# DOCUMENT

# User Requirements for the Querying and Serving of Highly Processed Data Products and Ancillary Data Products by the HSA

| | |
|---|---|
| **Prepared by** | **David Teyssier, Eva Verdugo** |
| **Reference** | **HERSCHEL-HSC-DOC-2100** |
| **Issue** | **1** |
| **Revision** | **3** |
| **Date of Issue** | **15 April 2016** |
| **Status** | **For review** |
| **Document Type** | **Requirement Document** |
| **Distribution** | **HSC, SAT** |

European Space Agency
Agence spatiale européenne

# APPROVAL

| Title   User Requirements for the Querying and Serving of HPDPs and Ancillary Data Products by the HSA | |
| --- | --- |
| Issue  1 | Revision  3 |
| Author   D. Teyssier, E. Verdugo | Date   15 April 2016 |
| Approved by: P. Garcia-Lario | Date |
| | |
| | |

# CHANGE LOG

| Reason for change | Draft Issue | Revision | Date |
| --- | --- | --- | --- |
| First version of document | 1 | 1 | 17 February 2016 |
| Changed "Ancillary Legacy Products" to "Ancillary Data Products" | 1 | 2 | 7 April 2016 |
| Implemented comments from K. Exter and I. Valtchanov. Re-organised section 2.2.1 to address delivery content definition and naming convention. | 1 | 3 | 15 April 2016 |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

**European Space Agency**
**Agence spatiale européenne**

## Acronyms:

| | |
|---|---|
| ADP | Ancillary Data Product |
| HELL | *Herschel* Explanatory Legacy Library |
| HK | HouseKeeping |
| HPDP | Highly Processed Data Product |
| HPSC | *Herschel* Point Source Catalogue |
| HSA | *Herschel* Science Archive |
| HUI | HSA User Interface |
| OBSID | Observation Identifier |
| PSF | Point Spread Function |
| SPG | Standard Product Generation |
| UPDP | User Provided Data Product |
| XSA | XMM-Newton Science Archive |

European Space Agency
Agence spatiale européenne

# Table of Contents

# 1 INTRODUCTION

This is the user requirement document compiling the specifications for the product query and serving of the *Herschel* Highly Processed Data Products (HPDPs) and Ancillary Data Products (ADPs) from the *Herschel* Science Archive (HSA). Although the HSA is expected to migrate from a Java App. interface into a web-based interface in its 8.x incarnation, the present requirements are presented in analogy with existing implementations using either of those technologies. Because some of the concepts considered here for HPDP are inherited from the current User Provided Data Product (UPDP) interface, but also try to circumvent some of the limitations of the latter, we recommend that the specifications phrased here be used to potentially improve as well on the functionalities to be used to query and serve UPDPs in the HSA.

# 2 DEFINITION OF HPDPs AND ADPs

## 2.1 General description

Highly Processed Data Products are by-products or alternative products to the outcome of the Standard Product Generation (SPG), whereby an instrument expert has applied further post-processing steps in order to make the product scientifically superior to that of the SPG. Most of the HPDPs therefore correspond to products attempting to either improve on the data quality (e.g. removal of residual instrument artefacts), or provide additional science products that require use of interactive scripts, achieving in some cases a more accurate flux calibration for those products. HPDPs can also be catalogues, or, especially in the case of the photometer sub-instruments, mergers of a larger number of overlapping observations that those contemplated by the HPDP. An exhaustive list of HPDPs can be found at:

http://herschel.esac.esa.int/twiki/bin/view/Public/HPDPDefinitionPage

Ancillary Data Products are products generated during the various phases of the mission (i.e. including pre-launch and post-operation), which are usually not attached to any particular *scheduled* observation. They are meant to allow long-term preservation of items that were not ingested by default in the HSA during the operational phases. This includes for example models used to derive the instrument calibration, or instrument PSFs, but also software, or graphical and tabulated databases of the instrument HK generated e.g. for health monitoring purpose. A list of the currently considered ADPs can be found at:

http://herschel.esac.esa.int/twiki/bin/view/Public/AncillaryDataProductsPage

Both products shall be queriable and servable by the HSA, but also made available in dedicated portals (i.e. web pages on Cosmos) to, typically, FTP repositories. We only discussed here the requirements on the former – the latter belongs to the design of HELL.

