Herschel Users' Group

RESULTS FROM THE HUG QUESTIONNAIRE 2014

22-23 May 2014

After the end-of-helium on 29 April 2013, *Herschel* entered its Post-Operations phase (POP), and the *Herschel* Science Archive (HSA) acquired a fundamental role in defining *Herschel*'s legacy. During the year from end-of-helium, the ongoing work by the HSA, the *Herschel* Science Center (HSC), and the Instrument Control Centers (ICCs) has been dominated by further refining instrument calibrations and data reduction algorithms in order to populate the archive with the best data possible. Now, in April, 2014, one year from the transition to POP, the time is ripe for a new questionnaire to ascertain community response to the HSA. Hence, as in 2012, the HUG organized a simple questionnaire that was hosted by Google:

https://docs.google.com/spreadsheet/viewform?formkey=dEJFWWU1TS1fQ1hsSmR2aE9xWXdaMmc6MA

Appendix A gives the content of the 2014 HUG Questionnaire. The questions regarded the frequency of HSA access, the ease and clarity of archive use, the use of User-Provided Data Products (UPDPs), and the importance attributed to features in the archive under development such as more sophisticated database queries and stand-alone data products. Answers were given as ratings (e.g., from 1="very difficult" to 5="very straightforward"), not applicable (rating 0), and a few optional text boxes for additional comments. The invitation to respond was sent to the subset of registered *Herschel* users (total users number roughly 2500 at the time of the survey) who had declared that they would be willing to receive communication from the *Herschel* Science Centre (HSC). A notice was also included in the HSC Newsletter and posted to the HSC main website. The questionnaire response took place from 1 April to 30 April 2014. At the 8th HUG meeting (HUG#8), 22-23 May 2014, we presented the results of this questionnaire.

General overview of results

96 users, or roughly 4% of the number of registered users in the HSC database answered the questionnaire[1]. Although this is a small fraction of total users, and thus there may be selection effects in the statistics, the respondents comprise the astronomical community that is interested in *Herschel* and its legacy. As in the previous 2012 questionnaire (HUG#5) the respondents of the 2014 questionnaire remain anonymous; the HUG did not trace, nor is able to trace, the specific identity of the respondents.

Table 1 gives the mean responses and their standard deviations categorized according to the frequency of HSA access defined by the questionnaire. Table 1 shows only the 11 questions regarding ratings and perceived importance of future developments. The values in parentheses in Columns 2-6 give the number of users who fell into the various frequency categories. Histograms of the responses are shown graphically in Figs. 1 and 2. Also illustrated in these figures is the *global mean and standard deviation,* 3.2 ± 0.4; this is the average rating for the six questions regarding the functionality and ease of use of the HSA (i.e., ease of HSA access, clarity

[1] The HUG 2012 questionnaire was answered by a similar number of users, 103. However, in 2012, the invitation to respond was sent to 338 PIs of accepted proposals (up to and including OT2), rather than the entire *Herschel* community as now for the 2014 version.

of front-end, ease of batch option, data organization, ease of format conversion, completeness of documentation). Some of the categories of responses may suffer from small-number statistics. Nevertheless, the overall impression is that users are generally satisfied.

Table 1. 2014 Questionnaire results: mean ratings, standard deviations (number of responders)

