Reinforcing the AI4EU Platform by Advancing Earth Observation Intelligence, Innovation and Adoption

D2.2: Consolidated technical survey and requirements report

Grant Agreement ID	101016798	**Acronym**	AI4COPERNICUS
Project Title	Reinforcing the AI4EU Platform by Advancing Earth Observation Intelligence, Innovation and Adoption		
Start Date	01/01/2021	**Duration**	36 Months
Project URL	https://ai4copernicus-project.eu/		
Contractual due date	30 June 2022	**Actual submission date**	
Nature	R = Document, report	**Dissemination Level**	PU = Public
Author(s)	B. Billière (TAS-F), M. Aubrun (TAS-F), P. Boza (INSEAD), T. Evgeniou (INSEAD), A. Chietera (THA), A.Koukourikos (NCSR-D), A. Troumpoukis (NCSR-D)		
Contributor(s)	M. Lazzarini (SATSEN), R.Hall (EQUINOR), X. Ziouvelou (NCSR-D)		
Reviewer(s)	M. Alexe (ECMWF), L. Bruzzone (UNITN), G.Weikmann (UNITN)		

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 101016798.

Document Revision History *(including peer reviewing & quality control)*

Version	Date	Changes	Contributor(s)
	14/03/22	Initialization (intro + TOC)	B.Billiere
	22/03/22	Take into account first feedback from M.Lazzarini, X.Ziouvelou, M.Aubrun => Update of intro (+ § thematic services examples)	B.Billiere
	25/03/22	Minor update	B.Billiere
	20/04/22	Interview and recommendations + First set of requirements	M.Aubrun, P.Boza, T.Evgeniou, B.Billiere
	13/5/22	new structure proposal	A. Koukourikos
	2/6/22	previous sections and diagrams integration + requirements update	A. Koukourikos, B.Billiere
	05/06/22	Existing services section	A. Troumpoukis, A.Chietera
	14/6/22	integration of interviews annex	B.Billiere (P. Boza, M. Aubrun)
	20/06/22	Final draft for review	B.Billiere, A.Koukourikos
V1	29/06/22	Update following reviewers comments. And final version for delivery. (Pending review by EC)	B.Billiere, A.Koukourikos

Executive Summary

In the frame of AI4Copernicus project, this report presents activities performed to capture from users, general expectations related to Earth Observation resources and services integrating Artificial Intelligence tools. The final goal being to make the AI-on-demand Platform, the platform of choice for users of Copernicus data along the value chain (scientists, SMEs, non-tech sector) and help to develop services with large positive economical, societal and environmental impact.

For that, two branches of activities has been followed with the intermediate objective been to build a set of requirements to be handled by the project. The first branch was to interview technical and non-technical users of Earth Observation data and Artificial Intelligence tools, while the second one was to analysis all proposals coming from the Open Call process reflecting general expectations from the market and the community.

These two activities allowed to build a requirements baseline for the project. Two types of requirements have been defined ; *Technical Requirements* related to “which resources-services are expected”, and *Operational Requirements* related to “how resources-services need to be proposed in the platform”.

The report then presents the analysis of existing resources and services provided by the AIEU project or already identified in the AI4Copernicus project baseline. This analysis allowed us to perform an assessment of Technical Requirements, meaning to identify remaining gaps corresponding to main areas to which the currently provided solutions do not fully cover the needs of the community.

Table of Contents

List of Figures

List of Tables

List of Terms & Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
AIDDL	AI Domain Definition Language
ACD	Amplitude Change Detection
AoI	Area of Interest
AIoD	Artificial Intelligence On Demand
API	Application Programming Interface
B2B	Business To Business
BR	Bootstrapping Resource
C3S	Copernicus Climate Change Service
CAMS	Copernicus Atmosphere Monitoring Service
CI/CD	Continuous Integration / Continuous Development
CORINE	Cordination of Information on the environment
CNN	Convolutional Neural Networks
CVA	Change Vector Analysis
DEMS	Digital Elevation Models
DIAS	Data and Information Access Services
DL	Deep Learning
EFFIS	European Forest Fire Information System
EO	Earth Observation
EuroGEO	European Association of Geographers
FAIR (data)	Findability, Accessibility, Interoperability, Reutilisability
GEOINT	GEOspatial INTelligence
GIS	Geographic information system
GNN	Graph Neural Networks
GUI	Graphical User Interface
HR-VPP	High-Resolution Vegetation Phenology and Productivity
IaaS	Infrastructure As A Service
IoT	Internet of Thinks
JRC	Joint Research Centre
JSON	JavaScript Object Notation
LSTM	Long Short Term-Memory
ML	Machine Learning
MTC	Multi-Temporal Coherence
NDVI	Normalized Difference Vegetation Index
OSM data	Open Street Map
RTDI	Research, Technology Development and Innovation
SaaS	Software As A Service
SWRL	Semantic Web Rule Language
TS	Thematic Service
TSO	Transmission System Operator

UNFCCC	United Nations' Framework Convention on Climate Change
UPF	Unified Planning Framework
UX	User Experience
W3C	World Wide Web Consortium

1 Introduction

1.1 Purpose and Scope

The objective of the deliverable is to establish the generic requirements framework for AI4Copernicus itself as well as its presence within the broader AI-on-demand Platform ecosystem. As such, the requirements elicitation process should cover two related but clearly distinct facets. On the one hand, the provision of meaningful, ready-to-use services and datasets that cover major needs for the development of impactful EO/AI services. On the other hand, the efficient discovery and usage of such services from both tech-savvy and non-savvy interested parties.

The present report summarises the applied approach for covering both requirement facets, taking also into account the dynamicity inherent in a relatively novel ecosystem, where service development is driven by application demand and, vice versa, application rely heavily on the pre-existence of services solving complex problems.

Summarising, the deliverable's main purpose is to steer the future development and deployment of generic and thematic services, in the context of the AI4Copernicus project. Furthermore, it aims to set the landscape for future developments and pinpoint the most critical technical and commercial directions that the EO/AI domain must follow in order to achieve significant and immediate impact.

1.2 Approach for Work Package and Relation to other Work Packages and Deliverables

Deliverable D2.2 is a direct outcome of tasks T2.1 and T2.2 of the AI4Copernicus workplan. Nevertheless, its developments are closely associated with activities in the project's technical work packages (WPs 3, 4 and 5), as well as, with the AI4Copernicus Open Calls (WP6).

1.3 Methodology and Structure of the Deliverable

The report initiates with the presentation of the methodology adopted to capture both technical and operational requirements for the AI4Copernicus services in Section 2. The two main branches of the methodology, namely the conducting of targeted interviews and the analysis of the solution proposals submitted in the context of the AI4Copernicus Open Calls, are presented in Sections 3 and 4 respectively. Section 5 summarises the findings from the two activity branches and presents the set of elicited requirements. Section 6 assesses the status of AI4Copernicus and, to some extent, the European AI-on-demand Platform with respect to the technical requirements and discusses on the perceived remaining gaps and most critical issues to be taken into account.

2 Requirements Elicitation Methodology

2.1 AI4Copernicus Framework

The methodology followed for eliciting requirements was aligned to the core objectives of WP2 and the implementation plan of the project as a whole. Based on the organisation of the project's technical approach, we retain the distinction between *Bootstrapping Resources* and *Thematic Services*, each addressing the needs of different users roles.

Bootstrapping resources are composed of data and computational services, such as data pre-processing and harmonisation to standard format. They can be either generic or related to a specific domain (see *D5.1, "Bootstrapping services and resources I"*). These services are used by *Bootstrapping Resources Users* for the development of *Thematic Services*.

Thematic Services are services falling explicitly under at least one of the 4 domains (Security, Energy, Agriculture, Health & environment) covered by the project. They are used by *Thematic Services Users* to provide at the end societal added value.

Respectively, ***Bootstrapping Resources users or Technical users*** are individuals or organisations with sufficient competences in IT, AI and EO analysis. In contrast, ***Thematic services users or End-users*** are not necessarily savvy in computational areas covered by the project but are positioned to deliver economic and societal impact through the provision and distribution of *Thematic Services*.

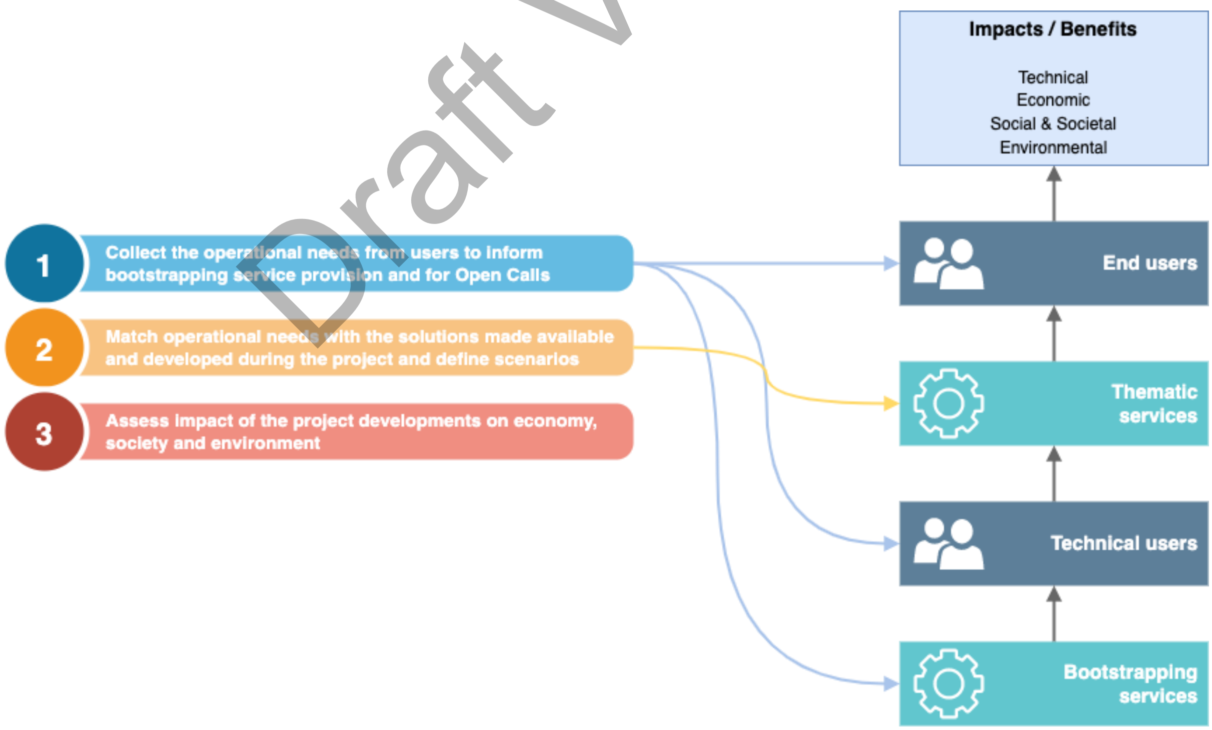

Figure 1: AI4Copernicus requirement analysis framework

The objective of the requirement elicitation process is, on one hand, to concretise generic needs from the end-user point of view for the development of thematic services. These are meant to be taken into account during the development of the services promoted by the project's Open Calls, as well as for steering efforts after the end of the project, thus contributing to the sustainability of the AI4Copernicus results. On the other hand, we also identify generic technical requirements for the development and provision of bootstrapping resources, i.e. tools and datasets that technical users need to build their thematic services efficiently and at a high quality. In the latter case, an exploration of existing bootstrapping services was carried out.

Given the above, the requirements elicitation methodology followed two branches.

The first branch dealt with the organisation and conducting of structured interviews for both *End-User* and *Technical-Users*. Each interview has been derived into a synthesis (available in Annex B), then a global synthesis is proposed collecting main recommendations. These recommendations have been used as an input to build formal user generic requirements.

The second branch dealt with the analysis of the proposals received in the context of the AI4Copernicus Open Calls. Early in the process, some examples of thematic services and associated useful bootstrapping resources has been created to support applicants building their proposal (available in Annex A). The proposals received have been then categorised under a multi-dimensional classification scheme in order to identify technological, sectorial, economical and societal focal points and needs of the submitting entities. This was subsequently used to build as an input for mapping out the current and future market needs and identify the technical points where users will probably need more urgent help from projects around the European AI-on-demand platform and AI4Copernicus in the immediate future.

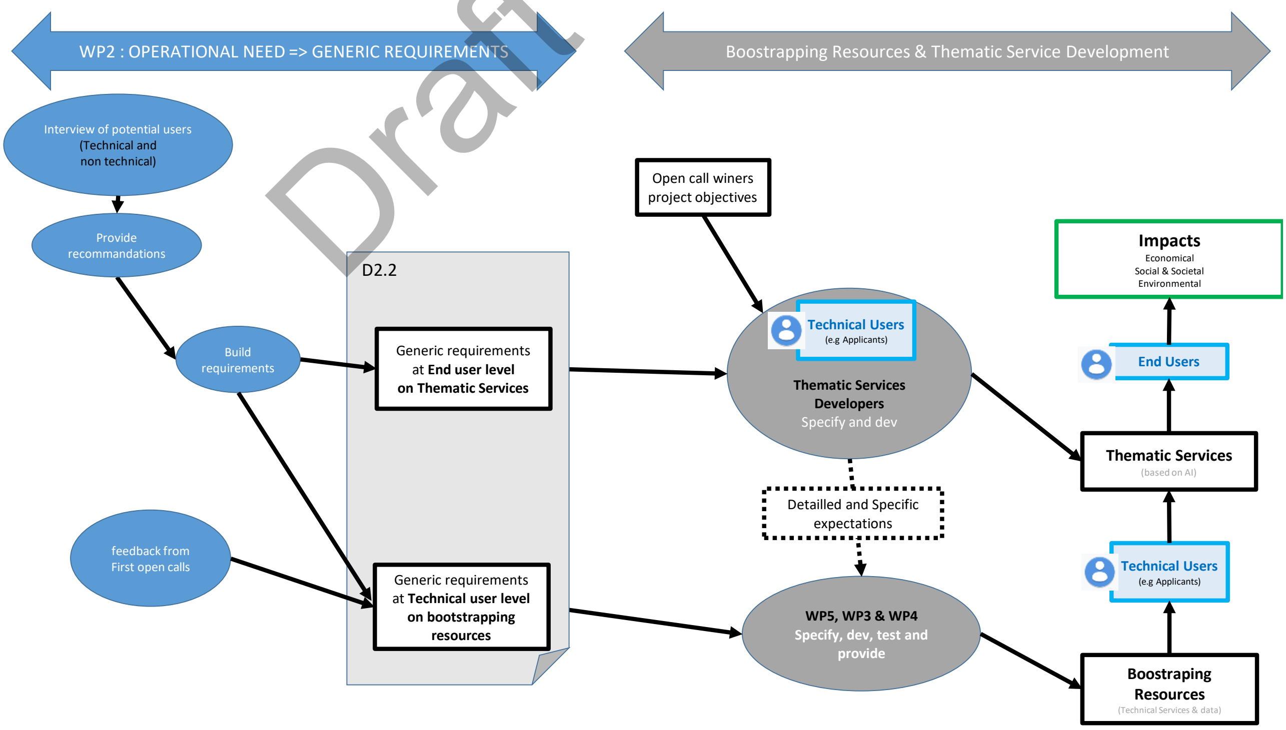

Figure 2: Requirements to Development

The last part of the methodology deals with the design of the impact assessment process for the project. The envisioned evaluation approach is based on a mapping of the currently available assets with the expressed or identified needs and an analysis of user's feedback after their operational use of resources and services. Feedbacks will be collected via a questionnaire derived from generic requirements. Questions will consider services usability but also benefit (positive impact) estimation.

The gap analysis, tentatively presented in this report, helps the steering of developing activities within AI4Copernicus and beyond the end of the project.

The user feedback analysis, will be handled in the frame of Task 2.3 and be presented in deliverables D2.3 and D2.4 to cover respectively, Bootstrapping Resources scope and Thematic Services Scope.

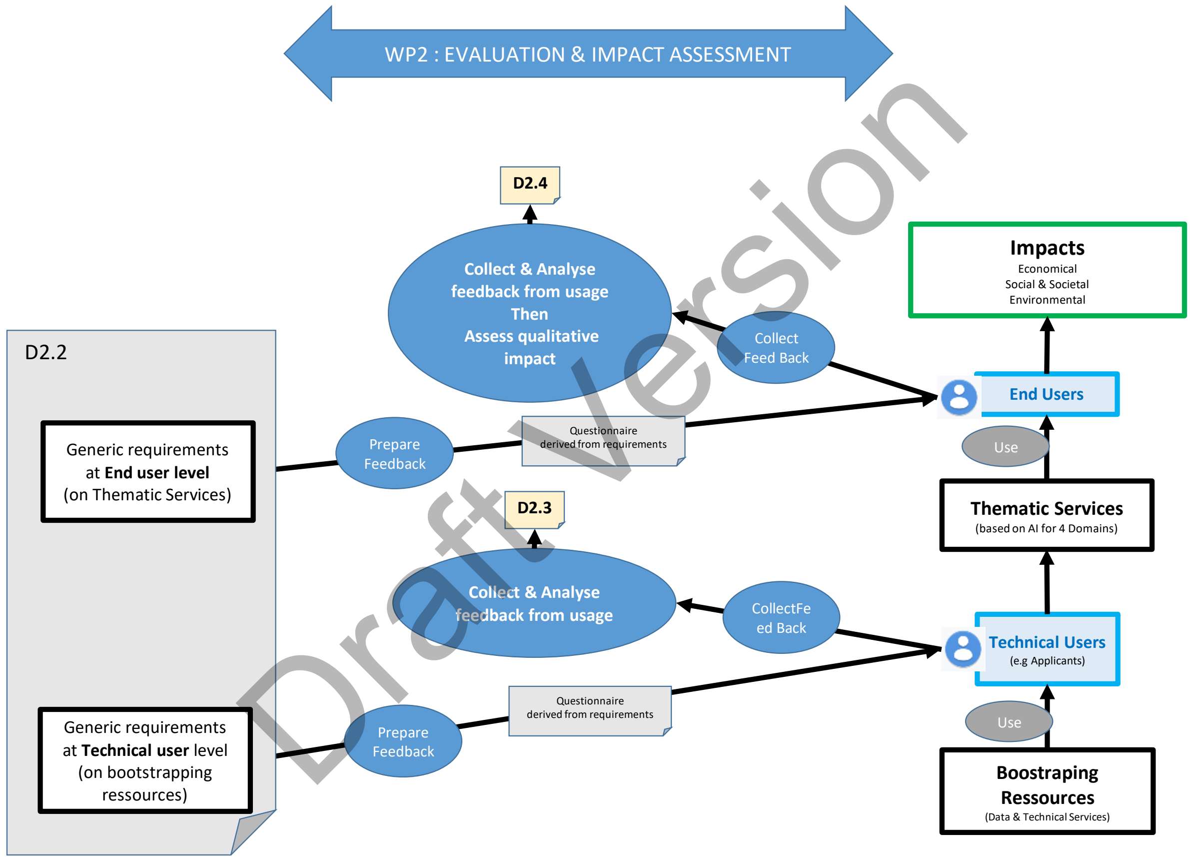

Figure 3: Evaluation Approach

The following subsections present in detail the different stages of the requirements elicitation methodology.

2.2 Organisation of interviews for capturing Operational Needs

With the contribution of AI4Copernicus partners, relevant stakeholders and end-users from the EO domain have been identified, with the objective to (1) better understand the main barriers and possible success factors related to the use of AI technology in the context of EO data and to (2) inform and derive clear requirements for the bootstrapping service and open call developers.

All together five interview-based use cases were developed, and an additional interview was conducted with an EO sector specialist (cf. Terrawatch) to have a more general view on the EO industry. The case studies included a small-medium sized company from the agricultural sector (VISTA), a major energy company using EO data and intelligent data-intensive tools (Equinor), a research centre dealing with both AI tools and EO data (CERTH), a non-profit organization specialised in providing services in the polar region (Polar View), and the Satellite Agency of the European Union (SATCEN) active in the geospatial intelligence industry.

