Ref. Ares(2017)3982358 - 09/08/2017

Deliverable D3.2

Final inventory of data on raw materials

Project:	Mineral Intelligence Capacity Analysis
Acronym:	MICA
Grant Agreement:	689468
Funding Scheme:	Horizon 2020
Webpage:	www.mica-project.eu
Work Package:	Work Package 3
Work Package Leader:	Evi Petavratzi
Deliverable Title:	Final inventory of data on raw materials
Deliverable Number:	D3.2
Deliverable Leader:	NERC (BGS)
Involved beneficiaries:	GEUS, Fraunhofer ISI, UL-CML, Minpol, BRGM, BGR, GTK, NTNU; Linked Third Parties: Swisstopo, GSI, GIR, LNEG
Dissemination level:	PU Public
Version:	Final
Status:	Submitted
Year:	2017
Authors:	Evi Petavratzi, Teresa Brown
Reviewed by:	Lauran van Oers, Murray Lark, Erika Machacek, Kisser Thorsøe
Approved by:	Erika Machacek

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 689648.

Notice

TABLE OF CONTENTS

PURPOSE

Deliverable 3.2 provides a summary report on the final inventory of data on mineral raw materials developed by WP3. This is a follow-up report from D3.1 *Draft inventory of data on raw materials* and as such it provides an overview of the progress made since the previous report. D3.1 and D3.2 should be read in combination to fully understand the development process of the metadata inventory undertaken in MICA (WP3).

EXECUTIVE SUMMARY

Deliverable 3.2 presents the final metadata inventory developed by WP3. D3.2 is a follow-up report from D3.1 and presents the progress made since the last report to finalise the MICA metadata inventory. This report does not focus only on data gathering and the development of metadata records, but also discusses the outcomes of the efforts undertaken to map the identified data to the methods (which were defined by WP4), summarises the results of two workshops undertaken on data uncertainty, provides an update of the progress made towards the online inventory and summarises the decisions and work progress made for linking the MICA metadata inventory to the RMICP (the MICA online platform). Key findings include:

- Data identification and gathering: A total of 408 records now form part of the MICA metadata inventory. During this second stage of data gathering several MICA partners have actively participated in the data collation process, provided their insight and assisted with the development of metadata records.
- Development of metadata records: Out of the 408 records identified, 370 metadata records have been fully prepared and the remaining 38 are in the pipeline and will be delivered shortly.
- Mapping data to methods: Identified datasets (selected from the WP3 dataset list) were mapped to methods (identified in WP4 and described in D4.1) with a group of experts from the MICA consortium. This proved to be a useful exercise as it identified data gaps that either could potentially be filled by expanding the current WP3 inventory list or where datasets do not exist and therefore future research is required to address them. It also identified gaps in the list of methods that may require further development by research or consideration for inclusion in WP4. Some of the findings and recommendations of this workshop will be discussed in more detail in the forthcoming deliverable D3.4 *Raw material data strategy*.
- Data uncertainty: The workshops provided insight on concerns about data uncertainty and how these are dealt by data users and data providers. It also provided additional information about the role and connections between metadata and data uncertainty and how uncertainty could be communicated more clearly to users.
- Online metadata inventory: The online metadata MICA inventory is well on its way. A beta-version has become publicly available, which contains a first batch of 101 records (http://metadata.bgs.ac.uk/mica/srv/eng/catalog.search#/home). All prepared records have now been transposed and are ready to be put online. Some further checks of the new metadata.mica-project.eu domain are currently under way, including testing the use of a new version of GeoNetwork, which will be used for the release of the next version of the online MICA metadata inventory.

- Linking the MICA metadata inventory with the RMICP (WP6): Several discussions have been held over the past few months on how we could effectively link the two systems. A procedure has been agreed whereby metadata records will be harvested by the RMICP for inclusion in the MICA online platform.

It is anticipated that as the project reaches its final stages and several more FactSheets, DocSheets and FlowSheets become available, some additional datasets may be identified and added to the inventory. Therefore the final list of records will not be available until the end of the project.

DELIVERABLE REPORT

1. Introduction

Work Package 3 (WP3) is concerned with the identification of relevant data on raw materials that could be used to satisfy the needs of various stakeholders, whether directly or through methods and tools (WP4) that may require such data. These datasets provide the evidence, information and knowledge required by stakeholders and respond to the identified stakeholder needs (WP2). Metadata records that describe the content, status and relevance of the identified datasets have been developed and form the final MICA metadata inventory.

The current report (D3.2) comprises a continuation of D3.1 and describes the work undertaken in Task 3.1 *Review of data availability, use and uncertainty* and Task 3.2 *Mapping data to tools and methods*. The following sections cover the subjects below:

- Identification of data and gathering of metadata information
- Description of the MICA metadata inventory
- Mapping of data to methods
- Exploring data uncertainty
- Links between the MICA metadata inventory and the RMICP (MICA online platform)

2. Identification of data and gathering of metadata information

The gathering of metadata is a core function of WP3 and the identification of relevant datasets was initiated from the start of this project. Deliverable 3.1 *Draft inventory of data on raw materials* presented the preliminary metadata gathering results, where 180 records were identified covering the whole range of the MICA Domains and Concepts. Since then, the hierarchy of the MICA ontology has been revised leading to seven instead of eight domains and the data identification and metadata collation process has progressed significantly resulting in 408 datasets identified by WP3.

The list of identified datasets is provided in the embedded spreadsheet in File 1.

WP3 Simplified list of records_Combine

File 1 Combined list of the identified MICA metadata records produced by July 2017. Access is provided to the list by clicking the above icon.

The data identification and gathering process is described in Figure 1. 'Known'[1] and identified data sources, for instance from desk-based research, were recorded in a spreadsheet and metadata records using the MICA metadata record template were generated. Then checks on the record and the validity of the information provided in the template form were undertaken by NERC (BGS), who were also responsible for transposing the information for inclusion in the MICA online metadata inventory using GeoNetwork software.

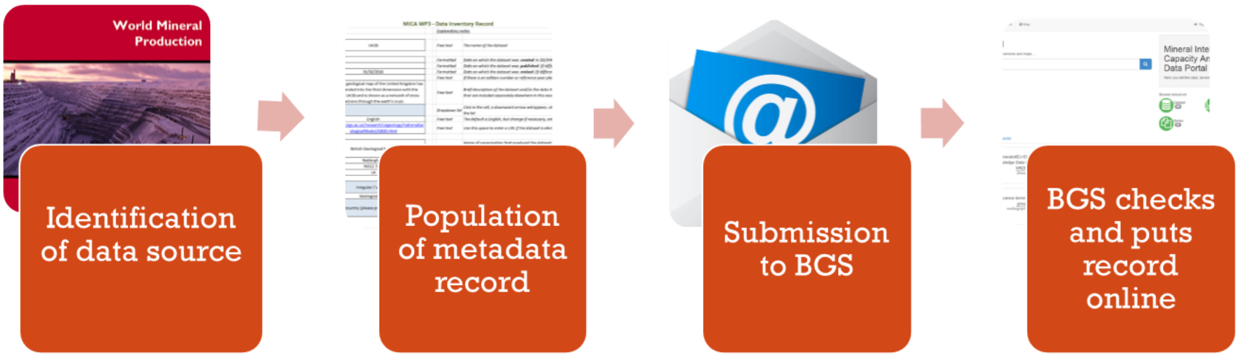

Figure 1 MICA data identification and gathering procedure.

File 1 already comprises an inventory of data sources with standard descriptors, such as the title of a datasets, the MICA Domain it comes under and a web link being included. In the second part of the data identification process, several MICA partners were involved and assisted in the development of the MICA metadata inventory; see Figure 2.

[1] Known data sources describe those produced by MICA partners, for example geological survey spatial databases.

Figure 2 MICA partners and linked third parties that assisted with the data collation process.

The content of the MICA metadata inventory varies considerably. It includes datasets, but also scientific articles, reports, websites (e.g. for trade associations), maps, project information, information about relevant legal documents and many more (Figure 3). Although the aim of WP3 is primarily to provide datasets, in many cases useful data are not found in existing databases and portals, but in reports and other contextual documents. Therefore, the decision made was to include such documents in the inventory. Substantial effort however has been put into minimising the volume of 'irrelevant' data sources, namely data sources that are only marginally relevant to the scope of this inventory. A thorough review of the inventory was undertaken and several records that were identified earlier in the project were removed, as they were found to be marginally relevant. At the same time a quality check was undertaken. Duplicated records and superseded data sources were removed and mistakes, omissions, inactive links and so on were corrected. This quality check process is ongoing and will continue until the end of the project.

