

# AN APP STORE FOR BIG DATA

## INGO SIMONIS REPORTS ON AN OGC INITIATIVE TO MAKE USING BIG DATA AS EASY AS DOWNLOADING AN APP FROM A STORE

For several years, OGC has been developing in collaboration with the European Space Agency (ESA), Natural Resources Canada and other OGC member sponsors a standards-based software architecture that enables the deployment and execution of data processing applications close to the physical location of Big Data, such as Earth observation data or outputs from models. More than this, by making Big Data more findable, accessible, interoperable and reusable (FAIR), the architecture enables a marketplace of data processing apps that will benefit developers, cloud infrastructure providers, data providers, and data consumers alike.

The primary goal of this architecture is to enable the analysis of truly Big Data while minimising data transfer between repositories and application processes. This is particularly important, as the amount and resolution of data is increasing significantly faster than internet bandwidth is – while the cost of cloud-based storage and processing is falling.

ESA's "EO Exploitation Platforms" initiative, which began in 2013, aimed to achieve a paradigm shift from "bring the data to the user" (the user downloads the data) to "bring the user to the data" (move user exploitation

to cloud-hosted environments with collocated computing and storage). This will lead to the development of a platform-based environment that provides infrastructure, data, computing and software as a service.

However, data from different providers is stored on different cloud systems. To unite existing and future data processing resources, there is a need to define standardised interfaces that facilitate the federation and interoperation of such scattered resources – enabling developers to create a single app that can run across the many different cloud systems with only minimal adaptation. This then allows the data consumer to efficiently access and consume the disparate services of different providers seamlessly, even chaining together the outputs of one data processing application to feed the input of the next.

We have therefore been working to create a standards-based architecture that enables this "application to the data" paradigm to be applied to diverse platforms, including those that focus not just on Earth observation data, but Big Data more broadly. Under OGC's Innovation Program – specifically, OGC Testbeds 13, 14, 15 and (currently) 16, as well as a number of pilots

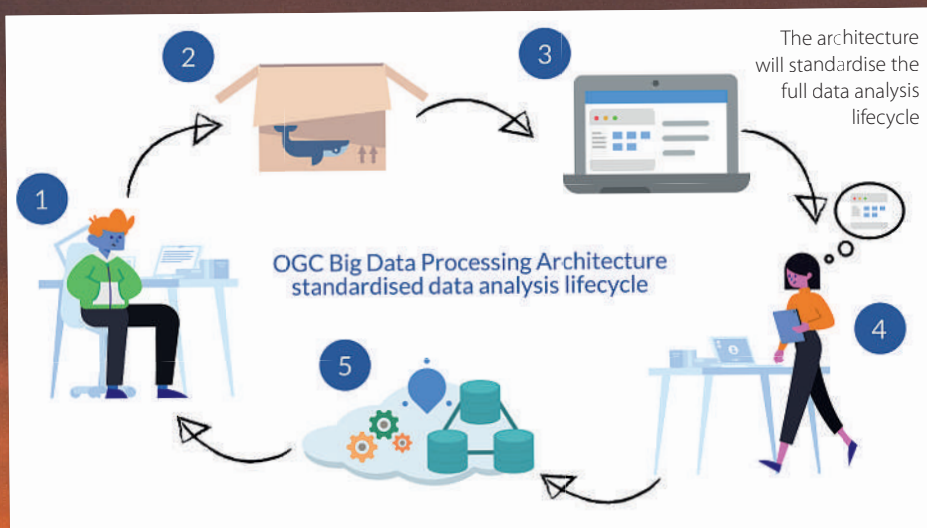
and other innovation initiatives – OGC members have been defining, developing and testing the required standardised interfaces and related solutions that comprise this Big Data Processing Architecture.

### A new architecture

The Big Data Processing Architecture features a set of emerging specifications that will standardise the full data-analysis life cycle, including: application development and description; containerisation; registration at app stores; discovery and on-request deployment in cloud environments; parameterised execution; and final result access. This lifecycle occurs in harmony with business functions such as authenticated user identity, access controls, quoting and billing for the resources consumed.

