INTEGRANO

MULTIDIMENSIONAL INTEGRATED QUANTITATIVE APPROACH TO ASSESS SAFETY AND SUSTAINABILITY OF NANOMATERIALS IN REAL CASE LIFE CYCLE SCENARIOS USING NANOSPECIFIC IMPACT CATEGORIES

This project has received funding from the European Union's Horizon Europe research and innovation programme under GA No 101138414

Project acronym	INTEGRANO
Work Package	1
Type of document	Report
Dissemination Level	PU
Deliverable	D1.1 First Data Management Plan
Lead Beneficiary	CNR
Contractual delivery date	31/03/2024 (M3)
Actual submission date	28/03/2024 (M3)
Author(s)	Murray B., Perucca M. (PRJ)

D1.1
First Data Management Plan

Revision and history chart

Version	Date	Main author(s)	Summary of changes
1	01/02/2024	Murray B.	First draft to be shared
1.1	25/03/2024	Murray B., Perucca M.	First review
1.2	28/03/2024	Costa A., Murray B., Perucca M.	Final review and upload to the F&T portal

Disclaimer

Acknowledgment

This document is a deliverable of the INTEGRANO project. This project has received funding from the European Union's Horizon Europe research and innovation program under grant agreement No 101138414.

Abstract

This deliverable presents the first Data Management Plan (DMP) of the INTEGRANO project. The plan is based on the 'Data Management Plan' template document issued by the European Commission [1] and aims to provide answers to the questions addressed by the Open and Findable, Accessible, Interoperable and Reusable (F.A.I.R.) criteria. This preliminary DMP encompasses a significant part of the life cycle of the data produced within INTEGRANO. It captures how data will be captured and handled during the project, how data will be made F.A.I.R. [2], and how data will be deposited in the long-term to ensure access to current and future nanosafety projects and other stakeholders.
The INTEGRANO project relies on completeness and harmonisation of datasets generated within the project. Since this is an essential component to the project success, a dedicated platform called INTEGRANO Data Management plaTform (DMT) is being implemented on purpose. This is as a step forward w.r.t. to other projects as the IDMP will not only enable generating and maintaining F.A.I.R. data but also to support tasks management planning of activities, monitoring tasks progress, defining and tracing internal project deliverables that will guarantee the reference project deliveries.

List of abbreviations and acronyms

Abbreviation / acronym	Meaning
DMP	Data Management Plan
DMT	(INTEGRANO) Data Management plaTform
f2f	"face to face" (referred to meetings)
F.A.I.R.	(referred to data and datasets) Findable, Accessible, Interoperable, Reusable
IC	Impact Category
KDFs	Key Decision Factors
KPIs	Key Performance Indicators
LCS	Life Cycle Stage
NF	Nanoform
NM	Nanomaterial

List of figures

Table of Contents

1. General Description of the INTEGRANO DMP and DMT

The INTEGRANO project addresses shortcomings and gaps in current safe and sustainability assessment approaches of Nano Materials (NMs) and Nano-enable products (NEPs). The project proposes a general assessment approach based on quantitative evidence to be applied in practice for specific Nano Materials (NMs) design cases - inorganic, organic and carbon NMs. The effort in the development of NMs novel impact categories (ICs) for nanotoxicity and eco-nano-toxicity assessment aims at the integrated application of standardised assessment methodologies. The synthesis, incorporation, use phase and end-of-life NMs life cycle stages (LCS) will be addressed for targeted case studies. The collected/generated data will be used to meet the project's objectives to formulate design hypotheses and make design decisions by applying a data-driven approach and methodology, providing stakeholders with a digital supported decision process to tackle the Safe and Sustainable by Design (SSbD) challenge in the NMs context.

1.1 Data Generation and data management platform

During the project, INTEGRANO will generate data across a range of activities, including physio-chemical characterization of nanoparticles and nano-enabled products, human hazard assessment, exposure assessment, environmental assessment, life cycle assessment and computed data derived from in silico or computational modelling (life cycle assessment, risk assessment).

Generated data type will be:

- Alphanumeric strings
- List of numerical values for continuous and discrete variables
- Tables including alphanumeric data
- Diagrams (bar charts, 2D-, 3D-plots, contour plots diagrams, spectrograms, Gantt diagrams, network and flow diagrams)
- Images
- Documents (e.g. pdf)
- Database queries
- Data templates

The collection, storage and sharing of data within the INTEGRANO project will be mediated through the INTEGRANO Data Management Platform (DMT, figures 1 and 2) and associated INTEGRANO database. The DMT, which has already been released to partners and will be periodically updated, is bespoke software with the aim of managing the demand for resources within the project to streamline data collection. The DMT facilitates the design of each of the chosen design case studies through a sequential process where the Case Study manager will input the requirements of the work to be performed to build a complete picture of each life cycle stage to be constructed in terms of experimental timescales, number of samples to be produced, and resource demands particularly those associated with characterisation techniques. For example, the DMT will enable each case study manager to request provision of data collection resources for characterisation of the NMs and/or NEPs, enabling the demand for each service to be managed and allocated in a proportionate manner according to the time and financial resources allocated to each data collection element. The DMT provides a structured interface to ensure that data collection across the life cycle stages is complete with respect to what is required to achieve the goals of the project, and as described in the following pages, ensures that for each case study a harmonised dataset is produced which adheres to F.A.I.R. data principles. The data to be collected will be stored within the INTEGRANO database. As listed in

the previous paragraph, INTEGRANO will generate a broad range of data across the LCS of each NM/NEM. The INTEGRANO database will also integrate data from existing EU databases. For each case study, Key Decision Factors (KDFs) will be specified, and Key Performance Indicators (KPIs) related to the five dimensions of safety, sustainability and functionality will be defined with values obtained from experimental results and computational modelling. A Decision Support Tool (DST) will be developed for the INTEGRANO project which, for each case study, will assess the correlation between set KDFs values in the DoE matrices and the obtained KPIs values, and will utilise a Multi-Objective Optimisation Design (MOOD) algorithm to identify from all possible design cases the multi-optimal ones, simultaneously complying with safety, sustainability and functional KPI requirements, providing the user with viable SSbD design cases from which to select. The INTEGRANO project will also establish a Think Tank platform for stakeholder engagement which is dedicated to sharing data for integrated sustainability assessment and SSbD methodology co-creation.

