Earth Science Collaborative Open Development Environment

EarthCODE

Industry Information Day

Dr Chandra Taposeea-Fisher – Telespazio/Project Manager
Garin Smith – Telespazio/Technical Lead
Dr Anca Anghelea – ESA/Technical Officer

Virtual

Wednesday 17th April 2024

EarthCODE Industry Information Day

Agenda

- Procurement Introduction
- Overview of Work Streams and Reference Architecture
- Approach to Selection
- Intended ITT Documentation
- Contractual / Financial Aspects
- Intended ITT Schedule

There will be no individual briefings or questions & answers

EarthCODE Industry Information Day

Agenda

- **Procurement Introduction**
- Overview of Work Streams and Reference Architecture
- Approach to Selection
- Intended ITT Documentation
- Contractual / Financial Aspects
- Intended ITT Schedule

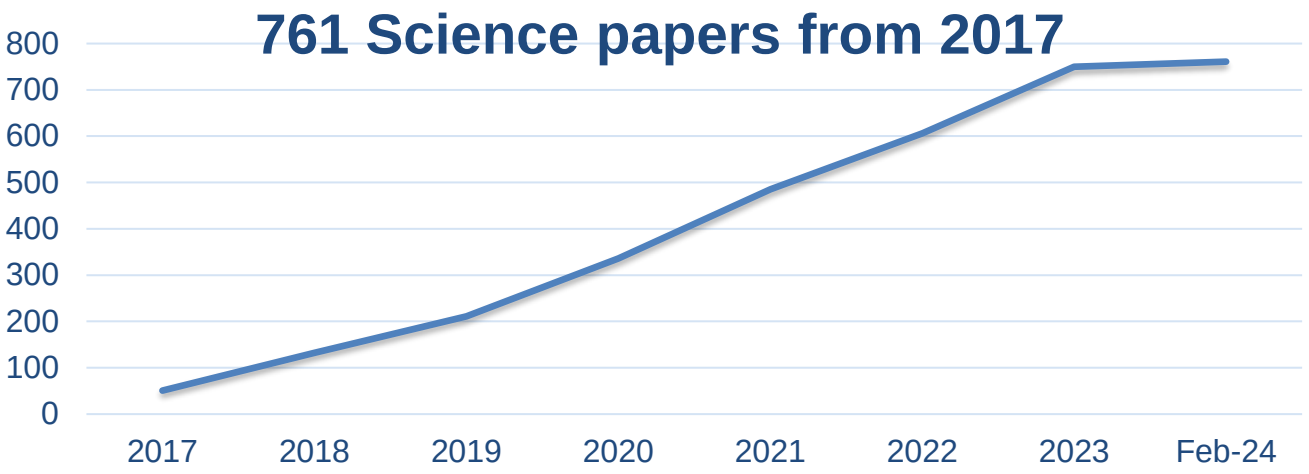
Earth Science Collaborative Open Development Environment

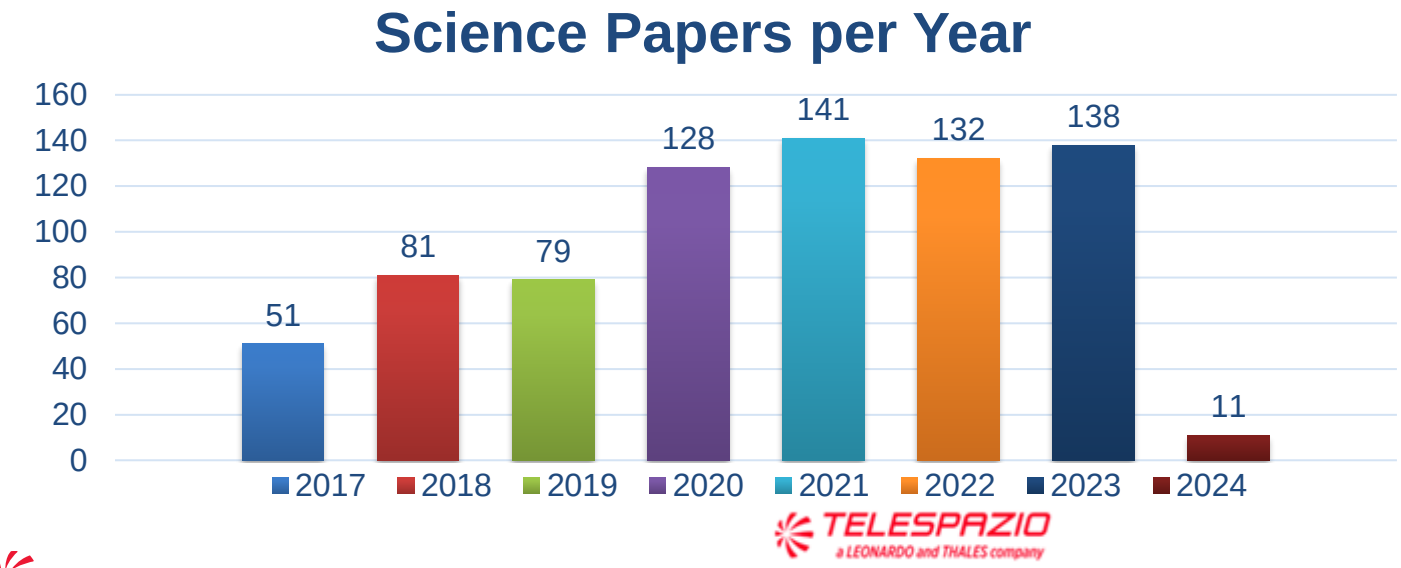
EarthCODE

Common Architecture

Dr Anca Anghelea

Science Clusters & Scientific output





TELESPAZIO
a LEONARDO and THALES company

Atmosphere Science Cluster

Ocean Science Cluster

Polar Science Cluster

Carbon Science Cluster

Water Cycle Science

Biodiversity

Extremes and natural disasters

Land and agriculture

Solid & magnetic Earth

FutureEO Independent Science Review 2022 recommendations:

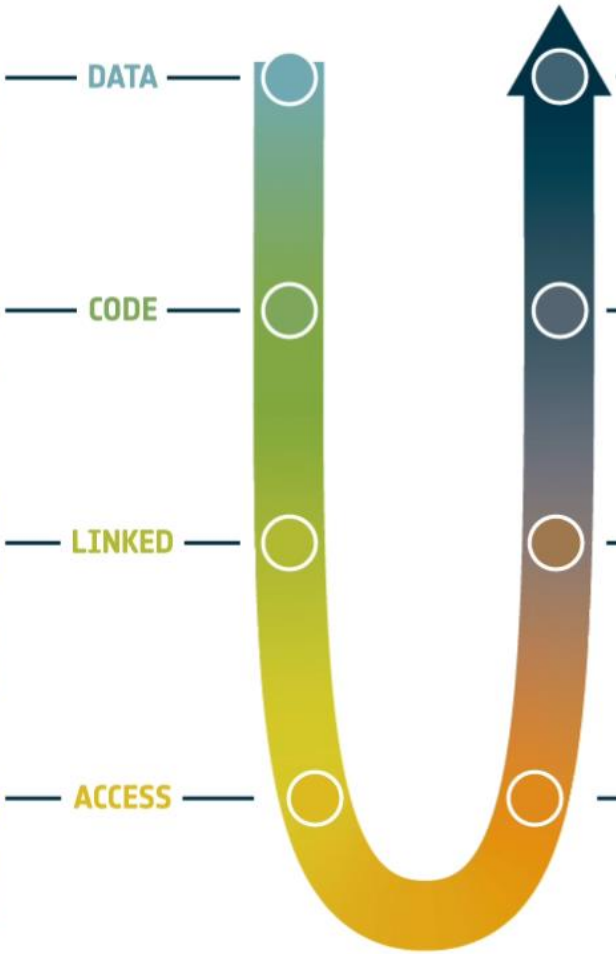
more visibility, discoverability and useability of science results, leveraging interoperability and Open Science Tools, and improving communication and community building.

Scientific community feedback at Science Strategy workshop (2023):

ESA to support FAIR and reproducible Open Science practices and complementing the scientific process by Software Development best practice.

