nffa.eu
research infrastructure

# A Lightweight Introduction to FAIR Digital Objects

Nicolas Blumenröhr – nicolas.blumenroehr@kit.edu

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

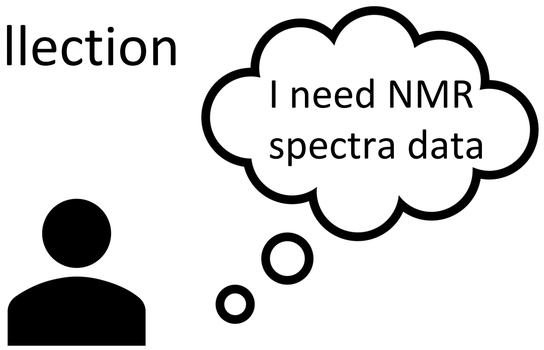
www.nffa.eu
secretariat@nffa.eu

# The Struggle of (Meta)data Management

- Scientists waste time with data wrangling

Let’s consider an example for NMR data acquisition:

Data collection




Searching different storage systems

Dealing with different metadata contents

Many steps done manually

- FAIR Principles provided guidelines for improved stewardship and management
- **But, their implementations are not fully aligned**

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

# Dealing with Storage Systems

- A variety of storage systems exist that implement FAIR principles
- Information retrieval works via access protocols and metadata descriptions
- **These are typically diverse**

For example, to retrieve license information for data reuse:

In the Zenodo repository:


In the NMRxiv database:




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

# Dealing with Metadata

- Metadata Schemas and Standards help to define, organize, and manage metadata
- Enable Machine-readability and interpretability
- Differ between- and within disciplines regarding structure, contents and formats
- **Metadata schemas often have different structures and vocabularies**


```
"rightsList":[{"lang":"en","rights":"Creative Commons Attribution 4.0
International","rightsUri":"https://creativecommons.org/licenses/by/4.0/
legalcode","schemeUri":"https://spdx.org/
licenses/","rightsIdentifier":"cc-
by-4.0","rightsIdentifierScheme":"SPDX"}]
```

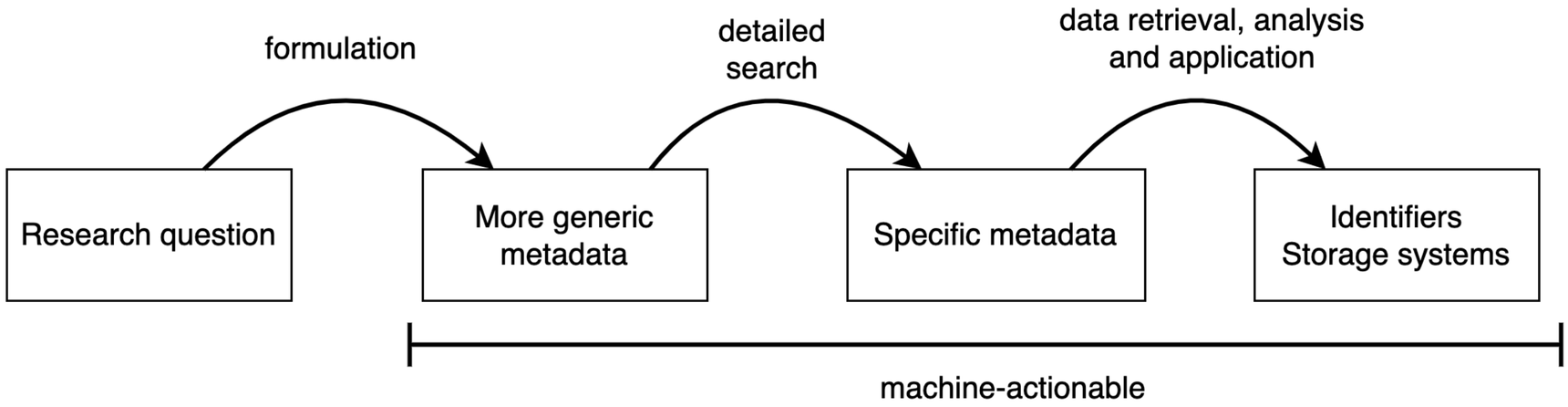
```
"license":{"title":"Creative Commons Attribution 4.0 International (CC BY
4.0)","slug":"cc-by-4.0","spdx_id":"CC-BY-4.0","url":"https:\/\/
creativecommons.org\/licenses\/by\/4.0\/…
```




