

**Project<sup>1</sup> Number:** [776517]

**Project Acronym:** [ORAMA]

**Project title:** [Optimising quality of information in RAW MAterials data collection across Europe]

## **DATA MANAGEMENT PLAN**

Version 2.0

28.6.2019

---

<sup>1</sup> The term 'project' used in this template equates to an 'action' in certain other Horizon 2020 documentation

## 1. Data Summary

### What is the purpose of the data collection/generation and its relation to the objectives of the project?

ORAMA collects information about how data on Primary Raw Materials (PRM) and Secondary Raw Materials (SRM) are collected and what datasets are available. Availability of datasets relating to social and environmental dimensions of primary mineral extraction will also be scrutinized. The objective of the project is to optimise collection of data within each country of Europe in order that pan-European datasets are of high quality and support the European Union Raw Materials Knowledge Base (EURMKB), which feeds the EC JRC's Raw Materials Information System version 2.0 (RMIS 2.0). To reach this objective, a thorough understanding on properties of the available datasets must be acquired. ORAMA does not produce new data. However, the applicability of recommendations will be demonstrated by harvesting new, improved datasets from selected data.

### What types and formats of data will the project generate/collect?

ORAMA will produce recommendations how datasets on PRM and SRM can be improved. Existing platforms such as the EU-MKDP (Minerals4EU – all types of mineral/ore deposits) and related platforms such as the IKMS (EURare – REE) and the CRMKDP (SCREEN – CRM) for PRM, and the EU-UMKDP (ProSUM – the urban mine and mining wastes) for SRM, will be used to distribute datasets and possibly new applications on top of the Central Diffusion DB will be created. ORAMA will test simplified XML model (e.g., ERML Lite) for data collection and delivery. Sankey diagrams will be produced on material and product stocks and flows. Guidelines and training materials for data providers related to collecting and distributing PRM and SRM statistics will be produced.

### Will you re-use any existing data and how?

Existing data will probably be re-used for the purpose of demonstration (see the different sub-tasks of T3.3, and also T3.4).

### What is the origin of the data?

Existing data will be used in for demonstrating the possibilities and benefits of recommendations and instructions provided. In respect to PRM the data include e.g. that currently available in the EU-MKDP (European Minerals Knowledge Data Platform) at <http://minerals4eu.brgm-rec.fr/>, which can be viewed in a map viewer at <http://minerals4eu.brgm-rec.fr/minerals4EU/>. These data originate from the Geological Surveys of many European countries.

The most important source for data related to SRM and used for demonstrative purposes in ORAMA will be EU-Urban Mining Knowledge Data Platform (EU-UMKDP) developed by the H2020 Project ProSUM and available at: [www.urbanmineplatform.eu](http://www.urbanmineplatform.eu). Other data sources that will be used can be named after their evaluation during the first year of the project.

### What is the expected size of the data?

In most cases, datasets are of limited size, the most important in size being the ProSUM dataset which contains around 100 million records (~ 12 Go) for raw (i.e., non-aggregated) data.

### To whom might it be useful ('data utility')?

The data handled will be useful to e.g. for following instances:

- political decision makers on matters related to PRM and SRM on both EU and national level;
- mining companies carrying out exploration or planning on investing in EU;

- geological research;
- companies evaluating the possibilities to invest in SRM sector within EU;
- individual EU citizens searching for information on raw material sources and flows.

## 2. FAIR data

### 2. 1. Making data findable, including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

Data on PRM and SRM used in the ORAMA are largely already available in public online services. Data users will be able to locate the original data sources, as links to these web services will be included in deliverables and the project website [www.orama-h2020.eu](http://www.orama-h2020.eu). The website will also contain all substance-related deliverables that will be produced in the project as well as the training materials. Data that will be provided are INSPIRE compliant.

INSPIRE identifiers and INSPIRE versioning numbers will be used for encoding structured data. Note that the MICA Raw Materials Expert System which will be extended in task T3.4 already use intensively CELEX numbers, DOI and URI for making the link between its information sheets and the external resources.

What naming conventions do you follow?

ORAMA will follow INSPIRE Mineral Resources data specification naming rules.