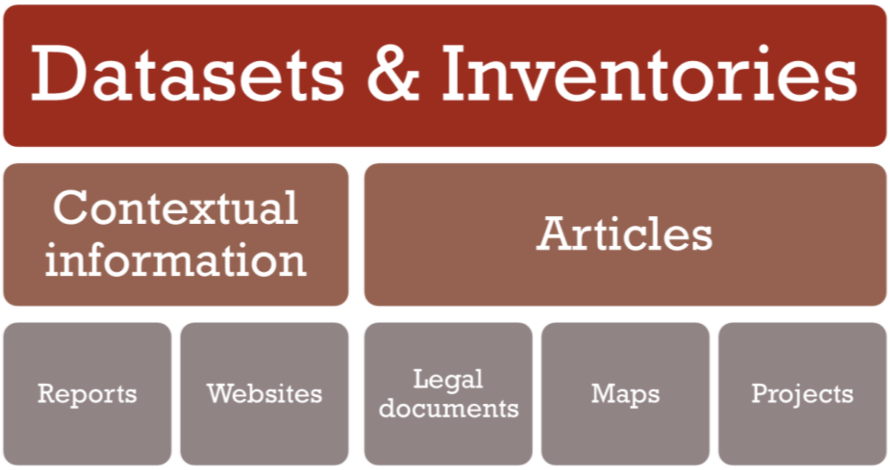

Figure 3 Content of the MICA metadata inventory.

The identified data sources cover a whole range of different themes or in MICA terminology Domains; see Figure 4. The RMICP ontology includes seven Domains and up to 4 levels of detailed concepts in a hierarchical order. Data sources identified cover all the Domains used by RMICP and the majority, if not all, of the concepts too. It is difficult at this stage to quantify whether datasets for all concepts are available, as the links between some records and Domains/Concepts are currently under development. Also, it is expected that additional metadata records will be produced until the end of the project as a consequence of the development of additional FactSheets and DocSheets.

Figure 4 MICA Domains.

3. Development of metadata records

3.1 MICA metadata template

During the first year of this project, substantial effort was put towards the development of a metadata record template that uses the principles of the standard ISO 19115-1. Extensive presentation and discussion of the work undertaken during the metadata template development was included in D3.1. In a nutshell, the template consists of five different sections and requests a variety of information to be recorded against the fields included in these sections that can be seen in Figure 5.

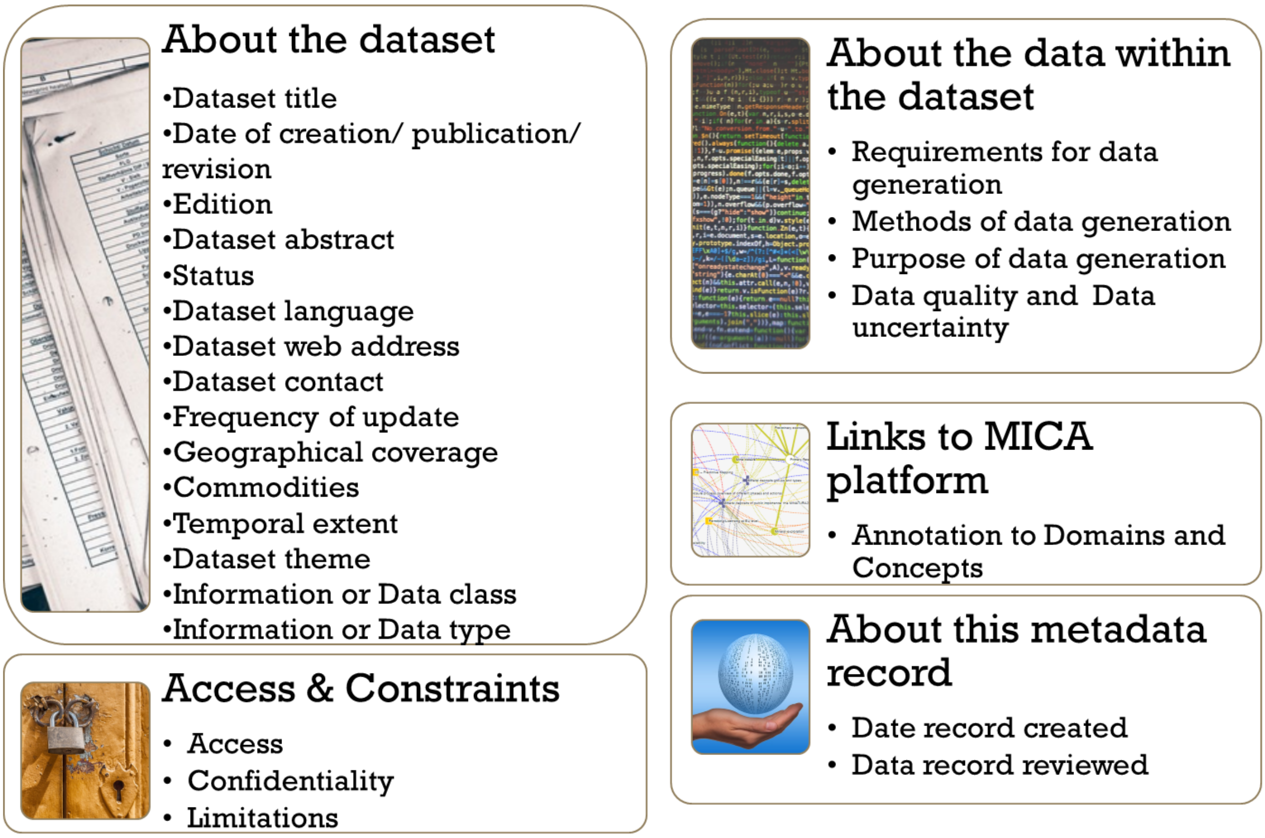

Figure 5 The MICA metadata template.

3.2 Metadata records

The datasets identified by MICA partners were developed into metadata records. Out of the 408 data sources identified, approximately 370 have been turned into metadata records with the remaining 38 being currently under development and due to be completed shortly.

A few examples of completed metadata records in the excel template are shown in File 2.

Minerals4EU Map Viewer.xlsx | CML_Life Cycle Inventory database | Croatian Bureau of Statistics - Industria

File 2 Examples of completed metadata records.

3.3 MICA online metadata inventory – development

The online metadata inventory is well on its way and a first batch of 101 records is publicly available and can be found at: http://metadata.bgs.ac.uk/mica/srv/eng/catalog.search#/home. This is a temporary link though, as the MICA metadata inventory and the RMICP will soon move into a '.eu' domain, under the MICA name. The new address for the MICA metadata inventory will be: http://metadata.mica-project.eu/. The establishment of this new '.eu' domain has happened only recently and therefore the move to it has not happened yet, but will take place shortly.

The move towards the online inventory required several additional development steps to be undertaken and issues to be resolved, such as mapping our terms and fields to specific vocabularies to ensure compliance with the INSPIRE directive and ISO 19115-1, the development of programming code and a procedure to transfer the spreadsheet records online, testing and alignment with the structure of the MICA ontology. Most of the work undertaken in the past few months covered the above activities and most of the issues faced during this period are now resolved. In detail:

- Mapping of terms and fields to ensure compliance with INSPIRE and ISO 19115-1: this included using a variety of vocabularies online for defining the terms and fields set in the metadata template.
- Development of programming code and a procedure to transfer spreadsheet records online: Spreadsheet records had to be transposed, requiring a methodology to be developed in excel, and then uploaded online using a programming code that was written to automate the transfer and ensure it was carried out in a consistent way.
- Testing to ensure that all information have been transposed online successfully.
- Alignment with the structure of the MICA ontology: MICA vocabularies were used for several of the fields in the metadata template; these are described in more detail in the section 4.

There are still tasks ongoing, but which are well progressed. They are listed below in a priority order:

1. Development of programming code and method to transpose online the annotations to Domains and Concepts for individual records.
2. Determine the final version of GeoNetwork to be used for the MICA metadata inventory.
3. Improve the visual presentation of records online and provide guidance to users about switching between the 'default' and 'full view' to ensure they have access to the whole record.
4. Make available a 'blank' metadata template for generation of records online in the future.

3.4 MICA online metadata inventory – Presentation of records

The MICA online metadata inventory introductory webpage is shown in Figure 6. The introductory webpage allows for users to search through records using the 'search box' or to look at the latest/most popular records available listed underneath the search box. It is also possible to browse through only datasets, non-geographic datasets or series, by clicking the icons highlighted in the yellow box of Figure 6. Once one of the resources (dataset, series or non-geographic dataset) is clicked, or a keyword is entered in the 'search box', then the 'search results' page appears, which includes some of the records identified by the system and several additional filters as shown in Figure 7.

