Pilots

Contents of this Page

Goals

Advances in nanotechnology research, development, translation and regulation are currently impeded by the lack of an informatics infrastructure that permits greater collaboration among the different disciplines, entities and stakeholders engaged in nanoscience and nanotechnology. Despite ample precedent in related fields, only recently has a consensus begun to emerge about specific needs in nanoinformatics and possible collaborative approaches to satisfy them. An exchange of ideas on informatics requirements among the various communities of interest would be helpful in developing a broader consensus. The purpose of this note is to initiate such a dialogue, but in the actionable context of identifying pilot projects in areas of critical need, critiquing the pilot's capabilities with respect to realistic user scenarios, and collaboratively iterating their design, functionality, and interfaces to satisfy the community of interest's requirements. The definition of "community of interest" is therefore the collection of stakeholders with a need for the pilot's capability with resources available to develop it and a willingness to engage collaboratively. This note will present a high-level view of some expressed needs and "seed" pilot projects which may be especially pertinent. Finally, this document is intended to be a working document which will be updated with input from others interested in nanoinformatics. The current document stresses nanobiology and nanomedicine, whereas the goal is to collaboratively produce a document with a more comprehensive scope. Please address any comments to Martin Fritts (frittsmj@mail.nih.gov) or Raul Cachau (cachaur@mail.nih.gov).

Background

Data

Currently there is fairly wide agreement that there is a lack of reliable, curated data in nanotechnology. Discussion of "minimum required characterization" for nanomaterial has continued over the last few years, and interlaboratory testing has been initiated to quantify the accuracy of current measurement methods and protocols. However, because of a previous lack of adequate characterization of nanomaterial, especially with respect to its structure, polydispersity, purity, stability and lot-to-lot variability, the amount of publicly available data that can be used to correlate the structure of a nanomaterial with its activity is small. Furthermore, there remains a lack of:

- reference materials to establish measurement repeatability,
- standard protocols with controls for interference by the nanomaterial, and
- measurements of the sensitivity of a nanomaterial's effects in realistic biological environments to variation in composition and environmental and test conditions.

Without such data concerning the linkage between the structure of nanomaterial and its activity, rational design of nanotechnology-enabled products is not possible.

The most urgent need for a nanoinformatics infrastructure is therefore to collect, curate, annotate, organize and archive the available data. There must be a common understanding of the characterization required to provide sufficient knowledge of both a material's structure and its activity in different biological environments, cell lines and animal models, interlaboratory testing to quantify the error, uncertainty and sensitivity of the data, and study materials to support those tests as well as instrument and method calibration. In addition to archiving the data, expert annotations and analysis regarding its quality and extent of validity, the infrastructure should allow for a federated system of public/private databases with adequate, layered access control to allow aggregation among public and private data where possible. For example, in nanomedicine this data would include the results of pre-clinical and (blinded) clinical trails to allow for establishing correlations of benefit and risk across populations and sub-populations.

Semantic Search

The second requirement for the informatics infrastructure is to provide a semantically rich mechanism to search for and retrieve data within a set of federated databases and systems. This implies methods to map among the keywords, vocabularies, taxonomies and ontologies describing the metadata employed by each of the database systems as well as advanced methods that allow for building logically linked, range-limited searches using valid values for nanomaterial properties and measurement data, and including curated annotation and comment. Ontology development for different namespaces is therefore a critical requirement. A key capability to specify nanomaterial structural motifs of interest would provide the capability to discern correlations among nanoparticle structures and functionalizations and their activity in specific environments. A nanomaterial registry would be mandatory to begin to establish the link between metadata descriptions and the actual nanomaterial tested, the effect of differences in similar structures on their activity, the lot-to-lot variation in manufacture and resultant difference in test results, and the establishment of unique identifiers for study material and standard reference material. The nanoinformatics system should also aid in establishing nanomaterial repositories to be used in storing and delivering study material for interlaboratory tests, in conducting those tests and analyzing their results. Again, in the example of nanomedicine, searches based on patient genetic and diagnostic information would enable the development of personalized, targeted treatments and a reduction in undesirable side-effects.

