Soho Long-term Archive (SOLAR)
Status Report 30th November 2009

Technical report n. 126
Author: A.Cora, S.Giordano

Table of Contents

1.0 INTRODUCTION

The SOHO Long-term Archive (SOLAR), developed at the INAF-OATo is one of the three European archives for the Solar and Heliospheric Observatory (SOHO) ESA/NASA data (Antonucci et al 2001). SOLAR is designed as a mirror of the NASA Goddard Space Flight Center (GSFC) archive. The other two mirrored archives are located at the Rutherford Appleton Laboratory (UK) and the Multi-Experiment Data Operation Centre (MEDOC) of the Institut d'Astrophysique Spatiale (France). The SOLAR archive stores the scientific data of all 12 instruments operating on board SOHO. According to the commitments assumed when the European Space Agency has granted the right to archive the SOHO data to the Torino solar physics group, SOLAR has to be operational during the SOHO mission (recently extended to 2010) and for the 10-year period following the mission end (2010-2020).
SOLAR stores the following types of data-sets and data products:

- Calibrated scientific data: the observational data of 12 instruments operating on board SOHO (see Table 1).
- Summary data (a simple collection of daily observations of the instruments).
- Software (packages developed by the instrumental teams to reduce and analyze the observations).

The archive development has been completed at the beginning of 2002. The archive, however, has not been declared formally on-line, due to problems related to the definitive settlement of the SOLAR hardware in the computer center facility, until August 6, 2003.

INSTRUMENT	START DATE	LATEST DATE
CDS	19-JAN-1996	07-DEC-2007
CELIAS	02-DEC-1995	06-DEC-2006
EIT	01-JAN-1996	27-NOV-2007
ERNE	08-MAY-1996	14-NOV-2007
GOLF	01-JAN-1996	18-DEC-2006
LASCO	08-DEC-1995	30-JAN-2007
MDI	01-MAY-1996	13-APR-2003
SUMER	10-JAN-1996	31-MAR-2007
SWAN	12-JAN-1996	28-JUL-2007
UVCS	20-JAN-1996	19-FEB-2007
COSTEP	07-DEC-1995	09-OCT-2006
VIRGO	06-DEC-1995	27-NOV-2007

TABLE 1. Data availability at 30th November 2009

At present it includes more than 1.500.000 data files, corresponding about to 1TB . SOLAR Catalog is based on ORACLE software, as a typical relational database, it consists of tables of highly structured records with fixed-length fields.
Keywords and fields are defined on the base of keywords fixed by the instrumental scientific teams.The Data Base Schema was described in the OATo Technical Publication n.70 (Cora, et al 2003).
The WEB server records the access to the Catalog (see Fig.1),and the number of the distributed files (see Fig.2).

