

“ **Ma** **tu** **ri** **ng** **th** **e** **pr** **od** **uc** **ti** **o** **n** **st** **an** **d** **ar** **d** **s** **o** **f** **ul** **tr** **a** **p** **o** **r** **o** **u** **s** **str** **uc** **t** **ur** **e** **s** **fo** **r** **hi** **gh** **de** **ns** **it** **y** **h** **ydro** **ge** **n** **st** **or** **a** **g** **e** **ba** **n** **k** **o** **pe** **ra** **ti** **ng** **o** **n** **sw** **in** **g** **te** **m** **pe** **r** **at** **ur** **e** **s** **an** **d** **lo** **w** **co** **m** **pr** **es** **s** **i** **o** **n** ” – **MA** **ST** **3** **R** **Bo** **o** **st**



## D1.3. Data management plan for the project

Due date of submission: 30/11/2022

Actual submission date: 24/02/2023



Funded by the  
European Union

## TABLE OF CONTENTS

|                                                |    |
|------------------------------------------------|----|
| TABLE OF CONTENTS .....                        | 2  |
| PROJECT INFORMATION.....                       | 3  |
| DELIVERABLE DETAILS.....                       | 4  |
| 1 Introduction.....                            | 5  |
| 2 Overview .....                               | 5  |
| 3 General Data Management Considerations ..... | 6  |
| 3.1 Utility of the data.....                   | 6  |
| 3.2 Accesibility of official and raw data..... | 6  |
| 3.3 Data sharing.....                          | 6  |
| 3.4 Publication policies .....                 | 6  |
| 3.5 Ethical considerations .....               | 7  |
| 4 LabCore Digital Notebook.....                | 7  |
| 4.1 Data Model.....                            | 8  |
| 4.2 Data Curation .....                        | 9  |
| 4.3 Data Interoperability.....                 | 9  |
| 4.4 Data Publication .....                     | 10 |
| 5 Data Safety and Conservation .....           | 10 |
| 6 Personal Data Treatment.....                 | 11 |
| 7 Datasets .....                               | 11 |

## PROJECT INFORMATION

**Project full title:** Maturing the production standards of ultraporous structures for high density hydrogen storage bank operating on swinging temperatures and low compression

**Acronym:** MAST3RBoost

**Call:** HORIZON-CL4-2021-RESILIENCE-01

**Topic:** HORIZON-CL4-2021-RESILIENCE-01-17

**Start date:** 1<sup>st</sup> June 2022

**Duration:** 48 months

**List of participants:**

| Number | Name of beneficiary                             | Acronym of beneficiary | Country      |
|--------|-------------------------------------------------|------------------------|--------------|
| 1      | ENVIROHEMP                                      | ENV                    | Spain        |
| 2      | CONTACTICA                                      | CTA                    | SPAIN        |
| 3      | Consejo Superior de Investigaciones Científicas | CSIC                   | Spain        |
| 4      | Spike Renewables Srl                            | SPIKE                  | Italy        |
| 5      | EDAG Engineering GmbH                           | EDAG                   | Germany      |
| 6      | Nanolayers                                      | NANO                   | Estonia      |
| 7      | FUNDACIÓN CIDETEC                               | CIDETEC                | Spain        |
| 8      | Leichtmetallkompetenzzentrum Ranshofen GmbH     | LKR                    | Austria      |
| 9      | University of Pretoria                          | UP                     | South Africa |
| 10     | Council for Scientific and Industrial Research  | CSIR                   | South Africa |
| 11     | PSA                                             | PSA                    | Portugal     |
| 12     | TWI Ltd                                         | TWI                    | UK           |
| 13     | University of Nottingham                        | UoN                    | UK           |

## DELIVERABLE DETAILS

|                            |                                                                                                                                                                                                                                                                                                                                                                                                |
|----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Document Number:</b>    | D1.3                                                                                                                                                                                                                                                                                                                                                                                           |
| <b>Document Title:</b>     | Data Management Plan for the project                                                                                                                                                                                                                                                                                                                                                           |
| <b>Dissemination level</b> | PU – Public                                                                                                                                                                                                                                                                                                                                                                                    |
| <b>Period:</b>             | PR1                                                                                                                                                                                                                                                                                                                                                                                            |
| <b>WP:</b>                 | 1                                                                                                                                                                                                                                                                                                                                                                                              |
| <b>Task:</b>               | T1.3                                                                                                                                                                                                                                                                                                                                                                                           |
| <b>Author:</b>             | Nanolayers OÜ                                                                                                                                                                                                                                                                                                                                                                                  |
| <b>Abstract:</b>           | This data management plan outlines the main strategy that MAST3RBoost will follow in order to ensure the project data is well documented, formatted and curated. Using the LabCore digital notebook platform the project guarantees that the data is abiding the F.A.I.R. principles required by the EU, and at the same time enforcing the data confidentiality agreements of the consortium. |

