“Maturing the production standards of ultraporous structures for high density hydrogen storage bank operating on swinging temperatures and low compression” – MAST3RBoost

MAST3RBoost

D1.4. Harmonized Data Gathering Methodology

Due date of submission: 10/11/2022

Actual submission date: 27/10/2022

Funded by the European Union

TABLE OF CONTENTS

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: 1st 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

Document Number:	D1.4
Document Title:	Harmonised Data Gathering Methodology
Dissemination level	PU – Public
Period:	PR1
WP:	WP1
Task:	T1.2
Author:	Nanolayers OÜ
Abstract:	Nanolayers has developed LabCore, a digital notebook platform specifically designed to tackle the challenges of storing and managing scientific data. By applying its formatting standards to all uploaded data, LabCore reduces many of the barriers that make the inherently heterogeneous scientific data difficult to access and reuse. The metadata system was enriched, thus providing an intuitive tagging system that make data clearly understandable and quickly sorted into queries for machine-learning applications.

Data Management Infrastructure

The project's data will be managed using LabCore Digital Notebook Platform developed by Nanolayers. This tool is specifically designed for scientists to store and curate their data into digital notebooks, perform analysis tasks, create visualisations and easily share it with collaborators. The server has been deployed (https://labcore.nanolayers.com) and is now accessible by all the consortium members that need to operate with the data in anyway.

The most important aspect of LabCore is its ability to parse raw-data files and extract the data into well-formatted data records (Figure 1). This way, a table column, for example, becomes a 1D array data record, formatted in the same way as all 1D array records regardless of their origin. LabCore understands only a few basic data types, such as arrays, matrixes, and images, however, its data architecture is modular and allows us to easily define new, compound data types that encapsulate multiple basic records into a single data entity. This is useful when a single experiment produces multiple pieces of data in a single measurement.

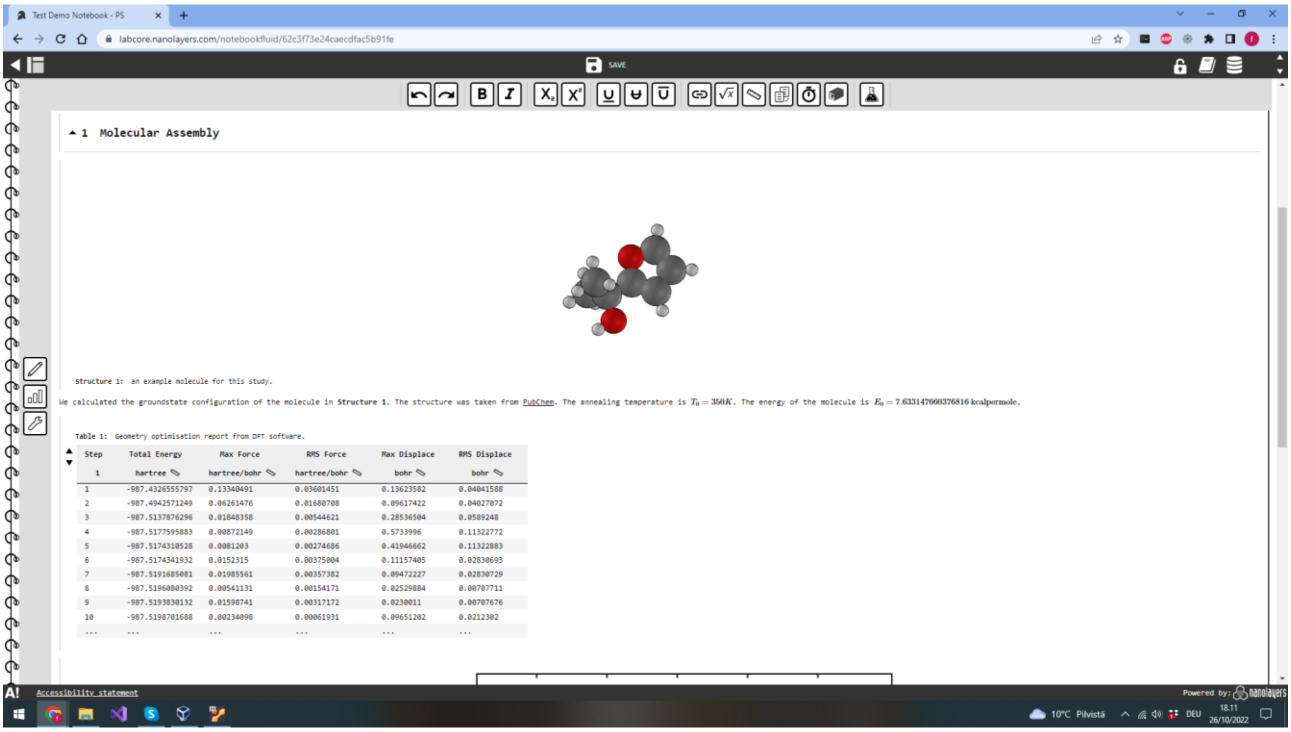

Figure 1: LabCore demonstration notebook showing a molecular structure and its calculation details imported directly from raw data files.

