EOF Open Source Roadmap

4-5 March 2025
Lisbon

→ THE EUROPEAN SPACE AGENCY

Agenda
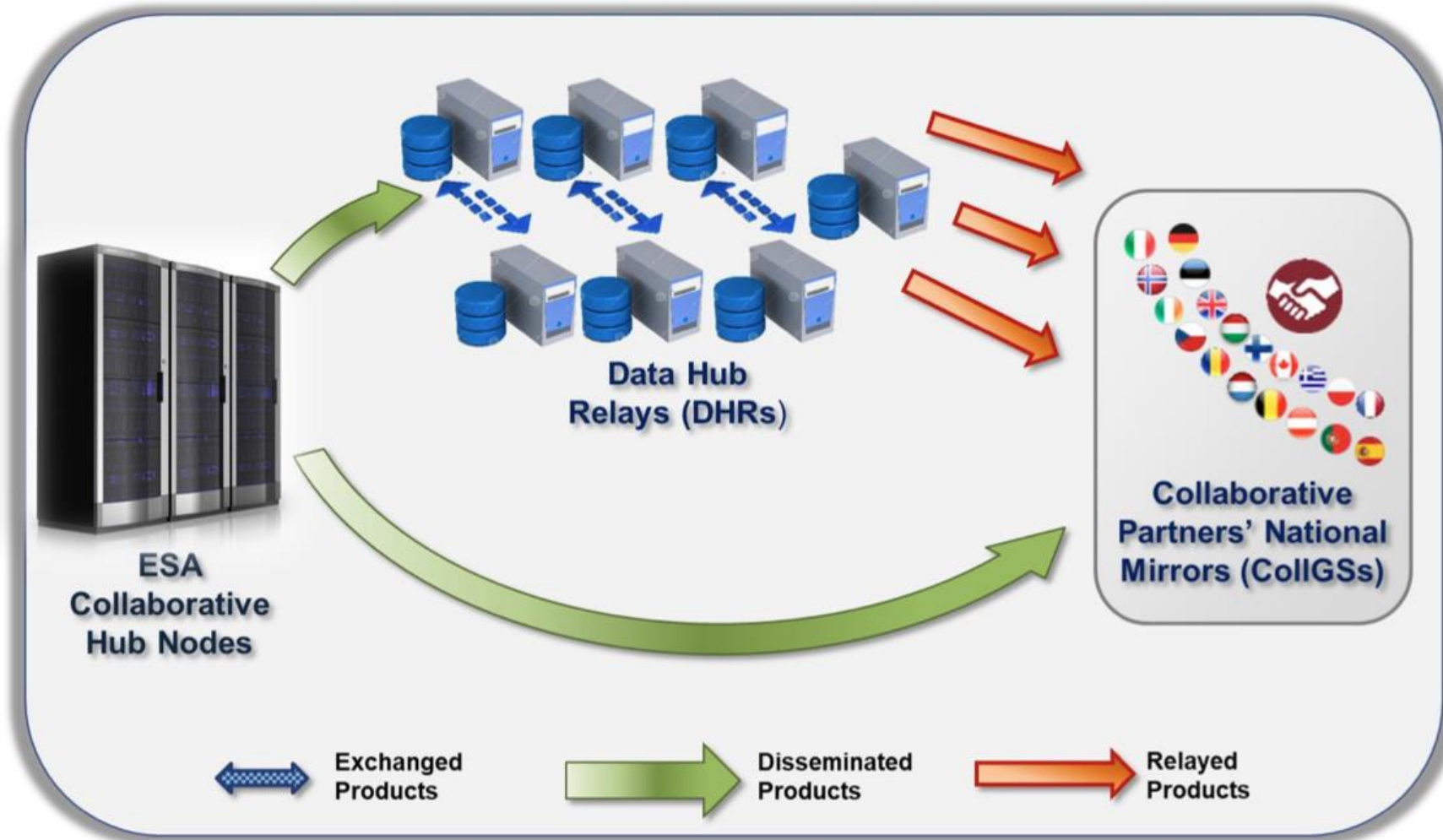
➢ Background

➢ Open Source "Data Hub Suite"

