Nanopublishing clinical diagnoses: tracking diagnostic knowledge base content and utilization

Alejandro Rodríguez-González
Centre for Biotechnology and Plant Genomics
Technical University of Madrid
Pozuelo de Alarcón, Madrid, Spain
alejandro.rodriguezg@upm.es

Marcos Martinez-Romero
IMEDIR Centre
University of A Coruña
A Coruña, Spain
marcosmartinez@udc.es

Mikel Egaña Aranguren
Centre for Biotechnology and Plant Genomics
Technical University of Madrid
Pozuelo de Alarcón, Madrid, Spain
mikel.egana.aranguren@gmail.com

Mark D. Wilkinson
Centre for Biotechnology and Plant Genomics
Technical University of Madrid
Pozuelo de Alarcón, Madrid, Spain
markw@illuminae.com

***Abstract*— Accurate and evidence-based diagnosis is a key step in clinical practice. High-quality diagnoses depend on several factors, including physician's training and experience. To assist physicians, medical diagnosis systems can be used, as part of clinical decision support systems (CDSS), to improve the accuracy of diagnoses, as well as inform the clinician regarding the bases of the diagnostic decisions in the context of prior knowledge. To support such CDSS systems, it is important to have accurate and well-formed knowledge bases with thoroughly-annotated diagnostic criteria, as well as models for representing clinical observations that allow them to more easily be analyzed by expert-systems. We propose the use of Nanopublications as a way to store provenance data related to the content of diagnostic knowledge bases, as well as the clinical diagnoses themselves. The primary goal is to be able to rigorously track the complete diagnostic process: from the knowledge base construction and its supporting evidence, to the clinical observations and the context within which they were made, through to the diagnosis itself, and the rationale behind it.**

Keywords— nanopublications; clinical observations; provenance; diagnosis; medical diagnosis; decision support systems

I. INTRODUCTION

Medical diagnosis is the process of determining the disease-state of a patient based on their clinical phenotype. The diagnostic process is complex, but can be simplified to, first, obtaining the case facts from the patient's history, physical examination, and laboratory tests (collection of clinical data); second, evaluating the relative importance of the different signs and symptoms within that clinical data, based on expert-knowledge, experience, and intuition; third, making a differential diagnosis based on all possible diseases that could be explained by the most relevant clinical symptoms; fourth, selecting the most likely diagnosis, or concluding that more evidence is required [1]. These procedures have been honed over centuries by physicians, based on their knowledge, training and experience. Recently, however, the quantity and granularity of clinical data routinely collected during patient encounters has resulted in the creation of computer based medical systems, namely diagnosis systems. These can assist physicians in the diagnostic process, utilizing the speed of computers and the wealth of global medical knowledge to obtain (arguably) more accurate diagnoses by taking into-consideration the full breadth of the available clinical observations.

In support of these automated systems, it is helpful to consider the design of clinical knowledge bases that can provide a maximum level of accuracy, while also providing rich contextual information to allow the clinical expert to assess the validity of the automated diagnosis. There are a wide range of sources of information that can be consulted to obtain information about diseases and their associated diagnostic criterion. These can be mined to generate rules for a basic Diagnostic Criterion Model (DCM model) [2], and such rules form the basis for most of the existing diagnosis systems. Given that rules can be derived from a wide variety of sources, it is desirable to provide some means by which the origin of any given rule can be traced to its source. This kind of "provenance" information currently is only recorded in an *ad hoc* manner, if at all, in existing diagnostic systems.

Similarly, during the diagnostic process itself, it is desirable to keep track of how and why a physician reaches a diagnosis based on the patient data. Each physician has different training and experience, and therefore may reach the same conclusion via different paths, or reach different conclusions from the same data. Provenance tracking of clinical decisions, therefore, is necessary for reproducibility and proper record-keeping.

