

Chapter 2

Standards and EO Data Platforms



Ingo Simonis and Karel Charvát

Abstract In the digital bio-economy like in many other sectors, standards play an important role. With “Standards”, we refer here to the protocols that describe how data and the data-exchange are defined to enable digital exchange of data between devices. This chapter evaluates how Big Data, cloud processing, and app stores together form a new market that allows exploiting the full potential of geospatial data. This chapter focuses on the essential cornerstones that help make Big Data processing a more seamless experience for bioeconomy data. The described approach is domain-independent, thus can be applied to agriculture, fisheries, and forestry as well as earth observation sciences, climate change research, or disaster management. This flexibility is essential when it comes to addressing real world complexities for any domain, as no single domain has sufficient data available within its own limits to tackle the major research challenges our world is facing.

2.1 Introduction

In the digital bio-economy like in many other sectors, standards play an important role. That is especially the case in exchanging digital data. With “Standards”, we refer here to the protocols that describe how data and the data-exchange are defined to enable digital exchange of data between devices. Such standards enable interoperability between all participating players and ensure compatibility. Standards reduce transaction costs of sharing data and often promote competition, as users can easily change suppliers. Users are not ‘locked in’ to a closed system. Standards often support innovation, or provide a foundational layer that new innovation is built on.

This chapter evaluates how Big Data, cloud processing, and app stores together form a new market that allows exploiting the full potential of geospatial data. There

I. Simonis
OGC, Wayland, USA

K. Charvát (✉)
Lesprojekt-sluzby, Ltd., Zaryby, Czech Republic
e-mail: charvat@lesprojekt.cz

is a growing standards landscape for Big Data and cloud processing. There are new standards and industry agreements to handle orthogonal aspects such as security or billing. Still, an interoperable, secure, and publicly available Big Data exploitation in the cloud remains a challenge. It requires a set of standards to work together, both on the interface as well as the product exchange side. Related technologies for workflow and process orchestration or data discovery and access come with their own set of best practices, as well as emerging or existing standards.

Within the knowledge-based or data-driven bioeconomy, data and information sharing is an important issue. The complexity is high, as long supply chains with a variety of influencing factors need to be integrated. Often, bioeconomy information systems lack standardization and show a poorly organized exchange of information over the whole value and supply chain. Although arable and livestock farming, forestry and fishery have their own specific needs, there are many similarities in the need for an integrated approach.

DataBio identified a set of relevant technologies and requirements for the domains of agriculture, fisheries, and forestry. There is an extensive list of interfaces, interaction patterns, data models and modelling best practices, constraint languages, or visualization approaches. Together with the Open Geospatial Consortium, the worldwide leading organization for geospatial data handling, DataBio contributed to the development of emerging standards that help forming new data markets as described above. These markets are important for everyone from the individual farmer up to the Big Data provider. They will allow the exploitation of available data in an efficient way, with new applications allowing targeted analysis of data from the farm, fishery, or forest level, all the way up to satellite data from Earth Observation missions.

The underlying technology shifts have been implemented mostly independent of the (bioeconomy) domain. They have been driven by mass-market requirements and now provide essential cornerstones for a new era of geospatial data handling. The emerging standards define how the generic cornerstones need to be applied to Earth observation data discovery, access, processing, and representation.

This chapter focuses on the essential cornerstones that help make Big Data processing a more seamless experience for bioeconomy data. The described approach is domain-independent, thus can be applied to agriculture, fisheries, and forestry as well as earth observation sciences, climate change research, or disaster management. This flexibility is essential when it comes to addressing real world complexities for any domain, as no single domain has sufficient data available within its own limits to tackle the major research challenges our world is facing.

2.2 Standardization Organizations and Initiatives

ISO

ISO is the International Organization for Standardization, which develops and publishes international standards. ISO standards ensure that products and services

are safe, reliable and of good quality. For businesses, they are strategic tools that reduce costs by minimising waste and errors and increasing productivity. They help companies to access new markets, level the playing field for developing countries and facilitate free and fair global trade. According to <https://www.iso.org>, “ISO standards for agriculture cover all aspects of farming, from irrigation and global positioning systems (GPS) to agricultural machinery, animal welfare and sustainable farm management. They help to promote effective farming methods while ensuring that everything in the supply chain—from farm to fork—meets adequate levels of safety and quality. By setting internationally agreed solutions to global challenges, ISO standards for agriculture also foster the sustainability and sound environmental management that contribute to a better future.”

W3C

The World Wide Web Consortium (W3C, <https://www.w3.org/>) is an international community where member organisations, a full-time staff, and the public work together to develop Web standards. The W3C mission is to lead the World Wide Web to its full potential by developing protocols and guidelines that ensure the long term growth of the Web. According to W3C, the initial mission of the Agriculture Community Group (<https://www.w3.org/community/agri/>) is to gather and categorise existing user scenarios, which use Web APIs and services, in the agriculture industry from around the world, and to serve as a portal which helps both web developers and agricultural stakeholders create smarter devices, Web applications & services, and to provide bird’s eye view map of this domain which enables W3C and other SDOs to find overlaps and gaps of user scenarios and the Open Web Platform.

OASIS

OASIS (Organization for the Advancement of Structured Information Standards, <https://www.oasis-open.org>) is a not-for-profit consortium that drives the development, convergence and adoption of open standards for the global information society. OASIS promotes industry consensus and produces worldwide standards for security, Cloud computing, SOA, Web services, the Smart Grid, electronic publishing, emergency management, and other areas. OASIS open standards offer the potential to lower costs, stimulate innovation, grow global markets, and protect the right of free choice of technology.