➢ The GSS Catalogue – a DHuS replacement

➢ Focusing a moment on STAC

➢ What Next

Background – DHuS Dismissal

Copernicus Data Access Service (ColHub 1,2,3)

Copernicus Data Access Service Continuity

Collaborative Hub Node 2 (OVH)

Collaborative Hub Node 3 (GRNET)

GSS Node (OVH)

GSS Node (OVH)

Col Hub Node 1 migration to CDSE

Collaborative Hub Node 1 operations on CDSE

EXTENSION TO JUNE 25

New Copernicus Data Access Service (CDSE)

2022   2023   2024   2025   2026

DHuS Maintenance   Extension   DH Suite Maintenance

→ THE EUROPEAN SPACE AGENCY

Recalling the success

Data Hub Relays (DHRs)

ESA Collaborative Hub Nodes

Collaborative Partners' National Mirrors (CollGSs)

Exchanged Products — Disseminated Products — Relayed Products

- From "1 copy" from the Core Ground Segment to sustained exchange and systematic synchronisation and redistribution

→ THE EUROPEAN SPACE AGENCY

Looking to the future

- With almost all Sentinel Data immediately available within CDSE or cached on DestinE the need for MS mirror sites is less evident

- Nevertheless, Member States advanced services and local infrastructures may require local copies of data (for legal requirements or technical performance or support to national initiatives)

→ THE EUROPEAN SPACE AGENCY

Collaborative Data Hub Suite – Open Source

Lisbon, 4-5 March 2025

4/5-03-2025

→ THE EUROPEAN SPACE AGENCY

Collaborative Data Hub Maintenance and Evolution Services Overview

SciHub *entering in operations*

GS Transition to Cloud

ColHub 1 migration on CloudFerro infrastructure

ColHub 2 migration from DHuS to GSS Instance

| 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 |
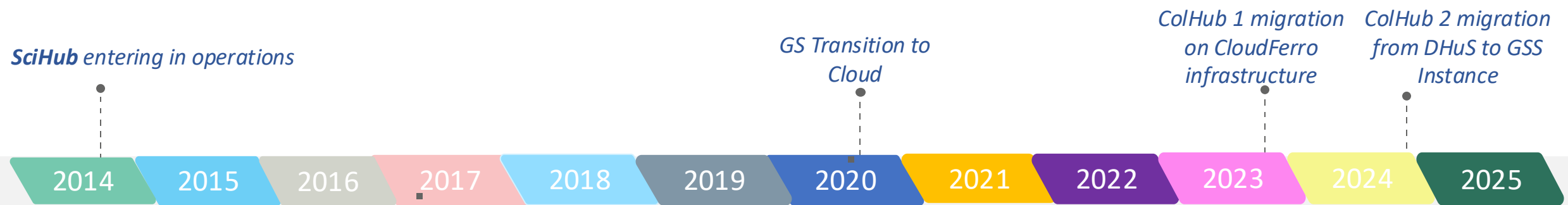
CollHub Nodes 1, 2, 3 *set up and Collaborative network definition*

Upgrade ColHub Node 1
Full historical dataset access

31st 2021 Kick Off Collaborative Data Hub Maintenance and Evolution services – Ready for DTE

Phase 2 Collaborative Data Hub Maintenance and Evolution services – Ready for DTE

Phase 3 Collaborative Data Hub Maintenance and Evolution services

Working to

*Support the evolution of Collaborative Users Network into an **ecosystem** of service providers, service consumers and data producers contributing to wider initiative*

Facilitate the management of Collaborative centre operations

Use open source and extendible solutions

7

Collaborative DHS Service
Data Hub Software suite (DHS)

- Opensource platform designed to support data discovery, distribution, processing and dataflow control for Earth Observation data

- Highly reliable, scalable and extendible ready to set up an EO data retrieval and distribution centre as a CollGS node.

- DHS includes frameworks dedicated to data ingestion, search and processing for multi-mission data management.

6 software applications supported by a set of state of art open-source solutions.

