nffa.eu
nanoscience foundries & fine analysis

The Challenge to define and use a metadata schema within the NFFA-EUROPE project

Stefano Cozzini

CNR – IOM

Istituto Officina dei Materiali

NOMAD-FAIRDI Workshop, Berlin 10.07.2019

NFFA-Europe has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654360

Agenda

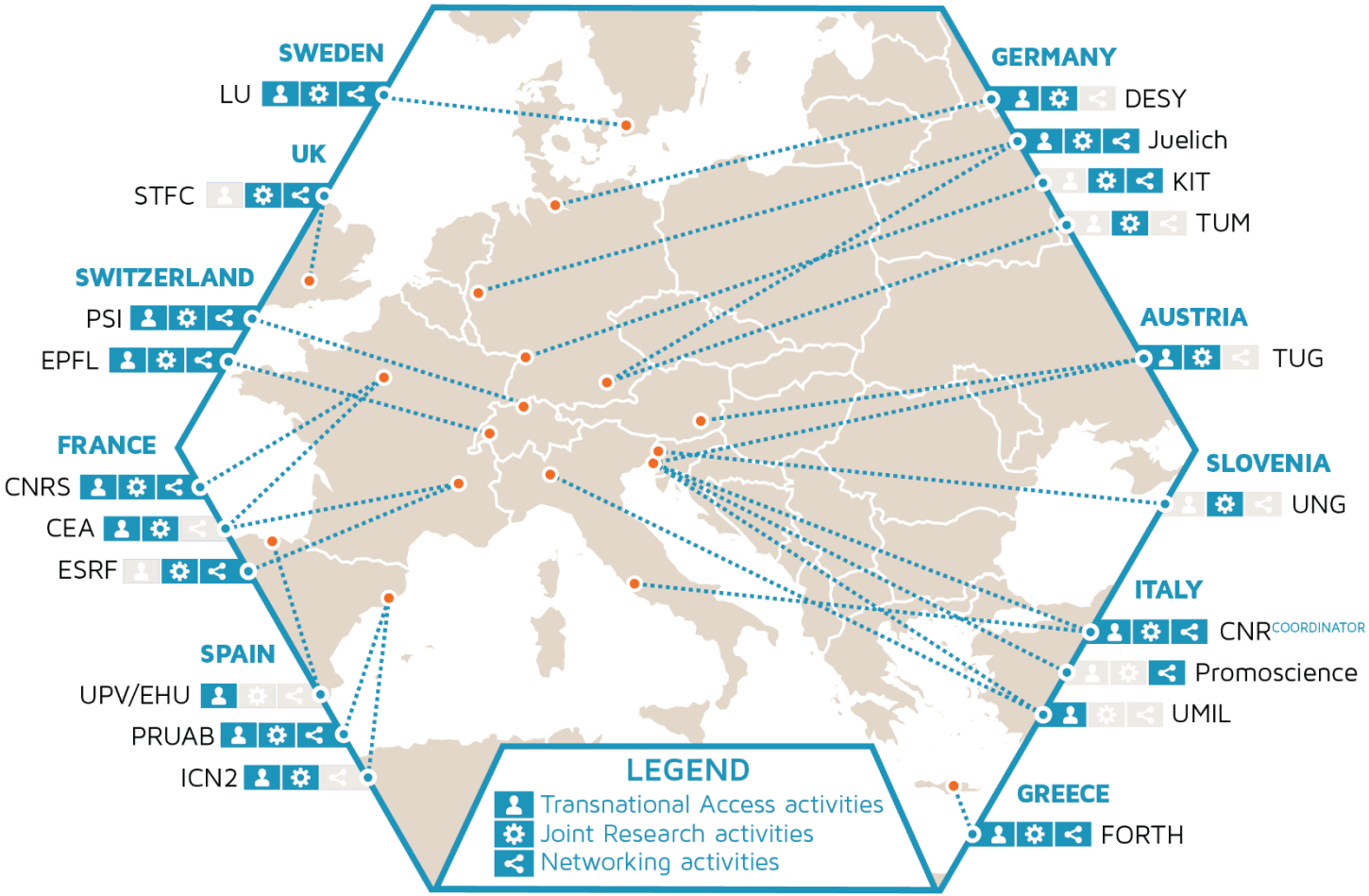
- Introduction: NFFA-EUROPE project
- NFFA Meta Data
- Data Repository for NFFA-EUROPE project
- Data service on the top of the repository:
  - Classify Scanning Electron Microscope (SEM) images at the nanoscale.
- Conclusions & perspectives

NFFA-EUROPE in a nutshell

- It provides the widest range of tools for research at the nanoscale
  - ~ 170 instruments/facilities
- Free transnational access to academia & industry
- 4 year EU Funded project
  - Sept 2015- Sep ~~2019~~ 2020
- 20 EU partners
  - Coordinated by CNR-IOM

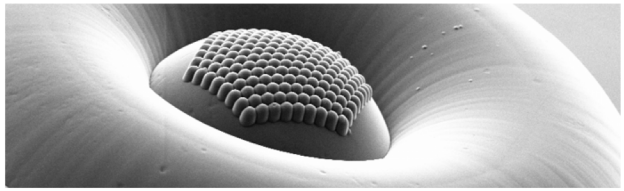
The consortium

Coordinated by CNR-IOM

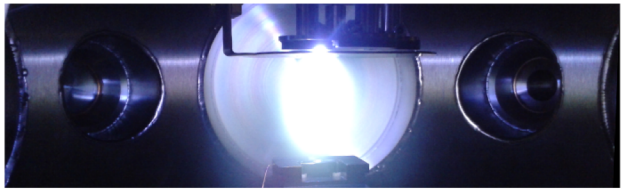


Open Access

- Regulated access: Users submit proposals and if accepted they use resources

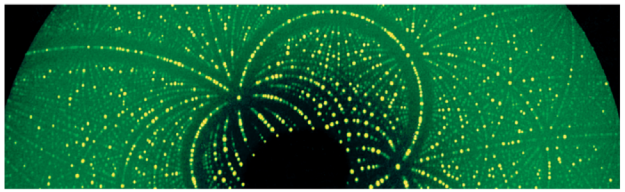


LITHOGRAPHY & PATTERNING



GROWTH & SYNTHESIS

THEORY & SIMULATION



CHARACTERISATION

NFFA-EU Data management

JRA3:

e-Infrastructure for data and information management

A transversal activity devoted to the setup of the first ***Information and Data Repository Platform*** (IDRP) for Nano science

- Definition of new metadata standards for data sharing in nanoscience
- Automatic acquisition of key metadata and create a data repository for future data access

Data infrastructure is complemented **by Data Analysis Services.**

Metadata..

- Metadata schema for nanoscience has been discussed and proposed:

WP11 - NA – Innovation and networking activities

D11.2 - Draft metadata standard for nanoscience data

22 Feb 2017

Vasily Bunakov, Tom Griffin, Brian Matthews (STFC), Stefano Cozzini (CNR-IOM)

