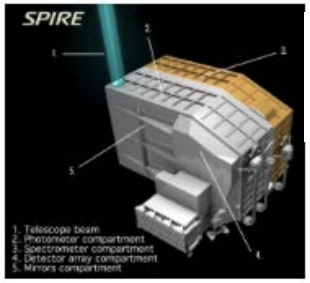
Pipeline Generated Products in a distributed Ground Segment: The example of Herschel

Eva Verdugo
Herschel Archive Scientist

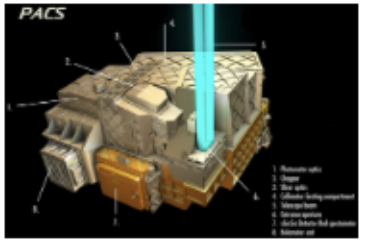
Science Operations 2017 – ESAC
17-20 October 2017

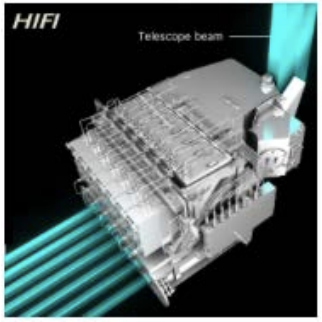
Pipeline developed in a broad number of institutions

- Three instruments: HIFI, PACS and SPIRE
- 14 different Observing Modes (photometry and spectroscopy) with slightly different pipelines developed at several different institutions
- Thousand of products generated per observation…



Consortium Institutes: Univ. Lethbridge (Canada); NAOC (China); CEA, LAM (France); IFSI, Univ. Padua (Italy); IAC (Spain); Stockholm Observatory (Sweden); Cardiff University, Imperial College London, RAL, UCL-MSSL, UKATC, Univ. Sussex (UK); Caltech, JPL, NHSC, Univ. Colorado (USA).





Consortium institutions: MPE (Germany); UVIE (Austria); KU Leuven, CSL, IMEC (Belgium); CEA, LAM (France); MPIA (Germany); INAFIFSI/OAA/OAP/OAT, LENS, SISSA (Italy); IAC (Spain).

Consortium institutions: CSA, U. Waterloo (Canada); CESR, LAB, LERMA, IRAM (France); KOSMA, MPIfR, MPS (Germany); NUI Maynooth (Ireland); ASI, IFSI-INAF, Osservatorio Astrofisico di Arcetri-INAF (Italy); SRON, TUD (The Netherlands); CAMK, CBK (Poland); Observatorio Astronómico Nacional (IGN), Centro de Astrobiología (CSIC-INTA) (Spain). Chalmers University of Technology – MC2, RSS & GARD, Onsala Space Observatory, Swedish National Space Board, Stockholm University – Stockholm Observatory (Sweden); ETH Zurich, FHNW (Switzerland); Caltech, JPL, NHSC (USA).

Coordination of Products Content and Format

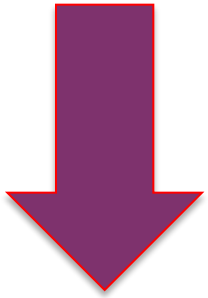
- Herschel Products Definition Group
- Led by A. Heras since the formation of the group in Sep. 2005 (4 years before launch!) till December 2008 when E. Verdugo took over the leadership
- Terms of reference:

  > "The Herschel Products Definition Group (HPDG) will enable the coordination of the definition of products across the Herschel Science Ground Segment"

- More than 30 presential meetings and telecons
- Formed by representatives of all areas of the ground segment: instruments scientists and developers, system engineers, data processing specialists, the archive scientist, the pointing calibration scientist, etc…

Problems I. Product Definition / Data model

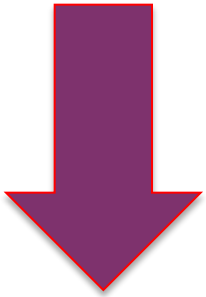
- A data model was required for the development of the Herschel Science Archive before the complete definition of pipelines and products (~April 2008) for the different tests of the Ground Segment before launch.



