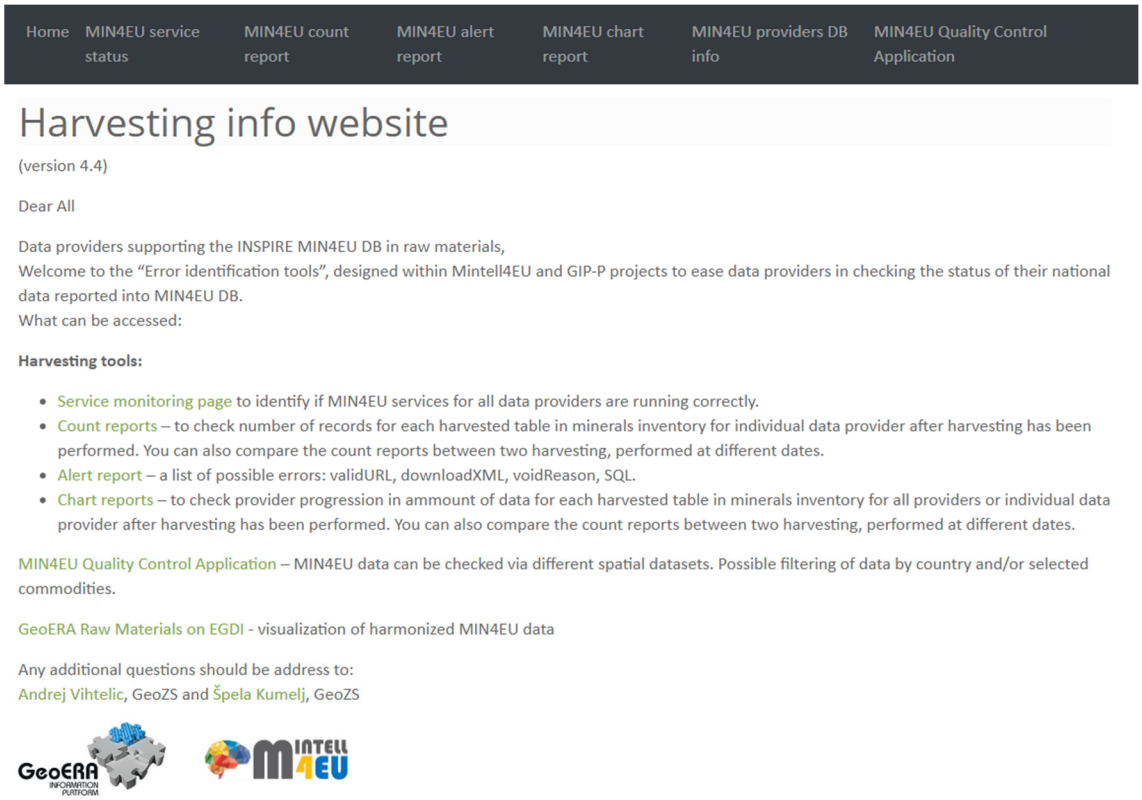

Figure 3. *Index page of harvesting info website*

2.1 MIN4EU service status

MIN4EU Service Status, available at https://data.geus.dk/egdi_monitor_smiley/MonitorSmileym4eu.html, monitors the MIN4EU data services for each data provider. It is a web application that displays the response status of MIN4EU data services from different sources (data providers) with

different coloured smileys. A green smiley means that the services are responding, a yellow smiley means that the response is slow, and a red smiley means that some or all services are not responding.

MIN4EU service status was developed in collaboration with the GeoERA GIP-P project and currently monitors the MIN4EU services for version 1.1.2 (Figure 4) and version 2020.08.02 (Figure 5). Both versions are available, because some data providers are not part of MINTELL4EU and their services will not be updated to a new version.

