RevMexAA (Serie de Conferencias), **24**, 119–124 (2005)

THE GTC DATA FACTORY: PRESENTATION AND CURRENT STATUS

P. Gómez-Álvarez[1] and C. E. García-Dabó[1]

RESUMEN

Se presenta el subsistema de reducción de datos del telescopio GTC, el así llamado *"Data Factory"*. Se describen, asimismo, sus objetivos principales, funcionalidad, arquitectura global y organización interna así como su estado actual de desarrollo.

ABSTRACT

We present the GTC data processing subsystem, hereinafter *"Data Factory"*. The main objectives, functionality and overall philosophy are described. We also briefly outlay the architectural description of the Data Factory and its organization. The current state of development is also presented as well as some real reduction examples.

Key Words: **METHODS: DATA ANALYSIS**

1. INTRODUCTION

Given that technical developments in the detector, optical and specially in the detector fields have allowed more, better and reliable data, the software tools to reduce, analyze and search data have increased their relative importance. We can consider the data reduction software as the "last link" in the proposal preparation, atmosphere, optics, mechanics, detector and software chain, therefore contributing multiplicatively to any observational effort.

Taking this principle in mind, the main objectives of the GTC Data Factory are:

- Provide reliable reduced scientific data to GTC users, thus increasing the GTC scientific return.
- Provide a programming framework which allows external instrumentation groups to develop their custom pipelines. Currently, three groups are developing their pipelines using the provided framework:
  - IAC group for the OSIRIS data pipeline
  - UCM group for the EMIR data pipeline, see Gallego et al. (2003)
  - LAOMP group for the EMIR simulator

  The following pipelines are being developed at the GTC Project Office, using the same framework:
  - ELMER pipeline, see García Vargas et al. (2003).
  - Commissioning camera, see Cuevas et al. (2003)
  - Acquisition, Guiding and wavefront sensing, see Devaney et al. (2000)
- To allow an homogeneous reduced data archive.
- To support the GTC commissioning phase, making it more efficient.

It can be objected, taking into account the myriad of good and proved Data Reduction Packages (DRPs) available, why another DRP need to be developed. This is not the place to make and exhaustive trade-off, however some aspects can be pointed out:

- Software reuse does not come from free. Significant effort is needed to wrap existing DRPs.
- The GTC Acquisition, Guiding and wavefront sensing (AGWFS) system, see Devaney et al. (2000) and the commissioning instrument demand significant data reduction functionality in a Real Time environment which is not sensible to implement using existing DRPs.
- AGWFS processing shares basic data processing functionality with the off-line processing. A portable framework is then needed.
- Error propagation resides in the Data Factory core, in contrast with other DRPs. See Cardiel et al. (2002).

[1] GTC Project Office, Instituto de Astrofísica de Canarias, La Laguna, Tenerife, Spain.

2. DATA FACTORY ORGANIZATION

The GTC data factory, as the rest of the GTC software services are being designed using a strict object-oriented approach and implemented in C++. This technique allows the existence of “layers” of software which represent different levels of complexity and abstraction, which eases and simplifies the implementation of data reduction pipelines (Reduction Templates using the Data Factory jargon).

The Data Factory answers two sorts of data reduction, those which need to run in a real time (RT) environment (normally in VME crates running the VXWorks operative system), typically closing a control loop and the offline functionality usually devoted to scientific data reduction.

From bottom to top, the Data Factory is organized in three software layers (See Fig. 1):

2.1. *The Data Processing Kit (DPK)*

The DPK is a software framework composed by those reusable elements used to implement the instrument-specific reduction templates. It also allows rapid building and prototyping of data reduction applications written in C++, encapsulating data reduction technicalities and thus allowing DPK users to concentrate on expressing data reduction procedures as directly as possible.

In order to cope with the portability requirement, the DPK is implemented using ANSI C++, and using some well-proved C libraries as cfitsio, gsl and wcslib, in order to implement the most basic functionality regarding FITS Input/Output, mathematical algorithms and World Coordinate Transformations respectively. See Pence (1999), Galassi et al. (2002) and Greisen & Calabretta (2002).

Only to give a hint of what the DPK is capable in terms of encapsulation, abstraction and expressiveness, the following is the needed code to generate a simulated image, similar of that in Figure 3:

```
//Define the size of the image
TinyVector<unsigned long,2> size(512, 512);

//Instantiate the 2D float data with the
//specified dimensions
DPKArray<float,2> data(size);

//Simulate some poissonian noise
data.rand(POISSON, 1000.0);

//Create the frame
Frame<float,2> image(data);

//Instantiate a filter which creates a
//simulated star
AddNoise<float,2> starMaker;

//We simulate a 50.000 counts, 1 pixel
//FWHM star in the middle of the image.
starMaker.run(image, 256, 256, 1.0, 5E4);

//We store the frame in FITS format.
image.toFits("example.fits");
```

For a detailed description of the Data Processing Kit and functionality, see Gómez-Álvarez and García-Dabó (2004) and the API HTML documentation included in the most recent GCS Software release.

