



Published in final edited form as:

J Biomed Inform. 2011 February ; 44(1): 59–74. doi:10.1016/j.jbi.2010.03.001.

NanoParticle Ontology for Cancer Nanotechnology Research

Dennis G. Thomas^{a,b}, Rohit V. Pappu^{b,c}, and Nathan A. Baker^{a,b,*}

^aDepartment of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110 USA

^bCenter for Computational Biology, Washington University in St. Louis, St. Louis, Missouri 63110 USA

^cDepartment of Biomedical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130 USA

Abstract

Data generated from cancer nanotechnology research are so diverse and large in volume that it is difficult to share and efficiently use them without informatics tools. In particular, ontologies that provide a unifying knowledge framework for annotating the data are required to facilitate the semantic integration, knowledge-based searching, unambiguous interpretation, mining and inferencing of the data using informatics methods. In this paper, we discuss the design and development of NanoParticle Ontology (NPO), which is developed within the framework of the Basic Formal Ontology (BFO), and implemented in the Ontology Web Language (OWL) using well-defined ontology design principles. The NPO was developed to represent knowledge underlying the preparation, chemical composition, and characterization of nanomaterials involved in cancer research. Public releases of the NPO are available through BioPortal website, maintained by the National Center for Biomedical Ontology. Mechanisms for editorial and governance processes are being developed for the maintenance, review, and growth of the NPO.

Keywords

ontology; cancer; nanotechnology; nanoparticle; nanomaterial; informatics; NPO; BFO; annotation; BioPortal

1. Introduction

Large and diverse types of data are being generated from experiments in *cancer nanotechnology research* – an intrinsically interdisciplinary field of research devoted to the development and application of nanotechnology-based methods in the treatment, diagnosis, and detection of cancer. To efficiently use these data, the cancer research community requires informatics methods that will help researchers to search, access, and analyze nanomedicine data, and thereby facilitate the realization of nanotechnology applications in personalized treatment methods. In particular, ontologies, which consist of a common vocabulary and logical (information) structure are necessary components of informatics because they provide the knowledge framework for scientific discourse, and for the annotation, semantic integration, knowledge-based searching, mining, inferencing and

*Corresponding author. Address: 700 S. Euclid Ave., Room 113, Campus Box 8036, St. Louis, MO 63110. Fax: 1-314-246-9166. baker@biochem.wustl.edu (N.A. Baker).

unambiguous interpretation of data [1]. In this work, we develop an ontology that represents the knowledge domain of cancer nanotechnology.

1.1. Ontologies and their uses

An ontology is a formal, explicit representation of knowledge belonging to a subject area: the knowledge is encoded and represented as multiple hierarchies of terms (or classes) that are described using attributes (e.g., preferred name, definition, synonyms, etc.), related to each other using associative relations (e.g., part_of, has_part, etc.), and may be formalized using logical axioms in a machine-interpretable language (e.g., Ontology Web Language (OWL)) [2]-[6]. Ontologies are used in informatics-supported collaborative research and databases in different ways. For example: (1) ontologies provide a common terminology that can be shared and reused among researchers; (2) ontologies are formally represented in a machine-readable language (e.g., OWL) which can be meaningfully interpreted by both subject domain experts as well as computers; (3) ontologies facilitate semantic sharing and integration of data stored in disparate resources by providing the logical and semantic relationships between different parts of information contained in data; (4) ontologies provide the logical structure for performing knowledge-based searches along with informatics tools and software in order to speed up the access to, retrieval of, and analysis of data for enabling knowledge discovery.

In addition to ontologies, there are controlled vocabularies (CVs) that serve as terminology sources. A controlled vocabulary (CV) provides a list of terms (organized in a hierarchy), textual descriptions of their meaning, and lexical terms corresponding to each entry [7]. Formally, vocabularies for informatics applications are developed and made available in the form of CVs and/or ontologies. Compared to ontologies, however, CVs are limited in their scope for informatics applications because they lack the expressiveness of ontologies in representing knowledge (an application of CVs is in indexing resources, such as records in a database).

1.2. Ontologies in biomedical research

In biomedical research, ontologies are used to represent the knowledge of a specific domain of interest in machine-processible form and to integrate experimental data that is annotated with terms from these ontologies. Biological and chemical ontologies have played significant roles in describing various characteristics of genes and gene products (e.g., Gene Ontology or GO [8]), and small molecular entities (e.g., Chemical Entities of Biological Interest or ChEBI [9]). For example, the Gene Ontology (GO) serves as a controlled vocabulary for describing gene and gene product attributes in any organism [8]. Annotation of genes using GO terms has provided ways to classify microarray experiments, thereby enhancing the analysis, querying, and mining of microarray gene expression data based on statistical methods [10]. Significant effort is being made to develop tools and methods to enable interoperability among different biomedical ontologies for integrative analysis of data from diverse experimental data sets in the biomedical knowledge-domain [11]. In this article, we will discuss the development of an ontology for representing knowledge underlying the diverse data arising from cancer nanotechnology research.

1.3. Importance of nanotechnology in cancer research

Nanotechnology involves the application of scientific knowledge from a variety of disciplines in science and engineering to understand, manipulate, and control the properties of matter at nanoscale (1-100 nm) size dimensions [12]. Nanotechnology holds tremendous potential for overcoming many of the problems that conventional methods face in the treatment, diagnosis and detection of cancer [13]. In particular, **nanoparticles** (nanoscale-sized particles) have been developed and investigated for **cancer diagnostics** and

therapeutics; these materials are hereafter referred to as NP-CDTs. Pre-clinical studies have shown that these NP-CDTs offer many advantages over small-molecule approaches. For example, NP-CDTs can ameliorate the problems of poor in vivo biodistribution and adverse side-effects associated with small-molecule agents (e.g., drugs, image contrast agents, etc.). These problems arise due to lack of specificity of the agents in targeting cancer cells and are due to a range of effects such as: rapid uptake by the reticuloendothelial system (RES); clearance by the macrophages in MPS (Mononuclear Phagocyte System) organs, if the intended target cells are not located in the MPS organs; and the presence of barriers (e.g., blood brain barrier (BBB)) [14]. Cancer cells share many features with normal cells, and therefore, agents lacking the desired target specificity will also target healthy normal cells and damage them, thereby, causing adverse side-effects in the body. In drug delivery, poor biodistribution of the drug can result in low drug concentration levels at the tumor site. These low concentration levels, in conjunction with dose-limiting toxic side effects, reduce the drug's overall therapeutic efficacy. NP-CDTs can increase the circulation times and efficacy of therapeutic and diagnostic agents [13] - [16]. Generally, circulation times increase if the agent of interest is attached to a small, hydrophilic nanoparticle [17]. While small sizes reduce the likelihood of uptake by the RES, hydrophilicity increases the overall solubility of the diagnostic/therapeutic agent. In addition, functionalized nanoparticles target specific receptors that are over expressed on surfaces of cancer cells, and this in turn facilitates the uptake of drug-loaded nanoparticles via endocytic pathways [13] - [16]. In the year 2005, the FDA (Food and Drug Administration) approved a paclitaxel-loaded albumin nanoparticle formulation for the treatment of metastatic breast cancer [18] – a positive advance for the therapeutic use of NP-CDTs. Additionally, several other NP-CDTs are being evaluated in clinical trials [19].

1.4. Informatics for advancing cancer nanotechnology research

Informatics methods are considered to be necessary tools for the advancement of cancer nanotechnology research [20]. A cursory evaluation of the literature reveals that NP-CDTs are diverse in nature and have a wide range of applications [21],[22]. This diversity is related to combinatorially large numbers of mechanisms by which the chemical composition of nanoparticles can be modified. Furthermore, small changes in the chemical composition can cause drastic changes in the physicochemical and functional/biological properties of NP-CDTs, further increasing the volume and diversity of nanomedicine datasets [23]. In fact, pre-clinical evaluation of every new NP-CDT formulation requires experimental characterizations to be done, and this in turn adds more volume and diversity to the nanomedicine datasets. As the variety of nanoparticles adapted for use as NP-CDTs has increased, the nanomedicine research community has experienced significant growth in NP-CDT data. While the volume of NP-CDT data may be smaller than the volume of genomic data, the complexity, richness, and direct therapeutic or diagnostic relevance of NP-CDT data leads to a combinatorial complexity that far exceeds genomic data. We propose that the richness of this data can be tapped to develop customized point-of-care treatment regimens based on NP-CDTs. Realization of this objective will require an integrative approach for culling data and drawing inferences that inform future *de novo* designs. A particularly relevant example is the “rescue” of drug candidates that have failed clinical trials. Often, failure is attributed to non-specific delivery and toxic side effects. With increased data sharing and integration, it should be possible to develop trials to “rescue” failed drugs by loading them onto or encapsulating them into targeted nanoparticles. It is expected that rational design of “rescue strategies” will need predictions obtained from data-driven models that have been tested and validated. It is desirable then to proceed rationally with reformulations of therapeutic or diagnostic agents using NP-CDTs. Informatics approaches are likely to be valuable for such reformulation efforts, especially if one has access to an

integrated resource that yields rich information regarding both the physicochemical and functional properties of NP-CDTs, as well as tumor physiological properties.