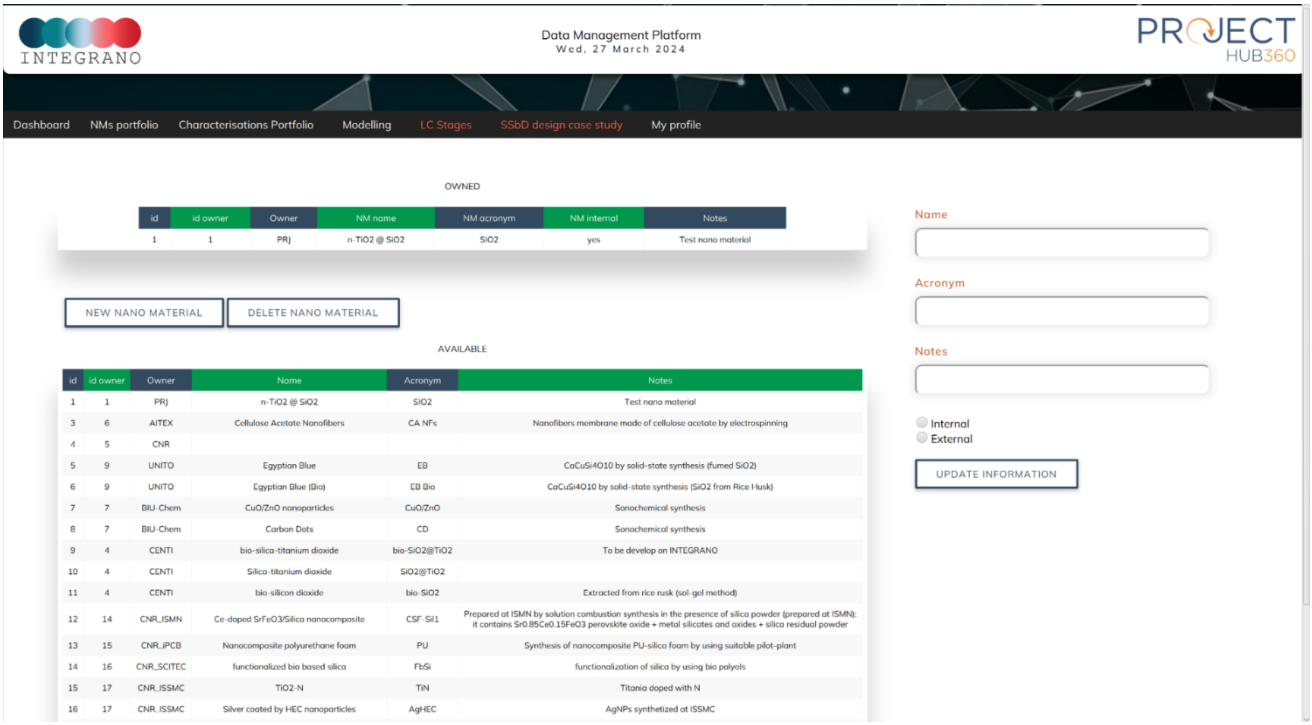


Figure 1 *Screenshot of INTEGRANO Data management platform (DMT) showing the INTEGRANO nanomaterials portfolio, with NM owner. NM ID is displayed as well as associated owner identification code associated to database record. Specification of NM source is also provided (if internal to the consortium or external source)*

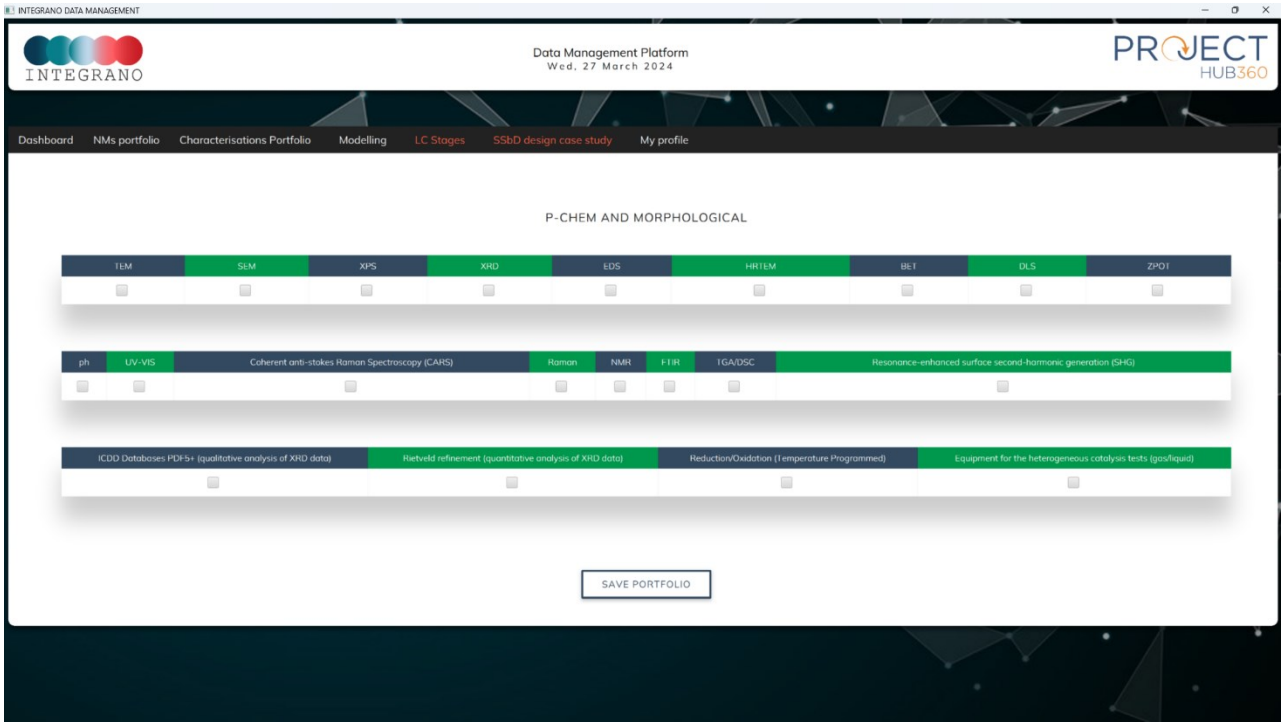


Figure 2 *INTEGRANO DMT interface. Screenshot of the INTEGRANO characterization portfolio, which enable defining which and how many characterizations are needed for a specific NM addressed in a defined case study.*

Think Tank activities will include e-learning courses, webinars, f2f meetings, matchmaking, and use of the INTEGRANO database for data sharing. The Think Tanks will be supported by the development of a dedicated IT platform.

The INTEGRANO website will also contain a private area accessible only to project partners, and through which documentation will be shared e.g. project documentation, technical and financial reporting.

This deliverable presents the initial First Data Management Plan (DMP) of the Multidimensional Integrated Quantitative Approach To Assess Safety And Sustainability Of Nanomaterials In Real Case Life Cycle Scenarios Using Nanospecific Impact Categories (INTEGRANO) project.

The plan is based on the 'Data Management Plan Template' document issued by the European Commission. It is a living document that will be updated over the course of the project, with the final plan to be delivered at month 18 of the project. The first version of the DMP will provide initial answers to the questions asked in the DMP template document, including those associated with Open and Findable, Accessible, Interoperable and Reusable (F.A.I.R.) criteria, with subsequent revisions providing finer levels of detail appropriate to the implementation stage of the project.

This preliminary DMP encompasses a significant part of the life cycle of the data produced within INTEGRANO. It includes how data will be captured and handled during the project, how data will be made F.A.I.R., and how data will be deposited in the long-term to ensure access to current and future nanosafety projects and other stakeholders.

