



ESA Sea Level CCI

## D3.1 System Prototype Description

Reference: CLS-DOS-NT-11-282

Nomenclature: SLCCI-SPD-016

Issue: 1.2

Date: Nov. 21, 12





Chronology Issues:			
Issue:	Date:	Reason for change:	Author
1.0	20/1/2012	First Issue	Anne Delamarche
1.1	17/7/2012	Re-issue following ESA review	Anne Delamarche
1.2	21/11/2012	Re-issue following complementary ESA review	Frédéric Briol Anne Delamarche Yannice Faugere

People involved in this issue:		
Written by (*):	A Delamarche (CLS)	Date + Initials:( visa or ref)
Checked by (*):	G Timms (Logica)	Date + Initial:( visa ou ref)
Approved by (*):	G Larnicol (CLS)	Date + Initial:( visa ou ref)
Application authorized by (*):	ESA	Date + Initial:( visa ou ref)

*\*In the opposite box: Last and First name of the person + company if different from CLS*

Index Sheet:	
Context:	Baghera tool, project ACT-OCEAN
Keywords:	Oceanography, sea level
Hyperlink:	

Distribution:		
Company	Means of distribution	Names
ESA	Email	Jerome Benveniste, Bruno Manuel Lucas, Salvatore Dinardo
CLS	Email	Michael Ablain, Yannice Faugere, Gilles Larnicol
Logica	Email	Gary Timms, Ed Pechorro
SLCCI Project FTP SLCCI Website		<a href="ftp.esa-sealevel-cci.org">ftp.esa-sealevel-cci.org</a> <a href="http://www.esa-sealevel-cci.org/documents">http://www.esa-sealevel-cci.org/documents</a>



List of tables and figures

List of tables:

No table of figures entries found.

List of figures:

Figure 1 - SLCCI Prototype Macro Functionalities .....2

Figure 2 - Input to the Altimetric Database .....9

Figure 3 - Output from the Altimetric Database..... 10

Figure 4 - Input to the Crossovers Database..... 11

Figure 5 - Output from the Crossovers Database ..... 12

Figure 6 - Input to the Statistics Database ..... 13

Figure 7 - Output from the Statistics Database ..... 14

Figure 8 - Input to the Temporary Product Files Database..... 15

Figure 9 - Output to the Temporary Product Files Database ..... 16

List of items to be confirmed or to be defined

Lists of TBC:

No table of contents entries found.

Lists of TBD:

No table of contents entries found.



#### Applicable documents

**AD 1** Sea level CCI project Management Plan  
CLS-DOS-NT-10-013

**AD 2** ESA Climate Change Initiative Phase 1: Scientific User consultation and Detailed Specification - Statement of Work (SOW)  
EOP-SEP/SOW/0031-09/SP issue 1.4 revision 1 - 09/11/2009

**AD 3** ESA CLIMATE CHANGE INITIATIVE - PHASE 1 ECV Sea Level CORALS - Technical Proposal  
CLS-DOS-PR-10-001 issue 2 revision 0 - 23/06/2010

#### Reference documents

**RD 1** Manuel du processus Documentation  
CLS-DOC

**RD 2** Preliminary System Analysis Document (PSAD), Issue 1.0, Reference CLS-DOS-NT-10-297, Nomenclature SLCCI-PSAD-006

**RD 3** System Requirements Document (SRD), Issue 1.1, Reference CLS-DOS-NT-11-080, Nomenclature SLCCI-SRD-012

**RD 4** System Specification Document V0 (SSD V0), Issue 1.0, Reference CLS-DOS-NT-11-159, Nomenclature SLCCI-SSD-016



List of Contents

- 1. Introduction..... 1**
- 2. Prototype ECV processing system ..... 1**
  - 2.1. Main architecture components..... 1**
  - 2.2. Performed processes ..... 3**
    - 2.2.1. Acquire Data .....3**
      - 2.2.1.1. Passive data delivery or FTP retrieval.....3
      - 2.2.1.2. Store the data acquired in a database .....3
    - 2.2.2. Pre-process data .....3**
      - 2.2.2.1. Apply consistent correction, ancillary data, models, references... ..3
    - 2.2.3. Perform input checks and Quality Control .....4**
      - 2.2.3.1. Edit raw data .....4
      - 2.2.3.2. Apply Specific algorithms (iono filtering, radiometer extrapolation...) .....4
      - 2.2.3.3. Compute mono and multi-missions crossovers .....4
      - 2.2.3.4. Validate crossovers .....4
    - 2.2.4. Inter calibrate & Unify.....4**
      - 2.2.4.1. Compute mono-satellite orbit error (OE) reduction.....4
      - 2.2.4.2. Compute multi-satellite OE reduction using a reference mission.....5
      - 2.2.4.3. Compute Long Wavelength Error (LWE) reduction with Optimal Interpolation (OI) 5
    - 2.2.5. Generate along track product.....5**
      - 2.2.5.1. Generate along track SLA product corrected by LWE and OE .....5
      - 2.2.5.2. Validate the SLA product with the previous one.....5
    - 2.2.6. Generate merged product .....5**
      - 2.2.6.1. Generate gridded SLA product (Map of SLA) .....5
    - 2.2.7. Perform output check and quality control .....5**
      - 2.2.7.1. Compute internal data quality control .....5
      - 2.2.7.2. Compute product quality control .....6
      - 2.2.7.3. Follow up data and products on long term .....6
    - 2.2.8. Do measures and built indicator.....6**
      - 2.2.8.1. Build scientific metric.....6
    - 2.2.9. Store the built product datasets.....6**
      - 2.2.9.1. Delivery product on FTP server.....6
    - 2.2.10. Archive / retrieve a “Product Dataset” .....6**
      - 2.2.10.1. Archive / retrieve .....6
  - 2.3. Interfaces ..... 7**
    - 2.3.1. External Inputs .....7**
      - 2.3.1.1. Level 2 altimetry products .....7
      - 2.3.1.2. Ancillary data.....7

Proprietary information: no part of this document may be reproduced divulged or used in any form without prior permission from CLS.



2.3.2. External Outputs.....8

2.3.3. Internal Inputs/Outputs .....8

2.3.4. Dataflows .....9

3. System characteristics..... 17

3.1. Main System characteristics ..... 17

3.2. Volumetrics ..... 17

3.3. Altimetric database..... 18

3.4. Main external libraries..... 18

Appendix A - List of acronyms..... 20



## 1. Introduction

This document provides a high level description of the end-to-end Prototype Production System for the generation of the required ECV data products.

It provides a description of:

- The components, functions, interfaces and modules that form the prototype ECV processing system,
- The Inputs/Outputs dependencies,
- The data flow between the processor components,
- The data throughput,
- The processing demand in term of CPU load,
- The processing volumes,
- The times resources.

The Prototype Production system relies on the DUACS system.

The DUACS system is made of two components: the Near Real Time (NRT) component, and the Delayed Time (DT) component. The production of the ECV is taken into account by the DT component of DUACS.

## 2. Prototype ECV processing system

This chapter gives an overview of the components, functions, interfaces and modules that form the prototype ECV processing system.

### 2.1. Main architecture components

This chapter presents an overview of the different processing steps activated to produce the SLCCI products.

