MACRAMÉ

Deliverable Report D3.1

Information Hub, Data Exchange Format Specifications, and initial Research Output Management Plan (ROMP)

Project & Programme Information	
Project Acronym	MACRAMÉ
Project Full Name	Advanced Characterisation Methodologies to assess and predict the Health and Environmental Risks of Advanced Materials
Grant Agreement No.	101092686
Work Programme	HORIZON.2.4 - Digital, Industry and Space
Call and Topic Name	HORIZON-CL4-2022-DIGITAL-EMERGING-01-35 - Advanced characterisation methodologies to assess and predict the health and environmental risks of nanomaterials
Instrument	Horizon Europe – Research and Innovation Action (RIA)
Project Start Date	01 December 2022
Duration	36 months

MACRAMÉ has received funding from:
the European Union's Horizon Europe Research and innovation programme under grant agreement No. 101092686.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

UKRI Innovate UK

Project & Programme Information	
Work Package	WP3
Deliverable Number	D3.1 (15)
Associated Task	Task 3.1 and Task 3.2
Type [1]	R
Dissemination Level [2]	PU
Version	V1.0
Actual Submission Date	31/05/2023 (M06)
Contractual Submission Date	31/05/2023 (M06)
Lead author (organisation)	Thomas Exner
Contributors	Maja Brajnik, Steffi Friedrichs, Joh Dokler, Rob Vandebriel
Reviewers	Steffi Friedrichs

Version	Date	Changes made	By	Sent to	Purpose
0.1	04.05.2023	Draft version	7P9-SI / 7P9-DE	AIST	Partner contribution
0.2	17.05.2023	Contribution	7P9-SI / 7P9-DE	All partners	Partner contribution
0.3	24.05.2023	Contribution	RIVM	7P9-DE	Review
1.0	31.05.2023	Finalisation & Submission	S. Friedrichs (AIST)	EC	Submission

[1] R= Document, report (excluding the periodic and final reports); DEM = Demonstrator, pilot, prototype, plan designs; DEC = Websites, patents filing, press & media actions, videos, etc.; DATA =Data sets, microdata, etc.; DMP = Data management plan; ETHICS = Deliverables related to ethics issues; SECURITY = Deliverables related to security issues; OTHER = Software, technical diagram, algorithms, models, etc.

[2] PU = Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page); SEN = Sensitive, limited under the conditions of the Grant Agreement; Classified R-UE/EU-R = EU RESTRICTED under the Commission Decision No2015/444; Classified C-UE/EU-C = EU CONFIDENTIAL under the Commission Decision No2015/444; Classified S-UE/EU-S = EU SECRET under the Commission Decision No2015/444

MACRAMÉ has received funding from:
the European Union's Horizon Europe Research and innovation programme under grant agreement No. 101092686.
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI
UKRI Innovate UK

Acronyms Listed in this Document	
AdMa@CMs	Advanced Materials in Complex Matrices
ALI	Air-liquid interface
CHADA	Characterisation Data (file format)
CML	Control Material Library
DOI	Digital Object Identifier
EC	European Commission
ELN	Electronic Lab Notebook
EoL	End of Life
FAIR	Findable, Accessible, Interoperable, and Re-usable
FAIR4RS	FAIR for Research Software ()
FER	FAIR Enabling Resource
FIP	FAIR Implementation Profile
FSR	FAIR Supporting Resource
GD	Guidance Document
LCA	Life Cycle Assessment
LCI	Life Cycle Inventory
PEC	Predicted Environmental Concentration
MaaS	Materials as a Service
MFA	Material Flow Analysis
MODA	Modelling Data (file format)
NIKC	NanoInformatics Knowledge Commons
RA	Risk Assessment
ROMP	Research Output Management Plan
SSbD	Safe-and-Sustainable-by-Design
TG	Test Guideline
WP	Work Package

Contents

Table of Figures

Table of Tables

Executive Summary

This report describes the knowledge management and sharing solutions of the MACRAMÉ project based on a mixed system of centralised and distributed components. The system is supporting (i) the planning of the experiments preformed as part of the method developments and the MACRAMÉ Use-Cases, (ii) organising the knowledge, information, and data as well as material flow between partners and work packages, and (iii) FAIRification and preparing the research outputs (sampling plans, study designs, *in vitro* and *in silico* method specifications, protocols, SOPs, and the data created, as well as guidelines, reports, training materials and publications) for long-term storage and public sharing. Concepts and approaches for harmonisation and FAIRification are outlined. They provide flexibility of the formats used to report the different research outputs and solutions for storing them and, at the same time, help to continuously improve FAIRness by data documentation and mapping to structured (meta)data schemas customised for specific types of research output. The initial Research Output Management Plan (ROMP), provided with this report, documents the recommendations resulting from the agreed concepts and the data management decisions taken; it also outlines the necessary steps to reach the high standards of FAIRness anticipated for all research outputs of the MACRAMÉ project. Constant updating, improvement and alignment with the ongoing community activities of the ROMP will ultimately lead to the final version at the end of the project giving full account of all research outputs generated and collected.

1 Introduction

As a part of its open science philosophy, MACRAMÉ is implementing high quality knowledge and data management using state-of-the-art data sharing concepts, approaches and tools and constantly improving research output/data documentation towards full implementation of the FAIR (findable, accessible, interoperable and re-usable) and FAIR for Research Software (FAIR4RS) principles. This is absolutely essential not only for public sharing for reuse but for the internal organisation and management of the complex knowledge flow between the work packages to be able to characterise the advanced materials in complex matrices (AdMa@CMs) investigated in the MACRAMÉ Use-Cases. As described in more detail below, MACRAMÉ's WP2 (Development and Advancement of Characterisation- & Test-Methods &-Protocols) and WP4 (Definition, Demonstration & Validation of MACRAMÉ Methods in Use-Cases) are intensively collaborating to define the needed characterisations, generate flows of information to satisfy the information requirements of downstream experiments, and bring all evidence together for performing Safe-and-Sustainable-by-Design evaluation of the materials along their complete life-cycle. Planning as well as execution of the studies is supported and documented by WP3, providing the tools for transfer and exchange of all data and knowledge. Due to the extreme variety of data types collected and generated in the studies, as well as the different starting points with respect to data management at the different MACRAMÉ partners, the data management tools (incl. FAIR Enabling and Supporting Resources (FERs and FSRs)) are selected and developed both with flexibility in mind and - more important still - with clear routes to improve the FAIRness of the research output over time at the partner institutions and in MACRAMÉ in general. This is not only limited to data but extended to all research output including - but not limited to - sampling plans, study designs, *in vitro* and *in silico* method specifications, protocols, SOPs, and the data created, as well as guidelines, reports, training materials and publications. Such selection and development of all underlying data will be realised, at its final stage, through a combination of public and internal knowledge and data resources and metadata-reach, semantically annotated data transfer formats.

This report describes the initial version of this data management system consisting of a central information hub, i.e. the **MACRAMÉ Registry**, a visualisation tool for the planning of information flows, and a concept for documenting the current state of data sharing documentation and sharing approaches and roadmaps towards harmonisation and interoperability. It also covers the first Research Output Management Plan (ROMP), an extension of the Data Management Plan covering all research outputs listed above, which will be constantly extended and improved as a living document to document all MACRAMÉ output and the taken steps to improve their FAIRness and integrate them into the MACRMÉ data system as well as the European digital materials ecosystem.

2 Knowledge Generation in MACRAMÉ

Knowledge and data is being generated as part of the development of (a) a **MACRAMÉ Control Material Library** (CML), including reference materials and AdMas, product matrices and their composites, (b) **MACRAMÉ Validated Protocols**, and (c) the **MACRAMÉ Use-Cases** on five industry-relevant materials (Use-Case 1: Graphene oxide (GO) flakes in drinking-water filters (Material Family: GRM); Use-Case 2: Few-layer graphene (FLG) in battery management systems (BMS) (Material Family: GRM); Use-Case 3: Graphene-related materials (GRM) in car polish consumer sprays (Material Family: GRM); Use-Case 4: Carbon nanotubes (i.e. CNTs) in car-seats (Material Family: carbon fibres); and Use-Case 5: Poly lactic-co-glycolic acid (PLGA) for inhalable antibiotics (Material Family: PLCG)), established in collaboration between WP2 and WP4.

WP4 is creating information and data by selecting sampling points, where potential inhalation exposure risks have been identified, along the value-chains of five market-relevant industrial MACRAMÉ Use-Cases, and collecting and delivering samples to WP2 for analysis and investigation.

WP2 is developing and harmonising methodological approaches for controlled aerosol generation, aerosol characterisation and *in vitro* exposure. All approaches are assessed and optimised for their capacity to characterise physicochemical properties of aerosol, dust and airborne material generated during the different phases of production, use and disposal of AdMas. State-of-the-art complementary qualitative and quantitative imaging methods and mass spectrometry approaches are combined, correlated and fully extended to characterise and image AdMas in their pristine or early state, or upon integration in complex matrices (incl. complex (biological) matrices and different physical/chemical states of the AdMas as a whole, or their individual components). The methods are applied to, benchmarked and validated against the MACRAMÉ Control Material Library and Use-Cases.