OGC

The Open Geospatial Consortium (OGC, <https://www.ogc.org>) is an international consortium of more than 500 businesses, government agencies, research organizations, and universities driven to make geospatial (location) information and services FAIR—Findable, Accessible, Interoperable, and Reusable. OGC’s member-driven consensus process creates royalty free, publicly available geospatial standards. Existing at the cutting edge, OGC actively analyzes and anticipates emerging tech trends, and runs an agile, collaborative Research and Development (R&D) lab that builds and tests innovative prototype solutions to members’ use cases. OGC members together form a global forum of experts and communities that use location to connect

people with technology and improve decision-making at all levels. OGC is committed to creating a sustainable future for us, our children, and future generations.

The Agriculture DWG will concern itself with technology and technology policy issues, focusing on geodata information and technology interests as related to agriculture as well as the means by which those issues can be appropriately factored into the OGC standards development process. The **mission** of the Agriculture Working Group is to identify geospatial interoperability issues and challenges within the agriculture domain, then examine ways in which those challenges can be met through application of existing OGC standards, or through development of new geospatial interoperability standards under the auspices of OGC. The **role** of the Agriculture Working Group is to serve as a forum within OGC for agricultural geo-informatics; to present, refine and focus interoperability-related agricultural issues to the Technical Committee; and to serve where appropriate as a liaison to other industry, government, independent, research, and standards organizations active within the agricultural domain.

IEEE

IEEE, <https://www.ieee.org/>, is the world's largest professional association dedicated to advancing technological innovation and excellence for the benefit of humanity. IEEE and its members inspire a global community through IEEE's highly cited publications, conferences, technology standards, and professional and educational activities. IEEE, pronounced "Eye-triple-E," stands for the Institute of Electrical and Electronics Engineers. The association is chartered under this name and it is the full legal name.

VDMA—ISOBUS

ISOBUS (<https://www.isobus.net/isobus/>) was managed by the ISOBUS group in VDMA. The VDMA (Verband Deutscher Maschinen und Anlagenbau—German Engineering Federation) is a network of around 3,000 engineering industry companies in Europe and 400 industry experts. The ISOBUS standard specifies a serial data network for control and communications on forestry or agricultural tractors. It consists of several parts: General standard for mobile data communication, Physical layer, Data link layer, Network layer, Network management, Virtual terminal, Implement messages applications layer, Power train messages, Tractor ECU, Task controller and management information system data interchange, Mobile data element dictionary, Diagnostic, File Server. The work for further parts is ongoing. It is currently ISO standard ISO 11783.

agroXML

agroXML (<https://195.37.233.20/about/>) is a markup language for agricultural issues providing elements and XML data types for representing data on work processes on the farm including accompanying operating supplies like fertilizers, pesticides, crops and the like. It is defined using W3C's XML Schema. agroRDF is an accompanying semantic model that is at the moment still under heavy development. It is built using the Resource Description Framework (RDF).

While there are other standards covering certain areas of agriculture like e.g., the ISOBUS data dictionary for data exchange between tractor and implement or ISOagriNet for communication between livestock farming equipment, the purposes of agroXML and agroRDF are:

- exchange between on-farm systems and external stakeholders
- high level documentation of farming processes
- data integration between different agricultural production branches
- semantic integration between different standards and vocabularies
- a means for standardized provision of data on operating supplies

INSPIRE

In Europe a major recent development has been the entering in force of the INSPIRE Directive in May 2007, establishing an infrastructure for spatial information in Europe to support Community environmental policies, and policies or activities which may have an impact on the environment. INSPIRE is based on the infrastructures for spatial information established and operated by the all Member States of the European Union. The Directive addresses 34 spatial data themes needed for environmental applications, with key components specified through technical implementing rules. This makes INSPIRE a unique example of a legislative “regional” approach. For more details, see <https://inspire.ec.europa.eu/about-inspire/563>.

2.2.1 The Role of Location in Bioeconomy

Few activities are more tied to location, geography, and the geospatial landscape than farming. The farm business, farm supply chain, and public agricultural policies are increasingly tied as well to quantitative data about crops, soils, water, weather, markets, energy, and biotechnology. These activities involve sensing, analyzing, and communicating larger and larger scale geospatial data streams. How does farming become more, not less, sustainable as a business and as a necessity for life in the face of climate change, growing populations, scarcity of water and energy. Matching precision agricultural machinery with precision agricultural knowledge and promoting crop resiliency at large and small scales are increasing global challenges. As food markets grow to a global scale, worldwide sharing of information about food traceability and provenance, as well as agricultural production, is becoming a necessity. The situation is not much different from fishery or forestry. Both are geospatial disciplines to a good extent and require integration of location data.

2.2.2 *The Role of Semantics in Bioeconomy*

“Semantic Interoperability is usually defined as the ability of services and systems to exchange data in a meaningful/useful way.” In practice, achieving semantic interoperability is a hard task, in part because the description of data (their meanings, methodologies of creation, relations with other data etc.) is difficult to separate from the contexts in which the data are produced. This problem is evident even when trying to use or compare data sets about seemingly unambiguous observations, such as the height of a given crop (depending on how height was measured, at which growth phase, under what cultural conditions, etc.). Another difficulty with achieving semantic interoperability is the lack of the appropriate set of tools and methodologies that allow people to produce and reuse semantically-rich data, while staying within the paradigm of open, distributed and linked data.