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

# Machine-actionability

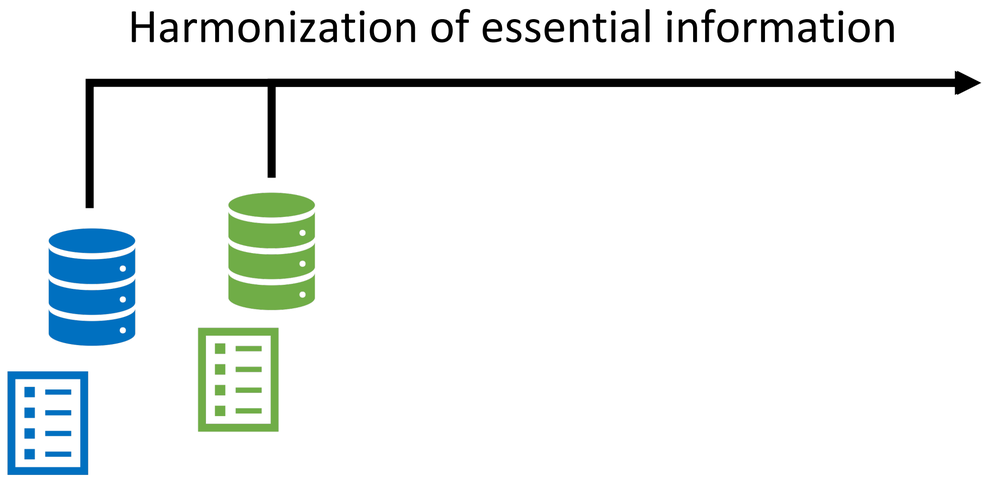
- Requires machine-readability and interpretability
- Automated systems act on digital resources and their metadata
- **No, or less human intervention is required**




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 Tackle?

- High-level information should be harmonized
- Underlying systems and standards must not be changed




Additional, but **uniform** representation format

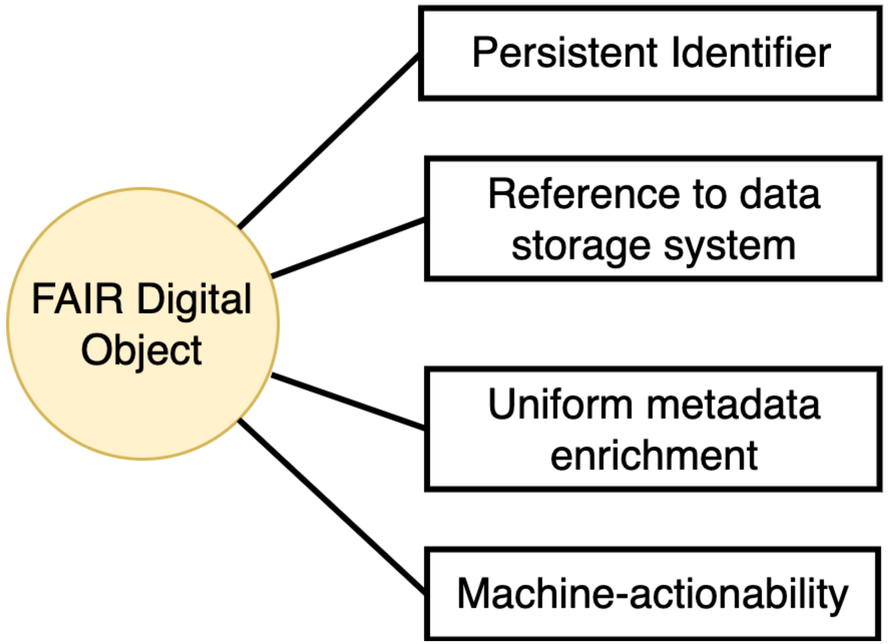
License example:

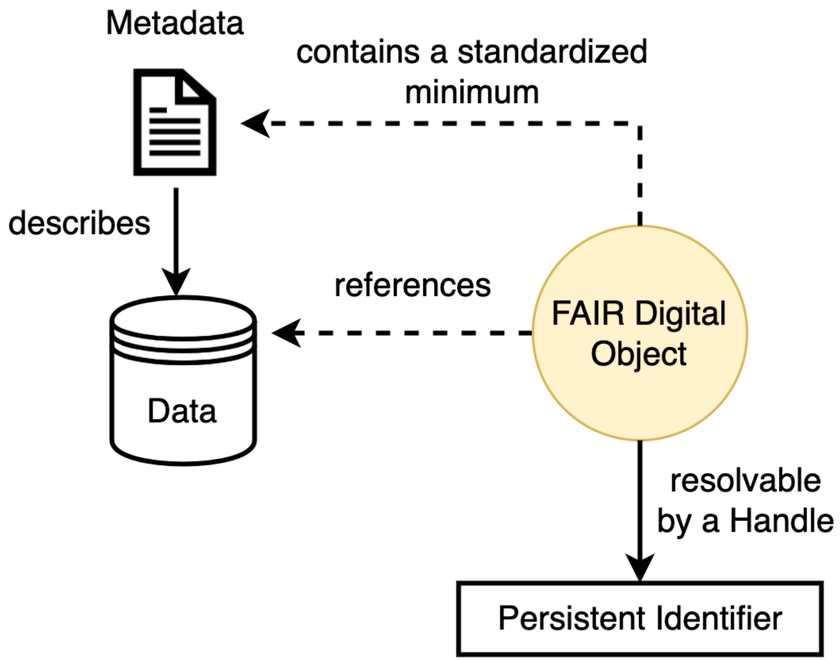
- Name: License
- Description: A URL referring to a license that defines the scope of use for a digital resource
- A unique PID
- Typing --> is a URL for a existing license from an enumeration list


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 Digital Objects Concept

- Representation of digital resources in a uniform way (Digital Object)
- Integrates the essential elements for FAIRness

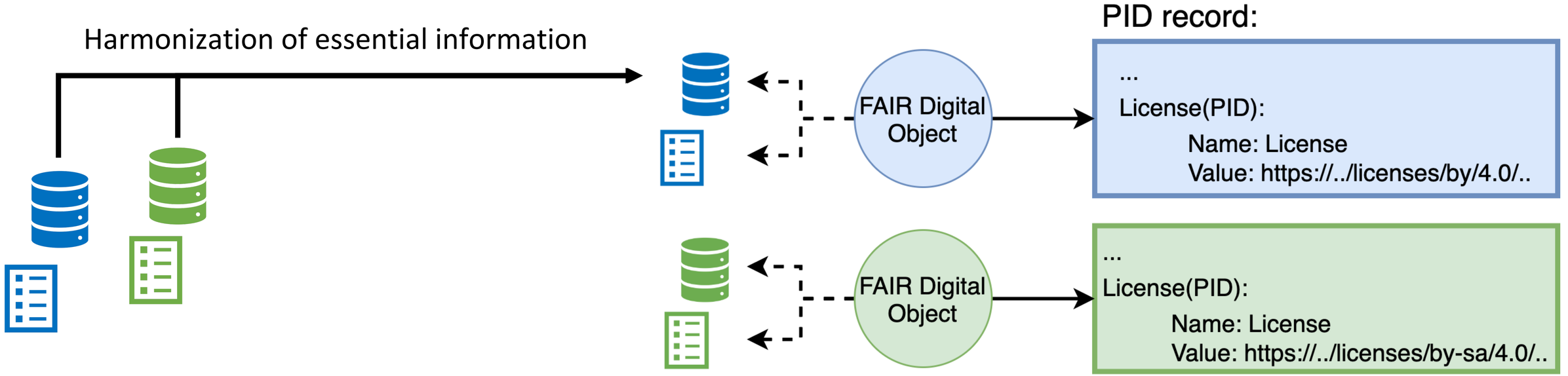



nffa.eu PILOT 2021 2026
This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