Figure 1 provides an overview of the DUACS system (generation of the ECV), where the processing sequences can be divided into the following key steps:

- Acquire Data
- Pre-process data
- Perform input checks and Quality Control
- Generate Products
- Performs output checks and Quality Control
- Do measures and built indicators

The product generation is divided into three sub-steps:

- Inter calibrate & Unify
- Generate along track product
- Generate merged product

Each step has its sequence of algorithms that can be activated to generate the ECV for the SLCCI.

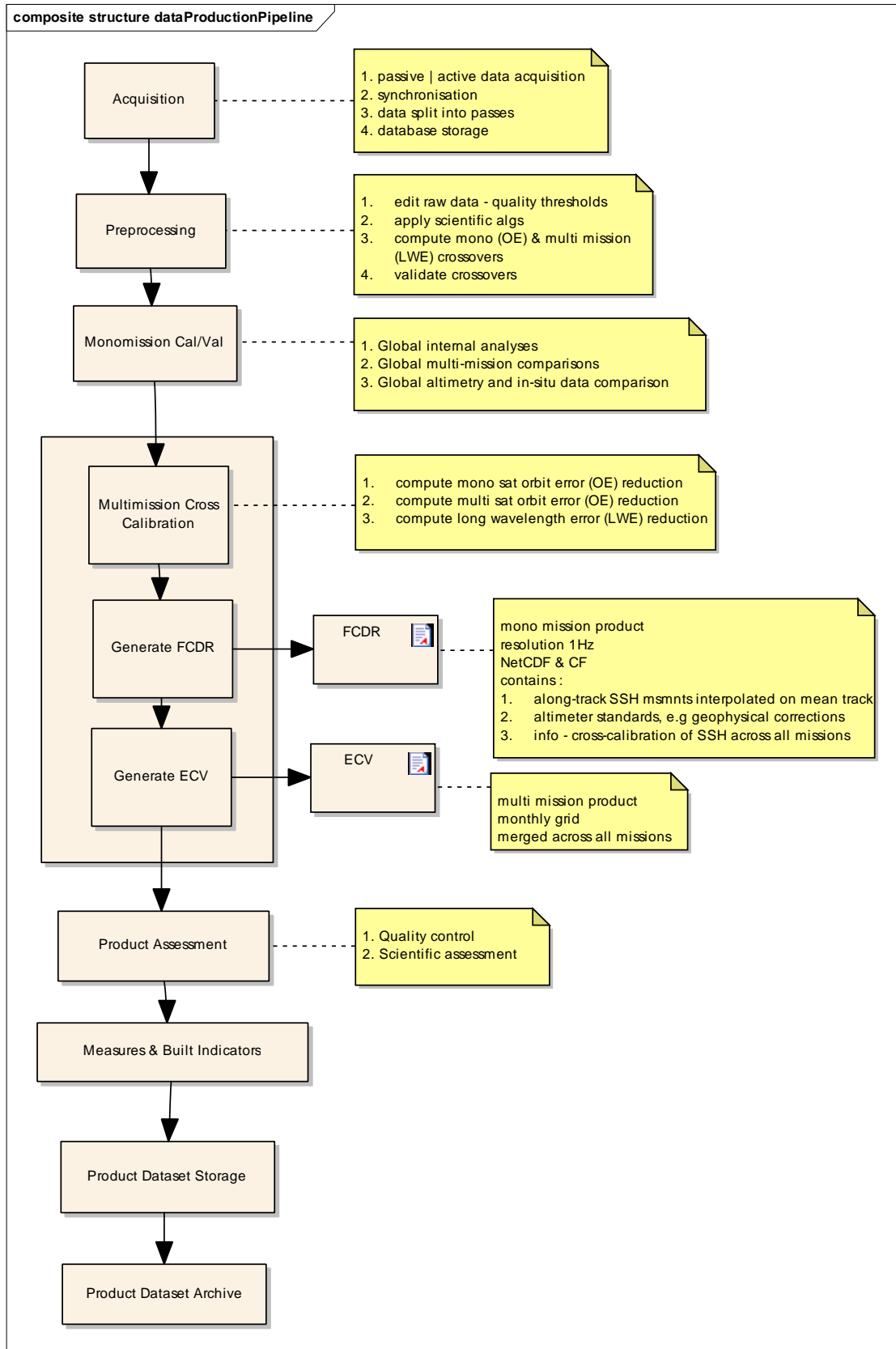


Figure 1 - SLCCI Prototype Macro Functionalities





## 2.2. Performed processes

---

This chapter describes each of the main components of the DUACS system, their relationships and the corresponding internal interfaces.

We also describe the sequence of algorithms that are activated in the frame of the SLCCI (only DT products).

### 2.2.1. Acquire Data

This component has two main functions:

- Acquisition
- Synchronization of dataflows.

The acquisition function allows for the downloading of altimetry data or ancillary data disseminated by the major space agencies such as CNES, ESA, NOAA, NASA, etc...

The synchronization function is synchronizing altimetry data with all ancillary data needed to process the data by the SLCCI.

#### 2.2.1.1. Passive data delivery or FTP retrieval

The acquisition function can be realized by two different methods:

- Passive
- Active

The passive method consists of scanning a given directory to determine if it contains new data to process. If so, the new data is moved to a working directory to be processed. In this case, the input data will be provided by a third party.

The active method follows the same steps as described in the passive method, but on a remote server. If the acquisition function finds a new file to be processed, this file will be downloaded into the working directory and then acquired. The interrogation of the remote server can be performed by different Internet Protocols: FTP, SSH, etc.

#### 2.2.1.2. Store the data acquired in a database

This operation allows storing in the "altimetric" database. This database is an internal data base to be used for all other operations of production of SLCCI.

### 2.2.2. Pre-process data

#### 2.2.2.1. Apply consistent correction, ancillary data, models, references...

The homogenization process allows applying the most recent corrections, the models and recommended standards for altimetry products to process the different missions in homogeneous ways in the SLCCI.



## 2.2.3. Perform input checks and Quality Control

### 2.2.3.1. Edit raw data

The validation of raw data is a critical process to ensure that the SLCCI uses only the most accurate altimetric data. The validation rejects a small percentage of height measurements. This is a necessary step to ensure the quality of the SLCCI products. The quality control is based on verification of the value of flags (rain, land / sea, etc.), quality thresholds of different parameters, and also on complex algorithms based on the detection of false objects.

Valid and selected altimetry measurements will be tagged with a field containing a flag. This field will contain a zero value in the case of the measure being valid and a nonzero value in the case of the measure being invalid. In the case of an invalid measure, the value of the flag will indicate the reason why this measure has been eliminated. Later in the process of generating products, only valid measures will be processed by the system.

### 2.2.3.2. Apply Specific algorithms (iono filtering, radiometer extrapolation...)

The valid measures selected are processed for applying specific algorithms calculating correction and bias necessary to calculate the Sea Level Anomaly (SLA) of the different missions processed by the system.

### 2.2.3.3. Compute mono and multi-missions crossovers

Crossovers are the intersection of two satellite ground tracks. Each crossover is composed of two measurements; one from an ascending pass and one from a descending pass. There are two types of crossovers: mono-missions crossover (intersection of arcs from a single sensor) and multi-missions crossovers (intersection of arcs from multiple sensors). These datasets are used to estimate the performance of each mission, or for the calculation of the orbit error.

### 2.2.3.4. Validate crossovers

Some crossovers are located in areas of high ocean variability, near the coast or are affected by errors of calculation or interpolation. Therefore this processing is to select only crossovers which are trustworthy. Moreover, this step gives the performance of the different altimetry missions processed by the system before bias correction or long wavelength error minimization and orbit error minimization.

