nffa.eu
PILOT 2021 2026

DELIVERABLE REPORT

WP9 Virtual Access

D9.3

Integration of the second set of the VA offer

Due date
M37

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
WP9	Virtual Access

WORK PACKAGE LEADER

Dr. Rossella Aversa (KIT)

DELIVERABLE DETAILS

DELIVERABLE ID	DELIVERABLE TITLE
D – D9.3	Integration of the second set of the VA offer

DELIVERABLE DESCRIPTION

This Deliverable presents the second set of Virtual Access (VA) services integrated into the NEP infrastructure.

DUE DATE	ACTUAL SUBMISSION DATE
M37 31/03/2024	12/04/2024

AUTHORS

Rossella Aversa (KIT), Nicolas Blumenröhr (KIT), Elias G.G. Vitali (KIT), Andrea Recchia (eXact lab)

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. Rossella Aversa (KIT)

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
12/04/2024	**14**

FOR MORE INFO PLEASE CONTACT

Rossella Aversa
Karlsruhe Institute of Technology
Scientific Computing Center
Kaiserstraße 12, 76131 Karlsruhe
Germany

email: rossella.aversa@kit.edu

VERSION	DATE	AUTHOR(S)	DESCRIPTION / REASON FOR MODIFICATION	STATUS
1	12/03/2024	R. Aversa	First draft	Draft
2	18/03/2024	R. Aversa, N. Blumenröhr, E. Vitali	Main text	Draft
3	19/03/2024	R. Aversa	Major review	Draft
4	27/03/2024	R. Aversa, N. Blumenröhr, E. Vitali	Finalization	Final version

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

This deliverable presents the second set of Virtual Access (VA) services integrated into the NEP infrastructure. These online services, all running on a virtual machine (4 CPUs, 16 GB RAM, 50 GB SDD, OS: Debian 11, SSL via Apache reverse proxy) hosted by the Karlsruhe Institute of Technology (KIT), were developed within the Work Package (WP) 16 and were designed to improve the data FAIRness and to facilitate the user experience on (meta)data generation, post-processing or exploration.

The services are authenticated upon the NEP Single Sign-On (SSO) system via Keycloak [14]. The usage is monitored by aggregating the Units of Access (UoA), which are established to be every single action made by a logged-in user on one of the services, and monitored: whenever a logged-in user performs an action, the service backend sends a REST request to the NEP backend including the service ID and increases the usage counter by 1 UoA.

No information about the users is handled or stored by the services. The Keycloak token, used by the Single Sign-On, is the only piece of information needed to grant access to the service.

The document consists of three sections describing one VA service each. For completeness, each section explicitly mentions the corresponding WP16 task in which the service was framed and the deliverable in which it was described, if applicable.

Nuclear Magnetic Resonance data curation

Data curation is essential to improve Nuclear Magnetic Resonance (NMR) correlation and spectral analysis. Many initiatives have already identified this need and have offered online databases [1-6], in addition to the ones locally maintained by individual institutes. However, these databases contain only a set of spectra related to a specific focus and often do not overlap with each other.

To overcome this issue, we developed the NMR Spectra Graph [7]. This service aims at offering a centralized Graphical User Interface (GUI) providing information in a uniform way about NMR resources and related data that are distributed over the internet. Using this service, the scientists can query for NMR spectral metadata on a single system and will no longer need to search in every separate database.

To deploy the service, we initially selected nmrXiv [1] and Biological Magnetic Resonance Data Bank [2], two online databases exposing a public Application Programming Interface (API) and thus allowing us to harvest the deposited NMR spectra. After a consultation with expert researchers in this field, the metadata of the spectra were then represented in a standardized format containing the most relevant information in terms of data location, experiment performed, molecules identified in the spectrum, provenance and related NMR resources.

The standardized format is based on the FAIR Digital Object (FDO) concept [8,9] and was adapted to an ontology based on W3C Semantic Web Standards [10]. The ontology was made actionable by transforming it into an RDF Triple-Store which can be queried using SPARQL [11]. To ease the users, we implemented and exposed a selected set of relevant queries on the service GUI. A description guides the user through the possible requests they can make on the NMR graph running in the backend. The required input is the set of metadata attributes based on which the graph should be queried (see Figure 1).

