Journal of Library Metadata
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/wjlm20

The COPD Ontology and Toward Empowering Clinical Scientists as Ontology Engineers

Jane Greenberg [a], Rahul Deshmukh [a], Lina Huang [a], Javed Mostafa [a], Lisa La Vange [b], Elizabeth Carretta [b] & Wanda O'Neal [c]

[a] School of Information and Library Science, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[b] Department of Biostatics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[c] School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Available online: 20 Oct 2010

To cite this article: Jane Greenberg, Rahul Deshmukh, Lina Huang, Javed Mostafa, Lisa La Vange, Elizabeth Carretta & Wanda O'Neal (2010): The COPD Ontology and Toward Empowering Clinical Scientists as Ontology Engineers, Journal of Library Metadata, 10:2-3, 173-187

To link to this article: http://dx.doi.org/10.1080/19386389.2010.520604

Journal of Library Metadata, 10:173–187, 2010

ISSN: 1938-6389 print / 1937-5034 online
DOI: 10.1080/19386389.2010.520604

The COPD Ontology and Toward Empowering Clinical Scientists as Ontology Engineers

JANE GREENBERG, RAHUL DESHMUKH, LINA HUANG, and JAVED MOSTAFA
School of Information and Library Science, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

LISA LA VANGE and ELIZABETH CARRETTA
Department of Biostatics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

WANDA O'NEAL
School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Ontology development and maintenance is a costly undertaking, despite known benefits. Empowerment via ownership may offer a solution to this problem. This article considers empowering clinical scientists as ontology engineers and illustrates this concept through an account of an ontology project in Chronic Obstruction Pulmonary Disease (COPD). The article begins with a brief review of ontology and clinical science. Next, the SPIROMICS project is introduced and the inter-workings (nuts-and-bolts) of the current COPD ontology are described, including an overview of the ontology development team's accomplishments. Following is an initial high-level proposal of steps to engage clinical scientists in the COPD work—as engineers, as well as some processing guidelines. The final section presents conclusions and highlights next steps.

KEYWORDS *ontology, ontological engineering, Chronic Obstruction Pulmonary Disease (COPD)*

Digital technology has had a tremendous impact on the complete scientific enterprise from data collection and preservation at the base level, to analysis

Address correspondence to Jane Greenberg, School of Information and Library Science, University of North Carolina at Chapel Hill, 205 Manning Hall, CB #3360, Chapel Hill, NC 27599-3360, USA. E-mail: janeg@email.unc.edu

and synthesis at higher-level goals (Greenberg, White, Carrier, & Scherle, 2009). These developments both feed and emanate from the swift progress underlying our cyberinfrastructure and contribute to an e-Science framework. A defining feature is the flood digital data (Carlson, 2006), a result that has encouraged the exploration of Knowledge Organization Systems (KOS) for information management along with innovative partnerships between scientists and informaticians.

Knowledge organization systems, in the form of controlled vocabularies, taxonomies, and ontologies, are sophisticated information systems that require intellectual analysis in both their development and use. These systems are used in conjunction with descriptive metadata standards that guide the recording object properties, such as authorship/creator, date of creation, and size. They are, however, more costly to develop and maintain than bibliographic descriptive systems. The reason is simple: Knowledge is complex and ever-changing. Cost issues are, in some respects, magnified with ontologies, given their granularity and the need for domain expert involvement.

As scientists continue to demand greater intellectual control of their digital stores and seek more sophisticated operations such as automatic data synthesis, there is a growing need to engage them more firmly in the ontology enterprise. This need is evident with the development of the Chronic Obstruction Pulmonary Disease (COPD) ontology, as part of the Subpopulations and Intermediate Outcome Measures In COPD Study (SPIROMICS). The work is in the early stages, although preliminary steps are underway to engage clinical scientists as domain experts through ownership and have them contribute to the COPD ontology as engineers.

This article provides an account of the initiation of the SPIROMICS COPD ontology project. The article begins with a brief review of ontology and clinical science. Next the SPIROMICS project is introduced and the inter-workings (nuts-and-bolts) of the current COPD ontology are described, including an overview of the ontology development team's accomplishments. Following is an initial high-level proposal of steps to engage clinical scientists in the COPD work, as well as some processing guidelines. The final section is a conclusion and highlights next steps.

