caNanoLab: data sharing to expedite the use of nanotechnology in biomedicine

caNanoLab: data sharing to expedite the use of nanotechnology in biomedicine

Sharon Gaheen[1]**, George W Hinkal**[2]**, Stephanie A Morris**[2]**, Michal Lijowski**[3]**, Mervi Heiskanen**[4]** and Juli D Klemm**[4,5]

[1] Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA

[2] Office of Cancer Nanotechnology Research, Center for Strategic Scientific Initiatives, National Cancer Institute, NIH, Bethesda, MD 20892, USA

[3] Essential Software, Inc., Potomac, MD 20854, USA

[4] Center for Biomedical Informatics and Information Technology, National Cancer Institute, NIH, Rockville, MD 20850, USA

E-mail: Juli.Klemm@nih.gov

Abstract. The use of nanotechnology in biomedicine involves the engineering of nanomaterials to act as therapeutic carriers, targeting agents and diagnostic imaging devices. The application of nanotechnology in cancer aims to transform early detection, targeted therapeutics and cancer prevention and control. To assist in expediting and validating the use of nanomaterials in biomedicine, the National Cancer Institute (NCI) Center for Biomedical Informatics and Information Technology, in collaboration with the NCI Alliance for Nanotechnology in Cancer (Alliance), has developed a data sharing portal called caNanoLab. caNanoLab provides access to experimental and literature curated data from the NCI Nanotechnology Characterization Laboratory, the Alliance and the greater cancer nanotechnology community.

Contents

[5] Author to whom any correspondence should be addressed.

1. Introduction

Progress in nanomedicine depends on making discoveries, integrating this new information into existing hypotheses, formulating new hypotheses and continuing experimentation to test and extend our knowledge. Nanomedicine is inherently interdisciplinary; as such, this process works most efficiently when researchers can integrate their results with those obtained from other groups and disciplines. Progress in the field has been impeded by the lack of a knowledge-management infrastructure as well as the lack of standards to describe the complexity of nanoparticles and their highly diverse nature. Having such standards would enable comparing and combining of results within and across disciplines.

Nanoparticle-based vehicles come in a wide variety of physical structures and chemical compositions, each with inherent dispersion. This polydispersity of nanoparticles leads to diverse effects on living organisms, and can impede the development and implementation of nanotherapies if not adequately characterized. Providing researchers with access to the nanoparticle characterization data and methodologies, as well as the raw data of subsequent experiments will expedite the use of nanoparticles in biomedicine. Informatics components essential to nanotechnology data-sharing include both terminology standardization and data handling. A functional informatics framework promotes interdisciplinary communication, allows data and protocol storage and facilitates search, retrieval and modeling of data output.

To address the challenges of data sharing, efforts are now underway by the National Cancer Institute (NCI) and collaborating organizations to define standards for representing nanoparticles and their characterizations via the establishment of a Nanotechnology Working Group (Nano WG) and the development of nanoinformatics resources, such as the cancer Nanotechnology Laboratory web portal (caNanoLab). The goal of caNanoLab is to provide a resource where primary nanotechnology research data are no longer disparate islands affiliated with their originators, but standardized and shared across the scientific and clinical community.

2. Use cases for data sharing

The need for data sharing and the complexities involved in the development of data standards for the application of nanotechnology in biomedicine is exemplified in several scientific scenarios developed by the Nano WG (Nano Scenarios). From these scenarios, a high-level use case diagram was derived representing the use cases associated with nanotechnology in biomedicine (figure 1).

In this high-level use case diagram, a nanomaterial scientist initiates a study involving nanomaterials and creates the nanomaterial model. The nanomaterial model consists of the nanomaterial itself and any functionalizing materials such as therapeutic, targeting and/or imaging agents. Once the scientist creates the model, the model is synthesized and characterized. Depending on the study, the scientist may perform physico-chemical, *in vitro* and *in vivo* characterization of the nanomaterial formulation, the nanomaterial itself and the functionalizing material (e.g. drug). The scientist then performs comparative analysis on the characterization results and publishes the research findings. Based on the results, the scientist may apply for an Investigational New Drug, work with a clinician to conduct clinical trials and work with a manufacturing laboratory to manufacture the nanomaterial formulation.

caNanoLab was developed to address several of these high-level use cases by supporting the creation of exhaustively described nanomaterials, the identification of standard protocols, the capture of data from diverse characterization studies and the publication of findings. Additionally, caNanoLab was designed as a resource with the potential to be interoperable with other nanotechnology, pre-clinical, clinical and imaging resources.

