Thoughts on Read Across, QSAR, and Data Completeness

John Rumble
jumbleusa@earthlink.net
25 September 2016

In thinking about Read Across, QSAR, and data completeness, it is useful to go back to what we are really trying to do by these efforts. The goal is simple:

> *To predict a property and functionality (behavior) of a substance, without making the necessary measurement, based on the similarity of that substance to other substances for which we have measurements.*
>
> *Further we wish to do this on the basis of scientific evidence, which includes reproducibility, accurate uncertainties, and sound physical principles.*

Let us start with "sound physical principles," and define a model for the measurement of a property of a nanomaterial, as shown in figure 1.

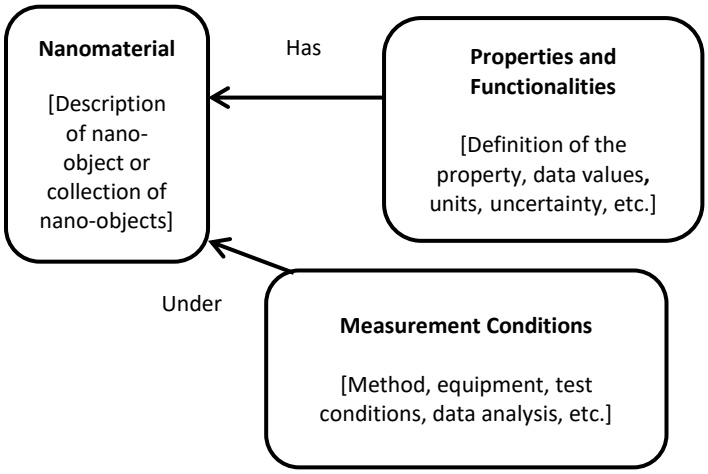

Figure 1. Types of information used to describe a nanomaterial and its properties

Based on this model, a property is dependent on the conditions of measurement and are specific to that nanomaterial. More formally,

> Property = Function (Nanomaterials, Measurement Conditions)

Then to achieve our goal of prediction, we need to answer the following questions.

1. Which nanomaterials features (nf) affect a property (P)?
2. Which measurement conditions (mc) affect the property
3. What is the function that links nanomaterials features and measurement conditions to the property, that is, how do they affect the property?

Again more formally, we can write

$$P = f\,(nf_1, nf_2, nf_3, \ldots, nf_{max}, mc_1, mc_2, mc_3, \ldots, mc_{max})$$

The number of possible independent nanomaterial features and measurement conditions is very large, and a sample set is given in Table 1. Other description systems, database schemas, and ontologies give similar sets of categories. Under each category, many detailed descriptors (independent variables and parameters) are involved.

Features of Nano-objects[1]	**Features of a Collection of Nano-objects**[1]	**Classes of Measurement Conditions**[2]
Shape	General features	Specimen preparation
Size	Size distribution	Measurement type
Chemical composition	Interfaces	Measurement procedure
Crystallographic structure	Surface characteristics	Instrument descriptions
Surface characteristics	Physical Structure	Test environment
Physical structure	Composition	Initial test results
Production method	Topology	Data analysis
Specifications	Association	
	Production method	
	Specifications	

Table 1. General categories of nanomaterials features and measurement conditions for a property measurement

[1] Uniform Description System for Materials on the Nanoscale (www.codata.org/nanomaterials)

[2] Based on Manual on the Building of Materials Databases (ASTM Manual 19) (Philadelphia 1993)

Answering the first two questions require careful measurements that identify the *nfs* and *mcs* that affect the property measurement. While all *nfs* and *mcs* have some effect, careful and systematic measurements are required to identify which are the most important. Standard test methods as developed by ISO, ASTM, and OECD have presumably been designed based on such analysis. The same must be done for answer the third question – are there synergistic effects due to a combination of two or more features, e.g. size and shape or pH and biomaterial concentrations.

The process of answering these questions and therefore putting prediction on a firm basis takes time. This is especially true for biological and environmental impact of nanomaterials.

- Nanomaterials ae complex and their size means that multiple nanomaterials that appear to be the same may differ in significant details. For example, the same general shape – star – may have varying size points.
- The surface reactivity of nanomaterials means that when prepare for testing and placed in media with biomolecules and other substances, coatings happen easily and usually randomly. Consequently testing is not being done on the original nanomaterial but one that is heavily coated.
- Actual mechanisms of actions can be difficult to discern, and relying on gross effects, while proving possible hazardous effects, may not say anything about what actually caused that effect.

Even reproducibility does not clarify cause and effect. As Krog has shown, most measurements showing gross effects have not control important gross level conditions (dosage, uniformity of nanomaterials, consistent test conditions) so that proposed nano-object mechanisms are suspect.

- Recent work by FutureNanoNeeds and others have cast doubts on the toxicity of nanomaterials themselves rather than the ions and other species the devolved from nanomaterials. If that is the case, toxic effects are not the result of a nanomaterial mechanism, but rather the result of increased exposure of materials (ions, etc.) already known to be toxic.
- Unlike organic and biochemistry, large data comprehensive data sets of nanomaterials properties and test results are not yet available. Those data sets that are available are sparse and do not contain the richness of data that years of testing of organic and bio-chemicals have produced.
- Present formal and informal standards for reporting test results, including description of nanomaterials, test procedures and deviations from standard procedures, and bio-assays, are not detailed enough to enable valid data analysis to determine cause and effect in most instances.
- The research culture in nanomaterials is still in the exploratory stage, with emphasis of finding new phenomena and gross effects. This is normal for a new scientific discipline, and nanomaterials research funding is concentrated on such efforts. Over time the need for more systematic exploration of cause and effect will evolve, but that takes time.

Based on the above, we can ask the following questions and try to answer them accurately.

1. Is the present body of data on the measurement of nanomaterials properties sufficiently complete and of high enough quality to support read-across and QSAR in the broad sense?
2. If not, is the body of data sufficiently complete with enough quality to support read-across and QSAR in specific areas, e.g. types of nanomaterials or specific biological outcomes?
3. If the data are not complete enough or of inadequate quality, or both, what can be done to improve the quantity and quality of needed nanomaterial property data?
4. How can one assess the quality of existing nanomaterial data, even if the data sets are not complete with respect to number of nanomaterials, properties measured, and metadata reported?
5. Are data recording formats for biological and environmental assays for nanomaterials, such as OECD templates, ISA-NANO-tab, eNanoMapper, etc., good enough to capture all relevant metadata (nanomaterials features and measurement conditions)? If not, how can they be improved?
6. If the body of data is not sufficient to support fully read-across and QSAR work, can scientifically useful approaches be developed to compensate for the lack of completeness or quality?
7. Grouping should still be possible in spite of inadequate data. How can preliminary grouping be done with existing data?
8. If grouping is done, how can it be made extensible as new data becomes available?