RMIB GERB Processing Data Flow

Luis Gonzalez

August 16, 2000

Contents

CHANGE RECORD

Issue	Date	Approved by	Reason for change
Draft	17/08/99		new document
Draft 2	25/08/99		§"RGP Processing" updated, § "Log and Error Tracking" added Operator Tasks
Draft 3	14/01/00		updated document
Version 1			

1 Introduction

1.1 Purpose of this document

The purpose of the document is to give an overview of:

1. the data flow between the several processes
2. the process management
3. the general structure of the global process

1.2 Scope of this document

The process internal data flows are not described here. Every process is considered as a black-box. The internal data flows are explained in the documents describing each process.

2 Acronyms, abbreviations and definitions

2.1 Acronyms and abbreviations

Not all these acronyms may appear in this document.

AD	Applicable Document
ADM	Angular Dependency Model
ASI	Agency Spatiale Italiano
BATS	Belgian Advanced Technology Systems
BB	Broad Band
BDRF	Bidirectional Reflection Function
CERES	Clouds and the Earth's Radiant Energy System
CNRS	Centre Nationale de la Recherche Scientifique
DCT	Discrete Cosine Transform
ERB	Earth Radiation Budget
ERBE	Earth Radiation Budget Experiment
EUMETSAT	EUropean organisation for the exploitation of METeteorological SATellites
FTP	File Transfer Protocol
Gb	Gigabit
GB	GigaByte
GERB	Geostationary Earth Radiation Budget
GGSPS	GERB Ground Segment Processing System
GIST	GERB International Science Team
GKSS	GKSS research centre
HD	High Density
HDF	Hierarchical Data Format
HRIT	High Rate Information Transmission
ICD	Interface Control Document
ICSTM	Imperial College of Science Technology and Medicine
IFR	Instantaneous Filtered Radiance
kb	kilobit
kB	kiloByte
LARC	LAngley Research Centre (NASA)
LU	Leicester University
LOS	Line Of Sight
LW	Long Wave
Mb	Megabit
MB	MegaByte
NANRG	Non Averaged, Non Rectified, Geolocated
NB	Narrow Band
NERC	Natural Environment Research Council
OSTC	Office for Scientific, Technical and Cultural Affairs

PSF	Point Spread Function
qlt	quick look time
RAL	Rutherford Appleton Laboratory
RD	Reference Document
RMIB	Royal Meteorological Institute of Belgium
RGP	RMIB Gerb Processing
ROLSS	RMIB On Line Short term Services
ScaRaB	Scanning Radiometer for radiation Budget studies
SEVIRI	Spinning Enhanced Visible and InfraRed Imager
SOL	Start Of Line
SBDART	Santa Barbara Discrete ordinate Atmospheric Radiative Transfer see http://arm.mrcsb.com/sbdart
SW	Short Wave
TAR	Tape ARchive
TBC	To Be Confirmed
TBD	To Be Determined
TOA	Top Of Atmosphere
TRMM	Tropical Rainfall Measurement Mission
UKMO	United Kingdom Meteorological Office
VIRS	VIsible and InfraRed Scanner
wrt	with respect to
WWW	World Wide Web

2.2 Definition of terms

Autorised Users

Authorised users will include RMIB, members of the GIST, GERB operations staff based at RAL, EUMETSAT and others. In order to become an Autorised user, users will be required to register with RMIB according to the rules established by the GERB Project Steering Group (TBC). Once the registration is approved the user will be able to obtain products from the ROLSS, and will be given their unique ROLSS account details, from which they will be able to get access to the products.

near real time (or NRT or nrt)

Between EUMETSAT and the GGSPS near real time means that the GGSPS at RAL will receive GERB data within 60 minutes (TBC) of the actual time at which the data was generated on board the GERB instrument.

quick look time (or QLT or qlt)

Between the GGSPS and RMIB, quick look time means that the RMIB will receive GERB data products within 4 hours of the actual time at which the corresponding data was generated on board the GERB instrument.

short term

Short term data usage refers to usage of the data within one month after their measurement.

long term

Long term data usage refers to usage of the data more than one month after their measurement.

Reference Earth Ellipsoid

(TBD)

GERB data loss

A GERB data loss is one for which the L1.5 radiances have not been obtained within the QLT.