Interactive data gathering and distribution service based on the CSC Copernicus STAC API
With GSS

New graphical Interface for data discovery and download
With COPSI

Strong dataflow visualization and Service control
With DAFNE

Extendible data processing workflow
With Transformation Framework

AI tool converting natural language queries into STAC/OData for Sentinel data search
With AI ChatBot (MVP)

Automatic DHS suite installation
With DHS Suite Easy Deploy
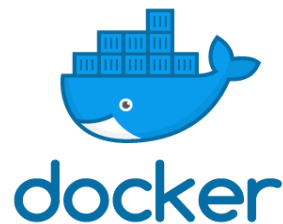
Collaborative DHS Service
GSS

GSS - Store Service is a set of components designed to **retrieve and expose products** stored in various sources.

GSS is a solution designed to replace **DHUS**, the reference software used for over 10 years to ingest, catalogue, and disseminate Earth Observation data.

Built on modern software and **cloud-native technologies**, **GSS** leverages years of operational experience to offer a **more flexible, scalable** service.

In **operation since 2023**, **GSS** is gradually taking over **DHUS** functions.

Collaborative DHS Service
DAFNE

DAFNE – Data Flow Network Environment is an intuitive and modern interface for managing data flows, reducing the need for log checking and manual intervention. It is designed as a plug-and-play visualization tool for harvesting and displaying data.

The system oversees:
- Archive Info
- Data Source Info
- Synchronization Info

A valuable tool for monitoring and reporting activities:
- Service Availability
- Products Timeliness
- Products Completeness
- Weekly and Monthly reports.

Collaborative DHS Service
COPSI

COPSI – Copernicus Space Component Interface is a web application based on angular, designed to offer **an intuitive and modern interface** for accessing Copernicus Sentinel products.

 Key Features
- **Intuitive Interface:** User-friendly GUI to easily access Copernicus data.
- **Centralized Access Management:** Integrated with Collaborative DHS IAM, no need for additional client configuration.
- **Product Discovery & Download:** Search, visualize, and download products via STAC/OData, with results shown as lists and map footprints.

Collaborative DHS Service

Transformation Framework

- Copernicus Sentinel user level data transformation on demand

- EOPF-cpm plug-in supporting different data format conversions (ZARR, COG, NetCDF).

- STAC Catalogue Support to integrate different data sources.

- Monitors & optimizes resource consumption (processing, downloads, user requests)

Collaborative DHS Service
Artificial Intelligence

Within the Collaborative DHS, an AI component is being implemented to simplify Sentinel data discovery by translating natural language queries into structured STAC formats.

- the user starts with a natural language query through the AI interface.
- the AI processes the query to understand the intent and the data required.
- the AI tool converts the query into STAC formats.
- the user receives these translated queries, which are ready for use with GSS.
- integration with COPSI enables the user to directly retrieve data

13

Collaborative DHS Service

The Data Hub Service is a modular software with several features that make it versatile:

- Microservice architecture - Docker images
- Scalable backend (Solr, Kafka)
- Exposes products through an OData API and a STAC API
- Storage abstraction – S3, OpenStack Swift, Disk
- OAuth2 authentication

Its specific adaptation has already established it as a solid and operational solution in areas such as:

- PRIP - Production Interface Delivery Points
- Collaborative Node 2 – Migrated to GSS in December '24