ONTOLOGY: A BRIEF INTRODUCTION

Ontologies, as structured semantic systems, are becoming more commonly discussed and understood in the digital library and repository community. The concept *ontology* is rooted in the field of philosophy, specifically metaphysics—a branch of philosophy that studies the nature of being (Smith & Welty, 2001). Ontological engineering as a computer science practice emerged in the 1980s. The initial context was artificial intelligence supporting

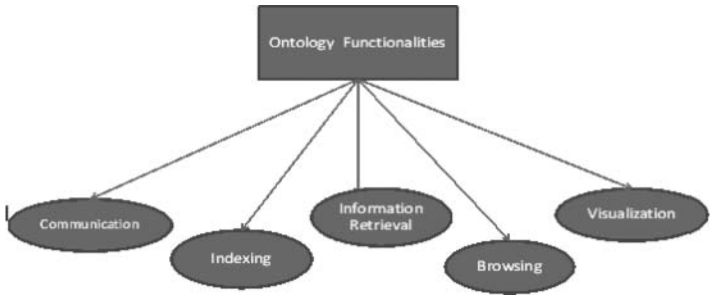

FIGURE 1 Ontology supported functions.

formal reasoning systems (McCarthy, 1980). Ontology made its way into the field of library and information science in the early-to-mid 1990s. This foray can be largely attributed to interdisciplinary collaborations sparked by the development of the networked technology and the World Wide Web.

Ontology is now understood through a range of definitions, extending from structured/rule-based systems to abstract representations of a world (Yin, Lee, & Yang, 2009). On the structured end, ontologies support formal reasoning; these semantic systems are promoted as a key component of the Semantic Web. The Web Ontology Language (OWL), a World Wide Web Consortium (W3C) standard for encoding ontologies, is fundamental to this work. On the more philosophical end, ontologies are studied as world views, relating to language and communication within a specific environment or community (Srinivasan, 2005, 2006).

Today, digital library ontology work is generally pursued for indexing and annotating, retrieving information, browsing, and communicating (see Figure 1). These common ontology-driven functionalities enable digital information systems to operate effectively. These desired functions point to precisely why initiatives inundated with an ever increasing amount of digital data adopt, modify, and even develop ontologies and ontology-like applications (e.g., controlled vocabularies, taxonomies, and classificatory systems) for managing information.

CLINICAL SCIENCE AND ONTOLOGIES

Clinical science, as with nearly any scientific research stream today, has embraced digital technology as a chief application for conducting and disseminating research results. A consequence of this evolution includes massive amounts of digital data, which has in turn sparked a growing interest in ontologies for data management. The following section provides insight into this change by defining clinical science and presenting examples of ontology work in this realm.

Clinical Science

Clinical science is a research approach that targets human health and usually involves experimental design. The work is generally evidence-based, and measurement emphasizes outcomes. Clinical science integrates with transactional science activities, where there is growing interest on moving bench science to the bed-side (National Institutes of Health, 2009).

Clinical scientists rely on large amounts of data associated with research literature and stored in data archives. They also continuously generate data via experimental work and research activities. Clinical science initiatives seek digital technologies for recording, synthesizing, and managing data. Clinical Data Management Systems (CDMS) have been developed to help with growing data management needs. These systems also help scientists to comply with federal regulations, aiding them with documentation and accountability.

Ontologies Take Hold in Clinical Science

Interest in ontologies for clinical science has increased in concert with the growth in digital data and emergence of CDMS. In fact, in 2007 the National Center for Biomedical Ontology (NCBO) hosted a two-day workshop to engage researchers in creating a reference ontology for annotating clinical trial results.[1] Although to date there is no single, shared ontology for annotating clinical science research, the NCBO Bioportal[2] has become a leading pathway providing access to a wide variety of domain ontologies useful for managing clinical science data.

