**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.*

These seem pretty complete yet concise to me as definitions. I might just add a caveat that extensive characterisation of pristine nanomaterials is probably not sufficient alone, without understanding of the transformations that would like occur and their timescales, in the relevant exposure environment. From my perspective, characterisation under the exposure conditions should be a central caveat of both data completeness and data quality. As datasets emerge, it might be possible in the future to predict characteristics under the exposure conditions from the pristine physical-chemical characteristics, but we are no-where near there as yet.

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 are essential in order to demonstrate a cause-effect relationship or to rule out such a relationship. Thus, if for example no toxicity is observed with a specific NM in a specific assay, without appropriate characterisation of the NMs under the exposure conditions (as well as ideally imaging of the cells/organisms to demonstrate lack of uptake/interaction) it is not possible to conclusively confirm that the NM dose was actually applied to the cells/organism as the particles may have agglomerated and thus been too big to be taken up in significant amounts.

Additionally, as learned painfully through inter-laboratory comparisons of even simple assays like DLS size determination or MTS assay for cytotoxicity, even very small details of the protocol, that may not even be documented in an early iteration as they seem “obvious” can have dramatic effects on the outcome. For example order of mixing during dilution of NP stocks (NPs in medium or medium added to NPs?) can play an important role in terms of reproducibility. Thus, a much greater emphasis on publication of detailed protocols and troubleshooting of protocols (such as via the Nature Protocols format which includes timing of steps and troubleshooting) should be encouraged in this community – if not in nature, when in other protocols journals such as Analyst and others.

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. Most (all?) need a much stronger emphasis on what is actually happening at the surface in different contexts (e.g. human biofluids, environmental matrices etc.) such as degree of charge dissolution (e.g. pKa), binding affinity of an capping agents etc., and a much stronger link between physico-chemical measurands and those parameters of use in models such as QSARs.

* 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.*

Not aware of more recent ones that those in Stefaniak et al.

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.*

See attached spreadsheet. I don’t think that my answer is fully comprehensive, but rather providing some pointers as to additional levels of detail / specification that is required for NMs in order to approach data completeness and data quality and facilitate interoperability.

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.*

As 4. above

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”? At present in NanoMILE, this is a compromise between what would be considered ideal and what is realistic within the timeframe and budget of the project, (i.e. characterisation of materials at selected timepoints rather than all timepoints, and in “representative” media rather than in every variation of the media in terms of additives and serum treatments. However, the baseline is full physic-chemical characterisation of all NMs in their pristine state and a minimal set of characterisation in the test media for each of the test organisms (cell lines, algae, water flea, worms and zebrafish / zebrafish embryo). More detailed characterisation studies are then performed ad hoc as specific needs arise (i.e. linked to data that are not easily explained without additional NMs characterisation, such as NP dissolution under exposure conditions).
   2. data “quality”? At present in NanoMILE, this has been defined within each WP for their endpoints and approaches, and is generally compliant with international best practice in each of the areas (NMs synthesis and characterisation, protein corona determination, toxicity (*in vitro* and *in vivo* and via High Content Screening), ecotoxicity (*in vitro* and *in vivo*) and “omics” assessment as the consortium partners are world-leading in these areas. Where best practice doesn’t exist, NanoMILE partners are developing this, i.e. in terms of protocols for ageing NMs under air, water and food conditions. We have a deliverable at Month 24 (March 2015) to describe our data quality considerations, and as such are working to bring the information together and harmonise across WPs. This will be

*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?

Not able to answer this question as yet, but will certainly include such considerations in our deliverable report as they are extremely useful.

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.*

It’s not really my area of expertise. But I found this quite helpful in terms of a specific case-study that put it into a specific context: http://www.ncri.ie/publications/statistical-reports/data-quality-and-completeness-irish-national-cancer-registry

This makes a good case for pushing/forcing a change in mid-set of researchers so that we are capturing and documenting key data as it is generated rather than retrospectively after the experiments have been performed!

Acad Emerg Med. 2005 Sep;12(9):884-95.

The accuracy and completeness of data collected by prospective and retrospective methods.

Nagurney JT1, Brown DF, Sane S, Weiner JB, Wang AC, Chang Y.

Lu FC. Safety assessments of chemicals with thresholded effects. Regulatory Toxicology and Pharmacology, 1985, 5(4): 460–464. doi:10.1016/0273-2300(85)90009-1

I like this one especially for the line at the end of the abstract “In addition, the possible inadvisability of overemphasizing the importance of dose-response relationship at the expense of in-depth, relevant studies is discussed.” This is especially relevant in light of the recent paper by Harold Krug stating that majority of nanosafety literature is useless!