1.5. Ontologies in cancer nanotechnology informatics

Ontologies will play an important role in the emerging field of cancer nanotechnology informatics. Recent efforts have focused on the development of databases for storing cancer nanotechnology data, with the goal of catalyzing discovery by bringing together the key knowledge in nanotechnology into coalesced resources. One such resource is the caNanoLab project [24], which has been developed to store, search, and share data generated from a variety of characterization studies for nanomaterials (e.g., NP-CDTs) used in cancer nanotechnology research. However, to maximize utility, databases must be complemented by a common vocabulary (or “terminology”) that can facilitate data sharing and integration of nanotechnology databases with each other and with other cancer-related databases. A few terminologies for nanotechnology are available in controlled vocabularies like the NCI (National Cancer Institute) Thesaurus [25] and through the efforts of caNanoLab, ASTM (<http://www.astm.org>), ANSI (<http://www.ansi.org/>), the International Standards Organization (ISO; <http://www.iso.org>), and the present work. However, with the exception of the current work, most of these nanomedicine vocabulary efforts are at early stages of development. In developing nanomedicine vocabularies, one can also reuse other existing vocabulary sources to obtain terms related to cancer biology (cancer cell lines, tumor types, animal models, genes, gene products, post translational modifications, and signaling pathways) and cancer medicine. Relevant vocabularies for such terms include GO [8], ChEBI [9], NCI Thesaurus (NCIT) [25], Medical Subject Headings (MeSH) [26], SNOMED_CT (<http://www.ihtsdo.org/our-standards/snomed-ct/>), etc. However, ontologies and controlled vocabularies focused specifically on cancer nanotechnology have not been developed. In particular, ontologies – with their rich machine-interpretable logical structure – are needed to facilitate communication between cancer nanotechnology researchers from different scientific fields, to ensure semantic interoperability between applications and databases, and to provide the foundation for new analytical tools in nanomedicine research. In addition to using the ontology to provide terms and definitions to help annotate data, we require ontologies to represent knowledge that is specific to the cancer nanotechnology domain such that they can be used for semantic integration of different parts of data, knowledge-based searching and for drawing inferences, based on the logical classification of terms and relations between terms provided by the ontology. Consider the following scenario as an example: a chemist has synthesized a dextran-coated nanoparticle, but wants to make a rough prediction about its *in vitro* and *in vivo* properties. So, s/he plans to compare it with nanoparticles that have characterization data available in a database such as caNanoLab. To make the best predictions, the researcher must identify that nanoparticle which most closely correlates with the dextran-coated nanoparticle. For this, the researcher must know what descriptors to choose from for comparing the nanoparticles, and also know the optimal descriptors needed to help determine the type of nanoparticle that highly correlates with the dextran-coated nanoparticle. These descriptors can be provided by the ontology. At the simplest level, if the descriptor is type of coating material, then by classifying nanoparticles based on this type of coating material will help identify the highly correlated classes of nanoparticles that are either the sibling classes or child classes of dextran-coated nanoparticles. In this way, the researcher only needs to look at nanoparticle data annotated with the ontology classes, and to compare results of the different nanoparticles identified from the classification in the ontology. Better predictive models such as Structure Activity Relationship (SAR) models can be developed using data annotated and integrated by an optimized set of ontology-based descriptors for every case in question.

To address the need for ontologies in cancer nanotechnology informatics, we have developed an ontology called the **NanoParticle Ontology (NPO)**. The NPO has been designed to represent knowledge underlying the preparation, chemical composition, physicochemical and functional/biological characterization of nanomaterials (e.g., NP-CDTs), which are formulated and tested for applications in cancer diagnostics and therapeutics. The remainder of this manuscript details the development of the NPO.

2. Method

2.1. Initial collection of terms

We started the construction of NPO by creating an initial list of terms pertaining to a general description of NP-CDTs in the literature. In the scientific literature, one finds that nanoparticles have inherently very complex types of descriptions and therefore, require a wide range of classes that are defined and logically related in the NPO. The following examples illustrate the complexity with which NP-CDTs are described in literature, and a few instances of the types of classes and relationships that are needed for describing nanoparticles:

1. *Entrapped gold dendrimers*: These are nanoparticle formulations that consist of gold nanoparticles entrapped in the core of poly(amidoamine) dendrimers. In these formulations, some of the amine surface groups of the dendrimer are covalently linked to folic acid, which acts as a targeting agent for targeting folate receptors that are over-expressed on the surfaces of certain types of cancer cells (e.g., human epithelial carcinoma cells). To reduce the toxicity of gold-dendrimer nanoparticles some amine groups on the dendrimer surface are covalently linked to acetic anhydride. Fluorescein isothiocyanate (FITC) is also covalently linked to a fraction of the amine groups to enable fluorescence imaging of the nanoparticles [27]. This example illustrates the different physical ways in which nanoparticles can be functionalized with other compounds and ligands for targeting and imaging.
2. *Polymer micelles*: Examples of these formulations include anticancer drugs such as adriamycin (doxorubicin) conjugated to polymer micelles through a pH-sensitive linker such as hydrazone. When these nanoparticles are exposed to the acidic pH, the linkage is destabilized, which consequently triggers the release of the drugs in therapeutic applications [28]. This example illustrates how nanoparticles can be functionalized using molecules with specific roles for certain applications, based on a priori knowledge of molecular properties (e.g, sensitivity of hydrazone to acidic pH range) and physiological/tumor properties (e.g., presence of acidic pH in solid tumor).
3. *An iron oxide nanoparticle formulation designed for imaging*: This example describes a crosslinked dextran-coated iron oxide (superparamagnetic magnetic resonance contrast agent) nanoparticle with a very complex composition [23],[29]. The primary amine groups of dextran are linked to *N*-succinimidyl 3-(2-pyridyldithio)propionate (SPDP), a heterobifunctional amine- and sulfhydryl-reactive crosslinker. The SPDP itself is linked to the C-terminal cysteine residue of peptide (Arg)₄-(Gly)-(Cys). Finally, the *N*-terminal Arg residue of the peptide is linked to a fluorescent dye, a *N*-hydroxylsuccinimidyl (NHS) ester of Cys5.5. This example illustrates the extremely complicated formulation of some multifunctional nanoparticle platforms.

Based on examples like the aforementioned ones, NP-CDTs can be generally described in the following way: In general, a nanoparticle formulation consists of **chemical components** that can be enumerated as 1) nanoparticles, 2) active chemical constituents, which are part of

the chemical makeup of the nanoparticle, and 3) active chemical components which functionalize the nanoparticle. There can be one or more types of nanoparticle in a nanoparticle formulation, depending upon the **nanoparticle structure, function or chemical composition**. All of the chemical components can be described by their **molecular structure, biochemical role, or function**. Besides enumerating and describing the chemical components, one needs to describe the types of **chemical linkages** (e.g., amide linkage, disulphide linkage, encapsulation, etc.) that exist between them, and thereby provide an overall qualitative description of the chemical composition in a nanoparticle formulation. Other descriptions include: **physical locations of chemical components** in a nanoparticle (e.g., core, surface, etc.); **shape** of the nanoparticle (e.g., spherical, cylindrical, etc.); **physical state** of the formulation (e.g., emulsion, hydrogel, etc.); **physical, chemical or functional properties** of the active chemical components / constituents (e.g., organic, hydrophilic, magnetic, etc.); intended **functions and applications** of chemical components for cancer treatment, diagnosis and therapy; **underlying mechanisms** guiding the design of the nanoparticle formulation (e.g., endocytosis, active targeting, etc.), and; **type of stimulus** that may be required for activating nanoparticles (e.g., magnetic field, ultrasound, pH change, etc.) and the **nanoparticle's response to the stimulus** (e.g., drug release from nanoparticle in response to magnetic field stimulus, heat generation from nanoparticle in response to stimulus of infrared light, etc.).

We generated the initial list of terms from the literature. For each type of information (listed in the previous paragraph), we identified the header terms and relationships associating them. This initial collection of terms and relationships provided the basis for structuring the information content of NP-CDTs from the literature, based on which we collected more terms and organized them in a taxonomic “is a” hierarchy. We adapted this initial hierarchy to systematically and formally construct the NPO, using the well-defined design principles and best practices in ontology development described below.

2.2. Design principles for the NPO

In this section, we discuss the design principles that were followed in the development of the NPO. We selected the Basic Formal Ontology (BFO – version 1.1) as the upper-level ontology for developing a structured classification of NPO terms. BFO is a formal ontology based on tested principles for biomedical ontologies [30]. There are several advantages to using the BFO framework for constructing the NPO: (1) the BFO provides a formal structure for classification of domain terms, (2) the BFO offers a set of well-defined principles known for best ontology practices in the biomedical arena, (3) the BFO helps make the NPO interoperable with other ontologies that have the formal structure of BFO, and (4) the BFO will make the information represented in the ontology remain clear, rigorous and unambiguous as it is expanded through collaborative development. For a comprehensive account of the BFO, the reader is directed to the BFO manual [30]. Only relevant BFO terms are currently used in the NPO; however, other top-level BFO terms may be added if needed.

We selected OWL-DL as the language for encoding the NPO because: (1) OWL has formal semantics and additional vocabularies which facilitate machine interoperability, (2) OWL is designed for use in applications that process and present information to humans, and (3) due to the availability of the Protégé-OWL [31] ontology editing software, which is intuitively designed for editing OWL files and it greatly facilitates collaborative ontology development by both ontologists and domain experts.

In the interest of completeness, the main design principles used in developing the NPO are listed below. These design precepts are based on BFO and Open Biomedical Ontologies

(OBO) Foundry principles (<http://www.obofoundry.org/crit.shtml>) as well as our review of other OWL-encoded ontologies and controlled vocabularies:

1. *Principle of unbiased representation*: Following BFO design principles, any term in the ontology should represent an entity as known in reality and not represent it from the biased view of an individual.
2. *Principle of asserted single “is a” inheritance*: Again following BFO principles, each term should have no more than one parent term in the asserted OWL hierarchy. This principle offers the advantages of making the ontology easily extensible and interoperable with other ontologies that have a formal structure. This single-parent structure also helps to build the ontology in a modular fashion whereby different parts of the ontology can be worked on independently [32].
3. *Principle of inferred multiple “is a” inheritances*: Multiple parent-child relationships for a term are not present in the asserted hierarchy. However, a term can have more than one parent in the inferred hierarchy that is constructed by invoking an appropriate “OWL reasoner” (e.g., Pellet [33]) on the asserted hierarchy. Rules for inferring these relationships are expressed using OWL description logics and specified as OWL *necessary and sufficient* or *necessary* conditions, in the ontology. The OWL reasoner uses the OWL expressions to create the inferred hierarchy.
4. *Sibling disjointness*: Unlike in the BFO, disjoint axioms for sibling classes are not enforced at all levels in the asserted OWL hierarchy. The disjointness is maintained at the upper level of the ontology formed by the BFO classes. If sibling disjointness is applied at a level of the asserted OWL hierarchy, then the following principles are considered:
 - a. Disjoint axioms are applied only between primitive sibling classes.
 - b. Disjoint axioms are applied to primitive sibling classes only after the hierarchical level containing the classes is exhausted, such that any class added later will not have instances that overlap with the instances of existing sibling classes.
5. *Preferred name and textual definition*: Every OWL class and OWL property (object, datatype) must have a preferred name and a textual definition using the NPO's OWL annotation properties: “preferred name” and “definition”.
6. *Synonym*: If a class or OWL property has multiple names, these names must be provided as synonyms using the NPO's “synonym” OWL annotation property.
7. *External class reference*: Classes borrowed from external sources should be given their external reference ID using the “dBXrefID” annotation property. A term may have multiple dBXrefIDs if it has mappings to more than one terminology.
8. *Code*: Every class must have an identification code that starts with the prefix “NPO_” (e.g., NPO_100)
9. *Rdf ID and rdf:Label*: Every class specifically defined in the NPO must have its NPO code as its rdf:ID. The rdf:ID of every class borrowed from an external ontology found in the OBO Foundry list, must be preserved in the NPO. Every class in the NPO must also have its preferred name as its rdf:Label.

