31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Ontologies for Cancer Nanotechnology Research

Dennis G. Thomas, Rohit V. Pappu, and Nathan A. Baker

Abstract—**Cancer nanotechnology research data are diverse. Ontologies that provide a unifying knowledge framework for annotation of data are necessary to facilitate the sharing and semantic integration of data for advancing the research via informatics methods. In this work, we report the development of NanoParticle Ontology (NPO) to support the terminological and informatics needs of cancer nanotechnology. The NPO is developed within the framework of the Basic Formal Ontology (BFO) using well-defined principles, and implemented in the Ontology Web Language (OWL). The NPO currently represents entities related to physical, chemical and functional descriptions of nanoparticles that are formulated and tested for applications in cancer diagnostics and therapeutics. Public releases of the NPO are available through the BioPortal web site, maintained by the National Center for Biomedical Ontology. Expansion of the scope and application of the NPO will depend on the needs of and feedback from the user community, and its adoption in nanoparticle database applications. As the NPO continues to grow, it will require a governance structure and well-organized community effort for the maintenance, review and development of the NPO.**

I. Introduction

Large volumes and diverse types of experimental data have been generated in cancer nanotechnology research, which is an interdisciplinary field that deals with the development and application of nanotechnology-based methods for the detection, diagnosis, and therapy of cancer. Most of these data characterize the physicochemical and functional properties related to the in vitro / in vivo behavior of nanoparticles that are formulated for applications in cancer diagnostics and therapeutics. Small changes in chemical composition can cause drastic changes in the properties of nanoparticles. Since there are many combinatorial ways by which the chemical composition can be modified, one can formulate diverse types of nanoparticles with varying properties and applications. Each new formulation will require experimental characterizations and this in turn adds more volume and diversity to the data. Additionally, the data and the underlying knowledge in cancer nanotechnology are complex due to the integration of information from multidisciplinary areas such as chemistry, material science, biology, and cancer medicine.

Most of the experimental results are found in disparate sources like journal articles. When scientists search for related journal articles, they are faced with the problem of searching unfamiliar journals, and this problem is compounded by variation in terminology between disciplines. It is also manually difficult to process large amounts of information from textual sources. Cancer nanotechnology data sets are rich in information and these can be mined for structure-activity relationships, and to seek correlations between different characteristic nanoparticle properties (e.g., correlation between in vitro and in vivo properties [1]). Mining of existing literature data can provide useful information to guide the re-purposing or de novo design of nanoparticles. There are database resources such as caNanoLab (http://gforge.nci.nih.gov/projects/calab/), which are being developed for storing, searching and sharing data generated from characterization experiments, with the goal of enabling knowledge discovery. But databases must be complemented by a common vocabulary to facilitate semantic interoperability among them.

In this work, we focus on the development of ontologies for cancer nanotechnology research. An ontology is a formal, explicit representation of knowledge belonging to a subject area: the knowledge is encoded and represented as a hierarchy of terms (classes) that are described using attributes (e.g., metadata such as preferred name, definition, synonyms, etc.), related using associative relations, and formalized using logical axioms in a machine-interpretable language [2]-[6]. Ontologies are used as common vocabularies, which researchers from different disciplines can share for annotating data in texts as well as in databases. There are several advantages to using ontologies: 1) the explicit definitions of the terms help avoid ambiguities in the usage of terminologies and interpretation of results; 2) the logical relationships among the terms help to semantically integrate different parts of the annotated data, and to perform knowledge-based searches for accessing and retrieving the relevant data.

There are ontologies / controlled vocabularies (e.g., Gene Ontology (GO) [7], Chemical Entities of Biological Interest (ChEBI) [8], NCI (National Cancer Institute) Thesaurus [9], etc.), which represent some parts of the knowledge within the cancer nanotechnology domain. But, there is no ontology, which has the terms and logical relationships that provide a unifying knowledge framework for supporting the annotation, semantic integration, mining and inferencing of

Manuscript received June 20, 2009. This work was supported by the National Institutes of Health under Grants U54 CA119342 and U54 HG004028.