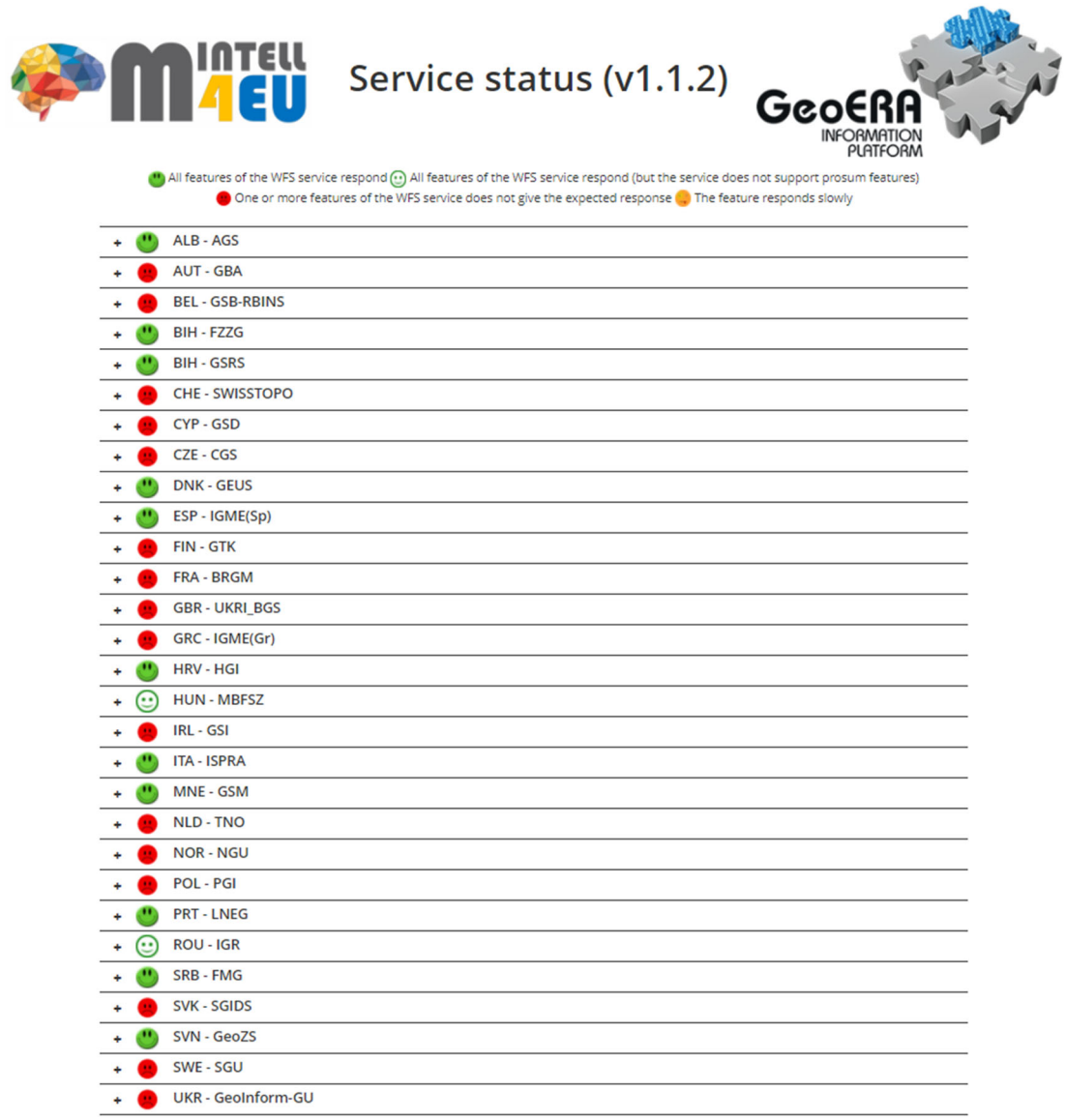

Figure 4. *Status of the MIN4EU data services for v1.1.2*

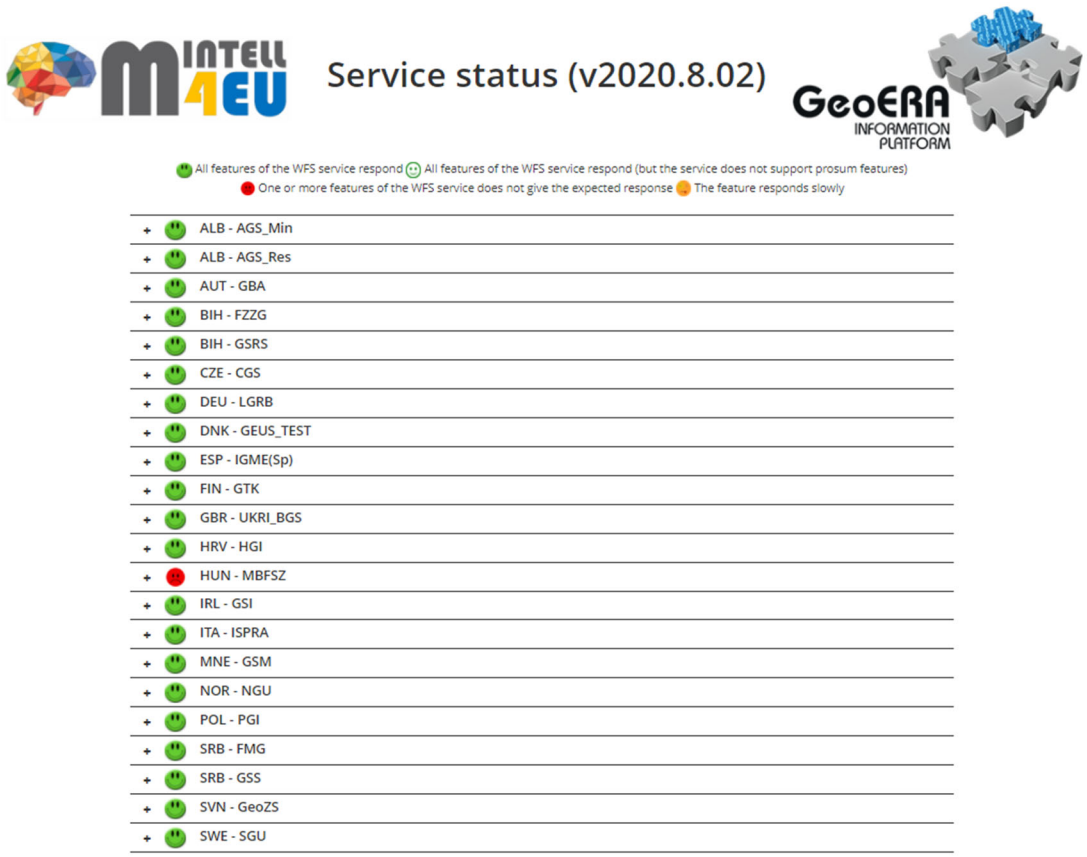

Figure 5. *Status of the MIN4EU data services for v2020.8.02*

By clicking on the + sign, next to the smiley, more detailed information about a particular service can be viewed, as displayed on Figure 6.

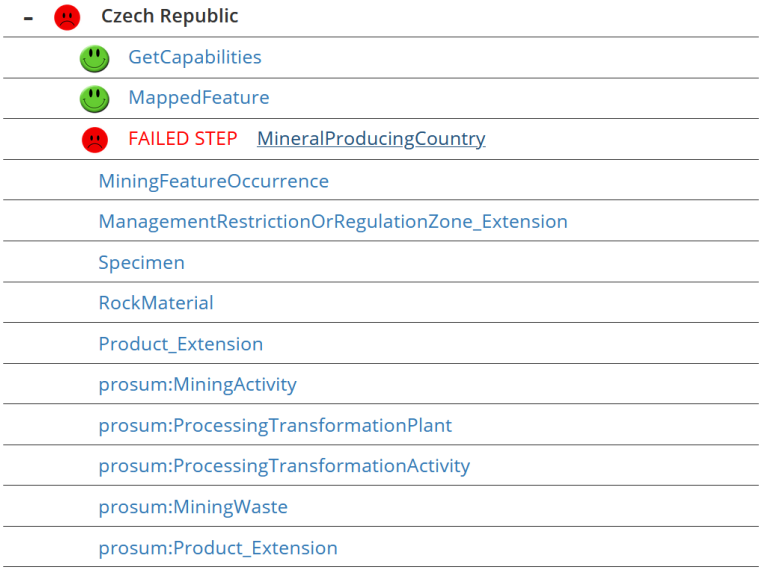

Figure 6. *Information about a particular service*

By clicking on the name of the service, also the time service needed to respond, and a response code over time can be viewed, as displayed on Figure 7.