The preceding list is included to demonstrate the clear set of principles used for the initial development and continued growth of the NPO. Interested readers should refer to the BFO documentation [30], the NPO documentation (<http://npo.wustl.edu/>), as well as other basic references on ontology development [4]. The initial set of NPO terms were organized based

on these design principles and then expanded based on literature surveys, data curation, and our own research. Terms that are found in other relevant ontologies / CVs like GO [8], ChEBI [9], NCI Thesaurus [24], and MeSH [26] are re-used while developing the NPO.

3. Results and Discussion

In this section, we present the salient features of the NPO using its OWL version ‘2010-01-29’ as available from BioPortal (<http://bioportal.bioontology.org/>). The namespace URI for accessing the NPO is <http://purl.bioontology.org/ontology/npo>, and the corresponding namespace is ‘npo’. The OWL DL expressivity of NPO is SHIQ(D). The NPO has 1564 OWL classes, 45 OWL object properties for specifying class-level associations, and 5 OWL annotation properties for annotating the classes and object properties. The NPO covers 10 namespaces that uniquely identify all the classes in the NPO, and these along with the number of classes (given in parenthesis) are: bfo (1), snap (10), span (6), CHEBI(264), GO(83), FIX(17), REX(8), UO(104), PATO(54), and npo (1017). In this paper, we indicate the names of object properties in boldfaced letters as in **object property**, and the names of classes and annotation properties in italicized letters as in *Class Name* and *annotation property*, respectively.

3.1. OWL annotation properties in the NPO

The five OWL annotation properties of NPO used for annotating classes and object properties are *preferred name*, *definition*, *synonym*, *code*, and *dBXrefID*.

3.1.1. preferred name—The property *preferred name* is used to indicate the preferred name of a class or a class-level association in the NPO. For example, the *preferred name* of class *Quantum Dot* is “quantum dot”.

3.1.2. definition—The property *definition* is used to indicate the meaning or definition of a class or a class-level association in the NPO. For example, the *definition* of class *Quantum Dot* is “A three-dimensional nanoparticle which is made of material that exhibits quantum effects”.

3.1.3. synonym—The property *synonym* is used to indicate an alternate name of a class or a class-level association in the NPO. For example, a *synonym* of class *Quantum Dot* is “QD”.

3.1.4. code—The property *code* is used to indicate a unique identification code for every class in the NPO. Every code value begins with the prefix “NPO_” followed by a number. For example, the *code* for class *Quantum Dot* is ‘NPO_589’. The smallest code value is NPO_100 of class *Cylinder*, while the highest code value in the current version is “NPO_1572” of class *Core-Shell Silica Nanoparticle*.

3.1.5. dBXrefID—The property *dBXrefID* is used to indicate the reference ID of a class from an external CV / ontology / database. For example, a *dBXrefID* of class *Polymer* is CHEBI:33839.

3.2. OWL Object Properties (class-level associations) in the NPO

The OWL object properties are used to specify associations between classes, and these are represented in the NPO as shown in Fig. 1. Also shown in the figure, are the number of times each of these properties are used in the NPO. These properties are described in Table 1.

3.3. Types of domain-specific entities represented in the NPO

In Fig. 2, we present the asserted upper-level hierarchy of NPO that shows how the top-level domain-specific classes are classified under the BFO classes (indicated with asterisks). *Entity* is the top-most class in the NPO, and all instances of Entity (entities) are broadly classified into two groups – continuants and occurrents. A continuant (endurant) has no temporal parts and it endures through time by maintaining its identity at any time in which it exists (e.g., nanoparticle, dendrimer, scanning electron microscope, etc.). On the other hand, an occurrent (perdurant) has temporal parts and it happens, unfolds, or develops through time [30] (e.g., assay, active targeting, endocytosis, angiogenesis inhibition, etc.). By default, each child class (subclass) in the NPO is related to its parent class through the **is a** (inheritance) relation. Among continuant classes, the **is a** relation is interpreted as follows: If *C* and *CI* are continuant classes, then *C is a CI* means that every instance of *C*, at any time, is an instance of *CI* at the same time [34] (e.g., *Nanoparticle is a Nanomaterial*, *Iron Oxide Nanoparticle is a Nanoparticle*, etc.). Among occurrent classes, the **is a** relation is interpreted as follows: If *P* and *PI* are occurrent classes, then *P is a PI* means that every instance of *P* is an instance of *PI* [34] (e.g., *Amide Linkage is a Covalent Linkage*, *Targeting is a Process*, etc.). As shown in Fig. 2., continuants are further classified as independent continuants and dependent continuants. For a detailed description of all the BFO classes, the reader is referred to the BFO manual [30]. In the following subsections, we describe the main types of domain-specific entities represented in the NPO, based on the examples shown in Fig. 2.

3.3.1. Types of entities represented as independent continuants—Entities that are represented as independent continuants, include the following types:

1. Material entities / objects that are synthesized, characterized and distinguished at the nanoscale (1 – 100 nm) length dimension: These entities are represented by subclasses of *Nanomaterial* as shown in Fig. 3A.
2. Material entities / objects that are distinguished at the molecular level: These entities are represented by subclasses of *Atom*, *Element*, *Isotope*, and *Molecular Entity* as shown in Fig. 3B. Most of the classes in this category can be obtained from the ChEBI ontology. The ChEBI asserted OWL hierarchy has multiple **is_a** inheritances, which are forbidden in the asserted hierarchy of the NPO through our adoption of the single inheritance principle from BFO. Thus, we cannot directly import the ChEBI ontology into the NPO. At the same time, we note that multiple inheritances are inevitable in the description of chemical entities as it is clearly magnified in ChEBI. Therefore, when using ChEBI terms in the NPO, we ensure these terms are included in the NPO according to NPO design principles stated in section 2.2. Currently, there are 264 ChEBI classes in the NPO. More terms from ChEBI will be added, as we need them to make restrictions for other terms in the NPO to represent knowledge that is specific to the cancer nanotechnology domain. However, the ChEBI branch will continue to grow as more terms from ChEBI are added during the development of the NPO.
3. Entities which represent instruments used for the preparation and characterization of nanomaterials: These entities are represented by subclasses of *Instrument* as shown in Fig. 3C.
4. Entities which describe physical sites in a material entity: These entities are represented under class *Material Site* as shown in Fig. 3D.
5. Entities which describe surfaces of material entities: These entities are represented under the class *Material Boundary* as shown in Fig. 3E.

6. Entities which are fiat parts of material entities: These entities include different types of molecular parts (e.g., functional groups) in molecules; linkages or chemical associations between chemical components in nanoparticle formulations, samples, etc.. These are represented under the class *Fiat Material Part* as shown in Fig. 4A. In many cases, nanoparticles contain molecules with specific functional groups, which serve as reactive sites for chemical conjugation with other molecules. Thus, in the ontology, it becomes important to represent the type of functional groups that are present in a given organic compound / molecule [36]. This is achieved in the NPO by associating “organic compound” classes with “functional group” classes using parthood relations, such as, **part_of** and **has_part** (see Table 1 for examples). The covalent linkages (shown in Fig. 4B) in nanoparticle formulations are part of functional groups, and are derived from functional groups of molecules that are covalently linked during the synthesis of nanoparticles. Fig. 4C depicts an example of how an amide linkage formed between a primary amine and a carboxylic acid is represented in the NPO using **part_of** and **derives_from** relations. This knowledge is useful to determine ways to link different types of molecules for functionalizing nanoparticles.

3.3.2. Types of entities represented as dependent continuants—Entities that are dependent on the existence of other entities are considered as dependent continuants, and these include the following types:

1. Entities which describe the quality inhering in material entities: These entities are represented under the class *Quality* as shown in Fig. 5A.
2. Entities which describe roles of material entities at the molecular level in a particular context: These entities are represented under class *Molecular Role* as shown in Fig. 5B.
3. Entities which describe functions and dispositions inhering in a continuant entity: These entities are represented under the class *Function* and *Disposition*, respectively, as shown in Fig. 5B.
4. Entities which describe the units of measurements: These entities are represented under the class *Unit Of Measurement* as shown in Fig. 5C. We re-use classes from the unit ontology (<http://purl.org/obo/owl/UO>) to represent many of the units of measurements in the NPO.
5. Entities which describe the types of chemical interactions inhering in linkages, and these are represented under the class *Chemical Interaction*, as shown in Fig. 5D.

3.3.3. Types of entities represented as occurrents—Entities that are processes or parts of processes are considered as occurrents, and these include the following types:

1. Entities which describe functions of molecular entities that are realized as processes. As shown in Fig. 6A, these entities are represented under classes like *Antineoplastic Activity* and *Molecular Function* – a top-level GO class, which describes the activity of a gene product at the molecular level.
2. Entities which describe processes occurring in integrated living units such as cells, tissues, organs and organisms. As shown in Fig. 6B, these entities are represented under the class *Biological Process* – another top-level GO class, which describes a biological process in which gene products participate.

