Popular (Enhanced) Products: Data Flow, Database Design, SSC Archival and IRSA Integration

Version 1.0

Frank Masci
April 7, 2005

Purpose

Popular products are defined as *Spitzer* data products that have been enhanced by the user community, for example by a Legacy Science Team (LST). This document outlines high level requirements on file delivery formats, storage in the SSC archive, loading of *meta*-data in database tables, and integration with the Infra-Red Science Archive (IRSA). ***This document is very preliminary as we have just started to scratch the surface on how to store and manage enhanced data products.***

High-Level Requirements (Some of these are currently posed as questions)

1. We will want to retain the original, native directory structure in the ZIP file(s) provided by the LST.
2. Each delivery from a LST shall provide in a README file a list of primary FITS files (e.g., mosaics) which represent distinct fields or observations. Each *primary* file shall have associated with it an ID numbered from from 1..*N*, where *N* is the maximum number of primary files in the delivery. We will refer to this ID as the *primary file ID* (PFID).
3. The LST shall also provide a list of the *ancillary files* or *ancillary directories* associated with each primary file in their directory tree?
4. The Spitzer Proposal ID (PID) shall be used to distinguish/classify each LST set of enhanced products.
5. Master copies of enhanced data products from all LSTs shall be permanently stored in the SSC science archive under a partition distinct from where regular SSC pipeline products reside, e.g., under /archive/enhanced. See Archive Structure plan in Figure 2 below.
6. This new archive partition shall be backed up routinely along with other partions.
7. Archival and metadata table loading software shall be able to replace an entire delivery with a new version. Old versions will be deleted and it is assumed that the current available version is the latest and best version for public distribution.
8. Archival and metadata table loading software shall be able to replace single files at any time, e.g., either due to file corruption or updates.
9. Archival and metadata table loading software shall be able to add new product files to existing deliveries.
10. Any update to an existing delivery for a given PID, or replacement thereof shall be tracked via an SSC (internally) managed version identifier.
11. There will be three metadata tables: one for primary FITS files (e.g., “EnhancedPDPs”), one for ancillary files associated with each primary file (e.g., “EnhancedADPs”) and another exclusively for source catalogs.

12. The EnhancedPDPs and EnhancedADPs tables will be cross-referenced via the LST PID and the primary FITS file ID (PFID; see requirement 2). Source catalog metadata tables will be referenced just by the LST PID. This is because delivered catalogs are likely to cover multiple PFIDs. Once loaded, the source catalogs can in principle be cross-referenced with the image metadata via WCS positions.
13. The schema for these tables will adhere closely to the format used/required by IRSA for the image and catalog metadata tables. There will be additional columns specific to Legacy enhanced products (see requirement 14).
14. The following additional table columns are envisaged (still TBD): Enhanced data product FITS filename, specific field name for primary product?, delivery date, archive date, *Spitzer* instrument/channel, SSC internal version ID, path/filename of primary documentation and/or README files, file-type or flavor of product - e.g., this should transparent in each ancillary filename or directory name in the directory tree?
15. The tables defined in requirement 11 will reside in the SSC SODB to facilitate advanced (SSC-internal) cross-referencing with specific AORs or raw products through the LST PID. These tables will also be replicated to the archive database to provide visibility by SSC staff.