MACRAMÉ WP2 is also optimising and developing biological assays/systems with increasing complexity for inhalation *in vitro* toxicology, including single cells, bi-culture systems, advanced *in vitro* and *ex vivo* models, representative of upper (bronchi) and lower (alveoli) airways, for regulatory hazard assessment and/or exploratory studies, thus enabling the formulation of new data-driven hypotheses. The methods are organised in a hazard assessment framework that allows toxicological assessment of AdMas life cycles, by harmonising different throughputs and biological readouts (e.g. early signs of toxicity or detection of long-term effects such as genotoxicity). Air-liquid-interface (ALI) exposure, a critical aspect of *in vitro* inhalation studies, are harmonised to ensure that the different methods for exposure of biological systems provide similar and comparable exposure conditions. The *in vitro*/*ex vivo* methods are benchmarked against the MACRAMÉ Control Material Library, comprising materials with different physicochemical properties and known biological effects.

A quantitative hazard and risk assessment (incl. occupational, consumer and environmental RA) on the most relevant stages of the life cycle of Use-Cases is being performed by WP4. Data from the characterisation methods of WP2 are used in and complemented by a **Material Flow Analysis** (MFA) to provide an overall assessment of the releases of materials at each life-cycle stage, based on production amounts and transfer coefficients from one stage of the life cycle to another. MACRAMÉ overcomes the current limitation of MFAs by linking to the advanced, quantitative imaging techniques (developed and applied in WP2) to generate specific data on the forms and quantities of AdMa(@CM)s. Focus is put on evaluating the end-of-life (EoL), where uncontrolled release and exposure is likely to occur. The MFA analysis is then used to generate data on potential occupational and consumer exposure and fate models to obtain predicted environmental concentrations (PECs). Currently missing elements within the Life-Cycle Assessment (LCA) for AdMas are addressed by Characterisation Factor (CFs) for the impact categories of human toxicity and ecotoxicity and life cycle inventories (LCIs), for each industrial Use-Case and for each life-cycle stage (manufacturing, use and end of life (EoL)).

Data from the MACRAMÉ material characterisation techniques is being enriched with information from a multitude of other sources and aggregated for knowledge discovery (WP3). Collected data is then also applied in existing approaches for grouping and read-across (WP3). This assesses the relevance of the data produced in MACRAMÉ for setting the boundaries between nanoforms.

3 Knowledge Transfer between Work Packages

The data generation process is depicted in Figure 1. It requires a very intense two-way exchange (of information and knowledge) between the WPs and is based on the development of new characterisation methods and their validation of them *via* industry Use-Cases. This exchange occurs in the form of materials and sampling plans, study designs translating needs into actions, method specifications describing the measurement principles, protocols and SOPs, and finally the data (raw, transformed and robust summary); the underlying data of this entire process is referred to as **MACRAMÉ (Meta)Data** in this report. For each Use-Case, information is initially generated by WP4 (incl. material flow analyses (MFAs), exposure points, sampling points and protocols, and sample preparation SOPs). Then, the data generation is started with the sampling protocols in WP4, and then continued with characterisation and human toxicity testing (WP2) and with ecotoxicity and life-cycle-based assessments, MFAs and risk assessment (RA) (WP4). This is further complemented by gap filling, using existing data from public resources identified by WP1, mined, curated and integrated by WP3, and subsequently processed in the modelling approaches of WP3 that generate computational data. All of this information is being integrated, harmonised and handed back to the Use-Cases (WP4) for analysis against LCC criteria, and in a condensed (summary) form to WP5 and WP6 as basis for the development of updated guidelines and standards, and for dissemination of the Project results to target stakeholders. All information is managed and exchanged *via* WP3, as depicted in Figure 1, guaranteeing safeguarding of all MACRAMÉ (Meta)Data with full provenance trails, and creating a first implementation example of the concept of 'Materials as a Service'-concept (MaaS) for the safety & sustainability of AdMa@CMs. Data to be shared includes not only data generated by the fully developed method as used in the Use-Cases of WP4, but every produced data point with their specific protocols from early development, calibration and control measurements, thus collecting knowledge about the applicability domain and reliability of the methods, and forming the basis for the (pre-)evaluation of the methods and for evaluation of their applicability in risk assessment, grouping and read-across approaches.

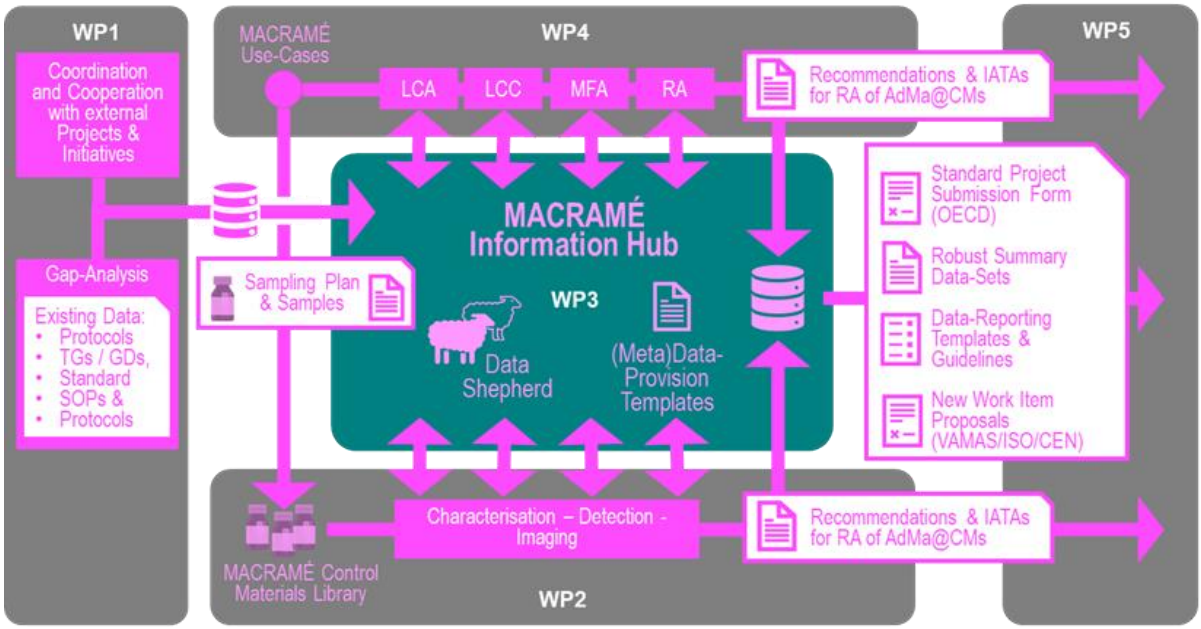

Figure 1: Schematic illustration of the flow of and value-add to data within the MACRAMÉ Project.

4 Data Lifecycle and Data Roles in MACRAMÉ

FAIRification of data, i.e. making data findable, accessible, interoperable and re-usable, is generally accepted as essential for being able to fully exploit the knowledge contained in the immense mass of produced data and is made mandatory in all EU funded projects. However, FAIRification is often still seen as something needed only when the data is shared for reuse, leading to two issues extremely reducing the amount of available FAIR data:

1) FAIRification is an external process outside the normal handling of data in the research laboratories, often depicted as the sixth stage of the data management life cycle; it leads to extra work added to the already heavy workload of the researchers.
2) First analysis of the data and internal data management cannot profit from the benefits FAIR data brings with it. Benefits such as full access to all relevant information (data and metadata like study designs, method description, protocols, as well as description of quality control, analysis and uncertainty estimation procedures) allow the direct use by other partners and WPs and applying existing standard tools and workflows for further processing (e.g. categorisation and grouping) without the need of time-consuming data transformation to fit the required input format.

In contrast, the MACRAMÉ consortium already agreed at the proposal preparation stage that high-quality data management and FAIRification tasks need to be budgeted for at the proposal stage, in order to ensure internal data sharing and implemented from the study design phase onwards. This was seen as essential to be able to plan the collection of all information needed for characterisation of the AdMa@CMs in the Use-Cases and to realise the complex information flows between the WPs described above. The commitment was formalised by assigning part of the person-months budgets of all data producing and collecting partners to WP3 for covering all data management related tasks. This way, moving the FAIRification task into the earliest stage of the data management life cycle helps to minimise the overhead, even if extra work is required.

In MACRAMÉ, harmonisation and sharing of (Meta)Data is done as part of the normal experimental workflow. Figure 2 (adapted from Papadiamantis et al.(1)) shows how such an early integration maps onto the data management life cycle, where information is reported in a FAIR way when it is produced, enhanced and completed, while it moves through the life-cycle stages. Clear documentation of the study design in the life-cycle phase 1 "Plan & Design" increases the FAIRness of the data and specifically supports the re-use part of the FAIR concept, since this information gives the data customer (i.e. the first user, as well as a re-user in the form of other partners) important details to understand the (Meta)Data and its reliability. Collection of the data and its analysis in phases 2 and 3 (Collection & Creation and Process & Analyze) are subsequently utilising harmonised and interoperable data reporting formats and workflows directly or are semi-automatically translated into these; this – in turn - guides the provision of the needed metadata. The remaining steps then only need to add more technical information like unique, persistent identifiers (e.g. DOIs), licences and other biographical metadata for improved findability and accessibility.