## 2.2   Assumptions on the products

### 2.2.1    Deliverable format, structure and naming convention

In the remainder of the document, we make the following assumptions (some can be considered as requirements on the products themselves):

*Deliverable content:*

1. HPDPs can take the form of FITS files, ASCII files, or images (aka postcards, usually PNGs or PDFs)
2. ADPs can take a wider variety of format (FITS, xls, csv, doc, pdf, any image format, any plain text format, binary, etc)
3. It is expected that several postcards could be associated with a given OBSID within a given HPDP. We propose that a "***prime postcard***" is identified, that would be most illustrative of the representative data quality proposed by the product.
4. Because some of the HPDPs will most likely supersede the science quality of some of the SPG outputs, they may end up being promoted to "Legacy Products" (although the implications of that concept reg. HSA design are yet to be assessed). When this happens, the postcard originally generated by the SPG as browse product may also be substituted by one representing the improved quality of the HPDP. Following the scenario described in the previous bullet, it is assumed that the ***prime postcard*** will be selected and proposed to serve as new postcard. This ***prime postcard*** will also be relevant for the particular requirement described in **HPDP-REQ8**
5. Both HPDPs and ADPs shall come accompanied with a dedicated **release note**, as well as an abstract, an author and a title. A possible template for such release notes is proposed at the following link. It is also possible to pass more OBSID-specific information via dedicated README files. This might be particularly relevant for those HPDPs relying on an ad hoc application of *ipipe* scripts, with possibly different parameters than those featured by default. One could argue whether the script itself shall be provided as well – this needs to be looked at again when more experience is gained on those particular products.
6. Both HPDPs and ADPs come in a collection of zipped archives where products are grouped according to a given general product name (e.g. the "HIFI Expert Reduced Spectral Scans" HPDP, or the "Instrument PSFs" ADP). **In this document, the terms "*HPDP*" or "*ADP*" represent one of those archives, while the term "*product*" represents one component of those archives**

*Header content:*

1. The HPDP FITS files shall always hold one or more OBSID meta-data keywords in their header. The keyword **OBS_ID** will be used when only one observation is concerned, while keywords **OBSIDxxx** (xxx from 001 to 999) will be used when more than one observation is concerned
2. The HPDP headers may, however, not contain some of the keywords necessary for advanced queries – as such the OBSID is the only discriminant keyword that can be used for filtered queries
3. ADPs will in general not hold any OBSID keyword in their header (when header applies), with possible exceptions e.g. for calibrator model deliveries. In this latter case, the easiest approach is to have those particular products also delivered in the framework of an UPDP, and so mechanisms to associate products to OBSIDs are already available

*Naming convention:*

1. The products and postcards naming convention needs to allow straightforward linking of them to a given OBSID (done via a dedicated xml file). We are not proposing a rigid/unique naming convention to follow, however the strict minimum should be to have the OBSID. A possible convention could be:
   <div align="center">

   **&lt;OBSID&gt;_&lt;HPDP&gt;_&lt;category&gt;.&lt;extension&gt;**
   </div>
   where &lt;HPDP&gt; is the top-level HPDP nickname, &lt;category&gt; represents the various flavours that could apply to the product (e.g. cubes, averaged maps, postcard, README, etc), and &lt;extension&gt; simply the file type. Recommendation reg. the naming convention of the **prime postcard** is addressed in the following bullet
2. Following the naming convention mentioned in the 3rd bullet above, the **prime postcard** needs to be singled out from all other postcards applying to a given HPDP. A possible way would to use as &lt;category&gt; "*prime-postcard*" instead of "*postcard*".