D. G. Thomas is with the Department of Biomedical Engineering, Center for Computational Biology, Washington University in St. Louis, St. Louis, MO 63130 USA (corresponding author; phone: 314-362-1938; e-mail: dgt1@ ccb.wustl.edu).

R. V. Pappu, is with the Department of Biomedical Engineering, Center for Computational Biology, Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail: pappu@wustl.edu).

N. A. Baker is with the Department of Biochemistry and Molecular Biophysics, Center for Computational Biology, Washington University School of Medicine, St. Louis, MO 63110 USA (e-mail: baker@biochem.wustl.edu).

cancer nanotechnology data.

To this end, we have developed an ontology called the **N**ano**P**article **O**ntology (NPO). The NPO is constructed in the Ontology Web Language (OWL) using the Protégé-OWL editor [10]. The upper-level of the NPO contains terms from the Basic Formal Ontology (BFO; http://www.ifomis.org/bfo/manual), which provide a formal framework for classifying domain terms in the NPO. In the following sections, we present the design and development of the NPO.

II. METHOD

A. Scope and development of the NPO

To construct the NPO, we created an initial list of terms using the descriptions of nanoparticle formulations in the literature. These terms were obtained using information related to the: 1) type of chemical components of a nanoparticle formulation which include the nanoparticle, active chemical constituents of the nanoparticle, and functionalizing agents; 2) molecular structure, biochemical role or function of these chemical components; 3) type of nanoparticle based on its structure, function or chemical composition; 4) chemical linkages between chemical components; 5) physical locations of chemical components within a nanoparticle; 6) nanoparticle shape; 7) physical state of the formulation; 8) physical, chemical, or functional properties of the chemical constituents and functionalizing agents; 9) applications in cancer diagnosis, therapy, and treatment; 10) underlying mechanisms guiding the design for the formulation; 11) type of stimulus for activating the function of nanoparticles, and the response to that stimulus. Specifically, for each type of information, we identified the header terms and relationships associating these terms. These terms and relationships provided a structure for organizing the information content in the literature, based on which we collected more terms and organized them in the form of a taxonomic "is_a" hierarchy. For formal and systematic development of the NPO, we re-factored this hierarchy of terms using terms from the Basic Formal Ontology (BFO) at the upper-level of NPO, and constructed the NPO in the Ontology Web Language (OWL) using well-defined design principles. Terms that are found in other relevant ontologies / controlled vocabularies like GO, ChEBI, and NCI Thesaurus (NCIT) are re-used in the NPO.

B. Design factors for the NPO

The reasons for using BFO as the upper-level ontology are as follows: 1) it provides a formal structure for the classification of domain terms; 2) it offers well-defined design principles that are known for best ontology practices in the biomedical area; 3) it facilitates interoperability with other ontologies having formal structure of BFO; and 4) it allows a clear, unambiguous and rigorous expansion of the ontology via collaborative development.

We use OWL-DL to express the NPO because 1) OWL has formal semantics and additional vocabularies that facilitate machine interoperability; 2) it is designed for use in applications that process information as well as to present information to humans (http://www.w3.org/TR/owl-features/), and 3) of the availability of Protégé-OWL editor, which has an intuitive design for editing OWL files and greatly facilitates collaborative ontology development by both ontologists and domain experts.

The main design principles that we used for developing NPO – based on BFO, Open Biomedical Ontologies (OBO) Foundry principles, and our review of other OWL-encoded ontologies / controlled vocabularies – are as follows:

1. *Unbiased representation (based on BFO)*: Any term in the NPO should represent an entity as known in reality and not represent it from the biased view of an individual.
2. *Asserted single "is_a" inheritance (based on BFO)*: Each term should have only one parent term in the asserted OWL hierarchy. This principle helps make the ontology easily extensible and interoperable with other ontologies that have the formal BFO structure. The single parent structure also helps to build the ontology in a modular fashion whereby different parts of the ontology can be worked on independently.
3. *Inferred multiple "is_a" inheritance*: If a child term requires more than one parent term to be represented in the NPO (e.g., classification of compounds based on chemical composition), then the multiple parent-child relationships are inferred by using appropriate OWL conditions for the classes and an OWL reasoner (e.g., Pellet), and then represented in the inferred OWL hierarchy.
4. *Sibling disjointedness*: Unlike in the BFO, disjoint axioms for sibling classes are not stated at all levels in the asserted OWL hierarchy. But, sibling disjointedness is maintained for the BFO classes in the upper-level part of the NPO. If sibling disjointedness is applied at a level in the asserted OWL hierarchy, then the following principles are considered:
    a. Disjoint axioms are applied to sibling classes only after the hierarchical level containing these classes is exhausted, such that any class added later will not cause overlap with the existing sibling classes.
    b. Disjointedness is not applied between a pair of sibling classes when both sibling classes are defined classes or when one class is primitive and the other is a defined one.
5. *Preferred name and definition:* Every term has to be given a preferred name and a textual definition through the NPO's OWL annotation properties: "preferred_Name" and "definition".
6. *Synonym:* If there are multiple names for a term, then synonyms should be provided using the

NPO's "synonym" OWL annotation property.

7. *Term reference:* If a term is borrowed from an external source, use the "dBXrefID" property to enter its external reference ID.
8. *Code:* Every term must have an identification code starting with the prefix "NPO_" (e.g., NPO_100).

Since GO, ChEBI, and NCIT do not follow all the above design principles, we do not import them into the NPO OWL file. Instead, we borrow terms from these ontologies / controlled vocabularies and represent them in the NPO according to the above design principles.

III. RESULTS

To date, there are 919 terms (or classes; including BFO classes), 6 OWL annotation properties, and 21 OWL object properties (associative relationships) in the NPO. The current version (2009-04-02) of the NPO is available through BioPortal (http://www.tinyurl.com/npo-bioportal). In Table 1, we list the main (top-level) domain terms representing the different types of entities in the NPO. These domain terms are organized under the BFO terms as shown in Figure 1.

TABLE I
TYPES OF ENTITIES REPRESENTED IN THE NPO

Type of entity	Term(s)
Material entity which is synthesized, characterized and distinguished at the nanoscale (1-100) nm size range	Nanomaterial
Material entity which is distinguished at the molecular level	Molecular Structure
Material entity which is part of a cell or its extracellular environment	Cellular Component [7]
Physical site in a material entity	Material Site
Surface of material entity	Material Boundary
Quality or property inhering in a material entity	Quality
Role of material entity at the molecular level	Molecule Role
Stimulus for activating the function of a nanoparticle	Stimulus For Nanoparticle Function
Response to stimulus	Nanoparticle Response to Stimulus
Tumor targeting method	Tumor Targeting
Function of molecular entity that is realized as a process	Molecular Function [7], Antineoplastic Activity
Process occurring in integrated living units such as cells, tissues, organs and organisms	Biological Process [7]
Process which occurs during a chemical synthesis or reaction	Linkage, Chemical Interaction

In the following two examples, we show how entities related to nanoparticle descriptions, are represented in the NPO.

1) Example 1: A nanoparticle formulation has gold quantum dot entrapped in the core of poly(amidoamine) dendrimer. The amine surface group of the dendrimer is attached to folic acid via an amide linkage. The relevant classes and relationships are:

- Gold Quantum Dot **is_a** Quantum Dot; **has_part** Gold
- Quantum Dot **is_a** Three-dimensional Nanoparticle; **is_a** Nanosphere ; **has_quality** Sphere
- Nanosphere **is_a** (Three-dimensional Nanoparticle *and* **has_quality** Sphere)
- Three-dimensional Nanoparticle **is_a** Three-dimensional Nanostructure; **is_a** Nanoparticle
- Three-dimensional Nanostructure **is_a** Nanostructure
- Poly(amidoamine) Dendrimer **is_a** Dendrimer; **has_part** (Carboxamide *and* **has_role** Dendrimer Repeat Unit); **has_part** (Primary Amine Group *and* **has_role** Dendrimer End Group); **is_a** Carboxamide; **is_a** Primary Amine
- Folic Acid **is_a** Pterin Compound; **has_part** Carboxyl Group; **has_part** Pterin; **is_a** Carboxylic Acid; **is_a** Organic Hydroxy Compound; **is_a** Primary Amine