## 1 INTRODUCTION

This document describes the data management strategy for the MAST3RBoost as a set of guidelines for the project activities, and it will be updated to its final form in M48. MAST3RBoost is a very experimentally-oriented project and the data management plan focuses on partners involved in measurements and their data. Since the project has a component of machine learning activities based on experimental data, it is of outmost importance that the data is well curated and formatted in order to make it understandable and usable without the technical know-how of the experiments.

The general guidelines are the following:

- The research data policy will be supervised by the project coordinator in close cooperation with NANO.
- Partners are encouraged to store their data in LabCore digital notebooks, including a thorough description of their methods, raw data from instruments and processed results from their software of choice.
- All notebooks are treated as confidential from their creation; owner(s) can extend access rights to other LabCore users, or publish their notebooks.
- Secure, direct access to Nanolayers server for data storage will be arranged on request.
- Sensitive data is not disclosed.
- New knowledge will be published after agreed IP protection in peer-reviewed scientific journals and made available through open-access mechanisms or research repositories of the academic partners.
- Specific data on actual industrial matters cannot be made available due to commercial sensitivity.

## 2 OVERVIEW

MAST3RBoost will produce mostly experimental data, including adsorption isotherms, spectroscopy, microscopy and chemical characterization measurements, as well as data resulting from post-processing methods. Details of raw data formats and ontology were listed in D1.1 and D1.4 respectively. All collected data will be transferred to the centralized MAST3RBoost database, operated by NANO, where it can be mined and used for training machine learning algorithms. Two main data management frameworks have been deployed in order to achieve this.

NANO guarantees access to their LabCore digital notebook platform to all project partners who need to work with the project data, and strongly encourages its use for all their data. The platform was designed specifically to handle scientific data and enable users to properly document and curate it, thus making the data truly understandable and reusable. This should be the preferred data management framework throughout the duration of MAST3RBoost. More details of the platform design and capabilities will be discussed in section 4.

Additionally, project partners can request direct access to the NANO data server with secure file transfer protocol (SFTP). This way they can quickly store data and eventually share it with other members. The server is effectively a conventional file storage, where users can manage their own personal and group folder, and share data with the consortium with a global project folder. This offers no tools for parsing, labelling or documenting the uploaded files, although it can be achieved to a lesser degree if the users upload convenient descriptions alongside their data, for example with *readme* files describing folder structure and files. Project partners are recommended to use this system only for quick raw data backup, and possibly to share files with other members that know how to use it, for example members of the same research group, or using the same software for analysis.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [name of the granting authority]. Neither the European Union nor the granting authority can be held responsible for them.

All uploaded data, either via LabCore or SFTP, is not publicly accessible by default, however partners will be able to publish LabCore notebooks and data that are not considered confidential. Please note that secondary exploitation opportunities may arise as the project runs, possibly going beyond the project focus, and data confidentiality classification might need to be adjusted.

### **3 GENERAL DATA MANAGEMENT CONSIDERATIONS**

#### **3.1 Utility of the data**

The results obtained within MAST3RBoost project will be useful for several industries (engineering and manufacturing automotive and road and non-road transport, refueling stations, OEMs, etc.) as well as researchers and agencies in the field of H<sub>2</sub> storage and Hydrogen Fuel Cells. Moreover, data obtained from the experimental tests will be useful for the project to cover data gaps in the literature such as in-detail risk assessment and LCC of cryo-compressed technologies and the mechanical properties of materials under H<sub>2</sub> atmosphere and cryogenic temperatures.

The main outcomes will include key indicators of the technology, a type of processed data that it is expected to have widespread distribution among other H<sub>2</sub> initiatives such as projects and clusters.