Will search keywords be provided that optimize possibilities for re-use?

In the project website, keywords will be used as tags to ensure that search engines will easily find them. Note here that an OpenSearch API has been recently developed using the synergy between the SCRREEN and Mintell4EU projects, allowing end users to fully parametrize their search in the Minerals4EU, EURare, ProSUM and SCRREEN Knowledge bases. The idea here is also to allow the RMIS 2.0 to perform dedicated searches on corpus of thematic data already constituted (REE, CRM, the urban mine [WEEE, ELV and spent batteries]), in addition to classical Web searches.

Do you provide clear version numbers?

Clear version numbers will be provided, whenever the deliverable/data is planned to be updated.

Each record of the primary deposit database will be accompanied by its INSPIRE ID and versioning number. Several applications like data augmentation need indeed stable identifiers and the elimination of outdated records.

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

In ORAMA metadata will be created following the INSPIRE directive in all possible instances. Each dataset or piece of information is accompanied by its own metadata based on the Dublin Core model (<http://www.dublincore.org/specifications/dublin-core/dces/>).

## 2.2. Making data openly accessible

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

All data produced/used will be available at least for viewing. Some of the data will be available for downloading, for example the public deliverables and training materials. It may not be possible to enable downloading of all of the data on mineral deposits due to copyright and IP issues. E.g. part of the data are owned by the Geological Survey in each country.

How will the data be made accessible (e.g. by deposition in a repository)?

Data will be accessible via the project website [www.orama-h2020.eu](http://www.orama-h2020.eu) and data portals to which there will be links from the project website. Most relevant portals are:

<http://www.minerals4eu.eu/>

<http://www.prosumproject.eu/>

<http://www.europe-geology.eu/>

What methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? Is there a need for a data access committee?

Data can be accessed (i.e., visualized) through Internet without the need for special software. Standard free Internet browser (e.g., Google Chrome) and PDF reader will be sufficient.

Some data is downloadable:

- The SCRREEN dataset is downloadable in the form of a classical CSV (Comma-Separated Value) file which can easily be re-organized using Microsoft Excel.
- The EURare validated dataset can be downloaded either as a CSV file or a 'ESRI' Shapefile that can be open in ArcGIS or QGIS...
- Regarding ProSUM data, a restful API has been developed taking advantage of the synergy between the SCRREEN and Mintell4EU projects. This API, the ProSUM (or UrbanMinePlatform) API allows end users – including the JRC's RMIS 2.0 - to fully parametrize their download of the ProSUM dataset.

The Minerals4EU dataset, because of its complexity, is currently not downloadable. The development of a dedicated API, based on a full-fledged data model, relatively simplified, but nevertheless very complete and fully operational such as ERML-Lite seems absolutely possible. This would notably allow the ScoreBoard developers to carefully select the data they need. This development is foreseen in the Mintell4EU project, Workpackage 5.

As the data can be viewed and accessed with standard tools there will be no need for specific software documentation or specific software to be downloadable with the data. Neither is there any need to set up a data access committee.

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible. Have you explored appropriate arrangements with the identified repository?

The Metadata Catalogue of the European Minerals Knowledge Data Platform can be found online at <http://m4eu.geology.cz/metadata/>. What comes specifically to ORAMA project, the guidelines for achieving INSPIRE compliance, including recommendations for metadata population, will be included in Deliverable 4.1 Technical Guidelines, which is due M17. This deliverable will cover the issue on the metadata repository.

If there are restrictions on use, how will access be provided? Are there well described conditions for access (i.e. a machine readable license)? How will the identity of the person accessing the data be ascertained?

All the data from the KDPs cited above are public. There is no restriction to access the data for viewing and downloading. However, we should make sure that it is not possible for anyone to start selling the data and therefore we may have to consider restrictions for downloading for example the pan-European datasets on mineral deposits.

### 2.3. Making data interoperable

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

Improving data interoperability is one of the main targets in the project: data will be (or are already) INSPIRE compliant.

