nffa.eu

PILOT 2021 2026

DELIVERABLE REPORT

WP16 JA6 - Implementing FAIR data approach within NEP

D16.2

Report on the first data services

Due date

M18

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

PROJECT DETAILS

PROJECT ACRONYM	PROJECT TITLE
NEP	Nanoscience Foundries and Fine Analysis - Europe|PILOT

GRANT AGREEMENT NO:	FUNDING SCHEME
101007417	RIA - Research and Innovation action

START DATE

01/03/2021

WORK PACKAGE DETAILS

WORK PACKAGE ID	WORK PACKAGE TITLE
WP16	JA6 - Implementing FAIR data approach within NEP

WORK PACKAGE LEADER

Giuseppe Piero Brandino (eXact lab)

DELIVERABLE DETAILS

DELIVERABLE ID	DELIVERABLE TITLE
D16.2	Report on the first data services

DELIVERABLE DESCRIPTION

This document describes the initial set of data services available in NFFA Europe Pilot.

DUE DATE	ACTUAL SUBMISSION DATE
M18 31/08/2022	31/08/2022

AUTHORS

Marco Franzon (eXact lab), Elda Osmenaj (CNR-IOM), Mirco Panighel (CNR-IOM), Tommaso Rodani (AREA Science Park), Giuseppe Piero Brandino (eXact lab), Rossella Aversa (KIT)

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

PERSON RESPONSIBLE FOR THE DELIVERABLE

Dr. Marco Franzon (eXact lab)

NATURE

☒ R - Report

☐ P - Prototype

☐ DEC - Websites, Patent filing, Press & media actions, Videos, etc

☐ O - Other

DISSEMINATION LEVEL

☒ P - Public

☐ PP - Restricted to other programme participants & EC: (Specify)

☐ RE - Restricted to a group (Specify)

☐ CO - Confidential, only for members of the consortium

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

REPORT DETAILS

ACTUAL SUBMISSION DATE	NUMBER OF PAGES
31/08/2022	**15**

FOR MORE INFO PLEASE CONTACT

Giuseppe Piero Brandino,
eXact lab
via Crispi, 56 – 34126 Trieste
(Italy)

email: brandino@exact-lab.it

VERSION	DATE	AUTHOR(S)	DESCRIPTION / REASON FOR MODIFICATION	STATUS
1	08/08/2022	Marco Franzon Elda Osmenaj Mirco Panighel Tommaso Rodani Giuseppe Piero Brandino Rossella Aversa		Draft
2	26/08/2022	Marco Franzon Elda Osmenaj Mirco Panighel Tommaso Rodani Giuseppe Piero Brandino Rossella Aversa		Final version
				Choose an item.

CONTENTS

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

INTRODUCTION

The present deliverable describes the initial set of data services made available to the NEP community. It is structured in four main sections: Section 1 describes the Data Stewardship Wizard (DSW) and its functionalities, Section 2 the STM Neural Network, Section 3 the STM Metadata Explorer and Section 4 the SEM Classifier.

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

DATA SERVICES

1. The Data Stewardship Wizard (DSW)

The Data Stewardship Wizard (DSW) is an open-source questionnaire tool that ensembles all the answers in a Data Management Plan (DMP). It is based on FAIR data stewardship, covering all the different aspects of data management and FAIR principles from data collection to publication and preservation. Through the smart questionnaire, users are guided by resources, recommendations and definitions when filling the answers.

A self-hosted DSW instance was deployed for the NEP project to support researchers on creating the DMP for their laboratories. The instance can be reached at: https://dsw.nffa.eu/ and accessed using NFFA-Europe credentials and login as reported in Figure 1. It has three types of users: Admin, Data steward and Researcher.

Figure 1. DS Wizard Login page

After login, the instance provides a dashboard page with a sidebar menu (Figure 2), containing:

- **Knowledge Models**: consisting of templates of DS Wizard and the personalized templates created for the NEP project
- **Projects**: contain the list of the DMPs created using the instance
- **Create new DMP**: by using the last version of the Knowledge Model here the user can start creating the DMP by answering the questionnaire
- **Quickstart Guide**[1]: a pdf file with step by step instructions for using the tool
- **Video tutorial**[2]: a brief video demonstration on how to fill the DMP

[1] Quickstart Guide document: https://datashare.nffa.eu/index.php/s/e534WHsR8wBLpyr
[2] Video tutorial: https://datashare.nffa.eu/s/Tf9LRcBqqwFHQFW

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

Figure 2. DS Wizard dashboard

The Knowledge model (KM) represents the content of the DSW that is used to create a Questionnaire that researchers fill in to generate a DMP. It contains different chapters with questions, external resources, definitions and recommendations.