3. National Cancer Institute Alliance for Nanotechnology in Cancer

The NCI Alliance for Nanotechnology in Cancer (Alliance), initiated in 2005, is NIH's flagship biomedical nanotechnology program supported by the NCI Office of Cancer Nanotechnology Research (OCNR) (figure 2). Over two 5-year phases, the Alliance has developed into a large research and training network that leverages techniques, equipment, personnel and information across many institutions to strengthen their capacities
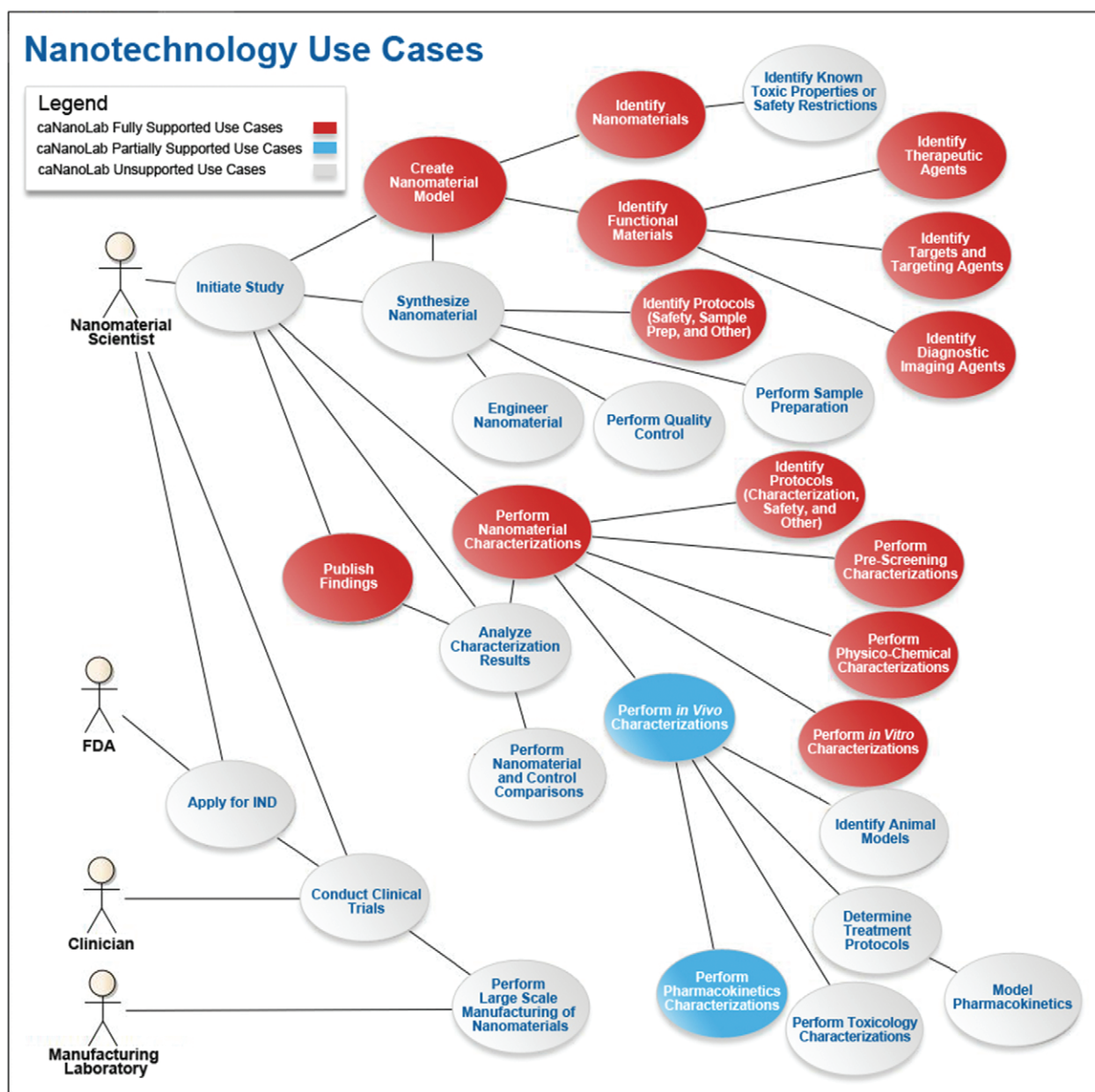
Figure 1. High-level biomedical nanotechnology use case diagram.