Question (see Appendix A)	Frequency of HSA access					All
	Never	< 1 time/month	1-5 times/month	5-10 times/month	> 10 times/month	
(3) HSA ease of access	3.0 (1)	3.35 1.15 (37)	3.79 1.05 (42)	3.29 0.97 (9)	4.57 0.79 (7)	3.62 1.10 (96)
(4) HSA ease of UPDP access	2.0 (1)	2.27 1.03 (15)	3.06 1.25 (17)	2.33 1.53 (3)	3.67 0.58 (3)	2.72 1.19 (39)
(5) US-based IRSA ease of UPDP access	2.0 (1)	2.64 1.03 (11)	3.50 1.35 (10)	3.00 1.41 (2)	3.50 0.71 (2)	3.04 1.18 (26)
(6) Clarity of HSA front-end startup page	3.0 (1)	3.29 0.96 (35)	3.90 0.89 (41)	3.11 0.60 (9)	3.43 0.53 (7)	3.55 0.92 (93)
(7) Ease of batch option	3.0 (1)	3.10 0.84 (30)	3.49 0.96 (37)	2.33 1.32 (9)	3.33 1.03 (6)	3.21 1.01 (83)
(8) Importance of single photometric/ spectroscopic bands	4.0 (1)	3.91 1.01 (35)	3.90 1.00 (41)	4.33 0.87 (9)	3.57 1.62 (7)	3.93 1.04 (93)
(9) Importance of more sophisticated batch queries	4.0 (1)	4.17 0.88 (36)	3.76 1.23 (42)	4.22 0.83 (9)	4.71 0.49 (7)	4.03 1.05 (95)
(11) Data organization	5.0 (1)	2.65 0.73 (34)	2.77 1.00 (40)	2.38 1.06 (8)	3.33 0.52 (6)	2.75 0.92 (89)
(12) Ease of format conversion	2.0 (1)	3.22 1.05 (36)	2.76 1.09 (41)	2.89 1.05 (9)	4.17 0.75 (6)	3.03 1.10 (93)
(13) Importance of stand-alone data products	3.0 (1)	3.47 1.06 (36)	3.38 0.96 (39)	4.22 0.83 (9)	4.00 1.55 (6)	3.54 1.05 (91)
(14) Documentation available within HSA	3.0 (1)	3.31 0.90 (35)	3.26 1.04 (39)	2.89 0.78 (9)	3.17 1.33 (6)	3.23 0.97 (90)