Knowledge Model Editor: the DSW contains a system to maintain Knowledge Models and to adapt or create new ones from scratch. This functionality is available for Data Steward users. Using this system, a personalized Knowledge Model was customized for NEP Laboratories with specific questions, references and recommendations in line with FAIR principles and Open Access obligations (Figure 3). By using the KM editor, Knowledge Models can be edited directly in the wizard.

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

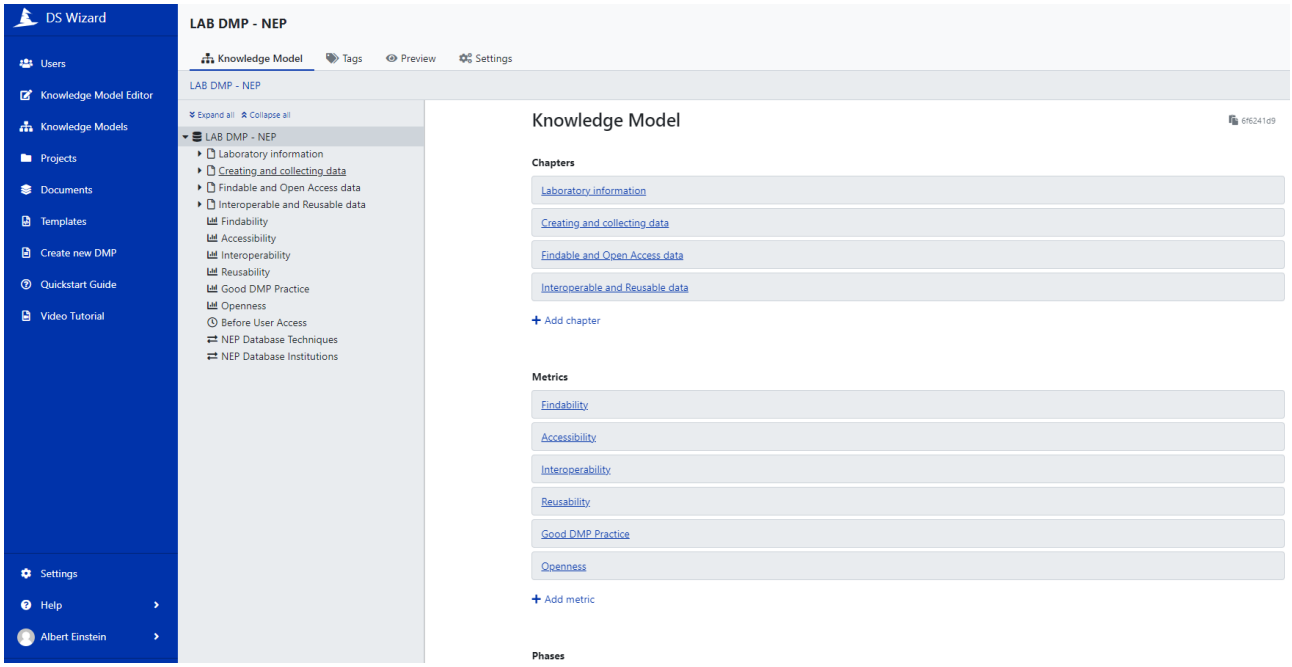


Figure 3. DS Wizard Knowledge Model Editor used to create Lab-NEP Knowledge Model

DS Wizard Questionnaire: contains questions and provides integrations in form of recommendations, definitions and resources. Most of the questions are closed, with options or lists of items, while a limited number are open questions (Figure 4).

