MegaPrime

Interface Control Document

12 May 2003

Luc Simard, David Schade, Séverin Gaudet, Norman Hill, Daniel Durand, David Bohlender

Canadian Astronomy Data Center
Herzberg Institute of Astrophysics
National Research Council of Canada

Eugene Magnier, Jean-Charles Cuillandre, Christian Veillet, Gregory Fahlman

Canada France Hawaii Telescope Corporation

Yannick Mellier, Emmanuel Bertin, Laurent Domisse, Gérard Tissier, Frédéric Magnard, Gilles Missionnier, Bertrand Morin, Pierre Didelon, Mireille Dantel-Fort

Projet TERAPIX
Institut d'Astrophysique de Paris
Centre National de la Recherche Scientifique

Created: November 14, 2001
Modified: October 15, 2002

Canada
France
Hawaii
Telescope
Legacy
Survey

Chapter 1

Introduction

1. Introduction

The success of the Canada-France-Hawaii Telescope Legacy Survey is a key factoring the success of all three partners in the data flow: CFHT, TERAPIX, and CADC. The future of our groups is tied to this ambitious project. The CFHTLS is going to require a level of cooperation between the three partners that does not yet exist. It appears to CADC that agreements such as those that will be produced by the ICD process are necessary to ensure success. The interactions between the partners will be technically complex and must ultimately take place in such a way that will satisfy the scientific desires of the communities. CADC feels that this level of cooperation is not going to evolve by itself; we believe strongly that a set of documents like the ICD are fundamental to producing effective interactions and that the work that all three partners will need to put into the ICD process is both worthwhile and necessary. Our sole motivation is our desire that the CFHT Legacy Surveys are scientifically successful.

This latest draft of a proposed Interface Control Document (ICD) is another step that will lead us to Memoranda of Understanding (MOUs) between the partners (CFHT, CADC and TERAPIX) in MegaPrime and the CFHT Legacy Survey (CFHTLS) projects.

The purpose of this draft is to stimulate continuing discussion and lead to the development of the next draft that we hope will be much closer than the present draft to defining the agreements that we need to make with each other. If there are suggestions for major or minor structural changes, then they will be accommodated.

The areas that we identify here as areas where agreement is required are neither complete nor fully accurate. We expect CFHT and TERAPIX to add new areas, delete others, and change most or all. The subjects of concern presented in this document are only proposals, and the CADC expects the other partners to suggest changes to most or all of these proposals. Through this iterative process we hope to reach consensus quickly.

2. Purpose

This document describes the interfaces between CFHT, CADC, and TERAPIX that will be needed for the MegaPrime instrument and the CFHTLS project.

This document is presented in five chapters:

- Chapter 1 is this introductory chapter.
- Chapter 2 describes the interfaces between CFHT and CADC.
- Chapter 3 describes the interfaces between CFHT and TERAPIX
- Chapter 4 describes the interfaces between CADC and TERAPIX
- Chapter 5 describes the general communications that must be in place between the partners

3. Abbreviations and Acronyms

CADC	□□	Canadian Astronomy Data Centre
CFHTLS	□□	Canada-France-Hawaii Telescope Legacy Survey
CFHT	□□	Canada-France-Hawaii Telescope
DLT	□□	Digital Linear Tape
FITS	□□	Flexible Image Transport System
ICD	□□	Interface Control Document
ICS	□□	Instrument Control System
ISU	□□	□egaCam Image Stabilization Unit
MEF	□□	Multi-Extension FITS file (See FITS and [3])
MOU	□□	Memorandum of Understanding
MSWG	□□	MegaPrime Science Working Group
QSO	□□	Queued Service Observing
SDLT	□□	Super Digital Linear Tape
SG	□□	Steering Group
TCS	□□	Telescope Control System
TERAPIX	□□	Traitement Elémentaire, Analyse et Réduction des Pixels de MegaPrime
XML	□□	Extensible Markup Language

4. Glossary

Detrend Data □□ Calibration files used by Elixir to remove instrument signatures from MegaPrime data. See also *Detrended Data* and *Elixir*.

Detrended Data □□ Data files that have been processed with Elixir to remove instrumental signatures. See also Detrend Data and Elixir

Tar □□ A standard UNIX file archiving utility

FITS header □□ The ASCII portion of a FITS file containing meta-data in a human readable format.

Primary Header Unit □□ The first FITS header in a MEF file that contains meta-data for the entire mosaic. Each extension within the file also has its own header unit.

Elixir □□ The data analysis and processing system produced by CFHT. The Elixir system at CFHT consists of several related components for the production of master detrend frames, the data qualification and calibration, and the application of these data products to the raw science data.

Elixir Distribution System □□ The subset of Elixir components needed to perform the distribution steps including application of the appropriate detrend frames and the updating of calibration keywords.

Processing Record □□ A recipe specifying processing steps applied to raw data to obtain a specific data product. Each processing step is defined by date, input/output parameter values, software name and version number.

Machine-readable □□ Format in which the information is properly field-delimited for easy parsing by a software input module.

5. References

[1] MSWG report
[2] CADC/CFHT archive MOU
[3] FITS file format
[4] FITS WCS proposal
[5] FITS checksum proposal
[6] The CADC Data Processing Architecture
[7] Gemini Science Archive Conceptual Design Documents

6. Baseline Configuration

This is the initial configuration for the software used for CFHTLS software development. The versions listed here will be updated over time, with the mutual agreement of the partners.

6.1 Solaris Version

Version 2.8

6.2 Linux Version

Red Hat 7.2

6.3 Sybase Version

Version 12.5

Created: November 14, 2001
Modified: May 12, 2003

Canada
France
Hawaii
Telescope
Legacy
Survey

Chapter 2
CFHT to CADC Interfaces

1 Introduction

This chapter describes the interfaces between CFHT and the CADC for the MegaPrime instrument, where practical, the interfaces between CFHT and CADC will be the same as those needed to maintain the non-CFHTLS CFHT archive. The interfaces required between CFHT and the CADC are:

- The Bulk Data (pixel data + headers and auxiliary files) interface is described in Section 2 on page 5.
- The Meta-Data interface is described in Section 3 on page 8.
- The transfer of software between CFHT and CADC is described in Section 4 on page 14.
- Data Management is described in Section 5 on page 15.

