

# eSDO Algorithms, Data Centre and Visualization Tools for the UK Virtual Observatory

Elizabeth Auden<sup>1</sup>, J.L. Culhane<sup>1</sup>, Y.P. Elsworth<sup>2</sup>, A. Fludra<sup>3</sup>, M.J. Thompson<sup>4</sup>

<sup>1</sup>Mullard Space Science Laboratory, University College London.

<sup>2</sup>Physics Department, University of Birmingham.

<sup>3</sup>Rutherford Appleton Laboratory

<sup>4</sup>Department of Applied Mathematics, University of Sheffield

## Abstract

The eSDO project is funded by PPARC to make data, algorithms and visualization techniques from the Solar Dynamics Observatory mission available to the UK solar community through the virtual observatory. Scientists and developers based at MSSL, RAL and the Universities of Sheffield and Birmingham form a consortium that has spent one year assessing the UK solar community's requirements for SDO access. Developers are now engaged in a two year implementation phase that will deliver three workpackages that encompass ten solar algorithms, UK data centre designs and visualization tools in preparation for SDO launch in August 2008.

## 1. Introduction

The Solar Dynamics Observatory (SDO) mission will be launched on 31 August 2008 with a payload of three instruments: the Atmospheric Imaging Assembly (AIA), the Helioseismic and Magnetic Imager (HMI) and the EUV Variability Experiment (EVE) [1]. The data rate from these instruments will approach 2.5 TB / day; this volume of data presents new challenges to the global solar community. Although the primary SDO data centre will be maintained in the US by the Joint Science Operations Center (JSOC), a secondary UK data centre will allow UK solar scientists to search for and retrieve data through the AstroGrid virtual observatory (VO) [2]. Visualization tools such as catalogues, streaming image viewers and movie makers will allow scientists to identify pertinent datasets more rapidly; once selected, these datasets can be processed with VO-enabled solar algorithms.

The ten algorithms under development by the eSDO project fall into two categories: event / feature recognition and helioseismology. The first category of algorithms will be primarily used by event driven solar physicists studying flares, coronal mass ejections (CMEs) and other

solar behaviours displaying rapid evolution. The second category of algorithms, helioseismology, will be used by scientists examining the interior of the Sun.

The needs of these two user groups direct the design of the UK SDO data centre. Event driven scientists expect access to data as soon as an interesting solar event has occurred, so rapid searching and avoidance of network bottlenecks are priorities for this user group. In contrast, helioseismologists study solar oscillations over much longer periods of time. While they do not require immediate access to data following solar events, scientists in the helioseismology community typically process months or years of solar data at a time. This group's priorities are storage and processing capabilities for large blocks of data.

## 2. Algorithms

### 2.1 Solar Algorithms

The eSDO project will deliver C code and documentation for four helioseismology algorithms and six feature / event recognition algorithms. Developers at MSSL and RAL are concentrating on the first category of algorithms, which will primarily process data

from the AIA instrument. These algorithms include non-linear magnetic field extrapolation, coronal loop recognition, helicity computation, small event detection and recognition of coronal mass ejection (CME) dimming regions. The magnetic field extrapolation algorithm will particularly benefit from grid computing techniques. Wiegelmann's "optimization" method of 3-D magnetic field extrapolation is used to calculate magnetic field strength over a user-defined solar volume [3]. Computations of volume integrals are distributed with the MPICH 1 or 2 parallel processing protocol; current investigations are exploring the execution of this code on National Grid Service (NGS) facilities through an AstroGrid application interface.

The universities of Birmingham and Sheffield are examining global and local helioseismology respectively, and these algorithms will process data from the HMI instrument. The Birmingham group is implementing mode frequency analysis and mode asymmetry analysis algorithms. The Sheffield group is developing subsurface flow analysis, perturbation map generation and computation of local helioseismology inversion algorithms. Helioseismology algorithms analyse low frequency solar oscillations with a periodicities that range from 5 minutes for  $p$  mode (acoustic) waves to hours or days for  $g$  mode (gravity) waves [4]. These oscillations are studied over long periods of time, so helioseismology algorithms require correspondingly long data sequences as input. Because of the large input series, these algorithms should be hosted to locally to helioseismology data caches in the US or UK to ensure efficient completion of data processing.

