Datacubes: Exploiting Big Earth Data Better

With the unprecedented increase of orbital sensor, in-situ measurement, and simulation data there is a rich, yet not leveraged potential for getting insights from dissecting datasets and rejoining them with other datasets, effectively establishing a "datacube" paradigm with the ultimate goal to allow users to "ask any question, any time" thereby enabling them to "build their own product on the go".

One of the most influential initiatives in Big Geo Data is EarthServer which is demonstrating new directions for flexible, scalable datacube services based on innovative NewSQL technology. Researchers from Europe, the US and recently Australia have teamed up to rigourously materialize the concept of the datacube. Such a datacube may have spatial and temporal dimensions (such as a satellite image time series) and may unite an unlimited number of individual images. Independently from whatever efficient data structuring a server network may perform internally, users will always see just a few datacubes they can slice and dice. EarthServer has established client and server technology for such spatio-temporal datacubes. The underlying scalable array engine, rasdaman, enables direct interaction, including 3-D visualization, what-if scenarios, common EO data processing, and general analytics. Conversely, EarthServer has significantly shaped and advanced the OGC Big Geo Data standards landscape based on the experience gained.

Phase 1 of EarthServer has advanced scalable array database technology into 100+ TB services; in phase 2, a federation of Petabyte datacubes is being built in Europe and Australia to perform ad-hoc querying and merging. We present core results and outline the impact on the OGC, ISO, and INSPIRE datacube standards landscape.