Introduction to FAIR Principles

Rossella Aversa

KARLSRUHE INSTITUTE OF TECHNOLOGY

nffa.eu PILOT 2021 2026

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

www.nffa.eu
secretariat@nffa.eu

Motivation

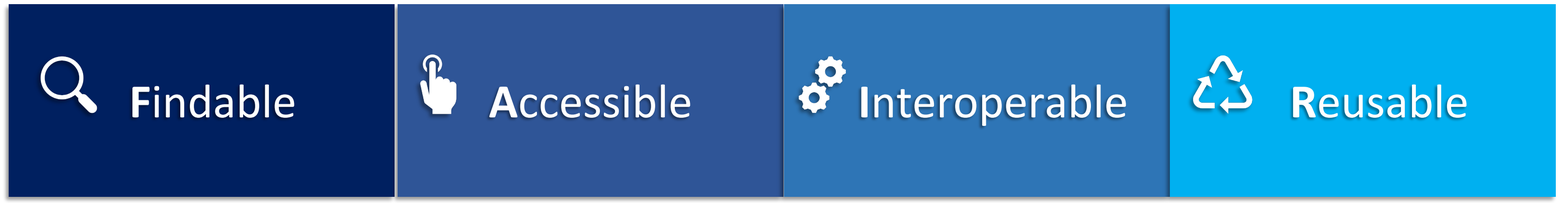
Powered by Bing Image Creator

nffa.eu PILOT 2021 2026

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

www.nffa.eu
secretariat@nffa.eu

The FAIR Guiding Principles



https://www.go-fair.org/fair-principles/

nffa.eu PILOT 2021 2026

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 FAIR Guiding Principles

(Meta)data should be easy to find for both humans and computers

Globally unique persistent identifiers (PID)

https://www.go-fair.org/fair-principles/

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 FAIR Guiding Principles

It should be known how (meta)data can be accessed

(Meta)data repositories, authorization & authentication

https://www.go-fair.org/fair-principles/

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

www.nffa.eu
secretariat@nffa.eu

The FAIR Guiding Principles

Data should be exchanged and interpreted by humans and computers

Structured metadata (schemas, vocabularies)

https://www.go-fair.org/fair-principles/

nffa.eu PILOT 2021 2026

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 FAIR Guiding Principles

It should be clear how data can be reused and/or replicated

Licences, rich (provenance) metadata

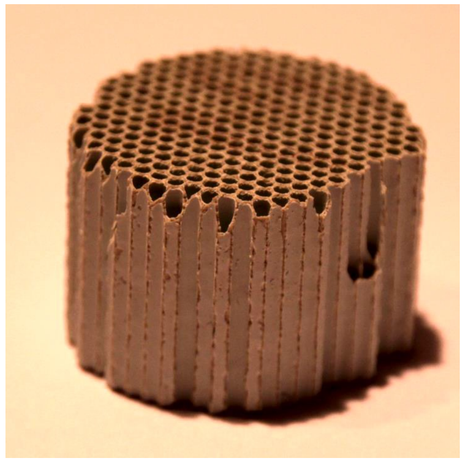
https://www.go-fair.org/fair-principles/

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

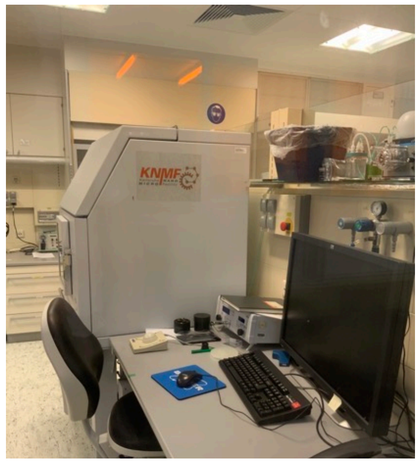
What to describe?

R1: Metadata should richly describe the data with a plurality of accurate and relevant attributes.