- Inefficient data model implemented: more than 15 millions of products in a single database table
- The data model required a complete re-definition in 2015! After a continuous degradation of the system was observed during the different bulk reprocessing exercises, with queries lasting forever (time-outs) and archive crashes

Problems I. Product Definition / Late additions

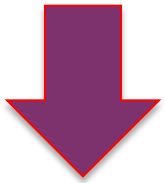
- New products (level 2.5 and level3) by the combination of several observations were produced in 2010 which were not contemplated originally in the Archive Data Model



- The Archive granularity is by observation, so again an inefficient data model. In this case to cope with multi-observations products
- The solution was to attach these products to all the observations involved in their production
- A better solution was far too expensive at that phase of the mission

Problems II. Products Content / Metadata

- Metadata requirements:
  - Definition of the mandatory set of metadata in all products
  - Homogeneous (in name and content) across all instruments and modes
- Very difficult to follow the correct implementation across hundreds of values in thousands of products
- A simple change of name could imply that the Archive queries were wrong or incomplete



- Repetition of metadata inside the products
- Large number of metadata with value NULL in the archive
- The consolidation of the metadata was not possible till ~2015
- Still so, a patch version of the archive in Feb. 2016 was required to disabled a query because the name of one metadatum was changed. It was enabled in October 2016 after a bulk reprocessing exercise which recovered all the values

Problems II. Products Naming convention

- The system incorporated an automatic generation of names for thousands of products (by using the metadata inside the products)
- Meaningful for external users
- Contain useful information to identify them: instrument, mode, type, level, …
- Avoiding duplication of names
- We implemented a convention like:

h<product/instrument><subinst><obsid/od>_<bbid>_<level><type>_<slice>_<timestamp> where

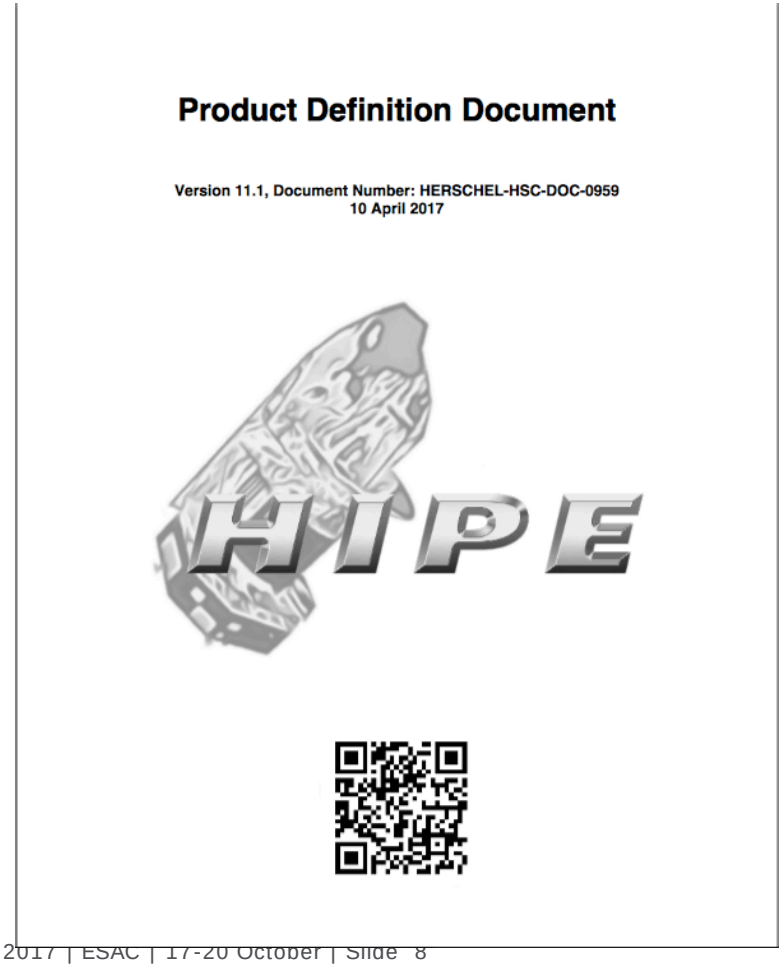
- h stands for Herschel
- <product/instrument>: is the product type such as *aux* for aux products or the instrument
- <subinst>: subinstrument used, the detector, polarisation mode etc...
- <obsid/od>: The observation ID given in decimal format.
- <bbid>: building blocks.
- <level>: The level of the product
- <type>: This indicates the type of product (maps, cubes, etc…)
- <slice>: When data from an observation need to be split up further than by building block

