

---

## An Information Model for Biomarker Research

J. Steven Hughes<sup>1</sup>, Deanna Stelling<sup>2</sup>, Dan Crichton<sup>1</sup>, Chris Mattmann<sup>1</sup>,  
Greg Warnick<sup>2</sup>, Suzanna Reid<sup>2</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

The Information Model is the foundation on which an Information System is built. It defines the entities to be processed, their attributes, and the relationships that add meaning. The development and subsequent management of the Information Model is the single most significant factor for the development of a successful information system.

The mission of the Early Detection Research Network (EDRN), a program managed by the National Cancer Institute, is the discovery and validation of biomarkers. In partnership with NASA's Jet Propulsion Laboratory (JPL) the program has developed the EDRN Knowledge Environment to provide integrated access to data across the EDRN enterprise that is captured during the biomarker discovery process. The infrastructure is a set of information system services that allow distributed data repositories to be integrated and accessed real-time. At the heart of the EDRN Knowledge Environment is an information model that describes the data entities involved in biomarker research and how they relate. The scope of this model ranges from tissue banking, to biomarker information, to validation studies, and scientific results. The information model integrates the data entities and enables researchers to navigate through the data repositories in the EDRN community using state-of-the-art search mechanisms.

Over the past year, the EDRN informatics team has made tremendous progress in developing the EDRN Information Model. Using a framework of data modeling tools, the information model has been captured from three sources, the EDRN registry of Common Data Elements (CDE), information specialists familiar with study design and processes, and the working subsystems. The model was captured in an ontology modeling tool and a specification document has been written with the model expressed using object-oriented class notation. The information model continues to be refined in concert with systems development.

The goal of the information model is to unify biospecimen data, scientific data, study specific data, and biomarker data for the EDRN Knowledge Environment and enable the EDRN public portal to provide access to information in data repositories throughout the EDRN community.

**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.

---