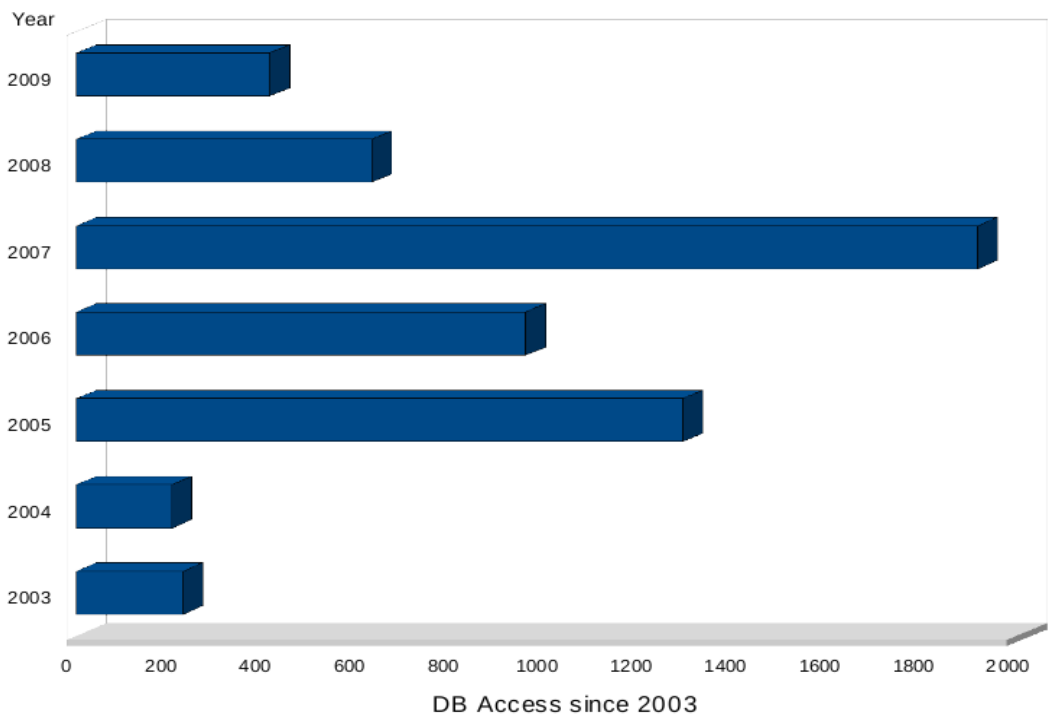

Fig.1 Number of access to the SOLAR data base since 2003

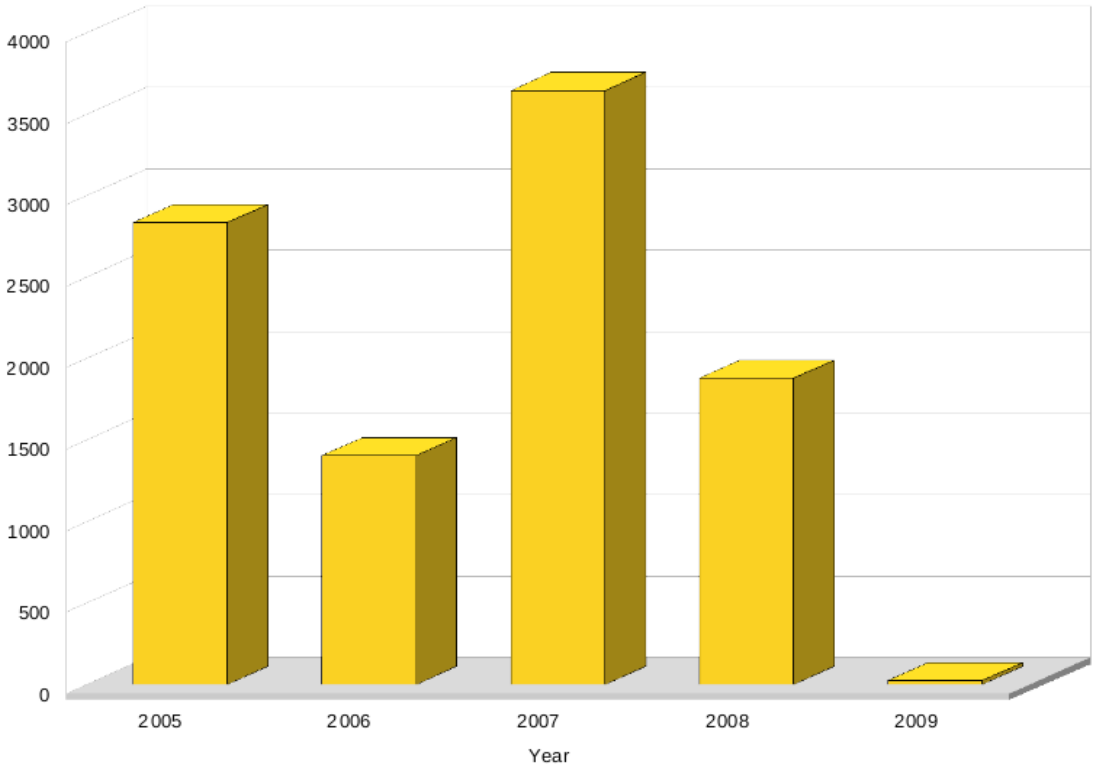

Fig.2 Number of data sets distributed since 2005

1.1 MAJOR IMPROVEMENTS

During the 2009, we have spent a lot of efforts to change ORACLE platform, from ORACLE 8i to ORACLE 10g. This was a critical task, necessary to maintain the catalog, and bit complicated because our installation uses Linux-Fedora as operative system (this solution is officially unsupported by ORACLE).
The new Data Base Management System based on ORACLE 10g will be available at the community in the early months of 2010.
The major improvements, developed on SOLAR since 2002, were

