Herschel Legacy Science
Phase Readiness Review

Data Processing

Jorgo Bakker, on behalf of DP team

10/05/2016
Contents

- Recap: post-operations plan 2012
- Releases: cycle and roadmap
- Software: trackers and outlets
- Handover: transfer of knowledge
- The Unexpected: what and how we deal with it
- Current challenges: towards legacy infrastructure
Recap: post-operations plan

Nutshell, for Data Processing group:

- Support creation of the best possible products
- No Herschel resources after post-ops

Thus, data processing legacy:

- Legacy Science Products stored in Archive
- Software stored in legacy infrastructure
- Other data in legacy infrastructure
Given resources allocated at minimum level:

- **1\textsuperscript{st}: support bulk processing & quality control**
  - resulting in high quality legacy products from pipelines

- **2\textsuperscript{nd}: maintenance of interactive data processing**
  - supporting further exploitation of Herschel data in detail

- **3\textsuperscript{rd}: software documentation**
  - supporting further exploitation Herschel data without availability of a help desk

- **other wishes, in general**
  - best effort basis or cannot be addressed
  - virtual machine thought to be addressed as part of under geo return project (DASLT)
Releases: version roadmap

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>10.1</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>10.2</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>10.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.0</td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.1</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.0</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Development
- Testing & release
- Bulk reprocessing
Releases: Key features

Key features of each release described in Wiki

- version 10 – see: HipeWhatsNew10x
- version 11 – see: HipeWhatsNew11x
- version 12 – see: HipeWhatsNew12x
- version 13 – see: HipeWhatsNew13x
- version 14 – see: HipeWhatsNew14x
- version 15 – to be written
Releases: cycle control & prioritization

- All stakeholders represented in configuration control boards (CCBs)
- CCB follows MoSCoW prioritization on tickets
  - Must have – *will block release*
  - Should have – *important, but won’t block*
  - Could have – *desirable, if time permitting*
  - Would have – *nice, but re-planned for next release*

<table>
<thead>
<tr>
<th>Phase of specific version</th>
<th>Controlling configuration board</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development</td>
<td>Common and Instrument CCBs</td>
</tr>
<tr>
<td>Verification and validation</td>
<td>System CCB</td>
</tr>
<tr>
<td>Operations</td>
<td>Core CCB</td>
</tr>
</tbody>
</table>
Releases: versions and software builds

Continuous integration:

- software builds created per version
- as of version 13, the development cycle increased from ½ a year to 1 year
- builds for version 15 is a projection
- **spr** - bug fixes
- **scr** - change requests
- **patches** - fixes resulting from validation process
- As of version 13, the development cycle increased from ½ a year to 1 year
- Progressive reduction resources reflected in number of deliverables
Used for software development, validation and handover software:

- **CIB – Continuous Integration Build system**
  - http://herschel.esac.esa.int/hcss/build.php

- **JIRA – issue tracker**
  - http://herschel.esac.esa.int/jira/

- **Sonar – Code quality tracker**
  - http://herschel.esac.esa.int/sonar/

- **AT – Automatic Tester pipelines & userscripts**
  - http://herschel.esac.esa.int/at/
Software: outlets

- **User installers**
  - [http://www.cosmos.esa.int/web/herschel/hipe-download](http://www.cosmos.esa.int/web/herschel/hipe-download)

- **Developer installers**
  - [http://herschel.esac.esa.int/hcss/build.php](http://herschel.esac.esa.int/hcss/build.php)

- **Online documentation**
  - [http://herschel.esac.esa.int/hcss-doc-15.0/](http://herschel.esac.esa.int/hcss-doc-15.0/)
  - [http://herschel.esac.esa.int/hcss-doc-14.0/](http://herschel.esac.esa.int/hcss-doc-14.0/)
  - :

- **Platform support**
  - Microsoft Windows 32/64-bit: Vista onwards
  - Apple OSX: Mountain Lion onwards
  - Linux 32/64-bit
Consortia leaving at different pace

- **HIFI:** March 2016
  - and last hand over work about to be completed
- **SPIRE:** June 2016
  - Spectrometer group already left March 2016
- **PACS:** December 2016
  - Two key developers already leave in July 2016

Managing hand over process

- procedure same for all consortia
Handover: HIFI

- Kick-off meeting October 2015
  - hand over procedure
  - roadmap definition

- 10 common modules
  - 8 handed over
  - 2 remain in hands of HIFI

- 17 HIFI specific modules
  - 10 handed over
  - 2 in validation
  - 5 in progress

- 81% completed – see: JIRA
Handover: SPIRE

- **Kick-off meeting November 2015**
  - hand over procedure
  - roadmap definition

