**[EarthServer](http://www.earthserver.eu/): Data Fusion on Distributed Petascale Datacubes**

## Use Case

* **Data sets:** 
  + 3D x/y/t hyperspectral satellite imagery served by European Space Agency (ESA, maintained by MEEO); Plymouth Marine Laboratory (PML); Jacobs University;
  + 4D x/y/z/t weather forecast data served by European Centre for Medium-Range Weather Forecast (ECMWF).
* **Data volumes:** between 20 TB and 1 PB per cube
* **Challenges:**
  + extract spatial/temporal subsets from datacubes, process datacubes, and fuse datacubes sitting at intercontinental distances.
  + Do so (i) simple & user friendly and (ii) through OGC Big Geo Data standards.
  + Provide integrated retrieval capabilities where data and metadata can be mixed in queries, thereby achieving a common information space.
* Advance over the State of the Art:
  + Use [Array Databases](http://en.wikipedia.org/wiki/Array_DBMS) with their datacube query languages; concretely, in EarthServer [rasdaman](http://www.rasdaman.org) (“raster data manager”) is used which has coined the research field of Array Databases.
  + Automatically distribute incoming queries across a rasdaman peer network while optimizing according to various criteria like data location and network traffic minimization.
  + Define, implement, and standardize query languages capable of integrating data and metadata retrieval (concretely: OGC WCPS and ISO SQL).

## Background

EarthServer is an ongoing intercontinental initiative advancing Big Earth Data access and analytics through rigorous application of the datacube paradigm. After its first phase (2011-1014) it has now entered its next phase (2015-2018), adding Australia to the group of European and US partners.

EarthServer-2 makes**Agile Analytics on Big Earth Data Cubes** of **sensor, image, simulation, and statistics data** a com­mod­ity for non-experts and experts alike through

* navigation, extraction, aggregation, and recombining of any-size space/time data cubes;
* easy to install & maintain value-adding services extending the existing portfolio of data and compute centers;
* based on open standards, in particular: the OGC Big Data standards and the forthcoming ISO SQL/MDA (“Multi-Dimensional Arrays”) standard.

In the **Joint Research Activity**, the project will advance the existing, world-leading **rasdaman Array Database** technology wrt. query functionality, inter-federation data processing with auto­matic data and query distribution, tape archive integration, and 3D/4D visualization based on NASA’s virtual globe technology.

In the **Services Activity**, large data centers (ECMWF, PML, MEEO/ESA, Ja­cobsUni, NCI Australia) will set up **water, air, weather, and planetary services** on 3D & 4D **Pe­tabyte-size** data cubes with user-tailored clients for both visual and textual queries for ad-hoc mix&match.

In the **Networking Activity**, the project will **advance open Big Data standards in OGC, RDA, and ISO** (in particular: **write ISO SQL/MDA**). Further, all adequate channels will be used for strong dis­semination & exploitation, specifically: writing a monograph explaining OGC Big Geo Data stan­dards; scientific publications & active conference organization; Earth science data user work­shops for each domain addressed; actively contributing technology & experience to GEO / GEOSS and further bodies; establish a standardized **Big Geo Data benchmark** and run it against Earth­Server-2 and further relevant systems.