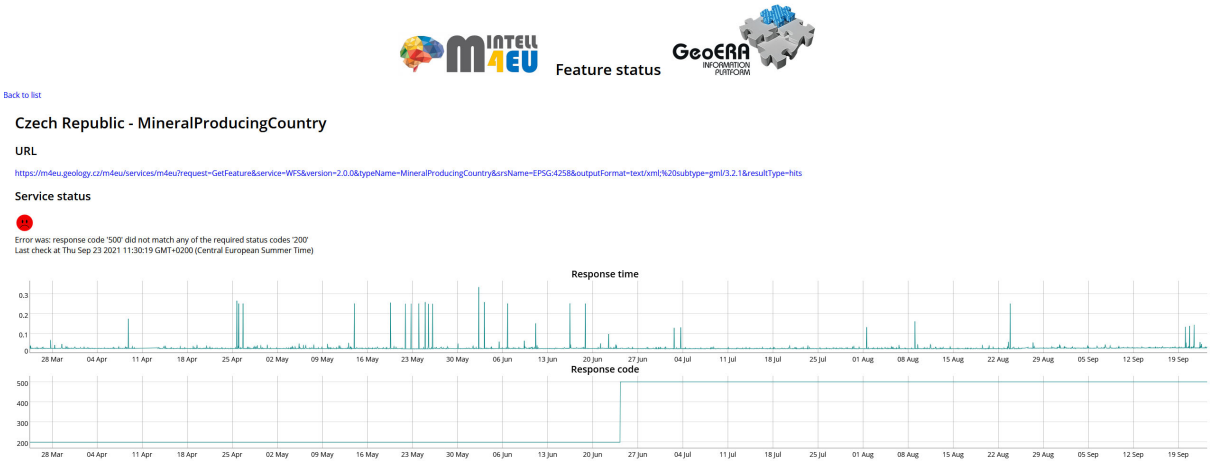

Figure 7. *A service's response code and time*

2.2 MIN4EU count report

The MIN4EU count report represents a number of harvested records for specific provider/country (in columns) and harvested table (in rows with table name in first column).

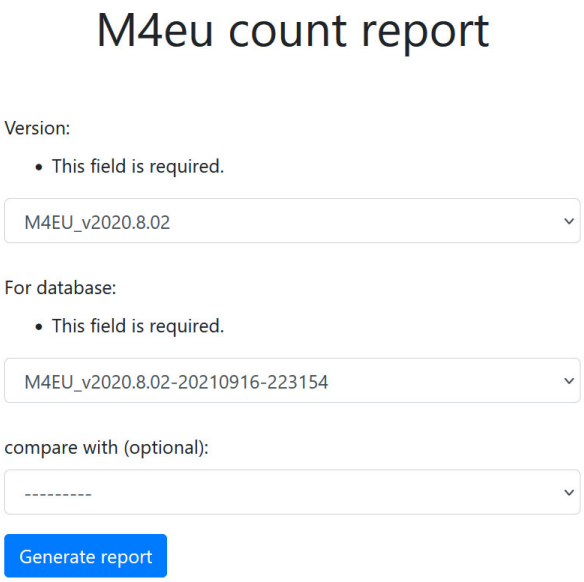

Figure 8. *A form for displaying a count report, example 1*

The last-successful database is the database without gaps, meaning the database contains data from all data providers. For those data providers, for which the harvesting process failed, data were built automatically from their referential databases (the last successful harvesting for that data provider).

To display the count report, first the version (for exampleM4EU_v2020.8.02) and the resulting database of the harvesting process must be chosen. Figure 8 shows an example (example 1) how to set a count report for the resulting database of the harvesting process ran on September 16th at 22:31.

Optionally, it is possible to compare the number of records in the resulting harvested database with the number of records in the data providers local database (v. 2020.8.02 only) – displayed on Figure 9 or with the number of records in the resulting database of another harvesting process.

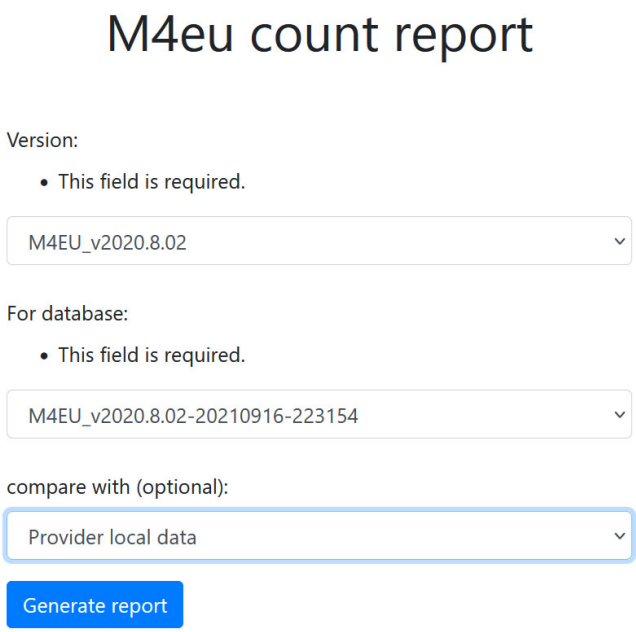

Figure 9. *A form for displaying a count report, example 2*

A count report of selected examples in Figure 8 and Figure 9 is returned in tabular form, as displayed at Figure 10. On the far-left side of the count report, you can find information about:

- Date - the date when the harvesting process finished is displayed for each data provider,
- Status - shows if harvesting process finished successfully or not: *finished*, (without any warnings), *warning* (some warnings occurred) or *failed* (harvesting did not finish successfully) and
- SUM – sum of records over all data tables for each data provider

The count report can also be saved as pdf.