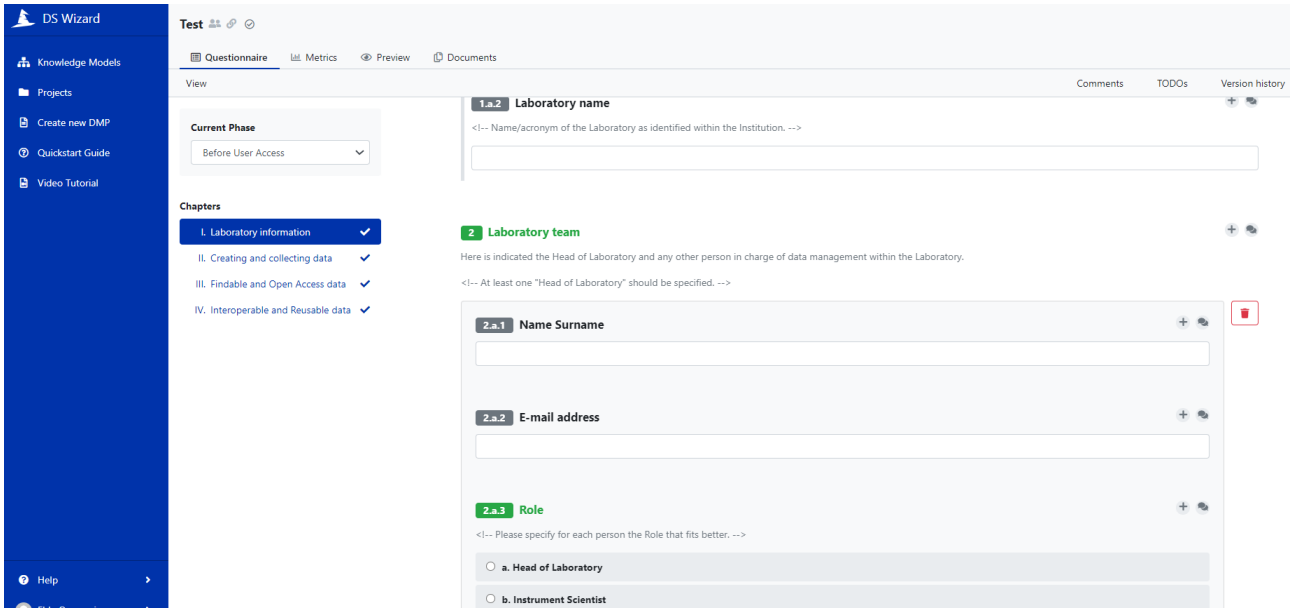


Figure 4. DS Wizard questionnaire based on the customized Knowledge Model created for the NEP project

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

For some questions, FAIR metrics were used as indicators to give an estimation of the FAIRness of data as by the summary report in Figure 5.

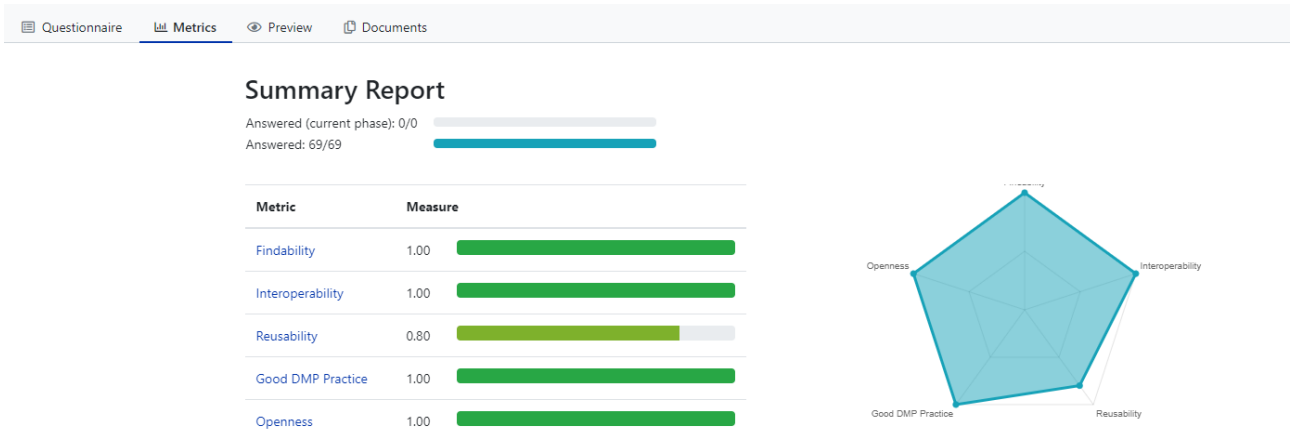


Figure 5. DS Wizard FAIR metrics

Data Management Plan (DMP): DSW assembles all the answers and generates a DMP document by using a selected template. The DMP can be exported in PDF, MS Word, HTML, LaTeX or JSON format, allowing the machine-actionability of the DMPs as reported in Figure 6. The DMP generated in the selected format can be downloaded and shared.

