Horizon Europe

Data Management Plan Template

Version 1.0
05 May 2021

HISTORY OF CHANGES		
Version	**Publication date**	**Changes**
1.0	05.05.2021	▪ Initial version
		▪

Action Number: [101046909]

Action Acronym: REMAP
Action title: Reusable mask patterning

Date: [31-7-2022]

DMP version: 1.1

1. Data Summary

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

Data re-use is a practice endorsed by REMAP consortium, because it is in line with REMAP's green philosophy. Whenever possible, existing data will be re-used from published articles belonging to the consortium members and from the literature for the pursual of T1.1, T1.2, T2.1, T3.3, T4.4, and T5.4. Proper citation will be given to such data, in accordance with the European code for research integrity. The pursual of T1.3, T2.2, T2.3 may rely on the re-use of data in the form of confidential protocols and procedures generated by the consortium partners before the start of REMAP. Data that are not disaggregated by gender and other categories relevant to REMAP will be disaggregated, where possible, or collected ex-novo.

What types and formats of data will the project generate or re-use?

REMAP will re-use or generate data of different types, depending on the nature of each deliverable. Namely: experimental, observational, textual, and numerical data. It is estimated that, on average, each deliverable's data composition will be approximately as follows: 35 % experimental, 15 % observational, 15 % textual, and 35 % numerical. Given the wide range of experimental and theoretical methods employed within REMAP, the number of data formats is high. The following extensions will be employed: .pdf, .txt, .opj, .cif, .cpp, .f90, .vtu, .mph, .cdb, .dxf, .h5p, .png, .mp3, .mp4. Additional formats may arise during the implementation, and this list will be updated accordingly.

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

The purpose of data generation and re-use within REMAP is directly related to the pursual of the project objectives. This is reflected in the data types and format composition estimated in the previous point. Namely, approximately 35 % of the data will be experimental and will relate mainly to the achievement of D1.1, D1.2, D2.2, D2.3, D3.1, and D3.3. Approximately 15 % of the data will be observational and will relate mostly to the achievement of D4.4, D4.8 and D4.9, and to the pursual of T5.4. Approximately 15 % of the data will be textual and will relate mainly to the achievement of D4.1, D4.2, D4.3, D4.5, D4.6, D4.7, D4.9, and D5.1 to D5.11, and to the pursual of T5.4. Approximately 35 % of the data will be numerical and will relate mainly to the achievement of D1.3, D2.1, D3.2, and to the pursual of T5.4. Where appropriate, data generated will be disaggregated by gender and other categories relevant to REMAP.

What is the expected size of the data that you intend to generate or re-use?

At this stage of the project, the estimated size of the data generated and re-used is in the order of 1.0 to 1.5 TB in total, of which around 300 GB correspond to data re-use.

What is the origin/provenance of the data, either generated or re-used?

Re-used data originates from the literature and from secure servers belonging to the consortium members. The generated data will originate from various sources, in line with the nature of said data. Namely, approximately 35 % of the data will originate from experimental activities performed mainly to achieve D1.1, D1.2, D2.2, D2.3, D3.1, and D3.3. Approximately 15 % of the data will originate from observations performed mainly to achieve D4.4, D4.8 and D4.9, and to pursue T5.4. Approximately 15 % of the data will originate from the consortium members' own intellect, e.g. during the evaluation of REMAP's technology benchmarking, during the periodic meetings of REMAP's governing board and the technical and scientific meetings (D4.1, D4.2, D4.3, D4.5, D4.6, D4.7, D4.9, and D5.1 to D5.11), but also for the pursual of T5.4. Approximately 35 % of the data will originate from theoretical calculations and models performed mainly to achieve D1.3, D2.1, D3.2, and to pursue T5.4. Data generated in relation to the evaluation of REMAP's technology benchmarking, during the periodic meetings of REMAP's governing board and the technical and scientific meetings will be disaggregated by gender and other categories relevant to REMAP, when appropriate.

To whom might your data be useful ('data utility'), outside your project?

