EuroGEO Showcases: Applications Powered by Europe

e-shape Best Practices Draft

e-shape

The e-shape project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 820852

ABSTRACT

This document is a **Proof Of Concept (POC)** provided as a **first draft** of the e-shape Guide development.

The e-shape guide for development does not attempt to develop expertise on all topics but rather to share a holistic approach and understanding, to support all the EO stakeholders' profiles in their global process from data to product delivery and in their interactions with their multidisciplinary teams. Its goal is to mainstream a baseline of knowledge to develop the community, intensify the connections between the different expertise and skills, and speed up the development process from the concept to a prototype and operations in a domain that constantly evolves, and is resource and knowledge-intensive.

The concept is to summarized succinctly the different topics and exemplify them as often as possible with the experience of selected e-shape pilot. The value is less in the technical details than in the holistic approach from an idea to results exploitation, taking the best of each e-shape pilot, whatever their thematic domain to support cross domain and cross expertise fertilization. This POC provides a draft table of content and the development of 2 or 3 selected paragraphs to illustrate how the final draft practices will be designed.

The strategy to bring this document to the final deliverable involves:

1. All the paragraphs will be documented,
2. Relevant paragraph should be exemplified with some pilot's experience.
3. The document will be reviewed to make it as simple as possible to read,
4. More external references to deepen each topic will be added,
5. External experts will be invited to co-write an executive summary with key messages relevant to the EuroGEO forum.

Document Type	Report
Document Name:	e-shape Best Practices Draft
Version:	Final
Date:	Jan 31st, 2023
Status:	Final
Dissemination level:	PU

Authors, Reviewers				
Author(s):	Marie-Françoise Voidrot			
Affiliation(s):	OGCB			
Further Authors:				
Peer Reviewers:	PMT			
Review approval:	PMT			
Remarks / Improvements:				

Version History (preliminary)			
Version:	Date:	Comments, changes, Status:	Person(s) / Organisation Short Name:
V0.1	5 December 2022	1st draft version	Marie-Françoise Voidrot
V1.0	13 December 2022	1st draft version reviewed	PMT
V2.0	25 Jan. 2023	Implementation of review comments	Marie-Françoise Voidrot
vfinal	31 Jan. 2023	Final version for submission to the EC	PMT

VERSION NUMBERING	
v0.x	draft before peer-review approval
v1.x	After the first review
v2.x	After the second review
Vfinal	Deliverable ready to be submitted

STATUS / DISSEMINATION LEVEL			
STATUS		DISSEMINATION LEVEL	
S0	Approved/Released/Ready to be submitted	**PU**	Public
S1	Reviewed	**CO**	Confidential, restricted under conditions set out in the Grant Agreement
S2	Pending for review		
S3	Draft for comments	**CI**	Classified, information as referred to in Commission Decision 2001/844/EC.
S4	Under preparation		

1 ABSTRACT OF THE BEST PRACTICES

The e-shape H2020 Project brings together decades of public investment in Earth Observation and more recently in cloud capabilities into services to the citizens, the industry, the decision-makers, and researchers. e-shape promotes the development and uptake of 37 cloud-based pilot applications, addressing the Sustainable Development Goals, The Paris Agreement, and the Sendaï Framework.

The pilots address seven different thematic domains, building on GEOSS and on the Copernicus data pool as well as diverse computational infrastructures. The consortium gathers 68 members from 17 European countries, Ethiopia, Egypt, and Israel. It is a major European contribution to EuroGEO.

This Guide for European Earth Observation application developers, decision-makers, and experts, delivering best practices to use Earth Observation resources is based on the experience collected during the project and shares such knowledge in an accessible way. It provides a unique source and guidelines to increase the usage and exploitation of Earth Observation in the thematic domains addressed by e-shape. It aligns as much as possible with the European Commission's digital strategy supporting "the digital transformation journey broadening the scope to the move from IT to digital transformation, from digital or EO scientific skills to EO digital culture and from technology as a service provider to digitalization and digital ready policymaking." This is done by mixing digital background, scientific challenges and policy-making support to make them accessible and shareable in a common baseline of understanding that is, maybe, the essence of an Earth Observation “digital culture”.

e-shape has captured the requirements and lessons learned out of the implementation of the 37 pilots over more than 70 platforms. It identified all essential elements to develop a successful Earth Observation application that builds on top of the available European Earth Observation resources and how to publish its results to make the Findable, Accessible, Interoperable and Reusable on the web.

e-shape has generated a large amount of complex information and a major challenge of the Guide is to address the concerns in a progressive, logical, and comprehensive way, making complex technical issues and challenges accessible to all. Sharing this baseline of knowledge and understanding will enable further, broader and faster collaborations.

The wealth of knowledge collected is organized via an abstract generic and reproducible workflow from data discovery to results publication, from the simplest to the most complex issues. The report is articulated along the steps listed in Figure 1.

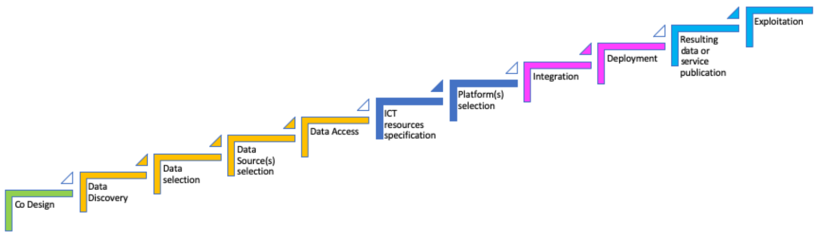


Figure 1: Generic reproducible Pilot development workflow. This scheme defines the structure of this report.

This approach is based on 11 steps, inspired by the proven and successful 5-step-user experience defined by the NextGEOSS H2020 project, that has been extended and detailed to structure and organize the topics identified by the pilots in the e-shape initial assessment as issues or challenges. The NextGEOSS 5-step-user experience was itself an adaptation of Terradue corporate Platform Operations proved procedures for the Ellip Solutions.

References:

- NextGEOSS 5-step user experience: https://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-46/3.%20Wednesday/2018.10.24_13.30_NextGEOSS.pdf
- Ellip solutions: https://www.earthobservations.org/uploads/wp23_25_global_wildfire_information_system_implementation_plan.pdf

2 Target Table of Content for the Final Best Practices

To date the target table of content is provided below. This could evolve based on the remaining work.

In this document, only the paragraphs in bold and followed by an asterix are provided as a Proof of Concept.

They provide a theoretical description of the topic, a feedback from one or more pilots and some lessons learned. This is the target for each paragraph except for some introductions that are general to help the reading. The “Introduction to the Best Practices” and the paragraph on the “different types of Earth Observation Data” are provided here as an example for these general introductions who do not provide examples and lessons learned.

3 INTRODUCTION TO THE BEST PRACTICES

Earth Observation (EO) involves a large range of data sources: not only from satellites, but also from ground-based and oceanographic observations collected through in situ, radars, and citizen sciences. From the angle of data science, EO includes data acquisition, data quality monitoring, data processing, science, statistics, and all types of computer sciences such as telecommunications, from data management to very big data management, cloud, web, mobile technologies, web analytics, computer graphics and visualization, and more. From the skills angle, EO includes soft skills such as human factors to face the complexity of the challenges it can support, the complexity of the Earth Observation products themselves, and of their production processes from data acquisition to product delivery. Human factors refer to environmental, organizational and workflow factors, human and individual characteristics which influence behavior at work in a way that can affect health, safety and efficiency. To provide the right product at the right time with the right skills for the right decision, EO mobilizes multidisciplinary teams including experts from data technologies, sciences, informatics, human factors, business, communication, and legal issues, who need to share a common understanding framework and a baseline of knowledge to cooperate efficiently.