All interviews were based on a questionnaire initially elaborated by partners participating in WP2 and validated by the partnership. The aim was to elaborate the use cases relying on the input of both technical and business profiles based on approximately 1-hour long interviews from the above-mentioned organisations. Based on the interviews a first version of the use cases was written which was later validated by the organisation involved in the case study. The conclusions of the five use cases and the one interview are included in Annex B and their insights are summarised in Section 3.

2.3 Open Calls Proposal Analysis

The second requirements elicitation branch of the adopted methodology refers to the qualitative and quantitative analysis of the proposals received in the context of the AI4Copernicus Open Calls. In total, 103 proposals were received and analysed, with the following distribution between calls. The table also showcases their distribution across the main domains tackled by AI4Copernicus, with the remaining proposals falling under additional domains. Prevalent other domains were Environment and Maritime.

Table 1: Distribution of proposals per call and thematic area

Call	Agriculture	Energy	Health	Security
1st Open Call	13	7	4	10
2nd Open Call	1	-	1	-
3rd Open Call	20	8	6	4
4th Open Call	2	-	1	1
TOTAL	36	15	12	15

The proposals were analysed on the basis of six dimensions.

1. **Targeted domain problems**: The specific domain problem(s) addressed by the solution
2. **Application type**: The computational problem targeted by the solution, where EO and AI technologies can be applied

3. **Business model**: The intended means of distribution and monetisation of the solution
4. **Scope**: Geographical range of the market to which the solution is targeted and/or applicable
5. **Data needs**: Additional data sources (beyond EO data) that the solution needs to function
6. **Core AI Challenges**: Major AI challenges inherent in the solution to which specialised technologies must be applied and integrated in the solution

The goal of the first analysis stage was to determine the degree to which these dimensions are covered by the proposals and – as a second step – to identify the generalization potential and the respective classification schemes to treat the dimensions as categorical.

Following this, each proposal's categorical value on each dimension was added to the relevant score of the class, to identify the most common/critical/interesting classes.

Additionally, the analysis of the proposals indicated additional aspects that should be taken into account when assessing the feasibility and impact of a proposed solution. Such aspects are not easily quantifiable or generalized, but in principle should be taken into account as factors affecting the impact and priority of a service to be provided by AI4Copernicus.

3 Operational needs capture Interviews summary

Based on four main questions, this paragraph provides a summary of answers provided during interviews, while all interviews records are available in annex B.

1. What are the main barriers for the adoption of AI technology and cloud services in the space domain in a non-technical point of view ?
2. What are the main successes for the adoption of AI technology and cloud services in the space domain in a non-technical point of view?
3. What are the main technical barriers and successes for the adaptation of AI technology and cloud services in the space domain?
4. What are the recommendations for the adaptation of AI technology and cloud services in the space domain?

3.1 Main barriers for adoption (from non-technical point of view)

What are the main barriers for the adoption of AI technology and cloud services in the space domain in a non-technical point of view ?

The cost to develop AI tools - The criterion has been several times mentioned in these interviews.

To develop an AI tool, entities need:

- Earth Observation (EO) data – The price of EO data has decreased dramatically in the past years, and for instance EO data is even provided for free through the Copernicus program. However, very high-resolution EO data (required for some applications) has an important cost and can require a significant budget from small companies.

- Annotated data – This is an important issue as there are few specific datasets available, they can be expensive, and it is a real challenge to develop robust datasets for different use cases while this requires a huge effort.
- Computational resources of the EO platforms, above all the DIASes – Companies are ready to pay for the EO platforms services. Moreover, there are also European projects that provide support for such services.
- Pre-existing AI tools –In order to create value for a specific organisation third-party AI tools need to be well tailored to the specific use cases. This also limits in some cases the option of buying pre-existing AI tools. However, if an AI tool corresponds exactly to the company needs and is accurate enough, it will allow to support the research and reduce the time of the development process.

The time to develop AI solutions - The criterion has been mentioned few times, mainly by developers of AI tools.

It takes significant amount of time and human resources to technically develop and integrate AI tools (several months). Furthermore there taking business needs into account while developing these tools is an even longer and complex process.

The evangelization of AI technology – The criterion has been mentioned by several organisations.

SMEs are in general open to use AI tools for future projects, but it is important for them to first to understand (1) the benefits of such tools compared to their current solution, and (2) the limits and constraints of this technology, which is rather complex to develop and operate. Moreover, there are different types, examples of AI tools used for different use cases, which makes the technology really complex to understand.

The knowledge accumulated within the research institutions concerning this technology, is usually brought to market in limited cases. So, it is important to support collaboration and technology transfer between research related organisation entities and private entities to adopt this technology.

The confidence in AI results – The criterion has been mentioned several times, mainly by end-users (potential or not) of AI tools.

In the EO domain, AI tools are currently very rarely used for operational decision-making processes. They are mainly providing assistance for humans because decision-makers, while the level of trust in AI tools still needs to evolve. Three reasons have been highlighted to explain this lack of trust in AI solutions:

- New technology – Technology needs still to mature;

- Robustness – This robustness of AI tools is still not always satisfactory. Models are often too specific to their training dataset (e.g. areas, objects, season), and thus cannot have a wide operational use;
- Explainability – The results are often hardly explainable, which can undermine consumer confidence. AI systems are often compared to black boxes.

The competence to develop AI solution in the space domain – This criterion has been mentioned several times.

According to most participants, developing AI tools required specific computing skills, but generally those working in the spatial domain are scientists and engineers, having usually some programming knowledge and experience that can also be used for AI tools. However, what is more complex is to find people who have a dual competence for the specific use-case and EO data, such as polar regions, security, etc.

The complexity of the space domain – This criterion has been often mentioned in the interviews by entities both outside and inside the EO sector.

Most of the companies (outside the spatial domain) are not fully aware of what can possibly be achieved with EO data and technology.

Huge amount of EO data is available, some of them freely, leading to significant competition on the market. Furthermore, the number of existing EO platforms has also considerably increased during the past years (e.g., Amazon Web Services, Google Earth or the DIASes), but end users do not necessarily understand the differences between the different platforms. Thus, end-users need to invest time and do not know where to start, but most of the time prefer to give up.

The time to change the mindset concerning cloud platforms – This criterion has been mentioned few times.

Few decades ago, the mindset was to “own the entire solution”, that is mean have the data locally available and work on the internal infrastructure. There is an increasing acceptance that collaboration in the spacesector is important but it takes time.

The sustainability of EO platforms – This criterion has mainly been mentioned by entities from the spatial domain.

EO platforms are currently often financed by grants from EU or other governmental agencies, which have a limited time period (about 2-3 years). So, there is a significant risk that some EO platforms (without long-term sustainable funding) might have to leave the market. This sustainability issue motivates end-users to buy their own infrastructure or to work with larger players.

The security of cloud solutions – This criterion has been mentioned several times, mainly by end-users (potential or not) of EO platforms.

Cloud solutions raise several safety issues:

- Data protection – Work with classified information limits the number of "trusted" third-party service providers, mainly in the defense & intelligence sector;
- Corporate espionage – Having access to business related information is a very sensitive issue (e.g. areas of interest on which AI tools have been used, objects on which models have been trained, types of AI tools that have been tested).

This is why, most of the European companies do not work with non-EU data providers such as Google.

3.2 Main successes for adoption (from non-technical point of view)

What are the main successes for the adoption of AI technology and cloud services in the space domain in a non-technical point of view?

The automation of the processing chain using AI tools – This criterion has been often mentioned in these interviews.

The main value of AI tools comes from the automation of highly human-intensive processes. Indeed, AI tools mainly replace human hand-drawn or displacement tasks (e.g. produce ice maps in the polar region, check the proportion of damaged crops in the fields). The advantage to automate these tasks with AI tools is double:

- Time and resources saving – With the increasing number of EO data and the huge pressure to produce more, human solutions are not competitive anymore. Thus, AI tools allow companies to save time and resources.
- Robustness – Human performance is not always equal between two different people or even the same people at two different times. Moreover, human communication can lead to interpretation issues (e.g. if the rule is to detect the cars, do I take into account the trucks?, if I transfer my notes to the analyst, does he understand them?)

The efficacy of using EO platform – This criterion has been mentioned few times, mainly by the end-users.

EO platforms offer computational power and storage resources that do not have to be built and maintained locally.

Moreover, having all the functionality at the same place (e.g. data access, data processing and data visualization) is a big time saving for the analyst.

Thus, operational EO platforms allow data analysts to better focus on what they are paid for.

<u>*The control over EO data using EO platform*</u> – This criterion has been mentioned only once.

Being closer to the source of EO data, that is mean downloading the data from EO platforms rather than asking external providers to deliver pre-processed EO data, increasing the confidence of the end-users in the data itself, and allowing to access more up to date information.

3.3 Main technical barriers

What are the main technical barriers and successes for the adaptation of AI technology and cloud services in the space domain?

<u>*Availability of EO data*</u> – This criterion has been mentioned few time, mainly by developers of AI tools.

There are two situations in which the availability of EO data can be an issue:

- For use cases which require a quick response, such as flood detection, because of the few days revisit time;
- For use cases which require long-term analysis because Sentinel data are made available for a limited period of time.

<u>*Request of EO data*</u> – This criterion has been often mentioned in these interviews.

All the participants agreed on the fact that the request of EO data is quite intuitive, and it is important to keep it easy. Only CERTH has suggested that the searching capability could be improved by using some semantic and metadata criteria.

<u>*Accessibility of EO data*</u> – This criterion has been mentioned several times.

Most of the participants agreed on the fact that the processing of EO data is more efficient on EO platforms, above all the DIASes, when working with large-scale datasets because of the ESA limitation concerning the download of multiple data at once and the size of EO data.

<u>*Understanding of EO data*</u> – This criterion has been sometimes mentioned.

For spatial domain companies, EO data is generally easy to understand, but for non-expert companies, it is more complex. However, ESA online manuals are well-elaborated and can provide explanations.

Moreover, the exact processing levels applied to raw EO data are not always known, which can decrease the user's confidence.

Processing of EO data – This criterion has been mentioned few time.

Processing standard EO data formats (e.g. geoTiff, shapefile or png) is quite easy, but specific data formats can require a lot of effort and time.

Integration of AI tools – This criterion has been mentioned few time.

It requires a little bit of effort at the beginning, mainly to understand Docker technology, but it is not a technical barrier.

Complexity of AI tools – This criterion has been often mentioned in these interviews.

Most of the participants agreed on the fact that AI tools are often too “technical” to use. Several related propositions have been made:

- Provide concrete examples (e.g. Jupyter notebooks) on how to use the tool is really important;
- Since there is a very large range of end-user profiles, propose different versions of the tool to target all end-users;
- Make the infrastructure resource allocation (e.g. CPU, memory) totally transparent.

Understanding of AI tools – This criterion has been mentioned several times.

It is not always easy to understand what the AI tool is exactly doing and what result it is providing. One solution that has been proposed several times, is to provide a visualization layer to look at the results.

Fuse-ability of EO data – This criterion has been mentioned once.

It is really complex to know how to fuse the different types of EO or no-EO data in order to make something useful.

3.4 Recommendations

What are the recommendations for the adaptation of AI technology and cloud services in the space domain?

- In order not to be limited by data availability, it is recommended not to select use cases which require real-time answer or long-term analysis (more than few months);
- To be relevant, it is recommended to develop AI tools for big data use cases, because the value-added of AI technology is relevant above all when dealing with a big amount of data, otherwise companies prefer to work with more traditional tools;
- To target a wide audience, it is recommended to work with open source EO data because their cost seems to be a main criterion for several potential end-users. Moreover, it is recommended not to work with no-European data for security, and use as much as possible standard format, such as geoTiff, shapefile or png, to facilitate the work of the analysts that will use the tool;
- To have robust performance, it is recommended to create a large training dataset, or at least to specify or describe on which dataset the model has been trained to understand if it is useful to train a new one;
- A second advice to target a wide audience is to develop generic AI tools, because specific AI tools provide data-specific AI models which need to be redesigned and fine-tuned, what is time-consuming;
- A third advice to target a wide audience is to propose different versions of the tool because there is a very large range of end-user profiles;
- To maintain the user's confidence in data, it is recommended to provide a layer to access data without downloading them and a layer to visualize data, mainly the results. These layers must be on the same platforms that the processing tools in order to facilitate the work of the analysts;
- It is recommended to use Docker technology to encapsulate the code, because it is the one that is mainly used
- Do not forget to provide a "Read me" that explains the processing applied to data in order to keep the user's confidence. And provide concrete examples, such as Jupyter notebooks, on how to use the tool in order to get started easily and fast for the analysts.

4 Open calls proposals Analysis

4.1 Classification Schemes

The resulting classification schemes for each dimension defined in the methodology are presented in the following tables.

Table 2: Classes of domain problems targeted by proposals

Targeted Domain Problems
Land Monitoring (sub-problems: Land usage, Land characterization)
Entity recognition (plant identification, building identification, territory monitoring)
Resource usage optimization
Food security

Environmental Impact Assessment
Emergency event prediction and management
Production optimization (facilities distribution, vessel monitoring)
Urban conditions monitoring (air quality, waste)

Table 3: Application types proposed

Application Types
Decision support
Predictive analytics
Process optimization
Low-code platforms
Aggregate dashboards
Assessment frameworks
Planning automation and support

Table 4: Business Models foreseen by proposals

Business Models
SaaS (software-as-a-service)
Platform (in combination with Hardware)
B2B (integration with other systems) / B2BB
Data products (only the data are delivered)
IaaS (Insights-as-service, only analytical results are delivered)

Table 5: Geographical scopes covered by proposed solutions

Geographical Scopes
Regional / National
European / Global
Local

Table 6: Data sources required by proposed solutions

Data Sources
Soil data

Weather data
Climate data
Environmental data
Vital signals
Socio-economic
Metadata

Table 7: AI Challenges faced by proposed solutions

Core AI Challenges
Image processing
Data downscaling
Multi-factor prediction
Automated configuration and fine-tuning
Explainability / Democratization

4.2 Proposal Categorization

The following tables present the results of the classification process of all submitted proposals in the four AI4Copernicus Open Calls, with respect to the described dimensions.

Table 8: Proposal distribution by domain problem

Targeted Domain Problem	**1st OC**	**2nd OC**	**3rd OC**	**4th OC**
Land Monitoring (sub-problems: Land usage, Land characterization)	7	4	9	3
Entity recognition (plant identification, building identification, territory monitoring)	12	2	17	2
Resource usage optimization	6	1	13	-
Food security	3	-	8	-
Environmental Impact Assessment	2	2	11	1
Emergency event prediction and management	6	-	5	2
Production optimization (facilities distribution, vessel monitoring)	2	-	2	-

Urban conditions monitoring (air quality, waste)	5	2	7	2

Table 9: Proposal distribution by Application Type

Application Type	1st OC	2nd OC	3rd OC	4th OC
Decision support	13	5	21	6
Predictive analytics	8	2	15	2
Process optimization	4	1	6	-
Low-code platforms	1	-	-	-
Aggregate dashboards	4	2	7	-
Assessment frameworks	2	1	2	-
Planning automation and support	2	-	-	-

Table 10 : Proposal distribution by Business Model

Business Model	1st OC	2nd OC	3rd OC	4th OC
SaaS (software-as-a-service)	28	9	37	8
Platform (in combination with Hardware)	2	-	1	-
B2B (integration with other systems) / B2BB	2	1	8	-
Data products (only the data are delivered)	1	-	1	-
IaaS (Insights-as-service, only analytical results are delivered)	1	1	4	-

Table 11 : Proposal distribution by geographical scope

Geographical Scope	1st OC	2nd OC	3rd OC	4th OC
Regional / National	24	5	26	5
European / Global	8	2	21	1
Local	2	4	4	2

Table 12 : Proposal distribution by geographical scope

Data Source	1st OC	2nd OC	3rd OC	4th OC
Soil data	4	-	3	1
Weather data	4	2	4	1

Climate data	3	1	8	2
Environmental data	2	-	3	2
Vital signals	1	-	2	-
Socio-economic	3	1	4	1
Metadata	1	-	-	-

Table 13 : Proposal distribution by AI Challenge

AI Challenge	1st OC	2nd OC	3rd OC	4th OC
Image processing	21	8	38	7
Data downscaling	4	-	4	1
Multi-factor prediction	14	4	18	2
Automated configuration and fine-tuning	3	-	2	-
Explainability / Democratization	2	-	-	-

5 Requirements Specification

Given the results of the aforementioned requirements elicitation processes, the following subsections summarise our findings for operational and technical requirements.

5.1 Operational Requirements (from the user perspective)

Two different levels of requirements are proposed, related to Thematic Services and Bootstrapping Resources. Requirements are integrated in the following frame.

Req_title	Req	Comment

Where:

Req_title is a unique identifier following the TS_<nnnn> and BR_<nnnn> formats for Thematic Services and Bootstrapping Resources respectively.

Req is the definition of the requirement. The requirements can operate on two levels, with their descriptions using corresponding, non-normative terms for expressing the strength of the requirement. Using “Shall” for stating the requirement means that compliance with this requirement is strongly recommended. A non-compliance would require a formal justification, while a partial compliance could be acceptable. Conversely, using “Should” means that compliance with the requirement is not mandatory or critical, but desirable.

5.1.1 Operational requirements at Thematic Services level (End-User view point)

Requirements here after are related to thematic services, and to be taken into account by applicants developing their service, and also by thematic services developers after the end of AI4Copernicus project in a frame of a sustainability approach. They are generic requirements common to any thematic service.

Requirements are organized along the main tasks for a thematic service user and listed here below.

1. Task - Access the Thematic Service
2. Task - Discover the Thematic Service
3. Task - Buy the Thematic Service
4. Task - Configure the Thematic Service and execute
5. Task - Consult results (core use)
6. Task - Manage results

For each task, a description is proposed and then associated requirements.

Then transversal requirements are proposed.

Task - Access the Thematic Service

Platform_Find_Service	The platform shall provide thematic services search capability based of semantic and metadata criteria.	
Platform_EO_Manuals	The platform shall provide link to Earth Observation introduction and basic manuals for users not from the EO domain.	For example ESA manuals considered as relevant.
Platform_Service_Access	The service shall be accessible via at least one DIAS and also via the European’s AIoD platform.	Corresponds to one of the key expectation from AI4Copernicus project.
Platform_Service_in_Catalogue	In the catalogue, the thematic service shall provide its presentation including items mentioned in requirement “TS_Discover_At_A_Glance”.	

Task - Discover the Thematic Service

To support *Discover* task, the challenge for the system is to answer quickly to such user’s questions : Is the service fit my needs ? is it free of charge or not ? what are the concrete output ? Can I see an example ? what are pre-requisite for service or data use ? Can I export data or only usage in the platform? Does it require specific skills or tools ?

TS_Discover_At_A_Glance	The service shall provide a short description with an illustration of the proposal : *If the service, is free of charge or not. What are potential pre-requisite for service use. What is the usage domain of the service and typical performance. What are input and output including format constraints. Level of skill required for usage (technical or not). Terms and conditions for using the service. Indication that the AI4Copernicus project funded the service. The version of the service. Maintenance/improvement history and strategy.*	
TS_Discover_Share	The service URL should be shared quickly with anyone.	Applicable to the platform
TS_Discover_Tutorials	The service shall provide links to documentation and video tutorials.	
TS_Discover_Webinars	The service should be presented via webinars ; live or recorded.	
TS_Discover_Users_Community	The service should provide links to an users forum and community.	
TS_Discover_Example	Without any subscription, the service should provide an interactive example of result with integrated description.	Interactive if relevant
TS_Discover_inputs	The service shall be described with clear references of input/data used and a simplified description for non-Earth Observation expert user.	
TS_ Discover_Usage_Domain	The service shall provide the performance level and associated usage domain. In particular, for AI service, indicate the dataset used for the training and validation.	
TS_Discover_Favorite	The platform should propose to add a service as a favourite in user's account.	
TS_Discover_Subscription	The service should propose to subscribe to news on service or results update.	
TS_Discover_Contact	The service shall indicate contacts for technical support.	
TS_Discover_User's_Evaluation	The platform shall allow users to leave user's feedback on a service.	To ensure an improvement loop, and direct link between users and developers. It is also a way to collect feedback from users in the frame of WP2 evaluation task.