As CDMS grow in size and number, there is an increased need for system functionalities supporting information discoverability, access, and use. Ontologies, as already reviewed (see Figure 1), can meet these needs. The following are three examples of ontology work in clinical science:

- *Information retrieval and reasoning*. Researchers from Columbia University Medical Center and IBM Watson Research Center modified the SNOMED CT (Systematized Nomenclature of Medicine-Clinical Terms—Clinical Terms) ontology to support reasoning and automate common clinical tasks, such as retrieving patient records (Patel et al., 2007). The work was an effort to bridge the 'semantic gulf' between raw data and the way a clinician interprets the data.
- *Communication*. Researchers from Philips Research Europe and Aachen University developed an ontology-based system for clinical trial data management. The prototype system includes a 'reference' ontology for generating clinical trial data and integrating data from other sources (Geisler, Brauers, Quix, & Schmeink, 2007). The ontology serves as a common knowledge base—at an abstract level, and aids database design.

- *Indexing and annotation.* Trial Bank Project[3] researchers introduced an ontology effort focusing on mother-to-child transmission (MTCT) prevention and structured treatment interruptions (STI) by structuring, summarizing, and visualizing trial data. The project supports annotations for HIV/AIDS clinical trials through two ontology-driven applications—CTSearch and CTExplorer.

These projects, including the NCBO Bioportal, confirm scientists' awareness of ontologies. One might even argue that ontology development relating to clinical science has been significant, as there is a growing interest in integrating these semantic systems into the scientific workflow. As this growth continues, basic design requirements are easier to pinpoint, including the need for an infrastructure accommodating the highly collaborative nature of successful ontology work. This is the case with the University of North Carolina's COPD ontology, pursued via the SPIROMICS, and with a strong partnership involving *informaticians* and *domain experts*.

SPIROMICS (SUBPOPULATIONS AND INTERMEDIATE OUTCOME MEASURES IN COPD STUDY) AND THE COPD ONTOLOGY

SPIROMICS provides the impetus and foundation for the COPD ontology. The following section introduces SPIROMICS and provides contextual information relating to the COPD ontology.

SPIROMICS

SPIROMICS is a prospective cohort study funded by the National Heart, Lung, and Blood Institute (NHLBI) to identify COPD subpopulations for targeted enrollment in future clinical trials and intermediate outcome measures for use as endpoints in clinical trials. SPIROMICS will enroll approximately 3,200 subjects at six clinical centers over three years. Subjects will be distributed across four enrollment strata (i.e., Non-smokers, Smokers without COPD, Mild/Moderate COPD patients, and Severe COPD patients).

The study is coordinated by the Genomics and Informatics Center (GIC) established at the Collaborative Studies Coordinating Center (CSCC) in the Department of Biostatics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill. The GIC is comprised of a cross-campus partnership involving researchers from the School of Medicine and the School of Information and Library Science in addition to faculty in the Biostatics Department.

SPIROMICS includes extensive data collection of phenotypic, biomarker, genetic and genomic data. These data are necessary for the attainment of

SPIROMICS' overarching goal of helping to accelerate the development of new COPD therapies in future clinical trials. Key to accomplishing the study's research goals is the availability of bioinformatics resources.

The COPD Ontology

The COPD ontology is a key aspect of the bioinformatics activity. The COPD ontology is critical for data and resource sharing within the current study. Additionally, the ontology has long-term implications for longitudinal research, as well as external COPD investigations.

As already noted, COPD ontology development has been supported via a partnership involving informaticians and domain experts. The informaticians have been responsible for constructing the technical infrastructure and vocabulary application, and for seeding the vocabulary with concepts drawn from standard COPD resources and important literature. Domain experts are beginning to play a crucial role in the ontology development process by verifying foundational work and contributing terminology. As the ontology is integrated into workflow, domain experts will need to become further engaged in ontology maintenance.

The informatician/domain expert partnership is a stronghold of the ontology work, and has been both reinforced by stakeholder interest and nourished by GIC team leadership. The core GIC team meets bi-monthly at UNC, and generally includes an ontology update, presented by informatician researchers. The COPD ontology progress has also been shared with the study's Steering Committee, which includes members from clinical centers pursuing related research as part of the larger NHLBI SPIROMICS initiative. Steering Committee members are principal investigators of clinical centers at Columbia University, University of California at Los Angeles, University of California at San Francisco, University of Michigan, University of Utah, Wake Forest University, and a Radiology Reading Center at the University of Iowa as well as an external Steering Committee chair and the NHLBI project officer.