## 2.2.4. Inter calibrate & Unify

### 2.2.4.1. Compute mono-satellite orbit error (OE) reduction

This algorithm estimates the orbit error deduced from the differences in Sea Surface Height (SSH) at mono-mission crossovers. This assessment is made on the reference mission of the system.

This process is based on the sinusoidal regression model on the observations derived from SSH differences on crossovers at 10 days.

The calculated orbit error is applied for the processed period to all data from the reference mission (altimetric database, crossovers mono and multi-mission).



#### 2.2.4.2. Compute multi-satellite OE reduction using a reference mission

The algorithm is based on the assumption that there is an altimetry mission more accurate than other available missions. This mission is considered the reference mission of the system. The orbit error is deduced from differences at the crossovers of multi-missions.

Note that this process is based on the minimization by the method of least squares splines on observations derived from SSH differences at crossovers in 10 days.

The calculated multi-mission orbit error is applied for the processed period on all the secondary missions of the system (altimetric database, crossovers).

#### 2.2.4.3. Compute Long Wavelength Error (LWE) reduction with Optimal Interpolation (OI)

LWE is mostly due to high frequency ocean signals (residual tidal or barotropic errors). The optimal interpolation (OI) is used for LWE estimation. The aim is to estimate LWE values on along track positions and at a fixed time tag, given along track SLA observations are unequally distributed.

#### 2.2.5. Generate along track product

##### 2.2.5.1. Generate along track SLA product corrected by LWE and OE

This step deals with SLA (anomaly compared to a mean profile) calculation and validation. The along track SSH measurements are interpolated on a mean track and the temporal mean of the along track SSH (the so called mean profile) is removed in order to work with anomalies of sea level.

A sophisticated process of filtering is then used, in order to remove the outliers that are still present in the data and which can be detected thanks to the removal of the mean profile.

##### 2.2.5.2. Validate the SLA product with the previous one

This along track SLA is validated by comparing the SLA measurement of the current production with the latest map of SLA (MSLA).

#### 2.2.6. Generate merged product

##### 2.2.6.1. Generate gridded SLA product (Map of SLA)

A mapping procedure using optimal interpolation with realistic correlation functions is applied to produce SLA maps at a given date. The procedure generates a combined map merging measurements from all available altimeter missions.

#### 2.2.7. Perform output check and quality control

##### 2.2.7.1. Compute internal data quality control

The purpose of this module is to calculate statistics on the internal data to evaluate the altimetric measures performance. This process will calculate for example, the crossovers statistics to estimate the performance of the altimetric missions. These calculations may also be performed on the correction of EO and LWE.



### 2.2.7.2. Compute product quality control

This module will calculate statistics on the various products of SLCCI: SLA, MSLA. For example, this process will calculate the signal oceanic variability rebuilt the kinetic energy of MSLA maps, etc.

### 2.2.7.3. Follow up data and products on long term

The different statistical results obtained in previous stages will be inserted in a database. This information will be retrieved from the database to see the change in these various statistical values over time.

## 2.2.8. Do measures and built indicator

### 2.2.8.1. Build scientific metric

The methodology is described at this address: <http://www.aviso.oceanobs.com/en/news/ocean-indicators/mean-sea-level/processing-corrections.html>

The numerical values calculated in the steps described in §2.2.7.1 and §2.2.7.2 will be expressed simply in order to create a scientific metric.

## 2.2.9. Store the built product datasets

### 2.2.9.1. Delivery product on FTP server

The SLCCI products are delivered on an authenticated FTP server.

This FTP server can be used by users wishing to download the Level-4 product.

## 2.2.10. Archive / retrieve a "Product Dataset"

### 2.2.10.1. Archive / retrieve

SL-TAC products are versioned and archived for 24 months. They can be retrieved upon request. The system automatically handles only the archiving process (on a monthly basis). The retrieval process is operator-activated.



## 2.3. Interfaces

---

Note: The Interfaces shall be refined and updated in this document after the 'expert' selection meeting in May 2012.

### 2.3.1. External Inputs

The external inputs are:

- Level 2 altimetry products
- Ancillary data

These inputs are stored in the Altimetry Database (see §2.3.3 Internal Inputs/Outputs)

#### 2.3.1.1. Level 2 altimetry products

Delayed Time SSALTO/DUACS products are generated from:

- Aviso GDR products for T/P, Jason-1, Jason-2 and Envisat (GDR-A: cycles 1 to 22 / GDR-B: cycles 23 to 85 / GDR-C: from cycle 86),
- NOAA GDR for GFO and from CERSAT (IFREMER) OPR for ERS-1 and ERS-2 (phases C (*1st 35-day repeat orbit period*), phase E and F (geodetic phases), phase G for ERS-1 (*last 35-day repeat orbit period*, tandem phase with ERS-2 ; phase A for ERS-2 (*1st 35-day repeat orbit period*, tandem phase with ERS-1)).

All GDR products are computed with a Precise Orbit Ephemeris (POE) and are delivered within 2 months depending on the mission. For several missions, an updated orbit is used:

- For ERS-1&2, the orbit used is DGME-04 provided by Delft Institute (<http://www.deos.tudelft.nl/>) until June 2003,
- For Topex/Poseidon, the orbit used is GSFC (std0809) for the whole mission,
- For Envisat, CNES POE of GDR-C standard is used for the whole mission,
- For the whole GFO mission, the orbit used is GSFC (std0809) and when not available, NASA POE is used.

#### 2.3.1.2. Ancillary data

Data for range calibration:

- Auxiliary instrumental calibration (ex: USO for ERS2 and Envisat)

Data for Orbit calculation:

- GSFC (ITRF2005, Grace last standards) for TP and GFO
- DGME-04 [Scharroo and Visser, 1998] for ERS1/2

Data for Ionospheric correction:

- GPS Ionosphere Maps
- Solar activity coefficient

Data for Dry troposphere and Invert barometer correction

- ECMWF meteo files



Data for high frequency correction

- MOG2D/AVISO

Static ancillary data to generate geophysical corrections (Tide GOT4.8 model, Non parametric SSB)

### 2.3.2. External Outputs

The external outputs are:

- ECV
- FCDR

### 2.3.3. Internal Inputs/Outputs

The following (internal) databases exist:

- Altimetric Database: This database contains all altimetry data exploited from Level 2 datasets. It contains the parameters needed to build the Sea Level content itself and all correction and environmental parameters needed for their processing and product generation.
- Crossovers Database: This database contains all the crossover-related information (position, date, content of two measurements (one from each arc). It is primarily used by the orbit error reduction step and for the monitoring of the input data quality.
- Statistics and follow-ups Database: This database is used for the long term information memory of mission performance and metrics.
- Temporary production files: These files are generated during the production process. For practical reasons, they remain online for two weeks or more (to avoid effort duplication if datasets must be reprocessed). They are not strictly necessary (no archive needed) as a production re-run would be able to generate this content again.



