**Lessons Learned: The Herschel Science Archive (HSA) - Moderated by David R. Ardila**

The lessons learned regarding the HSA fall within three broad areas:

1-The Science Archive Team (SAT) and the HSC
2-Management of the HSA within the HSC
3-Accessing Herschel Data

**1- The Science Archive Team (SAT) and the HSC**

**a - Lesson: Two SOC-SAT interfaces required for archives in operations: Archive Scientist and System Engineer.**

The HSA was an integral part of HSC and Science Operations, being a fundamental for all tests prior to the mission launch, and used everyday during operations to support other SOC subsystems. However, the Project Management Plan of the SAT did not include a human interface with the operations system engineer. An important lesson learned is that for archives to be used in operations, two human interfaces SOC-SAT are needed: 1) the traditional Archive Scientist, in charge of gathering the user needs of the archive from the scientific point of view; 2) the SOC System Engineer, in charge of gathering the needs of the archive from the SOC-subsystems point of view.

**b - Lesson: Weakness of loose data model for the archive.**

The HSA has been suffering the lack in definition of a proper data model of the data and metadata produced by Herschel mission; in the future, an overall conceptual data model of the different subsystems, including the data, should take higher priority and be tackled early on in the missions' initial phases.

**c - Lesson: The challenges of having an archive within the operational chain.**

The fact that the archive is part of the operational chain makes the performance and reliability key aspects. This implies the archive must be running in a 24 x 7 basis and that every downtime must be carefully planned. The archive has to process a huge number of petitions and it needs to have monitoring tools in order to know the health of the system and report it to other teams.

**d - Verification tests should include external applications.**

The archive should be tested against external clients and DP applications.

**e - Clear and documented interfaces are crucial.**

Thanks to a clear, brief and well-documented ICD, the archive is fully decoupled from HSC applications and libraries. It is efficient to involve in the normal archive activities not only the experience of the Archive Scientist but also all the project teams making use of the archive.

**2-Management of the HSA within the HSC**

**a - Lesson: An explicit and open process is necessary to determine scientific improvements to the Archive.**

Within the project, requests for improvements to the Archive were managed by informal communication with the Archive Scientist. This resulted in an opaque prioritization process, which was reactive and not proactive when it came to improvements, which dumped too much control on a single person, and which resulted in "single point failures" when it came to upgrades.

To this day, there remain high priority updates that were already available in previous missions (band searches, for example) that are not available for Herschel. Here we should understand that some of the work that needs to be done for the archive takes place within HSC (+ICC+NHSC) and some within SAT. The latter is under pressure to respond to the challenges of having an archive in the operational chain.

The Archive Scientist needs to be supported by an advisory group that establishes priorities and a long-term vision. This advisory group can serve as configuration control board. An internal ticket tracking system needs to be established within HSC, or access to the SAT ticketing system needs to be provided.

**3-Accessing Herschel Data**

**a - Lesson: For the end user, minimize the number of data retrieval methods**

Getting the data from the archive to the disk remained problematic for most of the Herschel mission as there were too many data access methods, too many steps, and different recommended methods among the instruments.

As the mission progressed, it became harder and harder to implement change and the situation of the multiple access methods remains unsolved. This is a case in which 'more' is not 'better.'

**b - Lesson: Data Management methods within the DP software should conform to the practices of the community using the software**

The Herschel software adopted the concept of pools, storages, and the building of a local database to store the data locally. These concepts proved challenging to the users as they are not part of the general culture of astronomy. The system has not

been exploited by the common users as designed: little pool sharing, little use of versioning, little in terms of setting local databases.

At this point in the Herschel mission (roughly mid-way) we remain divided internally as to the wisdom of these data management tools. Clearly, however, the astronomical community has already voted overwhelmingly against them.
Some of the problems in convincing the community had to do with the fact that early versions of HIPE had problems inspecting and organizing pools. Essentially, we gave astronomers an immature tool that fulfilled a need they did not have.