SPACEINN

D3.16) 1st report on help-desk and user-support operations (SDO): description of the support granted to access SDO data.

This deliverable is provided by the Max-Planck-Institut für Sonnensystemforschung (MPG).

4th December 2013

GDC-SDO Help Desk for SPACEINN

The Help Desk for the German Data Center for the Solar Dynamics Observatory (GDC-SDO) for the SPACEINN project has been created and is located

http://www.mps.mpg.de/projects/seismo/SpaceInn/gdc-sdo-help-desk.html.

This website provides essential information on how to access SDO data from the GDC-SDO. It contains instructions on how to download, install, setup and operate the Data Record Management System (NetDRMS) that is being used to share SDO data for SPACEINN. It also specifies contact details for technical and user support. Figure 1 shows the layout of the main page.

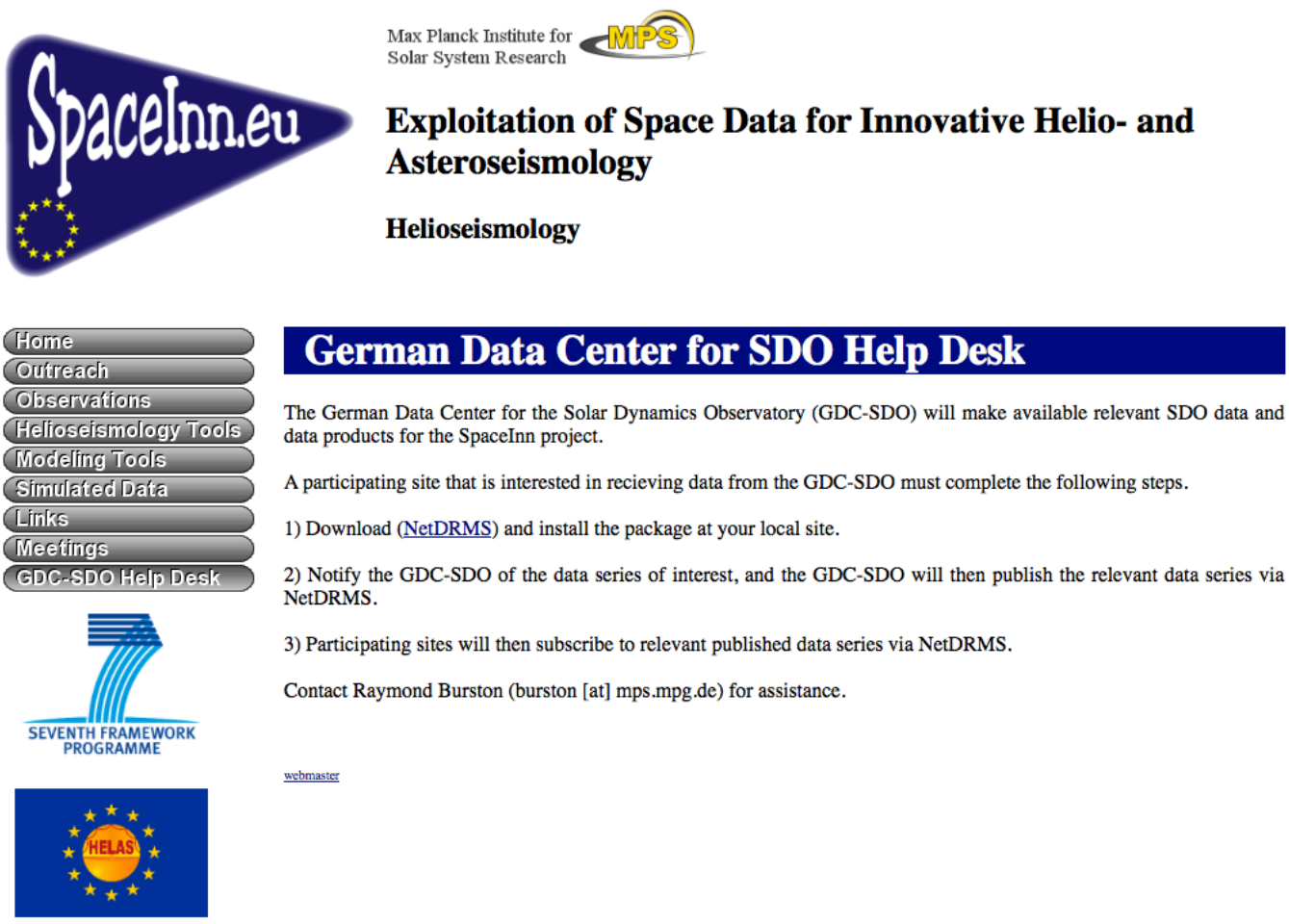

Figure 1. GDC-SDO Help Desk for SPACEINN main page

GDC-SDO

The German Data Center for the Solar Dynamics Observatory (GDC-SDO) is an IT infrastructure hosted by the MPG and funded by DLR, which collects, manages, and stores calibrated HMI data. It is a European master distribution center for HMI images. The GDC-SDO will ensure that HMI images are available for the SPACEINN Project. The core of the GDC-SDO is an online storage capacity of 660 TB plus a tape library/robot with 200 SDLT-4 slots providing an additional 160 TB of near-line archiving capacity. An InfiniBand network ensures sufficient bandwidth is available. The incredibly high-volume of HMI data requires such a dedicated center within Europe to facilitate rapid and transparent data access and scientific analysis to the European community. The Stanford University Joint Science Operations Centre (JSOC) automatically transfers all HMI data directly to the GDC-SDO within minutes of it becoming available at Stanford.

