



HAL
open science

Integration in and Between Earth Observation Research Centers for Achieving Sustainable Development

Todor Branzov, Krassimira Ivanova, Velichka Milousheva

► **To cite this version:**

Todor Branzov, Krassimira Ivanova, Velichka Milousheva. Integration in and Between Earth Observation Research Centers for Achieving Sustainable Development. 5th International Conference on Information Technology in Disaster Risk Reduction (ITDRR), Dec 2020, Sofia, Bulgaria. pp.206-215, 10.1007/978-3-030-81469-4_17. hal-03761635

HAL Id: hal-03761635

<https://inria.hal.science/hal-03761635>

Submitted on 26 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Integration in and Between Earth Observation Research Centers for Achieving Sustainable Development

Todor Branzov, Krassimira Ivanova and Velichka Milousheva

Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences,
Sofia, Bulgaria

todor.branzov@gmail.com, kivanova@math.bas.bg, vmil@math.bas.bg

Abstract. The digital transformation of society in recent years has led to a qualitatively new leap in the development of natural sciences. Currently, these new technologies allow the collection and processing of a huge amount of information and increase the growth of scientific production, parallel to fusing the boundaries between scientific disciplines. Hence, the created Earth observation centers are a natural prerequisite for facilitating these processes. This article presents our studies on the information system architecture of some of the leading Earth observation research centers, with a focus on capabilities for integration with other scientific infrastructures. We review an original solution based on hybrid service-microservice architecture being implemented in the National Geoinformation Center of Bulgaria, investigating some possible advantages it may provide for integration with other Earth observation centers.

Keywords: Earth Observation Centers, Information Systems Architecture, Microservices, Service Oriented Architecture, IT Service Management.

1 Introduction

The process of Earth observation (EO) gathers information about the physical, chemical, and biological systems of the planet using both remote-sensing technologies and in-situ surveying techniques. It covers all phases of data collection, data analysis, and data presentation aimed to support Earth ecosystem monitoring and management.

Currently, we live in the era of data-driven development. The EO data have already proven their value across many sectors of society, providing vital material for applying decision-making strategies in a broad range of sectors, which can be divided into three main groups – biospheric, economic and societal. Thus, numerous initiatives are currently working toward improving the accessibility, quality, applicability, and reliability of data as an important part of this process.

In 2015 the United Nations General Assembly adopted “The 2030 Agenda” as an urgent call for action by all countries in a global partnership for achieving “better and more sustainable future for all”. It is based on 17 global life-changing goals, called Sustainable Development Goals (SDGs) [1]. The integration of geospatial information, Earth observations, and other sources of Big Data, combined with modern descriptive

and predictive techniques and decision-making strategies are becoming the fundamental solution for the performance of many of the SDGs:

- directly (such as SDG 13: Climate Action);
- or indirectly (such as SDG 2: Zero Hunger).

Another point of view, for successful realization of this integration the Earth observation infrastructures are built in accordance with SDG 17 (Partnerships).

Figure 1 shows the impact of EO systems on the SDGs grouping to three main facets – Biosphere, Economy, and Society. EOs provide accurate and reliable information which is directly necessary for monitoring and management of Biosphere sectors, but also have an indirect influence on the economic sectors and many activities of the society.



Fig. 1. The impact of EO systems on the SDGs (in accordance with [2])

The benefits of EO as a valuable information source in support of many branches of science, industry, government, and society are already well understood. Because of this many national and international programs focus their efforts on establishing scientific infrastructures for integration of geospatial information and primary geoinformation data from different sources and providers in multidisciplinary, integrated data products

(IGP), using advantages of high-performance and grid computing, big data technologies and the techniques of the Artificial Intelligence.

The digital transformation of society in recent years has led to a qualitatively new leap in the development of natural sciences. Currently, these new technologies allow the collection and processing of a huge amount of information (its volume doubles every five to ten years) and increase the growth of scientific production, which complicates the exchange of scientific ideas. Hence, the created Earth observation centers are a natural prerequisite for facilitating these processes.

