The Sloan Digital Sky Survey (SDSS) was a pioneer in open scientific data in terms of its scope, quality, public access, and extent of uses and users. The Phase I and II survey began collecting data in 1998 and continued through 2008, covering over a quarter of the night sky with high quality photometric and spectroscopic imaging. SDSS Team fieldwork were conducted from 1990-2008, covering over a quarter of the night sky. This study examines two data-intensive astronomy sky surveys. These sky surveys are international projects that use a single set of instruments to generate a standardized record of the night sky over an extended period of time. The SDSS has, and the LSST plans to, release the images, spectra, and catalogs. The large amounts of standardized data generated and released by these sky surveys enables end-users to combine, manipulate, and explore vast datasets. To reach the full potential of discovery available in the era of data-intensive science, digital data must be continuously managed over extended periods of time.

With the goal of sustaining the usefulness of scientific astronomy data over time, how are astronomy data managed, for what purposes, and who does the work?

Astronomy is one of the fields transformed by big science and data-intensive research methods. This study examines two data-intensive astronomy sky surveys. These sky surveys are international projects that use a single set of instruments to generate a standardized record of the night sky over an extended period of time. The SDSS has, and the LSST plans to, release the images, spectra, and catalogs. The large amounts of standardized data generated and released by these sky surveys enables end-users to combine, manipulate, and explore vast datasets. To reach the full potential of discovery available in the era of data-intensive science, digital data must be continuously managed over extended periods of time.

SDSS and LSST Affiliation | Number of Interviews
--- | ---
SDSS Team | 22
LSST Team | 25
Both | 17
Neither | 16

**Methods**

This study employed three qualitative research methods: semi-structured interviews (n=80), ethnographic fieldwork (n=21 weeks), and document analysis. Interviews and fieldwork were conducted from Fall 2011 through Spring 2015.

**Long-term Infrastructures**

Sky survey funding is based on short and medium term grants, however sky survey data management may be needed for decades. Sustained funding is important because it ensures the stability of all the knowledge infrastructures that are critical to ensuring data usefulness for years to come. While funding for data collection and scientific investigation are important, additional support is necessary to retain the scientific efficacy of the data over time. For example, software and hardware require upgrades and migration, which necessitates funding to purchase new materials, write and update software, and transfer the data. A consistent, expert workforce, with a reliable salary, is necessary to ensure the appropriate long-term management of scientific data. For the SDSS and LSST data to remain valuable to scientific data end-users over the long-term, a full environment of sustainable expertise may be required.

**Invisible Work**

Those who perform the data management work are not always those who benefit. A large amount of infrastructure work is necessary before data can be used by end-user scientists. Infrastructure-builders may have worked on a project for the better part of a decade before most end-users gain access to the data. Data management work then continues to ensure the data remain accessible and usable. This early and continuous infrastructure work is in the critical path of the eventual data user; however, it is often under-valued and unappreciated. The SDSS and LSST are large-scale projects and require a high level of this "invisible" infrastructure.

The Large Synoptic Survey Telescope (LSST) is an international astronomy collaboration that is currently under construction. LSST planning began in the late 1990s and the collaboration expects to begin the ten year survey in 2022, gathering an expected 15TB of data each night.