It enables application developers and consumers to interact with simple APIs that abstract the underlying complexity of data handling, scheduling, resource allocation and infrastructure management. It consists of the following logical components:

- Application Developers that develop data processing and analysis applications.
- Application Consumers requesting the execution of these applications on remote data and processing platforms.
- One (or more) Docker Hubs that enable the storing of the processing applications, accessible to the Data and Processing Platform(s) and vice versa.
- One (or more) Exploitation Platforms to register applications, to chain these into workflows, and to request the deployment and execution on the Data and Processing Platforms.





App Developer



Docker Hub



Data & Processing Platform(s)



App Consumer



Exploitation Platform(s)



Components of the Big Data Processing Architecture

## TESTING

The maturity of the Big Data Processing Architecture is currently being tested as part of the OGC Earth Observations Applications Pilot, with results expected to be published this August. The outputs of this pilot and other related initiatives will soon move to OGC's Standards Program for future release as official OGC Standards. Information on the emerging specifications, as well as detailed reports from the work undertaken in previous Testbeds and other Innovation Initiatives, are available on [docs.opengeospatial.org/per](https://docs.opengeospatial.org/per)



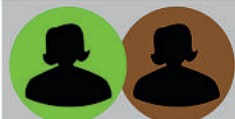
Alice & Ivan

Create apps to be run on Earth Observation Cloud App Architecture

Discovers and executes apps on the EO Cloud App Architecture



Carol



Lucy & Susan

Provide a Data & Processing Platform upon which Alice's app can be executed

Interested in viewing the results of Bob's processing activity



Oliver



Bob

Creates composite apps from those published by Alice & Ivan

Manages an Exploitation Platform through which Carol can chain Alice's or Ivan's apps, and Bob can access Alice's, Ivan's, or Carol's apps



Kami



Peter

Provides an online Database against which Ivan's app can be executed by Kami's Exploitation Platform

Has a large number of datasets and is interested in providing an alternative Cloud Processing Environment within which Alice's app can be executed



Thomas

The architecture is currently being tested in real-world multi-platform scenarios, like the one shown here

- Application developers can develop multi-platform data processing applications in their local environment and sell/publish them on an app store.
- Application consumers can discover available applications (and even chain them together), enabling them to access and process more data, more easily, using simple, consistent, reproducible and shareable workflows. This marketplace would not just benefit app developers and consumers: it would additionally benefit cloud infrastructure providers, as they can sell access to the processing and storage resources required, and data providers, as they can sell piecemeal access to their data for processing.

To simplify access, all communication on the platform is established in a web-friendly way, implementing the emerging next generation of OGC services built on top of the OGC APIs for features, coverages and processes.

Additionally, application consumers need only provide the desired area/time of interest (or other parameters) of the data that they want the application to process. The results are then returned through OGC standardised interfaces – Web Feature Service, Web Coverage Service instances or the latest APIs. In the case of workflows that execute a number of applications sequentially, the exploitation platform takes care of the transport of data from one process to the next. Upon completion, the application consumer is provided with an endpoint to retrieve the final results.

Such a marketplace is a great example of how by making location information more FAIR, OGC's combination of standards, innovation and the expertise of its members successfully connects people, communities, technology and decision-making for the good of society.

*Ingo Simonis is director, OGC Innovation Program & Science ([www.ogc.org](http://www.ogc.org))*

- One (or more) Data and Processing Platforms, where applications are executed in situ with the data. The basic idea is that each Data and Processing Platform provides a standardised interface that allows the deployment and parameterised execution of applications. The Exploitation Platform, then, allows the chaining of different applications (even across different Data Processing Platforms) into workflows with full support for quoting and billing.

### There's an app for that

Together, the components of the architecture allow a marketplace – a sort of 'app store' for Big Data processing – where: