Associations of NICMOS and STIS Exposures: an extension of the HST data processing pipeline

John W. MacKenty and Stefi Baum
February 19, 1996

ABSTRACT

Both of the Science Instruments scheduled for installation into HST during the Second Servicing Mission pose new expectations and requirements on the processing and archiving of their data. In this ISR we discuss a central aspect of the data handling for the NICMOS and STIS SIs -- the association of exposures. The association concept permits the use of multiple exposures in combination to produce data products. Examples include the identification and rejection of cosmic ray events, the wavelength calibration of STIS spectra, and the measurement of the background for thermally limited NICMOS images.

1. Introduction

Motivations

The basic or "atomic" element within the HST ground system has been the "exposure." The first generation HST Science Instruments (SIs) are commanded to generate single exposures, which result from a recognizably distinct sequence of commands to the SI, leading to the creation of a flow of data which is assembled into a single "dataset". Each dataset is given a unique 9 character identifier (an "ipppssoot" in STScI terminology) and is pipeline processed, calibrated, and archived separately from all other datasets.

An illustrative (partial) counter example to this procedure is the WFPC CR-SPLIT proposal instruction. This results in the execution of two WFPC exposures from a single line on the exposure logsheet (the observers interface to the process of commanding HST). However, the HST ground system treats a CR-SPLIT as two distinct exposures which are commanded, processed, calibrated, and archived separately. The pipeline does not perform the task of combining these two images (datasets) to create the single image without cos-

mic ray events *which was the observer's original intention*. Presently, the observers (and any future archival researchers) are left to perform this task on their own.

The second generation instruments present many instances in which the combination of data from two or more exposures is necessary to create a scientifically useful data product. The NICMOS and STIS SIs will need to combine exposures to remove cosmic rays and to improve flat fielding (by dithering or "stepping"). The STIS SI needs to combine nearly every spectrum with a "wavecal" observation since the repeatability of the grating wheel mechanism would otherwise limit the capability of the instrument. For NICMOS, the HST thermal background is expected to have significant temporal variations. Multiple exposures (dithered for small targets and offset onto blank sky --chopped-- for larger targets) will be necessary to measure and remove this background. While this has been standard practice for ground based infrared observations and is the basis of essentially all existing infrared data reduction schemes, it is a new paradigm for the HST ground system.

System Design Process

The need to combine multiple exposures during the data calibration phase of the HST pipeline was recognized early in the process of defining the pipeline changes necessary to support the Second Servicing Mission. A "systems engineering team" under the leadership of SMO with representatives from the commanding, OPUS, STSDAS, and archive groups was established early in 1995 to work through the many cross-system aspects of the SM97 pipeline, including the issue of associated exposures. A Preliminary Design Review, conducted in March 1995, described the philosophy of associated exposures and described the basic implementation scheme. Following the Preliminary Design Review in March, a 'SM97 Keyword Working Group' was established.[1] The KWG has conducted a detailed design of the implementation methodology for combining exposures into *associations*. This ISR explains this methodology and its implications for the various elements of the HST ground system. This ISR presents the view of the HST observer and is intended to communicate to the *users* of associations how they will work. It is outside the scope of this ISR to explore all of the implications of associations for the HST ground system or the details of their implementation with the various databases in OPUS and HDA. The KWG has made a concerted effort to assure that the design of associations will work with all of the elements of the HST pipeline, that the necessary information is passed at each interface, and that the view of associations presented to HST observers both workable and comprehensible. The technical details of the ground system implementation are contained in a set of memos and OPRs (Hopkins 1995, Swade, 1995, Rose 1995, Kutina 1995).

---

1.   The Keyword Working Group is led by Daryl Swade and includes Stefi Baum, Carl Biagetti, Ed Hopkins, Ray Kutina, Karen Lezon, John MacKenty, Jim Rose, and Mary Alice Rose. Ed Hopkins has taken the lead on defining the association keywords.

2. Usage of Associations

Associations exist to simplify the use of HST data by observers. This starts from the proposal phase, continues with a more complete calibration process than would be possible in the absence of associations, carries on into the archiving and retrieval of associated data, and includes the use of HST data by observers within the STSDAS system.

From a high level, an association is a means of identifying a set of exposures as belonging together and being, in some sense, dependent upon one another. The association concept permits these exposures to be calibrated, archived, retrieved, and reprocessed (within OPUS or STSDAS) as a set rather than as individual objects. In one sense, this is a bookkeeping operation which is being transferred from the observer to the HST data pipeline and archive.

Associations are defined by optional parameters on a single exposure logsheet line. That is, there is a one to one correspondence between proposal logsheet lines and associations (although it is possible to have exposures which are not in associations).

