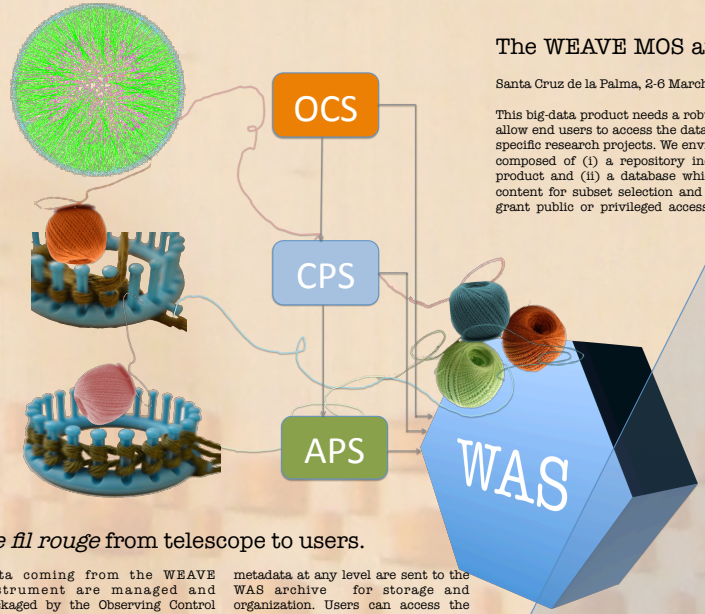
WAS: Textures of the WEAVE Science Archive

E. Molinari, M. Lodi, J. Guerra, C. Benn, L. Dominguez
TNG – Fundación Galileo Galilei, Rambla José Ana Fernández Pérez, 7 – Breña Baja – Spain
INAF - IASF Milano, Italy; Isaac Newton Group of Telescopes, Santa Cruz de la Palma, Spain
(emilio.molinari@tng.iac.es)

The WEAVE MOS and IFU spectrograph will produce millions of spectra

Santa Cruz de la Palma, 2-6 March 2015.

This big-data product needs a robust high-performance tool to allow end users to access the data reliably and easily for their specific research projects. We envisage here a Science Archive composed of (i) a repository including every WEAVE data product and (ii) a database which intelligently indexes the content for subset selection and data retrieval. A GUI will grant public or privileged access to the archive, giving the possibility of data mining, selection and retrieval of the data from different configurations of the WEAVE instrument, i.e. the spectra or IFU meta-images produced by the WEAVE data pipelines CPS and APS. The underlying database structure and performance will be guaranteed with a redundant network of disks and CPUs, and robust open-source software for which troubleshooting support is available for a high fraction of the time.

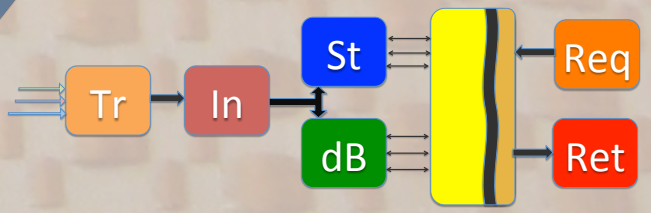


Le fil rouge from telescope to users.

Data coming from the WEAVE instrument are managed and packaged by the Observing Control Software (OCS) and then sent to the archive (WAS) and to the Core Processing Software (CPS) in order for spectra to be extracted and calibrated. In the next phase the APS (Advanced Processing Software) turns the spectra into physical observable quantities. All the metadata at any level are sent to the WAS archive for storage and organization. Users can access the wealth of data (from raw to final astrophysical quantities and IFU maps) through the Web interface to the archive. Regular, public releases of data will be overseen by the WEAVE project board.

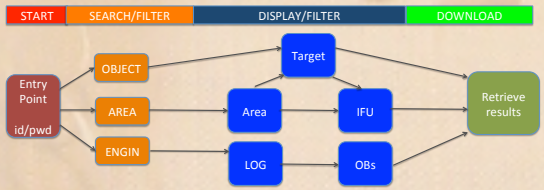
WAS inside: the data flow in its subsystems.

Data arrive to WAS (WEAVE Archive System) through a suitable and secure transportation system [Tr] which may make use of VPNs or more sophisticated techniques like GlusterFS in order to simplify the management on the producers' side (OCS, CPS and APS). The subsequent ingestion process [In] drives all FITS file and tables into a redundant repository storage [St] and populates the series of databases [dB] which will organize data and metadata in an indexed order suitable for any astrophysical research.
A layer of organic processes will have access to storage and databases and will allow other (external) processes to send requests (Req) and get results (Ret). Among plug-in programs we foresee a web based common GUI interface, a prerelease organizer used before public releases and the possibility to develop ad-hoc analysis which need to access the whole amount of data.



Users' GUI access to the archive.

A fairly complex gui is required to search for data obtained with the different observing modes (fibers, IFUs). The process involves first a search by e.g. area or object, then a display phase(target, area, IFU, observing log) which will offer the second filtering selection before the retrieval of data or frames.



Architecture down to the foundations.

The WEAVE archive architecture will sit atop of the latest technologies in work load distribution and data management in order to minimize the points of failure and at the same time, increasing its reliability and scalability.
The archive will be able to ingest (according to the initial estimations) more than 50GB of data for around 10,000 targets on daily basis. The raw FITS files and their sub-products, will be stored on file-system with a previous quality check that conforms to the WEAVE standard. The ingestion process continues and sends the files into the database.

On the other hand, the GUI will be built on web technology for the maximum flexibility to the scientist community. Web technologies such as HTML5 and Ajax will be part of the design to increase of the user experience on the browser. Initially, the web front end will be used as a search engine to display the target and IFU features. Later on, the GUI could also give access to the database Analysis adding new features such as Data Mining, Graphs and Machine Learning which are an extra bonus to capacity of the existing scientist data.

WEAVE stands for WHT Enhanced Area Velocity Explorer. The name also reflects the challenge of positioning correctly in a 2-dimensional plane a large number of fibres. Weaving also happens to be a long-standing traditional handicraft on the island of La Palma! The background image comes from the Silk Museum in El Paso, on the western side of the island.

Multi-Object Spectroscopy in the Next Decade: Big Questions, Large Surveys and Wide Fields - Santa Cruz de La Palma, Canary Islands 2-6 March 2015