**Background information**

*The following, broad definitions of data “completeness” and data “quality” are provided to clarify the scope of this questionnaire. A key aim of this questionnaire is to identify how these concepts are (or should be) treated in practice i.e. which specific issues are (or should be) considered.*

*Data “completeness” is a measure of the availability of all ‘necessary’ kinds of (meta)data / information which may include the extent of nanomaterial characterisation, both physicochemical/structural and biological, the degree to which experimental details are described as well as the availability of raw data, processed data, or derived data from the assays used for nanomaterial characterisation. N.B. [1] For the purposes of this questionnaire, data may be considered to be “complete” if they are compliant with some set of “minimum information criteria” - although definitions of “completeness” which go beyond “minimum information criteria” are also of interest. [2] The use of the term “completeness” is not meant to suggest that we fully appreciate all necessary independent variables which determine, say, a given result obtained from a particular biological assay - it is understood that definitions of completeness will evolve in tandem with our scientific understanding.*

*Data “quality” is a measure of the usefulness of data which encompasses both their inherent “reliability” (i.e. clarity regarding exactly what is being reported and trustworthiness/reproducibility) as well as their “relevance” (i.e. usefulness for a particular purpose). It may be considered related to data “completeness” and may encompass issues such as precision, error, sufficiency of metadata for reproducibility etc. N.B. This concept may be considered both in a qualitative and/or quantitative sense.*

*N.B. It is recognised that the information which might be required for both “completeness” and “quality” might be use case or data type specific. Where your responses may be considered specific to a given use case and/or type of data/study, please indicate this in your responses.*

**Questions**

Section A: What is meant by data “completeness” and “quality” and why are these issues important?

1. Do you have any comments regarding the broad definitions of (a) data “completeness” and (b) data “quality” provided above? *Please note that detailed considerations (such as which physicochemical parameters to measure or checklists for assigning data quality) should be addressed in subsequent questions.*

In our project data completeness is a parameter of data quality.

To assess data quality only the reliability of the data is taken into account independent of its use (relevance).

1. Please briefly comment on why (a) data “completeness” and (b) data “quality” is important in the context of nanomaterials. *For example, data completeness may be a prerequisite for database interoperability?*

Importance is not different in context of nanomaterials compared to chemicals. (Garbage in –> garbage out)

* Integration of data
* Interpretation of data
* Structure-activity relations
* Predictive toxicology

1. Please tell us about important, existing proposals for data “completeness” (or “minimum information criteria”) for nanomaterials that you are familiar with. *N.B. Existing proposals would include the* [*MINChar Initiative Parameters List*](http://characterizationmatters.org/parameters/)*, those implied by recent OECD reports such as* [*ENV/JM/MONO(2012)40*](http://www.oecd.org/officialdocuments/displaydocument/?cote=env/jm/mono(2012)40&doclanguage=en) *and* [*ENV/JM/MONO(2009)20/REV*](http://www.oecd.org/science/nanosafety/guidancemanualforthetestingofmanufacturednanomaterialsoecdsponsorshipprogrammefirstrevision.htm)*, as well as the proposed consolidated lists of physicochemical properties presented in* [*Stefaniak et al. 2013*](http://dx.doi.org/10.3109/17435390.2012.739664)*.*
   1. Are you familiar with the mentioned proposals? I am familiar with the reports of the OECD
   2. Are there any others you feel are particularly important? *N.B. Feel free to simply provide a set of links to references without additional comments.*

[Nanomaterial registry: database that captures the minimal information about nanomaterial physico-chemical characteristics - Springer](http://link.springer.com/article/10.1007%2Fs11051-013-2219-8)

[Evaluation criteria for the quality of published experimental data ... - PubMed - NCBI](http://www.ncbi.nlm.nih.gov/pubmed/24313439?dopt=Citation)

1. Taking into account the proposals discussed above, which specific (meta)data, or “items of information”, are necessary for nanomaterial data to be considered “complete”? For each “(meta)datum” or “item of information”, please indicate
   1. why is the “(meta)datum” or “item of information” important?
   2. is the “(meta)datum” or “item of information” only important in particular contexts e.g. for particular experiments, nanomaterials or use cases? *N.B. Different use cases might be an experimentalist wishing to reproduce the results vs. a QSAR modeller etc.*
   3. is it possible to rank the “(meta)datum” or “item of information” as more or less important than any other in your list?
   4. whether the “(meta)datum” or “item of information” should be considered “essential” i.e. a necessary component of “minimum information criteria” as opposed to merely being required for “data completeness”?