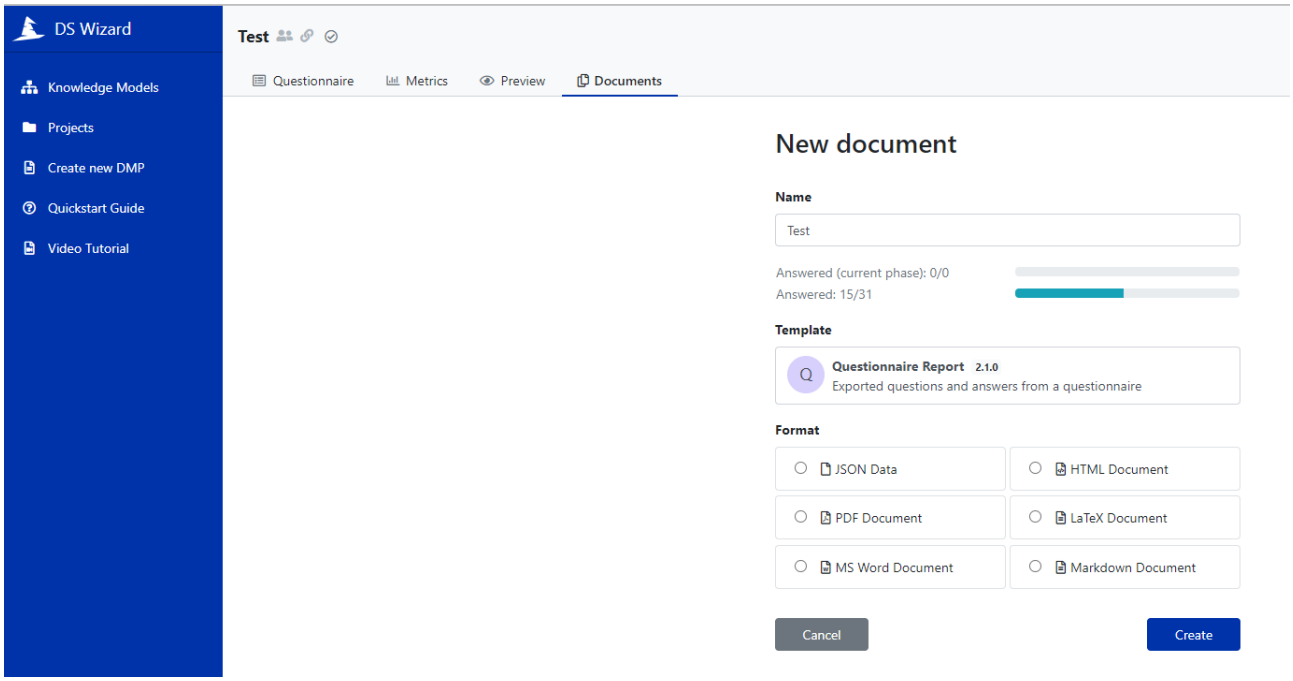


Figure 6. DS Wizard DMP generation formats

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

2. STM Neural Network

STM Neural Network is a machine learning based service for content-based image retrieval (CBIR). Content-Based Image Retrieval (CBIR) aims to find the similar images from a large scale dataset against a query image. The similarity between the representative features of the query image and dataset images is used to rank the images for retrieval. Convolutional Neural Networks (CNN) are especially suitable feature extractors for CBIR. Rather than using the classification output of the network, the information is taken from a layer prior to the output layer. These vectors contain image descriptive properties and are therefore used as a feature for CBIR.

