

# Summary of Data Management Principles

## Planck

### **Experiment description:**

Planck is a European Space Agency (ESA) satellite mission to measure the temperature and polarization of the Cosmic Microwave Background (CMB) over the entire sky, and use these to constrain the fundamental parameters of cosmology. Planck carries two instruments; the Low Frequency Instrument (LFI) carries 22 radiometers at 30, 44 and 70GHz, while the High Frequency Instrument (HFI) carries 52 bolometers at 100, 143, 217, 353, 545 and 857GHz. Launched in May 2009, Planck started recording data from the second Lagrange point in August 2009 and completed a survey of the entire sky every 6 months until it was decommissioned in October 2013.

### **DOE's roles in the experiment:**

DOE's role in Planck is limited to the provision of high performance computing (HPC) resources at the National Energy Research Scientific Computing (NERSC) Center at Berkeley Lab.

### **Partnerships:**

The provision of NERSC resources to the US Planck team is governed by an MOU between NASA and DOE. In return for these NERSC resources, NASA funds 2.5FTEs in Berkeley to develop and deploy appropriate massively parallel simulation and analysis software.

### **Organization – Agency/Lab level**

ESA is the agency with overall responsibility for all aspects of Planck, with NASA providing instrumentation and data analysis support.

### **Organization – Experiment level**

The organization of the Planck experiment is described in detail here:

<http://www.cosmos.esa.int/web/planck/planck-collaboration>

### **Collaboration:**

The Planck collaboration has approximately 600 members (scientists and engineers) in 50 institutions in 15 countries. The Planck Science Team, comprising 8 permanent members, governs the collaboration.

### **Data policy management:**

Data policy is set by the Planck Science Team so as to meet the ESA Data Management requirements. Data are distributed by both ESA and NASA archives as well as through NERSC, with certain data (specifically the full massive Monte Carlo simulation sets required to quantify uncertainties in the data) being provided exclusively at NERSC.

### **Data Description & Processing:**

The basic Planck data comprise raw time-ordered data from the instruments, and cleaned, calibrated time-ordered data produced by the two instrument Data Processing Centers (DPCs).

These data are then progressively reduced to:

- Sky maps at each observing frequency,
- CMB and foreground component maps,
- Angular power spectra of the CMB maps, and
- Spectral likelihood functions.

The satellite data are first transmitted to the Mission Operations Center. Each instrument's raw data are then sent to the corresponding DPC for pre-processing. Cleaned, calibrated data are then exchanged between the DPCs and sent to NERSC every 6 months. Further data reduction is then performed both at the DPCs and at NERSC.

### **Data Products and Releases:**

Public Planck data releases have occurred alongside the major scientific paper publications in 2013 and 2015, with a final (legacy) release anticipated in 2016. Such releases include the frequency and component maps, catalogs of compact sources in these maps, together with the power spectra and associated likelihood functions. The release of cleaned, calibrated time-ordered data is still under discussion.

### **Plan for Serving Data to the Collaboration and Community:**

Data are released to the public with the major Planck science publication releases. Both ESA and NASA have dedicated archives to distribute the data, along with complete descriptions of their contents. There is no formal plan to release the raw data processing pipelines since their use depends on very detailed instrument knowledge that will not survive the collaboration.

### **Plan for Archiving Data:**

All of the official data products will be archived at the ESA Planck Legacy Archive

<http://pla.esac.esa.int/pla/>

and the NASA Planck Archive

<http://irsa.ipac.caltech.edu/Missions/planck.html>

Additionally, the Planck maps – both from the real data and from the massive Monte Carlo simulation sets – will be kept spinning at NERSC for 5 years after the end of the mission.

### **Plan for Making Data Used in Publications Available:**

The Planck likelihood functions used to derive all of the major scientific results from the reduced Planck data will also be made available through the archives.

### **Responsiveness to SC Statement on Digital Data Management**

This data management plan fully follows SC Statement on Digital Data Management insofar as this is consistent with the Data Management requirements of ESA as the lead agency.