When clicking on a record, either from the introductory page or the 'search results' page, then a summary record becomes available to a user that includes some of the key fields populated by WP3. This is a 'default view' provided by GeoNetwork and only small modifications to the appearance of the fields can be undertaken (Figure 8). The 'full view' of the record is available by clicking on the 'eye' icon at the right top corner and selecting 'full view' (Figure 9). The 'full view' includes the whole spectrum of information and fields developed by WP3 (Figure 10). Registered users (for example, WP3 MICA partners) will also be able to sign in and add new records or amend the content of existing one (red box in Figure 6). This functionality is not active at the moment, but it will become available in the near future.

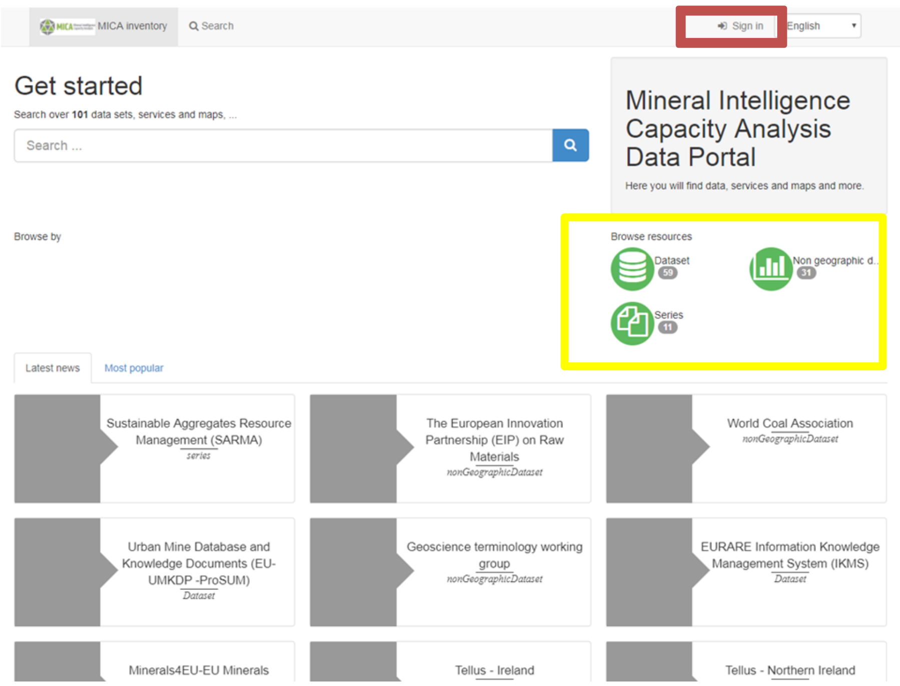

Figure 6 MICA metadata online inventory – introductory page.

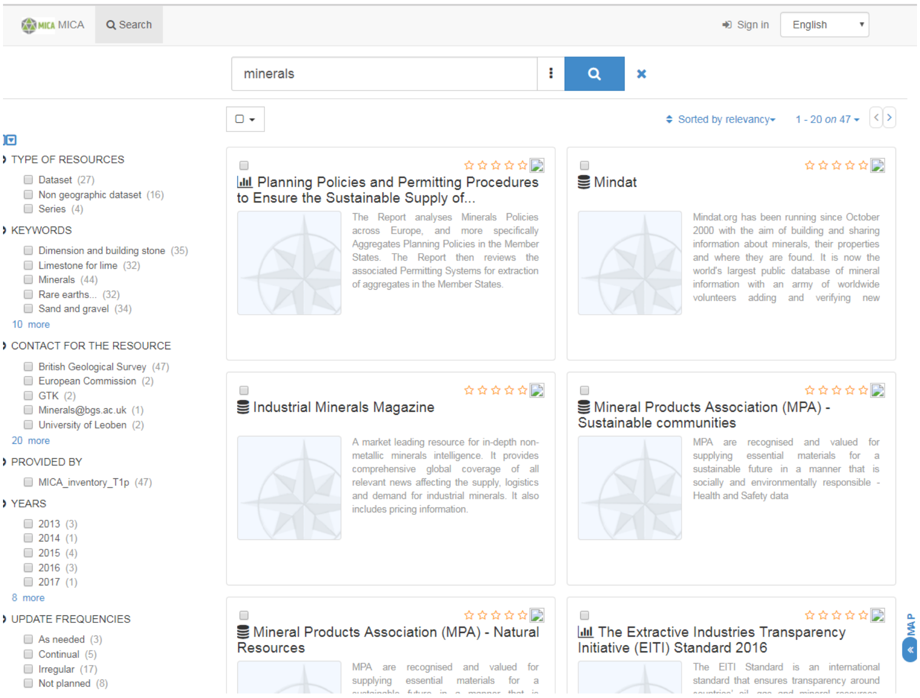

Figure 7 MICA online metadata inventory 'Search results' page and additional filters.

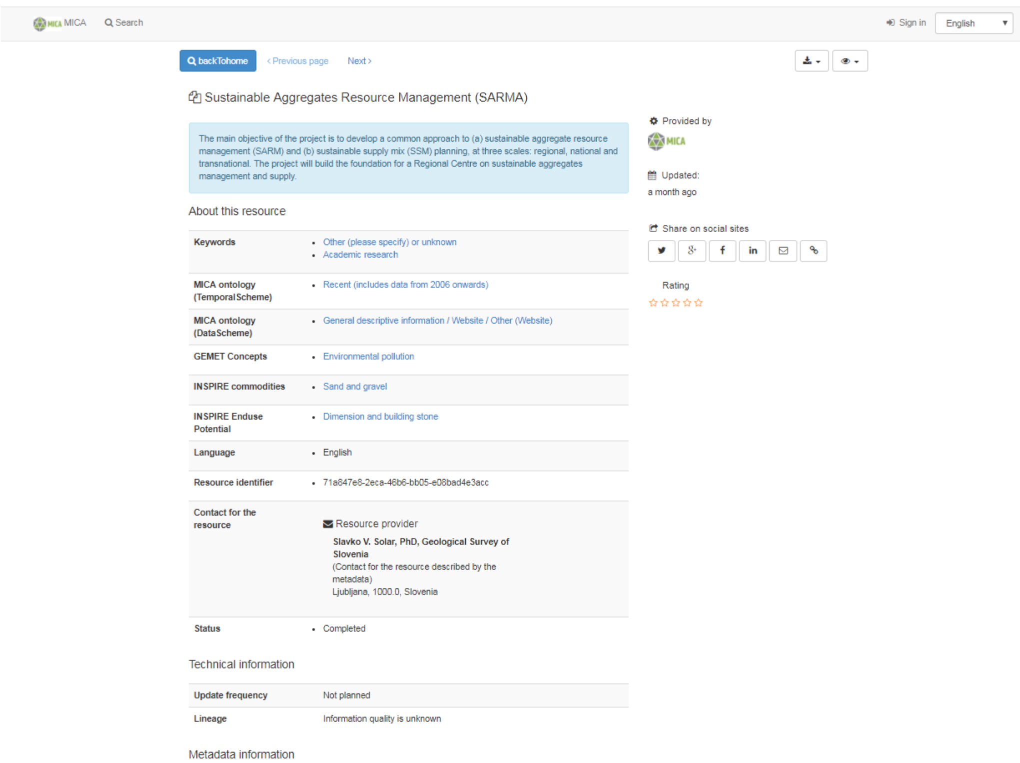

Figure 8 'Default view' of a metadata record presenting summary information only.