Sample



Instrument



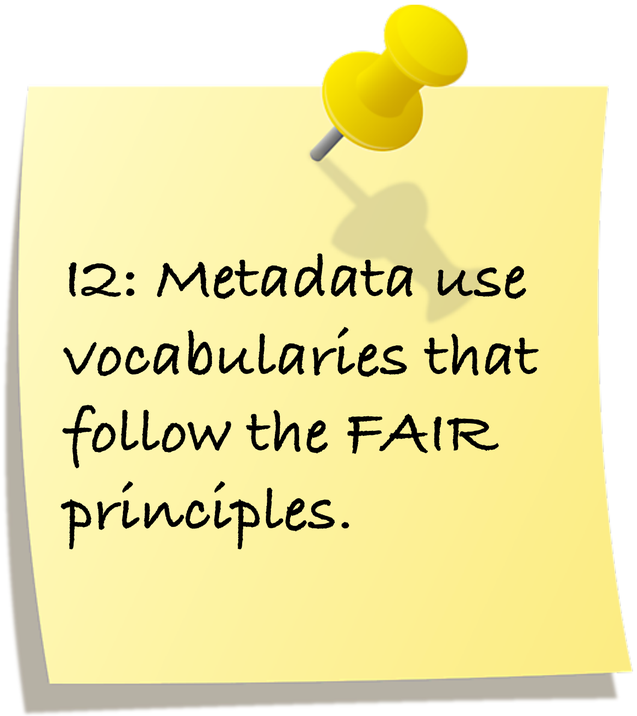
Research data

Images: courtesy of R. Thelen and M. Mail. Powered by Bing Image Creator

nffa.eu PILOT 2021 2026

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

How to describe data?



Vocabulary: set of terms pertaining to a particular domain + definitions. Useful to ensure that the data is described consistently.

Taxonomy: hierarchical (tree-like) structure of the terms. Useful to organize data into categories which are meaningful in a particular domain.

Ontology: formal description of the terms, their properties and their relationships within a particular domain. Useful to consistently represent the knowledge about a domain.

nffa.eu PILOT 2021 2026

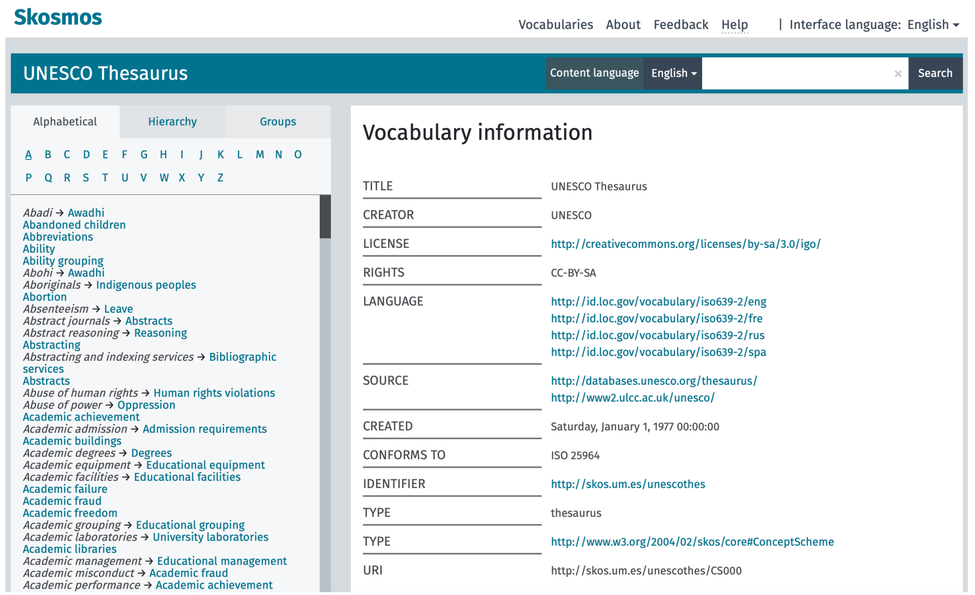
This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

How to describe data?

12: Metadata use vocabularies that follow the FAIR principles.

Minimal requirements:

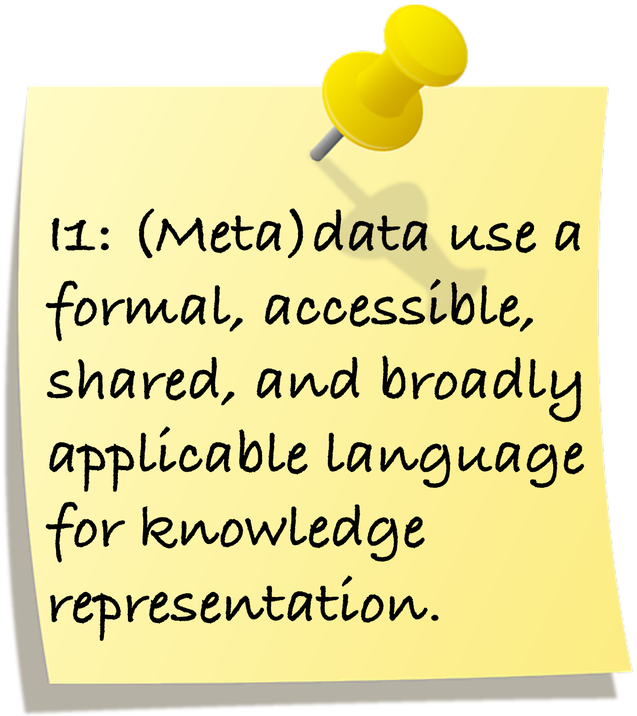
- The vocabulary and its terms have globally unique PIDs
- The vocabulary and its terms are documented
- The documentation is findable and accessible by users



https://skosmos.org

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

How to represent metadata?