In order to achieve the aforementioned aims, we propose that Nanopublications [3] could be used to store the provenance information associated with the clinical diagnostic process, as well as capturing the provenance of the rules contained within diagnostic knowledge bases.

1063-7125/14 $31.00 © 2014 IEEE
DOI 10.1109/CBMS.2014.82

The remainder of this paper is structured as follows: Section 2 presents related work. Section 3 presents the proposed model to represent the diagnostic information. Section 4 presents the results of our research. Finally, Section 5 outlines our conclusions and proposed future work.

II. STATE OF THE ART

A variety of approaches have been proposed in recent years to improve, using semantic technologies, the representation, processing and reuse of diagnosis information in medical information systems. These initiatives can be classified according to the specific application they were designed for, including the optimization of the diagnosis process [4], [5], the semantic representation of electronic health records (EHR) [6], and decision support systems [7]–[9].

Nevertheless, most of the systems developed to date do not store provenance information, or store it in a way that is difficult to access and interpret. Making electronic systems provenance-aware enables users to trace how a particular result has been produced, ensuring the reproducibility of scientific analysis and processes. As a consequence, considerable effort has been recently directed to developing models that can support provenance-aware applications.

A variety of provenance models and provenance-publishing frameworks have recently emerged, including the Open Provenance Model (OPM) [10], the Proof Markup Language (PML) [11], the PrIME methodology [12], the Provenir ontology [13], OvoPub [14], MicroPub [15] and finally the Nanopublication model [16], [17], which will be the focus of this paper. Some provenance models have been successfully applied to build new models and/or systems that can be used to solve specific problems in bioinformatics. A relevant example is the Biologic-Experiment-Result model (BERT) [18], which allows researchers to trace the experimental process flow in genomic databases. Another relevant example is the work by McCusker and McGuinness on how to add provenance information to data from high throughput experiments expressed using the MAGE (MicroArray and Gene Expression) standard [19]. The Chemical Information Ontology (CHEMINF) [20] has similarly been designed to take into account the provenance and reproducibility of data in computational experiments. The work presented here is novel in being the first to address the need to formalize the capture and representation of provenance information in diagnostic knowledge bases and processes.

III. MODELS

In this section, we first briefly present the Nanopublication model. We then introduce the Diagnosis Definition Ontology and discuss how it can be used to capture diagnostic rules. We go on to discuss a novel model for how these diagnostic rules can be "compartmentalized" within different Nanopublications. Importantly, this compartmentalization allows the same clinical disorder to be defined by differing rule-sets (e.g. if two clinicians disagree on the diagnostic criterion for a given disease) while avoiding logical conflicts within the provenance data; moreover, the Nanopublication metadata then allows these differing rule-sets to be interpreted in the context from which they were derived. Finally, we show how Nanopublications can also be used as a model for capturing diagnostic data from clinical encounters, together with the associated metadata relating to how, why, and when those observations were made.

1. Nanopublication Schema

Nanopublications are a semantic data model intended to capture the smallest unit of scientific fact: an assertion about anything that can be uniquely identified and attributed to its author [21]. Nanopublications have been created to support fine-grained attribution to authors and institutions with the aim of encouraging the reuse of data [3], for example, within the OpenPHACTS [22] consortium. Assertions are made using domain-semantics drawn from community ontologies and other information models. These are then "wrapped" within a named-graph [23] representing the Nanopublication itself. Additional named-graphs are linked to the Nanopublication, allowing provenance, annotation, attribution and citation to be associated with the assertion [21]. The basic elements of a Nanopublication are:

Assertion. This named graph contains statements that represent the scientific assertion being made by the author(s). Assertions consist of one or more semantic triples that form a single, indivisible unit of scientific thought.

Provenance. This named graph contains the authorship or origin of the assertion, and how it “came to be”. For example, by direct experiment, or by in silico prediction, when, and by whom? Statistical p-values and other indicators of validity should also be recorded here.