2 CFHT to CADC Bulk Data Interface

This section describes the transfer of bulk data from the CFHT to the CADC.

2.1 Content

The bulk data transferred from CFHT to CADC will consist of data files generated during telescope operations. The bulk data falls into four categories:

1. Raw FITS files (pixel data and headers of science and calibration images) collected by the telescopes during both QSO and classical MegaPrime operations.
2. Processed detrend FITS files.
3. Detrended science FITS files.
4. Auxiliary files: electronic observing logs, weather logs and images and SkyProbe data and images. All auxiliary files will be FITS files.

All FITS files will be NOAO/NOST-compliant (see [3]). All FITS files shall include checksums using the proposed FITS checksum standard (see [5]). Raw and detrended FITS files will be in 16-bit integer format. Detrend FITS files will be in 16-bit floating point format. Multi-extension FITS files will be used. There will be one chip per extension, and there will be one raw file per exposure.

All FITS files of MegaPrime data produced by CFHT will be transferred from CFHT to CADC. There will be no filtering by CFHT as the CADC copy represents CFHT's offsite backup copy of their data.

Even though headers should be included in bulk data transfers, they are technically considered to be meta-data. Raw FITS file headers, detrend FITS file headers and detrended FITS file headers are described in sections 3.1.1, 3.1.2 and 3.1.3 respectively.

2.1.1 Raw FITS File Names

File names will be in the form %d%c.fits, where %d is the exposure number, and %c is a single character indicating the exposure type. Exposure numbers will continue in the current CFHT exposure number sequence. Valid exposure types are:

- **o** — Science exposure
- **b** — Bias frame exposure
- **f** — Flat-field exposure
- **d** — Dark current exposure
- **x** — Focus sequence

2.1.2 Detrend FITS File Names

All detrend FITS file names must be unique. If a file is replaced or modified, it must be given a different name. Names shall never be reused.

The format of the name of a MEF master detrend file will be CRUNID.TYPE.FILTER|EXPTIME.VERSION where CRUNID is the camera run id (e.g., 01Ak01 = 2001/01/27 – 2001/02/06), TYPE is the detrend type (e.g., flat), FILTER is the filter used (e.g., B), and VERSION is a version number to force the image names to be unique. FILTER will be specified for TYPE = flat, modes, fringe, frpts, scatter, and EXPTIME (exposure time) will be specified when TYPE = bias, dark and mask.

The format of the name of a SPLIT master detrend file will be CRUNID.TYPE.FILTER|EXPTIME.VERSION with the same rules as for MEF names.

2.1.3 Detrended FITS File Names

All detrended FITS file names must be unique. If a file is replaced or modified, it must be given a different name. Names shall never be reused. The format of the name will be of the form %dp.fits, where %d is the exposure number.

2.1.4 Auxiliary FITS File Names

The multi-extension FITS file containing the meta-data for a camera run described in Sections 3.1 and 3.2 will have a name of the form mcmd.RUNID.date+time where RUNID is the program run id (e.g., 2003AL01).

The tar file containing the auxiliary files CFHT normally includes in their data distributions to principal investigators will have a name of the form mccd.RUNID.date+time.

2.2 Mechanism

The considerations in selection of the transfer mechanism are:

- Data should be transferred as soon as possible since the MSWG report requires that raw data be put on-line as soon as possible to allow science exploitation to begin immediately.
- The MegaPrime data rate is estimated to be 100-150 Gigabytes per night.
- Final detrend files will only be available at the end of each QSO run.

- A single data path from CFHT to CADC should be used for all CFHT instruments, and all observing programs regardless of whether they are CFHT Legacy or PI programs. We should avoid having separate data handling procedures for CFHTLS and the other data produced by CFHT.
- Elixir processing can only be done after all calibration files for a QSO run are available.

- Elixir processing of CFHTLS data will be done initially only at CFHT, and those processed data will be the data that are distributed to users with the quality guaranteed by CFHT.
- Elixir processing of MegaPrime data that are NOT part of the CFHTLS will be done by CFHT, and CADC will distribute the data to public users after the release date.
- Processed data represent approximately 70% to 80% of the volume of unprocessed data.
- Data compression should be used where possible.

There will two data streams from CFHT to CADC. The first stream will be raw FITS files and auxiliary files produced at the time of observation. The second stream will be detrend FITS files, detrended FITS files and auxiliary files that require processing after the end of the run.

The raw image meta-data FITS file will be part of the automatic nightly download of image information from CFHT to CADC currently in operation.

The transfer mechanism will likely evolve over the course of the project. Periodic (at least annual) discussions of the data transfer mechanisms will take place between CADC and CFHT.

2.2.1 Media Type

Initially, the transfer media for bulk data transport from CFHT to CADC will be Super DLT tapes. With a capacity of 110GB, the capacity of these tapes should be adequate for both transport and archival storage. CFHT is responsible for purchasing the tapes sent to the CADC.

The transfer of CFHT bulk data will continue to be on permanent media (tape initially) regardless of whether electronic transfer is implemented at some point. This is so that the CADC media continue to perform the function of CFHT data backup.

The data on the tapes will be formatted in the same way as the DLT tapes currently used to transfer data from CFHT to CADC i.e.,

- The first file on the tape is a tar file containing a single empty file. The name of the empty file is interpreted as the name of the tape.
- After the first tape name tar file there will be one (or more) tar files, each containing one or more data files.

Note: Network transfer is the ideal solution to this problem because it is cheap and can be made to be highly automatic. However, adequate network speed between CFHT and CADC does not exist at the present time. We should investigate the cost of enhancing existing networks.

