# Project Title

Clinical Ontology for Translational Research Affinity Group (COnTRA-G)

# Project Leader

William Hogan, MD, MS (University of Arkansas for Medical Sciences)

# Brief Project Description

The goal of the affinity group is to bring together various groups throughout the consortium who are developing and using ontologies in support of translational science. It will serve as a forum for communication and dissemination, as well as to identify unmet needs for new ontology development and integration. Its purpose is to bring together a wide variety of expertise on the topic of ontologies to (1) foster and initiate sharing, re-use, and development of ontologies across the group, (2) inform ontology users on ongoing ontology research and ontology evaluation criteria, and, finally, (3) provide a large and inclusive basis for scientific debate regarding research questions in ontologies. It will not serve as an infrastructure for developing one specific “CTSA ontology” or as a committee to endorse specific ontologies for use in either in the Consortium or for translational research in general.

Examples of ongoing efforts in the Consortium, the participants of which support the formation of this group, include the Ontology of Biomedical Investigations (OBI), the Ontology for General Medical Science (OGMS), the Human Disease Ontology (DO), the Infectious Disease Ontology (IDO), the Ontology of Medically Related Social Entities (OMRSE), and the CTSAconnect project (eagle-i/VIVO). The group is also doing formative work on ontologies of procedures, biobanking, representation of human studies, clinical trials and CT.gov data extraction, as well as work on improving ontological relations and their definition and use in existing ontology efforts. Participants from other ontology development efforts are free and welcome to join the group.

# CTSA Strategic Goal(s) Addressed by Project

Goal 1: Build National Clinical and Translational Research Capability

*This goal involves improving “…all processes related to the development, approval, activation, enrollment, enhancement, and completion of clinical trials.” The proposed group will identify common definitions and standards for use in informatics systems that support these processes. In addition, this goal relies on cross-Consortium studies as addressed by Goal 3. The goal is to unify and enhance existing efforts.*

Goal 3: Enhancing Consortium-wide Collaborations

*Collaborating across two or more CTSA institutions on research requires informatics systems that implement common definitions, standards, etc. The proposed affinity group will identify appropriate definitions and standards across the translational spectrum and catalyze the CTSA Informatics community to address areas where existing ontologies are absent, insufficiently developed, or insufficiently integrated. These ontologies are critical for cross-institutional research collaborations.*

Goal 5: Advance T1 Translational Research

*Bridging the language of bench researchers and clinicians will be essential to the progress of T1 Translational Research. The proposed group will have an important role in advocating for the incorporation and uptake by the CTSA community of appropriate ontologies that enable translational research. For example, advocating for disease ontologies following the Open Biomedical Ontologies principles in coupling clinical and basic science data, a la the Gene Ontology that enables the annotation of genes along biological process, molecular function, and cellular component axes to enable semantic precision and enhance computational analysis.*

# Other KFCs/Expertise Required

Knowledge of, or interest in, ontologies to support translational science and a willingness to consider the best manner of their integration with basic science ontologies such as GO, PRO, ChEBI, etc.

# Background/Significance

Increasingly, data aggregated across research groups, institutions, disciplines, settings (the clinic, the hospital, the research lab, the classroom, the policymaker’s office, and so on), and continents are necessary to the progress of translational science. The lack of standard representations of diseases, treatments, outcomes, risk factors, environments, cells, proteins, genes, DNA sequences, researchers, patients, clinicians, educators, and so on is significantly inhibiting this aggregation and thus constitutes a substantial drag on the progress of translational science.

Furthermore, the size of the datasets, even before aggregation, is growing exponentially. Sifting through these data for the purposes of scientific discovery is now fully reliant on computational analysis. However, there is still much more progress that is possible if we can impart more knowledge to the computer and its algorithms.

Finally, these two problems are interrelated: algorithms that understand one representation of disease will not work on data collected using a different representation. Thus, without standard representations, duplicative work on algorithms must be done to enable the same kinds of discovery from two disparate data sets.

One major and necessary component of the solution to these problems, even for purposes other than translational science, is the development and use of standard ontologies. An ontology, per Smith, is “…a representational artifact, comprising a taxonomy as proper part, whose representational units are intended to designate some combination of universals, defined classes, and certain relations between them.”[[1]](#footnote-1) Rather going into too much detail with this definition, we instead summarize it by saying that an ontology captures the classifications that scientists, educators, healthcare practitioners, policy makers, and so on, use in their daily work. The goal is then that these individuals use standard codes for the types captured in the ontologies when recording their data, so that aggregation becomes possible. As these individuals make new discoveries and subsequently revise their classifications of the types of things in the world, the job of ontologists is to revise the ontologies to capture these changes.