EarthCODE responds to these recommendations

EARTH OBSERVATION OPEN SCIENCE AND INNOVATION



Open Data — DATA

Input and output scientific project data

Open-Source code — CODE

Code developed in Earth Science Projects

Linked and executable code & data — LINKED

Data & code used and produced in scientific projects

Open Access — ACCESS

End-to-end open access Workflows and Documentation

COMMUNITY — Community

Open Competitions, Citizen science, new business models, open innovation

PRACTICE — Open Science Practice

Shared best practice across ESA-funded Science and Innovation activities

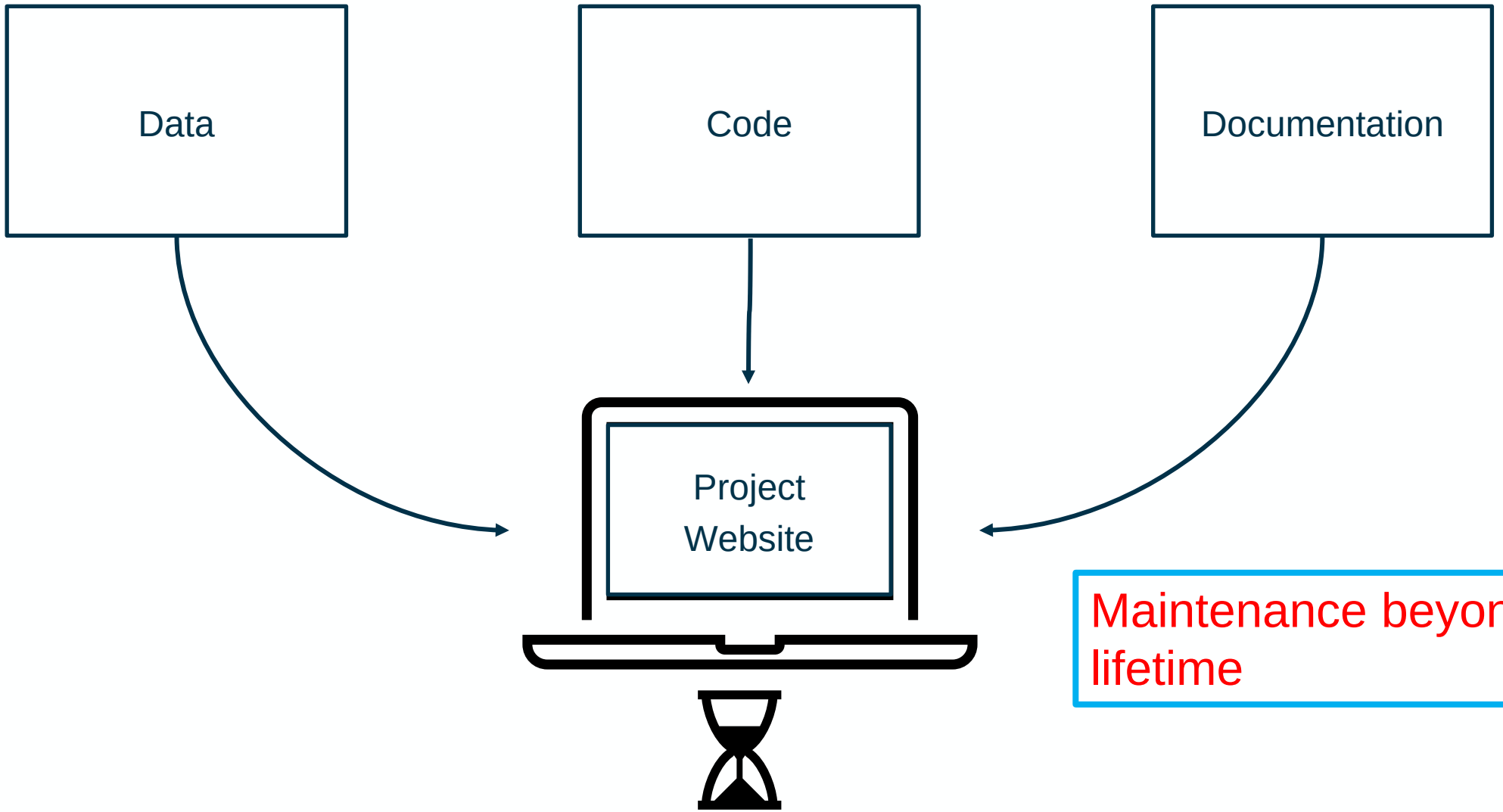
EDUCATION — Open Science Education

Education and best prectices for Open Science and Open Source

FAIR — Reproducible on Platforms

End-to-end Reproducible Workflows on various infrastructures

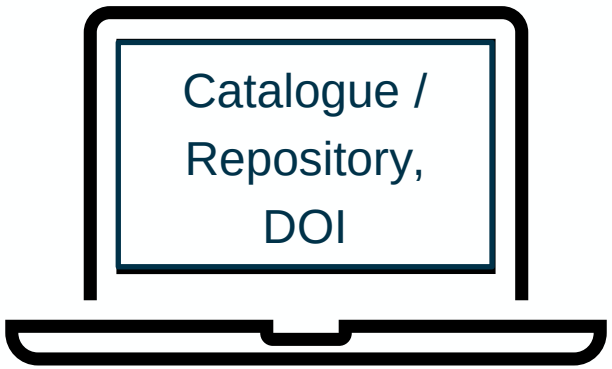
Challenge 1: Maintenance



Maintenance beyond project lifetime

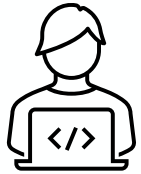
Challenge 2: FAIR

Data



Non-standard formats & metadata, various means of access to products

Code



GitHub

GitLab

Incomplete, broken, undocumented, inaccessible

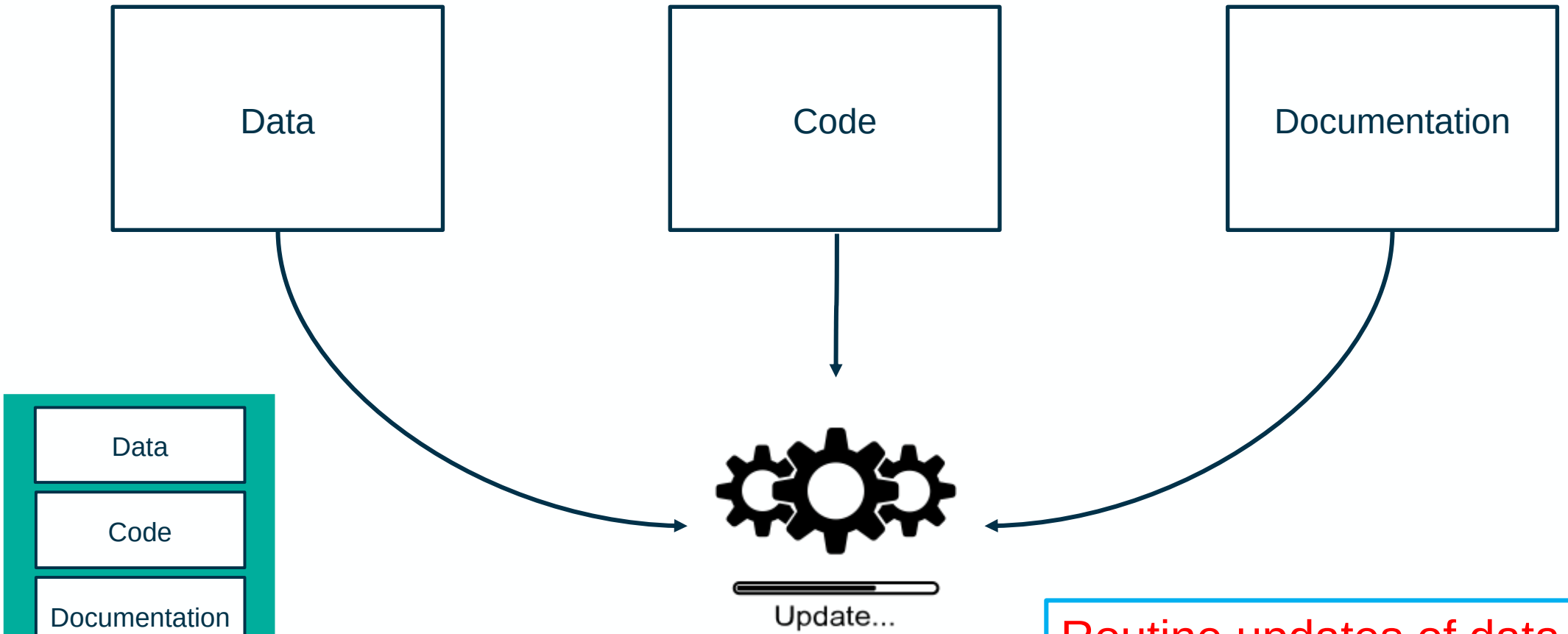
Documentation

OPEN ACCESS

Not linked to data & code

Discoverability, Accessibility and (Re)Useability

Challenge 3: Production cost



Data
Code
Documentation

RESEARCH DATA

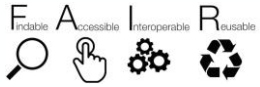
One-off production and delivery

Routine updates of data products

Objectives

Enable adoption of FAIR Open Science Principles throughout ESA-funded Earth System Science activities, to deliver long-term persistence of research data, aiding to its reproducibility, reuse and consumption by a wider community.

EarthCODE brings existing pieces together in a single, open-access solution for ESA scientific activities:



ESA

Open access cloud-based development environment leveraging on federated EO platforms, with:

- Scalable computing, tools for FAIR management of open-source research data
- Guidelines, community management, and support for FAIR Open Science including for activities that use own institutional computing resources
- Persistent storage, cataloguing and discovery services for research data

EarthCODE development and evolution

- EarthCODE is developed with existing platforms and tools
- It integrates relevant services, does not build new ones
- Service providers are engaged to develop new features according to science needs
- Building blocks are provided by EOEPCA+, services are provided by platforms from NoR
- Makes use of APEX Services (e.g. algorithm optimisation, code maturation, etc.)

EarthCODE in relation to other intiatives

EOEPCA+	*Open-source building blocks*	Network of Resources	*Platform Services*	Science Clusters	*Scientific users*
APEX	*Code maturation services*	Open Science Catalogue	*Catalogue services*	ESA Cloud	*Research data hosting*

Utilisation scenarios

	SCIENCE	PUBLISH	INTERACT
DEVELOPER / TECH	*Integrate Platform Services*	STAC *Develop Open Science tools*	*Develop community tools*
SCIENTIST - CONTRIBUTOR	*Compute on Platform*	STAC *Publish in Catalogue*	*Interact with community*
SCIENTIST - USER	STAC *Discover in Catalogue*	*(Re)Use on Platform*	*Interact with community*

Contractual Implementation

- **Prime contract: (2+3 years): Telespazio UK, EOX (AU) - KO Dec 2023**
  - Team coordination and management
  - Promotion, engagement and science communication

- **ESA Best Practices for Work Stream Development**
  - Possible second Best Practices for the selection of subcontractors in 2025
  - One or more subcontractors possible for the same Work Stream

EarthCODE

- How to Open Science? - Learning resources
- Development of Reproducible scientific workflows
- Community and collaborative tools
- Federated EO platform services for scientific development
- Catalogue & Repository services for FAIR & Open
  - Data
  - Code
  - Documentation
- Visualisation tools

Open Source as Open Governance

- **Adoption Oriented**
  - System architecture based on standards or community adoptions
  - Building Blocks design based on Decoupled Components
  - Public or joint open source development for each component
  - Based on FAIR and Open Science principles

- **Community Building**
  - Open Geospatial Consortium (OGC) for international promotion of the european strategy
  - Open Source Geospatial Foundation (OSGeo) to promote a more inclusive and community-oriented open source strategy
  - Shaping the promotion/collaboration with a larger audience
    - CDSE, DestinE, EC projects, GAIA-X, MS national assets

EO OPEN SCIENCE

EarthCODE Industry Information Day

Agenda

- Procurement Introduction
- **Overview of Work Streams and Reference Architecture**
- Approach to Selection
- Intended ITT Documentation
- Contractual / Financial Aspects
- Intended ITT Schedule

EarthCODE Vision

Objectives

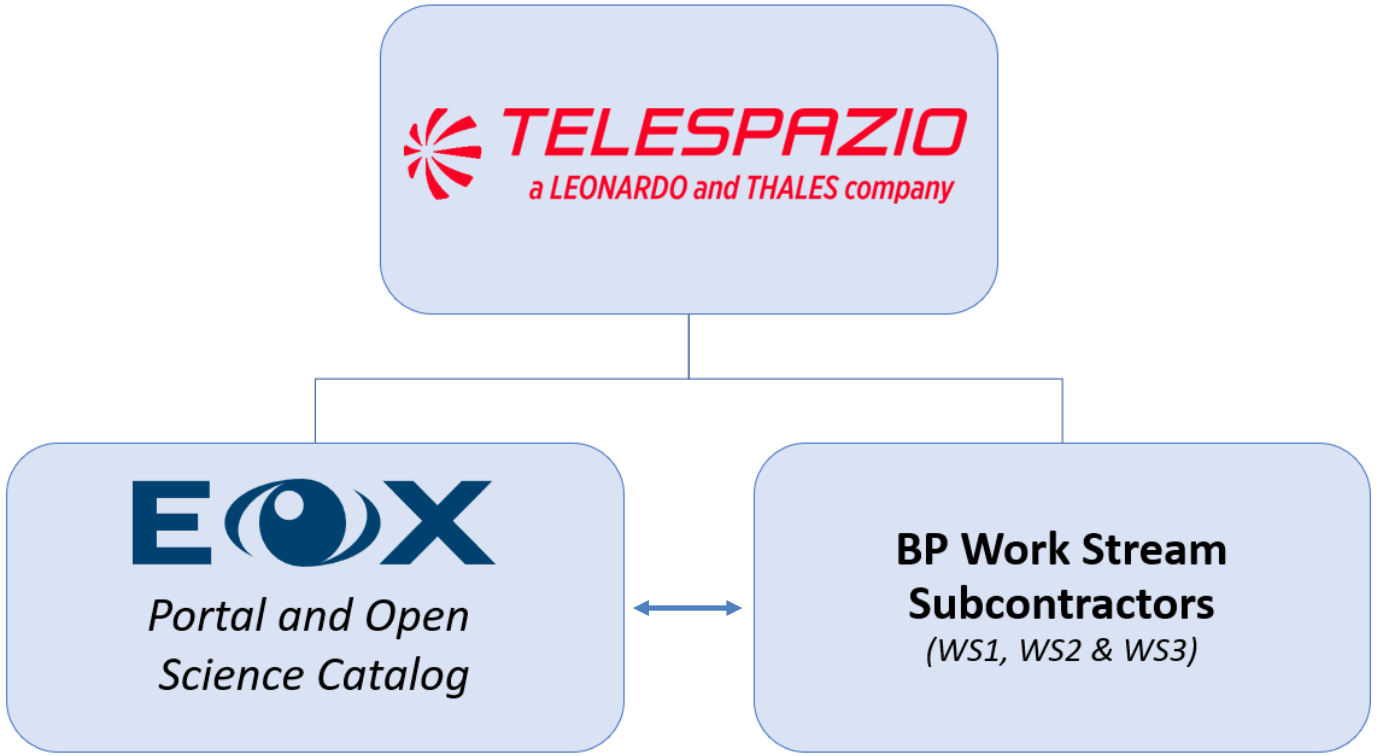
- Enable adoption of FAIR Open Science Principles throughout ESA-funded Earth System Science activities
- Re-use of existing systems
- Deliver long-term persistence of data, code and documentation
- Aid reproducibility, reuse and consumption of research outputs by a wider community

Open access cloud-based development environment

- Scalable computing
- Tools for FAIR management of open data, open-source code, documentation
- Guidelines, community management, and support to use the tools and apply the principles in practice
- Persistent storage, cataloguing and discovery services for the activities' research outputs

Overall, we want to help users create scientific workflows for their experiments and publish them to a repository after they have been verified, so that experiments can seamlessly be re-used with the specified data and necessary infrastructure according to FAIR Open Science Principles, engaging with the relevant communities to further science.