beyond that of a single entity. To enable the sharing of diverse nanomaterial data, the OCNR has supported the development of caNanoLab with input from the Alliance network.

The research pillars of the Alliance are the nine Centers of Cancer Nanotechnology Excellence (CCNEs). The CCNEs are charged with completing pre-clinical work and securing resources for further translation into clinical practice. This includes an expectation of close collaboration with industry. In addition, the CCNEs are meant to be models for interdisciplinary, transformative biomedical research groups where the scientific and social interactions engendered by the center-based structure lead to greater innovation than comparable individual awards. In addition to the CCNEs, the OCNR currently funds 25 other awards in the Alliance network: 12 Cancer Nanotechnology Platform Partnerships, 6 R25-Cancer Nanotechnology Training Centers and 7 K99/R00 Pathway to Independence awards [1] Additionally, the OCNR supports, in partnership with the National Institute of Standards and Technology and the United States Food and Drug Administration, the Nanotechnology Characterization Laboratory (NCL) at Frederick National Laboratory for Cancer Research.
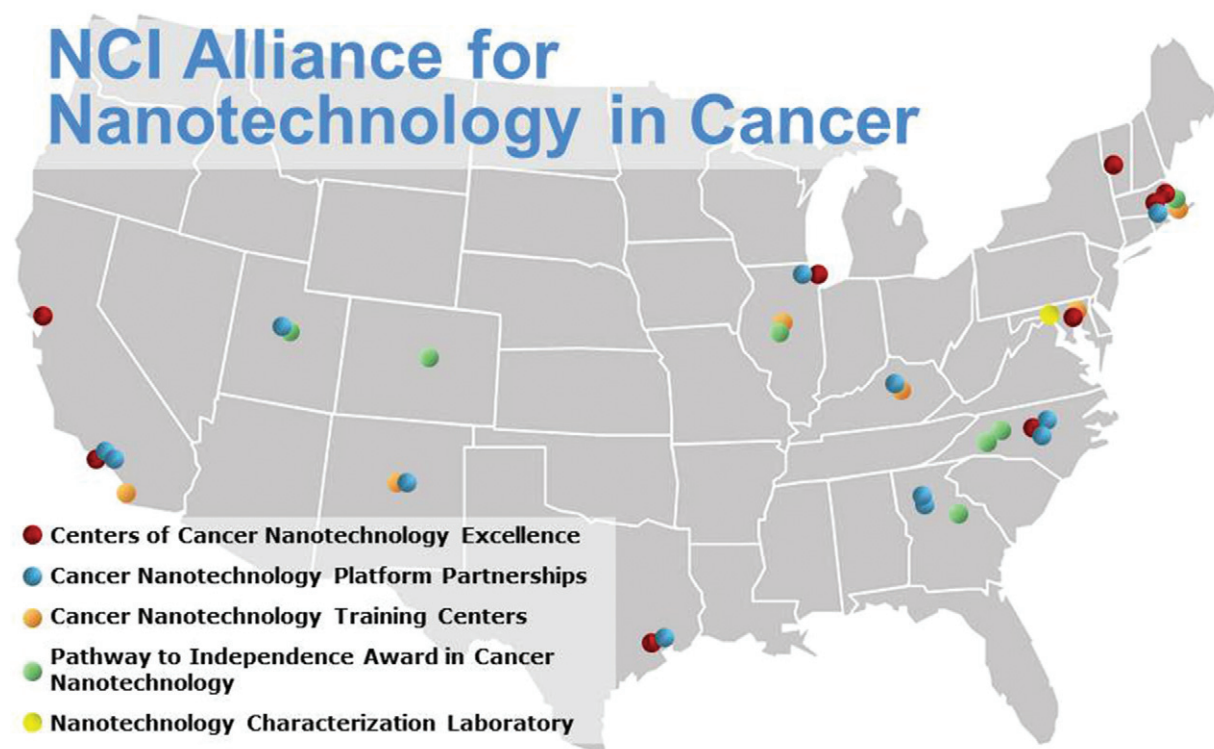
Figure 2. Map of institutions currently involved in the NCI Alliance.

