

1.2.3 Canadian Astronomy Data Centre

Report on the Survey Data Management and Usage of the Legacy Survey data

The data management for the Legacy Survey is defined in a Memorandum of Understanding as a collaboration between three parties. CFHT executes the observations and is responsible for pre-processing (removal of instrumental signatures) of the data. The Canadian Astronomy Data Centre (CADC) is responsible for distributing the CFHT LS data to science users and to the Terapix group. Terapix is responsible for processing of advanced products which are then returned to CADC for distribution.

The CFHT Board defined two objectives for the Legacy Survey: a) to accomplish selected large scale community-defined scientific programs and b) provide the community with an extensive, homogeneous datasets to enable a wide range of long lasting scientific applications. The production of advanced data products at Terapix and their distribution to the community through CADC was intended to enable both of these objectives.

At the end of two years of operations we report the following:

- CADC was fully prepared on schedule to receive and distribute the data products from CFHT which now comprise 65 Terabytes.
- Raw data from the Legacy Survey typically is catalogued, ingested and is available to users from CADC during the day following the observation.
- Pre-processing at CFHT begins at the end of an observing run. For pre-processed data the lags between observation, processing at CFHT, and transfer to CADC have been reduced from a median of about 40 days during the first half of 2004 to a median lag of about 14 days (measured from the end of an observing run) in the second half of 2004.
- All Legacy Survey data transfers are done by network. Raw and pre-processed data can be seen and retrieved by users a few minutes after it arrives at CADC.
- CADC distributed a total of 20,827.0 gigabytes of Legacy Survey data to 34 distinct users in the period February 2004 to February 2005. These totals include transfers to Terapix (20% of the total).

- Data pre-processed by CFHT accounted for 90.7% of the downloads from CADC whereas raw data was 8.4% of those downloads.
- Terapix delivered 30 gigabytes of stacked images and other data products to CADC in November and December 2004. These are observations of the four Deep Fields and they account for 171 gigabytes of downloads (1% of the total for Legacy Survey) from CADC by users. A total of 39 distinct users browsed the Terapix data products and the downloaded volume is equivalent to the 30 gigabyte dataset being download 5.7 times since its release in November.
- CFHT SAC has made a number of recommendations that recognize the importance of well-defined and well-planned data releases and recognizes the importance of both CFHT pre-processed products and the advanced data products from Terapix. A mechanism is now in place that gets raw and pre-processed Legacy survey data to users quickly. Deadlines for a major Terapix data release were set for June 2004 and February 2005 but, at this time, CADC has not received a major release of Terapix data to distribute to users.