

<http://cra.org/ccc/theimpactofnitrd>

**Data-Intensive Discovery in Science:
The Fourth Paradigm
Alexander Szalay
Johns Hopkins University**

The exponentially growing amount of data, attributed in part to advances enabled by the Federal investment in networking and information technology research and development over the last 20 years, is fundamentally changing science. Discoveries are increasingly driven by data; we derive more and more of our hypotheses from existing data collections. Data-intensive scientific analysis has been called the fourth paradigm of science because it is an entirely new way of doing science. While still in its infancy, it has the potential to become the next major advance in the evolution of the scientific method. All sciences, both physical and social, as well as some humanities, are increasingly data-intensive.

As datasets grow ever larger, the ability to analyze the information hidden in them becomes increasingly daunting. Developing the “Science of Big Data” is an integrative task unlike any in science before. It requires the collaboration of statisticians, computer scientists, and disciplinary experts who can merge their skills to create new tools (computer hardware and software), new techniques (algorithms and statistics), and a new methodology that yields discovery through data-driven rather than hypothesis-driven inquiry.

Astronomy has traditionally been data-driven, and thus the community has been at the forefront to embrace this new approach to science. Data from the Sloan Digital Sky Survey has been used by more than a million people outside the professional communities, yielding some unexpected discoveries, and demonstrating how the Internet and advanced computing technologies have led to the emergence of a “long tail” of scientists.