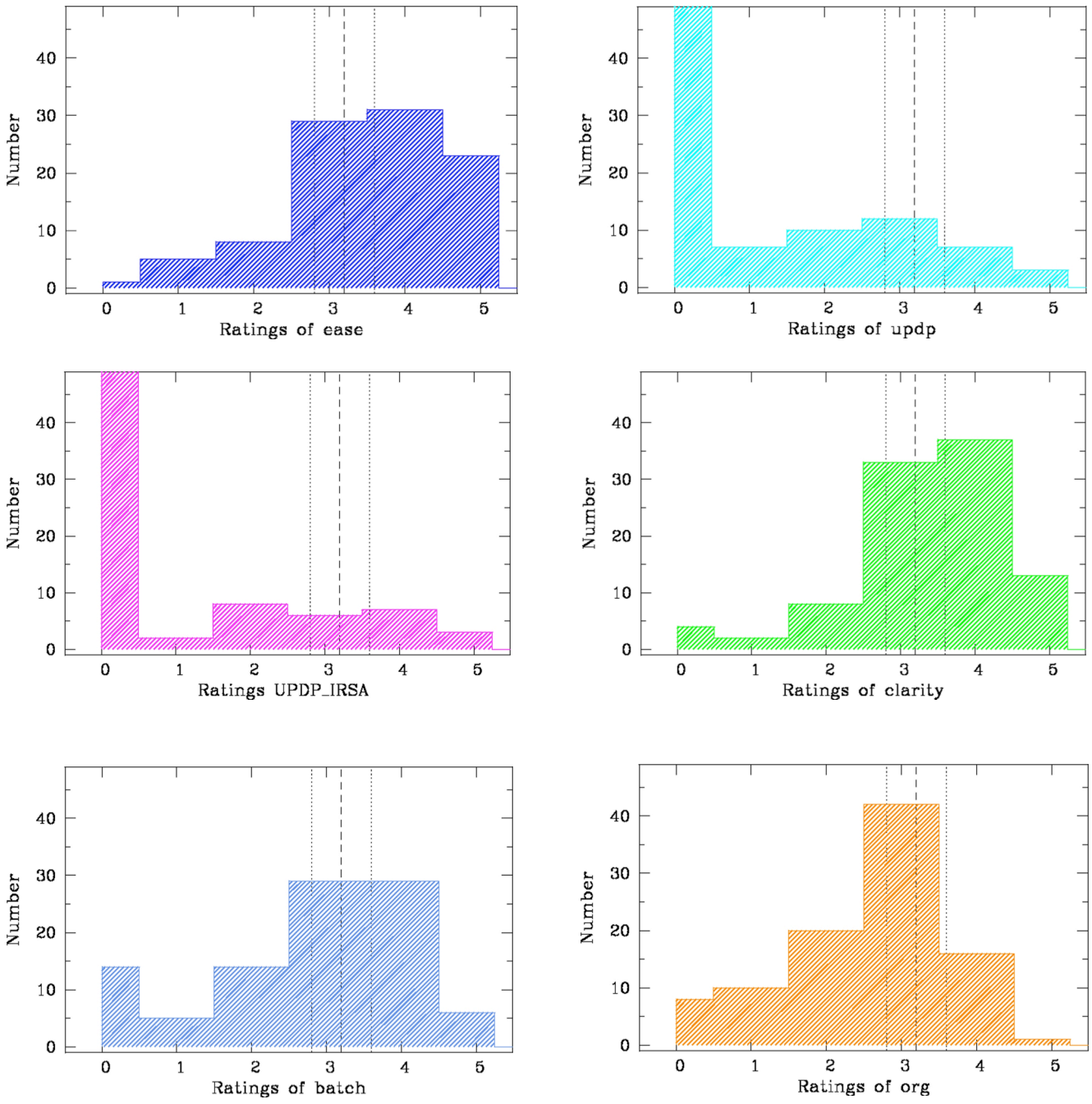

Figure 1. Histograms of ratings of HSA ease of access (upper left panel), HSA ease of access to UPDPs (upper right), IRSA ease of access to Herschel UPDPs (middle left), clarity of HSA front-end (middle right), HSA ease of batch queries (lower left), and data organization after downloading (lower right). These responses correspond, respectively, to questions 3, 4, 5, 6, 7 and 11 in Table 1. In all panels, the global mean rating (3.2 ± 0.4) for questions not regarding UPDP or importance for future developments is shown as a vertical dashed line, with associated standard deviations shown as vertical dotted lines. For the UPDP questions (upper right and middle left), the 0 rating corresponds to "not applicable" (never used).