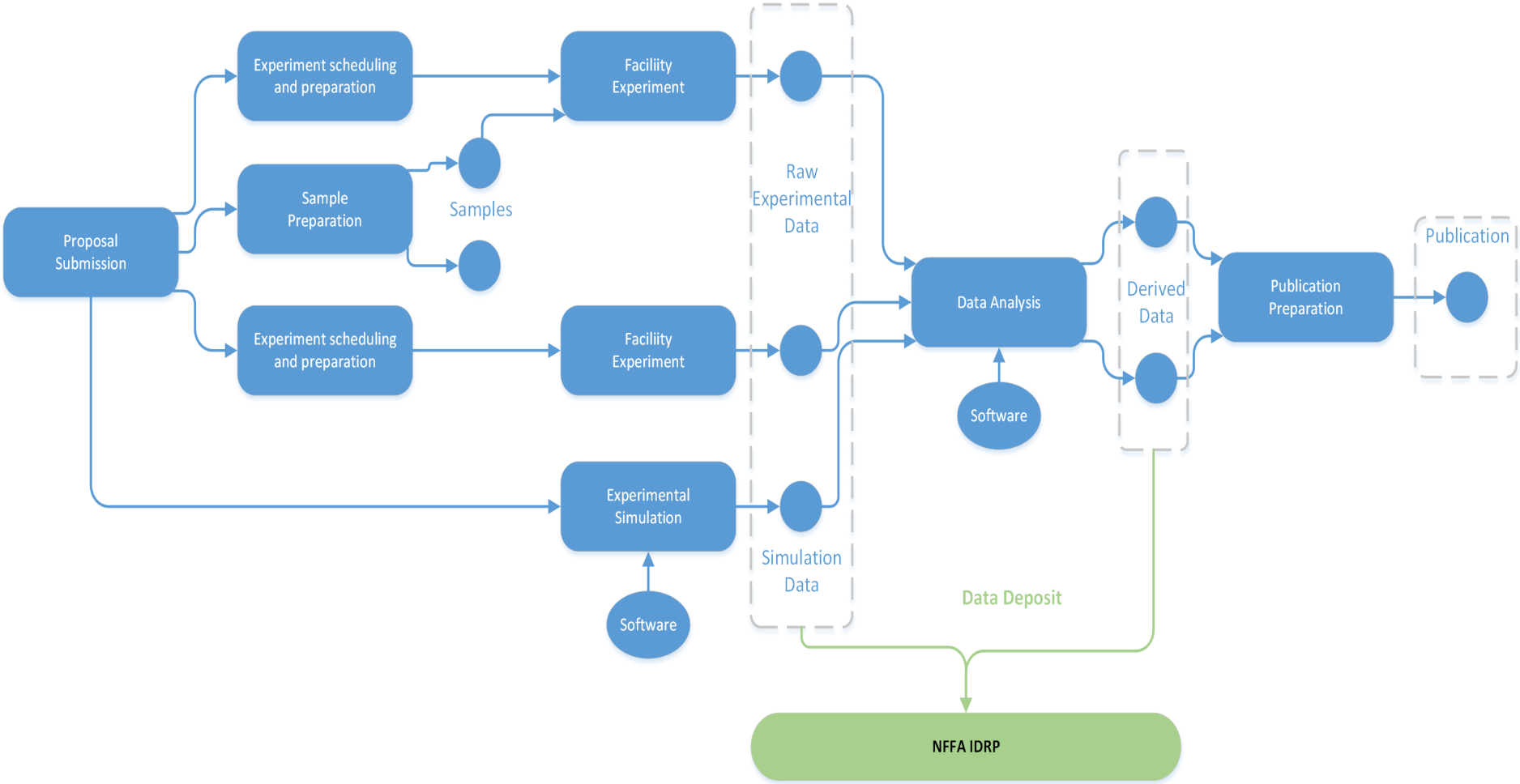
D11.14 - Final metadata standard for nanoscience data

28 Feb 2018

Vasily Bunakov (STFC), Brian Matthews (STFC), Thomas Jejkal (KIT), Rossella Aversa (CNR-IOM), Stefano Cozzini (CNR-IOM)

- currently under discussion within the community (beyond NFFA-EUROPE)

