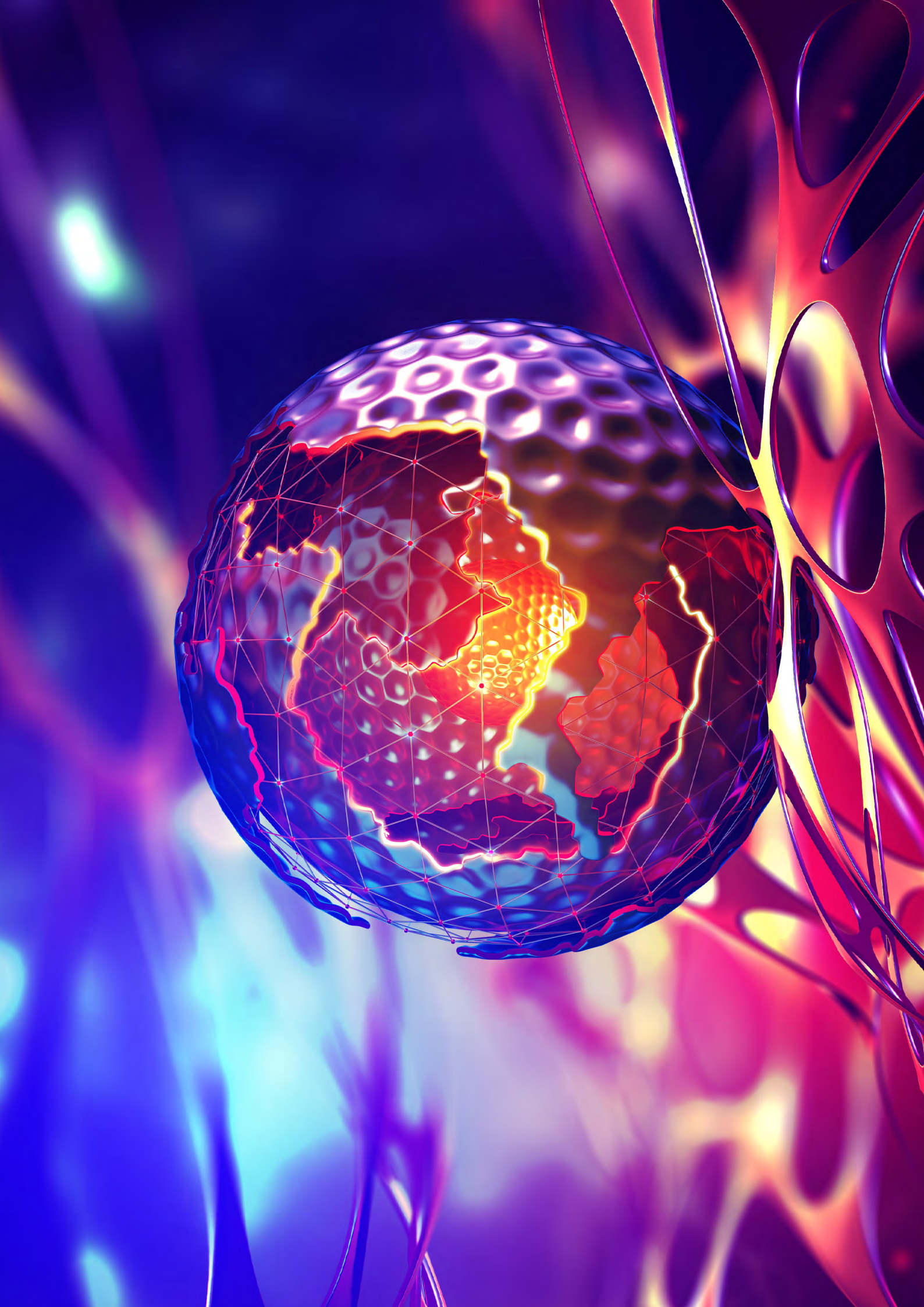
Governance brief 2

IMPLEMENTATION OF FAIR PRINCIPLES IN NANOSAFETY DATA MANAGEMENT

Gaps and recommendations for environmental, health and safety research on nanomaterials



INTRODUCTION

This policy brief provides challenges and proposed lines of actions for reusing nanosafety data in the perspective of the FAIR principles. The most important needs in the FAIRification process are highlighted and recommendations are included to improve the reuse of data with the aim of better risk governance on nanomaterials.