The data generated in REMAP will find utility by all the targeted scientific readership and audiences foreseen as part of T5.4. Namely, the data should find utility by the readers of various scientific journals such as Chemistry of Materials, J. of Magnetism and Magnetic Materials, Soft Matter, J. Phys. Chem. C, Applied Materials Today, RSC Chemical Communications, Journal of Micromechanics and Microengineering, Scientific Reports, Electrochimica Acta, J. Chemical Education, Angewandte Chemie, Environmental Innovations and Societal Transitions, Progress in PV Res. Appl., Nature Communications, NanoToday, Energy and Environmental Science, and Science. Likewise, the audience of various conferences will have access to the data and may find it useful: International Conference on Magnetic fluids, International Conference on Magnetism (ICM), Computational Structures Tech., Eur. Physical Society, EUChemS, SPIE Advanced Lithography, ElectroChemical Society (ECS), Greenering meeting, IUPAC Conference on Chemical Education, EUCHEMSIL, ManuFuture-EU, EUPVSEC, European Materials Research Society. Furthermore, REMAP's data should find utility by all the targeted communication audiences foreseen as part of T4.1. Namely, the data should find utility by Job seekers, R&D actors, SME actors, Schools, trainees, journalists, Advisory Board members and their networks, EU Circular Economy Stakeholder Platform, NGOs. Local communities, families and youngsters, School kids, Teenagers, YouTube viewers, the European Commission, the University-Industry Innovation Network, Members of the Greenering COST Action 18224, R&D pioneers in business greening, SusNanoFab partners, ManuFuture-EU partners, and Venture capitalists.

2. FAIR data

2.1. Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

To ease public retrieval, raw data leading to all publications will be appended directly as supplementary information files (S.I.) to the mother articles, with persistent digital object identifiers (doi). However, given that not all of the generated data will constitute scientific publications, all data that is not classified as sensitive will be uploaded to the project website or to the national or institutional repositories such as: HAL(FR), IRIS(IT), HELIX(GR), and ORBI(LU) and will be assigned permalinks. REMAP will also make use of Open Research Europe (ORE) services to make its research freely available.

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

The stated repositories allow for several metadata to be compiled (some even automatically) and be associated with the uploaded material, easing searchability, indexability, visibility and discovery by any interested user worldwide. The metadata included are the standard ones included by the major academic repositories in Europe: Title, Keywords, Author(s), Medium, Date, Place, Abstract, doi, ISBN. REMAP will monitor the linguistic biases in metadata, to promote inclusive language in metadata

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

Yes. Likewise, REMAP will also monitor the linguistic biases in keywords, to promote inclusive language.

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

Metadata supplied for all scientific publications can be automatically harvested and indexed with both proprietary and open-source reference management software. Likewise, metadata embedded into some of the national and institutional repositories (e.g. HAL) can also be harvested and indexed automatically. Those repositories that currently only offer manual indexing (e.g. ORBILU, IRIS) are likely to be upgraded soon.

2.2. Making data accessible

Repository:

Will the data be deposited in a trusted repository?

Yes.

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

Default arrangements are in place between REMAP's academic/research partners and the respective national or institutional repositories. According to such arrangements, the partners are obliged to create data and metadata entries within said repositories and upload the data, unless such data is deemed sensitive.

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Yes. The repository will assign a permalink to the data entry.

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

The data generated within REMAP will be open access as the default option unless such data is deemed sensitive. Data may be classified as sensitive if protection of such data is necessary as a measure to protect patentable intellectual property, in accordance with the Grant Agreement. Sensitive data corresponds to less than 5 % of the estimated data generated within REMAP and shall be published with an embargo period.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Sensitive data subject to embargo will be made available as soon as possible, i.e. as soon as the appropriate IP protection is put into place. This time generally amounts between 12 and 24 months, according to best practices.

Will the data be accessible through a free and standardized access protocol?

The access to the data will be free of charge through the standard protocols established by the publishers and by the administrators of the national and institutional repositories.

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

No restrictions on data use is foreseen, except for the sensitive data. Sensitive data use is restricted inasmuch as the data itself is unavailable during the embargo period mentioned above.

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

Access to data in the form of supplementary information of scientific articles published by REMAP's partners in gold open access is completely free and unconditional to the retrieval of the requester's identity. The same applies to the data stored in national and institutional repositories employed within REMAP, except for ORBILU for which the requester may be asked to disclose his/her name and surname in order to be granted access.

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

REMAP's data access committee is composed of all the members of REMAP's governing board. The approval of access requests to sensitive data is discretional and subject to the evaluation of REMAP's governing board upon optional consultation of experts in data management and protection as the need arises. Additionally, the Coordinator's Data Protection Officer will offer REMAP's consortium full support in managing GDPR related issues, national and institutional policies on data management and protection.

Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

Yes, all metadata will be made openly available under CC0 license. Where appropriate, metadata will contain information to enable a user to access the data itself (e.g. journal name, volume, issue, page and/or article number).

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

There are no reasons why REMAP's data, once made available, should be withdrawn. Hence, the data and the associated metadata will remain available and findable for at least as long as the depositories in which it is located will last (i.e. decades).