Besides capturing the types and their taxonomy from a particular field of study and/or practice, the ontology also captures other relations such parthood, ontological dependence, location, and so on. When structured as axioms in a logical formalism, the taxonomic and relational knowledge captured in an ontology supports computer algorithms in their advanced analysis of the data, allowing them to draw new conclusions that were not possible before the axioms existed. Furthermore, adoption of standard ontologies allows the new algorithms to analyze any data set collected in accordance with the ontologies.

The net effect is that we vastly improve the efficiency with which humans can make discoveries from data, and maximize the potential of the data collected. Given the great cost to society in the collection of these data, this objective is of utmost importance and value.

It should be noted that the best way to develop and use ontologies in support of this objective is itself not a settled scientific question. Thus, ontology itself is a scientific field of inquiry that seeks the best methods and approaches to ontology development, use, and reuse.

We note that the OBO Foundry is the only effort of which we are aware that has developed a conceptual and practical framework to support the interface between ontology for “bench” and that for “beside”. Researchers at UT Southwestern Medical Center, Northwestern University, Oregon Health & Science University (OHSU), and UAMS in Arkansas have been leaders in the development of clinical ontologies that interface and interoperate with basic science ontologies like GO and PRO. One significant effort in this regard has been the participation of Drs. Scheuermann and Hogan in the creation, development, and extension of the Ontology for General Medical Science and the leadership of Dr. Kibbe in developing the DO. In addition, Dr. Scheuermann and his colleague at UT Southwestern, Lindsay Cowell, are among the creators and active developers of the Infectious Disease Ontology. Finally, Dr. Haendel and her colleague Dr. Torniai have initiated a cross-CTSA ontology project (CTSAconnect) to infer expertise based on researcher and clinician activities and products of research for the purposes of promoting translational research in cross-disciplinary teams of scientists.

However, whether these efforts are the best approach to ontology for translational science remains an open research question in the field of ontology. This forum will serve as a mechanism to communicate and disseminate various scientific results in this regard.

We also have engaged the support of leaders in the field of ontology, including representatives from the National Center for Biomedical Ontology, who not yet part of the CTSA Consortium for this group, and obtained their commitment to participate. In particular, Barry Smith is organizing the “CTSA Ontology Workshop” in April, 2012 (<http://bioontology.stanford.edu/wiki/index.php/CTSA_Ontology_Workshop>).

# Project Description

The aims of this project are (1) to raise awareness of efforts across the CTSA Consortium to create, maintain, and use high-quality ontologies to transform clinical and translational science, (2) to disseminate results of these efforts, and (3) to identify needs for future efforts. It will also serve as a link from the Consortium to the National Center for Biomedical Ontology. We will raise awareness of various ontology development efforts without prejudice to any particular approach. In addition the group is meant to trigger participants to contribute to development, enhancement, and integration of ontologies needed within CTSA efforts.

# Project Goals/Deliverables and Expected Outcomes

1. Identify and list ontology development efforts in the Consortium aimed at translational research
2. Describe the state-of-the-art in ontology evaluation by cataloging and summarizing ontology evaluation criteria from the literature
3. Identify needs for future ontology development
4. Identify best practices for integrating ontologies into software applications

# Links to Other CTSA Projects

We will maintain close communications and relationships with the data standards and interoperability affinity group (DIAG) and Omics Data Standards Working Group of the IKFC. We have the enthusiastic support for the COnTRA-G of Jesse Tennenbaum, leader of the Omics Data Standards Working Group. Given our focus on a smaller, well-circumscribed subset of the DIAG focus, we believe it appropriate that this AG exist separate from DIAG in a similar manner to the OMICS WG.

# Participating Institutions

|  |  |  |
| --- | --- | --- |
| **Institution name** | **Role in project** | **Contact (name)** |
| University of Arkansas for Medical Sciences | Lead | William Hogan |
| University of Texas Southwestern Medical Center | OGMS Representative | Richard Scheuermann |
| University of Texas Southwestern Medical Center | IDO Lead | Lindsay Cowell |
| Northwestern University | DO Lead | Warren Kibbe |
| Oregon Health & Science University | OBI Representative, CTSAconnect lead | Melissa Haendel |
| University of Arkansas for Medical Sciences | OMRSE lead | Mathias Brochhausen |
| University at Buffalo\* | OBO Foundry and NCBO Representative | Barry Smith |

\*University at Buffalo is not (yet) a CTSA Consortium member

# Last Update

April 24, 2011

# Last Update Author(s)

William Hogan, Mathias Brochhausen, Warren Kibbe, Melissa Haendel

1. Smith B, Kusnierczyk W, Schober D, Ceusters W, editors. Towards a reference terminology for ontology research and development in the biomedical domain. The Second International Workshop on Formal Biomedical Knowledge Representation: "Biomedical Ontology in Action" (KR-MED 2006); 2006; Baltimore, MD. [↑](#footnote-ref-1)