Collaborative DHS Service Future Evolution

Collaborative Service phase 3
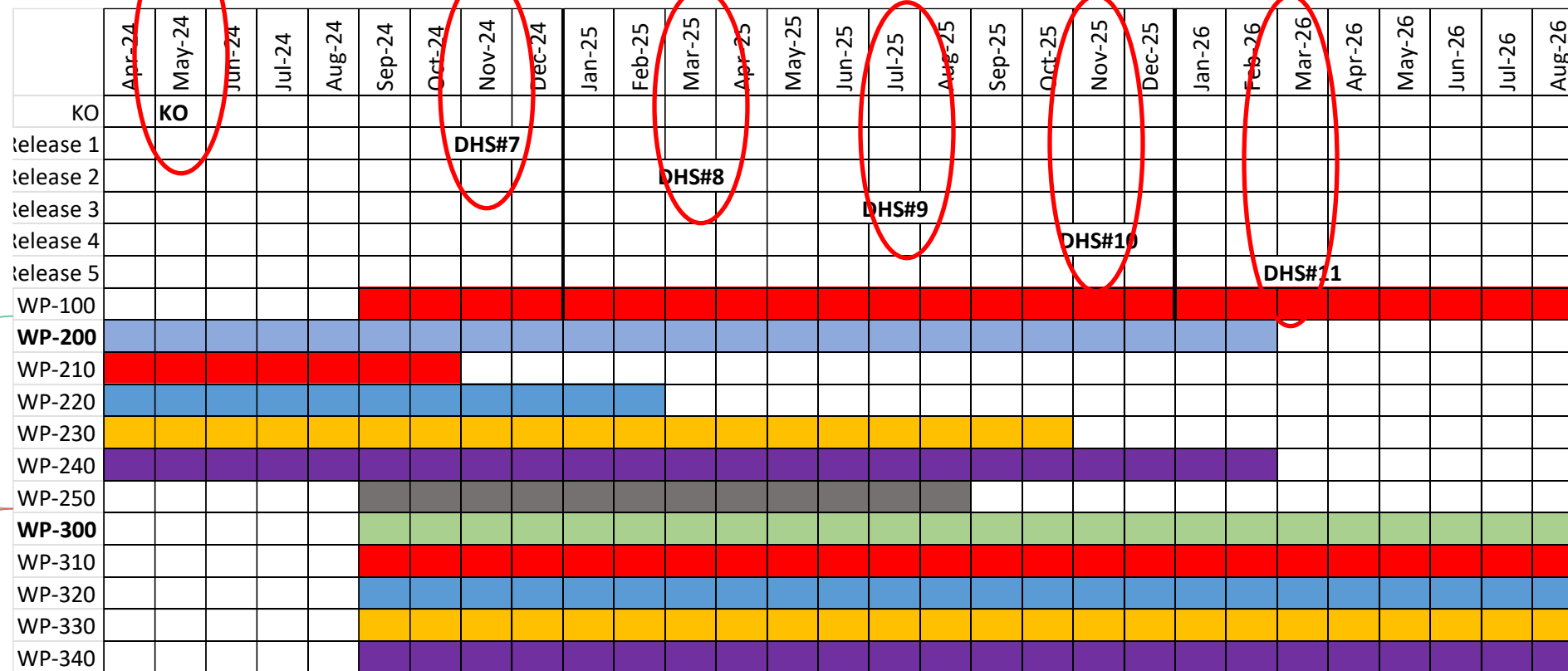
As part of the Collaborative DHS, SERCO organized a dedicated workshop for DHR and Collaborative users, held in a hybrid format, with both in-person and remote participation. The goal was to provide an in-depth overview of the new DHS components, with a special focus on operations and the new GSS.

The workshop also offered a hands-on demonstration, showcasing how the software is managed and operated, bringing participants closer to the technical and operational aspects most relevant to them.

A well-attended event, with **52 participants** (DHR, Collaborative users and partners) joining on-site and remotely

Thank you !

16

GSS Workshop

GSS Features Presentation

Lisbon – 4th and 5th of March 2025

Benoît Pironnet

Product manager

benoit.pironnet@gael.fr

Rosine Nyemeck *(remote)*

rosine.nyemeck@gael.fr

Sami Ben Rhouma *(remote)*

sami.benrhouma@gael.fr

GAEL SYSTEMS

Earth Observation Data Solutions

www.gael-systems.com

CONFIDENTIAL
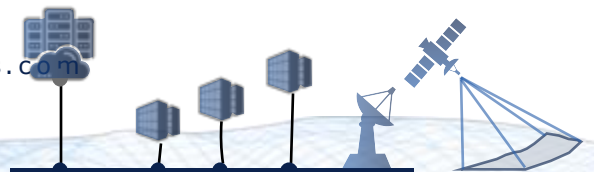
Agenda

EO Data Solutions

- **Introduction**
- **Use cases**
- **GSS key features**
- **DHUS vs GSS**
- **Architecture overview**
- **Last release features**
- **Next release features**
- **Q/A**

GAEL SYSTEMS

The **GSS (GAEL Store Service** is an open source software solution, to retrieve and ingest products from any data source, the solution is scalable, configurable and cloud agnostic.