The DPK is composed by three modules:

2.1.1. *The DPK framework*

The data processing kit framework is composed by those entities which represent the fundamental concepts involved in any data reduction process.

It is composed by the basic entities (classes) which represent and manipulates the data and its associated errors, see Fig 2.

Amongst the functionality included are:

- High-level manipulation of N-dimensional data (DPKArray and related classes).
- Region-of-Interest (ROI) manipulation (Roi, Shape and related classes): Given an Array, arbitrary shaped regions can be defined and manipulated.
- Frame manipulation. This and related entities represent an arbitrary scientific data and associated errors produced by any of the instruments of the GTC. It frequently represents a two-dimensional CCD image but it can also represent any possible HDU data present in the FITS context. It is also a fundamental concept within the GTC data reduction, it represents the raw input data, the intermediate reduced data as well as the final reduced data. Amongst its capabilities are FITS Input/Output, keyword manipulation and error management.

2.1.2. *The filter library*

Currently (release 0.7) it is composed by more that 40 multipurpose filters, which implement the most basic data reduction procedures as: cosmic rays cleaning, frames combination, bias computation, defringing, spectra extraction, flatfielding, error generation, basic arithmetical manipulations as well as data simulation capabilities.

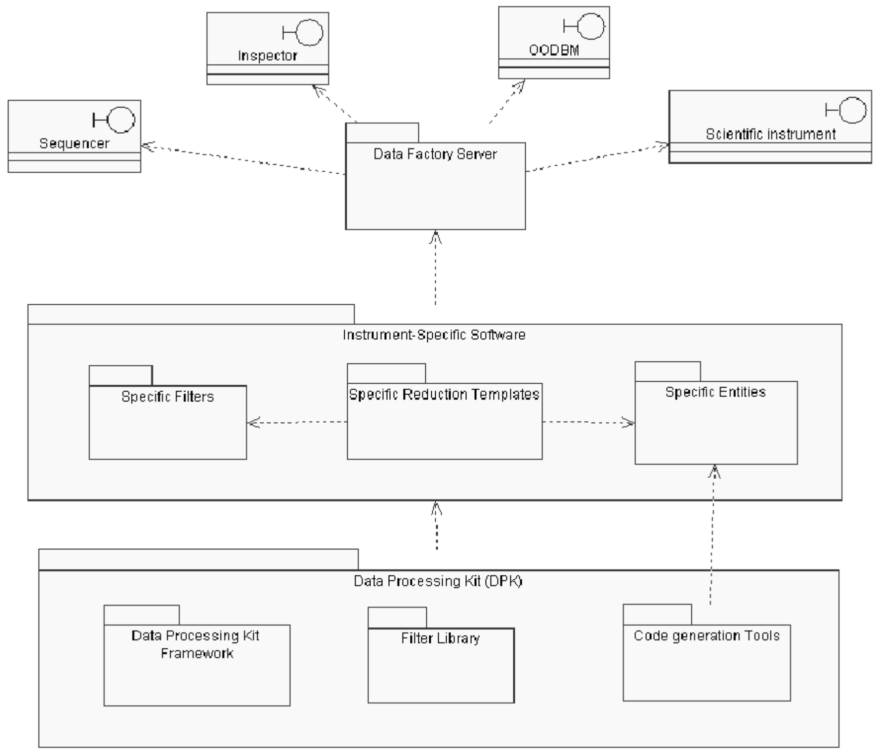

Fig. 1. Data Factory layered structure.

2.1.3. *Code generation tools*

This package contains utilities to quickly generate new types in the specific entities (See figure 1 and 2.2) which are exclusive of the instruments. Once generated, more functionality could be implemented in these classes.

2.2. *The instrument-specific software*

The instrument specific software is developed by the instruments groups and is composed by three types of entities:

- The classes needed to represent the reduction concepts which are particular of the instrument data reduction.
- The specific filters which are needed for the reduction process.
- The reduction templates for each observing mode of the instrument of interest. Each observing mode in the GTC produces a specifically structured set of data, which is canonically reduced in a predefined way which is called "Reduction Template". These reduction templates are implemented by means of a concatenated set of filters (both specific and generic) which operates over those entities involved in the reduction process.

