

---

## **A Distributed Informatics Knowledge Environment for Biomarker Research**

Dan Crichton<sup>1</sup>, Mark Thornquist<sup>2</sup>, Sean Kelly<sup>1</sup>, Chris Mattmann<sup>1</sup>, Don Johnsey<sup>3</sup>, Jackie Dahlgren<sup>2</sup>, Deanna Stelling<sup>2</sup>, Greg Warnick<sup>2</sup>, Suzanna Reid<sup>2</sup>, Cim Edelstein<sup>2</sup>, Andrew Hart<sup>1</sup>, Heather Kincaid<sup>1</sup>

<sup>1</sup>Jet Propulsion Laboratory, California Institute of Technology Pasadena, CA 91109, USA

<sup>2</sup>Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

<sup>3</sup>National Cancer Institute, Bethesda, MD 20892, USA

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” concept for bioinformatics having partnered with NASA’s Jet Propulsion Laboratory to deploy JPL’s data grid software tools for virtualized access to biospecimen data repositories distributed at several EDRN sites back in 2002.

EDRN’s core mission is the discovery and validation of biomarkers. Critical to that mission is having an informatics infrastructure that is “biomarker-centric”. The EDRN Knowledge Environment provides integrated access to data across the EDRN enterprise that is captured during the biomarker discovery process. The EDRN Knowledge Environment is built on the notion of a “data grid” allowing loosely related items across a highly heterogeneous, distributed environment to be linked together. The infrastructure of the data grid is a set of information system services that allow distributed databases to be integrated and accessed real-time. 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.

Over the past year, the EDRN informatics team has made tremendous progress in unifying biospecimens, scientific data, study specific data, and biomarker data into the EDRN public portal, providing virtual access to information repositories through the EDRN grid infrastructure and realizing the vision of building the EDRN Knowledge Environment.

**Acknowledgement:** This material is based upon work supported by the Jet Propulsion Laboratory, managed by the California Institute of Technology. Effort also supported by the National Cancer Institute’s Early Detection Research Network (EDRN) program.

---