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

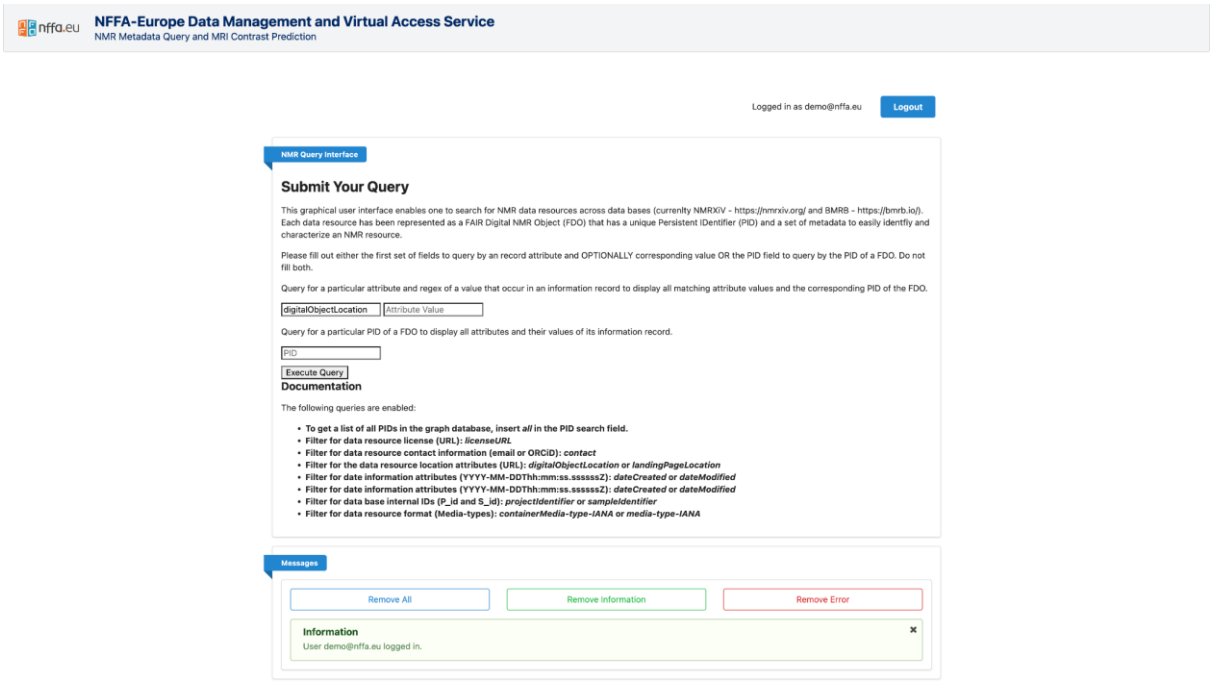

Figure 1: NMR Graph query input

For example, a user would like to search for NMR data which have been stored across several storage systems; in this case, the user can insert the standardized metadata attribute "digitalObjectLocation", available for all NMR resources, in the Record Attribute field. The result of the query will be all the available NMR resources, as shown in Figure 2.