Thanks to its standardised formatting policy, all data uploaded into LabCore notebooks can then be consumed by the notebook elements to create visualisations and processed results (also in the form of data records), and used by collaborators without requiring the original instrument software (Figure 2). Moreover, a powerful Python API was also implemented, giving advanced computational users the ability to query the data, and downloading it already neatly formatted for consumption by machine-learning workflows.

The rich metadata system allows user to add tags to their data. This can be used to embed information about measurement methods and conditions into the data records, as well as labelling them with their designated ontology, making clear to any consumer (user or application) what the data represents.

As such, LabCore is an optimal solution for collecting the inherently heterogeneous scientific data, harmonising and making it truly **F**indable, **A**ccessible, **I**nteroperable and **R**eusable.

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.

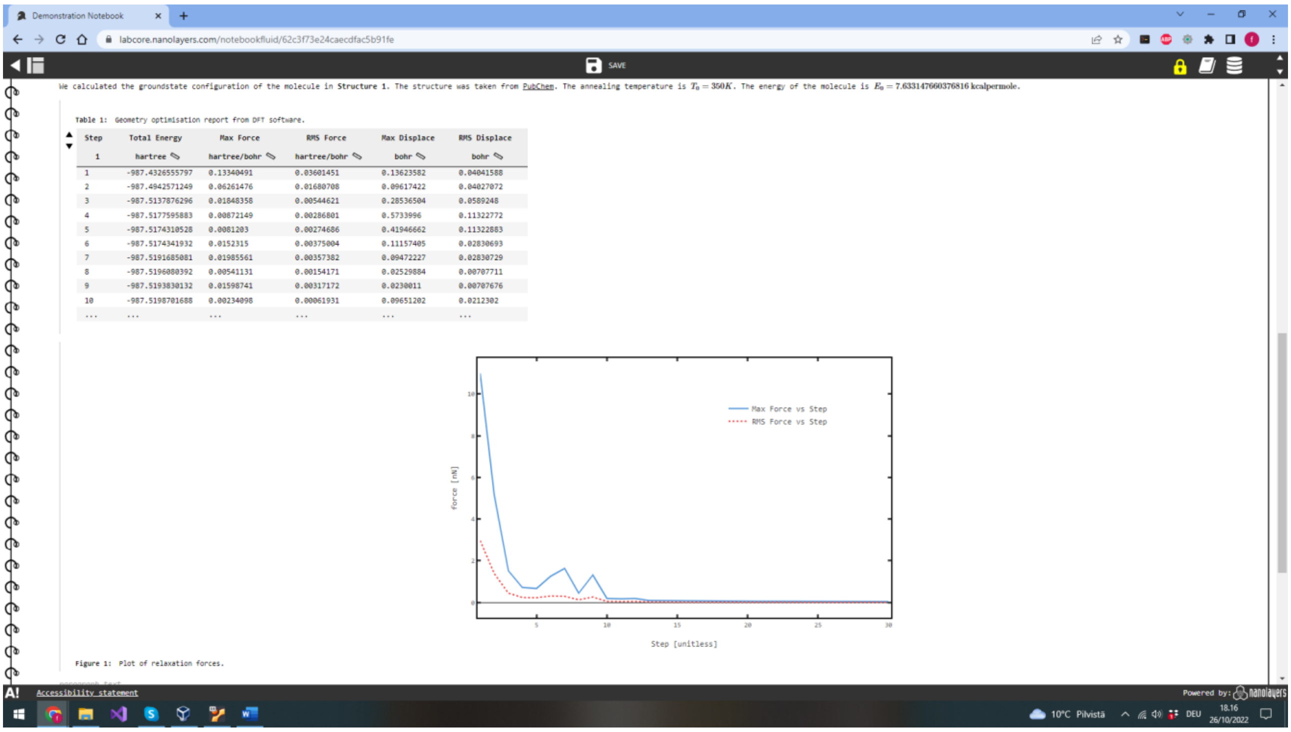

Figure 2: plotting tools in LabCore notebooks let users create visualisations from uploaded data, keeping track of how the plot was constructed.

Nanolayers ran tutorial sessions in July 2022 for all project partners that potentially need to upload and share, or receive, data. Since the partners are expected to hire researchers throughout the duration of the project, more general use, as well as targeted use-case, tutorial sessions will be organised upon request.

1. Data Ontology

Combining the input from all Mast3rBOOST partners, Nanolayers has compiled a list of data that users are expected to obtain from their instruments (Table 1).

Table 1: list of data produced by the various Mast3rBOOST experiments. The same experiment might appear multiple times with different data output when more partners are performing it with different equipment.

