The Herschel Data Products in the HSA

D. Teyssier

ESAC

25/10/2016
Outline

1. Products from the Standard Product Generation (SPG)
2. User-Provided Data Products (UPDPs)
3. Highly-Processed Data Products (HPDPs)
4. Ancillary Data Products (ADPs)
Standard Product Generation (SPG – aka pipeline data products)

- **Data Products generated by the respective instrument pipeline software**
  - Automatic product generation – no fine tuning
  - State of the art output as of applicable HCSS release
  - Best-ever pipeline products achieved by HIPE 14.1 for most of HIFI and SPIRE, and HIPE 14.2 for PACS – note that level 3 products for SPIRE-P and PACS-S will however come as version 14.2.1

- **Their data quality, but also limitations and caveats from a science-readiness perspective, will be described in the individual instrument presentations**

- **Served exclusively through the HSA**
  - All processing levels are made available, allowing ad hoc reprocessing by interested users (pipeline tasks only available within HIPE)
  - Products with older versions (13.0 and 14.0) also available through command line queries (within HIPE) – in contrast only one version available (latest one) through the HSA interface
User-Provided Data Products (UPDPs)

- **Data Products essentially stemming from Key Project consortia**
  - Formal commitment from KP teams was to at least provide all their data from the Science Demonstration Phase (SDP)
  - Rest of their core programme to be delivered on best effort basis, linked to accepted publications – several catalogues came with those
  - In addition, welcoming any initiative from other non-KP (guaranteed or open time) projects

- **To date, UPDPs from 22 programs/consortia have been ingested into the HSA**
  - First deliveries started as early as 2010 and HIPE 6/7 processing, implying data quality inferior to most recent HSA pipeline products, or in the worse-case scenario (rare) erroneous calibration
    (http://www.cosmos.esa.int/web/herschel/user-provided-data-products)

- **On-going activity to populate further the HSA with more UPDPs – focus will be put especially on catalogue and spectroscopy data – for photometry, pipeline products considered state-of-the-art**
Part of the pipeline products are not fully science-ready for some of the instruments – overall $\sim>10\%$ of all observations ($\sim20\%$ of spectroscopy data) could fall into that category.

Numerous dedicated software tasks and “useful scripts” exist in HIPE in order to circumvent those caveats and allow archive users to improve the products quality to full science-ready level.

The Herschel Ground Segment has taken the initiative to look into sub-sets of unruly products that could be systematically improved using some of the above tools, and provide them to the archive as additional or alternative products.

- These expert-generated products are called **Highly-Processed Data Products (HPDPs)**
- About 20+ HPDPs categories presently identified – see details in individual instrument presentations later this week.
- Will be served through the HSA, similar to UPDPs, or as a whole, see: [http://www.cosmos.esa.int/web/herschel/highly-processed-data-products](http://www.cosmos.esa.int/web/herschel/highly-processed-data-products)
Beyond SPG: Highly-Processed Data Products (HPDPs)

- **Typical targets for HPDP generation:**
  - Correction of remaining instrument artifacts (e.g. HIFI or PACS-S unchopped baseline distortion)
  - Source-coupling correction for SPIRE-S and PACS-S products
  - Red- and blue-leak spectral ranges for PACS-S products

- **Other added-value HPDPs considered:**
  - Optimised product merging in spatial or spectral domain, maximising Signal to Noise Ratio
  - Catalogues (PACS and SPIRE Point Source Catalogues, spectral line feature catalogues for each spectrometer)
Highly-Processed Data Products
HIFI examples

Baseline distortion correction in Spectral Scans

SPG 14.1

After baseline cleaning

SPG 14.1

After baseline cleaning

Mapping Obsid merging

OMC-1 [CII] 158 µm

Orion Bar [CII] 158 µm
16 Obsids

Obsid 1342250415

Obsid 1342250412

Obsid 1342250414
Highly-Processed Data Products
PACS-P examples

**Merge of overlapping fields**

- **LMC at 100 µm**
- **SMC at 100 µm**

**North Galactic Pole**

**Galactic Plane (±115 degree of Gal. Centre)**

- Unimap HPDP
- Pipeline Level 2

Courtesy of J. Graciá-Carpio, MPE
Highly-Processed Data Products
SPIRE-S examples

Various calibration corrections in SPIRE-S

Pointing offset correction

Before

After

Semi-extended source

Before

After

Background emission

Before

After

Each case requires ad hoc tuning of correction tool by instrument expert
Highly-Processed Data Products

PACS-S example

Continuum subtraction in un-chopped PACS-S observations

The continuum in un-chopped observations is usually not well calibrated and unreliable – the most practical products is that with continuum-subtracted spectra.
Highly-Processed Data Products
PACS/SPIRE Point source catalogues