The ambition of this policy brief is twofold: 1) to raise awareness and inform about the importance of FAIR data for nano risk governance and the reuse of nanosafety data, and 2) to appeal to the European Commission, the nano and advanced materials research community and other stakeholders such as industry to actively support the identified lines of action.

The brief is an outcome of four years of research activity on the topic of three European projects, including consultation of a wide range of experts in the field and stakeholders (e.g, data generators). Details on the methodology are reported in Gov4Nano deliverable 5.6: Report on case studies for Risk Governance available on **www.gov4nano.eu**.

It is part of a series of initiatives organized by the **Gov4Nano project**, in cooperation with the **RiskGONE** and **NANORIGO** projects, to discuss governance issues in the development of nanomaterials, with the goal to promote the safety and sustainability of innovation in nanotechnology in Europe.

BACKGROUND

Nanotechnology is a key enabling technology with significant global investment from public and private players.

The widespread use of nanomaterials and nano-related products can only be sustainable if comprehensive strategies to deal with potential environmental, health and safety (EHS) issues are adopted, in line with existing regulations. Relevant research efforts on these aspects, including funding of large collaborative projects, have been put in place since more than a decade. In this picture, the reuse of accumulated data is clearly lagging behind. Analysis of the large EHS data supply is paramount to maximize the potential offered by the acquired knowledge and lead to widely applicable policies and regulations.

Data sharing is a first essential step. Indeed, making data, in every scientific discipline, publicly available is among the main objectives of open science advocates. Data availability is not enough, however, unless proper curation can aid its reuse and analysis. A remarkable effort towards the efficient reuse of scientific data was the publication, in 2016, of the FAIR principles (Wilkinson et al 2016) by a large consortium of authors. The acronym FAIR stands for 'findable, accessible, interoperable and reusable'. Briefly, this means that each set of data must have an identifier, that enough information is provided for the data and the metadata to be accessed, that a set of data can be integrated with other sets and can be analysed by a range of software, and that permissions for reuse are clearly specified (FAIR Principles).

A noticeable feature of the FAIR principles is that they aim at "enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals". This is becoming more urgent, especially when considering the potential of machine-learning procedures to analyse large datasets and provide guidance for further scientific developments, but also for a larger improvement of the application of grouping and read-across and the development of QSARs methodologies.

In the EU, the FAIR principles have been recognized at policy level (Collins et al 2018) and FAIR data are considered key components of the new EU Chemicals Strategy for Sustainability (CSS). In fact, FAIR data management is a crucial requirement for many CSS actions to be carried out effectively including:

- the development and application of Safe and Sustainable by Design approaches and pre-regulatory tools
- the development of a Common open Data Platform on Chemicals
- the innovation and speeding up of hazard and risk assessment (e.g., through New Approaches Methodologies, grouping approaches and supporting "one substance, one assessment")
- the accessibility and reuse of academic data for regulatory purposes
- the streamlining of chemical data flow between EU and national authorities
- improving availability of information on chemical content and safe use of products and materials.

The Horizon 2020 projects Gov4Nano, NANORIGO and RiskGONE are committed to improving the FAIRness of data on the safety of nanomaterials for humans and ecosystems, to maximize their availability, understanding, exchange and ultimately their reuse, which is necessary to develop an efficient and effective risk governance process for nanotechnologies.

THE PROBLEM AT STAKE

The risk assessment of nanomaterials (NMs) implies an extensive characterization of a series of physico-chemical parameters, usually not relevant or not applicable in the case of bulk substances, whose variation could have an impact on their reactivity and eventually on their toxicity. Unlike conventional chemicals, the complexity in the characterization of NMs only begins with the chemical composition, and further comprises the use of other parameters, such as crystallinity, particle size, particle shapre, surface chemistry and specific surface area.

Moreover, in the evaluation of their potential toxicological effects there is the need to consider the dynamics of NMs in media. It is already established that some physicochemical properties of NMs (e.g., surface chemistry, aggregation/agglomeration state, dispersibility) can drastically change in the interaction with the medium, affecting NMs kinetics, bioavailability, and eventually their toxicity. Furthermore, NMs have an inherently ambiguous nature, since their structural properties are characterized by a distribution of values, rather than a single, well-determinable number. These peculiar characteristics add complexity to the challenge of determining the (eco)toxicological effects of NMs.