2.3. *The data factory server*

The Data Factory Server is responsible of integrating the Data Factory in the GTC control system. It implements a CORBA, see Penataro et al. (2000), interface allowing the access to the Data Factory functionality by the other subsystems.

It main responsibilities comprises:

- Exporting an interface visible from the GTC graphical user interface (Inspector).
- Identifying the incoming raw Frames to be reduced.
- Retrieving from the Object-Oriented data base (OODB) the most recent calibration data.
- Instantiating and execute the appropriate Reduction Template.
- Storing raw and processed data in the OODB and in FITS format for distribution to the GTC users.

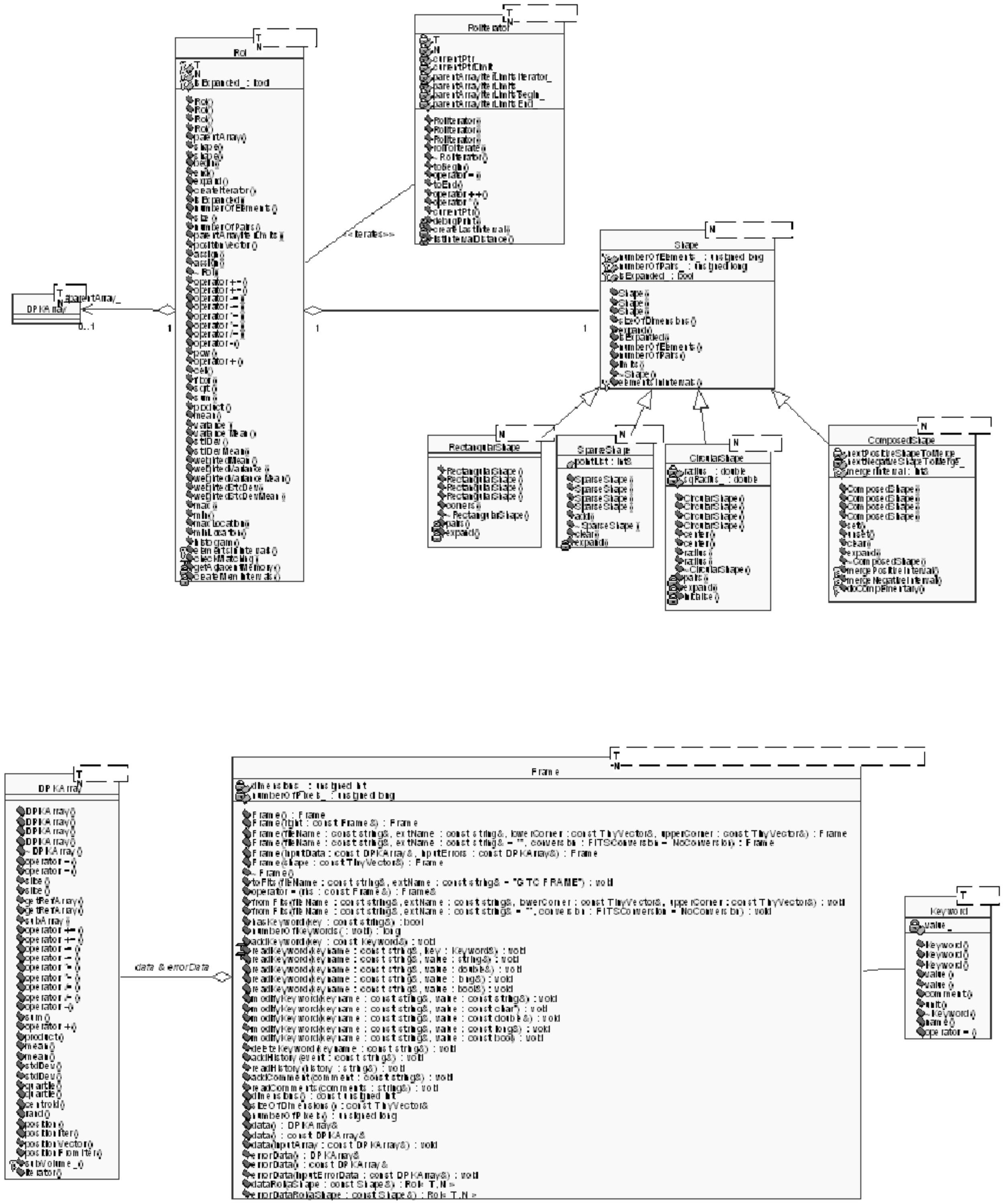

Fig. 2. The Data Processing Kit(DPK) most relevant classes.

