HAL
open science

LOOKING FOR REPRODUCIBILITY FOR EARTH OBSERVATION APPLICATIONS AT THE ABSTRACT LEVEL

Marie-Francoise Voidrot, Ingo Simonis, Raphaëlle Barbier, Pascal Le Masson, Nicolas Fichaux, Thierry Ranchin

To cite this version:

Marie-Francoise Voidrot, Ingo Simonis, Raphaëlle Barbier, Pascal Le Masson, Nicolas Fichaux, et al.. LOOKING FOR REPRODUCIBILITY FOR EARTH OBSERVATION APPLICATIONS AT THE ABSTRACT LEVEL. IEEE International Symposium on Geoscience and Remote Sensing (IGARSS), Jul 2021, Brussels, Belgium. hal-03335895

HAL Id: hal-03335895

https://minesparis-psl.hal.science/hal-03335895v1

Submitted on 6 Sep 2021

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.

LOOKING FOR REPRODUCIBILITY FOR EARTH OBSERVATION APPLICATIONS AT THE ABSTRACT LEVEL

Marie-Francoise Voidrot [(1)], Ingo Simonis [(1)], Raphaëlle Barbier [(2)], Pascal Le Masson [(2)], Nicolas Fichaux [(2)], Thierry Ranchin [(2)]

[(1)] OGCE, [(2)] Mines ParisTech, PSL University / ARMINES

ABSTRACT

The e-shape H2020 Project brings together decades of public investment in earth observation and in cloud capabilities into services to the citizens, the industry, the decision-makers and the researchers. e-shape promotes the development and uptake of a number of cloud-based pilot applications (27 at the start and up to 10 additional through an onboarding process), addressing the Sustainable Development Goals, The Paris Agreement and the Sendaï Framework. The pilots, which address 7 different thematic domains, build on GEOSS and on the Copernicus data pool and many computational infrastructures. The consortium gathers 55 members from 17 European countries, Ethiopia, Egypt and Israel. It is a European contribution to GEOSS [1].

A major deliverable at the end of the project in 2023, will be a Guide for European Earth Observation application developers, decision-makers, and experts delivering best practices to use Earth Observation resources based on the experience collected during the project. This guide will provide a unique source and guidelines to increase the usage and exploitation of Earth Observation in the thematic domains addressed by e-shape.

Index Terms— Reproducibility, Earth Observation, Workflow, Best Practices, EuroGEO, Horizon 2020, GEOSS, Copernicus, e-shape, NextGEOSS

1. THE CHALLENGE

e-shape will capture the requirements and lessons learned out of the implementation of big number of pilots over more than 40 platforms. It will capture all essential elements to develop successful Earth Observation application(s) that builds on top of the available European Earth Observation resources. e-shape will generate a large amount of complex information and a major challenge of the Guide will be to address the concerns in a progressive, logical and comprehensive way, making complex technical issues and challenges accessible to the greatest number. An abstract generic and reproducible workflow is therefore needed to structure this wealth of knowledge from the simplest to the most complex issues to support the uptake and increase the benefits for a broad community.

2. AN INITIAL ASSESSMENT AS A FOUNDATION

Based on an initial assessment of the e-shape pilots, the project has gathered an extensive understanding of the pilot's expectations and needs, on one hand, on the European Earth Observation resources they were using or planning to use and on another hand on the challenges they are planning to address. Considering the number of participants and their diversity in terms of nationality, background, gender and age, e-shape partners can be seen as a representative sample of the Earth Observation community. The expectations and issues identified during this initial assessment have been organized into a draft conceptual workflow presented in figure 1, which will be consolidated as needed during the project. The presentation will introduce the current state of this workflow based on the initial assessment, the lessons learned from the first implementation sprint and the questions collected from the Pilots. This approach is inspired by the proved and successful 5-step user experience defined by the NextGEOSS H2020 project:

3. GENERIC REPRODUCIBLE IMPLEMENTATION WORKFLOW

The conceptual workflow initialized by the first works is used to capture the lessons learned out of the pilots' implementations with a structured process. Each step is introduced below to illustrate the type of content which will be documented. This will be enriched and developed during the whole project.

3.1. Step 1: Co Design

The e-shape project is putting a strong emphasis on co-design to characterize different types of co design needs, and defining a specific taxonomy of co-design methods [2]. The co-design method is already being demonstrated over

the 27 initial e-shape pilots and soon to the 5 new ones. e-shape is already contributing to capacity development on this important topic, fertilizing towards other H2020 projects.

3.2. Step 2: Data discovery

Catalogues have improved a lot in the last years. But their discovery can remain a problem. Federation of catalogues, data hubs, data brokers may address this issue.