*N.B. [1] Please enter this information in the Excel spreadsheet you should have been provided with. [2] You may wish to refer to Tables II and III in* [*Stefaniak et al. 2013*](http://dx.doi.org/10.3109/17435390.2012.739664) *, or the provisional NanoSafety Cluster Databases Working Group “*[*minimal standard for reporting*](http://www.nanosafetycluster.eu/working-groups/4-database-wg/tasks-2/2013-2.html)*” as a starting point with regard to physicochemical parameters and/or experimental variables (for in vitro studies) respectively. Table II of* [*Stefaniak et al. 2013*](http://dx.doi.org/10.3109/17435390.2012.739664) *might be considered a starting point for a “data completeness” proposal whilst Table III might be considered a starting point for “minimum information criteria” i.e. a specification of the highest priority “(meta)data”. [3] Feel free to comment on (meta)data or information which would be valuable even if it is not currently available to you at the moment. [4] Please provide some additional, free text comments if you feel this spreadsheet cannot adequately cover the points you wish to make. [5] Please indicate whether you feel your response has been comprehensive or not.*

1. What additional/alternative considerations need to be accounted for when assessing nanomaterial data “quality”?

*N.B. [1] Insofar as possible, please provide this information in the same Excel spreadsheet you should have been provided with for question A.4. [2] Please provide some additional, free text comments if you feel this spreadsheet cannot adequately cover the points you wish to make.*

Section B: How does the purpose of your specific nanomaterial data curation effort impact how you define/assess data “completeness” and/or data “quality”? We try to make a high quality database independent of the purpose of the database

1. As part of your curation efforts, what formalised definition of/set of assessment criteria are you using for

I don’t think we have this formalised definition. But I am not sure I understand the question very well. I think all criteria we consider are in the excel table (A.4.).

* 1. data “completeness”?
  2. data “quality”?

*N.B. [1] If you do not have formalised definition(s)/set of assessment criteria, please state “None” for these questions. [2] If the definition(s)/criteria are publicly available, please provide a link i.e. a web address.*

1. If you are using a (set of) formalised definition(s) or assessment criteria, or your organisation is, please provide the following details. *N.B. Please provide details for both your data “completeness” and data “quality” definitions if applicable.*
   1. What is the basis for your definitions/assessment criteria? *For example, were they informed by, say, evidence in favour of the critical significance of a particular experimental condition in toxicology studies?* Variables that determine the toxicity outcome are considered. These variables are known from different toxicological studies and reports. In practice we chose only the variables which are often described in papers. (for example: confluence of cells upon exposure determines toxicity but is almost never specified in papers 🡪 is not a criterion)
   2. Are you aware of any limitations of your definitions/assessment criteria?

The minimum information criteria are very mild; otherwise there would be no data in our database.

* 1. To what extent do the specific goals of your organisation and/or purpose of your data resource/tool affect your definitions/assessment criteria?
  2. To what extent could your definitions/assessment criteria be generalised for use by other organisations involved in curating nanomaterial data?

Section C: What are established handling methods for addressing data “completeness” and “quality” in mature fields (e.g. biocuration)?

1. Are you familiar with specific established approaches, that exist in mature fields (e.g. biocuration/bioinformatics), for addressing:
   1. data “completeness”?
   2. data “quality”?

[Assessing toxicological data quality: basic principles, existing sc... - PubMed - NCBI](http://www.ncbi.nlm.nih.gov/pubmed/22507180?dopt=Citation)

[A systematic approach for evaluating the quality of experimental to... - PubMed - NCBI](http://www.ncbi.nlm.nih.gov/pubmed/9056496?dopt=Citation)

["ToxRTool", a new tool to assess the reliability of toxicological d... - PubMed - NCBI](http://www.ncbi.nlm.nih.gov/pubmed/19477248?dopt=Citation)

*N.B. Please feel free to simply provide links to publications you feel are of particular significance.*

Section D: What are the key challenges related to the “completeness” and “quality” of nanomaterial data?