W3C Semantic Web

Resource Description Framework: metadata model to represent interconnected data.
https://www.w3.org/RDF/

Simple Knowledge Organization System: standard to represent knowledge organization systems using RDF
https://www.w3.org/2004/02/skos/

Web Ontology Language: computational logic-based language to represent complex knowledge.
https://www.w3.org/OWL/

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

How to structure metadata?

R1.3: Metadata meet domain-relevant community standards or best practices.

General purpose

DublinCore http://dublincore.org/schemas/

DataCite http://schema.datacite.org

Schema.org https://schema.org

```
▼<xs:sequence>
 ▼<xs:choice minOccurs="0" maxOccurs="unbounded">
    <xs:element ref="any"/>
  </xs:choice>
</xs:sequence>

<xs:element name="givenName" minOccurs="0"/>
<xs:element name="familyName" minOccurs="0"/>
```

```
"colleague": [
  "http://www.xyz.edu/students/alicejones.html",
  "http://www.xyz.edu/students/bobsmith.html"
],
```

Neutron, x-ray, muon

NeXus http://www.nexusformat.org

```
entry:NXentry
   raw:NXsubentry
      definition="NXsas"
   reduced:NXsubentry
      definition="NXcanSAS"
   fluo:NXsubentry
      definition="NXfluo"
```

Crystallography

CIF https://www.iucr.org/resources/cif

```
_cell.entry_id           3ADK
_cell.length_a           48.500
_cell.length_b           48.500
_cell.length_c           141.000
_cell.angle_alpha        90.00
_cell.angle_beta         90.00
_cell.angle_gamma        120.00
_cell.Z_PDB              6
```


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

How to represent structured metadata?

I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

XML (eXtensible Markup Language)

```
<person>
  <firstName>John</firstName>
  <lastName>Doe</lastName>
</person>
```

JSON (JavaScript Object Notation)

```
{"person":
 {
  "firstName":"John",
  "lastName":"Doe"
 }
}
```

JSON-LD (JSON for Linked Data)

```
{
  "@context": "http://schema.org/",
  "person":
  {
    "@type": "person",
    "firstName": "John",
    "lastName": "Doe"
  }
}
```


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

How to publish (meta)data?

F1: (Meta)data are assigned globally unique and persistent identifiers.

Identifier: any label used to name an object uniquely (URL, serial number, personal name)

Persistent identifier (PID): long-lasting reference to locate and identify an object, even if it changes over time → connected to a set of metadata describing an object rather than to the object itself. They allow different platforms to exchange information consistently and unambiguously, e.g. to track citations and reuse.

nffa.eu PILOT 2021 2026

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

How to publish (meta)data?

F1: (Meta)data are assigned globally unique and persistent identifiers.

