Capturing Descriptions of Human Studies for Federated Data Sharing: Frontline Experiences from Three Institutions

Journal:	*2012 AMIA Clinical Research Informatics Summit*
Manuscript ID:	AMIA-072-C2012
Manuscript Type:	Panel
Date Submitted by the Author:	21-Oct-2011
Complete List of Authors:	Sim, Ida; UCSF, Medicine Chakraborty, Swati; Duke University Lehmann, Harold; Johns Hopkins Mollah, Shamin
Classification for Review Purposes:	Track 1: Research and resource discovery, collaboration and sharing

Capturing Descriptions of Human Studies for Federated Data Sharing: Frontline Experiences from Three Institutions

Ida Sim, MD, PhD[1], Swati Chakraborty, MS[2], Harold P. Lehmann, MD, PhD[3], Shamim A. Mollah, MA[4] for the Human Studies Database Project

[1]University of California, San Francisco, CA; [2]Duke University, Durham, NC; [3]Johns Hopkins University, Baltimore, MD; [4]The Rockefeller University, New York, NY;

Abstract

Human studies, encompassing interventional and observational studies, are the most important source of evidence for advancing our understanding of health, disease, and treatment options. To promote discovery, the design and results of these studies should be made machine-readable for large-scale data mining, synthesis, and re-analysis. The Human Studies Database Project aims to define and implement a semantic web–based infrastructure for research institutions to share the design of their human studies, using the Ontology of Clinical Research (OCRe) as the reference semantics. OCRe models study features such as design type, interventions, and outcomes to support scientific query and analysis. As with most every data sharing effort, this project faces both social and technical challenges. This panel will give an overview of our technical approach to semantically standardized data federation of human study protocols, and will present the sociotechnical experiences of three academic research institutions as they move towards sharing human studies information.

General Description

Human studies, encompassing interventional and observational studies, are one of the most central and valuable activities in biomedical research. To advance clinical and translational research, the design and results of human studies should be federated and shared for large-scale data mining, synthesis, and re-analysis. Because human studies data and information reside in many different information systems (e.g., electronic IRBs, clinical trial management systems, paper protocol documents), a shared semantics is necessary for effective data sharing and computation. These shared semantics must capture the critical features of clinical study design so that researchers can search and analyze across disparate studies with clinical and methodological precision.

The Human Studies Database Project (HSDB)[1] is a multi-institutional CTSA-affiliated project to build a federated database of the design of human studies being conducted at participating institutions. We have defined the Ontology of Clinical Research (OCRe)[2] – an OWL 2.0 project – to model a typology of study designs, as well as key features of the interventions, outcomes, analyses, and basic administration of quantitative human studies. We have defined a data sharing architecture using OCRe as the reference semantics, in which 1) the OCRe OWL model is automatically transformed into an XML Schema that refers back to OCRe entities using the Semantic Annotations for SWDL (SASWDL) technology; 2) the individual study protocols are instantiated as XML instances; 3) the instances are made available on the Internet from the respective institutions' web servers; and 4) the Query Integrator from the University of Washington is then used to simultaneously query over the instances as well as OCRe and clinical terminologies (e.g., SNOMED) through subsumption and other queries to BioPortal [http://bioportal.bioontology.org/],. We have previously presented an end-to-end demonstration of this data sharing approach[3], with UCSF, Rockefeller, and Johns Hopkins demonstrating different but compatible approaches to data acquisition from paper and electronic sources (e.g., an electronic IRB system).

This panel will discuss the sociotechnical approaches being taken by three CTSA institutions as they secure institutional buy-in and define local policies and procedures necessary for sustainable adoption of the HSDB vision. The first presentation will describe the goals and shared technical infrastructure for HSDB

data sharing and federated query. Presentations will then follow from Rockefeller University, Johns Hopkins University, and Duke University on how the HSDB vision fits with their broader institutional goals, how they are currently implementing human studies data acquisition and curation processes, and the key success factors and challenges they have faced so far. All presentations will identify and draw out crosscutting themes and lessons for discussion on data sharing of human studies descriptions in particular, and for data sharing in clinical and translational research in general.