The use and reuse of accurate semantics for the description of data, datasets and services, and to provide interoperable content (e.g., column headings, and data values) should be supported as community resources at an infrastructural level. Such an infrastructure should enable data producers to find, access and reuse the appropriate semantic resources for their data, and produce new ones when no reusable resource is available.

2.3 Architecture Building Blocks for Cloud Based Services

To fully understand the architecture outlined below, this chapter introduces high level concepts for future data exploitation platforms and corresponding applications markets first. There is a growing number of easily accessible Big Data repositories hosted on cloud infrastructures. Most commonly known are probably earth observation satellite data repositories, with petabyte-sized data volumes, that are accessible to the public. These repositories currently transform from pure data access platforms towards platforms that offer additional sets of cloud-based products/services such as compute, storage, or analytic services. Experiences have shown that the combination of data and corresponding services is a key enabler for efficient Big Data processing. When the transport of large amounts of data is not feasible or cost-efficient anymore, processes (or applications) need to be deployed and executed as closely as possible to the actual data. These processes can either be pre-deployed, or deployed ad-hoc at runtime in the form of containers that can be loaded and executed safely. Key is to develop standards that allow packing any type of application or multi-application-based workflow into a container that can be dynamically deployed on any type of cloud environment. Consumers can discover these containers, provide the necessary parameterization and execute them online even easier than on their local machines, because no software installation, data download, or complex configuration is necessary.

Figure 2.1 illustrates the main elements of such an architecture. Data providers on

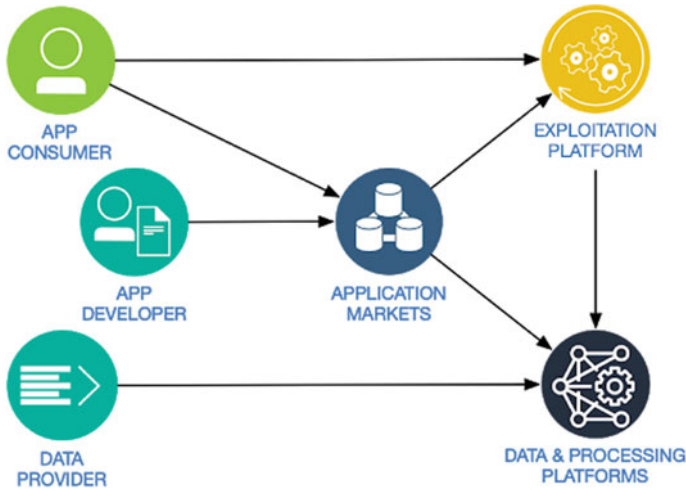


Fig. 2.1 High level architecture (Source [1])

the lower left make their data available at publicly accessible Data and Processing Platforms in the cloud. Ideally, these platforms provide access to larger sets of raw data and data products from multiple data providers. Application consumers (upper left), i.e. customers with specific needs that can be served by processing the data, identify the appropriate application(s) that produces the required results by processing (Big) data. The applications are produced by application developers and offered on application markets that work pretty similar to smart phone markets, with the difference that applications are deployed on demand on cloud platforms rather than downloaded and installed on smartphones. Exploitation platforms support the application consumers with single sign on, facilitate application chaining even across multiple Data and Processing Platforms, and ensure the most seamless user experience possible.

2.4 Principles of an Earth Observation Cloud Architecture for Bioeconomy

“*Earth Observation Cloud Architecture*” standardization efforts are underway that fulfill the aforementioned requirements to establish marketplaces for domain-specific and cross-domain Big data processing in the cloud. The architecture supports the “application to the data” paradigm for Big data that is stored and distributed on independent Data and Processing Platforms. The basic idea is that each platform provides a standardized interface that allows the deployment and parameterized execution of applications that are packaged as software containers. A logically second type of



Fig. 2.2 Earth observation cloud architecture platforms (*Source* [1])

platform is called Exploitation Platform and allows chaining containers/applications into workflows with full support for quoting and billing.

Exploitation and Data & Processing platforms are built using a number of components to provide all required functionality. As illustrated in Fig. 2.2, any number of these platforms can co-exist. Both types of platform can be implemented within a single cloud environment. Given that they all support the same interface standards, applications can be deployed and chained into complex workflows as necessary.

Standards define key components, interaction patterns, and communication messages that allow the ad-hoc deployment and execution of arbitrary applications close to the physical storage location of data. The application developer can be fully independent of the data provider or data host. The applications become part of an application market similar to what is currently common practice for mobile phone applications. The major difference is that applications are not downloaded to cell phones, but deployed and executed on cloud platforms. This is fully transparent to the user, who selects and pays an application and only needs to wait for the results to appear.

The above-mentioned standardization efforts are mainly driven by the Open Geospatial Consortium (OGC). These standards are made through a consensus process and are freely available for anyone to use to improve sharing of the world's

geospatial data. OGC standards are used in a wide variety of domains including Environment, Defense, Health, Agriculture, Meteorology, Sustainable Development and many more. OGC members come from government, commercial organizations, NGOs, academic and research organizations.

The OGC has worked for the last three years on a set of standards and software design principles that allow a vendor and platform neutral secure Big Data processing architecture. Supported by the space agencies ESA and NASA, the European Commission through H2020 co-funded projects (DataBio being one of them), and Natural Resources Canada, OGC has developed a software architecture that decouples the data and cloud operators from Earth Observation data application developers and end-consumers and provides all the essential elements for standards-based Big Data processing across domains and disciplines.

