Poster presentation (ID 103, board S4P9), at The 9th IVS General Meeting "New Horizons with VGOS", 13-17 March 2016, Johannesburg, South Africa

The new Geodetic Research Data Management System at HartRAO

Glenda Coetzer[1], Roelf Botha[1]
[1] *Hartebeesthoek Radio Astronomy Observatory, South Africa*

Corresponding Author: glenda@hartrao.ac.za

Introduction

The Hartebeesthoek Radio Astronomy Observatory (HartRAO) (see *Figure 1*) participates in radio astronomy and space geodesy research activities. The Radio Astronomy Programme techniques include astronomic, astrometric and geodetic Very Long Baseline Interferometry (VLBI) observations using both 26- and 15-m diameter radio telescopes. The Space Geodesy Programme utilizes a Satellite Laser Ranger (SLR), numerous Global Navigation Satellite System (GNSS) reference stations and geodetic VLBI. A Lunar Laser Ranger will become operational within 2 years, and together with the recent addition of seismic stations these further enhances the geodetic component. HartRAO also participates in the African VLBI Network (AVN) and is responsible for the data correlation and storage for the joint experiments, due to start in the near future.

The expansion of the Space Geodesy Programme with the addition of seismic measurement equipment will generate huge volumes of additional data. We procured equipment to build and install 12 remote GeoStations consisting of seismic, GNSS and metrology measurement systems (see *Figure 2*). The assembly of 12 new GeoStation boxes have recently been completed and these will be installed in Southern African regions during the next 2 years. Systems have already been installed at HartRAO, Matjiesfontein and Marion Island.

The data volumes will be stored and archived at HartRAO (*Table 1*) for research purposes and will also be made available to the scientific community. This necessitated the design and implementation of a new, next-generation Geodetic Research Data Management System (GRDMS) which combines all the datasets into one database, to cater for current and future data volume requirements (see *Figure 3*). Professionals from different disciplines are working together in the design and implementation of the HartRAO GRDMS, as depicted in the interaction model (*Figure 4*).

In terms of processing, HartRAO is collaborating with the CSIR Centre for High Performance Computing (CHPC) and have obtained and will soon operate a Dell Westmere cluster (*Figure 5*) containing 2880 CPU cores (37.1 TFlops) and 8640 GB RAM. This cluster provides for ample processing capacity for AVN VLBI correlation, GNSS data products as well as for Metrology, SLR and LLR analysis. The new seismic network has a specially designed data processing and storage cluster, which arrived at HartRAO during February 2016, and will be commissioned soon.

