Herschel Science Archive User requirements

Herschel/HSC/DOC/1081

Issue 1.0
25 July 2008

Eva Verdugo (ESA/HSC)
## Document Approval

<table>
<thead>
<tr>
<th>Prepared by</th>
<th>Organization</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eva Verdugo</td>
<td>HSC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Approved by</th>
<th>Organization</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Christophe Arviset</td>
<td>SAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leo Metcalfe</td>
<td>HSC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Document Status Sheet

<table>
<thead>
<tr>
<th>1. Document Title: Herschel Science Archive User requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2. Issue</strong></td>
</tr>
<tr>
<td>Draft 0</td>
</tr>
<tr>
<td>0.9</td>
</tr>
<tr>
<td>1.0</td>
</tr>
</tbody>
</table>
**Document Change Record**

1. **Document Title:** Herschel Science Archive User requirements  
2. **Document Reference Number:** Herschel/HSC/DOC/1081  
3. **Document Issue/Revision Number:**

<table>
<thead>
<tr>
<th>3.1 Section</th>
<th>3.2 Reason For Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Table of Contents

1 Introduction........................................................................................................................................1
1.1 Purpose........................................................................................................................................1
1.2 Overview.......................................................................................................................................1
1.3 Assumptions...................................................................................................................................1
1.4 Glossary.........................................................................................................................................1
    1.4.1 Acronyms .........................................................................................................................1
    1.4.2 Definition of terms...............................................................................................................2
1.5 References.....................................................................................................................................2
    1.5.1 Applicable Documents......................................................................................................2
    1.5.2 Reference Documents....................................................................................................3
1.6 Document Overview......................................................................................................................3
    Transitory Note...............................................................................................................................4
2 Functional user requirements ........................................................................................................5
    2.1 High level requirements.........................................................................................................5
    2.2 Content requirements.............................................................................................................6
    2.3 Data access process requirements........................................................................................7
    2.4 Data query requirements.......................................................................................................7
    2.5 Data query results requirements............................................................................................9
    2.6 Data retrieval requirements..................................................................................................10
    2.7 Security requirements...........................................................................................................11
3 Constraint requirements..............................................................................................................12
    Examples of queries....................................................................................................................14
1 Introduction

1.1 Purpose

This document presents the User Requirements on the Herschel Science Archive (HSA).

1.2 Overview

The requirements collected in this document cover all the HSA requirements for the Astronomer, Calibration Scientist and Instrument Engineer actors following the recommendations made by the Rapid Response Force group on HCSS Data Access in its final report [AD-9].

The start of the HSA URD is the Herschel Data Access URD [AD-10] from which the initial requirements have been extracted. The HSA URD is intended to be a "living document". After each new public release of the HSA, the URD content will be reviewed, and a new Issue agreed and published. The custodian of the HSA URD is the Archive Scientist (AS).

1.3 Assumptions

The requirements in this document assume that the HSA will be similar to the other ESA Science astronomical archives (IDA, XSA, etc...) with respect to:

- General capabilities: an HSA User Interface which is the GUI front-end of the complete HSA system, an HAIO (Herschel Archive Inter-Operability) which allows users to query on metadata and retrieve data using scriptable interfaces, and access to Herschel data through Virtual Observatory specific standards and protocols.

- Security: HSA shall respect data access rights and allows public data access for anybody and proprietary data access for authorized privilege users (e.g., Calibration Scientists, Instrument Experts).

- Content: all Herschel data shall be stored in the HSA.

In general, all HSA capabilities (and in particular the HSA User Interface functionalities) common to other ESA Science archives should be implemented like in those archives unless the contrary is explicitly stated.

1.4 Glossary

1.4.1 Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>Applicable Document</td>
</tr>
<tr>
<td>AOR</td>
<td>Astronomical Observation Request</td>
</tr>
<tr>
<td>AOT</td>
<td>Astronomical Observation Template</td>
</tr>
<tr>
<td>ASCII</td>
<td>American Standard Code for Information Interchange</td>
</tr>
<tr>
<td>BSM</td>
<td>Beam Steering Mirror</td>
</tr>
<tr>
<td>CSDT</td>
<td>Common Software Development Team</td>
</tr>
<tr>
<td>CUS</td>
<td>Common Uplink System</td>
</tr>
<tr>
<td>DPUG</td>
<td>Data Processing Users Group</td>
</tr>
<tr>
<td>ESAC</td>
<td>European Space Astronomy Centre</td>
</tr>
<tr>
<td>FITS</td>
<td>Flexible Image Transport System</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>HAIO</td>
<td>Herschel Archive Inter-Operability</td>
</tr>
</tbody>
</table>
1.4.2 Definition of terms

Herschel Common Science System (HCSS): The software system (tools and databases) required to carry out end-to-end Herschel Science Operations, and which is defined by the requirements in the HCSS URD [AD-3].