2.3.4. Dataflows

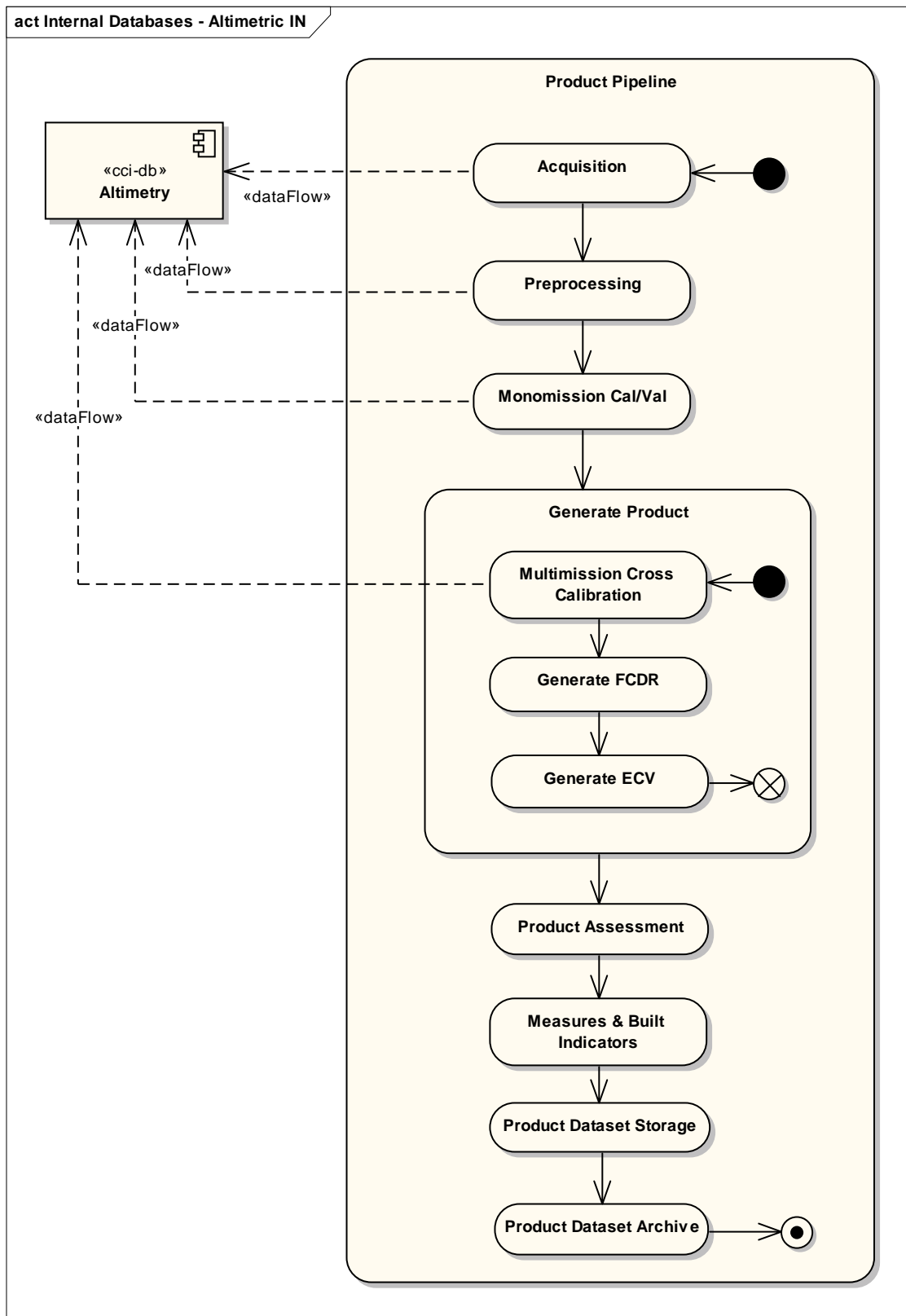


Figure 2 - Input to the Altimetric Database

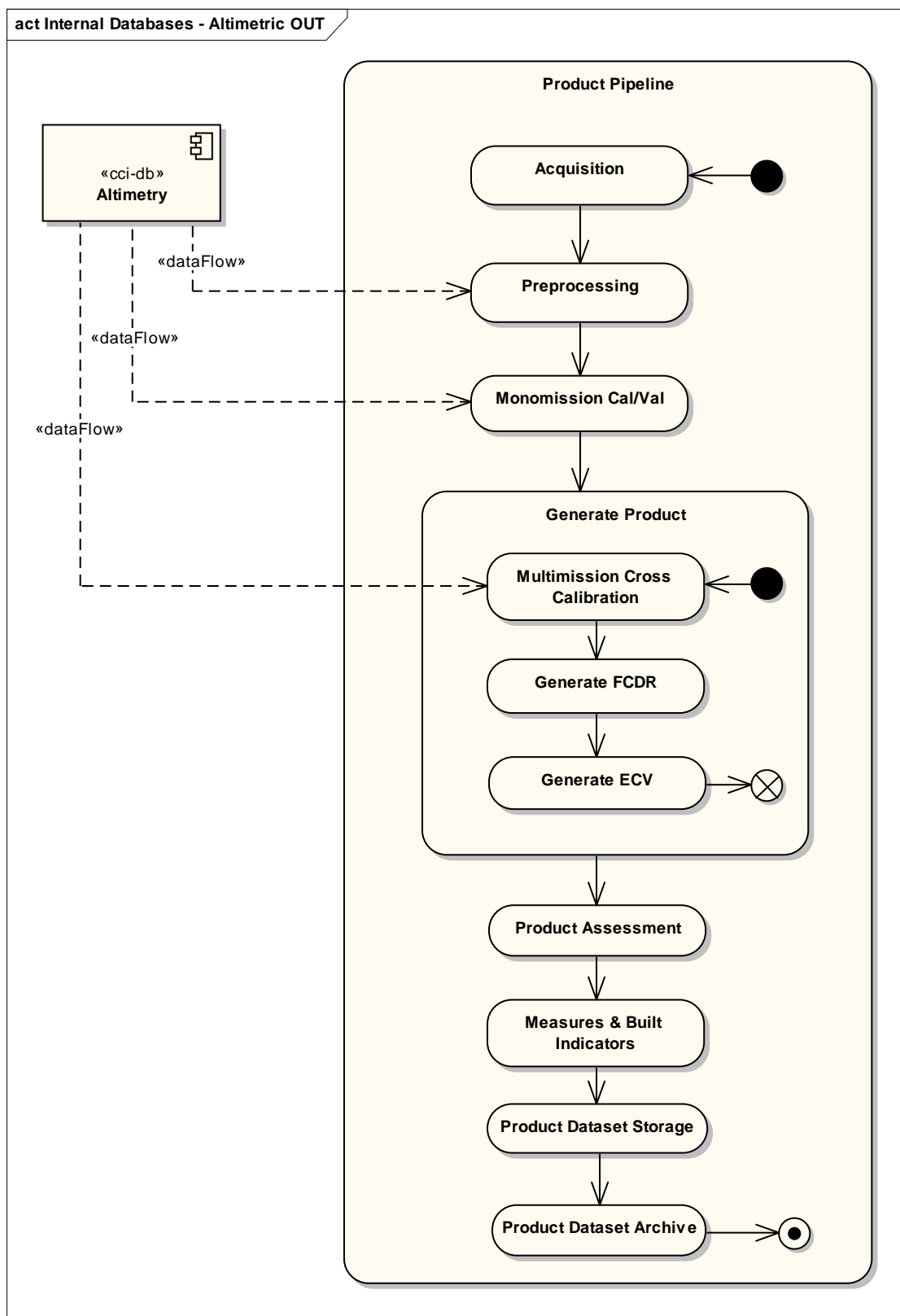


Figure 3 - Output from the Altimetric Database



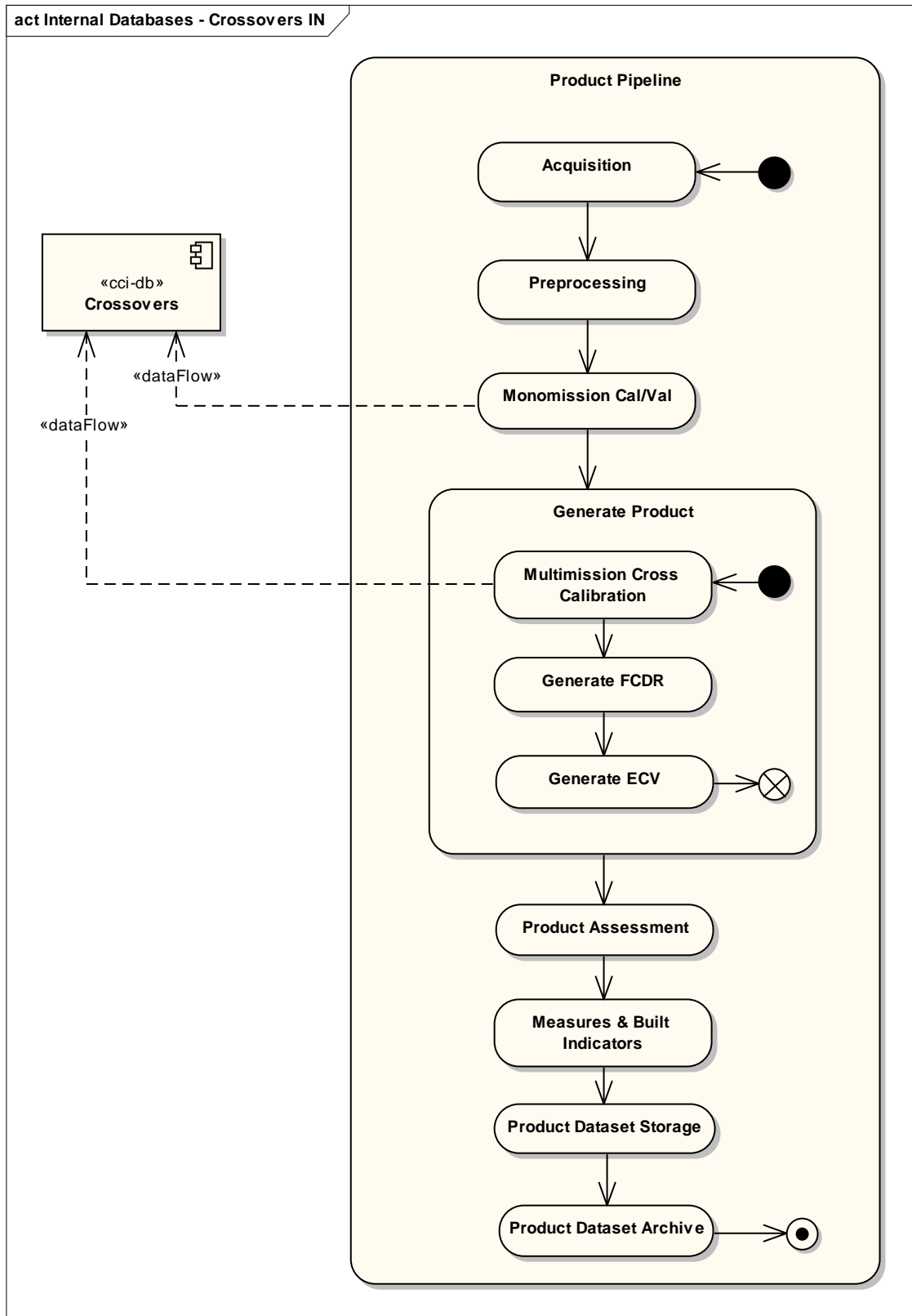


Figure 4 - Input to the Crossovers Database

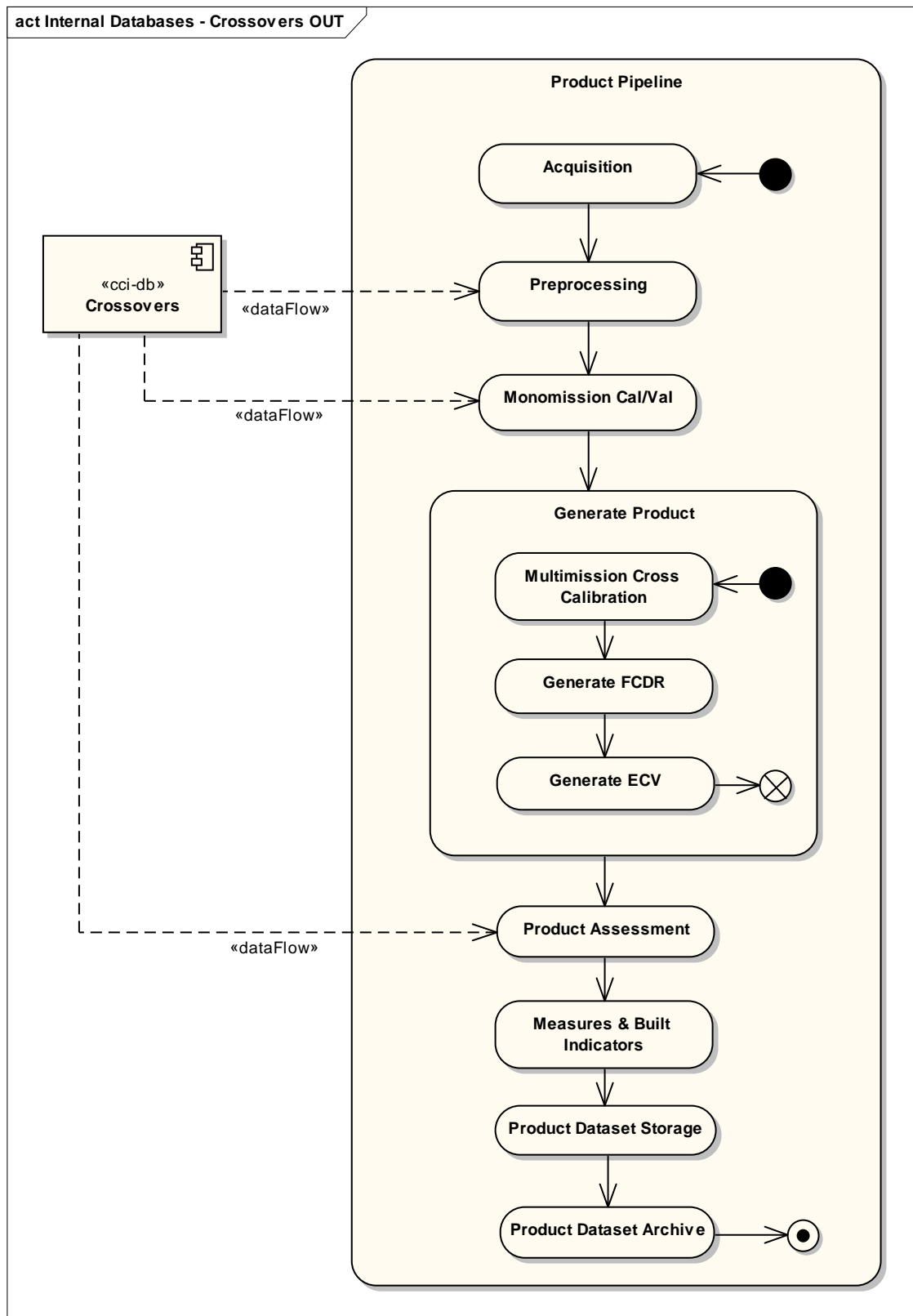


Figure 5 - Output from the Crossovers Database

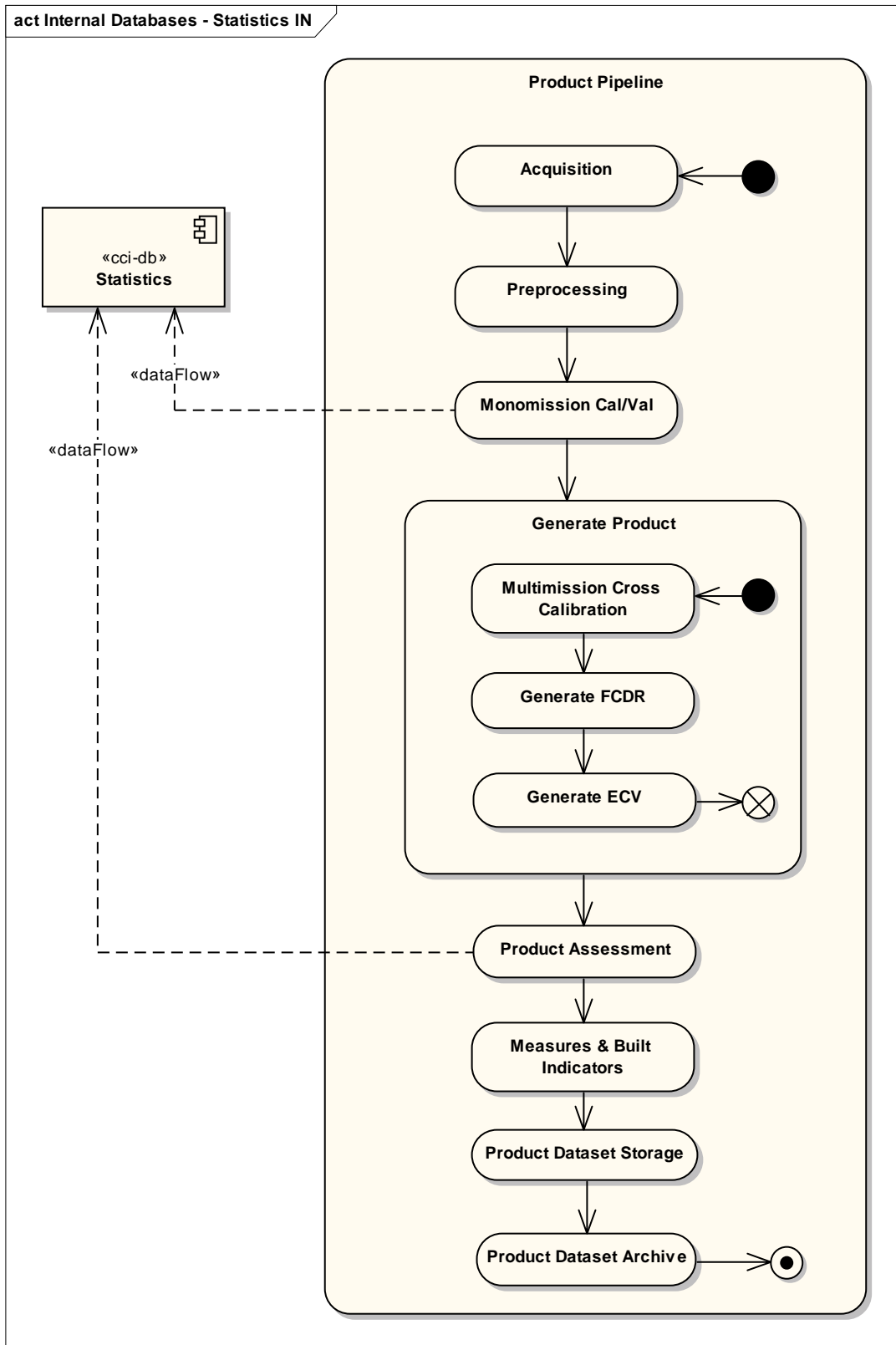


Figure 6 - Input to the Statistics Database

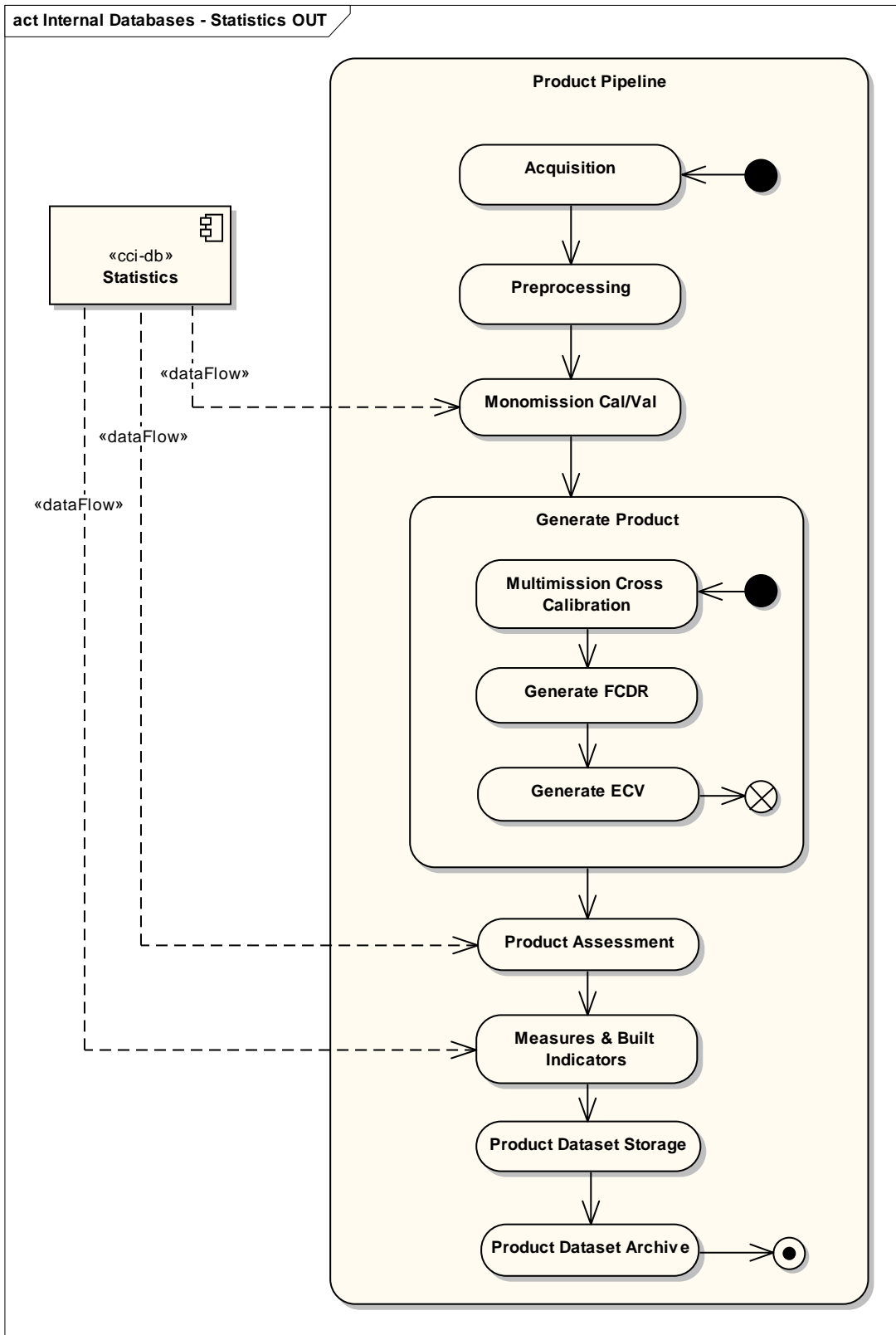


Figure 7 - Output from the Statistics Database

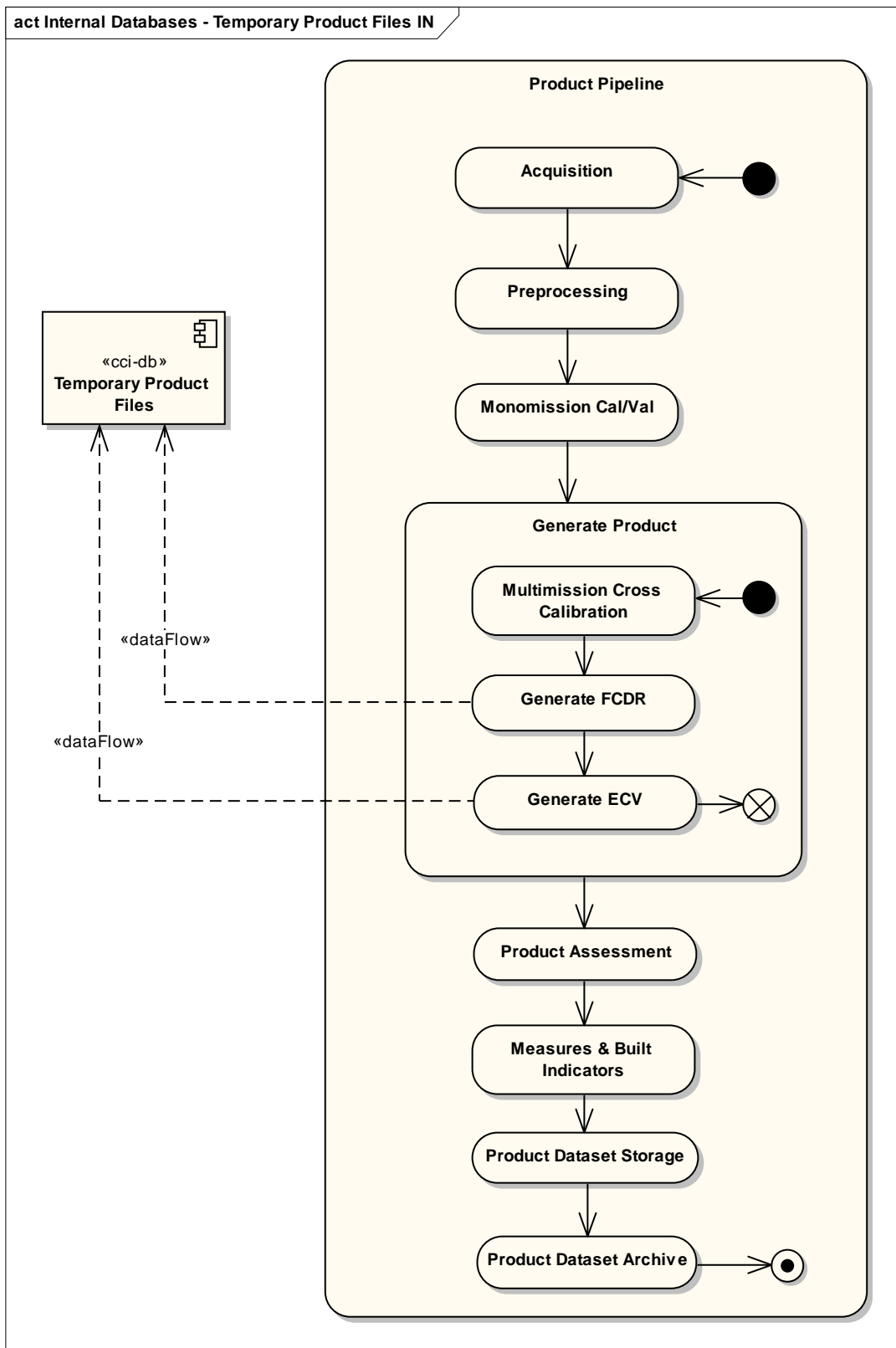


Figure 8 - Input to the Temporary Product Files Database

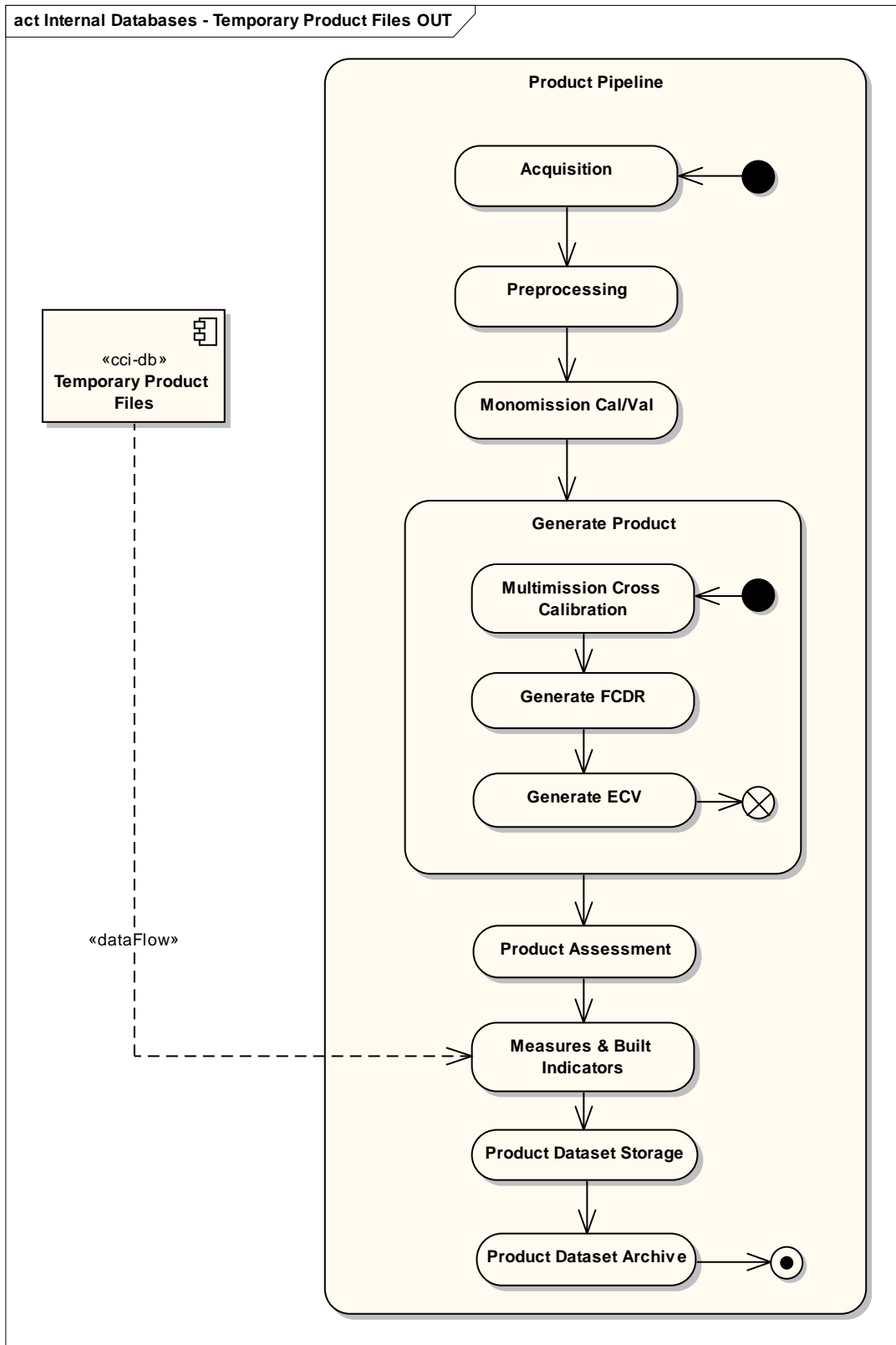


Figure 9 - Output to the Temporary Product Files Database



### 3. System characteristics

#### 3.1. Main System characteristics

The system is installed on the latest 32 bit version of Ubuntu server LT: 12.04.

The main languages used are:

- The C language for the core (algorithms, access to different databases, etc.)
- The languages Perl and Python for managing the tasks sequence of the prototype.

Versions of these programs provided by the operating system are:

- C language: gcc version 4.6.3
- Perl: 5.14.2
- Python: 2.7.3

All elements used by the prototype are "open source", but the prototype itself is the property of CLS.