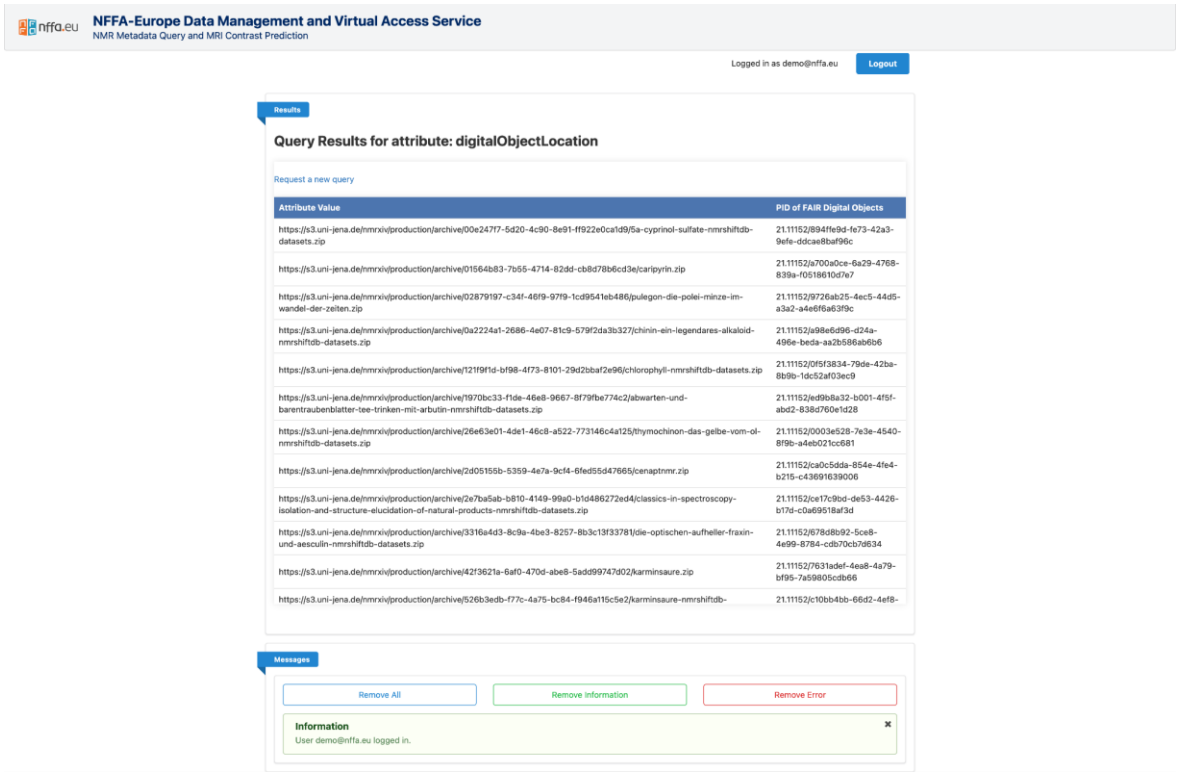

Figure 2: NMR Graph query results for "Attribute": "digitalObjectLocation"

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

Additional regex terms may be included to refine the search, e.g. to filter for specific storage systems. The resulting table then contains a list of URLs where the NMR resources are deposited, as well as the corresponding identifiers (PIDs) of the associated Graph entries containing additional metadata attributes and descriptions. These PIDs could be used in a subsequent query to further explore the NMR metadata description, as shown in Figure 3.

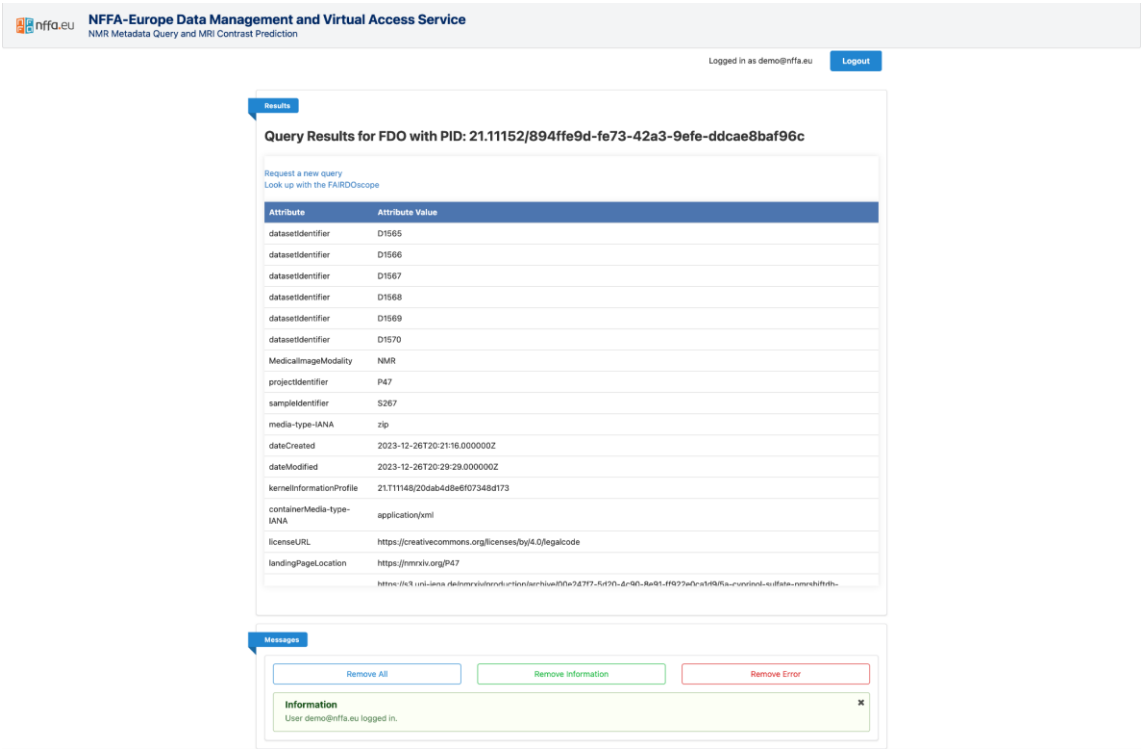

Figure 3: NMR Graph query results for "PID"

FDOs are primarily intended for machine-actionability and automation; nevertheless, the service provides the option to visualize the records using the FAIR-DOscope [12], a generic FDO viewer and browser which offers a tabular view of the contents in a human-readable format and a graphical representation of related FDOs. The resulting PIDs from a query (as the ones shown in Figure 3) can be used as input in the FAIR-DOscope [12] to visualize their content. Figure 4 shows how a given PID is intuitively represented in FAIR-DOscope.