Herschel Science Archive (HSA): The standard system developed by the Science Archives Team at ESAC for general data access and retrieval, and is equivalent to the ISO, XMM, Integral Science Archives. The HSA GUI is called the Herschel Science Archive User Interface (HUI).

1.5 References

1.5.1 Applicable Documents

AD-1 Herschel Data Access RRF Terms of Reference, issue 1.0, 9 January 2007, Herschel-HSC-
Chapter 2 of the document describes the functional user requirements for data access. Section 2.1 contains the high level user requirements, most of them directly extracted from the HCSS URD [AD-3]. These high level requirements are further specified and detailed in the next sections of the document. For example, what is meant by “Herschel data” is clarified in the requirements where the data to be queried and retrieved are listed in detail.

Section 2.2 contains the requirements on data content. Section 2.3 contains the requirements that are common to the data access process, that is, querying, selection and display of query results, and retrieval of data. Sections 2.4, 2.5 and 2.6 provide the respective detailed requirements for these steps. In Section 2.7 the requirements on security are listed.

Chapter 3 provides the constraint requirements.

Finally, the Appendix provides examples of queries.
Transitory Note

Each requirement has been assigned a "priority" on a numerical scale 1-3. This is used to formulate the schedule; it indicates how early in the development this capability must be present. The legend of priority rates is as follows:

Priority 1 .- HSA version 0.9beta3 (E2E 3.7)
Priority 2 .- HSA version 0.9 (E2E 4)
Priority 3 .- HSA version 1.0 (first public version)
Priority 3+ .- Later HSA versions
2 Functional user requirements

2.1 High level requirements

Most of the requirements in section 2.1 have been directly extracted from the HCSS URD. They are consequently high level requirements, established and agreed as a basis for the HCSS implementation. These requirements are expanded and detailed in the following sections of this document.

HSA-UR-0010  It shall be possible to access Herschel data by queries.  
It is assumed that regarding queries two main group of users can be defined: the external users (astronomer, proposer, observer, HOTAC) and the expert users (HSC, S/W maintenance, help-desk and ICCs). External users will perform “general” queries and expert users both “general” and “expert” queries. The access and proprietary rights will determine which HSC data will be accessible for retrieval.

Priority: 1

HSA-UR-0020  Queries against a criteria or a combination of criteria shall give as a result the list of Herschel data items that match the query.  
Here “data items” addresses both Herschel data entities, or the values of a queried parameter (from a list of predefined metadata keywords; see UR-0200) e.g. for Trend Analysis (Trend Analysis is a task for the ICCs, not for the astronomer).

Priority: 2

HSA-UR-0030  It shall be possible to retrieve any data associated with the result of a query, according to the user’s access rights.

Priority: 2

HSA-UR-0040  The user shall be able to query, select and retrieve data from the Herschel Science Archive.

Priority: 1

HSA-UR-0050  It shall be possible to run non-interactive queries through batch scripts able to generate and to prepare a (subset of) the corresponding data for download.  
(This refers to the HAIO system).

Priority: 1

HSA-UR-0060  The interoperability between the HSA and external archival facilities shall be ensured by the full compliance of the HSA development to VO protocols and standards.

Priority: 1
2.2 Content requirements

HSA-UR-0070 All Herschel data that will be accessible by the Astronomer shall be stored in the HSA.
Data accessible to the Astronomer shall be:
1. SPG Observation products and associated auxiliary data products (1)
2. Quality Control products (2)
3. Uplink and Dowlink calibration products (2)
4. Trend Analysis Products (3)
5. Highly processed products (level 3 products, catalogues, atlases, observer’s processed products) (3+)
6. Proposal data (Title, proposer names, abstract) (3)
7. POS and ICPF (3)
8. AOR Specification (3)
9. Browse products (icon, postcard and browse product) (3)
10. Telemetry product per observation. The infrastructure to generate this product is part of the system architecture, but it is not in the baseline that TM products will be routinely generated and distributed to the Astronomer. (3)
11. Dataframes product per observation: The infrastructure to generate this product is part of the system architecture, but it is not the baseline to provide Dataframe products to the community on regular basis, but only on emergency situations like a failure in the pre-processing step for PACS. (3)

Priority per item indicated

HSA-UR-0080 The following Herschel data that will be accessible by the Calibration Scientist/Instrument Engineer shall also be stored in the HSA:
1. Spacecraft Database (MIB): only downlink part as product
2. Telecommands: TC history product

Priority: 3

HSA-UR-0090 The HSA shall content all versions of products generated with the same SPG version.
Different versions of products generated with different SPG versions shall not be simultaneously stored in the HSA but a backup of older versions must be kept.

