EURAMET

DATA MANAGEMENT PLAN

Grant Agreement number	21GRD06
Project short name	MetCCUS
Project full title	Metrology support for carbon capture utilisation and storage
Data management plan	1st ☒ 2nd ☐ 3rd ☐

TABLE OF CONTENTS

1 Data management plan

1.1 Data summary

Questions	Answers
1 Will you re-use any existing data and what will you re-use them for? State the reasons if re-use of any existing data has been considered but discarded.	Yes. Data provided by some of the participants (including published data) and stakeholders will be used for: - Based on studies from other projects including previous EMPIR projects A1.2.5 will theoretically investigate the impact of impurities on the transfer standards which are used in A1.2.1 to A1.2.4. - Based on study from literature and existing data resources A1.3.1 will study into the current state of the art on the traceable measurement of liquid CO_2 - Data about fiscal metering for WP1 A1.4.1 - Previous data on performance of analysis methods for nitrosamines from external projects for Task 2.1 - Data on stack conditions for A2.1.1, existing performance data for methods A2.1.2 - Previous data on underwater emissions studies A2.3.2 - Data and literature about CO_2 composition for WP3 A3.1.1 from previous EMPIR project 20IND10 Decarb - Providing at least two simplified formulations for the prediction of the density of typical CO_2 mixtures (WP4, A4.2.3)
2 What types and formats of data will the project generate or re-use?	Measured data and associated uncertainties; typically, ASCII (American Standard Code for Information Interchange), MS office format, images in JPEG format, numerical data in CSV format, text description data in Markdown format. Research outputs will include new: software, calibration methods, protocols, materials. Other types and other formats (e.g. TXT, Python, R) may be considered – to be confirmed at a later date. Where possible, data formats are chosen that are supported by Open Source software, or software from different vendors to maximise the update of data sets by other groups.
3 What is the purpose of the data generation or re-use and its relation to the objectives of the project?	Data are generated and used for different purposes. Most of the data generated during the development of novel facilities and measurement standards will be used to demonstrate their capabilities and assess their performance. - In A1.1.4 data will be used to evaluate the transferability of the flowmeter calibration between CO_2 and alternative fluids. (objective 1) - In the comparisons planned in WP1 A1.2.2 and A1.2.4 to support calibration and measurement capabilities. (objective 1) - For WP2 data will be used to validate and understand the performance of methods and to inform uncertainty calculations. (objective 2) - Data is needed to have a summary on the existing CO_2 specification for CCUS in order to select key impurities in WP3 A3.1.3. (objective 3) - Data will be used to select the right cylinder type and sample bags for A3.1.3, A3.1.4, A3.2.3. (objective 3)

	- In WP4, A4.1.2, A4.1.3, A4.1.4, the results will be used to update the OES-CG formulation. (objective 4)
4 What is the expected size of the data that you intend to generate or re-use?	Although is not currently known, the expected size of the data is likely to be < 10 GB with individual files ≤ 1 MB.
5 What is the origin/provenance of the data, either generated or re-used?	*Data generated in the project* The data/research outputs generated will be from measurements, calibrations, testing, comparisons and validations. *Re-used data* The existing data will originate from several sources, which will include participant's pre-existing data, data from the scientific literature, real-world measurement data, and simulation data.
6 To whom might your data be useful ('data utility'), outside your project?	The data might be useful to JCGM, ISO and CEN technical committees, accreditation bodies, research institutes, calibration and testing laboratories, regulators, industry.