Cite this article

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Details

DOI
DOI 10.5281/zenodo.7778338

Resource type
Dataset

Publisher
Zenodo

Languages
English

ORCID
Connecting research and researchers

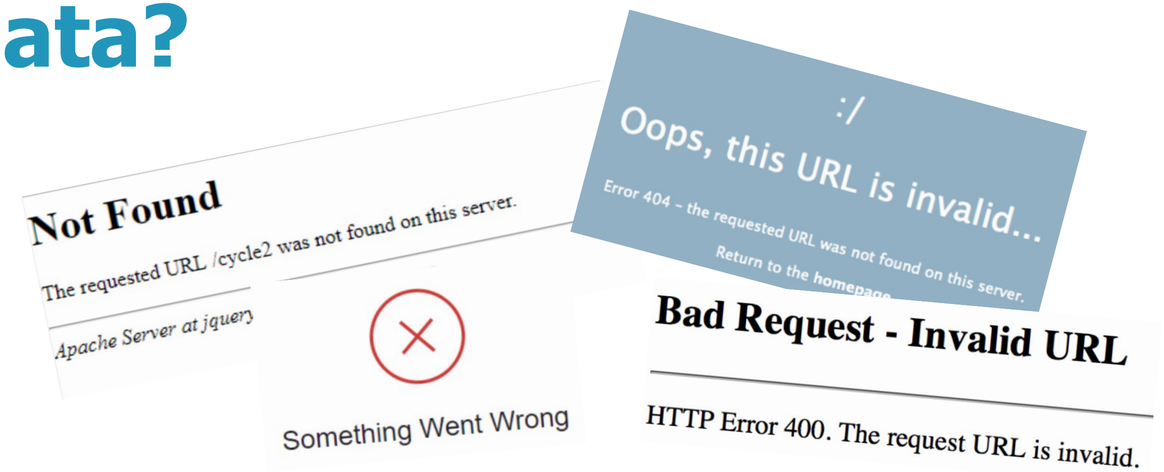
https://orcid.org/
0000-0003-2534-0063

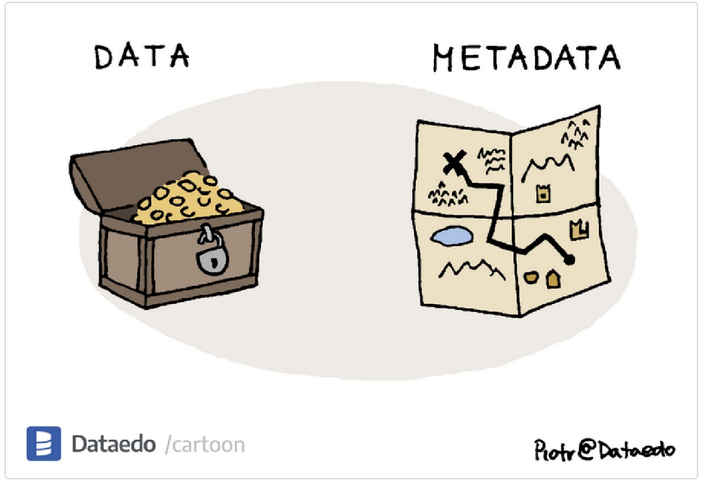
Preview public record

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

How to publish metadata?

A2: Metadata should be accessible even when the data is no longer available.





Metadata Repositories!

https://dataedo.com/cartoon/data-vs-metadata-2

nffa.eu PILOT 2021 2026

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

How to find data from metadata?

F3: Metadata clearly and explicitly include the identifier of the data they describe.

Details

DOI

DOI 10.5281/zenodo.7778338

Resource type

Dataset

Publisher

Zenodo

Languages

English

nffa.eu PILOT 2021 2026

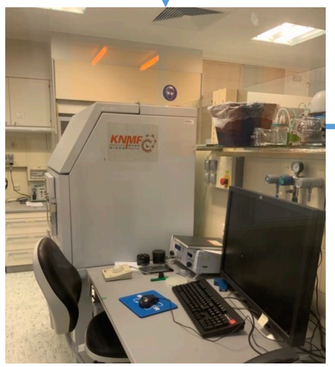
This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

www.nffa.eu
secretariat@nffa.eu

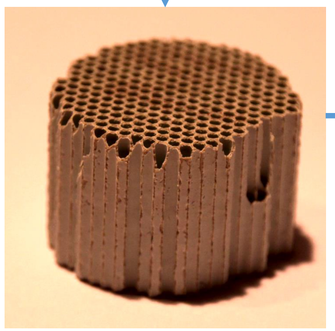
How to reproduce scientific results/measurements?

R1.2: Metadata are associated with detailed provenance.