This guide does not attempt to develop expertise on all topics but rather to share a holistic approach and understanding, to support the EO stakeholders in their global process from data to product delivery and in their interactions with their diverse background colleagues. Its goal is to mainstream a baseline of knowledge to support the community, intensify the connections between the different expertise and skills, and speed up the development process from the concept to a prototype and operations in a domain where the data constantly evolves and is very expensive to produce and process. The report might help specify some of the new or evolving terminologies and buzzwords resulting from the quick dynamics of the Earth Observation domain.

The report links to many external technical references to keep the holistic vision and understanding more readable and avoid quick obsolescence in a domain that is constantly and quickly evolving in terms of data volumes and data types, technological capacities, challenges, and impacts addressed, industrial networks and market structure, users' maturity. Information that can face quicker obsolescence is gathered in annexes and referred to in a more generic or conceptual way into the body of the document.

Based on the initial assessment of the e-shape pilots, the project has gathered an extensive understanding of the Pilot’s teams' expectations and needs on one hand, and of the European Earth Observation resources they were using or planning to use on another hand. Considering the number of participants and their diversity in terms of nationality, scientific or technical background, gender, and age, e-shape partners can be seen as a representative sample of the Earth Observation community and the issues or expectations that they have expressed can be used as a catalog of topics to address into the best practices. As a matter of fact, they cover all the workflow from data to product or service.

The development of an Earth Observation application requires addressing the 3 canonical scenarios:

1. how the users discover, access the data, or run the pilot
2. the new or improved EO service scope and development needs to clarify the interactions with the EO resources (platforms and data) used as an external resource
3. the publication, dissemination ... of the results of the new or improved service.

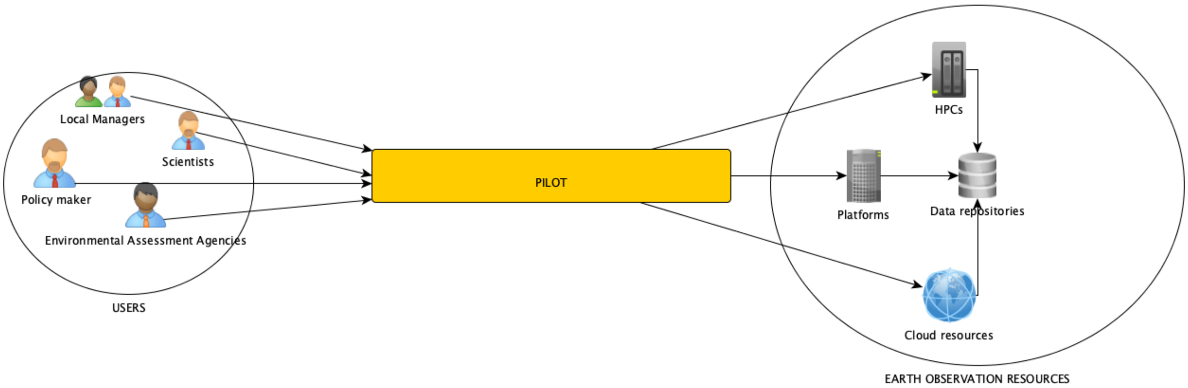


Figure 2: How do the user discover, access or run the pilot and how does the pilot interact with EO resources?

4 Co-Design

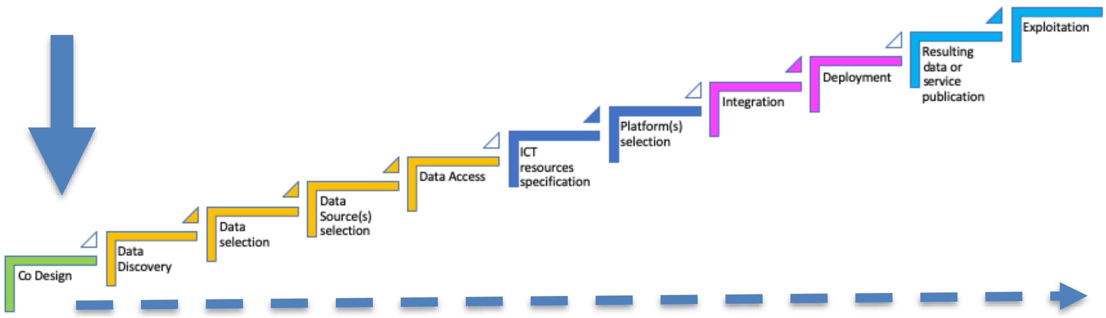


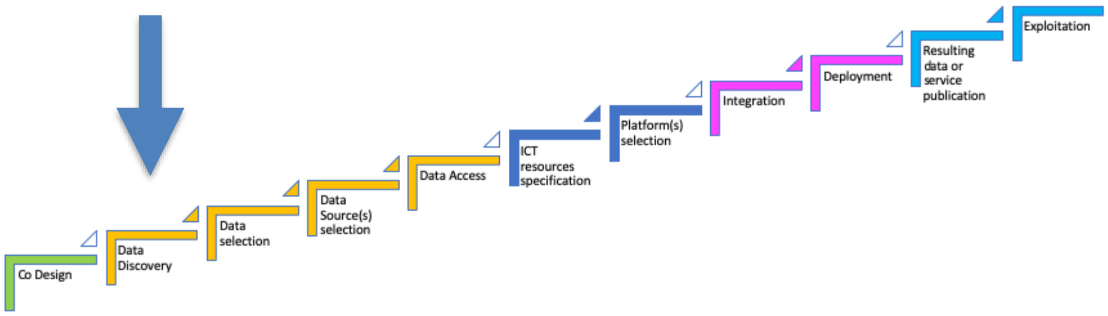
Figure 3 : Co Design activities in the Development workflow

4.1 Introduction

4.1.1 Best practice #1: A diagnosis process to help the pilots to better structure their co-design strategy

4.2 On-going reflection on further co-design routinization

5 DATA, PROCESS, AND APPLICATIONS DISCOVERY



5.1 From Data to Online applications

5.2 Community Portal, GEOSS, Google Data Search

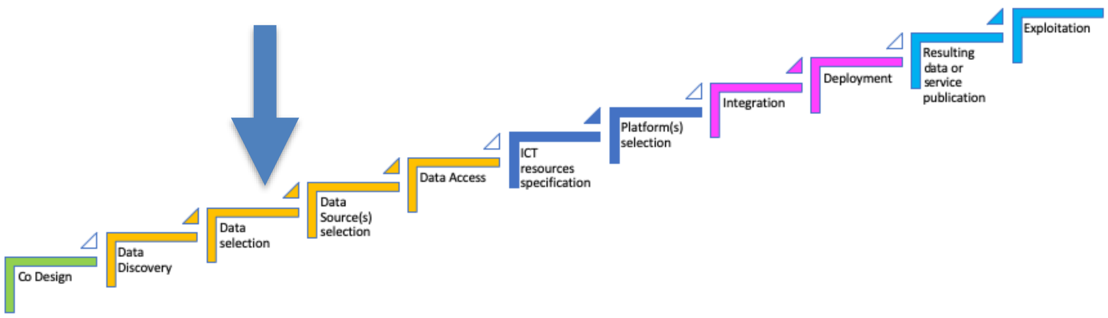
5.3 Standards for Data Discovery

5.3.1 Opensearch

5.3.2 Schema.org

5.3.3 The OGC Definition Server

6 DATA SELECTION



6.1 Data assets analysis

Some e-shape partners have raised the issue of the Data assets analysis in the initial assessment.