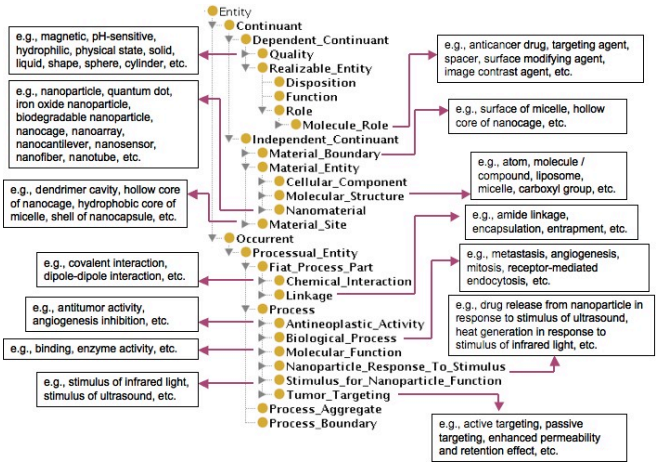

Fig. 1. Asserted OWL hierarchy showing the BFO classification of upper-level classes in the NPO

- Pterin **is_a** Pterin Compound; **has_part** Hydroxyl Group; **has_part** Primary Amine Group; **is_a** Organic Hydroxy Compound; **is_a** Primary Amine
- Amide Linkage Between Primary Amine And Carboxylic Acid **is_a** Amide Linkage; **has_participant** Carboxyl Group; **has_participant** Primary Amine Group; **has_output_participant** Carboxamide Group
- Carboxamide **is_a** (Amide *and* **has_part** Carboxamide Group); **has_part** Acyl Group; **is_a** Acyl Compound
- Primary Amine **is_a** (Amine *and* **has_part** Primary Amine Group); **is_a** Amine
- Amide Linkage **is_a** Linkage
- Entrapment **is_a** Linkage
- Gold Quantum Dot – Entrapped Poly(amidoamine) Dendrimer Nanoparticle **is_a** (Nanoparticle *and* **has_component_part** (Poly(amidoamine) Dendrimer and **participates_in** Entrapment) *and*

has_entrapped_component_part (Gold Quantum Dot *and* **participates_in** Entrapment)); **is_a** Nanoparticle

2) Example 2: A nanoparticle is loaded with doxorubicin. The relevant classes and relationships are:

- Doxorubicin **is_a** Anthracycline; **has_role** Antineoplastic Antibiotic; **has_role** DNA-RNA Transcription Regulator; **has_role** DNA Intercalating Agent; **has_role** Topoisomerase-II Inhibitor
- Doxorubicin-Loaded Nanoparticle **is_a** (Nanoparticle *and* **has_component_part** (Doxorubicin *and* **has_role** Nanoparticle Payload Agent); **is_a** Drug-Loaded Nanoparticle
- Drug-Loaded Nanoparticle **is_a** (Nanoparticle *and* **has_component_part** (Molecular Entity *and* **has_role** Anticancer Drug *and* **has_role** Nanoparticle Payload Agent)); **is_a** Nanoparticle
- Topoisomerase-II Inhibitor **is_a** Topoisomerase Inhibitor; **inhibits** DNA Topoisomerase Type II Activity
- DNA-RNA Transcription Regulator **is_a** (Antineoplastic Agent *and* **regulates** DNA-Dependent Transcription);
- Antineoplastic Agent **is_a** Anticancer Drug; **has_function_realized_as_process** Antineoplastic Activity
- Antineoplastic Activity **is_a** Process

Note that in the asserted hierarchy, Doxorubicin-Loaded Nanoparticle and Drug-Loaded Nanoparticle are sibling classes, with definitions containing the class-level associations: **has_component_part** and **has_role** (as shown above). Based on these definitions and the assertions of the associated classes in these definitions, the reasoner infers that Doxorubicin-Loaded Nanoparticle **is_a** Drug-Loaded Nanoparticle **is_a** Nanoparticle.