NICMOS

Observers may obtain one or more exposures at each of one or more position on the sky using the NICMOS proposal grammar. Typically usage will be:

*   to obtain a sequence of slightly offset exposures (dithering) to improve the flat fielding, avoid bad pixels and cosmic rays, and, for sufficiently compact targets, to remove the background illumination
*   to obtain a sequence of observations in which the beam is chopped between the target and one or more offset regions of (hopefully blank) sky.

A set of pre-defined patterns are provided in the proposal instructions for these types of observations or a combination of both types. The TRANS system will expand the observers single line request into multiple exposures each with its own identifying name (ipppssoot) and populate the necessary database relations to describe this association for the OPUS system.

STIS

To date, STIS associations have three uses:

*   to allow pipeline combination of crsplit CCD exposures producing a cosmic ray rejected output
*   to allow pipeline calibration of the zero point of the wavelength scale and aperture location on the detector, via the use of a wavelength calibration spectrum taken contemporaneously with the science data
*   to keep in a single file multiple time resolved images/spectra (so called 'repeatobs' observations).

STIS associations and file structures are explained in more detail in STIS ISR 95-006 (S.A. Baum and P. Hodge).

3. Associations as a Data Structure

Common Design

An association is a set of one or more exposures along with an association table and, optionally, one or more products. While NICMOS and STIS utilize different modes of packaging their data, the logical structure of their associations is identical. (We believe that WFPC2 and ACE could adopt this association design without modification.)

It is important to define some terms rather precisely:

- An *exposure* is the atomic unit of HST data.
- A *dataset* is a collection of files having a common rootname (first 9 characters).
- A *product* is a dataset derived from one or more exposures.

The first generation SIs all have a one to one correspondence between exposures and datasets. They do not have products. NICMOS and STIS use the association structure as a meta-dataset. Further, they use the information in multiple exposures during the calibration process to create products.

OPUS Flow and Elements

The present sequence of post-observation processing consists of:

1. Data Partitioning
2. Data Editing
3. Data Validation
4. Generic Conversion
5. Calibration
6. Archiving

To support associations, a new step is added to the data processing between Data Validation and Generic Conversion: *Data Collection*. This step assembles all of the exposures which constitute the association and holds further calibration processing until all exposures are present.

NICMOS

NICMOS calibration is divided into two stages ("A" and "B"). The first stage (CALNICA) calibrates individual exposures without information from other members of the associa-

tion. It is basically equivalent to the WFPC2 pipeline. Starting with the "RAW" datasets produced by Generic Conversion, CALNICA creates "CAL" datasets on a one to one basis. The second stage of calibration (CALNICB) creates combined images (products) for each distinct (non-overlapping) pointing of HST within the association. These products have names derived from the association name and the suffix "MOS".

STIS

STIS assembles all exposures from an association to create a single dataset with the name of the association as the output from generic conversion. The various STIS calibration modules all operate on this single dataset which contains all exposures in the association.

4. Packaging

There are two forms of "packaging" of data which must be considered: logical and physical. The first generation SIs group all of the information associated with an exposure into a logical package under a single rootname (ipppssoot) called a dataset. The physical packaging of this information is contained in a (generally rather large) number of files where each file contains only one type of object. For example, a single WFPC2 observation yields a single logical dataset with separate files for the uncalibrated data (.d0f, .q0f, .x0f, .q1f, .shf, .trl), and the calibrated data (.c0f, .c1f, .c3t), and supporting files (...). If these files are converted from FITS to GEIS format, each one consists of a header (e.g. .d0h) and a data (e.g. .d0d) file. A simple FOS observation can result in more than 25 files per dataset.

An important starting point in the design of associations was to greatly reduce the number of files generated from each exposure and simplify the use of those files by the GO. At the same time, it is not desirable for files to become too large since data transfer over the internet becomes more difficult and, if the contents of the file must be changed frequently, the processing may become rather inefficient.

NICMOS

NICMOS packages each exposure into a single dataset. Each NICMOS exposure results in a dataset consisting of three files (plus OMS and OPUS reporting files). There is a SPT file containing the standard header packet (SHP) and unique data log (UDL) data, a raw data file (RAW) and a calibrated data file (CAL). The second stage of NICMOS calibration (CALNICB) creates one or more mosaiced data files (MOS). These are the "products" of the association. The RAW, CAL, and MOS files have a common format and contain image, error, data quality, samples, and integration time arrays as defined in NICMOS ISR-002 (MacKenty et al. 1995). These files are all FITS files with image extensions. A description of the association is contained in an association table (a FITS table file).