EarthCODE Work Streams (WS)



- WS1: Infrastructure
- WS2: FAIR Open Science
- WS3: Community

EarthCODE Work Streams (WS) - Best Practice Procurement

WS1: Infrastructure

- Support execution of scientific workflow solutions created in WS2
- Provides data needed by workflows and processing capability
- Will have a prime (lead developer) for each platform

WS2: FAIR Open Science

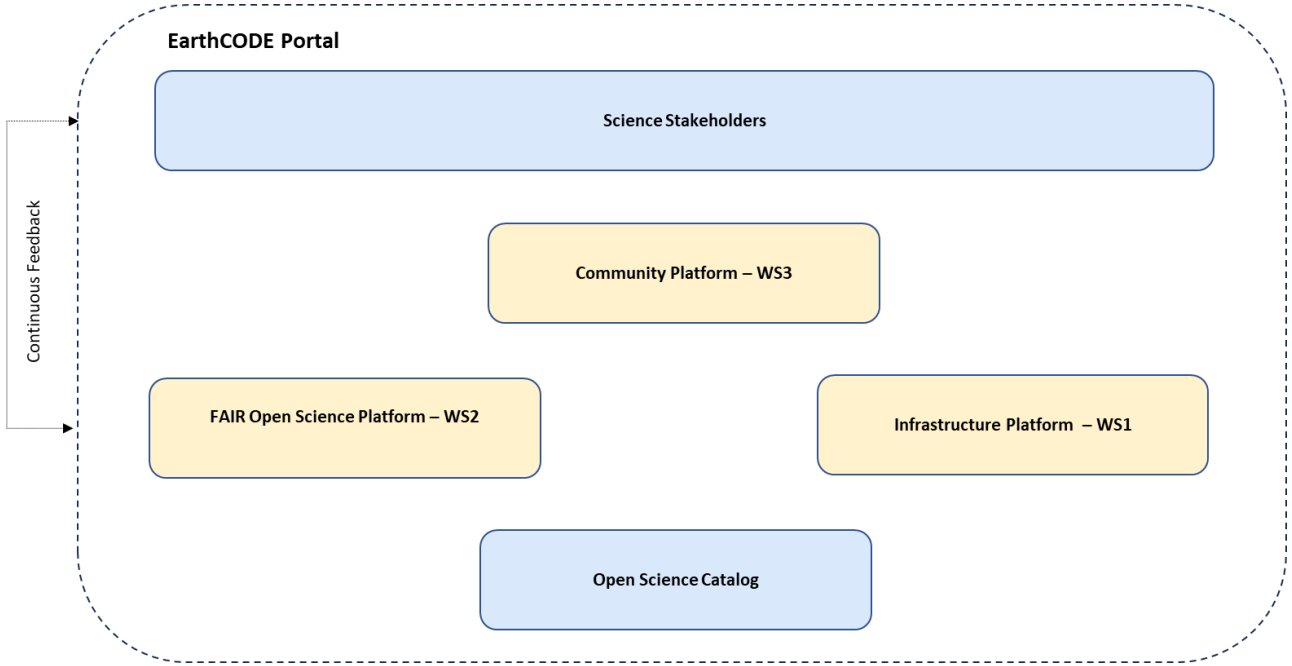
- Support development of workflow solutions to be used by WS1
- Workflow solutions may for instance use OpenEO, OGC API Processes or MLflow (AI Modelling) etc.
- Will have a prime (lead developer) for each platform
- Possible that WS1 and WS2 are implemented by the same consortium but not essential.

WS3: Community

- Provision of community support to use WS1 and WS2
- WS3 could be provided by a different consortium to WS1 or WS2

The above work streams, along with the Open Science Catalog, will be provided through the EarthCODE Portal.

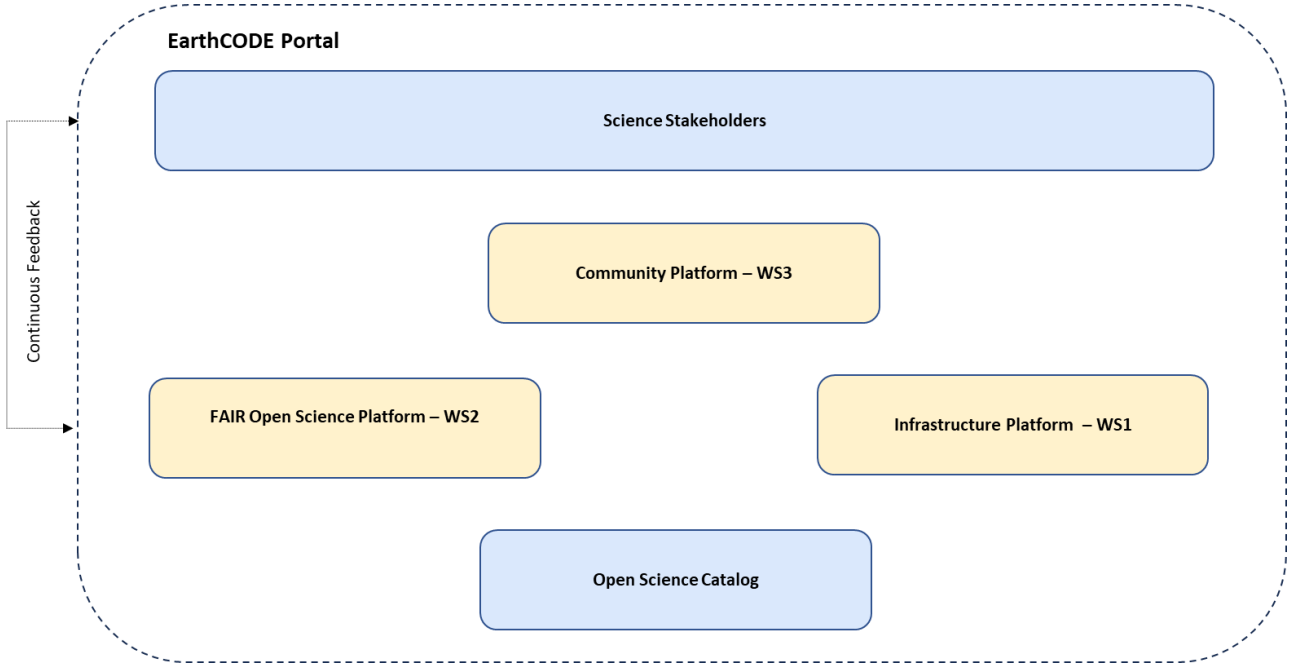
It is possible/likely that there will be more than one platform for WS1 and WS2.



EarthCODE Work Streams (WS) - Best Practice Procurement

Phase 1: First 12 months

- Requirements and Roadmap Refinement
- Platform integration within EarthCODE portal including single sign on
- Publish experiments and experimental output to Open Science Catalog
- Provision operational platforms within EarthCODE portal
- Scrum team for each Work Stream
- Scrum of Scrums Agile approach to support collaboration between Work Streams

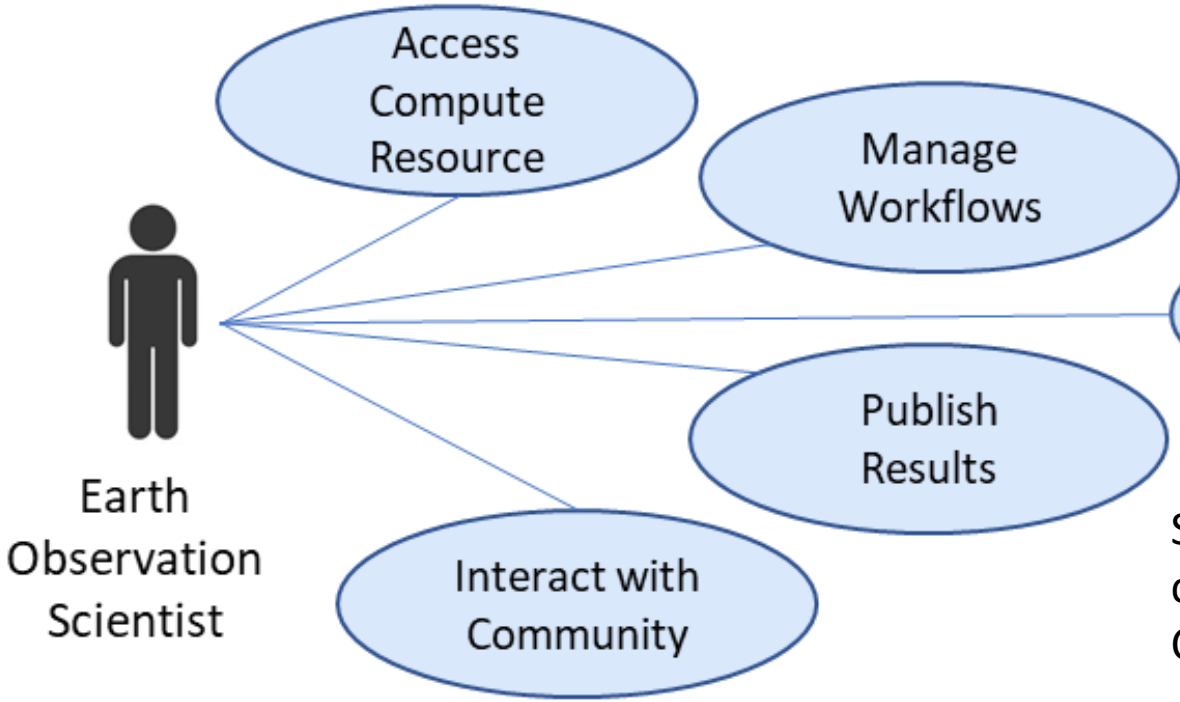


Phase 2:

- Focus on further integration
- Potential integration using EOEPCA+ Building Blocks
- Achieve greater interoperability between platforms

EarthCODE Use Cases

Access to platforms to develop and execute science experiments (workflows)
Work Stream 1, Work Stream 2 and Work Stream 3



Manage development of experimental workflows on a platform
Work Stream 2

Execution workflows on a platform with sufficient compute, near large volumes of data to create products
Work Stream 1

Share workflows and experimental results (products) created on platforms with others using Open Science Catalog. Work Stream 1 and Work Stream 2

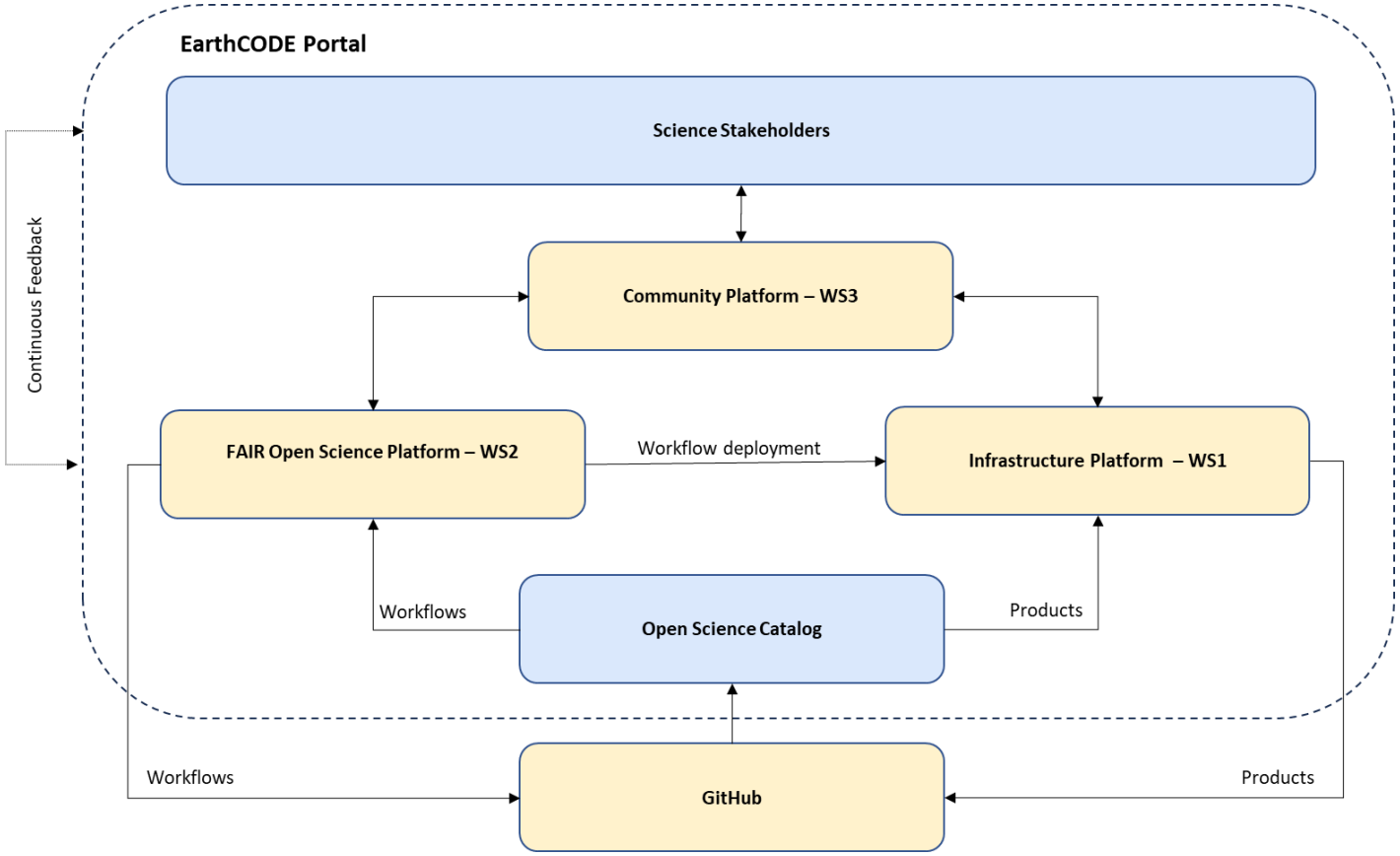
Support use cases above, sharing ideas and providing help to exploit the technical solutions. Work Stream 3