Publication Information. In the "PubInfo" named graph, important contextual information regarding the Nanopublication itself can be added. When was the Nanopublication produced, and by whom? Who owns the rights?

Nanopublications will soon include other features such as integrity hashes and versioning; however those issues will not be discussed further here. The one upcoming feature that is relevant, however, is the (as-yet not ratified) Nanopublication Collection model. This extension to the existing Nanopublication schema will allow Nanopublications to be grouped, and thereby share common metadata. In the context of this manuscript, we utilize this feature to allow individual diagnostic assertions to have independent statements of metadata (e.g. P-values), while still being part of a common overall model with its own metadata. As indicated by Paul Groth on the Nanopublication mailing list, collections are anticipated to become part of the formal specification in the next release.

2. Diagnosis Definition Ontology

In our previous work, semantic technologies were applied to build medical diagnosis systems for general practitioners [2], [5], [9], [24], [25], and later adapted to psychological disorders [26], [27]. The Diagnosis Definition Ontology (DDxDO)[1], was designed based on these previous efforts. It is a small ontology containing the core set of entity/relations that describe the

[1] http://purl.org/DDXDO/

diagnostic process. These include concepts such as diseases, disorders, clinical findings (signs and laboratory tests) and the relationships between them. The ontology is written in OWL [28] and is designed to encode the definitions of diseases, as well as represent the data coming from clinical observations during the diagnostic process. More detailed information about the DDxDO ontology is available on the project website[2].

3. *Nanopublishing disease definitions*

Though there are well-established and widely published clinical diagnostic guidelines (e.g. the Framingham 5 and 10-year risk guides), it has been shown that these guidelines are not rigorously followed by individual clinicians when making diagnoses or intervention decisions [29]. As such, there is a need to "personalize" diagnostic guidelines, such that not only are the rules themselves transparent and explicit, but the contextual information surrounding that novel set of rules is also explicit. The same is true when deriving treatment or diagnostic rule-sets from widely differing sources, not only physicians with different training/experience, but also text-mining from books or the Web, or through rule-discovery within data by machine-learning processes [30], [31].

Here, we propose a semantic model based on Nanopublications that both "compartmentalizes" the varying rule-sets associated with diagnosis of the same disease, and keeps-track of the contextual information surrounding the derivation of that specific set of diagnostic/intervention rules. The model is depicted in Figure 1, and consists of two Nanopublications - one representing a disease definition, and the other representing a clinical finding - that are linked-together in a one-to-many relationship. The disease definition consists of assertions that link disease names to their associated clinical findings; the clinical findings flesh-out the exact criterion that would cause that finding to be true. For example, ["Typhoid fever" has clinical finding "Fever"] is an assertion within the disease definition Nanopublication; and ["Fever" sublingual body temperature" > 39ºC"] is an assertion within the Clinical Finding Nanopublication. We will now go into further detail on how these Nanopublications should be used in our model.

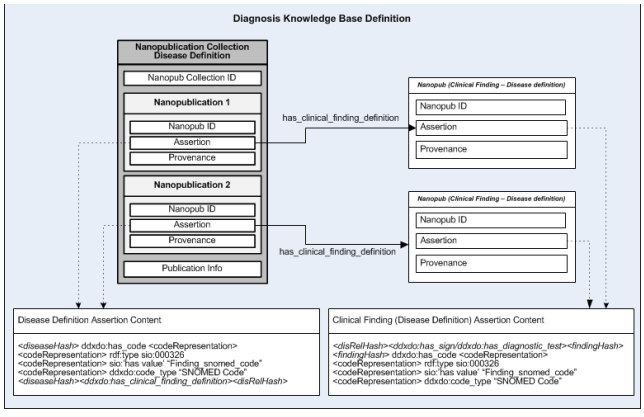

Figure 1. Nanopublication of disease definition.