SEVIRI data loss

A SEVIRI data loss is one for which the L1.5 SEVIRI images have not been received through the HRIT system within the QLT.

climate data

GERB climate data means the GERB products that are generated at RMIB for long term archival and data distribution at RAL.

near real time data

Near real data means the GERB products that are generated at RMIB and distributed for short term usage through the ROLSS server.

ROLSS server

The ROLSS server will be the RMIB based ftp server for near real time data distribution.

near real time processing

Near real time processing will be the processing that has to be done at RMIB as soon as a GERB or SEVIRI data file arrives.

daily, weekly, monthly processing

Daily, weekly, monthly processings will be the processings that have to be done at RMIB less frequently than the near real time processing, but that still need to be done on a regular basis.

casual processing

Casual processings will be those that have to be done at RMIB infrequently and irregularly, e.g. calibration table updates.

solar

Solar means relating to reflected solar radiation only.

thermal

Thermal means relating to emitted thermal radiation only

short wave

Short wave means radiation from the wavelengths below a cutoff wavelength of approximatively 4 micron.

long wave

Long wave means radiation from the wavelengths above a cutoff wavelength of approximatively 4 micron.

total wave

Total wave radiation means combined short wave and long wave radiation.

unfiltered

Unfiltered refers to a spectral integral of either solar or thermal radiation without spectral attenuation.

'GERB filtered' or simply 'filtered'

GERB filtered refers to a spectral integral, either over the short wave or over the long wave or over the total wave spectral interval, of radiation multiplied (filtered) with the GERB spectral response.

imager resolution

Imager resolution refers to SEVIRI pixel resolution

high resolution

HIgh resolution refers to 3 x 3 SEVIRI pixel resolution, this the highest resolution at which fluxes are derived

GERB resolution

GERB resolution refers to GERB footprint resolution

SEVIRI based products

SEVIRI based products refers to broadband radiance or flux estimates estimated from SEVIRI narrowband pixel data only.

level 2 fluxes

Level 2 fluxes refers to fluxes derived from combined GERB and SEVIRI data.

spectral unfilter factor

Ratio between the unfiltered broadband radiance and the filtered broad band radiance.

angular conversion factor

Ratio between the broad band flux and the unfiltered broad band radiance.

correction factor of SEVIRI by GERB

Ratio of the filtered broad band radiance measured by GERB and the same quantity model based on SEVIRI narrow band radiances.

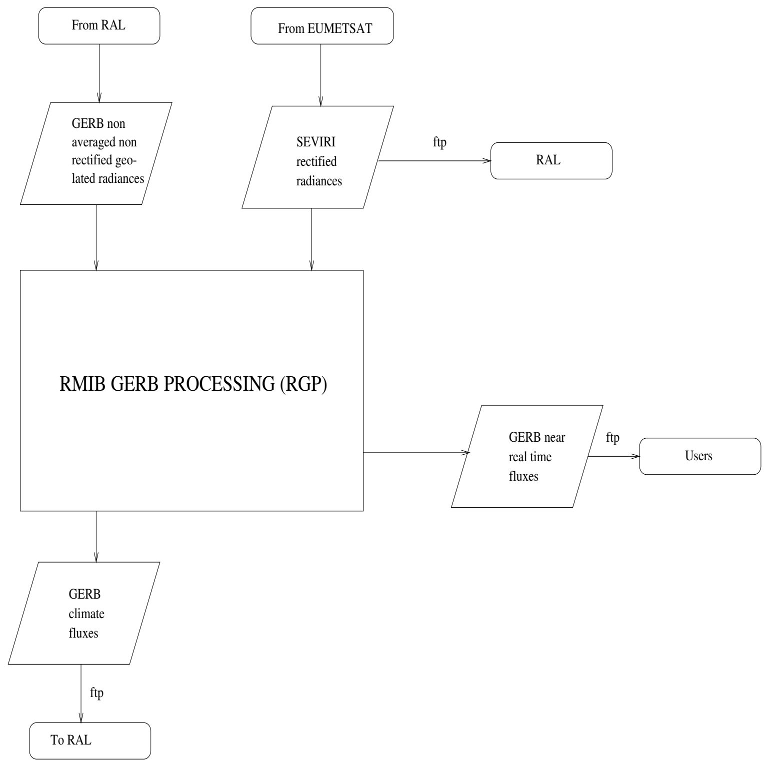

Figure 1: Overview of RMIB processing

3 Processing Overview

There are two sets of data coming from two different ways (see figure 1): the GERB data from RAL by Internet or ISDN connection and the SEVIRI data furnished by the electronic laboratory of the RMIB which receives the data from EUMETSAT. There are two products furnished to users: the climate data, which is a 24 hours product, and the near real time data, which is a 15 minutes product. The last product must be furnished every 15 minutes in order to be used in weather forecasting programs. The data send to RAL is a monthly archive of all the products i.e. the 15 minutes product and the 24 hours product stored in a tape. The near real time data is furnished by Internet or ftp.