1.2   *Findable, Accessible, Interoperable and Re-usable (FAIR) Data*

1.2.1   Making data findable, including provisions for metadata

Questions	Answers
7 Will data be identified by a persistent identifier?	Yes. Published data will have a DOI obtained from an open access repository.
8 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.	Published data will have a DOI obtained from an open access repository. Meta data will describe the quantities, units, methods of measurement (if known) and if possible or relevant, the origin of the data. The meta data will also contain references to the project, and where appropriate, to the owner(s) of the data. The metadata will provide information on the following: datasets (description, date of deposit, author(s), venue and embargo); the European Partnership on Metrology funding; grant project name, acronym and number; licensing terms; persistent identifiers, the authors involved. Where applicable, the metadata will include persistent identifiers for related publications and research outputs. The metadata created for all of the project's deposited datasets will be open under a Creative Commons Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles.
9 Will search keywords be provided in the metadata to optimise the possibility for discovery and then potential re-use?	Key words will be provided to enhance the discovery of the data by search engines, such as the following: Carbon Capture, Utilisation, Storage, Carbon dioxide, Metrology.
10 Will metadata be offered in such a way that it can be harvested and indexed?	Zenodo complies with FAIR principles (https://about.zenodo.org/principles/). The metadata are indexed in a searchable resource. Metadata are licensed under CC0, except for email addresses. All metadata are exported via Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and can be harvested.

1.2.2 Making data accessible

Questions	Answers
Repository:	
11 Will the data be deposited in a trusted repository?	Yes. The data will be deposited in the trusted open access repository Zenodo (https://zenodo.org)
12 Have you explored appropriate arrangements with the identified repository where your data will be deposited?	Zenodo is a general-purpose repository that provides a persistent digital object identifier (DOI) for each submission. Exploration of the arrangements with other repositories is an ongoing activity.
13 Does the repository ensure that the data are assigned an identifier? Will the repository resolve the identifier to a digital object?	Repositories will be used that provide a DOI, such as, Zenodo, but alternatives will be considered in the course of the project. Zenodo will assign an identifier (DOI) to each of the project's deposited datasets. The repository will resolve the identifier to a digital object.
Data:	
14 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.	All the data from scientific publications will be made openly available by default, unless there is a specific reason (to be discussed between participants and coordinator) not to publish the data. Other data may be made available on a case-by-case basis if it is relevant for third parties. Data obtained with the permission of third parties where the third parties have not agreed to make the data publicly available will not be made publicly available.
15 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.	Data will be made publicly accessible before, at, or shortly after the time of publication as appropriate. Data associated with other project outputs (e.g., reports, journal papers, conference contributions) will be released for publication in the same effort, so that the data set can be referenced in the project output and vice versa.
16 Will the data be accessible through a free and standardised access protocol?	Yes. Zenodo provides well described conditions for free and standardised access (see http://about.zenodo.org/policies/). All data associated with the project activities will be made available, with the possible exceptions of data obtained from organisations external to the project consortium. Where possible, permission to release the data will be obtained. Examples containing data sets may be reviewed by reviewers carefully selected by participants and coordinator. Some data sets will be released for general access, for example, to end-users of guides and standards, as soon as the data sets are complete, and quality is assured. Other data will more appropriately be

Questions	Answers
	made generally available at the time articles reporting on them are accepted for publication.
17 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 the use of the published data are envisaged. Access will simply require a standard user's registration.
18 How will the identity of the person accessing the data be ascertained?	Registered users of the research data repository.
19 Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?	No, the project participants and coordinator will ensure that registered users of repositories can access the data.
Metadata:	
20 Will metadata be made openly available and licensed 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?	In Zenodo, metadata are licensed under CC0, except for email addresses. All metadata are exported via OAI-PMH and can be harvested.
21 How long will the data remain available and findable? Will metadata be guaranteed to remain available after data are no longer available?	The data will remain available and findable for the lifetime of the Zenodo repository, which is expected to be a minimum of 20 years. If data are withdrawn from Zenodo, the DOI and the URL of the original object are retained. In case of closure of the Zenodo repository, best efforts will be made by Zenodo to integrate all content into suitable alternative institutional and/or subject based repositories.
22 Will documentation or reference about any software be needed to access or read the data and will this be included? Will it be possible to include the relevant software (e.g. in open source code)?	No additional documentation is needed for the software. The types and formats of files uploaded is to be determined; but will include: ASCII and MS Office. Other types may include: table, graphics, video, and other formats (e.g. TXT, MATLAB, Python, R) may be considered – to be confirmed as/when required.