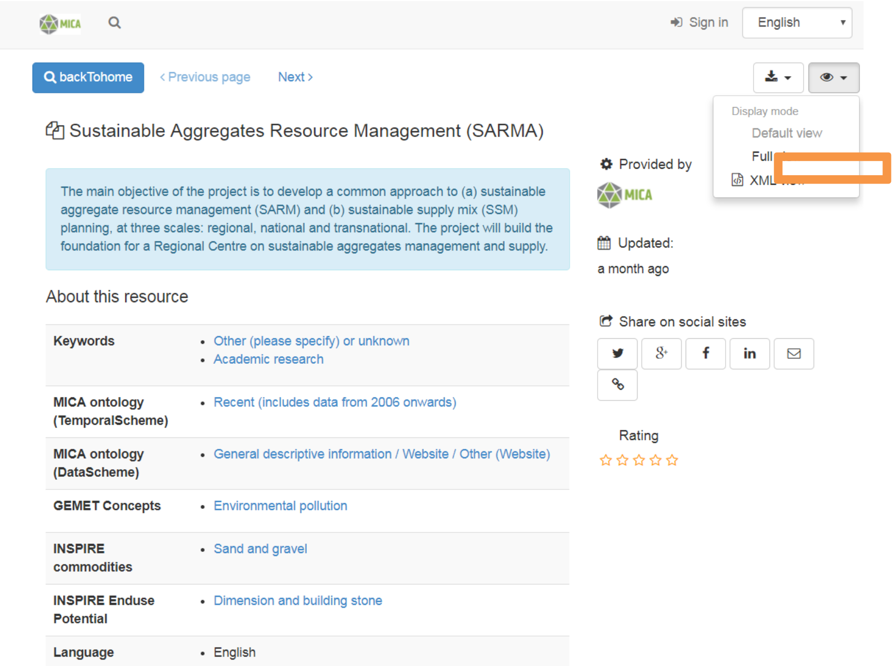

Figure 9 Route to selecting the 'Full view' of a record.

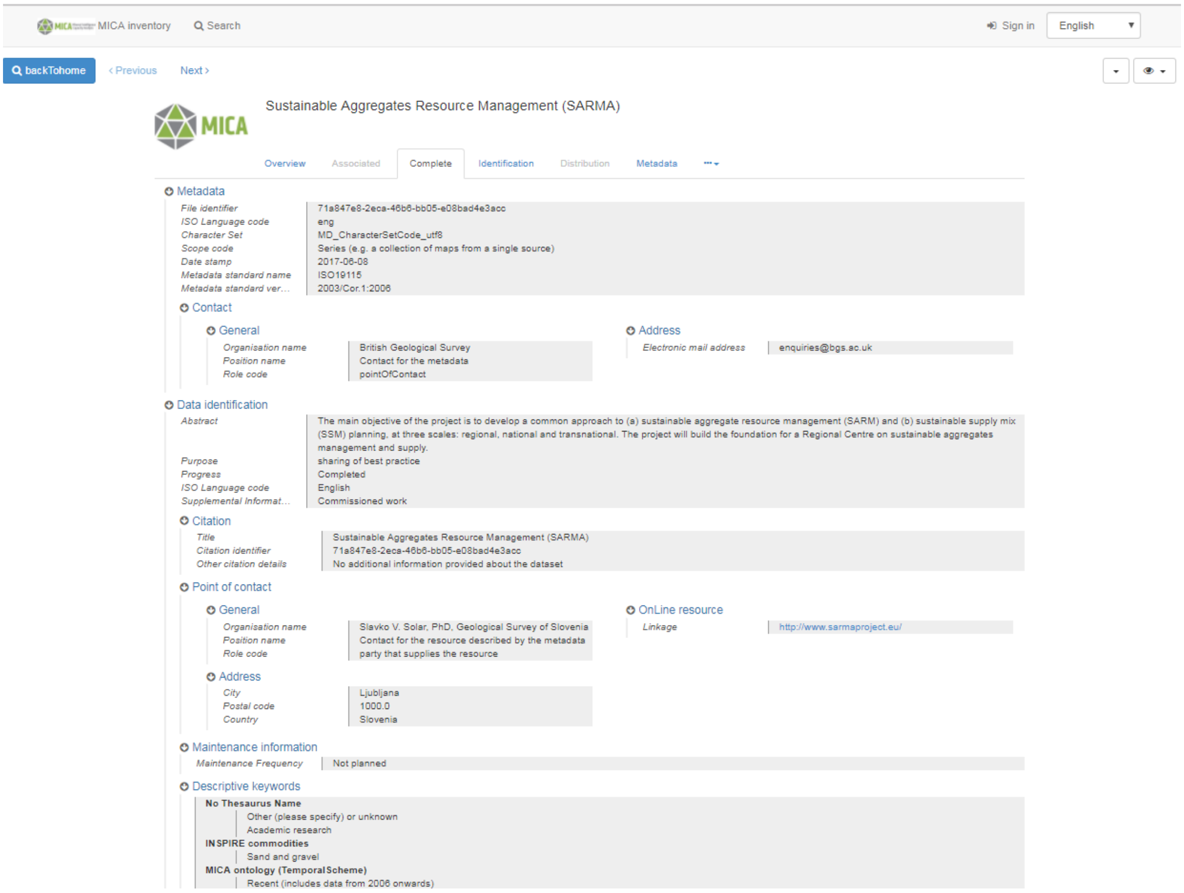

Figure 10 The 'full view' of the metadata record.

4. Linking the MICA metadata inventory to the RMICP (MICA online platform)

Many of the fields included in the metadata records utilise the MICA vocabularies and terms of the MICA ontologies; see Figure 11. This enables the RMICP (being developed by WP6) to provide users with access to relevant data when the linked fields are selected. Since the last WP3 deliverable report (D3.1), the MICA ontologies were finalised and WP3 was able to finalise the terms used in the relevant fields of the metadata template. For all the records developed, the links to the MICA Temporal Scheme, the Data Scheme and the Commodity Scheme have been completed. The links to the Domain Scheme has also been completed for the majority of the records, with a limited number remaining to be completed by September 2017. The links to the MICA Schemes are presented along with other fields in a metadata record.

The way the MICA online metadata inventory and the RMICP are going to be connected together is shown in Figure 12. Metadata records will be harvested by the RMICP and included in the Triple Store, which in turn will make them available to the user through the MICA visualisation application.

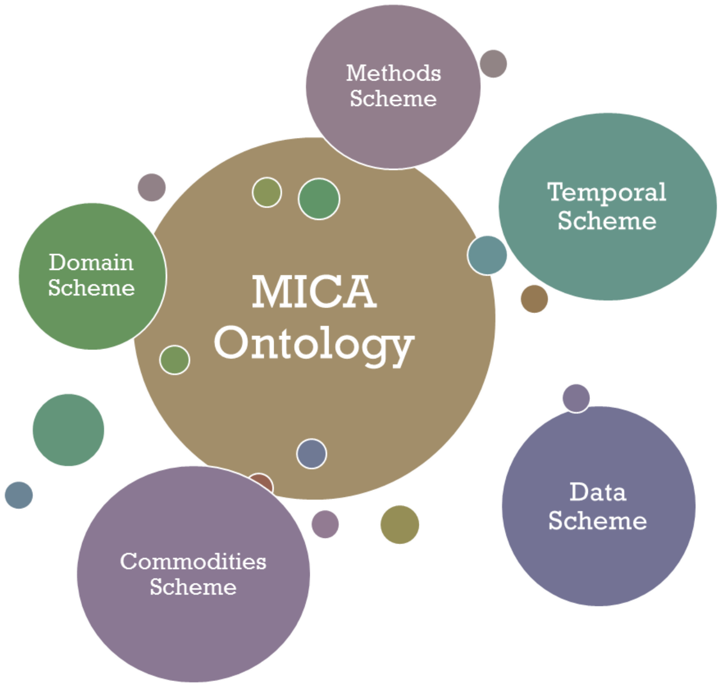

Figure 11 The MICA ontology Schemes represent fields of the metadata records completed to ensure compatibility with the RMICP.

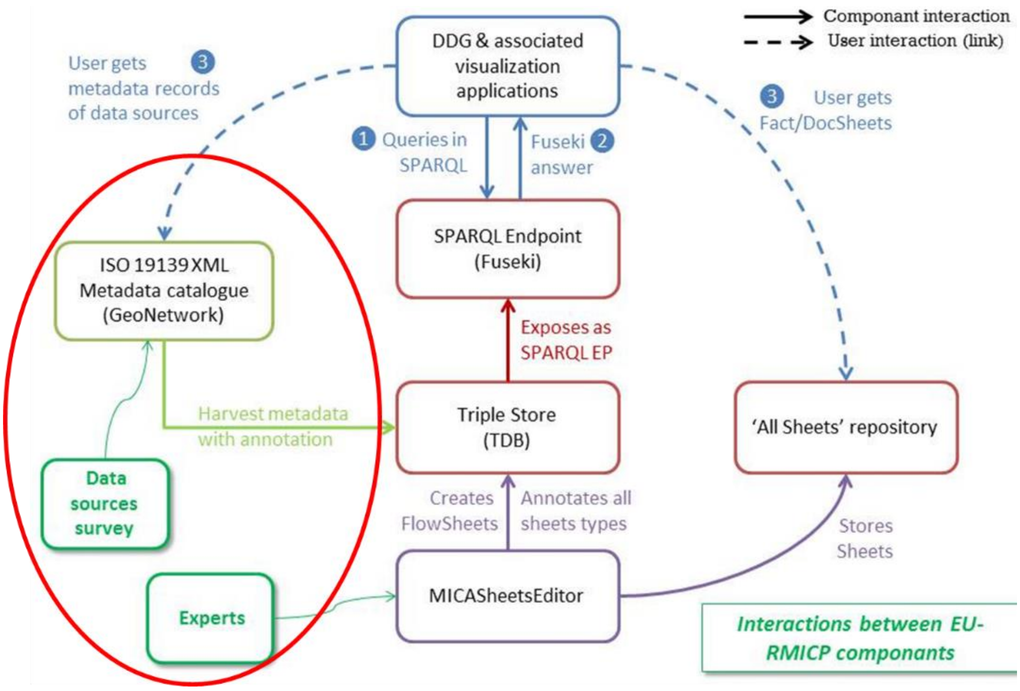

Figure 12 Interactions between the RMICP components. The MICA metadata catalogue links are presented in the red oval area.