Since the asserted classification of GO terms under *Molecular Function* and *Biological Process* does not adhere to the single **is a** inheritance principle, we

cannot directly import the Gene Ontology into the NPO. Therefore, when using GO terms, we ensure these terms are included in the NPO according to NPO design principles stated in section 2.2. Currently in the NPO, there are 83 GO classes, and their hierarchical structures do not resemble those of GO. More terms from GO will be added as we need them to make restrictions for other terms in the NPO to represent knowledge that is specific to the cancer nanotechnology domain. As more terms from GO are added during the development of the NPO, the branches of *Molecular Function* and *Biological Process* will continue to expand.

3. Entities which describe the stimulus for activating the function of a nanoparticle, and response to the stimulus. As shown in Fig. 7A, these entities are represented under classes *Stimulus* and *Nanoparticle Response To Stimulus*.
4. Entities which describe tumor targeting mechanisms; these entities are represented under the class *Tumor Targeting*, as shown in Fig. 7A.
5. *Entities which describe studies, experiments and techniques*; these entities are represented under classes *Study* and *Procedure*, as shown in Fig. 7B.

3.4. Inferred hierarchy of the NPO

So far, we have discussed the different parts of the asserted hierarchy and the class-level relationships of the NPO. The main feature of the asserted hierarchy is that every class is “**is a**” related to not more than one parent class (single **is a** inheritance), unlike that of the inferred hierarchy which contains classes with more than one parent class (multiple **is a** inheritance). We elucidate this by comparing the asserted and inferred hierarchies of the subclasses of *Nanotube*. As shown in Fig. 8, the following three classes – *Double-Walled Carbon Nanotube*, *Multi-Walled Carbon Nanotube*, and *Single-Walled Carbon Nanotube*, have an additional parent class *Carbon Nanotube*, which is inferred based on the criteria that all nanotubes composed of carbon are also instances of *Carbon Nanotube*. Rules for these criteria are expressed in the following OWL restrictions:

1. Necessary and sufficient condition for *Carbon Nanotube*: *Carbon Nanotube is a (Nanotube and has part Carbon)*
2. Necessary condition for *Double-Walled Carbon Nanotube*: *Double-Walled Carbon Nanotube is a (Double-Walled Nanotube and has part Carbon)*
3. Necessary condition for *Multi-Walled Carbon Nanotube*: *Multi-Walled Carbon Nanotube is a (Multi-Walled Nanotube and has part Carbon)*
4. Necessary condition for *Single-Walled Carbon Nanotube*: *Single-Walled Carbon Nanotube is a (Single-Walled Nanotube and has part Carbon)*

One could also infer “**is a**” relationships among sibling classes without having to assert them explicitly. For example, the subclasses of *Nanoparticle* (see Fig. 3A) are sibling classes, which are defined, in the NPO, using necessary and sufficient conditions in relation to subclasses in other branches of the ontology such as *Molecular Entity*, *Realizable Entity* or *Quality*. When a reasoner reads the instructions in these OWL definitions, the nanoparticle classes re-arrange based on how its related classes are arranged in other branches of the ontology. For example, the reasoner infers that *Dextran-Coated Nanoparticle* “**is a**” subclass of *Biopolymer-Coated Nanoparticle*, as depicted in Fig. 9.

3.5 Applications of the NPO

The NPO can be applied to solving several problems facing interdisciplinary research in cancer nanotechnology. Here, we mention a few applications of the NPO:

1. One particularly challenging aspect of interdisciplinary research such as nanomedicine is the extreme diversity of venues for publication of scientific results. The relevant literature is distributed through numerous collections of medical, chemical, physical and engineering journals. When scientists search for related articles, they are faced with the problem of searching unfamiliar journals, and this problem is compounded by variation in terminology between disciplines. This is particularly true for the field of nanomedicine where a typical research project might require information about the chemical synthesis, physicochemical characterization, *in vitro* behavior, and *in vivo* testing of a nanomaterial. Ontologies such as the NPO can be used to solve this problem by providing (a) a common set of terminology to help standardize nomenclature for indexing and expand searches to include synonymous terms or topics, and (b) semantic search capability: the ability to further expand or narrow searches based on inheritance relationships and associations provided by the ontology. Using such capabilities, users can include related terms in their search without the need for detailed knowledge of the terminology particular to a specific area of cancer nanotechnology research. Such semantic search capability can significantly reduce barriers to searching data across disciplines, and this is very much needed in cancer nanotechnology research.
2. Another important application of the NPO will be in data annotation, which can help indexing, integrating and retrieving data from disparate sources, and eventually facilitate data mining and knowledge discovery. In the past, annotation of data using ontologies has led to several discoveries through the use of logical relationships provided by the structure of the ontology; e.g., GO [37]-[39]. By providing the necessary definitions and relationships, the NPO will facilitate semantic interoperability among different applications and resources developed to store, search and analyze research data. Because NPO contains terms from GO and ChEBI it can provide a unifying knowledge framework to associate nanoparticle data with external data annotated using GO and ChEBI terms. Thus, applications using the NPO can also draw inferences from the wide range of annotated data associated with GO and ChEBI, thereby, facilitating data mining and inferencing. Preliminary steps for such annotation have been undertaken through the caOBR tool (manuscript in preparation) which uses the NPO to index several online resources, including caNanoLab, via the Open Biomedical Resources [40] technology. Initial caOBR results are currently available through the Biportal resource.

3.6. Documentation, availability and download of NPO

The latest information and documentation for NPO are made available through a wiki that can be accessed at <http://npo.wustl.edu>. Public releases of NPO can be accessed, visualized and downloaded through BioPortal (<http://purl.bioontology.org/ontology/npo>) – a web portal developed by the National Center for Biomedical Ontology. The NPO is also uploaded at the Biomed Grid Terminology (BiomedGT) Ontology Development Wiki (http://biomedgt.nci.nih.gov/wiki/index.php/Main_Page), which serves as a tool for collaborative development and maintenance of ontologies and other terminologies. Being founded upon well-defined design principles, the current version of the NPO provides a structure suitable for collaborative development and future growth of the NPO.

3.7 NPO as a caBIG[®] vocabulary standard

To facilitate the adoption of NPO into caBIG[®] applications, it is necessary to get the NPO approved as a caBIG[®] vocabulary standard by the Vocabulary and Common Data Elements (VCDE) workspace. Initial review by the VCDE in early 2009 indicated that NPO met more

than 70% of the criteria for structure and content but the NPO needs a better documentation, a clear editorial process and a clear governance structure. The NPO met most of the structure criteria related to the following areas: concept orientation, non-semantic term identifiers, polyhierarchical organization, explicitness of relations, recognition of redundancy and extensibility. The NPO, however, did not have rules for ensuring term permanence. NPO met most of the content criteria related to polyhierarchy, rejection of “not elsewhere classified” (NEC) terms, context representation and textual definitions. NPO partially met the criteria related to content coverage, and some terms were not sufficiently well defined to distinguish them from other terms. A detailed account of the VCDE review of NPO can be accessed from the link http://gforge.nci.nih.gov/frs/?group_id=471. Progress has been made in terms of improving the documentation of NPO through the establishment of the NPO wiki (<http://npo.wustl.edu>). A working draft for the editorial process can be viewed at the NPO wiki (http://npo.wustl.edu/index.php/Editorial_Process).

3.8 NPO governance

To date, a community of interest for NPO has been established through the activities of caBIG[®] Nanotechnology Working Group, our participation in NCI Nanotechnology Alliance events, collaboration with the caNanoLab development team, collaboration with NCBO, interaction with the BiomedGT Semantic MediaWiki Project, and discussions with the Nanotechnology Characterization Lab (NCL). While this current community of interactions and collaborations provides the basis for long-term growth of NPO, it lacks any formal organizational structure or structured plans for incorporation of feedback into ontology development and review cycles (see next paragraph on governance structure). The community could grow in many ways, some of the natural interactions include agencies with existing nanotechnology informatics efforts, such as Environmental Protection Agency (NBI Knowledgebase <http://oregonstate.edu/nbi/pages>) and National Science Foundation (InterNano <http://www.internano.org/> and NanoHub <http://www.nanohub.org/>).

With continued growth, there is the need for careful maintenance of the ontology. To ensure that the best practices are followed in the growth and maintenance of the ontology, future growth and development of the NPO must include a governance structure. The governance structure should establish well-defined development and review cycles, enforce ontology design principles, ensure term permanence, and handle extensions and revisions in a systematic manner. Such a governance structure requires long-term support for ontology development to minimize governance turnover and provide a continuous set of ontology engineers to maintain the vocabulary. Plans are underway to establish a governance structure for the NPO through interactions with VCDE workspace members, and through the activities of the caBIG[®] Nanotechnology Working Group.

In the longer term, the NPO development will benefit from more systematic harmonization with other ontologies and vocabularies. Existing ontology development tools such as BiomedGT Semantic MediaWiki (<http://www.biomedgt.org/>) and BioPortal can facilitate such harmonization. In effect, this harmonization will make it easier to integrate developing informatics resources throughout the nanotechnology community, to share terms and relationships among different disciplines, to facilitate the broader exchange of nanotechnology-related information between data sources. Harmonization, however, is a difficult undertaking that requires a sufficiently broad community of research groups and organizations to motivate participants to engage and contribute. Therefore, community building is an essential factor for driving the harmonization process.

4. Conclusions

In this paper, we have reported the development, design principles, and salient features of the NPO. The NPO is implemented in OWL and uses BFO as the upper-level ontology to formally and systematically classify the domain terms. The present scope of the NPO encompasses knowledge underlying the preparation, chemical composition, physicochemical characterization, in vitro characterization and in vivo characterization of nanomaterials studied in cancer nanotechnology research. The NPO shares terms from other ontologies / CVs such as CHEBI, GO, FIX, REX, UO, PATO, and NCI Thesaurus.