## 2.2 Algorithm Distribution

Solar algorithms developed by eSDO will be made available to UK users through the JSOC pipeline, AstroGrid, and SolarSoft. The JSOC pipeline systems at Stanford University and Lockheed Martin will contain processing modules that either run automatically or when invoked by user requests. Most data products created through the automated modules will be made available to the public through the virtual observatory; however, authorized pipeline users can execute optional modules if higher level processing or an alternative implementation of an algorithm is required. Some eSDO code, such as the global helioseismology mode parameters algorithm, will be integrated with the JSOC pipeline as automated modules. Other

eSDO algorithms will be designated by JSOC as optional user-invoked modules. UK SDO co-investigators and their teams will be able to access the JSOC pipeline directly through accounts at Stanford University. The coronal loop recognition algorithm will be the eSDO test case for JSOC pipeline integration.

For the wider UK solar community, access to eSDO algorithms will be available through AstroGrid or SolarSoft. Algorithms deployed as UK hosted AstroGridT CEA web services will be accessible through the AstroGrid workbench via the application launcher and parameterized workflows. Emerging VO access to facilities such as NGS will allow scientists to execute computationally intensive algorithms on powerful remote machines. Algorithms that are not computationally intensive will be wrapped in IDL for SolarSoft distribution through the MSSSL gateway. CEA deployment will begin in late September 2006.

In addition to the algorithms under development by eSDO consortium institutions, the project will also work with UK coronal seismologists to deploy wave power analysis algorithms as JSOC pipeline modules and AstroGrid CEA applications.

## 3. Data Visualization

The huge volume of SDO data collected every day makes it imperative to search the data archive efficiently. The visualization techniques including streaming tools, catalogues, thumbnail extraction and movie generation will aid scientists in archive navigation. By enabling scientists to identify relevant datasets quickly, these tools will reduce network traffic and save research time.

### 3.1 Browse Products

Browse products are files such as thumbnail images and movies that scientists can view quickly to identify interesting datasets that merit further processing. Thumbnail images of selected AIA and HMI datasets will be available through VSO, and the JSOC will produce movies of "regions of interest". The browse product tools developed by eSDO will allow scientists to generate their own thumbnail image galleries and movies through a web interface. Users will specify start time, end time, cadence, and data products along with outputs of either an image gallery or a movie. A web service will

locate the datasets, extract images from FITS files, label the images with relevant metadata from the FITS headers, and then display an image gallery in the user's web browser or create an MPEG movie that the user can view or download. Following advice from the eSDO Phase A review with PPARC in November 2005, the movie maker will be enhanced to evaluate a data sequence's suitability for wavelet analysis. Prototypes of the thumbnail maker and image gallery tool are now available [5,6].

### 3.2 Catalogues

Two science catalogues will be generated from eSDO algorithms. One catalogue will store statistical information about small solar events and CME dimming regions. The catalogue will be generated in the UK and annotated continuously from the UK data centre's cached AIA datasets. The second catalogue will provide a "GONG month" [7] of helioseismology information produced by the mode parameters analysis algorithm. This catalogue will be generated in the US using eSDO modules integrated with the JSOC pipeline. Both catalogues will be searchable through the virtual observatory with instances of the AstroGrid DataSet Access (DSA) software.