System leans on a cluster. A node is a unit for performing the altimetry data processing for 7 days of measurements. To calculate the reanalysis, the computation is distributed over all available nodes in the cluster. The network connection is used to exchange information needed for the computation (read the altimetric database, writing the results, etc.). A simple network enterprise connection (100Mb) is sufficient to meet these requirements. The used cluster has 70 nodes, fulfilling the minimum requirements described in following table:

Ethernet connection	100 MB/s or more
Computer memory	4096 MB or more
Computer processor	Dual 2 GHz Intel Core Duo or more. SSLATO/DUACS will take advantage of multiple processors and processing cores, if available.
Hard disk space	300 GB or more

Table 1 - System requirements

#### 3.2. Volumetrics

The volumetrics have a bearing on the processing metrics, so we describe these here in the same location.

- Disk storage needed for altimetric database (including input and intermediate data): 300 Gbytes
- Disk storage needed for output data: 50 Gbytes Power Processing:
  - 3 hours to process 7 days of altimetric data (L3 and L4 products) on a standard computer (Core Duo 2Ghz, 4GB RAM)
  - So, for 20 years = 3 200 hours on one computer = 4 months

Proprietary information: no part of this document may be reproduced divulged or used in any form without prior permission from CLS.



- So for 20 years on the RAN production cluster (50 computers shared with others projects) = 2 months