STM Neural Network is used to retrieve a subset of Scanning Tunneling Microscopy (STM) images similar to an input STM image. Researchers use this service to efficiently query a dataset of STM images to quickly find images relevant for their purposes.

STM Neural Network is implemented in Python, uses FastAI Machine Learning library and spym [3] package to load, process and plot images, and FASTAPI for the creation of REST endpoints. Users can programmatically use the service within a jupyter notebook in a few lines of code, as shown in Figure 7.

```
# single image similarity example:
# 85404 is the image unique identifier, 24 is the number of similar images to extract
similar_images_df = get_similar_images_euclidean(img_repr_df, 85804, 24)

# plot similar images with input image in the center
show_similar_images(similar_images_df)
```

Figure 7: STM Neural Network example query

In the example query above, the image with a unique identifier of 85804 is used as input to search for 24 similar images in the STM dataset. Images retrieved are presented in a plot with the input image in the center (Figure 8).

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

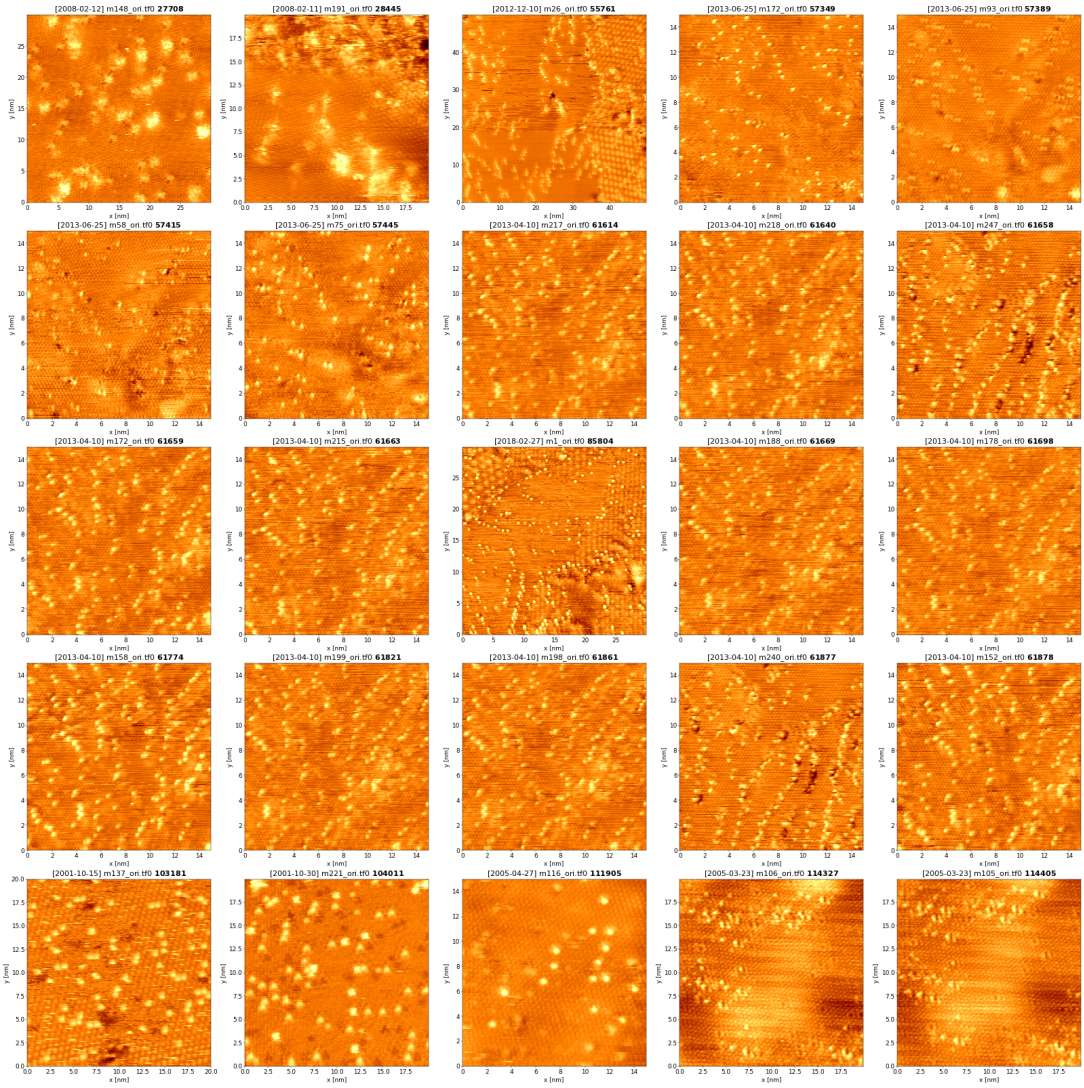


Figure 8: STM Neural Network extracted images from example query. The input image is the one in the center.