The NCL performs and standardizes the pre-clinical characterization of nanomaterials intended for cancer therapeutics and diagnostics developed by researchers from academia, government and industry. In particular, the NCL was pivotal in the design of the caNanoLab characterization data elements.

The OCNR strongly believes in the benefits of leveraging knowledge, techniques, instrumentation and information across the nanotechnology research community. Collaboration across the Alliance network is encouraged by an internal funding mechanism and facilitated networking opportunities. To manage the varied and significant amount of data expected of a multidisciplinary network, the CCNEs were encouraged to develop core facilities dedicated to bioinformatics and/or nanoinformatics. In addition to internal data management, these cores would drive the input of data from CCNE projects into caNanoLab. This would yield publicly available datasets that are fully characterized (i.e. physicochemical properties, *in vitro* biological properties and *in vivo* biocompatibility) and in a format that could be openly managed and interrogated by other informatics professionals. In support of this goal, the OCNR has sponsored the curation of data from publications into caNanoLab (see Data Curation and Archiving section, below). This effort prioritizes cutting edge publications and shared, unpublished datasets from the Alliance and the greater cancer nanotechnology community. Curation entails the extraction and aggregation of data from publications, supplementary information and author responses to capture the details of the planning, execution and output of all experiments involved in the development of nanotechnology in biomedicine, i.e. cancer. This coordinated effort has significantly enlarged the caNanoLab dataset and drawn the interest of other groups toward contributing and collaborating.

4. caNanoLab

The caNanoLab collaborative environment promotes data sharing and analysis across the cancer nanotechnology community to expedite and validate its use in biomedicine. caNanoLab is a web-based portal (figure 3) and data repository that allows researchers to submit and retrieve information on nanoparticles including their composition, function (e.g. therapeutic, targeting, diagnostic imaging), physical (e.g. size,

Figure 3. caNanoLab web portal, https://cananolab.nci.nih.gov/.

molecular weight) and *in vitro* (e.g. cytotoxicity, immunotoxicity) experimental characterizations. A given nanoparticle's entry also includes information on the protocols used for these characterizations and any related publications. Web-based forms are available to facilitate data submission and submitters can customize the visibility of their data to range from private, to share with specified collaboration groups, to fully public.

caNanoLab can also be used for discovery purposes through searching the results of all the publicly available physical and *in vitro* characterizations as well as providing access to the associated publications. In addition to obtaining webform-based query results, researchers are also able to download reports in a spreadsheet-based format. Based on user IP address web statistics, most of the caNanoLab users are from the United States, but approximately 10% of all users are international including countries from Europe, Asia and South America. Moreover, programmatic access to the data is available through web services accessible on caGrid.

caNanoLab is based on a nanotechnology information object model (nano-OM) which provides a standard representation of nanomaterial concepts and their physical (e.g. size, molecular weight) and *in vitro* (e.g. cytotoxicity, immunotoxicity) characterizations (figure 4). The nano-OM uses and extends concepts from the NCI Thesaurus and the NanoParticle Ontology (NPO). caNanoLab supports capture of parameters recommended by the Minimum Physical and Chemical Parameters for Characterizing Nanomaterials on Toxicology Initiative. Future features under consideration include support for standardized *in vivo* characterizations of nanomaterials and their functionalizing entities, which are analogous to those required for small molecules and other medical devices. These characterizations involve rigorous testing to determine toxicity and pharmacokinetics properties. Members of the caNanoLab project team are involved in the ISA-TAB-Nano effort, which enables the submission and exchange of information across nanotechnology resources by use of a defined data elements and vocabulary [2]. This involves input and support across the international informatics community. Future features of caNanoLab may also include support for the validation, import and export of ISA-TAB-Nano files using customized ISA-TAB tools.