1.2 SSbD Workflow and related data generation

Data generation in INTEGRANO follows a specific protocol which is aligned to the guidelines of SSbD framework for chemicals and materials [3] which finally enables to operate the Decision support Tool supporting safe and sustainable by design analysis. The general framework in which data are generated is exemplified in the following chart (see Fig.3).

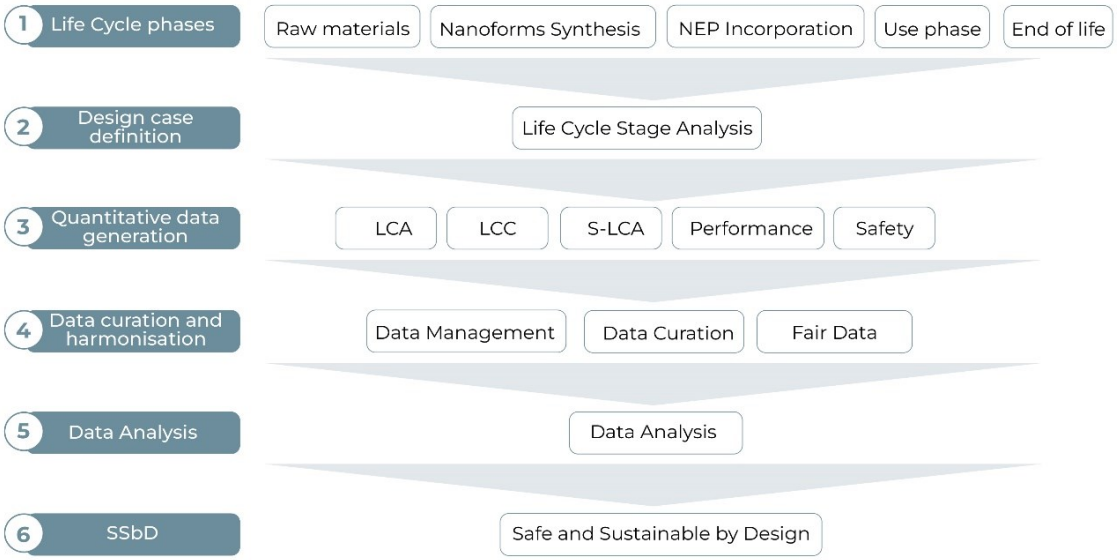


Figure 3 *General view of data generation from each Life Cycle Phase towards the generation of the Safe and Sustainable by Design solutions by Integrano DST.*

A specific protocol for the NMs and NEP SSbD analysis is put in place and essential to the activities of the case study designer (or team of designers). The SSbD implies data generation along the following workflow, and data streaming. Since INTEGRANO is aligned with and follows already existing reference standards the SSbD framework has been taken as a reference paradigm and the ISO14040-44 standard for Life cycle Assessment as a reference for conducting the study, for primary and secondary data collection and generation, for subsequent new data generation by computation and elaboration through the SSbD analysis. The SSbD workflow and related data generation and processing is organised in four stages:

i. Goal and scope definition.
ii. Data generation and Inventory by experimental and modelling work, data curation and F.A.I.R.ness compliance assessment.
iii. SSbD impact assessment: digital computation of the set of SSbD solutions with INTEGRANO DST.
iv. Interpretation of results, possible protocol reiteration, or final decision on the SSbD selected case operated by the NM (NEP) designer (decision maker).

SSbD Ph1, goal and scope definition. Reference design cases KDFs values data generation.

The first phase of the human-centric SSbD workflow is the goal and scope definition (see Fig.4), which implies specific and fundamental decision are taken by the designer (or design team):

- Goal of the assessment.

- Scope of the SSbD analysis: extent of the functional system and setting of the system boundaries to be analysed.
- Definition of the functional unit (FU) (in connection to the KPIs).
- Setting of hypotheses and assumptions.
- Definition of cut offs, data precision and accuracy.
- Definition of addressed (quantitative) KPIs.
- Identification of the Key decision Factors (KDFs) affecting the addressed KPIs:
  - Definition of the KDF types.
  - Definition of the total number of KDFs.
  - Definition of the variability range of KDFs.
- Definition of the Design of Experiment Matrix, setting the experimental and modelling plan.

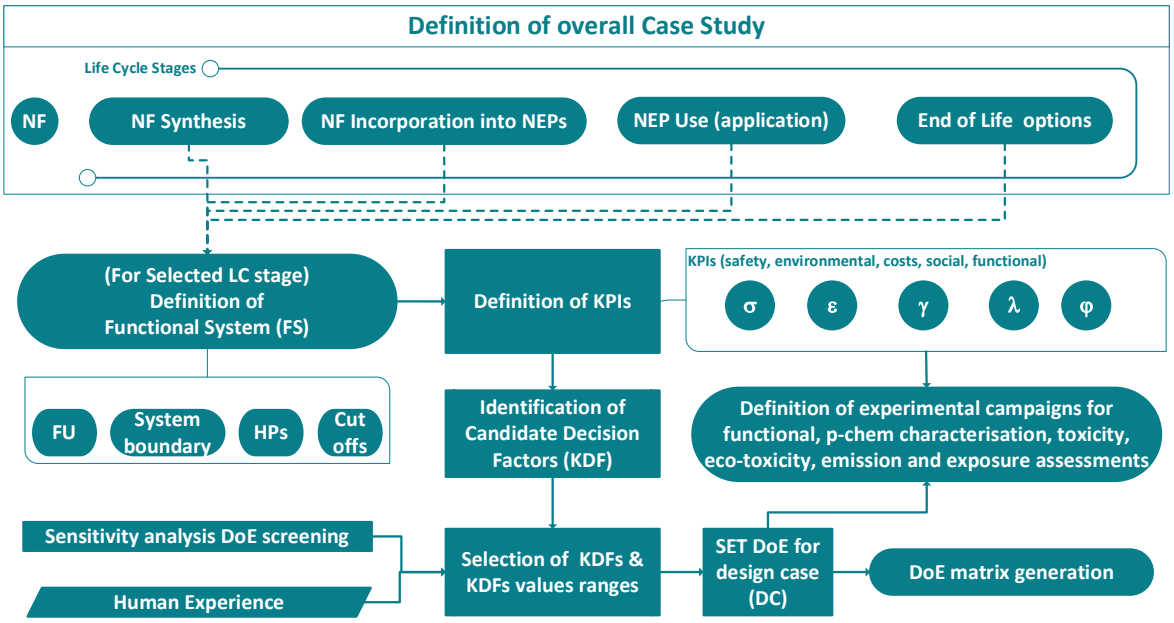
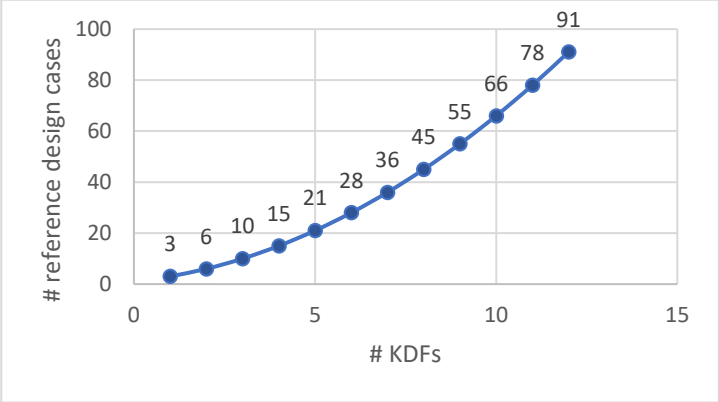


Figure 4 *Case study goal and scope definition and definition of the Design of Experiment (DoE) matrix, which the case study is based on. KPIs definition involves the following dimensions, which the identified KPIs relate to: (σ) safety, (ε) environment, (γ) economic-costs, (φ) functional and (λ) social.*

Definition of the number of KDFs is quite crucial since it strongly affects data generation associated to the definition of the minimum number of reference design cases (RDCs) in the DoE matrix to be considered. In the following chart the scaling of the RDCs as a function of the KDFs number is reported.