SPIRO-V: THE NUTS-AND-BOLTS OF THE COPD ONTOLOGY

The COPD ontology is being developed via a web-based system called SPIRO-V (SPIROMICS-Vocabulary). The SPIRO-V interface has been designed to support domain expert contributions, and includes a feedback feature for commentary on candidate terms. SPIRO-V's interactive front end is powered by Google Web Toolkit (GWT; Hanson & Tacy, 2007), and a MySQL relational database stores terms, relationships, domain expert commentary and feedback.

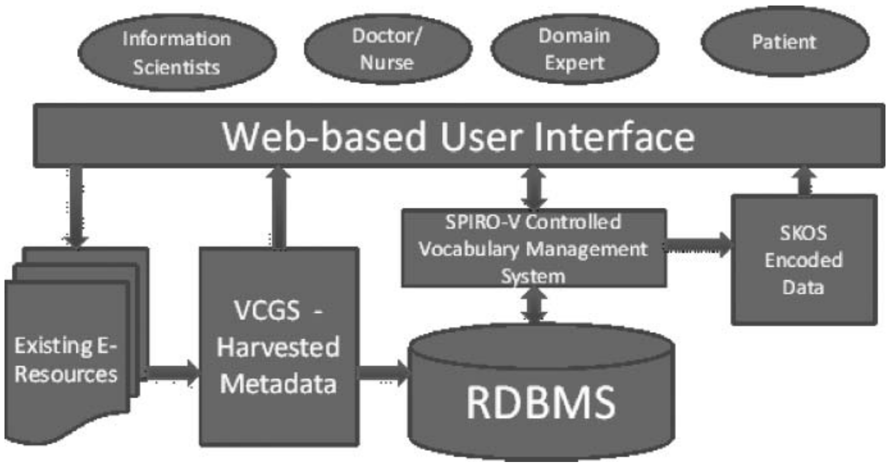

FIGURE 2 SPIRO-V architecture.

The SPIRO-V architecture plans to employ the Simple Knowledge Organization System (SKOS),[4] a World Wide Web consortium (W3C) encoding standard supporting interoperability among semantic systems. SKOS provides widespread potential for sharing the COPD ontology in domain-related initiatives beyond the immediate SPIROMICS initiative. Use of SKOS will also permit the COPD ontology to interoperate with other SKOS and OWL encoded vocabularies registered at the NCBO. And, on some level, the use of SKOS will permit the COPD ontology to be viewed as a Semantic Web contribution, supporting the construction of a more intelligent web of *linked data* (Bizer, Heath, & Berners-Lee, 2009). The system will also include a Vocabulary Cluster Generation System (VCGS; Fu et al., 2002)—an algorithm that automatically extracts concepts from selected sources for inclusion in the COPD ontology (Huang, Deshmukh, Mostafa, & Greenberg, 2010). Figure 2 shows the SPIRO-V architecture.

The first release of SPIRO-V system was completed in April 2010. The system can be accessed via the web. Currently a password can be requested, although the longterm plan is to make it publicly accessible. This current SPIRO-V release includes a series of functions (Table 1) for assisting domain experts in using, contributing to, and maintaining the COPD ontology.

TABLE 1 Summary of SPIRO-V Functionalities

- Browse and search all terms in current vocabulary set.
- Submit candidate terms to be included in the COPD ontology.
- Modify, add, and delete term relationships (e.g., equivalent, hierarchical, and associative term relationships) for all registered terms.
- Record annotations for selected terms, such as commentary suggesting new wording to improve a term definition.