A natural specificity of modern sciences is their mathematization, which increases the degree of abstraction of scientific knowledge, expands the possibilities for concentrating scientific information, and greatly increases its capacity. Therefore, the digital mathematical models and data processing tools are an essential part of the service package of each EO center.

The growing complexity of the processes that fall within the scope of research requires an interdisciplinary approach, leading to the emergence of “borderline” and “joint” sciences. This process of integration, i.e. synthesis of sciences and scientific disciplines, combining their methods and blurring the boundary between them, gives the opportunity to build a holistic picture of the world, combining natural sciences, mathematical, philosophical, and general scientific knowledge.

Nowadays, the cooperation between applied mathematics, informatics and information technologies, and another science or scientific discipline gives rise to new disciplines in which obtaining the results from the application of models is impossible without the use of high-performance technology or distributed calculations. In our case, the role of the third part of this cooperation is not one, but a group of sciences and scientific disciplines covered under the common name “*geosciences*”, which makes the integration process even more difficult. Thus, this collaboration gave birth to “*Computational Earth Sciences*”, which deals with a predictive understanding of Earth systems through the development and application of models, quantitative methods, and computational tools that exploit high-performance computing resources.

At an even higher level of integration, to understand the impacts and relationships between humans as a society and the natural processes of the Earth's system, Guo et al. [3] propose the introduction of a new scientific discipline: “*Big Earth Data Science*”, which aims to use data from the observations of Earth and social observation and develop theories to understand the mechanisms of how such a social-physical system operates and evolves.

The paper presents our studies on the information system (IS) architecture of some of the leading EO research centers, with a focus on capabilities for integration with other scientific infrastructures. Having in mind all of the above, we review an original solution based on hybrid service-microservice architecture being implemented in the National Geoinformation Center of Bulgaria, investigating some possible advantages it may provide.

2 The EO Ecosystem

According to [4] “*research infrastructure* means facilities, resources, and related services that are used by the scientific community to conduct top-level research in their respective fields and covers major scientific equipment or sets of instruments; knowledge-based resources such as collections, archives or structures for scientific information; enabling ICT-based infrastructures such as Grid, computing, software and communication, or any other entity of a unique nature essential to achieve excellence in research”.

So, one of the key elements in the research infrastructure are knowledge-based resources. The specificity of Earth Observation data is that it comes from a variety of:

- branches – climate, waters, air, soils, geological phenomena, and so on;
- sources – satellite, airborne, and in-situ measurements, crowdsourcing activities records, as well as historical data, stored in earlier databases;
- providers – international agencies, specialized national institutions, scientific organizations, business organizations, NGOs, citizens.

Depending on the degree of complexity of the processing to which the data is subjected, according to [5] the data can be categorized in the following levels (so-called EPOS data levels categorization):

- Level 0: raw data, or basic data (example: seismograms, accelerograms, time series);
- Level 1: data products coming from nearly automated procedures (earthquake locations, magnitudes, focal mechanism, shakemaps);
- Level 2: data products resulting from scientists' investigations (crustal models, strain maps, earthquake source models, etc.);
- Level 3: integrated data products coming from complex analyses or community shared products (hazards maps, catalogue of active faults, etc.).

Except for Level 0, where only software agents for capturing information and storing in a predefined standard form is used, the output of the next levels is called Data Products – where some processing of raw data has to be done, and for which additional modelling services are needed.

These elements – data, data products, software, and services (DDSS) are the production that can be served to the users through the infrastructure of the EO center (Fig. 2).



Fig. 2. DDSS Flow from Data Providers to Users

By its nature, DDSS can be given from the providers (DDSS') or produced by the center (DDSS''). We can say that attracting more providers allows the growing of the EO center horizontally – more kinds of data, from more destinations, etc. On the other hand, often geoscientists are generating products through their research activities, using the resources of the center or in combination with third party resources, and these data products are entered as new resources in the EO center. This pseudo-cycling nature of the resources allows increasing the potential of the center vertically, because:

- for data products – often the new data product is on the next level of complexity in respect to the input data that are used as components;
- for software and services – the situation is similar: more sophisticated modelling tools are built on the base of already existing components.