Task - Buy the Thematic Service

TS_Buy_adapted_pricing	The service price shall be adapted to the user's profile.	When the service could be an interesting asset for searchers and students, the service should be proposed with an adapted price, even free.
TS_Buy_open_Source_EO_Data	The service should use only open source EO data.	

Task - Configure the Thematic Service and execute

TS_Adaptability	The service should be easily configurable (perimeter of application and information presentation).	
TS_Estimation	Once configured, and before execution, the service should provide an estimation of processing duration.	

Task - Consult results (core use)

TS_Consult_Trustability	The service shall provide the origin of the service development, references and technical contacts.	
TS_Consult_Accuracy	The service shall provide the level of accuracy and confidence for proposed data.	
TS_Consult_Responsiveness	The service should provide a good level of responsiveness. For complex computations not responsive, the service shall provide a clear feedback on execution status and duration.	
TS_Consult_interactivity	The service shall have a good level of UX (User Experience), especially a minimum training should be required.	

Task - Manage results

TS_Results_Export	The service should use standardized format for export (e.g. geoTiff, shapefile or png).	
TS_Rest_API	If API are proposed, they should respect the REST standard.	
TS_Log_An_Issue	The platform should allow users to log an issue on a service.	

Transversal requirements

TS_platform_compliant	The service shall comply with major platform requirements such as cyber security, personal data management…	
TS_European_Data	The service shall use only European data.	

TS_Service_Definitive_Stop	The service shall announce the period during which it will be provided and maintained. Service definitive stop shall be announced to user's at least 6 months before.	

5.1.2 Operational requirements at Bootstrapping resources level (from Technical user view point)

Requirements are related to bootstrapping resources, and to be taken into account by future developers during the AI4Copernicus project and after the end of the project in a frame of a sustainability approach. They are generic requirements common to any bootstrapping service. They include transversal aspect like usability, performance, accuracy, and trustworthiness.

Requirements are structured along the main technical user's tasks related to Bootstrapping resources usage:

1. Task – Find and access the Bootstrapping Resource
2. Task - Discover the Bootstrapping Resource
3. Task - Buy the Bootstrapping Resource
4. Task – Use the Bootstrapping Resource

Task - Find the Bootstrapping Resource

Platform_Find_Bootstrapping _Resource	The platform shall provide a centralized Bootstrapping Resource search. The search shall scan AI4EU and all DIAS.	One of the main requirements for the project
Platform_Service_in_Catalogue	In the catalogue, the Bootstrapping resource shall provide its presentation including items mentioned in requirement "BR_Discover_At_A_Glance".	Applicable to the platform

Task - Discover the Bootstrapping Resource

BR_Discover_At_A_Glance"	The Bootstrapping resource shall provide a short description with an illustration of the resource.	
BR_Discover_Share	The Bootstrapping resource URL should be shared quickly with anyone by e-mail	Applicable to the platform
BR_Discover_Tutorials	The Bootstrapping resource shall provide links to video tutorials.	
BR_Discover_Technical_Manual	The Bootstrapping resource shall be provided with a technical user's manual. Including especially the recommended environment and tools.	
BR_ Discover_Example_of_use	The Bootstrapping resource shall be provided with concrete usage examples (e.g. with Jupyter notebooks).	

BR_ Discover_Required_skill	The Bootstrapping service shall indicate specific required skill for service usage.	
BR_ Discover_Result_viewer	The Bootstrapping service should provide input and results viewer integrated to the platform.	Without the necessity to download data to visualize them.
BR_ Discover_Version_Info	The Bootstrapping service shall provide information on the version (novelties, known limitation…)	
BR_ Discover_Usage_Domain	The Bootstrapping service shall provide performance level and usage domain. In particular, for AI service, indicate the dataset used for the training and validation.	Usage domain of a provided resource is a key.
BR_ Discover_Technical_Prerequisite	The Bootstrapping Resource shall advise on technical pre-requisite such as resource sizing for execution (CPU, GPU…)	This could be a real pain in using the resource.
BR_Discover_Favorite	The platform should propose to add the bootstrapping resource as a favourite in the user's account.	To ensure UX
BR_Discover_Subscription	The platform should allow to subscribe to news and update of Bootstrapping Resources.	To ensure platform vitality
BR_Discover_Contacts	The Bootstrapping resource shall indicate contacts for technical support.	
BR_Discover_User's_evaluation	The platform shall allow users to leave user's feedback on Bootstrapping Resources.	To ensure an improvement loop, and direct link between users and developers. It could also a way to collect feedback from users in the frame of WP2 evaluation task.

Task - Buy the Bootstrapping Resource

BR_Price	The Bootstrapping resource should be free of charge, or propose a pricing adapted to user's profile.	

Task – Use the Bootstrapping resource

BR_Specific_format	The Bootstrapping resource should avoid specific data format and favour standard format. In case of specific data use, the associated converter shall be provided.	Otherwise, it is time consuming.
BR_Access_To_Data	The Bootstrapping resource should be provided with results data access without downloading them and a layer to visualize data.	To maintain the user's confidence in data

BR_Dockerized_Service	When the Bootstrapping resource is an application, it shall be provided as a Docker.	To ease its usage.
BR_Simplified_Version	The Bootstrapping resource should be proposed into different versions, one simplified, adapted to a non-specialist user.	To target large audience.
BR_Infrastructure_Transparent	The bootstrapping resource use should not require resource allocation expertise (allocation of CPU, or GPU…)..	This should be totally transparent for the user

5.2 Technical Requirements

These requirements stem from the analysis of the proposals submitted in the project's Open Calls and refer to the assets and facilities developers would most critically need for developing AI/EO applications.

5.2.1 Functional Requirements

Table 14 : Functional requirements

ID	Requirement description	Type
FR01	Provide access to image analysis technologies	Software
FR02	Provide imaging information in different bands (pre-separated)	Data
FR03	Provide access to trusted soil data resources	Data
FR04	Provide access to climatic data resources	Data
FR05	Provide access to weather data in different intervals	Data
FR06	Provide access to socioeconomic data	Data
FR07	Provide algorithms for data downscaling	Algorithms/Software
FR08	Provide methods for entity recognition (generic)	Algorithms/Software
FR09	Provide methods for entity recognition (agricultural settings)	Algorithms/Software
FR10	Provide methods for entity recognition (urban settings)	Algorithms/Software
FR11	Provide methods for entity recognition (maritime settings)	Algorithms/Software
FR12	Provide implementations of optimization methods	Algorithms/Software
FR13	Provide pre-trained prediction models for testing	Algorithms/Software
FR14	Provide methods and tools for explaining AI models and results	Algorithms/Software
FR15	Provide methods and tools for IoT data management	Algorithms/Software

FR16	Provide methods and tools for timeseries analysis	Algorithms/Software
FR17	Provide event/change detection tools	Algorithms/Software
FR18	Provide libraries for data visualization	Software
FR19	Provide tools and frameworks for metadata management	Software
FR20	Provide tools for FAIR data publishing and standardization	Software

5.2.2 Non-functional Requirements

Table 15 :Non-functional requirements

ID	Requirement description	Type
NR01	Provide tools and guidance for application deployment	Systems
NR02	Provide CI/CD frameworks applicable to the EO domain	Systems

6 Technical Requirements Coverage Assessment

In the section, we firstly present currently available services from AI4Copernicus and the AI-on-demand Platform that adhere to the set requirements. Subsequently, we identify the gaps remaining for fulfilling all requirements.

6.1 Existing Services

In this subsection we present all available services from AI4Copernicus and its DIAS platforms, as well as relevant services from the AI-on-Demand Platform. The two platforms are the main resource distributors in the envisioned EO/AI ecosystem, thus their exploration highlights the existing resources useful for developing bootstrapping and thematic services under the ecosystem.

Briefly, the AI-on-Demand Platform is a service-oriented web platform that offers a catalogue of reusable AI resources, which can be incorporated as modules of more complex AI applications in a variety of domains. Also, it is used to bring together a variety of stakeholder types, strengthening this way cross-sector collaboration and promoting the integration of AI in real-world business and industry processes.

Respectively, the DIAS platforms are sophisticated and robust large-scale platforms integrating huge amount of observation and satellite data in real time datasets using cloud-based data systems. These data are static, dynamic and multi-layered, making this integration difficult. In addition, the DIASes have enormous data storage capacity. Real-time data and archived data are stored in a suitable analysis environment for high-speed processing. Incorporating rich APIs, including metadata management systems, to ensure high-quality data archives and efficient data utilization DIAS has prepared various APIs for processing, archiving huge amounts of data quickly and efficiently, and visualising them.

The information collected for each service included its application domain, its required data inputs, the AI discipline that it mainly falls under, and provenance information to ensure availability.

In the following table, we summarize the information for each service and we display its id, its name, a short description, its scope (whether it is a bootstrapping service, a thematic service, or a service of general purpose), and the ids of all relevant requirements. In Annex C we present detailed information about these services in order to support the "relevant requirements" column of these tables.

6.1.1 AI4Copernicus Services

Table 16 : AI4Copernicus services and their relevant requirements

ID	Service name	Service Description	Scope	Relevant requirements
A01	Sentinel-1 GRD pre-processing	This pipeline processes a S1 GRD product in native format to generate a terrain corrected image representing the calibrated backscatter in GeoTiff format.	Bootstrapping	FR01, FR02
A02	Sentinel-1 SLC pre-processing	This pipeline processes a S1 SLC product in native format to generate a terrain corrected image representing the calibrated backscatter in GeoTiff format.	Bootstrapping	FR01, FR02
A03	Sentinel-2 pre-processing	This pipeline processes a S2 product in native format to generate a product with a common resolution for all the bands in GeoTiff format. The process allows to apply a land/sea mask and a cloud mask in order to have an output product ready for analysis.	Bootstrapping	FR01, FR02
A04	Sentinel-1 Change detection– Amplitude Change Detection and Multi-temporal Coherence	This pipeline processes pairs of S1 SLC products in native format to generate a series of products to assess the changes between both images. These products include: - the coherence (the amplitude of correlation between the images), - the ACD (Amplitude Change Detection), which is a RGB composite of the backscatter of the input images	Bootstrapping	FR01, FR02, FR17

		- the MTC (Multi-Temporal Coherence), which is a RGB composite of the backscatters and the coherence - binary mask of changes		
A05	Sentinel-2 Change Detection	This pipeline computes (and classifies) the changes using as input a pair of S2-L2A products by using the Change Vector Analysis approach.	Bootstrapping	FR01, FR02, FR17
A06	Vector data of human features	SatCen has pre-processed and ingested several OSM data layers and can provide the data as a service in the scope of the project.	Bootstrapping	
A07	Deep network for pixel-level classification of S2 patches	This service provides functionality for users to train a custom pixel-level classifier of Sentinel 2 patches. For example users can train a classifier for crop types (corn, sunflower, wheat, etc), land cover (urban vs. natural, water vs land), road extraction (road vs other).	Bootstrapping	FR08-FR011
A08	TimeSen2Crop	TimeSen2Crop is a pixel-based dataset made up of more than 1 million crop type samples of Sentinel-2 time series. The dataset includes atmospherically corrected images and reports the snow, shadows, and clouds information per labelled unit, as well as the spectral signature of the samples of nine Sentinel-2 spectral bands at 10m of spatial resolution.	Bootstrapping	FR02, FR08, FR09
A09	Harmonization of pre-processed Time Series of Sentinel-2 data	The harmonization of pre-processed time series of Sentinel-2 data considers a statistic-based approach that computes the median for each pixel in the different images acquired in a particular month. The pixel composite approach to mosaic generation provides consistent results at large scale, allowing the processing of harmonized acquisitions.	Bootstrapping	FR01, FR02
A10	Long Short-Term Memory Neural Network for Sentinel-2	The Long Short Term-Memory architecture can be trained using samples selected by the user. The service exploits the data given by the user to train from scratch an LSTM and stores the resulting weights. Several parameters are exposed to allow the user to custom the model.	Bootstrapping	FR12, FR13
A11	Pre-Trained Long Short-Term Memory	The pre-trained Long Short Term-Memory architecture is already trained using the TimeSen2Crop database and is available in .h5 format. The service exploits a pre-trained architecture to classify the specified tile	Bootstrapping	FR12, FR13

		harmonized using the monthly composite approach.		
A12	Energy datasets	Meteorological data: ERA5 Example: offshore wind farms are located (training data). JRC Open Power Plants Database (JRC-PPDB-OPEN). Open data from the floating offshore wind farm, Hywind Scotland.	Bootstrapping	FR04, FR05
A13	Probabilistic downscaling of CAMS air quality model data	This service generates high-resolution (currently ~ 10km) air quality maps from low-resolution (~40 - 80km) CAMS model (re)analysis and/or forecast output.	Bootstrapping	FR04, FR07
A14	soCAP	Delivers detailed satellite-derived crop analytics (NDVI, FCOVER, LAI)	Bootstrapping	FR01, FR02, FR07
A15	MUSCATE	Offers reliable, ready-to-use data products for land monitoring	Bootstrapping	FR01, FR02
A16	Coastal TEP	Automatic monitoring and early warning system for pollution discharges, harmful algal blooms and storm surges	Bootstrapping	FR09, FR17
A17	Forestry TEP	Provides simple value-added products such as forest maps and vegetation indices	Bootstrapping	FR02, FR09
A18	Geohazards TEP	Give access to innovative processing chains for earthquakes, volcanoes, landslides and subsidences monitoring and prevention	Bootstrapping	FR17
A20	Polar TEP	Modeling and monitoring of change in the polar regions	Bootstrapping	FR02, FR09, FR17
A21	Urban Green	Combines average surface temperature and proportion vegetation to identify areas with high revegetation potential and follow the impact	Bootstrapping	FR09, FR10
A22	Mundi Cloud Mask	Create cloud mask for Sentinel-2 data	Bootstrapping	FR01

A23	Grassland monitoring	Provide information about grassland coverage	Bootstrapping	FR02, FR09
A24	RECAP	Crop Identification Service	Bootstrapping	FR09
A25	I.MODI	Monitor the stability of buildings in large urban areas	Bootstrapping	FR10
B01	GeoTriples	GeoTriples is a tool for transforming geospatial data from their original formats into RDF.	General	FR19, FR20
B02	JedAI	The Force Behind Entity Resolution. Perform State Of The Art Entity Resolution With The Java Generic Data Integration Toolkit.	General	TR08-TR011
B03	Strabon	Strabon is a spatiotemporal RDF store.	General	FR19, FR20
B04	Semagrow	Semagrow is a SPARQL query federator of heterogeneous data sources.	General	FR19, FR20
B05	Sextant	A web-based and mobile ready platform for visualizing time-evolving linked geospatial data.	General	FR18-FR20

6.1.2 AI-on-demand Platform Services

Table 17 : AI-on-Demand platform relevant services and their relevant requirements

ID	Service name	Service Description	Scope	Relevant requirements
C01	AI for Visual Vehicles Counting	Monitoring vehicle flows in cities by counting cars from images acquired from smart cameras.	Thematic	FR10
C02	AI4Agri Knowledge Graph	A dataset from Earth Observation, Machine Learning models and vineyard data, in the form of a Knowledge Graph.	Thematic	
C03	AI4Agri pilot GUI	A docker container including the GUI of the AI4AGRI pilot, providing access to all the different components developed in that pilot.	Thematic	

C04	AI4Agriculture Counting Model	A computer vision model to automatically detect grape bunches in vineyard pictures.	Thematic	FR09
C05	AI4Agriculture Grape Dataset	A collection of 250 images taken in a vineyard in Ribera de Duero, annotated using bounding boxes, to train and validate object detection models.	Thematic	FR01, FR09
C06	Ai4agricuture NDVI component	An algorithm to rectify NDVI measurements from sentinel 2 images.	Thematic	FR01
C07	AI4IoT Data Source	This asset was developed for the AI4IoT Pilot, with the goal of providing a tool to access raw data from a network low-cost sensors deployed in Trondheim, Norway.	Thematic	FR15
C08	AI4IoT Sensor Calibration	This asset was developed for the AI4IoT Pilot, with the goal of providing a tool to calibrate the measurements of low-cost sensors. The model is pre-trained for the low-cost sensor deployed in Trondheim, Norway.	Thematic	FR15
C09	AI4IoT/PhysicalAI pollution dataset	Trondheim data for pollutants PM10,PM2.5,NO2 from 4 urban stations Jan 2014-June 2019	Thematic	FR04, FR05
C10	GTA Dataset	Grand Traffic Auto Dataset	Thematic	FR10
C11	Hub'eau Hydrometry	Rest API to retrieve real-time French hydrometric data (one month history)	Thematic	FR04
C12	IoT data analysis model	Analysis of IoT data in the logistic industry	Thematic	TR15
C13	Portuguese Transmission System Aggregated Load Time Series and Encoded Time Covariates	Aggregated load time series of the Portuguese TSO (Transmission System Operator) for 2018 and 2019 (15 minute resolution) accompanied by encoded (in a cyclical manner) time covariates. Useful for TSO demand forecasting experimentation.	Thematic	FR13
C14	SUMO-RL	Traffic Control in a Simulated Environment for Pollution Reduction	Thematic	FR12
C15	Yield prediction	A docker container to get the quality and yield	Thematic	FR13

	& Quality Model	predictions for the AI4AEU agriculture pilot		
D01	PDDL Planners	Dockerized/ACUMOS Ready PDDL planners (ff, fd, popf, optic-clp)	General	FR12
D02	Pyperplan Unified Planning Interface	This asset provides the interfacing between the Pyperplan planner and the Unified Planning Framework developed by the AIPlan4EU project.	General	FR12
D03	TAMER Unified Planning Interface	This asset provides the interfacing between the TAMER temporal planner and the Unified Planning Framework developed by the AIPlan4EU project.	General	FR12
D04	Unified Planning Framework	The Unified Planning Framework (UPF) library makes it easy to formulate planning problems and to invoke automated planners.	General	FR12
E01	ABELE	Adversarial Black box Explainer generating Latent Exemplars	General	FR14
E02	Agresia	Agresia is a software metasystem with a graphical user interface and automated pipelines for the identification, labeling, and transformation of PII found in structured and unstructured data	General	FR19
E03	AIDDL	AI Domain Definition Language (AIDDL) and libraries for fast prototyping of integrated AI systems	General	FR08-FR17
E04	Altruist	Argumentative Explanations through Local Interpretations of Predictive Models	General	FR14
E05	CKRew - CKR datalog rewriter	Datalog rewriter for defeasible multi-relational Contextualized Knowledge Repositories in OWL RL	General	FR20
E06	Clingo Answer Set Solver	State-of-the-art solver for logic programming under the answer set semantics.	General	FR08-FR17
E07	Corese	Corese is a Semantic Web Factory that implements W3C standards	General	FR19, FR20
E08	Diversity Algorithms	While reinforcement learning algorithms converge towards a single policy, it may be useful to generate multiple policies instead of just one.	General	FR08-FR17
E09	DTW Mean	Matlab library for time series averaging and k-	General	FR16

		means clustering of time series with missing values.		
E10	EMLlib	A Python library to embed ML models in declarative optimization models	General	FR12
E11	Hexlite	A solver the HEX language which combines Logic Programming with Python-based plugins.	General	FR12, FR13
E12	ITE for SICStus Prolog	A lightweight approach for implementing constructive disjunction for in SICStus Prolog.	General	FR12, FR13
E13	Katie-Ann	Multi-purpose Artificial Neural Networks Software Library	General	FR08-FR17
E14	KENN-Docker	A python 2.7 library for Neural-Symbolic integration	General	TR08-TR17
E15	Linked Open Reasoning	S-LOR (Sensor-based Linked Open Rules)	General	FR12, FR13, FR15
E16	LioNets	A Neural-Specific Local Interpretation Technique Exploiting Penultimate Layer Information	General	FR14
E17	LionForests	Local Interpretation of Random Forests	General	FR14
E18	Locally Private Graph Neural Networks	Federated training of Graph Neural Networks (GNNs) with Local Differential Privacy	General	FR08-FR17
E19	Logic Tensor Networks	A neurosymbolic framework that supports querying, learning and reasoning with both rich data and rich abstract knowledge about the world	General	FR18-FR20
E20	LORE	LORE (LOcal Rule-based Explanations) is a model-agnostic explanator for tabular data	General	FR08-FR12, FR14
E21	Lyrics	The Lyrics framework injects logic knowledge into a learner using Semantic Based Regularization	General	FR08-FR17
E22	Memory-based Multi-Source Meta-Learning (M3L)	A framework to train a generalizable model for unseen domains.	General	FR08-FR17
E23	Meta-CSP Framework	A Java API for meta-constraint reasoning	General	FR12-FR17