2002 - SOLAR implement also GSFC user interface and the data summary
2003 – EIT data quick look is available at the SOLAR site
2004 – LASCO, MDI and UVCS data quick look
2005 - change the extract meta data procedure
2005 – SOLAR is accessible also through SOLARNET
2006 - new URL to: solar.oato.inaf.it
2007 – dismiss the old file system server (pico.oato.inaf.it)
2009 – Updated the RDBMS at ORACLE 10g

1.2 HARDWARE ARCHITECTURAL OVERVIEW

The SOLAR architecture has been developed to provide a flexible storage and data retrieval. His hardware and his architecture was continuously modified in these years.
The native configuration was based on 3 servers connected via a Fiber Channel (FC) line at 1 GB/sec in a private network. The main server (a SUN Entreprise server with 2 CPU) hosted the RDBMS and was connected through a SCSI bus to a 254 GB RAID disk array dedicated to the storage of the RDBMS software and the SOLAR catalog data. A 1 TB DLT library was connected to the server over the FC line and was used to backup the entire system.
Furthermore a hierarchical storage management software developed by SUN was used to increase the global amount of the available file system space putting near-on-line a part of data contained into the Data cartridges.
This software automatically retrieved from the DLT cartridge in the Library and cached in the on-line file system; except for the delay to perform the retrieval, the operation was transparent to users and to other programs interacting with the system, but this kind of application and its license is expensive.
The present architecture continue to be based on 3 servers. In order to reduce costs the main server (RDBM Server), is now a HP Proliant using open source software (Linux Fedora). The observation files (in FITS format) are stored in a low costs Network Attached System (File Server). The NAS has been a suitable solution because of its RAID support and overall lower cost. The third server have mounted Apache HTTP server and contains the web site (WEB server).

1.3 REMOTE CATALOG ACCESS

It is widely recognized that there is a need for good methods for disseminating large amounts of digital data generated by space missions.
The SOHO archives adopted the World-Wide-Web (WWW) as the platform to disseminate mission information and allow access for data retrieval.
The Graphic User Interfaces (GUI) developed by the ESA/GSFC team allows to build complex operations within a single query.
Furthermore through the WWW interface is possible to do a multi-step query process in which queries are progressively refined. At the end of the process the user decides to either save the information related to the requested data

files or retrieve them; in this case the requested observations are compressed into a cache area and then made available for download via a temporary web address, specific to the request.
SOLAR Archive was operating from the beginnings of March 2002 at the URL: http://solar.to.astro.it , and since 2006 at the new URL: http://solar.oato.inaf.it

3.0 UPGRADE DATA BASE PROCEDURE

In this paragraph we summarize, the work in progress to develop automatic upgrade procedures.

The GSFC developed for SOLAR and the other European SOHO Archives the software that extract the meta-data, defines, creates and populates the SOHO archive databases. It is not secret that the Archives shown different update dates and consequently they make available on the WEB different data sets.

The different update status, was due to extract-metadata procedure which populates the DataBase. This procedure take long time and for SOLAR was consisted of four steps:
- download and uncopress the FITS file from data cartridge on a temporary area
- read and extract the keyword from the FITS files
- populate the Tables of the DataBase
– move and compress the FITS files in a permanent storage called "staging area".

SOLAR from April 2005 adopts a new procedure to replicate the main Archive sites at the GSFC.
The pipeline of this procedure consists of the following operation
- download of the ORACLE dump file of the main Archive thought an FTP site.
- replicate the GSFC in our site
- selects files non available in SOLAR archive and download them through GSFC HTTP site
- modify the replica of the GSFC and make it available at SOLAR.