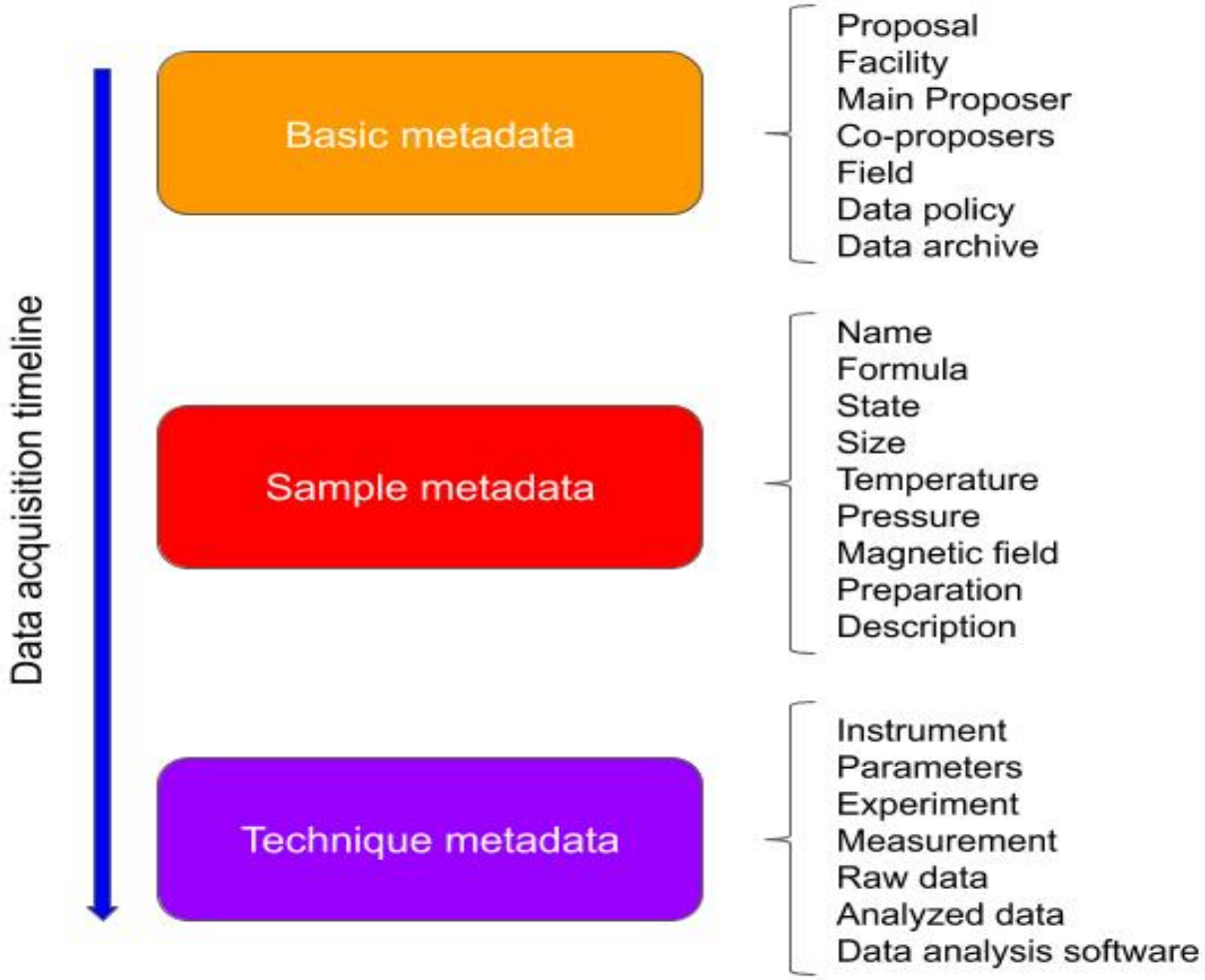
NFFA workflow



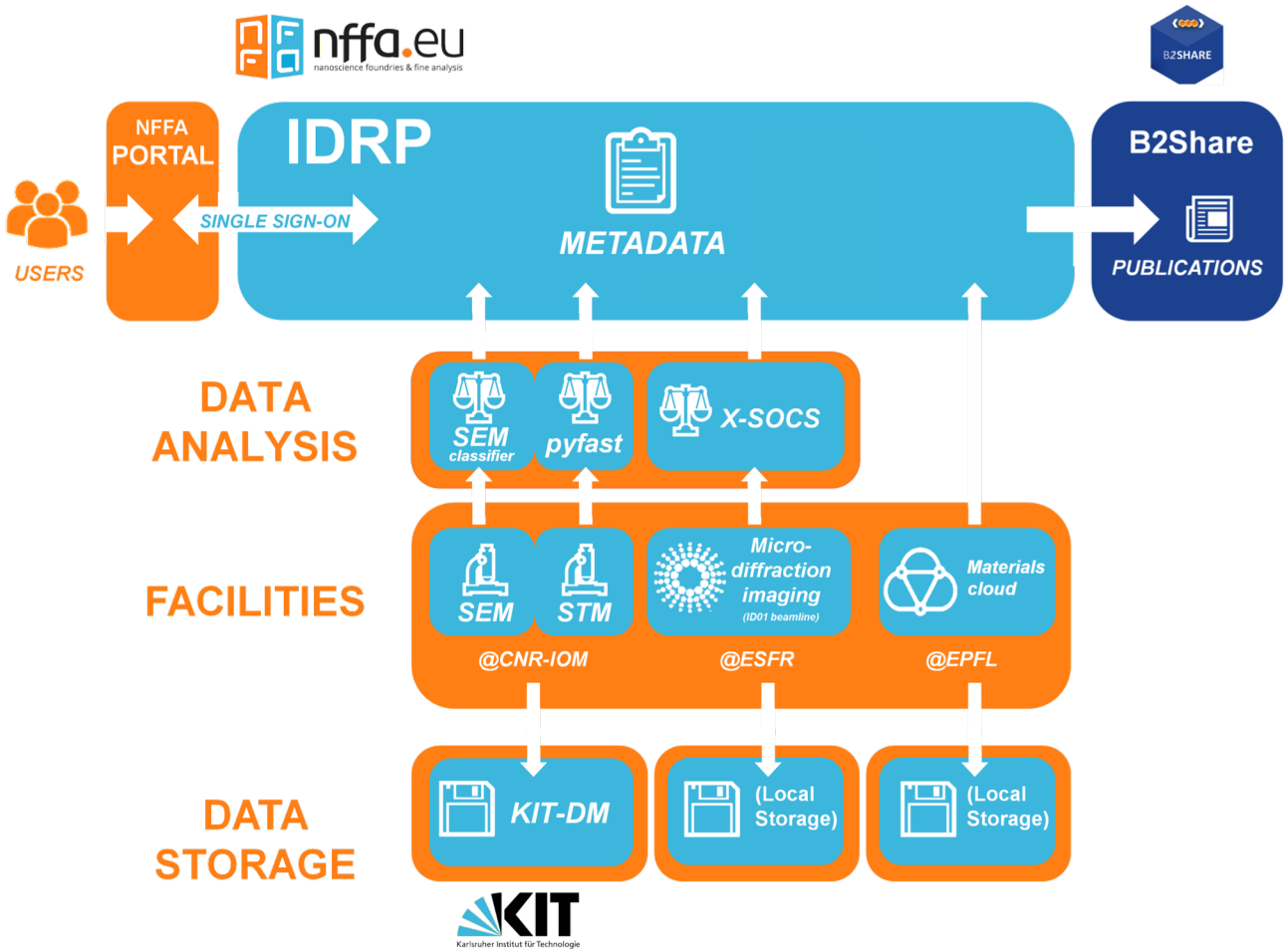
Our approach

- **pull-based import** where the IDRP takes care of mapping external metadata models into the NFFA metadata model.
- PRO:
  - minimization of the effort for the users and maintainers of local data archives.
- Cons:
  - Need to map all models

Elements of the Metadata schema



NFFA IDRP infrastructure



The portal: portal.nffa.eu

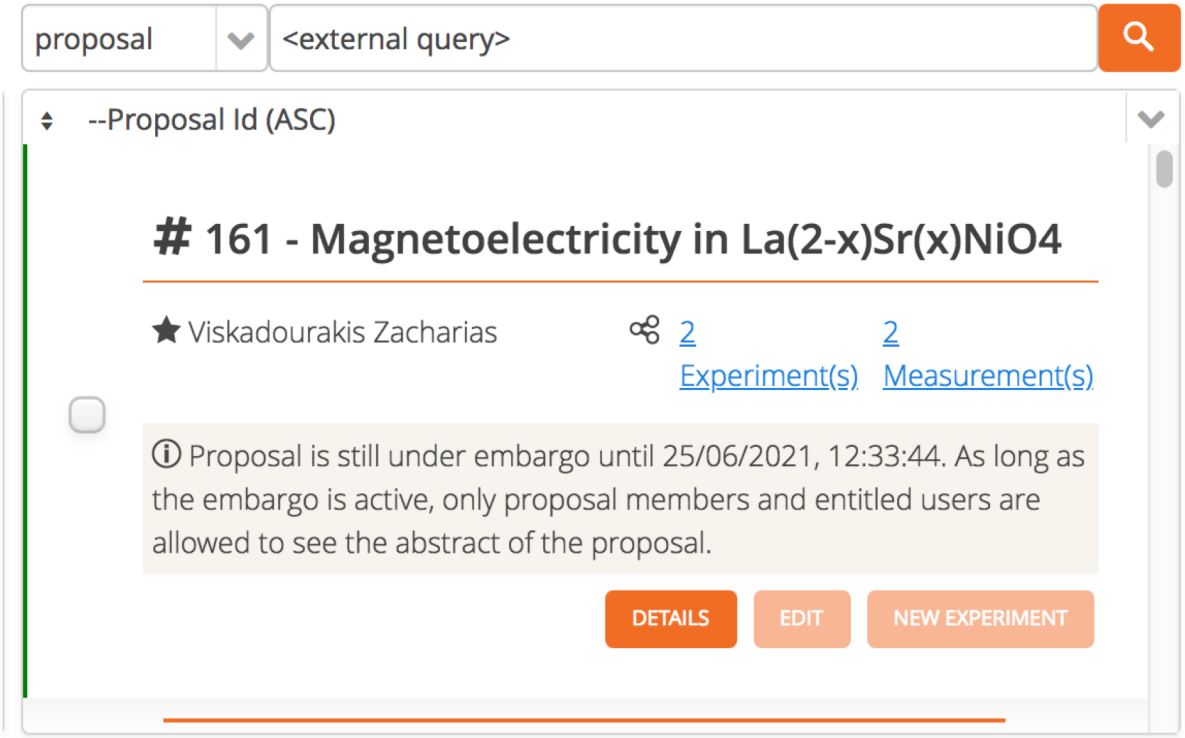
R. Aversa ▾ | IDRP | your applications | your wishlist | submit your application

ABOUT the project | **OFFER** tools catalogue | **NEWS** events & highlights | **APPLY** guidelines

The IDRP: idrp.nffa.eu

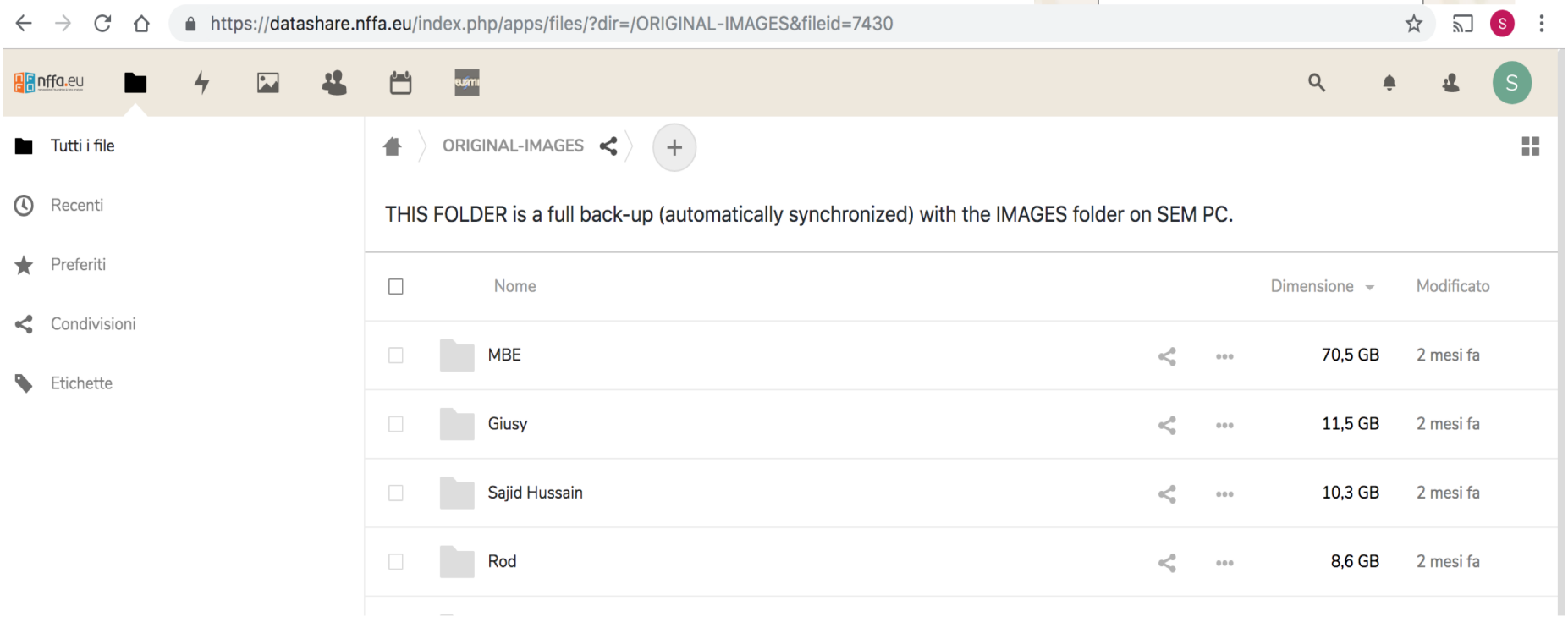
Register, manage and retrieve metadata and data stored in the local repositories

Manage authorization of metadata and data access



IOM Repo: datashare.nffa.eu

The Data Management System at the Local Facility adopted at CNR-IOM is **NextCloud based**



The challenge/ issues

- Make aware users of all this
- Do not scare users on using what we are proposing
  - Overhead cost some time underestimated

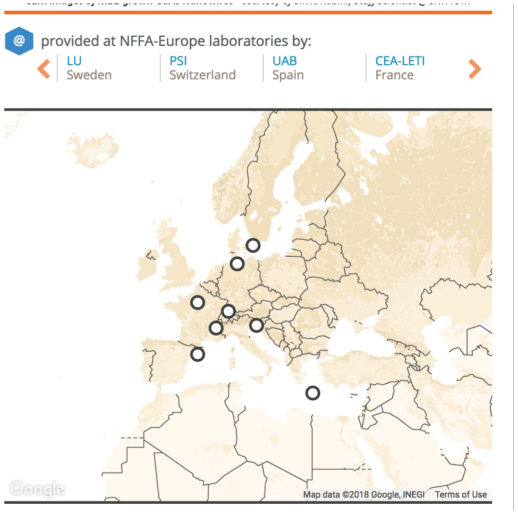
- Are we all ready to FAIR data ?
  - Small communities/large communities
  - Computational/ experimental links
  - Shared data/ closed data

A case study: classifying SEM images by Neural networks

Our Issue: SEM images

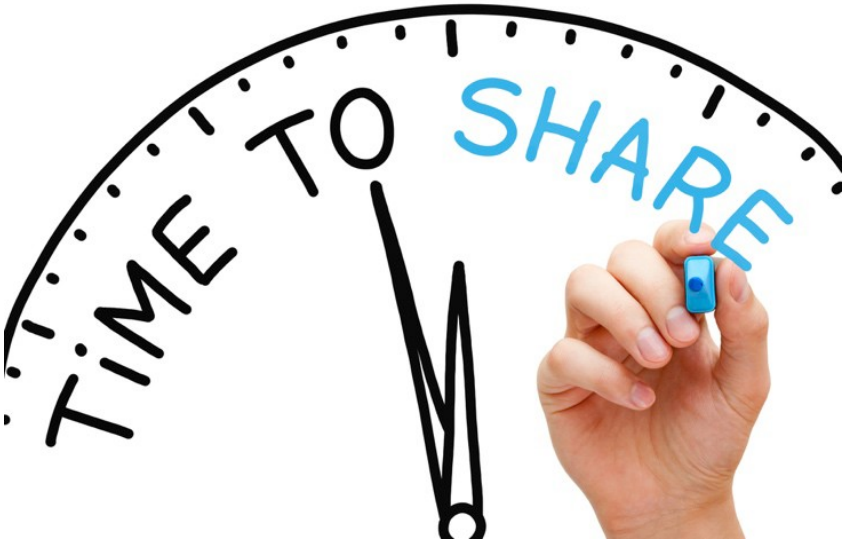
- One SEM Available at CNR-IOM Trieste with 150,000 images NOT classified

- 10 SEM across European partners … a sizeable amount of images



Sharing images is nice..

- A couple of million nano images can be of some help for some nanoscience..



But before doing that we need to start classifying them…

SEM images classification steps

- STEP 1: Classify images (scientific skills)
- STEP 2: Train a neural network (deep learning)
- STEP 3: Use the network as classifier (inference)
  - Semi - Automatic tool for SEM users
  - Massive process of all the images

Step I: classify images..

We created and manually annotated **the first dataset** of classified SEM images (18,577 images).

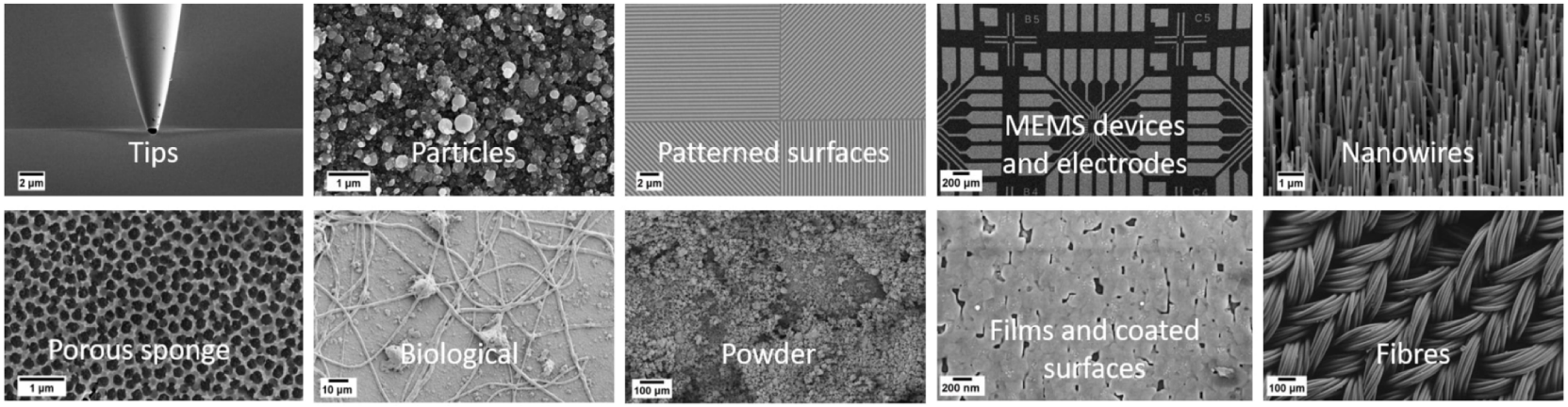
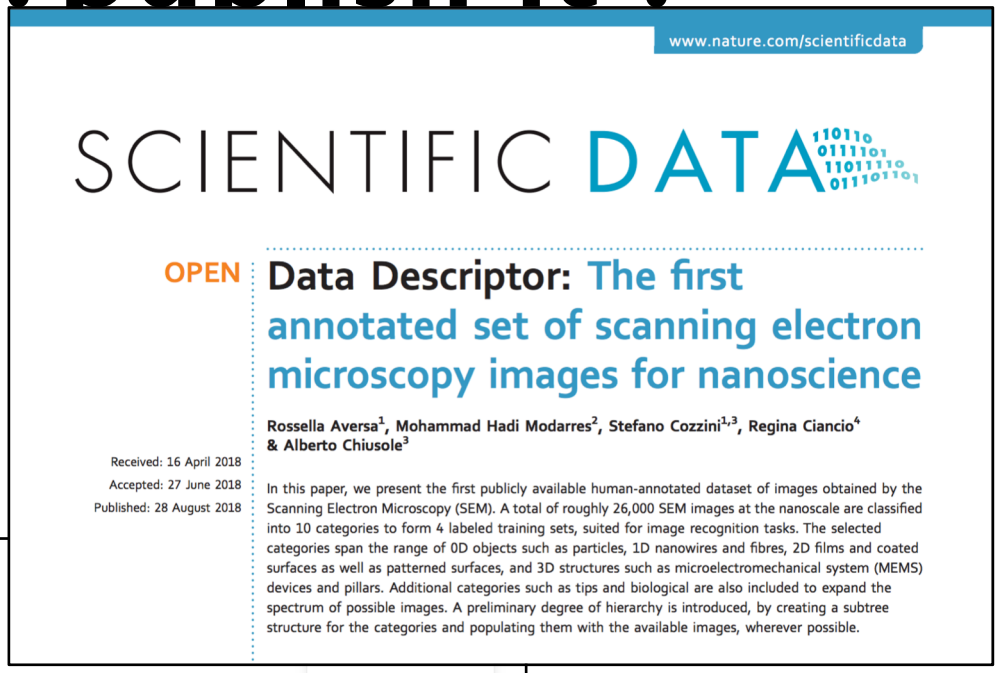


Figure 1. Categories chosen for SEM images. The dimensionality of nanoscience objects provided the basis for the choice. Other categories, such as `Biological` and `Tips` were added as these were common images found in the SEM database.