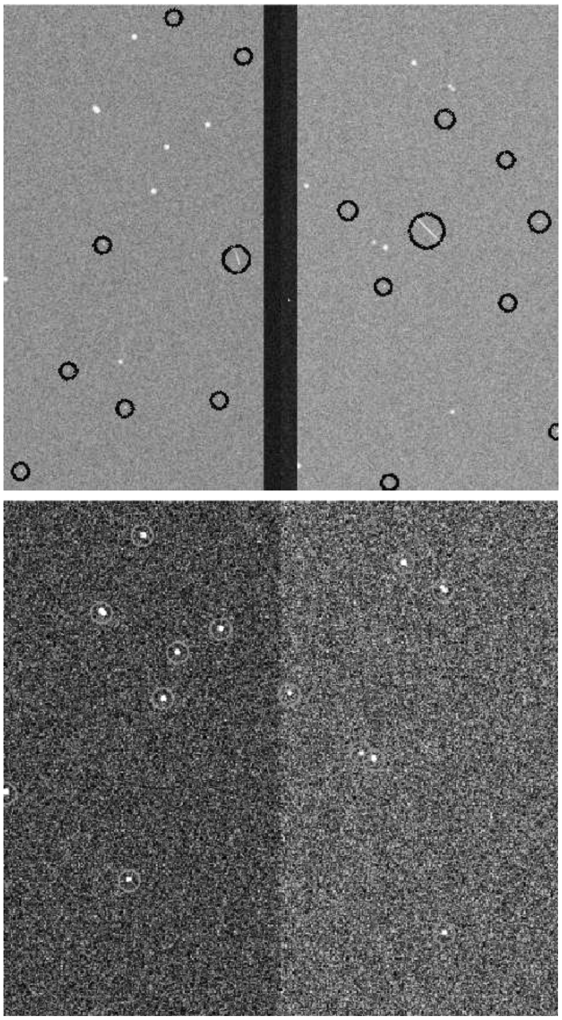

Fig. 3. The first image is a simulation with artificial noise, stars and cosmic rays added and detected. The second image is the processed one, in which the cosmic rays have been removed and the present stars detected.

3. REDUCTION TEMPLATES DEVELOPED IN-HOUSE

The Reduction templates developed at the GTC Project Office are those pertaining to the ELMER instrument, the Acquisition and Guiding Camera and the commissioning instrument. The first part of any reduction is the so-called *basic reduction*, which is composed by: trimming, bias/dark subtraction, flat correction, cosmic rays and cosmetic defects cleaning.

3.1. *ELMER RTs*

The pipeline of ELMER comprises reduction templates for all the observing modes that ELMER implements, which can be grouped in three types:

- Imaging modes. The reduction steps for these modes include the basic reduction, image restoration and flux calibration.
- Spectroscopic modes. The reduction steps include basic reduction, sky subtraction, extraction of spectra, wavelength calibration and flux calibration.
- Multiobject spectroscopy modes. The reduction steps for these modes include basic reduction, identification of spectral sources in the image by means of a correlation with a direct image, extraction of spectra, wavelength calibration and flux calibration.

3.2. *Acquisition, guiding and Wavefront sensing RTs*

The AGWFS pipeline has been developed with the same tools and software infrastructure that the scientific pipelines. This constitutes a novel approach and has been possible thanks to the DPK and GTC control system flexibility and portability.

The type of data produced by the AGWFS subsystem is divided into three main categories:

- Acquisition and slow guiding data. This is produced mainly by the ASG camera. For the observing modes defined, a basic reduction is performed which includes bias/dark subtraction, flatfielding and cosmic rays removal. A star-detection algorithm is then used to find objects in the field. The objects are cross-correlated with known catalog positions to obtain WCS parameters and identify the stars in the field[2]. In the case of acquisition, an interactive procedure allows the astronomer to select the observation target and guiding star. In the case of slow guiding, the guiding star is identified. A major goal is to obtain photometric/seeing monitoring during the guiding process. Note that all these routines are implemented in a real time environment. Figure 3 shows a simulated frame for the ASG camera, including stars and cosmic rays.
- Fast guiding data. The Fast Guider produces images at a rate up to 500 Hz. The data reduction steps are the same that in the slow guiding data, however, it is expected that due to performance issues basic reduction is to be

[2]This automatic field recognition might not be present on first light, instead the acquisition process must be through the graphical interface (Inspector).