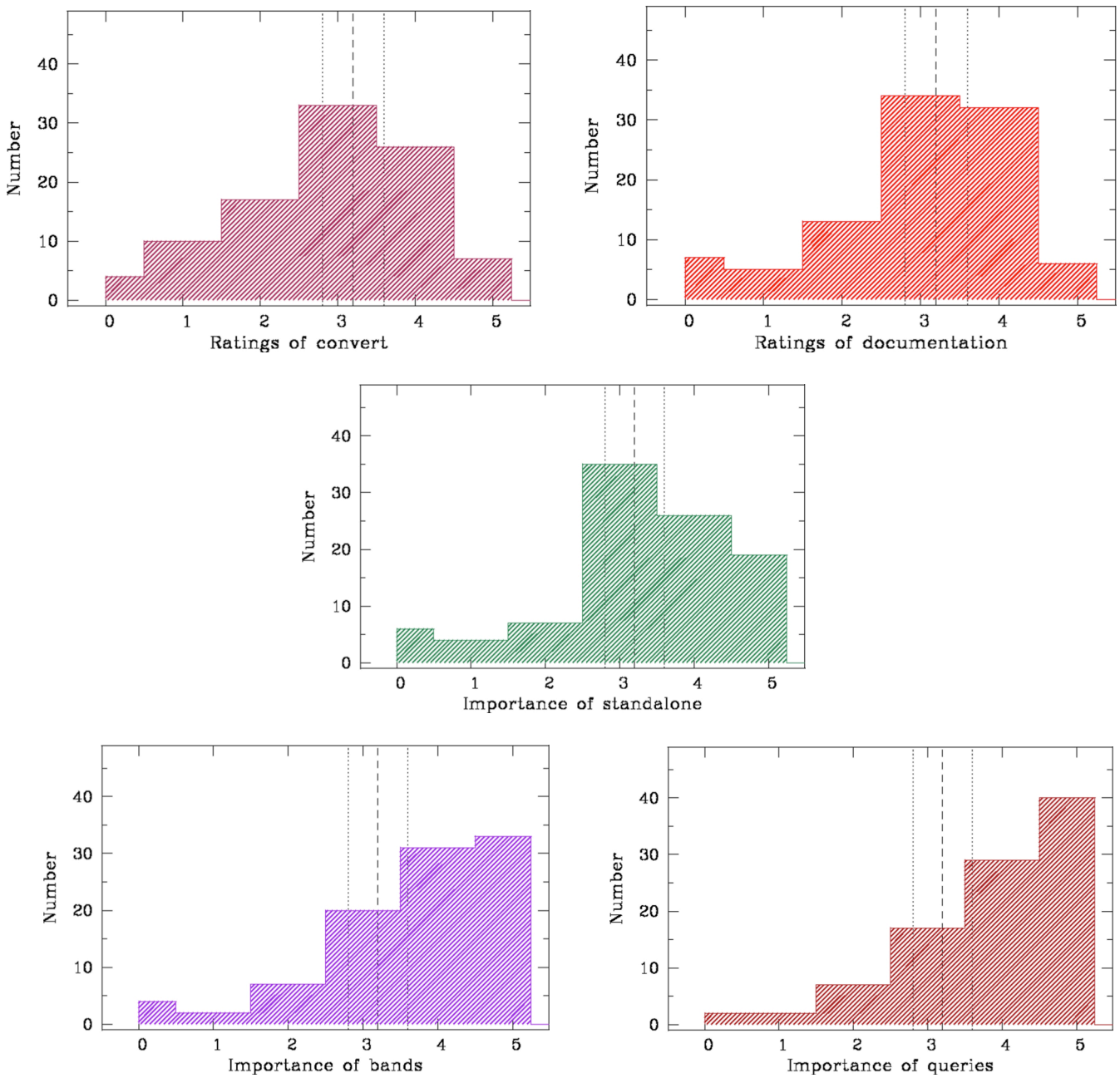

Figure 2. Histograms of ratings of HSA ease of format conversion (upper left panel) and quality of documentation available within the HSA (upper right). The remaining panels regard the perceived importance of stand-alone data products (middle panel), the importance of the possibility to query for individual bands or spectroscopic wavelength intervals (lower left), and the importance of more sophisticated batch queries (lower right). These responses correspond, respectively, to questions 12, 14, 13, 8, and 9 in Table 1. As in Fig. 1, the global mean rating for questions not regarding UPDP or importance for future developments is shown as a vertical dashed line, with associated standard deviations shown as vertical dotted lines.

Preferred data product = stand-alone

Question 10 (see Appendix A) asked users what products they access. These include tarballs (for input to HIPE), stand-alone data products (which do not require HIPE for analysis), the jpeg "postage stamps" available from the Postcard Gallery, UPDPs, and "other". At the time of the questionnaire (April, 2014), barely one month after the HSA v5.2 roll out that enabled access to stand-alone browse products (17 Mar 2014), the most popular products are stand-alone; 69% of the respondents use the stand-alone products while 53% use the tarball. 38% of the respondents use *only* the stand-alone data compared to 18% who use *only* the tarball.

Given the short time span between the questionnaire and the availability of the stand-alone data products (which did not include PACS spectroscopy or HIFI products at the time of the questionnaire), their popularity is perhaps surprising. Indeed, this result strongly supports the efforts of the *Herschel* project scientist, the HSA, the HSC, and the ICCs in making the stand-alone products the best possible. Given current trends, users will favor archival Legacy products that can be further exploited with software packages and tools used by the wider astronomical community.

The jpeg postage stamps are also relatively popular (40% of the respondents accessed them), but they are usually used in conjunction with other forms of downloads. Nevertheless, a small fraction of the users uses only these (5/96). The jpeg quicklook images in the Postcard Gallery are apparently quite appreciated by the average *Herschel* user.

General satisfaction but still room for improvement

Figures 3-5 plot the means and standard deviations for each question as given in Table 1. As in previous figures, the average and 1σ interval (standard deviation) are shown by dashed and dotted lines respectively. In general, the HSA is considered ***easy to access and its front-end interface presents a clear functionality*** (overall means = 3.62±1.10, 3.55±0.92, respectively, see Table 1 and Figs. 3, 5).

Within each question or mean rating, there is some trend of more frequent HSA users giving higher ratings. This is not unexpected given that with any software, habitual usage engenders familiarity and thus ease of execution. A possible example of such a trend is the rating of the ***data organization***. All users except those most frequent (>10 times/month) considered the data organization after downloading to be slightly below "adequate" (overall mean = 2.75±0.92, see Table 1 and Fig. 4); for this question, the alternatives spanned 1 (very bad), 2 (bad), 3 (adequate), 4 (good), and 5 (excellent). This view is reflected also in the comments (see below) where there is a general feeling that the arbitrary naming conventions and directory structure could be improved.

Another example of a rating which gets better the more often the user accesses the HSA is with the ***ease of format conversion***; only very frequent users find it "straightforward" (numerical grade 4.17), while the rest are slightly less satisfied (overall mean = 3.03±1.1, see Table 1 and Fig. 4).