Step 1b: publish it !



www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

Data Descriptor: The first annotated set of scanning electron microscopy images for nanoscience

Rossella Aversa[1], Mohammad Hadi Modarres[2], Stefano Cozzini[1,3], Regina Ciancio[4] & Alberto Chiusole[3]

Received: 16 April 2018
Accepted: 27 June 2018
Published: 28 August 2018

In this paper, we present the first publicly available human-annotated dataset of images obtained by the Scanning Electron Microscopy (SEM). A total of roughly 26,000 SEM images at the nanoscale are classified into 10 categories to form 4 labeled training sets, suited for image recognition tasks. The selected categories span the range of 0D objects such as particles, 1D nanowires and fibres, 2D films and coated surfaces as well as patterned surfaces, and 3D structures such as microelectromechanical system (MEMS) devices and pillars. Additional categories such as tips and biological are also included to expand the spectrum of possible images. A preliminary degree of hierarchy is introduced, by creating a subtree structure for the categories and populating them with the available images, wherever possible.

NFFA-EUROPE - SEM Dataset

by Aversa, Rossella; Modarres, Mohammad Hadi; Cozzini, Stefano; Ciancio, Regina;

Feb 19, 2018

Last updated at Apr 16, 2018

SeriesInformation: Dataset of 18,577 SEM images produced at CNR-IOM (Trieste, Italy). Images are classified into 10 categories in a folder structure, which have been used for convolutional neural network training. Results obtained from this dataset have been published in Modarres et al., Scientific Reports volume 7, Article number: 13282 (2017), doi:10.1038/s41598-017-13565-z The dataset is appropriate for the purposes of this study and in general for visual object recognition software research. Any scientific metadata associated to the measure is not present in the images. The dataset is therefore relevant as a whole, being the single images entirely detached from any specific information or scientific detail related to the displayed subject. This work has been done within the NFFA-EUROPE project (www.nffa.eu) and has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 654360 NFFA-Europe.

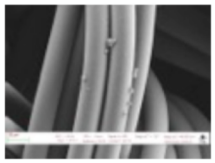
Disciplines: 4.1.12.1 → Computer graphics → Image processing; 4.1.16.3 → Information science → Database; 4.1.17 → Computer sciences → Artificial intelligence; 4.1.17.1.2 → Cognitive science → Machine learning; 5.6.37 → Engineering → Nanomaterials; 3.4.7 → Physics → Condensed matter physics;

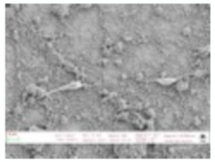
Keywords: Machine Learning; Convolutional Neural Networks; Image Classification; Nanoscience; Scanning Electron Microscopy;

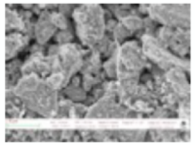
DOI: 10.23728/b2share.19cc2afd23e34b92b36a1dfd0113a89f Copy
PID: 11304/7f138d60-c8ab-4209-9ed5-18f47323f8e6 Copy

EUDAT

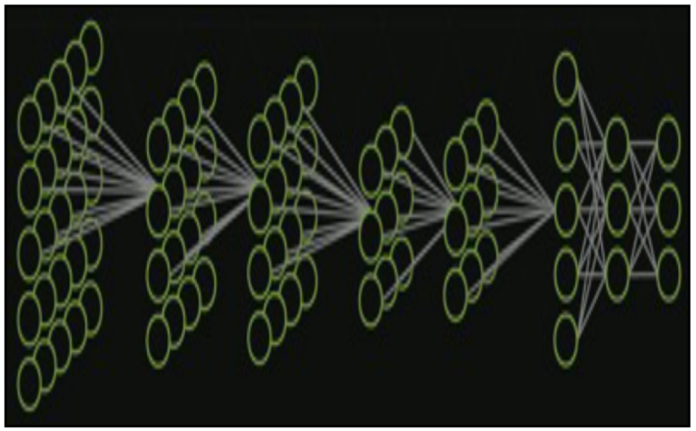
Step 2: train the network







Deep Neural Network



Cantilever Tips

Biological

Powder

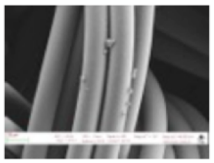
(10 classes)

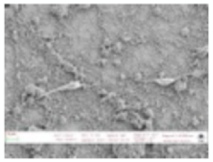
Unknown Images

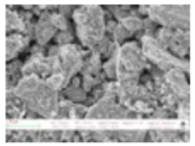
Something happens

Images with labels!!

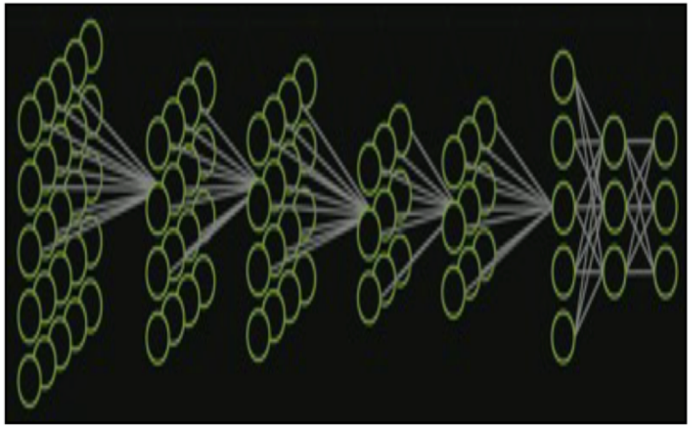
Step2 : the tools/infrastructure…







Deep Neural Network



Cantilever Tips

Biological

Powder

(10 classes)