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

3. STM Metadata Explorer

The STM Metadata Explorer is an easy-to-use and interactive web service integrated within the Trieste Advanced Data Services (TriDAS https://tridas.nffa.eu/). By selecting a small subset of metadata that provides significant information about the measurement settings, it allows to visually explore the STM images through interactive and downloadable plots.

The workflow is summarized in Figure 9 and explained in detail in a guidelines document [1]. Users can visually explore images metadata through a quantile plot for a single metadata field or a scatter plot showing the distribution of images between two chosen metadata fields. In these interactive plots, some information about the metadata pop up by hovering on top of them. In particular, from scatter plots it is possible to retrieve a table containing metadata fields for each image in the subset, which can be selected, ordered and filtered based on their metadata values. These features are useful for a first exploration of the dataset but images can also be downloaded for further analysis and image visualization: by clicking on the unique identifier, the corresponding STM image is shown, allowing to download the related data, metadata and provenance.

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

Figure 9. The workflow of the webservice on the TriDAS website

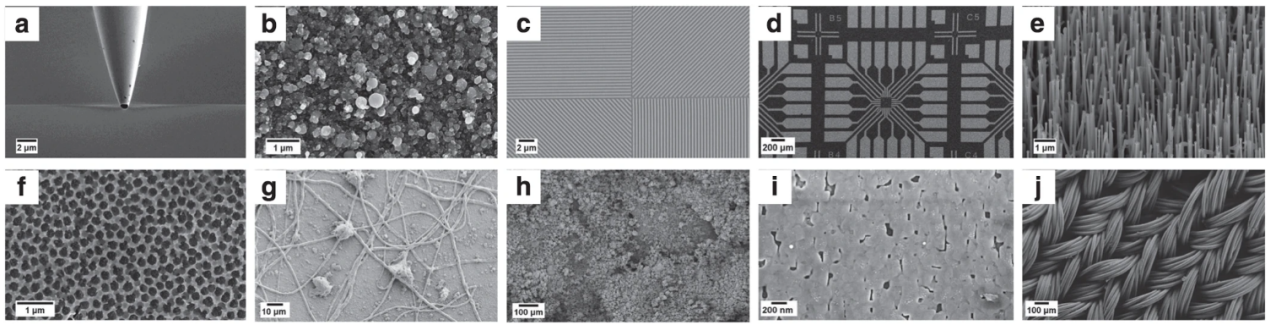
TriDAS is implemented in Python, and uses Bokeh [2] for data visualization, spym [3] package to process and plot images, and Flask [4] framework as backend. The dataset of STM images together with their provenance description [5] and all the source code of the software developed [6] are publicly available.

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

4. SEM Classifier

Scanning Electron Microscopy (SEM) classifier is a data service web service integrated within the Trieste Advanced Data Services (TriDAS https://tridas.nffa.eu/) that allows researchers to automatically classify and tag SEM images in a uniform way. After the classification done in collaboration with SEM researchers, 10 categories that cover a broad range of SEM images were identified: Biological, Fibres, MEMS devices and electrodes, Particles, Porous Sponge, Tips, Films and Coated Surfaces, Nanowires, Patterned surfaces, Powder (Figure 10). The classification is performed using convolutional neural networks (CNN) [7].

Figure 10. Representative images for each of the ten categories



The researcher can upload an image or a folder and after a few seconds per image can obtain the resulting categories, together with the rate (a measure of "certainty").