In the last 10-15 years, a wide range of experimental data on the variables mentioned above have been collected for different nanomaterials. Reuse of these data is a crucial aspect of moving towards efficient data-driven risk assessment, for the development of predictive models and in general for advancing knowledge on NMs mode and mechanisms of toxic action.

Nevertheless, the reuse of existing data is currently limited, hampered by several obstacles, such as poorly described (meta)data, non-standard terminology and a lack of harmonized reporting formats and criteria.
In response to the realization that research data is not reused to its full potential, the FAIR principles have been developed, describing and guiding data generators in the key aspects of data handling to make data Findable, Accessible, Interoperable and overall Reusable.

CHALLENGES AND BARRIERS

Major technical barriers for the reuse of nanosafety data from the FAIR principles perspective were identified and are listed in the figure and text below (Jeliazkova2021, Bossa2021):

FINDABILITY	Lack of persistent IDs for data and metadata and lack of indexed searchable databases
ACCESSIBILITY	Difficulties in identifying and retrieving data;
INTEROPERABILITY	Lack of harmonized terminology, reporting formats and standards and supporting information and tools;
REUSABILITY	Poorly described metadata, different levels of data pre-processing, difficulties in results interpretation, unclarity of data licenses.

FIGURE 2. TECHNICAL AND CULTURAL OBSTACLES FOR THE REUSE OF NANOSAFETY DATA.

Technical barriers

FINDABILITY
Lack of persistent IDs for (meta)data, lack of indexed searchable databases

ACCESSIBILITY
Difficulties in identifying and retrieving data

INTEROPERABILITY
Lack of harmonized terminology, reporting formats & criteria, supporting information and easy-to-use FAIRification tools

REUSABILITY
Poorly described metadata, different levels of data pre-processing, difficulties in results interpretation, unclarity of data licenses

Cultural barriers

LACK OF AWARENESS OF FAIR-PRINCIPLES

LACK OF SKILLS HOW TO FAIRIFY DATA

LACK OF INCENTIVES FOR DATA PROVIDERS

LACK OF TIME AND RESOURCES

DATA MISUSE AND PRIVACY CONCERNS

In addition, implementation of the FAIRification process that would lead to the resolution of these above issues is hampered by several barriers (Gov4nano Del. 1.2), such as:

LACK OF AWARENESS	The majority of stakeholders were unaware of FAIR and did not apply FAIR in their work. It is clear that increased awareness and provision of guidance regarding the FAIR principles could help to encourage more researchers to make their data available more widely.
SKILLS AND/OR TOOLS REQUIRED	The need for suitable tools, training and guidance was consistently highlighted across the surveys and reviews.
LACK OF INCENTIVES FOR DATA GENERATORS	As data management is perceived as a very time-consuming operation, researchers for the most part believe that incentives are needed for scientist to prioritize data-sharing activities over other activities.
TIME AND/OR RESOURCE REQUIREMENTS	Lack of time and resources to make data FAIR compliant were seen as the greatest barriers to its implementation.
DATA MISUSE AND PRIVACY CONCERNS	Researchers are concerned about the potential for inappropriate use of their data, and the possibility for data to be misunderstood. Concerns about privacy and confidentiality were also expressed, particularly those who worked with human subjects. Researchers are also concerned that visibility of data ownership may be lost due to the involvement of multiple people and institutions.

WAY FORWARD

The key to move forward is to establish what research and policy actions should be undertaken to overcome the gaps and roadblocks identified and who should be responsible for their implementation (policy, research, industry, other actors).

DEVELOPMENT OF PRACTICAL DATA MANAGEMENT TOOLS

The lack of skills identified with regard to the application of the FAIR principles indicate a need for simple-to-use data FAIRification tools. A range of user-friendly solutions to overcome the major technical barriers for FAIR adoption in nanosafety data management were implemented during the Gov4Nano project in the Nanosafety Data Interface, available at https://search.data.enanomapper.net/ (as described in Gov4Nano deliverable 1.3.). The Nanosafety Data Interface, originally established to collect data and solve reuse needs of the nanosafety community, provides nanosafety-community-specific solutions aligned with all the FAIR principles (Hastings et al. 2015).

- Automatically generated templates and related services and user interface, to minimize the efforts by data providers. An online

FIGURE 3. LOWERING THE BARRIERS TO THE REUSE OF NANOSAFETY DATA BY THE GOV4NANO, NANORIGO AND RISKGONE PROJECTS AND HOW TO MOVE FORWARD.