Experiment	Data ontology	Data Type	Physical units
N_2 adsorption isotherms	relative pressure	1D array	adimensional
	amount of N_2 adsorbed	1D array	cm^3/g STP
thermogravimetric analysis	time	1D array	s
	temperature	1D array	ºC
	weight	1D array	mg
	weight derivative	1D array	adimensional
N_2 adsorption isotherms	relative pressure	1D array	adimensional
	volume N_2 absorbed @ STP	1D array	cm^3/g
thermogravimetric analysis	time	1D array	min
	temperature	1D array	ºC

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.

Experiment	Data ontology	Data Type	Physical units
	weight	1D array	mg
	heat flow	1D array	mW
	temperature difference [°C]	1D array	ºC
	temperature difference [raw]	1D array	µV
	sample purge flow	1D array	mL/min
	derivative weight	1D array	%/ºC
powder X-ray diffraction	2θ angle	1D array	degrees
	PSD or Intensity	1D array	counts
FTIR spectroscopy	wavenumber	1D array	cm^{-1}
	transmittance	1D array	adimensional
scanning electron microscopy	SEM scan	3D array	adimensional
tensile testing	elapsed time	1D array	s
	stress	1D array	MPa
	strain	1D array	%
	ram displacement	1D array	mm
	force	1D array	kN
fatigue testing	time	1D array	s
	No. of cycles	1D array	adimensional
	min load	1D array	kN
	max load	1D array	kN
	min displacement	1D array	mm
	max displacement	1D array	mm
	min strain	1D array	strains (adim.)
	max strain	1D array	strains (adim.)
	min temperature	1D array	°C
	max temperature	1D array	°C
optical microscopy	micrograph	2D array	adimensional
scanning electron microscopy	secondary electron / backscatter electron	2D array	adimensional
EBSD	orientation maps	2D array	adimensional
EDX	EDX column A	1D array	keV
	EDX column B	1D array	cps/eV
Hardness measurements	indent number (multiple columns)	1D array	adimensional
	HV number (multiple columns)	1D array	Vickers hardness (HV)

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.

Experiment	Data ontology	Data Type	Physical units
alicona topography	surface image	2D array	adimensional
	height map	2D array	adimensional
	name	1D array	adimensional
	height	1D array	µm
adhesion test	adhesion test image	image	adimensional
scanning electron microscopy	SEM micrograph	image	adimensional
field emission SEM	FE-SEM micrograph	image	adimensional
EDX	EDX column A	1D array	keV
	EDX column B	1D array	cps/eV
Particle size measurement	particle size	1D array	nm
	intensity	1D array	%
zeta potential measurement	zeta potential	1D array	mV
	total counts	1D array	adimensional
OCV	time	1D array	s
	potential	1D array	V
EIS	frequency	1D array	Hz
	impedance	1D array	ohm
	phase	1D array	degrees
LP	current	1D array	mA
	potential	1D array	V
confocal	topographic image	3D array	adimensional

LabCore database has been seeded with metadata tags mirroring the required ontology of the project's output data. Each definition includes the keyword "Mast3rBOOST", making it easier for the project partners to find and apply them to their data. Additionally, inventory metadata tags will also be added into the database. These carry information about the instruments and samples that ultimately produced the data records.

Through the Python API, these metadata tags will also improve the process of designing and testing material descriptors for machine-learning applications.

This list of metadata is meant to be only a starting point. As the project continues, it will be most likely necessary to extend it with new metadata designed to label processed results from new analysis methods and machine-learning descriptors. Moreover, the project data ontology shall be formalised with EMMO standards when appropriate.

2. Methodology for harmonized data entering

It is expected that users will upload their data into digital notebooks, preferably including only one sample and its results into each one. Users have been instructed on how to operate the basic functionalities of LabCore in a tutorial session (https://www.dropbox.com/sh/lctpzijjgbnze2z/AADsnnUPtiMfynWca1Q5L_E-a?dl=0). However, Nanolayers remains available to provide tutorials and support to new staff members when needed.

Half of the data harmonization is automatically done by LabCore when users upload their results in digital notebooks, since those are reformatted using the platform internal standards. This ensures that all data of the same type will have the same format, regardless of the original source and raw data format, thus making it readable and reusable.

Harmonization can be completed by applying the metadata ontology, which cannot be inferred when uploading data. The users have to diligently label their data, using the metadata tags listed in the previous section. Additionally, it will be necessary to label data with special tags for sample that originated it. These are the minimal set of the metadata required to make the data intelligible, however, users are free to add more tags, for example to clarify measurement conditions, and instrument sources. Users will also write details of their procedures in the notebook text, even if it is not strictly necessary and only for their own benefit.

With the minimal metadata tags (sample tag and ontology) it will be possible to pull the data systematically into tabular datasets, where the sample tag identifies row, and the ontology tag determines the column location of each entry.