



Design, implementation and production upscaling of novel, high-performance, cluster-based catalysts for CO₂ hydrogenation

Deliverable D.8.4

Data Management Plan (DMP)



Grant Agreement No:	955650
Project start date:	01.11.2020
Duration of the project:	48 months
Deliverable number	8.4
Deliverable leader	KU Leuven
Due date:	30/04/2021 (M6)
Actual submission date:	27/10/2021
Dissemination level:	CO
Author(s):	Didier Grandjean with input from all partners

Contents

- 1. Introduction 3
- 2. CATCHY Data Summary..... 3
 - 2.1. Purpose of data generation and collection 3
 - 2.2. Types, formats and origin of data generated/collected..... 4
 - 2.3. Metadata 5
 - 2.4. Re-use of existing data..... 5
- 3. FAIR data 6
 - 3.1. Making data findable 6
 - 3.1.1 Data set description, reference and name..... 6
 - 3.1.2 Metadata 6
 - 3.2. Making data openly accessible..... 6
 - 3.2.1 Non-Open research data 6
 - 3.2.2 Open research data 6
 - 3.3. Making data interoperable..... 7
 - 3.4. Increase data re-use 7
- 4. Allocation of resources 8
- 5. Data security 8
- 6. Ethical aspects..... 8
 - 6.1. Confidentiality..... 9
 - 6.2. Involvement of non-EU countries..... 9
- 7. References 9

1. Introduction

The CATCHY Data Management Plan (DMP) gives an overview of the data and information collected throughout the project and shows the interrelation of the data collecting activities within and between the work packages. The DMP also links these activities to the CATCHY partners and discusses their responsibilities with respect to data handling within the network.

The CATCHY DMP lays out the procedure for data collection, consent procedure, storage, protection of data, and confirmation that they comply with national and EU legislations. The DMP will ensure that the exchange of data of companies and industries is in full compliance with the participating companies and industries internal data protection strategies. Thereby and wherever trade secrets of the participating companies and industries are not violated, CATCHY strives to comply with the open access policy of Horizon 2020.

This is the first version of CATCHY DMP that is intended to be a living document which will be adjusted whenever appropriate to the specific needs of network throughout the project's runtime. This plan will establish the measures for promoting the findings during CATCHY's lifecycle and will set the procedures for sharing the data of the project following the FAIR principle for research data management (Findable, Accessible, Interoperable and Re-usable) in line with the HORIZON 2020 FAIR DATA MANAGEMENT PLAN TEMPLATE and the Guidelines to FAIR data management in Horizon 2020 [1] and the GDPR (Regulation (EU) 2016/679).

The following document addresses for each data set collected, processed and/or generated in the project the following characteristics: dataset description, reference and name, standards and metadata, data sharing, archiving and preservation.

2. CATCHY Data Summary

During the Catchy project the size of the data collected/generated is expected to vary significantly according to the partner's research activity. From 1 or 2 Gb/year for the partners involved in the theoretical simulations (BME, FETI) and free cluster research (UU), to few hundreds of Gb/year for the partners involved in experimental laboratory research (KUL, DTU, VITO) and conventional synchrotron (PSI, KUL) or Free Electron Laser (RU) work and up to several Tb/year for the advanced in situ XAFs (PSI) and Electron microscopy (UA) measurements.

2.1. Purpose of data generation and collection

The purpose of data generation and collection resulting from the experiments carried out by Catchy ESRs is listed by the S&T WPs, namely WP1 to WP5: after analysis these data will support Catchy S&T deliverables and form the basis of publications arising from the project.

WP1-Cluster Deposition: The data will be used to identify the different cluster-based catalyst samples produced by ESR by detailing the sample preparation (including fabrication parameters) and to routinely characterize their physical and chemical properties.

WP2-Characterization of Deposited Clusters: The data generated from the ex and in situ characterization will contribute to the fundamental understanding of the structure-activity relationships and further development of in situ characterization tools. These will be of use for the whole consortium in the interpretation and understanding of catalysis experiments and ongoing physico-chemical processes. Furthermore, these data will be used as feedback to improve the synthesis of systems with desired catalytic properties.

WP3-Theoretical Modelling: Computed optimized geometries of the clusters, surfaces, reaction intermediates and products will be collected, deposited and published to ensure their reproducibility and enhance their future applicability as models for the novel catalysts developed in Catchy.

WP4-Free Cluster Reactivity: Data will be collected to detail the sample preparation and reactivity as well as to characterize their physical and chemical properties.

WP5-Catalysts Testing and Prototyping of a Model Gas Diffusion Electrode: The generated data will lead to a better understanding of the functioning of the clusters whether they are used as thermo-catalysts or electro-catalysts and eventually will help us develop a better product and process for CO₂ conversion.