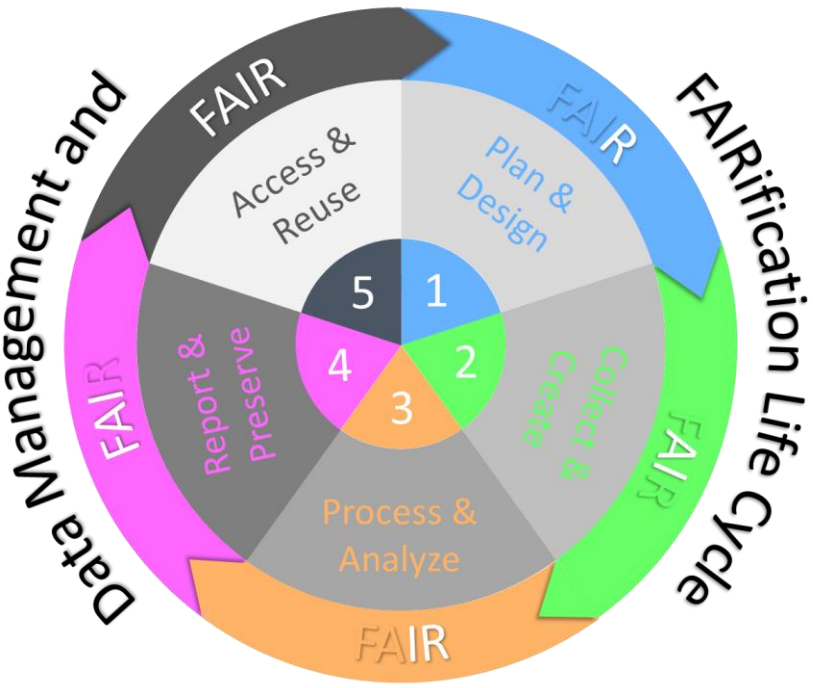

Figure 2: Five phases of the data management life cycle. In contrast to other depictions where FAIRification is a sixth phase, FAIRification in MACRAMÉ is starting in the Plan & Design phase and then continues throughout all other phases continuously enhancing the FAIRness of the data (on-the-fly FAIR).

Distributing data management and FAIRification across the data management life cycle stages brings with it the distribution between different roles. Papadiamantis *et al.*[1] defined these roles as 'data creators', 'analysts', 'curators', 'managers', 'customers' and 'shepherds', respectively, and described their responsibilities as summarised in Table 1. All these roles are present in MACRAMÉ with the data creators and curators mainly represented by WP2, data analysts and customers by WP4 and WP5, and data managers and shepherds by WP3. More information on the data shepherding services implemented in MACRAMÉ are given below (see Section 8).

Table 1: Data roles, responsibilities and interactions (reproduced from Papadiamantis et al.[1]).

	Set objectives	Design Approach	Collect	Processing	Modelling/Analysis	Validate	Store	Share	Quality Control	Annotation	Determine Relevance	Apply	Confirm Effectiveness	Generalise	Communication and Education
Creators	X	X	X	X		X			X	X	X		X		X
Analysts		X		X	X	X			X	X	X		X	X	X
Curators				X		X			X	X	X		X		X
Managers							X	X	X				X		X
Customers	X								X		X	X	X	X	X
Shepherds	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

5 The MACRAMÉ Registry and Pathway to harmonised Research Output Documentation

As described above, the MACRAMÉ Control Material Library development and the Use-Cases generate and collect many pieces of information and knowledge as research outputs (e.g. sampling plans, study designs, method specifications, protocols, SOPs, and the data created by them, as well as guidelines, reports, training materials and publications) coming from multiple partners, which need to be shared with the consortium in a timely manner and in parts collaboratively edited. To avoid delays caused by waiting times resulting from the need for agreeing file formats and data models between by all partners, adoption by all partners, or the transfer of data into these formats, the FAIRification actions at this early stage of the project concentrate on findability and accessibility. This is achieved by the combination of a distributed data storage system with a central MACRAMÉ Registry (see Figure 3), which indexes high-level metadata, like name, type of resource, short description, provider, and documentation completion status, as well as the link to where the resource is hosted (see Figure 4). Statuses of resources are ranging from "planned", over different levels of completeness (i.e. "under development", "internal review", and "external review") to "publicly shared". In this way, the registry can already document the need to generate a specific resource identified during the design stage of the Use-Cases or method development and the responsible partner. By changing the status, all partners can then be informed about the progress of creating the resource (e.g. ordering and arrival of starting material, development of a SOP, or collection of data, and location and access routes to the current, most up-to-date version are clearly defined).

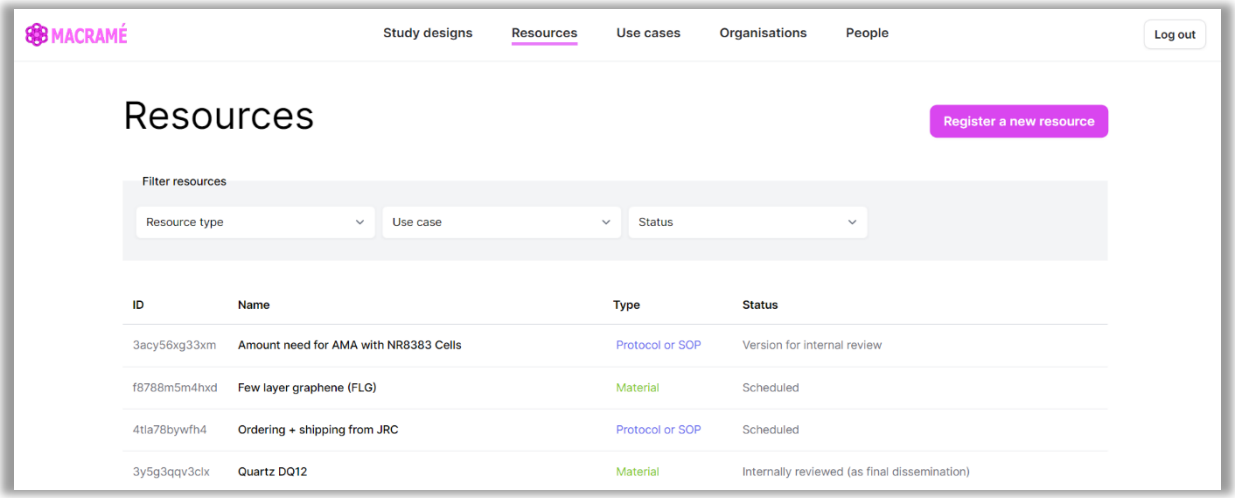

Figure 3 :Screenshot of the MACRAMÉ Registry showing four resources (2 materials and 2 protocols) with different statuses.

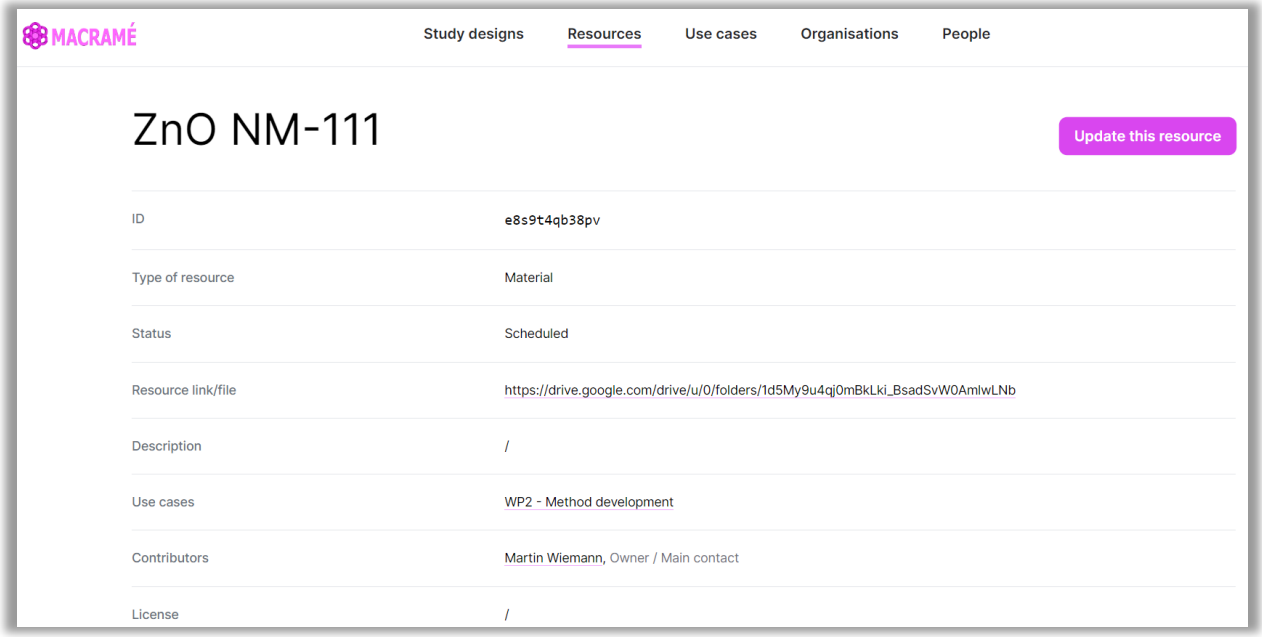

Figure 4: Screenshot of the high-level description of a MACRAMÉ material resource.