- Project started in late 2013 in order to extract point sources from the PACS and SPIRE photometer maps in all 6 available bands (∼10% of whole sky)
  - Effort co-executed by HSC (ESA), NHSC (NASA) and Konkoly obs. (HU)
- Homogeneous source extraction allowing unbiased sensitivity comparison - significant effort dedicated to simulation for spurious detection characterisation
- Status as of October 2016:
  - Source candidate extraction completed for both instruments
  - SPIRE-P expects ∼6.7M objects (3.2M, 2.6M, 0.9M at 250, 350, 500 µm). Most likely separated into high/low flux reliability catalogues
  - PACS-P expects ∼8.6M candidates, however high level of spurious detection (∼85%) – typically ∼<1M unique sources (all bands)
  - First catalogue version will be delivered in December 2016, together with an explanatory supplement
  - Catalogue entries will be served by the HSA, either as full catalogue, or on an obsid-per-obsid basis, based on the output of queries
  - No band merging planned – in total, 6 individual catalogues
Highly-Processed Data Products

Line catalogues

SPIRE-S: spectral feature catalogue
AFGL2688 Observation ID 1342188198

HIFI: line list catalogue
Orion_S: 479.501 - 489.563 GHz (20 lines)

HIFI: line transition assignment table

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Orion_S</td>
<td>483.46070148866235</td>
<td>483.461908656317</td>
<td>3.194957781782164-4</td>
<td>4.85883927881915-4</td>
<td>0.4660577581008531-4</td>
</tr>
<tr>
<td>2</td>
<td>Orion_S</td>
<td>491.92710065762265</td>
<td>491.92720065762265</td>
<td>2.1385532134721016-4</td>
<td>4.52312652504505-4</td>
<td>0.3061965231935435-4</td>
</tr>
<tr>
<td>3</td>
<td>Orion_S</td>
<td>492.775961161346</td>
<td>492.77610135192489</td>
<td>5.41143645535314-4</td>
<td>5.68135534899013-4</td>
<td>0.9185264738264691-4</td>
</tr>
<tr>
<td>4</td>
<td>Orion_S</td>
<td>494.913394865374</td>
<td>494.913594865374</td>
<td>0.04571534753076458-4</td>
<td>0.0739193510936536-4</td>
<td>0.9357597105205832-4</td>
</tr>
<tr>
<td>5</td>
<td>Orion_S</td>
<td>503.04068476643153</td>
<td>503.04083476643153</td>
<td>3.52622198387137-4</td>
<td>0.87622590188214-4</td>
<td>0.4965295320014114-4</td>
</tr>
<tr>
<td>6</td>
<td>Orion_S</td>
<td>508.23231579820535</td>
<td>508.23246579820535</td>
<td>3.57456535531647-4</td>
<td>4.09689058415819-4</td>
<td>0.4903193234287465-4</td>
</tr>
<tr>
<td>7</td>
<td>Orion_S</td>
<td>510.89795216530335</td>
<td>510.89810216530335</td>
<td>2.79511158753573-4</td>
<td>4.77213862860112-4</td>
<td>0.5533994239037586-4</td>
</tr>
<tr>
<td>8</td>
<td>Orion_S</td>
<td>511.402917615055</td>
<td>511.403067615055</td>
<td>0.4098114527955358-4</td>
<td>0.51378645867278-4</td>
<td>0.58645162731053-4</td>
</tr>
<tr>
<td>9</td>
<td>Orion_S</td>
<td>511.706976732732-4</td>
<td>511.707026732732-4</td>
<td>3.50720581234567-4</td>
<td>6.02717627120425-4</td>
<td>0.4944953769059625-4</td>
</tr>
<tr>
<td>10</td>
<td>Orion_S</td>
<td>512.066340883639</td>
<td>512.066480883639</td>
<td>1.10404106268708-7</td>
<td>1.67257690616224-4</td>
<td>0.194450370358654-4</td>
</tr>
<tr>
<td>11</td>
<td>Orion_S</td>
<td>514.336167293659</td>
<td>514.336317293659</td>
<td>0.59616116436668-4</td>
<td>0.86615779795629-4</td>
<td>0.90650092792992-4</td>
</tr>
<tr>
<td>12</td>
<td>Orion_S</td>
<td>519.180762924829-4</td>
<td>519.180912924829-4</td>
<td>5.819986426786776-4</td>
<td>8.42961532572821-4</td>
<td>0.903225982080835-4</td>
</tr>
<tr>
<td>13</td>
<td>Orion_S</td>
<td>520.816680027104</td>
<td>520.816830027104</td>
<td>3.6315159634707146-4</td>
<td>7.30664262595899-4</td>
<td>0.49279480120071-4</td>
</tr>
<tr>
<td>14</td>
<td>Orion_S</td>
<td>520.888139355255</td>
<td>520.888289355255</td>
<td>3.57290610403125</td>
<td>1.18486301604152-4</td>
<td>0.1906458081988986</td>
</tr>
<tr>
<td>15</td>
<td>Orion_S</td>
<td>524.255794053741</td>
<td>524.255944053741</td>
<td>3.4573822550355-4</td>
<td>8.21535425308293</td>
<td>0.3129186390044777-4</td>
</tr>
<tr>
<td>16</td>
<td>Orion_S</td>
<td>524.372197417252-4</td>
<td>524.372347417252-4</td>
<td>3.07594111612571-4</td>
<td>3.10150354040835-4</td>
<td>0.41024527017971-4</td>
</tr>
<tr>
<td>17</td>
<td>Orion_S</td>
<td>527.532798600880-4</td>
<td>527.532948600880-4</td>
<td>4.52355042606103-4</td>
<td>2.53855496621827-4</td>
<td>0.45145026480535-3</td>
</tr>
<tr>
<td>18</td>
<td>Orion_S</td>
<td>525.044978061637</td>
<td>525.044978061637</td>
<td>4.713470120595972-4</td>
<td>2.35748350035305</td>
<td>0.64181690936841-3</td>
</tr>
</tbody>
</table>
### Summary: the legacy science