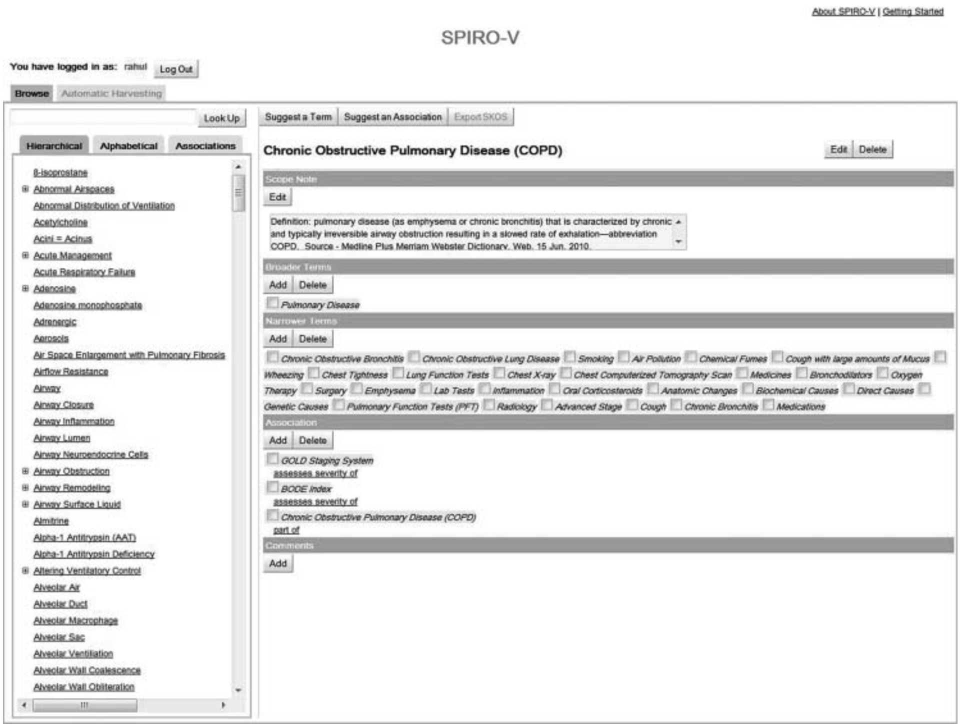

FIGURE 3 SPIRO-V/COPD ontology entry for *Chronic Obstructive Pulmonary Disease.*

The most recent COPD ontology development efforts have focused on adding scope notes containing definitional information for individual ontology concepts, and targeting domain expert for review of the notes and concepts. Figure 3 is an example of a COPD entry for the term *Chronic Obstructive Pulmonary Disease* with a scope note.

SPIRO-V and COPD ontology development priorities, outlined for the next year, include:

1. Integrating the VCGS automatic algorithm into core system.
2. Establishing guidelines to aid domain experts in contributing, maintaining, and using SPIRO-V.

The SPIRO-V and COPD ontology work accomplished to date (in a little over a year's time) is noteworthy, given that the effort was pursued without a base-level ontology. In other words, the effort was initiated with a blank slate. Currently, SPIRO-V can boast the implementation of an operational and highly interoperable ontology containing over 500 authorized concepts and close to 800 conceptual relationships.

Progress aside, there are well-known human resource and system maintenance challenges that need to be addressed in order to continue the COPD ontological work. By integrating the VCGS, we will take advantage of machine capabilities, and ease the human burden required for growing the COPD ontology. Furthermore, by providing guidelines, we seek to empower the domain experts as ontology owners to ensure sustainability. This latter solution represents a growing shift in science, where subject matter experts are more frequently being called on to help build and maintain vocabularies. We express this goal as a step toward empowering domain experts as ontological engineers, and present initial activities in this area in the remainder of this article.

EMPOWERING CLINICAL SCIENTISTS AS ONTOLOGICAL ENGINEERS

Urban studies and social science research provide evidence of a connection between *property ownership*, *empowerment*, and *maintenance* (Somerville, 1998). Empowerment via ownership, including co-ownership, has inspired the development of co-housing communities. The environs engage individuals in shared care of community centers, gardens, and other facilities. Although an ontology is not a physical entity like a dwelling, an ontology has a tremendous amount of intellectual value. In fact, an ontology might be viewed as an *object of collective intelligence*, particularly given that ontologies are generally the product of multiple contributors.

Following the notion that an ontology is an object of sorts, it seems reasonable to suggest that ontology development and maintenance can benefit from co-ownership. Ontologies are knowledge organization systems, and reflect conceptual relationships within a domain or discipline. Knowledge is dynamic, particularly in scientific disciplines, where the discovery and the presentation of new knowledge is the chief objective. Maintenance is therefore required for an ontology to remain relevant. Indeed, a commitment is required not only at Gruber's emphasis on strict adherence and use (Gruber, 1992), but active commitment feeding the development and maintenance. It is the active commitment by body or group of individuals who, in essence, are *engineers* performing tasks required to ensure the ontology's vitality.