Formats, in which the actual resources are prepared, are currently very different from one partner to another. For example, protocols/SOPs and data are often available as text documents (Microsoft Word, Google docs or pdf) and spreadsheets (Microsoft Excel) but some partners also use electronic lab notebooks (ELN) and automatic data capturing workflows or proprietary file formats, respectively. Again, at this early stage of the project, it is most important to not cause time delays by not sharing (Meta)Data between partners and work packages. Therefore, instead of forcing all resources into a predefined template, the Registry supports all these different reporting styles. To provide the needed flexibility, a resource is just specified by an URL, which in many cases links to a folder in the MACRAMÉ Google Shared Drive but could also redirect to e.g. a page of an ELN. Using the paid Google service instead of the free version provides higher data security and access restrictions including forcing the hosting of the data on European servers.

MACRAMÉ partners have started to populate the Registry with information specifically on material production and transfer into the complex product matrices, as well as on previous experiments to be used as reference data. In this way, the status of data management at each partner, used management tools and file formats are documented in these examples. This is now being used to identify commonalities in data documentation, define minimal data and metadata to be reported for each type of research output, and find (meta)data gaps hindering the reuse of the data internally and later public sharing. Based on this analysis, recommendations for improvements of data management at each partner to fill the identified gaps will be suggested, harmonised (Meta)Data schemas and data transfer formats will be proposed and aligned to community standards, still allowing flexibility to adapt to specific data types. Mappings between the data schemas used internally by a partner to these harmonised transfer formats and (semi-)automatic data transformation procedure will be created. This continuous improvement of the data management practices will finally generate data of highest FAIRness levels for use in categorisation and grouping as well as public sharing with the least amount of changes in the partner-internal workflows with respect to used data management tools and formats and with manual data transformation limited to the absolute minimum.

6 Documenting Experimental Design and Execution

Besides data documentation and access *via* the MACRAMÉ Registry, design of the complex studies and experimental workflows and organisation of the virtual transfer of data but also physical transfer of materials and samples between work packages and partners is supported by visualisation of the flows in the form of instance maps and integrating these maps into the Registry.

Instance maps were first used as organisational structure in the data curation efforts for the NIKC (NanoInformatics Knowledge Commons); they enable users to follow nanomaterial transformations, while capturing necessary metadata. An instance is defined as the nanomaterial in a medium at a specific moment in time. An instance map then represents a flow chart of the nanomaterial fate represented as a directed, often tree-like graph built out of nodes connected by edges represented as arrows to show the directionality. To support the quick generation of such maps, a specific software tool was developed in the NanoCommons project. The tool introduced a couple of modifications and new features:

1) Besides the five node categories (instance, material, medium, property, and supplementary) of the original approach, transformation protocol nodes explicitly describing the processes leading from one instance to the next, protocols/SOP nodes for experimental details and data nodes were added.
2) Information resources can be associated with each node.

These new features especially make instance maps an optimal planning and monitoring tool for the experimental work by creating "mind maps" of the Use-Cases showing material flows, dependencies of experiments, which information is integrated in the SSbD assessment and how.

Nodes can be linked to the resources in the Registry providing structure to the many resources e.g. associated with a Use-Case, giving easy access to all information needed to start the next phase of the study, and in this way, supports the management of the studies and of the overall project.

Within the MACRAMÉ Project, the planning of the experimental work and workflows using the tool has been started as a co-creative training session resulting in partial instance maps with one example shown in Figure 5.

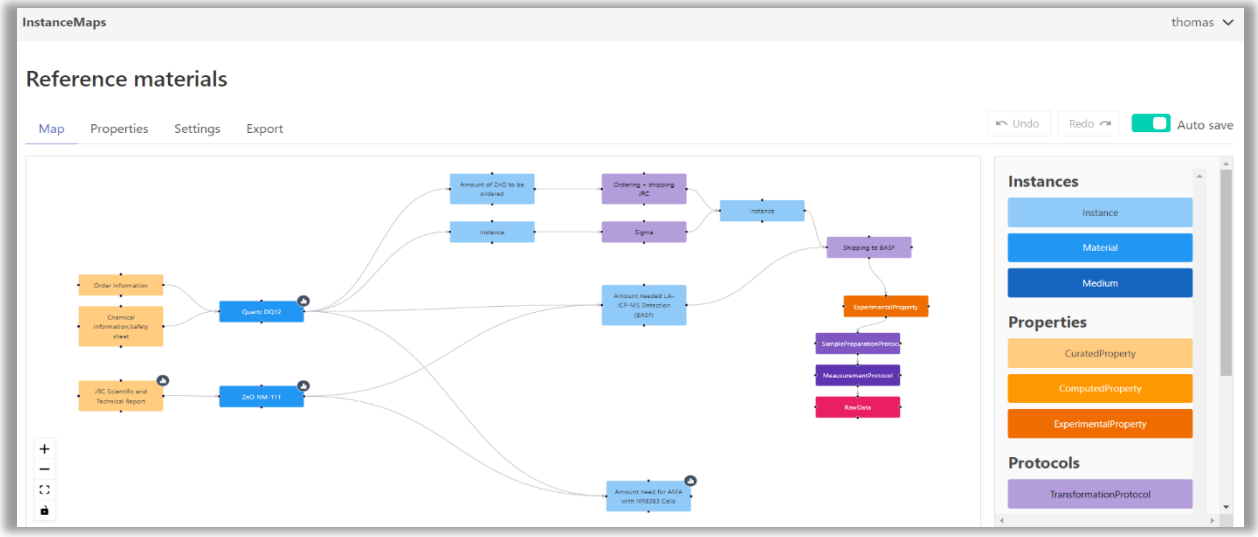

Figure 5: Screenshot of an instance map representing the current planning status of the central control material ordering and shipping to the partners.

7 Research Output Management Plan

More specific details of and guidance on the adopted management principles, FAIRification approaches, data security and ethics considerations are described in the first version of the research output management plan (ROMP) provided as ANNEX A1 to this report. These details are being applied to all digital objects produced by the project including, but not limited to, sampling plans, study designs, method specifications, protocols, SOPs, and the data created as well as guidelines, reports, training materials and publications. Besides the initial version presented here, we anticipate that at least two additional stable versions of the plan (intermediate and final) will be compiled and publicly shared, to accommodate reporting of the data generation at the midterm and end of the project but also be able to adapt to the fast-changing field of Open and FAIR data in Europe. These will then also represent the improvements in the data documentation and FAIRness described in Section 5.

According to the EC recommendation to make data findable, accessible, interoperable and re-usable (FAIR), a Data Management Plan and, thus, also its extension, the ROMP, includes information on the following details:

1) handling of research outputs (RO) during & after the end of the Project;
2) what RO are collected, processed and/or generated;
3) which methodology and standards are applied;
4) whether RO are shared/made Open Access;
5) how RO are curated and preserved (including after project end).

The initial version of the MACRAMÉ ROMP mainly documents decisions made by the consortium and guidelines to be implemented by the data providers and the data managers in collaboration with and under supervision of the MACRAMÉ data shepherd. This will subsequently be complemented throughout the project with specific information on the research outputs generated and resources reused in the project as well as updates and refinements of the used FAIR Enabling and Supporting Resources (FERs and FSRs) and improvements in harmonisation and interoperability. Thus, the ROMP has to be understood as a living document, in which changes are clearly tracked to show additions as well as decisions, which had to be changed,

revised or replaced as reaction to emerging new standards, tools and guidelines. In cases where the content of the document needs larger adaptations for a specific type of research output, additional individual ROMPs for this type or even of a specific research output will be created and referred to in this general ROMP covering the project globally.

For creating the ROMP, three different online DMP tools were evaluated according to criteria that evaluate if they covered all necessary aspects and provided the appropriate structure needed to define the management approach and the generalisation to all research output. The tools evaluated are:

1) DMPonline from the Digital Curation Centre,
2) DS Wizard and
3) argos.

DMPonline provides a quite general structure with questions associated with specific categories like FAIR data, Data security and Ethics. However, most of the questions are related to FAIR and are specific to data. Other research outputs (e.g. protocols, SOPs, software) can only be addressed by answering two very generic questions. Additionally, there is no way to describe the reasons for data collection and reuse or the overall concept, how links between different research outputs are established.

DS Wizard uses the concept of following the data life cycle from creation to processing to preservation and sharing. The advantage is that the user gets more guidance by the detailed structure of questions, follow-up questions based on previous answers and suggestions for answers for specific details like file formats and ontologies. However, some of the questions are very specific and should, in our opinion, be answered outside of a DMP or ROMP and other aspects cannot be provided at all. One example are the questions about quality control addressing important points but are not able to represent the complex quality control needed for SSbD projects, which is better covered outside of the DMP/ROMP as part of the protocols and SOPs.