M4EU_v2020.8.02-20210916-223154 m4eu count report (database with gaps)

Save as pdf | Get help

Number of records per data provider for minerals inventory data

	Table	Sum	ALB AGS_Min	ALB AGS_Res	AUT GBA	BIH FZZG	BIH GSRS	CZE CGS	DEU LGRB	DNK GEUS_TEST	ESP IGME(Sp)	FIN GTK	GBR UKRI_BGS	HRV HGI	IRL GSI	MNE GSM	NOR NGU	POL PGI
Data			2021-09-16	2021-09-16	2021-09-16	2021-09-16	2021-09-16	2021-09-16	2021-09-16	2021-09-16	2021-09-17	2021-09-16	2021-09-17	2021-09-16	2021-09-16	2021-09-16	2021-09-17	2021-09-17
Status			FINISHED	FINISHED	FINISHED	FINISHED	FINISHED	FINISHED	FINISHED	FAILED	FINISHED	WARNING	FINISHED	FINISHED	FINISHED	WARNING	WARNING	WARNING
Sum		814,021	9,555	2,152	25,858	2,085	1,283	25,383	12,597	3,244	230,966	27,333	278,206	4,546	202	1,433	33,595	63,643
1	alterationdescription																	
2	alterationdistribution																	
3	analyticalprocess	327	221													51		
4	commodity	127,565	551	133	9,318	89	53	4,177	2,150	852	37,112	2,522	25,107	187	37	60	17,612	18,492
5	commoditymeasure	1,669				49				76		1,519	6		12		7	
6	constituentpart																	
7	documentcitation	89,005	221	133	5,622	53	53	7,888		72	11,402	8,323	40,921			47		
8	earthmaterial	36,083	551	133		90	53			76			21,928	183		59	9	
9	earthresourcematerial	41,599	1,102	266		182	106		2,169	76		1,598	21,928	366		102	9	
10	endowment	42								42								
11	endusepotential	78,717	551			81	53			110	18,696		30,082	183		51		28,370
12	environmentaldomain	656	144	40		25	32							94		38		
13	environmentalimpact	99						99										
14	explorationactivity	8,113	221			81	53			15		6,943	34	183	25	51		
15	explorationresult	1,153	221			81	53			22			32	183	25	51		
16	geologicevent	37,675	221	18		81	52		2,074	498	6,798	707	21,886	180		51		

Figure 10. *Count report of harvesting, performed on September 16th at 22:31*

By clicking on *warning* and *failed* status, the reason(s) for getting such status is/are displayed. An example can be seen on Figure 11. The most common reason for *warning* status is incorrectly entered void-reason data (explaining why certain information is not reported or available). The reasons for getting *failed* status are usually non-working URLs and/or errors that occur with incorrect data entries in(to) the database.

voidReason (8)

Count	Description
1407x	child of "endusePotential" for table: "MINERALOCCURRENCE" is not set and therefore VOIDREASON (schema: nilReason) must exist
30x	child of "area" for table: "MINERALOCCURRENCE" is not set and therefore VOIDREASON (schema: nilReason) must exist
1730x	child of "depth" for table: "MINERALOCCURRENCE" is not set and therefore VOIDREASON (schema: nilReason) must exist
16781x	child of "endlifespanversion" for table: "MINERALOCCURRENCE" is not set and therefore VOIDREASON (schema: nilReason) must exist
16781x	child of "beginlifespanversion" for table: "MINERALOCCURRENCE" is not set and therefore VOIDREASON (schema: nilReason) must exist
16781x	child of "classification" for table: "MINERALOCCURRENCE" is not set and therefore VOIDREASON (schema: nilReason) must exist
83905x	child of "oreAmount" for table: "MINERALOCCURRENCE" is not set and therefore VOIDREASON (schema: nilReason) must exist
40x	child of "dimension" for table: "MINERALOCCURRENCE" is not set and therefore VOIDREASON (schema: nilReason) must exist

Figure 11. *An example of warning status after harvesting was performed*

When two count reports of different harvesting processes are compared, there is an additional option to display the count report for the first harvesting process, for the second harvesting process or the difference between them, as displayed on Figure 12.

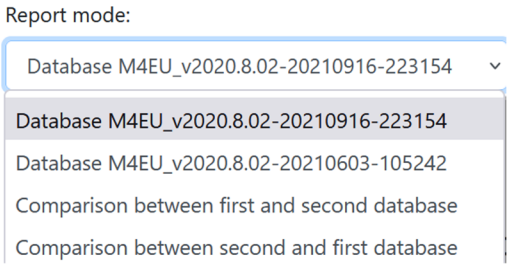

Figure 12. *Options to display count report when comparing resulting databases of two harvesting processes*

When comparing two harvesting processes, the count report is a bit different. The green coloured cell denotes that at the chosen harvesting process more records were harvested, compared to any previous harvesting process, while the red coloured cell denotes that less records were harvested. Example of such count report is shown in Figure 13.