The proposition here is that by joining ontology users and empowering them with a voice, a collective group can help pave a direction for ontology maintenance and long-term sustainability. Clinical scientists, as domain experts, possess in-depth discipline knowledge that is of important value for an ontology; and through ownership they should, on some level, view the ontology as a property of collective responsibility. Using this concept, the

GIC team is taking steps to engage COPD ontology users as "engineers" so that they may actively participate in the continued development and maintenance of this resource.

The SPIRO-V platform has been developed in-line with this goal. The SPIROMICS Steering Committee has convened a subcommittee comprised of investigators from each study site (domain experts) to take on COPD ontology responsibilities on a rotating basis and operate as ontological engineers. To move forward, the GIC ontology team has proposed a high-level governing structure comprised of an initial set of individual roles and responsibilities for ensuring the quality of COPD ontology. System privileges are based on individual role. As of now, all roles are partially supported by the SPIRO-V application. We plan to make them fully functional by the next release. The three initial roles include: contributor, editor, and system administrator. A brief explanation of these three roles follows:

- **Contributor**: A contributor creates and edits the term records. Approval by an editor is required to finalize the contributor's work. The contributor cannot delete term records.
- **Editor**: An editor reviews and edits ontology records, approves contributor work, and can delete ontology records. An editor can also request that the system administrator archive term records.
- **System administrator**: A system administrator has the highest level of access and rights to perform all administrative functions such as assigning and maintaining user roles, and approving, rejecting, archiving and deleting data. A system administrator can edit and delete access to all items in the ontology.

These roles identified here form a governing structure. Proposed qualifications, number of participants, and the duration of service are presented in Table 2. It should be made clear, however, that this is the very first iteration of this work, and any plan endorsed by the SPIROMICS Steering Committee

TABLE 2 Proposed Role, Qualification, and Service for Ontology Engineering

ROLE*	QUALIFICATION	SERVICE: Participants and duration
Contributor	Clinical scientist/ domain expert	2 to 4 contributors, with service for a 6 month period
Editor	Clinical scientist/ domain expert	1 to 2 editors, acting as co-editors, for a 6 month period
System Administrator	Informatician (e.g., information architecture, database personnel, information scientists)	1 expert, on a contractual basis, or integrated with other annual projects

*Responsibilities outlined above.

may look very different. Even so, it is useful to record the initial thinking of the GIC SPIRO-V development team, and the ideas that will be proposed in response to discussion from the March 2010 SPIROMICS Steering Committee meeting.

The initial proposal, summarized in Table 2, proposes between two to four participants take on contributor responsibilities, and between one to two participants take on editorial responsibilities. Over the last year, two informaticians working in the partner School of Information and Library Science, University of North Carolina, Metadata Research Center, and the Laboratory of Applied Informatics Research (LAIR), have been working as both contributors and editors. As contributors, they have drawn terminology from resources addressing relevant COPD topics, created scope notes, and established terminological relationships; and as editors, they have worked collaboratively to review all contributions, making necessary modifications. Two domain experts have assisted in reviewing this initial work that has seeded the vocabulary. The long-term plan is to more fully transfer the contributor and editorship responsibility to the domain experts and consider appointment durations for individuals taking on these roles.

System administration functionalities are also currently being overseen by the informaticians. Even as COPD ontology ownership becomes the province of the domain experts, it will be important to continue a relationship with an informatician to ensure long term system stability, perform general system updates maintenance, and provide important knowledge relating to ontology standards, registry protocols, and other systematic and conceptual issues regarding knowledge organization and ontology development. This type of professional relationship is something that is being worked out as we see development like NCBO and partnerships across disciplines where scientific research, including clinical scientists, embracing wanting to fully benefit from digital technology.

