

The EDRN Knowledge Environment: Providing Integrated Access to Scientific Information and Knowledge in Cancer Biomarker Research

Dan Crichton, MS, Jet Propulsion Laboratory

Authors

Dan Crichton (JPL), Mark Thornquist (DMCC), Sean Kelly (JPL), Andrew Hart (JPL), Heather Kincaid (JPL), Chris Mattmann (JPL), Jackie Dahlgren (DMCC), Kristen Anton (Dartmouth), Deanna Stelling (DMCC), Greg Warnick (DMCC), Susanna Reid (DMCC), Cim Edelstein (DMCC), John Tran (JPL), Christos Patriotis (NCI), and Sudhir Srivastava (NCI)

Abstract

Informatics in biomedicine is becoming increasingly interconnected via distributed information services, interdisciplinary correlation, and cross-institutional collaboration. The Early Detection Research Network (EDRN), a program managed by the National Cancer Institute, has been a pioneer at developing the data grid infrastructure for biomarker research. The distributed data grid infrastructure is a model-driven knowledge environment that enables distributed access and computation to biomarker research data and results across the virtual EDRN network.

The EDRN informatics team, consisting of experts in computer science, bioinformatics and statistics, has made significant progress in building a semantic knowledge portal that unifies biomarker information across EDRN. The EDRN Knowledge Environment provides integrated access to data across the EDRN enterprise that is captured during the biomarker discovery process. This includes information regarding biomarkers, studies, research results, and published work that can serve as a knowledge base to support scientific research. A principal feature of the EDRN informatics architecture is that it promotes a “loosely coupled” environment which allows it to unify different databases and systems together to support analytical research and discovery. At the heart of the EDRN Knowledge Environment is a core ontology model that describes the concepts of biomarker research and how they relate to existing data from tissue banking, to managing information about proposed biomarkers, to validation studies and scientific results. Defining this model allows for the data to be integrated into a unified science portal allowing researchers to navigate through the EDRN data using state-of-the-art search mechanisms including a Google-like search as well as facet-based navigation

Over the past year, the EDRN informatics team has made tremendous progress in building and curating new biomarker information and integrating that into the portal. An upgrade science portal with access to the data has now been fully deployed to the National Cancer Institute and review of the information by the research community is underway. The informatics team continues to work closely with the scientific community to ensure that the informatics vision can be fully realized and can demonstrate the power of building a unified informatics infrastructure to enable biomarker research at a national scale.