AIA Data processing and Distribution: from Telemetry to Science data

AIA and HMI impose novel requirements on the data processing and distribution. The volume of data places constraints on the frequency of reprocessing and redistribution of data. The uniform character of the data products permit users to access data by "observing the archive" -- making data requests for specific location resolution, cadence, etc. much like the observing plans for previous high-resolution instruments such as TRACE and Hinode/SOT. Here we layout the plan for data processing and distribution that combines these two elements. When referring to AIA data, we include HMI intensity and longitudinal magnetograms, since they are integral parts of the AIA investigation.

AIA data levels

Data levels for AIA are shown in Table 1. Raw telemetry is converted into Level 0 image files. All possible pixel-level calibrations are applied to these to generate Level 1 data. FITS files containing lists of bad pixels and cosmic ray hits are created and the images are repaired using the iterative method used by the TRACE browser software. The original values of the pixels and their positions are saved in the cosmic ray file for possible reversal. All higher products are based on Level 1 data. Levels 1.5 and 1.6 have all geometric corrections with all images sharing common plate scales and centers and rotation angle from Solar North. In addition, Level 1.6 has a quick FFT-based PSF correction to provide the cleanest images for browse products.

Level	Processing	Notes
TLM	Capture	Raw telemetry
0	Depacketized	Raw images with housekeeping, Level 0 Keywords
1	Overscan removal	
	Flat fielding & Dark current	Combined gain, pixel, entrance filter & vignette corrections
	Create bad pixel and cosmic ray index files; fix bad pixels	Iterative repair using TRACE-browser method; replaced pixel values saved in index files
	Rescale to integer; saved as 32bit Int	In DN, exposure & photon conversion as keywords
1.5	Plate scale, rotation & shift	Match HMI center & rotation to solar north. Scale HMI to match AIA
	Rescale to integer	In DN – exposure & photon conversion as keywords
1.6	PSF and diffraction correction	Quick FFT method
	Bad Pixel replacement	Nearest Neighbor replacement
	Plate scale, rotation & sub-pixel shifts	Match HMi center & rotation. Scale HMI to match AIA Image distortion correction
	Rescale to integer; saved as 32bit Int	In DN – exposure & photon conversion as keywords
Browse	Compress to JPEG2000 ~30 sec	Derived from Level 1.5 or 1.6 Cadence may vary with solar activity and age
Synoptic	Resample to 1kx1k, 90 sec	Derived from Level 1.6
Research DataCube	Cropped & re-sampled using HEK	Cutouts of Level 1.5, with selected cadence and resolution
Higher Products	TBD	PFSS coefficients, DEM….

Table 1. AIA data levels

Series Name	Prime Keys	Record structure	Segments	New SU	Retention	Archive	Notes
AIA.tlm	filename	Raw	?	Yes	30d	Yes	Internal to JSOC: no public access
AIA.hk	?	?	None	No			Match HMI
AIA.lev0	FSN	Level 0	Image	Yes	30d	Yes	Internal to JSOC: no public access
AIA.dark	DATE_OBS, FSN	Level 0	Image	Yes	1900d	No	Distributed via DRMS/SUMS
AIA.flat	DATE_OBS, FSN	Level 0	Image	Yes	1900d	No	Distributed via DRMS/SUMS
AIA.cal	TBD	TBD	None	No	1900d	No	Calibration values Distributed via DRMS/SUMS
AIA.lev1_nrt	T_REC or T_OBS, FSN	Level 1	Images, badpix, dropouts, despike map	Yes	7d	No	Preliminary science data Generated when Level 0 is complete
AIA.lev1	T_OBS, FSN	Level 1	Images Badpix, dropouts, despike map	Yes	60d	No	Production science data using best available darks and flats Distributed via DRMS/SUMS
AIA.lev1_5	T_OBS, FSN	Level 1	Channels	Yes	5d	No	Derived from AIA.lev1 or lev1_nrt; Temporary series
AIA.lev1_6	T_OBS, FSN	Level 1	Channels	Yes	5d	No	Derived from AIA.lev1 or lev1.nrt; Temporary Series
AIA.synoptic	T_OBS, FSN	Level 1	Channels	Yes	2000d	Yes	FITS format, resampled from AIA.lev1_6
AIA.browse	T_OBS, FSN	Level 1	photo, chromo, corona, flare	Yes	2000d	Yes	JPEG2000 Store as RGB as…? Photo: {aia_wl,hmi_cont, hmi_mgram}/90sec Chromo:{1600,304,1700}/ 30sec Corona:{171,211,335}/30sec Flare:{94,131,193}/30sec Derived from AIA.lev1_6
AIA.cubes	HCR ivorn or JSOCrequest	Level 1	Channels	Yes	30d	Yes	cutout & resampled AIA.lev1.5+
AIA.roi							treated as cube?