Data acquired from measurement



Measurement performed on sample



Sample placed on holder

nffa.eu PILOT 2021 2026

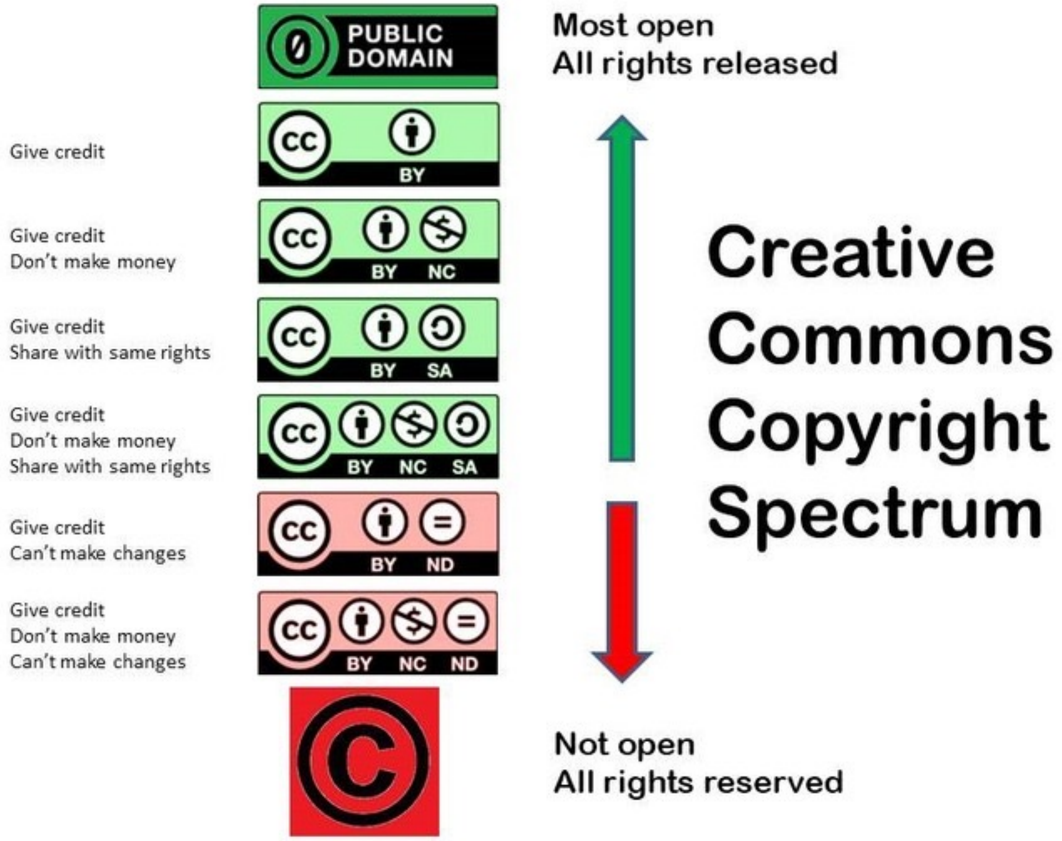
This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

www.nffa.eu
secretariat@nffa.eu

How to reuse the data?

R1.1: (Meta)data are released with a clear and accessible data usage licence.

A data licence is a legal arrangement between the creator of the data and the end-user, or the place where the data will be deposited, specifying what users can do with the data.



https://creativecommons.org/share-your-work/cclicenses/

nffa.eu PILOT 2021 2026

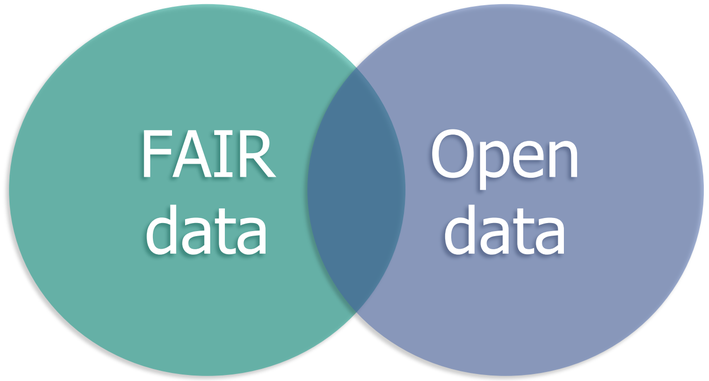
This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

Should FAIR data be open?

A1.2: The protocol allows for an authentication and authorization procedure where necessary.

Open data: "can be freely used, modified, and shared by anyone for any purpose"
https://opendefinition.org

FAIR data: "as open as possible, as closed as necessary".
It should be specified who can access the data under what conditions or whom to ask for permission to access the data.



nffa.eu PILOT 2021 2026

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

www.nffa.eu
secretariat@nffa.eu

FAIR or open?

My data is copyright protected	FAIR	
My dataset can be used only by a specific group of scientists	FAIR	
An image is shared on a public website	Open	
A dataset is published on Zenodo with an open licence	FAIR	Open
A data file is on my Dropbox	None	

nffa.eu PILOT 2021 2026

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

Contact us

Rossella Aversa – rossella.aversa@kit.edu

www.nffa.eu
secretariat@nffa.eu

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

Acknowledgements

- The Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – NFDI 38/1 – project number 460247524
- The Joint Laboratory Model and Data driven Materials Characterization (JL MDMC), a cross-centre platform of the Helmholtz Association
- NFFA-Europe-Pilot (EU H2020 – n. 101007417)
- The research program “Engineering Digital Futures” of the Helmholtz Association of German Research Centers
- The Helmholtz Metadata Collaboration Platform

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