Analyzing the assets of data is multi-dimensional. It can be related to its variety to mitigate the limitations of each EO data type, its volume (ex: Remote sensing data), a specific value it brings to a community (ex: a single observation in the middle of the oceans), its preprocessing done by the data provider to hide part of the complexity of the measurement process (Analysis Ready Data - ARD), its relevance for a specific Domain (EVs), its operational reliable delivery in a user centric way (ex: Copernicus), its coverage, its time frequency, its density or any of its characteristics and last but not least: its frequency of use.

The following paragraphs will raise awareness on part of these dimensions.

6.1.1 Different types of Earth Observation data:

Earth Observation refers to many different types of data that have all assets and weaknesses. The good news is that they are fully consistent by essence as they all capture some measurement of the same reality. and they are complementary as their strengths and weaknesses are not the same, but their variety introduces a physical and technical complexity that requires expertise and processing to extract the information and signals that can be relevant for decision-making.

We will not develop here full documentation on the different data types but just introduce some characteristics to highlight, the diversity, complementarity, and complexity of the data preparation to hide the specificities of the measurement process to the users and then better highlight their different assets.

The e-shape pilots have used all types of data demonstrating that all of them are needed.

6.1.1.1 Remote Sensing

6.1.1.2 Geostationary satellites and polar orbital satellites

6.1.1.3 In situ data

6.1.1.4 Citizen science data

6.1.1.5 Numerical models

6.1.1.6 Historical data and Time series

6.1.1.7 Reanalysis, Forecast, Seasonal and Sub-seasonal climate data

6.1.2 Analysis Ready Data

CEOS has been the first to formalize the concept of Analysis Ready Data ARD as "CEOS Analysis Ready Data (CEOS-ARD) are satellite data that have been processed to a minimum set of requirements and organized into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets."

The goal is to make satellite data more usable by users who are not high experts in raw data satellite technologies and cannot download big amounts of data due to bandwidth or infrastructure limitations. It also avoids running many times the same basic processing that is needed by all. ARD is a major enabler for mainstreaming the use of satellite data.

The OGC report "OGC Testbed-16: Analysis Ready Data Engineering Report" (OGC 20-041) generalizes the ARD concept and studies its implications for the OGC Standards baseline. In particular, the ER analyses how modern federated data processing architectures applying data cubes and Docker packages can take advantage of the existence of ARD. The main characteristics of the CEOS ARD for

Land (CARD4L) can be extended to other dataset sources including in-situ measurements. More work will follow on this topic in the OGC community.

The PROBA-V satellite was launched in May 2013 and reached the end of its operational lifetime end of June 2020.

But late Nov 2020, the pilot S1P3, was still using the non-corrected PROBA-V products for Africa (the processing line was temporarily re-established for this purpose) because there was no alternative : the Sentinel-3 SYNERGY data from Copernicus that was planned to be used at the start of e-shape, had been rejected by the pilot since the atmospheric correction is of poor quality; this proved to be specifically true for the highlands of Ethiopia. This is "the" reason the pilot needed Bidirectional Reflectance Distribution Function-BRDF adjusted data.

VITO stepped in to fill the data gap with PROBA-V

Past BRDF-adjusted 1 km resolution data based on Spot-VGT and Proba-V were made available on request (courtesy of VITO), and this includes the technical and user manuals (including release candidate versions). Our pilot was beta tester of these data. So far the experiences are excellent. This data archive covers 20 years (1999 onwards), which provides the variability (impacts by weather) as needed to proper establish the thresholds by zone of the lower extreme NDVI-ranges. The pilot (eShape; UTwente and Mekelle University) processed the past BRDF-data in anticipation of, and preparation for continuity of the VICI insurance in 2021.

VITO provided later BRDF-adjusted data at 300 meters resolution, derived from Sentinel-3 imagery. The provisioning has met some delays that could be mitigated.

Authors: Andy Nelson and Kees de Bie, UTwente

Lessons learned on ARD:

The concept of ARD is fully adopted by the users and several pilots have expressed the need to have more ARD products available to save time, and processing costs and benefit from the upstream expertise.

ARD products are defined by data providers and should converge, at some point, with Essential Variables being defined by the users 'communities.

References on ARD:

- CEOS ARD: CEOS Analysis Ready Data: https://ceos.org/ard/
- OGC Testbed-16: Analysis Ready Data Engineering Report: https://docs.ogc.org/per/20-041.html
- OGC ARD Working Group Charter under public review: https://www.ogc.org/standards/requests/259?utm_source=phplist893&utm_medium=email&utm_content=HTML&utm_campaign=Public+Comment+Requested+on+Draft+Charter+for+new+OGC+Analysis+Ready+Data+Standards+Working+Group

6.1.3 Identifying Essential Variables to assess the upstream data assets

6.1.4 The Copernicus Data and Services

Copernicus is the European Union's Earth observation programme, looking at our planet and its environment to benefit all European citizens. Copernicus is comprised of three components: Space, Services and In Situ.

- Copernicus In situ: https://insitu.copernicus.eu/about
- Copernicus Space: https://www.copernicus.eu/en/copernicus-satellite-data-access

The 6 information services draw from satellite Earth Observation and in-situ (non-space) data.(source https://www.copernicus.eu/en/about-copernicus and https://insitu.copernicus.eu/about)

- Copernicus Information Services:
    - Copernicus Atmosphere Monitoring Service - CAMS: https://www.copernicus.eu/services/atmosphere
    - Copernicus Land Monitoring Service - CLMS: https://www.copernicus.eu/services/land
    - the Emergency Management Service – Mapping - EMS: https://www.copernicus.eu/services/emergency; The Emergency and Management Service can be activated only by designated authorized users.
    - Copernicus Marine Environment Monitoring Service- CMEMS: https://www.copernicus.eu/services/marine
    - Security: https://www.copernicus.eu/services/security
    - Climate Change Service - C3S: https://www.copernicus.eu/services/climate-change

All these data and services are free of charge and their license only requests basic crediting. They are very operational 24/7.

The Copernicus Programme implements a user-centric development process and organizes many events to strengthen the connection with its users' community and share knowledge on their products and services.

Lessons learned on Copernicus data:

The Copernicus data and services are very valuable and 34 of the 37 e-shape pilots used Copernicus data (See: Annex 4: Copernicus Services used by the e-shape pilots).

6.1.5 Web Data Analytics

Usually, products are co-designed with users in face-to-face meetings and validated by the primary users. Developers know the users. After a web publication via one to several portals, user interfaces, search engines, links in documents or social media will support the outreach and upscale of the product's usage. At this stage, the developers don't know the users anymore. The only way to learn about the new users will be via web analytics services and tools. **Website analytics tools** allow us to track and analyze user behavior on a website to optimize the website for the target audience. There are many analytics tools available in the market such as Matomo, Mixpanel, Hotjar, Woopra...They can be compared to Google Analytics which is the most popular but suffers some critics related to GDPR.

These tools allow to know the number of visitors at any time, the number of new visitors or returning visitors, the number of visitors per country, the language of the visitors, the devices they are using (I.e. desktop, mobile, or tablet), the acquisition channels (did they reach your website from a social media, a direct link, an organic search...)? which keyword did they use for their search? , the portal or data catalog, and much more.

