ACEnano

Analytical and Characterisation Excellence

Analytical and Characterisation Excellence in nanomaterial risk assessment: A tiered approach

Grant Agreement No 720952

Deliverable Report 4.1

Deliverable:	D4.1 Design Concept for the Knowledge Warehouse
Work Package:	WP4: Linking ontology and methodology
Delivery date:	M6 - 30 June 2017
Lead Beneficiary:	Douglas Connect (DC)
Nature of Deliverable:	Demonstrator
Dissemination Level:	Confidential (CO)

Submitted by:	Thomas Exner and Lucian Farcal (DC)
Revised by:	Iseult Lynch (UoB), Geert Cornelis (SLU), Barry Hardy (DC)
Approved by:	Eva Valsami-Jones (UoB)

Table of contents

List of Abbreviations 3

1. Introduction 4

2. FAIR Principles 6

3. Knowledge infrastructure concept 8

3.1. Data Provider Layer 9

3.1.1. Protocols database 9

3.1.2. Data warehouse 13

3.1.3. Data templates 15

3.2. Data User Layer 17

3.2.1. Analysis, modelling and visualisation tools 17

3.2.2. Knowledge sharing tools 17

4. Conclusions 18

References 19

Annex I - Case study on NMs size characterisation data by DLS 20

List of Abbreviations

AOP - Adverse Outcome Pathway

API - Application Programming Interface

DLS - Dynamic Light Scattering

FAIR - Findable, Accessible, Interoperable, and Re-usable

KI - Knowledge Infrastructure

KW - Knowledge warehouse

NM - Nanomaterial

NSC - NanoSafety Cluster

OECD - Organisation for Economic Cooperation and Development

REST- Representational state transfer

WP - Workpackage

1. Introduction

ACEnano will enhance confidence, adaptability and clarity into nanomaterial (NM) risk assessment by developing a widely implementable and robust tiered approach to NMs physicochemical characterisation that will simplify and facilitate contextual (hazard or exposure) description, provide guidelines in terms of quality assessment of a particular study, and its transcription into a reliable NMs grouping framework. More specifically, ACEnano will produce enhanced hardware and software to improve, automate and speed up NM characterisation. The knowledge infrastructure, whose design concepts are outlined here, will provide a central place to showcase these improvements by providing access to state-of-the-art protocols for the developed and optimized test methods, including quality control guidelines and applicability domain considerations for the type, as well as properties-like size range for the NM, which can be studied by the method. These protocols will directly be linked to corresponding datasets of the highest quality achievable by the method at a current state provided by the partners in a format comparable to existing datasets to evaluate the enhancements over time and to be able to perform a cost/benefit analysis for a specific characterization problem at hand. Additionally, the results from the round robins performed in the ACEnano project will be documented with access to the datasets produced by the participating groups to demonstrate the interlaboratory reproducibility of the methods. This will lead to a larger confidence in the new methods and faster adoption of these by all stakeholders including academic and industry researchers and risk assessors and will lead to better validated methods towards the final goal of regulatory acceptance. The knowledge warehouse and infrastructure will be made openly accessible not only for searching, browsing, downloading and reusing the protocols and data, but also for uploading variants of protocols to enlarge the applicability domain of a method and datasets produced according to ACEnano protocols by third parties. In this way, the ACEnano knowledge warehouse will become a central resource of high quality NM characterization, which can then be used for the in-depth assessment of NMs for safety, including investigations of hazard as well as exposure, distribution and fate of the NM in biological systems. An additional goal of this central repository is to provide the necessary amount of data to allow for the first time to generate models for NM properties based on the conditions these were produced under, which will allow for better predictions of the characteristics of new NM and safe-by-design considerations.

The data made publicly available in the ACEnano knowledge warehouse (KW) will be harmonised and made interoperable with other related data sources for NM risk assessment. Data management will comply to the FAIR data concepts (Findable, Accessible, Interoperable, and Re-usable). Additionally, ACEnano aims to contribute to creating a harmonised NM ontology by covering the specific requirements of methodological advances in NM characterisation, improved analytical setups, and refined physico-chemical characteristics.

Workpackage (WP) 4 in ACEnano coordinates these efforts by creating a robust data management system, which will deliver a centralised management of all ACEnano data, and harmonise other major existing data management systems and resources, towards:

- streamlining and facilitating data collection/management;
- providing a quantitative quality assessment tool of NM characterisation data that can be usable in e.g. risk assessment, quality control and the like;
- consolidating a common language within the nanosafety community and supporting/integrating existing platforms for communication.

More specifically, Task 4.3 of WP4 includes the work to be performed on the ACEnano Knowledge Warehouse, which will store**:**

- NM Characterisation Protocols;
- Raw data and processed data generated within the project;
- Other related resources, support tools and processes for preparing and inputting datasets and templates.

These activities are based on the concepts developed in eNanoMapper [1], ToxBank [2] and more recently in OpenRiskNet [3], but will also take the specific requirements of the ACEnano project into account. These requirements are discussed within the consortium to identify the ACEnano needs, but also from similar requirement analyses for other projects via surveys and interviews.

Since the goal of the ACEnano project is the development of new, improved instrumentation, special attention will be put on the documentation of the associated scientific evolution, benchmarking and validation of methods and data:

- Procedures and file formats will change over time and might eventually lead to inconsistencies, when data from different development stages are combined for analysis or modelling, if the differences are not accounted for appropriately.
- Unique identifiers will be used for specific development stages of the used instruments and descriptions of the data and metadata file formats will be linked to the database entries.
- In this way, comparison of data produced at different time points and/or coming from different partners from inside and outside ACEnano, now and in the future, will be facilitated. This approach will also enable development of specific troubleshooting guides for each method, and for compilation of quality checks for datasets acquired using each method for use by regulators in determining data quality for decision making purposes.

The implemented data structure is aligned to scientific research and computing standards from other ongoing or previous activities (eNanoMapper, OpenRiskNet, OpenTox, NANoReg, NanoMILE, NanoFASE and other NanoSafety Cluster (NSC) data generation and management activities) and considers additional requirements for regulatory reporting (e.g. OECD harmonised templates) and Adverse Outcome Pathway (AOP) development. In this way, the warehouse will facilitate the data transfer to and from other databases. The standards guarantee interoperability with algorithm and modelling components from inside and outside the project.

2. FAIR Principles

To pursue optimal data usage, the implementation of the KW will be assisted by FAIR data guidelines. These guidelines consider the establishment of FAIR principles: Findability, Accessibility, Interoperability and Reusability of data and the algorithms, tools and workflows that operate on it [4],[5]**:**

To be Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

F2. data are described with rich metadata.

F3. (meta)data are registered or indexed in a searchable resource.

F4. metadata specify the data identifier.

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 Re-usable:

R1. meta(data) have 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 their provenance.

R1.3. (meta)data meet domain-relevant community standards.

Besides integrating these more technical concepts and guidelines regarding the KW design and standard for data representation, ACEnano is also committed to best practices of data management, which have proven sensible in projects like EU-ToxRisk [6], and will further develop them especially for physico-chemical characterization of NMs for safety and risk assessment. It is clear that data quality is mainly influenced by procedures adopted during data collection and data treatment, and documenting in detail how data are collected and treated (e.g. mathematically) provides evidence of such quality. The ACEnano protocol database (described below) will be an integral part of the KI and will support the structured documentation of the experimental but also computational procedures. Measures to guarantee quality, which will be documented with the protocols as well as the

metadata provided with each dataset, include:

- calibration of instruments to check the precision, bias and/or scale of measurement;
- taking multiple measurements, observations or samples;
- Showing a clearly recorded data treatment procedure that has been verified in the peer-reviewed literature;
- checking the truth of the record with an expert;
- using standardized methods and protocols for capturing observations, alongside data recording forms with clear instructions.

However, also the digitization and entering of data into the database needs to follow high quality standards. Errors during input and unintentional modification or disruption can be avoided or at least minimized by standardized and consistent procedures with clear instructions. These should include:

- setting up validation rules or input masks in the data entry software;
- using controlled vocabularies, code lists and choice lists to minimize manual data entry;
- detailed labelling of variable and record names to avoid confusion;
- designing a purpose-built database structure to organize metadata and data files;
- having clear guidance as to what constitutes raw and processed data, and procedures for when outliers should be included / excluded from raw and processed datasets;
- and especially working together with the hardware developers in the extraction of important data and metadata directly from the output produced by the experimental equipment in the format expected by the database removing the need of manual copy/pasting of data.

Integrity of data is maintained by restricting permission to modify data to specific persons, controlling the entry of specific fields like dates and information of the operator by automatic procedures (authentication) and documenting every modification and annotation of data. Even if in a research project like ACEnano the enforcement of all rules and strict schemata used in regulatory settings is probably counter-productive and will slow down discoveries, knowing the building blocks of good documentation practice and how to manage the risks with respect to accuracy, completeness, consistency and reliability of data throughout the entire period of its usefulness will still be helpful to establish a culture of responsible data management.

3. Knowledge infrastructure concept

The development activities to support data collection, management, interpretation and delivery to a data warehouse for safe use & storage will focus on the following components:

- Development of a KW to store all raw data, protocols and processed data generated by all ACEnano partners;
- Development of a quality assessment of NM characterisation data;
- Development of a NMs' ontology extending the current eNanoMapper ontology [7];
- Linking to NMs/nanosafety modelling projects by standardised data and protocols.

Therefore, the following components are envisaged for development within ACEnano (see also Figure 1):

- NM characterisation protocols and their structured documentation in the ACEnano protocol database, including links to the decision support tool developed in WP3 to guide users through the selection of the most appropriate method to answer specific characterisation questions;
- Data warehouse for handling of metadata and the experimental data for:
    - Raw data
    - Processed data
    - Computationally generated data (e.g. NM descriptors from crystal structure or quantum mechanical calculations).

  This will also include:
    - Data input tools
    - Data and metadata schema and templates
    - Direct linking of the data to the protocols in the protocol database describing its generation, as well as linking to publications describing / utilising the datasets / protocols as they become available.
- Integration of external data sources and harmonization with data from ACEnano;
- Preparation of data for their direct usage in analysis, modelling and visualisation tools;
- Knowledge sharing platform to present and discuss protocols under development, temporary results and outcomes of the round-robins of interlaboratory comparison;
- Other support tools (e.g. NM and biological system identifier converters, gathering of working documents, repository of templates, etc.).

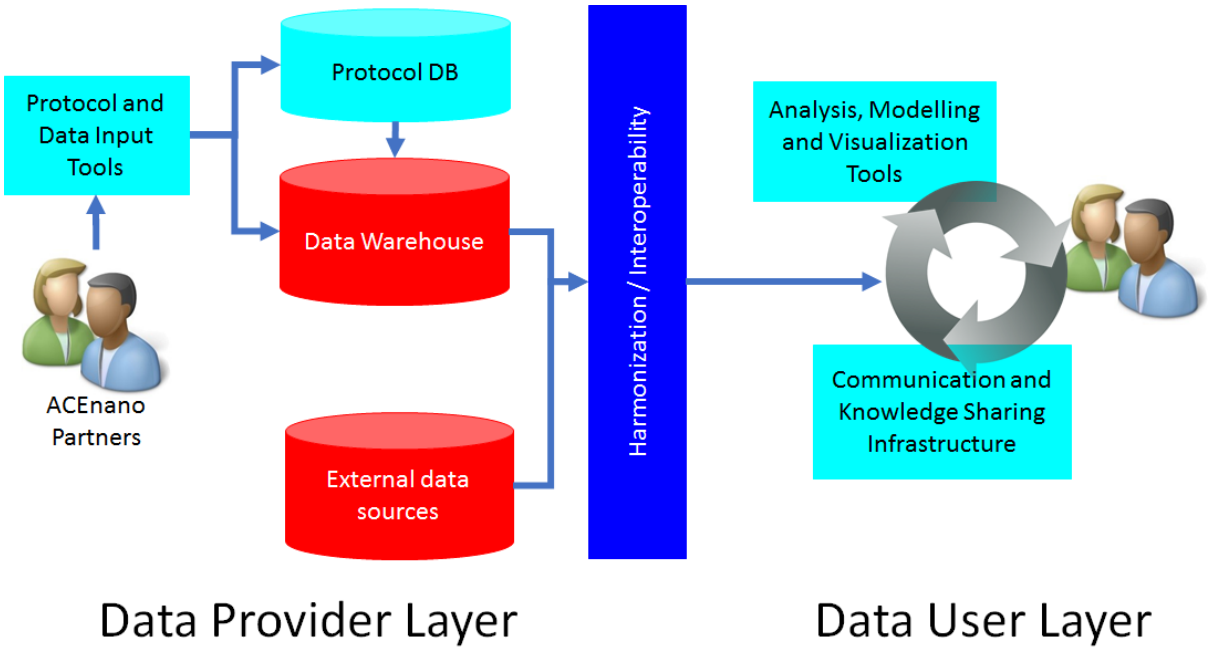

Figure 1. Schematic of the ACEnano knowledge infrastructure

3.1. Data Provider Layer

3.1.1. Protocols database

The aim of the protocol database is to facilitate access to and sharing of NM characterisation protocols first within the ACEnano program and later with the community, making them easy to browse and link to experimental datasets (the main functionalities and components of protocols database are shown in Figure 2). NM characterisation protocols from the ACEnano partners together with existing resources (publicly available) from the nanosafety community (e.g. NANOMMUNE [8], DaNa 2.0 [9]) and from NanoMILE and NanoREG will be used to start building the protocols inventory. For ACEnano and other new protocols, a structured interface will guide the protocol input. By asking the user to fill in information for the specific fields listed below, a more harmonized documentation will be achieved and quality standards can be enforced e.g. by making detailed descriptions of calibration procedures and negative and positive controls as well as of data processing required fields in the interface. The interface will also strongly encourage the stepwise description of the experimental procedure as spearheaded by protocols.io [https://www.protocols.io/] optionally supported by video protocols similar to JoVE [https://www.jove.com/]. Protocols will be categorized according to their level of development (experimental, under development, finalized, internally and externally validated) as well as accompanied by a discussion of the applicability domain based on the successful and unsuccessful usage of the protocol with respect to the type of NM and its characteristics (e.g. the accuracy of the method for a specific size range). These quality measures will be provided, on one hand, by the protocol provider as the expert knowing the

development status of the method as well as how it compares to other methods for the same endpoint and, on the other hand, by a panel of experts during the selection of methods for transfer to WP2 and WP3 for optimisation and benchmarking, respectively. Protocols will also be complemented by associated metadata schemata (see below) and direct bidirectional links between protocols and datasets will be integrated in the knowledge warehouse design. In this way, the protocols can be kept general but all important information to reproduce a specific dataset can still be documented by including parameters changing from one experiment to the next as metadata for this dataset. The combination of detailed protocols and all associated data will also be very helpful for defining the domain of applicability of the experimental method since the data will show for which type of nano material and e.g. which size range a protocol was successfully applied. Possibilities to comment and link to external data will be provided at a later stage to allow for better definitions of the applicability by the nanosafety community even when the data cannot be uploaded or no data could be produced since the protocol was not applicable in the described way. Small modifications could then also be documented in this way or, in the case larger variations are needed, derived protocols can be generated and the relationship will be documented in the protocol database.

Protocols from other projects will also be transferred to the ACEnano structure and stored in the database. For larger repositories of protocols available e.g. from NANOMMUNE, DaNa 2.0, NanoMILE and NanoREG, automatic workflows will be developed to extract and translate the information to the ACEnano schema followed by a manual validation and adding of missing information supported by the original protocol providers if possible.