IV. Conclusion

The NPO has been developed to serve as a common vocabulary for annotating data in the cancer nanotechnology domain. Since the NPO contains GO / ChEBI terms, it can provide a unifying knowledge framework to associate nanoparticle data with data annotated using GO / ChEBI terms. Currently, NPO contains a small set of GO / ChEBI terms, and more terms will be added as NPO develops. These terms can be mapped and linked (via URL links) to their counterparts in GO and ChEBI on BioPortal.

Terms from the literature are manually collected in the NPO. In the long-term, however, it will be necessary to develop and apply semi-automated text-mining methods to enrich the NPO, and to evaluate the terms – i.e., their preferred name, synonyms, definition, is_a classification, and relationships to other terms in the NPO. Also, long-term growth and success of the NPO will require a governance structure and well-organized community effort to ensure proper maintenance, review and development of the ontology.

The current scope of the NPO includes entities related to the physical, chemical and functional descriptions of nanoparticles in the cancer domain. Extension of the scope will depend upon the needs of and feedback from the user community, and its adoption in nanoparticle database applications. For example, terms and definitions in the NPO have helped define some model elements of the caNanoLab database. Continued interactions with caNanoLab developers will help identify the terminological needs of caNanoLab, which will enable the expansion of the scope and application of the NPO.

Public releases of the NPO are available through BioPortal, maintained by the National Center for Biomedical Ontology. The NPO is also uploaded at the BiomedGT (http://tinyurl.com/npo-biomedgt) semantic media wiki that enables collaborative development of the ontology.

Acknowledgment

The authors would like to thank Daniel Rubin, Sharon Gaheen, Liz Hahn-Dantona, Frank Hartel, and Gilberto Fragoso for their helpful comments.

References

[1] S. Y Shaw, E. C. Westly, M. J. Pittet, A. Subramanian, S. L. Schreiber and R Weissleder, "Perturbational profiling of nanomaterial biologic activity," *Proc Natl Acad Sci U S A*, vol. 105, pp. 7387-7392. "

[2] M. A. Musen, "Dimensions of knowledge sharing and reuse," Comput Biomed Res 1992, vol. 25, pp. 435–467.

[3] T. R. Gruber, "A translational approach to portable ontology specifications," *Knowl Aquis* 1993, vol. 5, pp. 199–220.

[4] N. F. Noy and D. L. McGuinness, "Ontology development 101: a guide to creating your first ontology," Stanford Knowledge Systems Laboratory Technical Report and Stanford Medical Informatics Technical Report 2001.

[5] L. Smith and W. J. Wilbur, "Retrieving definitional content for ontology development," *Comput Biol Chem* 2004, vol. **28**, pp. 387-391.

[6] R. Stevens, C. A. Goble and S. Bechhofer, "Ontology-based knowledge representation for bioinformatics," *Brief Bioinform* 2000, vol. **1**, pp. 398-414.

[7] The Gene Ontology Consortium, "Gene Ontology: tool for the unification of biology," *Nat Genet* 2000, vol. **25**, pp. 25-29.

[8] K. Degtyarenko, P. de Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcántara, M. Darsow, M. Guedj and M. Ashburner, "ChEBI: a database and ontology for chemical entities of biological interest," *Nucleic Acids Res* 2007, vol **36**, pp. 344-350.

[9] S. De Coronado, M. W. Haber, N. Sioutos, M. S. Tuttle, L. W. Wright, "NCI Thesaurus: using science-based terminology to integrate cancer research results," *Stud Health Technol Inform* 2004, vol. **107**, pp. 33-37.

[10] N. F. Noy, M. Crubézy, R. W. Fergerson, H. Knublauch, S. W. Tu, J. Vendetti and M. A. Musen, "Protégé-2000: An Open-Source Ontology-Development and Knowledge-Acquisition Environment," *AMIA Annu Symp Proc* 2003, vol. **2003**, pp. 953.