Without knowing personally his users, the developer can discover where he can expand his market, in which language it is worth translating the website, what are the keyword he should emphasize in his website because they are the most searched, what data is the most searched, accessed and retrieved, which data is never accessed revealing either a lack of interest for this data or a problem in its publication made with inappropriate keywords. The developer is then able to consolidate its publication, to emphasize the products that are the most popular, rank the search results using these popularity criteria to help the user find more easily the data of interest for them, and for instance optimize its online/offline catalogs to make the most searched data more easily accessible. He can communicate about the products that looked to be underused.

Of course, the data needs to have a minimum audience to collect significant Web Data analytics and there can be a syndrome of the chicken and the egg in the initial phase after the publication.

These Web Data analytics can be the foundation for Technical and Marketing strategy as well as for a complete business model. They reveal the societal and business value of the data or products, being these data open or commercial.

Pilot S7P3, a service supporting the forestry sector (harvesterseasons.com), has disseminated service promotion via several platforms and portals. The service has been using Google Analytics in the background to analyze the impact of all dissemination activities respectively. It is important to state that the Analytics service requires a statement in the privacy policy, which user information is collected for statistical analysis. In the case of S7P3 it is only basic information on location, IP and acquisition channel. But even with only basic information Google Analytics provides a good understanding about users, service usage and impact of dissemination activities.

Throughout the project timeline it could be clearly seen that each promotion activity has impacted in a rise in service usage. Most importantly the service is findable directly and prominently via Google search. Additionally to just name a few promotion activities, S7P3 had the climate showcase webinar series, individual service webinar, participation at user events for the Finnish forestry sector and presentation at conferences. Each of this activity helped to raise awareness of the forestry service and directly visible in click numbers.

Through Google Analytics it was also possible to investigate various user acquisition channels. Pilot S7P3 has been promoted through various channels. Harvester Seasons' own LinkedIn and Youtube profile, proving monthly service updates and information, helped most effectively to bring users to the service platform. Articles at the portals of Finnish Meteorological Institute, Copernius, E-shape as well as WEkEO DIAS (Harvester Seasons is one of the WEkEO use cases) is as well clearly visible as strong user acquisition channels according to Google Analytics. Last but not least pilot S7P3 Harvester Season is findable via GEOSS as well as the national GEOdata portal https://kartta.paikkatietoikkuna.fi/. Even though those platforms are not yet well known in the forestry community it could be shown via Google Analytics that the Harvester Seasons service acquainted at least some visits via those platforms.

Author: Miriam Kosmale, FMI

<u>Lessons learned on Web analytics</u>

- Open data value can be revealed by the use of Web Analytics tools.
- These data are an opportunity to optimize the catalogs.
- Unfortunately, the most current free tools are US and their data is not open.

- As these data reveal the marketing value of open data, after some first analysis with open tools, developers should consider moving to paying web analytic tools to preserve these data confidentiality.
- The use of these data analytics should be developed as a powerful tool to upscale. At the time being, they are underused and we can worry that some GAFAs make better use of them than the real data producer.

6.2 Data quality

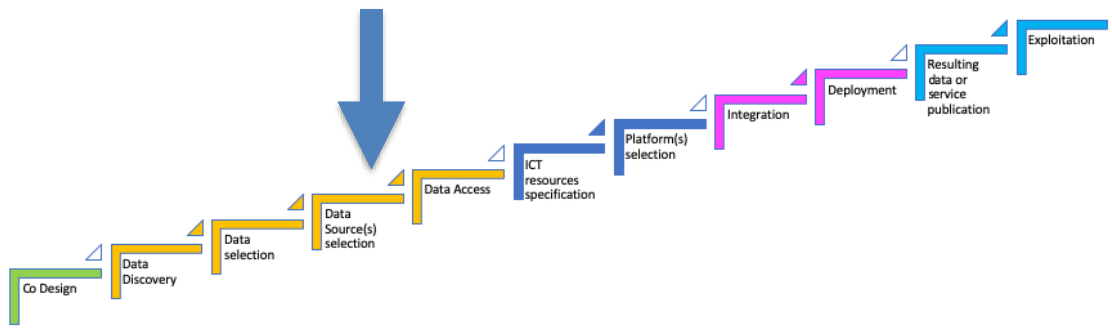
6.2.1 Data skill assessment

6.2.2 Assessing input data skills versus output data skills

6.2.3 Provenance Metadata Description

6.2.4 Gapfill based on AI or Deep Learning

7 DATA SOURCE(S) SELECTION



7.1 Data Cubes Data organization and services

The same data sets with the same or different coverage and resolutions in space and time can be accessible from different sources. The temporal depth of the online or offline archive can be different. They can be served in different formats and with different APIs. They can be accessed from Data Cubes to make the multidimensions exploration easier and more efficient.

The OGC® Discussion Paper "OGC: Towards Data Cube Interoperability" defines a data cube as *" a discretized model of the earth that offers estimated values of certain variables for each partition of the Earth's surface called a cell. A data cube instance may provide data for the whole Earth or a subset thereof. Ideally, a data cube is dense (i.e., does not include empty cells) with regular cell distance for its spatial and temporal dimensions. A data cube describes its basic structure, i.e., its spatial and temporal characteristics and its supported variables (also known as 'properties'), as metadata. It is further defined by a set of functions. These functions describe the available discovery, access, view, analytical, and processing methods that are supported to interact with the data cube."* It is a 1 to n dimension. Data cubes are often implemented by platforms serving data as a means to facilitate multidimensional exploration and offer good performances over all of them. The meteorology community has used Data cubes to serve Numerical models' outputs in the 90ies. The Numerical Model datacubes are not always dense nor regular in space and time: there are more layers near the ground than in the upper atmosphere, and more time steps in short range than in longer ranges. Some products are only processed on the levels where they can occur such as the risk of turbulence or icing. In this case, data cubes fit the production process and are not as regular and dense as the user could dream of. The concept has been adopted by the Space community to address the need to explore time series as well as spatial data exploration. In this case, the data cubes allow to go over the production process and prepare the data access in a more user-centric way than just spatial product collections for a given time that stick to the production process.

As the Datacube concept is a means to bridge the production process specificities to easy and efficient data access for the user, there can be several implementations depending on the data they served and on/or the user community they target. The OGC® Discussion Paper "OGC: Towards Data Cube Interoperability" provides a figure to illustrate several existing options.