Priority: 1
2.3 Data access process requirements

“Data access process” covers querying, query results display, selection of data to be retrieved, and retrieval of data.

HSA-UR-0100 Two levels of access to the HSA User Interface shall exist: one for general users (a simple or basic interface), and one for expert users.

*Priority: 3*

HSA-UR-0110 It shall be possible to switch from the basic to the expert level with a minimum number of steps (e.g., 2).

*Priority: 3*

HSA-UR-0120 It shall be possible to ask for on-demand processing from the HSA to get the products generated with the latest version of the pipeline.

*Priority: 3*

HSA-UR-0130 It shall be possible to disable on-demand processing from the HSA at:

1. Observation level
2. AOT level
3. Processing level

*Priority: 3*

HSA-UR-0140 The HSA User Interface shall be able to send data references to other applications using beforehand-agreed protocols (e.g., Application Intercommunication protocol, Message Brokers, etc. like, for instance, PLASTIC).

*Priority: 2*

2.4 Data query requirements

HSA-UR-0150 The HSA User Interface shall support queries on the following items:

1. Observation ID (1)
2. Observation ID list (1)
3. Observation proprietary rights status (proprietary, public, when it is public) (3)
4. Observation proprietary rights expire date (3)
5. Object. Name resolution should be allowed using SIMBAD or NED, to be selected by the user. (3)
6. Target position (including radius around a central position or box), given in equatorial, ecliptic or galactic coordinates. (3)
7. Target list (2)
8. Astronomical object type (3+)
9. Observation duration (2)
10. Observation date and time (1)
11 Observation OD (1)  
12 Instrument (1)  
13 Instrument Subsystem (1)  
14 AOT ID (3)  
15 AOR label (3)  
16 Wavelength range (2)  
17 Frequency range (2)  
18 $V_{LSR}$ (3)  
19 Proposal ID (3)  
20 Proposal category (3)  
21 Proposal abstract contents (3)  
22 Observer ID (3)  
23 Quality flag (3)  
24 Pointing mode (line scan, raster, fix pointing, others) (3)  
25 Entries in catalogue products (e.g., Point source list, Spectral line list, observer's generated catalogues or atlas) (3+)  
26 Building Block ID (2)  
27 Types of sub-measurements of Observations (to be specified) (3)  
28 Instrument specific data contents (e.g. Fluxes, glitch rates, to be further specified) (3)  

Priority per item indicated

HSA-UR-0160 In the HSA User Interface, it must be possible to combine queries on different items with AND and OR logic.  
Priority: 1

HSA-UR-0170 In the HSA User Interface, Unix-like wild cards shall be allowed in the character fields.  
Priority: 1

HSA-UR-0180 In the HSA User Interface, the following operators shall be available to specify the conditions with respect a certain value: “=”, “=>”, “<=”, “>”, “<”.  
Priority: 1

HSA-UR-0190 It shall be possible to write queries to the HSA through a scriptable interface.  
Priority: 1

HSA-UR-0200 It shall be possible to write queries to the HSA (through the scriptable interface) on a predefined set of standard products metadata keywords. The list of metadata keywords to query on will be specified by the Calibration Scientists/Instrument Engineers.  
Priority: 2
2.5 Data query results requirements

HSA-UR-0210 The user shall be able to select the items to retrieve from the list of query results.

*Priority: 1*

HSA-UR-0220 It should be possible to print or save to a file the text part of a query result.

*Priority: 1*

HSA-UR-0230 Each result of a search in the HSA User Interface shall be displayed with the following information:
1. Observation ID (1)
2. AOT ID (2)
3. Object (1)
4. Target coordinates (1)
5. Distance (from the query input coordinates) (2)
6. Observation OD (1)
7. Start time of observation (1)
8. End time of observation (1)
9. Observation duration (1)
10. Observer ID (2)
11. Proposal ID (1)
12. AOR label (1)
13. Proposal category (1)
14. Astronomical object type (3+)
15. Status of observation (proprietary, public) (1)
16. Proprietary period expired date (1)
17. Image stamp showing the Quick-look observation product (3)
18. A link to a displayable version of the QC Report Summary (3)

Proposal and mission planning data are linked to the observation, viewable and down loadable. The Quality Control Report summary product are both viewable and retrievable in the set of products.