Table 2. JSOC series for AIA. Subscriptions to series in grey are internal to SU/LM. Those in blue need further study.

Higher-level products for general distribution are reductions of these three series. Browse and synoptic products are generated from Level 1.5. (The possibility of deriving them from 1.6 is being investigated.) Research products are subsets of Level 1.5. These levels are maintained as data series in the JSOC DRMS/SUMS environment.

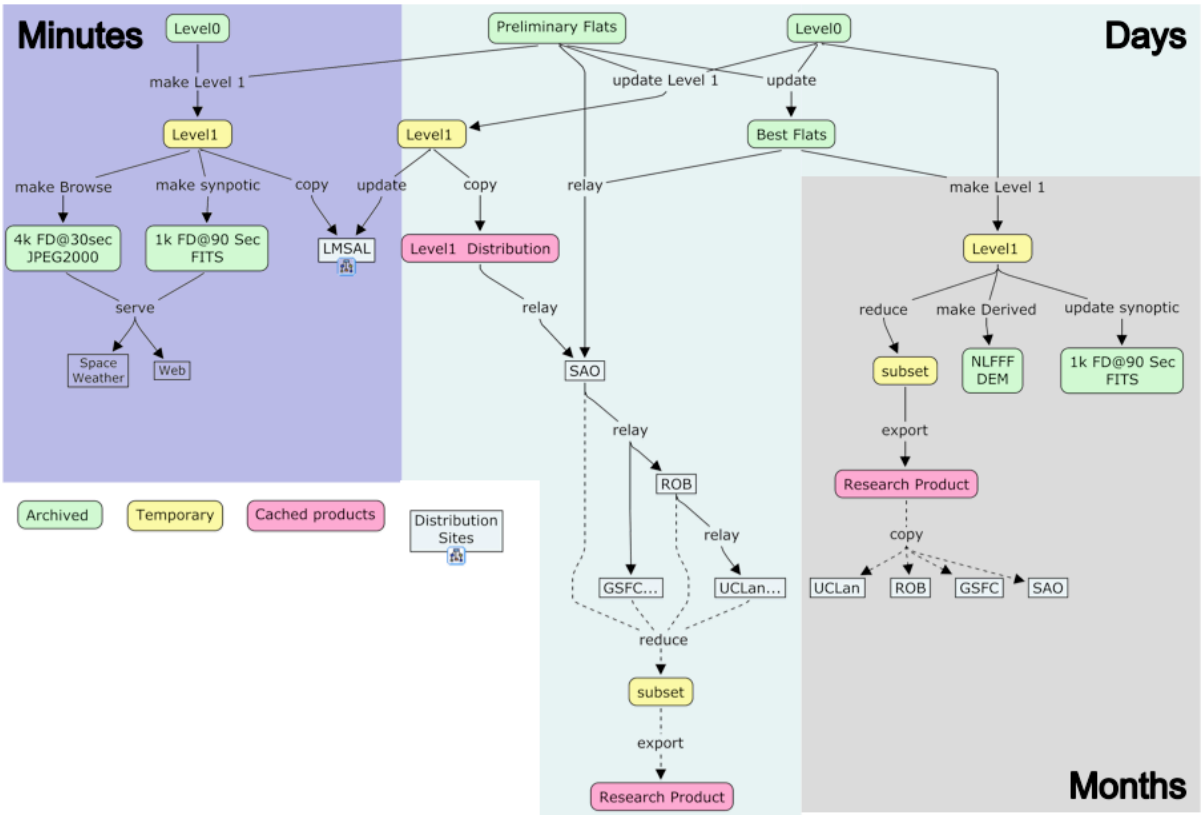

Figure 1. Data Distribution of AIA data. Data flows from the top down three distinct paths on three time scales as indicated by their background color, all of which are managed within the DRMS/SUMS environment. Items outside of the shaded areas are independent of DRMS/SUMS. The near real time path generates browse and synoptic products within minutes of data receipt. Once all data is received, it flows though the distribution path and is relayed to other distribution sites for temporary storage. (Details beyond relaying to SAO are TBD). As these temporary products become unavailable, or outdated, final level 1 products are created. These are used to update the synoptic images as well as to be sources for extracting subsets of research-grade products for distribution to the community.

AIA Data distribution

NOTE: All these products will become available after spacecraft and instrument commissioning and the initial calibration and data processing refinements are put in place. We expect these to be complete, and data to start flowing out of the JSOC by late May/early June.

There are competing needs for AIA data that our plan attempts to meet. Space weather forecasters and mission operators need near-realtime (NRT) data products. These do not need a high level of calibration or resolution in space and time. Full calibration requires both complete and final telemetry, which can only be confirmed a day after acquisition, and processed flat-fields, which require data taken at six months intervals for final calibrations. Science-grade data will be available once data acquisition is complete, but the quality will continue to evolve at least until after the subsequent six-month calibration run.

To avoid confusion, we define these terms for use in the following discussion:

- Subscribe/subscription. These are used in reference to the JSOC DRMS/SUMS system. Distributions sites and external users can follow data products within a DRMS series through subscription.
- Relay. To reduce distribute network load data will be transferred to the collaborating distribution sites via a chain of successive copies. These may be through a chain of subscriptions, or other mechanisms.