5. Mapping data to methods

Methods, models and tools all require data to operate and produce results. Work Package 4 in MICA has identified several methods related to different aspects of raw material studies and Task 3.2 *Mapping data to tools and methods* in WP3 aims to identify the links between the identified methods and the datasets in the MICA metadata inventory. This will:

1. Validate the completeness of the MICA metadata inventory;
2. identify data gaps, e.g. missing datasets from the inventory or genius data gaps;
3. identify methods missing, e.g. for generating data or for interpreting data towards a specific objective.

Two actions for mapping data to methods have taken place. The first one was at a WP3 internal workshop in the MICA Paris progress meeting, where partners from WP3 and WP4 worked in groups to map datasets to methods. This has resulted in a matrix 'data2methods' in which methods and relevant metadata records are mapped, either as input or output (Appendix 1, page 32). The result of this action has also been used as guidance for the second action.

The second action of mapping data to methods is undertaken during the metadata record development process. Methods represent one of the MICA ontologies (Methods Scheme) and records are required to link to the appropriate method, so that information not just on data, but also on methods are presented to users of the RMICP. The mapping of data to methods is not done for every records of the MICA metadata inventory, as several of them represent contextual information, reports, articles and websites rather than actual datasets. Only datasets and databases are mapped to methods.

5.1 WP3 technical workshop – mapping data to methods

The records selected from the MICA metadata inventory for the mapping exercise are shown in (Appendix 1, page 32). Datasets were grouped in 13 data types as shown in Table 1. Project participants were split into three groups and each group was given the task to map the datasets to the methods identified by WP4 (Table 2). The outcomes of this workshop are summarised in the subsequent sections and the matrix produced during the workshop is available in (Appendix 1, page 32).

The completeness of the MICA metadata inventory

Mapping between the selected datasets and methods was possible in all cases, which suggests that there is no method which does not require one/many datasets identified by WP3. This does not imply that data gaps do not exist, but that the identified methods are supported by the datasets found in the inventory.

Most datasets present a relationship with several methods, however their status may change. For instance, commodity statistics produced from trade associations is often the output of their attempt to quantify supply and demand patterns of a commodity. These data are then used as input to methods that assess society's metabolism to develop for example material flow analysis models, or in methods that assess the economic aspects of resource use. Similarly, many of the geoportals available from geological surveys represent output data, but such data may be used as input in resource estimation or risk assessment models. Methods and data therefore are interlinked in many different ways forming a network, where output data generated from a method are transformed into input data in another method. The mapping exercise has identified that the MICA inventory includes both 'input' and 'output' data. Also, the methods included in the MICA project reflect both methods that generate data (e.g. geological mapping), but also methods that utilise data to develop models and additional subsequent data.

Table 1 Data types used in the mapping exercise. The detailed datasets are shown in Appendix 1, page 32.

Group	Data type descriptor
1	Commodity sector data on production, trade, (reserves and emissions)
2	Commodity country data (profiles, incl. production, trade, (reserves and emissions) etc.
3	Commodity world data (profiles, incl. production, trade, applications, reserves etc.)
4	Continent or world data on production and trade of total economy (incl. emissions)
5	Footprint data of consumption by total economy
6a	Spatial – geological data
6b	Spatial – monitoring state indicators (concentrations, stocks)
7	Emission registration, total economy, breakdown in sectors
8	Waste and recycling data
9a	LCA Impact assessment factors of emissions
9b	LCA process data, production and emission of process
10	Risks and safety
11	Chemicals and properties
12	Market and prices
13	Statistics at Country level

Table 2 Methods used in the mapping exercise.

Method category	Method
Methods to identify and assess geological and anthropogenic (urban) stocks	Geological mapping
	Remote sensing
	Geochemical analysis
	Ground investigation
	Resource estimation
	Measuring input and output of industrial processes
Methods to assess society's metabolism and its environmental impacts	Economy wide material flow accounting
	Substance / Material flow modelling
	Life cycle assessment
	Environmentally extended Input Output Analysis
	Risk Assessment e.g. as in REACH
	Foot printing
Methods to assess the economic aspects of the use of resources	Cost Benefit Analysis
	Life cycle costing
	Input output analysis
	Criticality assessment
	Econometrics
	Computable Equilibrium Modelling

Depending on the research question in place, the use of a method may not always be essential, as data could provide the essential background information required. For instance, if the objective of a project or question is to understand the market of a commodity, then statistical data from public authorities (e.g. statistic offices, geological surveys) are sufficient to provide such insight. In that instance, primary data (statistics) are transformed into information (data in context) or knowledge (following expert analysis) without necessarily having to use a model or method.

Some additional remarks were recorded during the workshop. These are summarised below:

- The MICA metadata inventory includes records that are not just datasets. Although, they are considered important, as they often hold information or snapshot data that are not available from other sources, they may get quickly outdated. This may be a problem in the future and for the sustainability of this inventory, as it may not contain the latest information. However, at present it is considered important that they are highlighted to users of the MICA online platform as the only relevant data available.
- The question whether specific articles should be replaced by links to journals that often publish articles in related topics was raised. Due to time limitations, this will have to be a recommendation to be taken further by any future updates of the MICA metadata inventory.

Data gaps in the MICA metadata inventory

Data gaps identified during this workshop fall into two categories:

- Missing datasets that could be included in the MICA metadata inventory (Table 3), and
- Data gaps, namely datasets that do not exist, but their development is important (Table 4).

Table 3 Missing datasets from the MICA metadata inventory.

Missing datasets	Action to be taken
Coverage of all important trade associations	Several are already included, but some additional associations have been identified and will be added to the inventory list. For example, IMA (Industrial Minerals Association), Eurofer, European Aluminium and so on. Trade associations that provide datasets will be the primary focus of this exercise.
National Statistics Offices to cover the EU28	Several are already in the inventory, but some missing. These have been identified and will be added.
Private databases e.g. Granta MI database, or the S&P (Metals and Mining data)	Some are already part of the inventory, but some more will be added. The MICA metadata inventory aim is to cover primarily public sources, rather than private data, therefore only a small number of highly relevant data from private sources will be added.
World governance indicators (World Bank)	This will be added to the inventory list.

Table 4 Data gaps identified during the mapping data to methods workshop.

Data gap	Description
Waste and recycling data	Data currently available are considered poor and incomplete. They cannot be incorporated into methods without making additional assumptions, which affect subsequent analyses and models developed.
World Emission Registration data	These are not available for every country. In order to calculate normalisation factors that find use in LCA models, emission registration data from some countries are used and extrapolated to a World-wide level. This is an important gap and can introduce significant uncertainties in LCA models.
Urban stock data (built up or accumulation)	These are not currently available, but are essential for assessing the urban environment and for quantifying resources that may become available in the future.
Composition of goods (e.g. metal content of ores, materials, components, products, waste)	Essential data used by several methods but are currently only partially available from various dispersed sources.
Data on dissipative losses	The European Pollutant Release and Transfer Register (E-PRTR) reports emissions from installations. A start has been made to also include dissipative emissions e.g. from livestock and fertilizer use etc.). Dissipative losses are required by several methods. Essential in quantifying environmental impacts and undertaking mass balance exercises.
Data on the lifetime of goods	Data are only partially available from various dispersed sources. They are very important when quantifying the resource potential from secondary resources.
Data on the production of secondary raw materials	A comprehensive dataset that addresses several commodities is missing. Partially available for selected commodities alongside the mineral statistics.
Monitoring concentrations in soils	Partial data may be available for some countries, but often are deemed of insufficient detail or/and are not updated frequently.
Social factors and policy related data	These are often available from reports rather than databases. They may be partially present in national statistics for some countries. Overall comprehensive and standardised datasets are missing.
Mining waste data	Some may be available through national statistics or public authorities, but there is no comprehensive dataset that holds such data at EU/ World level. Again an important dataset used for assessing society's metabolism and environmental impacts.