E24	Moving Targets via AIDDL	A Python library that integrates machine learning with constraint optimization using the AI Domain Definition Language.	General	FR12
E25	NegDis	Tool for the discovery of temporal-logic patterns as a binary supervised learning problem	General	FR12-FR17
E26	ngames	A framework to bridge the norMAS and game theory communities.	General	
E27	pygrank	A Python library bringing together fast recommendation algorithms for large graphs based on link analysis.	General	FR12
E28	Relational Neural Machines	The Relational Neural Machines framework is a general neuro-symbolic methodology to inject logic knowledge into a machine learner.	General	FR12, FR13
E29	Semantic Middleware	The Semantic Middleware is a solution of middleware that acts as a connector mediating data exchanges between shop-floor data producers (publishers) and data consumers (subscribers), thus contributing to enhance the level of interoperability of the modern production resources distributed within the shop floor.	General	FR19, FR20
E30	Stable Baselines	Stable Baselines 3 provides open-source implementations of deep reinforcement learning (RL) algorithms in Python.	General	FR12, FR13
E31	SUNNY-CP	sunny-cp is a parallel portfolio solver for solving Constraint Satisfaction/Optimization Problems	General	FR12
E32	SWIRL2SPIN	A tool for transforming SWRL rule bases in OWL ontologies to object-oriented SPIN rules in SWI-Prolog	General	FR12-FR17
E33	The Devil is in the GAN: Defending Deep Generative Models Against Backdoor Attacks	Training time procedures to produce secret adversarial backdoors in Deep Generative Models	General	FR12, FR13
E34	VERIFAI	Verifiable and Explainable RIsk Forecasting	General	FR14

		Artificial Intelligence Framework		
E35	Whitening for Self-Supervised Representation Learning	A new loss function for self-supervised representation learning, which is based on the whitening of the latent space features.	General	FR12, FR13
E36	Zephyrus2	zephyrus2 is a tool that allows computing the optimal configuration of applications in a cloud	General	NR01

6.2 Gaps and Risks

The comparative analysis of the elicited technical requirements and the identified relevant services allows us to determine the main areas to which the currently provided solutions do not fully cover the needs of the community (at least as represented by the Open Call applicants).

In broad terms, the provided bootstrapping and general services cover most of the basic functional needs of the proposed applications. The prominent use of Sentinel and Copernicus data, and their high availability via the DIAS platforms constitute a solid baseline for building AI-empowered EO services with a relatively broad reach. Additionally, critical data sources like environmental and climate data are incorporated and offered via AI4Coperniucs. Soil data is one area where more sources must become available via the AIoD platform, as well as a richer collection of socioeconomic data.

In terms of data management technologies, AI4Copernicus and by extension the AIoD platform offer sufficient resources for managing, annotating and publishing data under the FAIR principles and thus allowing its further reuse and exploitation, and facilitating interoperability with other services. However, at the pre-processing field, a more extensive and coherent tool palette should be built for handling IoT data and timeseries.

In terms of AI technologies, there is a relatively extensive collection of basic, general-purpose solutions for typical AI problems. In addition, there is a variety of more refined solutions that are nevertheless built and tested for domain or case-specific problems. Their immediate applicability in adjunct problems is not certain; thus a more elaborate testbed for assessing the applicability of a model in other domains, and the ability to parameterize and finetune a model for repurposing it is a requirement that should be taken into account in the AIoD platform.

In the same fashion, visualization and explainability is two factors that are critical for the success of a service and thus require the provision of a more extensive toolset.

7 Conclusions and Next Steps

Deliverable D2.2 presents in detail the methodology followed and the produced outcomes for eliciting and concretizing a broad set of requirements for data and services that are useful and impactful for the AI/EO community. It has identified the main technical and operational needs as stemmed from the requirement elicitation process, and conducted a first analysis of the current landscape and its relevance to the goals of the community.

The analysis will naturally serve as the guide for further developments and offering in the context of AI4Copernicus. The identified gaps will be communicated, prioritized and assessed with the technical developments work packages, to better focus their work towards the provision of solution for critical community needs.

What's more, the report will be a first point of reference for better directing future developments within the broader AI-on-Demand ecosystem as it forms in the coming years, as it offers a technical and operational framework for the ways to build, deploy and distribute services in a way that maximizes their reach and impact in the economy and the society.

8 Annexes

8.1 Annex A : Example of thematic Services

8.1.1 Anomaly detection on critical infrastructures

[1] Scenario title: Anomaly detection on critical infrastructures	**Domain:** Security
Context: Challenge: Critical infrastructure is defined as systems and assets so vital that the disruption of the service provided would have a debilitating or catastrophic impact on security or safety of a nation or a group of nations. Examples of critical infrastructure are dams, water treatment facilities, oil fields, pipelines, pumping stations, airports, highways and governmental buildings. These infrastructures face certain risks associated with different threats, regardless of whether they are a result of accidents, natural events or international acts. The identification of critical infrastructure under threat enables crisis management officials and first responders to prepare for possible damages. However, this work requires a complex analysis of the influential factors, both spatial and non-spatial, which is currently performed with an intense human effort due to the lack of automatized means to detect and/or identify features of interest. Scope: The solution in response to this challenge shall be to provide innovative AI models to detect and/or identify automatically man-made infrastructures (e.g. roads, buildings, ports) and estimate if these infrastructures present anomalies (e.g. road blockages, infrastructure dismantling) or relevant changes, considering time series.	
Dataset sources: Sentinel-1 and/or Sentinel-2 data (mandatory), VHR data (asset), navigation/localization data, meteorological data Any other data can be used if relevant for the call scope. In this case, the data must be accessible (free of charge or not) so a possible operational use of the solution is guaranteed.	
Output and coverage: The expected output is a service (available through API) or a tool (dockerized) that can be executed in different AoIs to generate "anomaly maps" (as vector and/or raster). Note: AoI can be selected by the bidders but the methodology shall be applicable in other AoI. Although it is expected a higher performance in specific areas, a good performance in any area will be considered an asset.	

AI4Copernicus services: The following bootstrapping services will be made available: ● Pre-processing chains for Sentinel-1 and Sentinel-2 data products in order to obtain ready to use images for algorithms based on AI ● Processing pipelines for Sentinel-1 and Sentinel-2 data products (e.g. customizable S2 change detection pipeline based on CVA) ● OpenStreetMap-derived vector data (e.g. roads, buildings) ● Ground truth data (for validation, according to the area proposed and its availability).
Expected Impact (technical, social, economic, environmental, policy etc.): <u>Technical</u> ● Quantitative step forward in the use of AI technologies in EO-based applications for security scenarios; ● Interoperable solutions to be able to work in multi-disciplinary environments; ● Increased processing and analysis capacity for large volumes of Copernicus Earth Observation data; ● Develop operational tools for feature detection and anomaly detection; ● Increased performance and/or automation of image processing, in order to have pre-operational services to be used in GEOINT tasks. <u>Socio-Economic</u> ● Enlargement of the number of solution providers in security; ● Foster the use of Sentinel data and other data sources; ● Enhance the understanding of security threats for the general public, including EU citizens and international actors. <u>Policy</u> ● Foster the use of EO data analysis for policy and decision-making in EU domain (e.g. EU Common Foreign Security Policy); ● Contribute with the results to the different active initiatives in the EO domain, including the GEO Space and Security Community Activity and EuroGEO.

8.1.2 Exploring optimal locations for renewable energy infrastructures

[2] Scenario title: Exploring optimal locations for renewable energy infrastructures	**Domain:** Energy

Context:

Challenge:

Renewables made up 26.2 percent of global electricity generation in 2018, and it is expected to rise to 45 percent by 2040. Most of the increase will likely come from solar, wind, and hydropower. In this context, energy companies urgently need to explore optimal locations in terms of renewable energy resources in order to get to a more sustainable future.
Currently, geospatial analytics are used to find these locations. However, Earth Observation data can provide additional, complementary and valuable layers of information.
For instance:

- In the case of wind farms, wind parameters can be estimated from LiDAR or SAR data. These spot measurements can be used to validate and improve existing models, to increase their granularity. For example, radar measurements can be converted to wind speed by measuring wave height generated by the wind
- In the case of solar panels, the main obstacle to capturing the solar energy comes from cloud shadows (from the terrain) or aerosols (particles in the atmosphere). These can all be measured or modelled based on Earth Observation data and Digital Elevation Models (DEMs)
- In the case of hydropower plant, their potential can be estimated from DEMs

Scope:

The solution in response to this challenge shall be to provide methods based on geospatial analytics and AI models to explore the optimal locations in Europe for the various types of renewable energy infrastructure.

Dataset sources: Sentinel data, Commercial EO data, Concentration map of air pollutants, Shuttle Radar Topography Mission (DEM), Meteorological data and geo-localisation data.

Output and coverage:

The expected output will be a service (available through API and Dockers), which can be integrated as a layer in a GIS system for further analysis or directly into a web map. The data will be in an open geospatial format (e.g. the layers delivered in [Rapid Action on coronavirus and EO (esa.int)](esa.int))

AI4Copernicus Services:

The following bootstrapping services will be made available:

- Pre-processing tools for Sentinel-1 and Sentinel-2 data products, as well as optimized tiling techniques for reconstructing time series

● Air-quality five-day forecasts provided the Copernicus Programme, through its CAMS service ● ERA5 reanalysis data to extract local weather observations ● Ground truth data of onshore and offshore locations - https://www.emodnet-humanactivities.eu/ and SETIS - SET Plan Information System (europa.eu)
Expected Impact (technical, social, economic, environmental, policy etc.): Technical ● Develop AI techniques and products based on Earth Observation data to provide complementarity and value added to existing geospatial data layers Socio-economic ● Reducing the investment costs and operational costs related to renewable energy sources ● Better location of renewables (e.g. solar and wind farms) can reduce their intermittency thus leading to smaller costs to integrate renewables into the power system Environmental ● More efficient use of renewable energy resources (wind, sun, water, etc.) can increase their penetration reducing the reliance on fossil fuels Policy level ● Provide improved tools for policy level planning of energy supply of renewable energy sources ● Demonstrated adoption of the results of using AI in an area that is central in the policy of the European Commission

8.1.3 Pollution predictions

[3] Scenario title: Pollution predictions	**Domain:** Health and Environment
Context: Challenge: Air pollution affects world-wide populations with an estimate of 7 million premature deaths every year according to the World Health Organization, and close to half a million in Europe alone, according to the European Environmental Agency. Air of insufficient quality can not only be the direct cause of these deaths but it can also exacerbate existing diseases in an individual, and the range of diseases where it is thought that air pollution is a driver or a factor is ever increasing including lung cancer, asthma, heart problems, pulmonary issues or dementia. Reliable and timely information about air pollution, especially forecasts, can help in two complementary ways: informing citizens about their exposure to pollution can help them adapt their behaviour; informing decision makers can support them for taking emission mitigation decisions (such as banning or reducing traffic or certain activities). However, the current most reliable operational air quality forecasts provided by the Copernicus Atmosphere Monitoring Service (CAMS) have a spatial resolution of 10 km, which is similar to the resolution of meteorological forecasts. This is not always sufficient, in particular for big cities with heterogeneous air quality and there is a need for downscaling information to finer spatial resolutions. Scope: The solution in response to this challenge shall be to provide innovative machine learning models to predict the air quality at sub-city district level using various data sources that provide global and local information relevant to pollution. Solutions can be global or local for a specific area or large city or conurbation.	
Dataset sources: Concentration map of air pollutants, Sentinel data, land use and topography data, meteorological data and local monitoring data (such as local pollution ground measurements and traffic information). Some of the datasets can be accessed here: ● API interface for CAMS and C3S datasets from http://cds.climate.copernicus.eu (C3S products including ERA5, seasonal forecasts) and http://ads.atmosphere.copernicus.eu (CAMS products including reanalyses, analyses and forecasts of air quality in Europe and worldwide). ● For satellite data (mostly from S-3 and S-5P), this can be used through WEkEO or one of the other DIASes. ● CORINE Land cover product provided the Copernicus Land Monitoring Service over Europe, as well as global land cover products.	

Output and coverage:

The expected output would be a containerized service accessible through standard APIs, that can be hosted on different hardware and execution environments. This service would be able to be plugged into existing workflows and to generate improved estimates of pollution forecasts on local and peripheral level. The estimates should be re-trainable and adjustable for new data inputs and datasets that will become available in the future.

AI4Copernicus Services:

As part of AI4Copernicus, bootstrapping services will be made available to support the challenge and the use of machine learning tools in the context of Copernicus data and the DIASes.

Expected Impact (technical, social, economic, environmental, policy etc.):

Technical

- Development of deep learning techniques to merge large scale prediction information with local information on a finer (potentially unstructured) grid.
- Development of machine learning techniques that take multi-scale behaviour of atmospheric dynamics into account in both space and time, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks.
- Support scientific researches with vanilla solution for machine learning down-scaling applications on the DIASes.

Social

- Improve awareness of exposure of society to pollution and the related impacts on health and well-being.
- Warn individual citizens with personalised risks estimated according to their location and activities.
- Highlight best practices in the world to tackle air pollution.

Economical

- Stimulate tourism in small cities or the countryside with better air quality.
- Support development of real estate in cleaner areas.

Environmental

- Identify critical areas in a region.

Policy

- Advice policy makers on pollution related questions regarding city developments and traffic regulations.
- Advice policy makers on health impact of various sources of pollution.
- Advice policy makers on urban and suburban design plans.

8.1.4 Crop type mapping

[4] Scenario title: Crop type mapping	**Domain:** Agriculture
Context: Challenge: Mapping crop type allows national and multinational agricultural agencies to make inventories in order to predict yields, collect crop production statistics, facilitate crop rotation records, map soil productivity, etc… However, this information is generally collected from census and ground surveys that generate multiple databases with different information extraction strategies which are not easy to integrate.. In this context, remote sensing offers an efficient and reliable means for collecting the information required. Moreover, due to the increase of the number of Earth Observation data available, crop identification can benefit from the use of multi-temporal and multi-sensor data. Indeed, certain crops are easier to discriminate at some specific periods, and different sensors can extract complementary information. For instance, optical data provide information on the chlorophyll content of the plants and radar provides information relating to plant structure and moisture. **Scope:** The solution in response to this challenge shall be to provide innovative AI models to extract crop boundaries and/or identify crop type at country or continental level from time series of satellite images. This can be achieved by leveraging the promising activities developed in this domain in the past two years.	
Data sources: Sentinel-2 and Sentinel-1 data, Landsat data, HR-VPP data from VITO, Land Cover, Meteorological data, Administrative Divisions dataset (GADM). Any other data can be used if relevant for the call scope. In this case, the data must be accessible (free of charge or not) so a possible operational use of the solution is guaranteed.	
Output and coverage: The expected output is a service (available through API) or a tool (dockerized) that can be executed in different AoIs to generate "crop type maps" and/or "crop boundary maps" (as vector and/or raster). Note: AoI can be selected by the bidders but the methodology shall be applicable in other AoIs. Although it is expected to have higher performance in specific areas, a good generalization capability to other areas will be considered an asset.	
AI4Copernicus Services: The following bootstrapping services will be made available: ● Standard pre-processing chains for time-series of Sentinel-2 data products	

● Gap filling techniques for reconstructing time series taking into account cloud cover and irregular acquisitions in different areas. ● Classification tool based on Long-Short Term Memory (LSTM) deep neural network from time-series Sentinel-2 data ● Large dataset with crop-type labelled samples coming from H2020 ExtremeEarth project (useful to train deep learning models)
Expected Impact (technical, social, economic, environmental, policy etc.): Technical ● Develop AI techniques to map crops and their boundaries with Sentinel-2 and/or Sentinel 1image times series to significantly enhance the accuracy of present systems. ● Generate accurate products based on the trend of vegetation, water and fertilization indices at crop level (considering different use of the same field in the same agronomic season) and large scale. Socio-economic ● An improved capability of extracting semantic information for supporting mapping and management of agricultural areas at large scale. ● Improving crops monitoring and management has huge effects on food security. Environmental ● Proper management of crops can have a dramatic impact on the environment (e.g. in terms of rationalization of the use of chemical additives or pesticides). ● Management of resources (e.g. water for irrigation) is crucial for the environment. Policy ● Demonstrated adoption of the results of using AI in an area that is central in the policy of the European Commission and has not benefit yet of the huge potential of AI. ● Contribute with the results to the different active initiatives in the use of EO for agriculture, enhancing the extraction of the semantic information from data with customized machine learning technologies.

8.1.5 Maritime situational awareness identification

[5] Scenario title: Maritime situational awareness identification	**Domain:** Security
Context: Challenge: Current maritime navigation, surveillance, and communications systems, combined with today's network technologies, create enormous amounts of data collected providing information of vessels operating worldwide. This information is crucial to enhance maritime situational awareness, to plan safe and rescue operations for coast guards and to detect illegal activities. However, the available data come from different domains and data providers (e.g. AIS, Copernicus, Search and Rescue, coast cameras), what makes the fusion of different data sources quite complicated. Thus, the sources are often used separately. Moreover, the vast amounts of data cause that one specific event of interest is often lost in the regular traffic patterns. All these problems result in difficult, delayed or impaired decision-making in front of a real scenario. Scope: The solution in response to this challenge shall provide innovative AI models to extract relevant information to support safe and secure activity in maritime domain (e.g. vessel direction, vessel speed, route anomalies), using one or more data sources.	
Dataset sources: Sentinel-1 and/or Sentinel-2 data (mandatory), VHR images, AIS data (asset), meteorological data. Any other data can be used if relevant for the call scope. In this case, the data must be accessible (free of charge or not) so a possible operational use of the solution is guaranteed.	
Output and coverage: The expected output is a service (available through API and Dockers), which can include models (e.g. machine learning tools) and generate images (as vectors and raster) and databases (e.g. csv files with coordinates). AoI can be selected by the bidders but the methodology shall be applicable to other AoI.	
AI4Copernicus services: The following bootstrapping services will be made available: ● Pre-processing chains for Sentinel-1 and Sentinel-2 data products in order to obtain ready to use images for algorithms based on AI	

● Processing pipelines for Sentinel-1 and Sentinel-2 data products (e.g. change detection pipeline)
Expected Impact (technical, social, economic, environmental, policy etc.): Technical ● Foster the use of new data sources relevant for maritime awareness and the identification of data or technology gaps to be filled; ● Quantitative step forward in the use of AI technologies in EO-based applications for security scenarios; ● Interoperable solutions to be able to work in multi-disciplinary environments; ● Increased processing and analysis capacity for large volumes of Copernicus Earth Observation data; ● Develop operational tools for increase the capacity in the maritime awareness domain; ● Increased performance and/or automation of image processing, in order to have pre-operational services to be used in GEOINT tasks. Socio-Economic ● Enlargement of the number of solution providers in security; ● Foster the use of Sentinel data and other data sources; ● Enhance the understanding of security threats for the general public, including EU citizens and international actors. Policy ● Foster the use of EO data analysis for policy and decision-making in EU domain (e.g. EU Common Foreign Security Policy); ● Contribute with the results to the different active initiatives in the EO domain, including the GEO Space and Security Community Activity and EuroGEO.