2.2.2 Media Names

Names of media containing raw data files will continue in the same sequence as the tapes CFHT is currently sending to the CADC (i.e., CFHT-CADCnnnn). Names of media containing detrend

and detrended data will be in the form CFHT-CADC-DETnnnn, and the sequence number will start at 0001.

2.2.3 Handling

Bulk data SDLT tapes will be sent to CADC once a week or in batches of 4, whichever comes first. There are time-critical issues involved in this area of transfer. The MegaPrime Science

Working Group (MSWG) report requires that the raw data be put on-line as soon as possible so that science exploitation can begin immediately. At the end of a MegaPrime run, tapes will be sent immediately to CADC regardless of how full they are.

Two identical copies of the raw FITS file stream will be made by the CFHT, one copy kept at the CFHT and the other sent to the CADC as described above and in existing agreements with respect to data not related to MegaPrime. The tapes will be identical.

Both CADC and CFHT need to be capable of creating copies of raw FITS file stream tapes if needed by the other site.

Only one copy of media containing detrend and detrended FITS files will be made by the CFHT and will be sent to the CADC at the time of each data release described in section 2.2. CFHT shall preserve the ability to re-generate these files until they have been ingested in the CADC systems.

3 Meta-Data Interface

This section describes the transfer of meta-data from the CFHT to the CADC. The purpose of the meta-data is to fully characterize the telescope, instrument and detector status, environmental conditions, astrometric and photometric calibrations, and software subsystem version numbers. The meta-data content must be sufficiently complete to:

(1) Produce processed data products satisfying or exceeding the scientific requirements set by the CFHTLS Steering Group and Survey Coordinators.
(2) Process raw data to reproduce or improve a specific data product, and
(3) Understand the origins of any problems found in processed data products

Meta-data include everything but pixel data themselves. Therefore, headers are technically part of the meta-data.

Default values should not be used in FITS file headers and meta-data FITS table files. When a value is unknown, it should be identified as such. Fields in the meta-data FITS tables will take on the special values of NULL and NA. NULL means the CADC database should be updated to the appropriate NULL value while NA means the CADC database should not be updated for that field. The tables will have the keyword TNULL*n* that defines the NULL value for a table value and the keyword TNVAL*n* to define the value of NA for each column. It is necessary to define these special values on a case-by-case basis because, for example, a numeric column can only contain numeric values.

All date and time fields in the image headers and meta-data tables will be in Universal Time (UT) and will have the format YYYY/MM/DD,HH:MM:SS.

3.1 Meta-Data Content

The sources of the meta-data required in Sections 3.1.1 to 3.1.4 are given for each item. “H” stands for FITS headers and means that the information will be extracted from the header of the relevant image. “T” stands for FITS table and means that the information will be extracted from the tables in the multi-extension FITS meta-data file produced by CFHT and described in Section 3.2.

3.1.1 Raw FITS File

- Public release date will be one of the values in the meta-data. Release date is set by CFHT and may be modified at any time. (H)
- WCS information will be present, correct and conform to the NOAO WCS model (see [4]). (H)
- Run ID for each observation. (H)
- Camera Run ID for each observation. (H)
- Observation type. (H)
- Environmental information (temperatures, wind speed/direction, humidity level, etc.). (T)
- Filter name (must refer to actual physical filter ID and must be unique). (H)
- Telescope focus setting. (H)
- CCD description. Example: Marconi/EEV CCD42-90. (H)
- Maximum pixel values for linearity and saturation. (H)
- Dome azimuth. (H)
- Image Stabilization Unit Status (ON/OFF). (H)
- ISU guide star magnitude. (H)
- Shutter status (OPEN/SHUT). (H)
- Data quality estimates. Example: fringe amplitude, etc.
- Survey Program IDs as defined by the CFHTLS Steering Group. Example: Ultra-Deep, Shallow-Wide, Deep, etc. (H)
- Association (or “dither group”) name. May be generated by QSO. (H)
- Names and version numbers of all software subsystems (TCS, ICS, QSO, etc.) used to generate meta-data described in this section. (H)

3.1.2 Detrend FITS File

- Public release date will be one of the values in the meta-data. Release date is set by CFHT and may be modified at any time. (H)
- Detrending purpose (bias, dark, dome flat-field, twilight flat-field, night sky flat-field, fringe, etc.). (T)
- Processing record in machine-readable form. (T)
- Flat field images used if detrending purpose is flat-field. (H)
- Bias images used if detrending purpose is bias. (H)
- Dark images used if detrending purpose is dark. (H)
- Fringe images used if detrending purpose is fringe. (H)
- Fringe point map name used if detrending purpose is fringe. (H)
- Names and version numbers of all software subsystems (Elixir, FLIPS, etc.) used to generate meta-data described in this section. (H)

3.1.3 Detrended FITS File

- Raw FITS file header content carried through unmodified to all later processed FITS files. However, raw WCS will be modified by Elixir processing. (H)
- Public release date will be one of the values in the meta-data. Release date is set by CFHT and may be modified at any time. (H)
- Detrend flat field image used. (H)
- Detrend bias image used. (H)
- Detrend dark image used. (H)
- Detrend fringe image used. (H)
- Post-Elixir data quality estimates. Example: fringe residual amplitude, etc. (H)
- Elixir will write the processing record to the detrended file headers. The processing record includes input data, calibration files, processing location and configuration information for the processing system. The processing record will be machine-readable. (H)
- Names and version numbers of all software subsystems (Elixir, FLIPS, etc.) used to generate meta-data described in this section. (H)

3.1.4 Other Meta-Data

- Sky brightness in magnitude per square arcsec (T)
- An electronic log covering the period of each observation in a machine-readable format with timestamp entries to allow linking to a particular observation. May be generated by QSO.
- Associations (sets of frames that can be stacked and sets of filters also) will be identified in a machine-readable way. (H)
- Image quality data (seeing, etc). (H for raw, T for raw and T for detrended)
- Skyprobe transparency measurements. (T)
- Photometric meta-data (type of photometric solution, coefficients and residuals for each filter and each observation at median air mass). (H, T)
- Source of extinction value. (T)
- Filter transmission curve name. (T)
- Elixir fringe point maps. (H)
- Value of colour term with error (T)
- Value of zeropoint with error (T)
- Units of photometric solution (electrons or DU). (H)
- Astrometric meta-data (type [form, local/global] of astrometric solution, number of stars, epoch, coefficients and residuals). (H)
- Table of associations of detrend FITS files with raw FITS files. (H)
- Table of associations of astrometric and photometric with raw FITS files. (H)
- Names and version numbers of all software subsystems (Elixir, FLIPS, etc.) used to generate meta-data described in this section. (H)

3.2 Meta-Data Table Formats

There will be eleven (11) meta-data tables. Detailed descriptions of these meta-data table are given in Sections 3.2.1 - 3.2.11.