2.2 Types, formats and origin of data generated/collected

Scientific data will consist of the raw data collected by the experimental measurements from the different S&T fields represented in Catchy as well as theoretical models, metrics in time, frequency or space domains. S&T work packages will rely on a large variety of data types (measurements-based, theoretical modelling, etc.) resulting from the different software tools used either to control the numerous technical instruments available in the Catchy network or to simulate theoretically the data. The different types of data that will be collected/generated in Catchy are listed below by WP:

WP1-Cluster Deposition:

Raw data measured by different instruments: *.txt, .dat*, format of the softwares used and arise from these measurements: *SEM, XRD, EDX*.

Written reports and scientific publications *.doc, .pdf*

WP2-Characterization of Deposited Clusters:

Raw data measured by different instruments:

- *.txt, .dat, .*
- *.emi, .ser, .dm3, .dm4, .bcf, .mrc*, (Softwares: Multem, StatStem, Esprit, TIA, Velox, Digital Micrograph).
- *.h5 files* (HDF5 data archives at RU)

Analysed formats:

.tif, .png, .jpeg, .rec, .txt

Processed data produced by:

Origin, C++, Python, Matlab,...

Written reports and scientific publications

.doc, .docx, .pdf

WP3-Theoretical Modelling:

Optimised geometries produced by quantum chemical software (e.g. Q-Chem, jDFTx, Gaussian, GPAW etc) are collected in standard *.XYZ* formats.

Written reports and scientific publications: *.doc, .pdf*

WP4-Free Cluster Reactivity:

Raw data measured by different instruments: *.dat, .h5, or .txt* files

Processed data produced by Origin, Python, Matlab,...

Written reports and scientific publications: *.doc, .pdf*

WP5-Catalysts Testing and Prototyping of a Model Gas Diffusion Electrode:

Raw data will be in the format of the softwares used and arise from these measurements- *LSV, CV, EIS, chronoamperometry, HPLC, SEM, XRD, EDX, GC data, pH*

Written reports and scientific publications: *.doc, .pdf*

2.3. Metadata

Metadata to be created within Catchy refers to any lab books, logbooks and ReadMe files for datasets to be produced within the project (measurements, SW, simulated data, analytical/mathematical model, etc.). For each experiment or simulation there will be clear set of metadata, in line with metadata standards that are commonly used in the specific domain of each partner as well as standards that come with the instruments / techniques. In this respect, we will be guided by the FAIR principles.

2.4. Re-use of existing data

Published data available in the literature, in data repositories (RU), in lab archives (PSI, KUL, DTU), collected data in future computations as starting geometries (BME, FETI) as well as some of the existing data about the gas diffusion electrodes and their production process (VITO) will be re-used.

3. FAIR data

3.1. Making data findable

3.1.1 Data set description, reference and name

- Each data set will be assigned with a unique name to distinguish and easily identify data sets; this name can also be used as the identifier of the data sets.
- Names of data files produced, including emails, should begin with the term “CATCHY”, followed by file name which briefly describes its content, followed by a version number (or the term “FINAL”), followed by the short name of the organization which prepared the document (if relevant).
- Each data set that will be collected, processed or generated within the project will be accompanied by a brief description in the form of a header or a separate ReadMe file.

3.1.2 Metadata

All the file names and descriptions of raw data and metadata will be documented in logbooks and/or directly be stored on the respective CATCHY partners storage servers. All data generated within large facility such as Synchrotrons (PSI) will immediately be stored on local storage servers or data repositories (FELIX laboratory, RU), and grouped per experimental instrument and generation day, clearly labeled in filename and directory structure for which day the data are generated. The storage is backed-up incrementally every day and kept for at least five years after the end of the project. All datasets supporting a publication from the Catchy network will be deposited on the Zenodo repository or equivalent repositories that automatically provides a DOI for each data entry that will be directly discoverable through the platform search system.

3.2 Making data openly accessible

The digital data created by the project will be diversely curated depending on the sharing policies attached to it. For both open and non-open data, the aim is to preserve the data and make it readily available to the interested parties for the whole duration of the project and beyond.

3.2.1 Non-Open research data

All data created by the project will be first scrutinized for any IP possibility. When the choice of filing a patent will be made, the research data will be kept confidential and stored on the Catchy partner local server until the patent becomes public at which point it will be made openly available.

3.2.2 Open research data

To the extent possible, data related to the publications will be made openly available in the form of processed images (png, tiff) and scripts (log files, matlab) used in the processing of

raw data. The raw data such as the microscopy images could be made available after a careful analysis of the valorization potential.

This open research data will be archived on the **Zenodo platform** [2]. Zenodo is a EU-backed portal based on the well-established GIT version control system [3] and the Digital Object Identifier (DOI) system [4]. The portal's aims are inspired by the same principles that the EU sets for the pilot; Zenodo represents thus a very suitable and natural choice in this context. The repository services offered by Zenodo are free of charge and enable peers to share and preserve research data and other research outputs in any size and format: datasets, images, presentations, publications and software. The digital data and the associated metadata are preserved through well-established practices such as mirroring and periodic backups. Each uploaded dataset is assigned a unique DOI rendering each submission uniquely identifiable and thus traceable and referenceable.