Finally, argos seems to be the closest match for MACRAMÉ's requirements. Unfortunately, the tool expects that a DMP is created for each research output individually. It is, therefore, more useful for documentation of the approaches taken during or after generation but less for conceptualisation and giving guidance to the project partners at the beginning of a project. For that, the description should cover groups of research outputs and should be more in the form of a guidance document. Therefore, it was decided to not use one of the online tools but to generate the ROMP using the structure from argos as the starting point. This was then amended and complemented by adding individual questions or topics from the other two systems and considering the FAIR Principles for Research Software (FAIR4RS) besides the original FAIR principle for data since topics like versioning and qualified references to other objects are also relevant for protocols, SOPs and guiding documents.

8 Data Shepherding

All MACRAMÉ partners are working together on the definition of reporting standards and corresponding SOP/protocol and (Meta)Data reporting formats and corresponding datasets. However, for harmonisation and increasing interoperability both within the project, and with the relevant community activities and the forming European digital materials ecosystem, advanced understanding and knowledge on data management and FAIRification concepts and the available FAIR Enabling and Supporting Resources is required. This requirement resulted in the definition and establishment of data stewards[2] and recently of data shepherds[1]. The latter can be described as an enhanced version of the first, who not only oversees the data management, handling and quality control processes, but can communicate in a clear and

simple language with all parties and resolve any misunderstandings[1]. Data shepherds lead the data quality control and FAIRness efforts, selection and/or development of the appropriate data management toolset according to both the data types and the requirements of specific partners, as well as the continuous optimisation of data workflows, including technical developments to facilitate data curation, annotation, and cleansing. To provide these services to all partners, a continuous, centralised (Meta)Data Shepherding process was established in MACRAMÉ. This started with multiple trainings and workshops on the use of the MACRAMÉ Registry and the instance map tool (i.e. during the project kick-off meeting & 1st project meeting, and through workshops on 15th March and 29th March 2023) and will now continue with separate workshops for the five individual Use-Cases, in which instance maps will be created to describe the study design and information flow between the partners and document the existing information on the starting materials and products as well as their production processes. New resources in the MACRAMÉ Registry (materials, methods, protocols and data) will then be continuously linked to the instance maps, as soon as they are generated ensuring timely sharing of data as the route of knowledge transfer between the WPs and then for open and FAIR sharing with all stakeholders. The data shepherd is also continually supporting all partners by analysing the current data management practises at individual partners and proposing options and tools for the harmonisation of data, improving interoperability and FAIRness. Frequent re-evaluation of the completeness of the provided information for the Use-Cases and for standardisation of the MACRAMÉ methods and output will be performed, and improved reporting guidelines resulting from these activities (e.g. implementing the concept of method-specific reporting) will be proposed to the scientific community and standardisation bodies (VAMAS, OECD).

9 Conclusion

At this early stage, the knowledge management and sharing system is focusing on supporting the planning of the experiments supporting the MACRAMÉ Use-Cases and organising the complex sample and information flow between partners and work packages required during their execution. Additionally, the concepts and approaches for harmonisation and FAIRification of all research output are outlined and first steps have been implemented for the high-level data documentation based on a system of unique identifiers and a limited amount of metadata describing provenance, accessibility and status of the research outputs. The data providers are currently given full flexibility of the formats used to report the different research outputs and solutions for storing them, in order to allow direct sharing within the project without delay caused by manual data transformation. The provided preliminary research outputs are now being analysed to generate further detailed reporting guidelines, create highly structured (meta)data schemas for types of research output, and start harmonisation and further FAIRification. The Research Output Management Plan (ROMP), provided in its initial version as ANNEX A1 to this report, documents the data management decisions taken and outlines the necessary steps to reach the high standards of FAIRness anticipated for all research outputs of the MACRAMÉ project. It will be constantly updated, improved and aligned with the ongoing community activities to give full account of all data generated and collected including measures to guarantee FAIRness at the end of the project.

10 Bibliography

1. Papadiamantis A.G. *et al.*, Metadata Stewardship in Nanosafety Research: Community-Driven Organisation of Metadata Schemas to Support FAIR Nanoscience Data. Nanomaterials, 2020 Oct.; 10(10):2033; https://doi.org/10.3390/nano10102033

2. Mons B. Data Stewardship for Open Science: Implementing FAIR Principles. New York: Chapman and Hall/CRC; 2018, 244 p., https://doi.org/10.1201/9781315380711

ANNEX A1 – Initial Version of the MACRAMÉ Research Output Management Plan (ROMP)

This ROMP is based on the Horizon Europe template taken from the argos DMP platform with extensions by MACRAMÉ (marked in pink), needed to be applicable to all research outputs.

0 Main information

0.1 Project

MACRAMÉ (Advanced Characterisation Methodologies to assess and predict the Health and Environmental Risks of Advanced Materials)

0.1.1 Grant information

Project Start: 01.12.2022

Project End: 30.11.2025

Funder: European Commission - Horizon Europe

Grant url: https://doi.org/10.3030/101092686

0.1.2 Contributors

Name	ORCID	Affiliation	Role
Thomas Exner	0000-0002-1849-5246	7P9	Data Manager Principal Investigator
Joh Dokler	0000-0002-8053-8198	7P9	Data Manager
Maja Brajnik	0000-0002-7420-1388	7P9	Data Manager
Iseult Lynch	0000-0003-4250-4584	UoB	Data Manager Principal Investigator
Steffi Friedrichs	0000-0002-7276-892X	AIST	Principal Investigator Project Administrator

0.2 Description

0.2.1 Objective of the project

MACRAMÉ aims to widen the development of harmonised test guidelines (TGs) and guidance documents (GDs) and standards to market-relevant Advanced Materials in their complex product matrices (AdMa@CMs). This is achieved by defining the R&I Strategy through life-

cycle assessment for five market-relevant industrial MACRAMÉ Use-Cases. By applying, combining and evaluating both established and novel inhalation toxicity tests, a tiered approach to toxicity testing is being developed that provides data on state-of-the-art characterised control materials for the MACRAMÉ Control Material Library and will be proposed as topic for harmonisation and (pre-)standardisation projects provided to and further elaborated with relevant regulatory bodies.

0.2.2 Purpose of data generation and re-use

Knowledge and data is being generated as part of the development of (a) a MACRAMÉ Control Material Library (CML), including reference materials and AdMas, product matrices and their composites and (b) MACRAMÉ Validated Protocols, and (c) the MACRAMÉ Use-Cases on five industry-relevant materials (Use-Case 1: Graphene oxide (GO) flakes in drinking-water filters (Material Family: GRM); Use-Case 2: Few-layer graphene (FLG) in battery management systems (BMS) (Material Family: GRM); Use-Case 3: Graphene-related materials (GRM) in car polish consumer sprays (Material Family: GRM); Use-Case 4: Carbon nanotubes (i.e. CNTs) in car-seats (Material Family: carbon fibres); and Use-Case 5: Poly lactic-co-glycolic acid (PLGA) for inhalable antibiotics (Material Family: PLCG)). The resulting robust summary datasets, scientific documents and recommendations for hazard- and risk-assessment methodologies for AdMa@CMs will be provided as input for the harmonisation and (pre-)standardisation projects to be started by standardisation and regulatory bodies (i.e. OECD, VAMAS/CEN/ISO) as follow-ups of the MACRAMÉ project.

0.3 Tags

0.4 Template

This ROMP is based on the Horizon Europe template taken from the **argos DMP** platform with extensions by MACRAMÉ (marked in pink), needed to be applicable to all research outputs.

0.5 FAIR Implementation Profile (FIP) associated with the project

Selection of the FAIR Enabling and Supporting Resources (FERs and FSRs) is based on the general FIP for nanomaterial/nanosafety research developed by the WorldFAIR project. A FIP more specific to the MACRAMÉ project is under development.

0.6 National / funder / sectorial / departmental policies and procedures for data management relevant to research output management

With this ROMP, MACRAMÉ is fully implementing the transition from the Open Data initiative of H2020 to the Open Science initiative of HE especially by extending the management concepts from data to all research outputs including but not limited to study designs, method descriptions, protocols/SOPs and software.

0.7 Support provided

Creation and implementation of the ROMP is supported by the (Meta)Data Shepherding service established as part of the MACRAMÉ project. Additional support is provided by the WorldFAIR project as well as the GO FAIR AdvancedNano Implementation Network.

1 Summary

1.1 Brief description of the described research output

Areas of data collection, generation and curation:

1. MACRAMÉ Control Material Library (CML), including reference materials and AdMas, product matrices and their composites
2. MACRAMÉ Validated Protocols
3. MACRAMÉ Use-Cases on five industry-relevant materials
    a. Use-Case 1: Graphene oxide (GO) flakes in drinking-water filters (Material Family: GRM)
    b. Use-Case 2: Few-layer graphene (FLG) in battery management systems (BMS) (Material Family: GRM)
    c. Use-Case 3: Graphene-related materials (GRM) in car polish consumer sprays (Material Family: GRM)
    d. Use-Case 4: Carbon nanotubes (i.e. CNTs) in car-seats (Material Family: carbon fibres)
    e. Use-Case 5: Poly lactic-co-glycolic acid (PLGA) for inhalable antibiotics (Material Family: PLCG)

1.1.1 What kind of research output are you describing?

This ROMP is meant to give a general description of all research output produced by MACRAMÉ. This includes, besides raw, processed, and robust summary data, outputs like sampling plans, study designs, method specifications, protocols, SOPs, as well as guidelines, reports, training materials and publications and management and FAIRification are harmonised across all these kinds as much as possible. However, each of these outputs requires some specific treatment including e.g. selection of appropriate (Meta)Data structures, exchange file formats and public repositories for long-term storage. These are described either as separate sections for individual output kinds in this ROMP or in more specific future ROMPs if necessary to enhance readability.