TensorFlow

nervana

docker

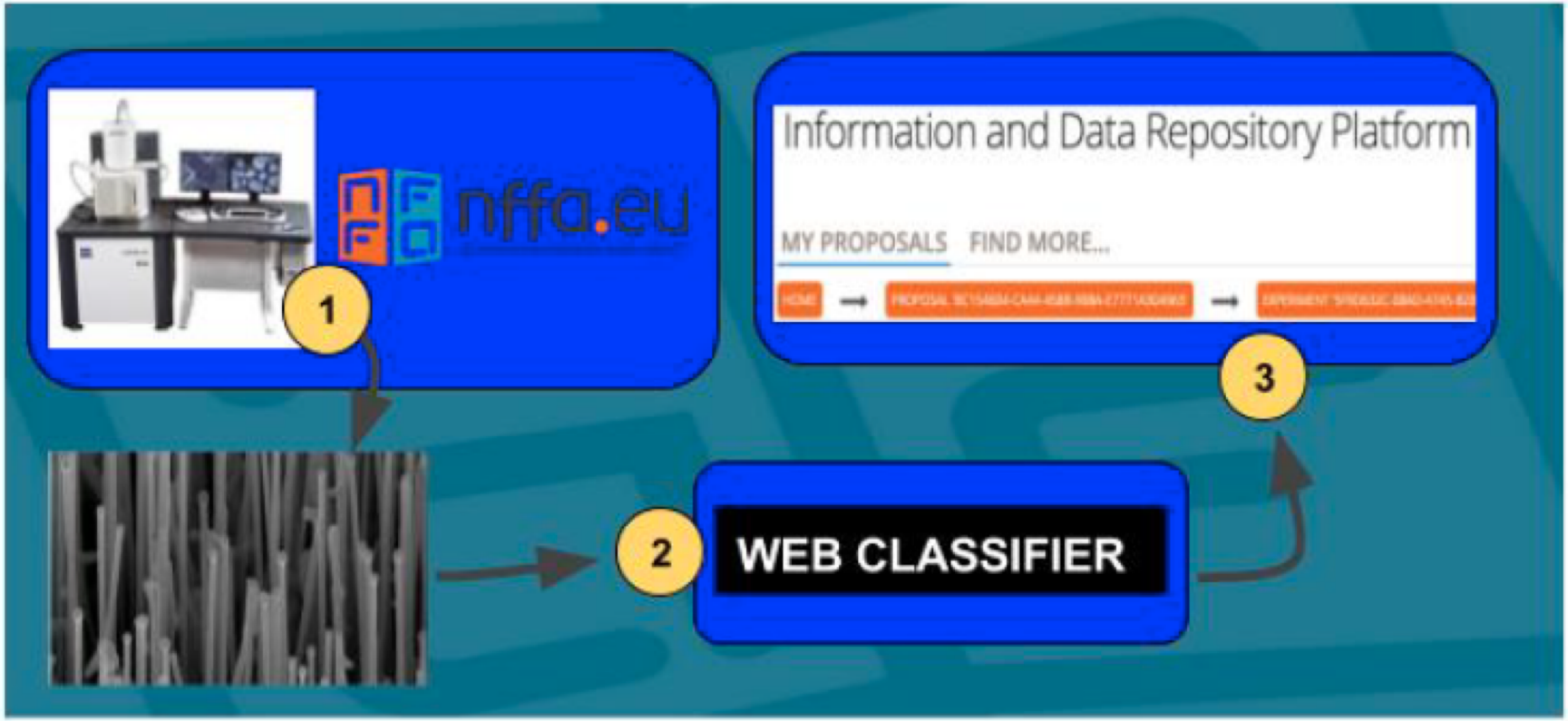
·l·u·s·t·r·e·
File System

HPC infrastructures

Project
HPC on-demand ;-)

CINECA

Step 3: data analysis services



sem-classifier.nffa.ue

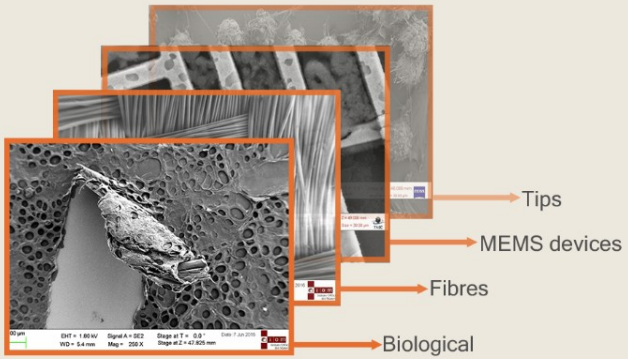
nffa.eu

ONLINE SEM IMAGE CLASSIFIER

Automatically tag your Scanning Electron Microscopy images

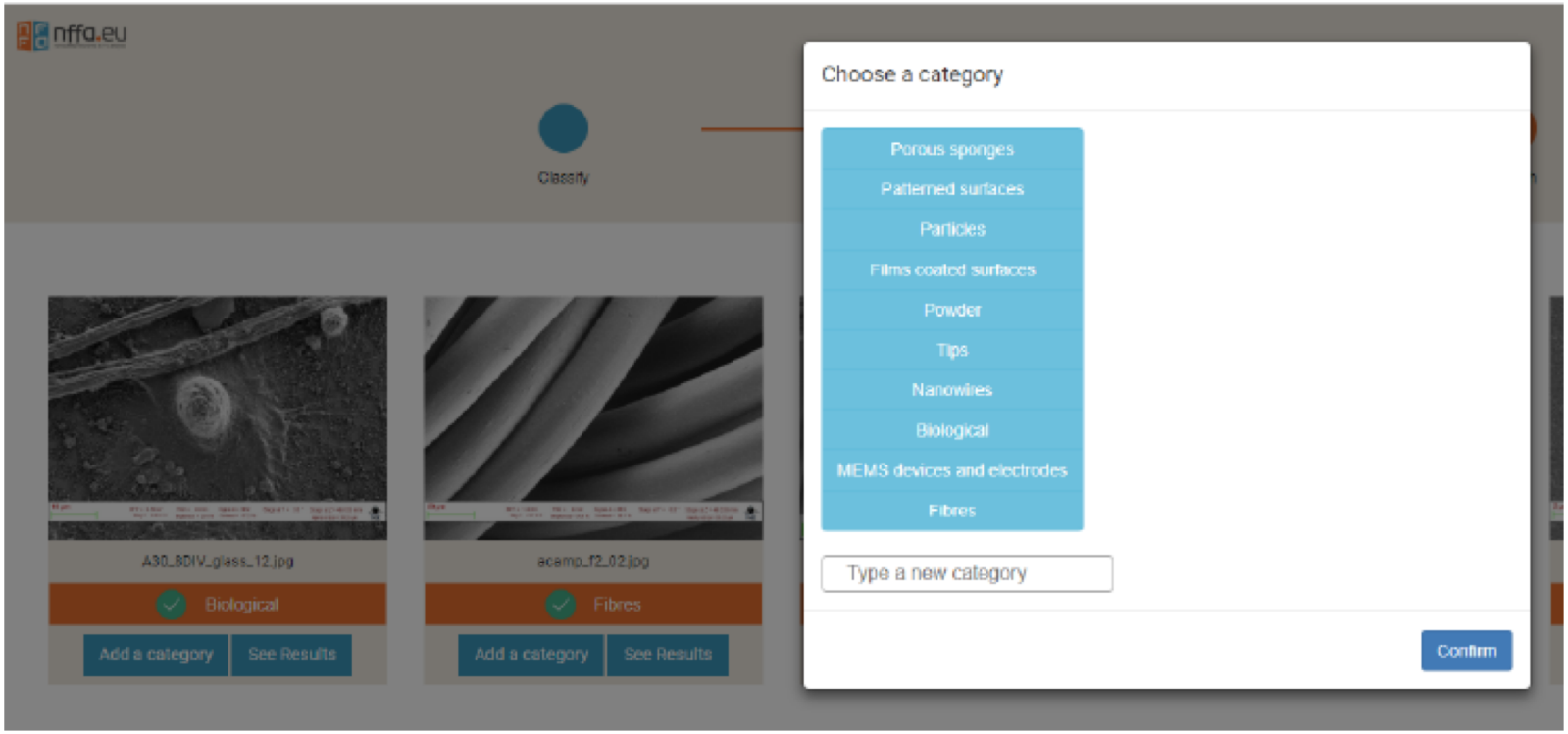
Effective nanoimages recognition

Welcome to the NFFA-EUROPE online SEM classifier. Use this service to automatically classify and tag your SEM images. If you are not satisfied with the result, you can manually insert a new category. Help us improving!
When you have finished, you can download the complete summary of your results.
Login with your NextCloud or NFFA-EUROPE account to immediately store your images on your personal NextCloud on datashare-iom or to publish to the IDRP.
The classification is performed by an Inception-v3 deep convolutional neural network, trained on this dataset.
Interested in details? Read our paper



Drag images file here or browse from your computer

SEM images classification



What next ?

- Unsupervised learning approach: can we automatically detect new categories without manual input ?
- We are exploring this approach on our SEM data set..