4 EUMETSAT Data Acquisition

12 images are received from EUMETSAT each 15 minutes. The total amount of data is about 233 MBytes/15' without compression. The expected compression is 2,5 without loss. This gives about 9 GBytes per day. The archiving of this amount of data will not be done.

The acquisition is done trough a dedicated hardware and software at the RMIB electronic loboratory (see [USM]). The material has to be purchased. The data is then available for the SEVIRI processing. The images are in PIF (Picture In File) format.

5 SEVIRI Header distribution to RAL

The header of SEVIRI images will be made available through ftp to the RAL (see [ICD]). The RAL is responsible of downloading the data.

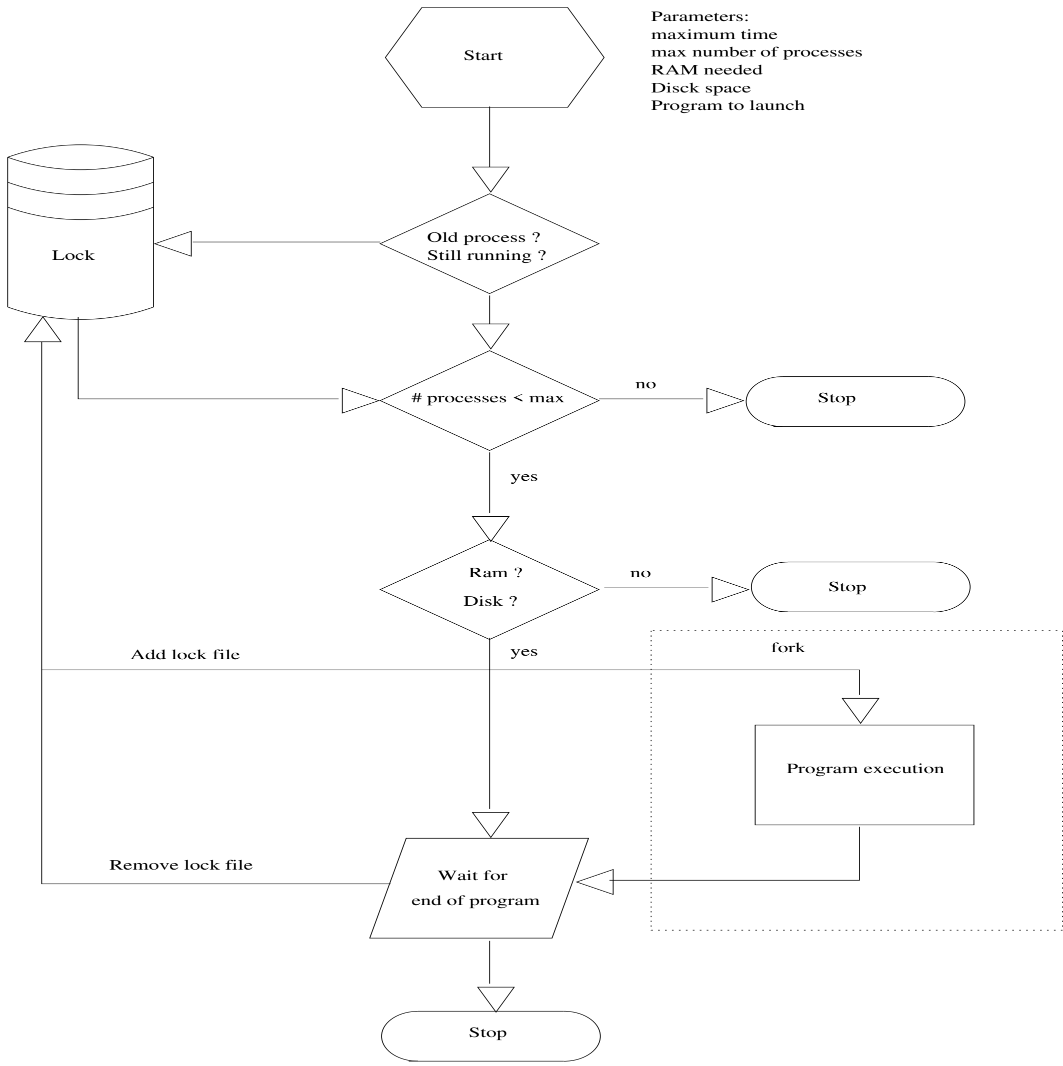

Figure 2: Launcher program

6 RAL Data Acquisition

The RAL will provide the L1.5 GERB data. The data is made available trough ftp. RMIB is responsible of downloading the data. The interface is described in [ICD].