[2] https://github.com/wilkinsonlab/DDxDO

Disease Definition

For each disease definition derived from a different information source (physician, researcher, textbook, guideline, etc.) a Nanopublication Collection will be created representing the diagnostic model of that disease according to that source.

Within these Nanopublication Collections, the Assertion named graphs will contain OWL property restrictions representing individual diagnostic criteria, according to that expert-source. Effectively, a set of [*Disease equivalent_to has_clinical_finding some Finding*] property restrictions, defined using the DDxDO ontology. Diagnostic criteria each have provenance information, including their likelihood of being associated with the disease. Shared provenance information, spanning all diagnostic criteria, includes the name of the physician, or the website from which the criteria were derived. This is encoded using ontologies such as PAV [32], Dublin Core [33] and the Nanopublication schema itself.

Clinical Finding

Each Diagnostic Model's restriction assertions are associated with a Clinical Finding, which is also defined via OWL property restrictions, following the DCM model [2]. This allows the models to be used by automated medical systems to execute a diagnosis by interpreting raw clinical data, or by physicians to explore detailed information about the symptomology of a disease. The Nanopublication of clinical findings will contain the following important data:

The Assertion named graph will contain restrictions on one of a limited number of properties, including *has_sign* (for direct clinical observations), *has_diagnostic_test*, or *has_disorder* (for prior/existing disorders or diseases). These restrictions might also include details such as levels of intensity, or value-ranges. For this purpose several ontologies such as Semanticscience Integrated Ontology (SIO)[3], Measurement Unit Ontology (MUO)[4], Units Ontology (UO)[5] or Quantities, Units, Dimensions and Data Types Ontologies (QUDT)[6] could be used to create these OWL restrictions, as was done in our previous work on encoding diagnostic rules in OWL [29], [34].

Provenance information associated with these Assertions will include all the information regarding the nature of the data introduced in the assertions such as the source of the definition (automated, manually-derived), author, version, etc.

What is most crucial to note about this model is that the disease definition can be re-assembled from these separate Nanopublications; the contents of the Assertion graphs combine to become a valid OWL-DL Class that can be utilized directly by logical reasoners to automatically interpret Nanopublished clinical data (as described below).

[3] https://code.google.com/p/semanticscience/wiki/SIO
[4] http://idi.fundacionctic.org/muo/
[5] https://code.google.com/p/unit-ontology/
[6] http://www.qudt.org/

4. Nanopublishing clinical data and diagnoses

To be able to utilize these OWL-encoded diagnostic models, we must have clinical data encoded in a format that is amenable to automated reasoning. Moreover, it would be desirable to also capture the metadata around these clinical observations as they pass through the diagnostic process; this includes not only metadata about how the measurements were made during the clinician-encounter, but also metadata about how the clinical measurements contributed to the final diagnosis. Once again, we propose the Nanopublication schema for this task. The model we propose is depicted in Figure 2.

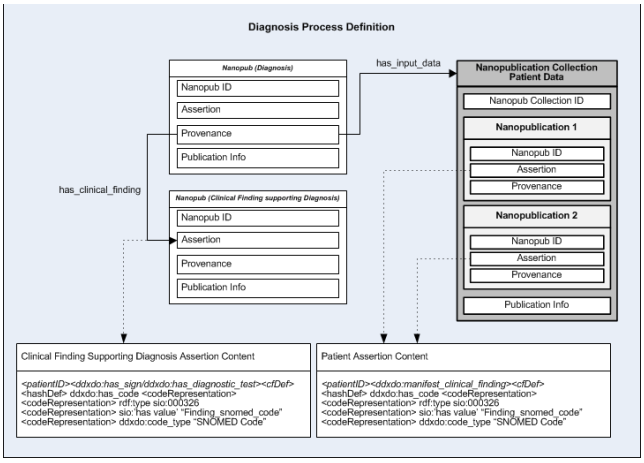

Figure 2. Nanopublication of the diagnostic decision.