WS1 Infrastructure

WS2 FAIR Open Science

WS3 Community

EarthCODE High-level Architecture – Platform Expectations



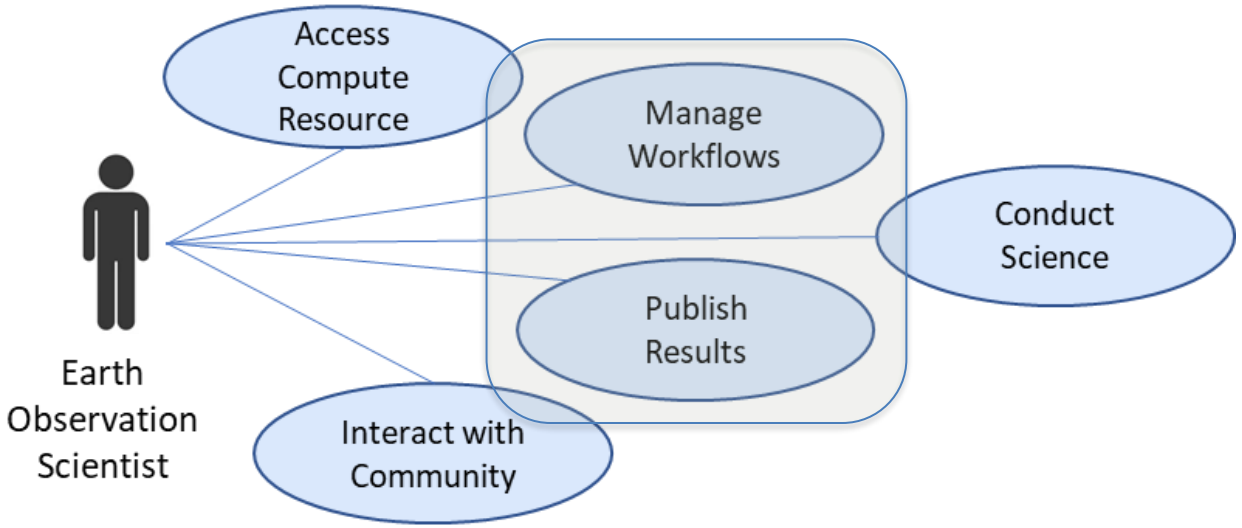
EarthCODE – Use Cases Minimum Viable Product

Manage Workflows

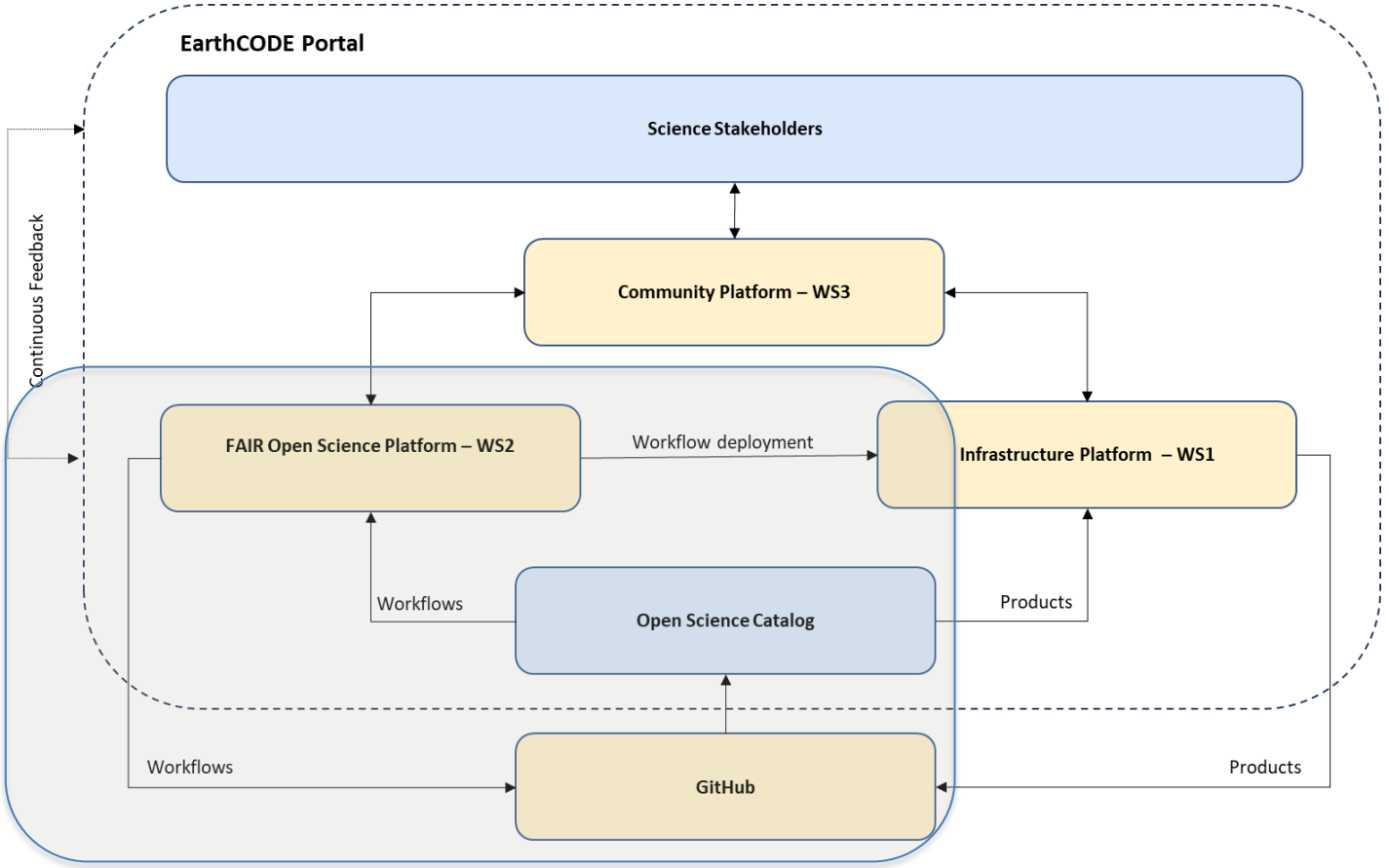
- As a user, I want to develop code for my experiment using my preferred language (E.g. Python, R etc) that can be described using a workflow, so that I can execute it on the cloud.

Publishing

- As a user, I want to publish my workflow (experiment), so that other users can discover it and re-use it.



EarthCODE High-level Architecture – Platform Expectations



- As a user, I want to develop code for my experiment using my preferred language (E.g. Python, R etc) that can be described using a workflow, so that I can execute it on the cloud.
- As a user, I want to publish my workflow (experiment), so that other users can discover it and re-use it.

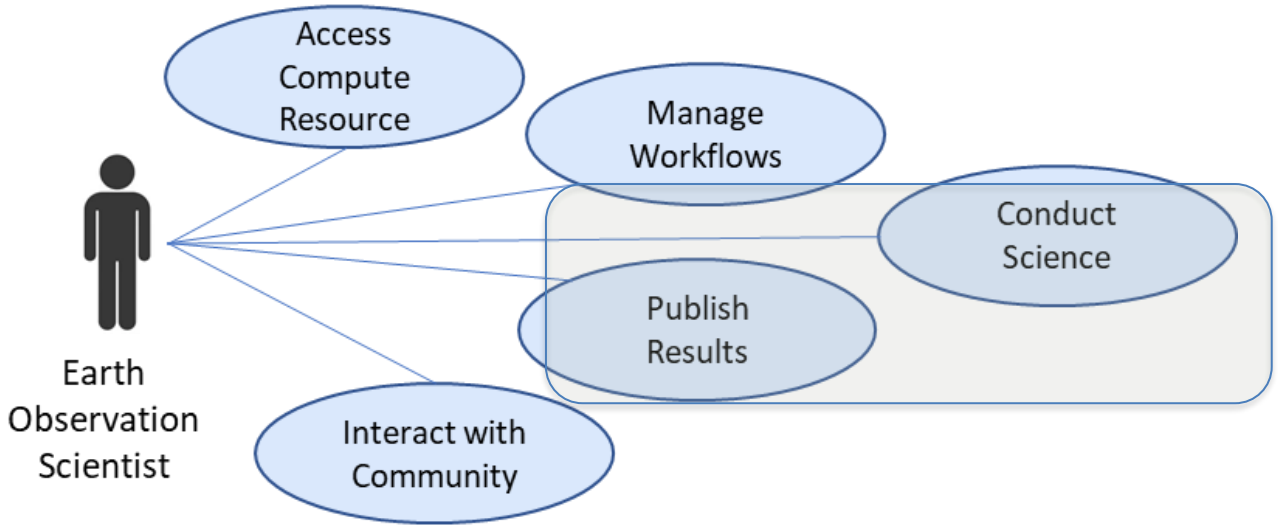
EarthCODE – Use Cases Minimum Viable Product

Conduct Science

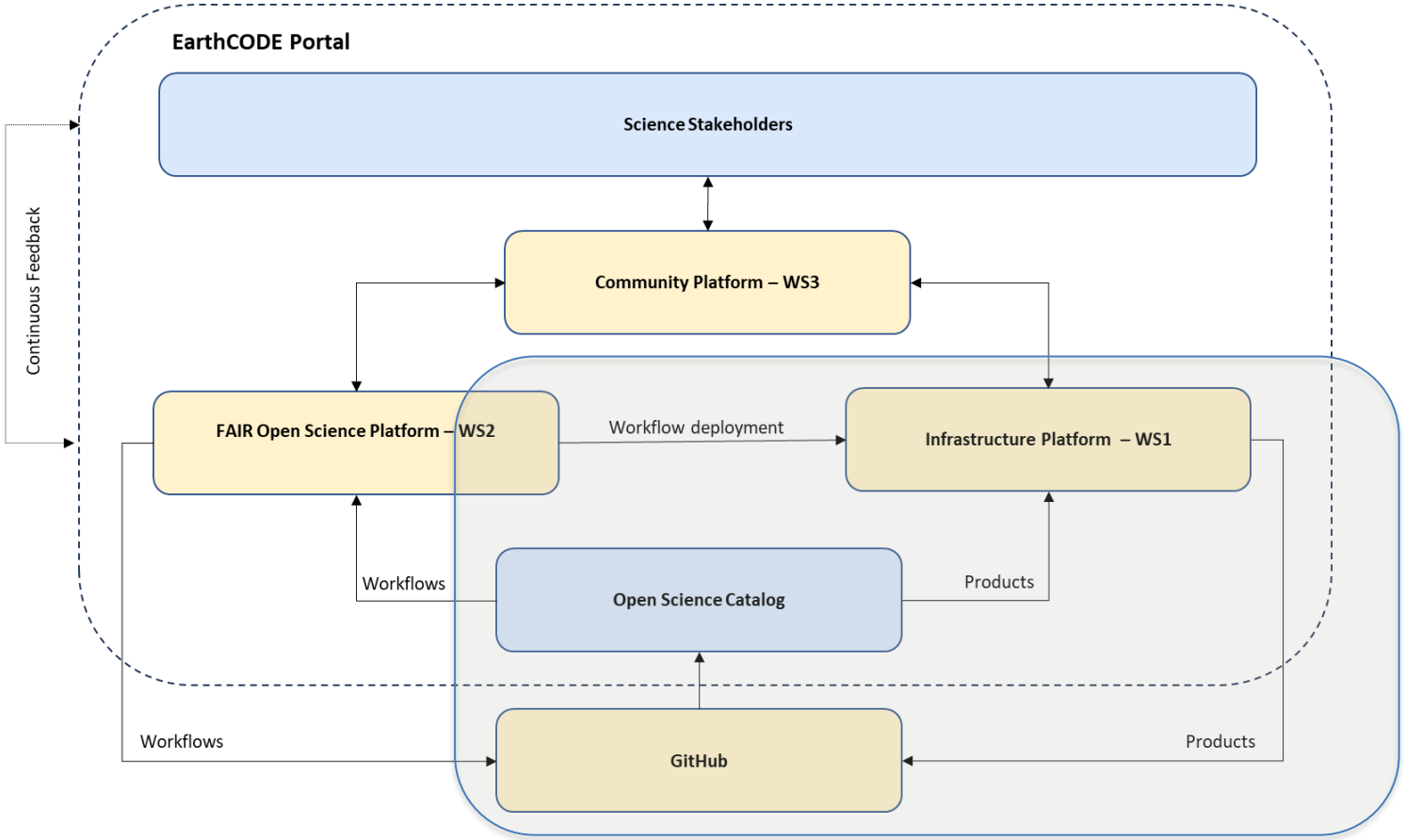
- As a user, I want to execute my workflow on a chosen Infrastructure Platform, so that I can see the results of my experiment on that platform. During execution I want my workflow to get any data that it needs.

Publishing

- As a user, I want to publish the results (product) of my workflow, so that other users can discover it and reproduce the same output.
- As a user, I want to instruct the Infrastructure Platform to copy my experimental output to a Long Term Archive, so that it is preserved.