TABLE 3 Protocol for Adding New Terms to SPIRO-V

1. Any new concept will be stored as a ‘Candidate Term’ in the system and the Domain Experts will be notified about its entry into the system. The concept can either be suggested by an authorized user or automatically harvested using the VCGS system.
2. The experts will then discuss the concept and give their comments about how and where in the ontology should it be placed. If needed, they will also provide references or ‘scope notes’ for the concept.
3. The moderators will then decide whether to add the concept to the ontology and if yes, at what location.
4. The concept will then be a part of the ontology and will also display the scope notes that were added to it during its candidacy.
5. Once the term is added, its associations with other existing terms can be suggested by the experts in the same way.

TABLE 4 Ontology Development Process

1. The contributor and the editor both have the right to propose the development of a new term. This step is initiated by suggesting a new term on SPIRO-V. Whenever a new term is suggested, a comment should also be added along with the new term to define why the term is needed.
2. When suggesting a new term, the record should at least contain the name of the term, the scope note, and the comment.
3. The status of the added term is marked as "Under Review" by the SPIRO-V system. And the editors will be notified that the new terms are available to review.
4. The editor should initiate the review process. The editor will propose additions, corrections and any other changes. If only minor modifications are needed, the editor may approve the changes with a notice to the contributor. If there are major modifications, the editor will reject the new terms and send a notice to the contributor.
5. The contributor revises the submission based on the notice sent by the editor. And within a short amount of time, the contributor can either abandon the submission or submit a revised version.
6. The editor will mark the accepted changes as "Approved" and make an announcement to the community.

In line with the proposal roles and responsibilities, the GIC ontology team has drafted the following three items:

1. A protocol for adding new terms to a SPIRO-V like system (Table 3)
2. A set of guidelines for ontology creation process (Table 4)
3. A set of guidelines for ontology maintenance (Table 5)

Work here has been inspired and informed by the significant work of the EPA[5] terminologies project, and is reflective of extensive work being conducted there to reach out to domain scientists and engage them in terminology registry across the agency.

This protocol calls for active participation from the domain experts, all the while making sure that the system complexities do not obscure work, and rather users are presented with a seamless display user-friendly access to a selection of desired functionalities.

Two processes have been articulated for ensuring quality of COPD ontology: 1. the development process (Table 4) for actualizing the protocol

TABLE 5 Ontology Maintenance Process

1. Identify the need of maintenance: by schedule or initiated by the system administrator.
2. The system administrator will initiate as well as oversee the maintenance process.
3. The contributors and editors involved can propose additions, corrections of terms during the maintenance process.
4. If the changes are minor, the system administrator can approve and take the proper action; If there are major changes, the system administrator should validate the changes with the community and take the further action based on community feedback.
5. The system administrator incorporates the changes and makes a list of maintenance items to announce to the community.

Next step will focus on the further refinement of the guidelines and development of user manuals for the contributor, editor and system administrator respectively.

outlined in Table 3; and the maintenance process (Table 5). The preceding tables presented high-level steps for these processes.

CONCLUSIONS AND FUTURE WORK

This article presented briefly the emerging collaboration between ontologies and clinical science, discussed in detail how the SPIRO-V system is being developed and shaping the COPD ontology with the help of domain experts. The proposed steps to take to ensure more engagement of clinical scientists in the COPD work were discussed as well.

We end our article with a discussion about the future course of work that SPIRO-V will take to ensure that the system is widely used and supports the proposed functionalities.

Ontology development is an important component of SPIROMICS and should be beneficial to other research projects in COPD. The vision for the SPIRO-V Clinical Study Application is to build upon the collaboratively created Controlled Vocabularies and create a knowledge base ontology that will map to the data stored in the SPIROMICS database. This mapping will allow for the easy access, search and retrieval of the study data by SPIROMICS investigators with the help of the vocabularies as the pointers to the data. The SPIRO-V interface is not designed to be limited to SPIROMICS. The architecture of this interface could be used by other studies to create controlled vocabularies and mapped the ontology in the same manner as will be done in SPIROMICS.

Other important functionality that the system will provide is the Ontology Visualization. This will give the prospective users a web interface to the underlying information in a structured way by means of the ontology and the controlled vocabulary.

Visualization can help reduce the confusion that arises when a domain is not clearly defined and when the same terms are often used to mean different things. It gives a platform where discussions can take place and one or more definitions and scopes for terminologies can be agreed upon. Visualization techniques can represent a larger picture and different scenarios enabling new means for managing and understanding the data.