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

Whenever possible, the data will be self-explanatory. When this is not possible, a documentation will inform about the data software needed for data access. Such documentation shall be embedded as metadata whenever possible. Whenever possible, the data will be provided in the form of a .txt file to ensure the widest interoperability. The inclusion of relevant software will not be possible unless such software is open-source, in which case a link to the open-source developer will be supplied.

2.3. Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

The attention to data interoperability is relatively new for the research communities to which REMAP belongs, and best practices are in the process of being established. The experimental and theoretical data generated in REMAP will typically consist of series of x-y independent-dependent variables, whereas the observational data will typically consist of categorical variables. Therefore, in all cases the variables will be duly labelled within the data themselves. In some cases, the data will be generated by software codes created ad-hoc that will also be made available and, whenever possible, enriched by comment strings. To the extent deemed appropriate, the associated metadata of all entries will contain a vocabulary attentive to diverse disciplines, with the intent to ensure the widest possible data interoperability.

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?

REMAP endorses open science practices, which entails a proactive and didactic approach towards the dissemination and communication of science including the data upon which it rests. As such, it will be a priority to ensure a proper contextualization of the data entries with adequate metadata information and reference to other sources from REMAP or elsewhere, if needed.

Will your data include qualified references1 to other data (e.g. other data from your project, or datasets from previous research)?

Whenever possible, a rich data contextualization will be supplied in the form of links to other data from REMAP or elsewhere.

2.4. Increase data re-use

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

Such documentation will be supplied as dedicated metadata whenever possible. Alternatively, it will be embedded in the data itself, e.g. as comment strings.

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Yes, the data will be made available on the public domain free of charge. The default licenses will be CCBY and CC0 unless the data is classified as sensitive as a measure to protect patentable intellectual property, in which case it will only be available after the embargo period.

Will the data produced in the project be useable by third parties, in particular after the end of the project?

The data will remain available from the publishers' websites indefinitely or from the national/institutional repositories for at least as long as these repositories will exist. As such, users interested to obtain the data can do so and reuse them under the CCBY or CC0 licenses at least for as long as the repositories will exist well beyond the end of the project.

Will the provenance of the data be thoroughly documented using the appropriate standards?

Such details are needed to contextualise the data and, as such, they will be included in the documentation, as described above.

Describe all relevant data quality assurance processes.

The quality assurance (QA) processes followed to implement REMAP's tasks are reported hereunder and kept updated. The QA principles are designed in line with the best practices of the relevant scientific communities. Therefore, when applicable, reference to appropriate peer-reviewed material is provided for each task, describing the most appropriate methodological approaches to data acquisition and management.

T1.1 Formulation and characterization of aqueous MREs. The preparation of both magnetic nanoparticles (MNPs) and Magneto Rheological Electrolytes (MRE) will start from state of art synthetic approaches that will be modified as to match REMAP's requirements The quality of MNPs and MREs will be checked using well established multi techniques characterization approaches. All the results will be stored on paper and elaborated into reports for partners involved in T1.2 and T1.3, and for the community.

T1.2 Multiscale Modelling of new MRE behaviour. Datasets (results) are produced by well-established techniques i.e., molecular dynamics and Monte Carlo. All simulations are performed for systems that are related to T1.1, T2.2 and T1.3. Metadata produced are plain .txt or .dat files.

T1.3 Formulation and characterization of LMS-based MREs. All LMS-based formulations will follow the quality control process in place at Solvionic. When relevant, an analysis report will be produced, assembling product quality analyses. Physical characterisations will be made, and results will be collected and stored. Physical characterisations will implement a MSDS (material safety data sheet) that will be made available to project partners. Distribution of the data to external partners will be subjected to confidentiality.

T2.1 Finite element simulations and design. Datasets (results) are produced by well-established and tested methodologies of finite elements formulations. All simulations are performed for systems that are relevant to T2.2 and T2.3. Results are exported with established .vtu format. Metadata are produced are plain .txt or .dat files.

T2.2 Cleanroom fabrication and validation of 1st embodiment of magnetic micro-structuring device. The fabrication of the embodiment of magnetic micro-structuring device will be performed in INL's Micro- and Nanofabrication Facility, which is ISO 9001 certified. All process run sheets will be stored and any generated metrology data will be properly documented and stored.

T2.3 Micro-fabrication of 2nd embodiment. The same QA measures described for T2.2 shall apply.

T3.1 Establishment of KPI for MRE masking efficiency. This task will follow the QA practices described by J.R. Basore et al. in Chemical Communications 48, 1009 (2012) doi: 10.1039/C2CC16938J.