*Priority per item indicated*

HSA-UR-0240 It shall be possible to sort the results of a query by time of the observation, by distance to the position specified in the query, and by proprietary period expiration date.

*Priority: 2*

HSA-UR-0250 It shall be possible to search observations in other ESAC/External Archives using the coordinates of a selected observation.

*Priority: 3*
2.6 Data retrieval requirements

HSA-UR-0260 The user shall be offered the following options to retrieve the selected data from the HSA:
1 Send data references to other applications
2 Download to FTP

Priority: 2

HSA-UR-0270 The results of a retrieval action on the HSA are observation products, auxiliary products and calibration products.

Priority: 2

HSA-UR-0280 The HSA shall provide the user with an indication of the volume of data requested for retrieval.

Priority: 3

HSA-UR-0290 It shall be possible to trace back any retrieval of any data from the HSA.
Monitoring of access to Herschel data is not considered necessary. Note that a distinction between access (read, view) and retrieval (download) is made. In order to trace back, all retrieval requests shall be logged indicating who, what and when. The users will be informed of this policy.

Priority: 2

HSA-UR-0300 By default, only the latest version of data matching the query will be included for retrieval, unless a version number is specified.

Priority: 2

HSA-UR-0310 When retrieving data through the HSA User Interface, the user shall be given the option to select the type of data:
1 All observation associated data, that is, observation products of all levels, auxiliary data, proposal data, quality control data, calibration products
2 Level-0 observation products, and associated calibration, quality control and auxiliary products
3 Level-0.5 observation products, and associated calibration, quality control and auxiliary products
4 Level-1 observation products, and associated calibration, quality control and auxiliary products
5 Level-2 observation products, and associated calibration, quality control and auxiliary products
6 Highly processed products (level-3 products, catalogues, atlases, observer's processed products)
7 A product or products selected from the list of products associated with the observation.
8 AOR specification\textsuperscript{1}
9 Telemetry (see UR-0060)
10 Dataframes (see UR-0060)

\textsuperscript{1}The aim is to have a list of AOR specifications (even from different proposals) that can be read by HSpot.

*Priority: 1, 2, 3, 3+*

**HSA-UR-0320** Data retrieved from the HSA shall have the option to be compressed to facilitate rapid transfer.

*Priority: 2*

**HSA-UR-0330** The originator of a retrieval request that implies downloading of data shall be sent a notification as soon as the data are available for retrieval.

*Priority: 1*

**HSA-UR-0340** HSA User Interface users shall be given the option to know the status of their request.

*Priority: 3*

## 2.7 Security requirements

**HSA-UR-0350** The HSA shall make sure proprietary observation raw data and products are only accessible to and retrievable by the owner of that data during the corresponding proprietary time period.

*The calibration scientists will have access to all data of their instrument, and to proprietary data from other instruments on request to the Project Scientist.*

*Priority: 3*

**HSA-UR-0360** It shall be possible to assign the proprietary time period to observation raw data and products according to the policy guidelines established by the PS.

*Priority: 3*

**HSA-UR-0370** HSA User Interface shall be able to provide the functionality to login automatically at start-up if credentials are given.

*Priority: 2*

**HSA-UR-0380** Proprietary rights information shall be preserved in the interactions of the HSA with external applications using the agreed intercommunication protocols.

*Priority: 3*

**HSA-UR-0390** It shall be possible to assign proprietary rights to individual products.

*The Quality Control Report is a product only accessible (viewable and retrievable) to some expert users involved in quality analysis.*

*Priority: 3*
3 Constraint requirements

HSA-UR-0400  The HSA shall have a user interface that is compatible with generally used graphical user interfaces in astronomy.

*Priority: 1*

HSA-UR-0410  The HSA shall have a response time which is consistent with interactive use. *Nominally the response time should be less than a second. This response time is the time interval between entering a command or pushing a button and getting the action started by the system. It does not include network delays in case of remote access.*

*Priority: 1*

HSA-UR-0420  The HSA shall support a minimum of 30 (TBC by the HSC and ICCs operations plans) remote accesses in parallel for expert users (HSC and ICCs).

*Priority: 3*

HSA-UR-0430  The HSA shall support a minimum of 50 (TBC by the HSC and ICCs operations plans) local accesses in parallel for expert users (HSC and ICCs).