The data is then transformed in internal format for further processing. The data is written on a specific directory (TBD).

7 Resource Availability

All the data processing programs are launched by another program (the launcher - see figure 2) that tests resource availability and the number of processes (from the same program) that are running. If the resources are sufficient, a file in the directory “lock” is created with the name of the program to launch and the PID as suffix. So, scanning this directory the launcher knows how many programs are running. Once the program is finished, the lock file is removed. If the program fails, the lock file is not removed and all the available resources are not used. To avoid this problem, when started the launcher tests if the processes associated to the files in the lock directory are still running. If not, the lock file is changed to “s_kill_<lock name>”. A program can fail but still be in the process table (e.g. endless loop). To avoid this, if the process running time is higher than a maximum, the process is killed and the lock file name is changed to “kill_<lock name>”. The creation time of the lock file is used as reference for the process start time.

The available disk space and RAM memory can also be tested to be sure that the program will not fail due to a lack of available ressources. This is really important since some operating systems (Linux or AIX[1] for example) can crash.

8 Log and error tracking

Two kinds of log files can be generated during the process execution:

1. process log files
2. product log files

The process log files are temporary files used to store information about the process. The product log file has the following information:

1. Name and PID of the process
2. Parameters of the program
3. Time of processing (Start time and end time of processing)
4. Inputs of the process
5. Version of the program (date of compilation)

If an error is trapped, the information is also written on the process log file. The name of the log file is the name of the process with the PID added as suffix. This file is removed if the process is successful. Scanning the lock directory and the log directory allows to know what processes have failed. If a log file exists without its lock file, something wrong append.

The product log file is generated with the name of each result file and the suffix “.err” on the result file directory. Once the input has been successfully generated, the suffix will be changed to “.log”.

The processes are launched with a cron. If an error is trapped, a warning is sent by mail (this is automatically done using cron and directing the warning to the standard output).

9 Operator Tasks

If an error occurs, the error file is verified by an operator. If the data must be re-processed, removing the error file from the log directory and the output directory and feeding the input directories with the appropiate files is sufficient to re-run the process. Since not all the data will be archived (SEVIRI data), re-processing is not guaranted.

10 Archive Policy

The program that creates new data is responsible to make a copy of what must be archived in a specific directory. The archiving is done by another program.

The archiving hardware and software is TBD. It depends on the total amount of data that will be archived.

[1] In order to avoid useless allocation from fork, the data is only allocated when used. The amount of allocated memory can be higher than the physical available memory. If all the allocated memory is really used the system can crash. This feature can be set/unset on AIX but it is the default on Linux and no other option is available.

11 Data Sizes

Type	Resolution	Format	Size
SEVIRI	HR	<f>	3600*3600*4 = 52 MBytes
GERB	LR	<f>	256*256*4 = 260 KBytes
SEVIRI bands	HR	<f>	120*3600*4 = 1,7 MBytes
SEVIRI/3	MR	<f>	1200*1200*4 = 5,8 MBytes

12 RGP Processing

Each program is launched every minute from a cron (see 7). To avoid computer resource availability check at the same time, every program is delayed by some seconds. Then, it checks if the available data has already been processed (e.g. the existence of a log file of the product to be created). If not, the program is launched on the new data. The results are then stored in a buffer directory for further processing, on the archive directory for archiving, on the WWW directory for Internet availability and the log file is copied on the log files directory.

Each program is responsible to remove from buffer directory the data that is no more needed.

12.1 Data Transmission

12.1.1 SEVIRI data

Input from buffer:

1. -[2]

Output to buffers:

1. 10 SEVIRI images

Output to archive:

1. -

Output to WWW:

1. -

Output to log:

1. Log of processed files

Disk Resource for five days:

10 images * 4 per hour * 24 hours * 5 days = 4800 images => 250 GBytes (in float format) => 78 GBytes (in 10 bits format)

[2] Means nothing.