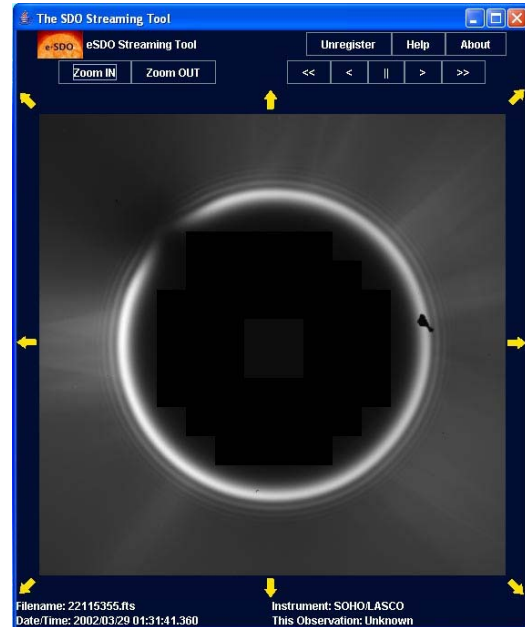
### 3.3 SDO Streaming Tool

The SDO streaming tool will display HMI and AIA science products in a web GUI. Scientists will be able to pan and zoom the images in both space and time. A user will launch the java webstart SDO streaming tool from a web browser and then specify a start time, stop time, cadence, and data product. Three types of SDO products will be available: AIA images from 10 channels, HMI continuum maps, and HMI line-of-sight magnetograms.

The user will be able to zoom in spatially from a full disk, low resolution image to a full resolution 512 by 512 pixel display of a solar area. Users will also be able to pan spatially zoomed images in eight directions. Once a user has selected a cadence (for instance, 1 image per hour), data products matching that cadence will be displayed on the screen; users will be able to "zoom" in time by increasing or decreasing this cadence, while rewind, fast forward and pause facilities will allow users to "pan" in time.

Collaboration with the UK coronal seismology community will enhance the streaming tool with a wave power analysis function.

A prototype of the eSDO streaming tool that can be used in conjunction with the AstroGrid Helioscope application is available now [8].



**Figure 1: The SDO streaming tool prototype loading a coronal image from the SOHO LASCO instrument, zoomed in one spatial level.**

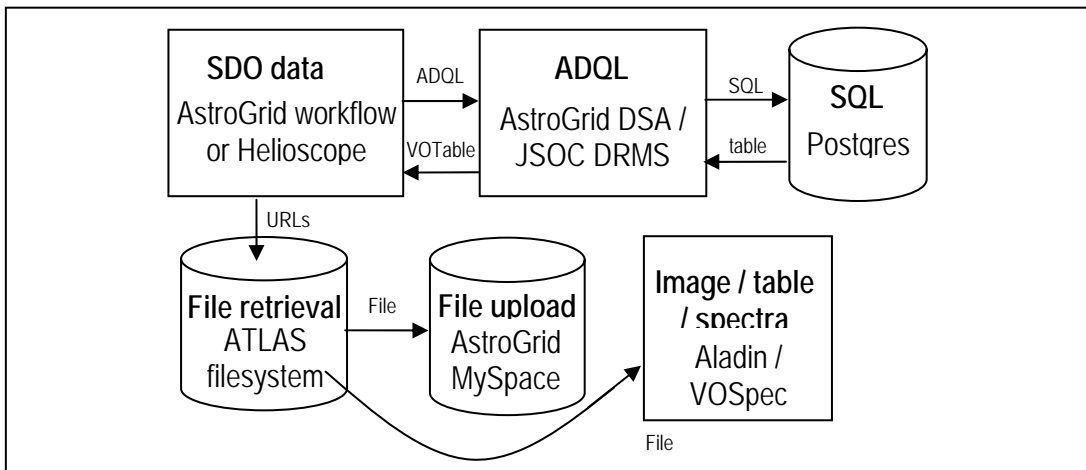
## 4. UK Data Centre

The primary SDO data centre will be based in the US at Stanford University and Lockheed Martin, and users will be able to access the data stored there directly through the Virtual Solar Observatory (VSO). A second data centre will be established in the UK at the ATLAS datastore facility at the Rutherford Appleton Laboratory. Rather than providing a full data mirror, this data centre will store all SDO metadata as well as holding a cache recent data products and popular data requests. The metadata storage will allow SDO data searches to be included in AstroGrid workflows while the local data cache will provide UK scientists with access to SDO data when the US data centre experiences decreases in network speed during periods of high user demand.