1.1.2 Is it physical or digital?

This ROMP covers the digital objects generated or re-used in MACRAMÉ. If required, physical objects produced by the project and managed using specific centralised and standardised resources (e.g. reference material repositories, biobanks) will be addressed in a specific ROMP on physical objects aligned to the ROMP on digital objects presented here as much as possible.

1.1.3 Are you generating or re-using it?

Digital research objects are generated as part of the MACRAMÉ method developments and their application for creating the MACRAMÉ Control Material Library and execution of the MACRAMÉ Use-Cases. These are complemented by data and accompanying metadata mined and curated from public databases and scientific literature.

1.1.4 What is the type of the described research output?

As described above, this ROMP covers all research output from MACRAMÉ. It can be grouped into the following general categories (more specific categories will be added during the continuous updating of this living document), which is used (including numbering) for the further structuring of the answers in this ROMP:

1. Sampling plans
2. Study designs
3. Method specifications
    a. *In vitro*
    b. *Ex vivo*
    c. *In silico*
4. Protocols / SOPs
    a. Material production
    b. Sample preparation
    c. Measurement
    d. Processing/analysis
5. Computational models, software and workflows
6. Data
    a. Raw
    b. Processed
    c. Robust summary
7. Guidelines
8. Reports
9. Training materials
    a. Manuals / tutorials
    b. Videos
    c. Handbooks / reference material
10. Publications
    a. Dissemination material
    b. White papers
    c. Peer-reviewed papers

1.1.5 What is its format?

The variety of research output types listed above clearly shows that it is not possible to cover all of them in one exchange format. Additionally, different MACRAMÉ partners entered the project with very different knowledge management approaches applied in their internal processes and workflows, which have to be considered when creating the recommendations for project-wide data management and public sharing. Therefore, this initial ROMP lists all data formats currently in use and uses the (Meta)Data reported in these formats to develop recommendations for improvement of the data management practices at specific partners, selects appropriate FAIRification tools for each setting, and develops (semi-)automated approaches to translate internal formats into harmonised and interoperable data documentation formats aligned with community and regulatory standards.

1. *Sampling plans*: text documents
2. *Study designs*: instance maps
3. *Method specifications*: text documents, data templates (MODA, CHADA)
4. *Protocols / SOPs*: text documents, electronic lab notebooks
5. *Computational models, software and workflows*:
6. *Data*: customised spreadsheets, data templates (eNanoMapper, NanoFASE, MODA, CHADA), proprietary formats, data serialisation formats (json, yaml)
7. *Guidelines*: text documents
8. *Reports*: text documents

9. *Training materials*: text documents, slides, videos
10. *Publications*: text documents

1.1.6 What is its expected size?

Limited information on expected size is available at the current state. Similar projects had relatively low requirements on data storage capacities. However, the distributed data storage approach adopted by MACRAMÉ will also be able to cover much higher demands if necessary. This can be achieved by e.g. storing large raw data files only locally at the data providers or using data storing solutions specialised for a specific data type (e.g. omics) or for general storage of big data as e.g. provided by the European Open Science Cloud (EOSC).

1.1.7 Why are you collecting/generating or re-using it?

Data is primarily generated and collected for re-use to satisfy the data needs of the MACRAMÉ method development, Control Material Library and Use-Cases. Additionally, the data will be made available for the application in grouping and read-across within the project and externally, and as input for the development of new computational models and integrated approaches to testing and assessment (IATAs). Finally, the robust summary datasets including metadata on study design, methods and protocols will be provided as input for the standardisation and regulatory validation project initiated at the relevant bodies (e.g. OECD, VAMAS, ISO, CEN) as follow-ups to the MACRAMÉ project.

1.1.8 What is its origin / provenance?

Each research output, both generated by MACRAMÉ and collected from third parties, will specifically specify its origin and provenance as part of the set of standardised metadata.

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

Primary users of the data will be the MACRAMÉ consortium partners performing the quantitative hazard, risk and life-cycle assessment (occupational, consumer, environmental) on Use-Case materials. Secondary users will then be developers of computational models, grouping and read-across approaches, and IATAs based on the advanced MACRAMÉ characterisation methods as well as labs (academia and industry) interested in adding the methods to their Safe-and-Sustainable-by-Design (SSbD) portfolio. The data will then also be made available for starting the standardisation and regulatory validation projects for the creation of SOPs for round-robin testing and test guidance and guidance document development.

2 Links Between Outputs

2.1 Publications

2.1.1 Does the described output support any scientific publication?

Due to the recent project start, MACRAMÉ has not yet produced scientific publications supported by the newly generated research outputs. For data collected and mined from databases and literature, the corresponding references are stored as metadata providing clear provenance trails for the extracted information.

2.1.2 Is there a data availability statement provided along with the publication?

Data availability statements will be provided in all future publications resulting from the MACRAMÉ project. These will include unique, persistent identifiers (DOIs or data-specific identifiers), licensing information, clear provenance trails, and references to the data models used if applicable.

2.2 Qualified references to other objects

2.2.1 Does the described output support any other research output?

MACRAMÉ has adopted an approach of providing the different research output types as individual resources. Instead of creating complex (meta)data templates storing information on the methods, protocols and the generated raw and processed data, as e.g. implemented in the NIKC-NanoFASE templates or the MODA/CHADA system, all these components are managed using individual tools customised and optimised for each type (e.g. electronic lab notebooks, method-specific data formats and databases, version control systems). However, this results in the fact that many research outputs are needed to provide all information a specific piece of evidence or conclusion is based on. Managing all research output using the harmonised approach described in this ROMP has the advantage that all different types can be handled individually and, at the same time, outputs supporting each other can be linked together giving full access to all information relevant for e.g. a specific method or a MACRAMÉ Use-Case. For example, data include references to the methods and protocols used to generate it as metadata and methods can list all dataset used for testing and validation.

2.2.2 Is the research output integrated into a system of qualified references cross-linking outputs?

The concept of individual resources has the advantage of access to optimal data management tools for each type. However, such a distributed system is putting more demands on keeping track of all the resources supporting each other and consistency of the resource references since losing links between the resources would destroy data completeness, understandability, interpretability, and, thus, trust and ultimately re-usability. MACRAMÉ has established an advanced system for research output cross-linking composed out of two services, the MACRAMÉ Registry and the instance map tool. The first assigns a unique, even if only internal identifier to each research output. Additionally, basic metadata on data provenance and accessibility, references to Use-Cases and project partners are provided. The second service is then providing ways to build further cross-links between the research outputs based on the unique identifiers, e.g. linking protocols to the generated data, and to visualise these as instance maps representing life-cycle stages of the materials and the experimental workflow of production and assessment. For public sharing and long-term storage, all relevant research objects e.g. supporting a scientific publication can be extracted from the internal management solution by:

- Replacing the internal identifiers automatically by global and persistent identifiers from authorities like DataCite/Zendo (DOIs), European Registry for Materials (ERM), ORCID, Research Organisation Registry (ROR) or public data/protocol repositories;
- Packing all resources specified in one instance map into a data package following the frictionless data or RO-Crate specifications either as (Meta)Data files or as links to other public resources; and
- Publicly sharing of the data packages in FAIR data storing solutions.

In this way, the resulting self-contained datasets can be fully integrated into the European materials ecosystem of qualified references currently envisioned by the Advanced Materials Initiative 2030 with some components already under development in projects from the EU NanoSafety Cluster, the European Materials Modelling Council and Industry Commons.

2.3 Pre-existing data

2.3.1 Are you using any pre-existing research output?

Pre-existing research outputs will be re-used in the form of existing method descriptions, protocols and SOPs as well as data mined and curated from public databases and scientific literature.

2.3.2 Is the pre-existing research output handled according to this ROMP including additional FAIRification if necessary? If not, provide links to DMP describing the treatment of these pre-existing research outputs.

The pre-existing data will be handled in the same way as newly generated data as described in this ROMP. Additional data management steps will include integration into the MACRAMÉ Registry, provision of data provenance trails, and harmonisation of data models and transfer formats. Additionally FAIRification steps will be described here if they become necessary for specific pre-existing resources.

3 Quality control, FAIR Practices and openness

3.1 Making research outputs findable, including provisions for metadata

3.1.1. What type(s) of persistent identifier(s) are used for the described research output?

MACRAMÉ is using an internal set of identifiers, which are unique within the project and can be translated into globally unique and persistent identifiers. In this way, the different (future) research outputs can be clearly identified from the planning phase on and global uniqueness and also indexing in the relevant identifier services is achieved for the final versions of the outputs meant for public sharing and long-term storage. Persistent identifiers are currently available for:

1. Sampling plans:
2. Study designs:
3. Method specifications: DOI (Zenodo)
4. Protocols / SOPs: DOI (Zenodo, DataCite)
5. Computational models, software and workflows: Github
6. Data: DOI (Zenodo)
    a. Materials: ERM, NInChI
    b. Chemicals: CAS, InChI, (CAS)
    c. Provider: ORCID
    d. Institutions: ROR
    e. Projects: DOI (EU)
7. Guidelines, reports, training materials: DOI, TeSS
8. Publications: DOI (from publisher)

3.1.1.1 Are components of the research output representing levels of granularity (software modules, experimental steps, materials) assigned distinct identifiers?