#### **3.2 Accesibility of official and raw data**

According to good practises established in other EU projects, the official administrative data (i.e. consortium agreement, reports, deliverables, minutes, meeting presentations, publications, publication permissions, etc.) will be stored centrally on the Private Area of the MAST3RBoost website, where all the partners have access to the data. Raw data will be stored mainly in LabCore and Nanolayers' file server, but also in a conventional way by the owner of the results, i.e. locally at each partner's facilities and according to their respective internal rules. Datasets will be processed and analysed using relevant characterization and analysis software. Variable and quantity names will be constructed following general data processing principles. Used abbreviations, technical terms and standards will be explained or mentioned in the local Data Management Plans.

#### **3.3 Data sharing**

During the project, the data will be used by the consortium members. After the project is closed, all requests for further use of data will be considered carefully and whenever possible approved by the Coordinator and by the General Assembly. Permission for data use will be granted providing there are no IPR or confidentiality issues involved or any direct overlap of research questions with the primary research. Ownership of datasets will belong to the partners that generated them.

Official documents (e.g. deliverables, reports, minutes, contracts) are available within the consortium, using the Private Area of the project website. Partners using the system will first accept the Terms of Use, in order to receive their login information from the Coordinator.

#### **3.4 Publication policies**

The deliverables in the MAST3RBoost project are a mix of public and confidential documents. Project management documents, material deliverables, reports on novel processing methods, product formulation, scale up documentation, and LCA of the materials are typically in the form of confidential reports, whereas paper publications on optimisation of processes typically are public and will be published in peer-reviewed scientific journals. However, many other reports with key technical information will be made public. This is the case for all those deliverables related with the efforts made in literature review (D1.2, D4.1), harmonization of

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [name of the granting authority]. Neither the European Union nor the granting authority can be held responsible for them.

the material characterization methodology for Machine Learning (D1.1, D1.4, D1.5, D1.6, D1.7), catalogue of new structural materials (D4.2), prototypes and demonstrators (D3.5, D4.5, D5.4), design, simulation and benchmarking of the storage technology (D4.6, D5.1, D5.8) and risk assessment and LCA and LCCA of the complete system (D6.2, D6.4, D6.5).

The public papers and project information will be disseminated through several channels, as summarized in the DoA: dissemination and communication plan. The Coordinator in collaboration with the Steering Committee (SC) will take all the appropriate measures to make the data openly available and usable for third parties for study, teaching and research purposes via licensing or other means, whenever possible.

Open access publication will be ensured to all peer-reviewed scientific publications.

All publications are under the permission policy described in the GA: Prior notice of any planned publication and patent application shall be given to the other Parties at least 30 calendar days before the publication. Any objection to the planned publication shall be made in accordance with the Grant Agreement in writing to the Coordinator and to the Party or Parties proposing the dissemination within 20 calendar days after receipt of the notice.

### 3.5 Ethical considerations

The ethical principles of the project members agree with the highest standards of research integrity. The members will comply with the following principles: reliability and honesty when explaining the methods, analysis and results of the tasks carried out during the project development, respect for the partners, the environment and the society, and responsibility for research carried out from start to finish.

If personal data should be collected during any stage of the project, the guidelines of the GDPR will be met.

## 4 LABCORE DIGITAL NOTEBOOK

Unlike most digital notebook available, LabCore was designed with particular focus on scientific data, and its integration in the notebook. LabCore’s modular design allows scientists to upload raw data files directly from instruments, parse them, and directly use the data in the built-in visualisation and processing tools (Figure 1). As such, writing a notebook is akin to building and simultaneously documenting the data pipeline, from acquisition to processing and consumption.