The procedure

- Use a general Deep Learning Network to extract features (2048/4096)
  - Use of the ad-hoc algorithm to enhance the quality of features
- Dimensional Reduction (if needed)
- Clustering on the obtained dataset
- Look at the clusters (together with the scientists)

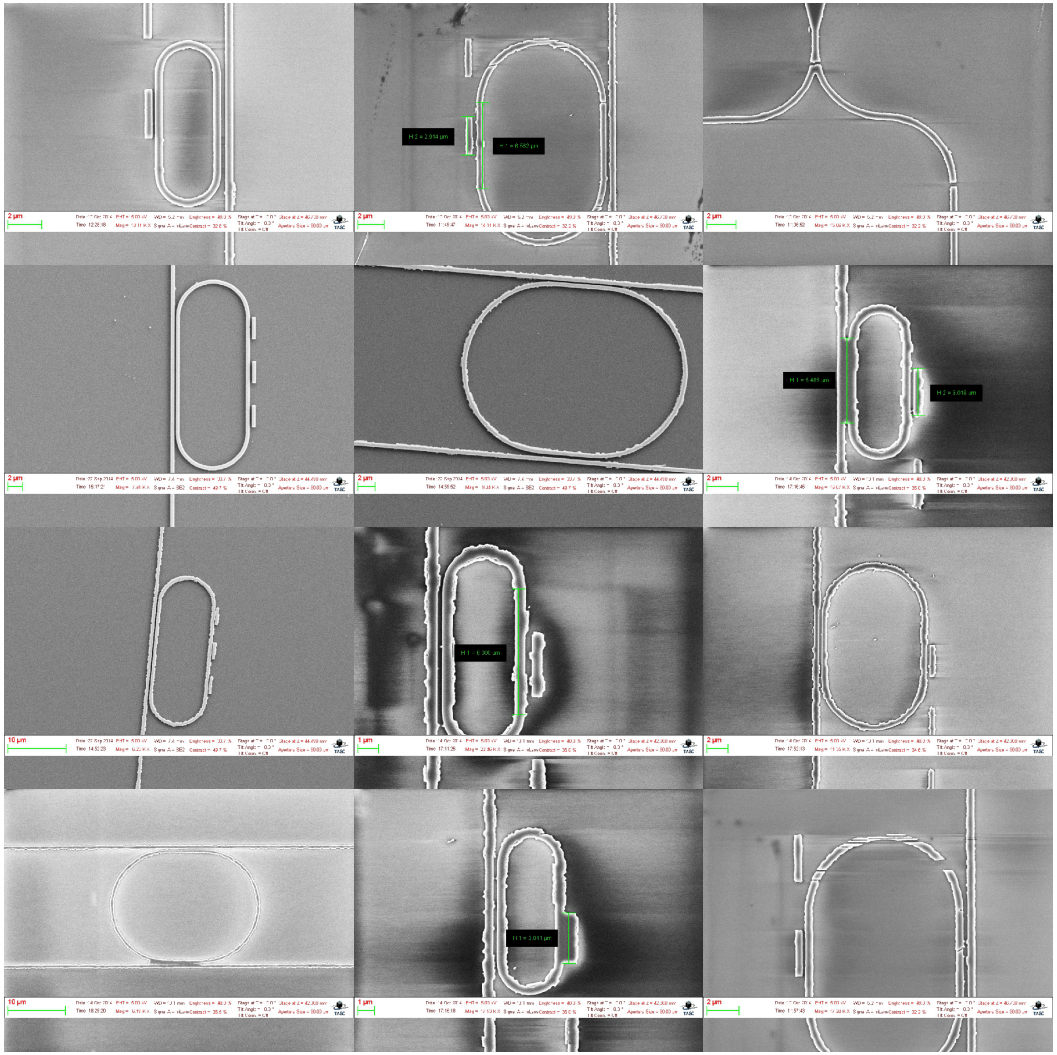
The tools

- Use a general Deep Learning Network to extract features (2048/4096)
  - Resnet/Alexnet/Inception v3
- Dimensional Reduction
  - PCA:
  - Autoencoders
- Cluster of the obtained dataset
  - HDBScan
  - Advanced Density Peaks

Result so far..

MEMS

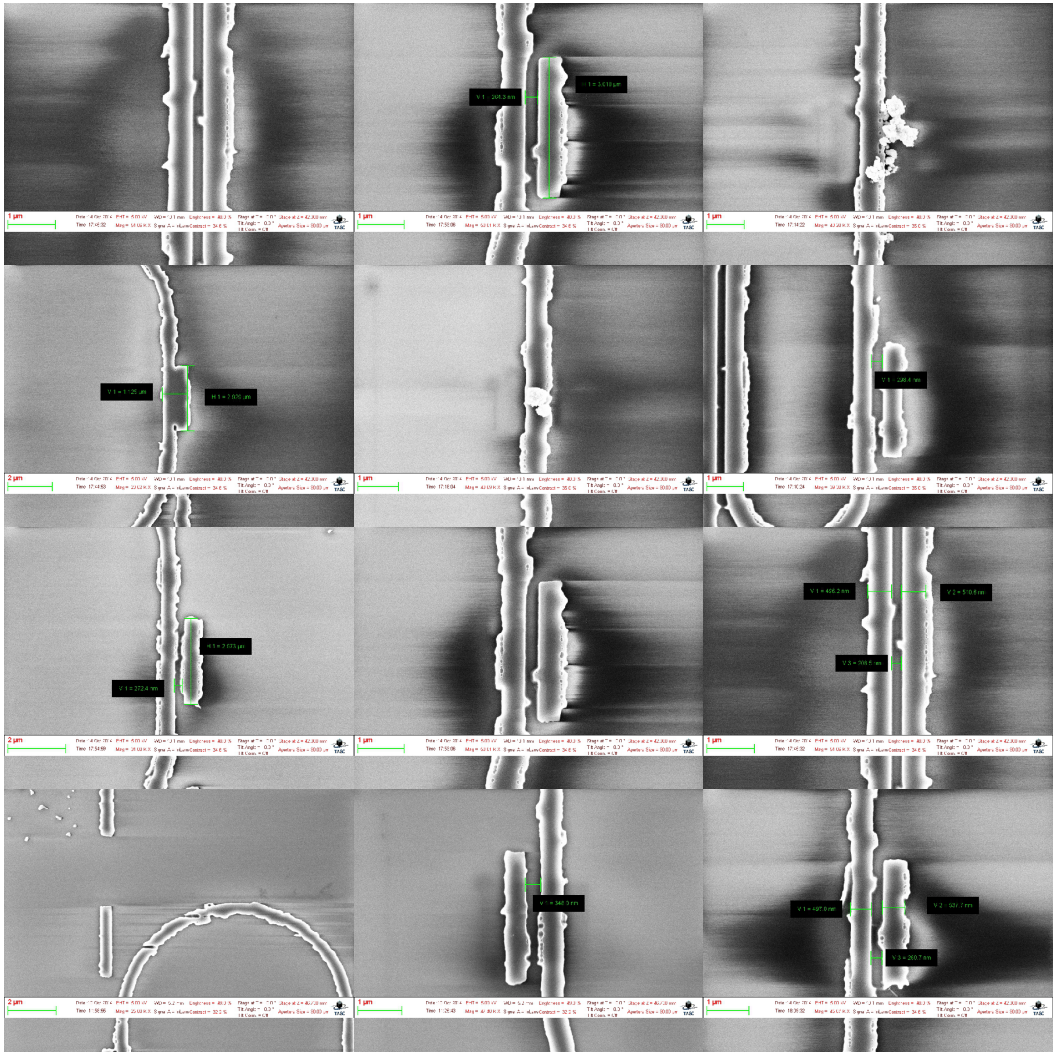
global wave guides



Result so far..