Two Nanopublications and one Nanopublication Collection are used to store clinical data and their links to diagnostic decisions, and these will now be described in detail.

Diagnosis

The Diagnosis Nanopublication describes the final diagnosis - a single disease diagnosed by a single physician.

The Assertion graph in the Diagnosis Nanopublication contains a single triple identifying the diagnosed disease by its SNOMED code. We use the diagnosis relationship of DDxDO to make this assertion as follows:

<patientID> ddxdo:diagnosis <disease>.

<disease> ddxdo:has_code <codeRepresentation>

<codeRepresentation> sio:'has value' "Code" .

<codeRepresentation> ddxdo:code_type "SNOMED Code".

The Provenance graph describes how the physician or automated system reached the conclusion. The core of the provenance are two triples; one describing the *has_input_data* relationship, which links the Diagnosis Nanopublication to the Patient Data Nanopublication, containing the raw data used to make that diagnosis; the second describes the *has_clinical_finding* relationship, which links the Diagnosis with Clinical Findings that support the performed diagnosis (described below). The Publication Info graph contains information about the physician who performed the diagnosis, the date, etc.

Clinical Finding supporting Diagnosis

The Clinical Finding Assertion graph describes the set of filtered/interpreted clinical data that was deemed to be relevant to reaching the conclusion in the Diagnosis. Clinical Findings can be re-used by multiple Diagnosis Nanopublications. This covers two common scenarios: 1) the patient is suffering more than one disease, or 2) diagnosis has been performed by multiple physicians. Information regarding how that raw clinical data was interpreted/filtered in order to generate the Clinical Finding is stored in the Provenance named graph. Finally, the Publication Info graph describes the provenance of the Clinical Finding Nanopublication itself. It is important to note that the Nanopublications supporting the diagnosis could be either positive (symptom is present) or negative (symptom is absent), thus these are not precise duplicates of the Patient Data Nanopublications.

Patient Data

The Patient Data Nanopublication Collection captures the patient's raw clinical data - the data that is "filtered" to become Clinical Findings supporting the Diagnosis. Patient Data Assertions are triples where the *manifest_clinical_finding* predicate of DDxDO is used to link a patient with a symptom that they are exhibiting, coded using SNOMED. Assertions could contain more information about the symptom such as intensity, duration or any other valuable information regarding the manifestation. Provenance metadata might include, for example, the method or machine used to take the measurement.

More information

All the models presented in section III are fully explained with more detailed figures in the DDx2NP webpage [36, p. 2].

IV. RESULTS

A prototype of the models described above has been created. In [35], we describe a multi-level diagnosis system based on [2], [5], [24], [25]. The output from that investigation includes both clinician and automated diagnoses; however, provenance data was not formally associated with these outputs, for example, how the clinician or diagnostic system reached its conclusion. We therefore re-used the data and infrastructure from this earlier analysis to demonstrate the utility of our model, using the process described below. The source code and binaries to reproduce the analysis, as well as the resulting nanopublications, are available at the DDx2NP webpage [36, p. 2].

Knowledge base creation

The software first loads information about the disease into the knowledge base from a file containing the relations between diseases and their necessary Clinical Findings (the *has_clinical_finding* relationship in Figure 2). For each disease, a Nanopublication Collection file is created, defining the collection of disease-associated clinical findings (assertions) and provenance data such as probability of association (in the current version of the software, this is randomly generated for

demonstrative purposes). Subsequently, Clinical Finding Nanopublications (Figure 2) are created, containing the specific clinical definition of that finding (e.g. the boundaries of the temperature).

Diagnosis creation

Our software first creates a Nanopublication Collection file containing the clinical data of the patients from [35]. This includes the Clinical Findings manifested by the patient and provenance data related with those assertions. In that earlier analysis, for each clinical case, at least 3 physicians plus the automated system proposed a set of diseases as possible diagnoses. Thus, our software generates individual Nanopublications describing these independent diagnoses. Since the original study did not capture the true provenance of how a diagnostic decision was made, our prototype executes a random algorithm to use clinical manifestations of the patient as supporting information for the diagnosis. For each supporting finding selected by the random algorithm a new Nanopublication file containing that information is created.