The Earth Observation Cloud Architecture defines a set of interface specifications and data models working on top of the HTTP layer. The architecture allows application developers and consumers to interact with Web services that abstract from the underlying complexity of data handling, scheduling, resource allocation, or infrastructure management.

2.4.1 Paradigm Shift: From SOA to Web API

Standards are the key pillar of any exchange or processing of information on the World Wide Web. Offering geospatial data and processing on the Web is often referred to as Spatial Data Infrastructure (SDI). These SDIs have been built following the Service Oriented Architecture (SOA) software paradigm. Nowadays, the focus is shifting towards Web Application Programming Interfaces (Web APIs). The differences for the end users are almost negligible, as client applications handle all protocol specific interactions. To the end user, the client may look the same, even though the underlying technology has changed.

At the moment, both approaches work next to each other to acknowledge the large number of existing operational SOA-based services. However, in the long run, Web APIs offer significant benefits, which is also reflected in OGCs Open API development activities. The architecture described in the following two sections, defines two ‘logical’ types of platforms. Both can be implemented using SOA-style Web services or Web-API-style REST interfaces. To the end user, it is most likely irrelevant.

2.4.2 Data and Processing Platform

The Data and Processing Platform illustrated in Fig. 2.3 has six major components:

In addition to the actual data repository, the platform offers the application deployment and execution API. The API allows deployment, discovery, and execution of



Fig. 2.3 Data and processing platform (Source [1])

applications or to perform quoting requests. All applications are packaged as Docker containers to allow easy and secure deployment and execution within foreign environments (though alternative solutions based on other container technology are currently explored). The Docker daemon provides a Docker environment to instantiate and run Docker containers. The Billing and Quoting component allows obtaining quotes and final bills. This is important because the price of an application run is not necessarily easily calculated. Some applications feature a simple price model that only depends on parameters such as area of interest or time period. Other applications, or even more complex entire workflows with many applications, may require heuristics to calculate the full price of execution. The workflow runner can start the Docker container applications. It manages dynamic data loading and result persistency in a volatile container environment. Identity and Access Management provide user management functionalities.

2.4.3 Exploitation Platform

The Exploitation Platform is responsible for registration and management of applications and the deployment and execution of applications on Data and Processing Platforms. It further supports workflow creation based on registered applications, and aggregates quoting and billing elements that are part of these workflows. Ideally, the Exploitation Platform selects the best suited Data and Processing Platform based on consumer's needs. As illustrated in Fig. 2.4, the Exploitation Platform itself consists of seven major components.

The *Execution Management Service API* provides a Web interface to application developers to register their applications and to build workflows from registered applications. The *application registry* implementation (i.e. application catalog) allows managing registered applications (with create, read, update, and delete options), whereas the optional *workflow builder* supports the application developer to build workflows from registered applications. The *workflow runner* executes workflows and handles the necessary data transfers from one application to the other.

The *Application Deployment and Execution Client* interacts with the data and processing environments that expose the corresponding Application Deployment and Execution Service API. The *Billing & Quoting* component aggregates



Fig. 2.4 Exploitation platform (Source [1])

billing and quoting elements from the data and processing environments that are part of a workflow. *Identity and Access Management* provides user management functionalities.

2.5 Standards for an Earth Observation Cloud Architecture

The architecture described above builds primarily on three key elements: The Application Deployment and Execution Service (ADES), the Execution Management Service (EMS), and the Application Package (AP). The specifications for all three have been initially developed in OGC Innovation Program initiatives and are handed over gradually after maturation to the OGC Standards Program for further consideration. Applications are shared as Docker containers. All application details required to deploy and run an application are provided as part of the metadata container called Application Package. The following diagram illustrates the high-level view on the two separated loops application development (left) and application consumption (right) (Fig. 2.5).

The left loop shows the application developer, who puts the application into a container and provides all necessary information in the Application Package. The application will be made available at the cloud platform using the Application Deployment and Execution Service (ADES). Using the Execution Management Service (EMS), application developers can chain existing applications into processing chains. The right loop shows the application consumer, who uses the EMS to request an application to be deployed and executed. Results are made available through additional standards-based service interfaces such as OGC API-Features, -Maps, -Coverages, or web service such as Web Map Service, Web Feature Service,

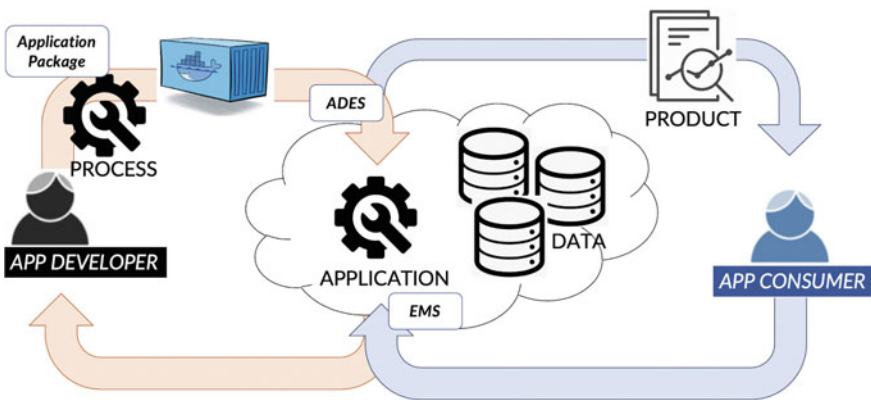


Fig. 2.5 Architecture elements in context

or Web Coverage Service. Alternatively, results can be provided at direct download links.