- GSS is based on a **micro-services** architecture.

A brief history:

- For more than 10 years, DHUS has been a reference software for ingesting, cataloguing and disseminate earth observation data.

- years of experience have enabled the refinement of the solution adapted to cloud infrastructure technologies.

- Since 2023, when it first went into operation, GSS has been gradually replacing the DHUS.

GAEL SYSTEMS
Earth Observation Data Solutions
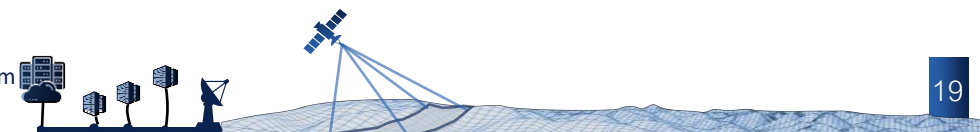
Why GSS ?

- The DHUS is a one-piece architecture and application

- The GSS micro-service architecture offers modularity and enables services to be mutualized
    - e.g. a single Identity Access Management instance like Keycloak for several instances of an application.

- GSS works with **Docker** and **Kubernetes**.

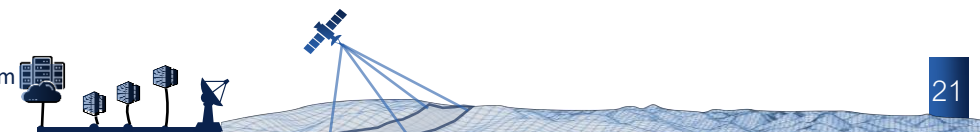- The DHUS will no longer be maintained in the **second half of 2025**.

GSS compliant with :
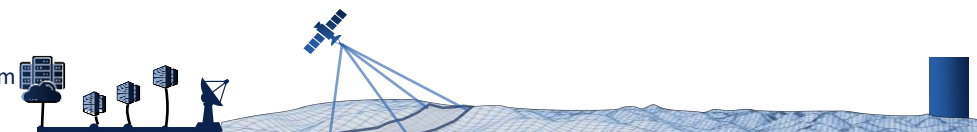
Services
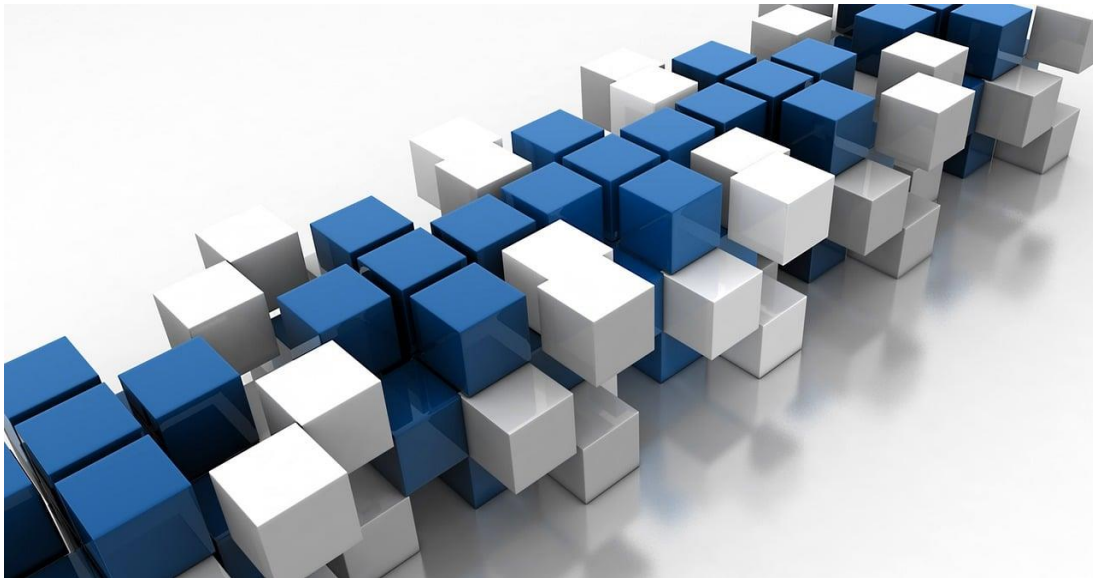
- **Data retrieval and ingestion**