The protocol template could contain the following fields and will be extended according to the requirements collected during the method and protocol standardization efforts:

General information

Status: `<choices>`
- In progress (=*visible only to owner*)
- Under review (=*visible to reviewers*)
- Accepted (=*visible to all users*)
- Rejected (=*visible to owner*)

Development status: `<choices>`
- Under development
- Finalized
- Internally validated
- Externally validated

Project: `<choices>`
- ACEnano
- NanoMILE
- NanoFASE
- NANOMMUNE
- [*add more projects*]

Organisation: `<choices>`
- Organisation Full Name, Country (Abbreviation)
  [*all ACEnano partners*]
  [*add new organisation*]

Created by: `<choices>`
- First Name, Last Name (Organisation Abbreviation) (=*all ACEnano members imported from* [*http://www.acenano-project.eu/private-area/who-is-who*](http://www.acenano-project.eu/private-area/who-is-who))

Contact information

First name: `<CharField>`
Last name: `<TextField>`
Email address: `<TextField>`
[*multiple contacts*]

Protocol name

Unique code: `<CharField>`
[=*automatic identifier will be generated*]
Derived from: `<choices>`
[derived from an existing protocol in the database]
Version: `<CharField>`
Assigned name for ACEnano repository: `<CharField>`
- Project_measurement_method_nanomaterial_version_variant
- e.g. ACEnano_hydrodynamic particle size_DLS_titanium dioxide_1_a

Original protocol name: <CharField>
[=*any descriptive name, trade name, etc.*]

Protocol description

Objective: <TextField>
[*Define the central objective of the protocol*]

Definitions: <TextField>
[*Explanation of terms and abbreviations used in the protocol*]

Area of application: <TextField>
[*Limitation of the method to a certain domain area area*]

Procedure:
Principle: <TextField>
Materials and devices: <TextField>
Controls: <TextField>
Sample preparation: <TextField>
Detailed description of the procedure
Step 1: <CharField>
Step 2: <CharField>
[*add new step*]

Data Management: <TextField>
[*Detailed description of data processing and analysis*]
[*Upload of related templates for data collection and processing*]

Health, safety and environmental considerations

<TextField>
[*Relevant instructions in regard to related health, safety and environmental conditions and protection*]

References

<TextField>
[*list and link to referenced materials, articles*]
[*Author, Title, Journal, Year, DOI*]

Since some of the methods of the ACEnano partners are still at an early stage and under intense development, we expect that the protocols will change over time. To protocol the progress, new versions of a protocol can be created including new metadata/data schemata. These versions will provide a clear change log to the previous versions and a description of why the changes were made

and how they enhance the reproducibility or quality of the resulting data. Additionally, any protocol can be used to create derived protocols, depending on the user needs, e.g. for describing related experiments. By storing the relationships between protocols (derived from), similar protocols can easily be identified, which will support the search for variants if a protocol is not applicable to a specific NM type or size, or where additional considerations are included such as, for example, time-resolved characterisation. The change log and relationships between derived protocols will help to generate troubleshooting guides since they highlight parts of the protocol which can be varied depending on the nano material at hand and will also point regulators to the key areas to consider when evaluating the quality of the data for decision making. Finally, protocols will be tagged to different categories of techniques or classes of NMs (if applicable), and to stages of development/validation, in order to group them during browsing and searching of the database based on the tagging, as well as linked to the decision tree to support method selection for answering specific characterisation questions.

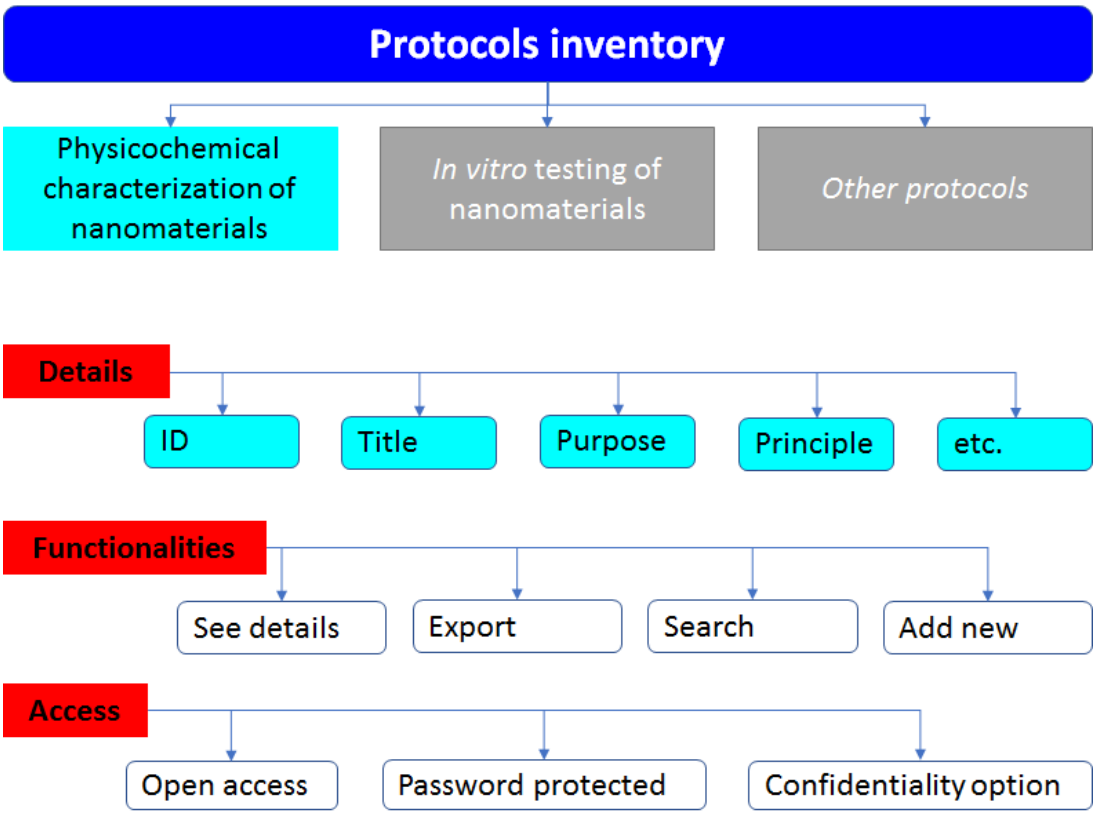

Figure 2. Main functionalities and components of the ACEnano protocols inventory (draft scheme)