2.5.1 Applications and Application Packages

Any application can be executed as a Docker Container in a Docker environment that needs to be provided by the platform. The application developer needs to build the container with all libraries and other resources required to execute the application. This includes all data that will not be provided in the form of runtime parameters or be dynamic mounted from the platform’s Big data repository. The Docker container image itself can be built from a Docker Build Context stored in a repository following the standard manual or Dockerfile-based scripting processes. To allow standards-based application deployment and execution, the application should be wrapped with a start-up script.

As described in Ref. [2], the Application Package (AP) serves as the application metadata container that describes all essential elements of an application, such as its functionality, required processing data, auxiliary data, input runtime parameters, or result types and formats. It stores a reference to the actual container that is hosted on a Docker hub independently of the Application Package. The Application Package describes the input/output data and defines mount points to allow the execution environment to serve data to an application that is actually executed in a secure memory space; and to allow for persistent storage of results before a container is terminated (Fig. 2.6).

The OGC has defined the OGC Web Services Context Document (OWS Context Document) as a container for metadata for service instances [3]. The context document allows to exchange any type of metadata for geospatial services and data offerings. Thus, the context document is perfectly qualified to serve as a basis for the

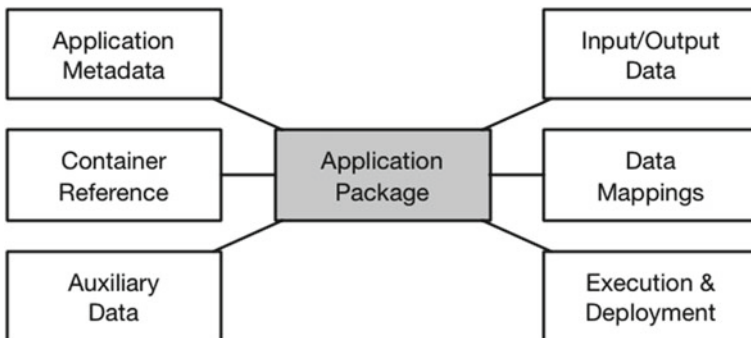


Fig. 2.6 Application package elements

Application Package. It can be used to define all application specific details required to deploy and execute an application in remote cloud environments.

2.5.2 Application Deployment and Execution Service (ADES)

Once application consumers request the execution of an app, the Exploitation forwards the execution request to the processing clouds and makes final results available at standardized interfaces again, e.g. at Web Feature Service (WFS) or Web Coverage Service (WCS) instances. In the case of workflows that execute a number of applications sequentially, the Exploitation realizes the transport of data from one process to the other. Upon completion, the application consumer is provided a data access service endpoint to retrieve the final results. All communication is established in a web-friendly way implementing the emerging next generation of OGC services known as WPS, WFS, and WCS 3.0.

2.5.3 Execution Management Service (EMS)

The execution platform, which offers EMS functionality to application developers and consumers, acts itself as a client to the Application Deployment and Execution Services (ADES) offered by the data storing cloud platforms. The cloud platforms support the ad-hoc deployment and execution of Docker images that are pulled from the Docker hubs using the references made available in the deployment request.

2.5.4 AP, ADES, and EMS Interaction

As illustrated in Fig. 2.7, the Execution Management Service (EMS) represents the front-end to both application developers and consumers. It makes available an OGC Web Processing Service interface that implements the new resource-oriented paradigm, i.e. provides a Web API. The API supports the registration of new applications. The applications themselves are made available by reference in the form of containerized Docker images that are uploaded to Docker Hubs. These hubs may be operated centrally by Docker itself, by the cloud providers, or as private instances that only serve a very limited set of applications.

The EMS represents a workflow environment that allows application developers to re-use existing applications and orchestrate them into sequential work-flows that can be made available as new applications again. This process is transparent to the application consumer.