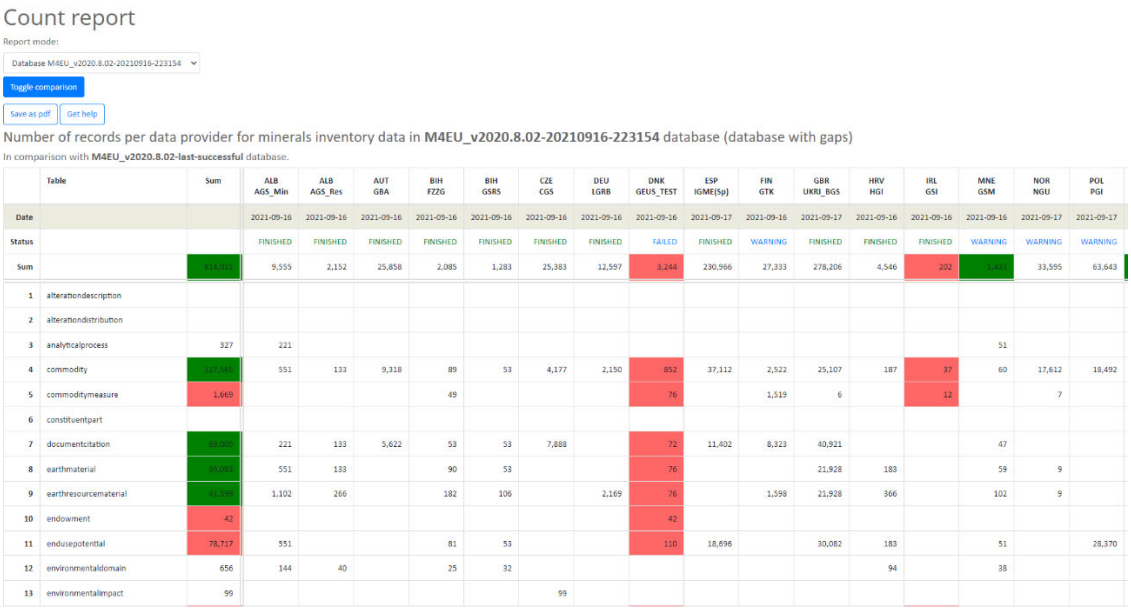

	Table	Sum	ALB AGS_Min	ALB AGS_Res	AUT GBA	BIH FZZG	BIH GSRS	CZE CGS	DEU LGRB	DNK GEUS_TEST	ESP IGME(Sp)	FIN GTK	GBR UKRI_BGS	HRV HGI	IRL GSI	MNE GSM	NOR NGU	POL PGI
Date			2021-09-16	2021-09-16	2021-09-16	2021-09-16	2021-09-16	2021-09-16	2021-09-16	2021-09-16	2021-09-17	2021-09-16	2021-09-17	2021-09-16	2021-09-16	2021-09-16	2021-09-17	2021-09-17
Status			FINISHED	FINISHED	FINISHED	FINISHED	FINISHED	FINISHED	FINISHED	FAILED	FINISHED	WARNING	FINISHED	FINISHED	FINISHED	WARNING	WARNING	WARNING
Sum		914,021	9,555	2,152	25,858	2,085	1,283	25,383	12,597	3,244	230,966	27,333	278,206	4,546	202	1,242	33,595	63,643
1	alterationdescription																	
2	alterationdistribution																	
3	analyticalprocess	327	221													51		
4	commodity	127,345	551	133	9,318	89	53	4,177	2,150	852	37,112	2,522	25,107	187	37	60	17,612	18,492
5	commoditymeasure	1,669				49				76		1,519	6		12		7	
6	constituentpart																	
7	documentcitation	89,005	221	133	5,622	53	53	7,888		72	11,402	8,323	40,921			47		
8	earthmaterial	85,083	551	133		90	53			76			21,928	183		59	9	
9	earthresourcematerial	43,599	1,102	266		182	106		2,169	76		1,598	21,928	366		102	9	
10	endowment	42								42								
11	endusepotential	78,717	551			81	53			110	18,696		30,082	183		51		28,370
12	environmentaldomain	656	144	40		25	32							94		38		
13	environmentalimpact	99						99										

Figure 13. *Count report that enables comparison*

2.3 MIN4EU alert report

The MIN4EU alert report contains all alerts that occurred during harvesting for a single data provider, and generates a list of possible errors: validURL, downloadXML, voidReason, SQL, etc. To display the alert report, first the resulting database of the harvesting process and the data provider must be chosen, as displayed on Figure 14.

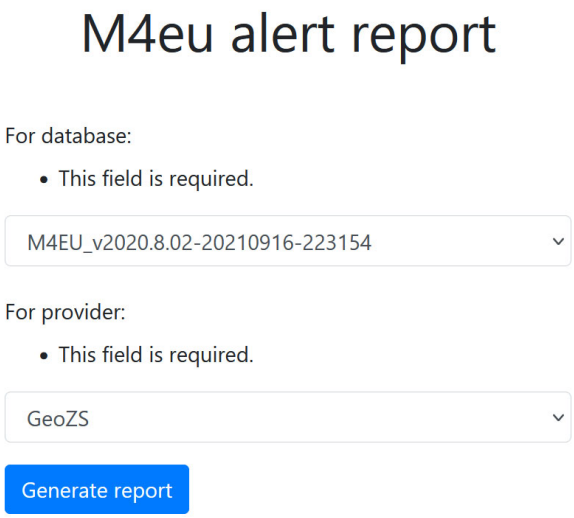

Figure 14. *A form for displaying alert report*

After clicking on Generate report the warnings/errors that occurred during the chosen harvesting process for a particular data provider will be shown (see Figure 11).

2.4 MIN4EU chart report

The MIN4EU chart report shows the number of records in one or multiple tables for one or all data providers for a given period, displayed as a chart. To display the chart report, first the database version, the data provider, the table, and the starting and ending date must be chosen. It is also possible to display the chart for all data providers and/or all tables. If a certain data provider is chosen in the form (as displayed on Figure 15), the chart will show the number of records in selected tables for this data provider (see Figure 16). If all data providers are chosen in the form and a certain table (see Figure 17), the chart will show the number of records in selected table per data provider (see Figure 18).

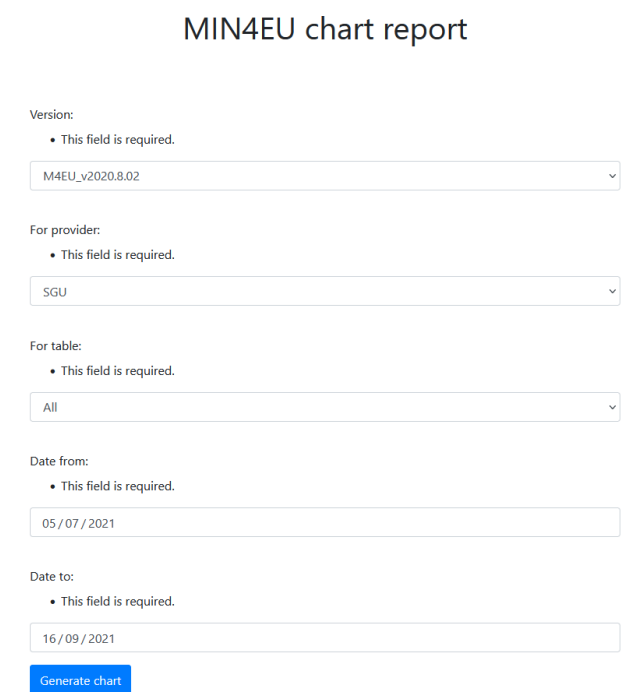

Figure 15. *A form for displaying chart report, example 1*

Figure 15 shows an example of a chart report form, where the number of records per data table(s) for a selected data provider in a selected time frame is chosen.
Figure 17 shows an example of a chart report form, where the number of records per data provider for a selected table in a selected time frame is chosen.

The result is a chart which displays the changes (if any) in the number of records in the database resulting from the harvesting processes for the chosen time interval, data provider(s) and/or data table(s). Examples can be seen at Figure 16 and Figure 18. By clicking on table names/data provider names, the number of records for that table/data provider is drawn or removed from the chart. Date range can be selected by dragging the sliders on the bottom of the screen. By hovering over the point on the chart, the date and the number of records is displayed.