SSbd Ph2: Experimental and modelling Data Generation: reference design cases KPIs

Figure 5 *scaling of the number of reference design cases with respect to the number of key decision factors. It is important to observe that for each design case more measurements need to be done in order to obtain an average value and uncertainty of the experimental measurement. Therefore, it is quite evident that the selection of the minimum and sufficient number of KDFs to describe the impact on KPIs is key.*

Data generation consists in obtaining measured quantities in harmonized way associated to the reference design cases (specified in the DoE matrix). Data are generated by:

- Experimental campaigns, including characterization, end point toxicity and eco-toxicity assessments and emission to different environmental compartment sampling.
- Modelling (through Virtual design of experiment).

The types of generated data are:

- NMs pchem attributes, such as nanoparticles size, morphology (aspect ratio), pH, Z-potential, etc.
- KPIs levels associated to the reference design cases.

NMs pchem attributes are obtained by characterisation of NMs (e.g. TEM, HRTEM, SEM, XRD, Raman, UV-VIS, resonance-enhanced surface second-harmonic generation -SHG, NMR, etc.).

KPIs referred to the addressed SSbD dimensions are obtained by:

- Experimental campaign:
  - tests on NMs functionality (e.g. antibacterial or antiviral activity, photocatalytic activity, oxidative stress reduction, photoluminescence etc.)
  - process yields (to assess process efficiency)
  - toxicity assays (e.g. cytotoxicity, oxidative stress, genotoxicity, etc.)
  - emission sampling (e.g. emission to atmosphere: particle mass concentration, OPC, emission to water compartment: leaching tests, etc.)

- Modelling:
  - Lung-deposited surface area (LDSA) for workers and human exposure.
  - Life Cycle Assessment-LCA (for environmental eco-toxicity and toxicity mid-point impacts), this assessment excludes nano-specific impacts and addresses the production processes of NMs and their related impacts.
  - Life Cycle Costing -LCC assessment (for economic impacts).
  - Social Life Cycle Assessment SLCA (for assessing positive and negative social impacts).

Experimental and modelling data are generated in accordance with the experimental and modelling plan referred to the DoE matrix.

- A suitable number of reference samples (corresponding to a specified sufficient amount of NMs or NFs incorporated in NEPs) associated to the DoE determined combinations of KDFs (see fig.6) is generated to enable the number of non-destructive and destructive characterization of p-Chem, functionality performance assessment, eco-toxicology and toxicology assessments, emission sampling campaign, emission to specific environmental compartment testing and assessment. (It should be noticed that this is part of the data generated to enable running the design case study).
- Life Cycle Assessment (LCA) and Life Cycle Costing (LCC) is performed to assess the environmental and economic performance associated to the generated reference samples. This is attained by running a Virtual Design of Experiments (V-DoEs). The defined implementation framework includes the possible data generation aimed at assessing social impacts.

Within the same data generation phase, data curation and compliance assessment with F.A.I.R. principles are performed.

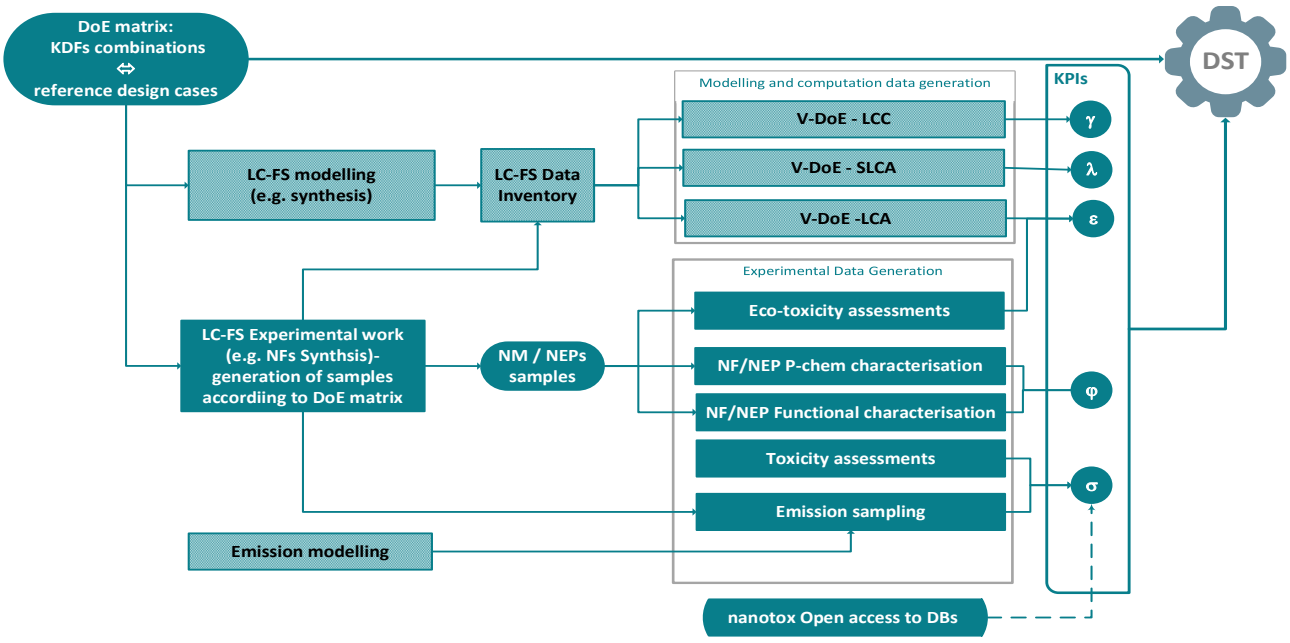


Figure 6 *Data generation workflow for providing input to the INTEGRANO decision support tool (DST)*

SSbD Ph3: SSbD impact assessment. Set of sorted SSbD cases data generation: KDFs and KPIs arrays.