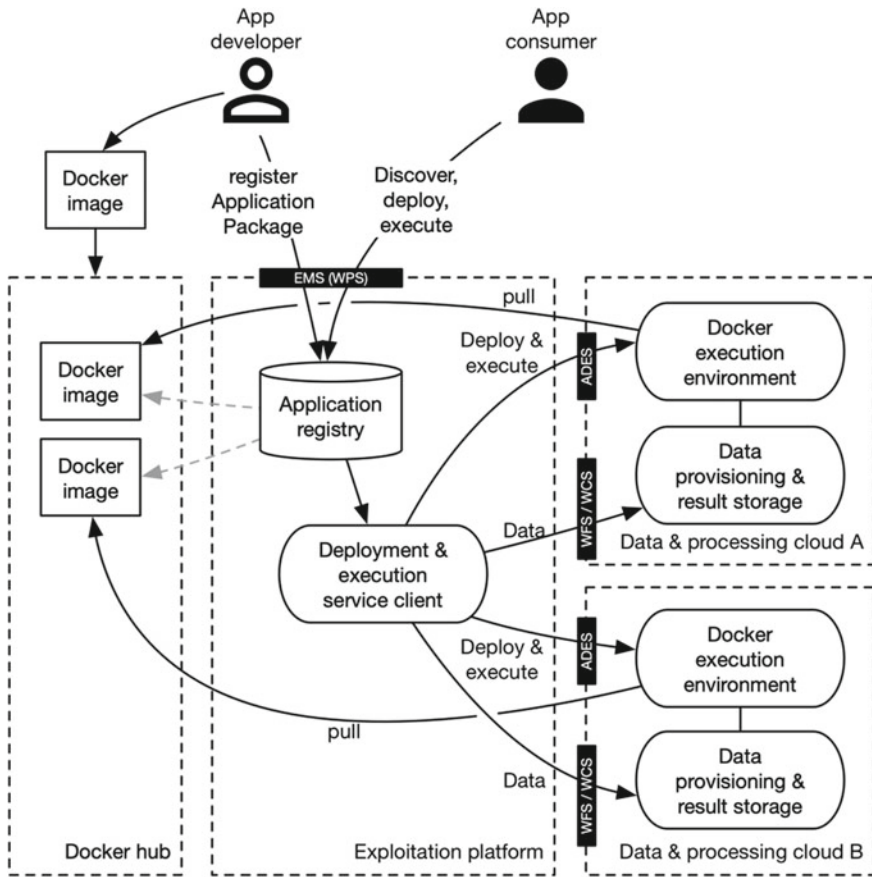


Fig. 2.7 Detailed software architecture (Source [4])

2.6 Standards for Billing and Quoting

Currently, lots of Big data and in particular satellite image processing still happens to a large extent on the physical machine of the end-user. This approach allows the end-user to understand all processing costs upfront. The hardware is purchased, prices per data product are known in advance, and actual processing costs are defined by the user’s time required to supervise the process. The approach is even reflected in procurement rules and policies at most organizations that often require a number of quotes before an actual procurement is authorized.

The new approach outlined here requires a complete change of thinking. No hardware other than any machine with a browser (which could even be a cell phone) needs to be purchased. Satellite imagery is not purchased or downloaded anymore, but rented just for the time of processing using the architecture described above, and the final processing costs are set by the computational resource requirements of the

process. Thus, most of the cost factors are hidden from the end-user, who does not necessarily know if his/her request results in a single satellite image process that can run on a tiny virtual machine, or a massive amount of satellite images that are processed in parallel on a 100+ machines cluster. The currently ongoing efforts to store Earth Observation data in data cubes adds to the complexity to estimate the actual data consumption, because the old unit “satellite image” is blurred with data stored in multidimensional structures not made transparent to the user. Often, it is even difficult for the cloud operator to calculate exact costs prior to the completed execution of a process. This leads to the difficult situation for both cloud operators that have to calculate costs upfront, and end-users that do not want to be negatively surprised by the final invoice for their processing request.

The OGC has started the integration of quoting and billing services into the cloud processing architecture illustrated in Fig. 2.8. The goal is to complement service interfaces and defined resources with billing and quoting information. These allow a user to understand upfront what costs may occur for a given service call, and they allow execution platforms to identify the most cost-effective cloud platform for any given application execution request.

Quoting and Billing information has been added to the Execution Management Service (EMS) and the Application Deployment and Execution Service (ADES). Both service types (or their corresponding APIs) allow posting quota requests against dedicated endpoints. A JSON-encoded response is returned with all quote related data. The sequence diagram in figure below illustrates the workflow.

A user sends an HTTP POST request to provide a quasi-execution request to the EMS/quotation endpoint. The EMS now uses the same mechanism to obtain

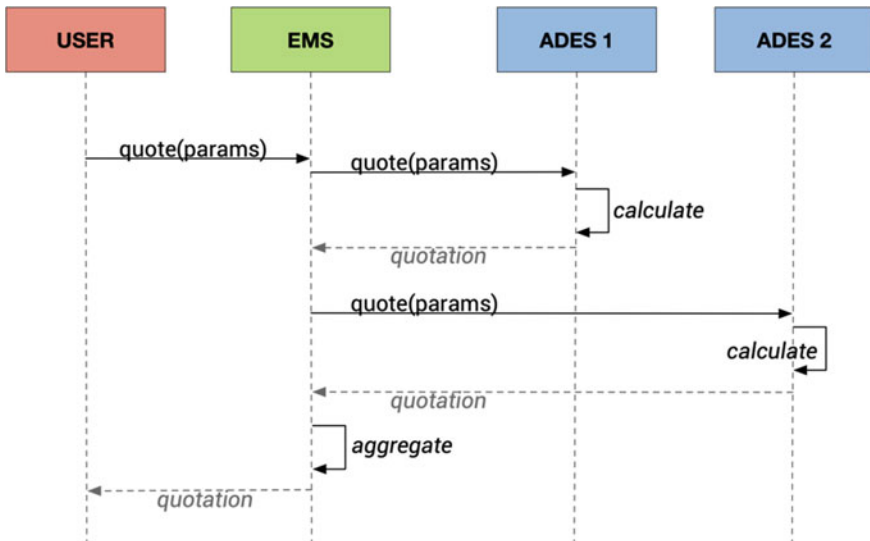


Fig. 2.8 Quoting process

quotes from all cloud platforms that offer deployment and execution for the requested application. In case of a single application that is deployed and executed on a single cloud only, the EMS uses the approach to identify the most cost-efficient platform. In case of a workflow that includes multiple applications being executed in sequence, the EMS aggregates involved cloud platforms to generate a quote for the full request. Identification of the most cost-efficient execution is not straightforward in this case, as cost efficiency can be considered a function of processing time and monetary costs involved. In all cases, a quote is returned to the user. The quote model is intentionally simple. In addition to some identification and description details, it only contains information about its creation and expiration date, currency and price-tag, and an optional processing time element. It further repeats all user-defined parameters for reference and optionally includes quotations for alternatives, e.g. at higher costs but reduced processing time or vice versa. These can for example include longer estimated processing times at reduced costs.