# Information is Reduced and Standardized

- Each FAIR Digital Object (FDO) is based on the same structure
- Information at this level is unified and can be treated equally




- **Readable and interpretable for humans and machines**

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

# Original Handle Record Example

Handle.Net®

**Handle Values for: 21.11152/865d3383-55a4-4620-b4ef-e806382e7e09**

| Index | Type | Timestamp | Data |
|---|---|---|---|
| 1 | 21.T11148/076759916209e5d62bd5 | 2024-05-31 19:42:07Z | 21.T11148/631080d008dfbf1ec49e |
| 2 | 21.T11148/f3f0cbaa39fa9966b279 | 2024-05-31 19:42:07Z | HMDB0001149 |
| 3 | 21.T11148/6ae999552a0d2dca14d6 | 2024-05-31 19:42:07Z | Aminolevulinic Acid |
| 4 | 21.T11148/aafd5fb4c7222e2d950a | 2024-05-31 19:42:07Z | 2004-09-16T00:00:00.000000Z |
| 5 | 21.T11148/397d831aa3a9d18eb52c | 2024-05-31 19:42:07Z | 2024-05-25T00:00:00.000000Z |
| 6 | 21.T11148/2f314c8fe5fb6a0063a8 | 2024-05-31 19:42:07Z | https://creativecommons.org/licenses/by/4.0/deed.en |
| 7 | 21.T11148/b8457812905b83046284 | 2024-05-31 19:42:07Z | http://moldb.wishartlab.com/system/documents/files/000/035/104/origi 1354674735 |
| 8 | 21.T11148/8710d753ad10f371189b | 2024-05-31 19:42:07Z | https://hmdb.ca/spectra/nmr_two_d/1591 |
| 9 | 21.T11148/c83481d4bf467110e7c9 | 2024-05-31 19:42:07Z | application/zip |
| 10 | 21.T11148/82e2503c49209e987740 | 2024-05-31 19:42:07Z | { "sha256sum": "5174fd6992c4a6c1f718711a19d2c6314d6908402488 |
| 11 | 21.T11148/68aed8017b345bf87643 | 2024-05-31 19:42:07Z | [1H, 13C]-HSQC NMR Spectrum (2D, 600 MHz, H2O, experimental) |
| 12 | 21.T11148/1c699a5d1b4ad3ba4956 | 2024-05-31 19:42:07Z | 21.T11148/fe078f6951993ca0b829 |
| 13 | 21.T11148/6f0d1c34a6ab5d67049f | 2024-05-31 19:42:07Z | C5H9NO3 |
| 100 | HS_ADMIN | 2024-05-31 19:42:07Z | handle=21.11152/USER01; index=300; [create hdl,delete hdl,create der admin,del admin,add admin,list] |

## Enables machine-readability and interpretability

nffa.eu PILOT 2021 2026

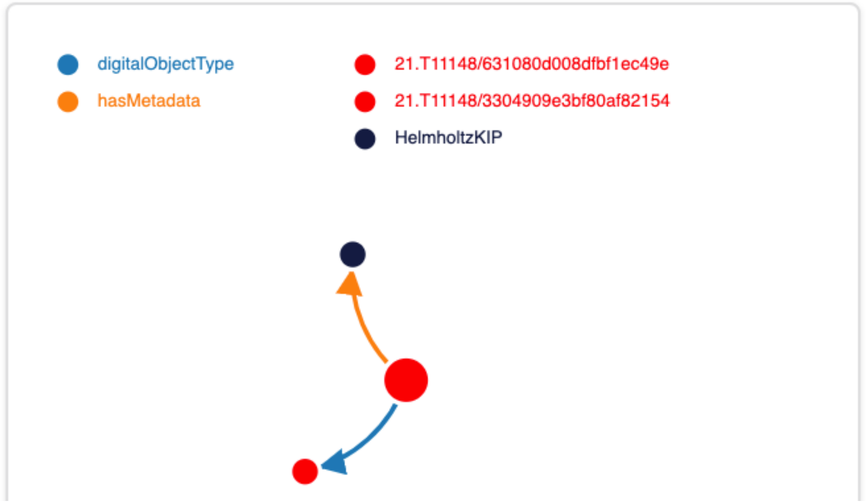
This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