The NPO is being developed to meet the terminological and informatics needs in cancer nanotechnology research, such as: facilitating interdisciplinary discourse among diverse research groups, enabling semantic interoperability among applications and resources that store and exchange nanomaterial data, and providing knowledge support for data annotation in order to facilitate semantic integration, knowledge-based searching, unambiguous interpretation, mining and inferencing of data.

In this paper, we have described the different types of entities and the relations between entities that are relevant to the domain of cancer nanotechnology. We also showed where the domain-specific entities are classified under the BFO, and how NPO handles the representation of multiple inheritances using OWL defined classes. We note that the NPO is far from being complete and is subject to corrections, as the field of nanomedicine and nanotechnology informatics continues to develop. Presently, we have laid a foundation for future growth of the NPO. In the path forward, we will continue to evaluate the NPO to assess its suitability for applications in different user scenarios (e.g., annotation of data, answering specific scientific questions, etc.), and to identify gaps and errors in the ontology. We also intend to review the dependent continuant branch as many of the terms in this branch are under-represented. It is clear that NPO has overlap with other domain ontologies (especially GO and ChEBI). We anticipate that more terms will be added from these external domain ontologies as we find use for them in the cancer nanotechnology domain. We intend to focus most of our efforts on representing knowledge that is not represented in these or other domain ontologies but that is essential to the cancer nanotechnology domain and to the types of applications that are of interest to the NPO user community. For example, we have represented covalent linkages in the NPO by relating them to functional groups of organic molecules. This type of representation enables one to identify the different ways of functionalizing nanoparticles using organic compounds, based on knowledge about the covalent linkages that can be derived from any two functional groups.

As the NPO continues to grow, its scope will expand depending upon the needs of and feedback from the user community. The NPO has significantly impacted development of the caNanoLab caBIG[®] database application by helping to identify and define important terms and relationships to model the description of nanoparticles. Continued collaboration with the caNanoLab developers helps ongoing identification of the terminological needs of caNanoLab and lead to expansion of the scope and application of the NPO. Currently, we are manually collecting and relating terms based on literary information to develop the NPO. However, in the long-term development of the NPO, we plan to develop and apply semi-automated text-mining methods to evaluate and enrich the NPO.

Acknowledgments

The authors would like to thank the current and past members of the caBIG[®] Nanotechnology Working Group for helpful comments. We appreciate the feedback from Daniel Rubin during the initial stages of NPO development, that led to a more formal development of the NPO using BFO and OBO Foundry based design principles. We thank Frank Hartel, Liz Hahn-Dantona, Gilberto Fragoso and Sherri De Coronado, for valuable discussions on the NPO

design and development, and also for providing support for collaborative development of NPO using the BiomedGT Semantic Media Wiki. We also thank Nigam Shah and Natasha Noy for helpful advice and support to facilitate access to NPO OWL files at the NCBO BioPortal via purl links. We had very useful discussions with Sharon Gaheen and Sue Pan - lead developers of caNanoLab, that helped determine the current scope of NPO. The authors are grateful for the reviews and feedback about the NPO from Brian Davis, Kristel Dobratz and Stuart Turner of the VCDE workspace of caBIG®. The authors would also like to thank the anonymous referees for their valuable comments and suggestions, which has greatly affected the development of NPO. We gratefully acknowledge the NIH for funding this work through grants U54 CA119342 and U54 HG004028.

References

1. Blake J. Bio-ontologies - fast and furious. *Nat Biotech* 2004;22:773–4.
2. Musen MA. Dimensions of knowledge sharing and reuse. *Comput Biomed Res* 1992;25(5):435–67. [PubMed: 1395522]
3. Gruber TR. A translational approach to portable ontology specifications. *Knowl Aquis* 1993;5(2): 199–220.
4. Noy, NF.; McGuinness, DL. *Ontology development 101: a guide to creating your first ontology*. Stanford Knowledge Systems Laboratory Technical Report and Stanford Medical Informatics Technical Report. 2001. Available from: Knowledge Systems Laboratory Stanford University. http://www.ksl.stanford.edu/KSL_Abstracts/KSL-01-05.html
5. Smith L, Wilbur WJ. Retrieving definitional content for ontology development. *Comput Biol Chem* 2004;28(5-6):387–91. [PubMed: 15556479]
6. Stevens R, Goble CA, Bechhofer S. Ontology-based knowledge representation for bioinformatics. *Brief Bioinform* 2000;1(4):398–414. [PubMed: 11465057]
7. Rubin DL, Shah NH, Noy NF. Biomedical Ontologies: a functional perspective. *Brief Bioinform* 2008;9(1):75–90. [PubMed: 18077472]
8. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000;25(1):25–9. [PubMed: 10802651]
9. Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, et al. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res* 2008;36:344–50.
10. Li S, Becich MJ, Gilbertson J. Microarray data mining using gene ontology. *Stud Health Technol Inform* 2004;107:778–82. [PubMed: 15360918]
11. Rubin DL, Lewis SE, Mungall CJ, Misra S, Westerfield M, Ashburner M, et al. National Center for biomedical ontology: advancing biomedicine through structured organization of scientific knowledge. *OMICS* 2006;10(2):185–98. [PubMed: 16901225]
12. Nanoscale Science, Engineering and Technology Subcommittee, Committee on Technology, National Science and Technology Council. *The National Nanotechnology Initiative Strategic Plan*. 2004. Available from: National Nanotechnology Initiative. <http://nano.gov/html/res/pubs.html>
13. Nie S, Xing Y, Kim GJ, Simons JW. Nanotechnology applications in cancer. *Annu Rev Biomed Eng* 2007;9:257–88. [PubMed: 17439359]
14. Moghimi SM, Hunter AC, Murray JC. Long-circulating and target-specific nanoparticles: theory and practice. *Pharmacol Rev* 2001;53(2):283–318. [PubMed: 11356986]
15. Allen TM, Cullis PR. Drug delivery systems: Entering the mainstream. *Science* 2004;303(5665): 1818–22. [PubMed: 15031496]
16. Brigger I, Dubernet C, Couvreur P. Nanoparticles in cancer therapy and diagnosis. *Adv Drug Deliv Rev* 2002;54(5):631–51. [PubMed: 12204596]
17. Gaur U, Sahoo SK, De TK, Ghosh PC, Maitra A, Ghosh PK. Biodistribution of fluoresceinated dextran using novel nanoparticles evading reticuloendothelial system. *Int J Pharm* 2000;202(1-2): 1–10. [PubMed: 10915921]
18. Gradishar WJ, Tjulandin S, Davidson N, Shaw H, Desai N, Bhar P, et al. Phase III Trial of nanoparticle albumin-bound paclitaxel compared with polyethylated castor oil-based paclitaxel in women with breast cancer. *J Clin Oncol* 2005;23(31):7794–803. [PubMed: 16172456]
19. Registry of clinical trials. Available from: ClinicalTrials.gov. <http://clinicaltrials.gov/ct2/results?term=cancer+nanoarticle>

20. Baker, NA.; Fritts, M.; Guccione, S.; Paik, D.; Pappu, RV.; Patri, A., et al. Nanotechnology Informatics White Paper [document on the internet]. NCI GForge site. 2009. [updated 2009 February 9]. Available from: <http://gforge.nci.nih.gov/projects/nano/>
21. Ferrari M. Cancer nanotechnology: opportunities and challenges. *Nat Rev Cancer* 2005;5(3):161–71. [PubMed: 15738981]
22. Ferrari M. The mathematical engines of nanomedicine. *Small* 2008;4(1):20–5. [PubMed: 18165947]
23. Shaw SY, Westly EC, Pittet MJ, Subramanian A, Schreiber SL, Weissleder R. Perturbational profiling of nanomaterial biologic activity. *Proc Natl Acad Sci U S A* 2008;105(21):7387–92. [PubMed: 18492802]
24. National Cancer Institute. The caNanoLab database. Available from: <http://gforge.nci.nih.gov/projects/calab/>
25. De Coronado S, Haber MW, Sioutos N, Tuttle MS, Wright LW. NCI Thesaurus: using science-based terminology to integrate cancer research results. *Stud Health Technol Inform* 2004;107:33–7. [PubMed: 15360769]
26. Nelson SJ, Schopen M, Savage AG, Schulman JL, Arluk N. The MeSH translation maintenance system: structure, interface design, and implementation. *Stud Health Technol Inform* 2004;107:67–9. [PubMed: 15360776]
27. Shi X, Wang S, Meshinchi S, Antwerp MEV, Bi X, Lee I, et al. Dendrimer-entrapped gold nanoparticles as a platform for cancer-cell targeting and imaging. *Small* 2007;3(7):1245–52. [PubMed: 17523182]
28. Bae Y, Jang W, Nishiyama N, Fukushima S, Kataoka K. Multifunctional polymeric micelles with folate-mediated cancer cell targeting and pH-triggered drug releasing properties for active intracellular drug delivery. *Mol Biosyst* 2005;1(3):242–50. [PubMed: 16880988]
29. Josephson L, Kircher MF, Mahmood U, Tang Y, Weissleder R. Near-Infrared fluorescent nanoparticles as combined MR/Optical imaging probes. *Bioconjugate Chem* 2002;13(3):554–60.
30. Spear, AD. Institute for Formal Ontology and Medical Information Science (IFOMIS); 2006. Ontology for the twenty first century: an introduction with recommendations. Basic Formal Ontology (BFO) [document on the internet]. Available from: <http://www.ifomis.org/bfo/manual>
31. Noy NF, Crubézy M, Ferguson RW, Knublauch H, Tu SW, Vendetti J, et al. Protégé-2000: An Open-Source Ontology-Development and Knowledge-Acquisition Environment. *AMIA Annu Symp Proc* 2003 2003:953.
32. Rector, AL. Modularisation of domain ontologies implemented in description logics and related formalisms including OWL. *Proceedings of the 2nd International Conference on Knowledge Capture; 2002 October 23-25; Sanibel Island, Florida, USA. 2003. p. 121-8.*
33. Sirin, E.; Parsia, B.; Grau, BC.; Kalyanpur, A.; Katz, Y. Clark & Parsia, LLC; Pellet: A practical OWL-DL reasoner [document on the internet]. Available from: <http://clarksparsia.com/parsia/#res>
34. Smith B, Ceusters W, Klagges B, Kohler J, Kumar A, Lomax J, et al. Relations in biomedical ontologies. *Genome Biol* 2005;6(5):R46. [PubMed: 15892874]
35. Schulz S, Kumar A, Bittner T. Biomedical ontologies: What part-of is and isn't. *J Biomed Inform* 2006;39(3):350–61. [PubMed: 16442850]
36. Villanueva-Rosales, N.; Dumontier, M. Describing chemical functional groups in OWL-DL for the classification of chemical compounds. *Proceedings of the 3rd International Workshop on OWL: Experiences and Directions, co-located with European Semantic Web Conference; 2007 June 6-7; Innsbruck, Austria. 2007. Available from: OWLED 2007. <http://www.webont.org/owlled/2007/Proceedings.html>*
37. Tuikkala J, Elo L, Olli SN, Aittokallio T. Improving missing value estimation in microarray data with gene ontology. *Bioinform* 2006;22(5):566–72.
38. Hsing M, Byler KG, Cherkasov A. The use of Gene Ontology terms for predicting highly-connected 'hub' nodes in protein-protein interaction networks. *BMC Syst Biol* 2008;2:80. [PubMed: 18796161]
39. Tao Y, Sam L, Li J, Friedman C, Lussier YA. Information theory applied to the sparse gene ontology annotation network to predict novel gene function. *Bioinformatics* 2007;23(13):i529–38. [PubMed: 17646340]

40. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucl Acids Res Adv* 2009:1–4.

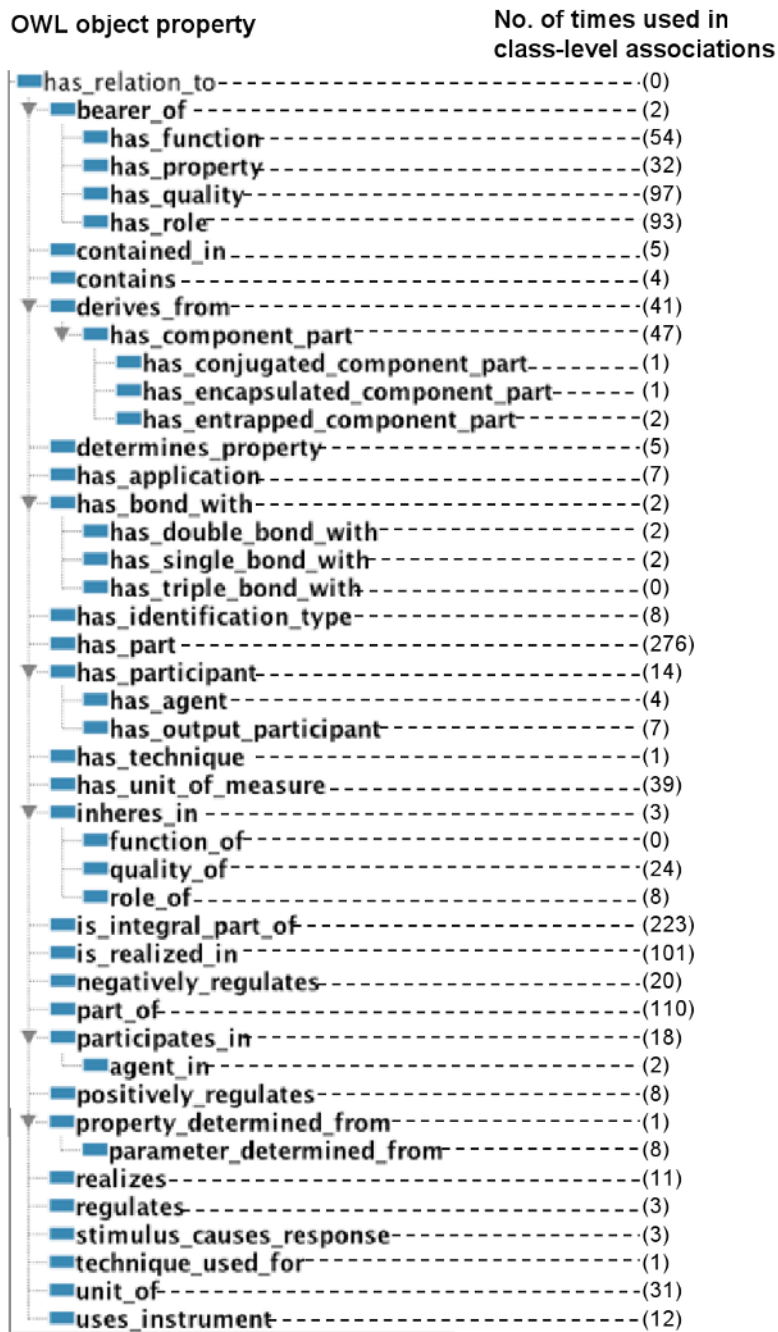


Fig. 1. OWL object properties and the number of times they are used for asserting class-level associations in the NPO.

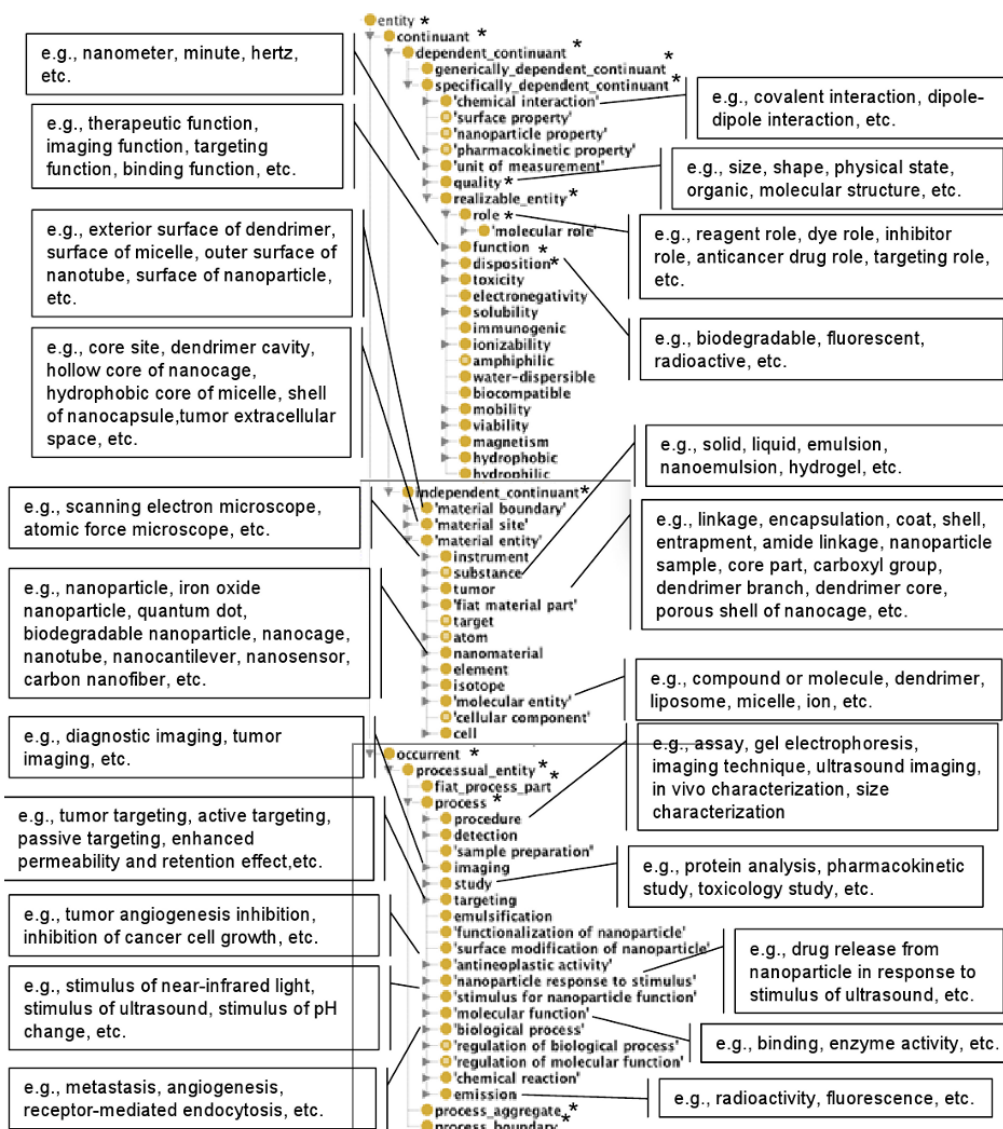


Fig. 2. Asserted OWL hierarchy showing the BFO-based upper-level classification of domain terms in NPO with selected examples.