### 3.3. Altimetric database

---

The disk storage needed for the altimetric database is 300 Gbytes (see §3.2).

The altimetric database is based on a binary format managing tables.

### 3.4. Main external libraries

---

The following table lists the packages to install on the system after installing a basic Ubuntu-Server distribution and describes the minimum requirement necessary for the execution of the prototype on a node:

Package name	Description
openssh-server	Secure shell (SSH) server, for secure access from remote machines.
lesstif2	OSF/Motif 2.1 implementation released under LGPL.
libncurses5-dev	Developer's libraries and docs for ncurses.
libboost1.42-dev	Boost C++ Libraries development files.
libmagick++-dev	Object-oriented C++ interface to ImageMagick - development files.
libnetcdf-dev	Development kit for NetCDF.
libboost-regex1.42-dev	regular expression library for C++.
libboost-python1.42-dev	Boost.Python Library development files.
libdate-manip-perl	Perl module for manipulating dates.
libxml-simple-perl	Perl module for reading and writing XML.
libdigest-sha1-perl	NIST SHA-1 message digest algorithm.
libxml-dumper-perl	Perl module for dumping Perl objects from/to XML.
libparams-validate-perl	Perl module to validate parameters to Perl method/function calls.
libreadonly-perl	Perl module for creating read-only scalars, arrays and hashes.