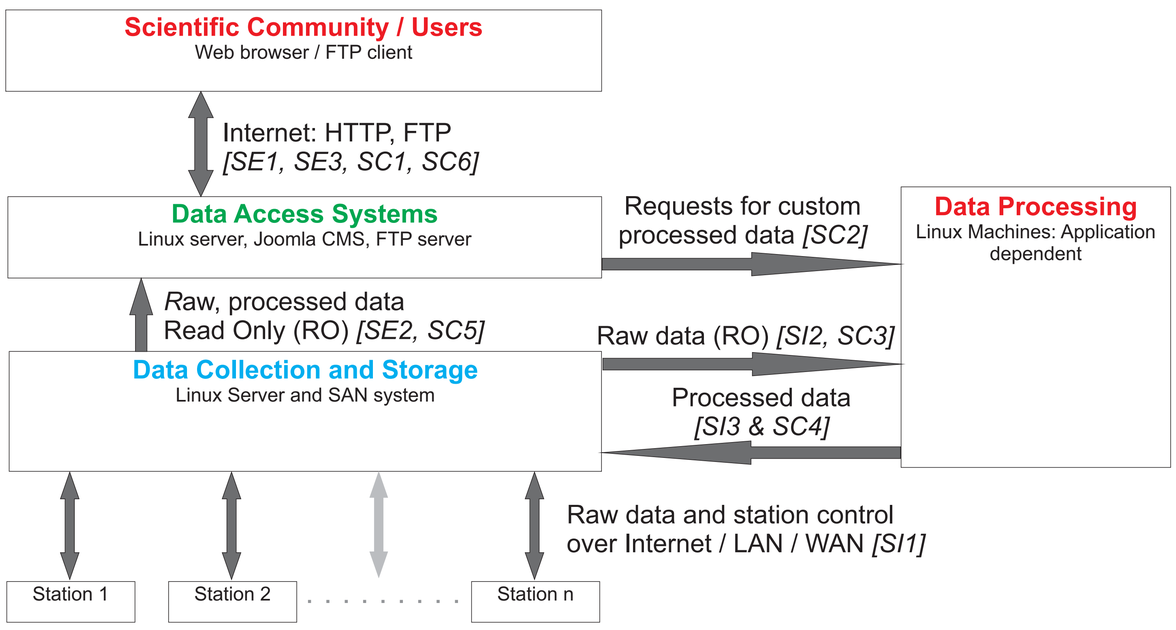

Figure 3: Conceptual system design of the GRDMS (Brackets denote steps in data cycle, as addressed by Coetzer 2014).

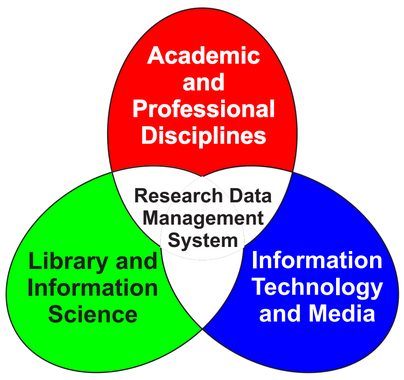

Figure 4: Interdisciplinary interaction between role players in the data management process (adapted from Corrall 2008; slide 6).

Design of the GRDMS

Figure 3 depicts the conceptual design of the GRDMS. Progress has been made towards implementation of this:

- HartRAO has fundamental geodetic equipment, namely the VLBI radio telescopes, GNSS reference stations and the SLR.
- Twelve newly assembled GeoStations are to be installed at remote sites in the Southern African region.
- Recently obtained a Dell Westmere processing cluster from the CSIR CHPC which will be set up in collaboration with the CSIR CHPC, for processing most geodetic data.
- A specialized processing and storage cluster for use with the seismic component of the equipment network.

The Data Collection and Storage server (a new system, in planning phase) will contain the data archive content, which will include:

- Observational data (real-time, usually irreplaceable raw data)
- Metadata (non-standard metadata, data type specific, extracted from incoming files). Metadata are required for archive management, report generation, and data distribution reporting.
- Products (pre- & post-processed data)
- The archive structure and software will allow for the extraction of the required metadata and move files to the appropriate archive structure.

The Data Access Systems server (in planning phase), will provide for data discovery with various access channels to the data archive:

- Data access (real-time and static) will require user authentication, which will be handled through a User Registration System (URS), allowing the gathering of metrics for usage reporting.
- DSpace (open source software) will be used as the main repository management system
- An intelligent data structure translator will cater for all standardised structures as specified by international service providers (1 copy – many access paths). Data in these files will also available in the "standard" format (SSSSDDD#.YYT.Z). This access method provides flexibility for the user community.
- Web application for display of data holdings
- Query interfaces will allow users to enter spacial and temporal parameters to determine sites of interest