All FITS meta-data tables will come with their own meta-data. These table meta-data will consist of the table type and creation date. The table type defines which meta-data table is in it, and the creation date is a time-stamp needed to re-ingest meta-data tables.

The meta-data tables will consist of rows of data values with a consistent number of columns per row. The meta-data tables should be self-documenting: the column and column range corresponding to a given data type should be indicated in the meta-data tables' meta-data.

The meta-data tables should provide a complete description of the distributed data. All other data types (images, image headers, graphics files) will be identified and listed in one of the meta-data tables.

CADC will only verify and ingest meta-data specified by the "core" meta-table columns described here. CFHT may add/remove extra columns to the tables, but these extra columns will be ignored by the CADC ingestion tools.

The core meta-data table columns should be static. No changes should be made to the allowed fields, their valid data ranges or their formats without prior communication between CFHT, CADC and Terapix.

3.2.1 Processed Science Images

This table contains entries relevant to a specific image not found in the image header keywords. The extension name EXTNAME will be PROCESSED_IMAGES.

TTYPE	TUNIT	TFORM	COMMENT
FILENAME		A99	Filename
HDR_FILENAME		A99	Filename of FITS header
OBSID		I10	Unique ID number
INSTRUMENT		A20	Acquisition instrument
TAPE_ID		I6	Tape ID number
TARSET		I3	Tar set on tape
DETYPE		A10	Descriptive name of image type
DESTATE		A30	Processed level
DMTYPE		A10	Distribution media type
QSO_COMMENT		A255	QSO coordinator comment
QSO_EVAL		I2	QSO evaluation
QSO_STATUS		A30	QSO status
Elixir_VALID		A10	Elixir validation string
REQ_IQ_MIN	arcsec	F4.2	Minimum requested image quality
REQ_IQ_MAX	arcsec	F4.2	Maximum requested image quality
REQ_BG_NAME		A10	Requested background band
REQ_AIRM_MAX	airmass	F5.2	Requested maximum airmass
OBS_IQ_REFCCD	arcsec	F5.2	Observed image quality (REF CCD)
OBS_IQ_CENTER	arcsec	F5.2	Observed image quality (center region)
OBS_IQ_R_RATIO		F5.2	Observed IQ ratio (outer / center)
OBS_IQ_X_RATIO		F5.2	Observed IQ ratio (right / left)
OBS_IQ_Y_RATIO		F5.2	Observed IQ ratio (top / bottom)
OBS_BG_VAL	counts / pixel	F9.3	Observed background
OBS_BG_NAME		A10	Observed background band
IS_PHOT	boolean	I2	Observer evaluation of transparency
JPG_LARGE_NAME		A99	Filename of JPEG image
JPG_SMALL_NAME		A99	Filename of JPEG thumbnail image

3.2.2 Detrend Images

This table describes the Elixir detrend images used to process the distributed science images and included on the distribution tape. The extension name EXTNAME will be MASTER_DETREND.

TTYPE	TUNIT	TFORM	COMMENT
KEY		A32	Primary key ID
START_TIME	yyyy/mm/dd,hh:mm:ss	A20	Start time of validity range
STOP TIME	yyyy/mm/dd,hh:mm:ss	A20	Stop time of validity range
REG_TIME	yyyy/mm/dd,hh:mm:ss	A20	Time of Elixir registration
EXPTIME	seconds	F7.1	Exposure time
IMAGETYP		A10	Detrend type
FILTER		A10	Filter name
CCDINFO		A7	CCD information (Name or Number)
MODE		A7	Data format mode (SPLIT or MEF)
VERSION		I3	Image version number
ORDER		I3	Selection order
LABEL		A64	Data label
PATH		A256	Filename in db
TAPE_ID		I6	Tape ID number
TARSET		I3	Tar set on tape
JPG_LARGE_FILE		A99	Filename of JPEG image

3.2.3 Standard Star Photometry

This table lists the zeropoints measured from each standard star frame as processed by the Elixir photometry system. The extension name EXTNAME will be IMAGE_ZPTS.

TTYPE	TUNIT	TFORM	COMMENT
ZP_OBS	mag	F8.4	Measured zero point
ZP_REF	mag	F8.4	Nominal zero point
ZP_ERR	mag	F7.4	Error on zero point
C_AIRMASS	mag / airmass	F7.3	Airmass coefficient
C_COLOR	mag / mag	F6.3	Color coefficient
START_TIME	yyyy/mm/dd,hh:mm:ss	A20	Start time of measurement
STOP_TIME	yyyy/mm/dd,hh:mm:ss	A20	Stop time of measurement
C1_NAME		A12	Filter 1 name for color
C2_NAME		A12	Filter 2 name for color
FILTER		A12	Filter name for zero point
LABEL		A70	Data label

3.2.4 Standard Star Transparency

This table lists the collected zeropoint statistics for a period of time (generally a night) as determined by the Elixir photometry system. The extension name EXTNAME will be SUMMARY_ZPTS.