3.3. Making data interoperable

In order to make the data interoperable in Catchy we will use data standards and metadata standards that are commonly used in the specific domain of each partner as well as standards that come with the instruments / techniques. Proprietary software and language-dependent formats will be avoided where possible (e.g., during industrial secondments).

3.4. Increase data re-use

Raw data originating from Catchy will be useful to other specialists of the same technique inside and outside the project. Processed data will be useful for consortium partners for the interpretation of ongoing experiments and as feedback to the synthesis tasks and more broadly to physicists and chemists working in the field of material science and catalysis. After publication the data supporting the work will be useful to the whole research community.

Third parties will be encouraged to re-use Catchy data to drive future research. IP restricted data will be made available for re-use after IP checks and patent filing. Data attached to a scientific paper will become available at publication time. Raw data will become available at the end of the project, or after an embargo time of 1 to 5 years, subject to internal evaluation of the valorization potential and internal re-use. To permit their widest reuse possible, data protected by IP will be licensed where possible using Creative Commons 4.0 BY or free licenses.

For most of Catchy partners the data that will not be subject to IP protection will remain re-useable for a minimum period of 10 years. The long-term lifetime will depend on that of the Zenodo repository, hosted and maintained by CERN. All data recorded at the FELIX Laboratory (RU) will be made available publicly at the Radboud Repository, where they are stored for a minimum of 10 years, with a 2-year embargo after the date of creation to allow the ESR to publish the data in a peer-reviewed scientific journal.

4. Allocation of resources

FAIR data management in CATCHY is part of the WP7 and KUL, as project coordinator, will be responsible for its implementation. Costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions). Resources for long term preservation, associated costs and potential value, as well as how data will be kept beyond the project and how long, are the responsibility of the partner who generated/collected the data. For network-wide data, where relevant, the resource allocation for long term preservation of data will be discussed by the whole consortium during the Supervisory Board meetings.

5. Data security

Every partner is responsible to ensure that the data are stored safely and securely and in full compliance with the European Union data protection laws on its own storage system for the duration of the project. After the completion of the project, the responsibilities concerning data recovery and secure storage will go to the repository storing the dataset. All data files will be transferred via secure connections and in encrypted and password-protected form (for example with the open source 7-zip tool providing full AES-256 encryption [5] or the encryption options implemented in MS Windows or MS Excel). At a network level, the sharing platforms for the consortium data is the intranet of Catchy website (see Catchy deliverable D7.3 for details) complemented by Catchy folders in the Microsoft OneDrive of the university infrastructure at KUL shared between the relevant ESRs. These vital components of hosting and sharing the data are expected to stay in place in the long term, thus ensuring continued access to the collected data in a secure way.

6. Ethical aspects

This section deals with ethical and legal compliance issues, like the consent for data preservation and sharing, protection of the identity of individuals and companies and how sensitive data will be handled to ensure they are stored and transferred securely. Data protection and good research ethics are major topics for the consortium to guarantee that sensitive information will not be misused. Furthermore, all processes of data generation and data sharing have to be documented and approved by the consortium to guarantee highest standards of data protection.

CATCHY partners have to comply with the ethical principles as set out in Article 34 of the Grant Agreement, which states that all activities must be carried out in compliance with:

- ethical principles (including the highest standards of research integrity — as set out, for instance, in the European Code of Conduct for Research Integrity including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct) and
- applicable international, EU and national law (in particular, EU Directive 95/46/EC).

6.1. Confidentiality

CATCHY partners must retain any data, documents or other material as confidential during the implementation for the project. Further details on confidentiality can be found in Article 36 of the Grant Agreement along with the obligation to protect results in Article 27.

6.2. Involvement of non-EU countries

CATCHY non-EU partners (TCL and PSI) have confirmed that the ethical standards and guidelines of Horizon2020 will be rigorously applied, regardless of the country in which the research is carried out. Activities carried out outside the EU will be executed in compliance with the legal obligations in the country where they are carried out, with an extra condition that the activities must also be allowed in at least one EU Member State.

CATCHY data will be transferred between the named non-EU country (UK and Switzerland) and countries in the European Union to allow for joined analyses and storage of all data in the common database. All data transferred between project partners (within or outside the EU) will be restricted to pseudonymized or anonymized data and transfer will only be made in encrypted form via secured channels.

7. References

- [1] European Commission, Participant Portal H2020 Online Manual. https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm
- [2] Zenodo platform. <https://zenodo.org>
- [3] GIT version control system. <https://git-scm.com>
- [4] Digital Object Identifier (DOI) system. <http://www.doi.org>
- [5] 7-zip tool providing full AES-256 encryption: <http://www.7-zip.org/>