This SSbD phase involves SSbD impact assessment obtained by digital computation through the INTEGRANO DST Decision Support Tool (DST). In this phase the following workflow steps are performed by the INTEGRANO DST:

1. The designer inputs all primary generated data to the DST obtained by the experimental characterizations, functional tests, emission sampling campaigns, modelling and computation referred to the KDFs and KPIs this can be done by:

a. Manual data input
b. Upload of standard template file

2. INTEGRANO DST maps Design space into the performance space by capturing second-order non-linear effects.

3. INTEGRANO DST kernel solves the Multi-Objective Optimization Problem (MOOP) via artificial intelligence algorithm of Multi-Criteria Decision Analysis (MCDA) by discarding all design options that do not comply with the simultaneous maximization (minimization) of all addressed KPIs.

4. INTEGRANO DST sorts out the multi-optimal design cases (MODCs): these are generated data identifying the SSbD cases.

5. DST provides quantitative performance of each MODCs in graphical form (Graphical data generated).

6. DST stays in prompt mode to get input from the decision maker.

SSbD Ph4: Interpretation of results and selection of SSbD case. Data of selected SSbD case(s).

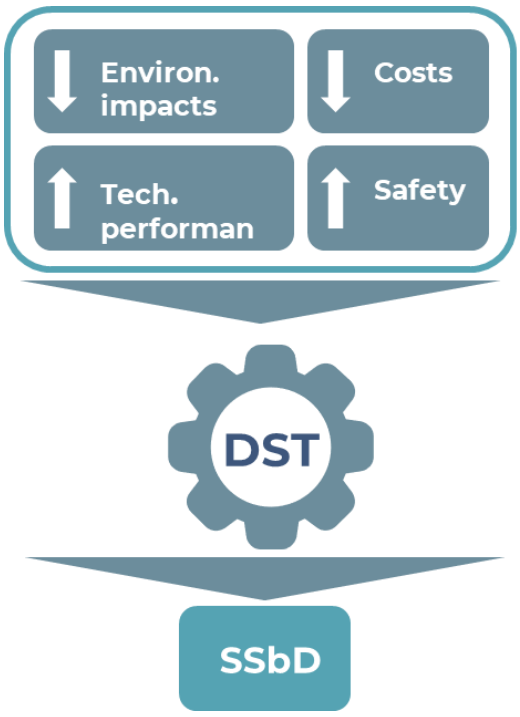


Figure 7 *Scheme representing target for maximisation or minimisation of the four dimensions-related KPIs, the subsequent SSbD impact assessment and the provision of the SSbD solutions simultaneously complying with functionality and safety maximisation, and environmental and economic impacts minimisation.*

Interpretation of results, possible protocol reiteration, or final decision on the SSbD selected case operated by the NM (NEP) designer (decision maker).

SSbD Ph4: interpretation and selection of SSbD case. In the last phase of the INTEGRANO DST implementation the human machine interaction is implied, with a focus on the human-centric approach. INTEGRANO DST enables the best user experience to facilitate the decision maker in the interpretation and final decision on the best SSbD case to be selected. Two final steps are implemented in the SSbD workflow:

1. The human decision maker (HDM) selects the most suitable combination of performance levels complying with SSbD specifications.
2. DST finds out the combination of KDFs values that identify the selected design case based on its performance profile, thus the KDFs values associated to the selected SSbD, this generates the array of selected KDFs combination values and the corresponding array of KPIs values.

In this last phase it is essential that the decision makers are aware of the KPIs levels when selecting the best KPIs level combination: this allows informed decisions supported by quantitative evidence, which then could be motivated also by commenting the selection, by generating data in text form.

2. Data Summary

Q1-Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

INTEGRANO aims to use data collected/generated through the project, as well as high-quality peer-reviewed published data if required.

INTEGRANO will contribute to the ongoing effort in data integration through incorporation of data from outside the project within the INTEGRANO database. INTEGRANO will implement a data mining process to search reference literature and open access databases for quantitative data and for the availability of NM samples. For example, engineered nanomaterial (ENM) data generated within previous EC H2020 projects already providing developed F.A.I.R. datasets (e.g. NANoREG, NanoMILE, NanoReg2, ASINA, PROTECT, REINVENT, Gov4Nano, NANORIGO, Gov4Nano, NanoInformaTIX, NanoSolveIT) and ongoing EC projects (e.g. BIOMAT, BIONANOPOLYS ORIENTING); as well as EC HORIZON EUROPE projects (TOXBOX, CheMatSustain, CHIASMA, SSbD4CheM, BEST-COST, MARCHES, UBDPOLICY, M ISTRAL, VALESOR, ANALYST, INSIGHT, SUNRISE, POTENTIAL, nanoPASS, MACRAME, DigiPass, iCare, IRISS). Links with "sister projects" and projects funded under HORIZON-CL4-2023-RESILIENCE-01-21, CE-NMBP-42-2020, HORIZON-HLTH-2022-ENVHLTH-04-01 and HORIZON-CL4-2023-RESILIENCE-01-39 are currently being explored and established. It is envisaged that INTEGRANO will utilise data from some of these projects and that will contribute to disseminate best practices on data generation, on management and elaboration towards attaining SSbD solutions.

INTEGRANO will also obtain secondary data from open access NMs databases (e.g. caNanoLab, eNanoMapper, NR, NBIK, NKB, NIL, Nanowerk, and PubVINAS) for p-chem characterisation data, and toxicity data, Eurostat (for regionalised economic data), NanoFASE and SimpleBox4nano for environmental fate and exposure predictions, as well as licenced databases (Ecoinvent v.3.9.1 Db for LCA, Psilca v.4 DBs for LCA and SLCA).

Data will be obtained from licenced data bases and open access databases via data mining process to provide INTEGRANO with a broader set of data related to NMs (features and properties). Results will be associated with NMs exhibiting common features (e.g. environmental interactions, toxicity effects), and data mining will be utilised to support the identification of suitable NMs reference impact categories (ICs) based on NMs-specific effect factors. Data from licenced databases will be used in accordance with the licence associated with the access and use of the data – licenced data is envisaged to be utilised within INTEGRANO to perform LCA and SLCA analyses.

Q2-What types and formats of data will the project generate or re-use?

There are distinct types of data which will be generated and utilized within INTEGRANO:

1. (primary) Raw or experimental data derived from inventories associated to the functional systems and from experimental direct and indirect measurements through instruments (physico-chemical characterization of nanoparticles and nano-enabled products, functionality indicators, end-point eco-toxicity and human toxicity assessment, emission and exposure assessment).
2. Computed data derived from in silico or computational modelling (life cycle assessment, risk assessment, environmental, social, economic impact assessments, exposure modelling).

3. Curated (secondary) data associated with formal publications (literature curated data) derived from journals such as tables, as PDFs, spreadsheets, etc., or data derived from databases.
4. Interactive data, whereby the data is put in the context of other data.

Data of types 1 and 2 will be collected in a harmonised and database-compatible manner via the development and application of the bespoke INTEGRANO repository, to provide INTEGRANO with coded data records which will associate NMs samples, related synthesis and incorporation processes, p-chem and functional characterisation data, emission sampling and safety assessment data. The repository will integrate data management with data generation from the outset of the project to ensure that raw scientific data which is collected is transformed into transparent and interoperable datasets which are findable, accessible, and reusable. This approach contrasts with data management being a secondary consideration e.g. applied only at the point of publication where datasets are required to be deposited in appropriate databases.