3 Integration IN and BETWEEN the EO Centers

The analysis of existing geoinformation centers (another term for Earth Observation Centers), made by us in [6] showed a tendency of establishing national centers aimed to gather the information from various sources in the country, along with international agreements for enriching these source data with global observations. On the upper level of the process of integration of the research information and knowledge is the establishing the global aggregation centers, such as GEOSS (<https://www.geoportal.org/>) and EPOS (<https://www.epos-ip.org/>), which form sets of coordinated, independent Earth observation, information, and processing systems for strengthening the monitoring of the state of the Earth, where the national and regional centers come as data providers. Especially in EPOS, the data providers are only national and trans-national research infrastructures thus ensuring access to quality-checked data and products, while in GEOSS the providers are more than 150 institutions – international, regional, and national, and from the private sector.

According to the specificity of organizational structure (see [6]) a variety of approaches are used for the achievement of information integration and interoperability. We may distinguish three approaches:

- In the case when a center is governed by a state organization and its operations are closely connected with everyday life (such as prognosis, warnings, etc.), while the research activities play a secondary role as support of the main activities, the centers are built using monolithic architecture. The typical example is GEARS (Ground Enterprise Architecture Services) [7], implementation of which transforms the NESDIS Ground Enterprise from a set of stand-alone capabilities to an Integrated Ground Enterprise based on common ground services [7]. Another example is ETRIS DZZ – Russian Unified Geographically Distributed Information System for Earth Remote Sensing [8] – the project, which is evaluated by Roscosmos as one of the largest in the field of using the results of space activities in Russia and in the world. ETRIS DZZ unite the entire ground infrastructure within the framework of a new hierarchy with unified technical standards, which ensures the management of the target use of

Russian Earth remote sensing spacecraft, the reception of information, its processing and transmission to consumers;

- When a national or trans-national center is built on the basis of an agreement of independent bodies (universities, research centers, business partners) usually some kinds of federated structures are established. The reason is that the federated organization is a hybrid that combines characteristics of centralized (e.g. centralized planning, standardization, etc.) and decentralized organizations (e.g. local leadership, competitive local objectives, etc.) [9] and specifics helps to unravel the situation in which the center has no ownership over data sources, however, it serves as a unifier. As it is mentioned in [10], in this case, the interoperability is achieved through the development of own service buses by each disciplinary community, which cover common/federal “domain” specifications;
- In the case when the disciplinary diversity and the amount of data sources are too many, the center is built as an independent entity. The constituent bodies are not necessarily DDSS providers or this is not their main role. In this case, the realization of cross-domain interoperability using federated architecture is not appropriate, so a middleware structure takes up the tasks for mediation, adaptation, distribution, semantic mapping, and quality checks. Such a “brokering” approach is implemented in GEOSS infrastructure [10].