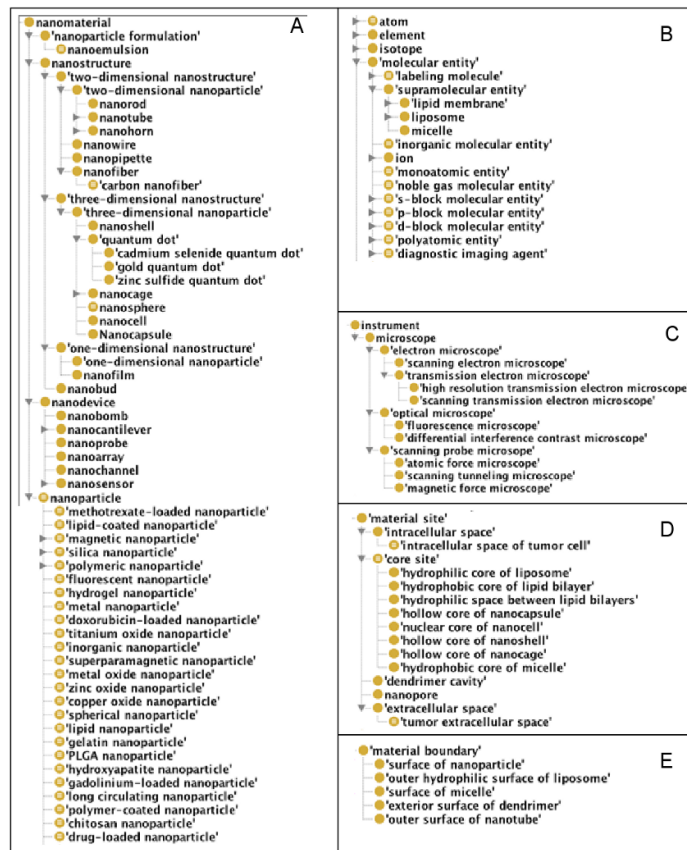


Fig. 3. Asserted OWL hierarchy of a subset of NPO classes. (A) Subclasses of *Nanomaterial*. (B) Subclasses of *Molecular Entity*, and other classes representing atoms, elements, and isotopes. (C) Subclasses of *Instrument*. (D) Subclasses of *Material Site*. (E) Subclasses of *Material Boundary*.

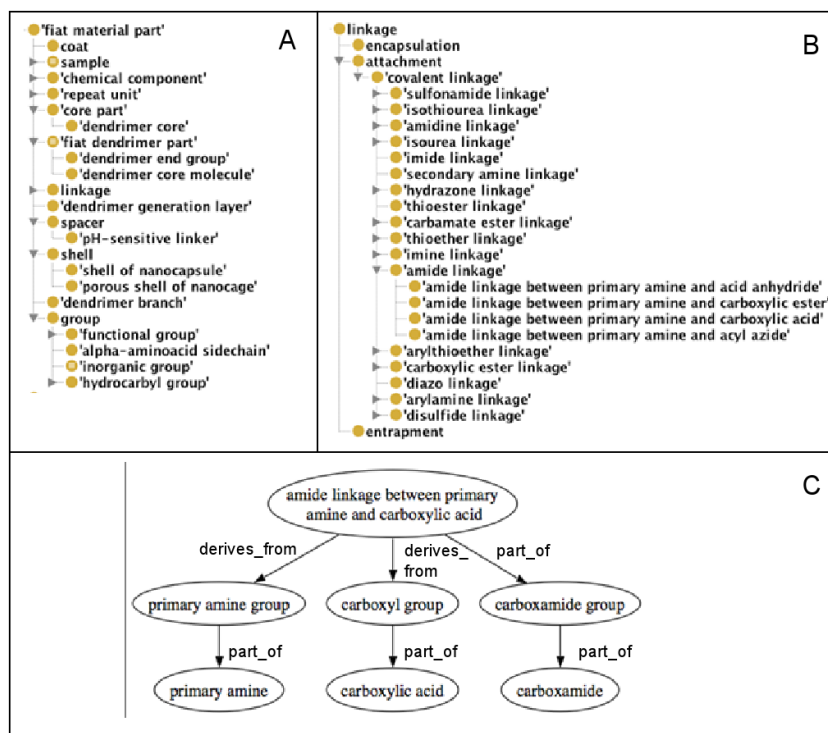


Fig. 4. (A) Asserted OWL hierarchy of a subset of classes under *Material Fiat Part*. (B) Asserted OWL hierarchy of subclasses of *Linkage*. (C) An example showing how an amide linkage formed between a primary amine and a carboxylic acid are associated to the respective functional groups in the NPO.



Fig. 5. Asserted OWL hierarchy of a subset of NPO classes. (A) Subclasses of *Quality*. (B) Subclasses of *Role*, *Function*, and *Disposition*. (C) Subclasses of *Unit Of Measurement*. (D) Subclasses of *Chemical Interaction*.

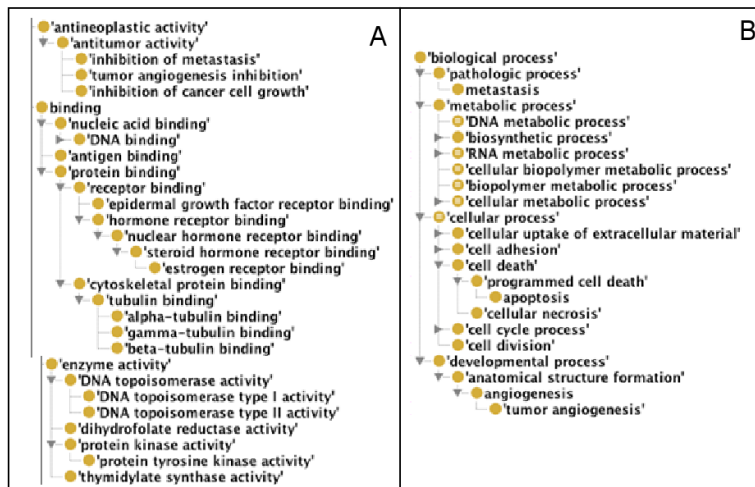
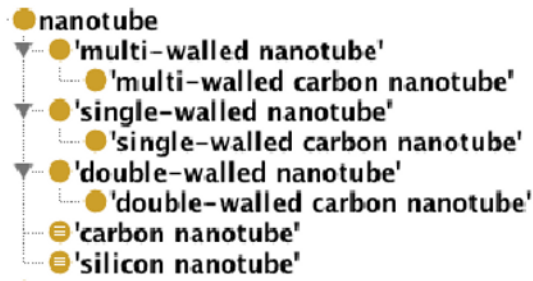


Fig. 6. Asserted OWL hierarchy of a subset of NPO classes. (A) Subclasses of *Antineoplastic Activity* and *Molecular Function*. (B) Subclasses of *Biological Process*.



Fig. 7. Asserted OWL hierarchy of a subset of NPO classes. (A) Subclasses of *Stimulus For Nanoparticle Function*, *Nanoparticle Response To Stimulus*, and *Tumor Targeting*. (B) Subclasses of *Study*, *Characterization*, and *Procedure*.

Asserted is_a hierarchy



Inferred is_a hierarchy

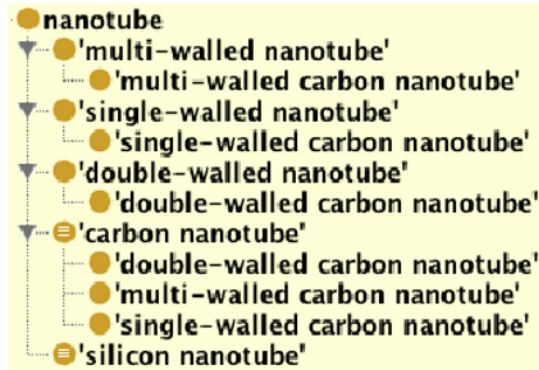


Fig. 8. Single and multiple inheritances of subclasses of *Nanotube*, as shown in the asserted and inferred hierarchies, respectively.

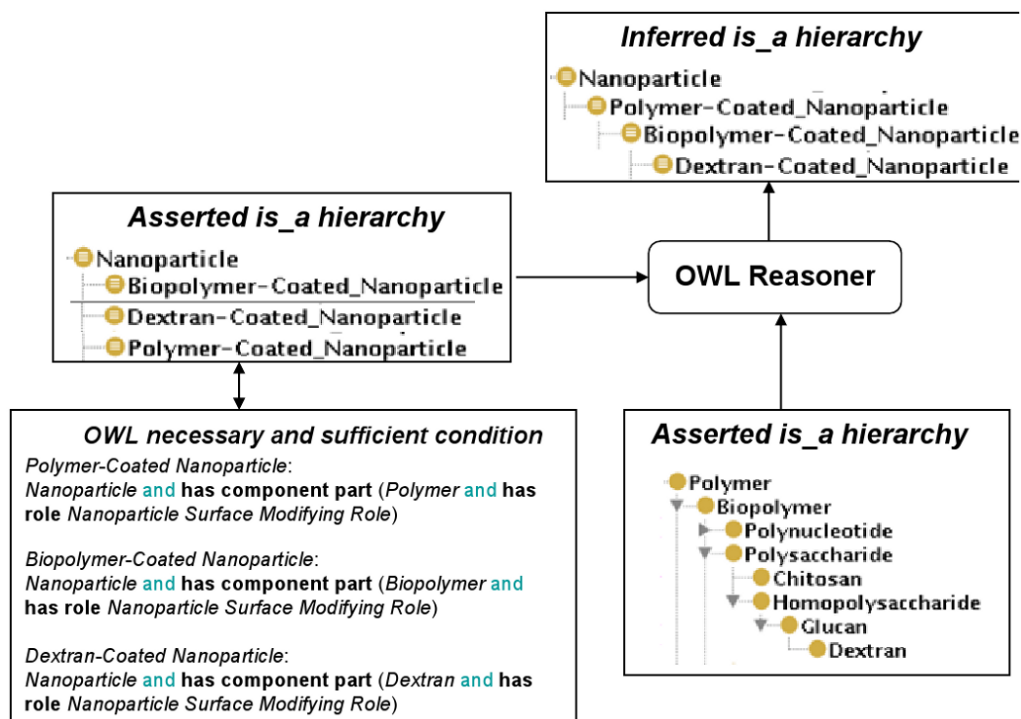


Fig. 9. Example showing how an OWL reasoner infers parent-child relationships among *Polymer-Coated Nanoparticle*, *Biopolymer-Coated Nanoparticle*, and *Dextran-Coated Nanoparticle* based on their asserted associations to and based on the asserted hierarchy of *Polymer*, *Biopolymer*, and *Dextran* respectively.

Table 1

OWL object properties (class-level associations) in the NPO.