2.7 Standards for Security

Reliable communication within business environments requires some level of security. This includes all public interfaces as well as data being secured during transport. As shown in 4, the system uses identity providers to retrieve access tokens that can be used in all future communication between the application consumer, EMS, and ADES. The authentication loop is required to handle multiple protocols to support existing, e.g. eduGAIN, as well as emerging identity federations. Once an authentication token has been received, all future communication is handled over HTTPS and handles authorization based on the provided access token. Full details on the security solution are provided in OGC document OGC Testbed-14: Authorisation, Authentication, and Billing Engineering Report; OGC document OGC 18-057).

2.8 Standards for Discovery, Cataloging, and Metadata

DataBio's contribution to OGC standardization further includes metadata and service interfaces for service discovery. This includes Earth Observation (EO) products, services providing on-demand processing capabilities, and applications that are not deployed yet but waiting in an application store for their ad-hoc deployment and execution. The aforementioned OGC Innovation Program has developed an architecture that allows the containerization of any type of application. These applications can be deployed on demand and executed in cloud environments close to the physical location of the data.

From a catalog/discovery perspective, the following questions arise: How to discover EO applications? How to understand what data an application can be applied to? How to chain applications? How to combine applications with already deployed

services that provide data and data processing capabilities? The following paragraphs provide a short overview of standardization efforts currently underway.

Catalog Service Specification

The discovery solution proposed by OGC comprises building blocks through which applications and related services can be exposed through a Catalogue service. It consists of the following interfaces:

- Service Interface: providing the call interface through which a catalogue client or another application can discover applications and services through faceted search and textual search, and then retrieve application/service metadata providing more detail.
- Service Management Interface: providing the call interface through which a catalog client or any other application can create, update and delete information about applications/services.

Each of the above interfaces is discussed in full detail in the [OGC Testbed-15: Catalogue and Discovery Engineering Report \[5\]](#). This discussion includes the metadata model that provides the data structure through which the application and/or service is described and presented as a resource in the catalog.

The current standardization work builds on a series of existing standards as illustrated below (Figs. 2.9, 2.10 and 2.11).

These standards provide robust models and encodings for EO products and collections.

Now extended by OpenSearch specifications as illustrated below.

And integrated into a set of specifications as shown in figure below.