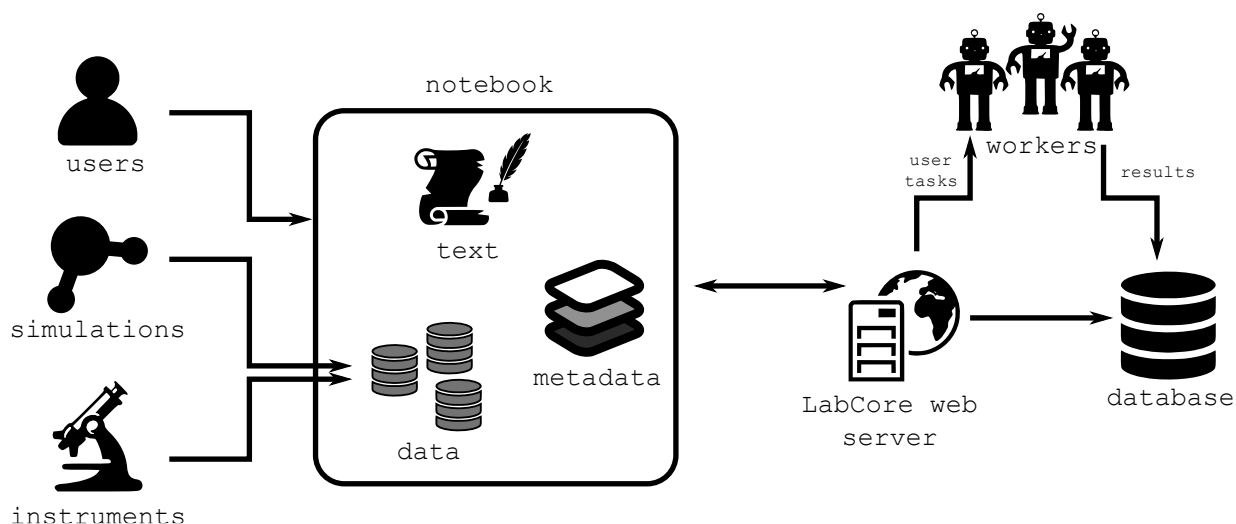


Figure 1: schematic representation of the LabCore infrastructure and workflow.

LabCore notebooks go well beyond being just hypertext documents in front of a file server, and, when properly written, will make the project data truly compliant with the F.A.I.R. principles. Moreover, the platform design is extremely modular, making it easy to add support for new data types and formats, and keeping LabCore relevant as new methods are developed.

#### 4.1 Data Model

The main difference between LabCore and other commercially available digital notebook solutions is in the data model. Data files uploaded into notebooks are not simply stored in a file server, but they are parsed first, in order to separate data and metadata from the rest. Parsing is almost completely automated, and only requires the uploader to know the raw-data format to select the appropriate parser.

The data is then reformatted according to its basic type (number, array, matrix, image, ...) and placed into *data records*. Regardless of their origin and raw format, all records of the same type are stored in the database with the same format. A lot of the heterogeneity in scientific data is due to the multiplicity of formats for the same types of data, mirroring the variety of instruments and vendors; LabCore's data model solves this issue, making the data reusable and interoperable.

The fundamental data types are listed in Table 1.

Table 1: List of LabCore's basic data types.

| Name      | Description                       |
|-----------|-----------------------------------|
| DATA0D    | A single value (scalar or string) |
| DATA1D    | One-dimensional array data        |
| DATA2D    | Two-dimensional matrix data       |
| DATA3D    | Three-dimensional tensor          |
| DATA4D    | Four-dimensional tensor           |
| IMAGE     | Image file                        |
| FILE      | Unparsed raw file                 |
| COMPOSITE | Custom structure of other basic   |

NANO has noticed that most scientific data is well represented already by fewer types, namely arrays and matrixes. Other types are however necessary in specialised cases, and are thus supported by the system to improve its general applicability. It should be noted that the modular nature of LabCore allows developers to add other basic data types should the need arise, without breaking the existing database. The IMAGE type is used to store conventional image files and display them in the notebook. FILE type records are only meant to store files that cannot or should not be parsed, such as documents (instrument/software manual, scientific publications, ...), or to place in the database an unparsed copy of raw-data files for safe-keeping.

While single pieces of scientific data fit well within one basic type, a measurement often provides multiple pieces simultaneously that should be viewed as a single entity. For example, a scanning probe experiment outputs several acquisition channels at once, in the form of matrixes or tensors, depending on the operation mode, combined with the physical height and width of the scan area as scalar values. It is often possible to process the data only when all these pieces are combined. Fitting the whole experiment into a tailored basic

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [name of the granting authority]. Neither the European Union nor the granting authority can be held responsible for them.



type is just not feasible, since even slight differences in the experimental method will require the definition of a new data type, thus reintroducing the heterogeneity in the data that LabCore purposed to attenuate. This issue is solved by the COMPOSITE type, which holds no data itself but defines a hierarchical structure of references to other data records and internal metadata describing the record sub-type and how the platform tools can interact with it. The constituent parts of the composite are still accessible to the users for analysis or visualisation in the notebook, but they are tagged as belonging to a composite master record, avoiding data fragmentation. New composite sub-types are easy to define and implement, should the need arise.