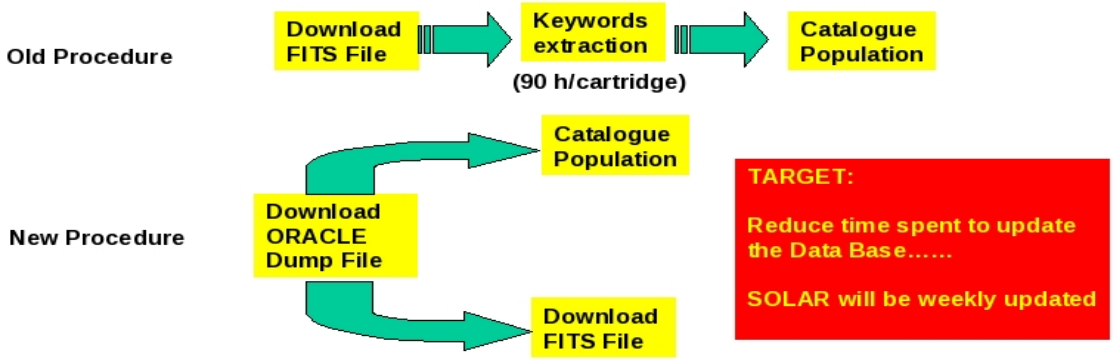

FIG.3 Old and New Procedures schema

This procedure reduce the time spent to download and uncompress FITS file and in general the time requests by the extract-metada procedure.
But the new procedures have a costs in term of synchronization of the RDBM Server with the main server hosted to the GSFC. The ORACLE import/export facilities are correlated to the software release and for this reason we have, in this year, update the RDBM Server.
At to day SOLAR archive shows a considerable delay from the principal Archive (GSFC); in the future with the fully automation of the upgrade procedure SOLAR will be updated every week and it should distribute the same data set available on the main Archive

4.0 WEB SITE AND DATA QUICK LOOK

The conceptual model for organizing data files within the SOHO archive is as follows. We make a data file equivalent to an 'observation'. This is the smallest piece of data we deal with.
Several observations can be grouped in a 'study'. Typically, all the observations of a study share some parameters, like a scientific objective, or the object being observed. The only requirement we impose is that all observations of a particular study should have been taken with the same instrument. Some instruments organize observations in studies, and some do not. The third and final level is the 'campaign'.
Observations related to a campaign can be from different instruments. Also, an observation can be related to more than one campaign.
The system should provide quick access to the desired documentation and guiding the user through the catalog (meta-data) to choice the data.
The data retrievals approach is conditioned firstly on the type of information that is stored; secondly, by the way the database is structured (relational or object oriented Database, DB schema, user interface, and so on); and finally by the tools used to search it.

The EGSO (European Grid for Solar Observations) project has put in evidence the request by Scientific Community of a data display to to refine the selection and the retrieval data (Reardon 2002).
The Data Quick look is the first tool implemented on SOLAR, belonging to this category; it permits to see the raw data of four instruments: EIT, LASCO, MDI and UVCS. At the site of the archive, after the first query, the User clicking on the question mark in the table of the results, can obtain the Observational Parameters and for four instruments a data quick look (a small image with a size of 250x250 pixels about).
For the instrument which produces images like EIT, LASCO and MDI, is also possible visualize the full size picture clicking on the quick look image.
For the UVCS the quick look shown the first 3 exposures of the data recorder by the Lyman and OVI channel (a visualization of the white light channel is under development), and the User should click on this data preview to have the full data set of the raw exposures.