- **14 Common modules**
  - 8 handed over
  - 4 remain in hands of SPIRE
  - 2 in progress

- **53 SPIRE specific modules**
  - 16 handed over
  - 21 in validation
  - 16 in progress

- **73% Completed – see: JIRA**
Handover: PACS

- **Kick-off meeting March 2016**
  - hand over procedure
  - roadmap definition

- **11 Common modules**
  - initiated

- **17 PACS specific modules**
  - not started yet
The unexpected: Removing versant

- **Versant: object oriented database**
  - Single commercial dependency since start of development
  - License and support costs shared between HSC and ICCs
  - Faced serious upgrade and maintenance issues
  - License issues for post-operations and legacy phase

- **Refactor process**
  - Decided to remove dependency at Paris Meeting (May 2014)
  - Localized and controlled but serious refactoring effort
  - Completed and validated as of HCSS version 14 (Dec 2015)
Performance issues

- very poor query performance
- stability issues ingestion products and querying archive (AIO)
- seriously hampered bulk data reprocessing
- impacted user experience during bulk processing as well

Joint Archive/HSC effort

- definition of new hardware
- definition of deployment roadmap
- execution test & roll-out campaign
- significantly improved situation
The unexpected: Network infrastructure

- **Performance issues**
  - network disruptions and inexplicable slowness
  - affected progress with v11, v12 and v13
  - impacted bulk processing activities of above versions
  - drained resources to mitigate the effects

- **Joint CSG/HSC effort**
  - HSC providing continuous analysis and reports
  - CSG/Network investigations specific to HSC
  - general/major upgrades of network infrastructure in 2015
  - significantly improved situation mid development v14 onwards
The unexpected: New bulk processing needs

Original plan
- HCSS 14 release December 2015
  - bulk processing January-February 2016
- HCSS 15 release December 2016
  - bulk processing January-February 2017

Adaptation of plan needed
- instrument consortia expertise leaving earlier than v15
- increase of required extended processing efforts
- in tension with validation and acceptance efforts
- need for accommodating plans accordingly
Revised approach (summary)

- **HCSS 14.x**: Primary focus on improving legacy products
- **HCSS-14.0**: released mid December 2015
  - full validation and acceptance testing
  - bulk reprocessing all instruments
  - planned to be last processing for HIFI, but ...
- **HCSS-14.1**: released early April 2016
  - last bulk reprocessing SPIRE
  - in addition unforeseen bulk reprocessing HIFI needs
- **HCSS-14.2**: release planned early July 2016
  - bulk reprocessing PACS
  - placeholder for unforeseen SPIRE needs
- **HCSS-15.0**: release planned December 2016
  - interactive environment updates only
1\textsuperscript{st} time expected: 2011-2012
- from Sun GRID by Complutense University of Madrid
- to Univa GRID implementation
- migration hardware and software to new GRID

2\textsuperscript{nd} time unexpected: 2015-2016
- imposed by ESAC infrastructure
- migration to new ESAC GRID6 environment
- pros: upgraded OS, faster I/O, expanded capacity
- cons: extra efforts into adaptation, validation and migration
Current challenges: human resources

- **Controlling transfer of knowledge**
  - people leaving earlier than planned
  - both at Instrument Consortia and Herschel Science Centre
  - NAOC (China) could not find funding as assumed in POPS 2012

- **Addressed by**
  - advancing hand over of software modules where needed
  - minimizing maintenance by test automation
  - minimizing maintenance by code quality control
  - introducing SCRUM approach as of February 2015
Software Code repository

- Herschel uses CVS (antiquated) and will not be maintained
- SOCCI (under geo-return) or Github.com (public repository)?
  - either migration is non-trivial
  - migration not possible while consortia still connected
  - no resources available for actual migration
- Open source licenses and 3rd party libraries
  - a few libraries are in tension with each other
  - HSC has no resources to address/replace them

Legacy destination

- Main source code shipped as part of installer
- CVS dump on COSMOS (utility server)
- Migration to GIT on best-effort basis only
Current challenges: software binaries

- **Preserving medium term usage (<5 years)**
  - direct installation of software
  - direct reuse software components

- **Preserving long term usage (beyond)**
  - provision of pre-installed Virtual Machine
  - unclear status DAS-LT (under geo-return)
  - at minimum HSC will provide own Virtual Machine

- **Legacy destination**
  - Planned to use a COSMOS utility server
  - For both virtual machine and direct installation
Current challenges: raw data

- **Legacy raw data**
  - Consolidated raw telemetry
  - Raw auxiliary, raw ancillary data etc.

- **RAWDAR (under geo-return)**
  - Work in progress
  - New approach and not ready in time for Herschel

- **Legacy destination**
  - Planned to use a COSMOS utility server instead