1. To what extent do the approaches established in mature fields (e.g. biocuration/bioinformatics) need to be modified in order to be applied in the context of nanomaterial data? Please comment on this with regards to
   1. data “completeness”

* physico-chemical characterisation of the nanomaterials
* measurement techniques used to characterize the nanomaterials
* instance of characterisation
* testing interference of nanomaterials with assays

* 1. data “quality”
* appropriateness of measurement technique for physico-chemical characterisation
* appropriateness of in vitro assay for nanomaterial testing
* instance of characterisation

1. What are the outstanding challenges for nanomaterial data with regard to
   1. data “completeness”?

* For the moment it is not completely clear which variables are determining the toxicity outcome so we don’t know what completeness is. As much detail as possible should be curated.
* A lot of published data is far from complete. A lot of data on characterisation of nanomaterials and the toxicity outcome is missing. Usefulness of these data?
  1. data “quality”?
* Standard methods (characterisation techniques and in vitro assays) used for the toxicity assessment of chemicals must be revalidated for nanomaterials.

1. Which of these key challenges are specific to the goals of your organisation and/or data resource with regard to
   1. data “completeness”
   2. data “quality”? no

Section E: Are there any specific use cases to illustrate these issues and make them tangible?

1. Please provide examples of case studies and/or scenarios which illustrate the key challenges (as noted above) for nanomaterial data with regard to
   1. data “completeness”
   2. data “quality”

Section F: Recommendations. What are some practical next steps for individual stakeholders or the community as a whole?

1. To what extent do you feel there is (a) redundancy in nanomaterial data and (b) how best can this be addressed? *For example, might computational predictions be employed to substitute for missing physicochemical characteristics (PCCs) based on a subset of measured PCCs?*

All data on specific groups of nanomaterials should be collected and integrated to be analysed. Results of the analysis could reveal redundancy and guide future experiments.

We don’t think there is a redundancy because data completeness and quality of the current studies is very low. A lot of existing information is not very useful for analysis.

1. Taking into account all of the previous questions, are you able to recommend specific definitions/assessment criteria, which should be adopted by the community as a whole, or for specific scenarios, for
   1. nanomaterial data “completeness”?
   2. nanomaterial data “quality”?

*N.B. You may wish to simply propose either that (a) your own organisation’s criteria should be universally adopted (or adopted in specific contexts) or (b) that a scheme based upon compliance with the “wish list” you proposed in response to questions A.4 and A.5 should be applied.*

1. Are you able to recommend how to best capture the information required for “complete” data within, say, a database? *For example, might templates based upon pre-defined fields in* [*ISA-TAB-Nano*](https://wiki.nci.nih.gov/x/MwGGAg) *files be employed in some fashion?*

Within our project pre-defined templates in ISA-TAB-Nano were developed. I think this is very helpful. Standardized templates should be available and maybe obligatory to fill out when publishing data.

1. Are you able to make any recommendations regarding
   1. how best to implement a scoring scheme for “quality” and/or “completeness” of nanomaterial data in practice? *For example, a human expert might (1) read a nanotoxicology study and score the data using a set of predefined questions as (2) implemented in an extension of the* [*ToxRTool*](http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/archive-publications/toxrtool) *program.*

I think pre-defined ISA-TAB-Nano files should be filled out when publishing data. Everything needed for a data quality or completeness evaluation should be included in the files. The scores can be calculated automatically from the ISA-TAB-Nano files.

* 1. what challenges would be associated with implementing such a scheme?

Development of pre-defined ISA-TAB-Nano files for in vitro/in vivo and physicochemical characterisation of nanomaterials

* 1. existing approaches which could be extended?

Extension of ToxRTool is a good idea but then automatically in the ISA-TAB-Nano file.

1. Do you have any suggestions regarding a dataset or data resource which would be suitable for a pilot study for addressing the issues raised in this questionnaire?

The nanosafety cluster on modelling is building a database on the in vitro toxicity of amorphous silica nanoparticles

1. Do you have any recommendations from a higher-level perspective (i.e. to funding agencies, researcher associations, publishers etc) on how to move forward? no