



Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by **Battelle** Since 1965

Frontiers in Computing

Lecture Series

The Center for Expanded Data Annotation and Retrieval: Making Data 'FAIR'



Mark A. Musen, M.D., Ph.D.

Professor, Biomedical Informatics and Biomedical Data Science, and Director, Stanford Center for Biomedical Informatics Research, Stanford University

Monday, March 13, 2017

11 AM ♦ BSF Crick Room (2008)

Dr. Musen conducts research related to intelligent systems, reusable ontologies, metadata for publication of scientific data sets, 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 NIH. He also serves as PI of CEDAR. 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)*. He received the Donald A. B. Lindberg Award for Innovation in Informatics from the American Medical Informatics Association in 2006 and was elected to the American College of Medical Informatics, Association of American Physicians, and National Academy of Medicine. He is founding co-editor-in-chief of the journal, *Applied Ontology*.

In the past few years, there has been considerable buzz about making data sets created through publicly funded scientific investigation findable, accessible, interoperable, and reusable, or FAIR. The FAIR principles increasingly guide policy at the National Institutes of Health and at other federal agencies, although the scientific community still struggles to identify a means to measure FAIRness. Stanford University's Center for Expanded Data Annotation and Retrieval (CEDAR), a center of excellence in the NIH Big Data to Knowledge Program, has the goal of enhancing the authoring of experimental metadata to make online data sets more FAIR. Metadata are the "data about the data" that enable investigators to search data repositories and explore online data sets. CEDAR technology includes methods for managing a library of templates for representing metadata and interoperability with a biomedical ontologies repository that standardize the way templates can be filled out. The approach uses a repository of previously authored metadata from which it is possible to learn patterns in the metadata that can guide the instantiation of templates prospectively. In this talk, Dr. Musen will discuss the FAIR principles and the ways in which CEDAR may ease access to and reuse of scientific data sets stored in public repositories.

Host: Nathan Baker (nathan.baker@pnnl.gov), ACMD Division Director