Modeling and Simulation

The third requirement is for an infrastructure to support modeling and simulation. A predictive capability based on the nanomaterial's structure and the underlying physical, chemical and biological mechanisms of its interaction with its environment provides methods to test hypotheses and to advance both the science and the technology. Without that capability we have at best good observation, but no method other than testing to guide development. Previous examples of scientific and technology development clearly indicate that combining characterization and test with modeling and simulation accelerates the development cycle and translation to market. There is no indication that nanotechnology is an exception to that paradigm. What is need, therefore, are freely available structural models for nanomaterial as well as free, open-source methods for model sharing and development. The current cycle for duplicating a modeling result reported through the literature is measured in years: sharing code establishes collaborations in days. Again, to use the field of nanomedicine, the development of small-molecule drugs accelerated only when the basic structural motifs for those molecules could be identified and correlated with their effects so that modeling of mechanism could be used to augment test. The required nanoinformatics infrastructure must support such collaborative efforts.as well as links to required computational resources.

Communities of Interest

By satisfying these three basic requirements for the nanoinformatics infrastructure we can establish the larger communities of interest in different applications of nanotechnology. If we continue to progress application by application, agency by agency, and institution by institution we will necessarily realize slow progress. But if we establish viable communities of interest – by definition, each participant actively supporting their collaborative ventures – then we will accrue great benefit. In addition to more rapid progress due to more facile teaming, access to extensive resources, and shared expertise, we will be able to benefit from shared results and expertise among related collaborations. For example, partnering in instrument development would make available both broader markets and more specific requirements by linking requirements for research with those for nanomaterial development, for quality assurance in large scale manufacture, and for regulatory and field testing, thereby providing longer term planning for increasing market size with maturing capability.

Existing Capability and Pilots

In the following sections examples of existing capability and promising pilots are mentioned. In addition, there have been attempts at establishing different communities of interest in these project areas. They are referenced in connection with ongoing or evolving projects rather than being listed separately.

Data

Definition: In this document "characterization" includes physico-chemical, in vitro and in vivo testing to determine the structure and properties of nanomaterials and their effects in relevant biological environments.

Most existing public nanomaterial databases gather, organize and archive published papers on nanomaterial characterization or references to those papers. Although publications provide an overview of an experiment, analysis or computation, they rarely provide sufficient data and information to allow rapid duplication of the published results. This is particularly true in nanotechnology: for example, nanoparticles are generally both polymorphic and polydisperse, and detailed knowledge of the sensitivity of the results of a characterization on the variation in structure of the subpopulations is generally not reported. Instead, results are frequently attributed to an ideal, monodisperse structure. Finally, there currently exists no generally accepted mechanism for sharing proprietary nanomaterial data. As a result, the raw data and supplementary data that would be so useful in evaluating error, uncertainty and sensitivity are lacking.

There have been some efforts made to rectify these shortcomings in the available data.

Standard Protocols

Standard measurement protocols for nanomaterials are now being becoming available through the standards development organizations (SDOs). Standards have been completed and are under development at both ATSM and ISO. Although many laboratories have developed protocols which include controls to test for interference by different classes of nanomaterials, there is generally little incentive for providing the resources necessary to turn those protocols into a viable standard. NCI's Nanotechnology Characterization Laboratory (NCL) is one of the exceptions to that rule: NCL has published its protocols on its website and many of the developing standards at ISO and ASTM are based on those protocols. However, the resources available to develop standard protocols are limited. Although the pool of experts in standard development is generally shared among the different SDOs, collaborative standard development is hindered by competition among the SDOs. A pilot effort has been proposed to establish a collaborative "pre-standards" protocol development using shared electronic tools, but that concept has been embraced solely by ASTM at present.

Interlaboratory tests

The utility of a given standard characterization protocol is limited if the error and uncertainty associated with the result in unknown. The determination of the error and uncertainty of the test results for a new standard are provided by conducting an interlaboratory study (ILS) through a sponsoring organization, although ASTM provides the management and financial resources for an ILS for any of the standard protocols it develops. A new organization, the International Alliance for Harmonization in Nanotechnology (IANH), has recently been formed to close the gap in sponsorship of ILS and to provide a quantitative measure of the reliability and repeatability of standard protocols. Several ILS studies are currently underway through sponsorship by the IANH, ASTM, and other participating organizations.