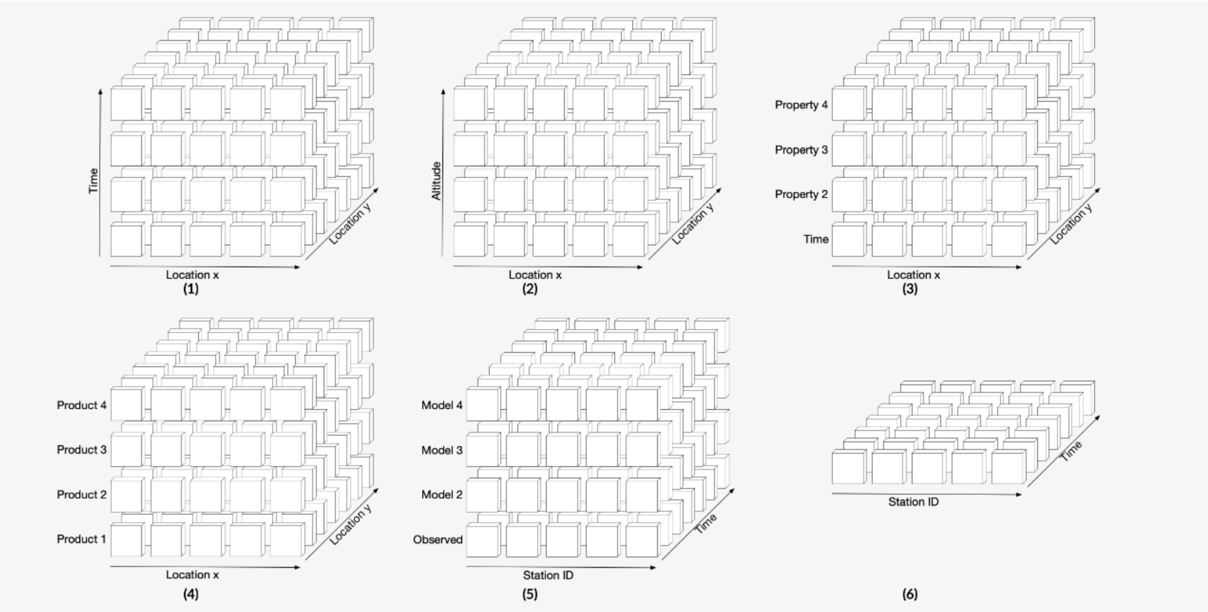


Figure :Different options for Data Cubes implementations

The OGC Discussion paper is based on a workshop co-organized by GEO and OGC that gathered data providers and data users. Interestingly they presented different types of data cubes organizations as the data providers suggested data cubes homogeneous in terms of data types when the users implemented their own data cubes merging heterogeneous data types but optimized to serve their decision process. This could be mapped with an architecture of Analysis Ready Data Cubes used to build Decision Ready Data cubes.

As a data provider or as a data user, you can get benefits from implementing a data cube.

When selecting a Data source, a Data cube service can be a good asset but it is highly probable that only benchmarking can determine what source is the most efficient for a given scenario need.

Some Data cubes used by e-shape pilots:

- EuroDataCube (https://eurodatacube.com/):

The EuroData Cube homogenizes the access to Data from different sources to make them available in one place (one-stop-shop for EO). Combined with the Cloud capacities provided by PaaS Platforms such as DIAS or AWS, it allows processing near the data with a unified API and standards. It is Cloud-agnostic and currently operational at AWS, CreoDIAS and Mundi. It allows adding custom data by implementing the "Bring Your Own Data" concept including non EO Data via xcube or to generate "data on demand". it implements several OGC standards such as OGC WMS and EO-WMS profile, WCS and EO-WCS profile. It provides an environment to run Jupyter Notebooks and to host your own applications, and a Marketplace for free or revenue-generating options to share data, applications, and algorithms via github enabling cooperation (https://github.com/eurodatacube). These are some of the major capacities, please check the website for updated and comprehensive capacities.

- Smartmet Data Cube:

SmartMet Server is a data and product server for MetOcean data based on data cubes concept, developed by the Finnish Meteorological Institute (FMI). It provides a high capacity and high availability data and product server for MetOcean data. The main feature it carries is being able to return any grid product (same as raster) in another grid with a different resolution/projection. Data from different producers and variables can always be computed in a common grid. Visualization is performed on-demand into images or OGC (tiled) web maps. Gridded data can also be served as OGC web features or downloaded into NetCDF or Grib files. The most powerful interface is the time series plugin. It is a way to query and compute on demand to a point, route, or area several variables allowing equations between them. With lua function, even more complicated processing can be performed. The Pilot 3 of the climate showcase https://harvesterseasons.com/ is using this feature to compute from 51 values of a seasonal forecast with 3 variables an index of values 0, 1 or 2. It seeks if 90% of the ensemble members are above a certain threshold for a 2. 0 is the case if 90% of values for this variable are below another threshold. 1 is for the case that good 2 or bad 0 are not reached. See the codes at https://github.com/fmidev/harvesterseasons-site as a single request to smartmet-server time-series API. The API is described at https://github.com/fmidev/smartmet-plugin-timeseries with usage examples and the API syntax. The server data in this pilot is also hosting Sentinel-3 NDVI data in addition to bias-adjusted C3S seasonal forecasts ECBSF and ERA5 reanalysis or FMI weather forecasts and postprocessing model data. The data available can be discovered with the grid-GUI plugin at https://sm.harvesterseasons.com/grid-gui . The SmartMet server system is cluster ready to enable the latency low enough for web users to be ready to wait for. Splitting the data backends to end-user serving frontends, even large data computations can be served in a short time. The pilot service is not needing this functionality but is still performing with only some latency for the data queries. Powerful caching is enabled to reuse existing queries and speed up the service.

- MEEO Data Cube

The ongoing e-shape FRIEND S6P5 pilot (https://e-shape.eu/index.php/showcases/pilot-6-5-friend), developed in the context of the climate security domain, relies on the ADAM - Advanced geospatial Data Management technology (https://adamplatform.eu/). This technology fully implements the data cube concept, with a pixel based access to multi-dimensional data that are distributed on different data cubes. In particular, the ADAM platform Data Access Service (DAS) represents the component in charge of applying the data cube approach: its implementation in front of each data source enables effective access services.

The FRIEND service is accessible through a web platform which allows the user to assess the flood risk (by mainly exploiting CMEMS, GloFAS, IMERG and ERA5 Land data as risk products), as well as to assess the flood event impact (through the GEO-DAMP products), for three pre-selected areas (Char-Piya, Australia and Darfur) with a spatial resolution of 10 m. Mainly data related to relevant historical events are considered, but the platform provides also near real time data for the risk assessment. Thanks to the ADAM data cube technology, the FRIEND pilot implements the Digital Earth concept and the data are stored on distributed systems available via standardized Open Geospatial Consortium (OGC)-compliant interfaces.

The FRIEND pilot, in particular, is based on two data cubes (both based in MEEO-ADAM technology): a MEEO customized DAS (ADAM DAS), for the risk assessment products which are pre-processed and ingested by MEEO, and a SatCen GEO-DAMP DAS which provides access to flood and water masks generated from Sentinel-1 and Sentinel-2 imageries. The ADAM technology offers WMS/WMTS services for all FRIEND raster datasets, providing subset maps that can be interactively navigated, and producing time series over a selected location, while the Web Coverage Service (WCS) is the core part of the DAS module. WCS can be queried directly via REST queries. The data cube technology is used by the FRIEND service for all the crucial operations in terms of discovery, visualization of maps and time series, processing and download of the data related to the selected specific location.

Lessons learned on Geospatial Data Cubes

- Data Cubes optimize the access to multidimensional data but they can be implemented with differences in design, interfaces or dimensions characteristics leading to interoperability issues when there is a need to interact with several data cubes.

References on Data Cubes:

- **Euro Data Cube: https://eurodatacube.com/documentation/about**
- **Simonis, I. 2021:** OGC® Towards Data Cube Interoperability 21-067
- **OGC initiative Data Cube Interoperability:** https://www.ogc.org/projects/initiatives/gdc

7.2 Umbrella Sentinel Access Point (USAP)

NOA has developed the so-called Umbrella Sentinel Hub (USH), which acts as a broker among multiple Sentinel Access points (i.e. Open Access Hub, Hellenic Mirror Site, Finish Mirror Site), Sentinel 5P hub etc.). This applications allows the user to access data from **all Sentinel missions** through a single API. Additionally, continuously harvesting metadata from multiple hubs allows for reduced latency and increased download speed (through the USH's scoring mechanism of the connected Hubs). It should be noted that USH accesses only Sentinel data and no other data from GEOS and GEO,

USH allows for the seamless access of Sentinel data through a single access point. This provides **complete global coverage**, **reduced download speeds**, **reduced latency of ingestion** and **complete availability to all Sentinel missions**. USH follows the same API logic, as the DHuS based Sentinel Hubs (e.g. Open Access Hub) to allow for the seamless transition of other Sentinel data access scripts and applications from the existing Hubs to USH.