EarthCODE High-level Architecture – Platform Expectations

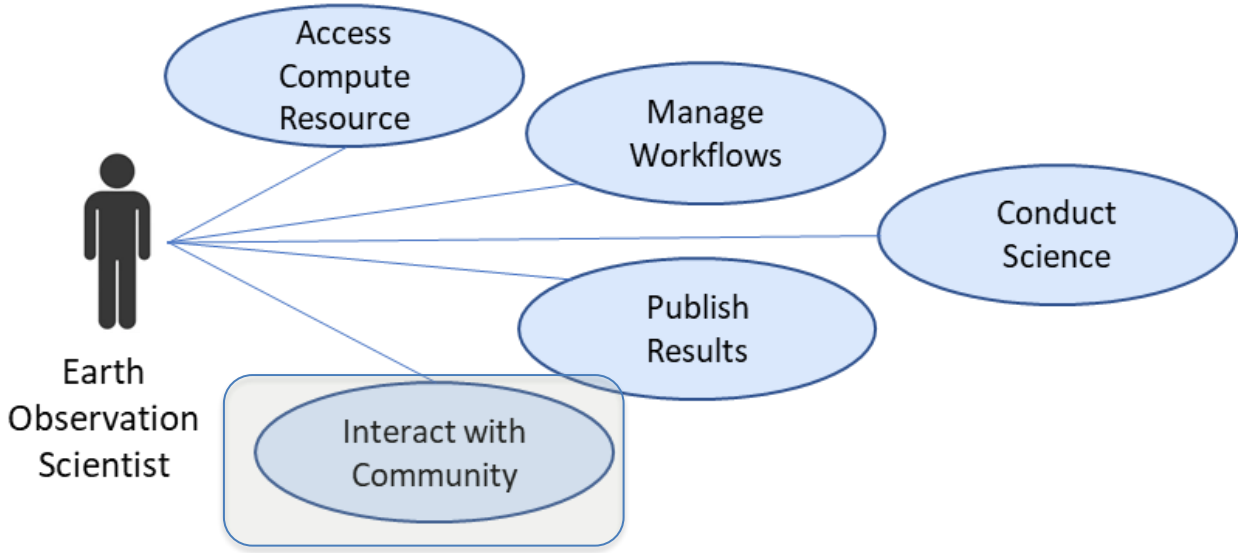


- As a user, I want to execute my workflow on a chosen Infrastructure Platform, so that I can see the results of my experiment on that platform. During execution I want my workflow to get any data that it needs.
- As a user, I want to publish the results (product) of my workflow, so that other users can discover it and reproduce the same output.
- As a user, I want to instruct the Infrastructure Platform to copy my experimental output to a Long Term Archive, so that it is preserved.

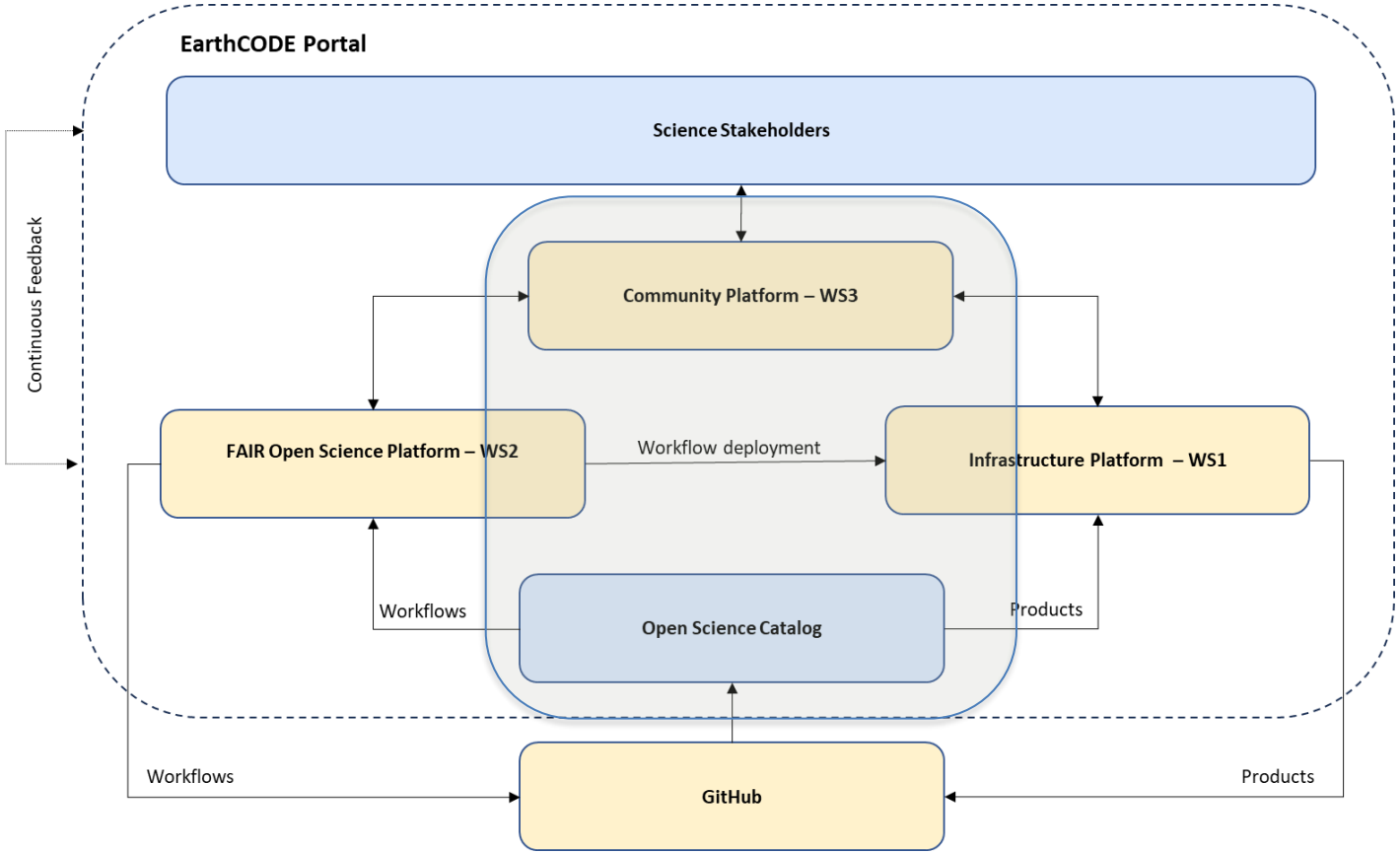
EarthCODE – Use Cases Minimum Viable Product

Interaction With The Community

- As a user, I want to discover an existing experiment in Open Science Catalog, so that I can use it to progress science.
- As a user, I want to help the community find my science, so that we can collaborate.



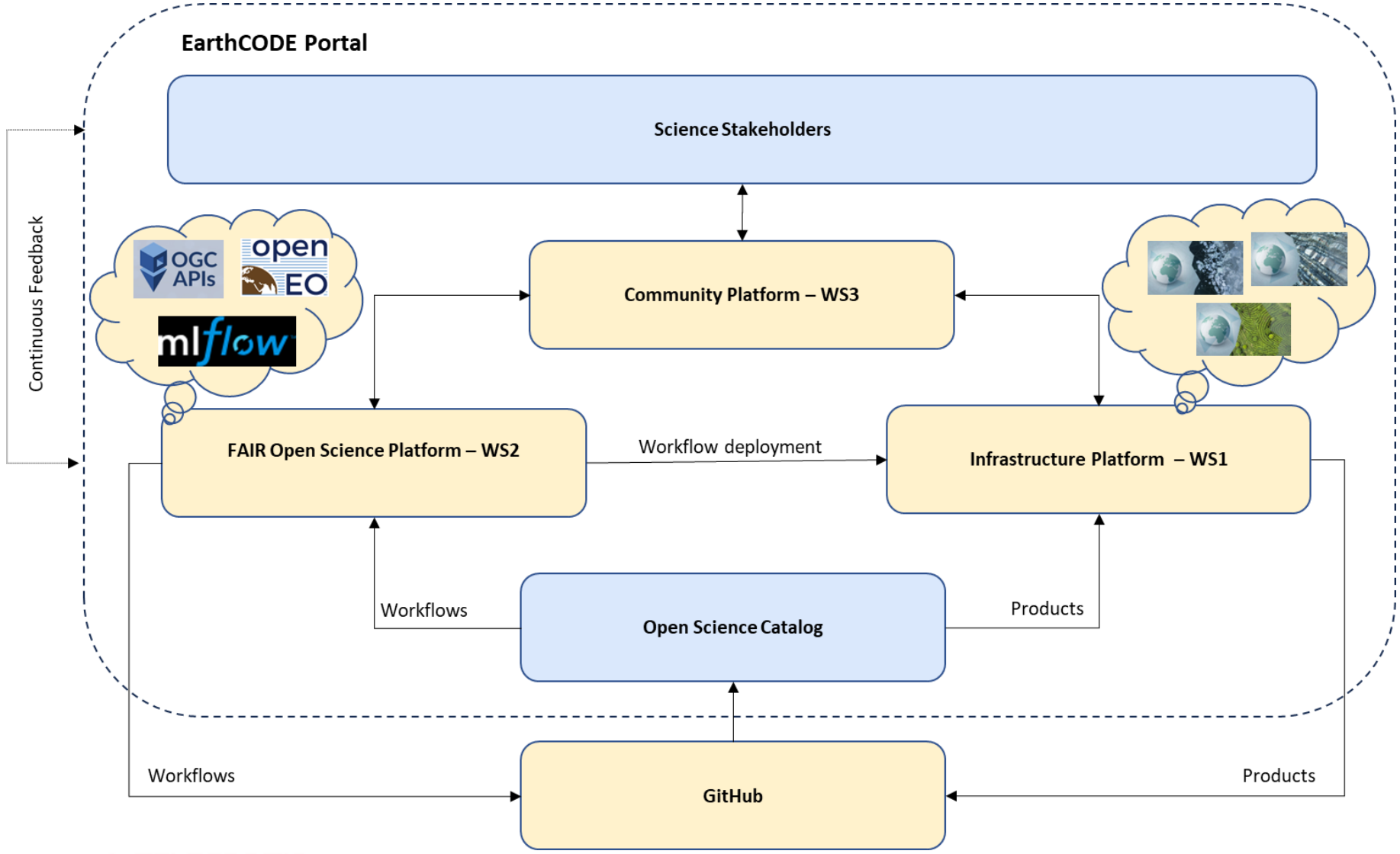
EarthCODE High-level Architecture – Platform Expectations



- As a user, I want to discover an existing experiment in Open Science Catalog, so that I can use it to progress science.
- As a user, I want to help the community find my science, so that we can collaborate.

EarthCODE High-level Architecture – Implementation Approach

- Select one or more platforms that allow Experiments (workflows) to be developed
- Select one or more platforms that allow Workflows to be executed
- Publish products and experiments using Open Science Catalog
- GitHub Single Sign On for integrated user journey
- Expose Service Layers
- Federated Solutions
- Community Standards



EarthCODE WS1 Infrastructure (Execution and Data)

- Data discovery and access to metadata describing that data.
- Data retrieval – cloud-optimised where possible
- Data visualisation
- Access to data through data cube semantics
- Execution of scientific processing algorithms/workflows – including OGC API application packages and openEO
- Training of Machine Learning models
- Execution of Machine Learning models
- Version management service for resources
- Storage services for long-term preservation of resources, including data/results, code and documentation
- Assigning Digital Object Identifiers (DOIs) to resources in long-term storage repositories
- User/project resource management services to support storage and publishing of data
- User/project services for hosting of Interactive Dashboards

EarthCODE WS2 FAIR Open Science (Develop, publish and deploy)