## 4.2 Data Curation

LabCore offers a simple, yet effective metadata system for labelling data records. There are three types of metadata: tags, value-tags and inventory tags. All simple and value- tags are created by the platform users, and stored in the LabCore database, and the definitions are shared with all other users.

Tags are very simple labels consisting of name and description, as provided by the user who defined it. Their purpose is to attach additional information to a data record, like specifying whether it is raw or processed, how it was acquired or processed. More importantly, it can be used to define the ontology of a record, informing the consumers about what the data represents.

Value-tags are defined in the same way as simple tags, but require users to specify a value when the metadata is attached to a record. A common use case is to provide information about measurement conditions such as temperature or pressure.

Inventory tags are related to LabCore inventory management system (IMS), and are defined indirectly by the users. In the IMS, user can create labs and list their equipment, such as instruments, tools, consumables and samples. Special metadata tags are then created automatically for each one of these, and can be used to specify which sample or instrument was used to get the data.

If properly used, this metadata system makes it to search through large amounts of data, and understand what the data is and how it is supposed to be used. Combined with LabCore's python API, it is also easy to programmatically aggregate records into well-defined datasets for consumption in machine-learning algorithms.

## 4.3 Data Interoperability

LabCore's web interface offers several common visualisation tools for plotting data: these range from simple scatter plots and histograms, to two-dimensional density and contour plots. Additionally, NANO implemented a more flexible plotting element that relies on python scripts from the user to generate custom plots. This feature makes use of the open-source Matplotlib package, allowing users to create any desired custom plot. In all cases, the platform records not only the resulting plot, but the sequence of instructions that generated it, making clear to everyone reading the notebook how to get from data to visualisation.

Data processing tools in the web interface are few, and tailored to specific experimental techniques, such as scanning probe microscopy and nuclear magnetic resonance spectroscopy, developed and funded by previous projects. However, similarly to the custom plotting feature, LabCore also has a custom python script element, where users can execute their own processing routines that consume the data in the notebook to create new data records with the results.

The custom script and plot element bypass the computational limitations of the web browser by not being executed in the browser at all. Instead, they send the LabCore server a request to execute the computation and

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [name of the granting authority]. Neither the European Union nor the granting authority can be held responsible for them.

wait for the results. The server submits the tasks to a queue, where worker processes draw them, and return the output. The workers can execute longer and more demanding calculations than the visualisation/processing tools built in the web interface, simply because the web browser imposes quite severe limitations on the hardware resources available to the software. On the server side, these limitations do not apply, and the processing can make use of the whole hardware. However, for security reasons, this system is not intended for truly computationally heavy tasks that could take several days, or more than one computer core.

For those situations, LabCore's python API can be used to intuitively access notebooks and data and perform custom, computationally heavy processing from a dedicated high-performance computing infrastructure. Many of the basic data functionalities in the web interface have a corresponding API implementation, so that users can programmatically download, upload and label even large sets of data records with minimal effort. The API is particularly useful since it can aggregate records from multiple notebooks into a machine-learning-ready dataset: this feature is required by MAST3RBoost WP1 activities.

#### 4.4 Data Publication

Among other features on the LabCore development roadmap, there is the software framework that allows the users to publish their digital notebooks, with all the included data, into an open space within LabCore. As planned, this space will let any person, even non-registered platform users, search for notebooks, read them and download their data. Development efforts will focus on two main publication mechanisms: notebooks and datasets.

Publishing a notebook will simply move it to the open space, together with its data, and make it immutable, so that no further edits can be made to the notebook or its data. This feature is almost completely implemented, and only the web interface for public notebook search remains to be designed.

However, it is often useful to collect data from multiple source notebooks into a single dataset and make it publicly accessible. In this case, the dataset creator will have to publish an additional, templated notebook listing all data sources and providing a summary of the data acquisition methods involved. This feature is more difficult to implement as the owners of all source notebooks will have to agree on the publication of the dataset, and the platform should have a way to update the dataset if new data becomes available.

The publication system will fully realise the vision of open-science and open-data. NANO is currently exploring the possibility of joining the DOI network, thus being able to assign unique, permanent identifiers to published notebooks and datasets. However, if successful NANO will have to guarantee their long-term storage, possibly even permanent.