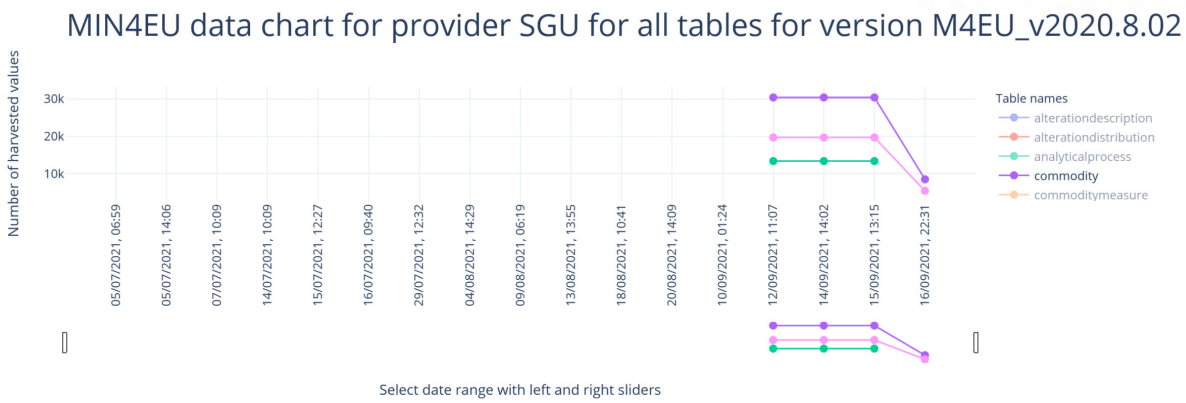

Figure 16. *Chart report per tables for one selected data provider and a selected time frame*

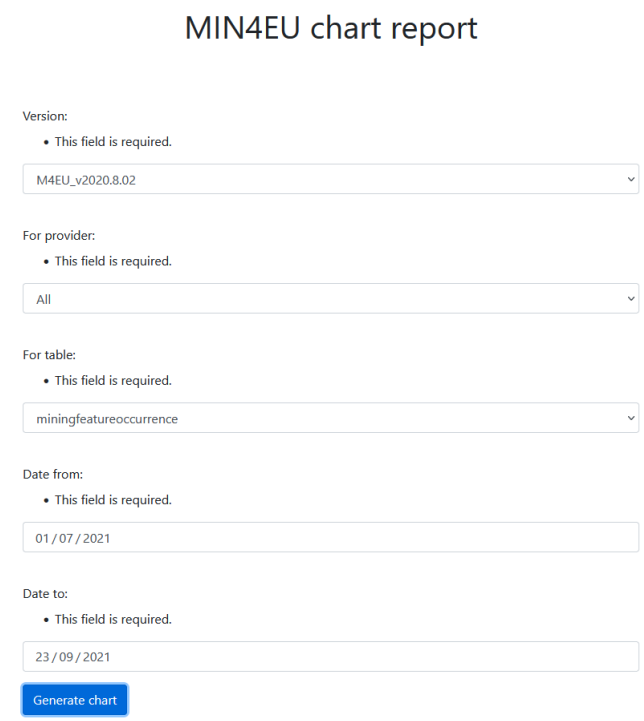

Figure 17. *A form for displaying chart report, example 2*

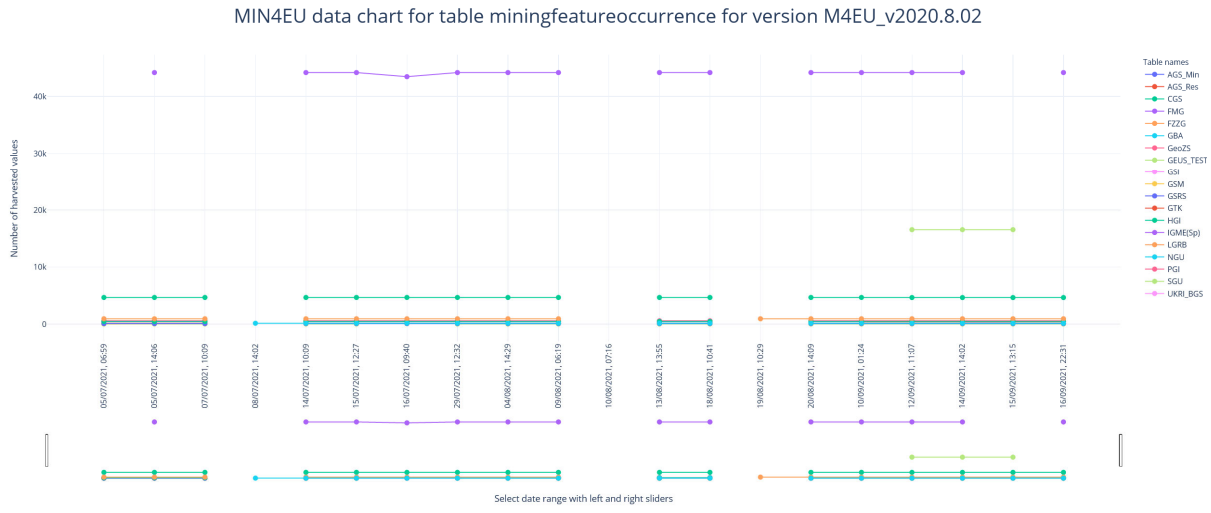

Figure 18. *Chart report per data providers for a selected table and a selected time frame*

2.5 MIN4EU providers DB info

The MIN4EU providers DB info page displays information about data providers local database and some information about the harvesting processes for that data provider.

First, the database version has to be chosen, as shown on Figure 19.

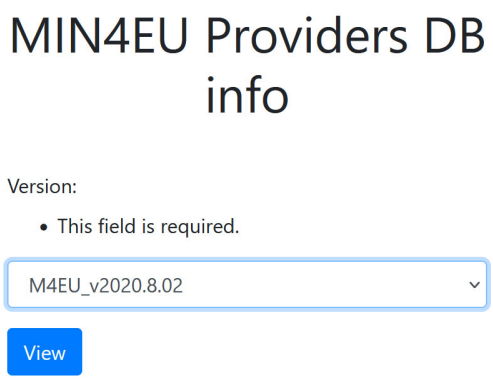

Figure 19. *A form for displaying additional information concerning the harvesting process*

For v. 1.1.2, there is information for each data provider regarding when the last harvesting attempt was triggered, and when the latest harvesting process was successful. Even though the harvesting as such was successful, it doesn't mean that there were any changes in data. That is why there is also information on the latest successful harvesting when the final updates of data were included. An example is shown on Figure 20, where for Albania there was a successful harvesting performed on 13th of September 2021, but the latest updates of data were made on 4th of August 2021.