- **Flexible storage**

- **Indexing and metadata management**

- **Compatible APIs**

- **Scalable architecture**

- **Catalog et metrics**

- **Subscription / Notification**

GAEL SYSTEMS
Earth Observation Data Solutions

www.gael-systems.com

GSS is operated as part of the Collaborative Node2 :

- ESA ColHub Node2 since third trimester 2024

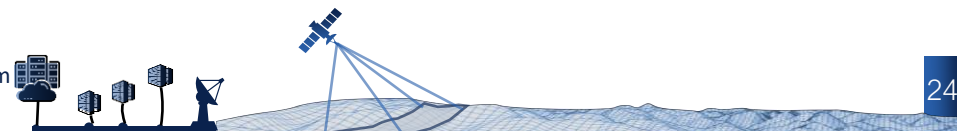Transition from GSS to DHUS has been performed with:

- CESNET

- AIRBUS

- DHR AT

GAEL SYSTEMS
Earth Observation Data Solutions
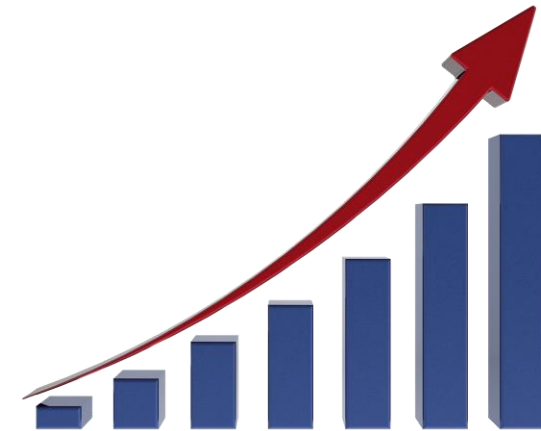
- Micro service architecture - **docker images**

- Ingest EO products from multiple sources in a horizontally scalable way

- **Scalable backend** (Solr, Kafka)

- Expose products on an **OData API (v4 CSC)** and a **STAC API (1.1.0)**

- Storage abstraction – **S3**, Openstack Swift, Disk

- Automatic management of containers/buckets

- **Customizable ingestion workflow**

- **Powerful product searching**

- Automatic eviction + product deletion jobs

- Configurable user quotas (**download volume** + parallel downloads)

- **OAuth2 authentication** (Keycloak) with roles (centralized user management)

- **Subscription/Notification** mechanism

GAEL SYSTEMS
Earth Observation Data Solutions

GSS vs DHuS
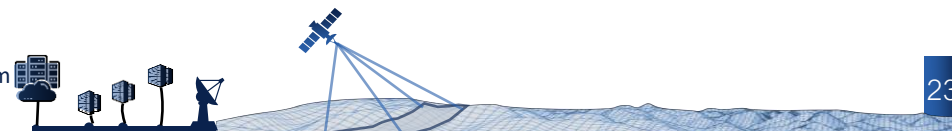
Features	DHuS	GSS
Ingest from HFS	✅	✅
Ingest from DHuS	✅	✅
Ingest from GSS	✅	✅
ODATA API (V2, V4)	✅	✅
Micro-services	❌	✅
STAC specifications v1.1.0	❌	✅
OAuth2 / SSO compliant	❌	✅
Admin API	❌	✅
Notification	❌	✅

GAEL SYSTEMS
Earth Observation Data Solutions

www.gael-systems.com

Operate the GSS

Some open resources:

- Compliant with STAC specifications v1.1.0

- Sortby extension support (STAC API)

- Products, quicklooks and files (such as OData nodes) can be downloaded (STAC API)

- **Ingester Management in Admin API**

- **Eviction Management is now in Admin API**

- Possibility to aggregate logs in Kafka topic(s)