USH is already deployed and operating on NOA owned infrastructure and can be easily deployed on other cloud environments. Nonetheless, there is no particular interest on where exactly USH is deployed, as it is a broker of metadata and not a database that physically stores the data"

A SINGLE POINT OF ACCESS FOR SENTINEL DATA, CONNECTING TO MULTIPLE SENTINEL DATA SOURCES

Searching for Sentinel data is often a complicated process due to the different missions available and the different hubs that host the data, but also to the different performances of the hubs in terms of download speed and latency (at both the inter and intra level). Thus, the Operational Unit BEYOND of IAASARS/NOA has developed the Umbrella Sentinel Access Point that brings them all together (Umbrella Sentinel Access Point (beyond-eocenter.eu).

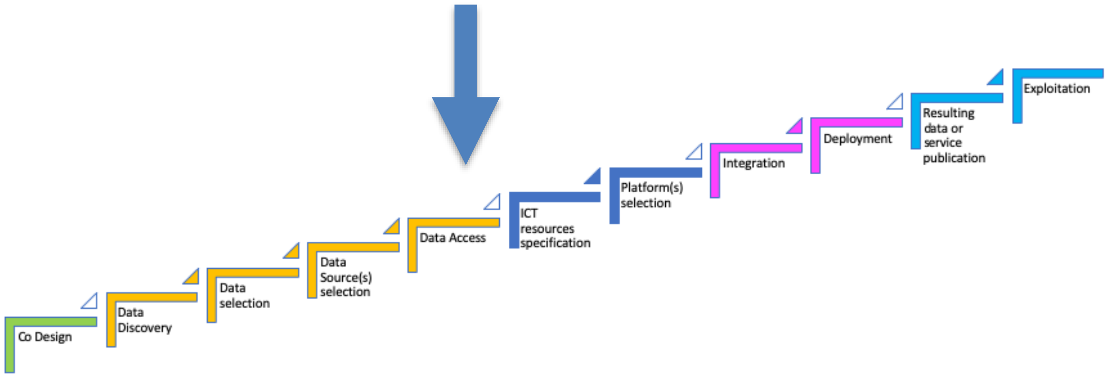
The Umbrella Sentinel Access Point acts as a single data access point which:
1. links Sentinel data hubs, regardless their back-end architecture, to a single data hub
2. provides access to all Sentinel mission data and better performance on downloading products, as products are chosen from the most appropriate data hub at any time instance based on integrity, speed and availability tests.

Currently, the Umbrella is being extended in order to allow the potential users to blend also meteorological parameters, such as temperature, aiming at decreasing the number of satellite images provided based on this parameter.

The Umbrella provides the aforementioned services via an API, an example of which is presented below:
http://umbrella.beyond-eocenter.eu/api/products/sentinel2?in_bbox=20.8,38.41,23.82,40&sensing_date__gte=2020-05-01&sensing_date__lte=2020-05-31

Author: Alexia Tsouni, NOA

8 DATA ACCESS



8.1 User interaction: accessing the Data or running the application

8.2 EO Data or EO Service, Cloud or High-Performance Computing (HPC)?

8.2.1 Data privacy

When dealing with Citizen observations or in situ in Europe, the General Data Protection Regulation (GDPR) is a sensitive issue and has to be addressed properly. Other countries also have their own regulations. GDPR is a regulation in EU law on data protection and privacy in the European Union and the European Economic Area. It also addresses the transfer of personal data outside the EU and EEA areas. The GDPR's primary aim is to give control to individuals over their personal data and to simplify the regulatory environment for international business by unifying the regulation within the EU.

The French National Institute for Information & Rights protection (Commission Nationale Informatique et Liberté - CNIL) provides a toolkit to support the compliance to GDPR. (https://www.cnil.fr/en/gdpr-toolkit) that includes a tool to carry out a Privacy Impact Assessment -PIA (https://www.cnil.fr/en/privacy-impact-assessment-pia)

When working with personal data out of Europe these considerations have to be reviewed and adapted to local regulations. This approach will typically differ by country and depends on current legislation and data-sharing policies: e.g. Tanzania is extremely strict, some organizations elsewhere require signed contracts and/or payments, while most nations/organizations require basically the presence of local staff (approved presence) during fieldwork and interviews.

Concerning S1P3 (VICI-Ethiopia), GDPR management is basically catered for through the established Public-Private-Partnership architecture. Locally, the role of ITC is limited to academic work/advice regarding the use and quality of remotely sensed imagery to capture impacts by perils (mainly of droughts) and calibration of VICI-thresholds. Implementation is fully carried out through Ethiopian public & private partners of the consortium. Any action/intervention/training/sale that is carried out in Ethiopia, always takes place "through" or at least "with" (under the responsibilities of) one of our local PPP-members. These national partners guarantee that we (ITC) properly adhere to local GDPR-aspects. In 2021, working locally on VICI-validation and perils-enquiries, will mainly be carried out through the ICIP-project that has in turn direct connections with (oversight by) MoAgr, ATA, and Kifiya. In short: we work with local partners whom locally fully carry all responsibilities regarding regulations etc.

Authors: Kees de Bie and Andy Nelson UTwente

Lessons learned on Data privacy

- Data Privacy is a sensitive issue that has to be addressed properly as soon as possible to be sure of the usability of the datasets
- It can require the support from an expert, hopefully local to the country where the measurements are done

References on Data privacy:

- Complete guide to GDPR compliance https://gdpr.eu/
- CNIL's GDPR Compliance toolkit: https://www.cnil.fr/en/gdpr-toolkit
- CNIL's Privacy Impact Assessment Software: https://www.cnil.fr/en/gdpr-toolkit

8.3 Running the pilot

The pilot application can be triggered in different ways depending on the needs and possible flexibility given to the user.

8.3.1 Login service

8.3.2 On demand processing

8.3.2.1 Graphical User Interface (GUI)

8.3.2.2 Jupyter Notebook

8.3.3 Scheduled processing

8.3.3.1 On time condition

8.3.3.2 On resources availability

8.3.4 Batch processing

8.3.5 Integrated system back-end/front-end

9 New Processing Technologies: Artificial Intelligence, Machine Learning, Deep Learning

10 Cloud technologies for Earth Observation

10.1 SWOT Analysis of Cloud Technologies for Earth Observations - Theory and Practice

10.1.1 Strengths

10.1.1.1 Theory

10.1.1.2 Return of experience from the e-shape pilots

10.1.2 Weaknesses

10.1.2.1 Theory

10.1.2.2 Return of experience from the e-shape pilots

10.1.3 Opportunities

10.1.3.1 Theory

10.1.3.2 Return of experience from the e-shape pilots

10.1.4 Threats

10.1.4.1 Theory

10.1.4.2 Return of experience from the e-shape pilots

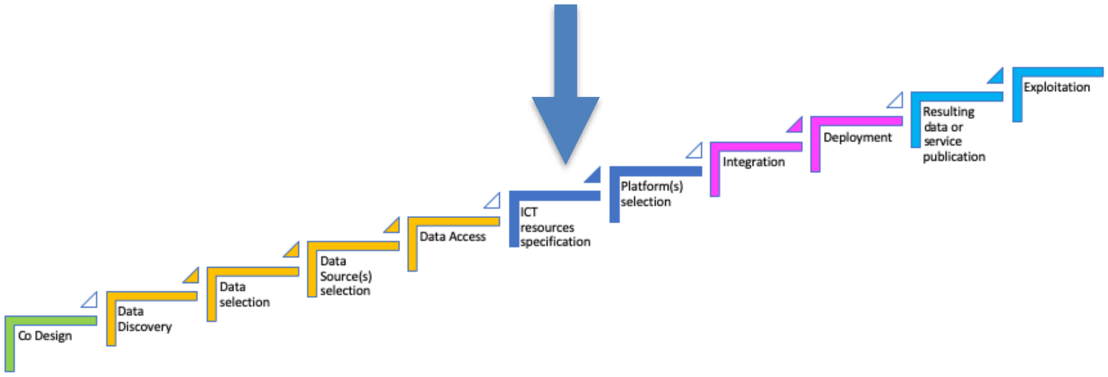
10.2 Synthesis on the usability of Cloud Technologies for Earth Observations - Theory and Practice Status