V. Conclusions and Future work

Representation of clinical data with provenance information is crucial in order to achieve high-quality, auditable, and reproducible clinical practice. Using the proposed models, the diagnostic process - the core of clinical practice - can now be rigorously tracked. The models enable the capture and representation of provenance information regarding two distinct but related processes: The rules describing the diagnostic framework, and provenance information about the diagnostic event itself. The first describes how the diagnostic knowledge base was created/curated, including information about where the knowledge originated, and who selected the diagnostic criteria. These knowledge bases can be explored to obtain precise information about diseases definitions, according to different sources; moreover, the desired rules can then be individually selected, resulting in a personalized decision support system containing diagnostic rules generated by experts (or expert-systems) with differing opinions. The second keeps track of how and why a clinician or expert-system reached a diagnosis from the patient data - what rules did they apply, and why. This will allow the capture of detailed information about expert diagnostic procedures, potentially enabling the development of more accurate decision support systems through learning from past errors, or simply recording information for auditing purposes.

Future work will be focused on the extension of the current approach to include additional knowledge related to the diagnosis process and the elements that are part of it such as treatments, drug usage or prognosis, each with its own provenance information. As part of the future efforts, an analysis of the expressiveness of our model in comparison with other approaches will also be taken into account as well as a performance evaluation of the use of this data.

Acknowledgment

This work was supported by the Isaac Peral and Marie Curie COFUND Programmes of Polytechnic University of Madrid, undertaken at the Centre for Plant Biotechnology and Genomics UPM-INIA.

References

[1] R. S. Ledley and L. B. Lusted, "Reasoning Foundations of Medical Diagnosis Symbolic logic, probability, and value theory aid our understanding of how physicians reason," *Science*, vol. 130, no. 3366, pp. 9–21, Jul. 1959.

[2] A. Rodríguez-González and G. Alor-Hernández, "An approach for solving multi-level diagnosis in high sensitivity medical diagnosis systems through the application of semantic technologies," *Computers in Biology and Medicine*, vol. 43, no. 1, pp. 51–62, Jan. 2013.

[3] B. Mons, H. van Haagen, C. Chichester, P.-B. 't Hoen, J. T. den Dunnen, G. van Ommen, E. van Mulligen, B. Singh, R. Hooft, M. Roos, J. Hammond, B. Kiesel, B. Giardine, J. Velterop, P. Groth, and E. Schultes, "The value of data," *Nat Genet*, vol. 43, no. 4, pp. 281–283, Apr. 2011.

[4] V. Podgorelec, B. Grašič, and L. Pavlič, "Medical diagnostic process optimization through the semantic integration of data resources," *Comput. Methods Prog. Biomed.*, vol. 95, no. 2, pp. S55–S67, Aug. 2009.

[5] Á. García-Crespo, A. Rodríguez, M. Mencke, J. M. Gómez-Berbís, and R. Colomo-Palacios, "ODDIN: Ontology-driven differential diagnosis based on logical inference and probabilistic refinements," *Expert Systems with Applications*, vol. 37, no. 3, pp. 2621–2628, Mar. 2010.

[6] C. Tao, G. Jiang, T. A. Oniki, R. R. Freimuth, Q. Zhu, D. Sharma, J. Pathak, S. M. Huff, and C. G. Chute, "A semantic-web oriented representation of the clinical element model for secondary use of electronic health records data," *J Am Med Inform Assoc*, vol. 20, no. 3, pp. 554–562, May 2013.