In conclusion, the SPIRO-V team is continuing ontology development to help users successfully navigate SPIROMICS clinical study stystem, and with a long-term goal making the COPD ontology widely accessible to researchers via an open portal.

ACKNOWLEDGMENT

The Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS) is funded by contract from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (HHSN268200900020C).

NOTES

1. National Center for Biomedical Ontology (NCBO) Clinical Trial Ontology workshop: http://www.bioontology.org/wiki/index.php/Workshop_on_Clinical_Trial_Ontology
2. NCBO Bioportal: http://bioportal.bioontology.org/
3. Trial Bank Project. http://rctbank.ucsf.edu/
4. Miles, A., Bechhofer, S.: SKOS Simple Knowledge Organization System—Reference. http://www.w3.org/TR/2009/REC-skos-reference-20090818/ (*W3C Recommendation* 2009/).
5. EPA cite: http://iaspub.epa.gov/sor_internet/registry/sysofreg/home/overview/home.do

REFERENCES

Bizer, C., Heath, T. & Berners-Lee, T. (2009). Linked data—the story so far. (World Wide Web). *International Journal on Semantic Web and Information Systems*, *5*(3), 1.

Carlson, S. (2006). Lost in a sea of science data. *Chronicle of Higher Education*, *52*(42), A35.

Fu, Y., Bauer, T., Mostafa, J., Palakal, M., & Mukhopadhyay, S. (2002). Concept extraction and association from cancer literature. In *Proceedings of the 4th International Workshop on Web Information and Data Management* (pp. 100–103). McLean, Virginia.

Geisler, S., Brauers, A., Quix, C., & Schmeink, A. (2007). Ontology-Based System for Clinical Trial Data Management. IEEE Benelux EMBS Symposium. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.69.2629

Greenberg, J., White, H., Carrier, S., & Scherle, R. (2009). A metadata best practice for a scientific data repository. *Journal of Library Metadata*, *9*(3/4), 194–212.

Gruber, T. (1992). *What is an ontology?* Retrieved from http://www-ksl.stanford.edu/kst/what-is-an-ontology.html

Hanson, R., & Tacy, A. (2007). *GWT in action: Easy ajax with the google web toolkit*. Greenwich, CT: Dreamtech Press.

Huang, L., Deshmukh, R., Mostafa, J., & Greenberg, J. (2010). SPIRO-V: a collaborative approach to controlled vocabularies gathering and management. *Proceedings of the 10th Annual Joint Conference on Digital libraries*, Gold Coast, Queensland, Australia, June 21–25, 2010, pp. 371–372.

McCarthy, J. (1980). Circumscription—A form of non-monotonic reasoning. *Artificial Intelligence*, *13*(1-2), 27–39.

National Institutes of Health. (2009). *Re-engineering the clinical research enterprise. Translational research*. Retrieved from http://nihroadmap.nih.gov/clinicalresearch/overview-translational.asp.

Patel, C., Cimino, J., Dolby, J., Fokoue, A., Kalyanpur, A., Kershenbaum, A., et al. (2007). Matching patient records to clinical trials using ontologies. In *Proc. of ISWC07*, volume 4825 of *LNCS*, (pp. 816–829). Berlin, Germany: Springer.

Smith, B., & Welty, C. (2001). Ontology: Towards a new synthesis. FOIS'01, (pp. 3–9). October 17–19, 2001, Ogunquit, Maine.

Somerville, P. (1998). Empowerment through residence. *Housing Studies*, *13*(2), 233–257.

Srinivasan, R. (2005). Weaving spatial, digital and ethnographic processes in community-driven media design. Doctoral Dissertation, Graduate School of Design, Harvard University.

Srinivasan, R. (2006). Indigenous, ethnic and cultural articulations of new media. *International Journal of Cultural Studies*, *9*(4), 497–518.

Yin, C., Lee, Y., & Yang, J. (2009). Ontology: The historical review and literature productivity analysis using bibliometric methodology from 1956 to 2008. *International Conference on Interaction Sciences* (ICIS '09; Vol. 403, pp. 1346–1350). Seoul, Korea, November 24–26.