This design is motivated by the expectation that observers will frequently want to modify the contents of an association by the deletion, addition, or editing of the individual expo-

sures which constitute the association. This is easily accomplished by modifying the association table and (if appropriate) the files containing the changed individual exposures. A typical NICMOS association may contain images of distinct regions of the sky and therefore will create distinct products with unique dataset names.

STIS

STIS packages all of the exposures in an association into a single dataset. The basic components of a STIS dataset as produced by generic conversion are: the SPT file containing the standard header packet and unique data log data, the raw science data file (SCI) containing all the associated science exposures and, a wavelength calibration file (WAV) containing the raw wavelength calibration lamp exposure. A description of the association is contained both in the image headers and in an association table (a FITS table file). As for NICMOS, all STIS files are FITS.

This design is motivated by the expectation that the elements of a STIS association rarely change and are tightly bound together. The grouping of the entire association under a single dataset name follows from the understanding that the dataset represents a pointing at a single HST target and contains a primary component (e.g. the spectra) and a subordinate component (e.g. the wavecal). The packaging of associated STIS exposures into a single dataset allows the simplification for users that the dataset is a complete unit, with easily recognizable parts and all science data collected in a single file.

5. Keywords

An association consists of one or more datasets and an association table.

Data File Header Keywords

ASN_ID -- contains the unique identifier assigned to the association by the HST ground system. If the dataset is not a member of an association, this keyword will be set to "NONE." The value of this keyword is in the form "ipppss_nn" where:

    i = N (for NICMOS) or O (for STIS)
    ppp = Program ID using the current base36 system
    ss = Obset ID using the current system
    nn = Association ID incremented for each association in the Obset.

ASN_TAB -- contains the name of the association table. Observers may replace this with the name of their own table to reprocess with different exposures for NICMOS (for STIS this table is a permanent record of the planned exposures). The possible values of this keyword are:

    Npppss_nn_ASN.FITS -- Generic Conversion generated association table
    Npppss_nn_ASC.FITS -- Updated association table generated during calibration, produced only for NICMOS

' ' -- If the dataset is not a member of an association.

ASN_STAT -- Status of the association. An association is "complete" if all of its exposures were received; otherwise it is "incomplete."

ASN_MTYP -- The type of the member of the association. This refers to the entire dataset for NICMOS and to individual image extensions within the file for STIS (e.g. CRSPLIT, SINGSCI, RPTOBS, WAVE).

ASN_ORPH -- The dataset contains one or more exposures which were intended to be part of an association but were processed separately.

Association Table Header Keywords

ASN_ID -- same as data file headers

ASN_STAT -- same as data file headers

ASN_PROD -- Status of the generation of products from the association. If all products are present its value is "complete"; otherwise it is "incomplete."

Association Table Require Columns

memname -- name of association member (includes both exposures and products)

memtype -- type of member

memprsnt -- flag to indicate if member if present

These keywords are populated by Generic Conversion for both STIS and NICMOS. For NICMOS, CALNICB updates the association table to indicate the successful creation of the products. CALSTIS does not operate on the association keywords.

6. Archive Considerations

Screens and User Views

Associations simplify the storage of observations in the Hubble Data Archive (HDA) and the visibility of these observations to users. There are two general rules for NICMOS and STIS data in the HDA.

• Each exposure results in an entry in the Instrument Table.

• Each product from an association (or an un-associated exposure) results in an entry in the Science Table.

The HDA will also maintain an Association Table based on the contents of the association FITS tables created in OPUS. This will provide a cross reference between the association ID and its constituent exposures and products.

The user will therefore be able to search the Science Table for observations of particular targets or positions on the sky. For example, if NICMOS chops between a target and an adjacent region on the sky, two products will be created during calibration and each will result in an entry in the Science Table.

For information related to the operation of the SIs (primarily for internal STScI use) the complete detailed Engineering information on each commanded exposure will be available in the Instrument Table.

For users who need to re-calibrate their data, or in the case of reprocessing, the inclusion of the Association Table enables the identification of all of the component parts of an association.

Retrieval Requests

Three types of retrieval requests are supported.

- Retrieval of Product (as identified from the Science Table). This is the most common type of retrieval and should be sufficient for users who do not need to re-calibrate the data.

- Retrieval of all of the members of an association given the name of a Product. This is the same as (1) for STIS. For NICMOS this results in the retrieval of all of the products and all of the individual exposures which make up the association.

- Retrieval of all of the members of an association given the name of an exposure. This results in the retrieval of all of the products and all of the individual exposures which make up the association in whatever form they are packaged at the point of ingest into HDA. External users should not need this capability.