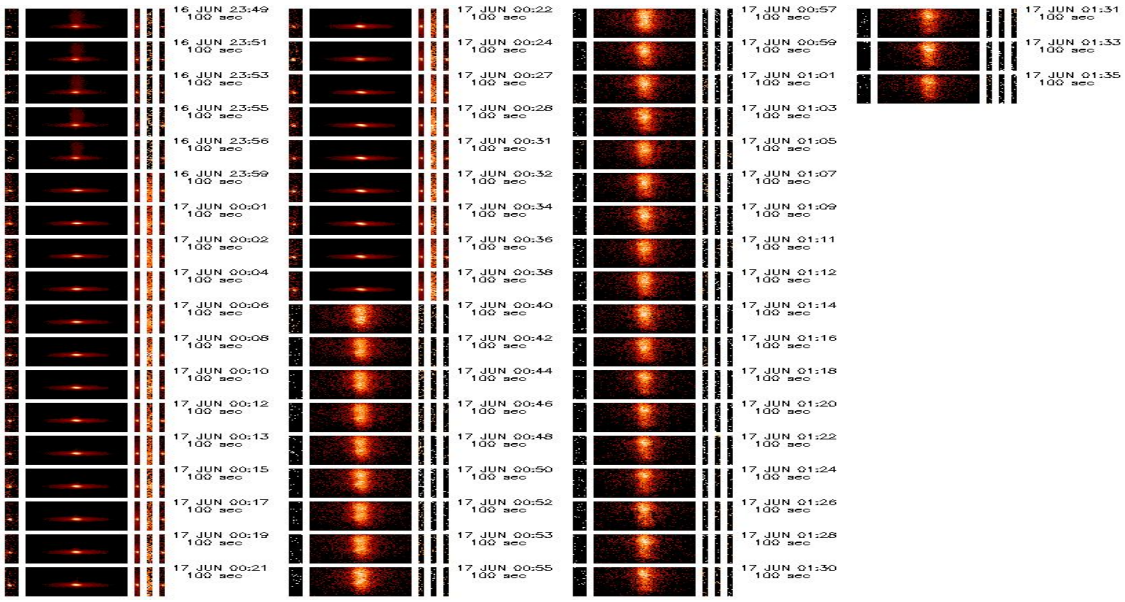

FIG.4 – UVCS full quick look an example

This increase the information distributed by the archive and in this way the User can test the intermediate results and make judgment on the quality of the data and eventually visibility of the phenomena.
Furthermore, clicking on the "Data Summary" button (under the quick-look picture), is also available a sort of multi-instrumental quick look composed by images take from different instrument; so the Researcher is able to make an opinion on possible events/details visible by other instruments.

5.0 SOLAR and SOLARNET

During the 2005 the data base schema of SOLAR was modify to introduce two tables containing a brief description of the fields and the primary tables. This operation was finalized to improve its accessibility by SOAP methods and integrate it in SOLar ARchives NETwork: SOLARNET (Volpicelli 2006).
SOLARNET is a federation of heterogeneous Italian solar archives into a VO framework; currently the SOLARNET federation includes, besides SOLAR, 4 other data archives:
TSRS telescope archive SOLRA (SOLar Radio Archive) at Trieste Observatory, archive of RISE/PSPT (Precision Solar Photometric Telescope) network of Rome and Mauna Loa (the Mauna Loa data is not currently available), a small subset of VAMOS at Napoli Observatory and in the next future CATANIA SOLAR DATA ARCHIVE at Catania Observatory.
Information and data set distributed by SOLAR are available through SOLARNET like a single integrated database (http://solarnet.to.astro.it:8080).

6.0 CONCLUSIONS

SOLAR was design to be operational during the SOHO mission and for the 10-years period following the mission end. The mission end originally foreseen to 1998 is continuously extends (at the present up to 2010). This mean that SOLAR should be operative for the 10-years period following the mission end (2010-2020).
As shown in figure 2, the number of data sets distributed by SOLAR during this year is considerably reduced. This is happened because during 2009 we had update the RDBM Server, and we have neglect the archive upgrade procedure. Furthermore we need to increase the space disk on the File server to store the new data sets.
In the future with the fully automation of the upgrade procedure SOLAR will be updated every week and it should distribute the same data set available on the main Archive, in this way we hope to increase the SOLAR users.

