Data Management Policy

1. General Principles

1.1 The objectives of the Data Management Policy are to provide guidelines to ECCSEL ERIC community (Operation Centre, Members, Access Providers and Users) for ensuring that ECCSEL ERIC's data and information assets are managed consistently and used properly.

1.2 The DMP is namely intending to follow and help implementing in ECCSEL ERIC the following principles:

-FAIR principles: set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable.

-Open science: approach to make scientific research (including publications, data, physical samples, and software) and its dissemination accessible to all levels of an inquiring society, amateur or professional.

-GDPR: regulation aiming to create more consistent protection of consumer and personal data across EU nations.

FAIR principles, when fully respected, can be considered as very constraining and expensive. For that reason, the present and upcoming DMP versions will intend to provide guidelines to assist ECCSEL ERIC in progressively reaching a full implementation of those principles. The implementation steps will be selected by considering ECCSEL ERIC capacity to reasonably carry them out.

1.3 Acceptance of the DMP by Users is mandatory before any access and will be notified in the corresponding Access Agreement.

1.4 Acceptance of the DMP by the Access Providers is mandatory and is required for each SLA.

1.5 Deliberate infringements of the DMP may lead to denial of further Access and/or contracting with ECCSEL ERIC.

1.6 If Data contains personal data, the European legislation GDPR will be applicable to those portions. The GDPR compliance and implementation is considered as being part of the DMP Data management.

2. Distinction between public and external data

2.1 Public Data will be fully submitted to DMP guidelines while External Data will be relieved from the obligation to follow DMP guidelines 4.2 to 4.6. Beside the collection stage (guideline 4.1), Data

Owner(s) will independently be responsible of their External Data fate namely in terms of security, integrity, confidentiality and GDPR compliance.

The distinction between Public and External Data will be done by the access provider according to the following rules:

2.2 Research that has been entirely publicly funded will be necessarily considered as public data. However, exception will be granted if the access provider can deliver a data owner statement certifying that those data belong to the following exception list:

> 2.2.1 Data involved in national/international security matters and considered as classified according to the law(s) and treaty(ies) of the country(ies) involved in the research.
>
> 2.2.2 Data being used to generate a patent and any IP based or commercial products. However, the user has to be authorized by the public funding party(ies) to undertake such data valorization. It includes data that may fall under the scope of the Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets).
>
> 2.2.3 Data being used for an ongoing ECCSEL ERIC publication. This process goes from the drafting process to the publication itself, however the embargo cannot last more than 5 years after the data were produced.
>
> 2.2.4 Scientific data that are considered as more suitably reattached to another RI or data management network. Such data might be rightly defined as public data (sic) but will be considered as external data in the present DMP. A precise list of the data categories that fall under that scope will be provided and updated by the OC.

2.3 It is the data owner duty and responsibility to certify if its publicly funded data truly belong to at least one of section 2.2 categories and to notify any status change to the access provider. The data falling in these categories will be considered as External for the time period where the conditions are met. In case of status change and if they do not belong anymore to one of these categories, they will then be considered as Public Data and will have to be treated as such and comply with DMP guidelines.

2.4 All other research (partially publicly funded or entirely funded by private funds) will have the possibility to be declared by their owner as External Data if required. If not, the data will be considered as Public Data by default.

3. Ownership and Responsibilities

3.1 The Data ownership of the Data produced during Access will be précised in the corresponding Access Agreement, and if not, the Access Provider will be considered as the sole Data Owner. Regarding Data produced outside Access, ECCSEL will be considered as the sole Data Owner.

3.2 If not agreed differently before (see section 3-3), the Data Owner(s) will be considered as jointly and fully responsible for the Data management in accordance with the DMP.

3.3 Beside the ownership, a possible agreement between the Access Provider and the Users can be established in the Access Agreement about any specific sharing of the Data management responsibility. This will be done at their discretion, but the Access Provider will have to communicate the sharing conditions as well as their status (Public or External Data) to ECCSEL OC.

3.4 Considering Accesses, the Access Provider will be the contact point for the Data Owner and will provide the guidelines and methods to fulfil DMP requirements.

3.5 The Access Provider will be responsible for checking the implementation of the DMP, for collecting the final Public Data addresses and identifications and for providing them to the ECCSEL ERIC OC.