8.1.6 Greenhouse gases emissions

[6] Scenario title: Greenhouse gases emissions	**Domain:** Health and Environment

Context:

Challenge:

As part of the Paris Agreement under the auspices of the United Nations' Framework Convention on Climate Change (UNFCCC), the governments from all over the world agreed to establish political measures in order to reduce the accumulation of greenhouse gases in the atmosphere in order to limit global warming to +1.5 to +2 degrees compared to the pre-industrial period. In particular, the Green Deal aims to transform Europe into the first carbon-neutral continent in the world by 2050.

Estimating greenhouse gases emissions is essential to monitor progress towards this ambitious target and more generally of the efforts of all the countries in the world. However, the current state of the art to estimate these emissions generally relies on accounting of emission sources which are often uncertain. As part of Copernicus, a new element of the Copernicus Atmosphere Monitoring Service (CAMS) is being developed to enable monitoring of anthropogenic emissions of CO_2. It will rely on new satellite observations as well as on very advanced processing techniques, which are required for disentangling human emissions from the other sources and sinks in the carbon cycle

Scope:

The solution in response to this challenge shall be to provide innovative machine learning models to support the estimation of greenhouse gases concentrations and emissions by combining Earth observation data combined and other types of information and building on the information in particular available from CAMS. Solutions should focus on specific areas or large cities and aiming at refining the information that is already available from Copernicus.

Dataset sources: Atmospheric concentrations and emissions inventories of greenhouse gases, Sentinel data, land use and topography, meteorological data and local monitoring data (such as local greenhouse gases ground measurements, traffic information, factory localisation).

Some of the datasets can be accessed here:

- API interface for CAMS and C3S datasets from http://cds.climate.copernicus.eu (C3S products including ERA5, seasonal forecasts) and http://ads.atmosphere.copernicus.eu (CAMS products including emissions inventories of greenhouse gases, as well as reanalyses, analyses and forecasts of CO2 and CH4 worldwide).
- For satellite data (methane from S-5P), this can be used through WEkEO or one of the other DIASes. Other satellite data such as from OCO-2 (NASA) or GOSAT (JAXA) may also be used.

● CORINE Land cover product provided the Copernicus Land Monitoring Service over Europe, as well as global land cover products. ● European Forest Fire Information System (EFFIS).
Output and coverage: The expected output would be a containerized service accessible through standard APIs, that can be hosted on different hardware and execution environments. This service would be able to be plugged into existing workflows and to generate improved estimates of greenhouse gas concentration and emission. The estimates should be re-trainable and adjustable for new data inputs and datasets that will become available in the future.
AI4Copernicus Services: As part of AI4Copernicu,s bootstrapping services will be made available to support the challenge and the use of machine learning tools in the context of Copernicus data and the DIASes.
Expected Impact (technical, social, economic, environmental, policy etc.): Technical ● Development of deep learning techniques to fuse datasets to enhance local information. ● Development of machine learning techniques that take multi-scale behaviour of atmospheric dynamics into account in both space and time, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) features) networks. ● Support scientific researches with vanilla solution for machine learning down-scaling and data fusion applications on the DIASes. Social ● Improve knowledge about greenhouse gas emissions that are driving anthropogenic climate change. ● Inform individual citizens and policy makers and allow them to attribute greenhouse gas emissions and sinks to local features. Economical ● Identify sources and sinks of greenhouse gases to provide additional information for emission trading. Environmental ● Identify sources of greenhouse gases within a region. Policy ● Advice policy makers on local greenhouse gas emissions. Monitor the impact of private companies on greenhouse gases emissions

- Monitor progress towards regional or national-level targets (such as Nationally Determined Contributions) and contribution to the stocktake exercises of the Paris Agreement.

8.1.7 Precision maintenance support on energy infrastructures

[7] Scenario title: Precision maintenance support on energy infrastructures	**Domain:** Energy

Context:

Challenge:

An efficient and reliable energy infrastructure is vital to society. With pipeline and electrical corridors stretching across vast expanses of land, city-sized solar farms and supply chains that demand ever-increasing security, there is a strong energy industry demand for monitoring their infrastructure. Traditional methods are often based on expensive, systematic and labour-intensive human inspection.

Earth Observation can provide an innovative and cost-effective way to support asset management (e.g. information about when to clean solar panels, cutting vegetation that disturb electrical installations) and risk management (e.g. natural hazards such as fire, flooding and land subsidence). Accidents on energy infrastructures (e.g. oil and gas pipeline) tend to be high-impact events that affect public safety, environmental protection and entire energy markets, so it is crucial to quickly detect potential risks before they have an impact.

Scope:

The solution in response to this challenge shall be to provide innovative AI models to detect anomalies on energy infrastructures, which require human intervention.

Dataset sources: Sentinel-2 and Sentinel-1 data, Commercial EO data, Land use and Meteorological data

Output and coverage:

The expected output will be a service (available through API and Dockers), which can be integrated as a layer in a GIS system for further analysis or directly into a web map. The data will be in an open geospatial format (e.g. the layers delivered in Rapid Action on coronavirus and EO (esa.int))

AI4Copernicus Services:

The following bootstrapping services will be made available:

- Pre-processing tools for Sentinel-1 and Sentinel-2 data products, as well as optimized tiling techniques for reconstructing time series
- ERA5 reanalysis data to extract local weather observations
- CORINE Land cover product provided the Copernicus Land Monitoring Service

- Ground truth data of onshore and offshore locations - https://www.emodnet-humanactivities.eu/ and SETIS - SET Plan Information System (europa.eu)

Expected Impact (technical, social, economic, environmental, policy etc.):

Technical

- Develop AI techniques and products based on Earth Observation data to detect anomalies on energy infrastructures, which require human intervention

Socio-economic

- Reduction of operational costs (e.g. labour cost, material cost) related to energy infrastructures
- Improved energy supply security through improving the management of energy infrastructures
- Reduced risks of accidents related to energy infrastructure increasing public safety and fluctuation on energy markets.

Environmental

- Reduced risks of accidents related to energy infrastructure decreasing possible negative impact on the environment

Policy level

- Demonstrated adoption of the results of using AI in an area that is important in the policy of the European Commission

8.1.8 Crop dynamic monitoring

<table>
<tr><td>[8] Scenario title: Crop dynamic monitoring</td><td>Domain: Agriculture</td></tr>
<tr><td colspan="2">Context:
<u>Challenge:</u>
The main source of food for the population of the world is agriculture. Today, the latter faces two major challenges, the demographic explosion and climate change. That is why intensive farming practices which are thought to be sustainable have been developed to slow the deterioration of agricultural land and even regenerate soil health.
In this context, there is a strong requirement for monitoring crop growth in near-real time. This information is valuable for a combination of purposes: react to extreme climatic events, anticipate food production shortages to ensure food security in the most vulnerable regions, identify factors influencing crop stress, highlight the best practices in terms of sustainability, etc…

<u>Scope:</u>
The solution in response to this challenge shall be to provide innovative AI models to extract phonological parameters of vegetation from complete time-series satellite images in order to monitor the full cycle at field level.</td></tr>
<tr><td colspan="2">Data sources: Sentinel-2 and/or Sentinel 1 data, other satellite data, in situ measurements for calibration, HR-VPP data from VITO, Land Cover, Meteorological data, Administrative Divisions dataset (GADM). Any other data can be used if relevant for the call scope. In this case, the data must be accessible (free of charge or not) so a possible operational use of the solution is guaranteed.</td></tr>
<tr><td colspan="2">Output and coverage:
The expected output is a service (available through API) or a tool (dockerized) that can be executed in different AoIs to generate maps that present one or more parameters related to crops (e.g. phenological parameters, fertilization parameters, crop stress parameters, etc.) (as vector and/or raster).
Note: AoI can be selected by the bidders but the methodology shall be applicable in other AoIs. Although it is expected to have higher performance in specific areas, a good generalization capability to other areas will be considered an asset.</td></tr>
<tr><td colspan="2">AI4Copernicus Services:
The following bootstrapping services will be made available:
Standard pre-processing chains for time-series of Sentinel-2 data products</td></tr>
</table>

- Gap filling techniques for reconstructing time series taking into account cloud cover and irregular acquisitions in different areas.
- Classification tool based on Long-Short Term Memory (LSTM) deep neural network from time-series Sentinel-2 data that can used to have an initial identification of crop type and field boundaries

Expected Impact (technical, social, economic, environmental, policy etc.):

Technical

- Develop AI techniques to monitor crops growth and phonological parameters in near-real time to significantly enhance the accuracy of present systems.
- Generate accurate products based on the trend of vegetation, water and fertilization indices at crop level (considering different use of the same field in the same agronomic season) and large scale.

Socio-economic

- Information regularly extracted and comparisons among the productivities of different fields enable the definition of best practice for improving productivity.
- Improving crops monitoring and management has huge effects on food security and provides information for company insurance
- Identification of the different factors influencing crop stress to improve productivity.

Environmental

- Proper monitoring and management of crops can have a dramatic impact on the environment (e.g. in terms of rationalization of the use of chemical additives or pesticides).
- Providing best practices in terms of sustainability

Policy

- Demonstrated adoption of the results of using AI in an area that is central in the policy of the European Commission and has not benefited yet from the huge potential of AI.

8.1.9 Disease spread forecasting

[9] Scenario title: Disease spread forecasting	**Domain:** Health and Environment
Context: Challenge: Air pollution and particular weather conditions can play a role in the spread of some diseases, such as Malaria, the Flu, or Covid both directly and indirectly by affecting human behaviour (indoor vs outdoor, etc..). Forecasting the probability of occurrence of these diseases can help the management of outbreaks and health service resources. However, conventional approaches used to predict these probabilities do not allow to "learn" the behaviour of these diseases as they appear and evolve in time. Scope: The solution in response to this challenge shall be to provide innovative machine learning models to forecast risk for development and spread of certain diseases using pollution and weather past (analyses, reanalyses) and forecasts data.	
Dataset sources: Concentration map of air pollutants, Sentinel data, land use and topography, meteorological data and local monitoring data (such as local pollution, ground measurements, local weather observations, health information). Some of the datasets can be accessed here: ● API interface for CAMS and C3S datasets from http://cds.climate.copernicus.eu (C3S products including ERA5, seasonal forecasts) and http://ads.atmosphere.copernicus.eu (CAMS products including analyses and forecasts of air quality in Europe and worldwide). ● For satellite data (mostly from S-3 and S-5P), this can be used through WEkEO or one of the other DIASes. ● CORINE Land cover product provided the Copernicus Land Monitoring Service over Europe, as well as global land cover products.	
Output and coverage: The expected output would be a containerized service accessible through standard APIs, that can be hosted on different hardware and execution environments. This service would be able to be plugged into existing workflows and to generate improved estimates of disease spreads based on pollution and weather (climatological and forecast) data. The estimates should be re-trainable and adjustable for new data inputs and datasets that will become available in the future.	

AI4Copernicus Services: As part of AI4Copernicus, bootstrapping services will be made available to support the challenge and the use of machine learning tools in the context of Copernicus data and the DIASes.
Expected Impact (technical, social, economic, environmental, policy etc.): Technical ● Development of deep learning techniques to fuse datasets from environmental science and health applications. ● Support scientific researchers with vanilla solutions for machine learning for health applications. Social ● Improve knowledge of the impact of environmental data on the spread of diseases. ● Inform individual citizens and policy makers about areas at risk for the spreading of diseases. Economical ● Allow to respond to risk alerts for specific diseases and to adapt travel plans, for example for tourism. Policy ● Help policy makers in the management and prevention of outbreaks of diseases and to allow for the timely distribution of medical resources.

8.1.10 Climate security understanding

[10] Scenario title: Climate security understanding	**Domain:** Security
Context: Challenge: The society continuously faces security threats. In the last decade, these threads have increased and proceed from causes not considered a priority in the past. For instance, changes in climate patterns can introduce major risks in some regions, extreme weather events can lead to forced population displacements, natural or man-made hazards can result in air pollution, illegal activities can damage ecosystems, uncontrolled fires can endanger citizens and damage ecosystems, unforeseen measures can bring about socio-economic hardships for citizens. Documenting and understanding Earth's changing climate allows to support decision making for the security of our planet. As climate security is a quite new topic, AI tools are not completely exploited in this domain. These model analyses are even more important as climate security is a cross-domains, that regroups different expertise. Scope: The solutions in response to this call shall be able to extract relevant information that allow to address scenarios in which climate change is affecting safety and security of people (e.g. forced migration caused by environmental hazards, food insecurity due to climate variations in vulnerable areas, scarcity of resources due to climate changes (e.g. water)).	
Dataset sources: Sentinel data (mandatory), meteorological data (asset), statistical data -e.g. population density, economic indicators or number of conflicts- (asset), other remote sensing data (e.g. Landsat), navigation/localisation data. It is important to use at least two different data sources. Any other data can be used if relevant for the call scope. In this case, the data must be accessible (free of charge or not) so a possible operational use of the solution is guaranteed	
Output and coverage: ● Services (available through API and Dockers) ● Datacube (e.g. from heterogeneous datasets) ● Models (e.g. statistical relationship between different variables) AoI shall be selected outside EU territory.	
AI4Copernicus services: The following bootstrapping services will be made available:	

- Pre-processing chains for Sentinel-1 and Sentinel-2 data products in order to obtain ready to use images for algorithms based on AI   - Processing pipelines for Sentinel-1 and Sentinel-2 data products (e.g. change detection pipeline)   - OpenStreetMap-derived vector data (e.g. line of communications, buildings)
Expected Impact (technical, social, economic, environmental, policy etc.): Technical   - Quantitative step forward in the use of AI technologies in EO-based applications for security scenarios;   - Interoperable solutions to be able to work in multi-disciplinary environments;   - Increased capacity of processing and analysing large volumes of Copernicus Earth Observation data;   - Develop statistical tool to infer about the relationship between climate and security;   - Create and/or enlarge existing data-cubes with heterogeneous data for climate and security (e.g. EO, statistical, meteo);   - Increased performance and/or automation of image processing, in order to have pre-operational services to be used in GEOINT tasks.   Socio-Economic   - Enlargement of the number of solution providers in security;   - Foster the use of Sentinel data and other data sources;   - Enhance the understanding of security threats for the general public, including EU citizens and international actors.   Policy   - Foster the use of EO data analysis for policy and decision-making in global (e.g. UN Sustainable Development Goals) and EU domain (e.g. EU Common Foreign Security Policy, EU Green Deal);   - Contribute with the results to the different active initiatives in the EO domain, including the GEO Space and Security Community Activity and EuroGEO.

8.1.11 Energy consumption understanding

[11] Scenario title: Energy consumption understanding	**Domain:** Energy
Context: Challenge: Urban areas presently consume around 75% of global primary energy supply, which is expected to significantly increase in the future due to urban growth. Understanding urban energy consumption patterns may help to address the challenges to urban sustainability and energy security. However, urban energy analyses are severely limited by the lack of urban energy data. To overcome the scarcity of urban-level energy data, Earth Observation data can be a first answer to monitor and quantify urban energy utilization patterns. For instance, High-resolution satellite images are used to identify and classify various urban settlement types based on their physical and textural analysis. The resulting classification can be correlated with energy consumption in order to generate statistics. Another means to extract patterns and estimate energy consumption at suburb-level is to use night-time images, which capture light emission from human settlements only . Scope: The solution in response to this challenge shall be to provide innovative AI models to extract energy consumption patterns in order to better understand energy distribution.	
Dataset sources: Sentinel-2 and Sentinel-1 data, Commercial EO data, Land use, Meteorological data and local energy consumption data	
Output and coverage: The expected output will be a service (available through API and Dockers), which can be integrated as a layer in a GIS system for further analysis or directly into a web map. The data will be in an open geospatial format (e.g. the layers delivered in Rapid Action on coronavirus and EO (esa.int))	
AI4Copernicus Services: The following bootstrapping services will be made available: ● Pre-processing tools for Sentinel-1 and Sentinel-2 data products, as well as optimized tiling techniques for reconstructing time series ● CORINE Land cover product provided the Copernicus Land Monitoring Service	

Expected Impact (technical, social, economic, environmental, policy etc.):

Technical

- Develop AI techniques and products based on Earth observation data to estimate energy consumption at suburb-level

Socio-economic

- Improved forecasting of consumption at the settlement level
- Increased efficiency of demand side policies through better information energy consumption of settlements
- Improved forecasting of the demand at the settlement level can decrease the integration cost of renewables into the electricity system

Environmental

- Reduced integration costs of renewables into the electricity system will increase their competitiveness compared to fossil fuels and reduce CO_2 emissions

Policy level

- Improved policy decision as a result of better monitoring and quantification of urban energy utilization patterns
- Demonstrated adoption of the results of using AI in an area that is important in the policy of the European Commission

8.1.12 Support irrigation management

[12] Scenario title: Support irrigation management	**Domain:** Agriculture

Context:

Challenge:

While most water services are allocated, regulated and measured, it is not the case for the irrigation sector. However, with the expansion of irrigation systems and the increasing trend towards more frequent and severe droughts, there is a real need to know the amount of water consumed and whether it is used efficiently.

Monitoring the amount of water in agricultural soils allows not only to increase water savings but also to get better crop production. Currently, this information is generally extracted from flow meters and soil moisture probes, but with the impressive proportion of land dedicated to agriculture, it is not possible to install these systems everywhere. In this context, remote sensing offers an efficient and reliable means of collecting data, above all soil moisture and plant evapotranspiration.

Scope:

The solution in response to this challenge shall be to provide innovative AI models to support irrigation management from Earth Observation data.

Data sources: Sentinel-1 and Sentinel-2 data, other satellite data, in situ measurements for calibration, Land Cover, Meteorological data, Administrative Divisions dataset (GADM). Any other data can be used if relevant for the call scope. In this case, the data must be accessible (free of charge or not) so a possible operational use of the solution is guaranteed.

Output and coverage:

The expected output is a service (available through API) or a tool (dockerized) that can be executed in different AoIs to generate maps presenting variables related to irrigation(e.g., soil moisture, evapotranspiration, etc.) (as vector and/or raster).

Note: AoI can be selected by the bidders but the methodology shall be applicable in other AoIs. Although it is expected to have higher performance in specific areas, a good generalization capability to other areas will be considered an asset.

AI4Copernicus Services:

The following bootstrapping services will be made available:

- Standard pre-processing chains for time-series of Sentinel-2 data products
- Gap filling techniques for reconstructing time series taking into account cloud cover and irregular acquisitions in different areas.

Expected Impact (technical, social, economic, environmental, policy etc.):

Technical

- Develop AI techniques to estimate and monitor water use at single field level to significantly enhance the accuracy of present systems.
- Generate accurate, efficient and reliable products to collect data based on remote sensing techniques in the agricultural sector to support irrigation systems

Socio-economic

- Information regularly extracted on the water use and comparisons among the productivities of different fields with respect to the irrigation strategy enable the definition of best practice for improving productivity.
- Improving the effectiveness of the irrigation system has huge effects on food security and on the productivity of producing crops.
- Increased investment in micro-irrigation technologies

Environmental

- Better management of irrigation, water resources
- Increased awareness in relation to water resources

Policy

- Demonstrated adoption of the results of using data incentive technologies AI in an area that is central in the policy of the European Commission
- Help farmers to be aware of the quantity of water consumed and to optimize crop production versus the amount of used water.
- Water use and related impact on the production is a fundamental information for irrigation departments to manage periods of water scarcity, provide equity in water supplies, and to guide investment in micro-irrigation technologies for better irrigation management.
- Enhance informed decision making capabilities of public authorities

8.1.13 Synergy of EO and geolocation data for security applications

[13] Scenario title: Synergy of EO and geolocation data for security applications	**Domain:** Security

Context:

Challenge:

The evolving space and security environment is facing the rise of new technologies and business models in different domains including Space, ICT, Security and Communication. Collateral data sources could be used to complement the evidence derived by satellite data and enhance the information provided to users / decision-makers in the Space and Security domain.

In recent years, the use of mobile applications has arisen significantly to the point that millions of users all over the globe make large and heterogeneous volumes of georeferenced data available. These data, fused with conventional EO imagery, can enhance the results by providing new elements to be considered in the analysis.

Synergies between EO data and geo-location data are not yet completely exploited (e.g. data availability, licensing issues, privacy issues) and their fusion could add a value to the standard practises in the security domain.

Scope:

The solutions in response to this call shall extract information of human and activity patterns using EO and geolocation data (e.g. from GNSS) to detect potential risks in critical situations (e.g. hot-spot in migration routes, identification of road blockages, disruption in supply chains, etc.).

Dataset sources: GNSS data or derived geo-location data (mandatory), Sentinel data, VHR images, other remote sensing data (e.g. Landsat) and open source data (e.g. population density, economic indicators or number of conflicts).

Any other data can be inserted if relevant for the call scope. In this case, the data must be accessible (free of charge or not) so a possible operational use of the solution is guaranteed

Output and coverage:

- Services (available through API and Dockers)
- Models (e.g. statistical relationship between different variables)
- Datasets (e.g. geo-location files)

AoI can be selected by the bidders but the methodology shall be applicable in other AoI.

AI4Copernicus services:

The following bootstrapping services will be made available:

- Pre-processing chains for Sentinel-1 and Sentinel-2 data products in order to obtain ready to use images for algorithms based on AI
- Processing pipelines for Sentinel-1 and Sentinel-2 data products (e.g. change detection pipeline)

Expected Impact (technical, social, economic, environmental, policy etc.):

<u>Technical</u>

- Quantitative step forward in the use of AI technologies in EO-based applications for security scenarios;
- Interoperable solutions to be able to work in multi-disciplinary environments;
- Increased capacity of processing and analysing large volumes of Copernicus Earth Observation data
- Develop operational tools that demonstrate the use of geo-location data with EO data in the Space and Security domain;
- Increased performance and/or automation of image processing, in order to have pre-operational services to be used in GEOINT tasks.

<u>Socio-Economic</u>

- Enlargement of the number of solution providers in security;
- Foster the use of Sentinel data and other data sources;
- Enhance the understanding of security threats for the general public, including EU citizens and international actors.

<u>Policy</u>

- Foster the use of EO data analysis for policy and decision-making in EU domain (e.g. EU Common Foreign Security Policy);
- Contribute with the results to the different active initiatives in the EO domain, including the GEO Space and Security Community Activity and EuroGEO.

8.2 Annex B: Entity interviews

8.2.1 Equinor interview

Equinor Case Study

Theodoros Evgeniou (INSEAD)
Pal Boza (INSEAD)
Michelle Aubrun (Thales Alenia Space)

Equinor is an energy company with more than 20,000 collaborators working in the field of oil, gas, wind and solar energy in more than 30 countries worldwide. It is the largest energy operator in Norway and a leading international offshore operator with a growing portfolio of renewables.

1. **In what practices are Earth Observation (EO) data being used by the organization through Artificial Intelligence? Is it common for other companies from your sector to use EO data?**

A key usage of Earth Observation (EO) data at Equinor relates to image analysis based surveillance routines to detect important objects, such as icebergs or oil spills, in order to monitor risks potentially affecting operations. Satellite images can be efficiently used to support change detection because of repetitive coverage at small intervals, and reliable image quality. Equinor needs to constantly monitor and notice changes or potential hazards – icebergs and oil spills being examples of these – affecting its energy infrastructures. For example, icebergs can potentially damage oil platforms and lead to important material and environmental damages. For instance, only from Greenland 20.000 to 40.000 icebergs are formed during a year and travel to the Northern part of the Atlantic Ocean. Oil spills are also hazards that Equinor treats as emergency events and aims to detect. In both cases, early detection is key for prevention. Similarly to other players in the sector, Equinor uses EO data to prevent or diminish critical events.

Another case for relying on EO data is weather forecasting, which is very important for the whole industry. However, in this case EO data is not being "processed" by energy companies, as they are mainly buyers of the final weather forecasts, without dealing with the raw EO data and related tools. Weather forecasting is important for Equinor concerning both its oil infrastructure (e.g., safety issues) and renewable infrastructure (e.g., production forecasting).

The number of potential use cases of EO data and tools is increasing at Equinor and in general in the energy sector. For instance, Equinor is investigating how satellite data can support conducting environmental impact assessments while real-life needs such as the Covid situation has also led to new use cases. Moreover, EO data can be used for making other forecasts – EO data and tools providers need to understand the key forecasts of the users such as Equinor.

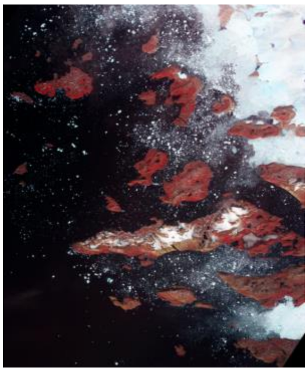

Figure 1 : Satellite view of icebergs (1)

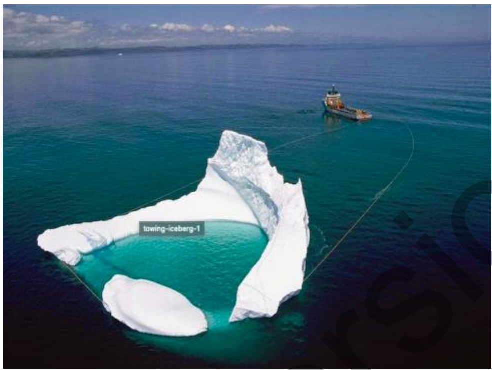

Figure 2 : Towing an iceberg from a collision course (2)

2. **How do EO data and AI tools create business value for the organization, enable you to innovate in your business models and value chain, and/or economic value for the society at large? What factors do you consider when making an investment case using AI tools to leverage EO data? What costs and benefits do you typically consider?**

Equinor can use two different channels for procuring EO related information. Satellite images have been traditionally purchased from trusted external suppliers who organize it after sourcing it from the original sources. This is the main channel used for procurement today. Meanwhile, Equinor sees benefit in directly getting information from the original EO sources through remotely accessing online platforms that provide raw EO data, tools and services. The company is currently experimenting with this second option and is testing exiting remote platforms and tools. For the moment Equinor has not tested extensively Artificial Intelligence (AI) tools on EO online platforms, which is something to be considered in the coming period.

EO online platforms can provide value in the following forms:

- *Reduced operational cost of dealing with EO data*. The processing of raw satellite images is time-consuming, but the tools provided by online EO platforms can considerably reduce this processing time. EO platforms also offer computational power that does not have to be built and maintained locally. It is more convenient and cost-effective to process and analyze data on an EO platform than to having to first download it locally and then analyze it – given the size of this data. This also supports collaboration and sharing across multiple users through the platform. The ability for remote access and analysis is considered to be key, assuming security is man[1]aged properly.

[1] Source: *https://eoimages.gsfc.nasa.gov/images/imagerecords/78000/78479/greenland_ast_2005198_lrg.jpg*

[2] Source: Photo by Randy Olson, *https://www.amusingplanet.com/2014/05/towing-icebergs-away-from-oil-platforms.html*

- *Control over data*. External providers will deliver satellite images that are pre-processed based on initial demand on behalf of customers (such as Equinor). However, the same *raw* data (Sentinel data, etc.) used by external providers are also accessible through online EO platforms. Being closer to the source of data gives control to the user organization (e.g., Equinor), and access to more up to date information while increasing the confidence in the data itself.
-

There is no standard process for defining investment cases. Nevertheless the *cost of purchasing* satellite images has decreased dramatically in the past years as the technology has matured, and because most satellite raw data is now freely available. This also makes EO data and tools much more accessible.

However the price sensitivity is not very high when choosing between different sources of information and tools to mitigate possible damages. For instance, an iceberg can cause serious damage to an offshore oil platform. Since the investment cost of an oil platform is equal to several hundred million USD, the price to be paid for EO based information to avoid a possible damage is often not the main issue. Similarly for oil spills, the environmental damage can be immense.

3. Who are the main users of EO data, users (not developers) of AI tools in your organization?

Who is involved in the implementation of these technologies? What are the roles needed and what are their responsibilities? Are specific skills or trainings required?

Depending on the use cases, different parts of Equinor's organization can be considered as users. There is a dedicated team (currently one person) inside Equinor to serve the demand of these users from inside the company for EO data („Our job is to tell our colleagues where things are when"). There are specific computational skills and training required for being able to use online EO platforms. There is also important work to be done to increase awareness across the organization of what can be achieved with EO data as well as tools such as AI (what can be called "evangelization" of EO data and tools).

4. What are the main barriers for the adoption in the sector (in general) and in your organization (more specifically)? How could these barriers be mitigated? What are the success factors for the adoption, are there any best practices or lessons from your experience?

There is an important *organizational element of change* to be overcome for the adoption of EO online platforms and tools. A generation gap exists between those in their fifties taking managerial decisions and younger generations who have a better understanding and awareness of how new technologies can be used. This is the case in the energy sector in general and at Equinor specifically. Moreover, several colleagues have a background in AI at Equinor, but often they are not aware of where the data is available and what can be achieved with it, especially in the EO space. This is why "evangelization" can be critical.

Furthermore, there is also an important difference between individually developing an EO infrastructure or remotely and collaboratively using it without effectively "owning" it. A few decades ago the mindset at Equinor was more to have such infrastructure developed internally, without having to share the data and the infrastructure itself with external partners. This was very resource-consuming but was also adequate from a security point of view, compared to cloud solutions that raise safety issues. However, use cases are shifting today towards remote data processing and collaboration across companies/organizations. There is increasing acceptance that collaboration in this space is important. Leveraging these new use cases in the EO space needs both a *novel mindset* ("sharing data is good for everyone") about the infrastructure and data ownership on behalf of the organization and *well-defined roles and responsibilities*. The trend seems to be from "in-house and proprietary" EO data systems and tools to "on the cloud and collaboratively developed" ones.

Finally, there is some lack of clarity not only about what can be achieved with EO data and technologies such as AI, but also what the main differences between different platforms are – for example users sometimes cannot tell the difference also across the different DIAS platforms. Lack of clarity by end users such as Equinor is also naturally a barrier for the development of new businesses that can serve these users.

5. What are the technical barriers for the adoption?

At this stage Equinor has been testing only the Euro Data Cube application. The following technical barriers and issues have been identified: incentives motivations issues

- It was very helpful to have some information (e.g. jupyter-notebook examples) on Euro Data Cube about how to use the application, though having more concrete examples would be important,
- Few IT skills are necessary to use this application,
- There is a real interest to have visualization and processing tools on the same platform because it is time-consuming and not scalable to switch from one platform to another, that is why Equinor will test an EO processing library, named EO learn,
- The use of regular data format is an important issue,
- It is important not to use resources, such as Google based information, in order not to be tracked (e.g. look which areas Equinor is interested in). This raises safety and security issues for Equinor,
- Knowing how the data is produced can give the users more confidence. Hence such information may be important to share,
- The infrastructure resource allocation (e.g. CPU, memory) were totally transparent. For free trial account, there is capacity limitation, but if users need more resources, they just have to upgrade their license. Equinor might need more resources and would be probably open to purchase such a version.

6. Possible takeaways for the AI4Copernicus project

- Many companies, both SMEs and larger corporations – leverage EO data and many platforms provide EO data and related tools. The AI tools integrated into the AI4Copernicus platform can be a differentiator for AI4Copernicus,
- However, potential users of these platforms – and of EO data platforms – are currently not clear about what differentiates the platforms from each other. Frameworks and metrics (e.g., comparing different offerings) that can help improve clarity in the market may prove useful,
- Evangelization needs to be done to increase awareness of what can be achieved with EO data and technologies such as AI and eventually reach potential future users. Facilitating such awareness building – for example through demos, events, presentations, etc. – can be fruitful,
- EO data and tools providers need to understand the key forecasts ("what does the user want to know") of the users,
- Trust in the data sources and understanding of how the data is produced needs to be considered. This can be achieved for example by developing appropriate information resources or potential training modules (e.g., videos online, tutorials, etc.),
- Price is not only the most important factor for users of the platform, ease of use is also a very important aspect to take into consideration.

We thank Richard Hall and *Claire Bernard-Grand'Maison* for their valuable contribution and participation in the interview on behalf of Equinor.

8.2.2 SatCen interview

SatCen Case Study

Theodoros Evgeniou (INSEAD)
Pal Boza (INSEAD)
Michelle Aubrun (Thales Alenia Space)

The European Union Satellite Centre (SatCen) is the agency of the European Union (EU) that supports the EU's decision-making in the field of the Common Foreign and Security Policy. It provides geospatial intelligence products and services based on the exploitation of relevant space assets and data, including satellite and aerial imagery.

1. In what practices are Earth Observation (EO) data being used by the organisation through Artificial Intelligence?

SatCen has a broad field of action, covering all features, activities and phenomena related to security that can be analysed and monitored from space. These include humanitarian aid, support to evacuation operations, military capabilities, monitoring of weapons of mass destruction, critical infrastructures monitoring and security related to border surveillance and climate change. SatCen is a so-called operational centre, meaning that it executes direct tasks following the request of its stakeholders (mainly Member States of the EU and other EU entities). In practice, a task consist in (1) receiving a request to extract geospatial information from a given geographical area for a specific time period, (2) processes geo-located data (coming mostly from satellite VHR data, with the support of collateral data) and (3) provides a product consisting of analysed image(s) and a report. SatCen performs also R & I activities within the Research, Technology Development and Innovation (RTDI) Unit (the Unit involved in AI4Copernicus), to maintain the centre capabilities at the state-of-the-art.

The Centre explores the use of Artificial Intelligence (AI) solutions to execute its activities related to Geospatial Intelligence. SatCen participates to different R&I projects based on AI (e.g. GEM, where the aim is creating a Conflict Pre-Warning Map based on the use of EO and collateral data, and CALLISTO, a project to build services related to border surveillance -e.g. new barriers, new roads and cart tracks- on top of the ONDA DIAS infrastructure). An example for using AI concerning EO activities at SatCen, is the detection, analysis and monitoring of roads in desert regions (Figure 1), derived from a collaboration between SatCen and ESA phi-lab.

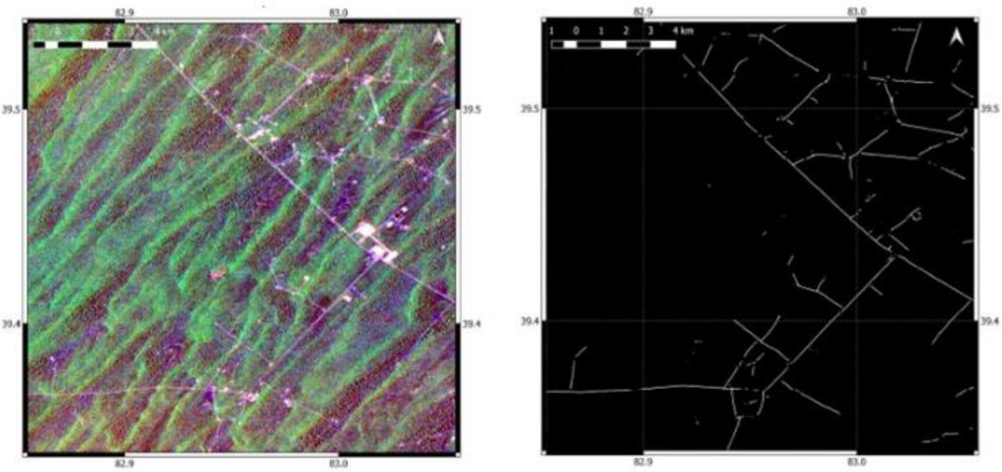

Figure 3 : The result of the ESA-SatCen article

(The result of the ESA-SatCen article on the use of Artificial Intelligence on SAR imagery to identify roads in desert regions. On the left side, the processed SAR image is used as input to the model, on the right side the roads detected by the model, June 2020[2])

Another application of the use of AI solutions at SatCen relates to the identification of occupied buildings from space. The United Nations was evaluating the quality of local governance in the different regions of Lybia and to do so, it has concluded that changes in the population could be a good indirect indicator. SatCen built a set of methodologies based on the accurate digitalisation and classification of occupied buildings, supported by AI algorithms and specific night-time imagery. Using these methodologies, SatCen was able to provide a precise estimation of the population in different Libyan regions.[3]

2. How do EO data and AI tools create business value for the organization, enable you to innovate in your business models and value chain, and/or economic value for the society at large? What factors do you consider when making an investment case using AI tools to leverage EO data? What costs and benefits do you typically consider?

SatCen is in a special "intermediary" position in the EO value chain. The Centre does not operate satellites itself; it secures EO data both from commercial sources and freely available Sentinel satellite data. After processing them, it provides information and services to its stakeholders (the end-users). As part of the processing, SatCen uses AI techniques to extract information from the gathered data. One of the main application domain of AI solutions at SatCen is the automation of human tasks to save time, resources and to increase the accuracy of the provided information. The future aim is to further increase the number of use cases where AI tools can efficiently create value compared to human-intensive processes.

SatCen also explores external means to purchase EO related AI tools (e.g., data sets for training algorithms, pre-trained algorithms, etc.), which are similar to the services AI4Copernicus intends to develop. These external means are important since they allow to support the research and development activities and reduce the time and human resource-intensive development processes of AI tools. Consequently, SatCen intends to find the right balance between procuring AI tools from external sources and developing them internally.

There is no a specific standard process for defining when AI tools should be purchased from outside or developed inside the organisation. The main trade-off is between price, availability and accuracy of the AI tools. The purchase of AI tool eventually affects SatCen annual budget (where also a dedicated budget line for

[2] https://www.satcen.europa.eu/page/study-on-sar-data-processing-published-by-satcen-staff
[3] SatCen Annual Report 2020.

purchasing satellite data is included). Moreover, further funding are available from the EC (e.g., from R & I projects and other Delegation Agreements).

3. Who are the main users of EO data, users (not developers) of AI tools in your organization? Who is involved in the implementation of these technologies? What are the roles needed and what are their responsibilities? Are specific skills or training required?

SatCen has the necessary AI talents and ML skills inside the organisation to develop and use AI tools, hiring is not so much of an issue. What is mainly missing inside the organisation is the time and the label datasets to train the algorithms.

Final users of the final information produced with the support of AI tools are outside of SatCen's organisation. These are in most cases the Member States of the EU, who also provide the budget for SatCen on an annual basis.

4. What are the main barriers to the adoption in the sector (in general) and in your organization (more specifically)? How could these barriers be mitigated? What are the success factors for the adoption, are there any best practices or lessons from your experience?

SatCen is using EO related AI tools in its day-to-day operations. From this perspective, there are no major organisational barriers for adoption and there is a willingness to increase the use of AI tools for use cases when these can be competitive compared to traditional 'manual' techniques both from a price and accuracy perspective.

However, some barriers limit the external use of EO related AI tools and services that are similar to the ones offered by the AI4Copernicus platform:

- The security sector is particular since it often involves dealing with classified information, this also limits the number of "trusted" third-party service providers,
- When dealing with restricted information, most stages of the AI tool's development process needs to be kept in-house for security reasons. For instance, the storage of any type of related information on the cloud would be a very sensitive issue (e.g. areas of interest on which the AI tool has been run, or objects on which the AI tool has been trained). Therefore, developing and/or storing such a tool in-house would be a good option;
- Use cases in the security sector often require special data and knowledge. An example use case could be the automatic recognition and distinction of military airplanes from space. The development of such automatized AI tool is complex since it needs to be trained based on a very special and sensitive database, which will be very expensive, and requires both knowledge of the data science domain and familiarity with the security sector.