libmime-lite-perl	Perl module for convenient MIME message creating.
libgd-gd2-perl	Perl module wrapper for libgd - gd2 variant.
libproc-processtable-perl	Perl library for accessing process table information.
unzip	De-archiver for ".zip" files.
libregexp-common-perl	Perl module with common regular expressions.
libgrib-api-0d-0	GRIB decoding/encoding software library.
libudunits2-0	Library for handling of units of physical quantities.
netcdf-bin	Programs for reading and writing NetCDF files?
pdl	perl data language: Perl extensions for numerics.
libpdl-netcdf-perl	Netcdf-IO for PDL.





nco	Command-line operators to analyze netCDF files.
libyaml-perl	Perl module for YAML Ain't Markup Language
libkml0	A library to manipulate KML 2.2 OGC standard files.
python-kml	Python extension for libkml0.
python-yaml	YAML parser and emitter for Python.
python-scientific	Python modules useful for scientific computing.
plplot9-driver-cairo	Scientific plotting library (Cairo driver)
texlive-latex-base	TeX Live: Basic LaTeX packages.
texlive-latex-recommended	TeX Live: LaTeX recommended packages.
texlive-fonts-recommended	TeX Live: Recommended fonts.
texlive-latex-extra	TeX Live: LaTeX supplementary packages.
matplotlib	Python based plotting system in a style similar to Matlab
basemap	The matplotlib basemap toolkit is a library for plotting 2D data on maps in Python.

Table 2 - Main external libraries