Finally, ***batch queries and the quality of documentation*** within the HSA are considered slightly better than "adequate" (overall means = 3.21±1.01, 3.23±0.97, respectively). There is no clear trend with frequency of HSA access. This would imply that the work by the archive team has been well spent, but there would still be some room for improvement.

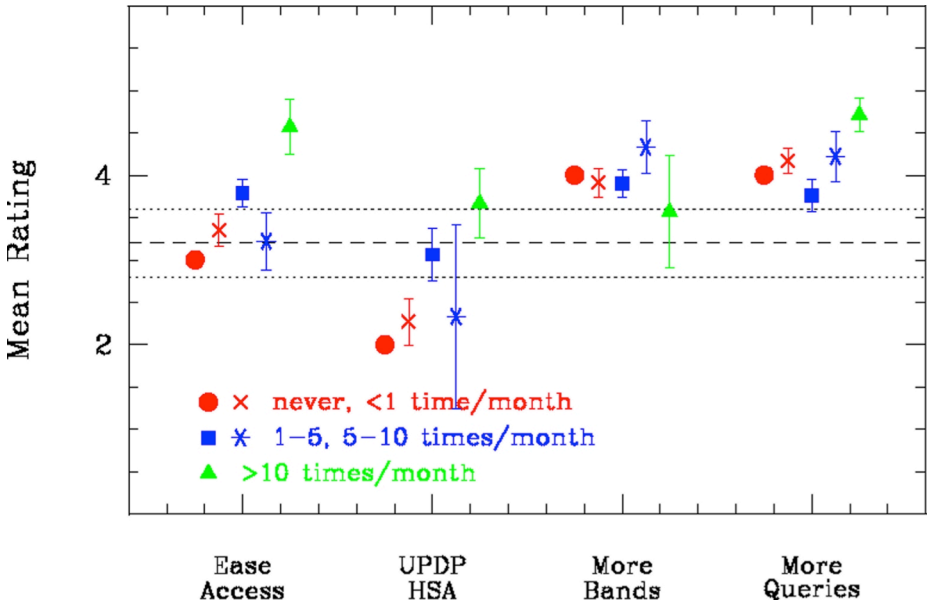

Figure 3. Mean questionnaire ratings for HSA ease of access, the ease of HSA access to UPDPs, the importance of being able to query for bands/spectroscopic intervals, and the importance of more sophisticated queries (questions 3, 4, 8, 9 in Table 1). The different frequencies of access to HSA are shown by different symbols: rarely (never, <1 time/month) shown as filled circles and X's; frequent access (1-5, 5-10 times/month) as filled squares; and six-sided asterisks; and very frequent access as filled triangles. The horizontal dashed and dotted lines correspond to the global mean of 3.4±0.4, averaged over those six questions that do not regard either UPDP or future developments.

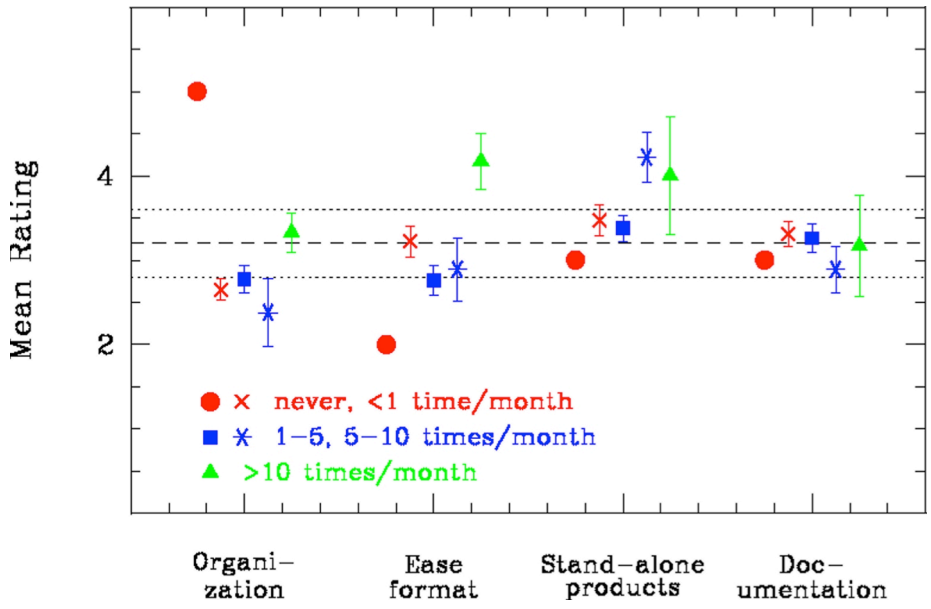

Figure 4. Mean questionnaire ratings for HSA data organization after download, ease of format conversion, the importance of stand-alone data products, and the documentation available within the HSA (questions 11, 12, 13, 14 in Table 1). Symbol coding and meaning of horizontal lines are as in Fig. 3.