7. Exceptions

When OPUS gets an exposure which is a member of an association it will wait a pre-defined period of time for all the members of that association to arrive. OPUS is nearly always expected to receive "COMPLETE" associations. However, rules must exist to allow the smooth processing of the data in those rare cases when an exposure from an association is lost or delayed beyond the time out period. The KWG has adopted a philosophy that assures the smooth operation of OPUS in those cases. The design allows for the repair of such INCOMPLETE or broken associations in a reprocessing pipeline.

Rules for OPUS Processing of Exceptions

The adopted design for handling exception conditions is identical for STIS and NICMOS and conforms to the following set of rules:

- The status of an association is set by the first exposures processed within the time-out period for that association, and the association table is created at that time. If not all exposures are present (the exception condition considered here), then ASSN_STAT is set to INCOMPLETE.

- Whenever OPUS cannot generate a *product* from an association, OPUS will process the individual exposures through as single exposures (i.e., as "orphans"). Orphans retain their original IPPPSSOOTs. They are cataloged fully in both the HDA science and instrument tables as well as the HDA orphan table. They are not cataloged in the HDA association table.

- Exposures received subsequent to the original processing of the association will be processed as single exposures (orphans).

The ramifications for NICMOS and STIS users are as follows.

NICMOS

If the ASN_ID keyword is not equal to "NONE", then the exposure or product was intended to be part of an association. NICMOS follows the simple rule that associations either succeed completely (all data were received and processed successfully) or fail and are then archived as if they were single (un-associated) exposures. If only some of the expected exposures were received during the time-out period, then ASN_STAT will equal "INCOMPLETE" and ASN_ORPH will be "T" for every member of the intended association. These data will not be processed with CALNICB (therefore no products will be created) and each exposure will be archived as if they were obtained as individual exposures. The association table will be saved to document the intended association and to facilitate potential future reprocessing. If all of the exposures were received but it was not possible to perform pipeline calibration, then ASN_STAT will equal "COMPLETE" and ASN_ORPH will equal "F" but ASN_PROD will equal "F". The exposures will again be archived individually. CALNICB will assure that in all cases it will create all or none of the products for an association.

STIS

If the user receives a dataset with ASN_ID not equal to "NONE", then that dataset contains exposures which are part of an association. In the overwhelming majority of cases, the association will have processed through the pipeline as a unit and ASN_STAT will equal "COMPLETE" and ASN_ORPH will equal "F". When exceptions occur (e.g. all exposures have not been received within the time-out period), the diagnostics for the user will be as follows. If ASN_STAT is "INCOMPLETE", then not all the members of the exposure were received when the association was first processed. If ASN_ORPH is "F", the dataset still contains multiple exposures which were processed as a product, if ASN_ORPH is "T" the dataset contains a single exposure which has been processed through by itself.

3. DATA REDUCTION. Unlike the NICMOS instrument with well-constrained lters as in the ALICE project for archival analysis,38 the STIS coronagraphic imaging mode has a broad bandpass (âˆ¼ 2000ËšA â€" âˆ¼ 10, 000ËšA).51 Therefore in classical. Â§This is an empirical upper limit we set: beyond 5 pixel the relative movement of the star and wedge will create signicant inner working angle change, which causes PSF change consequently.Â Figure 7: Three circumstellar disks examples in the STIS archive processed with KLIP, the disk morphologies are partially changed from their previous discoveries because of the over-tting as expected.Â More exposures of the PSF stars matching the color of GJ 809 is needed to better reveal its possible circumstellar disk. NICMOS Calibration Pipeline: Processing Associations of Exposures. January 1997. Project: NICMOS calibration. Howard Bushouse. J. MacKenty.Â The processing of these associated images presents new challenges to the HST data processing pipelines, which heretofore operated only on single images. In this paper we describe the concept of associated observations and how they will be processed collectively by the NICMOS calibration pipeline. Discover the world's research. 17+ million members. and Data Handling unit Flow of science data in the Hubble Space Telescope Flight Support System configuration Flight Support System Berthing and Positioning System ring pivoted up with Telescope berthed Orbital Replacement Unit Carrier Second Axial Carrier Portable Foot Restraint. Page.Â Space Telescope Imaging Spectrograph (STIS). Weight Dimensions. 825 lb (374 kg) 3 x 3 x 7 ft (0.9 x 0.9 x 2.2 m).Â The roles of NASA centers and contractors for the fifth phase of the HST program, on-orbit servicing, are: (1) GSFC â€" Responsible for overall manage-ment of daily on-orbit operations of HST; development, integration and test of replace-ment hardware, space support equipment, and crew aids and tools.