10.2.1 Lessons learned on usability of Cloud Technologies for all stakeholders

10.2.2 Lessons learned on usability of Cloud Technologies for EO applications developers

10.2.3 Lessons learned on usability of Cloud Technologies for EO Cloud Platforms providers

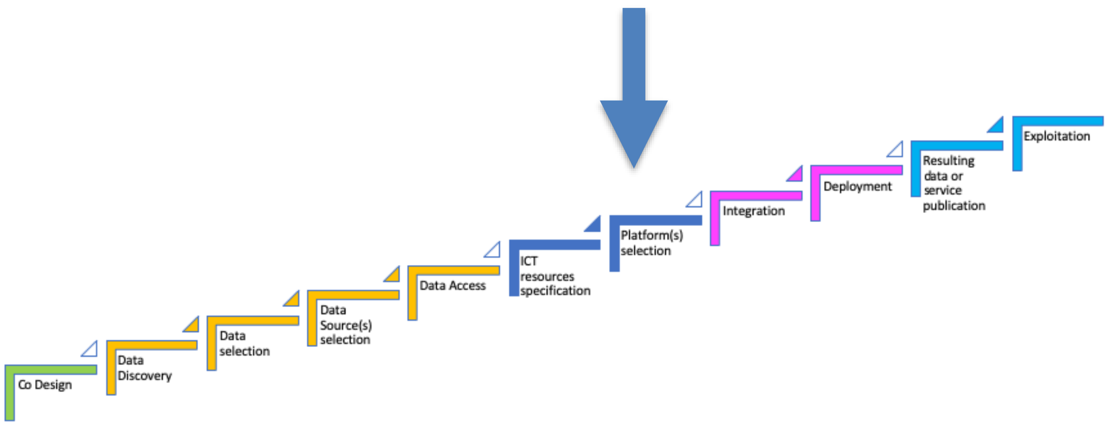
10.3 ICT resources specification



10.4 A change of paradigm

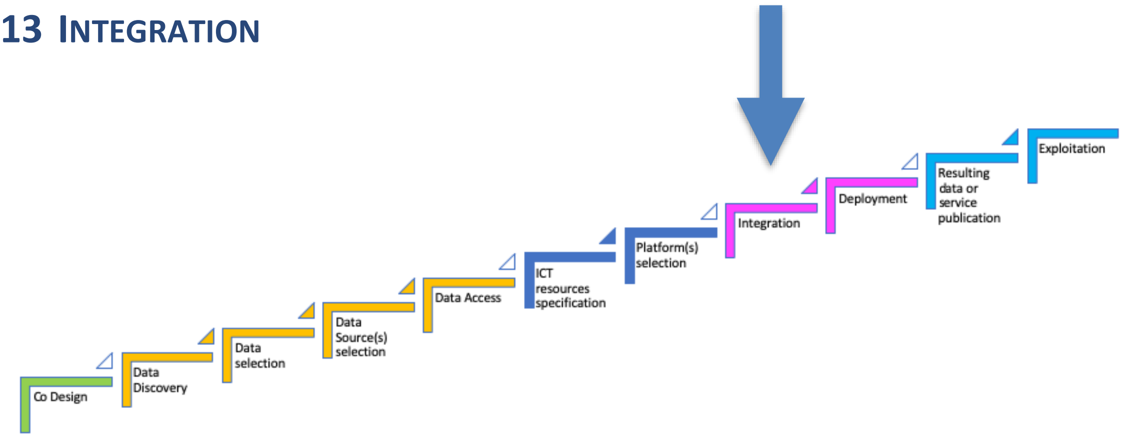
11 Sponsorship for Cloud Resources

12 Platform Selection



12.1 Contributions expected:

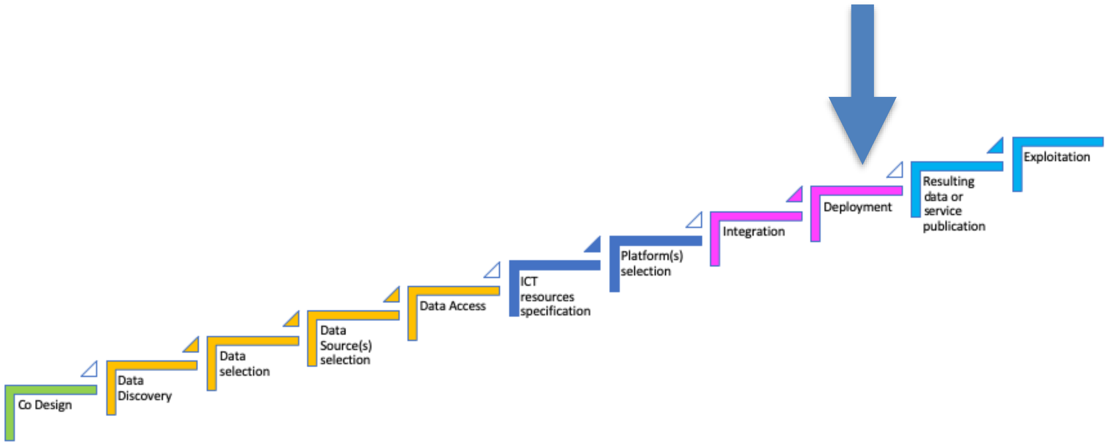
13 INTEGRATION



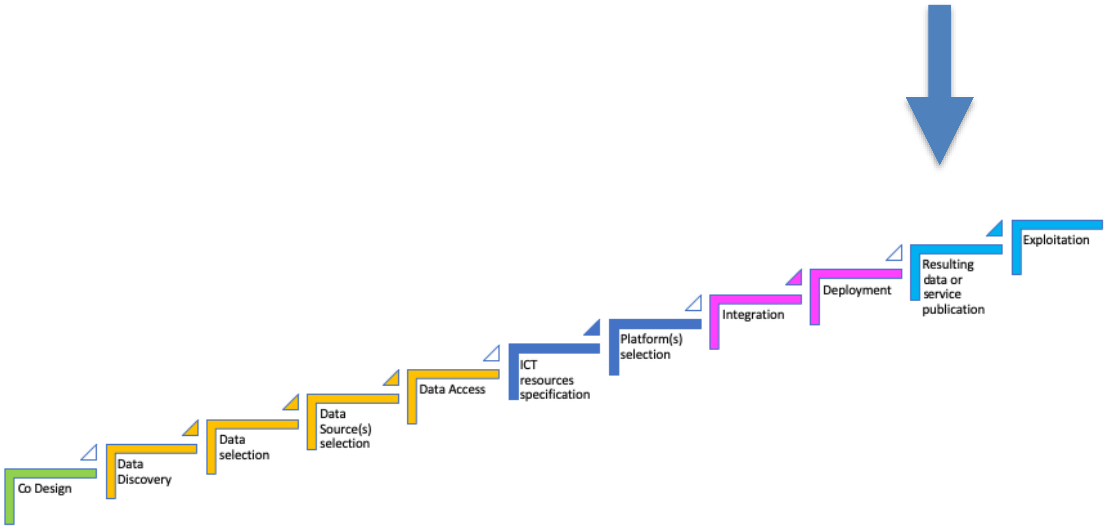
13.1 Open source tools

13.2 Containers

14 DEPLOYMENT



15 RESULTING IN DATA OR SERVICE PUBLICATION OR DISSEMINATION



15.1 Publication platforms overview

15.1.1 Data and service metadata.

15.2 Defining a data license

15.3 Publishing on the web via data portals

15.3.1 Publication platforms details

15.3.2 Publishing on the GEOSS platform

15.3.3 GEOSS features basics

15.3.4 GEOSS publication process

15.3.5 Further information

15.3.6 S3P2 published its resources as 3 services and 1 data resource.

15.3.7 Publishing on the GEO Knowledge Hub

15.3.8 Publishing on NextGEOSS

15.3.9 GEOSS features basics

15.3.10 NextGEOSS publication process

15.3.11 NextGEOSS further information

15.3.12 Link your data

15.4 Google Datasets

15.5 Publishing on the GEO-CRADLE

15.6 Publishing on the DIASs

15.6.1 Publishing on the DIASs

15.6.2 Publishing on EoMall

15.6.3 Publishing on EoWiki (knowledge component)

15.7 Metadata or standards Best practices for results publication

15.8 Making data accessible via ftp

15.9 Disseminating data via Satellite dissemination

Data can be disseminated (pushed) via ground network and/or via satellite dissemination. Under the footprint of the satellite dissemination, the reception is done via a satellite antenna that is usually quite inexpensive.

GEONETCast is a global network of satellite-based data dissemination systems providing environmental data to users around the world. It is part of the GEO System of Systems - GEOSS. It gathers several broadcast streams: EUNETCast, CMACast and GEONETCast.

As these dissemination channels are inexpensive, timely and very operational, they are also very intensively used and it can be difficult to add a new product in the dissemination catalogue.

Delivery Media	**Description**
EUMETCast	EUMETCast is the EUMETSAT contribution to GEONETCast with coverage over Europe, Africa and the Americas. Established in 2004, EUMETCast has more than 4,000 registered reception stations with more than 3,200 users benefiting from the environmental data it provides.
CMACast	CMACast is the China Meteorological Administration's contribution to GEONETCast. CMACast utilises the AsiaSat 4 satellite beam to broadcast data and products to a user community in the Asia Pacific region.
GEONETCast Americas	Broadcast covering the Americas, managed by NOAA.