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

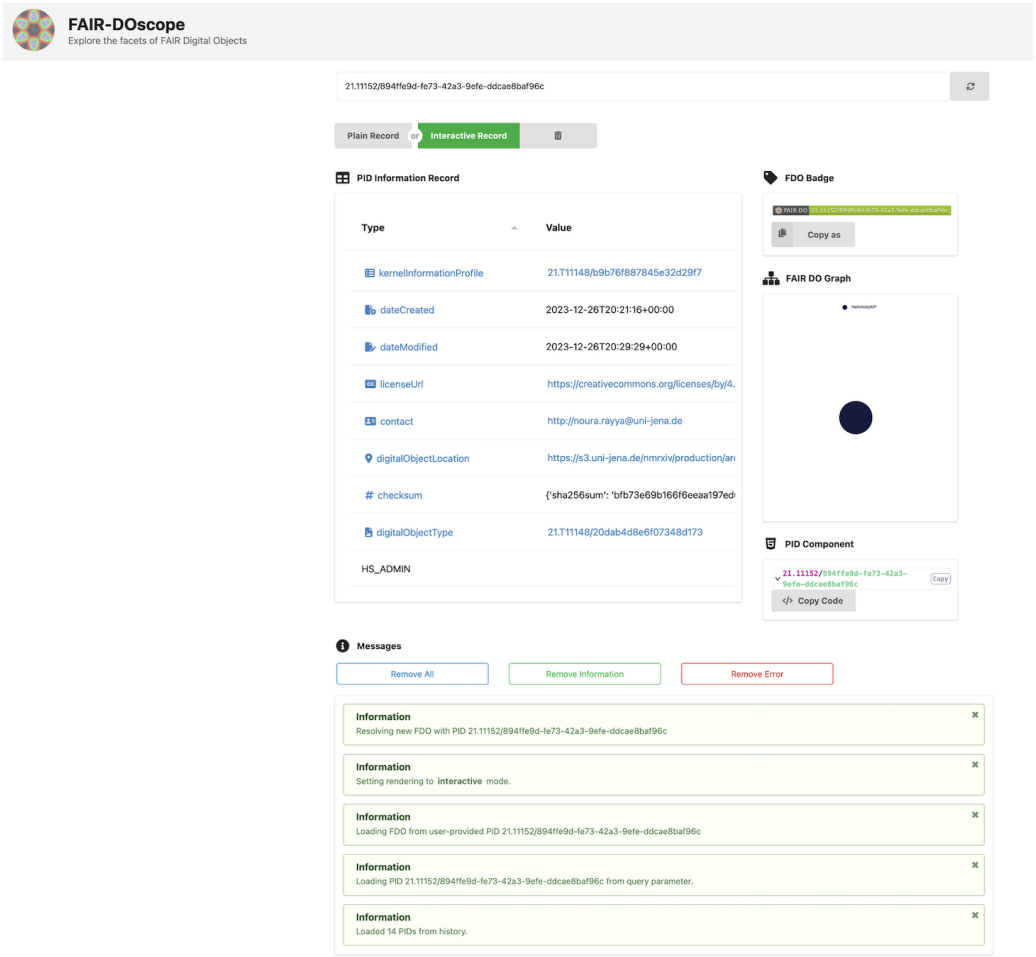

Figure 4: FAIR-DOscope tabular view of the information record in human-readable form (left) and the graphical representation of the related FDOs (right) for a given PID.

In the first months after the release of the service, we expect to collect feedback from the community: for instance, the current queries are provided as a starting point and new functionalities may be easily integrated upon request. Moreover, the NMR Graph can possibly be extended to integrate new databases (e.g., the Human Metabolome Database [3]) as well as internal NMR archives locally stored by individual institutes (e.g. at KIT-IMT). The chosen metadata format is appropriate to extend the service for potential future needs, e.g. to link NMR resources to both the experimental metadata and the software needed to analyze the NMR spectra.

The NMR Spectra Graph was developed as foreseen in the Subtask 16.2.4 "NMR Data Curation" of the WP16, and was described in the deliverable D16.4 [13].