Study designs, protocols, and materials/chemicals are referred to by individual identifiers in the internal system. These can thus be mapped to global identifiers. The system also encourages splitting the experimental procedure into multiple protocols, e.g. for sample preparation, measurement, processing, all with their own identifiers. Splitting into even smaller parts (individual protocol steps) is in principle also possible but currently not envisioned. This decision will be periodically reviewed.

3.1.1.2 Are different versions of the method descriptions, protocols, software assigned distinct identifiers?

Development of all research output, which might exist in different versions, are managed in solutions with automatic version control (github, Google documents). Stable versions used to generate results for specific studies will be specifically marked (named versions) and will be provided with a distinct identifier.

3.1.2 Will you provide metadata for the described research output? What metadata will be created?

All research outputs will be accompanied by metadata, which will be standardised and made richer over the time of the project. At the current stage, the following high-level metadata fields are mandatory for all research output:

1. Unique internal identifier
2. Name
3. Type of resource (e.g. Material)
4. Status (e.g. Scheduled)
5. Resource link (access path)
6. Short description
7. Contributors and their roles
8. Licence

3.1.2.1 What disciplinary or general standards will be followed?

The final metadata schema of MACRAMÉ will be composed out of archetypes, describing specific aspects like contributors, publications, (meta)data schema, and method-specific metadata. These archetypes will be constructed following existing standards. For example, contributors, institutions and publications will use the DataCite and Dublin Core specifications. For disciplinary components, standards are currently under development or revision (MODA/CHADA, eNanoMapper-based templates) and will be integrated when available.

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

The new methods developed in MACRAMÉ will require method-specific metadata to fulfil minimum reporting requirements. These are currently under development based on the expertise from earlier projects and the first results provided by the MACRAMÉ partners. They will be based on existing minimal reporting guidelines and regulatory requirements but will provide the flexibility to report metadata specific to the method. To guarantee interoperability to the highest possible extent, the metadata models/schemas implemented in the reporting

guidelines will be provided as high-level metadata to the research output (data but also structured information from e.g. protocols, study designs).

3.1.3 Will search keywords be provided in the metadata to optimise the possibility for discovery and then potential re-use?

Provision of structured (meta)data with well defined (meta)data schemas can be used for data discovery for use in the MACRAMÉ and potential reuse using advanced searching/browsing features. These will heavily rely on the unique identifiers for materials, methods and protocols as well as semantic annotated metadata. Additional full text search will provide additional means to find relevant research outputs. If this does not satisfy all needed for information discovery, search keywords will be additionally provided.

3.1.4 Will metadata be offered in such a way that it can be harvested and indexed?

Metadata is available in a structured format and by providing the (meta)data schemas as high-level metadata, options on how to perform harvesting and indexing will be provided as part of the metadata accessible via the application programming interfaces (APIs) of the Registry and potentially of the long-term storage solutions.

3.2 Making research outputs accessible

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

The research output is currently managed and stored in the internal MACRAMÉ Registry and different data storage solutions. Options for long-term storage are currently evaluated and selected based on their fitness for the specific output and their FAIRness.

3.2.1.2 Is the selected repository a trusted source?

Long-term storage solutions will be selected based on community recommendations at the point of time, the output is prepared to the high MACRAMÉ FAIR standards and ready to be publicly released. This selection will only consider trusted sources.

3.2.1.4 Are appropriate arrangements made with the repository(ies) where the described dataset will be deposited

Arrangement will be made when the selection of the long-term storage solution is finalised.

3.2.1.5 Does the repository(ies) assign research outputs with persistent identifiers?

Only solutions providing persistent identifiers will be considered in the selection process.

3.2.1.7 Does the repository support versioning?

Only solutions providing versioning when needed will be considered.

3.2.2 Data

This section will be complemented whenever new research outputs become available since the answers need to be specific to these outputs. Currently, only general aspects of the internal data management system (MACRAMÉ Registry and instance maps) are given.

3.2.2.1 What is the described research output title?

"MACRAMÉ (Meta)Data". Additional and more specific titles will be added during integration of specific research outputs.

3.2.2.2 How is the research output shared? Specify reasons for the type of sharing selected (fully open, restricted, confidential) and embargo period, if applicable, separating legal and contractual reasons from intentional restrictions.

Currently, all research output is only internally shared via the MACRAMÉ Registry. This is the case since none of the outputs have already matured to their final version. When this status is reached for a specific output, sharing decisions are made on a case-by-case basis with preference for fully open sharing and licences allowing reuse in most situations.

3.2.2.4 Will the research output be accessible through a free and standardised access protocol?

Final, long-term access will be provided using existing web-based FAIR data solutions with free and standardised access protocols (http, ftp and specific API endpoints).

3.2.2.5 Are there any methods or tools required to access the research output?

All outputs will be provided in a way not requiring any specific methods or tools other than standard applications (e.g. pdf). However, if relevant, raw data will be provided in proprietary formats since they offer reuse in advanced, data-type-specific analysis software.

3.2.2.8 Is the described research output supported by a data/research output access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

The MACRAMÉ General Assembly will function as the research output access committee guaranteeing that no confidential data is shared unauthorised. Since no sensitive personal data is meant to be collected in MACRAMÉ, an additional external committee is deemed non-essential.

3.2.2.9 Please specify how the research output will be accessed during and after the project ends especially if restrictions are in use

TBD

3.2.2.10 Please specify how long after the project has ended the research output will be made accessible for

TBD on a case-by-case basis.

3.2.2.11 How will the identity of the person accessing the data be ascertained?

The MACRAMÉ Registry, instance map tool and the Google shared drive used for temporal data storage are all protected by state-of-the-art authentication and authorisation management and accessible only after logging in using a personal account. In this way, ascertaining and protocolling of the identity of persons accessing research outputs is assured.

3.2.3 Metadata

3.2.3.1. Will you provide metadata even if the described research output cannot be openly shared?

Metadata for all research outputs are collected independent of their final usage and planned sharing options. This will also include information on how to get access to the output and the person responsible to handle all requests.

3.2.3.2. Under which licence will metadata be provided? Specify reasons for selecting the licence separating legal and contractual reasons from intentional restrictions.

Metadata will be shared under the Creative Commons Attribution 4.0 License (CC-BY-4.0) if not prohibited by legal or contractual reasons. If such cases become relevant in the future, the reasons will be specified here.

3.2.3.3. Do metadata provide information about how to access the described research output?

Information on the access routes and the person responsible for handling all requests will be provided as metadata to every research output. Currently, this is managed by the MACRAMÉ registry providing links to the resources and responsible partners. This will be replaced with information on the long-term storage solutions once the research output is publicly shared.

3.2.3.4. Do metadata provide a full description of the data model used for the research output (including input and output formats) or a link to such a description?

It is anticipated that all research outputs provide the underlying (meta)data model as part of their high-level data documentation. This will be established as part of the additional FAIRification following the currently ongoing evaluation of the reporting and management approaches at the individual MACRAMÉ partners.

3.2.3.5. Will metadata remain available after the dataset / output is no longer available?

Indexing of the metadata in standard FAIR Supporting Resources like Zenodo is planned, which will guarantee availability even after the output is no longer available.

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

Controlled vocabularies will be used whenever available. On one hand, the DataModel Ontology and/or the Information Artefact Ontology will be used for the high-level data documentation including the description and semantic annotation of the (meta)data model. On the other hand, for the low-level annotation of method-specific metadata, different ontologies like the eNM ontology and the EMMO as well as multiple chemical and biological ontologies are available. However, it is expected that these will not cover all relevant aspects and MACRAMÉ will collaborate with other projects to increase the ontological coverage.

3.3.2 If you created the vocabulary, where can it be found?

Terminology resulting from MACRAMÉ's ontology work will be integrated into existing ontologies and will therefore be available from the services providing these ontologies (e.g. BioPortal).

3.3.3 Have you applied a standard schema for your (meta)data?

As described above, harmonisation has been started with standardisation of high-level metadata reusing (parts of) existing standards (DataCite, Dublin Core). This will be complemented by schemas for the method-specific (meta)data documented as part of the high-level metadata. These are based on existing standards and minimum reporting guidelines and the enhanced versions created by MACRAMÉ (in intensive collaboration and alignment with other projects) will be proposed for standardisation.

3.3.5 What is the methodology followed?

The following concepts, partly already described in previous sections, define the MACRAMÉ methodology:

- Individual management of research outputs according to their types allows optimal selection of tools for curation, documentation and sharing (e.g. electronic lab notebooks for protocols).
- Separation of high- (biographical metadata, access options, licences, documentation of data model) and low-level (method-specific metadata) data documentation provides interoperability and computer-actionability for different applications (data discovery vs. data integration into computational workflows)
- Metadata is structured into sections described by archetypes based on existing standards (e.g. DataCite for biographical metadata)
- FAIRification will be continuously improved over the runtime of the project by aligning with and integrating emerging concepts and standards and proposing new or improved standards when necessary.