Formats of Data: Data produced and exchanged within INTEGRANO is expected to be predominantly CSV format, which is a widely supported format, commonly used and machine-readable. Other formats such as images (JPEG, PNG) and videos (MP4) will be used. The size of individual digitized data files is not expected to be big (<100Mb, except for recordings of meetings) and spreadsheets will be used for data sharing among the partners and those data will be integrated and stored in the internal INTEGRANO repository individually and integrated in MySQL (Server Version: 5.7.42-0ubuntu0.18.04.1 - (Ubuntu), Protocol Version: 10).

Q3-What is the purpose of the data generation or re-use and its relation to the objectives of the project?

The INTEGRANO project addresses shortcomings and gaps in current safe and sustainability assessment approaches of Nano Materials (NMs) and Nano-enable products (NEPs). The project proposes a general assessment approach based on quantitative evidence to be applied in practice for specific Nano Materials (NMs) design cases - inorganic, organic and carbon NMs. The development of NMs novel impact categories (ICs) for nanotoxicity and eco-nano-toxicity assessment will enable the integrated application of standardised assessment methodologies. The synthesis, incorporation, use phase and end-of-life NMs life cycle stages (LCS) will be addressed. The collected/generated data will be used to meet the project's objectives to formulate design hypotheses and make design decisions by applying a data-driven approach and methodology, providing stakeholders with a digital supported decision process to tackle the Safe and Sustainable by Design (SSbD) challenge in the NMs context. Therefore, at each LCS INTEGRANO will address:

• Definition of: (a) goal and scope, (b) Key Decision Factors (KDFs) and (c) Key Performances Indicators (KPIs), (d) Design of Experiment (DoE) matrix.
• Measurement as part of the experimental data generation, p-chem and functional characterisation, emissions sampling, toxicity and eco-toxicity. assays, process primary data collection for LCA, LCC and SLCA data inventory,
• Analyse experimental data, process modelling and virtualisation (digital twin), computational data generation by Life Cycle Impact Assessment associated to LCA, LCC, SLCA.
• Design by assessing and evaluating multiple performances supported by digital Decision Support Tool (DST) to sort the set of multi-optimal design cases.
• Verify by selecting the best SSbD case among the set of multi-optimal design cases and submit it to experimental validation.

Q4-What is the expected size of the data that you intend to generate or re-use?

The data generated/collected through the project will consist of raw, analytical and metadata to support the project's actions and expected not to exceed 10 Tb.

Q5-What is the origin/provenance of the data, either generated or re-used?

Most data in INTEGRANO will be generated internally from project partners. For the external data sources (listed earlier) INTEGRANO will take into account the original licences when integrating within the INTEGRANO database.

Q6-To whom might your data be useful ('data utility'), outside your project?

The project aims to make the data accessible and beneficial to all stakeholders within the nanosafety community.
Potential beneficiaries of the data include:
• Academics, at all levels, working in all fields of nanosafety research and the wider toxicity community.
• International standardization bodies, regulatory agencies, EC agencies (e.g. ECHA, EMA, EFSA, ISO, JRC) and policy makers.
• National authorities (National Institutes of Health) and public agencies for health and safety research (European Agency for Safety and Health at Work (EU-OSHA).
• Sustainability impact assessment associations (Life Cycle Initiative, Global Compact, Rete Italiana LCA, EU Environment Agency (EEA), European Platform on LCA, EPD international). • Relevant industrial sectors and associations (e.g. European Chemical Industry Council (Cefic), Cosmetic Europe, European Apparel and Textile Confederation (EURATEX), Water Europe).
• SMEs that do not have the resources or the knowledge to develop and use in-house tools for SSbD approaches and risk assessment requirements.
• Consumers. As end users, consumers will have greater security surrounding the innovation process for the safe and sustainable use of NMs, resulting in safer and sustainable NEPs.

3. F.A.I.R. data

F.A.I.R., which stands for Findable, Accessible, Interoperable, and Reusable, are a set of principles that, if met, maximises the value of data – increasing discoverability and maximising research exposure and impact. Data are grouped into the four categories given earlier. F.A.I.R. attributes are adapted to each category. For example, raw data may not be findable to people outside INTEGRANO until publications are completed, and interoperability mostly applies to metadata. Data associated with formal publications must be publicly available and interoperable at a very high level to benefit the community. Similarly, when computational approaches are considered, F.A.I.R. raw data is essential for any beneficial scientific effects.

3.1 Making data findable, including provisions for metadata

Q7-Will data be identified by a persistent identifier?

Data will be identified by a persistent identifier – data will be saved within the INTEGRANO repository with a unique key.

For data to be findable, these principles will be followed:
• (Meta)data are assigned a globally unique and persistent identifier.
• Data are described with rich metadata.
• Metadata clearly and explicitly include the identifier of the data it describes.
• (Meta)data are registered or indexed in a searchable resource.

To meet these principles the project will use the centralised INTEGRANO repository to store the data generated/collected by the project. The repository will also include an appropriate licensing system to allow a layered access to the available data based on the accessibility decided by the data owners. Datasets will be complemented with the appropriate metadata and unique repository and/or DOI identifiers to allow users to easily query and reference.

Q8-Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Rich metadata will be created to allow discovery. Descriptive (e.g. provenance) and administrative (e.g. type, permissions, key dates) metadata will be created and will describe the process used to create the data. The metadata produced within INTEGRANO will largely be associated with data concerning physicochemical characterization, toxicological assays, life cycle assessment and monitoring exposure metadata will be created, alongside metadata covering data provenance (see Data Summary for more details). Metadata will follow standards set by internationally recognised databases e.g. JRC Publications Repository – the characterisation techniques to be utilised within INTEGRANO are mature and metadata will follow standards common to each community, as well as being internally consistent.

In particular, data will be associated to a unique (exclusive) identifier code, which will enable tracking data origin (time, place, data creator, data owner), as well as to explicit associations with other data. In such a way it will be easy to associate primary generated data to the specific associated case study, to the addressed life cycle stage analysed, and to previous or subsequent related data referred to the other life cycle stages. Moreover, data generated through subsequent elaboration and computations will be automatically associated to the dataset of primary data generated within the project without ambiguity and possibility of making any mistake in SSbD analyses outcomes.

Important to say that the availability of data unique identifier code in association with related metadata will leverage each quantitative datum by clarifying its full information content, including precision, accuracy, context in which of datum was generated (e.g. measurement instrument and measurement, protocol, specific experimental controlled conditions, reference to benchmark, baseline, and more).

Q9-Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Search keywords will be provided in every digital representation of the data; the repository will be readily searchable using strings. Data interconnection will facilitate discovery and re-use.

Q10-Will metadata be offered in such a way that it can be harvested and indexed?

Search fields will be offered within the repository to enable harvesting and indexing. Data within the repository will be a mix of primary and elaborated data – some data may not be made available due to licensing where this goes against the beneficiary's legitimate interests, including regarding commercial exploitation, as set out within the grant agreement, or due to explicit data (external) origin ownership (see also section Q8).