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” the context-specific nature of NMs, coupled with their potentially multi-component nature (where each component has different toxicity and probably different degradation rates etc.) makes them more difficult to characterise and describe completely. Also, unlike chemicals for example, there is no convention as yet to report the purity or list the potential impurities associated with NMs, although purity is included in the NSC minimal standards for reporting list, which is great, but there is no agreed method to do this as yet. Plus, it would need to be applied to each component (core, shell, coating, capping agent, functionalization) so how to do this? Does this then extent to under the exposure conditions listing all binding components (ecological or biological corona)? A similar approach should also include the speciation and proportion of different species over time under the exposure conditions Some early efforts towards this include:

Izak-Nau E, Voetz M, Eiden S, Duschl A, Puntes VF. Altered characteristics of silica nanoparticles in bovine and human serum: the importance of nanomaterial characterization prior to its toxicological evaluation. Part Fibre Toxicol. 2013 Nov 11;10(1):56. doi: 10.1186/1743-8977-10-56.

Domingos RF, Franco C, Pinheiro JP. The role of charged polymer coatings of nanoparticles on the speciation and fate of metal ions in the environment. Environ Sci Pollut Res Int. 2014 Sep 16. [Epub ahead of print]

Nanayakkara CE, Pettibone J, Grassian VH. Sulfur dioxide adsorption and photooxidation on isotopically-labeled titanium dioxide nanoparticle surfaces: roles of surface hydroxyl groups and adsorbed water in the formation and stability of adsorbed sulfite and sulfate. Phys Chem Chem Phys. 14, 6957-6966

DOI: 10.1039/C2CP23684B

Another divergence from chemicals at present is that many nanotoxicologists (and even many nanoparticle producers) have very little understanding of the surface chemistry of their materials (which functional groups, which bonds at the surface etc.). All these need to be addressed as part of a drive towards data completeness and again making the links between properties that are characterised and properties that are useful as the basis for QSARs.

* 1. data “quality” The combination of NMs inherent instability (i.e. most commercial dispersions are not guaranteed for more than 1 year) and their batch-to-batch irreproducibility require some additional considerations regarding data quality. Several of these are indicated in the excel file, and should over time become incorporated into good practice in this field. A checklist to complete as part of NMs data quality would include date-stamped characterisation of the pristine NMs periodically during the study duration, as well as periodic checks that characteristics / cytotoxicity in the exposure media were also consistent as often biological milieu more sensitive to small changes in characteristics, for example. This would also include information on, for example, the storage conditions once a new vial of NPs is opened and the number of times the bottle was opened and closed during the duration of the study (which could be calculated from the number of separate studies, repeats etc.).

1. What are the outstanding challenges for nanomaterial data with regard to
   1. data “completeness”? as above
   2. data “quality”? as above
2. Which of these key challenges are specific to the goals of your organisation and/or data resource with regard to
   1. data “completeness”? None are specific to our goals as apply generally, but NanoMILE is addressing in a systematic way issues of ageing, environmental transformation and linking these to observed (eco)toxicities (in vitro, in vivo, and HCS). NanoMILE also intends using this data for development of QSARs if possible (i.e. if relationships are determined to exist / suitable convergence between what experimentalists measure and what modellers can model is reached).
   2. data “quality”? None, but as with all projects we are compromising between what we would like in an ideal scenario and what is achievable within budget and time constraints as well as ensuring interesting PhD and postdoc experiences for the researchers employed.

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?*

This would be the optimal scenario, both for pristine NMs and to predict fate of pristine NMs under realistic exposure conditions. Modelling work is ongoing to study for example effects of hydration and ligands on Silver NP dissolution as a function of the cluster/particle size and the dominant crystal faces (which for very small particles are intimately linked) – work performed within EU FP7 ModNanoTox submitted for publication. Such approaches could in due course be used to predict fate and behaviour, and then be validated experimentally under specific conditions to refine the models. This would be an important and potentially very productive line for research.

Within NanoMILE, one of the approaches that we are investigating is to try to determine which physico-chemical parameters are inter-dependent and as such if one is varied the other changes also, in order to potentially reduce the overall characterisation required. Use of systematically varied NMs libraries is vital here but despite the multitude of synthetic routes, actually producing such libraries is still a significant challenge without introducing additional variants between members.