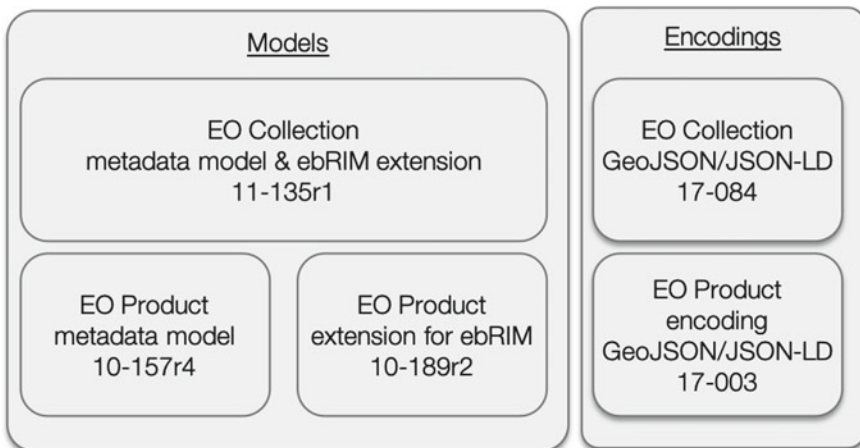


Fig. 2.9 Existing OGC Standards supporting discovery for EO data

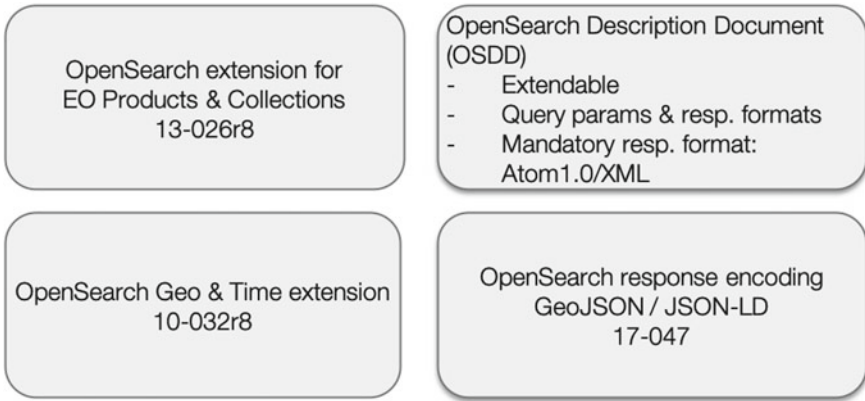


Fig. 2.10 OpenSearch extensions for existing OGC Standards

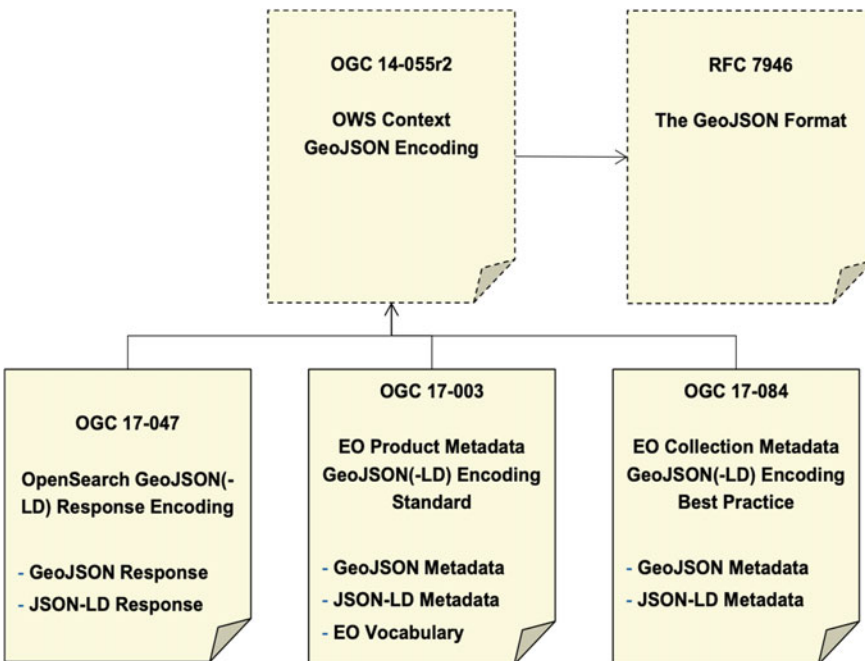


Fig. 2.11 Overview of OGC Standards for standards-based application discovery (Source [5])

2.9 Summary

This chapter provided an overview of currently ongoing standardization efforts executed by the Open Geospatial Consortium with support by DataBio to define

an application-to-the-data environment for Big geospatial data. All work till date has been documented in OGC Engineering Reports. As a more detailed discussion would go far beyond this book chapter, the interested reader is referred to the following documents:

- [OGC Testbed-15: Catalogue and Discovery Engineering Report \[5\]](#)
- [OGC Testbed-14: Application Package Engineering Report \[6\]](#)
- [OGC Testbed-14: ADES & EMS Results and Best Practices Engineering Report \[7\]](#)
- [OGC Testbed-14: Authorisation, Authentication, & Billing Engineering Report \[8\]](#)
- [OGC Earth Observation Exploitation Platform Hackathon 2018 Engineering Report \[9\]](#)
- [OGC Testbed-13: EP Application Package Engineering Report \[10\]](#)
- [OGC Testbed-13: Application Deployment and Execution Service Engineering Report \[11\]](#)
- [OGC Testbed-13: Cloud Engineering Report \[12\]](#)

Acknowledgements The authors would like to thank everybody who contributed to the work presented in this chapter. Particularly, we would like to thank the European Space Agency, ESA, and Natural Resources Canada for additional support in the development of the architecture documented herein.

References

1. Open Geospatial Consortium. (OGC). *OGC earth observation applications pilot: call for participation (CFP)*. Online publication: https://portal.ogc.org/files/?artifact_id=90733. Accessed January 5, 2020.
2. Simonis, I. Standardized Big Data processing in hybrid clouds. In *Proceedings of the 4th International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM 2018)* (pp. 205–210).
3. Open Geospatial Consortium. (OGC). OGC OWS context conceptual model. OGC Implementation Standard. OGC Document 12-080r2.
4. Simonis, I. Quoting and billing: commercialization of Big Data analytics. In *Proceedings of the 2019 conference on Big Data from Space (BiDS'19)* (pp. 53–56).
5. Open Geospatial Consortium (OGC). *OGC Testbed-15: Catalogue and discovery engineering report, OGC Document 19-020r1*. Online publication. Accessed January 5, 2020.
6. Open Geospatial Consortium (OGC). *OGC testbed-14: Application package engineering report, OGC Document 18-049r1*. Online publication: <https://docs.opengeospatial.org/per/18-049r1.html>. Accessed January 5, 2020
7. Open Geospatial Consortium (OGC). *OGC testbed-14: ADES & EMS results and best practices engineering report, OGC Document 18-050r1*. Online publication: <https://docs.opengeospatial.org/per/18-050r1.html>. Accessed January 5, 2020.
8. Open Geospatial Consortium (OGC). *OGC earth observation exploitation platform hackathon 2018 engineering report, OGC Document 18-057*. Online publication: <https://docs.opengeospatial.org/per/18-057.html>. Accessed January 5, 2020

9. Open Geospatial Consortium (OGC). *OGC earth observation exploitation platform Hackathon 2018 engineering report, OGC Document 18-046*. Online publication: <https://docs.opengeospatial.org/per/18-046.html>. Accessed January 5, 2020
10. Open Geospatial Consortium (OGC). *OGC testbed-13: EP Application package engineering report, OGC Document 17-023*. Online publication: <https://docs.opengeospatial.org/per/17-023.html>. Accessed January 5, 2020
11. Open Geospatial Consortium (OGC). *OGC testbed-13: Application deployment and execution service engineering report, OGC Document 17-024*. Online publication: <https://docs.opengeospatial.org/per/17-024.html>. Accessed January 5, 2020
12. Open Geospatial Consortium (OGC). *OGC Testbed-13: Cloud engineering report, OGC Document 17-035*. Online publication: <https://docs.opengeospatial.org/per/17-035.html>. Accessed January 5, 2020

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