[7] Martínez-Romero, Vázquez-Naya, M, J. Pereira, M. Pereira, A. Pazos, G. Baños, "The iOSC3 System: Using Ontologies and SWRL Rules for Intelligent Supervision and Care of Patients with Acute Cardiac Disorders," *Computational and Mathematical Methods in Medicine*, vol. 2013, Feb. 2013.

[8] C.-S. Lee and M.-H. Wang, "A Fuzzy Expert System for Diabetes Decision Support Application," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 1, pp. 139–153, 2011.

[9] A. R. Rodríguez González, G. H. Hernandez Chan, R. Colomo Palacios, J. M. Gomez Berbis, A. Garcia Crespo, G. Alor Hernandez, and R. Valencia Garcia, "Towards an Ontology to support semantics enabled Diagnostic Decision Support Systems," *Current Bioinformatics*, 2012.

[10] L. Moreau, J. Freire, J. Futrelle, R. E. McGrath, J. Myers, and P. Paulson, "The Open Provenance Model: An Overview," in *Provenance and Annotation of Data and Processes*, J. Freire, D. Koop, and L. Moreau, Eds. Springer Berlin Heidelberg, 2008, pp. 323–326.

[11] P. P. da Silva, D. L. McGuinness, and R. Fikes, "A Proof Markup Language for Semantic Web Services," *Inf. Syst.*, vol. 31, no. 4, pp. 381–395, Jun. 2006.

[12] S. Miles, P. Groth, S. Munroe, and L. Moreau, "PrIMe: A Methodology for Developing Provenance-aware Applications," *ACM Trans. Softw. Eng. Methodol.*, vol. 20, no. 3, pp. 8:1–8:42, Aug. 2011.

[13] Satya S. Sahoo and Amit Sheth, "Provenir ontology: Towards a Framework for eScience Provenance Management," presented at the Microsoft eScience Workshop, Pittsburgh, PA, USA, 2009.

[14] A. Callahan and M. Dumontier, "Ovopub: Modular data publication with minimal provenance," *arXiv:1305.6800 [cs]*, May 2013.

[15] T. Clark, P. N. Ciccarese, and C. A. Goble, "Micropublications: a Semantic Model for Claims, Evidence, Arguments and Annotations in Biomedical Communications," *arXiv:1305.3506 [cs]*, May 2013.

[16] P. Groth, A. Gibson, and J. Velterop, "The anatomy of a nanopublication," *Information Services and Use*, vol. 30, no. 1, pp. 51–56, Jan. 2010.

[17] G. P. Patrinos, D. N. Cooper, E. van Mulligen, V. Gkantouna, G. Tzimas, Z. Tatum, E. Schultes, M. Roos, and B. Mons, "Microattribution and nanopublication as means to incentivize the placement of human genome variation data into the public domain," *Hum. Mutat.*, vol. 33, no. 11, pp. 1503–1512, Nov. 2012.

[18] S. F. Saccone, J. Quan, and P. L. Jones, "BioQ: tracing experimental origins in public genomic databases using a novel data provenance model," *Bioinformatics*, vol. 28, no. 8, pp. 1189–1191, Apr. 2012.

[19] J. P. McCusker and D. L. McGuinness, "Explorations into the Provenance of High Throughput Biomedical Experiments," in *Provenance and Annotation of Data and Processes*, D. L. McGuinness, J. R. Michaelis, and L. Moreau, Eds. Springer Berlin Heidelberg, 2010, pp. 120–128.

[20] J. Hastings, L. Chepelev, E. Willighagen, N. Adams, C. Steinbeck, and M. Dumontier, "The Chemical Information Ontology: Provenance and Disambiguation for Chemical Data on the Biological Semantic Web," *PLoS ONE*, vol. 6, no. 10, p. e25513, Oct. 2011.

[21] The Open PHACTS RDF/Nanopublication Working Group, "Nanopub Guidelines (v1.8.1)." 26-Mar-2012.