Delivered in NMBP-13 projects

DEVELOPED TOOLS IN NANOSAFETY DATA INTERFACE

- Template wizard
- Visual summaries
- Data completeness tools
- A visual guide to FAIRness

LAUNCH OF GO FAIR ADVANCEDNANO IN

DEVELOPMENT OF FAIRIFICATION WORKFLOWS

Way forward

SUPPORT CITATION STANDARDS FOR DATA SHARING

STIMULATE TRAINING AND EDUCATION

DEVELOPMENT OF A LICENSING SCHEME

AVAILABILITY OF RESOURCES

Template Wizard, currently including 55 templates for a variety of physicochemical characterisations, in-vitro assays and omics metadata has been developed through a co-creation process. The templates are available through the **NanoSafety Data Interface**. As a step towards the automatic FAIRification, a template validator was additionally developed to verify if a user-template can be correctly interpreted (parsed) by the eNanoMapper FAIRification workflow.

- Visual summaries and visual queries to improve the findability and the overview of data in the Nanosafety Data Interface. To improve the findability and the overview of data in the Nanosafety Data Interface, a **dashboard**, featuring more than 25 different visual summaries of stored data, has been implemented and is available. By selecting a type of summary at the top right of the dashboard webpage (e.g., toxicity data, materials, data availability, dose-response information), an overview of this type of data currently included in the database is given.
- Data completeness tools to help in addressing data quality. It is one of the factors that underlies data quality (other factors are relevance and reliability, Marchese Robinson et al.).
  In the Nanosafety Data Interface, the completeness of the data for each category (physicochemical properties, toxicity, etc.) is presented as a percentage. The completeness percentage is calculated against a predefined list of minimum reporting standards for different experiment types (roughly following the OECD Harmonized Templates). This calculation is presently for demonstration purposes, and the list of minimum reporting standard should be discussed with domain experts. A complementary approach has been developed by the NanoCommons project where data completeness is accessed against different sets of specific criteria/reporting standards (a collaboration with the nanosafety Data Interface has been already started).

GUIDANCE AND TRAINING

- Lack of awareness and know-how of FAIRifying data can be addressed through the provision of training resources and guidance, with survey results indicating a need for introductory information and materials. In addition, completion of relevant training should lead to more efficient implementation of FAIR thus saving researchers time and resources.
- A **visual guide** to FAIRness was created early in the project to address awareness of the FAIR principles and their interpretation

PROMOTION OF AND ACTIVITIES BY GO FAIR ADVANCEDNANO IMPLEMENTATION NETWORK

- To smooth the process of data FAIRification for all stakeholders a GO FAIR Implementation Network called AdvancedNano was launched in 2020 (Dumit et al, submitted). This community brings together experts with the common intent of maximising the value of scientific data generated within nanosafety research, fostering the implementation of FAIR principles. The planned actions of AdvancedNano fall in four categories: people's awareness, FAIRification tools, harmonisation and infrastructure. The first three categories can be recognized in this governance brief.

AVAILABILITY OF PRACTICAL CASE STUDIES AND EXAMPLES, HARMONIZATION AND STANDARDIZATION OF DATA CURATION PRACTICES	Development of FAIRification workflows for improving data reuse in different case studies: Compliance with the FAIR principles does not automatically indicate that data is of high quality or fit for purpose (Jeliazkova et al. 2021). Nonetheless, activities such as quality assessment, curation (relating to what is known as completeness of data, referring to whether enough metadata is available) and translation, are all part of the main activity, i.e., reuse of the data (Jeliazkova 2021). Case studies for evaluating data reusability in specific reuse scenarios were carried out in the Gov4Nano project (deliverable 1.4). To support better interpretation of the results from each case study a harmonized approach was developed with a focus on each of the identified areas of activities, including quality assessment, curation and completeness assessment, and translation. These are all activities which depend on the intended reuse scenario.
SUPPORT CITATION STANDARDS FOR DATA SHARING	Incentives for data sharing could come from citation standards for data, and many journal publishers now have clear guidelines encouraging open access and providing guidance on data repositories and the citation process for data sets. Nonetheless, the vast majority of journals currently only 'encourage' or 'expect' rather than 'mandate' sharing of data associated with the papers that they publish. In addition, there remain a substantial number of journals that require high fees for open access which is an active disincentive to open sharing of research in general and data sharing in particular.
STIMULATE TRAINING AND EDUCATION	Further training and guidance on sharing data, including clarification that the data generator can still retain some control over shared data through placing restrictions on how their data is used (for example by detailing acceptable use terms or by requiring granting of specific permission prior to any data re-use) would help address worries about data misuse.
DEVELOPMENT OF A LICENSING SCHEME	Remembering that FAIR data does not equal open data, following the principle 'as open as possible and as closed as necessary', further development of a licensing scheme specifically for data sets could be envisaged.
AVAILABILITY OF RESOURCES	Measures to address concerns regarding availability of resources could also include a stronger recognition of the need for, and benefits of, FAIRification by funders and sponsors, and the inclusion of this topic within research calls, and invitations to tender. This would allow provision for FAIRification to be integrally included in study proposals and protocols and for resource time and funding to be explicitly costed into the bids.