*Priority: 3*

HSA-UR-0440  The HSA shall support 120 accesses for general users (non HSC or ICCs) in any 10 minute period.

*The maximum number of accesses to the ISO Data Archive in any 10 minute period has been 10. The maximum number of allowed simultaneous accesses is 120, but there is no record that it has ever happened.*

*Priority: 3*

HSA-UR-0450  The HSA shall allow several (authorised) users to read a data item in parallel.

*Priority: 1*

HSA-UR-0460  The HSA shall support a smooth transition between mission phases.

*This requirement covers the need for the HSA to support the retrieval and processing of data ingested in a previous phase of the mission (e.g., retrieval and processing of ILT data during the routine phase), and the re-usability of tools in different phases of the mission.*

*Priority: 1*

HSA-UR-0470  All data archived, as well as the latest versions of S/W and calibration files shall be ready to be delivered to authorised users within 5 minutes of a request.

*This requirement assumes a standard data request for a few Gbytes, and data to which no processing must be applied. That is, it excludes the time required for e.g. On Demand Processing.*

*Priority: 3*
HSA-UR-0480  The system shall return the results of a query on product metadata on 10 parameters in less than 5 seconds (one second is desirable).

Priority: 3

HSA-UR-0490  For Trend Analysis, it shall be possible to perform a query on a certain product type and return the data values for up to 10 parameters from a time period of up to 1 week in less than 30 minutes (this is a goal, since the query duration depends on the data volume queried).

Priority: 3
Appendix   Examples of queries

General queries

• Give me the list of public observations of good quality made on M51 with a radius of 5 arcminutes and sort them according to the distance of the nominal pointing.
• Give me the list of HIFI Solar System observations made by a certain observer between ODs 100 and 200.
• Give me all observations of a certain astronomical object type carried out with SPIRE or PACS photometer.
• Give me all observations within a frequency or wavelength spectral range performed with a certain AOT.
• Give me all observations belonging to the same proposal.
• Give me all calibration products generated by a certain observation or by observations performed in a certain time interval.
• Give me all calibration products used in the pipeline for the generation of a certain observation product.
• Get dataframes or level-n products of all observations associated with a certain calibration object (both uplink and downlink).
• Give me all dataframes for HIFI parameter WBS-H where the main flux level in band 1 is higher than X.
• Find calibration products that have a certain instrument configuration and certain pointing parameters (e.g. OFF position in the sky).
• Browse and extract all information associated with this calibration product, e.g. calibration procedure, documentation links, data used to generate it.
• Show me the uplink commands, POS and orbit files of observations with a certain bbid executed during this period of time.
• Browse and extract all observation on a specific target and show AOT used.

SPIRE Trend Analysis queries

• Return BSM rest position once per hour over the whole mission (needs an automated way of knowing that SPIRE is not performing an AOT where the BSM is moved). A query would need to say AOT not X, Y or Z, want BSM choppson and jiggposn, once per hour, start time and end time or give me all.
• Return SUBKTEMP once per minute over any given 24 or 48 hour period.
• Return MCUENG parameter as fast as generated from an engineering observation (very fast!)
• Return SMECLVDTPOSN when SMECENCPOSN is in a given range over a 24 hour period one year ago.
• Return SCAL2CURR and SCAL2TEMP, once per 10 minutes, when SCAL2TEMP is greater than 80K for the whole mission.
• Return all values of BSMCHOPPOSN and BSMJIGGPOSN when BSMCHOPPOSN is greater than 50,000.
• Return all values of PUMPHSTEMP one hour before and one hour after and during each occurrence of the PUMPHSTEMP reaching 20K (extracts cooler recycles).
• Return Flux on one pixel from PCAL flashes for one month/mission (extract derived fluxes).
• Return average temperature of thermistors during each day over the mission (extract all the derived temperatures then average it for the day).
• Return all the values of derived detector parameters from all the AOTs of a given type throughout the mission, at a rate of once per AOT, time plus 288 columns for the photometer file produced, separate time plus 72 columns for the spectrometer file produced.
• Return all products for a given AOT for a given time period.
• Return all level 1 products, (both photometer and spectrometer) for a six month period which match the criteria SUBKTEMP greater than 290 mK.
• Return all channel derived tables (intermediate product which should be stored) for a one month period.
• Return all Small or Large map products for a given RA and DEC range (area on sky).
• Return all high resolution spectrum products where the number of scans is greater than 100.
• Get all interferogram/spectrum products where the interferogram is flagged as being clipped/saturated at ZPD.