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

Magnetic Resonance Image reconstruction and contrast prediction

Magnetic Resonance Imaging (MRI) is applied in material sciences for non-invasive investigation of sample structure and composition, by leveraging the differences in tissue contrasts. However, every different type of contrast, encoded in the MR image, typically requires a separate measurement, which is a time-consuming task.

To tackle this issue, the MRI Prediction Service [15] was designed. The service offers a simple and intuitive GUI based on machine learning, which has the aim to optimize the information contained within the datasets measured in the same MRI experiment in order to predict an alternative contrast type from a given one. This decreases the measurement time by a factor n for each of the n contrast types that can be predicted.

The service, shown in Figure 5, allows users to upload a DICOM file obtained at a given contrast expressed in terms of the pulse sequence parameters TE and TR (echo time and repetition time, respectively) and to request a prediction of how the image would look like if measured with an alternative contrast, based on the given theoretical sequence parameters that would have been used.

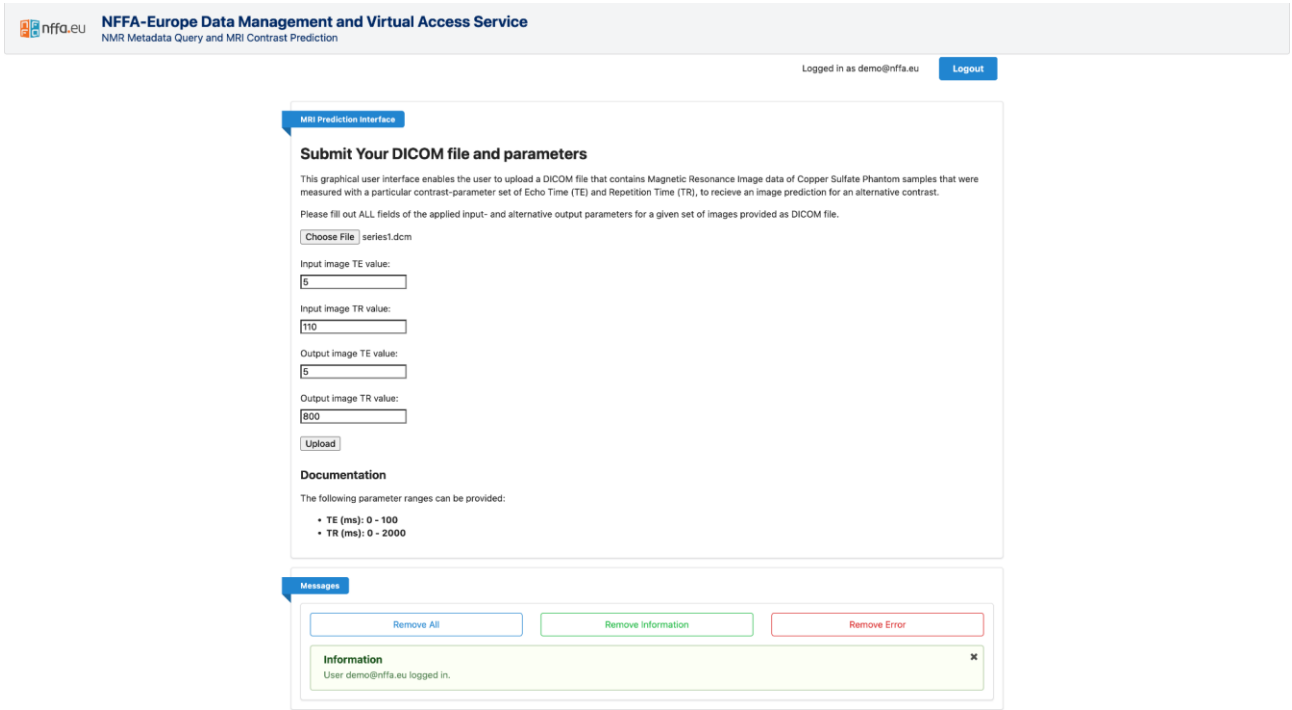

Figure 5: MRI Prediction Service input page

The resulting image is then displayed as in a preview and can be downloaded together with the corresponding theoretical sequence parameter used to generate it and the original DICOM file (Figure 6). In the case where one DICOM file contains multiple images, an alternative image is predicted for each, but only the first image in the stack is displayed to prevent possible conflicts with large datasets.