3.6 ECCSEL ERIC OC will be responsible for keeping track of all the Public Data that will be produced and will manage the frontend website and applications that will serve to structure and display the access to ECCSEL ERIC Database.

3.7 ECCSEL RICC will be responsible for establishing and updating the DMP guidelines.

3.8 As stated in section 3-3, the Data Owner(s) might contract out the Access Provider or a third party as a Data Manager to manage and process the Data as described in the DMP guidelines. The sharing can be done in all or part of the Data as well as all or part of the Data life cycle management. If such delegations are not specified in the Access Agreement, the Data Owner will be considered by default as the Data Manager.

4. Guidelines

The Data Managers are responsible of the implementation of the guidelines.

4.1 Data Acquisition

4.1.1 The Data should be collected and labelled in a clear and consistent way. The labelling should clearly state the status of the data and the conditions of their acquisition. It should at least precise for each Data the:

Access Number of the Access Agreement, for Data not provided by Access, the data can be referenced by the minutes reference where the initiation of the acquisition have been decided.

Sample number (Access Provider numbering).

Data status: External or Public Data, personal or non-personal Data.

Time of Data production by measurement or calculation (dd/mm/YYYY).

Origins if associated to natural samples (referenced type or description, geographical location) or location in an experimental site (geographical location and precise site inner location).

All parameters/conditions of acquisition and their corresponding standards units or references that would be required to reproduce the experiment or calculation. For example, this may include but not limited to experimental time, temperature, pressure, solution and material compositions, databases, equipment, programmes and standardized procedures that were employed.

4.1.2 Personal Data will require the procurement of the explicit and positive consent of the corresponding individuals prior to their acquisition. By explicit, we mean that the content and usage of the Data collected will be entirely and precisely presented together with a contact reference for further inquiries. By positive, we mean that the procurement should be actively

validated by the individuals and imply therefore a dedicated action from them (signature, digital signature, ticking action).

4.1.3 Pre-processing steps and format

The Data might be primarily processed in order to improve their quality, correct biases, enhance their readability and to convert them into a digital format. It expected that, at the end of this process, the Data can be digitally transmitted and directly used by a potential user.

The Data will also be evaluated and selected in regards of their final wanted usage. This step will namely allow discarding incomplete, low quality, perturbed or biased Data that would not be suitable for the intended study or project purposes.

4.1.4 Transmission

Pre-processed and formatted Data, as well as potential Personal Data procurements, will be transmitted to the Data Manager in charge of the Data Curation or to the Data Owner if the Data Manager is unknown.

4.2 Data Curation

Annotation: The pre-processed and formatted Data will be transcribed using RDF Triple architecture with Xtensible Markup Language (XML Version 1.0, UTF-8) as metalanguage (RDF/XML) and OWL Lite as ontology language.

Core model and Categorization: The Core Scientific Metadata Model (CSMD, V 4.0) core model will be used as base for ECCSEL metadata model (http://icatproject-contrib.github.io/CSMD/csmd-4.0.html). This model defines a hierarchical model of the structure of scientific research around studies and investigations, with their associated information, and also a generic model of the organisation of data sets into collections and files. The model will be adapted to the specific field of CCUS. The Data will be assigned to Data categories according "ECCSEL Ontology Rules and Taxonomy Guide" file. The corresponding file will be provided and updated by the OC to detail and make evolve those specifications. It will describe precisely and with ECCSEL vocabulary the correspondence between CSMD model and the collected Data.

The guide will namely contain:

- information about hierarchical correspondence:

    E.g: Study is referring to a specific contract and will be described by the Access Number of the Access Agreement in the "Study description" field.

- information about the use of a common categorization for ECCSEL investigations, with namely:

    - a list of application fields for CCUS with their respective definitions:

        E.g: Capture, Transport, Use, Storage and Global approaches…

    - a list of sub-categories for each field with their respective definitions.

        E.g: Transport: Flow regime, Leakage, Corrosion, Cost studies…

    - a list of run type with their respective definitions associated parameters.

E.g: Transport: Leakage: Leakage IR detection…

Both application fields and sub-categories will be used to characterize the type of investigation performed and will serve to reference “Investigation type” in CSMD.

- information about the required parameters descriptions for each run detailing namely the units to be used.

- indications about the Data sub-categories that should be entirely or partially excluded from Public Data (see section 2-2-4).