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

Data on mineral/ore deposits are interoperable, because standard/hierarchical lexicons on mineral deposits have been defined and further developed during previous projects such as ProMine and Minerals4EU (Schubert et al., 2014<sup>2</sup>), and more recently in the ProSUM project on the urban mine ([http://www.prosumproject.eu/sites/default/files/D5%205%20ProSUM\\_DataModels%26CodeLists\\_fin\\_al.pdf](http://www.prosumproject.eu/sites/default/files/D5%205%20ProSUM_DataModels%26CodeLists_fin_al.pdf)). A review of the INSPIRE Mineral Resource (MR) data model implementation is provided in ORAMA deliverable D3.1. The joint evolution (corrections, additions, code lists improvements...) of this data model and of the international (world-wide) counterpart is also presented

One important aspect in WP1 is to examine the possibility to select one reporting code for primary resources and reserves and start using it in a pan-European level. This harmonising needs to be done before it is possible to add up the resources in different deposits in order to assess the quantity of a specific commodity on national or pan-European level. Conversions from different reporting codes into the common one are described in bridging documents or, when they don't exist, have to be created in a relevant context.

Will you be using standard vocabularies for all data types present in your data set, to allow interdisciplinary interoperability?

Standard vocabularies will be used whenever possible. If there is need to create, modify or refine any vocabulary related to PRM or SRM, this will be addressed in the recommendations that will be produced in the project (see notably D3.1).

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

If there is a need to create ORAMA specific ontologies or vocabularies, mappings to more commonly used ontologies will be provided. Note here that the MICA Raw Materials Main Ontology which is - to our knowledge - the most complete ontology regarding the mineral resources domain will be improved for social and environmental dimensions

---

<sup>2</sup> Schubert C., Vuollo J., Tomas R., Cassard D., and WP5 Partners (2014). *Minerals Intelligence Network for Europe – Minerals4EU. WP5: Common terminology for Minerals4EU – Version 1.0. Minerals4EU Report, Public Document, 330 pp.*

## 2.4. Increase data re-use (through clarifying licences)

How will the data be licensed to permit the widest re-use possible? When will the data be made available for re-use?

The data used in ORAMA can be freely viewed and downloaded (see above) and the public reports and training material produced can be freely downloaded. There will be no licences concerning the data as all the data shown in KDPs is public.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why. How long is it intended that the data remains re-usable?

ORAMA website and the data portals serving the data will be in operation also after the end of the project. It has not been decided how long the ORAMA website will be online after the end of the project. The data portals are intended to be more or less permanent services, however it is also possible that they could be replaced by a comparable but more advanced service in the future, e.g., the EGDI which is currently under construction (<http://www.europe-geology.eu/>) or the EPOS infrastructure (<https://www.epos-ip.org/>). Also note that the communication gateways with the RMIS 2.0 currently developed by the DG JRC are under construction, notably within the Mintell4EU (GeoERA) project.

Are data quality assurance processes described?

Metadata of each dataset will describe the INSPIRE compatibility, which can be regarded as quality assurance.

## 3. Allocation of resources

What are the costs for making data FAIR in your project? How will these be covered?

ORAMA project has been designed so that the data are FAIR (findable, accessible, interoperable and re-usable), but it is not possible to express in numbers what is the cost of that. For that, ORAMA inherits from all the efforts that have been done in several former projects such as Minerals4EU, EURare, ProSUM which are fully INSPIRE compliant, with Dublin Core metadata, CS/W catalogs.

Who will be responsible for data management in your project?

In ORAMA, GTK as the Coordinator and Leader of Work Package 6 Communication, dissemination and exploitation has the main responsibility of data management, however the decisions will be made in the Executive Committee.

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

The issue of long term preservation of data will be discussed later in the project among the members of the Executive Committee. However the idea which is currently explored in association with EGS (EuroGeoSurveys) is to re-activate the Minerals4EU Permanent Body (a Brussels-based non-for-profit foundation) in order to manage/maintain/update the 'ORAMA's datasets' and develop the necessary applications on top of the Central Diffusion DB. Financing of such activities remains a real problem.

#### 4. Data security

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)? Is the data safely stored in certified repositories for long term preservation and curation?

The servers serving the data are secured by standard means such as firewalls and automatic backing up.