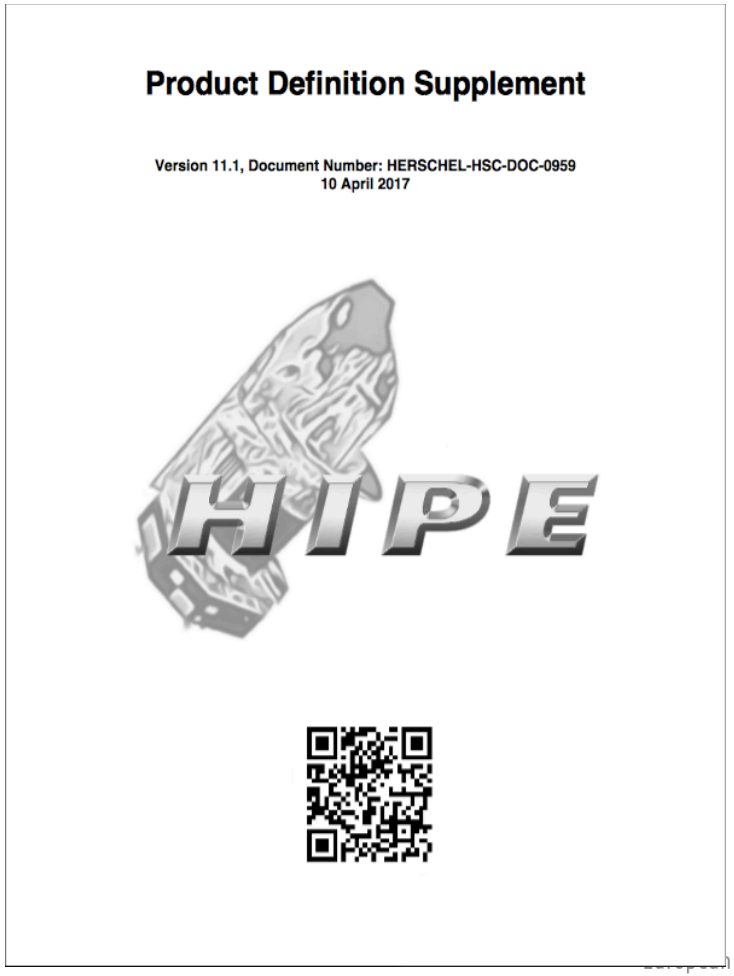
The ordering of the parameters is designed to give a logic ordering of the filenames when listed in a directory.

All product names also contain at the end a 13 digits number which is a <timestamp> that the system generates when the FITS product is created.

The Product Definition Document

All this was compiled in a document that from a document for developers (internal use) of ~ 3000 pages to a “legacy” document for archival users of 80 pages + a supplement with the detailed definition of each product





Conclusions and lessons learned

The consequences of a late or not complete definition of products may be unexpected! and fundamental! and crucial! and important!

- COORDINATION: Start very early in the mission
- COMMUNICATION: Implement a good system for communication, reporting and feedback
- LEADERSHIP: Many people to coordinate, a few to decide…
- FLEXIBILITY: Make the system as flexible as possible to cope with future unexpected events
- LOOK BACK: Learn from the lessons learned by other missions to know what problems you may need to face