Additional recommendations and actions regarding the identified data gaps will be included in D3.4 *Raw material data strategy*.

Gaps in methods

The mapping exercise also identified some gaps in methods that have not been included in the current list. These include:

- Mapping urban stocks: there is no specific method for developing such models, but the research community has been exploring this subject (see MICA case study produced under Task 4.2). The use of 4D-GIS data at urban scale is one of the approaches followed to map urban stocks. Good data at urban scale are required to apply this approach, which are often missing.
- Building Information Modelling: This method has not been included in the MICA project, but could be relevant especially in assessments of the urban environment and stocks.

6. Exploring data uncertainty

Two focused workshops have been held by NERC (BGS) in an attempt to gain a better understanding of the perspective of 'data users' and 'data providers' on the uncertainties contained in selected datasets. The first workshop explored data uncertainty related to geochemical data and the second workshop was tailored towards mineral statistics.

The objective of this exercise was to gain insight into the requirements of data users in the minerals sector with respect to the uncertainty in the data. In particular the workshops identified:

- whether data users were concerned about uncertainty;
- how uncertainty is currently managed;
- what the implications of data uncertainty are; and
- which new approaches might be useful.

This will allow us to make recommendations about how uncertainties in new data might be managed, how to communicate uncertainty effectively to end users and to identify what metadata information should be recorded about datasets if their value for the minerals sector is to be assessed.

The questionnaires used during the two workshops including the responses received from participants are included in Appendix 2 (page 36) and Appendix 3 (page 37). The following two sections discuss in more detail the two workshops individually with section 6.3 combining the findings and developing some preliminary recommendations. It is anticipated that key findings and recommendations from these two workshops will also be included in D3.4 *Raw material data strategy*.

6.1 Workshop I on the implications of uncertainty in geochemical data for applications connected to mineral resources

The dataset investigated and details about the workshop participants are described in Table 5. The questionnaire consisted of two different sections, one directly relevant to 'data users' and a second to 'data providers', but all participants engaged in discussion throughout. The questionnaire provided a general outline for discussion, but several additional points were raised and relevant ones were recorded. A summary of the questions presented to data users is provided in Table 6.

Table 5 Dataset and workshop I participants.

Dataset	Geochemical Baseline Survey of the Environment (GBASE) undertaken in Great Britain by BGS and the Tellus Survey undertaken in Northern Ireland by the Geological Survey of Northern Ireland and BGS
Data providers	Dr Louise Ander and Mr Bob Lister Geochemists at BGS with experience in the collection, management and analysis of geochemical data
Data users	Paul Lusty: Team Leader of the Ore Deposits and Commodities Team at BGS and Eimear Deady: Mineral Resource Geologist
Workshop moderators	Dr Murray Lark: Environmental Statistician and Dr Evi Petavratzi: Senior Mineral Commodity Geologist

Table 6 Workshop I in Geochemical data – Questions used in discussion.

Data user questions
What are the questions you use GBASE data to answer?
In what form do you use the data? i. Point values (e.g. one or a few values from close to a site of interest) ii. Interpolated point values iii. Sub-regional mean values iv. The 'general spatial pattern'
When using GBASE data are you interested primarily in i. Absolute values (mg kg^{-1}) ii. Values relative to a specific threshold iii. Relative values (high, medium low)
What are the implications of uncertainty in the data for your interpretation (systematic biases, random fluctuation (e.g. due to sampling error) or both).
What account do you currently take of uncertainties in GBASE data when interpreting them?
Data user questions (continue from previous list)
What ways of representing the uncertainty in GBASE data are or would be most useful[2]? i. Variance or standard deviation ii. Percentiles of the distribution iii. Confidence intervals iv. Probabilities relative to a threshold v. 'Reproducibility' of the spatial pattern
Have any issues raised by the data providers given you new insights or concerns about uncertainty in geochemical data such as GBASE?
Data provider questions
What do you regard as the principal sources of uncertainty in GBASE data? i. Spatial variability, both geogenic and anthropogenic ii. Field sampling iii. Post-collection management and treatment of material iv. Sample preparation and analysis v. Initial data management and levelling vi. Subsequent data base management vii. Other
Which of these uncertainties are most readily managed? Do any, which are now well-controlled, pose problems for the use of older data?
How is this uncertainty communicated to users?
What assumptions are made about how users will account for uncertainty in their use of data?
Are there examples of how interactions with users have changed the way in which you deal with uncertainty in the data? Have any new issues emerged from user responses today?
How do you think that uncertainty in GBASE data can be most effectively communicated to users?

[2] The first three approaches to visualising uncertainty are widely used in the geostatistical literature and can be found in texts such as Webster and Oliver (2007).The approach to visualizing probabilities in combination with a verbal scale is more novel and is illustrated by Lark et al. (2014). The approach based on reproducibility of the spatial pattern is presented by Lark and Lapworth (2013).

Answers from participants to the questions presented are shown in detail in Appendix 3 (page 37). The discussion led to the following key topics and observations.

Does uncertainty matter?

Uncertainty in geochemical data matters to both data users and data providers. Resolving uncertainty however may not always be possible. This depends on the stage of investigation and requirements of a project as the cost that would be incurred to minimise uncertainty, for example additional investigations or drilling, may be significant. Methods such as fuzzy logic are used to address interpretative[3] uncertainty.

What is done to manage uncertainty?

Data uncertainty[4] is an issue taken into account by mineral specialists during data interpretation. The geological setting, the values of more than one element and ancillary information are often used to support data interpretation and manage data uncertainty. An experienced geologist will be able to identify values that do not match the geological setting. Pathfinder elements[5] are often used to check unusual results.

Regarding analytical uncertainty, geological expertise is once again very important. Values such as detection limits for elements in question are examined and particular attention is paid to elements that are prone to analytical error (e.g. bismuth).

The role of metadata is also very important. They are used both by data users and data providers to identify errors and anomalies. Metadata information often include detection limits, details on analyses (e.g. analysis of variance to quantify analytical subsampling error and variability between duplicate sites), links to procedure manuals, and information on location accuracy for older data.

Uncertainty is discussed in reports and where appropriate the need for additional work is presented to commercial clients. For research purposes a more thorough discussion and interpretation of uncertainty may be undertaken depending on the project requirements.

Is there potential to do more?

There is the potential for additional statistical measures, such as confidence limits or probabilities that values exceed thresholds, and the novel measure of map reproducibility, which are seen as useful by data users, but several additional steps need to be undertaken before making these available. The use of verbal scales[6] has been seen as particularly useful for communicating uncertainty in a wider group of users. The limitations of datasets should be understood by data users to avoid mixing datasets that are not of the same quality.

---

[3] Interpretative uncertainty corresponds to the meaning of values.
[4] Data uncertainty refers to whether a value is correct.
[5] Pathfinder elements refer to relatively mobile elements that are closely associated with a commodity of interest and can be more easily found or detected by analytical methods. Pathfinder elements can assist investigators to a deposit or substance of interest.
[6] Verbal scales represent verbal descriptions of uncertainty (e.g. unlikely, likely etc.) to convey imprecision in results, predictions, conclusions.

6.2 Workshop II on the implications of uncertainty in mineral statistics

Mineral statistics find use in numerous different areas associated with raw materials including:

- Monitoring the physical economy using material flow analysis models.
- Investigating security of supply of raw materials to identify potential supply disruption issues.
- Strategic planning at government, sector or company level to enable the development of strategies on resource efficiency or sustainability and to set targets for production growth.
- Commodity market analysis to investigate a commodity market within different geographical boundaries and predict future market trends.
- Environmental assessments to identify environmental impacts associated with the mineral industries, but also to estimate waste generation and quantify the 'urban mine'.
- Other, e.g. the use of mineral statistics in scientific communication, research and education.

Mineral statistics comprise a broad range of different datasets including mineral production data, trade data, statistics on mineral resources and reserves, secondary production data, waste data, urban stocks data, exploration data and others. Mineral statistics may represent both the physical and monetary economies. However, data representing the physical economy are perceived as most important as they are the direct derivative values of minerals produced, sold and traded.

Information about the participants and the questions addressed to data providers and data users are shown in Table 7 and Table 8. The following discussion points summarise the workshop outcomes.

Table 7 Dataset and workshop II participants.

