September 13, Mark Musen, The Center for Expanded Data Annotation and Retrieval

When left to their own devices, scientists do a terrible job creating the metadata that describe the experimental datasets that make their way in online repositories. The lack of standardization makes it extremely difficult for other investigators to find relevant datasets, to perform secondary analyses, and to integrate those datasets with other data. At Stanford, we are leading the Center for Expanded Data Annotation and Retrieval (CEDAR), a center of excellence in the NIH Big Data to Knowledge Program, which has the goal of enhancing the authoring of experimental metadata to make online datasets more useful to the scientific community. CEDAR technology includes methods for managing a library of templates for representing metadata, and interoperability with a repository of biomedical ontologies that normalize the way in which the templates may be filled out. CEDAR uses a repository of previously authored metadata from which it learns patterns that drive predictive data entry, making it easier for metadata authors to perform their work. Ongoing collaborations with several major research projects are allowing us to explore how CEDAR may ease access to scientific data sets stored in public repositories and enhance the reuse of the data to drive new discoveries.

Session details...

BIO:

Dr. Musen is Professor of Biomedical Informatics and of Biomedical Data Science at Stanford University, where he is Director of the Stanford Center for Biomedical Informatics Research. Dr. Musen conducts research related to open science, metadata for enhanced annotation of scientific data sets, intelligent systems, reusable ontologies, and biomedical decision support. His group developed Protégé, the world's most widely used technology for building and managing terminologies and ontologies. He is principal investigator of the National Center for Biomedical Ontology, one of the original National Centers for Biomedical Computing created by the U.S. National Institutes of Heath (NIH). He is principal investigator of the Center for Expanded Data Annotation and Retrieval (CEDAR). CEDAR is a center of excellence supported by the NIH Big Data to Knowledge Initiative, with the goal of developing new technology to ease the authoring and management of biomedical experimental metadata. Dr. Musen directs the World Health Organization Collaborating Center for Classification, Terminology, and Standards at Stanford University, which has developed much of the information infrastructure for the authoring and management of the 11th edition of the International Classification of Diseases (ICD-11).

Dr. Musen was the recipient of the Donald A. B. Lindberg Award for Innovation in Informatics from the American Medical Informatics Association in 2006. He has been elected to the American College of Medical Informatics, the Association of American Physicians, the International Academy of Health Sciences Informatics, and the National Academy of Medicine. He is founding co-editor-in-chief of the journal *Applied Ontology*.

SUMMARY:

Topic: Making online data searchable, accessible, and reusable: The Center for Expanded Data Annotation and Retrieval

Speaker: Mark A. Musen, M.D., Ph.D., Stanford University

Date: Wednesday, September 13, 2017

Time: 11 AM – 12 PM ET

Room: Seminar 408/410

You are invited to listen to Dr. Musen's presentation in the NCI Shady Grove Building on Medical Center Drive or via WebEx. **Dr. Musen will give his presentation in person at NCI.**

Presentation: A screen cast of the presentation will be available for viewing after the event on the NCI CBIIT Speaker Series YouTube Playlist

About the NCI CBIIT Speaker Series:

The National Cancer Institute (NCI) Center for Biomedical Informatics and Information Technology (CBIIT) Speaker Series presents talks from innovators in the research and informatics communities. The biweekly presentations allow thought leaders to share their work and discuss trends across a diverse set of domains and interests. The goals of the Speaker Series are: to share leading edge research; to inform the community of new tools, trends, and ideas; to inspire innovation; and to provide a forum from which new collaborations can begin. For additional information, including past speaker series presentations, visit the CBIIT Speaker Series page.

Individuals with disabilities who need reasonable accommodation to participate in this program should contact the Office of Space and Facilities Management (OSFM) at 240-276-5900 or the Federal TTY Relay number 1-800-877-8339.