### 5 DATA SAFETY AND CONSERVATION

The LabCore's web interface is present on the public domain, with the proper restrictions. Outside users may be able to see only published notebooks, but cannot access any other content, nor register an account without NANO authorisation. The LabCore database is not visible outside the NANO network and can only be queried from the server machine, either by a user with admin credentials, or by the LabCore web server process. LabCore users can only access notebooks and data therein, if they own the notebooks, or have been granted read/write permissions by the owner. Our LabCore web-server has already been targeted several times by programs that scan automatically and repeatedly for systemic weaknesses and backdoors. These programs are designed to probe certain entry points that are found in common off-the-shelf web server solutions. Since our server is entirely custom-built, these standard administration entry points do not exist. All sensitive HTTP

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or [name of the granting authority]. Neither the European Union nor the granting authority can be held responsible for them.

requests (opening notebooks, downloading data, ...) sent to the server are expected to be signed with a X-CSRF token. The token is provided by the server to authenticated users, and frequently re-issued, thus making it extremely hard to counterfeit a legitimate user session. Finally, the web server uses only HTTPS, ensuring that all communication from users to the server is encrypted.

Regarding file system access via SFTP, this is also restricted. Only users with valid credentials are allowed in the system, and can only operate on their personal and project folders. Access permissions on each file can be extended to any other user by their owner, if necessary.

The database (LabCore), user personal folders and the project common folder (SFTP) are backed up weekly to a separate physical location. Simple data loss, such as accidental deletion by users, can be recovered quickly from the backup. More catastrophic events, such as physical destruction of the server infrastructure, will require redeployment of the software on a new infrastructure. This will take more time to reinstall all required software, however all the data can be recovered from the backup. Since backups are performed weekly, it is only possible that at most 7 days of data can be completely lost in the worst-case scenario. NANO will monitor incoming data traffic during the project and calculate the effect of possible data loss accidents at different times. This will help us establish whether weekly backups are enough to avoid significant data loss, or more frequent backups are advisable, possibly storing multiple successive ones.

All data and backups will be preserved for an additional five years beyond the end of the project. Access to the platform (LabCore and SFTP) will also be ensured for the same time frame. All notebooks and data access rights are in line with the consortium agreements.

## 6 PERSONAL DATA TREATMENT

Users can authenticate to LabCore using their email or username and their password. This is the only **personal information** required by the system, and the only one that is stored. Passwords are hashed with SHA512 algorithm and salted before entering the database. The platform does not collect data about the users' behaviour. Personal information is at no point sold or transferred to third-parties, and will be deleted upon user's request. All user accounts, and related personal data, will remain active until five years past the end of MAST3RBoost: at this point the accounts will be deactivated and the personal data permanently obfuscated unless users wish to maintain their access rights on the platform.

## 7 DATASETS

Here is a list of datasets that are expected to be collected during MAST3RBoost. This section will be updated at the end of the project with the actually produced datasets.

|                         |                                                                                                                                                                                                                                                                                                                                                                                                                               |
|-------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Dataset name</b>     | <b>Ultraporous carbons and MOFs</b>                                                                                                                                                                                                                                                                                                                                                                                           |
| <b>Corresponding WP</b> | 2                                                                                                                                                                                                                                                                                                                                                                                                                             |
| <b>Related partners</b> | CSIC, UoN, UP                                                                                                                                                                                                                                                                                                                                                                                                                 |
| <b>Description</b>      | Data collection with respect to WP2 generated carbon-based and MOFs ultraporous samples. The dataset includes the characterisation, via elemental analysis, of the precursor mixture used to make the sample, as well as of the resulting sample. Additionally there will be adsorption isotherms for each sample, and quantities calculated from post-processing, such as micropore area/volume estimates and H <sub>2</sub> |

