**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.*

The field of biological effect assessment of nanoparticles is rapidly growing at the moment. Also, the number of parameters available for characterising nanomaterials, is rapidly growing. Thridly, we do not (yet?) have sufficient insight in the main properties determining biological responses of wide arrays of nanomaterials that vary with regard some of their phys.chem. properties. These observations imply that it looks like there are almost infinite possibilities of collecting and reporting meta- and other type of data on top of the actual effect data. Commonly, the development of broadly agreed and validated procedures for generating new types of additional data, is lagging behind the development of new means of characterizing nanomaterials in their broadest sense. **Our experience hence is that although it is nowadays common to report numerous experimental details, the data are never to be considered as ‘complete’.** In common practise, this is also our experience when publishing data: the requests for additional experimental details and new methods of characterising particles and test solutions, seem to diverse more and more, depending on the journal and on the reviewers.

In short, our response to the question on data completeness is that the definition in itself is okay, but the reference by which the data completeness is to be judged, is dynamic – both in a quantitative and in a qualitative sense.

With regard to data quality, we support the definition and have no additional comments.

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 issues are of major importance. Data completeness is essential when comparing data for whatever purpose, or when using data for, for instance, risk assessment. In the latter case it is preferred that it is possible to have data available of similar completeness, and of similar quality. The same is true for the application of data within QSAR and other types of predictive modelling. In the latter example, it is common practise that data from different sources lack various key parameters that are essential elements of data completeness. In such data, the data cannot be merged for the purpose of QSAR development or QSAR validation.

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?

Yes, we have been involved in part in the development of some of these reports.

* 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.*

No, not really. I even doubt whether the list of minimum ‘requirements’ is currently still of use.

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?

I am afraid that, as stated above, data are never complete. Most proposal include detailed information on the pristine particles. Nowadays, it is increasingly recognised that characterisation of test solutions, assessment of transformation pathways and transformation products, as well as kinetic parameters of both solution stability as well as the dynamics of interactions of particles and biological surfaces, are increasingly important.

* 1. 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.*

*Oh no, this is virtually ‘always’ of importance, i.e. at least until we reach the stage in which it is possible to extrapolate on a higher scale/level, the results of data collection of different nature (like extrapolation of information on pristine particles towards the kinetics of aggregation in test solution, just to give a general example).*

* 1. is it possible to rank the “(meta)datum” or “item of information” as more or less important than any other in your list?

Not on forehand. This will depend on the application of the data.

* 1. 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”?

Again, this will depend on what the data are used for.

*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”?

An important consideration are the issues of standardization and validation of specific methods, and of the extend of experience obtained with specific methods: data quality is to be considered higher when these issues have been properly dealt with.

*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”?

1. As part of your curation efforts, what formalised definition of/set of assessment criteria are you using for
   1. data “completeness”?
   2. data “quality”?

All these issues are dealt with on a case by case basis. For instance, for the case of risk assessment of specific nanomaterials, it has turned out that we basically are not able to apply strict criteria with regard to data completeness and data quality. For pragmatic reasons (as we would otherwise be stuck with no data at all), we basically need to set aside most criteria defined on forehand. An example is the case of nanosilver: here one of the key issues, apart from nano-considerations, is dissolution of nanosilver in aquatic solutions. When closely inspecting the available data, it is virtually impossible to assess the extent of dissolution and the impact of silver ions on the toxicity profiles reported in literature.

*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?*

*Basically, we apply in general terms the same criteria as done for conventional substances. For pragmatic reasons we do not apply strict or formal definitions or assessment criteria.*

* 1. Are you aware of any limitations of your definitions/assessment criteria?

For sure: any criteria will have its (severe) limitations, dependent on the endpoint that is being assessed.

* 1. To what extent do the specific goals of your organisation and/or purpose of your data resource/tool affect your definitions/assessment criteria?

To a large extent, the key is that we report as much as possible the criteria used, and we report on the limitations when evaluating specific studies.

* 1. To what extent could your definitions/assessment criteria be generalised for use by other organisations involved in curating nanomaterial data?

Basically, this is possible by a large extent. Actually, as we are doing most evaluations within a European setting, the criteria are evaluated by other EU countries as well, and also by for instance ECHA and (to a limited extent) by industry.

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”?

Yes. What comes to mind first, are the criteria set by OECD for evaluation of specific test methods, and (probably more relevant) the criteria set by OECD for evaluating the merits of predictive (QSAR) models, including evaluation of the underlying experimental data.

*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”
   2. data “quality”

As stated above, the key challenge is to have sufficient flexibility in assessment of completeness and quality, given the fast developments with the field of nano-toxicity assessment, or nano-fate.

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

Assessing the linkages between various types of data/endpoints, generated by strongly different methodologies, and the extrapolation of data across (test) systems and across environmental media (like extrapolation across different surface waters).

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”?

Now of them are really specific, we try to deal with all of them: one way or the other

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”

I do not have specific, detailed, cases available

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?*

In itself there is a large extent of redundancy. On the other hand, it is important to note that often different methods yield information on different endpoints. To give the example of nanoparticle characterisation: there are different methods available to assess the size distribution. Each method will yield a different size distribution in numerical terms. It is very difficult to transform size distributions obtained by a specific methods into the distribution obtained by means of a different method.

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”?

No, not at this stage. I think that this is still one step too far.

*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?*

My proposal would be to start with an extensive set of pre-defined fields. However, it will be very important to leave blank fields that may be used to include data generated by novel methods, or data generated for ‘new’ endpoints. We have performed a survey of ecotoxicity data and found that at the end, you are left with a database that is basically empty when considering all possible means of characterising nanomaterials and test conditions.

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?

Again, the key would be to have a flexible system that is able to accommodate new endpoints and new methods. The scoring system should at least contain the assessment the number of endpoints that are different from basic considerations. For instance, two methods for determining size should be considered differently in terms of completeness than the case in which only one method is used for determining size but supplemented with a method to determine surface charge. The latter example would be ‘more complete’.

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?

A possibility would be the case study mentioned above of ecotoxicity data and underlying particle and medium characterisation.

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?

The best way to move forward, is to start with defining a minimum set of criteria. The set should be feasible in terms of experimental efforts and it should yield sufficient information to be of practical use. In addition it is important to recommend to redefine the minimum set on a regula basis in order to accommodate new scientific and regulatory developments.