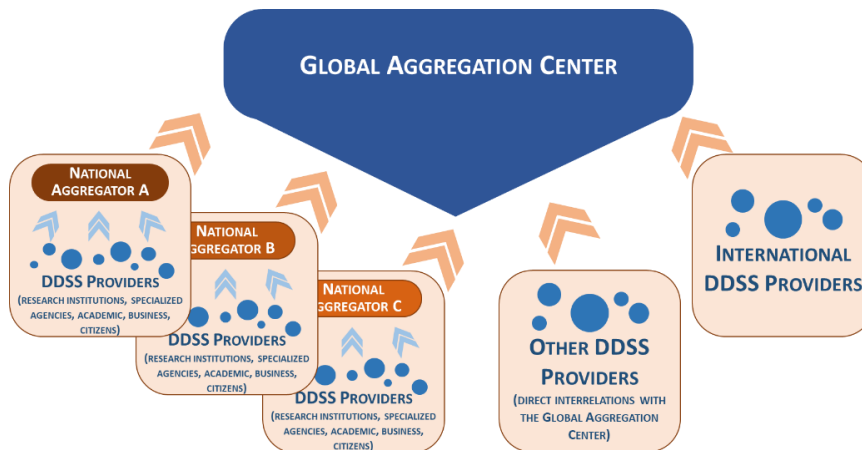


Fig. 3. Integration between centers

So, Figure 3 stresses this hierarchical process of integration. At the bottom are the primary sources - a variety of ICT and content providers. In terms of the complexity of given sources from the content providers, it varies from raw data (given directly from satellite, airborne, in-situ measurements) or pieces of data (for instance, from crowdsourcing channels that gather sensitive in-time event information) to complex data products (for instance, geomagnetic storm forecast for next 6 hours [11]).

In the middleware are the national and regional aggregators, who assimilate these data and data products from some of these primary sources and in turn add new data

products as a result of the work of the center. Such integration allows gathering data and models from different scientific disciplines, so, such results as integrated maps of geological, geophysical and seismological data for seismic hazard assessment [12] stand possible.

At the top level are the global aggregators, such as GEOSS and EPOS, which use the primary and middleware providers directly (like GEOSS) or through the national aggregators (like EPOS) as sources of DDSS and ensure integrated multi-disciplinary and trans-national science platforms.

4 Service-Microservice Hybrid Approach to Integration

The appearance of the concept of “microservice” has catalysed changes in enterprise system architecture and pushed for redesign of the software information systems of some of the leading EO centers. Notably, EPOS [13] included the microservices approach as one of the four principles of system development and declared that it will use web services “wherever possible”.

Recently, the software information system architecture of the Bulgarian National Geoinformation Center (NGIC) adopted both service and microservice concepts [6]. Since the NGIC aims to foster scientific cooperation between all major EO organizations in Bulgaria, one of the main goals of its information system is to provide the necessary mechanisms for the integration of DDSS that are managed or produced by the partners in NGIC and are needed in joint activities. The architecture employed three layers, with the middle acting as a smart interconnector between the sources of DDSS (bottom layer) and the production layer (top layer) where integrated data products (IDP) are being produced (Fig. 4). All of the “sources” are accessible through sets of microservices and all of the IDPs can be reached through services, which may also provide access to a set of microservices.

The service-microservice architecture used in NGIC provides several options for integration with partner EOC.

The most basic option for integration is accessing publicly available or restricted access value-added (VA) service. This option is by design provided to all users of NGIC. However, the architecture has capabilities for the accomplishment of some more sophisticated scenarios for the provision of outbound access to DDSS produced in NGIC or inbound flow of DDSS needed by NGIC.

The first scenario (1. on Fig 4.) is when the partner EOC has a blueprint for its “accessor” (we use the term as per GEOSS) - the protocol and all other means to receive DDSS from out of its boundaries. If the DDSS is an integrated data product (IDP), i.e. it is a product of some processing in NGIC, a service is built according to both partner’s accessor blueprint and internal policies and processes. The partner is provided with access to that service; in case such service exists, that may fulfill the requirements of the accessor, the service is provided, eventually with needed service level agreement alterations. An important subcase of this scenario is when the partner prefers to use microservices to access IDP - then the service provides the API specification to the

proper set of microservices; it may also contain some functionalities needed for the business goals of NGIC, e.g. usage accounting.