# Tooling for Content Assessment

**PID Information Record**

| Type | Value |
|---|---|
| kernelInformationProfile | 21.T11148/b9b76f887845e32d29f7 |
| dateModified | 2023-08-01T00:00:00+00:00 |
| # checksum | { "sha512sum": "f0a6e42dc67335e6857b6 |
| dateCreated | 2023-02-07T00:00:00+00:00 |

**FAIR DO Graph**

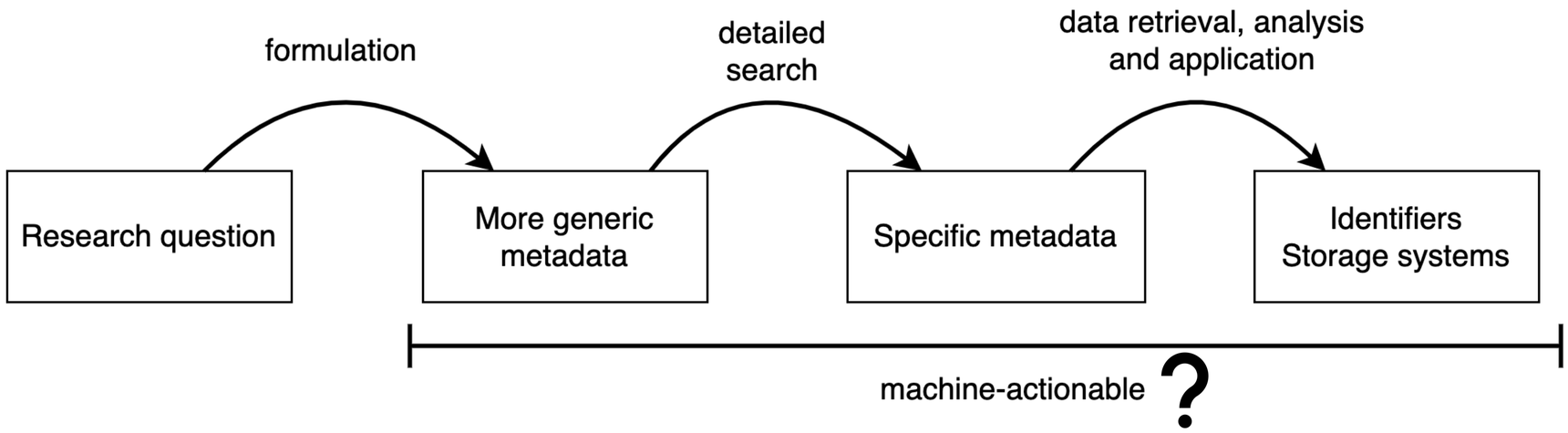



https://kit-data-manager.github.io/fairdoscope/?pid=21.11152/b0b5de04-6e11-480b-ab66-2d4a5f42ea9e

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

# Machine-Actionability as the Final Goal




- What do we need?
  - An entry point for the user
  - An infrastructure to implement the concept
  - Software to work with the components
  - A mechanism for machines to act on the components


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

# NEP Virtual Access Services

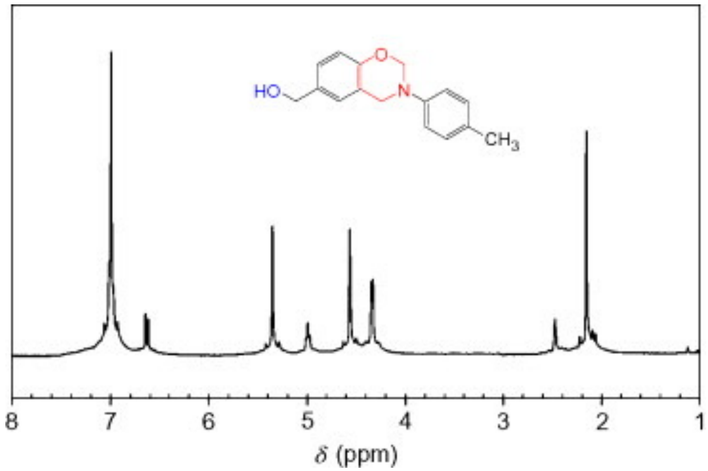
- NMR Graph for retrieval of NMR spectra resources
  - A service to provide a unified search interface of NMR spectra data
- MRI Prediction tool for the prediction of Magnetic Resonance Image data (DICOM format)
  - A service that uses AI software to estimate experimental outcomes by existing results