## 2.2.2 Product sub-categories and types

We also need to define so-called "*HPDP sub-categories*" and "*HPDP types*", in an analogy with what is currently used to interface with UPDPs.

*HPDP sub-categories:*

This term refers to a sub-group of products present in a given HPDP archive, that one would want to access without having to download the whole HPDP archive. We assume that the following applies to the *Herschel* HPDPs (same as current UPDP interface):
- *Full HPDP*
- *Catalogue*
- *Data*
- *Postcard*

- *Documentation*

Each HPDP does not necessarily contain products belonging to every single category – for example Catalogue HPDPs usually don't hold data (or postcards), and vice-versa. One noticeable exception to the above is the release note – i.e. documentation – which is mandatory for each HPDP.

*HPDP types:*

This distinction is meant at performing queries following more "science-oriented" criterions (e.g. spectroscopy *vs* photometry, etc). The types currently considered for the *Herschel* HPDPs are:
- *Photometry*
- *Spectroscopy*
- *Line Catalogue*
- *Source Catalogue*

What this implies for the HSA is that HPDPs need to be ingested with some of the above keywords being associated to it. This also implies that HPDP providers need to provide this information to the Archive Scientist when the delivery is made.

*ADP types:*

Similarly, Ancillary Data Products can be grouped into certain types. The types currently considered for the *Herschel* ADPs are:
- *Science data*
- *Models*
- *Instrument PSFs*
- *Software*
- *Instrument HK*
- *Documentation*

Unlike for HPDPs where the type is an attribute that can be used to filter queries, the ADP types are mostly relevant to the design of the page (typically in a thin-layer type of implementation) that will serve as interface to those products.

# 3 USER REQUIREMENTS ON HPDPS

## 3.1 HPDP-REQ1: HPDP summaries

It should be possible to visualize the description of an HPDP (typically its title, abstract, author, the type it belongs to, and its available sub-categories) prior to any query or action. Total HPDP download sizes should also be indicated

## 3.2   HPDP-REQ2: HPDP full download

Any HPDP shall be downloadable as a whole, i.e. independent on any selection filter other than that of the particular HPDP of interest. Several HPDPs should be accessible at a time, i.e. this particular interface should not be limited to one single HPDP. We assume in the following that this interface will be in the form of a so-called "*HPDP selection panel*" – even if this takes a different form in practice, this is how we will refer to it when necessary in the following specifications. The requirement is that full downloads should be possible from the selection panel already, i.e. the user should not need to do this via an additional window opened upon an HPDP selection of some sort.

## 3.3   HPDP-REQ3: HPDP sub-category download

It should be possible to select sub-categories (see Section **2.2**) of the HPDPs in order to retrieve only parts of the HPDP products. This selection should be available as a filter to the action specified in **HPDP-REQ2**.

## 3.4   HPDP-REQ4: HPDP search by type and title

Alternatively to the actions specified in **HPDP-REQ2** and **HPDP-REQ3**, it should be possible to query HPDPs according the type they have been linked to, as well as using semantic filters applying to the HPDP title and/or abstract. Unlike the action specified in **HPDP-REQ2**, this query will result in a list of OBSIDs. This requirement covers functionality similar to that already existing for UPDP under the "Search UPDP" tab.

One use case to warn against in the context of this requirement is that of products that will undergo several post-processing steps, that taken separately, will correspond to individual HPDPs – a typical example could be that of PACS-S data benefiting both from a background contamination correction (in itself a particular set of HPDPs) and of off-centred point-source offset correction (again, in itself a particular set of HPDPs). At that stage it is not clear how to handle those products, e.g. to which HPDP archive then should belong.

## 3.5   HPDP-REQ5: HPDP filtered queries – general

HPDP shall also be served according to any other filter applying to the HUI. This means that the list of OBSIDs returned from a particular query based on any filter other than those of the

dedicated *HPDP selection panel* (**HPDP-REQ1**) should take into account the availability of products from any HPDP for these OBSIDs and evidence it.

