

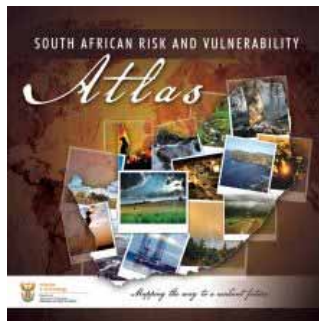


## SAEON and CHPC in collaboration agreement to develop data platforms

Wim Hugo, Systems Engineer, SAEON

The Centre for High Performance Computing (CHPC) and SAEON have entered into a collaboration agreement that will assist both organisations with the development of research and data platforms.

As a starting point, CHPC has provided a hosting platform for a test version of the SAEON Data Portal, with a view to evaluating the feasibility of a longer-term, physical home for the SAEON Data Portal and its associated sister sites. If successful, this would result in the South African Risk and Vulnerability Atlas, South African Earth Observation System, and World Data Centre for Biodiversity and Human Health all being hosted at CHPC, in addition to the SAEON Data Portal.



The South African Risk and Vulnerability Atlas is one of SAEON's sister sites that might end up being hosted at CHPC.

### Collaborative research

The potentially massive meta-data repository required by these systems and the size of the associated data sets both lead to potential future collaborative research:

1. Improved indexing and discovery mechanisms to increase the speed of searching for data sets. This includes the possibility of broadened, more effective searches through the real-time application of controlled vocabularies and the semantic web.
2. Given the computing power at CHPC, research can be done on the correlation and statistical analysis of large spatial data sets on demand.
3. Improved usage of large, multi-dimensional data sets: the climate and weather research domain, as well as Earth observation data sets in general, produce very large data sets for which management practice, visualisation, and real-time querying all present challenges. The opportunity also exists to test the aggregation of physically disparate, but logically related databases. These aspects align with the Very Large Databases (VLDB) project at CHPC.
4. Concepts of networked databases allow re-utilisation of data in distributed locations to fit many semantic models. The opportunity exists to evaluate and test this concept.
5. Several applications require the processing of spatial data that are physically distributed. SAEON can develop prototypes as reference implementations for the standardised implementation of Web Processing Services.

In terms of the agreement, CHPC will not only benefit from the contributions to its VLDB project, but will also be in a position to evaluate a reference implementation in the Earth and Environmental sciences domain.

A virtual server was made available at CHPC in September, and a test platform will be operational in October 2010.