1.2.3   Making data interoperable

Questions	Answers
23 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	Meta data will be provided in accordance with the stipulations of the Grant Agreement. Furthermore, in the meta data the quantities and units will be provided as well as relevant key words. The datasets will use the trusted repository's basic metadata schema for administrative data, which is compliant with the recommended standards used by DataCite (https://search.datacite.org/) and OpenAIRE (https://www.basesearch.net/).

	For individual datasets, the following discipline-specific vocabularies, standards, formats, and methodologies will be used: e.g. GUM, OBO foundry, DICOM, NetCDF, HDF5, CityGML, INSPEC, ISO 9001.
interoperability best practices? Which ones?	
24 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 their re-use, refinement or extension?	In the unlikely case that we will need to create our own ontologies or vocabularies, we will create mappings and publish the vocabularies.
25 Will your data include qualified references[1] to other data (e.g. other data from your project, or datasets from previous research)?	Yes, the project's datasets that will be deposited in the chosen repository (Zenodo) will include qualified references to other datasets from the same project and from previous research.

1.2.4 Increase data re-use

Questions	Answers
26 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.)?	Where documentation concerns data sets and code (e.g., scripts), this documentation will be primarily provided in the form of meta data and comments in the script text. A readme file will be provided with each submission describing the parts of that submission. Validation results, where applicable, will be provided also as part of the submission, preferably in PDF to ensure that this part of the documentation cannot unintentionally be changed.
27 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 re-use licenses, in line with the obligations set out in the Grant Agreement?	The data will either be licensed under the latest available version of the Creative Commons Public Domain Dedication License (CC 0) or a license with equivalent rights as set out in the Grant Agreement. Users will be required to acknowledge the consortium and the source of the data in any resulting publications.
28 Will the data produced in the project be useable by third parties, in particular after the end of the project?	Yes. The metadata will ensure that data will be reusable after the end of the project.
29 Will the provenance of the data be thoroughly documented using the appropriate standards?	Yes, the provenance and context of the data will be thoroughly documented to meet relevant standards using the Provenance and Context Content Standard (PCCS) Matrix. Data will be accompanied by information on how they were captured, processed, analysed, and

---

[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/)*

Questions	Answers
	validated. Other information that enables interpretation and use will also be provided.
30 Describe all relevant data quality assurance processes.	The quality assurance consists of three layers: (1) Data quality will be assured through both repeated and comparison measurements, adherence to standards for data recording and the use of controlled vocabularies and standard terminology. Moreover, the metrological characterisations of the measurement set-ups and the validation of the data collected will be used to ensure data quality. Test results, along with the data and the peer-review of publications based on these data, will constitute additional data quality assurance processes. (2) The data will undergo peer-review by participants and coordinator prior to release. The participants in the project will be requested to include in their review also the metadata and assess whether the data can be reused without further intervention by the creator or the consortium. (3) In case of issues with the data set, the participants and coordinator will address these jointly with the creator and issue an improved version of a data set.
31 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.	*Allocation of resources* The costs for making other research outputs FAIR are included in the project's budget and will be claimed if compliant with the Grant Agreement's conditions. Where feasible, long-term preservation will be ensured by depositing the other research outputs in repositories. The coordinator will decide on a case-by-case basis on which other research outputs will be deposited and for how long. *Security of other research outputs* The participants will store other research outputs on their organisations' networks, which are protected by firewall, backups etc. Other research outputs will also be stored in the project's SharePoint environment, with password-protected login. Deposition in public repositories will provide additional security as they have multiple replicas in a distributed file system which is backed up on a nightly basis. This project will not generate sensitive other research outputs. The other research outputs will be safely stored in open access repositories. *Ethical aspects* There are issues that could impact on the sharing of other research outputs. • Information relating to other research outputs acquired from third parties, e.g. manufacturers, will not be shared without their explicit consent. • Information relating to other research outputs collected by the consortium at commercial sites will not be shared without the site owner's explicit consent. Ethical issues will be addressed as the project will prepare an ethics report.