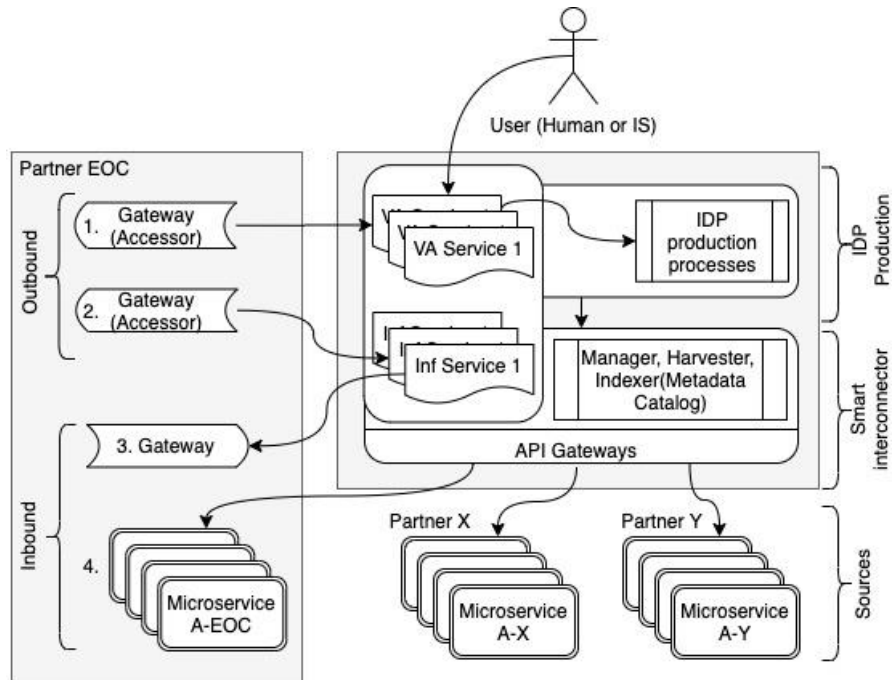


Fig. 4. Scenarios for integration using the service-microservice hybrid architecture

The second scenario (2. on Fig 4.) is when the partner EOC uses accessors and needs to receive DDSS that is produced by one of the partners in NGIC, i.e. there is no processing or adding value by NGIC. In that case, service is provided that uses the smart interconnector to locate, adapt and deliver to the partner EOC; the NGIC acts as a relay to the source (with little overhead, not related with the value, e.g. again usage accounting). Although from the point of the EOC partner this case is very similar to the previous, it is of importance for NGIC to differentiate, since much more significant resource usage is expected in the first scenario. A subcase of this scenario is when the partner EOC needs to access some module that is actually part of the smart interconnector. For example, one of the requirements for integration with EPOS is to maintain a CERIF metadata catalog (for abstraction we used “Indexer” to denote all of the technologies that have similar functionalities) that may be included in the hierarchy of theirs. That will be provided through a service in the setting of the current scenario.

The first two scenarios covered the outbound flow and the next two will cover the inbound flow. The third scenario (3. on Fig 4.) is when the partner EOC has a blueprint for the protocol and all other means to supply its products out of its boundaries. A service is built in NGIC according to both partner’s blueprint and internal policies and processes. That service is used by the smart interconnector to provide the production processes in NGIC with what is needed from the partner EOC. An option is to wrap

that service into a microservice, so to keep all of the source access methods consistent with the microservice paradigm. This scenario is the reverse of the first scenario.

The fourth scenario (4. on Fig 4.) is when the partner EOC provides access to its products through microservices. In that case, NGIC uses its API gateway (altered if needed) to access the microservices and to provide through the smart interconnector the production processes in NGIC with what is needed from the partner EOC.

5 Conclusion and Discussion

The reviewed three-layered service-microservice architecture allows clear separation of duties. All of the mechanisms needed to supply the NGIC with “raw materials” are within the smart interconnector, which employs technologies for data harvesting, adaptation, indexation, storage, and delivery. The production layer is responsible for the assimilation and processing of that materials to prepare value-added products, and to deliver them to the users. All of the outbound flows are managed as services. All of the inbound flows are managed either as microservices or services. All of the production is also managed as microservices or services.