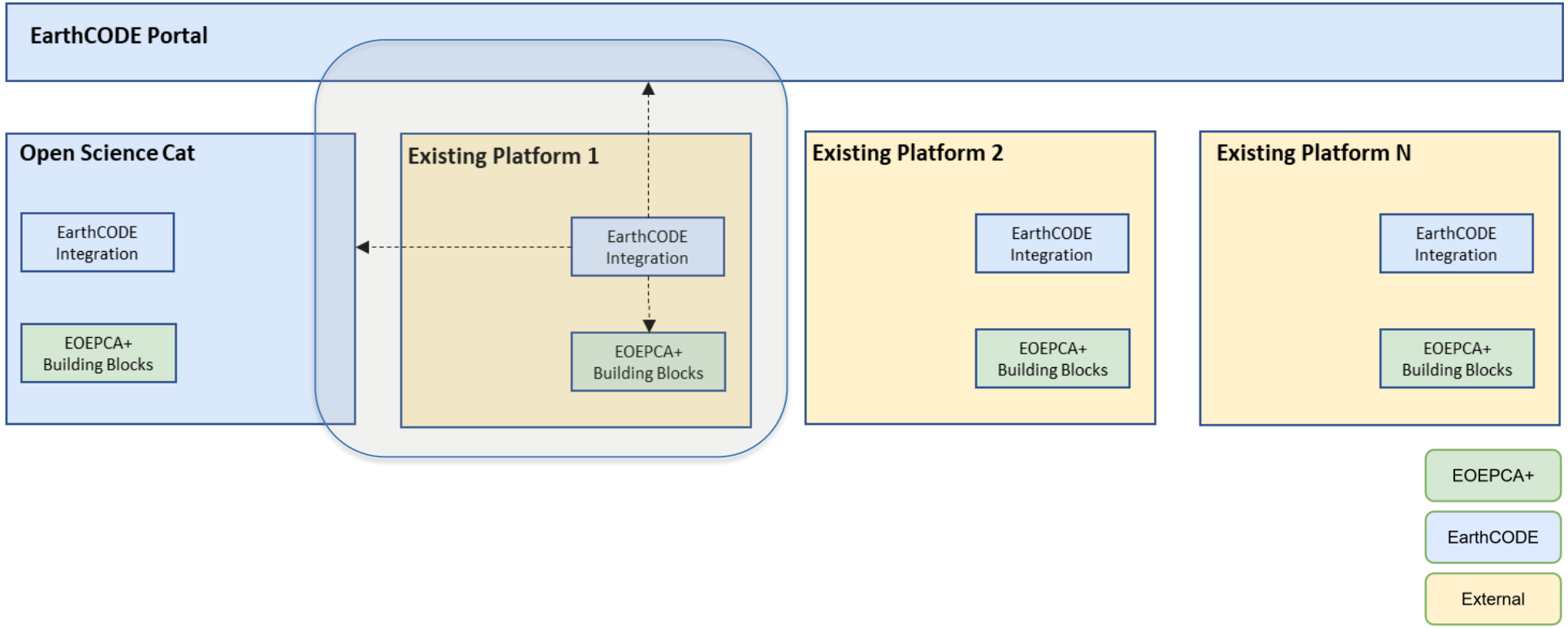
- Interactive graphical tooling to support workflow development
- Authoring/editing/validation of application packages and design of openEO DAGs etc
- Development of Interactive Dashboards to publish research results for convenient consumption
- Management of users to facilitate collaboration and sharing within projects, groups and communities
- Source code management and Release management
- Automated build pipelines and Automated deployment pipelines
- Packaged build artefacts – such as containers and automated code quality tooling
- Automated vulnerability scanning for code and containers
- Documentation – version controlled with automated publishing and Issue tracking
- Access to storage for data/results and publishing services

EarthCODE WS3 Community

- Best practice advice to develop experiments
- Support of FAIR Open Science principles
- Community forums encouraging best practise to describe experiments and products with STAC or other suitable standards.
- Support to publish experiments
- Best Practice advice to enable reproducibility
- ***About community engagement and not primarily a technical Work Stream***

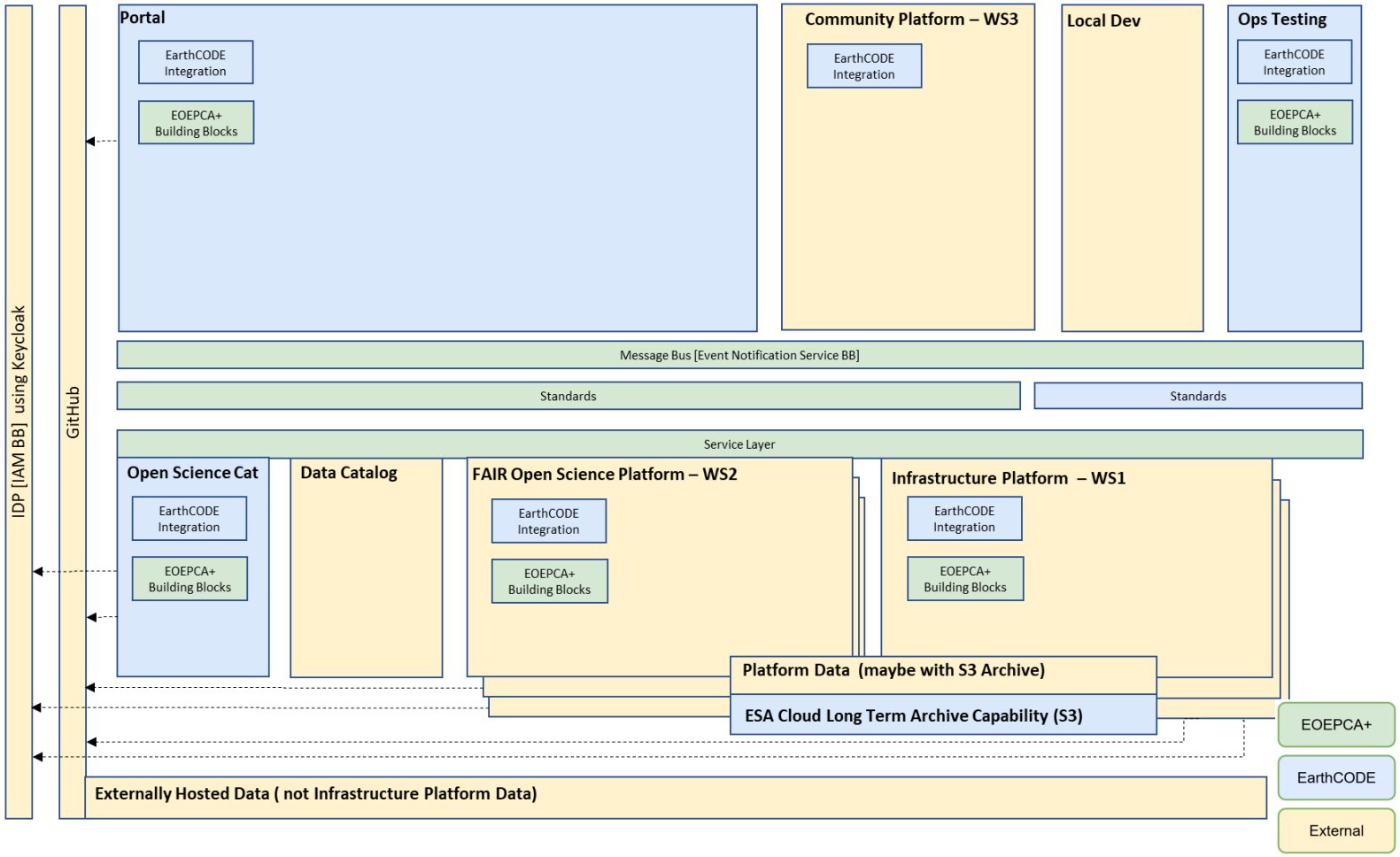
EarthCODE High-level Conceptual Architecture – Guiding Principles

- EarthCODE Portal (EoX)
- Re-use existing Platforms (WS1 and WS2)
- Re-use Open Science Catalog to store Experiment(workflow) metadata and Product metadata
- Integration rather than development
- Platforms maintained operationally by current operator
- Users can gain access platforms through the ESA NoR
- Will use EOEPCA+ building blocks if helpful



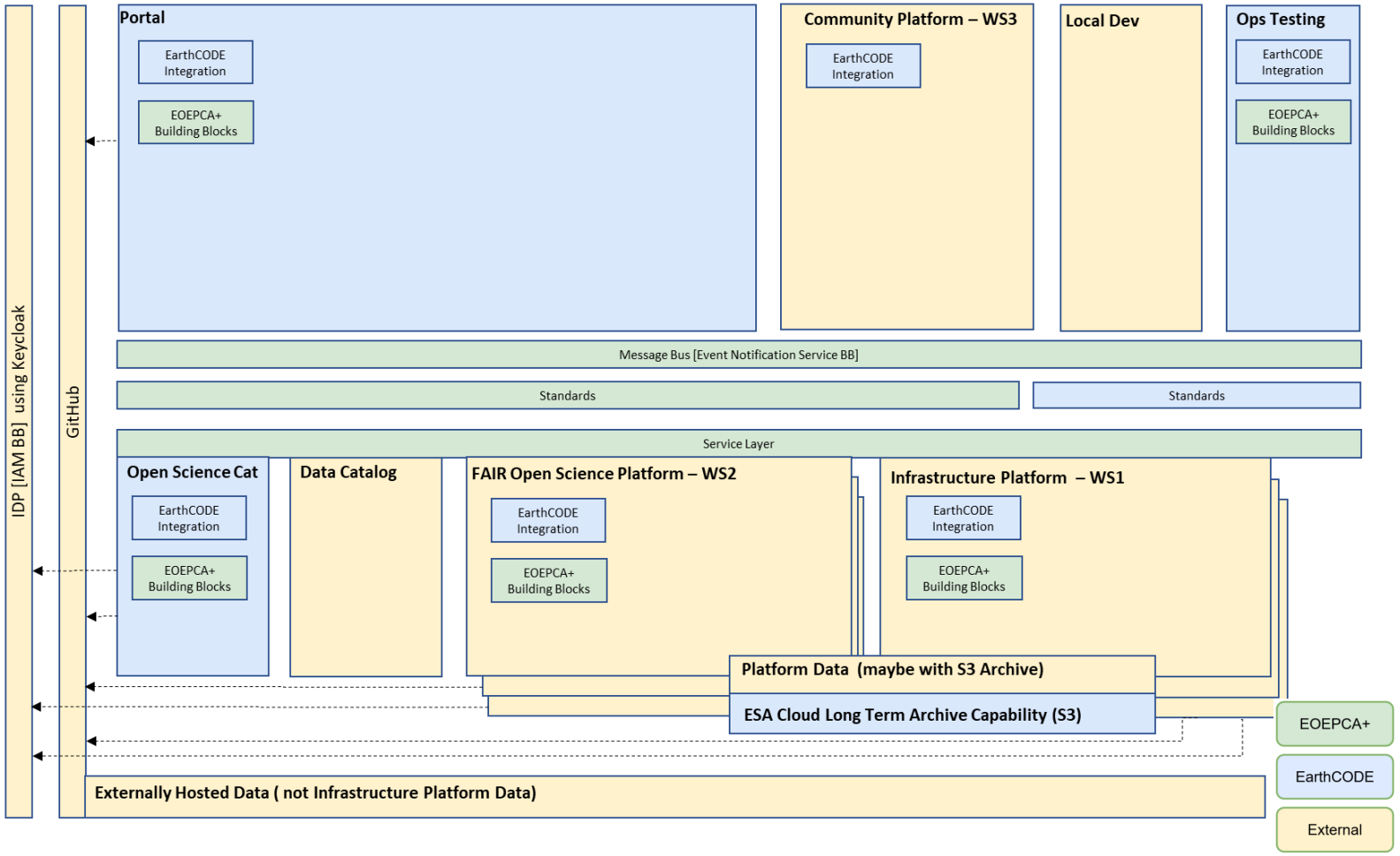
EarthCODE High-level Conceptual Architecture – Approach Phase 1

- Re-use EOxHub portal
- Re-use Open Science Catalog
- Re-use Workflow Execution Platforms (WS1)
- Re-use Workflow Development Platforms (WS2)
- Re-use Community Platforms (WS3)
- Single sign on – GitHub IDP (Identity Provider)
- WS1 Platforms are near the data
- WS2 Platforms can access small volumes of data
- Store common metadata in one place (GitHub)
  - Experiment (Workflow) Metadada
  - Experiment output (Product) Metadada
- STAC used to describe metadata
  - Led by Open Science Catalog
  - Reviews/contribution by WS2, WS1 and WS3
  - STAC may reference non STAC metadata
- Platforms (WS1/WS2) submit metadata to GitHub
- Open Science Catalog will Publish GitHub metadata
- WS1 Platforms - long term archive capability
- Scientists supported to create workflows themselves
- Can be more than one WS1 Platform
- Can be more than one WS2 Platform



