Summary of Data Management Principles
DESI Experiment

Experiment Description

The Dark Energy Spectroscopic Instrument (DESI) consists of 5000 robotically actuated fibers mounted at prime focus behind a wide-field corrector on the Mayall telescope in Kitt Peak, AZ, and read out with ten moderate-resolution spectrographs. The Mayall telescope is operated by the National Optical Astronomy Observatory (NOAO) under cooperative agreement with NSF. Over five years DESI will collect spectra and determine the redshifts of over 30 million objects. This data set will be used to detect baryon acoustic oscillations, thus measuring the expansion history of the Universe over a broad range in redshift with percent-level precision. DESI will be the first Stage IV dark energy experiment and will be able to rigorously test theories of dark energy – is it a cosmological constant, or does it vary with cosmic time? Are modifications needed to the Einstein's theory of gravity on cosmic scales? DESI will also place tight constraints on the sum of neutrino masses and on inflation parameters. DESI had first light in October 2019, followed by commissioning and survey validation, with the five-year survey starting in mid-2020.

DOE's roles in the experiment

DOE is responsible for designing, building, installing, commissioning, and operating DESI. NSF will provide the telescope, and will support telescope operations during a transition period prior to first light through NOAO. DOE funds will be used to support NOAO staff during the survey operations period, as well as the HEP lab and university groups who contribute to operations.

Partnerships

DOE and NSF have partnered on DESI, with NSF providing the telescope, following the NSF Astronomy Division Portfolio Review[1] panel's recommendation that NSF should divest of the Mayall operations. DOE and NSF have signed an MOU that describes the partnership and the funding responsibilities of each agency.

Additional partners include the UK, France, Spain, Switzerland and China, all of whom have made significant financial and/or in-kind contributions to the DESI project. In addition many institutions have contributed to the DESI Common Fund through their buy-in for participation (non-DOE funded participants only). Finally, DESI has benefited from philanthropic support from the Moore Foundation and the Heising-Simons Foundation. These foundations have been very important partners to enable the start of long-lead procurements.

---

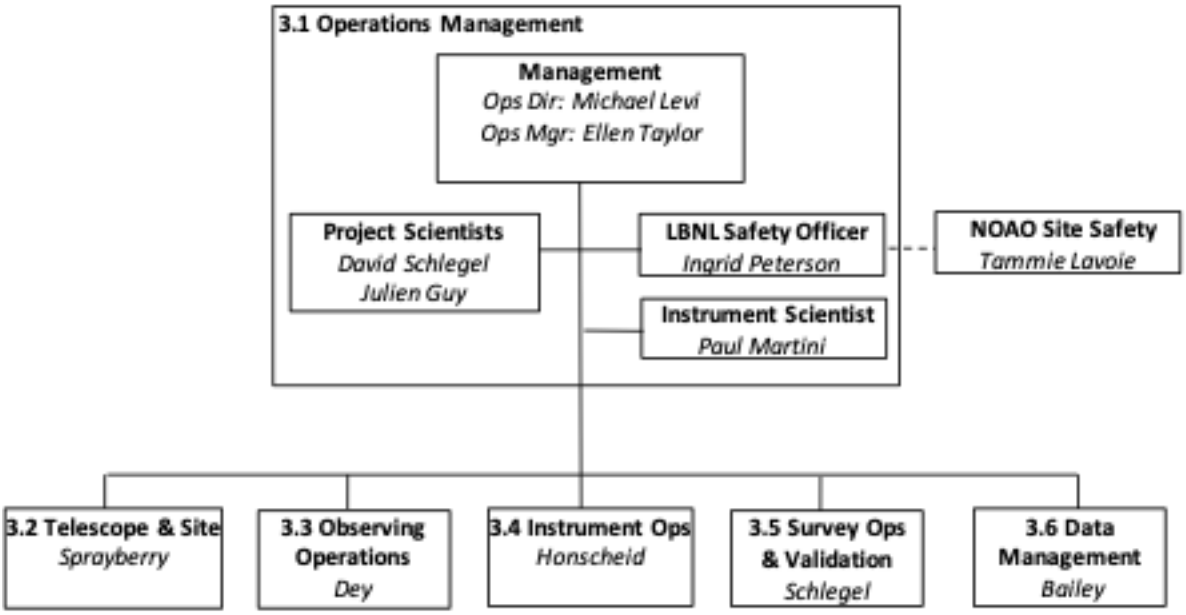
[1] http://www.nsf.gov/mps/ast/ast_portfolio_review.jsp

Organization – Agency/Lab level

Lawrence Berkeley National Lab (LBNL) is the lead lab for the DESI project. A separate DESI science collaboration is led by two spokespeople who are elected by the collaboration.

Organization – Experiment level

The DESI organization chart is shown below.



Collaboration

The DESI collaboration consists of over 600 scientists from 93 institutions and regional participation groups, including 5 US national labs and 32 US universities. The collaboration has elected 2 spokespeople to guide the scientific work.

Data Policy Management

Details of DESI data management are contained in the Final Design Report located at https://www.desi.lbl.gov/final-design-report/ .

The DESI project management team is responsible for setting data policy, in consultation with the collaboration and its spokespeople. The project will be responsible for archiving the data and for periodic data releases.

Data Description & Processing

Raw data from the DESI data acquisition system will be compressed and transferred daily from KPNO via NOAO to the central data repository at LBNL/NERSC and backed up to the HPSS tape storage system at NERSC. At NERSC the data are pre-processed, calibrated, and made available to software that performs spectral extraction and redshift determination. The output data products are hosted and served to the science collaboration through NERSC science gateways. In addition to the raw spectroscopic data, Data Systems will also archive operational data such as the Guide-Focus-Alignment camera output, temperature and pressure measurements, the history of commands sent to the fiber positioners, their recorded position, engineering monitoring data, and the images and analyzed data from the fiber view camera.

The outputs from the DESI processing will be in standardized file formats such as FITS and JSON and conform to the DESI Data Model for each file type.

Data Products and Releases

Catalog data will include, but will not be limited to, celestial coordinates, redshift, spectral classification (galaxy, QSO, etc.), spectral metadata (exposure date, fiber number, etc.) and targeting data. The data products will also contain the full spectral data (flux versus wavelength) and associated metadata. The data will be released to the collaboration for analysis, and will eventually also be made publicly available.

Plan for Serving Data to the Collaboration and Community

Tagged versions of the processed data and associated code will be released to the DESI science collaboration approximately yearly. These will serve as the standard datasets for publications for consistency across analyses and future traceability. Additionally, the most recent reductions of the data will always be available to the entire collaboration as soon as they are processed, typically within 24 hours of observation. After a proprietary period, these tagged data and software will also be publicly released, with the intention of producing high-quality science in a timely manner. Data releases will include software and additional documentation as needed.

Plan for Archiving Data

Copies of the raw data will be archived at both NOAO and NERSC to provide geographically separate backups. The raw data and reduced data products will be archived at NERSC for 5 years after the experiment ends subject to funding availability from DOE.

Plan for Making Data Used in Publications Available

All data points shown in the published graph will be also available in a machine-readable form on a website listed in the publication.

Responsiveness to Office of Science Statement on Digital Data Management

This data management plan fully follows the DOE Office of Science Statement on Digital Data Management.