|                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                  | <p>intake rate. UP will also provide X-ray diffraction, Infrared spectroscopy and scanning electron microscopy measurements.</p> <p>The experimental data from elemental analysis are generally scalar values. Adsorption isotherms and thermogravimetric analysis are given as one-dimensional arrays. Electron microscopy data is three-dimensional tensors.</p>                                                                                                 |
| <b>Standards and metadata</b>                                    | <p>The original characterization results, adsorption isotherms and thermogravimetric analysis are all stored in human-readable table files (.csv). Electron microscopy data is stored in TIFF image format.</p> <p>When uploaded into LabCore notebook, each value will be labelled with metadata corresponding to the sample of origin, and its ontology.</p>                                                                                                     |
| <b>Data Sharing</b>                                              | <p>Concerning the confidential levels, (a) project confidential, (b) consortium confidential and (c) public, our data is generally project confidential. Data provision to the public is limited according to the provisions of the signed G.A. and C.A. Access can be granted in individual cases by the data owners. Data that is classified as project confidential will only be available via the original producing partner or the Nanolayers run backup.</p> |
| <b>Archiving and preservation (including storage and backup)</b> | <p>Labelled data in LabCore digital notebooks and optionally in Nanolayers data server.</p> <p>Raw-data files in LabCore and optionally in Nanolayers data server or data owner local storage.</p> <p>LabCore database and Nanolayers data server are backed up on Nanolayers NAS.</p>                                                                                                                                                                             |

|                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Dataset name</b>           | <b>Ultraporous carbons and MOFs – Material Maps</b>                                                                                                                                                                                                                                                                                                                                                                                                                |
| <b>Corresponding WP</b>       | 1                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <b>Related partners</b>       | NANO                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <b>Description</b>            | <p>This dataset is the result of unsupervised machine-learning processing applied to the <b>Ultraporous carbons and MOFs</b> dataset. It contains the principal components calculated from the mixture composition descriptors of the samples (2D matrixes), and the mapping coordinates of the samples calculated from t-SNE (2D matrixes).</p>                                                                                                                   |
| <b>Standards and metadata</b> | <p>The data is generated directly in LabCore, and formatted with its built-in rules. Metadata will be applied to specify the ontology and computational methods used to obtain it.</p>                                                                                                                                                                                                                                                                             |
| <b>Data Sharing</b>           | <p>Concerning the confidential levels, (a) project confidential, (b) consortium confidential and (c) public, our data is generally project confidential. Data provision to the public is limited according to the provisions of the signed G.A. and C.A. Access can be granted in individual cases by the data owners. Data that is classified as project confidential will only be available via the original producing partner or the Nanolayers run backup.</p> |

|                                                                      |                                                                                                                                                                                                 |
|----------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Archiving and preservation<br/>(including storage and backup)</b> | <p>Labelled data is generated directly in LabCore digital notebooks and stored in the LabCore database.</p> <p>LabCore database and Nanolayers data server are backed up on Nanolayers NAS.</p> |
|----------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|----------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Dataset name</b>                                                  | <b>Storage vessel materials</b>                                                                                                                                                                                                                                                                                                                                                                                                                             |
| <b>Corresponding WP</b>                                              | 4                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <b>Related partners</b>                                              | LKR, TWI                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <b>Description</b>                                                   | Data collection of metal alloys and coating materials for hydrogen storage vessels, compiled by WP4. The dataset includes the characterisation of materials from tensile and fatigue testing (1D arrays), hardness measurement (1D arrays), optical and electron microscopy (2D matrixes), electron diffraction (both 1D and 2D arrays), and topography measurements (2D matrixes).                                                                         |
| <b>Standards and metadata</b>                                        | <p>Tensile, fatigue and hardness testing measurements and electron diffraction analysis are stored in human-readable table files (.txt, .csv). Optical micrographs, electron microscopy and topography data is stored as JPEG images.</p> <p>When uploaded into LabCore notebook, each value will be labelled with metadata corresponding to the sample of origin, and its ontology.</p>                                                                    |
| <b>Data Sharing</b>                                                  | Concerning the confidential levels, (a) project confidential, (b) consortium confidential and (c) public, our data is generally project confidential. Data provision to the public is limited according to the provisions of the signed G.A. and C.A. Access can be granted in individual cases by the data owners. Data that is classified as project confidential will only be available via the original producing partner or the Nanolayers run backup. |
| <b>Archiving and preservation<br/>(including storage and backup)</b> | <p>Labelled data in LabCore digital notebooks and optionally in Nanolayers data server.</p> <p>Raw-data files in LabCore and optionally in Nanolayers data server or data owner local storage.</p> <p>LabCore database and Nanolayers data server are backed up on Nanolayers NAS.</p>                                                                                                                                                                      |