



## EarthServer: Information Retrieval and Query Language

Thanassis Perperis, Panagiota Koltzida, and George Kakaletris

Research and Innovation Centre in Information, Communication and Knowledge Technologies "Athena" - Institute for the Management of Information Systems (IMIS), Athens, Greece

Establishing open, unified, seamless, access and ad-hoc analytics on cross-disciplinary, multi-source, multi-dimensional, spatiotemporal Earth Science data of extreme-size and their supporting metadata are the main challenges of the EarthServer project ([www.earthserver.eu](http://www.earthserver.eu)), funded by the European Commission under its Seventh Framework Program. One of EarthServer's main objectives is to provide users with higher level coverage and metadata search, retrieval and processing capabilities to multi-disciplinary Earth Science data. Six Lighthouse Applications are being established, each one providing access to Cryospheric, Airborne, Atmospheric, Geology, Oceanography and Planetary science raster data repositories through strictly WCS 2.0 standard based service endpoints. EarthServers' information retrieval subsystem aims towards exploiting the WCS endpoints through a physically and logically distributed service oriented architecture, foreseeing the collaboration of several standard compliant services, capable of exploiting modern large grid and cloud infrastructures and of dynamically responding to availability and capabilities of underlying resources. Towards furthering technology for integrated, coherent service provision based on WCS and WCPS the concept of a query language (QL), unifying coverage and metadata processing and retrieval is introduced. EarthServer's information retrieval subsystem receives QL requests involving high volumes of all Earth Science data categories, executes them on the services that reside on the infrastructure and sends the results back to the requester through a high performance pipeline.

In this contribution we briefly discuss EarthServer's service oriented coverage data and metadata search and retrieval architecture and further elaborate on the potentials of EarthServer's Query Language, called xWCPS (XQuery compliant WCPS). xWCPS aims towards merging the path that the two widely adopted standards (W3C XQuery, OGC WCPS) have paved, into a new construct allowing "mixed search" on both OGC coverages and XML-represented metadata and also returning "mixed results" further enabling seamless geospatial and array, combined data and metadata, processing under a familiar syntactic formalism. xWCPS is a superset of WCPS closely following XQuery's syntax and philosophy, further extending it with capabilities to handle coverages, array and multidimensional data, allowing different degrees of compliance to its results and opening new possibilities for data definition, processing and interoperability. Our long term vision for xWCPS is from the one hand to enable coverage and corresponding metadata retrieval, irrespective of their actual origin and form and from the other hand to offer syntactic constructs for data definition and data manipulation. Thus xWCPS queries could potentially employ distributed services to access diverse, cross-disciplinary and physically distributed data sources, data within them and metadata about them without directly specifying which coverages to employ, which parts come from metadata and which come from data processing (aggregates) functions. Respectively a Data Definition Language could allow schema definition and a Data Manipulation Language could enable updates, inserts, and deletes of data handled by an xWCPS system. xWCPS's specification is currently in draft form. We intend to initiate the corresponding OGC standardization activity with the finalization of the specification of the language.