Questions	Answers
	The project will not share other research outputs with identifiable personal information. Sensitive information relating to the other research outputs will be collected, separated as soon as possible and kept secure.

1.3 Other research outputs

Questions	Answers
32 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.).	The software developed in the project will be released under a GNU-GPL license. The new calibration methods, and protocols produced by the project will be stored in the Protocol Exchange repository. The management of the IP issues surrounding the new materials that will be developed in the project have been planned for in the project's consortium agreement. The consortium intends to seek patent protection. This project will only re-use existing data and will not re-use any other research outputs.
33 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.	As far as possible, the FAIR data approaches specified in questions 7-30 above will be applied to the management of this project's other research outputs. This commitment will be met by releasing the new software that will be developed in the project under license, by placing the new calibration methods, and protocols, in a repository and by patenting the new materials that will be developed in the project in line with the requirements of the project's consortium agreement.

1.4 Allocation of resources

Questions	Answers
34 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 estimated curation and storage/preservation costs for making the data and research outputs FAIR are 5000 € (personnel costs). These costs will be kept to a minimum by using i) suitable trusted repositories from the Registry of Research Data Repositories https://www.re3data.org/ where no additional costs are associated with long-term preservation, and ii) by making only relevant data and outputs FAIR.
35 How will these be covered? Note that costs related to research data/output management are eligible as part of the European participantship on metrology grant (if compliant	The costs for making the data FAIR are included in the project's budget and will be claimed in accordance with the Grant Agreement's conditions.

with the Grant Agreement conditions).	
36 Who will be responsible for data management in your project?	The coordinator, with support from the participants, will have overall responsibility for the management of data/research outputs and quality assurance. The coordinator will be responsible for coordinating updates to the data management plan and for deciding on a case-by-case basis which data/research outputs will be kept and for how long. The participant(s) that produced the data will be responsible for organising backup and storage, archiving, and for depositing the data/research outputs within the chosen repositories.
37 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)?	Data will be safely stored in trusted repositories for long term preservation and curation, such as, Zenodo. There are no expected costs associated with the long-term preservation of the data. This consortium will not establish a DAC. The coordinator will be responsible for deciding on a case-by-case basis which data/research outputs will be kept and for how long.

1.5 Data security

Questions	Answers
38 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)?	*Data recovery and secure storage* Deposition in the Zenodo public repository will provide additional security as it has multiple replicas in a distributed file system which is backed up on a nightly basis. Data will also be backed-up on the partner organisations' own servers, to mitigate the (very low) risk of loss of data from the central repository. *Transfer of sensitive data* Sensitive data will only be transferred in a depersonalised / deidentified or pseudonymised form.
39 Will the data be safely stored in trusted repositories for long term preservation and curation?	Yes, the data will be safely stored in the Zenodo open access repository. Zenodo and the underlying Invenio Framework for digital repositories were designed according to the Open Archival Information Systems (OAIS) reference model. Zenodo is working towards ISO 16363 certification. Note: pre-publication data will be held by VSL in secure memory, primarily on a SharePoint site with access by participants. Access is limited to representatives from the consortium participants appointed by the members of the Project Management Board.

1.6 Ethics

Questions	Answers
40 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	Potentially there are issues that could have an impact on data sharing relating to carbon dioxide metering data, which may be considered as commercially sensitive. Sharing of these data will be considered jointly with the respective owner(s). Data may be anonymised or not disclosed at all after these considerations.

Questions	Answers
references to ethics report(s) and the ethics section in the Annex 1.	
41 Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?	No personal data are involved.

1.7 Other issues

Questions	Answers
42 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)?	Data will be stored on the Zenodo research data repository, which is hosted by CERN and receives funding from the European Commission via the original OpenAIRE Open Access Infrastructure for Research in Europe and subsequent programmes. Alternatively, similar repositories will be used that meet the applicable objectives and requirements of the the European Partnership on Metrology programme. Data management will be compliant with the research data policy of the European Partnership on Metrology programme and with European laws about data security and the protection of privacy (e.g. GDPR).