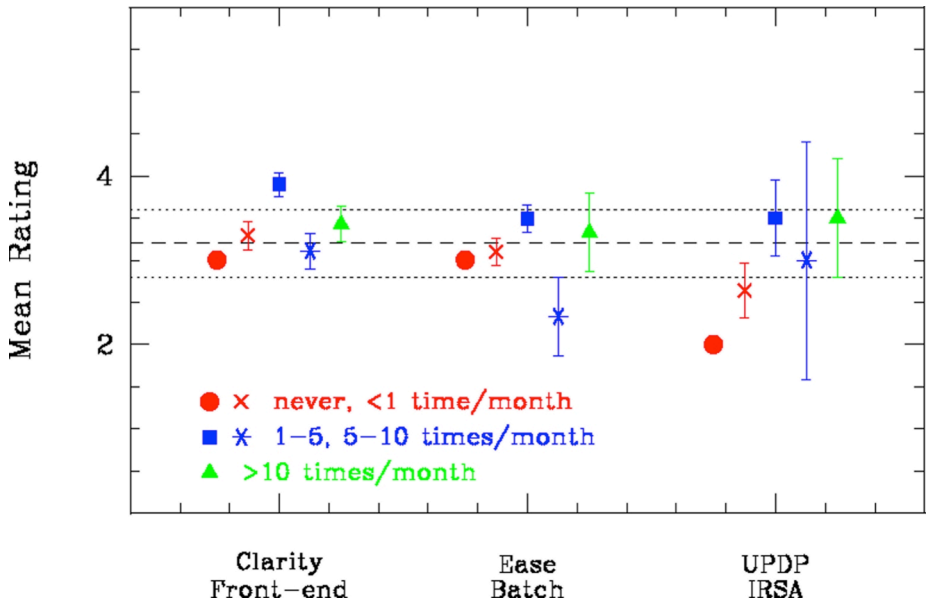

Figure 5. Mean questionnaire ratings for the clarity of the HSA front end, ease of batch queries, and the ease of NHSC/IRSA access to *Herschel* UPDPs (questions 6, 7, and 5 in Table 1). Symbol coding and meaning of horizontal lines are as in Fig. 3.

UPDP access from HSA and IRSA

Two questions regarded the ***ease of access of the UPDPs*** (questions 3, 4, see Table 1). Most of the respondents "never" accessed these data; only 41% used the HSA access to UPDPs and 27% the IRSA access. Ease of access to UPDPs was considered around "average", with a possible trend for the IRSA route to be "easier" than the HSA. Nevertheless, the number of respondents for these products is small, and the statistics undoubtedly suffer from low significance. Specific comments suggest that many users are unaware of the existence of these products; only 16% of the respondents listed the UPDPs among the products accessed. Some effort should be made now, as with new versions of HIPE these products will likely be replaced by better ones.

Importance of improved queries

The questionnaire gives a clear sense that the *Herschel* community places much importance on ***more sophisticated batch queries and the possibility to query the HSA according to wavelength window*** (either image filter, HIFI band, or spectroscopic wavelength interval). ***Stand-alone data products*** are also considered highly important. Much effort has already been spent in this direction, and the improvement of HSA v5.2 over previous versions is evident. Nevertheless, the questionnaire implies that users would appreciate even more improvement, much of which is already in preparation for future HSA releases. Because of the roll out of HSA 7.0 planned for December, 2014, such preparation is especially significant over the next six months which will be crucial for the lasting legacy value of the HSA.

Summary of numerical results

Perhaps the most significant result of the survey is the frequent usage of the stand-alone data products, only recently available from the archive. Such products tend to be preferred relative to the tarballs which necessitate the use of HIPE. Thus the legacy value of the HSA and indeed

the entire *Herschel* mission is intimately connected to making available stand-alone data products of the highest possible quality.