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

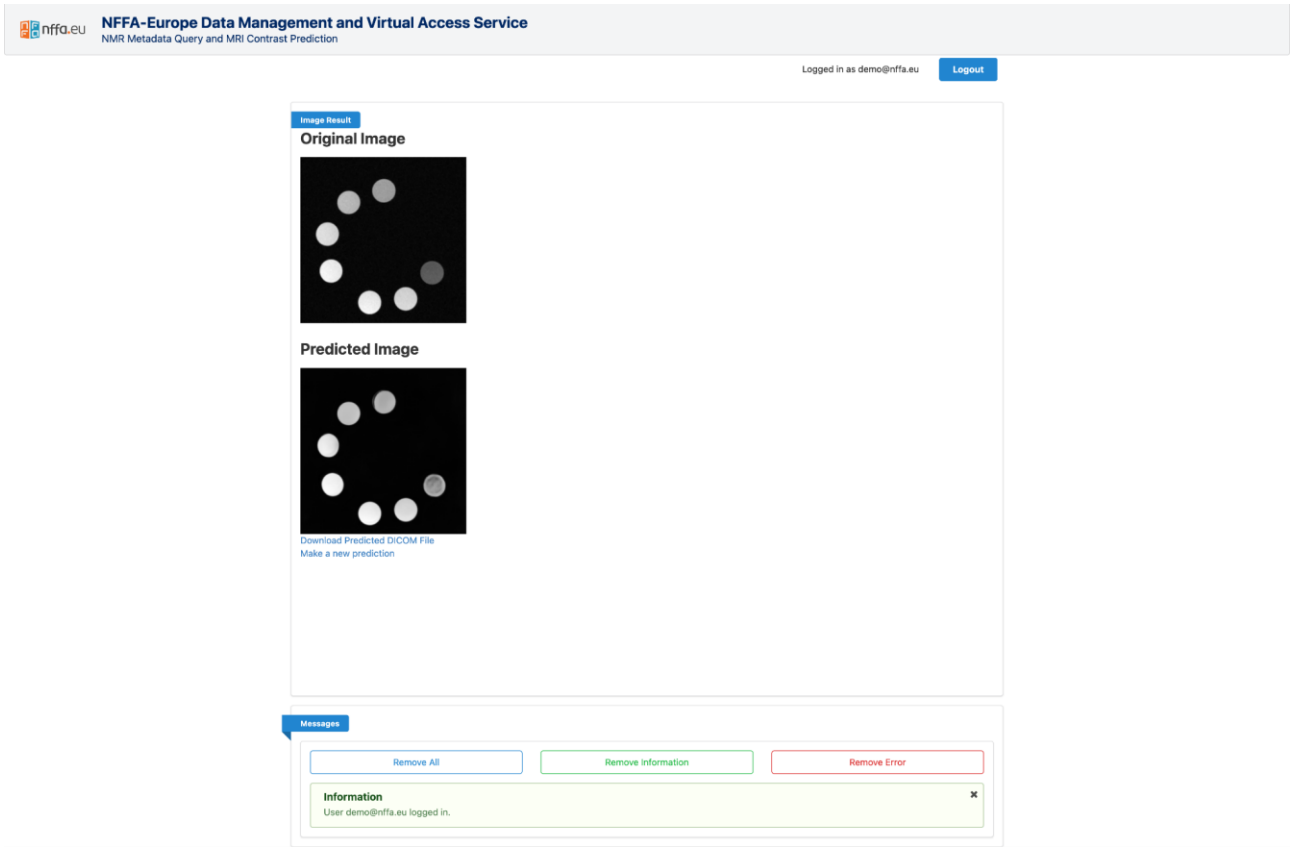

Figure 6: MRI prediction result

Currently, the machine learning model has been trained on predicting contrasts for a solution of $CuSO_4$ at different concentrations, as required by our target users. We are open to receive further input data from the community to train the model on different materials to further improve the applicability of the service.

The MRI Prediction Service was developed as foreseen in the Subtask 16.2.3 "MRI reconstruction and contrast prediction" of the WP16, and was described in the deliverable D16.4 [13].

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

Mapping Service

The Mapping Service [16] is a versatile tool designed to streamline the management of scientific research data. It addresses the critical need for standardized metadata by automatically extracting it from diverse experimental data files and mapping it to community-developed schemas which are registered in MetaRepo [17]. The Mapping Service features a user-friendly interface, pictured in Figure 7, and a modular plugin architecture designed to be easily extensible.