TTYPE	TUNIT	TFORM	COMMENT
ZP_OBS	mag	F8.4	Measured zero point
ZP_REF	mag	F8.4	Nominal zero point
ZP_ERR	mag	F7.4	Error on zero point
C_AIRMASS	mag / airmass	F7.3	Airmass coefficient
C_COLOR	mag / mag	F6.3	Color coefficient
START_TIME	yyyy/mm/dd,hh:mm:ss	A20	Start time of validity range
STOP_TIME	yyyy/mm/dd,hh:mm:ss	A20	Stop time of validity range
C1_NAME		A12	Filter 1 name for color
C2_NAME		A12	Filter 2 name for color
FILTER		A12	Filter name for zero point
LABEL		A70	Data label

3.2.5 SkyProbe Transparency

This table gives the transparency measurements from SkyProbe along with other relevant information. The extension name EXTNAME will be SKYPROBE_ZPTS.

TTYPE	TUNIT	TFORM	COMMENT
START_TIME	yyyy/mm/dd,hh:mm:ss	A20	Start time of measurement
FILTER		A10	Filter name
ZP_OBS	mag	F8.4	Measured zero point
ZP_ERR	mag	F7.4	Error on zero point
RA	dec. degrees	F11.6	RA (J2000)
DEC	dec. degrees	F11.6	DEC (J2000)
C_AIRMASS	mag / airmass	F7.3	Airmass coefficient
SKY	counts	F7.1	Median sky flux
NSTAR	stars	I6	Number of stars in image

3.2.6 QSO Weather Data

This table gives QSO weather comments from the observers. The extension name EXTNAME will be OBSERVER_WEATHER_LOG.

TTYPE	TUNIT	TFORM	COMMENT
TIME	yyyy/mm/dd,hh:mm:ss	A20	Time of measurement
COMMENTS		A255	Observed sky conditions
CONDITIONS		A20	Observer comments on weather

3.2.7 Weather Logger Data

This table gives the data from a collection of CFHT weather probes inside and outside the dome. The extension name EXTNAME will be RECORDED_WEATHER_LOG.

TTYPE	TUNIT	TFORM	COMMENT
TIME	yyyy/mm/dd,hh:mm:ss	A20	Time of measurement
PROBE19	deg celsius	F7.2	
PROBE84	knots	F7.2	Weather tower wind speed

PROBE85	deg azimuth	F7.2	Weather tower wind direction
PROBE86	deg celsius	F7.2	Weather tower air temp
PROBE87	percent	F7.2	Weather tower relative humidity

3.2.8 Graphical Data

This table lists the miscellaneous graphical data types provided with the data distribution and defines their data type. The extension name EXTNAME will be GRAPHICS.

TTYPE	TUNIT	TFORM	COMMENT
FILENAME		A99	Filename on distribution
FILE_TYPE		A3	Graphics type: GIF, JPG, PNG
DATA_TYPE		A99	Quantity represented by data: skyprobe, satellite, etc
START_TIME	yyyy/mm/dd,hh:mm:ss	A20	Start time for plotted period
STOP_TIME	yyyy/mm/dd,hh:mm:ss	A20	Stop time for plotted period

3.2.9 Weather Logger Probe Data

This table defines the collection of probes used in the weather logger data table and their meaning. The extension name EXTNAME will be WEATHER_LOG_PROBES.

TTYPE	TUNIT	TFORM	COMMENT
NUMBER		I3	Data probe number
DESCRIPTION		A50	Data probe description
TYPE		A8	Probe type
UNITS		A20	Measurement Units

3.2.10 Filters

This table defines the collection of MegaPrime filters with their identification tags and their transmission curves. It also includes the environmental conditions under which filter transmission measurements were made. The extension name EXTNAME will be FILTER_ID.

TTYPE	TUNIT	TFORM	COMMENT
NAME		A20	Filter name
CAMERA		A10	Camera
TRANSID		A99	Transmission Filter Curve Name
LAMBDA_CENTRAL	angstroms	I10	Filter central wavelength
BANDWIDTH	angstroms	I10	Filter bandwidth
TRANS_MAX	%	I3	Filter maximum transmission
CTEMP	deg celsius	F5.1	Temperature during filter curve measurement
CHUMID	percentage	I3	Humidity level during filter curve measurement
CPRESSURE		F6.2	Atmospheric pressure during filter curve measurement
START_TIME	yyyy/mm/dd,hh:mm:ss	A20	Start time of filter validity range

STOP_TIME	yyyy/mm/dd,hh:mm:ss	A20	Stop time of filter validity range

3.2.11 Raw Images

This table defines meta-data related to the raw images. These meta-data will be generated by the "quick-look" Elixir pipeline at the summit as images are acquired. The extension name EXTNAME will be CADC_RAW_IMAGES.

TTYPE	TUNIT	TFORM	COMMENT
FILENAME		A99	Filename
HDR_FILENAME		A99	Filename of FITS header
OBSID		I10	Unique ID number
OBS_IQ_REFCCD	arcsec	F5.2	Observed image quality (REF CCD)
OBS_IQ_CENTER	arcsec	F5.2	Observed image quality (center region)
OBS_IQ_R_RATIO		F5.2	Observed IQ ratio (outer / center)
OBS_IQ_X_RATIO		F5.2	Observed IQ ratio (right / left)
OBS_IQ_Y_RATIO		F5.2	Observed IQ ratio (top / bottom)
OBS_BG_VAL	Counts / pixel	F9.3	Observed background

3.3 Meta-Data Transfer Mechanism

Meta-data will be transferred within specific FITS image headers, FITS data tables and in the form of small graphics files (GIF, PNG, JPEG).

The FITS tables included as meta-data with processed MegaPrime science data will be PROCESSED_IMAGES, MASTER_DETREND, IMAGE_ZPTS, SUMMARY_ZPTS, SKYPROBE_ZPTS, OBSERVER_WEATHER_LOG, RECORDED_WEATHER_LOG, GRAPHICS, WEATHER_LOG_PROBES, and FILTER_ID. Each table will be an extension in a multi-extension FITS file, named as described in section 2.1.4, that will be shipped to CADC along with processed science data but on a separate media. Distributions are generated by RUNID, so CADC can expect separate distributions for each Legacy Survey component and each PI run when there are data ready to be processed. This design means that data in many of the tables, such as MASTER_DETREND, *_ZPTS, *_WEATHER_LOG, and GRAPHICS will be duplicate. WEATHER_LOG_PROBES and FILTER_ID will remain unchanged for long periods of time, but are being included with the non-static meta-data content. It will be the responsibility of the CADC to determine if an update to their database for these information is required based on the DATE keyword in the tables.

