

Where the Rubber Meets the Sky: The Semantic Gap between Data Producers and Data Consumers

J.Gray

Microsoft

gray@microsoft.com

Abstract

Historically, scientists gather and analyze their own data. But technology has created functional specialization where some scientists gather or generate data, and others analyze it. Technology allows us to easily capture vast amounts of empirical data and to generate vast amounts of simulated data. Technology also allows us to store these bytes almost indefinitely. But there are few tools to organize scientific data for easy access and query, few tools to curate the data, and few tools to federate science archives. Domain scientists, notably NCBI and the Virtual Observatory, are making heroic efforts to address these problems. But this is a generic problem that cuts across all scientific disciplines. It requires a coordinated effort by the computer science community to build generic tools that will help all the sciences. Our current database products are a start, but much more is needed.