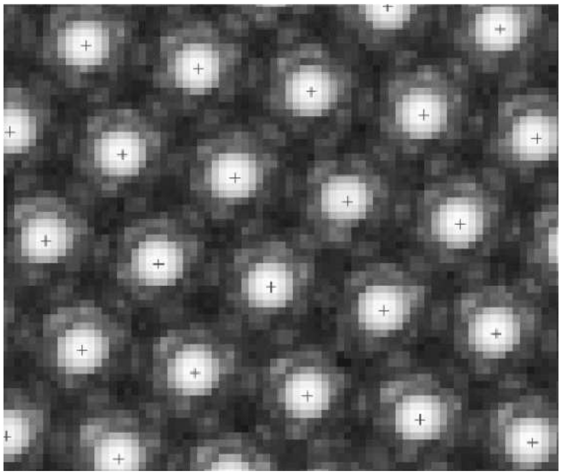

Fig. 4. Star detection as part of the Process Commissioning Frame reduction template.

skipped and only the centroid determination is done. The centroid information is then passed directly to the secondary mirror control system in order to close the fast guiding.

- Wavefront sensing data. This data is produced by the SFS and STS cameras. The images obtained are formed by a Shark-Hartmann wavefront sensor. The reduction steps include basic reduction and the precise determination of the spot positions and errors. Using a 2-D fitting procedure we have been able to compute very precise positions with typical errors of 0.03 pixels, as tested by means of simulations. The resulting data is then passed to the Observing Engine which reconstruct the incoming light phase and operates over the available degrees of freedom of the optics to compensate the observed errors, thus closing the loop for active optics.

3.3. *Commissioning instrument RTs*

The commissioning instrument will perform basic optical testing and will provide the first quality control of the telescope optics. It provides imaging and a high-resolution wavefront sensing capabilities. The reduction templates used to reduce this data are similar to those used by the AGWFS pipeline. Figure 4 shows the centroid determination for a frame taken with the commissioning camera detector and lenslet under laboratory conditions.

4. CONCLUSIONS

The GTC telescope will have a complete and reliable set of scientific pipelines from first light. In order to support data reduction pipelines developed outside the GTC PO, a developing framework is provided and updated every 6 months to the instrument groups. The data reduction pipeline of the ELMER instrument is being developed within the GTC project office allowing not only a working scientific instrument at GTC first light but also reliable reduced data from the very beginning. The engineering data reduction requirements for acquisition, guiding and wavefront sensing will be satisfied using the same framework used for scientific data reduction.

We expect that the GTC Data factory will be an useful tool which catalyze the GTC commissioning, allows a quick scientific return and produce and maintain an homogeneous data archive.

REFERENCES

Cardiel, N., Gorgas, J., Gallego, J., Serrano, A., Zamorano, J., García-Vargas, M., Gómez-Cambronero, P., & Filgueira, J. M. 2002, Proc. SPIE, 4847, 297

Cardiel, N., Gorgas, J., Gallego, J., Serrano, A., Zamorano, J., García-Vargas, M. L., Gómez-Cambronero, P., & Filgueira, J. M. 2003, RevMexAA Ser. Conf., 16, 73

Cuevas, S., et al. 2003, Proc. SPIE, 4840, 517

Devaney, N., Cavaller, L., Jochum, L., Bello, C. D., & Castro, J. 2000, Proc. SPIE, 4003, 146

Gallego, J., Zamorano, J., Serrano, A., Cardiel, N., Gorgas, J., García-Dabó, C. E., & Gil de Paz, A. 2003, RevMexAA Ser. Conf., 16, 275

García Vargas, M., Hammersley, P. L., Sánchez-Blanco, E., Cavaller, L., Martín-Fleitas, J., Kohley, R., & Medina, M. 2003, RevMexAA Ser. Conf., 16, 9

Gómez, P., García, C. E, 2004, GTC Document NTE/CTRL/0204-R

Greisen, E. W. & Calabretta, M. R. 2002, A&A, 395, 1061

López-Ruiz, J. C., Díaz, J. J., Gago, F., Garzón, F., Patrón, J., Pelló, R., & Gallego, J. 2002, Proc. SPIE, 4848, 474

Galassi, M., et al. GNU Scientific Library Reference Manual - Second Edition, ISBN 0954161734

Penataro, R., Filgueira, J. M., Gómez-Cambronero, P., González, M., & Puig, M. 2000, Proc. SPIE, 4009, 152

Pence, W. 1999, ASP Conf. Ser. 172: Astronomical Data Analysis Software and Systems VIII, 8, 487

P. Gómez-Álvarez and C. E. García-Dabó: GTC Project, Instituto de Astrofísica de Canarias, Vía Láctea s/n, 38200 La Laguna, Spain.