NetDRMS

The underlying foundation for the GDC-SDO is the Data Record Management System (NetDRMS). NetDRMS is software developed by Stanford University (with GSC-SDO support) and is used for the management and automatic transfer of SDO/HMI data. NetDRMS is the primary method for making HMI data available for the SPACEINN project.

NetDRMS consists of several important components. A Storage Unit Management System (SUMS) is in charge of retrieving and accessing the "physical data" (e.g. arrays in FITS files) that reside on the GDC-SDO 840 TB storage area network. A SUMS server listens continuously for requests from NetDRMS users wishing to retrieve or store data on the SUMS disks. It first determines how much space is available on each of the SUMS partitions, and subsequently, stores any specified data in the most appropriate partition. Communication with the SUMS server is typically achieved when a user runs a NetDRMS module or NetDRMS utility commands.

A very important feature of NetDRMS, and this is common practice for very large data sets, is that it separates the meta-data (e.g. keyword value pairs) from the physical data

(e.g. data arrays in FITS files). This is critical for very large data series, as it is significantly faster to search through the relatively small meta-data than e.g. to search through the headers of many FITS files. The information in the meta-data is stored in database tables. Thus another integral ingredient of NetDRMS is the advanced open source database software PostgreSQL (PSQL). This delivers rapid and flexible searches through SDO meta-data to NetDRMS users who are local to GDC-SDO and also to the European scientific community via the NetDRMS infrastructure. A PSQL server is also running and listening continuously, which handles search requests from NetDRMS users.

NetDRMS uses "publication" and "subscription" processes in order to manage SDO sharing. The Joint Science Operations Center (JSOC) in the USA publishes all HMI data since they own those data series. Once a series is published in the NetDRMS sense, then it possible for participating sites that also have a local instance of NetDRMS operational to subscribe to that series. In practice the subscription automatically transfers the relevant HMI meta-data from JSOC to the subscribing site. The subscribing site then has the option of whether to transfer the corresponding HMI images, and furthermore, can select only a subset of images, the latter useful in the case where disk space is limited. A Java Mirroring Daemon (JMD) is then used to transfer the requested images by reading unique record identifiers from a PSQL table. The meta-data is always transferred directly from the publishing site (e.g. JSOC for HMI data) so that there is always a unique source for all meta-data. The meta-data is typically on the order of GBs so there are no bandwidth problems on many subscribing sites obtaining the meta-data from a single source. However, the corresponding images will be transferred from which site is geographically closest to the subscribing site. The images are on the order of TBs, so it is very important to spread the load and reduce the bandwidth on JSOC.

The GDC-SDO has both the SUMS and PSQL servers listening on a single virtual machine. There are also many client machines where users may initiate NetDRMS sessions and communicate to the server programs. Communication is achieved using Remote Procedure Calls (RPCs) over the Local Area Network (LAN). Each of the SUMS disks are mounted on the server machine and the client machines using a Network File System (NFS) therefore allowing quick access to SDO data from any machine. There is

also an Application Programming Interface (API) written in C with about 250 functions, which gives programmers the opportunity to access data directly from within their C codes.

User Support Operations

Technical support was provided to the Kiepenheuer-Institut für Sonnenphysik (KIS) in Freiburg, Germany. KIS intend to subscribe to the HMI data series in the future and consequently they need to install a local instance of NetDRMS. Recall that all subscribers will receive meta-data from the unique source of whoever published the data series of interest, so in this case KIS will receive the HMI meta-data from JSOC, and this is on the order of GBs. However, the HMI images are on the order of many TBs and thus they will be transferred from the GDC-SDO to KIS using a high-performance secure transfer protocol. Two initial interface tests were conducted and completed successfully.

Test 1: Setup relevant Secure Shell (SSH) keys and confirmed they were valid and functional at both the MPG and KIS. This will enable the transfer of HMI images (using a high-performance secure transfer protocol) from the GDC-SDO to KIS without requiring passwords.

Test 2: Setup the NetDRMS program "jsoc_fetch" at the MPG that is used by the Java Mirroring Daemon for transferring images. Checked that KIS could execute this program remotely, hence confirming that correct permissions have been set. The Java Mirroring Daemon uses this program for transferring images.

Summarily, the interface tests between MPG and KIS are complete and were successful. The next steps include assisting KIS install NetDRMS locally, and ultimately help them subscribed to the HMI meta-data (from JSOC) plus corresponding images (from GDC-SDO).

User support was provided to users at IRAP in Toulouse, Paris, whereby the GDC-SDO Help Desk provided information regarding SDO/HMI data.