The guide is intended to be built in a collaborative way. Thus, the different Data Owners and Data Managers will be able to make propositions to modify and improve the guide. However, the proposal will have first to be approved by the OC (with potentially RICC assistance) before being implemented.

- Review and validation

Once the Data have been transcribed, a review will be accomplished in order to check that there are consistent with ECCSEL Ontology Rules and Taxonomy Guide (EORTG) requirements.

4.3 Data storage:

The storage of the Data into Databases will be performed by ensuring that:

- the Database storage is secured in a way it can reasonably sustain any usual trouble like power failure and establish adequate threats prevention. Among the different measures that could be taken, a separate and independent copy of the Data is mandatory.

- measures have been taken to kept the storage perpetual.

- access to sensitive Data, namely regarding GDPR compliance, is controlled and monitored.

4.4 Data Access and Use:

4.5.1 The Data management system (DMS) used for ECCSEL community Databases will be an open and free DMS that works on SQL. A front end will be displayed at ECCSEL scale with a portal that will ensure the inter-connexion of each Database as well as essential tools to target and find Data.

4.5.2 To fulfil its goal of giving a full access to its Public Data, ECCSEL will operate them under the status of the Open Data Commons Open Database License (ODBL).

4.5 Monitoring compliance:

4.6.1 The Data and corresponding Databases will have to be monitored to ensure that their integrity is preserved. The curation of the data will also have to be kept up to date following the last EORTG recommendations.

4.6.2 Misuse of Public Data regarding ODBL and GDPR requirement will have to be investigated and reported to ECCSEL OC.

4.6 Scientific publication

4.7.1 ECCSEL will contribute to achieve the goals set by the Amsterdam Call for actions on Open science namely considering scientific publications. All ECCSEL Publications will be open source publications (either gold or green access type).

4.7.2 For needs of traceability but also to improve outreach on ECCSEL activity, each ECCSEL Publications should also refer to "ECCSEL Facility(ies)" or "ECCSEL ERIC" in their text or acknowledgement(s).

5. Data management and DMP implementation

5.1 The Data governance committee is to be designated by the RICC. The participation will be opened to RICC members as well as personal from their respective institutions. This committee will have the duty to help and counsel the OC to accomplish its duty regarding DMP implementation, namely when it is updating the EORTG. The Data governance committee might also be consulted by the RICC on specific and technical issues related to Data management.

5.2 Any change in the DMP will have to be discussed and validated by the RICC before being published into a new official version by the OC. It is up to the ECCSEL Director to evaluate if the changes are substantial enough to delegate RICC validation power to the General Assembly.

5.3 Specific deviations from the DMP will possibly be authorised by the ECCSEL ERIC Director and notified in the corresponding SLA and Access Agreement if at least they fulfil the following requirements:

- It does not constitute a threat to (1) the global deployment of the last validated DMP version and (2) the long-term inclusion of the corresponding Public Data into ECCSEL ERIC Database ontology

- It will not last more than 3 years

Definitions:

CCUS: Carbon capture, utilisation and storage.

External Data: Data that has been granted the right to not be shared publicly by ECCSEL.

Data: ECCSEL ERIC Research and administrative data that are generated through the Access to ECCSEL ERIC Facilities and/or produced through ECCSEL boards activities (i.e. RICC, OC and other consultative committees).

Database(s): Databases made of data

Data Life Cycle: Set of processes that perform acquisition, curation, storage, publishing, processing and usage of the data. Management of Data legal and security issues are also included in the data life cycle.

Data Management Policy (DMP): The present document.

Data Manager(s): People or entity(ies) that will directly manage data life cycle. They can be the data owner(s) or be contracted out by the Data Owner(s) to perform all or part of the tasks associated to Data life cycle.

Data owner(s): People or entity(ies) that possess the right and control over acquisition, use and distribution of a Data set.

ECCSEL ERIC Ontology Rules and Taxonomy Guide (EORTG): Guide produced by the OC towards Data Owners and Managers in order to provide them with a set of common rules to harmonize curation procedures of the Data.

ECCSEL ERIC publication(s): Scientific publications (i.e. documents, posters, manuscripts, abstracts or the like, of a scientific nature) that are associated or performed with ECCSEL ERIC Research.

OC: Operations Centre – the management of ECCSEL ERIC.

Public Data: Interoperable and reusable data that has to be shared with the public.

RICC: Research Infrastructure Coordination Committee of ECCSEL ERIC.

SLA: Service Level Agreement.