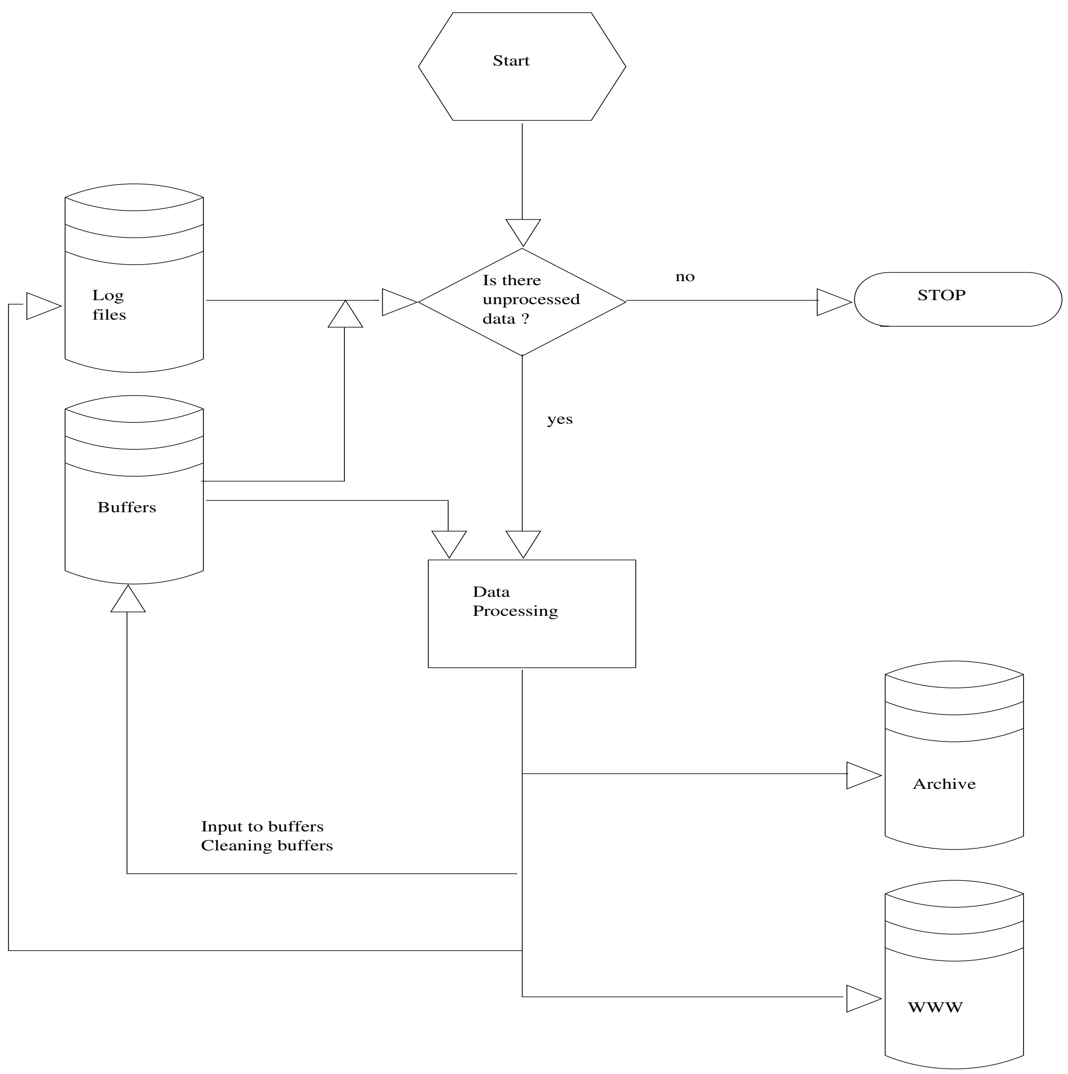

Figure 3: Processing Data Flow

12.1.2 GERB data

Input from buffer:

1. -

Output to buffers:

1. GERB images
2. Geolocation

Output to archive:

1. -

Output to WWW:

1. -

Output to log:

1. Log of processed files

Disk Resource for five days:
Negligeable

12.2 RGP SEVIRI Spectral Modelling

(see [SM])
Input from buffer:

1. Viewing angle images (SEVIRI/3 resolution)
2. Fitting coefficients
3. SEVIRI/3 images (3 for SW and 7 for LW)

Output to buffers:

1. 6 (or 4 TBD) SEVIRI/3 broad-band images

Output to archive:

1. 10 SEVIRI/3 images

Output to WWW:

1. -

Output to log:

1. Log of processed files

Disk Resource for five days:
6 SEVIRI/3 images/15’ minute => 17 GBytes/5 days ($\sum = 95$ Gbytes)

12.3 RGP SEVIRI Scene Identification

(see [ScId])

12.3.1 Reflectance Images

Input from buffer:

1. 3 SEVIRI images (visible bands)

Output to buffers:

1. 3 SEVIRI images (visible bands)

Output to archive:

1. -

Output to WWW:

1. -

Output to log:

1. Log of processed files

Disk Resource for five days:
3 images * 4 per hour * 13 hours * 5 days = 780 images => 40 GBytes ($\sum$ = 135 Gbytes)

12.3.2 Clear Sky (Weekly Processing)

Input from buffer:

1. reflectance images (3 bands*14 slots*60 days = ˜ 2500 SEVIRI resolution images in 2 bytes format, 39 images added per day during one week)

Output to buffers:

1. clear sky reflectance image (3 bands*14 slots * 4 interpolation = 168 SEVIRI resolution images in 2 bytes format)

Output to archive:

1. once per month without interpolation (3 bands*14 slots * 4 interpolation in 2 bytes = 1,1 GBytes)

Output to WWW:

1. -

Output to log:

1. Log of the processed data

Disk Resource for five days:
75 GBytes (about 2863 SEVIRI resolution images) (2500 + 5*39+168) ($\sum$ = 210 Gbytes)

12.3.3 Surface Identification (Weekly Processing)

Input from buffer:

1. Clear sky images

Output to buffers:

1. Pixel class (1 SEVIRI/3 image in byte)
2. NDVI (1 SEVIRI/3 image in byte)

Output to archive:

1. Pixel class (1 SEVIRI/3 image in byte)
2. NDVI (1 SEVIRI/3 image in byte)

Output to WWW:

1. -

Output to log:

1. Log of the processed data

Disk Resource for five days:
2 SEVIRI/3 image in byte format (negligeable)

12.3.4 Cloud Analysis

Input from buffer:

1. 3 reflectance image at SEVIRI resolution
2. 3 clear sky image
3. viewing angles (3 images at SEVIRI resolution)
4. Scene identification
5. 3 tables of overcast

Output to buffers:

1. cloud products (size: 6 SEVIRI/3 images)

Output to archive:

1. cloud products

Output to WWW:

1. -

Output to log:

1. Log of processed files

Disk Resource for five days:
6 images * 4 per hour * 24/2 hours * 5 days = 1440 images in byte format => 2 GBytes/5 days ($\sum = 137$ Gbytes)

12.4 RGP SEVIRI ADM

(see [ADM])

12.4.1 Thermal Flux

Input from buffer:

1. 7 SEVIRI/3 radiances images
2. viewing zenith angle

Output to buffers:

1. Thermal flux (1 SEVIRI/3 image in 2 bytes)

Output to archive:

1. Thermal flux

Output to WWW:

1. -

Output to log:

1. Log of processed file

Disk Resource for five days:
4 slots * 24 hours * 5 days = 480 SEVIRI/3 images in 2 bytes= 1,2 GBytes ($\sum = 138$ Gbytes)

12.4.2 Solar Flux

Input from buffer:

1. Solar BB SEVIRI image
2. viewing angles
3. angular models
4. surface identification

Output to buffers:

1. Solar Flux (1 SEVIRI/3 image)

Output to archive:

1. Solar Flux (1 SEVIRI/3 image)

Output to WWW:

1. -

Output to log:

1. Log of processed files

Disk Resource for five days:
4 slots * 13 hours * 5 days = 260 mid-resolution images in 2 bytes = 0,8 GBytes ($\sum = 139$ Gbytes)

12.5 RGP GERB Processing

(see [GP])

Input from buffer:

1. Filtered SEVIRI radiances and unfiltered SEVIRI fluxes (or radiances TBD) (4 SEVIRI/3 resolution images)
2. GERB Level 1.5 data
3. GERB Geolocation, A values, α
4. GERB detectors status
5. GERB LW and SW PSF

Output to buffer:

1. Level 2 fluxes (or radiances TBD) at GERB resolution (2 GERB resolution images)
2. SEVIRI images for each GERB acquisition time (256 images of SEVIRI image height and as width the PSF width multiply by $\sqrt{2}$[3] plus a margin- so the maximum width is about 120)
3. Filtered SEVIRI at GERB resolution
4. Correction factor

Output to archive:

1. Level 2 fluxes (or radiances TBD) at GERB resolution (2 GERB resolution images)

Output to WWW:

1. Level 2 fluxes (or radiances TBD) at GERB resolution (2 GERB resolution images)

Output to log:

1. Log of level 2.0 product processed

Disk Resource for five days:

256 SEVIRI/3 bands images/5' minutes = 48 MBytes/5' = 70 GBytes/5 days ($\sum = 209$ Gbytes)

12.6 RGP Resolution Enhancement

(see [RE])

Input from buffer:

1. level 2 fluxes (or radiances TBD) at GERB resolution (2 GERB resolution images)
2. SEVIRI for each GERB acquisition time (256 images of about PSF width and SEVIRI/3 image height)
3. filtered SEVIRI at GERB resolution
4. filtered SEVIRI radiances and unfiltered SEVIRI fluxes (or radiances TBD) (4 SEVIRI resolution images)

[3] For a maximal rotation of 90°.

5. GERB Geolocation
6. GERB detectors status
7. GERB LW and SW PSF
8. Correction factor (GERB resolution)

Output to buffers:

1. Correction factor (SEVIRI resolution)

Output to archive:

1. level 2 fluxes (or radiances TBD) at SEVIRI resolution (2 SEVIRI resolution images)

Output to WWW:

1. level 2 fluxes (or radiances TBD) at SEVIRI resolution (2 GERB resolution images)

Output to log:

1. Log of level 2.0 product processed

Disk Resource for five days:
2 SEVIRI/3 images/5' minutes = 17 GBytes/5 days ($\sum = 226$ Gbytes)

13 General structure

All the information necessary to run the complete process is stored in a directory. Two directories, the "batslib" and the "gerblib" directory, contains the process sources. The other directories are data directories:

- Archive: data stored in this directory will be archived
- Buffers: all the intermediate data
- Data: information for the processing that is seldom changed (detector PSF,...)
- WeeklyData: data stored for the weekly processing
- lock: directory where the lock files are stored
- log: directory where the process log files are stored

13.1 Standard library structure

All the processing is based on the use of a scientific library. The structure of this library is shown in figure 4.

The directories content at first level is:

- bin: all the executables
- depend: the dependency files
- include: include files
- lib: the library
- pack: additionnal libraries needed