### 4.1 Metadata and VO Searches

JSOC will store SDO metadata in a distributed Data Resource Management System (DRMS) while actual data files will be stored in a distributed Storage Unit Management System (SUMS). Each instance of DRMS will hold metadata for all SDO files, regardless of the fraction of SDO data stored in the local SUMS. When new SDO high level data sets are created in the UK, the metadata for those datasets will be entered into a DRMS database and disseminated globally to all DRMS instances. Instances of DRMS operate in conjunction with Postgres and Oracle databases to store keyword value pairs that are used to generate SDO FITS headers on the fly.

JSOC has designated the Virtual Solar Observatory (VSO) as the official front end for user access to SDO data, and UK scientists will be able to download SDO data from the VSO web interface. However, VSO is not compliant with the International Virtual Observatory Alliance (IVOA) web services that AstroGrid uses; therefore AstroGrid workflows cannot access SDO data through VSO. Instead, the eSDO project is configuring an instance of the AstroGrid DSA software to interface with the UK DRMS. This will allow scientists to set up an AstroGrid workflow that will search through the SDO metadata, identify the relevant SDO datasets, and retrieve the datasets as URLs.



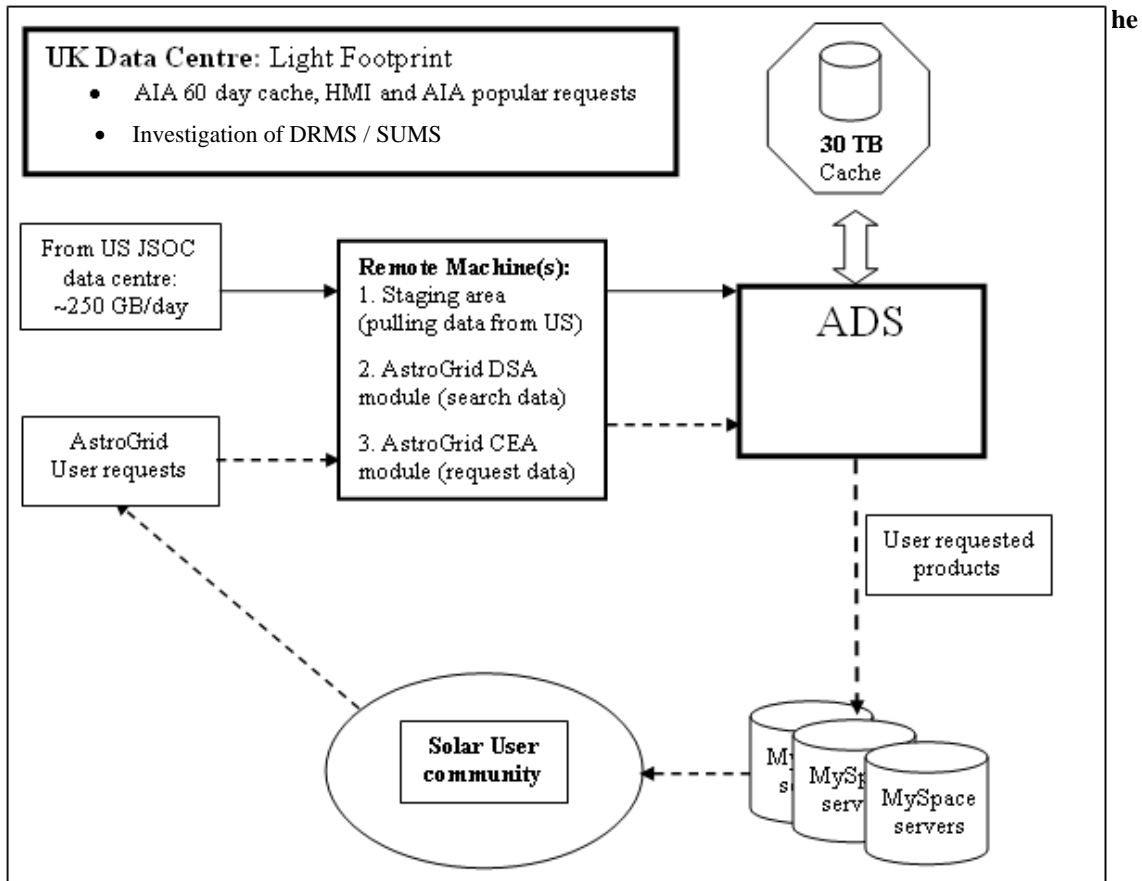
**Figure 2:** This figure demonstrates the flow of information when an AstroGrid user submits a request for AIA or HMI data to the UK SDO data centre.