No.	OWL object property	Definition and comments	Examples
1.	has relation to	A general relation, which indicates that a class is related to another class.	
2.	has part	A transitive parthood relation in which one entity has itself, or another entity as part, and the relation is valid for use between continuants or between occurments but not between a continuant and an occurrence [34], [35]. The has part relation between any two continuant classes is interpreted as follows: If <i>C</i> and <i>CI</i> are continuant classes, then <i>C has part CI</i> means that every instance of <i>C</i> , at any time, has some (one or more) instances of <i>CI</i> as parts at the same time [34]. The has part relation between occurrence classes is interpreted as follows: If <i>P</i> and <i>PI</i> are occurrence classes, then <i>P has part PI</i> means that every instance of <i>P</i> has some instances of <i>PI</i> as parts [34].	<i>Primary Amine has part Primary Amine Group</i> <i>Thiol has part Sulfhydryl Group</i>
3.	part of	A transitive parthood relation in which one entity is part of another entity, and the relation is valid for use between continuants or between occurments but not between a continuant and an occurrence. The part of relation between continuants is time-dependent while between occurments it is time-independent [34], [35]. The part of relation between any two continuant classes is interpreted as follows: If <i>C</i> and <i>CI</i> are continuant classes, then <i>C part of CI</i> means that every instance of <i>C</i> , at any time, is part of some (one or more) instances of <i>CI</i> at the same time [34]. The part of relation between occurrence classes is interpreted as follows: If <i>P</i> and <i>PI</i> are occurrence classes, then <i>P part of PI</i> means that every instance of <i>P</i> is part of some instances of <i>PI</i> [34].	<i>Carboxyl Group part of Carboxylic Acid</i> <i>Dendrimer Generation Layer part of Dendrimer Sample Preparation part of Study</i>
4.	determines property	A relation that indicates the property determined by a characterization.	<i>Size Characterization determines property Size</i>
5.	function of	A relation that indicates the bearer inhering a function.	<i>Targeting Function function of Targeting</i>
6.	has application	A relation that indicates the application or use of an entity.	<i>Western Blot Analysis has application Detection Of Proteins</i>
7.	has bond with	A relation, which indicates that one atom has a covalent bond with another atom in a polyatomic entity.	<i>Organofluorine Compound has part (Carbon Atom \cap has bond with Fluorine Atom)</i>
8.	has double bond with	A relation, which indicates that one atom has a double covalent bond with another atom in a polyatomic entity.	
9.	has single bond with	A relation which indicates that one atom has a single covalent bond with another atom in a polyatomic entity.	<i>Chlorambucil has part (Carbon Atom \cap has single bond with Fluorine Atom)</i>
10.	has triple bond with	A relation, which indicates that one atom has a triple covalent bond with another atom in a polyatomic entity.	
11.	has component part	A relation that is used between material entities where one is composed of the other. In the NPO, this relationship is used to describe the composition of nanomaterials. It is interpreted as follows: If <i>C</i> and <i>CI</i> are continuant classes, then <i>C has component part CI</i> means that every instance "c" of <i>C</i> is prepared from some instances of "c1" of <i>CI</i> and a significant material portion of c1 becomes part of the constituting material of "c" when "c" begins to exist.	<i>Gold Nanoparticle has component part Gold</i> <i>Solid Lipid Nanoparticle has component part Lipid</i>
12.	has conjugated component part	A relation which is used to infer that a particular chemical component of a nanoparticle formulation is attached or chemically conjugated to some other chemical components of the same formulation.	
13.	has encapsulated component part	A relation which is used to infer that a particular chemical component becomes enclosed – due to non-covalent interactions –	

No.	OWL object property	Definition and comments	Examples
		in a nanoparticle's material site that is fully bounded by the particle's constituting material.	
14.	has entrapped component part	A relation which is used to infer that a particular component is trapped – due to non-covalent interactions – in a nanoparticle's material site that is partially bounded by the particle's constituting material.	
15.	has function	A relation that indicates the function inhering in an entity.	<i>Targeting Agent has function Targeting Function</i>
16.	is realized in	A relation between a realizable entity and an occurrent.	<i>Emulsifying Role is realized in Emulsification</i>
17.	has participant	A relation between a continuant and an occurrent, which holds if and only if at every time an instance of the occurrent is created, it has at least one instance of the continuant participating in it at the same time. Therefore, if <i>P</i> is an occurrent class and <i>C</i> a continuant class, then <i>P has participant C</i> means that every instance of <i>P</i> involves some instances of <i>C</i> as participants [34].	<i>Western Blot Analysis has participant Antibody</i> <i>Folic Acid Metabolism has participant Folic Acid</i>
18.	has output participant	A relation between a continuant and an occurrent that has other participating continuants, which holds if and only if at every time <i>t</i> an instance of the occurrent is created, there is at least one instance of the continuant participating in it at time <i>tI</i> > <i>t</i> , and it is formed / derived from the instances of those continuants that participate at time <i>t</i> .	<i>Sample Preparation has output participant Sample</i>
19.	has quality	A relation that indicates the quality inhering in an entity. It is a general relation used between any class and a subclass of <i>Quality</i> .	<i>Solid Phase has quality Solid State</i> <i>Quantum Dot has quality Sphere</i>
20.	has role	A general relation between an independent continuant and role. It is interpreted as follows: if <i>C</i> is an independent continuant and <i>CI</i> a role continuant class, then <i>C has role CI</i> means that every instance of <i>C</i> fills some instances of role <i>CI</i> every time the role is realized in a particular context [34].	<i>Doxorubicin has role Anticancer Drug Role</i> <i>Targeting Agent has role Targeting Role</i>
21.	has technique	A relation that indicates the technique used for accomplishing a purpose.	<i>Imaging has technique Imaging Technique</i>
22.	has unit of measure	A relation that indicates the unit of a measured quantity.	<i>Length of Time has unit of measure Time Unit</i> <i>Transfection Efficiency has unit of measure Percent</i>
23.	negatively regulates	A relation between two occurrents where one negatively regulates the other.	<i>Tumor Angiogenesis Inhibition negatively regulates Tumor Angiogenesis</i>
24.	is integral part of	A parthood relation that holds when <i>C part_of CI</i> and <i>CI has part C</i> , where <i>C</i> and <i>CI</i> are continuant classes.	<i>Gold Atom is integral part of Gold</i> <i>Carbon Atom is integral part of Carbon</i>
25.	participates in	A relation between a continuant and an occurrent, which holds if and only if at every time an instance of the continuant is created, it participates in atleast one instance of the occurrent at the same time. Therefore, if <i>C</i> is a continuant class and <i>P</i> is an occurrent class, then <i>C participates in P</i> means that every instance of <i>C</i> participates in some instances of <i>P</i> [34].	<i>Scanning Electron Microscope participates in Scanning Electron Microscopy</i>
26.	property determined from	A relation, which indicates the type of process that determines a qualitative or quantitative measure of a property.	<i>Pharmacokinetic Property property determined from Pharmacokinetic Study</i>
27.	parameter determined from	A relation, which indicates the type of process that determines the value of a parameter.	<i>Elimination Half Life parameter determined from Pharmacokinetics Study</i>
28.	quality of	A relation between a quality and its bearer.	<i>Crystalline State quality of Crystal</i>

No.	OWL object property	Definition and comments	Examples
29.	regulates	A relation between two occursents where one regulates the other.	<i>Regulation Of DNA-RNA Transcription</i> regulates <i>DNA-Dependent Transcription</i>
30.	stimulus causes response	A relation that indicates the response caused by a stimulus. In the NPO, subclasses of <i>Nanoparticle Stimulus</i> and <i>Response To Stimulus</i> are associated using this relation.	<i>Stimulus Of Infrared Light</i> stimulus causes response <i>Nanoparticle Response To Stimulus Of Infrared Light</i>
31.	technique used for	A relation that indicates the process for which the technique is used.	<i>Imaging Technique</i> technique used for <i>Imaging</i>
32.	unit of	A relation that indicates the quantity associated with a unit of measure.	<i>Time Unit</i> unit of <i>Length Of Time</i>
33.	uses instrument	A relation that indicates the type of instrument used in a process.	<i>Scanning Electron Microscopy</i> uses instrument <i>Scanning Electron Microscope</i>
34.	positively regulates	A relation between two occursents where one positively regulates the other.	<i>Positive Regulation Of Mitosis</i> positively regulates <i>Mitosis</i>
35.	has property	A relation between an entity and a specifically dependent continuant	<i>Lipid</i> has property <i>Amphiphilic</i>
36.	bearer of	A relation between an entity and a dependent continuant	<i>Covalent Linkage</i> bearer of <i>Covalent Interaction</i>
37.	contained in	A relation between a material continuant and an immaterial continuant, when the material continuant is located in but shares no parts in common with the immaterial continuant at the same time [34].	<i>Tumor Extracellular Fluid</i> contained in <i>Tumor Extracellular Space</i>
38.	contains	A reciprocal relation of contained in , where given an immaterial continuant at some time, there is a material continuant contained in the immaterial continuant at the same time.	<i>Intracellular Space of Tumor Cell</i> contains <i>Intracellular Fluid In A Tumor Cell</i>
39.	derives from	A relation between distinct material continuants when one succeeds the other across a temporal divide in such a way that at least a significant portion of the matter of the earlier continuant is inherited by the later [34].	<i>Disulfide Linkage Between Two Thiols</i> derives from <i>Sulfhydryl Group</i>
40.	has identification type	A relation which indicates the name of the type of identification, as well as other attributes of an entity.	<i>Dendrimer End Group</i> has identification type <i>Functional Group</i>
41.	has agent	A relation which is between a process, a continuant and a time whenever the continuant is a participant in the process and is at the same time directly causally responsible for its occurrence [34].	<i>Emulsification</i> has agent <i>Emulsifier</i>
42.	agent in	A relation between a continuant and an occurrent in which the continuant participates in and causes the occurrence of that occurrent.	<i>DNA Crosslinking Agent</i> agent in <i>DNA Crosslinking</i>
43.	realizes	A relation which is between an occurrent and a realizable entity.	<i>Emulsification</i> realizes <i>Emulsifying Role</i>
44.	inheres in	A relation between a dependent continuant and an entity.	<i>Drug Toxicity</i> inheres in <i>Anticancer Drug</i>
45.	role of	A relation that indicates the bearer inhering a role.	<i>Dye Role</i> role of <i>Dye</i>