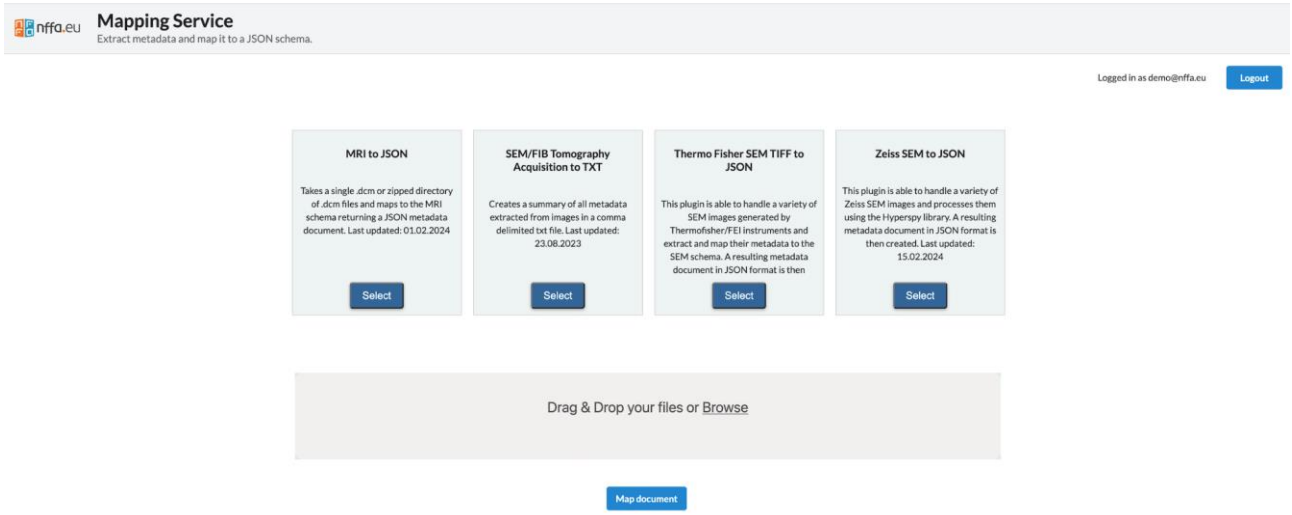

Figure 7: Mapping Service landing page.

The user is greeted with a short message which explains the main functions of the tool and links to the available documentation. The list of available mapping plugins is then presented and easily navigable with legible descriptions of each plugin. The user selects the plugin best suited to their data format and the generating instrument, is then prompted to upload their file, which is processed via the "Map document" button, triggering an automatic download of the results file. The output is a schema-adherent metadata document in JSON format, ensuring consistency, reusability, and interoperability of the research data and the metadata describing it.

The list of the available mapping plugins, together with the instrument vendors that are currently supported and the metadata schemas according to which the extracted metadata are mapped, are outlined in Table 1. In particular, the "MRI to JSON" plugin was presented and used in [18], while the "ThermoFisher SEM to JSON" was used to map the metadata of the SEM image [31] acquired and analyzed in [32], before registering it in MetaRepo [33].

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

Table 1: List of mapping plugins currently available in the Mapping Service, instrument vendors supported and references to the target metadata schemas to which the plugins map the metadata extracted from the data file.

Mapping Plugin	Supported Vendor	Related Metadata Schema
Zeiss SEM to JSON	Zeiss	[19]
SEM to TXT	Zeiss	[19]
ThermoFisher SEM to JSON	ThermoFisher	[19]
SEM-FIB Tomography Acquisition to JSON	ThermoFisher	[20-22]
SEM-FIB Tomography Acquisition to TXT	ThermoFisher	[20-22]
MRI to JSON	DICOM	[23]

Some of the key benefits of the Mapping Service include:

- Elimination, or at least reduction, of manual metadata compilation: by automating the process, the Mapping Service not only reduces the time and effort required for metadata management, but it also mitigates the potential for human errors
- Extensible design: the Mapping Service readily accommodates new schemas and mappings for different experimental techniques
- Easy adoption: A web-based architecture requires no local installation or dependencies from end-users.

The functionalities and interface have been showcased at several conferences during its development, namely in the form of posters and talks [24-27]. Thanks to the discussions that arose during these events, we were able to integrate feedback received from researchers in the materials science field to improve the service and its user interface. Nevertheless, new plugins can be developed, if required by the community.

The Mapping Service can be included in the Task 16.4 "Scouting activities for further data services" of the WP16, as it was developed in order to facilitate the management and FAIRification of research data and to make interaction with the MetaRepo [20] more efficient.