The RAW_IMAGES fits table will be copied to the distribution directory at CFHT where CADC will be responsible for retrieving it on a daily basis.

4 Software

This section describes the software that will be shared between CFHT and CADC.

4.1 Elixir (CFHT)

4.1.1 Description

The Elixir Processing System is used to create detrended, calibrated science images from the raw science frames. The system automatically updates the astrometry and photometric keywords for each image and generates grey-scale jpeg images at two different binning factors. The system requires the availability of the appropriate meta-data in specified locations.

- CFHT provides Elixir-detrending (includes photometric and astrometric modules).
- Elixir-detrending must run automatically at CADC with no visual assessment of the results required at CADC.
- How will we identify all inputs, configurations, parameters and software versions needed to produce a detrended file?
- CFHT will keep CADC informed of the correct software versions to apply to datasets as they are delivered to CADC.
- Elixir-detrending will always be backward compatible. The latest version will be capable of processing any file produced at any time by MegaPrime and will provide the best possible processing.
- When the need for an urgent patch is discovered, CADC will be notified as soon as possible of which datasets are affected by the problem.
- Elixir-detrending will stamp the data that it processes with an accurate version number.
- There will be a description of the operation of Elixir appearing in PASP in 2002.
- CFHT will provide a complete user's manual with the purpose of informing users of how Elixir works.

4.1.2 Installation and Maintenance

- CFHT will visit CADC to do the initial installation and configuration of Elixir-detrending at CADC and train CADC scientists into its use.
- Reference and configuration files needed will be located in directory at CFHT that will be remotely synchronized via rsync with the CADC
- CFHT will provide advance notification (30 days) of any new releases or patches to software except in the case of an urgent patch. CFHT will notify CADC in advance of the need to make a change in Elixir so that proper support can be provided.
- An urgent patch will be required when a software error is detected that has serious consequences on data quality. CFHT will provide urgent patches when they are necessary.
- Changes to software must be documented in a changes document, and if necessary the Elixir user and installation documentation will be updated.
- If re-processing is necessary, CFHT will provide and install the patch and notify CADC of the actions required to process the affected datasets.

5 Data Management

The meta-data database and the CADC data management system will contain information on content and distribution of MegaPrime products. CADC will provide interfaces to allow CFHT (and TERAPIX) to query this information for their own needs. The interfaces will be http proxy servers coupled to HTML web pages. The CFHT feedback page will allow CFHT to see which MegaPrime files have been successfully received, verified and ingested in the CADC systems and will also allow CFHT to manually or automatically download any file they may need to replace missing data or repair corrupted data in their own MegaPrime collection.

6 Distribution of Real-time Processing Results

6.1 Proprietary policy

It is believed at this time that the results of real-time processing will be proprietary to the communities (C and F) for one year but NOT proprietary to individuals or smaller groups. CADC will provide an interface to real-time results that protects their proprietary status.

6.2 Format of results

The results of real-time processing will be in the form of tables of information and postage stamps which will be in FITS format and will conform to the standards described above for FITS format files.

Tables of information will be in FITS table format.

6.3 Posting results at CADC

The results of real-time processing will be transferred from CFHT to CADC by network. CFHT will maintain a data area that can be synchronized with CADC to effect the transfer. The decision regarding which data will be placed there belongs to CFHT. CFHT will decide upon and enforce a standard format for real-time results.

(A second possibility is for CADC to clone a web site that is completely produced and managed by CFHT.)

7 Authentication System for CFHTLS users

CFHT is responsible for deciding who is authorized to access CFHTLS data.

CFHT will send the full list of CFHTLS users via e-mail two weeks before official data distribution is set to begin. Following the ingestion of this initial list into the CADC system, CFHT updates to the list of authorized CFHTLS users will be sent to CADC via e-mail when necessary. These updates will be incremental changes, i.e. they will only contain user entries that need to be changed in the list currently residing on the CADC system. Possible changes are: (1) addition and/or activation of a registered user, and (2) deletion and/or de-activation of a registered used. Each user entry in the CFHT updates should contain the following attributes: LAST_NAME, FIRST_NAME, TITLE, POSITION, AFFILIATION, EMAIL_ADDRESS, COMMUNITY, PASSWORD, ACTION. The ACTION attribute specifies which type of change the entry represents. To add/activate a registered user, ACTION should be set to “A”. To delete/de-activate a registered user, ACTION should be set to “D”.

Created: November 14, 2001
Modified: March 21, 2002

Canada
France
Hawaii
Telescope
Legacy
Survey

Chapter 3

CFHT to TERAPIX Interfaces

1. Introduction

3. Meta-Data Interface

The meta-data database described in Section 7 on page 7 will be used to transfer meta-data from CFHT to TERAPIX.

4. Software

4.1 Description

Elixir?

4.2 Installation and Maintenance

See section 4 on page 9

5. Others

Notification that data must be re-observed

Created: November 14, 2001
Modified: December 4, 2002

Canada France Hawaii Telescope Legacy Survey

Chapter 4

CADC to TERAPIX Interfaces

1. Introduction

This chapter describes the interfaces between CADC and TERAPIX for the CFHTLS project. The interfaces required between CADC and TERAPIX are:

- The bulk data interface which described in Section on page 14.
- The meta-data interface which is described in Section 3 on page 16. *(INCOMPLETE)*
- The transfer of software which is described in Section ?? on page ?? *(INCOMPLETE)*.
- Processing control which is described in Section ?? on page ?? *(INCOMPLETE)*.
- The transfer of TERAPIX created catalogs which is described in Section ?? on page ?? *(INCOMPLETE)*.

