D1.1. Implementation of raw-data parsers and common data processing into LABCORE notebook

Due date of submission: 31/08/2022

Actual submission date: 21/10/2022

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

Document Number:	D1.1
Document Title:	Implementation of raw-data parsers and common data processing into LABCORE notebook
Dissemination level	PU – Public
Period:	PR1
WP:	WP1
Task:	T1.1
Author:	Nanolayers OÜ
Abstract:	Nanolayers has extended the capabilities of its LabCore digital notebook platform in order to deal with MAST3RBoost partners data. Several parsers have been implemented, wherever possible, allowing users to extract data from raw instrument output files and upload it to LabCore notebooks, where it is stored in standardised and documented data records.

1 DATA PARSERS

With the help of all partners, Nanolayers has compiled a list of raw-data file sources provided by the experimental instrumentation used in the project, and the data to be found therein (Table 1).

Table 1: list of raw-data files produced by the various Mast3rBOOST experiments.

Experiment	File type
N_2 adsorption isotherms	smp
thermogravimetric analysis	xls
N_2 adsorption isotherms	txt
thermogravimetric analysis	txt
powder X-ray diffraction	txt
FTIR spectroscopy	dpt
scanning electron microscopy	tiff
tensile testing	xlsx
fatigue testing	txt/csv
optical microscopy	jpeg
scanning electron microscopy	jpeg
EBSD	jpeg
EDX	txt
Hardness measurements	xlsx
alicona topography	tiff
alicona topography	jpeg
alicona topography	xlsx

The vast majority of raw data consists of column-wise data tables, mostly stored in text files (*txt*, *csv*) or in Microsoft Excel sheets. LabCore has been extended with a flexible parser for data tables that can quickly parse text files and extract their data columns. Excel sheet (*xls*, *xlsx*) parsers are not yet available since there are multiple versions with different specifications. On the other hand, Excel can easily and reliably export its data sheets into text or *csv* files, thus making the implementation of dedicated parsers not a priority at this stage. Their development is however on its way to completion.

LabCore include several parsers for image files, such as output by optical and electron microscopes (*jpeg*, *tiff*), and extract image type data records. These can be displayed directly as they are in the notebooks, or processed as indicated in section 2.

There are only two file formats for which parsers could not be implemented: *smp* files from N2 adsorption isotherms, and *dpt* files from FTIR spectroscopy. As sometimes occurs, the instrument manufacturers did not release their raw-data file format specifications to the public, making their proprietary software the only one able to open those files. However, the data is again in tabular form, and the instruments software can export them to text files that LabCore can parse.

2   DATA PROCESSING

Most users are expected to continue using the analysis tools most familiar to them, and upload processed data into notebooks afterwards. However, LabCore was extended with fully customisable notebook elements where users can inject their own processing and visualisation functions, and operate arbitrarily on data upload into the notebooks.

Both the processing and visualisation elements accept python scripts from the users, providing all the flexibility and features available in its multitude of open-source libraries. User scripts are queued and executed on the LabCore server, and the clients show all results (or error messages) when the computation is completed.

Nanolayers will keep providing support to all interested project partners that are interested in porting existing analysis and visualisation tools into the digital notebook platform, or design novel ones from scratch.