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Non science-ready pipeline products</th>
<th>Fraction recovered by HPDPs, or added-value HPDPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIFI</td>
<td>20% (distorted baselines)</td>
<td><strong>Half of those</strong>, leaving ~10% of total observations as non strictly science-ready (mostly single-point observations)</td>
</tr>
<tr>
<td>PACS-S</td>
<td>&gt; 15 % (not counting TBD number of semi-ext. source cases)</td>
<td><strong>All</strong> but semi-extended source cases. Added-value HPDPs for ~30% of observations (un-chopped and/or featuring red leak range data)</td>
</tr>
<tr>
<td>SPIRE-S</td>
<td>~ 55% (semi-extended source and background contamination)</td>
<td>¾ of those, leaving just <strong>15%</strong> of all sparse mode observations un-corrected</td>
</tr>
<tr>
<td>PACS-P</td>
<td><strong>All science ready</strong></td>
<td>Added-value HPDPs for 75% of parallel mode and 4% of PACS-P maps (being only 7% of total PACS-P observations but ~30% of the total observing time)</td>
</tr>
<tr>
<td>SPIRE-P</td>
<td><strong>All science ready</strong></td>
<td>SPIRE Point Source Catalogue extracted from all SPIRE-P observations</td>
</tr>
</tbody>
</table>
Ancillary Data Products (ADPs) correspond to products not necessarily associated to any observation performed by the Observatory.

Used as Herschel’s long-term preservation storage of items that may be needed for users of the archive, but also as source of information for future missions – as such a lot of engineering-oriented items are contemplated there.

The bulk of the instrument ADPs is provided through the Instrument teams, or the corresponding HSC experts after instrument team has been formally dismantled – Observatory Ancillary Products are provided by the HSC.

The ADP deliverables can be summarised as:

- Calibration: calibrator models, instrument PSFs, photometer filters
- Software: on-board software images and release notes
- Observatory: all uplink-related products
- Engineering: trend-analysis data from instrument HouseKeeping
- Historical: Telemetry and Science data from pre-launch test campaigns

http://www.cosmos.esa.int/web/herschel/ancillary-data-products
Ancillary Data Products: pre-launch instrumental measurements

Unique dataset collected pre-launch: HIFI spectrum of CH$_3$OH

~6000 spectra from laboratory gas-cell
Covers ~22,000 lines of JPL catalogue
Conclusions

- A plethora of products will be served by the HSA on top of the pipeline outputs
- A key component of those deliverables are **Highly-Processed Data Products**, exquisitely curated by instrument experts:
  - Those dedicated to artefact or sub-optimal calibration correction will allow recovery of \( \sim \frac{2}{3} \) of the pipeline products identified as non science-ready (the vast majority being spectrometer observations)
  - Those providing added-value legacy products typically in the form of optimised (large) spatial merging will cover \( \sim 75\% \) of the parallel mode obs., and 4% of all PACS-P prime observations (30% of PACS-P time)
  - Line catalogues will apply to up to \( \sim 75\% \) of spectral data, and point source catalogues to the whole sky area mapped by the photometers
- The long-term preservation of other unique mission-related products will be ensured by the storage of **Ancillary Data Products**. Noticeable component of this archive subset are the **Herschel Calibrator models**, or the instrument **PSFs**
QUESTIONS?