

FAIR approach for nanomaterial databases

Cecilia Bossa

Chiara Laura Battistelli

Environment and Health Department - ISS



Nanoinnovation 2020
New approaches for safety evaluation of nanomaterials

«Good research data management is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse.»

EUROPEAN COMMISSION
Directorate-General for Research & Innovation

H2020 Programme

Guidelines on

FAIR Data Management in Horizon 2020

Version 3.0
26 July 2016



«The value of data lies in its use and reuse»

Brussels, 19.2.2020
COM(2020) 66 final

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN
PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL
COMMITTEE AND THE COMMITTEE OF THE REGIONS

A European strategy for data



The FAIR principles



guide the
improvement of the
infrastructure
supporting the
reuse of data

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

Wilkinson M.J. et al. 2016 *The FAIR Guiding Principles for scientific data management and stewardship*
<https://doi.org/10.1038/sdata.2016.18>

FAIR principles (simplified)

Findable → easy to find for both humans and machines

Accessible → possible to humans and machines to access data:
persistent IDs, IT protocols,...

Interoperable → possible to combine and exchange

Reusable → possible to **reuse** the data

nanosafety data

Reuse Scenario

ECHA
EUROPEAN CHEMICALS AGENCY

GUIDANCE

Guidance on information requirements and chemical safety assessment

Guidance assessment

Appendix R.6-1 for nanoforms applicable to the Guidance on QSARs and Grouping of Chemicals

Version 2.0
December 2019

Version 2.0
December 2019



- Step 1: Identification and Characterisation of the nanoforms and sets of nanoforms of the substance
- Step 2: Initial grouping of (sets of) nanoforms – Develop a read across hypothesis, identification of a group's boundaries and its members
- Step 3: **Gather the available data** for each group member and evaluate the data for adequacy and reliability
- Step 4: **Construct a matrix of data availability**
- Step 5: Assess the adequacy of the approach and fill data gaps within the group of nanoforms/sets of nanoforms
- Step 6: Perform and/or propose testing to fill the data gap for the whole group of nanoforms/sets of nanoforms
- Step 7: Document the finalised grouping approach and refine the grouping rationale

Nanoreg2 strategy for grouping

What they are?

- Chemical Identification
• (e.g. composition, impurities)
- Physical descriptors
• (e.g. size and aspect ratio, shape, porosity, surface area)
- Computational descriptors
• (e.g. quantum-mechanical properties, descriptors related to composition, aggregation tendencies)

Physico-Chemical
Characterization of pristine
materials (as synthesized)

Where they go?

- Stability
• (e.g. chemical (solubility), physical (agglomeration), thermodynamic (phase transition))
- Surface charge
- Surface reactivity
- Bio-nano interface
- Mobility in different media
- Toxicokinetics
- Cellular uptake

System-dep. characterization, NM
fate, Toxicokinetics

What they do?

- Physical hazards
• Flammability, explosivity
- Human toxicity (acute/ chronic)
• (Cyto-)toxicity
• Genotoxicity
• Oxidative stress
• Inflammation
- Ecotoxicity (acute/ chronic)
• (Cyto-)toxicity
• Genotoxicity
• Oxidative stress

Hazards

eNanoMapper database

NanoSafety data

The screenshot displays the eNanoMapper database interface. On the left, there is a sidebar with a tree view of 'Data sources (14384)' and 'Nanomaterial type (21892)'. The main content area shows a search results page with a 'Hits list' and 'Selection' tab. The first result is 'Multi-Walled Carbon Nanotubes (MWCNT), synthetic graphite in tubular shape (Multi-Walled Carbon Nanotubes (MWCNT), synthetic graphite in tubular shape) mono constituent substance'. Below it, there are two other results for gold nanoparticles: 'G15.AC (G15.AC) gold nanoparticle' and 'G15.AHT (G15.AHT) gold nanoparticle'. Each result includes a 'Results' section with toxicological data and a 'Material Composition Studies' link.

eNanoMapper database

The eNanoMapper database is part of the computational infrastructure for toxicological data management of engineered nanomaterials, developed within the EU FP7 eNanoMapper project. More information:

Data content

- eNanoMapper
- NANoREG
- FP7 NanoTest
- NANOGENTOX
- FP7 ENPRA

Data access

Public

eNanoMapper
NANoREG

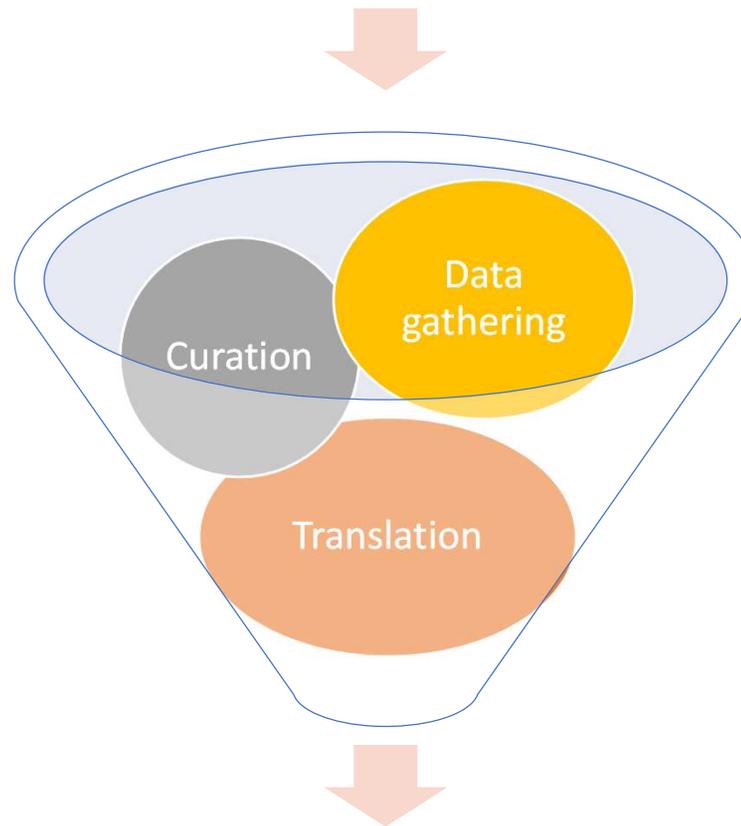
<https://search-legacy.data.enanomapper.net/>

What they do?

- Physical hazards
 - Flammability, explosivity
- Human toxicity (acute/ chronic)
 - (Cyto-)toxicity
 - Genotoxicity
 - Oxidative stress
 - Inflammation
- Ecotoxicity (acute/ chronic)
 - (Cyto-)toxicity
 - Genotoxicity
 - Oxidative stress

Hazards

many results available in
Nanoreg2 instance of eNanomapper DB



Grouping?

What problems have hampered optimal data reuse?

- Differences in SOPs
- Differences in templates
- Differences of information filled in the templates
(data entry procedure)

Differences in reporting between projects and within the same project

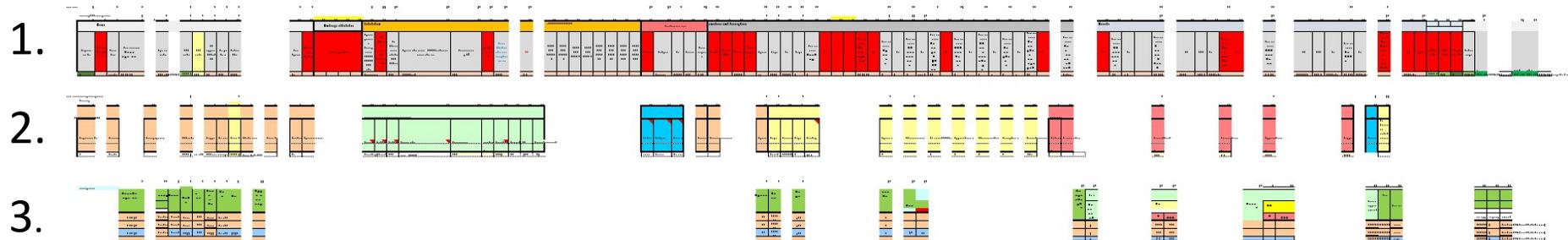
Example

In vitro Genotoxicity – COMET

3 templates

- Nanoreg (JRC original) adopted in Nanoreg2
- JRC modified by partner in Nanoreg
- Nanogenotox

Different templates: mapping fields between templates



1. JRC template, mod. Nanoreg
2. JRC template original (Nanoreg2)
3. Nanogenotox template

No. common fields	JRC Modified (Nanoreg)	JRC (Nanoreg2)	Nano genotox
JRC Modified (Nanoreg)	90	39	17
JRC (Nanoreg2)	39	46	11
Nano genotox	17	11	28

How to improve nano-data FAIRness?

Findable → easy to find for both humans and machines:
searchable resource

Accessible → possible to humans and machines to access data:
persistent IDs, IT protocols,... → **IT technical**

Interoperable → possible to combine and exchange:
relevant standards, controlled vocabularies...

Reusable → possible to reuse the data:
rich metadata, relevant attributes and relevant domain standards

Standardization of templates and data entry procedure

Templates

- To decide which is the minimum set of fields needed to characterize the experiment
- To harmonize the name of the fields (unique, universally understandable, etc)

Data entry procedure

- To use controlled vocabularies wherever possible (picklists?)
- To harmonize the field formats (numerical/text)
- Links to ontologies
- Documentation to guide the data entry / search procedures

→ Gov4nano & NMBP-13 projects

Acknowledgements

ISS Multidisciplinary Unit on Nanomaterials and Nanotechnologies

This work was performed within the EU project **Gov4Nano**, funded by the Horizon 2020 Research and Innovation Programme under Grant Agreement 814401