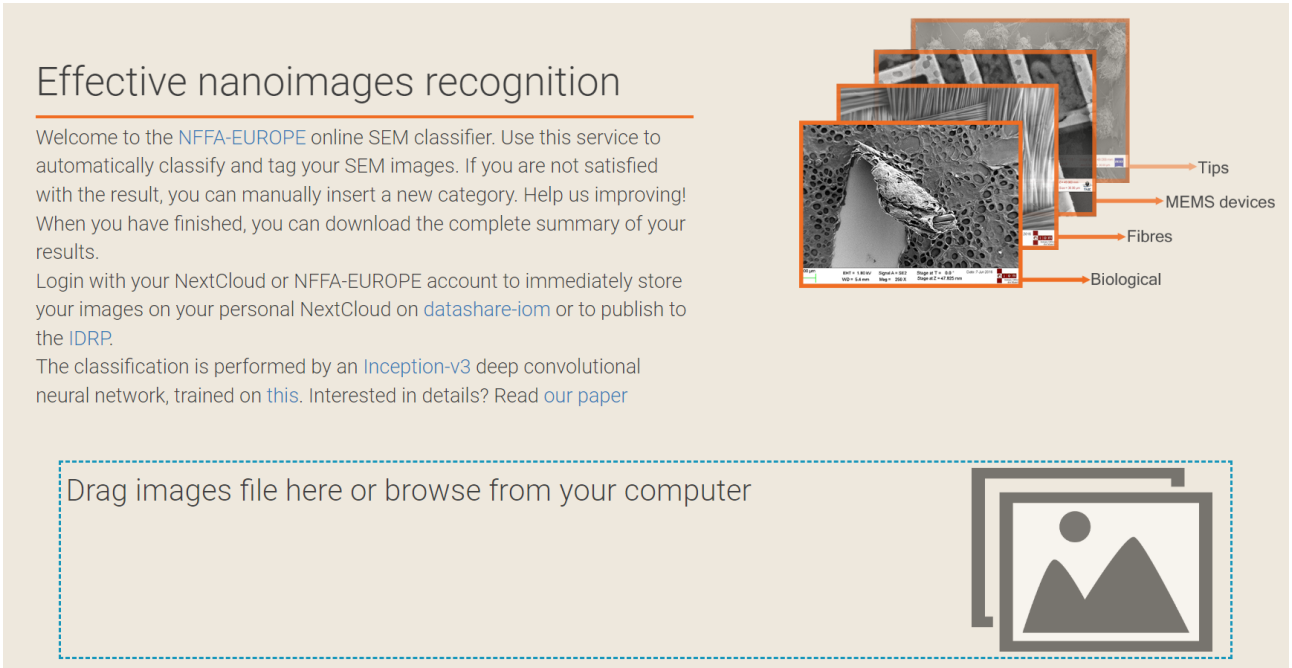


Figure 11. The SEM Classifier home page

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

The researcher can confirm and download the result or change it by adding new categories.

The SEM classifier is implemented in Python, uses TensorFlow Machine Learning library and FASTAPI for the creation of REST endpoints.

References

[1] Osmenaj, Elda, Antonio, Fabrizio, Kulüke, Marco, Andres, Pablo de, & Cozzini, Stefano. (2022). UC1 Guidelines for researchers: Defining Procedures and Services to enforce data provenance for thematic communities and beyond (1.0.0). Zenodo. 10.5281/zenodo.6827750.

[2] Bokeh Development Team, Bokeh: Python library for interactive visualization, 2020. [Online]. Available: https://bokeh.org/.

[3] Panighel, M., Rescipy-project/spym: V0.7.0, version v0.7.0, 2021. 10.5281/zenodo.5792909.

[4] Grinberg, M., Flask web development: developing web applications with python. "O'Reilly Media, Inc.", 2018.

[5] Rodani, T., Osmenaj, E., Cazzaniga, A., Panighel, M., Africh, C., and Cozzini, S., Dataset of Scanning Tunneling Microscopy (STM) images of graphene on nickel, version 1.0, Zenodo, 2021. 10.5281/zenodo.5799773.

[6] Tommaso Rodani. t0m-R/STM images 0.1.0. Version 0.1.0. Sept. 2020. 10.5281/zenodo.4019640.

[7] Modarres, M.H., Aversa, R., Cozzini, S. *et al.* Neural Network for Nanoscience Scanning Electron Microscope Image Recognition. *Sci Rep* 7, 13282 (2017). 10.1038/s41598-017-13565-z.

This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project