5
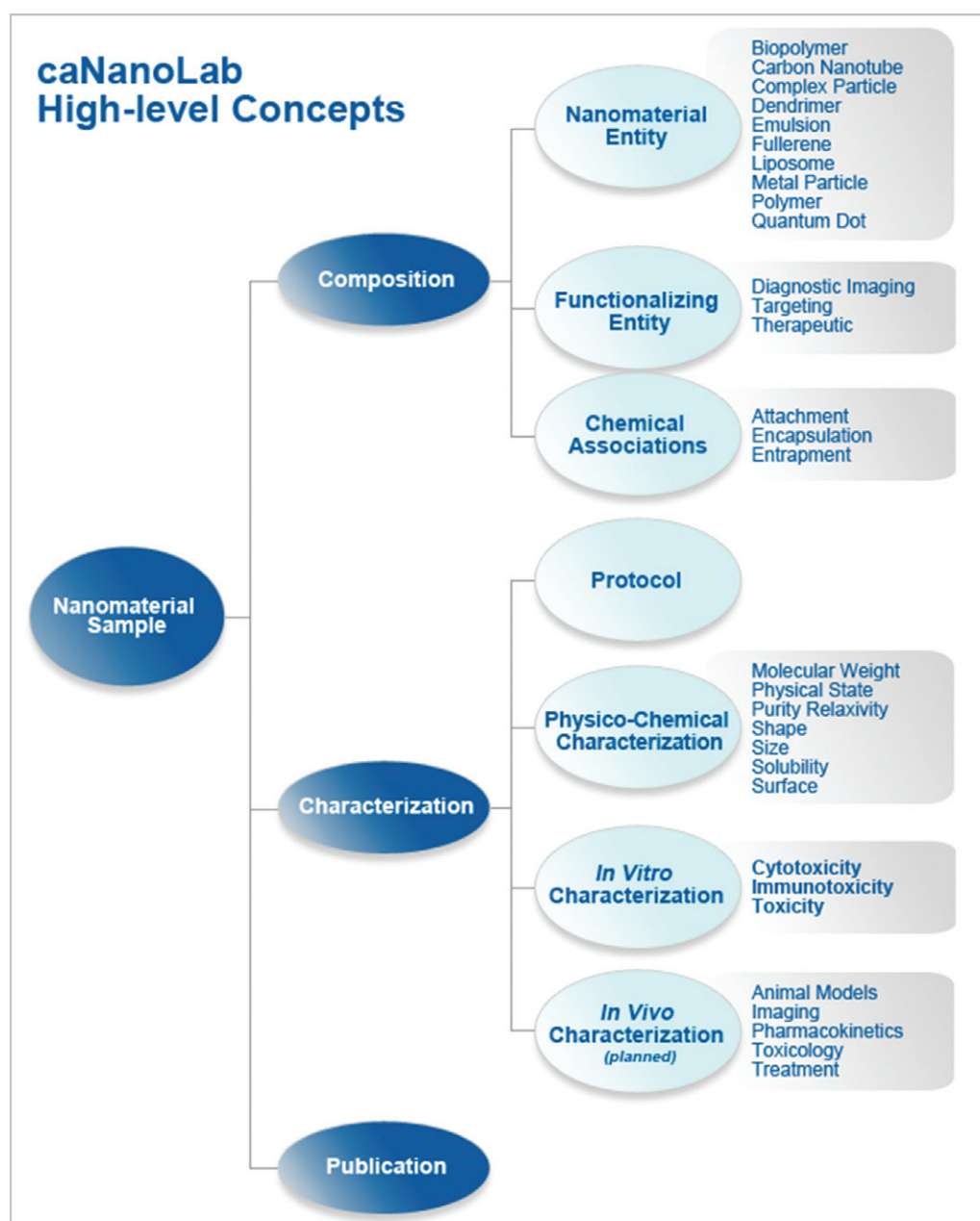
Figure 4. caNanoLab high-level concepts.