Several AI tools can be purchased commercially for object detection task (such as the distinction of military airplanes), however, the robustness of these tools is not always satisfactory. Often they are too much specific to their training dataset (areas, objects, season) and thus does not always reach a wide operational use. There are today several online platforms that offer different services in the EO space. However, for the moment there is no one general registry accessible concerning the services and data available on different platforms. Such information could help matching supply of services of online EO platforms with the demand for those services. Furthermore, the services related to AI tools are often very "technical", more user friendliness and different service levels could lover the barrier for adoption.

Furthermore, there are also cases when "hybrid" solutions are used between procuring AI tools from outside the organisation and the development of these tools by SatCen itself. For instance, pilot projects were developed on an external platform and once their methodology was tested and confirmed they were moved to a platform used internally at SatCen, to be further developed for a specific context both for sustainability and internal control (e.g. H2020 BETTER project outcome, SatCen-ESA phi lab study). This approach has proved to be successful by SatCen in the past.

5. What are the technical barriers to the adoption?

At the current stage, SatCen has tested several EO platforms and AI tools similar to the one proposed by AI4Copernicus. Based on its experience with the other online platforms, the technical barriers are:

- *Availability of raw data for long-term analysis* (e.g. commercial data for training need to be purchased, archived Sentinel data are made available for a limited period of time);
- *Accessibility of EO data (*e.g. SatCen has privileged access to the ESA Sentinel catalogue, which is faster than the normal service, while other players might have issues with speed), it is important to guarantee an easy and intuitive access to EO data (e.g. request data on an Area Of Interest for a certain period of time);
- *Availability of AI tools and their robustness*, having AI tools "close" to data is something to take into account and often models might need to be trained to do similar tasks in different contexts. For instance, detecting roads in two different desert areas of the world might require different AI models;
- *Complexity of AI tools*, using AI tools might not be straightforward and requires a technical knowledge (e. g. there are many hyper-parameters to set);
- *Results visualization*, it is important that cloud platforms have a visualisation layer to visualize these results and the input EO data.

6. Possible takeaways for the AI4Copernicus project

- Several EO related platforms are financed from grants. Such projects have a specific lifecycle (2-3 years), however, it is a challenging sustainability issue to maintain the platform beyond the end of the project's lifecycle once the grant financing is no more available;
- AI algorithms need to be trained for a very specific context to have a proper accuracy. For an AI4Copernicus type of online platform it would be important to offer algorithms that can be further transferred in function of the given context[4];
- A high-level inventory of services provided by different online platforms in EO (including AI tools) is something that does not exist yet and it would be very beneficial to have a general overview of the existing offers as orientation for the users;
- Since users of EO data have a different level of 'AI literacy', online platforms could have different service levels: (1) Services for 'standard' users without special AI knowledge which still allows the user to access EO derived products in a simple user-friendly manner (e.g. answering a request of "counting how many trucks are in an image"; (2) More specialised offer with AI tools included, that requires more data science knowledge (e.g. adapting the hyper parameters of Neural Networks to a specific context). For instance, in the internal SatCen GEO-DAMP platform there is a visualization part (to visualize data and output) and also a processing part (with the procedures to create output);
- The security sector is particular since it often involves dealing with restricted information, this also limits the number of "trusted" third-party service providers and most stages of the development

[4] https://www.satcen.europa.eu/page/study-on-sar-data-processing-published-by-satcen-staff

process needs to be kept in-house, having all the information stored on the cloud is often not an option.

We thank *Michele Lazzarini* for his valuable contribution and participation in the interview on behalf of SatCen.

8.2.3 Polar View interview

Polar View Case Study

Theodoros Evgeniou (INSEAD)
Pal Boza (INSEAD)
Michelle Aubrun (Thales Alenia Space)

Polar View is a non-profit organization providing operational, satellite-based monitoring for the polar regions and the cryosphere. It delivers services to support safe and cost-efficient marine operations, improved resource management, sustainable economic growth and risk protection across sectors.

1. In what practices are Earth Observation (EO) data being used by the organization through Artificial Intelligence? Is it common for other companies from your sector to use EO data?

Polar View is using satellite EO data in combination with different sophisticated models and automated tools, for most of its activities. The satellite images are converted into products that graphically illustrate the characteristics of the ice and snow in the polar regions. Polar View is overwhelmingly working with SAR and Sentinel 3 data products, these are the most suitable for the polar regions and the most reliable and are also free. Polar View's main services include enhanced sea ice information and ice-edge and iceberg monitoring.

Polar View's sea ice services provide forecasts on sea ice motion, concentration, thickness, ridges and deformations for the polar regions using numerous multi-category sea ice models. This service is being delivered for the shipping, scientific, tourism, emergency management, oil and gas, and fishing operators in the Arctic, Baltic Sea and the Southern Ocean.

Information concerning the location of ice-edge is crucial for Northern residents to navigate safely and efficiently when hunting, travelling or developing the tourism industry. It is increasingly difficult to rely on "traditional knowledge" in predicting ice conditions, which might be associated with the effects of climate change. This means that there is an additional need for information to navigate safely in the ice-edge area. Polar View's dedicated product combines satellite SAR imagery with a human analyst's interpretation of the location of the ice floe edge in the form of updated image maps. Delineation of the floe edge is now being automated and in the future, involvement of a human analyst will no longer be required.

Finally, icebergs are an important source of risk both for the shipping and the energy industry (e.g. for off-shore oil platforms, wind turbines, etc). Polar View supports the management of operational risk by providing information about the exact position of the iceberg in near-real time, as well as their forecasted and historical motion. Polar View's other services include monitoring for lake and river ice, snow cover maps and glacier monitoring and assessment.

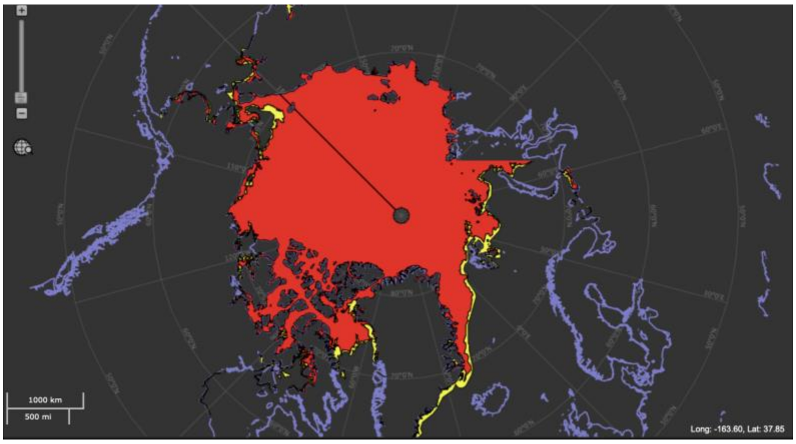

Figure 4 : Ice chart of the North Pole, source: www.polarview.aq

2. How do EO data and AI tools create business value for the organization, enable you to innovate in your business models and value chain, and/or economic value for the society at large? What factors do you consider when making an investment case using AI tools to leverage EO data? What costs and benefits do you typically consider?

In the practice of Polar View, the main value creation potential of intelligence tools comes from the automation of highly human-intensive processes. For instance, in the past, information related to the ice situation in the polar regions was provided on hand-drawn maps for the shipping industry or the local population.

This has now changed. On one hand, with the increasing use of EO data, human-intensive processes are now more competitive. On the other hand, there is a huge pressure to cut costs both from the private and the public sector, which also drives the industry towards experimenting with intelligent solutions for instance to classify sea ice by using machine learning techniques or to identify icebergs.

3. Who are the main users of EO data, users (not developers) of AI tools in your organization? Who is involved in the implementation of these technologies? What are the roles needed and what are their responsibilities? Are specific skills or trainings required?

Only a small number of professionals have skills both in AI tools and EO data while knowing the polar regions (e.g., there are approximately 30-40 people in the entire world who are applying machine learning to sea ice classification). Furthermore, users of the final information produced with the support of AI tools and EO data usually have no competence in these fields and are not aware how the information was created.

4. What are the main barriers for the adoption in the sector (in general) and in your organization (more specifically)? How could these barriers be mitigated? What are the success factors for the adoption, are there any best practices or lessons from your experience?

A huge amount of EO data is available from all types of sources, which leads to significant competition on the market. For instance, Amazon Web Services or Google Earth Observation are providing satellite images for free and the number of existing EO platforms has also considerably increased during the past years.

First, although EO platforms provide important value for society indirectly, it is difficult to build a viable business case for existing EO platforms to ensure long-term sustainability - they are currently often financed based on grants from the EU or other governmental agencies. Grants are provided for a limited period (e.g., 2-3- years) and, at least currently, once the financing period finishes new funding needs to be secured. This might also mean that some of the current market players (mainly the smaller ones, without access to long-term sustainable funding) might have to leave the market, which also leads to hesitation on behalf of users concerning their choice of EO platform. A common reaction is *"why should I store all of my data on a small platform that might have financial sustainability issues and might not exist in a few years? I might rather buy my own computer system, or work with the larger players, with less risk."*

Second, currently, AI tools are still not widely used in the operational decision-making processes in the EO sector, they are mainly providing assistance for humans because their accuracy is not yet reliable (or at least decision-makers do not trust them). Still, 5-10 more years are needed before these technologies can become reliable and safe and can be used operationally without a problem.

Third, the general end-users do not need to know how the information based on EO data and AI tools were created. Only a small number of professionals need to understand the tools in detail and to have access to the online EO platform, especially in the context of the polar regions, consequently, they do not represent an important demand on the market (e.g., there are approximately 30-40 people in the world who have the knowledge to apply machine learning to sea ice classification). Hence the user base may be very limited.

Fourth, there is a mindset of wanting to *"own computing power"* and have the *data locally available* instead of buying it from third-party service providers. This is especially true for academia, a sector that is a potential client of EO platforms. This phenomenon is also enforced by the fact that academia usually gets its financing *"in chunks"* so it is difficult to get them onboard for products that have recurring costs.

Finally, there is often a lack of willingness to build on existing tools and services that have been already delivered by other parties, *"since it is always more exciting to build something new".* For this reason, Polar View is trying to integrate different parts of existing solutions from other providers to its platform.

5. What are the technical barriers for the adoption?

The technical barriers are mainly based on Polar View's experience with the TEP polar platform:

- Access to EO data is relatively easy, mainly thanks to data interoperability standards. However, downloading large datasets is not an efficient way of working, so it is better to bring the analysis to the data in the cloud;
- Complexity of AI tools, once EO data is accessible, a certain technical level in programming is required to interact with the data;
- Adaptability of AI tools, there is a very large range of end-user profiles, so it is not possible to propose a unique tool to target all the users (e.g. expert users do not want to be overwhelmed with fancy code).

6. Possible takeaways for the AI4Copernicus project

- The use of *AI tools in the EO domain is still not mature* and it may take 5-10 years (or more) until these technologies will be widely used in this industry, mainly for the polar regions. The reasons for this are both technological and also cultural since some decision-makers still do not fully understand and trust data-intensive technologies. This means that the evangelisation of data-intensive technologies in the EO sector is an important issue to consider;

- Important results have been achieved in the past in the form of projects, tools and also platforms in the EO space. However, there is often a *lack of willingness to build on already existing results and to reuse them to created future products*. Channelling these into future offerings can save resources and also provide value for potential customers;
- There is a *general issue of sustainability* in the market of EO platforms. Although there is some demand for the services of EO platforms (e.g., in the case of Polar View, on behalf of local communities living in the polar regions) this is not sufficient to build a long-term sustainable business case, there are several reasons for this:

    - For instance, Big Tech players such as Google or Amazon are giving away EO data and tools to manipulate those for free, probably using the profitability of their other business lines, which means that the *supply of EO products and services on the market is heavily subsidised* on behalf of a few big players;
    - Some potential clients still have a preference towards wanting to *"own computing power" and have the data locally available* instead of buying it from third-party service providers. These clients have low willingness to pay for computational power or cloud data storage which is also part of the value proposition of several EO platforms;
    - Several EO related platforms are financed from *projects related to grants*. Such projects have a specific lifecycle (2-3 years), however, it is a challenging issue from a sustainability perspective to financially maintain the platform beyond the end of the project's lifecycle once the grant financing is no more available, which also leads to hesitation on behalf of users concerning their choice of EO platforms as there are risks that the platforms will close.

We thank *David Arthurs* for his valuable contribution and participation in the interview on behalf of Polar View.

8.2.4 CERTH interview

CERTH Case

Theodoros Evgeniou (INSEAD)
Pal Boza (INSEAD)
Michelle Aubrun (Thales Alenia Space)

1. **In what practices are Earth Observation (EO) data being used by the organization through Artificial Intelligence? Is it common for other companies from your sector to use EO data?**

CERTH is an internationally recognized research centre and one of the largest in Greece with more than 1,200 employees. The organization carries out a wide range of scientific activities, the most important areas of which include energy, environment, industry, mechatronics, information & communication, transportation & sustainable mobility, health, agro-biotechnology, smart farming, and safety & security research.

The applications based on EO data and AI tools represent a significant part of the activities covered by the research centre, such as water resource monitoring, oil spill and environmental disaster surveys, as well as civil protection plans. CERTH is also specialized in the fusion of Earth Observation multi-modal data with additional sources of data, such as in-situ observations from citizens through social media, along with their associated modalities (text, image, temporal, and spatial information).

Figure 5 : The 1989 Exxon Valdez Oil Spill from https://earsc-portal.eu/display/EOwiki/Detect+and+monitor+oil+slicks

2. **How do EO data and AI tools create business value for the organization, enable you to innovate in your business models and value chain, and/or economic value for the society at large? What factors do you consider when making an investment case using AI tools to leverage EO data? What costs and benefits do you typically consider?**

CERTH mainly uses free databases (e.g. satellite data from the Copernicus EU programme) and develops open-source AI algorithms to add value on top of raw data. State-of-the-art algorithms are developed at CERTH to address specific real-life EO related issues with high societal, economic and environmental impact. The organization provides an integrated and complete solution that involves the acquisition of satellite data to visualize the result of the analysis. CERTH defines itself as a solution provider and not as a simple AI algorithms developer, since it offers complex solutions with high Technology Readiness Levels to the wide community of users.

There have been no major investments to purchase commercial EO data, novel AI tools or additional computational (cloud) capacity in the past at CERTH, so currently, there is no specific process built internally to evaluate the economic feasibility of such investments. However, due to the increasing connections among CERTH and public/private partnerships and contracts, CERTH would invest money in purchasing EO related high-resolution data and computational capacity from third-party in relation to a privately financed large-scale project aiming to develop a novel algorithm and deliver a mature solution that meets the market's needs.

EU-funded projects provide a significant proportion of CERTH's revenue and allow the progressive development of internal expertise. Thanks to the diverse research fields covered by CERTH and the numerous opportunities, the organization can participate in several large-scale projects simultaneously and is not exposed at an organizational level to unexpected funding interruptions.

In addition to EU-funded projects, CERTH also works with companies from the private sector. In such a context CERTH typically provides answers to some well-defined research questions (for example oil spill detection software suite for inland waters) using free EO data and self-developed AI tools. Through these projects, CERTH enables private companies to gain scientific knowledge, use state-of-the-art scientific methodologies and more mature ICT solutions in their day-to-day operations.

3. **Who are the main users of EO data, users (not developers) of AI tools in your organization? Who is involved in the implementation of these technologies? What are roles needed and what are their responsibilities? Are specific skills or training required?**

Most of the people working in the research institute are scientists who are experts in a specific field. Scientists, engineers, and computer scientists in-house have the relevant qualifications and knowledge to work on the accurate and efficient design of AI tools.

The evangelization of EO data related AI tools could be improved not only by educating professionals but also by supporting collaborations both internally and externally. The realization of further projects jointly with private companies and research institutes could be mutually beneficial for all parties and could contribute to the broader adoption of these new technologies.

4. **What are the main barriers to the adaptation in the sector (in general) and in your organization (more specifically)? How could these barriers be mitigated? What are success factors for the adoption, are there any best practices or lessons from your experience?**

The development of AI tools requires annotated data and massive computational resources, which has **a** *significant cost*. An important aspect that CERTH might consider is the use of the DIAS platforms that can offer satellite data products and computational capacity.

Each use case can be very specific and might need completely *redesigned and fine-tuned tools* (e.g., databases and trained algorithms). However, it is often complex to financially justify the acquisition of new tools for each specific use case.

The lack of market-facing projects is an important barrier for using third-party services to provide AI tools in the EO space. The knowledge accumulated within the research institutions is usually not brought to market, it is only used in the framework of internally realized scientific projects, while the resources and time required to engage and integrate new innovative tools also discourages organizations from using third-party services. This problem can best be solved by supporting collaboration and technology transfer between research centres and private entities.

5. **What are the technical barriers for the adaptation?**

CERTH has a long experience related to AI tools and mainly related to two different EO platforms: Copernicus Open Access Hub (ESA platform) and ONDA (DIAS platform). Based on this experience, the technical barriers for successes are the following:

- *Availability of EO data.* For tasks that require a quick response, such as flood detection, the revisit time is a limitation (3-4 days);
- *The request of EO data.* The searching capability could be improved by using some semantic and metadata criteria;
- *Accessibility of EO data.* Not being able to perform large-scale downloads (download multiple products at once) is one of the main limitations. Thus, when CERTH needs to acquire only a few EO data, it downloads them using a “classical” platform, and when CERTH needs to acquire a larger set of EO data, it uses a DIAS platform and processes those directly online;
- *Understanding of EO data.* Satellite data products are not so easy to understand but the ESA manuals online are well-elaborated and can provide help and explanations;
- *Processing of EO data.* Standard data formats are quite easy to process (e.g. geoTiff, shapefile, png, …) but specific data formats can require a lot of effort for processing and visualization;
- *Development of AI tools.* CERTH prefers to develop its own solutions on its own servers;
- *Integration of AI tools.* CERTH can meet some technical issues concerning the availability of specific libraries and GPU, but the main issue remains the cost of using those. CERTH generally uses Docker containers for the integration step, which is a technology that requires some learning effort at the beginning;
- *The complexity of AI tools.* Understanding and reusing AI algorithms require significant efforts. This can be reduced by using Jupyter Notebook to illustrate what the algorithm is doing;
- *Understanding AI results.* Visualisation is the best way to understand the results.

6. Possible takeaways for the AI4Copernicus project

- Supporting collaboration between research institutes and companies can both help the evangelization of technology and stabilize revenue for scientific institutions,
- Research institutes need to be provided with a clear, transparent return on investment plan to be interested in innovative solutions. Market facing projects can significantly help to clarify financial feasibility,
- It is a common practice to use existing algorithms and further train them for a specific use-case. For an AI4Copernicus type of online platforms, it would be important also to offer algorithms that can be further trained in function of the given context,
- Research institutes operating in an academic environment often apply existing solutions through redesign and minor modifications. However, the resources and time required to engage and integrate new innovative tools discourages them from using external enhancements,
- Price is not necessarily the most important factor influencing the use of EO platforms, ease of use is also a very important aspect to take into consideration.

8.2.5 VISTA interview

VISTA case

Theodoros Evgeniou (INSEAD)
Pal Boza (INSEAD)
Michelle Aubrun (Thales Alenia Space)

VISTA Remote Sensing in Geosciences GmbH is a company with 23 employees located in Germany, active in the field of remote sensing and modelling and has over the last 20 years worked on translating state-of-the-art scientific methods into operational services.

1. In what practices are Earth Observation (EO) data being used by the organization through Artificial Intelligence? Is it common for other companies from your sector to use EO data?

VISTA mainly uses EO data, and especially Copernicus related EO data, for agricultural monitoring, such as yield forecasting and general crop development tracking. It has developed EO data-based smart-farming services to provide crop production information for professionals to facilitate decision-making. For instance, the company performs long-term growth analyses from the available data, to identify the ideal seeding densities and the areas with the best soil. Besides continuously monitoring crops and providing yield forecasting, it also provides down to field level information, e.g. fertilization and irrigation recommendations, that result in significant savings and environment friendly exploitation processes.

Based on EO tools provided by the Copernicus program, VISTA participated in the EU financed ExtremeEarth project, which deals with extreme environmental impacts. In the project, the company is helping to identify food security and water supply problems, based on EO data. In the project VISTA closely cooperates with the University of Trento to explore food security as one of the 17 UN-set sustainability goals.

A range of technologies rely on EO data in the agricultural sector, but the efficient and accurate use of big data and the full integration of Artificial Intelligence (AI) solutions is still among the areas under development, also in the case of VISTA.

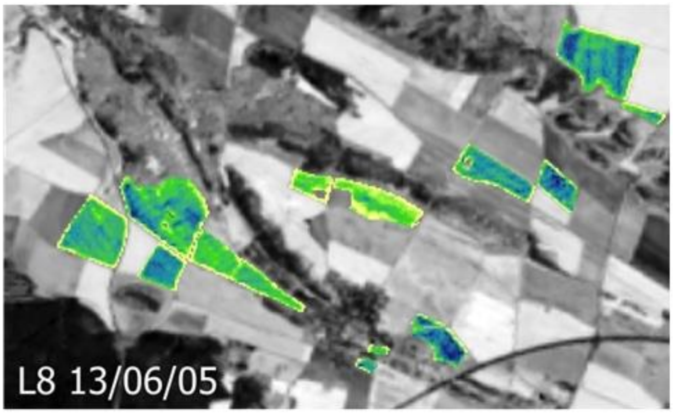

Figure 6 : Map of chlorophyll content

(Map of chlorophyll content for several winter wheat fields in mid-June (05.06.2013) at vegetation maximum from www.vista-geo.de)

2. How do EO data and AI tools create business value for the organization, enable you to innovate in your business models and value chain, and/or economic value for the society at large? What factors do you consider when making an investment case using AI tools to leverage EO data? What costs and benefits do you typically consider?

VISTA is almost uniquely using tools developed in-house and relies on free EO data platforms for its current activities. Today, most of the data used by the company are obtained through the Copernicus program. The data available free of charge is crucial for the company, as it contributes to affordable operating costs leading to more widespread adoption of innovative technology. However, VISTA considers that "it is part of the business" to pay for computing performances.

To provide smart farming services, experts work with sensitive data, some of it covered by the EU General Data Protection Regulation, from farmers and customers. To assure highest standards of data protection, VISTA only works with cloud platforms located within the EU. Therefore, the data security of the tools used is essential in any collaboration and potential investment.

Currently, EO data-based tools can provide value in the agricultural sector. However, for the moment VISTA has not been able to test whether EO-focused AI platforms can offer suitable solutions for the company to use AI techniques. VISTA is focused more on physically-based modelling, e.g. with radiative transfer models and mechanistic crop growth modelling, but has also developed and tested e.g. random forest models and CNNs. VISTA is also involved in several large-scale or direct collaborations which helps to understand and get views on the potential and limitations of AI tools in EO data analytics. Last but not least, VISTA has recently added AI tools to their cloud platform Food Security-TEP via a federation with the AI platform Hopsworks.

Before deploying AI solutions in services delivered to farmers, it is key for VISTA to understand the possible application of these solutions in its field of use and to compare their value creation potential with the currently applied solutions at the company. Accuracy and stability of the solutions as well as their ability to deal with extreme events have to be thoroughly tested and validated before operational use, as e.g. an inability to deal with an extreme weather year can endanger a whole harvest. Hence, for now AI tools are mostly used in a research context at VISTA. Within these research projects, also the compatibility of these new solutions with the currently used tools as well as possible synergy effects are being tested.

3. Who are the main users of EO data, users (not developers) of AI tools in your organization? Who is involved in the implementation of these technologies? What are roles needed and what are their responsibilities? Are specific skills or training required?

VISTA is structured in three teams: Data Analytics, DevOps and Application Development. VISTA's experts work almost exclusively with in-house developed tools, so employees need to master the operation of the software in addition to the basic knowledge needed to manage EO data. Programming knowledge is an extremely important skill, as employees are both users and developers of the tools used. AI technologies are tested in research projects by data analysts, in-house AI tools have been developed by VISTA's application development team and the federation of Food Security-TEP and Hopsworks is managed by DevOps as well as external experts.

4.What are the main barriers for the adaptation in the sector (in general) and in your organization (more specifically)? How could these barriers be mitigated? What are success factors for the adoption, are there any best practices or lessons from your experience?

From VISTA's perspective, three main issues hinder the use of AI tools related to EO data: The quality and size of the available (training) databases, the transferability in time and space of trained models and generally AI literacy are still relevant barriers in the agricultural industry.

The specificity of the *available databases to train algorithms* is a significant problem. Use cases in the agricultural sector can be very different in function of crops or the micro-climate, this makes it difficult to develop "one-type-fits all" solutions. Because of this, trained agricultural AI models can often not be transferred well in time and place, as different growing seasons and different regions have crop growth conditions not present in the original training.

An issue that goes beyond agriculture is the lack of knowledge concerning the "limits" of AI technology. Professionals, in general, working with EO data do not have access to practical information about building, operating AI systems, can only learn about these through their "own discovery" and need clarity about their effectiveness and constraints. This means, that the lack of know-how to develop and operate data-intensive solutions slows down the integration of the technology. Renowned experts with an in-depth knowledge of AI tools are often snatched up by large corporations, who have larger resources than the average EO business.

In addition to general barriers, consumer confidence and acceptance are also important considerations for VISTA. Understanding the results and interpreting causal relationships is essential for VISTA's customers as they are developing their strategy. The introduction of new methods, such as AI, may make the transparency and explainability of the results more difficult and thus can undermine consumer confidence. *"Most of the time clients want to know every aspect of how we came to a given result or conclusion. We have perfect answers for such questions concerning classic EO data and tools. However, for the moment AI technologies are a kind of black box, we might have difficulties explaining the results provided by such a system which might lead to a lack of trust on behalf of our clients."* Consumer attitudes are key to VISTA's long-term success.

5. What are the technical barriers for the adoption?

VISTA has already worked with a lot of EO platforms such as private platforms (e.g. Planet), DIASes (e.g. CREODIAS or MUNDI) and toolbox platform (e.g. SNAP), but has not tested many external AI tools. Based on this experience, the technical barriers or successes are:

- *Accessibility of EO data*. It is something understood and quite intuitive;
- *Understanding of EO data*. VISTA works in this domain since its creation, so many different types of EO data (e.g. multispectral, hyperspectral, radar) are well-understood, even if they know that it is not easy for non-expert end-users. Moreover, when the processing level of data is rising, it is getting more and more complex, above all to understand which processes exactly have been applied; to have full control over this, VISTA develops and utilizes very complex in-house tools and processing chains.
- *Integration of tools (not AI).* It requires a bit of effort at the beginning (e.g., to learn about Docker containers), but it is not complex;

- *Scalability of cloud platform*. It is possible to run thousands processes in parallel;
- *Security*. European cloud providers are preferred due to GDPR reasons. Results (e.g. pre-processing products) are downloaded in order not to have them in public clouds due to effectiveness; some products are provided via clouds;
- *Multiplicity of AI tools*. There are a lot of tools, types and examples, which makes the domain really complex to understand;
- *Availability of datasets* to train accurate and robust AI models. These datasets do not exist for agriculture and it is a huge effort to create them;
- *Explainability and Robustness of AI tools*. For two similar areas, the neural network model can possibly provide accurate results for one, and less accurate ones for the other. The reason is often not clear, contrary to physical models.

6. Possible takeaways for the AI4Copernicus project

- *Understanding and trusting new technology* is essential for the adoption of AI tools in the EO space, both for the users inside the organisation and for those taking decisions outside the organisation based on the recommendation of these systems,
- The *internal development of tools and the use of free data sources* is a cost-effective choice especially helping SMEs to be competitive in the EO space. Similar practices can be expected in the case of AI solutions especially on behalf of SMEs,
- *Awareness raising and evangelisation* concerning data-intensive technologies is essential and can foster the adoption of technologies among potential future users. For instance, SMEs need help in understanding what can be achieved with AI solutions and what are the main use cases concerning EO data,
- *SMEs might have difficulties* in implementing new use cases and adopting new data-intensive technologies compared to larger organisations with considerable resources and significant knowledge base,
- In the agricultural sector, it is challenging to build databases to train algorithms. Use cases can be specific for several reasons, this makes it difficult to develop “one-type-fits all” solutions.

We thank *Florian Appel* for his valuable contribution and participation in the interview on behalf of VISTA.

8.2.6 Terrawatch interview

Terrawatch interview

Michelle Aubrun (Thales Alenia Space)

What is your profile ?

Mr. Aravind comes from the software industry and has joined the space domain 6 years ago in around 2016. Since then, he works on Earth Observation (EO) projects for ESA, European Commission and Space companies, and he is specialized on the business and the strategy of the space domain. In parallel, he is adviser and writer at Terrawatch.

In which domain is it value-added to use EO data?

It is a huge market, there is a lot of sectors where the use of EO data can have benefits, but the three main sectors are:

- Defense & Intelligence;
- Agriculture;
- Assurance.

How EO data can create business in these domains?

The two main reasons to use EO data are:

- Scalability – EO data are really big so it is easy to look at different scales, unlike IoT *(Internet of Things)* sensors (e.g. fields in one region for crop resources, one field for crop yield or inside one field for soil moisture);
- Easy to integrate in a processing chain – Do not need to send humans somewhere and understand its notes (e.g. count the number of buildings with a risk for flood events).

Who are the main users of EO data?

The main users of EO data are the defense and small companies. But bigger companies using it more and more. In most of the cases, small companies use EO data, develop an EO tools and sell it to big companies, such as Climate Corporation bought by Bayer or Geosys with UrtheCast.

What is the EO data the most used?

Outside defense domain where the most used data are MAXAR in the United-States and PLEIADE in the Europe, the most used EO data are open data because they are free and easy to access.

Some sectors are getting more and more interested by SAR and multispectral data. According to the use case, it is better to use very high-resolution data or low-resolution data.

What is the main barriers to adopt EO data?

Lots of EO data and IA tools are open source and there are lots of EO platforms, so the barriers to entry in the space domain are really low for about 5 years.

However, there are two main barriers:

- Use-ability – Most of the companies (outside the spatial domain) are not aware about the different types of sensor and why they are used for, so they do not know what can be achieved with the EO data. Moreover, there are lots of EO data and platforms, what it is quite confusing for the end users, who need to invest time and do not where to start, but most of the time prefer give up. Finally, it is missing people who have the dual competence in space domain and another domain where EO data can help;
- Fuse-ability – It is really complex to know how to fuse the different types of EO or no-EO data in order to make something useful.

Among the companies that use EO data, what is the proportion that use IA tools?

There is a minority of companies that use IA tools. Indeed, most of the companies using EO data are startups that do not know exactly what their use cases is, and if it can be solved and if the customers are ready to pay for their solution, so it is a risk.

The development of an IA tool is costly and time-consuming. It requires:

- To buy training data which can be expensive if it is very high-resolution data. Moreover, robust training data is hard to find or complex to build;
- To put someone during 3/4 months to build a solution and train a model.

If the EO companies have few customers, they can begin to invest in AI technology and they have the skill internally for that.

We thank *Aravind* for his valuable contribution and participation in the interview.

8.3 Annex C: AI4Copernicus and AI-on-Demand platform services

In this annex we list the details of all services shown in Section 6.1, in order to support the definition of the relevant requirements columns in the services tables. The details are closely related to the classification of the technical characteristics of the proposals shown in Section 4.1. We list a different set of details of each service, depending on its scope.

Table 18 : Details for each service, according to its scope

Thematic services	Bootstrapping services	General services
-	-	Research area
Targeted Domain Problem	Targeted Domain	Domain
Application types	-	-
Data sources	Data sources	-
Geographic scope	Geographic scope	-
Core AI challenges	Core AI challenges	Technical area
Asset Type	Asset Type	Asset Type

In the following table we list the details of all bootstrapping services of the AI4Copernicus project.

Table 19 : Details all bootstrapping services offered by AI4Copernicus

ID	Domain	Data Sources	Geographic Scope	Core AI Challenges	Asset Type
A01	Security, General	Satellite data, Soil data	Global	Image processing, Fine-tuning	Docker Container
A02	Security, General	Satellite data, Soil data	Global	Image processing, Fine-tuning	Docker Container
A03	Security, General	Satellite data, Soil data	Global	Image processing, Fine-tuning	Docker Container
A04	Security, General	Satellite data, Soil data	Global	Image processing	Docker Container

A05	Security, General	Satellite data, Soil data	Global	Image processing	Docker Container
A06	Security, General	Soil data	Regional	Image processing, Fine-tuning	As-a-Service
A07	Agriculture, General	Soil data	Regional	Image processing	Docker Container
A08	Agriculture	Soil data	Regional	-	Dataset
A09	Agriculture	Soil data	Regional	Image processing	Docker Container
A10	Agriculture	Soil data	Regional	Image processing	Docker Container
A11	Agriculture	Soil data	Regional	Image processing	Docker Container
A12	Energy	Weather data, Socio-economic data	Regional	-	Dataset
A13	Health	Weather data	Global	Image processing, Data downscaling	Docker Container

In the following table we list the services that comprise the Semantic Web toolkit of the AI4Copernicus project. These services are classified as of "general purpose".

Table 20 : Details all the Semantic Web toolkit of AI4Copernicus

ID	Research Area	Technical Area	Application Domain	Asset Type
B01	Integrative AI	Knowledge Representation, Semantic web	Earth Observation	Executable
B02	Integrative AI	Knowledge Representation, Semantic web	Earth Observation	Executable
B03	Integrative AI	Knowledge Representation, Semantic web	Earth Observation	Docker Container
B04	Integrative AI	Knowledge Representation, Semantic web	Earth Observation	Executable
B05	Integrative AI	Knowledge Representation, Semantic web	Earth Observation	Executable

In the following table we list all relevant thematic services that can be currently found in the AI-on-demand platform. These services have targeted domain problems that are similar to the thematic services to be implemented by the open call winners.

Table 21 : Details all the thematic services found in the AI-on-demand platform

ID	Domain Problem	Application Types	Data Sources	Geographic Scope	Core AI Challenges	Asset Type
C01	Urban conditions monitoring, Entity recognition	Predictive analytics, Assessment frameworks	Socio-economic data	Local	Image processing	Docker Container
C02	Land Monitoring, Food Security	Predictive analytics, Decision support	Soil data	Local	-	Dataset
C03	Land Monitoring, Food Security, Entity recognition	Predictive analytics, Decision support	Soil data	-	-	Docker Container
C04	Entity recognition	Predictive analytics, Decision support	Soil data	Local	Image processing	Docker Container
C05	Land Monitoring, Entity recognition	Predictive analytics, Decision support	Soil data	Local	-	Dataset
C06	Land Monitoring, Food Security	Predictive analytics, Decision support	Soil data	Local	Image processing, Data downscaling	Docker Container
C07	Urban conditions monitoring, Environmental Impact Assessment	Predictive analytics, Assessment frameworks	Weather data	-	-	Docker Container
C08	Urban conditions monitoring	Predictive analytics, Assessment frameworks	Weather data	Local	Automated configuration & fine-tuning	Docker Container
C09	Urban conditions monitoring, Environment, Entity recognition	Predictive analytics, Assessment frameworks	Weather data	Local	-	Dataset
C10	Urban conditions monitoring, Entity recognition	Predictive analytics, Assessment frameworks	Socio-economic data	Local	-	Dataset
C11	Urban conditions monitoring, Environmental Impact Assessment	Aggregate dashboards	Weather data	Local	-	Dataset
C12	Production optimization	Predictive analytics, Decision support	Socio-economic	Local	Multi-factor prediction	ML Model

			data			
C13	Production optimization	Predictive analytics, Decision support	Socio-economic data	Local	-	Dataset
C14	Urban conditions monitoring, Environmental Impact Assessment, Entity recognition	Predictive analytics, Assessment frameworks	Socio-economic data	Local	Multi-factor prediction	Docker Container
C15	Land Monitoring, Food Security	Predictive analytics, Decision support	Soil data	Local	Multi-factor prediction	Dataset

In this table we list all general-purpose services that can be currently found in the AI-on-demand platform. These services are not always directly linked with the services to be implemented by the open call winners, since they provide research tools of abstract domains. We add them since they may be used by domain specific tools or AI pipelines in services relevant to the context of AI4Copernicus.

Table 22 : Details all the general purpose services found in the AI-on-demand platform

ID	Research Area	Technical Area	Application Domain	Asset Type
D01	-	Planning automation and support	-	Library
D02	-	Planning automation and support	-	Library
D03	-	Planning automation and support	-	Library
D04	-	Planning automation and support	-	Library
E01	Explainable AI	Machine Learning, AI Ethics	-	Library
E02	-	AI Services	Agriculture	As-a-Service
E03	Integrative AI	AI Services	-	Library
E04	Explainable AI	Machine Learning	-	Docker Container
E05	Integrative AI	Knowledge Representation and Reasoning	-	Executable
E06	Explainable AI, Verifiable AI	Knowledge Representation and Reasoning, Searching, Planning	-	Docker Container
E07	-	Knowledge Representation and Reasoning	-	Executable
E08	Physical AI	Machine Learning, Robotics	-	Library
E09	Physical AI	Machine Learning	-	Library
E10	Integrative AI	Machine Learning, Constraints	-	Library

E11	Integrative AI	Knowledge Representation and Reasoning, Constraints	-	Docker Container
E12	-	Knowledge Representation and Reasoning, Constraints	-	Executable
E13	-	Machine Learning	-	Library
E14	Integrative AI	Knowledge Representation and Reasoning, Machine Learning, Constraints	-	Docker Container
E15	-	Knowledge Representation and Reasoning	-	Dataset
E16	Explainable AI	Machine Learning	-	Docker Container
E17	Explainable AI	Machine Learning	Earth Observation	Docker Container
E18	Collaborative AI, Verifiable AI	Machine Learning, AI Ethics	Cloud, Edge, and Infrastructure	Jupyter Notebook
E19	Integrative AI	Knowledge Representation and Reasoning, Machine Learning, Constraints	-	Library
E20	Explainable AI	Machine Learning, AI Ethics	Earth Observation	Jupyter Notebook
E21	Integrative AI	Machine Learning, Knowledge Representation and Reasoning	-	Docker Container
E22	Integrative AI	Computer Vision, Machine Learning	-	ML Model
E23	Integrative AI	Constraints	-	Executable
E24	Integrative AI	Constraints, Machine Learning	-	Library
E25	Explainable AI	Knowledge Representation and Reasoning, Machine Learning	-	Executable
E26	Integrative AI	Knowledge Representation and Reasoning, Multi-agent systems	Public services	Library
E27	Integrative AI	AI Services	-	Library
E28	Integrative AI	Machine Learning	-	Library
E29	Collaborative AI	Knowledge Representation and Reasoning	Manufacturing	Executable
E30	Physical AI	Machine Learning, Robotics	-	As-a-Service
E31	-	Constraints	Manufacturing	Docker Container
E32	-	Knowledge Representation and Reasoning	-	Executable
E33	-	AI Ethics, Machine Learning	Cloud, Edge, and Infrastructure	Jupyter Notebook
E34	Verifiable AI	-	Healthcare	Docker Container
E35	Integrative AI	Computer Vision, Machine Learning	-	ML Model

E36	-	Constraints, Optimization	Cloud, Edge, and Infrastructure	Docker Container

------End of Document------