Providers DB info checker for M4EU_v1.1.2 version

Country	Provider	Last try for harvesting	Last successful harvesting	Successful harvested with no data changes from
ALB	AGS	2021-09-13 13:20:40	2021-09-13 13:20:40	2021-08-04 12:12:21
AUT	GBA	2021-04-09 20:17:48	2021-04-09 20:17:48	2021-04-09 20:43:54
BEL	GSB-RBINS	2021-01-12 15:38:11	2021-01-12 15:38:11	2021-01-12 15:53:21

Figure 20. *Harvesting information for every data provider (v. 1.1.2)*

For v. 2020.8.02, additional information is available – which PostgreSQL the data provider has installed, which PostGIS version they use, which database and codelist version they uses if is data being updated (*in progress*), which is also denoted with a yellow highlight, or not (*data updated*), is the prosum:DBVersion* running and also if

other services are running (highlighted with red colour if they are not). An example is shown on Figure 21.

Providers DB info checker for M4EU_v2020.8.02 version

Country	Provider	PostgreSQL version	PostGIS version	Codelist version	DB version	Provider local database status	Last try for harvesting	Last successful harvesting	Successful harvested with no data changes from	Localdbcount service active
ALB	AGS_Min	10.10	2.4.4	2020.8.02	2020.8.02	DATA UPDATED 25-06-2021 17:44:44	2021-09-16 22:31:54	2021-09-16 22:31:54	2021-07-29 12:41:02	✓
ALB	AGS_Res	10.10	2.4.4	2020.8.02	2020.8.02	DATA UPDATED 15-07-2021 11:55:28	2021-09-16 22:31:54	2021-09-16 22:31:54	2021-08-18 10:44:58	✓
AUT	GBA	11.13	2.5.5*	2020.8.02	2020.8.02	DATA UPDATED 03-09-2021 15:41:47	2021-09-16 22:31:54	2021-09-16 22:31:54	2021-09-16 23:14:26	✓
BIH	FZZG	10.10	2.4.4	2020.8.02	2020.8.02	DATA UPDATED 25-06-2021 17:47:08	2021-09-16 22:31:54	2021-09-16 22:31:54	2021-08-20 14:14:56	✓
BIH	GSRS	10.10	2.4.4	2020.8.02	2020.8.02	DATA UPDATED 25-06-2021 17:48:00	2021-09-16 22:31:54	2021-09-16 22:31:54	2021-08-18 10:42:37	✓
CZE	CGS	9.2.8	2.0.4	2020.8.02	2020.8.02	DATA UPDATED 16-09-2021 10:34:13	2021-09-16 22:31:54	2021-09-16 22:31:54	2021-09-16 23:03:11	✓
DEU	LGRB	12.7	3.0.0	2020.8.02	2020.8.02	DATA UPDATED 09-07-2021 17:18:14	2021-09-16 22:31:54	2021-09-16 22:31:54	2021-07-29 12:44:24	✓
DNK	GEUS_TEST	12.4	3.0.2	2020.8.02	2020.8.02	DATA UPDATED 07-07-2021 15:58:23	2021-09-15 13:15:50	2021-09-15 13:15:50	2021-07-29 12:58:44	✓
ESP	IGME(Sp)	10.15	POSTGIS=	2020.8.02	2020.8.02	DATA UPDATED 29-06-2021 09:18:44	2021-09-16 22:31:54	2021-09-16 22:31:54	2021-09-12 21:25:57	✓
FIN	GTK	10.13	2.5.3	2020.8.02	2020.8.02	DATA UPDATED 03-09-2021 13:32:39	2021-09-16 22:31:54	2021-09-16 22:31:54	2021-09-10 02:31:47	✓
GBR	UKRI_BGS						2021-09-16 22:31:54	2021-09-16 22:31:54	2021-08-09 10:00:42	ⓘ
HRV	HGI	10.10	2.4.4	2020.8.02	2020.8.02	DATA UPDATED 15-07-2021 12:45:13	2021-09-16 22:31:54	2021-09-16 22:31:54	2021-08-18 10:43:58	✓
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	DATA UPDATED 15-09-2021	[illegible]	[illegible]	[illegible]	[illegible]

Figure 21. *Harvesting information for every data provider (v. 2020.8.02)*

* The prosum:DBVersion service contains the mentioned additional information for v. 2020.8.02 and also the information about performed database updates, using sql scripts from GEUS (https://geusgitlab.geus.dk/m4eu/2020-m4eu/-/tree/master/09%20Database%20scripts/02%20m4eu%20DB%20m4eu%20UPDATE%20from%20v2020.8.01%20to%20v2020.8.02), number of records in each data table, and also if XOR constraints of the database are satisfied or not.

2.6 MIN4EU Quality Control Application Viewer

The MIN4EU Quality Control Viewer, https://mintell4eu-qca.geo-zs.si/ , is an application allowing quality control of MIN4EU data through different spatial datasets, displaying harvested data on mineral occurrences and mines at commodity level (Figure 22).

The data display allows viewing spatial information (point or polygon data) and attributes describing them (geological, management zone, mining activity or transformation plant data). This tool enables data providers to check their harvested data on a map by e.g. the location of mineral occurrences and mines before they are published at EGDI and made available for public.

There is an option to filter data by country and/or selected commodities or search through corresponding attribute data, displayed in tables below the map and also available on click (Figure 23).

The displayed status of the data on the QCA Viewer corresponds to the latest reported data from data providers.