To better understand the TERAPIX data products, a few words must be said about the way the TERAPIX pipeline works. Science images are managed in 2 different "contexts", which indeed represent 2 different pipelines that can operate asynchronously:

- The first context is that of data ingestion from CADC. At this level, images are handled on an observing-run basis; affiliation to a particular survey program is ignored. Each incoming file is first tested for integrity and content, and registered in the database. For each science exposure, image quality is automatically evaluated through various estimators describing background homogeneity, statistics of astronomical sources, and defects such as saturated pixels, cosmic ray impacts, glitches, trails, bright spikes and aureoles[1]. At the same time, a weight-map and a flag-map incorporating masks for the most obvious of these defects are assigned to each exposure. A shallow source catalog is then extracted, which uses the weight-map to reject image artifacts. Eventually, a PSF model is built for each CCD, and used for tracking image sharpness and elongation over the focal plane.
- The second context involves all images and their catalogs of a given filter over a given sky area, or a set of sky areas. The (non-linear) astrometric solution is computed using both a reference astrometric catalog and overlapping detection measurements. The more the overlapping detections, the better the solution. A similar method is applied for photometric calibration; in this case, exposures taken during photometric nights act as anchor points for the zero-point. Calibrated images can then be resampled and co-added around each survey region[2]. Finally, panchromatic source extraction for the final catalog starts once all available data in all available filters concerning a surveyed area have been co-added. Quality assessments concerning the final products are also performed at this stage.

En [1] Automatic identification of trails and aureoles will be implemented in the second half of 2003.
[2] PSF homogenization is planned for implementation at this position in the pipeline in late 2003/early 2004

Most of the survey programs extend over several MegaPrime runs, hence the solution for (inter-) calibration and the resulting co-added products will significantly improve with time (each month). This would make incremental releases of the "final" product suboptimum with most observing strategies. Waiting for several years before releasing the whole survey is also excluded. Hence it was decided to rerun the second part of the pipeline each month (or season, depending on the available computing power and transfer bandwidth) on all the incomplete data acquired so far. It is important to understand that with this production scheme each new version of a given observing program supersedes the previous one. It benefits from both an increased number of pointings and from software improvements/bug fixes. This means that the final science images will get deeper and/or grow in size from one release to the next. In what follows we shall simply call these releases``TERAPIX releases'', in contrast to the continuous creation of weight-maps and other products generated during ingestion of the data sent by the CADC. The decision whether one shall archive or not previous versions is left to the science steering group and CADC.

2. CADC to TERAPIX Bulk Data Interface

Bulk data transfer between CADC and TERAPIX goes in both directions. Detrended FITS images will be transferred to TERAPIX and several levels of processed data products will be transferred from TERAPIX back to CADC.

2.1 Content

The content of the bulk data transfer is FITS images, that is, pixels with header information.

Detrended data, processed detrends and all meta-data files will be transferred from CADC to TERAPIX. These files will be as described in Section 2.1 on page 5 of Chapter 2.

Processed data files will be transferred from TERAPIX to CADC. These processed files will be MEF files. The MEF headers will contain a history of the data processing used to create the files. The MEF files will be fully described by the headers.

Any calibration file generated by TERAPIX shall be transferred to CADC.

The bulk of data transferred from TERAPIX to CADC will consist of images, catalogs and auxiliary data generated by the pipeline modules.

The format will be FITS and MEF for images, binary FITS tables for catalogs, as well as binned PNG images[3] and Postscript for plots. More precisely, the files consist of:

1. Weight-maps and flag-maps generated for each science MEF file,
2. For each exposure, PNG thumbnails of the
    a. Binned CCD content,
    b. Background level (low surface-brightness features),
    c. Point Spread Function model.

The data above are produced "on the fly" as images are received from CADC. The following is created only for each TERAPIX release:

[3] PNG is preferred to JPEG in the context of quality control, to prevent compression artifacts from DCT quantization to affect the image thumbnails (lossless JPEG is not well supported and less efficient).

1. Updated headers for the detrended images ("pseudo-MEF" files, that is, updated MEFs without the data bodies),
2. Postscript plots related to the astrometric reduction,
3. Postscript plots related to the photometric reduction,
4. Co-added images[4] in each filter and their composite weight-maps (both in basic FITS)
5. The so-called "χ^2" images (see Szalay et al. 1999) combining all filters,
6. Multi-band source catalogs based on the "χ^2" images for detection, and the co-added images for measurement,
7. PNG thumbnails of the
    a. Co-added and "χ^2" images,
    b. Background level (low surface-brightness features),
8. Postscript plots related to the quality control of the "final" catalog (including star and galaxy counts, 2 point correlation functions, magnitude-color and color-color diagrams)

2.1.1 Processed data file names

The new headers associated to the detrended images will be in the form %d%c.head, where %d%c is as in the CFHT filename %d%c.fits (see 2.2.1.1). Example: The updated header for 5468901o.fits will be called 5468901o.head.
Following the CFHT convention, the filenames of combined images created at TERAPIX are built like serial numbers: they will be in the form %c%3s%06d%c.fits:

- The first character (a-z) refers to the physical processing pipeline (or generally the site). Images processed at TERAPIX will begin with a "t".
- The following set of 3 characters codes the release number, starting with "aaa", then "aab", and so on up to "zzz".
- The TERAPIX image number is a 6-digit decimal number, right-justified and padded with 0's. At each new release the first product is reset to 000001.
- The last character before the dot is the image type. Currently envisioned image types are
    - a — averaged[5] (co-added data)
    - s — subtracted (image subtraction)
    - c — combined with the "χ^2" method (see 4.2.1)

Example: tarb000237a.fits. Note that contrary to CFHT filenames, there is currently no explicit case dependency, although lower-case is assumed by default.