EarthCODE High-level Conceptual Architecture – Approach Phase 1

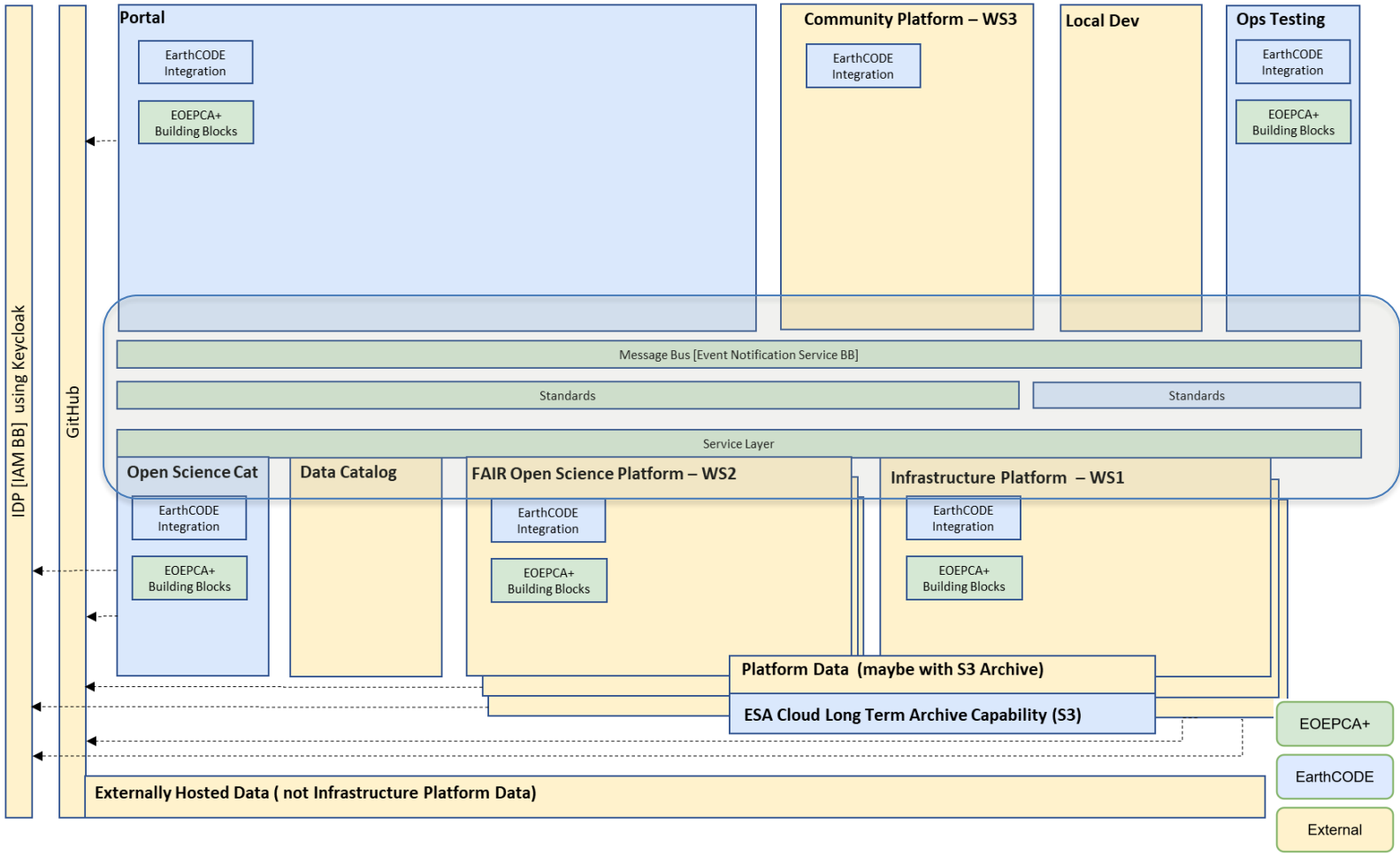
- Local Development Possible
- External Data Catalogues may still be used
- External Data may still be used
- Hope WS1 Data Catalogues become more popular
- Hope WS1 Data become more popular
- Consider use of EOEPCA+ building blocks



EarthCODE High-level Conceptual Architecture – Approach Phase 2

Further Integration

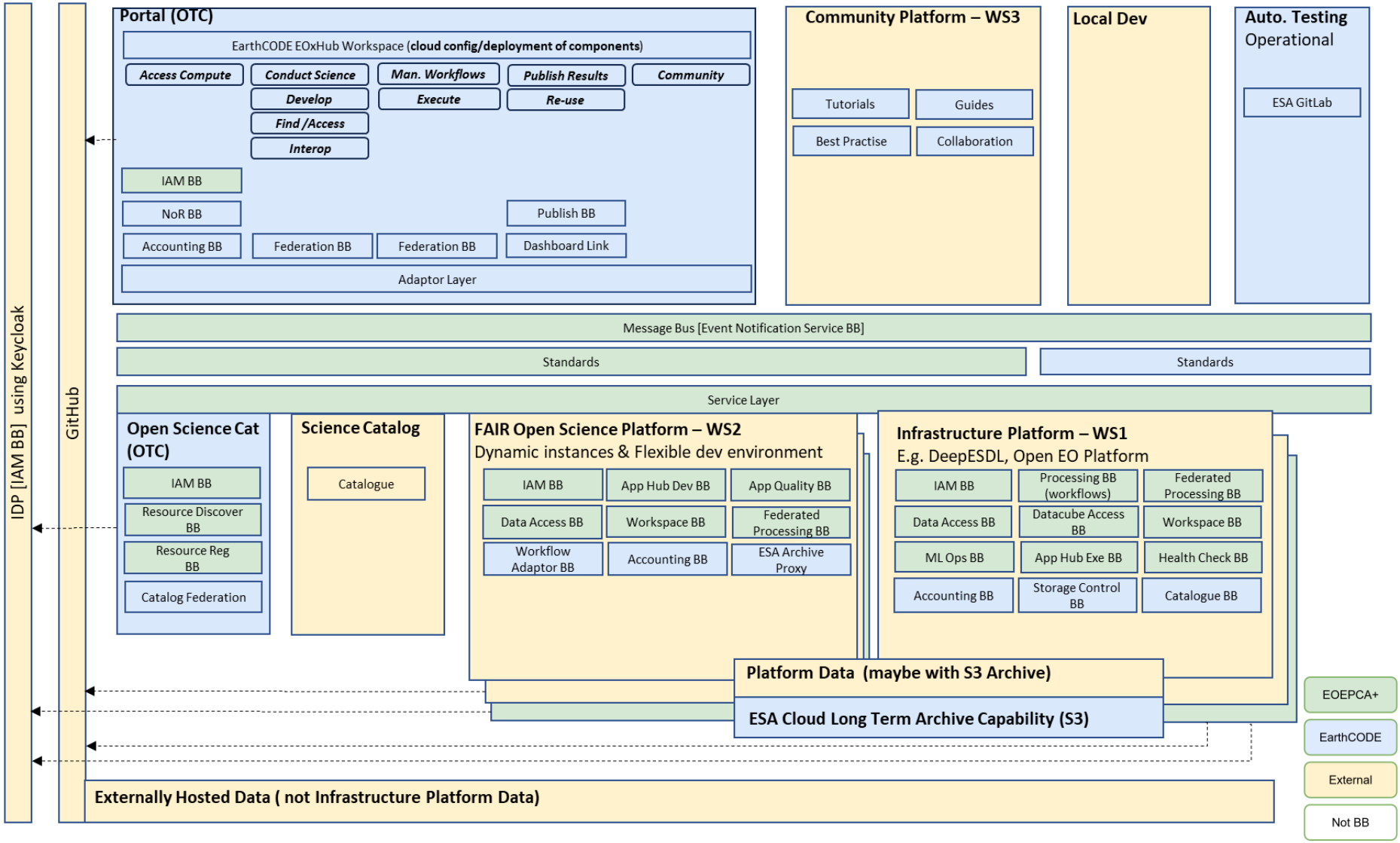
- Platforms expose a service layer
- Service layer could be used by Portal
- Platforms federation capability using service layer between platforms.
- WS1 federated catalogues
- WS1 handle multiple workflow formats
- WS2 utilise multiple WS1 capability (cross platform workflows)
- Open Science Catalog WS2 federation capability
- Standards based approach
  - Standards referenced by EOEPCA+
  - Other standards appropriate for space industry



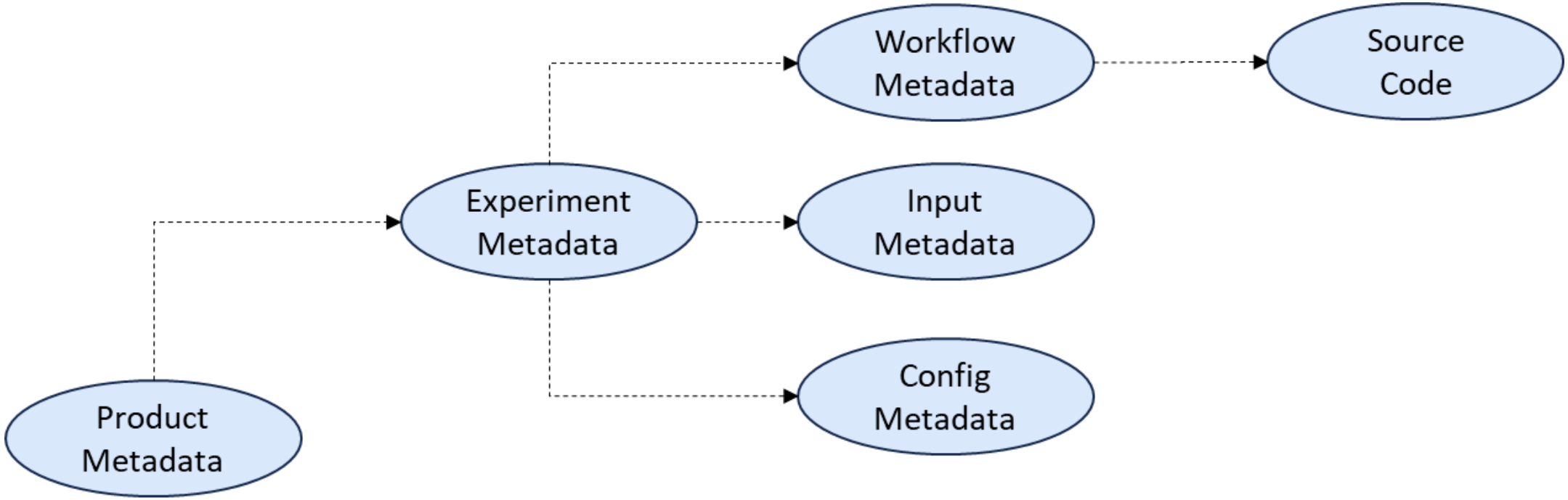
EarthCODE High-level Conceptual Architecture – EOEPCA+

EOEOPCA+ Building Blocks

- Optional
- EarthCODE is a Utilisation Domain for EOEPCA+ (requirements setting) ***Similar to APEx***
- EOEPCA+ Building Blocks
- EOEPCA+ Standards
- Use EOEPCA+ BBs (existing)
- Define EOEPCA+ BBs (new)
- Help expose Service Layers
- Help support Federated Solutions
- Help support emerging Community Standards



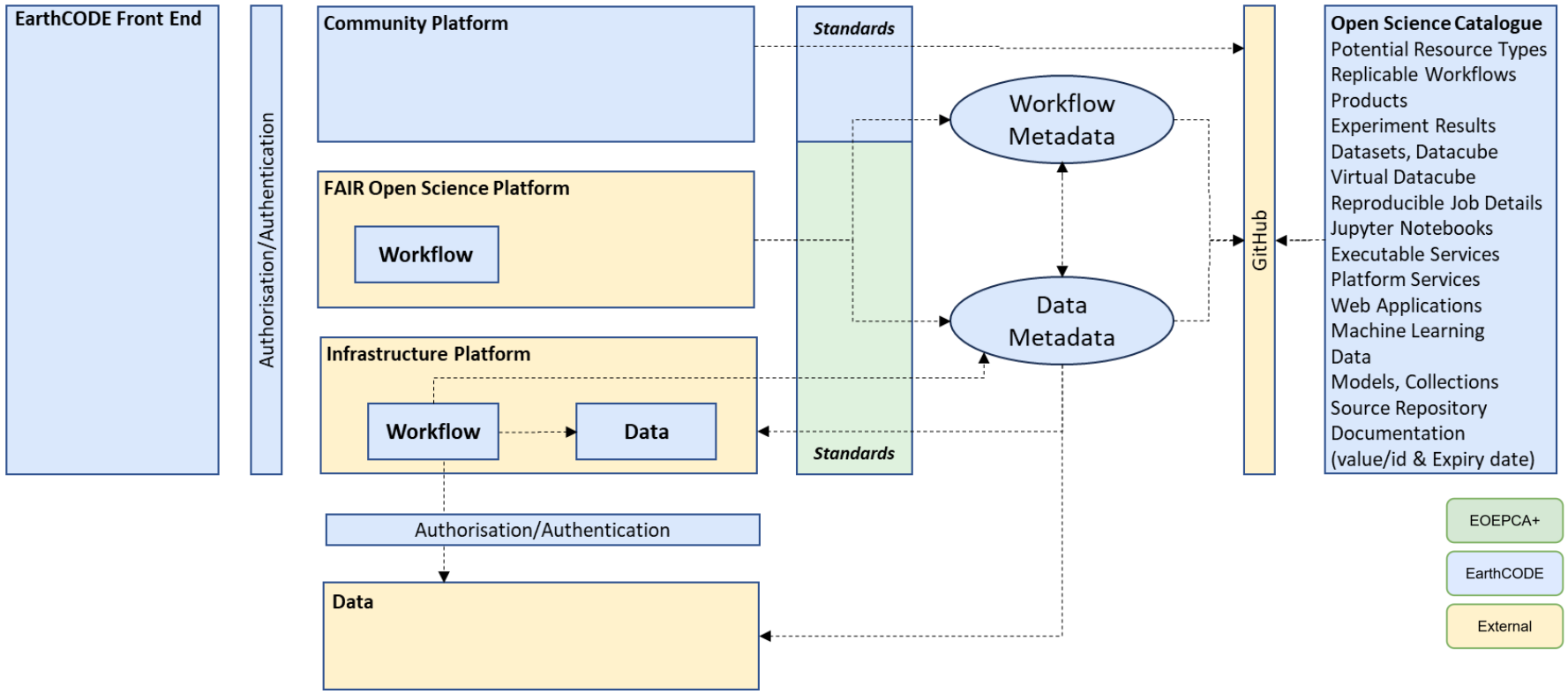
EarthCODE High-level Architecture - Metadata (Workflow and Product)



Products are the results from experiments that can be published, and they have:

- A unique reference id that can be used to identify them.
- A formal definition of the experiment used to create them.
  - A formal definition of the workflow used by the experiment.
  - A formal definition of the input data used by the experiment.
  - A formal definition of the workflow configuration used by the experiment.
- A reference to scientist's source code used by the Workflow (in some cases may be the same thing).