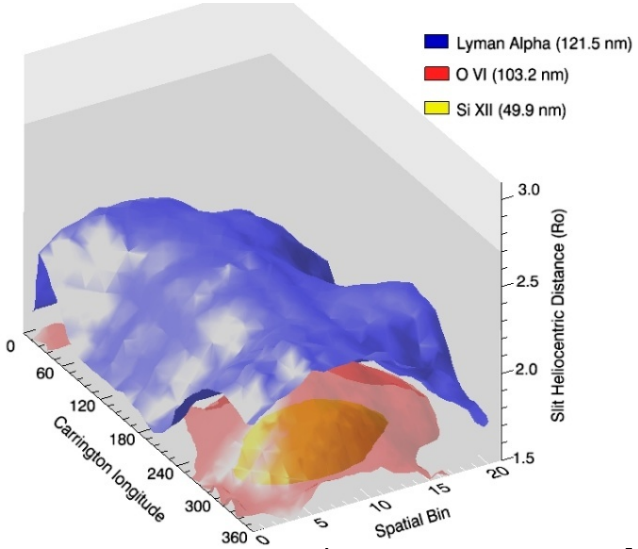

FIG.5 – UVCS 3D carrington map an example

In perspective, we can use SOLAR to distribute the UVCS 3D carrington maps developed by the local Solar Physics group,(CORA et al. 2009a) and test the graphical interface developed for the MINVIT: Multi Instrumental Visualization Tool (CORA et al. 2009b)

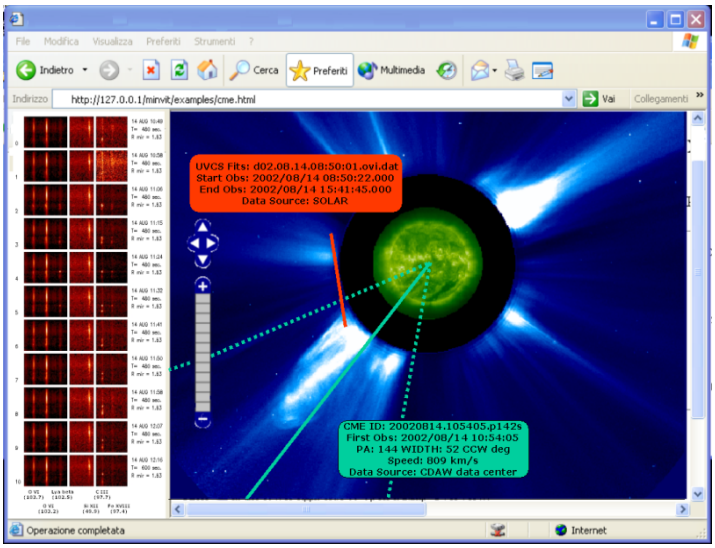

FIG.6 – The graphical user interface of MINVIT: Multi Instrumental Visualization Tool

REREFENCES

SOLAR archive : an introduction to SOho Long term ARchive (SOLAR) developed at the Astronomical Observatory of Turin- Antonucci, E.; Cora, A.; Volpicelli, C.A. -Rapporti Interni OATo ; 58 (2001)

EGSO: User and Science Requirements Document - Kevin Reardon, Silvio Giordano, Ester Antonucci et al - Document number : EGSO-WP1-D2-20021031 (2002)

SOHO Long term ARchive (SOLAR) data base schema- Cora, A.; Volpicelli, C.A.; Giordano, S.- Rapporti Interni OATo ; 70 (2003)

SOLARNET-Italian Solar Archive Federation. The First Italian Virtual Observatory Application - Volpicelli, C. A.; Antonucci, E.; Cora, A.; Giordano, S.; Messerotti, M.; Santin, A.; Zlobec, P.; Severino, G.; Oliviero, M.; DeMarino, I.; and 9 coauthors - Memorie della Societa Astronomica Italiana Supplement, v.9, p.129 (2006)

MINVIT: Multi Instrumental Visualization Tool for Solar Physics - Cora, A.; Giordano, S.; Volpicelli, C. A.; Mancuso, S. Earth, Moon, and Planets, Volume 104, Issue 1-4, pp. 131-133 (2009a)

3D Carrington Maps from UVCS Experiment on SOHO - A.Cora, S.Giordano and S. Mancuso - 3° Solar Orbiter Workshop– May 25-29 2009 (2009b)