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

# NMR Graph Service - Motivation




https://www.sciencedirect.com/topics/chemistry/1h-nmr-spectrum

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | . | 1 | 1 | 1 | 1 | PCA | H | H | 1 | 8.27 |
| 2 | . | 1 | 1 | 1 | 1 | PCA | HA | H | 1 | 4.09 |
| 3 | . | 1 | 1 | 1 | 1 | PCA | HB3 | H | 1 | 2.24 |
| 4 | . | 1 | 1 | 1 | 1 | PCA | HB2 | H | 1 | 1.94 |
| 5 | . | 1 | 1 | 1 | 1 | PCA | HG3 | H | 1 | 2.05 |
| 6 | . | 1 | 1 | 1 | 1 | PCA | HG2 | H | 1 | 2.05 |
| 7 | . | 1 | 1 | 2 | 2 | GLN | H | H | 1 | 8.23 |
| 8 | . | 1 | 1 | 2 | 2 | GLN | HA | H | 1 | 4.26 |
| 9 | . | 1 | 1 | 2 | 2 | GLN | HB2 | H | 1 | 1.75 |

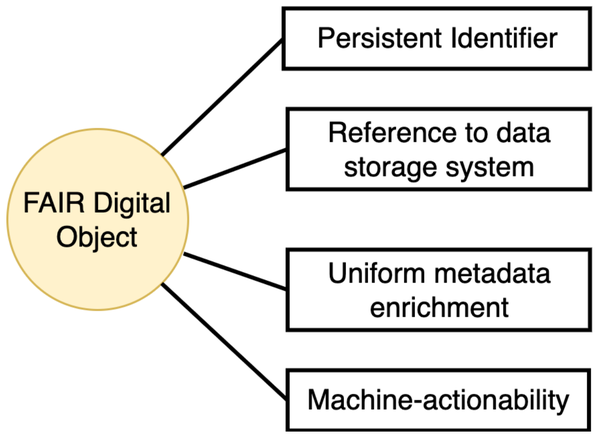
- NMR spectra are provided in different formats, are distributed over different storage systems and described using different metadata schemas
- Relations to related digital resources like publications, metadata documents or software are not easily assessable
- A common representation of various NMR spectra enables a unified search interface at this level
- Discovery, evaluation and retrieval is facilitated

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

# What is the Baseline?

- Certain types of metadata exists for all NMR spectra resources and is typically used for information retrieval
- This information can be unified and transferred into a machine-readable and interpretable format
- We used the concept of FAIR Digital Objects (FDOs) to describe various NMR spectra this way




NMRxiV

BMRB Biological Magnetic Resonance Data Bank

Biological Magnetic Resonance Data Bank

hmdb The Human Metabolome Database

Human Metabolome Database

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

# Graph Format for Extended Usability

- FDOs are entities that contain reusable, interconnected elements
- A graph representation enables the assessment of contents these FDOs describe
- Assessment by graph queries using SPARQL

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX fdoo: <http://anonymized-namespace/FDO-Graph>