Overall, the most highly rated aspects of the HSA are:

- ✓ Ease of access (overall mean = 3.62±1.10) and clarity of the front-end interface (overall mean = 3.55±0.92). Clearly, the enormous effort spent on the HSA in the two years from the last HUG questionnaire has paid off.

Two aspects of the HSA have roughly "average" ratings:

- ✓ Functionality of batch queries (overall mean = 3.21±1.01) and the quality of documentation within the HSA (overall mean = 3.23±0.97). On the one hand, this is already indicative of the success of the HSA; on the other, it means that more and better documentation of the downloadable products would be desirable. Improvements of the database queries are already planned for future HSA releases.

Two aspects of the HSA have slightly below average ratings:

- ✓ Data organization after download (overall mean = 2.75±0.92) and ease of format conversion (overall mean = 3.03±1.10). The rating of the data organization after download is the lowest of the six facets of the HSA that do not regard UPDPs or importance of future development. There is significant room for improvement in both data organization and format, especially since the ease of format conversion could be related to the importance of stand-alone products and their internal format.

The results of the questionnaire suggest that **the community considers highly important the future developments in the HSA (more sophisticated batch queries, the capability to select in wavelength interval and filter band, and stand-alone data products).** In the POP phase and throughout the long lifetime of the *Herschel* legacy, the HSA must ensure that the 37,141 individual science observations can be easily accessed and that the stand-alone products are of the highest quality. From the last HUG questionnaire in 2012, the HSA was seen to be critical by users; now in 2014 the past development of the HSA has been fruitful and is generally appreciated. Nevertheless, work cannot be concluded until the HSA query capabilities are the best possible and the optimized data products are available in stand-alone form (PACS spectroscopy products are foreseen for HCSS 13 while HIFI spectroscopy products are already available with the release of HCSS 12).

Specific comments

Specific comments were given in response to invitations for further text, and include opinions and suggestions that could not be addressed with the multiple-choice rating system with which the questionnaire was formulated (see Appendix A).

Acknowledgements

The HUG wishes to thank the Herschel Project Scientist, the HSC and the NHSC staff for their input on the questionnaire and the HSC Helpdesk for circulating the survey to the *Herschel* community. Particular thanks are due to the HSA team whose hard work has clearly made the archive a very useful tool. This report has been prepared by Leslie Hunt, Alberto Noriega-Crespo and Dimitra Rigopoulou on behalf of the HUG.

Appendix A. Questions in the HUG 2014 Questionnaire

1) How often do you access the Herschel Science Archive (hereafter HSA)?
2) When did you access the HSA last?
3) Rate the ease with which you were able to access the archive and find what you were looking for.
4) The HSA currently hosts some user provided data products (UPDPs) from Herschel key programs. Rate the ease with which you were able to search for Herschel UPDPs in the HSA archive.
5) The NASA/IPAC Infrared Science Archive (IRSA) currently hosts some user provided data products (UPDPs) from Herschel key programs (http://irsa.ipac.caltech.edu/Missions/herschel.html). Rate the ease with which you were able to search for Herschel UPDPs in the IRSA archive.
6) Rate the clarity of the front-end startup page for the HSA query.
7) Rate the ease of use of the batch option for searching the HSA for lists.
8) How important would it be to be able to access single photometric or spectroscopic bands (e.g., PACS) in the HSA query?
9) How important would it be to have the possibility to perform more sophisticated queries, e.g., coordinate regions (boxes defined by 4 vertices) combined with a single PACS, HIFI, or SPIRE wavelength/spectral mode?
10) What products did you access? [e.g., standalone products , entire tar library, jpeg postage stamps, UPDP (User Provided Data Products), other]
11) Once a target is selected and the data downloaded from the HSA, how would you rate the current organization (e.g., the sub-directory structure and naming protocols) of the data assuming that they are used outside of HIPE?
12) Did you find it easy to convert the data to your preferred format?
13) How important would you rate the availability of standalone products accessible directly from the query results (i.e., to be able to avoid selecting them from a choice in Retrieval option)?
14) Rate the documentation on the files accessible in the HSA.

The questionnaire also included options for the respondents to add comments on almost all of the questions. We received specific input for the ease of use of the HSA, the usefulness and ease of access of the UPDPs, additional options for HSA queries, product nomenclature, data format and organization, and, of course, general suggestions for the HSA. The comments are entirely consistent with the overall statistical results of the survey, and have been shared with the HSA team in their entirety.