**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.*

*(a) Using the term "data completeness" makes sense if the associated "minimum information criteria" are clearly communicated at the same time. Depending on what these criteria are and where they come from, "data completeness" can have very different meanings. Hence, the term should be used always together with the criteria on which it is based.*

*(b) I would associate the term "data quality" more with the technical quality of the data (precision, error, reproducibility, etc), and less with their relevance for a specific purpose. In my view, data of high technical can at the same time be of little relevance for a specific purpose. Therefore I would prefer to use "data relevance" as another criterion in addition to "data quality".*

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?*

*Both these terms are highly important, and not only in the context of nanomaterials. Data completeness and quality are essential for the safety assessment of nanomaterials, and for the comparation of different but similar nanomaterials in terms of hazard and risk assessment.*

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 all three of the proposals listed above. These proposals have a lot in common, and the same core properties are suggested in all three of them. Additional properties that are not common to all three can also be very useful for specific purposes.*

* 1. 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.*

*I am not aware of other particularly important proposals.*

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

*Please see attached Excel sheet*

* 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.*

*Important: Reference materials used? Instrument calibration done (when, how)? Are the instruments used adequate for the desired purpose? Are the raw data available/accessible? Why are the reported data considered relevant by the data provider?*

Section B: How does the purpose of your specific nanomaterial data curation effort impact how you define/assess data “completeness” and/or data “quality”?

1. As part of your curation efforts, what formalised definition of/set of assessment criteria are you using for
   1. data “completeness”? *None*
   2. data “quality”? *None*

*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?*
   2. Are you aware of any limitations of your definitions/assessment criteria?
   3. To what extent do the specific goals of your organisation and/or purpose of your data resource/tool affect your definitions/assessment criteria?
   4. 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”?

*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”

Reported data should reflect the fact that certain properties are thought to be of specific relevance for the safety assessment of nanomaterials. Please see question 3.

* 1. data “quality”

No fundamental modification necessary, but the methods (measurement and evaluation) should be recognized as being suitable for nanomaterials.

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

"Completeness" depends on the minimum information criteria used, but there is no generally accepted standard for that. So the challenge is to agree on such a general standard.

* 1. data “quality”?

*Characterisation of nanomaterials and nanotoxicology are both not yet very mature fields. Therefore it is not easy (a) to assess the quality of third party data and (b) to agree on an acceptable standard for data on 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”?

*All..*

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”

*Currently there are open cases with regard to the registration of nanomaterials for the purposes of the European Chemicals Regulation (REACH).*

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?*
2. 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”?

A good approach for the safety assessment of nanomaterials would be to follow the OECD proposals mentioned in question 3. This could be supplemented in specific cases with information according to the other two proposals.

* 1. 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?*

*It is very difficult for me to recommend a specific database. Experience shows that no matter how detailed and complete a suggested database is, specific projects and research groups still prefer to develop and use their own tailor-made databases and approaches.*

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.*
   2. what challenges would be associated with implementing such a scheme?
   3. existing approaches which could be extended?
2. 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?
3. Do you have any recommendations from a higher-level perspective (i.e. to funding agencies, researcher associations, publishers etc) on how to move forward?

*According to my experience, general approaches carried out by international (global?) organisations have the highest chances that their results will be accepted and used by a broad community (OECD, FAO,…). The drawback is that such approaches tend to be slow. Regional (US/EU/….) and sectorial (Environment/food/cosmetics/…) may be more successful in their fields, but at the same time more of specific use.*

*My recommendation is therefore to (i) get a clear vision in which direction to move forward (which sector) an then (ii) work on sector-specific approaches rather than on general approaches. In any case, I do not recommend to work towards regional solutions!*