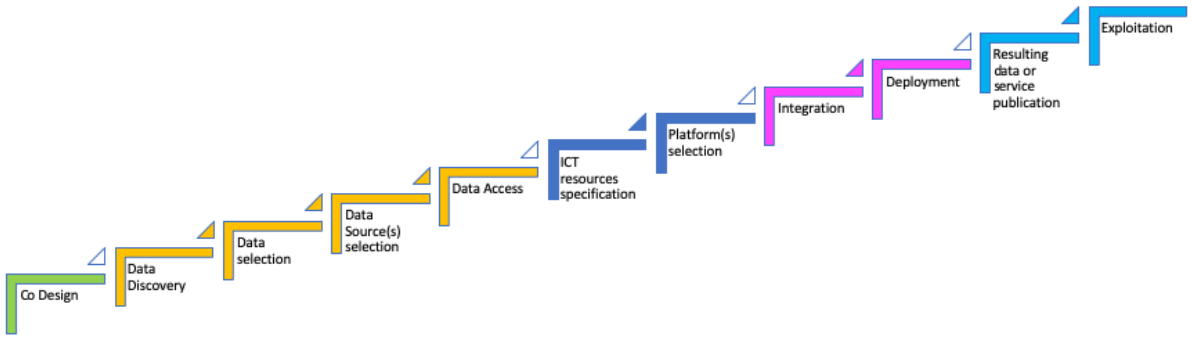

Figure 1: Generic reproducible Pilot development workflow

3.3. Step 3: Data selection

Several datasets produced with different sensors can be comparable or the same dataset can be accessible at different resolutions in space and time, with different update frequency. Very often these different resolutions are attached to different access rights: low resolution can be accessible as open data and high resolution as paying data.
Datasets can be post-processed to produce similar datasets with a better space or time resolution. This is the case for downscaling algorithms, or gap fill based on AI or Deep Learning for instance. This postprocessing can impact the data quality or data skills in case of numerical models. This impact should be assessed and documented.
The fitness for use of similar datasets can be characterized.

3.4. Step 4: Data source selection

The same data set with same or different resolutions in space and time can be accessible from different sources. They can be served in different formats and with different APIs. Criteria need to be established to decide on the strategy for sourcing data - producer, data hub or a data broker. Standards lower the risk, accelerate the connection to new data sources and facilitate architectures securing the data access from various sources.

3.5. Step 5: Data or application access

The access strategy should be defined for each specific situation. Data may be accessed a few times to run a one shot application over an archive or the application may run regularly on real time data, accessing the data via an API. Even when serving the same datasets, the depth for online data catalogue and the mechanisms to access the archived data vary from one platform to another.

Most of the pilots require to access data which are provided by the platform and others which are not. Strategies are needed to access data seamlessly and to combine resources which might only be available from certain providers.

3.6. Step 6: ICT Resources specification

This step involves specifying the resources required by the application, identifying if these are resources common to several cloud providers or are restricted to specific cloud platforms. The strategy requires defining if the requirements are critical, if response time is appropriate and if a limited scope of resources could be a risk to the Pilot.

3.7. Step 7: Platform selection

Most of the pilots have expressed the need to have more information about the Earth Observation platforms. Such precisions would allow assessing the risks of using such assets. Clarity is required regarding the basic service provided by the platforms, the APIs, compliance to open standards… Eventually, the question can be raised on the comparability or unique values of the DIAS and other platforms, to compare their services, performance, interoperability and ease of access. Each pilot can have its own criteria such as timeliness for Near Real Time applications, pricing, ease of deployment of the code, processing efficiency or ease of access for external users. Again here, the challenge is to clarify criteria for the fitness for use.

3.8. Step 8: Integration

Integration requires being aware of the level of dependence and risks to become platform dependent. The risk should be weighed against the possibility to include flexibility in the design, with the view to facilitate upscaling. With increased complexity, it might be safer to go through an integration environment implementing a baseline of standards that will allow deploying on several platforms. Standards are critical to lower the dependency and increase the agility in a young and evolving resources landscape.

3.9. Step 9: Deployment

For critical applications, anti-fragile strategies, or highly complex processes may command duplication, redundancy or deployment on multiple platforms. In such cases, the question might arise whether to deploy all the components of the application on the same platform or to chain components deployed over several platforms.

3.10. Step 10: Resulting data or service publication

Maximizing impact requires targeted dissemination to appropriate audiences, in the appropriate context, under the right format, with the right messages. Recommendations are needed to publish and disseminate Pilots over the different portals available.

3.11. Step 11: Exploitation

Data sources have a life duration and despite the high complexity of the developments, services providers have to be aware of the data sources lifetime and be prepared to input data changes.
For critical missions, Single point of failure analysis might be needed to guarantee a high reliability service.

4. CONCLUSION

Beside intensifying the capture of the lessons learned from the many pilots implementations which cover a broad field of applied domains, next year will bring the opportunity to review in each step the activities that can empower the upscaling activities via the access to knowledge, markets capital and technology that é-shape is developing [3]. This will include interoperability and reproducibility.

5. ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 820852

6. REFERENCES

[1] T. Ranchin, L. Ménard, N. Fichaux, M. Reboul and all the e-shape partners "e-shape – EuroGEO Showcases: application powered by Europe contribution to EuroGEO and to the development of the EO industry" ", IGARSS 2021, Brussels, Belgium, July 11 – 16, 2021

[2] R. Barbier, P. Le Masson, B. Weil. "How to develop a pilot for and with users? Co-design to grow an ecosystem of efficient service designers", IGARSS 2021, Brussels, Belgium, July 11 – 16, 2021

[3] E. Mamais, F. Piatto, E. Pajot, "Upscaling European earth observation solutions through a comprehensive portfolio of tools – the case of e-shape", IGARSS 2021, Brussels, Belgium, July 11 – 16, 2021