T3.2 REMAP demonstration of additive and subtractive patterning. Given that this task represents the core of REMAP's

[1] *A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)*

innovation demonstration, no QA standard can be merely transferred tout-court from the literature. Conversely, the QA standards in place at state-of-the-art nanofabrication facilities (such as that at INL) will serve as a starting guideline; these will be appropriately adapted to the lower resolution range foreseen in REMAP. In any case, all process run sheets will be stored and any generated metrology data will be properly documented and stored.

T3.3 Proof of concept of a micropatterned solar cell. This task will follow the QA practices described by M. O. Reese et al. in Nanostructured Materials for Type III Photovoltaics, Chapter 1: Reliably Measuring the Performance of Emerging Photovoltaic Solar Cells. RSC Nanoscience & Nanotechnology Series (2017), doi: 10.1039/9781782626749-00001.

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

3. Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

It is apparent that the considerations at points 1 to 2 already apply to all the data generated or re-used within REMAP.

4. Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?

The direct cost associated with making findable, accessible and re-usable the data that is part of main texts and supplementary information of scientific articles published under gold open access schemes typically ranges from 1500 to 6000 Eur per article. Such costs for each published submission through the ORE services are 780 Eur, with a maximum 5% increase for 2023 and 2024 prices. Conversely, archiving data in the national/institutional repositories has no direct costs. The costs for making the data interoperable are indirect and are difficult to quantify because they depend heavily on each entry. The indirect costs associated with storing, archiving and securing the data depend on the repository and on the operator.

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

Direct costs associated with making findable, accessible and re-usable the data that is part of main texts and supplementary information of scientific articles published under gold open access schemes are covered by the dedicated project budget. The direct costs associated with the publications through ORE services are covered by the European Commission that also covers a fraction of indirect costs through a four-year framework contract (ending in February 2024). All remaining indirect costs are covered by the project budget under the indirect costs category.

Who will be responsible for data management in your project?

Data management will be the responsibility of REMAP's governing board.

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

As stated at point 2.2, REMAP's consortium believes that there are no reasons why REMAP's data, once made available, should be withdrawn. Hence, the data and the associated metadata will remain findable, accessible, and re-usable, for at least as long as the depositories in which it is located will last (i.e. presumably well beyond the end of the project). A foreseeable exception concerns the case in which data is later deemed inaccurate, in which case it may be withdrawn to prevent its re-use.

5. Data security

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

The data stored in ORE is aligned to Horizon Europe requirements for research data management. The data stored at publishers' websites is secure inasmuch as it can only be accessed by the publisher through internal procedures, once published. The data stored is REMAP's website and in national/institutional repositories are only accessible by the consortium members or by the respective administrators upon insertion of a password.

Observational data classified as sensitive according to the GDPR (e.g. data collected in the framework of T4.1) will be stored in lockable storages at the premises of the coordinator institution and will not be stored in online repositories. Access will be given to

the governing board members and the related task leader. The anonymization of personal data will entail the removal of direct identifiers, the generalization of text variables to reduce identifiability, the anonymization or removal of indirect and contextual data. All personal data will be removed during the transcription phase and replaced by alias. All replacements will be recorded in an anonymization log that will be stored separately from the anonymized data files, whereas the latter may be stored in online repositories and as part of publications, as for other data. The sensitive data, including the anonymization log will be deleted as soon as allowed to ensure methodological reproducibility and, in any case, it shall not be retained longer than necessary for the purposes of the Project, within the limitations of applicable laws.

Will the data be safely stored in trusted repositories for long term preservation and curation?

Yes. ORE platform, the journal publishers' websites and the national/institutional repositories where the data will be stored are deemed trusted to ensure the safe and long-term preservation and curation of the data.

6. Ethics

Are there, or could there be, any ethics 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 project uses human participants and collects personal data in the framework of T4.1. Therefore, there are issues regarding collection and storage of personal data as well as voluntary participation. These issues are discussed in the ethics section. Briefly, the project will involve citizens who will be informed that their personal data will be gathered and processed in an entirely anonymous way. The informed consent/assent form will contain the information about the nature of data that will be collected (age, gender, cultural interests, school degree and education), how they will be stored, processed and protected and a detailed description on how the data will be used and for which purposes. On all these matters, REMAP will strictly adhere to the EU legal framework meeting GDPR standards. The collection of data itself will be pursued in an inclusive manner (ensuring that there are no systematic biases and that no vulnerable groups are systematically excluded, for example people with visual disability). Inclusive language will be used when collecting, analysing and reporting collected data. All forms will use an inclusive language and data will be collected to allow disaggregation.

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

Yes, these details will be included in the consent/assent forms, as described at the previous point.

7. Other issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

No other procedure applies nor shall be followed.