2.1.2 Weight-map and flag-map File Names

The weight-map associated to the image %s.fits will be %s.weight.fits, and the flag-map %s.flag.fits. This is true for both detrended images from CFHT and the co-added images created by TERAPIX. Example: The weight-map and flag-map for 5468901o.fits will be called 5468901o.weight.fits and 5468901o.flag.fits, respectively. Other example: the weight-map and flag-map for tarb00237a.fits will be called tarb00237a.weight.fits and tarb00237a.flag.fits, respectively. Note that the image number is not incremented for weight-maps and flag-maps: there are therefore more TERAPIX products for each release than image numbers.

2.1.3 Auxiliary FITS File Names

Meta-data files related to the detrended data are generated in the context of an observing ("camera") run. Hence their naming convention is very similar to that of CFHT files. The multi-extension FITS file containing the meta-data for a camera run will have a name of the form CRUNID.tmd.VERSION where CRUNID is the camera run id (e.g., 01Ak01 = 2001/01/27 – 2001/02/06) and VERSION is a version number to force the names to be unique.
I need more input from my colleagues: TBW

[4] weighted combination-only in a first time; with homogenized seeing later on (2004).
[5] May not necessarily be a linear combination of pixel values

2.2 Mechanism

2.2.1 Media Type

TBW

2.2.2 Media Names

TBW

2.2.3 Handling

TBW

END OF ICD UPDATES SENT BY E. BERTIN ON DECEMBER 3, 2002

3. Meta-Data Interface

CADC will inform TERAPIX when new data become available in the outgoing FTP area. By a mechanism TBD.

TERAPIX will supply a "data dictionary" describing the FITS keywords that are allowed and expected in each of the data product FITS files. The data dictionary will allow keywords to evolve over time. The data dictionary format will be based on the data dictionary used by the fitsIngest program described in Section 4.1 on page 9.

CADC will need an electronically parsable description of the data processing sequence that was used to create each processed data product. The information will include:

- The inputs to each data processing step. The inputs are either detrended data files, or the outputs from other data processing steps.
- The options and/or configuration of the data processing step.
- Software version used to do the data processing.

CADC has both Sybase tables and XML documents which describe data processing as a DAG. Ideally one of these formats would be used to describe the processing which produced a data product.

We need a mechanism to allow data to be re-processed when necessary.

4. CADC to TERAPIX software interface

4.1 TERAPIX

4.1.1 Description

TBD by TERAPIX

4.1.2 Installation and Maintenance

TERAPIX will install software at the CADC.

4.2 Other

CADC may provide software related to physical transfer of bulk data and ingestion of data and meta-data.

CADC may provide distributed processing software.

5. Processing Control

TERAPIX will be responsible for the supervision of the processing done with the TERAPIX software regardless of where the processing actually occurs. For example, TERAPIX will be responsible for informing CADC when processing or re-processing is appropriate or required, which versions of the TERAPIX software to use with which datasets, which inputs and outputs to use or expect and whether the outputs replace existing data products or are supplementary to existing products. If a centralized database is used, it might be possible to have TERAPIX detecting the "redo" switch automatically and (re) process the affected files. *In fact, we might want to say somewhere that these pipelines should be database driven in order to have a good tracking mechanism.*

TERAPIX is supervisor of processing.

TERAPIX will control processing, including choosing which inputs to use, and which software versions.

If TERAPIX processing is done at both TERAPIX and CADC, we will have to develop and document a method of sharing the data processing load. This will include:

- Deciding which data will be processed by each site.
- Possibly duplicating some data processing and automatically cross checking the data processing output.

There are several possible mechanisms for controlling the processing:

1. CADC is in the final stages of developing a distributed data processing infrastructure. This infrastructure could be used by both TERAPIX and CADC. TERAPIX would insert DAGs describing the desired data processing into the data processing infrastructure at both CADC and TERAPIX. TERAPIX would be responsible for creating and maintaining a "to-do" list.

   CADC would take primary responsibility for integrating the TERAPIX data processing tasks into the infrastructure.

2. CADC could use the CADC data processing infrastructure, and TERAPIX use another data processing infrastructure. TERAPIX would create and populate a "to-do" list of processing tasks. CADC and TERAPIX would take tasks from the to-do list and do the processing.

3. CADC would provide TERAPIX with access to computer resources which would be managed remotely by TERAPIX.

Data Processing control should be database driven.

6. Catalogues

If TERAPIX generates catalogues, CADC would like to create a database of these catalogues. This section will document what types of catalogues TERAPIX will generate, and how the catalogues will be transferred to CADC.

6.1 Content

This section will contain a detailed description of the catalogues, and the format of the files.

6.2 Mechanism

TERAPIX will include FITS tables in the data returned to CADC.

Created: November 14, 2001
Modified: March 24, 2002

Canada
France
Hawaii
Telescope
Legacy
Survey

Chapter 5

Other Interfaces

1. Contact

In addition to the formal interfaces between these groups, ongoing collaboration between CFHT, CADC and TERAPIX will be required during the course of the CFHTLS project. To facilitate this collaboration, the CADC suggests that:

1. Each site assigns a contact person responsible for coordinating the interactions with the other partners.
2. The partners will participate in annual coordination meetings.

2. Coordination

There should be a single point of contact at each institute who will be responsible for dealing with problems.

2.1 Re-observing

If CFHT, CADC, or TERAPIX detects the need to repeat observations, then there must be a mechanism to do so.

2.2 Problem Reports

Problem reports will need to be collected, archived and dealt with in some systematic way.

2.3 User Support

What if we get questions from users that we cannot answer? Pass them along to CFHT or TERAPIX? Could be diffused with Web pages.

What is role of SG versus us to serve user requests? How does this software work?

PI mode: problem reports go directly to CFHT.

CFHTLS: Problems go to SG to resolve them.

2.4 Reports to Survey Execution Authority

Each of the partners will make a regular report to the SG.

2.4.1 Changes to Meta-Data Database Schema

2.4.2 Changes to the FITS Data Dictionary

2.4.3 Reprocessing with Elixir