3.1.2. Data warehouse

The data warehouse is the part of the knowledge warehouse offering long-term storage of data produced by the ACEnano project or provided by the nanosafety community to be uploaded and harmonized towards the ACEnano goal of generating a reference resource for NMs characterisation

for nanosafety risk assessment. We will especially encourage other projects of the NSC to upload their data to the ACEnano data warehouse to support the harmonisation and FAIR data management goals of the European Commision. The data warehouse will be implemented as a set of highly decoupled components that communicate with each other via well-defined representational state transfer application programme interface (REST APIs). Through this separation, we will be able to provide a more secure and scalable system. Figure 3 shows the overall design and relationships between the individual components of the ACEnano KI and KW. The database system is the central component and will be composed out of one of the two different kinds of databases (relational and NoSQL or non-relational), both or possibly even more complex combinations to handle different data needs. The database system will be interlinked with the protocol database so that the protocol used to generate a specific dataset will be directly accessible from the dataset summary page and can be downloaded with the data and metadata.

A web portal will expose interfaces to end users to access the data but also all other services of the KI such as the protocol database described above as well as the knowledge sharing platform and, at a later stage, the integrated analysis, modelling and visualization services as outlined below. Powerful search features based on faceting will be implemented as part of the web portal to make it easy to search and browse the available datasets and protocols. These facets will be provided by the user during upload or, whenever possible, will be automatically populated by extracting the information from the metadata. All functionality that the user portal provides, including data up- and downloading, searching and browsing will also be available through APIs for automated computer access.

Even if this is not specific for the data warehouse but a central part of the overall KI, the authentication / authorisation system is also included in Figure 3 and will be shortly described here. It will provide a unified identity management and single-sign-on authentication system to all services developed in ACEnano, first internally. Whenever data, protocols or methods are published, they will be made accessible to users from the nanosafety community after registration. This registration will then also allow the user to upload her/his own data, and enable compilation of statistical information regarding users of the KB and KI in order to demonstrate utilisation and impact of the KI.

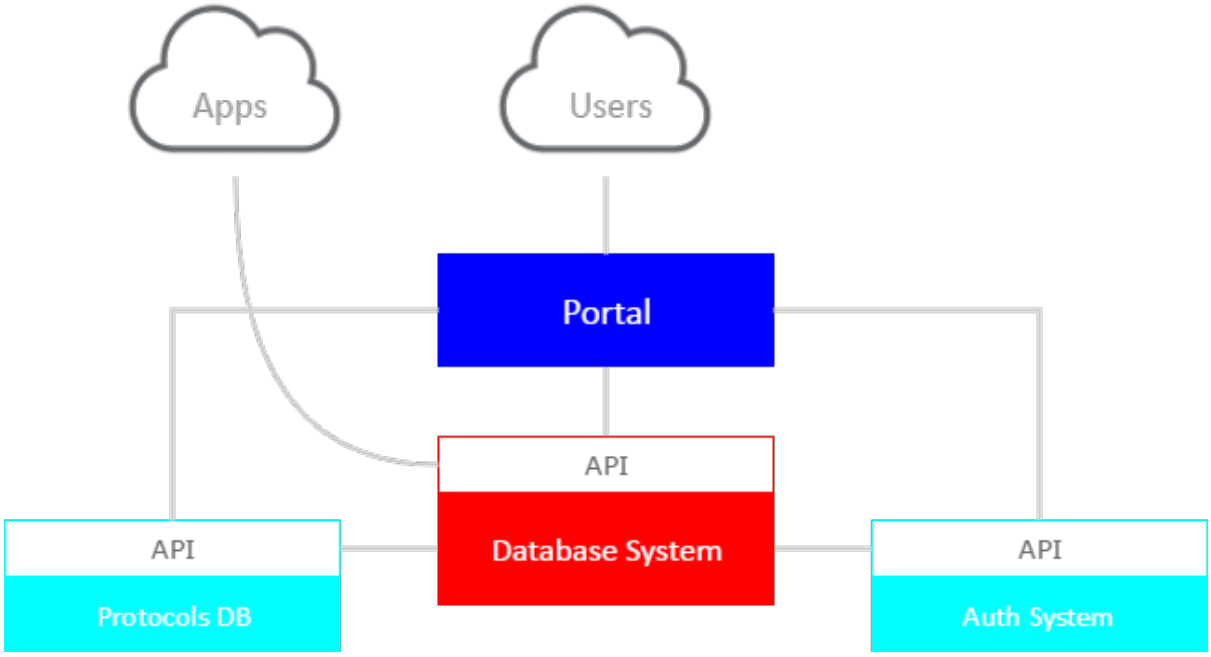

Figure 3. Schematic presentation of the data warehouse composed of the database system, the portal and the authentication/authorisation system as well as its link to the protocols database.