3.2 Making data accessible

This principle prescribes that data must be accompanied with metadata that explains how people get access to the data. This does not imply they always get access. To be accessible, the principles specify:

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

Q11-Repository:

Q11.a-Will the data be deposited in a trusted repository?

Yes. Data access will be facilitated by storing all data in the project's centrally managed datastore, i.e. the internal INTEGRANO server for raw data (hosted by Aruba cloud web service according to current GDPR and security standards, as well as data safety thanks to the compliance to disaster recovery protocols implying data redundance and mirroring, also enabling real time data accessibility in case of severe environmental events or hackers attacks. This approach mirrors and extends that of the ASINA [4] project, in which elaborated SSbD data were also stored using the Aruba cloud service offered by Project HUB360. Raw data and metadata will be stored within the data repository in formats that are Accessible and understandable by humans and machines. In parallel, the formats will be harmonised to make them Interoperable, and Reusable through a straightforward tiered licensing system. Web based public data will be available with secured access.

Q11.b-Have you explored appropriate arrangements with the identified repository where your data will be deposited?

The INTEGRANO repository is managed by partner PRJ as sole partner supporting service costs. The INTEGRANO repository is rendered available to the whole consortium free of charge and it is managed within the consortium, as described above.

Q11.c-Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Each record within the repository will be assigned a unique identifier. Data storage protocol is incremental thus avoiding any chance of data overwriting or deleting data.

Q12-Data:

Q12.a-Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

In summary, data related to the identified NMs physiochemical parameters and methods for measurement, human and eco-toxicity assessments, and NMs detection campaigns in real and simulated environments will be openly available. Data regarding the relationships among p-chem properties and performances attributes (techno-economic, hazard, exposure, LCA impacts) are confidential as these data are sensitive to commercial competition. The dissemination classifications are included in the grant agreement (GA).
Secondary data retrieved by licenced databases will be kept confidential as due by licencing conditions issued by the data-owner, nevertheless elaborated data deriving partially or entirely by secondary licenced data may be rendered open in compliance with permissions provided by data owners within the consortium and in accordance with their exploitation plans and ownership claims.

Q12.b-If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Most data will be available for re-use during the project lifetime and beyond. Regarding other sensitive data, a minimum of four years after the end of the project is foreseen to protect the partner intellectual property rights and interests, including ensuring future patent applications. In case of patent filing, the embargo duration will at least correspond to the time needed for publication (typically 18 months for patents granted by the European Patent Office).

Q12.c-Will the data be accessible through a free and standardized access protocol?

The data generated/collected/analysed through INTEGRANO will be provided with appropriate metadata and where possible will be made publicly available at zero cost (to the user) through a tiered make-accessible process, in collaboration with the data owners and their exploitation needs. In many cases, a data embargo may need to be imposed. Some data, particularly elaborated data, will be exploited and databases will be licenced.

Within the project timeframe data will be accessible free of charge to partners, the specified level of data confidentiality will restrict their use outside the consortium according to the relative clauses reported in the Grant Agreement.

Q12.d-If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

Data published in the project time frame will be findable and easily accessed by third parties. Where restrictions on data use exist, access to data will be at the discretion of the beneficiary until restrictions are lifted by the beneficiary or data is released as according to the grant agreement. INTEGRANO has committed to preserve data for four years after the lifetime of the project, where it will be hosted within the INTEGRANO repository, meaning there is a direct link between the data repository and data owners through which access requests can be made after the lifetime of the project.
In the longer term, it is envisaged that, with agreement of all partners, data could be transferred to third party secure file transfer facilities, such as the NanoCommons knowledge warehouse [5], Figshare [6] and OpenAIRE [7], to enable longer-term data sharing.

Q12.e-How will the identity of the person accessing the data be ascertained?

Any uploaded data will be linked to a user registered in the platform with specific access rights. (See also section Q2, above).

Q12.f-Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

Access rules have been defined and formalised within the INTEGRANO proposal and GA. Should they arise, unforeseen questions will be handled by the project management board. The application of access rules to specific data generated will be done by the relevant data beneficiaries and the Data manager.

Q13-Metadata:

Q13.a-Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

Metadata will be stored in the INTEGRANO data repository in formats that are Accessible and understandable by humans and machines. In parallel, the formats will be harmonised to make them Interoperable and Reusable through a straightforward tiered licensing system. It is envisaged that metadata will be made openly available, as per the grant agreement, with the potential for exceptions where the release of metadata would go against the beneficiary's legitimate interests, including regarding commercial exploitation, as set out within the grant agreement.

Q13.b-How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

The INTEGRANO repository will remain active for the project duration and for four years after the end of the project. However, in the longer-term it is anticipated that INTEGRANO datasets will be curated for and stored in certified repositories not dependent on the project funds for long term preservation, e.g. NanoCommons knowledge warehouse, Figshare and OpenAIRE. In the vast majority of cases, metadata will co-exist alongside the associated data and preservation of metadata will continue post-INTEGRANO through transfer to external certified repositories.

Q13.c-Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open-source code)?

All resources are accessible through HTML standards. It is envisaged that all software developed during INTEGRANO will be protected as this will be commercially exploited. Therefore, open-source code for software will not be made available. The software may be licenced to end-users. Reference documentation such as user manuals may be rendered available.

3.3 Making data interoperable

Q14-What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

The INTEGRANO goal is to make the data generated/collected during the project's lifecycle as interoperable as possible and to promote dataset combination. This means that the acquired data

will be captured using a harmonised approach. It is expected that most data will be in the CSV format, which is a widely supported format, commonly used and machine-readable. As the INTEGRANO consortium owns the database, data can be shared and exported in virtually any format required by interested parties.
An important aspect of data harmonisation is the use of a vocabulary (i.e. ontology) that will employ common agreed definitions for the terms used by all aspects of nanosafety research and will allow both qualitative and quantitative data combination and reusability. As a result, the mid to long-term plan of INTEGRANO is to use the single nano-wide NanoCommons ontology implementing already existing and established ontologies (e.g. OBI, ChEBI, PATO, UO, NCIT, EFO, OAE, eTOX, eNanoMapper, MPATH, etc). This common vocabulary will be used for both the data and metadata curation irrespective of the file format the data is being stored.

Q15-In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Yes, mapping is foreseen, if necessary. The format will be self-explanatory and will guide the user through more common ontologies.

Q16-Will your data include qualified references1 to other data (e.g. other data from your project, or datasets from previous research)?

Yes, qualified references are envisaged, linking to available open-source data, or data which other partners have access to, to enrich the context surrounding the data produced within INTEGRANO.

3.4. Increase data re-use

Q17-How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

The INTEGRANO data flow is declared openly and materials will be made available within the INTEGRANO repository to provide sufficient context to the freely accessible data hosted there. Analyses will be conducted in compliance with existing standards for which documentation is readily available – namely, ISO standards 14040-14044 for LCA will be applied. Protocols and standards for all standardised measurements and characterisations as well as toxicity and ecotoxicity assessments will be followed. Metadata will report such standards which will allow verifying data compliance to the related ones. (see also section Q8 above)

Q18-Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Aside from sensitive and confidential data as described within the GA and within the exploitation plan (to be generated along the project development), most INTEGRANO data will be freely available for re-use during and beyond the project lifetime. Data will be made publicly available through a tiered make-accessible process in collaboration with the data owners and their exploitation needs.