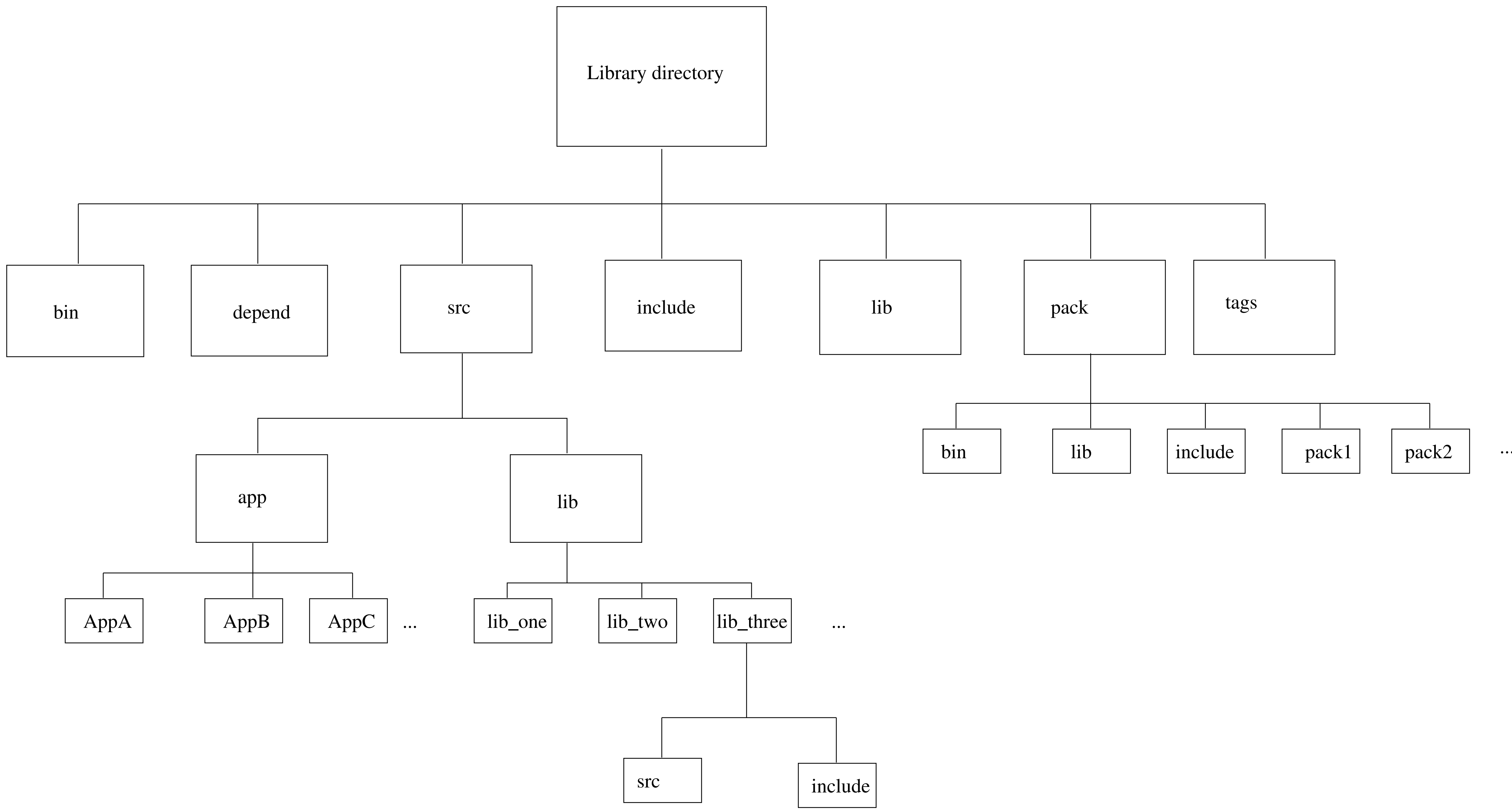

Figure 4: Standard library structure

- tags: information in order to locate easily library information
- src: all the information to build the library

The src directory has two directories:

- app: executables using the library
- lib: the library sources that are for convenience in different logical directories (one for image processing, one for matrix processing, ...)

The pack directories, that contains additionnal libraries as jpeg, hdf, gzip lib,... necessary to compile some
parts of the library, have a bin, include and lib directories where respectively, all the exectables, include files and libraries are
stored.

13.1.1 The BATS library

The BATS library is an image/signal processing oriented library. It follows the standard library structure.

13.1.2 The GERB library

Library containing all that is GERB specific. It follows the standard library structure.

13.2 The Archive directory

There are 2 sub-directories:

- Meteosat : data to archive from Meteosat 7 processing output and one day archive for debug purpose
- SEVIRI: data to archive from SEVIRI processing output and one day archive for debug purpose

The “one day” archive keeps some intermediate data for debug purpose. When new data is generated for a specific hour, it replaces the old data that comes from the previous day at the same hour. This data is not archived in tape.

An additionnal directory (logs) for each of the two archive sub-directories is used to store the logs of intermediate files that are not archived but needed to retrieve information about the story of the archived files.

14 ROLSS

The data is stored on disk for 40 days. The amount of data is about 180 GBytes (Solar and thermal fluxes at high resolution in floating point format) without compression and 45 GBytes with compression.

References

[USM] "Definition of User Stations for MSG", H. Wulf, MSG/VS/TR/004

[ICD] "GGSPS-IRMB Interface Control Document", D. Corney, MSG-RAL-GE-IF-0011

[ScId] "RMIB GERB Processing: Scene Identification", N. Clerbaux and S. Dewitte, MSG-RMIB-GE-TN-0007

[ADM] "RMIB GERB Processing: Angular Dependency Models", N. Clerbaux and S. Dewitte, MSG-RMIB-GE-TN-0008

[SM] "RMIB GERB Processing: Spectral Modelling", N. Clerbaux and S. Dewitte, MSG-RMIB-GE-TN-0005

[GP] "RMIB GERB Processing: GERB Processing", L. Gonzalez and A. Hermans, MSG-RMIB-GE-TN-0006

[RE] "RMIB GERB Processing: Resolution Enhancement", S. Dewitte and A. Hermans, MSG-RMIB-GE-TN-0003