3.1.3. Data templates

The flexibility needed for serving the large group of partners and at a later stage users from outside the consortium, as well as the different use cases of the data, will not be reached by enforcing specific data formats with strictly predefined structure but rather by adopting the concept of data/metadata schemata providing a way to, on one hand, define individual file template per specific characterisation endpoint (as long as the data and metadata is in tabular form) and, on the other hand, achieve interoperability by controlled vocabularies for the column names. Even if the exact formats of the data schema and the endpoint specific file templates are not completely defined since they have to be adopted to the requirements of the ACEnano partners collected and discussed in project-internal data management workshops (*the first was held in South Korea of 16 May, the second in Birmingham on 27 June 2017, and another is scheduled to be held in Basel in July 2017*), we expect that they will be based on sample-centered, tabular descriptions of the experimental metadata accompanied by files with the data in method-specific binary or text files or, whenever possible, also in harmonised tabular form. In the tabular files, one row will describe one specific sample with all the metadata/data in the columns (see Figure 4). This is a very similar approach to the ISA-tab standard, which was also adopted in the eNanoMapper data warehouse using the ISA-tab nano extension. However, we will additionally integrate a concept developed in cooperation with the OpenRiskNet project to semantically annotate the dataset structures. The content of metadata and data columns will be described in a human-readable form as well as based on ontologies and the combination of terms in these will facilitate description and implementation

of more complex concepts (see also Figure 4). This can then be used for reasoning based on the ontologies, for example, automatic validation of the contents and identification of corresponding data in different datasets even if the file formats are not the same. As already described for the protocols, one main goal of ACEnano is to support the harmonization of protocols and data from different projects of the NSC. Therefore, the development of the schemata will not start from scratch but will integrate approaches from previous and ongoing projects. The most important repository of data formats, especially for the physico-chemical characterization of NMs is coming from the NANoREG project (NANoREG templates for data logging – Towards a harmonised way of logging data within the NanoSafety Cluster community) [10], and from NanoMILE.

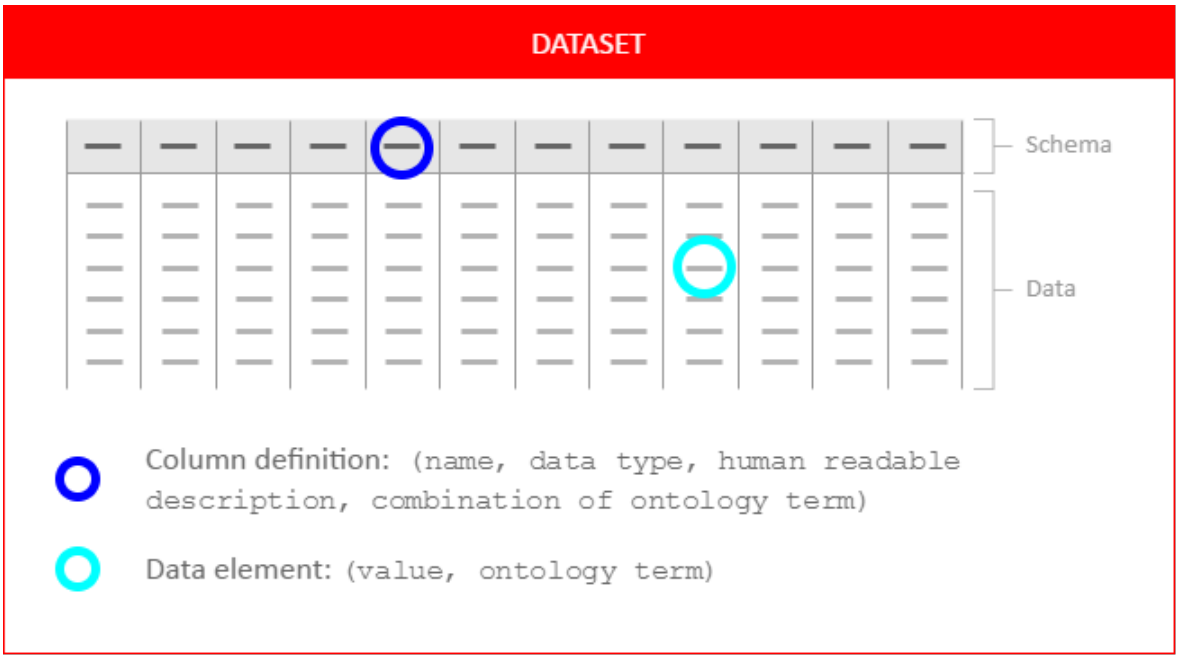

Figure 4. Schematic representation of metadata/data files with corresponding data schema. Where possible, expected value ranges and units will be specified as part of the ontology descriptions.

The NANoREG template for Dynamic Light Scattering (DLS) data was already used as a starting point to demonstrate the proposed concepts using an example of an ACEnano dataset, which was discussed in the ACEnano data management workshop as well as at a combined ACEnano-OpenTox-S2Nano workshop in Korea [11]. The original protocols and data files provided by the University of Birmingham, the NANoREG template as well as the adapted ACEnano template, will now be used to develop the data schema (*see Annex I*).

To support the goals of quality control especially with regards to regulatory acceptance of NMs caracterisation data, the data schemata will be revised by the consortium (especially in the period immediately after the launch) and additional metadata fields can be proposed / demanded for a better coverage of the experiment and the results throughout the time span of the project. These curated data schemata will be included in the Template Library to facilitate the upload of similar

datasets especially for the interlaboratory comparisons / round robins on the methods at sufficiently advances stages for benchmarking and validation.