Reference materials

At present it is difficult (in some instances impossible) to correlate test results found in the literature on similar nanomaterials. Any given type of nanomaterial may be manufactured through a number of different procedures involving different precursors, reagents and excipients, producing different products and contaminants. Even when a given procedure is repeated by the same hands, results may vary because of uncontrolled environmental conditions, variability in reagents, reactivities and timings, as well as instrument drift and calibration error. As a result, a given nanomaterial may have a very significant lot-to-lot variability even when produced by the same expert. In addition, material stability and storage conditions may increase its variability. For that reason large lots of reference materials are necessary for use in interlaboratory studies to ensure that, as nearly as possible, all tests are performed using aliquots of the same material. This is true for determining the physical and chemical properties of the materials themselves, but even more so for in vitro or in vivo tests which also must contend with a large variability in cell lines and animal models. As a result there is a great need for certified reference materials, standard reference materials and study materials to provide instrument calibration, protocol controls and reference measurements to track similarities and differences among lots. Without the reference points, correlating results on the "same" nanomaterial produced by different labs can be extremely difficult.

Just two years ago NIST produced its first batches of standard reference nanomaterial, gold colloids with nominal sizes of 10, 30 and 60 nanometers. That material was used for the ASTM ILS studies concluded one year ago and is available at moderate cost for other laboratory and interlaboratory studies. The OECD has called for the initiation of similar interlaboratory studies using large batches of material, and some of those studies are currently underway under the sponsorship of different participating labs as well as by the IANH. It should be noted that these materials are in general obtained through commercial vendors of the materials in large lots, perhaps by blending several smaller lots. An initial characterization is performed on the materials by the providers.

Data Characterization Standards

To provide for some uniformity in how nanomaterial characterization is performed, several new pilot efforts are underway to standardize the number and types of protocols that should be performed on nanomaterials to establish some meaningful measure of the quality and reliability of published and private data. These efforts tend to follow the spirit of the MIAME (Minimum Information About a Microarray Experiment) standard by the MGED Society which "specifies all the information necessary to interpret the results of the experiment unambiguously and to potentially reproduce the experiment." For example the Min Char (Minimum Information for Nanomaterial Characterization) Initiative has published a suggested minimum list of parameters necessary to characterize nanomaterial as well as some supplementary considerations that should be considered for completeness- http://characterizationmatters.org/ . Other organizations such as the OECD and several SDOs are considering other recommendations for additional characterizations. However, there is not yet movement toward an overall classification scheme for different levels of characterization that would be useful in annotating nanomaterial data as to its degree of quality and reliability. Some of the underlying issues, such as aggregating information on the interference produced by certain nanomaterial types, sizes and functionalizations, and the effects of sample preparation are actively being considered, and have initiated other efforts in data characterization standards such as that of caBIG's Nanotechnology Working Group. In particular ONAMI is developing a Nanomaterial-Biological Interaction Knowledgebase to aid in interpretation of the effects of nanomaterial exposures as well as innovative rapid in vivo assessments of potential toxicity at multiple levels of biological organization (molecular, cellular, system and organism) using embryonic zebrafish.

The "molecular" structure of nanomaterials

While structural motifs such as alpha helices and beta sheets and strands have provided insight into protein folding, function, and interaction, there exists no similar body of knowledge concerning structural motifs for nanomaterial and their interactions with other molecules and assemblies. Furthermore, there currently is no recognized repository for nanomaterial structures similar to the PDB for proteins. In addition, real nanomaterials are polydisperse and polymorphic, creating a further complexity due to the need to develop structures for each of the material's subpopulations. Finally, nanomaterial is frequently functionalized through the attachment of ligands, other molecules such as drugs or antibodies, or even other nanoparticles. Because of this inherent complexity, nanotechnology currently does not have a structural database to serve as a common resource and focus for research on the biological interactions of nanomaterial. Over the past few years Raul Cachau of the NCI's Advanced Biomedical Computing Center and his collaborators at the University of Talca in Chile, Bowie State University in Maryland have developed a pilot database for nanoparticles which an annotation facility to aid in the identification of structural motifs in nanomaterial and elucidation of their role in biological interactions. The pilot database is part of the Collaboratory for Structural Nanobiology (CSN) (aka. The Linnaeus Project) and serves as a focus for several projects in nanoinformatics and modeling discussed below. The CSN also incorporates a wiki to permit rapid annotation of the structures and to facilitate collaboration, an ISBN number to ensure permanent archiving of material developed on the site, and a capacity for storing and sharing computer models.