Panelists and Presentations

Ida Sim, MD, PhD is Professor of Medicine and Co-Director of Biomedical Informatics at UCSF's Clinical and Translational Sciences Institute (CTSI). She is an international leader knowledge representation and open approaches for clinical research and evidence-based practice. Besides leading the Ontology of Clinical Research (OCRe) and the Human Studies Database projects, she recently co-founded OpenmHealth.org to build an open software architecture supporting scalable evidence generation to improve of individual and population health through mobile technologies. In policy work, Dr. Sim was the founding Project Coordinator of the World Health Organization's International Clinical Trials Registry Platform, which sets global standards on clinical trial registration. Dr. Sim will present the first talk of the panel, which will be an overview of the OCRe and Human Studies Database Project's goals and technical approach. She will also moderate the discussion session.

Shamim Mollah, MA, is a Bioinformaticist with the Biomedical Informatics Department and with Biomedical Informatics at Rockefeller's Center for Clinical and Translational Science. She holds a Masters in Biomedical Informatics from Columbia University. Her expertise lies in computational methods for assessing and assuring data quality, analyzing and visualizing high-dimensional data, and analyzing heterogeneous data in biological and medical domains. She has developed a comprehensive, systematized Bleeding History Phenotype (BHP) registry system that collects, records, and stores detailed phenotypic information for patients with bleeding disorders for worldwide access by physicians and researchers. She leads Rockefeller's participation in the HSDB project, contributing both to the design of the HSDB data federation architecture as well as the administrative and technical processes for collecting and sharing human studies protocol information from Rockefeller information systems and investigators.

Swati Chakraborty, M. Eng is a Clinical Research Informatics project leader at the Duke Clinical Research Institute. She has 20 years of experience in research and technology, including 7 years organizing and managing academic and commercial clinical studies. Ms. Chakraborty has served as enterprise architect and project leader for multiple DCRI projects in addition to the HSDB project, including the M.U.R.D.O.C.K. Integrated Data Repository (MIDR) with Cardiology, Osteoarthritis, Obesity and Hepatitis studies, and projects with the Clinical Trials Transformation Initiative (CTTI). Ms. Chakraborty will describe and discuss Duke's operational and implementation strategy for cross-CTSA sharing of protocol information about human studies being conducted at Duke.

Harold Lehmann, MD, PhD is Associate Professor of Health Policy and Management, and Pediatrics, and is the Research and Training Director of the Division of Biomedical Information Sciences at Johns Hopkins University. Dr. Lehmann's expertise is in Bayesian decision-making, guideline development, evidence synthesis, and methods and informatics approaches to evidence-based medicine. He has conducted systematic reviews with Hopkins' AHRQ-funded Evidence-Based Practice Center, and is widely published in informatics and decision-making. Dr. Lehmann has contributed to the development and evaluation of OCRe's Study Design Typology, and is leading the effort at Hopkins to collect and curate study protocol information. He will discuss both practical issues of data acquisition, as well as policy issues such as investigator concerns about the effect of protocol sharing on research competitiveness and efficiency.

Statement

All listed persons have participated in the creation of this proposal and have agreed to participate in the panel presentation.

References

1. CTSA Human Studies Database Project. [Available at http://hsdbwiki.org/, accessed October 14, 2011]
2. Sim I, Carini S, Tu S, Wynden R, Pollock BH, Mollah SA, Gabriel D, Hagler HK, Scheuermann RH, Lehmann HP, Wittkowski KM, Nahm M, Bakken S. The Human Studies Database Project: Federating Human Studies Design Data Using the Ontology of Clinical Research. *AMIA CRI Summit* 2010, p. 51-55.
3. Query Integrator. [Available at http://www.si.washington.edu/projects/QI, accessed October 19, 2011]