## 3.6   HPDP-REQ6: HPDP filtered queries – quick-look

Using the UPDP analogy, there should be a dedicated column in the query result table indicating the availability or not of HPDP for a given OBSID. Clicking on this cell should provide information about all HPDP products available for this OBSID. The current equivalent UPDP implementation of this action returns a window providing the top-level summaries of each HPDP relevant for the corresponding OBSID. For HPDPs, this latter is covered by **HPDP-REQ1** and the requirement for the quick-look window is different. Instead, the window shall display the available HPDP products associated to the considered OBSID in the form of preferably only one postcards per HPDP (the **prime postcard**), in a similar fashion e.g. to what the XSA "Show details" icon provides (in this case, with a scroll-down panel – see http://nxsa.esac.esa.int/).

## 3.7   HPDP-REQ7: HPDP filtered queries – full retrieval

The outcome of **HPDP-REQ5** is a list of OBSIDs for which all available products shall be evidenced in the relevant columns. For HPDPs this implies that several HPDP products may be associated with a given OBSID. It shall be possible to download those products based on a dedicated HPDP "*retrieve*" action button. This action, however, should support retrievals according to **two** different use cases: the *first one*, contemplated in this requirement **HPDP-REQ7**, considers the retrieval of all HPDP products associated with a given OBSID selection – it is similar to what the UPDP interface currently offers via the "User Provided Data Product" icon at the top of the query result table. The *second one*, contemplated in the next requirement **HPDP-REQ8**, should allow a further selection of these different products, i.e. a user should not be forced to download **all** HPDP products associated to one or more OBSIDs.

## 3.8   HPDP-REQ8: HPDP filtered queries – selected retrieval

Following the results of a query as specified in **HPDP-REQ5,** users should also be able to download a sub-set of HPDP products associated to a given OBSID selection (i.e. one or more OBSIDs). For this a dedicated selection panel should be made available, where all individual HPDPs available for a given OBSID should be evidenced, e.g. in a tabulated form (typically all available HPDP *vs* OBSID). In order to keep the display simple, this table should ideally only display the "*prime postcard*" for each combination of HPDP and OBSID. It would correspond to the "*2-D*" (HPDP *vs* OBSID) version of the quick-look panel specified in **HPDP-REQ6** (which is contemplated for only one OBSID). Note that there is currently no such functionality existing for UPDPs.

## 3.9 HPDP-REQ9: *Herschel* Point Source Catalogues

The handling of the HPSC requires a special approach. The deliverable of this HPDP will consist both in the full catalogues of source IDs, and in all individual sub-catalogues corresponding to the list of source IDs associated to a given OBSID (i.e. one file per OBSID). While the former shall be served via the action specified in **HPDP-REQ2**, the latter is what shall be served via the queries specified in **HPDP-REQ4** and **HPDP-REQ5**. It is still TBD at this stage whether all 6 catalogues are treated as a whole (i.e. always provide catalogues for all 6 filters together) or individually – the former would greatly simplify the selection interface required in **HPDP-REQ6** and **HPDP-REQ8**.

## 4 USER REQUIREMENTS ON ADPs

## 4.1 ADP-REQ1: ADP summaries

It should be possible to visualize the description of an ADP (typically its title, abstract, author, the type it belongs to, and its available sub-categories) prior to any query or action. Total ADP download sizes should also be indicated.

## 4.2 ADP-REQ2: ADP full download

Unlike HPDPs or UPDPs, ADPs are not ingested in the HSA; instead they are simply stored in dedicated (FTP) repositories. In the context of a thin-layer design, this means that the HSA interface is mostly a dedicated web page allowing to perform downloads of the ADPs of interest (see e.g. the "Catalogues and tools" tab in the XSA pages). As such the requirement is simply that full download of ADP archives should be made available through this interface.

European Space Agency
Agence spatiale européenne