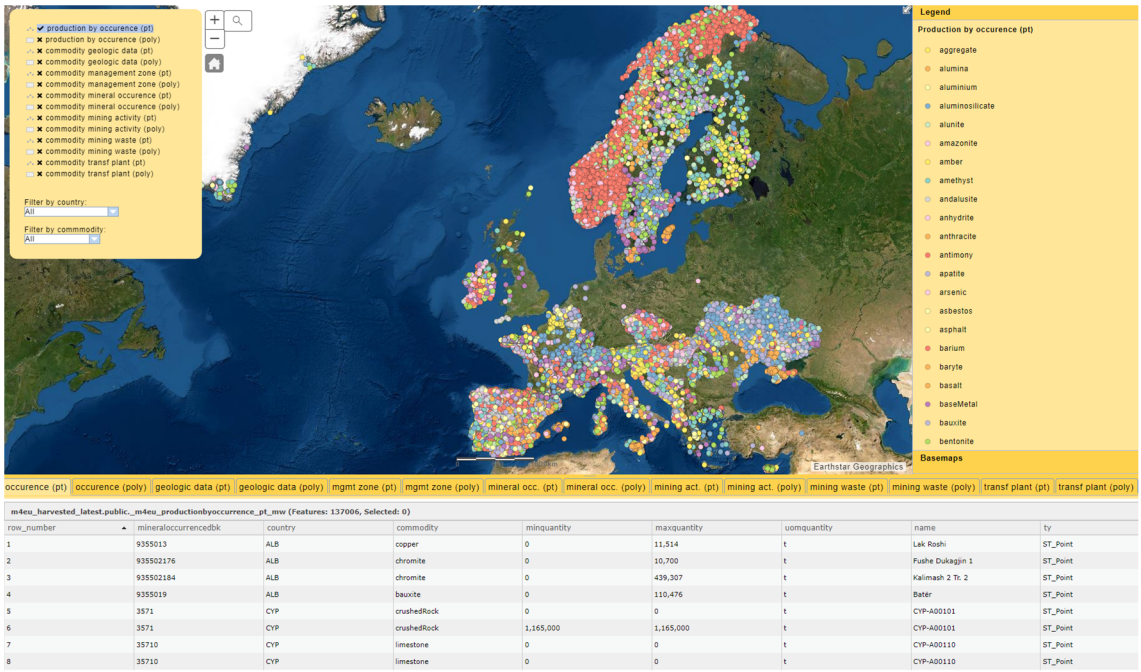

Figure 22. *MIN4EU Quality Control Application, showing the latest reported data from data providers.*

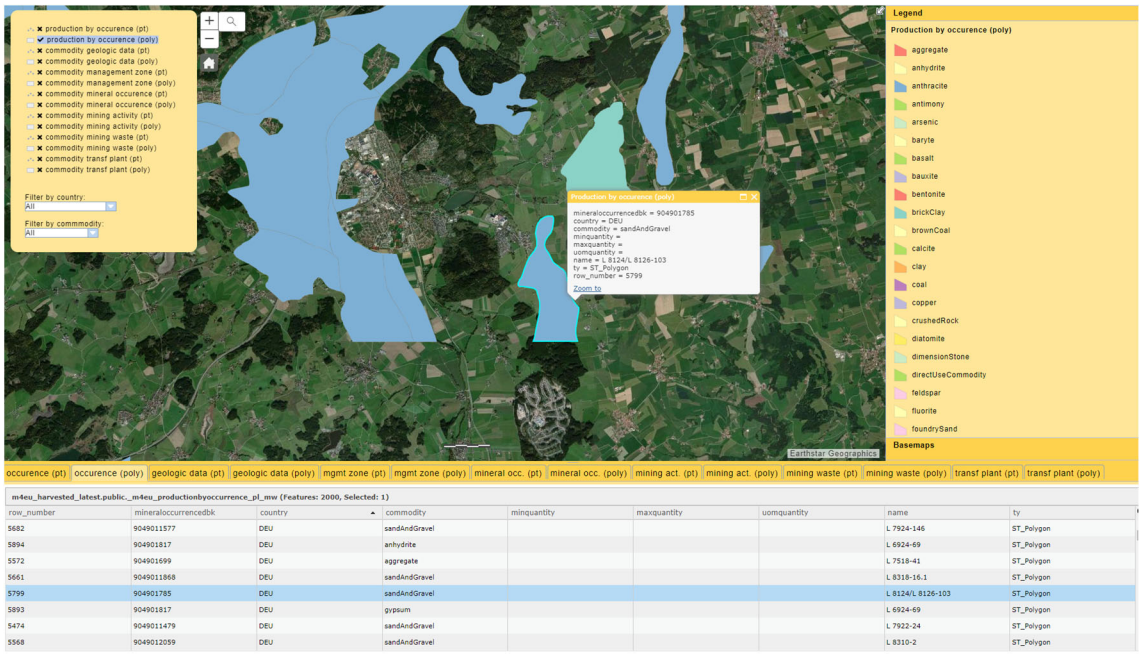

Figure 23. *QCA Application – an example of reported mineral occurrence areas by a data provider and the corresponding attribute data*

3 MAINTENANCE

GeoZS and GEUS are in direct contact with the relevant data providers. However, it is important to remember that the content of the database depends on the input from these data providers. The quality and content of national data sets transferred to MIN4EU is the sole responsibility of each individual data provider.

The routines of harvesting are automized, staying available and running also after the project end in October 2021. Support from GeoZS and GEUS will be given to existing and possible new data providers. If an existing data provider will update their national data, the data provider must send this information to the support e-mail harvesting@geo-zs.si, to notify GeoZS to run the harvesting. The EGDI will automatically be updated with these new data.

4 REFERENCES

Bahar; B., Bavdek, J., Šinigoj, M., Vihtelič, A., Pjetursson, B., Schjøth, F., Thyregod, J., Hansen, M., Søgaard Rasmussen, V., Abdelfettah, F., Jean-Baptiste, R., Hegen, D., Huisman, E., Sohier, P., Čápová D., Kramolišová, P., Prieto Martín, A., Sánchez Molinero, H. GeoERA, GIP-P project. Final version of Central database / harvesting. Deliverable D7.3, Public Document.
Link:https://geoera.eu/wp-content/uploads/2021/04/D.7.3-Central-database-harvesting.pdf

Šinigoj, J., Vihtelič, A., Kumelj, Š. GeoERA, Raw Materials Theme. MINTELL4EU project. Technical guidelines – status after 1st year. Deliverable D3.2, Public Document. Link: https://geoera.eu/wp-content/uploads/2019/11/D3.2-Technical-Guidelines.pdf