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

REFERENCES

[1] nmrXiv: https://nmrxiv.org/

[2] Biological Magnetic Resonance Data Bank: https://bmrb.io/

[3] The Human Metabolome Database: https://hmdb.ca/

[4] Spectral Database for Organic Compounds: https://sdbs.db.aist.go.jp/sdbs/cgi-bin/cre_index.cgi

[5] nmrschiftdb: https://nmrshiftdb.nmr.uni-koeln.de/

[6] SpectraBase: https://spectrabase.com/

[7] NMR Spectra Graph: https://metarepo.nffa.eu/frontend/nmr_query

[8] Ivonne, A., et al., FAIR Digital Object Forum - FDO Requirements Specifications (2023). DOI: 10.5281/zenodo.7781925

[9] FAIR DOs: https://helmholtz-metadaten.de/en/fair-data-commons/realizing-fair-digital-objects

[10] W3C Semantic Web standards: https://www.w3.org/2001/sw/wiki/Main_Page

[11] SPARQL: https://www.w3.org/TR/sparql11-query/

[12] FAIR-DOscope: https://kit-data-manager.github.io/fairdoscope/

[13] Aversa, R., Blumenröhr, N., Recchia, A., Tsibidis, G., Pantazis, Y., NFFA-Europe Pilot - D16.4: Report on additional data services. Zenodo (2023). DOI: 10.5281/zenodo.10201228

[14] Keycloak: https://www.keycloak.org

[15] MRI Prediction Service: https://metarepo.nffa.eu/frontend/mri_prediction

[16] Mapping Service: https://metarepo.nffa.eu/frontend/mapping-service-ui.html

[17] MetaRepo: https://metarepo.nffa.eu/frontend/schema-management.html

[18] Blumenröhr, N., MacKinnon, N., Aversa, R., FAIR Data Management Workflow for MRI Data (2023). DOI: 10.5445/IR/1000159552

[19] SEM metadata schema: https://metarepo.nffa.eu/api/v1/schemas/sem?version=15

[20] SEM-FIB Tomography Acquisition metadata schema: https://metarepo.nffa.eu/api/v1/schemas/sem_fib_tomography_acquisition?version=5

[21] SEM-FIB Tomography Dataset metadata schema: https://metarepo.nffa.eu/api/v1/schemas/sem_fib_tomography_dataset?version=5

[22] SEM-FIB Tomography Image metadata schema: https://metarepo.nffa.eu/api/v1/schemas/sem_fib_tomography_image?version=5

[23] MRI metadata schema: https://metarepo.nffa.eu/api/v1/schemas/mri_schema?version=8

[24] Joseph, R.E., Vitali, E.G.G., Aversa, R., Scanning Electron Microscopy (SEM) - Metadata extraction tool and schema mapper (2023). DOI: 10.5445/IR/1000156874

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

[25] Joseph, R.E., Vitali, E.G.G., Kirar, A., Aversa, R., Metadata Extraction and Mapping Service (2023). DOI: 10.5445/IR/1000157902

[26] Aversa, R., Joseph, R.E., Vitali, E.G.G., Kirar, A., Extracting, Mapping, Editing SEM Metadata (2023). DOI: 10.5445/IR/1000158866

[27] Vitali, E.G.G., Joseph, R.E., Aversa, R., Metadata Extraction Tool and Schema Mapper for Scanning Electron Microscopy (SEM) images (2023). DOI: 10.5445/IR/1000159338

[28] Vitali, E.G.G., Kirar, A., Joseph, R.E., Hartmann, V., Blumenröhr, N., Aversa, R., Stotzka, R., Metadata Extraction Tool and Schema Mapper for Scanning Electron Microscopy (SEM) images (2023). DOI: 10.5445/IR/1000160100

[29] Abdildina, G., Heinrichs, B., Joseph, R.E., Kirar, A., Vitali, E.G.G., Aversa, R., Applying Metadata Management: Extraction, Mapping & Visualisation (2023). DOI: 10.5445/IR/1000160126

[30] Vitali, E.G.G., Jejkal, T., Kirar, A., Joseph, R.E., Hartmann, V., Blumenröhr, N., Inckmann, M., Aversa, R., Stotzka, R., A Tool for Automatic Metadata Extraction and Schema Mapping for SEM Images (2023). DOI: 10.5445/IR/1000164968

[31] Keller, T.F., et al., STO crystal supported Pt nanodot arrays. Zenodo (2024). DOI: 10.5281/zenodo.10850111

[32] Keller, T.F., et al., Coherent x-ray diffraction of a semiregular Pt nanodot array, Phys. Rev. B 108, 134109 (2023). DOI: 10.1103/PhysRevB.108.134109

[33] Keller, T.F., Aversa R., Metadata document of SEM imaging of STO crystal supported Pt nanodot arrays. MetaRepo (2024). URI: https://metarepo.nffa.eu/api/v1/metadata/98a79b77-f55a-47a3-85bf-209df39cbb0e?version=1

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