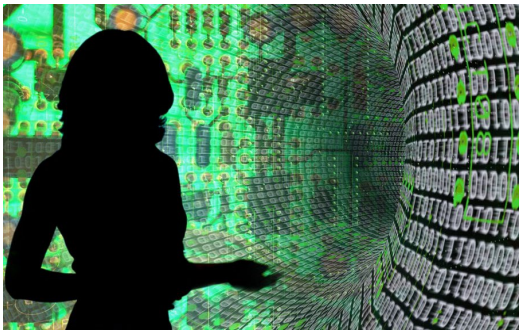
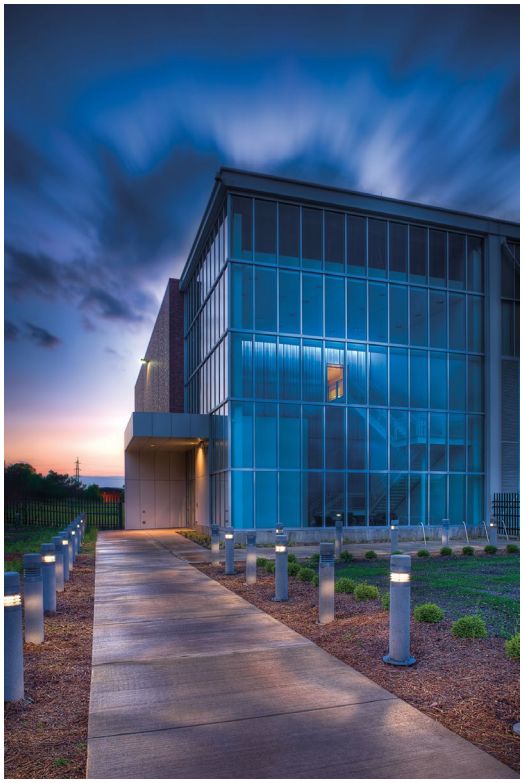
# LSST AND BIG DATA

LSST Corporation, 933 N. Cherry Avenue, Tucson, AZ 85721 520 881-2626 www.lsst.org

LSST will characterize billions of individual galaxies and stars, such as these produced with the LSST Image Simulator.





National Petascale Computing Facility (NCSA/University of Illinois)

## Era of Big Data Provides Opportunities and Challenges

The term "Big Data" refers not only to the volume of data, but the complexity of data processing necessary to extract knowledge from that data. The old method of gathering data to answer a question or validate a hypothesis is giving way to data-driven exploration and discovery. That is, complex datasets of high dimensionality are "mined" to discover connections and relationships you never knew existed - you have the information to answer the question before you know what question to ask. The data itself drives the discovery.

LSST will facilitate data-driven explorations of the most fundamental questions of the Universe. By measuring tiny distortions in billions of galaxies, LSST will provide information about the origin and distribution of Dark Matter. Mapping evidence of cosmic acceleration will provide answers about Dark Energy and resolve fundamental questions in our understanding of physics. The motions of billions of stars can be measured individually to construct an understanding of how our Galaxy was formed through a history of cosmic mergers. Detecting and then tracking the orbital paths of tens of thousands of moving objects will complete a census of potentially harmful Near Earth Objects. Rarely observed events will become commonplace, new and unanticipated events will be discovered, and the combination of LSST with contemporary space-based missions will provide powerful synergies.

## Characteristics of the LSST Database

To the user, LSST is a database, not a telescope. LSST will be like a giant "search engine" of the sky, digitizing and making available in a non-proprietary database the locations, motions, and characteristics of 20 billion galaxies and 20 billion stars. LSST will survey the visible night sky for ten years, generating a motion picture of the Universe. This relentless survey cadence will result in a complete dataset of high dimensionality, allowing exploration in both time and space. Discovery and sophisticated analysis of relationships buried in the data will require new algorithms that are scalable and affordable as well as novel database architectures. LSST scientists and engineers have already been recognized for advances in extremely large database technology.[1] These capabilities have value to fields far from astronomy and physics including defense, finance, medicine, and behavioral sciences. The open source and non-proprietary LSST database will be valuable as a training set for the education of the 21st century STEM workforce.

[1] http://go.usa.gov/Gsi