MEMS

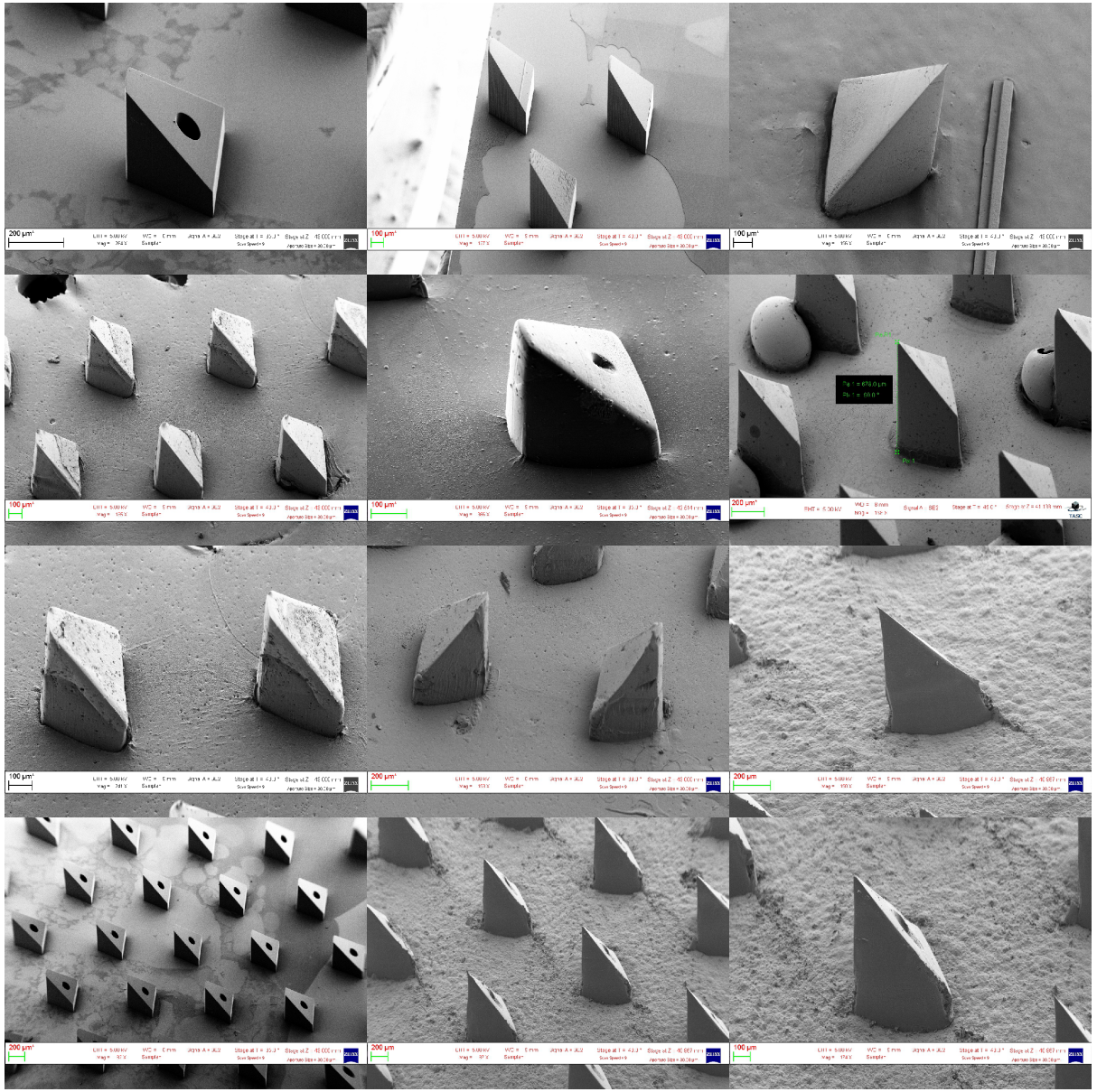
partial wave guides



Result so far..

PATTERN

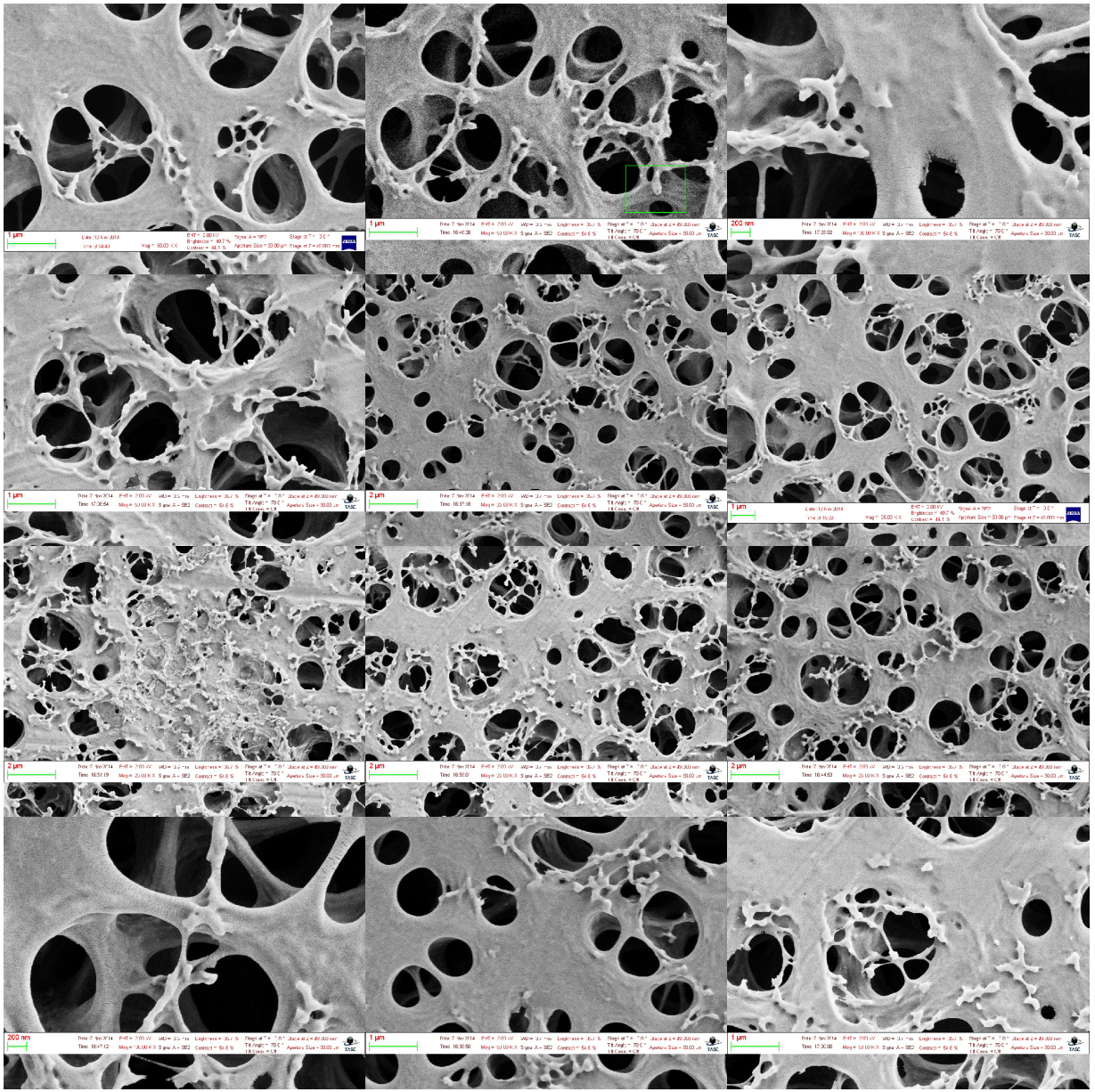
Triangular Pillar



Result so far..

Spounge

All images within this small category !



Considerations

- Results are encouraging; cluster well defined
  - Surprisingly closed the manual classification
- We can easily enlarge the SEMS dataset with new categories (hierarchically identified) and further enhance the trained classifier
- Method can be applied to any dataset (i.e. easy transferable to other experimental techniques TEM/ STM etc..)
  - Allows to explore the dataset
  - Labelling can be easily done in collaboration with scientist

Future actions/Perspectives

- Metadata/ IDRP
  - Integrate more experiment specific metadata to enable content related discovery
  - Identify cross-experiment relationships for improved search results
  - Integrate recent recommendations of Research Data Alliance and the European Commission Expert Group on FAIR data
- Data analysis services
  - New servicesv
- Envision seemless integration with European Open Science Cloud
    - EOSC-PILLAR: Task 6.1

Conclusions

- A Data management infrastructure/Meta-Data schema for NFFA-EUROPE up & running.
- Strong connections with other projects: EUSMI / EUDAT
- Looking for (natural) connections to projects that address FAIR data management principles, as quality metadata is FAIR enabler

Thanks !

References

- **H.M. Modarres, R. Aversa, S. Cozzini, et al.**, "*Neural Network for Nanoscience Scanning Electron Microscope Image Recognition*", *Scientific Reports* **7**, 13282(2017)

- **R. Aversa, S. Cozzini et al.** "*The first annotated set of Scanning Electroscope Microscopy images*", Scientific Data 2018

- **R. Aversa**, "*Scientific Image Processing within the NFFA-EUROPE Project*", MHPC thesis, 16-12-2016

- **P. Coronica**, "*Feature learning and clustering analysis for images classification*", MHPC thesis, 26-10-2018