Federation of Databases in Nanotechnology

An important area of activity involving new pilots is that of database federation. Although currently nanotechnology databases cite and link to each other, only recently have there been discussions of federating different databases to permit common searches through all the databases in a federation. Such federation would provide more than ease of search: with sufficient capability to safeguard proprietary data, federation would permit searches over both public and private data in a controlled manner, greatly expanding the amount of data available, and creating a mechanism for expert annotation and curation of data at its source. This is particularly important due to the fact that most nanomaterial data is available through publications and that it will be otherwise difficult to annotate this data with regard to its quality and reliability (as discussed above) lacking such a mechanism. The topic has been discussed at several recent workshops including the 2008 NanoHealth Enterprise Workshop (now the NanoHealth and Safety Enterprise) and the October 10, 2008 NIST Nanoinformatics Workshop, as well as ongoing work in caBIG;s Nanotechnology Working Group. Databases currently being considered for a pilot federation include NCI's caNanoLAB, the CSN, ONAMI, NNI's NanoHUB (http://nanohub.org/) and the National Nanomanufacturing Network (http://www2a.cdc.gov/niosh-nil/), and NIOSH's Nanoparticle Information Library (http://www2a.cdc.gov/niosh-nil/) with possible participation by both the EPA and FDA. Links to ICON (http://www.goodnanoguide.org/tiki-index.php?page=HomePage) and Nanowerk (http://www.nanowerk.com/phpscripts/n_dbsearch.php) are also being discussed.

Advanced Instrumentation

It is important to realize that there is a need for access to better instrumentation for characterization of nanomaterial and determination of its structure. Although NIH has traditionally provided access to US national facilities as Light and Neutron Sources necessary for diffraction studies, the usage of these facilities has been low. Raul Cachau (NCI, ABCC) has proposed a new effort to establish user centers to help scientists access these facilities by providing help with the detailed and sometime onerous demands required for sample preparation, remote operation and data archiving and analysis. Such resources could greatly increase the utility of these facilities for scientists involved in nanotechnology for the biological sciences and medicine. There is also a related need that has not yet resulted in a pilot application, and that is for collaborative development of new instrumentation for more detailed characterization of nanomaterial including ligand distributions. There are several very promising new methods which could aid biological; and medical applications of nanotechnology, especially low energy electron microscopy, terahertz spectroscopy, and single molecule spectroscopy. Collaborative development of such instrumentation between manufacturers and researchers could accelerate their availability. By developing comprehensive requirements for research, scale-up, full-scale manufacture, and regulation, manufacturers would be aware of the extent of the market for this instrumentation and the differing requirements for each stage of development while the field could benefit from early testing to advance promising research and applications more quickly

Semantic Search

A semantic search capability offers very significant advantages over keyword search. Because search terms are not only defined but also have defined relations to other search terms, the number of false returns is greatly reduced. Furthermore, since valid values are also defined, searches such as "pegylated gold nanoparticles with size between 20 and 60 nm" become possible. Although the infrastructure required to institute such capability is significant, the development of ontologies for nanomaterial is proceeding with several notable examples. As the Semantic Web comes closer to realization, "namespaces" consisting of practitioners with identifiable disciplines (e.g. toxicologists, cell biologists, ...) will begin to formulate their own ontologies which could be adapted freely. Indeed, nanomaterial ontologies have already reached some maturity with the Nanoparticle Ontology (NPO) at Washington University, caNanoLAB's ontology, and the Japanese Nanomateria Platform ontology and are being used in conjunction with pilot databases. These activities take advantage of existing expertise and organizational bodies such as Open Biomedical Ontologies (OBO) as well as new tools for mapping among ontologies such as Biomed GT. Other activities are also underway such as ASTM and ISO development of standard ontologies and Norway's new Ontolution Nano Project. Finally, the need for a nanomaterial registry to aid in relating specific nanomaterial lots to the ontology terms describing their composition and structure is now recognized with NIBIB issuing a new RFA for development of a nanomaterial registry.

Modeling and Simulation

Support for modeling and simulation in nanotechnology has largely been proceeding piecemeal with separate efforts within different agencies and institutions, The prime major exception to that rule is NanoHUB which hosts selected applications and tools for use on demand throughout the nanotechnology community and which has seen continually increasing usage of their resources. Another is the previously mentioned CSN which hosts over 150 structural nanoparticle models currently that are freely available for use in modeling and simulation codes. The CSN will also make available computer codes, run parameters and validation suites for open source codes in the future, together with the wiki forum for collaborative development of tools and software. The EPA's National Center for Computational Toxicology - http://www.epa.gov/ncct/ - provides relevant models but is not currently nano-related. The ACTION-Grid, a EU 7th Framework project involving nanoinformatics, personalized medicine and grid computation, is a recent promising addition in this area. As mentioned above, new pilot activities are currently being discussed for modeling and simulation activities ranging from structure /activity relationships through quantum-based first principles applications and models for exposure and risk.