Highlights of Astronomy, Volume 14
IAU XXVI General Assembly, 14-25 August 2006
Karel A. van der Hucht, ed.

doi:10.1017/S1743921307011878

‘Retooling’ data centre infrastructure to support the Virtual Observatory

Séverin Gaudet

Canadian Astronomy Data Centre, Herzberg Institute of Astrophysics,
National Research Council, Canada
email: severin.gaudet@nrc.gc.ca

The Canadian Astronomy Data Centre manages a heterogeneous collection of data from the following ground and space-based telescopes: CFHT, DRAO, *FUSE*, Gemini, *HST*, JCMT, and *MOST*. The archive data models implemented for these data collections are ten years old and pre-date two important developments: the Virtual Observatory and the systematic generation and management of data products. Three years ago, we began the process of supporting access to processed data products through IVOA protocols such as SIA by building a layer over the archive data models. Today, we now realise that this approach of layering VO models on archive models is not sufficient and that every archive must be re-tooled to properly support the VO – from the storage model through to the query, processing and access models. The CADC has begun an ambitious software development effort to implement a new infrastructure to serve both telescope archive and Virtual Observatory needs.

At the core of this new infrastructure is the Common Archive Data Model. This data model, to be implemented in each archive, will standardize the way observations are characterized and relationships are expressed. The design of this model is inspired from IVOA standards under discussion (Observation, Characterization, Simple Image Access, Simple Spectral Access) and leverages the CADC’s archive, data modeling and data engineering experience. The observation characterisation is based on FITS WCS papers I, II and III. The data model will become the sole metadata interface between the CADC’s archives and the CADC’s data warehouse upon which the VO services are built. The data model will also allow the CADC to adapt to the evolving VO standards.

The CAOM is only but one element of the retooling of the CADC to support the VO. The storage model is being modified to support caching of data products for synchronous retrieval, the generation of cutouts, and the decompression and recompression of files in streaming modes of access. The processing model is also being significantly re-designed to use the CAOM and to persist processing metadata and the relationships between complex multi-observation products and their simpler inputs. The retrieval model is being changed to support additional programmatic interfaces and the retrieval of complex data packages. An access control model is being developed to support authenticated access to proprietary data through VO interfaces.

These changes represent fundamental changes to a mature data centre infrastructure that, in 2005, archived more that 61 terabytes of data and distributed 38 terabytes to more than 2,500 distinct IP address. So what path has the CADC chosen? The design process began in the fall of 2005. This was followed by a prototype implementation for the JCMT archive with its new instrumentation scheduled for released at the end of 2006. This will be followed by an evaluation of the prototype (lessons learned) and planning for the conversion of the remaining CADC archives. The target completion date is set for the end of 2007.

https://doi.org/10.1017/S1743921307011878 Published online by Cambridge University Press