- Export occurs when data leaves the DRMS/SUMS environment as data files with complete headers. A series of transformations may be applied during the export process (pre-export), including image cropping and resampling, etc.

Since we cannot re-transmit the entire AIA dataset and none of our partners is in a position to act as a full-mirror site, we will distribute AIA science data in three stages as shown in Figure 1. Once science telemetry is captured by the Stanford Data Capture system, it is processed immediately into Level 0 images. These images are then calibrated using existing flat fields to Level 1, and these are used to generate Browse and Space Weather products. The Level 1 NRT images are also copied over to LMSAL for immediate science assessment.

The browse and synoptic products are registered as series within the JSOC that can be subscribed to by other distribution sites. Since they are also accessible on the web as individual URLs, they can also be copied via webget and other web-methods. Alternatively, browse products can be viewed using client applications such as HelioViewer, which communicates with JSOC servers using JPEG2000 streaming protocols.

Level 1 science data for distribution to AIA Co-Is will be generated as soon as the final telemetry is processed to Level 0, typically within 48 hours of acquisition. It will also be calibrated using existing, preliminary flat fields. These data will update the LMSAL cache and be relayed to SAO through DRMS as the series AIA.lev1 and then on to other distribution sites. The flat-fields and other calibration data will evolve on roughly a six month cadence, and these can be relayed to SAO and beyond by subscription to the AIA.dark, AIA.flats and AIA.cal series. These data will reside at the various distribution sites for as long as their resources permit, and the data can be served to the research community, as they are able. We present a plan for managing these services below.

As data arrives at SAO, it will be used to populate other VSO data caches. These caches will maintain the most recent AIA data for as long as resources permit.

Tentative Plans for Long-term AIA data distribution

After this initial stage, we envision that researchers will make focused data requests. The full archive will not be re-transmitted. Instead, the Heliophysics Event Knowledgebase (HEK) will guide researchers to the most useful times and locations and will *suggest* reasonable fields of view and durations (see Figure 2A). These events are primarily generated by the Event Detection System and by visual inspection at LMSAL. However, contributions from external sources are