## Appendix A - List of acronyms

AVISO	Archiving, Validation and Interpretation of Satellite Oceanographic data. Aviso distributes satellite altimetry data from Topex/Poseidon, Jason-1, ERS-1 and ERS-2, and EnviSat, and Doris precise orbit determination and positioning products.
CCI	Climate Change Initiative
CEO	Committee on Earth Observation
CERFACS	Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique
CERSAT	Centre d'Exploitation et de Recherche Satellitaire (IFREMER)
CLS	Collecte Localisation Satellites
CNES	Centre National d'Etudes Spatiales
DT	Delayed Time
DUACS	Developing Use of Altimetry for Climate Studies
ECMWF	European Centre for Medium-range Weather Forecasting
ECV	Essential Climate Variable
ESA	European Space Agency
FCDR	Fundamental Climate Data record
FTP	File Transfer Protocol
GDR	Geophysical Data Record
GFO	GEOSAT Follow-On Altimetric Satellite (US Navy).
IT	Information Technology
LWE	Low Wavelength Error
MSLA	Maps of Sea level Anomaly
NOAA	National Oceanic and Atmospheric Administration
NRT	Near Real Time
OE	Orbit Error
OI	Optimal Interpolation
QC	Quality Control
SEWG	System Engineering Working Group
SL	Sea Level
SLA	Sea level Anomaly
SSALTO	Segment Sol multissions d'ALTimétrie, d' Orbitographie et de



	localisation précise
SSH	Secure Shell or Sea Surface Height
T/P	Topex/Poseidon
UKMO	UK Met Office
TBC	To be confirmed
TBD	To be defined
AD	Applicable Document
RD	Reference Document