#### 5. Ethical aspects

Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

The data in the data portals is not sensitive in any way and does not contain any personal information. The IP rights are respected by acknowledging data sources. Public project deliverables contain names and organisational information of some of the people involved in producing the deliverable, but publishing that personal information does not create any risks to anyone.

Is informed consent for data sharing and long term preservation included in questionnaires dealing with personal data?

In the questionnaire used for collecting information on data on PRM there was option to fill in the questionnaire anonymously. All replies will be made anonymous after the information collected has been analysed, at latest by the end of the project.

In ORAMA, Lyyti registration system is used to collect personal information to GTK's register of participants of events. It can be used for registering participants of ORAMA stakeholder events and internal meetings. After the EU legislation on personal information comes into effect on 25 May 2018, it is not possible to submit Lyyti registration forms without accepting the privacy policy statement, which describes in detail how the submitted personal information will be managed. All information on the participants will be removed from the register at the latest after one month after the end of each event. In addition, the privacy policy statement explains that names and organisational information of participants may be permanently stored in reports describing each event.

#### 6. Other issues

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

#### 7. Further support in developing your DMP

The Research Data Alliance provides a [Metadata Standards Directory](#) that can be searched for discipline-specific standards and associated tools.

The [EUDAT B2SHARE](#) tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.

Useful listings of repositories include:

[Registry of Research Data Repositories](#)

Some repositories like [Zenodo](#), an OpenAIRE and CERN collaboration), allow researchers to deposit both publications and data, while providing tools to link them.

Other useful tools include [DMP online](#) and platforms for making individual scientific observations available such as [ScienceMatters](#).

## SUMMARY TABLE 1

### FAIR Data Management at a glance: issues to cover in your Horizon 2020 DMP

DMP component	Issues to be addressed
<p><b>1. Data summary</b></p>	<p>ORAMA collects information how data on Primary Raw Materials (PRM) and Secondary Raw Materials (SRM) are collected and what datasets are available. ORAMA does not produce new data. Existing data will be re-used for making demonstrations. However, the applicability of recommendations will be demonstrated by harvesting new, improved datasets from selected data.</p> <p>The objective of the project is to optimise collection of data within each country of Europe in order that pan-European datasets are of high quality and support of the European Union Raw Materials Knowledge Base (EURMKB), which feeds the EC JRC's Raw Materials Information System (RMIS). To reach this objective, a thorough understanding on properties of the available datasets must be acquired.</p> <p>Data originates from data producers that are responsible for collecting data on PRM and SRM.</p> <p>The data is useful for policy makers, mining companies, geologists, researchers, anyone who is interested in PRM or SRM.</p>
<p><b>2. FAIR Data</b></p> <p>2.1. Making data findable, including provisions for metadata</p>	<p>Data will be available via public web services. Data will be described in metadata. Relevant search keywords will be used as well as clear versioning, when applicable. The data and metadata are INSPIRE compliant.</p>



2.2 Making data openly accessible	Data that will be used in data portals or ORAMA website will be openly available. There are no intentions to keep any part of the raw material data closed. Standard web browser and pdf reader are required to access the data and therefore no documentation on software is needed. The locations where the data and metadata are deposited will become clear later in the project.
2.3. Making data interoperable	The data are INSPIRE compliant and therefore interoperable.
2.4. Increase data re-use (through clarifying licences)	There is no need for licencing, because the data can be publicly viewed, also after the project has ended. INSPIRE compliance can be regarded as quality assurance. The data in the webportals are planned to stay online for years to come, which is also supported by the development of the EGDI and of the JRC's Raw Material Information System.
3. Allocation of resources	ORAMA is designed to use or generate FAIR data. The costs are included in the project budget but the amount cannot be quantitatively estimated. Data management is the main responsibility of the Coordinator, but the issues will be decided in the Executive Committee or, if need, in General Assembly.
4. Data security	Data in servers can be recovered by standard recovery procedures. The project does not deal with sensitive data; any personal information collected during the project will be made anonymous or deleted by the end of the project.
5. Ethical aspects	The data in the data portals is not sensitive in any way and does not contain any personal information. The IP rights are respected by acknowledging data sources.
6. Other	