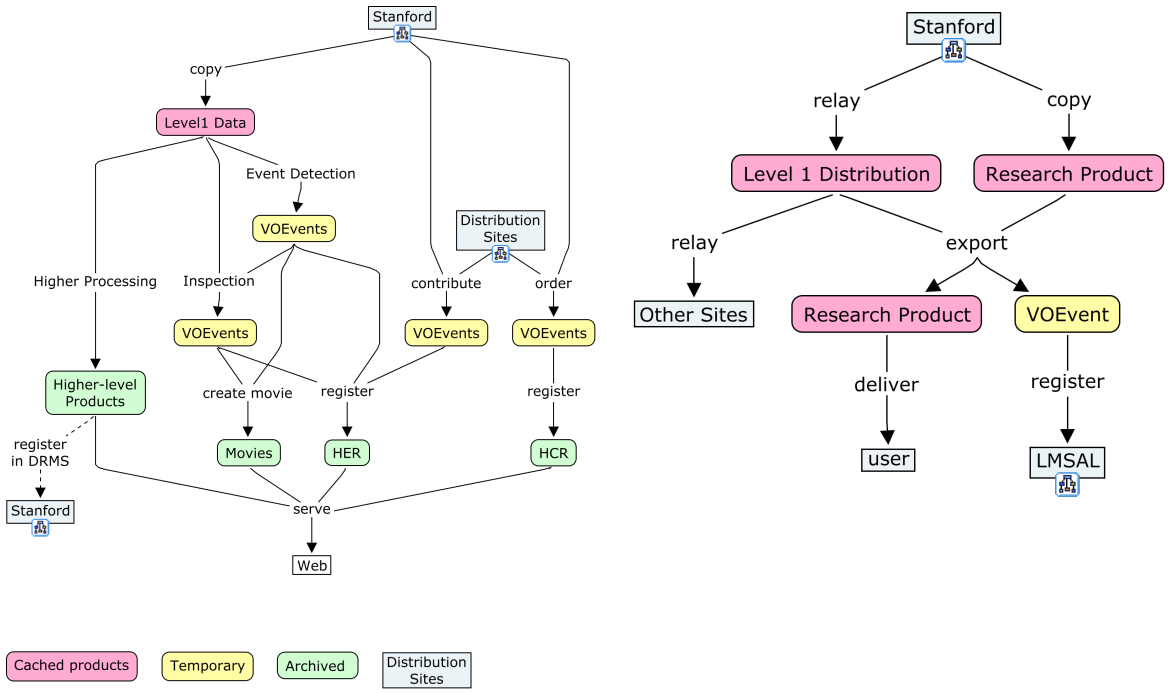

Figure 2. A) Actions within LMSAL. Level 1 data flows in for analysis, which results in both higher level products and event descriptions. The descriptions are registered into the HER. Independently Stanford and other collaborators can register events within the HER by providing appropriately structured submissions. They also report orders from external users, which is recorded into the HCR. B) Actions at distribution sites. They receive the intial Level 1 datastream to populate their local data caches, and serve this data to the local community and other distribution sites. As the data expires or is purged from the local cache, research products that have been requested by the community can be copied from Stanford to repopulate the cache with recently-accessed data. If the distribution site is configured to accept requests for new data products, they generate the product and register it with the HCR. The origin of research products is tracked by the HCR. Any products generated outside of the AIA science processing (Stanford and LMSAL) and any results based upon them will not be the responsibility of AIA team.

possible through an access-controlled web interface. With *guidance* from the HEK, researchers will access data by specifying cutouts, wavelengths and sampling rates in space-time. These research data products will be Level 1.5 data that uses the best flat fields and calibration data available at the time they are created. They will be tagged both in DRMS and in the Heliophysics Coverage Registry. Subsequent researchers can be directed to existing, local copies of research products through the HEK.

These products are tracked via the existing HCR system, if the distribution sites are willing to participate (Figure 2B). At a minimum this requires that each site report the presence of a copy of a product to the HCR, along with an expected expiration date for when the data will be purged. SolarSoft routines are available to support this reporting. This permits the HCR to identify where data products are available at all locations at any time.

We assume most users will be satisfied with research products derived from Level 1.5 data. However some may wish to apply specialized processing steps that require Level 1 data. To help them we will provide SolarSoft routines for AIA calibration and distribute the flatfields and calibration data with the SSWDB.