Currently, caNanoLab is deployed at the NCI as a resource supporting the use of nanotechnology in the biomedical community. caNanoLab compliments other nanotechnology information resources such as the Nanomaterial Registry, the Nanomaterial-Biological Interactions Knowledgebase (NBI), the National Nanomanufacturing Network InterNano, the National Institute of Occupational Safety and Health Nanoparticle Information Library (NIL) and the University of North Carolina at Chapel Hill ChemBench (table 1).

While the NCI-hosted instance of caNanoLab is broadly available for public use, the caNanoLab software is also available for download and installation at local institutions. caNanoLab software is open source and the code is being deposited to the National Cancer Informatics Program channel in the GitHub code repository to support open, community lead development. The goal is to encourage both end-users and developers to coalesce around caNanoLab to add new features and functionality.

Table 1. Comparison of nanotechnology resources.

Feature/resource	caNanoLab[a]	Nanomaterial Registry[b]	NBI[c]	InterNano[d]	NIL[e]	ChemBench[f]
Detailed curated formulation and characterization information	X	X	X		X	
Sources from primary literature	X	X	X		X	
Sources from other databases		X				X
Access to supporting protocols	X	X		X		
Access to supporting literature	X	X		X	X	
Community data submission and sharing through collaboration groups	X					
Data comparison tools for formulation, and biological/environmental implications		X	X			
Describes degree of nanoparticle characterization	X	X				
Access to nanomaterial biological interactions	X	X	X			X
Intended for biomedical community	X	X	X		X	X
Intended for nanotechnology community	X	X	X	X	X	
Intended for environmental community		X	X		X	X
Primary source of nanomaterial safety information					X	
Enables the development and application of quantitative nanostructure-activity relationships			X			X

[a] cananolab.nci.nih.gov/caNanoLab/
[b] www.nanomaterialregistry.org
[c] nbi.oregonstate.edu
[d] www.internano.org
[e] nanoparticlelibrary.net/index.asp
[f] chembench.mml.unc.edu

5. Data curation and archiving

The objective of the NCI OCNR-sponsored data curation project is to provide cancer nanotechnology data using standardized data elements and vocabularies to expedite the re-use of these data to support the development of novel agents for cancer diagnostics and therapeutics. The NCI instance of caNanoLab currently has 992 samples and 37 protocols in the public domain. Data curation is focused on papers published by NCI Alliance investigators (table 2), with some works produced by the broader international cancer nanotechnology community. The caNanoLab data curation team meets on a weekly basis to review progress and priorities, and maintains an active to-do list of at least three publications. Data submission is done using terminology from NPO and other ontology sources as appropriate.

The curation of nanotechnology information is accomplished by selecting relevant publications, manually extracting reported text and data and submitting extracted information in caNanoLab. The primary curation steps are as follows:

1. The NCI OCNR program staff provides a list of publications for curation. The criteria for determining whether a publication should be curated includes whether the publication is meaningful to the cancer nanotechnology field (cutting edge science); whether associated meaningful data are available in the

Table 2. Summary of data available in caNanoLab.

Nanomaterial type	Nanomaterial entities	Description
Biopolymer	43	A polymer formed by a living organism
Carbon black	2	A material produced by the incomplete combustion of carbon-rich organic fuels in low oxygen conditions
Carbon nanotube	50	A nanotube comprised of one or more graphite sheets (graphene) of hexagonal arrays of carbon rolled into seamless cylinders with capped ends
Carbon particle	1	An amorphous nanopowder formed by laser techniques
Dendrimer	74	A polymeric molecule that has a highly branched, three-dimensional tree-like architecture, synthesized with monomers where shells of branched molecules are added in discrete steps to a central core
Emulsion	88	A colloid in which both phases are liquids that are immiscible with each other
Fullerene	16	Any cagelike, hollow molecule composed of hexagonal and/or pentagonal groups of carbon atoms
Liposome	34	A supramolecular structure, which is a closed vesicle that forms on hydration of dry phospholipids above its transition temperature
Metal oxide	186	A nanomaterial composed of a metal oxide
Metal	132	A nanomaterial composed of a metal
Metalloid	36	A nanomaterial with properties between a metal or non-metal
Nanohorn	7	A single-walled carbon nanostructure with anirregular horn-like shape
Nanorod	33	A nanoscale rod composed of either metallic or semiconductor material or a mixture of both
Nanoshell	1	A three-dimensional nanostructure that is composed of a spherical core surrounded by a few nanometers in thickness. If the shell is made of metal, then it is called a metallic nanoshell
Polymer	188	A nanomaterial composed of single or multiple monomers
Quantum dot	73	A nanometer size fragment of semiconductor material, whose excitons (electron–hole pairs) are confined in three spatial dimensions
Silica	43	A nanomaterial composed of a silicon oxide