Q19-Will the data produced in the project be useable by third parties, in particular after the end of the project?

The INTEGRANO repository will remain active for the project duration and for four years after the end of the project. However, in the longer-term it is anticipated that INTEGRANO datasets will be curated for and stored in certified repositories not dependent on the project funds for long term preservation, e.g. NanoCommons knowledge warehouse, Figshare and OpenAIRE.

Q20-Will the provenance of the data be thoroughly documented using the appropriate standards?

(Meta)data will be richly described with a range of accurate and relevant attributes, including detailed data provenance, covering origin databases, p-chem characterisation equipment, operators, protocols utilised. New methods will also be described in detail.

Q21-Describe all relevant data quality assurance processes.

Reviews and moderation of data will be performed by different committees before publication. The INTEGRANO external advisory board will also be invited to review and advise upon the INTEGRANO workflow, and data acquisition, handling and processing methodologies. The checks envisaged include the evaluation of the experimental protocols used or to be used, reproducibility, the tools source code and abilities and the presence of potential gaps in datasets. In addition, INTEGRANO will adopt the practice of inter-consortium open review for any content to be published in open access using the dedicated INTEGRANO dissemination channels. The open review process will assign on a voluntary basis the referees from the consortium or make all partners be able to participate in the review.

Q22-Further to the F.A.I.R. principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

Further INTEGRANO outputs envisaged include: peer-reviewed publications, the MultiOptimal software demonstrator, samples of nanomaterials and patents. See the discussion in section 4 for further detail.

4. Other research outputs

Q23-In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.). Beneficiaries should consider which of the questions pertaining to F.A.I.R. data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the F.A.I.R. principles.

INTEGRANO expects to result in the publication of at least 20 open access publications. The application of F.A.I.R. principles to scientific and technical data associated with these has been addressed in the DoA, where publications themselves will be open access and associated data will be made publicly available within appropriate repositories. The European Open Science Cloud (EOSC) and Open Research Europe (ORE) will represent crucial platforms to allow easier replicability

of INTEGRANO results providing findability, interoperability and open access to research outputs. Open Access Infrastructure for Research in Europe (OpenAIRE), the Registry of Open Access Repositories (ROAR) and the Directory of Open Access Repositories (OpenDOAR) will be used for storing scientific and technical data. The storing of publications and associated data in these repositories ensures the output is furnished with a persistent identifier and associated with rich metadata, ensures the outputs are accessible, will be interoperable through adoption of community-accepted ontologies and vocabularies, and will ensure data re-use through standard reuse licences.

Samples of NMs and NEPs will be produced within the project. Unless it goes against the beneficiary's legitimate interests, including regarding commercial exploitation, these samples will be made openly available to the scientific community e.g. via the JRC Nanomaterials Repository.

Software and algorithms as part of the partners background knowledge as well as Software and algorithms produced within INTEGRANO are protected – access will be granted via licensing. Access to the DST prototype will be given to industrial partners within INTEGRANO for a limited time frame aiming at evaluation of user experience and testing in the design of future products.

Trademarks and patents will be handled by the relevant beneficiary according to well-established procedures. Each relevant partner is experienced in dealing with IP issues.

5. Allocation of resources

Q24-What will the costs be for making data or other research outputs F.A.I.R. in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.)?

The costs are the ones allocated at Tasks 1.1 and 1.3 of WP1, dedicated to Data management, the creation of the DMP and the data curation process.

Q25-How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions).

The related costs are to be covered as part of the EU Grant for the project.

Q26-Who will be responsible for data management in your project?

CNR (ISSMC unit) is responsible for data management, assisted by PRJ, the leader of WP5.

Q27-How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

The data produced as part of the project will be stored in the INTEGRANO database, and, in the longer term, it is envisaged that data will be migrated to external repositories, subject to the approval of INTEGRANO partners. The sustainability and accessibility of the data have been discussed and defined in the GA.

The project management board is in charge of arrangements related to the long-term preservation of these data. The long-term data storage costs will be discussed in the context of project resources. Free to low-cost data storage options are available e.g. Figshare, and these will be considered as long-term data repository options.

6. Data security

Q28-What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

SSL Encryption of all web content is applied by specific certificates. As defined as an initial constraint at the beginning of the INTEGRANO project, the data management platform does not use any system or digital service hosted outside Europe for all its administrative and platform management. Initially, the data generated by the INTEGRANO project will only be accessible to project partners. Technically, the data access will be restricted via role-based security and standard authentication and authorization protocols.

Q29-Will the data be safely stored in trusted repositories for long term preservation and curation?

Implementation has already begun with the INTEGRANO server having all necessary security protocols and tools (anti-hacking and malware plugins) added to prevent any malicious attacks.
The management team and the INTEGRANO Consortium will ensure that all partners and users will follow strict ethical guidelines covering all aspects of the project's infrastructure.
Public data sharing and transfer among persons or partners will be mediated through the INTEGRANO DMT, or, where appropriate, through third party secure file transfer facilities, such as the NanoCommons knowledge warehouse, Figshare and OpenAIRE. In the longer-term it is anticipated that INTEGRANO datasets will be curated for and stored in certified repositories not dependent on the project funds for long term preservation.

7. Ethics

Q30-Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

The ethics self-assessment is detailed within the DoA and no ethical issues arise from the project overall, or the data management aspects, that would impact on data sharing. However, a strong effort will be made to ensure that ethical information regarding the datasets integrated will have as a minimum a statement regarding the ethical approvals in place at the time the data were generated. Where data are generated, these will be generated within the ethical framework of INTEGRANO, and as above will have a clear declaration regarding the ethical approvals associated with any underlying datasets. Thus, we do not anticipate any ethical or legal issues relating to the datasets generated.

Q31-Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?

Not applicable to INTEGRANO.

8. Other issues

Q32-Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

No.

9. References

[1] https://www.openaire.eu/images/Guides/HORIZON_EUROPE_Data-Management-Plan-Template.pdf

[2] Wilkinson, M. D. et. al., The FAIR Guiding Principles for scientific data management and stewardship, Sci Data, 2016, 3, 160018.

[3] Caldeira, C., Farcal, R., Garmendia Aguirre, I., Mancini, L., Tosches, D., Amelio, A., Rasmussen, K., Rauscher, H., Riego Sintes, J. and Sala, S., Safe and sustainable by design chemicals and materials - Framework for the definition of criteria and evaluation procedure for chemicals and materials, EUR 31100 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53264-4, doi:10.2760/487955, JRC128591.

[4] ASINA, Anticipating Safety Issues at the Design Stage of NAno Product Development, European Union's Horizon 2020 research and innovation programme - grant agreement No 862444; (web reference: https://www.asina-project.eu/)

[5] https://ssl.biomax.de/nanocommons/cgi/login_bioxm_portal.cgi

[6] https://figshare.com/

[7] https://www.openaire.eu/