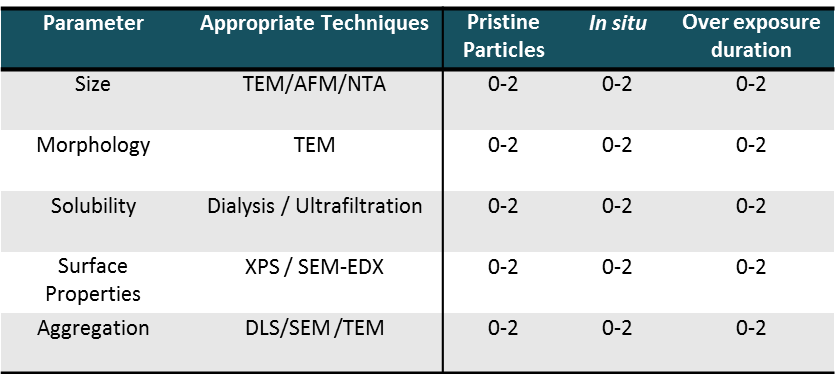
1. Taking into account all of the previous questions, are you able to recommend specific definitions/assessment criteria, which should be adopted by the community as a whole, or for specific scenarios, for
   1. nanomaterial data “completeness”?
   2. nanomaterial data “quality”?

*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?*
2. 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.*

Within EU FP7 ModNanoTox, the human expert approach was used, and while this produced a very high quality of extracted data from literature papers, it was very slow (about 1 paper/day) in order to fully verify the quality of the study and whether the findings were valid. We are currently writing up a paper on this that can be shared in due course. The scoring criteria were quick to determine if the study merited detailed analysis (i.e. had to pass a threshold in terms of the amount of phys-chem characterisation provided from the total maximum that considered pristine and under the exposure conditions as well as material-specific aspects such as dissolution, oxidation number and/or photostability.

**Table 1:** NP characterization scheme for silver NPs



0 = absent; 1 = qualitative; 2 = quantitative, sufficient for statistical analysis. Note study can score a maximum of 30 here if all characterization is performed.

Since not all end-points are relevant to all NMs, I would recommend that rules-based approach could be coded, that would define a maximum characterisation score (as per the ModNanoTox) approach.

In ModNanoTox the studies were specifically related to aquatic toxicity of AgNPs, and studies were then analysed to extract out details such as capping agent, LC or EC50 and other relevant outputs. The Quality Control of the biological assays and end-points was based on basic scientific principles – the studies were too variable to standardize. Typical parameters recorded regarding the biological assays included species used, gender / life-stage, maintenance and preparation, media conditions (pH, ionic strength, temperature), exposure route, exposure duration, endpoints measured, endpoint method and controls included etc.). However, to be widely applicable, this step would also need to be coded and some minimum set of data tabulated in order to give the data a *completen*ess and *quality* score. I would be really keen to support something like this.

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

Given the enormous variability of end-points and assays, and the divergence in the community as to whether standardisation at this point is essential, and in terms of toxicity versus mechanism (e.g. Harold Krug recent paper - Krug HF. Nanosafety research--are we on the right track? Angew Chem Int Ed Engl. 2014 Nov 10;53(46):12304-19.) a key first question would need to be the purpose of the study and the completeness and quality thresholds adjusted accordingly. These should ideally align with the different purposes for which the data could then be used – e.g. further research, modelling, safer-by-design modelling, versus risk assessment and formal decision making where clearly higher thresholds are needed.

One such approach could be based on the other (more constructive) recent paper from Krug and colleagues:

Rösslein M, Elliott JT, Salit M, Petersen EJ, Hirsch C, Krug HF, Wick P. Use of Cause-and-Effect Analysis to Design a High-Quality Nanocytotoxicology Assay. Chem Res Toxicol. 2015 Jan 6. [Epub ahead of print]

* 1. existing approaches which could be extended? See above (ModNanoTox and/or C&E analysis.

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 ModNanoTox dataset could be useful for this purpose, and we would be delighted to share it. It consists of data from ~130 publications specific to Aquatic toxicity of AgNPs and covers bacteria, algae, daphnia, fish etc.

The NanoMILE datasets are still too early in their data-generation phase to be of use, but could potentially be offered as a validation dataset at a later stage.

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?

Until the community has progressed a bit in terms of agreeing best practice and having somewhere to deposit data usefully, external pushing may be counter-productive and result in further fragmentation. Thus, I comment the community-lead approach taken here and am keen to support it in every way possible (despite being late filling out the questionnaires – this is not a lack of commitment but rather a symptom of being overstretched).