SELECT ?profile ?operation ?expectedOutput ?fdo
WHERE {
  VALUES ?profileName { "Profile1" "Profile2" ... }
  ?profile a fdoo:Profile ; rdfs:label ?profile .
  FILTER(?profileName IN (?profile))
  ?fdo a fdoo:FDO ; fdoo:hasProfile ?profile ;
  rdfs:label ?fdo .
  ?operation a fdoo:Operation ; fdoo:isOperationFor
  ?fdo ; rdfs:label ?operation .
  ?operation fdoo:returns ?attribute .
  ?attribute a fdoo:Attribute ; rdfs:label
  ?expectedOutput .
}
```

- Search interface for users (via the GUI) and query endpoint for machines

- Try it out:
  - Visit: https://metarepo.nffa.eu/start_query



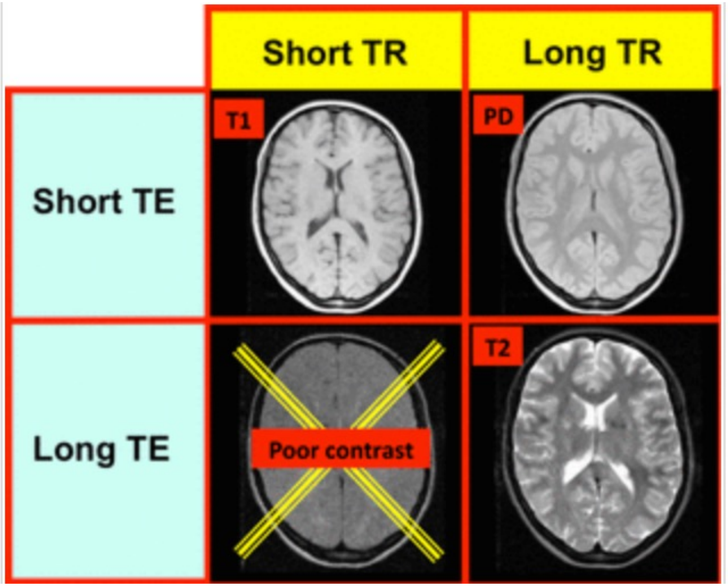
This initiative has received funding from the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe Pilot Project

# MRI Prediction Tool - Motivation

- Magnetic Resonance Imaging is a measurement technique mostly known from medical imaging – also applied in the materials science field
- Measurements take long time
- Often, many measurement sequences are required

for analytics

- The tissue contrast (T1, T2, PD) must be optimized




https://mriquestions.com/image-contrast-trte.html

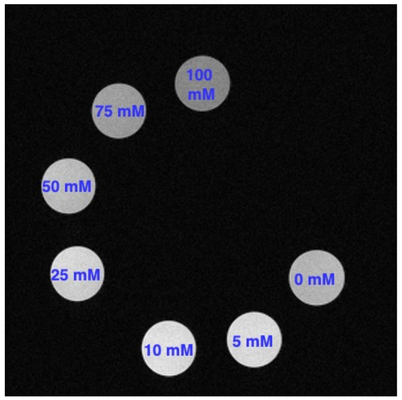
- **Can be reduced by digital acquisition of estimated measurement results**

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

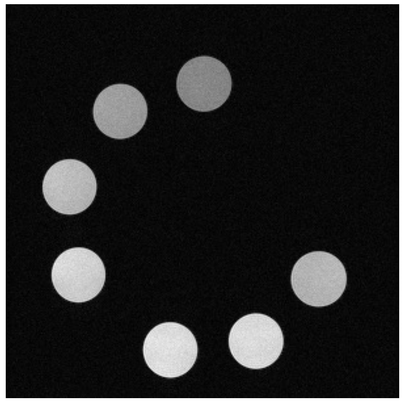
# The Approach

- Two main parameters that need to be adjusted – TE and TR
- Instead of measuring each parameter setting, a minimum of required experimental data is collected and applied to an AI model
- Model predicts the image of an alternative parameter setting

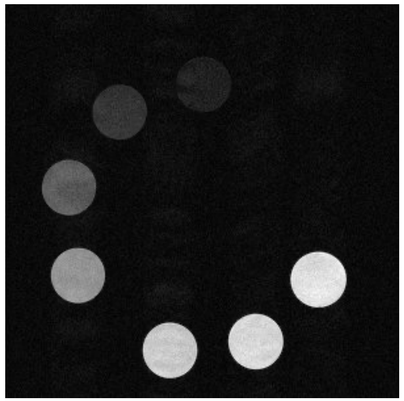
$CuSO_4$ in Millimolar (mM)




T1- weighted - TE: 5 ms
TR: 100 ms



T2- weighted - TE: 25 ms
TR: 5000 ms



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

# What is Possible?

- Currently, the model is specialized for a particular sample type
- Perspectives: prediction of images for a more versatile sample set

- Try it out:
  - Download a test file (DICOM format)
  - Visit: https://metarepo.nffa.eu/prediction

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