

DataBridge – A Sociometric System for Long-tail Science Data Collections (NSF/NIH)

Summary: There are currently thousands of scientists creating millions of data sets describing an increasingly diverse matrix of social and physical phenomena. This rapid increase in the both amounts and diversity of data implies a corresponding increase in the potential of data to empower important new collaborative research. However, the sheer volume and diversity of data presents a new set of challenges in locating all of the data relevant to a particular line of research. Maximizing the realization of the long-tail of science data requires tools specifically created to assist scientists in their search for relevant data sets. New tools specifically created to assist scientists in their search for relevant data sets and collaborators are required to realize the potential of the large and diverse data and enable important new multidisciplinary collaborative research on grand challenge scale problems. DataBridge is an e-science collaboration environment tool designed specifically for this purpose. We propose to explore a rich set of sociometric tools and the space of relevance algorithms, and adapt them to define different types of semantic bridges that link large quantities of diverse datasets into a sociometric network. We will develop a software tool, *DataBridge*, which will analyze data from several large NSF funded projects to provide relevance-based data discovery.