The advantage of that approach is its agility - microservices are meant to be as simple and autonomous as possible and give the partners freedom to organize their production as they want, using the technologies and methods they want. It's designed to require the minimum effort for altering the partner side for coupling with the center. Another feature, which in our opinion is an advantage, is the broad options that may be used in the design of the smart interconnector layer – in the case of NGIC we plan to use a publisher-subscriber approach and various data buffering solutions, which would allow us to implement our own data models and optimize data management. The use of both monolith services and small autonomous microservices in production would allow choosing an optimal approach in building each particular IDP.

The main disadvantage of the approach is that most or even all of the tasks for verification and integrity of the connection and the flow are a matter of the center, which generates overhead. Although the partners may prepare some relevant mechanisms, they are not required to do that. The basic case is - we have something useful here, connect and use it at your own risk.

Since that architecture is in process of implementation in the information system of NGIC, we are still approbating and optimizing some of its components.

Acknowledgements

This work is supported by Contract DO1-161/28.08.2018 “NGIC – National Geoinformation Center for monitoring, assessment, and prediction of natural and anthropogenic risks and disasters” under the Program “National Roadmap for Research Infrastructure 2017-2023”, funded by the Bulgarian Ministry of Education and Science.

References

1. United Nations: Transforming our world: the 2030 Agenda for Sustainable Development (A/RES/70/1). Resolution adopted by the General Assembly on 25.09.2015, 40 p. (2015).
2. Paganini, M. et al.: Satellite Earth Observations in Support of the Sustainable Development Goals. European Space Agency, 108 p. (2018).
3. Guo, H., Nativi, S., Liang, D., Craglia, M., Wang, L., Schade, S., Corban, C., He, G., Pesarasi, M., Li, J., Shirazi, Z., Liu, J., Annoni, A.: Big Earth Data Science: an information framework for a sustainable planet. *International Journal of Digital Earth*, 13(7), 743-767 (2020).
4. European Commission: Legal framework for a European Research Infrastructure Consortium – ERIC Practical Guidelines, DOI: 10.2777/79873, 40 p. (2010)
5. Bailo, D., Jeffery, K.G.: EPOS: a novel use of CERIF for data intensive science. *Procedia Computer Science*, 33, 3-10 (2014).
6. Branzov, T., Ivanova, Kr., Georgiev, M.: Service-microservice basic system architecture model for geoinformation centers. *Proceedings of XIX International Multidisciplinary Scientific GeoConference SGEM 2019*, 19(2.1), 587-594 (2019).
7. NESDIS Ground Enterprise Architecture Services (GEARS): Concept of Operations, Feb 2015, 48 p. (2015).
8. Romashkin, V. V., Loshkarev, P. A., Fedotkin, D. I., Tokhiyan, O. O., Arefyeva, T. A., Musiyenko, V. A.: ETRIS DZZ — sovremennyye resheniya v razvitii otechestvennoy nazemnoy kosmicheskoy infrastruktury distantsionnogo zondirovaniya Zemli iz kosmosa. *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 16(3), 220-227 (2019).
9. Rychkova, I., Zdravkovic, J., Speckert, T.: Challenges of EA methodologies facing progressive decentralization in modern organizations. *CEUR Workshop Proceedings*, vol. 1023, pp. 18-28 (2013).
10. Nativi, S., Craglia, M., Pearlman, J.: Earth Science Infrastructures Interoperability: The Brokering Approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(3), 1118-1129 (2013).
11. NIGGG – Geomagnetic Storm Forecast.
http://www.geophys.bas.bg/kp_for/kp_mod_en.php, last accessed 01.10.2020
12. Trifonova, P., Metodiev, M., Stavrev, P., Simeonova, S., Solakov, D.: Integration of Geological, Geophysical and Seismological Data for Seismic Hazard Assessment Using Spatial Matching Index. *Journal of Geographic Information System*, 11, 185-195, doi: 10.4236/jgis.2019.112013 (2019).
13. EPOS-IP WP6 & WP7 teams: ICS-TCS Integration Guidelines – Handbook for TCS integration: Level-2, 18 p. (2015).