EarthCODE High-level Architecture - Metadata Definitions (Scope)



- There is scope to model many resource types
  - Workflows (Experiments)
  - Products (Experimental results)
- We include other resource types including dashboards.

EarthCODE High-level Architecture – Success Factors

General

- Exploitation of existing operational services
- Demonstration of operational robustness
- **Development, Publishing** and **Re-use** of reproducible science workflows
- Platform collaboration and interoperability
- Community collaboration
- Integration with NoR
- Continual evolution and maturity of EarthCODE
- **Ease of use by Scientists**

WS3 – Community

- Community Engagement key to project

WS2 – FAIR Open Science

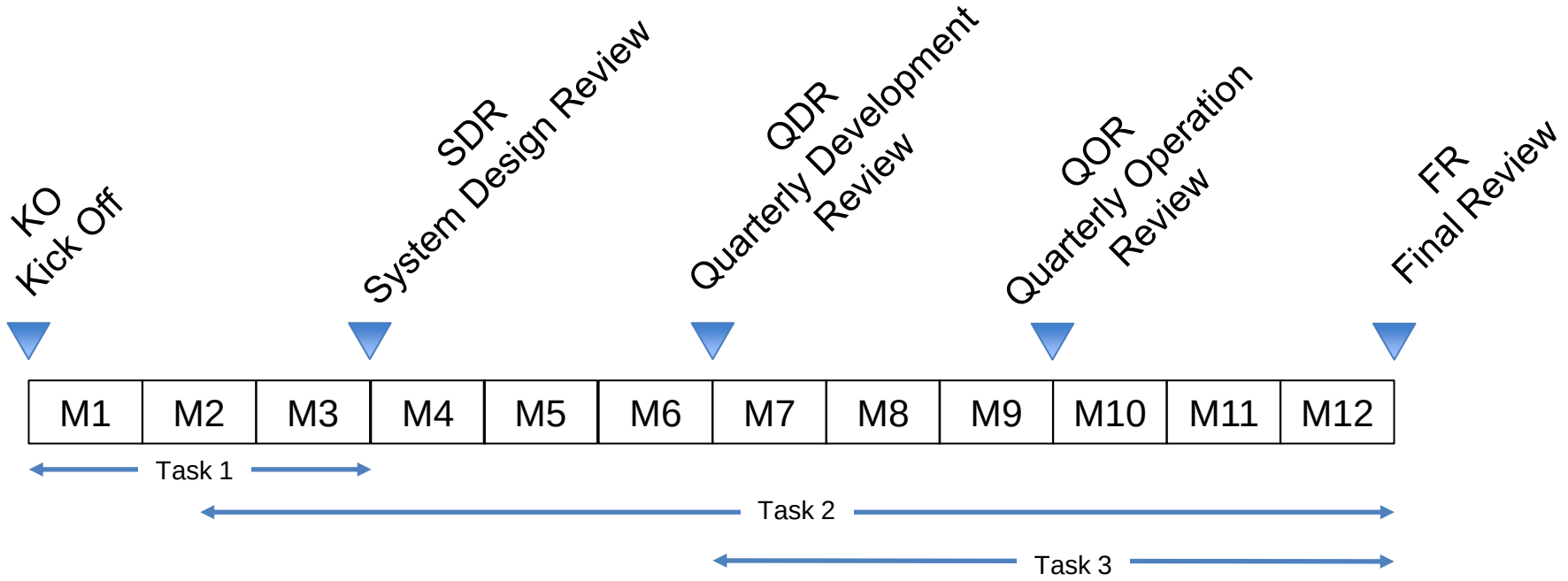
- Scientists can easily create reproducible workflows from their scientific code
- Workflow technical complexity is hidden from scientists unless they are interested
- Workflows can be developed that are compatible with input data available on target WS1 Infrastructure

WS1 – Infrastructure

- Ensure capability to reliably execute workflows developed on WS2

EarthCODE Development Schedule

Nominal Schedule (Quarters)



BP Main Work Streams (for phase 1 and phase 2)

BP Task 1 – Requirements Refinement and Initial Roadmap and Design – completes at SDR

BP Task 2 – Design, Implementation, Deployment and V&V – continues while development is ongoing

BP Task 3 – Operational Integration and Evolution – Nominally starts at QDR until end of contract

EarthCODE Industry Information Day

Agenda

- Procurement Introduction
- Overview of Work Streams and Reference Architecture
- **Approach to Selection**
- Intended ITT Documentation
- Contractual / Financial Aspects
- Intended ITT Schedule

EarthCODE Approach to Selection

Invitation to Tender

Negotiation Phase

Award of contract to Tenderers to develop negotiated WS

1. Invitation to Tender

In this ITT Telespazio UK seeks Proposals for EarthCODE Work Streams

Tenderers are invited to submit proposals based on WS1 Infrastructure, WS2 FAIR Open Science, WS3 Community. Tenderers can submit on more than one Work Stream. There can also be more than one platform per Work Stream.

Bidders are encouraged to:

- Demonstrate their understanding of the overall architecture
- Submit specific proposals to modifying and operating their platforms
- Describe their wider competencies as they may relate to other work streams
- See the future evolution of EarthCODE as it moves towards greater integration and interoperability

2. Negotiation Phase

TPZ UK system team will review proposals and may review and refine the reference architecture in the light of these proposals.

Work Streams are to be developed in collaborative manner. Developers will need to work collaboratively across the whole development team to deliver this objective (achieving interoperability between different subcontractors). TPZ UK will coordinate activities by selecting/allocating who does what to produce an open environment according to open governance.

3. Award of contract to Tenderers to develop negotiated Work Streams

Successful negotiation will result in award of contract to selected sub-contractors to join an integrated cross-company development work team.

EarthCODE Industry Information Day

Agenda

- Procurement Introduction
- Overview of Work Streams and Reference Architecture
- Approach to Selection
- **Intended ITT Documentation**
- Contractual / Financial Aspects
- Intended ITT Schedule

EarthCODE Intended ITT Documentation

https://esastar-publication.sso.esa.int/nonEsaTenderActions/details/13691

- Cover Letter
- Draft Contract
- Special Conditions of Tender
- Statement of Work
- Proposal Outline
- Applicable Documents:
  - EarthCODE Architecture Implementation Roadmap
  - Outreach and Communication Plan

EarthCODE Industry Information Day

Agenda

- Procurement Introduction
- Overview of Work Streams and Reference Architecture
- Approach to Selection
- Intended ITT Documentation
- **Contractual / Financial Aspects**
- Intended ITT Schedule

EarthCODE Contractual / Financial Aspects

The total budget to spread across all Work Streams is €750k over 12 months.

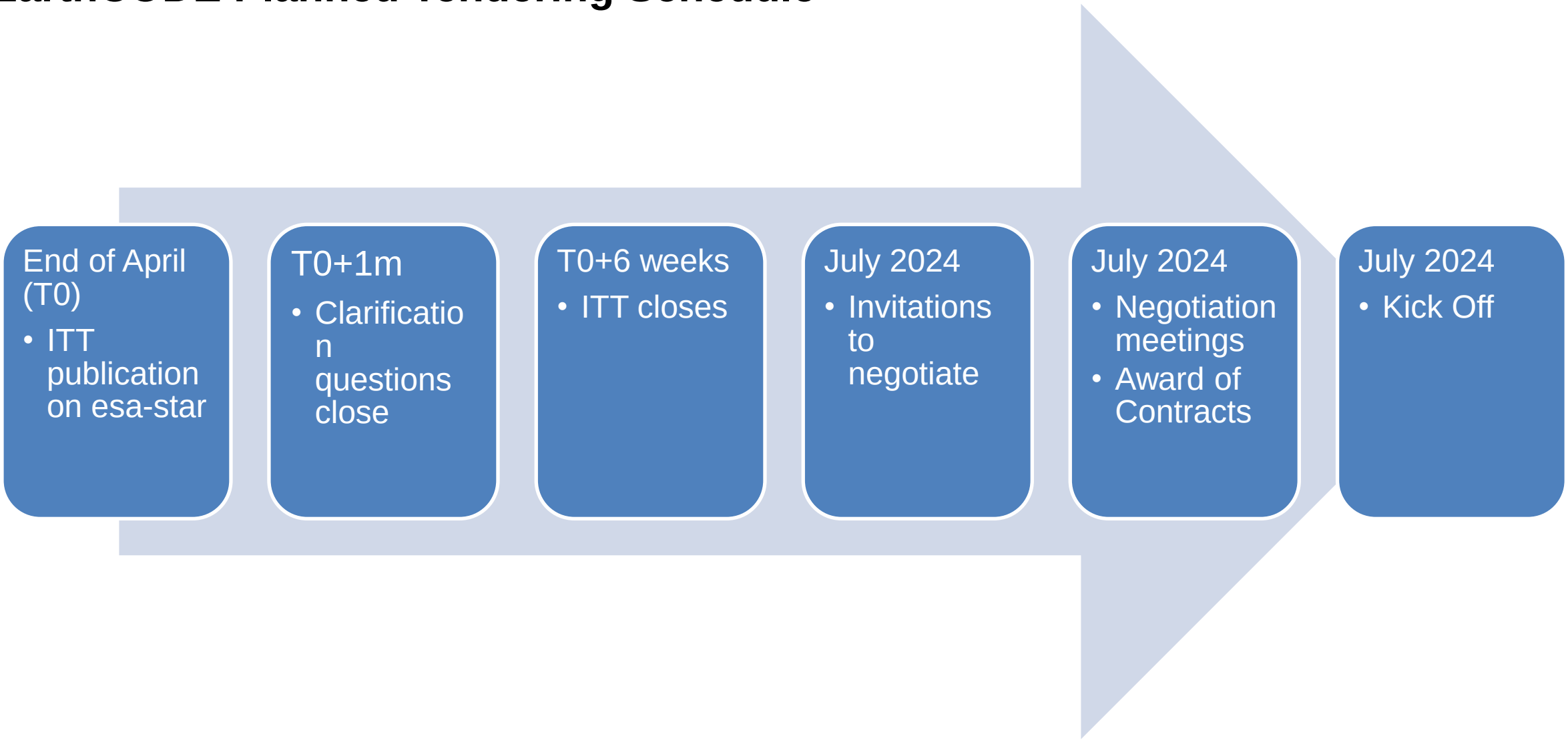
Tenderers will be registered in AT+BE+CZ+DE+DK+EE+GR+ES+FI+FR+GB+HU+IE+IT+LT+LU+LV+NL+PL+PT+RO+SE+SI+SK+CA+CH+NO

EarthCODE Industry Information Day

Agenda

- **Procurement Introduction**
- Overview of Work Streams and Reference Architecture
- Approach to Selection
- Intended ITT Documentation
- Contractual / Financial Aspects
- **Intended ITT Schedule**

EarthCODE Planned Tendering Schedule



esa-star registration

Registration on esa-star Tendering before submission of a proposal is mandatory

Esa-star registration website: https://esastar-emr.sso.esa.int/

Any entity, including any subcontractor in the Proposed Project team at the time when a tender is submitted, is required to at least have already completed the “Light Registration” as described in the following esa-star Registration User Manual:

https://esastar-emr.sso.esa.int/Account/DownloadFile

Intended ITT

THANK **YOU**
FOR YOUR ATTENTION

telespazio.co.uk

The Industry Information Day presentation will be made available through ESA-STAR (ESA News)