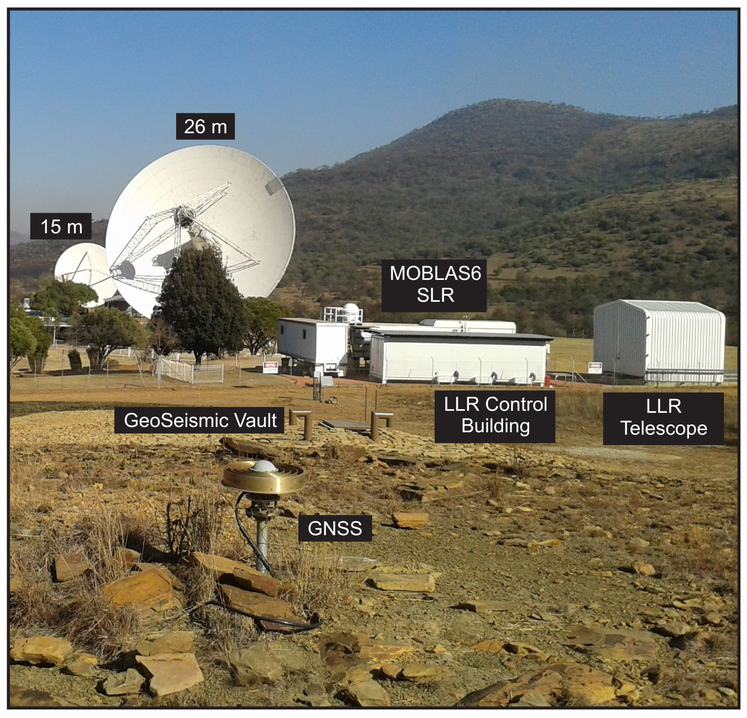

Figure 1: HartRAO's astronomy and geodetic equipment types. The first of the new-generation GeoStation systems is installed in the GeoSeismic vault, for which data capture commenced on the 22nd of September 2015.

Figure 2: A GeoStation control system assembly. Twelve stations will be installed throughout the Southern African region.

	Storage unit / frequency	Raw Data Type	Some products	MB / station / day	Current space GB	Required space GB *
SLR	Satellite pass	CPF file	Orbital data	10	1	37
LLR	Tracking session	Compressed text files	CPF and Orbital data	10	0	37
VLBI (products) currently 2, soon 3)	Per experiment	(NGS card files)	ITRF, ICRF, station position and motions	10	1	300
GNSS (12).	Daily 24hr	Rinex and Sinex files	ITRF, station position and motions, PWV	300	150	45150
Seismic (12)	Daily 24hr	Seedlink records	Seismic event data	300	30	45030
Gravimetric	Daily 24hr	Compressed text files	Gravimetric variations	3	0	12
Meteorological	Daily 24hr	Compressed text files	Variations, trends	5	1	18
Tide gauge	Daily 24hr	Compressed text files	Tidal periods, extremes	5	1	20
				643	**184**	**90604**

Table 1: Current data volumes and future requirements for the HartRAO Space Geodesy Programme

Figure 5: The Dell Westmere cluster front and back (left and right photos) as in operation at the CHPC, before disassembly for shipment to HartRAO.

Future Activities

Activities currently being planned for the near future include:

- GeoStations installations: at least 4 of the 12 stations will be installed this year: Ghana, Sani Pass, Necsa and MeerKAT
- CHPC Cluster commissioning for AVN and other processing: a storage room will be cleared out and converted to a server and data centre. The CHPC cluster and Geodetic processing servers will be installed here
- Identification and procurement of an expandable storage solution for the Data Collection and Storage system
- Implementation of Data Access Systems on virtual machines

Acknowledgements

The authors would like to acknowledge Funding awarded by the National Equipment Programme (NEP) of the NRF to start the development of the Co-located Academic Network. The support received from Prof. Ludwig Combrinck and the GeoStation assembly work done by Mr. Juan Grey is also acknowledged.

References

Coetzer, G. L. et al. 2014, Library and Information Services in Astronomy VII, ASP Conference Series, Vol. 492, p 22

Corrall, S. 2008, Research Data Management Forum, http://www.dcc.ac.uk/sites/default/files/documents/RDMF/RDMF2/07%20Corrall.pdf