3.3.6 What community-endorsed interoperability best practices are followed?

MACRAMÉ partners have been / are engaged in multiple activities defining interoperability best practices, which form the basis for this ROMP. This has started with the OpenRiskNet and NanoCommons projects defining FAIRification and FAIRness assessment approaches as well as the roles responsible for specific tasks and is now continued in the WorldFAIR project (cross-domain interoperability), PARC and GO FAIR AdvancedNano Implementation Network. Additionally, collaboration with the European Materials Modelling Council and the OntoCommons project have been started.

3.3.6.1 What domain-relevant community standards are used for reading, writing and exchanging data?

MODA/CHADA formats, eNanoMapper-based and NIKC/NanoFASE data curation templates, ISA-TAB-nano format as well as the OECD harmonised templates have been proposed as community standards for data exchange. However, all these formats are currently under evaluation by many projects to increase their FAIRness and computer-actionability. Therefore, MACRAMÉ decided to postpone the decision on using a specific standard and to use first project-internal, highly structured reporting formats with clearly defined (meta)data models provided as high-level metadata. Mapping of these internal (meta)data models to the models

implemented in the improved standard file formats will allow automatic translation into these new formats as soon as they have been endorsed by the community.

3.3.6.2 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? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Ontology and vocabulary development in MACRAMÉ will only be performed, if at all, in intensive collaboration with other projects (e.g. OntoCommons, ELIXIR), which have established solutions for ontology publishing, reusing and mapping.

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your research output?

A licence is assigned to each individual MACRAMÉ research output individually by the producer. They can choose between different Creative Commons licences.

3.4.1.1 How is the ownership of the collected data arranged?

Ownership of the generated research outputs is fully in the control of the producer. This includes licensing as well as selection of the sharing and long-term-storing options. Guidance will be given by the MACRAMÉ (Meta)Data Shepherd.

3.4.1.2 Is there a clear procedure for the decision process selecting the licence, sharing option and embargo periods?

The MACRAMÉ procedure for licensing and sharing is applied to each individual research output and contractual obligations are specified in the MACRAMÉ Consortium Agreement. A handbook to guide the decision process for selecting output-specific licences, sharing options and embargo periods is under development.

3.4.2 What reusability and / or reproducibility methods are followed?

Clear licences, rich metadata and especially the linking of all resources (study design, method description, protocols for sample preparation, measurement and processing, and raw and processed data) supporting each other via persistent identifiers or even combining all these resources into one self-contained data packages is dramatically increasing repeatability, reusability and reproducibility.

3.4.3 Will you provide the described research output in the public domain?

MACRAMÉ is following the policy of "as open as possible, as closed as necessary". All research output generated for the internal validation of the new MACRAMÉ methods will be publicly shared to reduce the need for replicating the work as part of the subsequent standardisation and regulatory validation efforts.

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

The MACRAMÉ Control Material Library and the MACRAMÉ Use-Cases including all generated and collected data/knowledge are meant to provide the starting point for standardisation and regulatory validation of the MACRAMÉ as well as third-party methods. Additionally, MACRAMÉ is showing how the generation data can be used for grouping and read-across, which will

inspire new computation prediction methods based on MACRAMÉ data as well as new data generated using the MACRAMÉ methods. Both guarantee that the MACRAMÉ research outputs will be reused in the future.

3.4.5 Is provenance well documented?

As already described above, provenance trails are important metadata associated with each research output, both generated in MACRAMÉ and reused in MACRAMÉ coming from database and text mining activities.

3.4.6 What documented procedures for quality assurance do you have in place?

All MACRAMÉ protocols and SOPs will provide detailed descriptions of quality control and assurance procedures applied and recommended when reusing them. This follows the high standards of Good Practices (GxP) and requirements from bodies like OECD and ECHA.

3.4.7 How will you provide documentation needed to validate data analysis and facilitate re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, training sets, unit tests, etc.)?

Protocols and SOPs are a central component of the MACRAMÉ research outputs and give detailed descriptions of all steps including but not limited to sample preparation, measurement, and processing / data cleaning, all with corresponding descriptions of quality control and assurance measures. Rich, structured metadata will be provided as part of the protocols, SOPs and data resources providing important experimental settings and parameters. Software used and customised as part of the MACRAMÉ project will be documented, shared as open source if possible, and accompanied by training and test sets. MACRAMÉ will provide functionality to pack all this information into self-containing data packages, which will provide all necessary information to facilitate validation and reuse.

3.4.8 Will you monitor data integrity once it has been collected?

During the generation of the self-contained data packages, integrity of the information is thoroughly checked. These will be used to provide the data to the long-term storage facility and indexing services and can be revisited if issues with integrity become evident (e.g. if a long-term-storage service is going out of business). They are also the primary place if further data curation is needed to improve the completeness of the data or correct errors identified during reuse of the data. All services, to which the data is uploaded or in which it is indexed, can then reinsert the data clearly marking the changes or providing it as a new version.

4 Allocation of Resources

4.1 Collaboration platform

4.1.1 Are you using a shared working space to work with your data?

The MACRAMÉ Collaboration Platform, MACRAMÉ Registry, instance map tool and Google shared drive form the basis for the shared working environment. This can be extended with additional shared services like electronic lab notebooks, version control systems (e.g. github), and database systems if required.

4.1.2. Are you using a centralised or distributed data storage system? How are the resources connected in the distributed system?

MACRAMÉ is using a mixture of centralised and distributed data storage. The MACRAMÉ Registry is the central place providing information on existing research outputs (both final and under development) and how this can be accessed. The actual resources are then stored in a distributed system allowing integration of optimal solutions for the different research output types and customisation reflecting the different data management approaches at the different partners.

4.1.3. Are you using centralised and harmonised approaches for developing or modifying the workflow for data processing and analysis?

No centralised approaches are used at the moment. This is necessary to not slow down the method development and sharing of preliminary data for the Use-Cases. The provided data and other research output is currently being analysed to find similarities in the used approaches and start harmonisation of the data reporting including processing and analysis. This analysis will then be used to generate centralised services whenever possible.

4.1.4 How are you monitoring progress?

The combination of the MACRAMÉ Registry and the instance maps are used to monitor progress. Instance maps are used to plan the needed experiments and other data collection activities. The resources (research outputs) documenting these activities are then accessible in the MACRAMÉ Registry by direct links from the instance maps and are clearly showing the status of completion.

4.2 Data management and FAIRification costs

4.2.1 What is the cost of making the described output FAIR?

MACRAMÉ approach is to base data management and FAIRification on the internal processes established at the partners and, in this way, minimise additional effort. However, to also show the importance of these tasks and especially making the research outputs interoperable, 15% of the budget reserved for experimental work has been assigned to research output documentation.

4.2.2 How is this cost covered?

Budget for the experimental work in MACRAMÉ is split between WP2, WP3, and WP4 with 15% of the budget of data producers reserved for data management in WP3. Additionally, budget for data management infrastructure provision and the data shepherding service are allocated.

4.3 Identify the people who are responsible and their role(s) in the management of the described output

- Data reporting, curation, and FAIRification: all partners of WP2, WP3 and WP4
- Data quality control: all partners of WP2, WP4 and WP5
- Data completeness evaluation: all partners of WP3 and WP5
- Data infrastructure provision, data transformation: 7P9-SI, 7P9-DE
- Data shepherding service: 7P9-DE, 7P9-SI

5 Security

5.1 What security measures are followed?

Strong data security measures are applied in MACRAMÉ. All collaborative services and the data stored in them are secured by state-of-the-art authentication and authorisation mechanisms. Hosting on European servers is also guaranteed. Data transfer is using secure and encrypted communication (https). Additionally, the distributed knowledge management approach allows local storage of highly sensitive information. Finally, backup solutions are implemented for disaster recovery.

5.2 What conditions do the security measures meet?

- Authentication and authorisation mechanisms
- Monitoring of access
- Privacy protection and clear conditions of use
- Encrypted data transfer
- Backup solutions
- Constant application of security updates

5.3 How will you preserve the described research output in the long term?

Long-term storage will be using existing and emerging standard solutions (e.g. NanoCommons KB, Zenodo, ELIXIR, and EOSC). Negotiations with the corresponding services providers are ongoing.

6 Ethical Aspects

6.1 Are there any ethical or legal issues that can have an impact on sharing the described research output?

Full ethical approval has and will be obtained for all experimental work performed within MACRAMÉ. Sharing of research outputs is not expected to add any additional ethical or legal issues.

6.2 Does the described research output contain sensitive information?

Confidential data might be included in research outputs supporting the industrial Use-Cases. If and when this data can be shared and with whom will be decided by the industry partners on a case-by-case basis.

6.3 Does the described research output contain personal data?

Personal data on the providers of the research outputs is collected as part of the data provenance trails. This includes names, addresses, and email. These will be handled according to EU's General Data Protection Regulation (GDPR).

7 Other Issues

7.1. Do you make use of other procedures for data management?