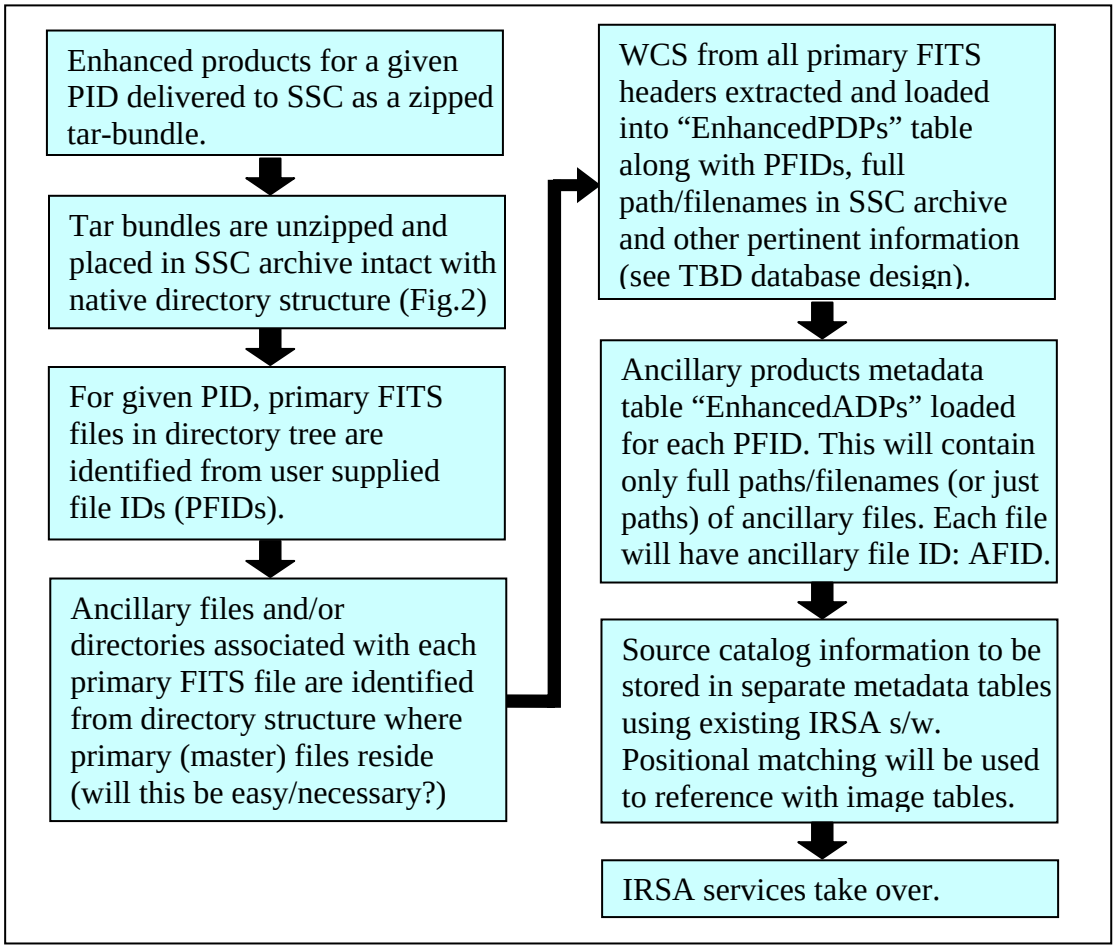

Figure 1. Operational flow of enhanced data products.

Archive Structure

Below is a schematic of the very top level of the Legacy enhanced products area in the SSC archive.

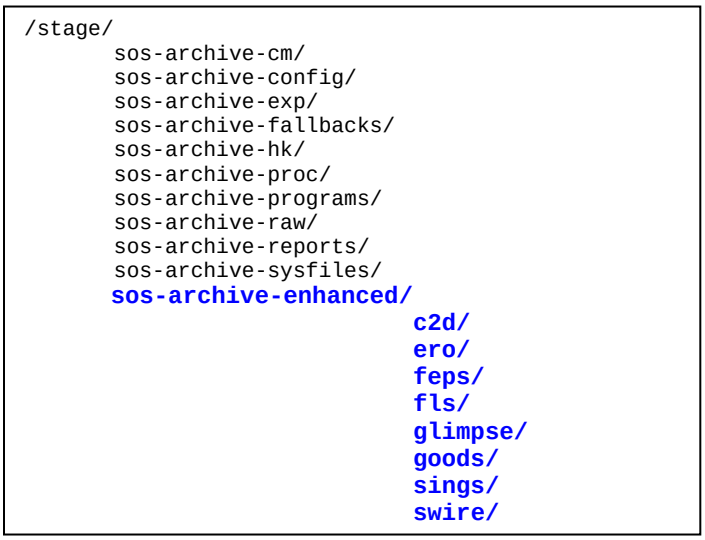

Figure 2. Archive structure plan

Database Design

To adhere as close as possible the existing IRSA metadata table design for serving generic SSC products (TBD). There will be additional column fields specific to Legacy enhanced products (e.g., see requirement 14 above). This section will contain the schema once more is known.