FINAL REMARKS

Although the scientific community is generally keen on sharing and reusing data, recognizing the benefits through publications, peer recognition and collaboration opportunities, as well as the general benefits through the exploitation of existing data, effective data reuse is hampered by several obstacles.

This policy brief provides an overview of issues and challenges in the reuse of data on the safety of nanomaterials from the perspective of the FAIR principles. In addition, roadblocks, and barriers to the FAIRification process are listed.

Practical solutions to facilitate the implementation of FAIR principles in nanosafety data management have been identified and implemented through the Nanosafety Data Interface infrastructure.

Further lines of action supporting the FAIRification process, with the aim of improving data reuse, are proposed to be addressed by different stakeholders, including the AdvancedNano GO FAIR IN, the research community, funding agencies, journal publishers, data managers and data providers.

KEY REFERENCES

- Bossa C, Andreoli C, Bakker M, Barone F, De Angelis I, Jeliazkova N, Nymark P. and Battistelli C.L. 2021. **Fairification of Nanosafety Data to Improve Applicability of (Q)Sar Approaches: A Case Study on in Vitro Comet Assay Genotoxicity Data.** Computational Tox. https://doi.org/10.1016/j.comtox.2021.
- Collins S, et al., Turning FAIR into Reality. **Final Report and Action Plan from the European Commission Expert Group on FAIR Data**, European Commission (2018), https://doi.org/10.2777/1524.
- Dumit V, et al. **From Principles to Reality. FAIR Implementation in the Nanosafety Community Comment** article submitted to Nature Nanotechnology December 2022
- **European Union Observatory for Nanomaterials (EUON),** hosted and maintained by the European Chemicals Agency (ECHA), https://euon.echa.europa.eu/uses
- Gov4Nano **Deliverable 1.2 Report and ready-to-use methods (tools, training sessions) on how to improve data findability and accessibility.**
- Gov4Nano **Deliverable 1.3 Operating interoperable databases and report on how to sustain them.**
- Gov4Nano **Deliverable 1.4 Report on how to improve data reusability, with case studies.**
- Hastings J, N. Jeliazkova, G. Owen, G. Tsiliki, C.R. Munteanu, C. Steinbeck, E. Willighagen, **eNanoMapper: harnessing ontologies to enable data integration for nanomaterial risk assessment,** J. Biomed. Semant. 6 (2015) 10, https://doi.org/ 10.1186/s13326-015-0005-5.
- Jeliazkova N et al. **Towards FAIR nanosafety data.** Nature Nanotechnology (2021) 16, 644–654, https://doi.org/10.1038/s41565-021-00911-6
- Marchese Robinson RL, et al. **How should the completeness and quality of curated nanomaterial data be evaluated?** Nanoscale 19 (2016) https://doi.org/10.1039/C5NR08944A
- **Making nano data FAIR enough.** Nat. Nanotechnol. 16, 607 (2021). https://doi.org/10.1038/s41565-021-00935-y
- Wilkinson M.J. et al. **The FAIR Guiding Principles for scientific data management and stewardship Scientific Data** (2016) 15;3:160018, https://doi.org/10.1038/sdata.2016.18

Brief: Implementation of FAIR principles in nanosafety data management

Main authors: Cecilia Bossa, Chiara L. Battistelli, National Institute of Health, (ISS) Italy;
Martine Bakker and Cornelle Noorlander, National Institute for Public Health
and the Environment (RIVM), The Netherlands

Other briefs of the series available on: www.gov4nano.eu

The Gov4Nano, NanoRigo, RiskGone
projects have each received funding under the EU's Horizon 2020 R&I Programme.