publication or from the investigator and whether the data are complete (e.g. contains material composition details and linkage information).

2. Data are extracted from the publication and samples are associated with characterizations based on information provided in text, tables, figures and figure legends.

3. Extracted information and numerical data are entered into caNanoLab.

4. The publication authors are contacted to request additional data to augment data curated from publications. These additional data are also entered into caNanoLab.

Curation of data from publications requires significant effort and domain expertise, thus it is almost always necessary to contact the authors for additional information. However, once the information has been carefully curated and entered in caNanoLab, it is easily accessible and available for re-use by both researchers and other data providers.

ISA-TAB-Nano documents are also created for datasets submitted to caNanoLab. This standard specifies the format for representing and sharing information about nanomaterials, small molecules and biological specimens along with their assay characterization data using spreadsheet or tab-delimited files [2, 3]. It is part of the ISA Commons community that uses the ISA metadata tracking framework to facilitate standards-compliant collection, curation, management and reuse of datasets. ISA-TAB-Nano files representing studies listed in table 2 are available for download from the ISA-TAB-Nano Wiki.

6. Conclusions and future directions

A key challenge in biomedical research is making the primary data supporting publications available to and re-usable by the research community. NCI is working to address this challenge in the field of cancer nanotechnology research by curating nanotechnology data from publications and submitting these data to caNanoLab. This manual, expert-driven approach to information extraction has led to the creation of a high quality, readily searchable and computable repository of nanoparticle characterization data suitable to support the structure–activity relationship analysis that is a current focus of the Nano WG. In the future, the caNanoLab team will be participating in inter-agency collaborations through the Nano WG toward the identification of information needed for obtaining regulatory approval on the use of nanotechnology in biomedicine.

It is important to emphasize that extracting data through the curation of publications is rather inefficient, as the publications are not written with interchangeable information extraction in mind and often do not include sufficient information for a detailed understanding of nanotechnology experiments described therein. Going forward, the caNanoLab team aims to emphasize policies and resources that promote and incentivize standards-based data capture directly by the data producers. There are many efforts underway to encourage primary data sharing in the scientific community (e.g. www.fged.org, www.force11.org, biosharing.org) and the team aims to adopt and support the best practices of these communities. The team is also working together with the ISA community (isacommons.org) to extend the ISA Tools software suite to support the nanotechnology data extensions to ISA-TAB (ISA-TAB-Nano). The goal of these efforts is to make it easier to share nanotechnology data among different data resources in a standards based manner. The caNanoLab team will continue to promote data sharing in the nanotechnology community, and is planning to continue curation of data in caNanoLab and providing the research community with access to comprehensive, high quality nanotechnology data sets.

Acknowledgments

References

[1] Hinkal G W, Farrell D, Hook S S, Panaro N J, Ptak K and Grodzinski P 2011 Cancer therapy through nanomedicine *IEEE Nanotechnol. Mag.* **5** 6–12

[2] Thomas D G *et al* 2013 ISA-TAB-Nano: a specification for sharing nanomaterial research data in spreadsheet-based format *BMC Biotechnol.* **13** 2

[3] Baker N A, Klemm J D, Harper S L, Gaheen S, Heiskanen M, Rocca-Serra P and Sansone S A 2013 Standardizing data *Nature Nanotechnol.* **8** 73–4