GeoNetcast (see https://www.earthobservations.org/geonetcast.php) is managed/operated collaboratively by China (CMA), EUMETSAT, and the US (NOAA). Eumetsat greatly facilitates the timely and fast distribution of trillions of image data to end-users, located anywhere, at near to zero costs. In locations where data volume, download speed, and internet speed/reliability are limited, GeoNetcast offers the (only) way to remain in business. GeoNetcast is specifically used for the distribution of meteorological data (globally). Experiences are perfect. Our partners in Mekelle University have experienced the use of GeoNetcast as their only means to obtain RS-Imagery. To date, e.g. we captured on a dekad basis, all Proba-V 10-day NDVI images (Africa-window) through GeoNetcast. Local processing of many captured images takes place through the Ilwis platform (software). For VICI, specific routines as required for insurance purposes were developed, and are operational both at the NMA in Addis, ad Mekelle Uni in Tigray.

Authors: Andy Nelson and Kees de Bie, ITC, University of Twente, Netherlands

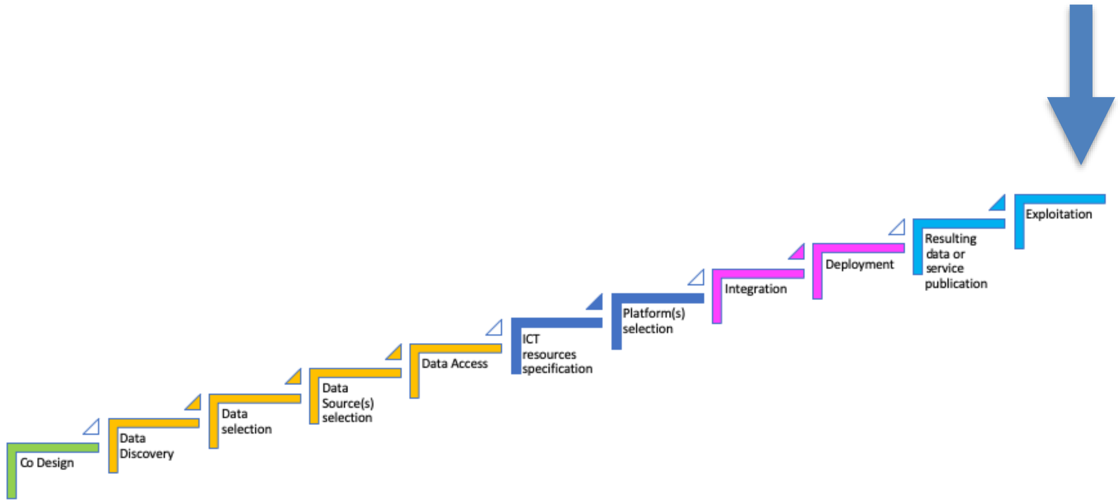
Lessons learned on EumetCAST:

- EumetCAST is a very reliable service and very valuable source of data(see above "*experiences are perfect* ") to provide data over the EumetCAST footprint
- EumetCAST provides mainly environmental data
- It has more than 4000 customers that are probably the major stakeholders in its footprint.

15.10 Publication standards

15.11 Data description best practice

16 EXPLOITATION



16.1 Common terminology ambiguities

Some words such as "Real-time" or "Operational" are commonly used and carry a lot of ambiguities. These concepts are relevant to most of the EO use cases and have to be explicited clearly in any new working context, new partners interaction or in any legal agreement to avoid any later problems.

Use cases are usually "real-time" or over past data. Real-time can mean in the second following the measurement or following the production process, to the following hours or following days. The definition refers to short-term use. It can be linked to the scientific temporal validity of the data, to the timeliness of the usage scenario, or to the technological capacities. EO Data are always potentially useful in the long term for climate studies impacts.

A platform or system offering a service can be operational because it provides sophisticated capacities and can be used for many productions and suffer interruptions of several hours for maintenance or other without impacting its "operational" qualification. This is not the case for emergency crisis situation, for businesses or for applications on the web or on mobile. In this case the mean time between failure, the recovery delay, the process in case of problem and the performance of the support will also have to be discussed.

16.2 Describing the fit for use and limits of operational products

16.3 Managing input data changes

16.4 Single point of failure analysis

16.5 Using Web Analytics tools or services to optimize the publication

16.6 Reproducibility

16.7 Billing

16.8 Data Management Plan

16.8.1 Capacity building resources on DMP:

16.8.2 References on DMP:

17 The value of Standards, Data Models and Best Practices

17.1 Introduction

18 Standards Compliance

19 Annexes

19.1 Annex 1: e-shape pilot applications

19.2 Annex 2: Abbreviations and Acronyms

19.3 Annex 3: Glossary

19.4 Annex 4: Bibliography

19.5 Annex 5: Copernicus Services used by the e-shape pilots

19.6 Annex 6: Open source software and packages used by the e-shape pilots

19.7 Annex 7: Platforms used by the e-shape pilots

19.8 Annex 8: Platforms description