## 4.2 ATLAS Data Cache

Although a UK instance of the DRMS server will provide a UK holding of all SDO metadata, the datasets themselves will be archived in the US. The eSDO project has proposed a 30 TB disk cache hosted at the ATLAS facility for holding a fraction of the SDO data. 15 TB will hold a rolling 60 day cache of recent AIA and HMI data while the other 15 TB will serve as a “true cache” for older datasets requested by UK users through the AstroGrid system. When a UK user requests an HMI or AIA science product through the AstroGrid system, the request will be sent to the UK data centre. If the data is not available there, the request will be redirected through an AstroGrid / VSO

interface, and the relevant data products will be returned to the UK in export format (FITS, JPEG, VOTable, etc).

In the event of a large flare or coronal mass ejection, UK scientists can download data from the rolling 60 day cache without impedence from the heavy network traffic of users accessing the JSOC archive. In addition, UK users of the SDO streaming tool will have more seamless access to images viewed at a high cadence. The true cache will reduce transatlantic data transfers for such user groups as the helioseismology community that process large blocks of data but are not concerned with recent events.



**Figure 3: The proposed architecture for a UK SDO data centre demonstrates the interaction between the full US SDO archive, the UK cache, and user requests.**

## 5. Conclusions

Although eSDO funding will cease before the SDO mission is launched in August 2008, the project will deliver completed designs, code and documentation for each of the three major workpackages. Ten algorithms concentrating on solar feature recognition and helioseismology will be ready for deployment through the JSOC pipeline, AstroGrid and SolarSoft. The UK data centre will provide interfaces between the JSOC data resource management system, the AstroGrid data set access software and the ATLAS storage facility, allowing execution of virtual observatory searches all SDO metadata with rapid access to SDO data cached in the UK. Finally, scientists will be able to identify relevant datasets through visualization tools such as catalogues, web-based image gallery and movie makers, and the SDO streaming tool.

## 6. References

- [1] "Solar Dynamics Observatory Mission Factsheet", 3 November 2005. [http://sdo.gsfc.nasa.gov/sdo\\_factsheet.pdf](http://sdo.gsfc.nasa.gov/sdo_factsheet.pdf)
- [2] Rixon, G. "Software Architecture of AstroGrid v1", 27 September 2005.
- [3] Wiegmann, T. 2004, Solar Physics, 219, 87-108
- [4] Lou, Y.Q. 2001, Astrophysical Journal, 556, 2, L1221-L125
- [5] Auden, E. eSDO Thumbnail Maker, 19 July 2006. [http://msslxx.mssl.ucl.ac.uk:8080/esdo/visualization/jnlp/esdo\\_thumbnailmaker.jnlp](http://msslxx.mssl.ucl.ac.uk:8080/esdo/visualization/jnlp/esdo_thumbnailmaker.jnlp)
- [6] Auden, E. eSDO Image Gallery, 18 July 2006. [http://msslxx.mssl.ucl.ac.uk:8080/esdo/visualization/jnlp/esdo\\_imagegallery.jnlp](http://msslxx.mssl.ucl.ac.uk:8080/esdo/visualization/jnlp/esdo_imagegallery.jnlp)
- [7] Howe, R., Komm, R., Hill, F. 2000, Solar Physics, 192, 2, 427-435
- [8] Auden, E. eSDO Streaming Tool, 14 July 2006. [http://msslxx.mssl.ucl.ac.uk:8080/esdo/visualization/jnlp/esdo\\_streamingtool.jnlp](http://msslxx.mssl.ucl.ac.uk:8080/esdo/visualization/jnlp/esdo_streamingtool.jnlp)