[22] A. J. Williams, L. Harland, P. Groth, S. Pettifer, C. Chichester, E. L. Willighagen, C. T. Evelo, N. Blomberg, G. Ecker, C. Goble, and B. Mons, "Open PHACTS: semantic interoperability for drug discovery," *Drug Discov. Today*, vol. 17, no. 21–22, pp. 1188–1198, Nov. 2012.

[23] J. J. Carroll, C. Bizer, P. Hayes, and P. Stickler, "Named Graphs, Provenance and Trust," in *Proceedings of the 14th International Conference on World Wide Web*, New York, NY, USA, 2005, pp. 613–622.

[24] A. Rodríguez-González, J. E. Labra-Gayo, R. Colomo-Palacios, M. A. Mayer, J. M. Gómez-Berbís, and A. García-Crespo, "SeDeLo: using semantics and description logics to support aided clinical diagnosis," *J Med Syst*, vol. 36, no. 4, pp. 2471–2481, Aug. 2012.

[25] A. Rodriguez, J. E. Labra, G. Alor-Hernandez, J. M. Gomez, and R. Posada-Gomez, "ADONIS: Automated diagnosis system based on sound and precise logical descriptions," in *Computer-Based Medical Systems, 2009. CBMS 2009. 22nd IEEE International Symposium on*, 2009, pp. 1 –8.

[26] C. Casado-Lumbreras, A. Rodríguez-González, J. M. Álvarez-Rodríguez, and R. Colomo-Palacios, "PsyDis: Towards a diagnosis support system for psychological disorders," *Expert Systems with Applications*, vol. 39, no. 13, pp. 11391–11403, Oct. 2012.

[27] A. R. González, J. M. Á. Rodríguez, C. Casado-Lumbreras, and R. Colomo-Palacios, "Towards an Ontology for Psychological Disorders," *Int. J. Metadata Semant. Ontologies*, vol. 7, no. 4, pp. 260–268, Jan. 2012.

[28] M. Dean and G. Schreiber, "OWL Web Ontology Language Reference," W3C, 2004.

[29] S. Samadian, B. McManus, and M. D. Wilkinson, "Extending and encoding existing biological terminologies and datasets for use in the reasoned semantic web," *J Biomed Semantics*, vol. 3, no. 1, p. 6, 2012.

[30] S. Samadian, B. M. Good, B. McManus, and M. D. Wilkinson, "A data-driven approach to automatic discovery of prescription drugs in cardiovascular risk management," *Bio-Ontologies 2012*, 04-Oct-2012. [Online]. Available: http://bio-ontologies.knowledgeblog.org/358. [Accessed: 21-Jan-2014].

[31] J. Lehmann, S. Auer, L. Bühmann, and S. Tramp, "Class expression learning for ontology engineering," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 9, no. 1, pp. 71–81, Mar. 2011.

[32] P. Ciccarese, S. Soiland-Reyes, K. Belhajjame, A. J. G. Gray, C. Goble, and T. Clark, "PAV ontology: Provenance, Authoring and Versioning," *Journal of Biomedical Semantics*, vol. 4, no. 1, p. 37, 2013.

[33] A. Powell, M. Nilsson, A. Naeve, and P. Johnston, "DCMI Abstract Model," Mar. 2005.

[34] Samadian, S., McManus, B., and Wilkinson, D., "Automatic detection and resolution of measurement-unit conflicts in aggregated data," presented at the The 3rd Annual Translational Bioinformatics Conference (TBC/ISCB-Asia 2013), Seoul, Korea, 2013.

[35] A. Rodríguez-González, J. Torres-Niño, M. A. Mayer, G. Alor-Hernandez, and M. D. Wilkinson, "Analysis of a Multilevel Diagnosis Decision Support System and Its Implications: A Case Study," *Computational and Mathematical Methods in Medicine*, vol. 2012, Dec. 2012.

[36] *DDx2NP webpage.* Available online at: http://wilkinsonlab.github.io/DDx2NP/ .2014.