Dataset	Mineral statistics including mineral production and trade data
Data providers	Teresa Brown: Mineral Commodity Geologist and Eimear Deady: Mineral Resource Geologist, both working on the production of the BGS World Mineral Production annual publication
Data users	Daniel Beat Mueller: Professor of Hydraulic and Environmental Engineering (Material Flow Analysis) at the Norwegian University of Science and Technology Dominique Guyonnet: Head of BRGM Campus, BRGM, actively involved in Material Flow Analysis.
Workshop moderators	Dr Murray Lark: Environmental Statistician and Dr Evi Petavratzi: Senior Mineral Commodity Geologist

Table 8 Workshop II in Mineral Statistics – Questions used in discussion.

Data user questions
For what purpose do you use mineral statistics?
When using mineral statistical data are you interested primarily in data at: a. Global level; b. National level; c. Regional level; d. Mine specific
Which data types do you currently use that are important to your work? Please prioritise.
What are your preferred sources of mineral statistical data and why?
What level of uncertainty you think there is for the identified data types?
What level of additional uncertainty is created by the use of classification systems?
What are the implications of uncertainty in the data?
How do you take uncertainty into account in your analysis?
What ways could be used to represent uncertainty? i. Explanatory notes; ii. ii. Use of verbal labels to identify uncertainty in reported data
How do you represent data uncertainty in your analysis and interpretation?
After discussion with data-provider group: Have any issues raised by the data providers given you new insights or concerns about uncertainty in mineral statistics?
Data provider questions
What are the principal sources of uncertainty in mineral statistics?
Which types of data uncertainty are most readily managed? Do any, which are now well-controlled, pose problems for: • older data (e.g. time series), or • across different datasets from different data providers, or • across data representing different life cycle stages of a commodity?
How is this uncertainty communicated to users?
What assumptions are made about how users will account for uncertainty in their use of data?
Are there examples of how interactions with users have changed the way in which you deal with uncertainty in the data? Have any new issues emerged from user responses today?
How can uncertainty in mineral statistical data be effectively communicated to users?

Answers from participants to the questions presented are shown in detail in Appendix 3 (page 37). The discussion led to the following key topics and observations.

Does uncertainty matter?

Uncertainty is important and the development of methods, such as on material flow analysis takes it into consideration. Overall two types of uncertainty related to mineral statistics are identified:

- Conceptual, interpretative or epistemic uncertainty regarding the definition of the data available. For instance, information about what process or life cycle stage data represent often is not clear to data users. During data collection, false interpretation of the survey questionnaire leads to misconception of the information and data request made. The aggregation of different commodities in trade classification systems is another example where systematic errors may be present. During data collection, the return of data in a different form to the one requested, without providing any supportive information (e.g. data return may just say copper without indicating the form).

- Data or stochastic uncertainty leading to random errors. Such errors often occur in data collection, or after data collection due to mistakes in the interpretation or publication stages. Another example could be a data return that includes both rounded and unrounded numbers.

The use of classification systems is common in some mineral statistics datasets (e.g. trade data). They are often compulsory and are in place to standardise and harmonise data from different providers, thereby minimising the related uncertainties. In many cases, however, systems of classification introduce additional imprecision because classes include aggregated data, may contain unclear explanations without sufficient metadata and even combine data that represent different life cycle stages or different commodities. Classification systems are often not fit-for-purpose.

Organisations that produce 'world data' have no control over data collection practices taking place in a country or company and often the methods followed at national/regional level are unknown. Therefore the level of uncertainty that is inherent from primary data surveys is also unknown.

What is done to manage uncertainty?

Any type of uncertainty in mineral statistics affects the reliability of subsequent analyses or methods and models employing them and therefore their results. As such, substantial time and effort is spent interpreting the meaning and values of mineral statistics prior to their use in any model. The collection of additional information to improve understanding on the meaning of data or random errors is often essential. Cross-checking complementary sources is commonly undertaken to reduce conceptual uncertainty.

Ongoing communication with primary data providers is crucial to enable the minimisation of errors and associated uncertainties. This is one of the key actions undertaken during data compilation and interpretation to manage uncertainty, including desk studies.

The annual meeting of the International Consultative Group (ICG) on metal statistics, where several data providers meet up and undertake a peer review of the data available for non-ferrous metals is a constructive way to manage data uncertainty. Unfortunately, this meeting only focuses on specific commodities and does not cover the whole range.

Data errors associated with time series are often tackled by requesting data for several years rather than a single year. Any corrections of errors normally appear in the next edition of mineral statistics publications. A number of explanatory notes accompany existing data, which provide some additional useful information. However, these were not deemed sufficiently detailed by data users.

Is there potential to do more?

Additional metadata information from data providers, for example on what the numbers represent including the publishing of mineral statistics in a system context, is perceived as highly beneficial by data users. It is expected that such an approach would minimise conceptual uncertainty substantially. The use of explanatory notes, for examples notes about the data sources, questionnaire returns, calculations undertaken and specific information on the data provided were deemed particularly useful. Equally the use of verbal scales to communicate uncertainty to users, for example by providing qualitative information on the level of confidence associated with data or including colour coded scales can provide a crude, but quick and visual expression of data uncertainty to users.

Data uncertainty associated with missing data could be minimised using the classical probability theory (e.g. normal distribution). Conceptual uncertainty can be handled using fuzzy numbers, namely intervals or possible sets of numbers rather than a single value. This is particularly suitable for representing expert knowledge.

At the moment errors are corrected in the next edition of data. However, establishing a blog or erratum online could speed up this process and enable additional communication between data users and providers. Overall, communication is key in resolving data issues and additional lines of communication between primary data providers (companies, national authorities) and 'World data providers', as well as between data users and data providers should be developed.

Development of a peer review group that extends beyond the work undertaken by ICG would be highly advantageous to improving mineral statistics and minimising uncertainty. This will require several data providers to work together towards a global mineral statistics network.

6.3 Summary conclusions from Workshop I and II

Several discussion points highlighted from the two Workshops are presented in sections 6.1 and 6.2 that are specific to the examined datasets. Many common conclusions have been identified from this work, which are applicable to any dataset.

- Overall two key types of uncertainty are associated with data on raw materials: conceptual uncertainty regarding the meaning of data, and data uncertainty caused by random errors.
- The comprehensiveness and quality of metadata is particularly important. Information and explanatory notes on detection limits, missing data, the data sources used to produce a dataset, sampling variability, location position, the type of survey undertaken and procedures followed and many more, are important and should accompany any dataset.
- Communication between data users and data providers to understand the data generation and data supply chain is crucial in optimising data collection and minimising uncertainty.
- Employed measures of managing uncertainty need to be effectively communicated to data users. Several propositions have been made in sections 6.1 and 6.2 including confidence levels and intervals, verbal labels that are clearly explained and the use of a system context approach to make data available to users could prove useful.
- Establishing peer review processes to address uncertainty is also a method that can be constructive and lead to good results.

The conclusion of both workshops was that additional research is required to address some of the above points and recommended actions to actively incorporate issues around uncertainty in datasets related to raw materials. In MICA an attempt to categorise in a qualitative way the datasets identified in the metadata inventory is ongoing. This work will be using the information recorded on uncertainty during the metadata record development stage. The outcome of this attempt will be presented within D3.4 *Raw materials data strategy*.

7. Conclusion

Deliverable 3.2 provides an overview of the components of the MICA metadata inventory and the progress made since the previous report D3.1 in producing a final version. The metadata inventory is fully populated, but some additional records are expected to be created as more MICA sheets (e.g. FactSheets) are produced. A procedure to link the records to the RMICP is now in place as well as procedures for incorporating references to metadata in FactSheets and DocSheets. The online MICA metadata inventory is well on its way and a second version under the '.eu' domain will become available shortly.

The 'mapping of data to methods' has also been completed and some of the findings of the technical workshop undertaken in the last MICA progress meeting are presented in section 5 of this report. Several data gaps and missing data have been identified. Mitigation actions for managing missing data are described in this report, whilst recommendations for addressing the data gaps will be included in D3.4 report and the 'Raw materials data strategy'.

The two workshops on uncertainty proved to be highly informative as they looked in more detail specific datasets. The outcomes of those two workshops are relevant to other raw materials datasets and as such some of the key observations and outcomes will be included in the D3.4 report. One of the key outcomes of these workshops, which are highly relevant to WP3, is the importance of metadata in identifying uncertainty and the importance of recording high quality information when developing metadata. Following this work, WP3 will attempt to qualitatively prioritise the inventory datasets using as criteria the uncertainty information collected during the metadata record development stage.

8. References

Webster, R., Oliver, M.A. 2007. Geostatistics for Environmental Scientists. John Wiley & Sons, Chichester.

