

Data Management

Your data is stored on the STRW data storage media which can be either the local disk of your workstation, the local-to-the-server direct attached disks of a server or on the bulk storage devices. As part of the data management cycle you need to keep track of your data and make sure that at the end of your employment at the Observatory you either copy your data to your new institute or provide the computer group with a list of pointers to allow them to store your data onto a save archive.

Sources of data

There are three main sources of data in the astronomical community at the Observatory, they are

- Data from the big international observatories
- Data generated in the Sackler and Optical labs
- Data produced by theoretical work and simulations

For the three categories, three different scenarios are in place with regards to data management.

Data from the big international observatories

At the big international observatories data acquisition takes place in a controlled fashion. Data processing is done in the head quarters computing centers and is stored and archived at the storage facilities of sites. In some cases the data are proprietary for at maximum one year, after which they become public.

Each international observatory has a portal that can be accessed to download observational data. So the processed 'raw' data are available always and to any user. Once data has been downloaded and will be used in further processing, it is the responsibility of the user to keep track of the data and to retain its integrity. Most of this functionality is already available in the standard data reduction packages provided by the big observatories. But data can be processed by personal code for which the user needs to keep track of the data products and the data integrity him/herself. The Linux operating system can help out through its file and directory ownerships and access rights.

Depending on the processing unit data can be stored on the local disks at your workstation or on the disks attached to the server that is used for processing. This provides the best data-to-cpu throughput. Data stored at the local workstation should be stored on the /data2 device as this is always a RAID (1 or 5) type of disk. There are no backups for those disks. Local disks in servers are always a RAID (5 through 60) type. Again no backups are made for those disks. It is therefore important that you keep the original data processing software save as it may be needed to reproduce intermediary data product in case of accidental loss.

In general final data products are published and made available to the community. This can be done in several ways, depending on the size of the data or the type of product. For small data set the Centre Données Stellaires at Strassbourg is the standard place when published in European journals. For larger data sets one might opt to make the data available through a project website (which can be hosted by the Observatory) or through the associated big international observatory.

Data coming out of local research will always be stored on local disks (bulk storage) at the Observatory. Just contact the computergroup so that they can copy the data to bulk storage and make it available for further access.

Data generated in the Sackler and Optical labs

At the two Observatory laboratories data is produced by the instruments acquisition board and stored locally on lab computers. Data that will be used in further processing can be stored on the Lab Sackler share. This share is located on the High Availability EqualLogic disk cluster, though a Windows file server.

This share is accessible both from the data acquisition equipment and from the personal desktops of the lab associates. The storage device is on a special secure medium and is snapshotted regularly as a form of backup.

Users are responsible for keeping track of the data and retaining its integrity. The Windows accounts and access rights can help in this task.

Data produced by theoretical work and simulations

As a result of theoretical calculations or model simulations, data may be produced that are further processed to allow astronomical interpretation. These calculations can take place on the desktop or on server levels, where storing the data will most efficiently be done on the local-to-the-cpu storage devices. So either the local disks of the desktop or the local attached disk devices for the servers.

Similarly to the data processing for data coming from the big observatories it is the users responsibility keep track of the data and its integrity. As most of this processing is done on the Linux computing environment, similar facilities can help out in this process.

From:

<https://helpdesk.physics.leidenuniv.nl/wiki/> - **Computer Documentation Wiki**

Permanent link:

<https://helpdesk.physics.leidenuniv.nl/wiki/doku.php?id=linux:datamanagement&rev=1511787845>

Last update: **2017/11/27 13:04**