Finally, the data warehouse will also store data which is derived from raw data e.g. by combining different raw data points by computational approaches (e.g. averaged values, correlation coefficients or data from more complex processing, analysis and modelling procedures). This will be helpful for comparing different processing approaches, for developing common standards and quality control measures, for reporting these approaches and, in this way, for improving the reproducibility of the research. The derived data storage will use a similar approach as used for storage of raw data, in which the metadata for the computational procedure will be reported based on the groups of raw data points combined in the process.

3.2. Data User Layer

3.2.1. Analysis, modelling and visualisation tools

ACEnano is a data-providing project for NMs characterization and the development of new analysis, modelling and visualisation tools is not anticipated. However, to be able to use methods from previous projects like eNanoMapper, as well as to profit from emerging approaches, ACEnano will intensively cooperate with other EU-funded projects like NanoFASE, OpenRiskNet and the new project funded under call NMBP-28-2017: Framework and strategies for NMs characterisation, classification, grouping and read-across for risk analysis and safe-by-design approach for the development of new NMs. To guarantee interoperability of the ACEnano data with the computational tools, a harmonization and interoperability layer of the knowledge warehouse will be developed to provide access to the data by means of an API, besides being able to download the data in specific file formats. The advantage of these data APIs is that they can be accessed directly from the data searching/browsing, visualisation applications, analysis and modelling applications as well as integrated into workflow management tools like KNIME (the Konstanz Information Miner) [12], always providing the most current version of the data without the need to download and update files. Simple statistical information like average, maximum and minimum values and distributions will also be provided directly by the API.

3.2.2. Knowledge sharing tools

All components of the knowledge infrastructure described so far are designed for the long-term storage of data and protocols. However, especially in the first stage of the ACEnano project, protocols and the corresponding data schemata will be under intense development and it is important to get instant feedback from other partners to assess the suitability of approaches taken, to evaluate the completeness and reproducibility of the documentation and to discuss possible improvements. Therefore, a knowledge sharing tool will be integrated in the knowledge warehouse. The main component of this tool will be a shared storage for temporary data files generated by a

new experimental method or a processing/analysis approach as well as protocols and data schemata under development. All partners of ACEnano will be able to see these files and to comment and discuss them directly in the tool. In this way, the newest version of such living documents, the context why they were uploaded and the full history of the discussion will be easily and completely accessible, which is very complicated to achieve with existing tools like sharepoint or when file sharing and discussions are done by e-mail exchange. Additionally, this tool can be used to plan and document the round robins, compare and discuss the results from the different groups, and document the interlaboratory (ir)reproducibility and approaches to improve on it.

4. Conclusions

The ACEnano knowledge warehouse concept envisions provision of services for data management arranged around a data warehouse storing all experimental data of the project and also acting as a central repository for data in the area of NMs physicochemical characterization, as well as exposure, fate and hazard data coming from other projects. The data warehouse will be interlinked with the protocols database, providing a structured approach for documenting the experimental procedures, and will be complemented with a portal providing interfaces for data and protocol upload, browsing, searching and download as well as the user management. Flexible data formats will be used, to allow for the collection of data from very different experiments, and interoperability will be improved by semantic annotation of the data as well as the data format. In this way, the data can be presented to the data user in an optimized form for the specific application at hand, e.g. it could be converted into a format expected as input for a specific analysis, visualization or modelling software. Finally, the KW and KI will support community activities of the project and later by the entire nano safety community and stakeholders (industry, regulators etc.) in the form of the knowledge sharing tool, which can be used to give feedback and discuss protocols under development and temporary experimental and modelling data, as well as documenting the results of the round robins and the steps to improve interlaboratory reproducibility. Finally, the decision tree to guide selection of the appropriate methods from those developed in WP3, and the troubleshooting guides for each method, will also be integrated into the KW and KI.

References

1. eNanoMapper | Computational infrastructure for toxicological data management of engineered nanomaterials [Internet]. [cited 8 Mar 2017]. Available: http://www.enanomapper.net/

2. ToxBank | Data warehouse for toxicity data management [Internet]. [cited 16 Jun 2017]. Available: http://toxbank.net/

3. OpenRiskNet: Open e-Infrastructure to Support Data Sharing, Knowledge Integration and in silico Analysis and Modelling in Risk Assessment [Internet]. [cited 8 Mar 2017]. Available: http://openrisknet.org/

4. Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0. In: FORCE11 [Internet]. 10 Sep 2014 [cited 29 May 2017]. Available: https://www.force11.org/fairprinciples

5. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3: 160018.