Lark, R.M., Ander, E.L., Cave, M.R., Knights, K.V., Glennon, M.M, Scanlon, R.P. 2014. Mapping trace element deficiency by cokriging from regional geochemical soil data: a case study on cobalt for grazing sheep in Ireland. Geoderma, 226–227, 64–78.

Lark, R.M., Lapworth, D.J. 2013. The offset correlation, a novel quality measure for planning geochemical surveys of the soil by kriging. Geoderma, 197–198, 27–35.

Appendix 1 – Mapping data to methods

Datasets selected from the MICA metadata inventory for the 'mapping data to methods' workshop undertaken in the MICA progress meeting in Paris (June 2017).

Database type	Dataset
1. Commodity sector data on production, trade, (reserves and emissions)	Cobalt Development Institute (CDI)
	International Copper Study Group (ICSG)
	International Lead and Zinc Study Group (ILZSG)
	International Nickel Study Group (INSG)
	World Gold Council
	World Steel Association
	International Copper Association
	Mineral Products Association (MPA) - data
	International Energy Agency - energy statistics - energy balances
	FAOSTAT - Food and agriculture data
2. Commodity country data (profiles, incl. production, trade, (reserves and emissions) etc.)	Report on the raw material situation in Germany
	Minerals Ireland, exploration and mining division
	Mindat
	The mineral resource potential of the Northern European Arctic
	The mineral resource potential of Greenland
	Mineral Resource Map of Cyprus
	Fennoscandia Mineral Deposits
	Finland - Mineral Deposits and Exploration
	Mineral reconnaissance programme reports
	The mineral raw materials of Switzerland
3. Commodity world data (profiles, incl. production, trade, applications, reserves etc.)	BGS Mineral Profile, Coal
	BGS Mineral Profile, Cobalt
	BGS Mineral Profile, Copper
	BGS Mineral Profile, Fluorspar
	BGS Mineral Profile, Lithium
	BGS Mineral Profile, Nickel
	BGS Mineral Profile, Niobium-Tantalum
	BGS Mineral Profile, Platinum Group Elements
	BGS Mineral Profile, Rare Earth Elements
	BGS Mineral Profile, Tungsten
	BGS Mineral Profile, Uranium
	Commodity Information - Bismuth
	Commodity Information - Platinum group metals
	Commodity Information - Zinc
	Commodity Information - Tin
	Commodity Information - Tungsten
	Commodity Information - Antimony
	Commodity Information - Copper
	Commodity Information - Zirconium
	Commodity-Economic Profile - Aluminium and Bauxite
	Commodity-Economic Profile - Antimony
	Commodity-Economic Profile - Bismuth
	Commodity-Economic Profile - Chromium
	Commodity-Economic Profile - Graphite
	Commodity-Economic Profile - Copper
	Commodity-Economic Profile - Palladium

	Commodity-Economic Profile - Platinum
	Commodity-Economic Profile - Phosphates
	Commodity-Economic Profile - Rare Earths
	Commodity-Economic Profile - Silicon
	Commodity-Economic Profile - Titanium
	Commodity-Economic Profile - Tungsten
	Commodity-Economic Profile - Zinc
	Commodity-Economic Profile - Tin
	Commodity-Economic Profile - Zirconium
	Raw materials for emerging technologies (Rohstoffe für Zukunftstechnologien)
	Factsheet zinc
	BGS, World Mineral Statistics
	BMWFW, World Mining Data
	USGS, Commodity Statistics
	USGS, Country Statistics
4. Continent or world data on production and trade of total economy (incl. emissions)	The World Bank, Energy & Extractives Open Data Platform
	Minerals4EU Yearbook
	Eurostat, International Trade
	Eurostat, Statistics on the production of manufactured goods (prom)
	United Nations Commodity Trade database
	Eurostat - Material flow accounts
	Eurostat - Resource productivity
	Eurostat - Consolidated supply, use and input-output tables - data 2008-2009
	Eurostat Manual of Supply, Use and Input-Output Tables
	EXIOBASE
	Eurostat - Material flow accounts - flows in raw material equivalents (RME)
	The World Input-Output Database
	EIONET - European Topic Centre on Sustainable Consumption and Production: Material Flows, Concepts and Methodology
	System of Environmental-Economic Accounting (SEEA)
	Global Material Flows database UNEP - CSIRO
5. Footprint data of consumption by total economy	2012 Data Tables Ecological Footprint and Biocapacity
	The Global Resource Footprint of Nations, Carbon, water, land and materials embodied in trade and final consumption calculated with EXIOBASE2 (release 2.1)
	Water Footprint Network
	Carbon Footprint of Nations
6a. Spatial - geological data	Tellus - Ireland
	Tellus - Northern Ireland
	Finland - Geochemical baselines
	Austrian geological survey on line data portal
	BGR Geo viewer
	Borehole Map (BGR)
	Czech geological survey on line data portal
	Geoindex
	Geological map of Croatia
	Geological map of Spain
	Geology of Britain Viewer
	Geus - geo data for Denmark and Greenland
	GSI - geological Survey of Ireland online data viewer
	GTK on line maps
	IGME 5000 international map of Europe

	Infoterre (BRGM data portal)
	NGU - geodata for Norway
	One Geology
	Polish geological survey data viewer
	Romanian geological survey on line data portal
	Slovakian geological survey on line data portal
	Swedish geological survey on line data portal
	Swiss geological survey on line data portal
	"Geoinform Ukraine, Interactive Geological Map of Ukraine"
	Finland - Geological map
	Geologische Bundesansalt (Austria)
	Minerals4EU Map Viewer
	USGS : Mineral resource data system
	MINVENTORY
	Promine
6b. Spatial - monitoring state indicators (concentrations, stocks)	Global Environmental Database
	Groundwater resources maps of Europe
	FOREGS - geochemical atlas of Europe
	Geochemical map of Spain
	INSPIRE Geoportal
	Atlas of Switzerland: interactive 3D atlas system
	EMODNET - European Observation and Data Network
7. Emission registration, total economy, breakdown in sectors	National emission accounts, WebDab - EMEP
	National emission accounts, UNFCCC
	National Air emission accounts by activity, EUROSTAT
8. Waste and recycling data	Environment data centre on waste - Eurostat
	European Commission - Environment - Waste
	Metal Recycling - Opportunities, Limits, Infrastructure
	Metal Stocks in Society - Scientific Synthesis
	Recycling rates of metals - A Status Report
	Urban Mine Knowledge Data Platform (EU-UMKDP)
	Substance dossier for the recovery of copper from selected waste compared to primary production
	Substance dossier for the recovery of gold from selected waste compared to primary production
	Material dossier for the recovery of SEM from selected waste in comparison to primary production
	Substance dossier for gypsum recovery from waste compared to primary production
	Substance dossier for the recovery of antimony from selected waste compared to primary production
9a. LCA Impact assessment factors of emissions	The Risk List [2015]
	Life Cycle Impact Assessment ILCD recommended factors
	Life Cycle Impact Assessment ReCiPe
	Life Cycle Impact Assessment CML2002
9b. LCA process data, production and emission of process	Life Cycle Inventory data ECOINVENT
	Life Cycle Inventory GABI
	Life Cycle Inventory ELCD
	UNEP SETAC Life Cycle Initiative
	OPEN LCA NEXUS
10. Risks and safety	CDC - NIOSH Mining
	European Agency for Safety and Health at Work

	The Minerva Portal of the Major Accident Hazards Bureau
11. Chemicals and properties	European Chemicals Agency
12. Market and prices	LME
	Metal Prices
	Shanghai Metals Market
	World Bank Commodity Price Data (KNOEMA ENGINE)
	World Bank Commodity Price Data
13. Statistics at Country level	Central Statistical Bureau of Latvia
	Central Statistical Office of Poland
	I.Stat
	Industrial production (annual PRODCOM results)
	StatBank Denmark
	Statistics Lithuania
	National Statistical Institute from Bulgaria
	National Bureau of Statistics of the Republic of Moldova
	State Statistics Service of Ukraine

Workshop results 'Mapping data to methods'

Data2Methods_fina
l2_filled_out_worksh

File 3 Geochemical data uncertainty workshop questionnaire.

Appendix 2 – Data uncertainty questionnaires

MICA
Handout_data unce

File 4 Geochemical data uncertainty workshop questionnaire.

Handout_data
uncertainty worksho

File 5 Mineral statistics uncertainty workshop questionnaire.

Appendix 3 – Data uncertainty workshop responses

Case_Study_Uncert
ainty in Geochemica

File 6 Geochemical data uncertainty workshop – responses to questionnaire.

Data uncertainty
workshop_Mineral S

File 7 Mineral Statistics uncertainty workshop - responses to questionnaire.