6. EU-ToxRisk - EU-ToxRisk – An Integrated European “Flagship” Programme Driving Mechanism-based Toxicity Testing and Risk Assessment for the 21st century [Internet]. [cited 16 Jun 2017]. Available: http://www.eu-toxrisk.eu/

7. eNanoMapper | NCBO BioPortal [Internet]. [cited 4 May 2017]. Available: http://bioportal.bioontology.org/ontologies/ENM/

8. Nanommune. Quality Handbook Standard Procedures for Nanoparticle Testing. 2011.

9. Operating Instructions [Internet]. [cited 4 May 2017]. Available: http://www.nanopartikel.info/en/nanoinfo/methods/992-operating-instructions

10. NANoREG templates for data logging – Towards a harmonised way of logging data within the NanoSafety Cluster community [Internet]. [cited 19 Jun 2017]. Available: http://www.nanoreg.eu/media-and-downloads/templates/269-templates-for-experimental-data-logging

11. ACEnano presented at OpenTox Asia 2017 [Internet]. [cited 19 Jun 2017]. Available: http://www.acenano-project.eu/news-events/12-acenano-presented-at-opentox-asia-2017

12. KNIME | Open for Innovation [Internet]. [cited 19 Jun 2017]. Available: https://www.knime.org/

Annex I - Case study on NMs size characterisation data by DLS

Aim: To organise protocol information and experimental results in order to make them usable by data management systems

Steps:

1. Evaluate the available information (protocol, raw data and processed data) on Dynamic Light Scattering (DLS) measurements:

Sample: 10 K PVP capped copper oxide NMs

Method Description:

Particle size analysis was carried out on a Malvern Zetasizer (nano ZS) using Zetasizer Software Version 7.10. A polystyrene cuvette was filled with about 1 cm of the filtered NMs dispersed in water to be analysed. This was then placed in the sample holder of the particle size analysis apparatus. A standard operating procedure was set up for the material and involved inputting the refractive index and absorption values of the material and the dispersant. Once the parameters were set using the software, measurement was begun. A minimum of five consecutive measurements were collected to ensure repeatability and averaged to calculate a Z-Average size. The results were obtained at 20°C with samples equilibrated for 2 minutes before measurements were started. The Stokes–Einstein algorithm was used to calculate the hydrodynamic diameter of analysed particles.

Procedure:

Material:

Refractive Index: 1.59

Absorption: 0.010

Dispersant: Water

Temperature: 20 °C

Equilibration Time: 120 seconds

Cell: Malvern DTS0012 Disposable Cuvette

Number of Measurements: 5

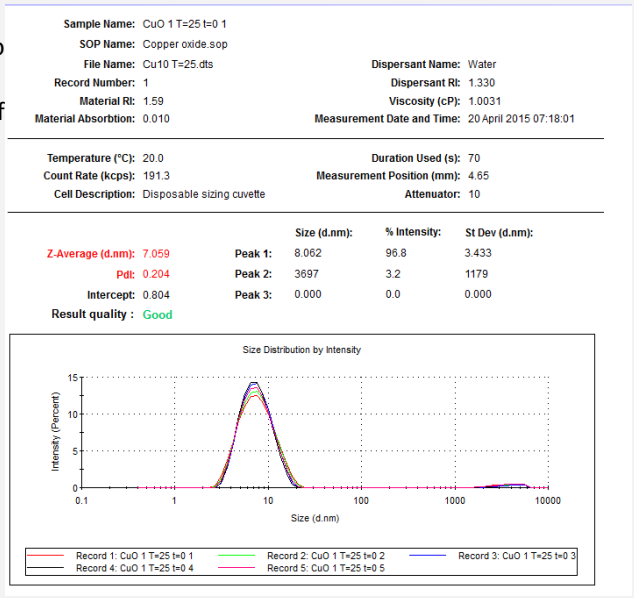

Figure 5. Example of protocol and results on size characterisation by (DLS) provided by UoB

2. Create a metadata file (e.g. in Excel) starting from the existing data collection file for size characterisation by DLS from the NANoREG templates [10]

3. In the metadata file:
    - Fill-in with the information extracted from the Protocol:
        - Sample information
        - Method and instrument information
        - Results
    - If needed, add or adapt the tabs with useful information from the protocol (e.g. 'Constant' data tab)
    - Add a new tab for raw data

Sample Information													
Replicate number	NM ID code (e.g. NM-300, JRCNM0100 1a, ...)	NM chemistry (core)	CAS Number	Vial number	NM supplier	Material State (liquid or fluid, fluid dispersion, powder)	Use of dispersant: Yes/No (e.g. for Ag NPs NM-300K or NM-302)	Dispersant	Dispersant reference (indicate here the reference number of the vial used)	Sample Name (internal reference)	Date of preparation	Date of analysis	File name
1	10 K PVP capped copper oxide	Copper oxide				Dispersion	Yes	Water		CuO 1 T=25 t=0		20-04-2017	Cu10T=25.dts

Method and instrument information														
Cell model	Resting time at room temperature	Temperature (°C)	Thermal equilibrium time (min)	Number of runs	Number of sub-runs	Delay between runs (s)	Duration of the run	Laser focus position	Laser attenuation	Scattering angle	Refractive index of the sample	Absorption index of the sample	Refractive index of the medium	Absorption index of the medium	Viscosity of the suspension
Malvern DTS0012 Disposable Cuvette		20	2	5	n/a						1.59	0.010	1.330		1.0031

Results								SOP
Z-Average (d.nm)	PdI	Hydrodynamic Diameter Z-ave peak 1 (nm)	SD	Hydrodynamic Diameter Z-ave peak 2 (nm)	SD	Hydrodynamic Diameter Z-ave peak 3 (nm)	SD	references to SOPs
7.059	0.204	8.062	3.433	3697	1179	0.000	0.000	

Figure 6. Example of a data schema using a NANoREG template for DLS and experimental information from ACEnano. Note that for some of the parameters, which are fixed to the instrument, ACEnano will define these as constant values and have them auto populate once the instrument details are inputted. Examples of these are the empty columns under "Method and instrument information" which we will subsequently extract from the instrument.