GAEL SYSTEMS
Earth Observation Data Solutions

www.gael-systems.com
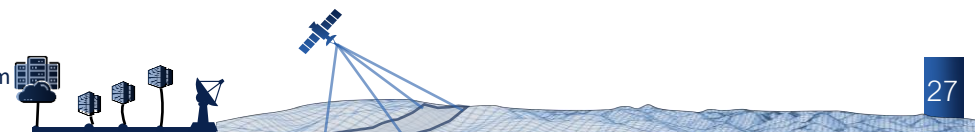
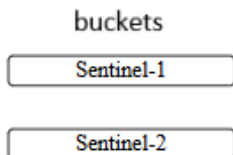**S3 storage**

S3DataStore
- S3 single bucket storage (automatically created)
- Unlimited number of products

S3DataStoreGroup
- Multiple S3 buckets storage
- Buckets are automatically created according to configuration

buckets

S1.*:Sentinel-1,S2.*:Sentinel-2 ⇒ | Sentinel-1 |
| Sentinel-2 |

**Ingestion from STAC source
+
STAC S3 links**

- New STAC source in producer and consumer
- compatibility with STAC sources
- First target is CDSE and GSS
- Configurable filter, sortby, limit and pivot date
- OAuth2 and basic auth support

Notifications for ingested products*

- Users can be notified when a product has been ingested in a ingestion job
- Users will be able to subscribe products events (deletion, ingestion, eviction)

*based on CESNET DHR improvement suggestion
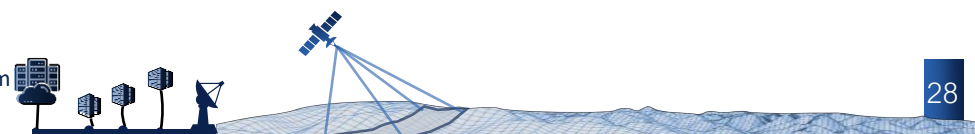
Thank you for your attention
/
Questions

https://www.gael-systems.com/

GAEL SYSTEMS
Earth Observation Data Solutions
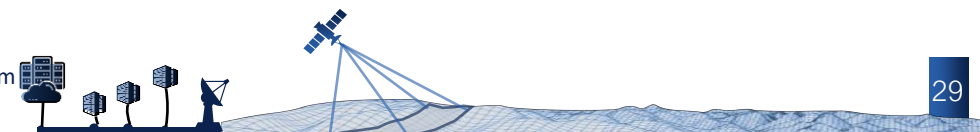
Spatio Temporal Asset Catalogs (STAC)
- an opportunity for improved interoperability

Reflections & Conclusions from a Workshop - ESA ESRIN

Jolyon.Martin@esa.int

06/06/2024

→ THE EUROPEAN SPACE AGENCY

Background

As evidenced during the *ESA EO Data Management and Operations Framework (EOF) Workshop* in April 2024 there is a current and large interest in the STAC standard for cataloguing and data access in the EO related domains

   - Space Component / Services / Member States and far beyond

Many initiatives are beginning to publish their "proof of concept" STAC services

Whilst some "Best Practices" are emerging, there is a clear window of opportunity to align choices of implementation to maximise interoperability to simplify user experience and facilitate federation in data access.

What is STAC?

- **STAC** stands for **S**patio **T**emporal **A**sset **C**atalog

- It is a specification to standardize the way geospatial assets are described and cataloged.
  - Open specification, started in 2018 by Radiant Earth, current version 1.0.0

- The specification is divided in two parts
  - STAC Core: A set of common data structures using JSON and GeoJSON
  - STAC API: A RESTful programming interface for interrogation of catalogues providing actionable information for subsequent data access

Purpose: To make it easier to find, share, and use geospatial data.

→ THE EUROPEAN SPACE AGENCY

Why STAC?

- **Challenges**:
  - Geospatial data is often scattered across different platforms.
  - Inconsistent metadata makes it hard to find relevant data.
  - Difficulties in sharing and integrating geospatial data.
- **Solution**:
  - STAC provides a standardized and consistent way to catalog geospatial data.
  - Enhances discoverability and usability of geospatial assets.
- **Benefits**:
  - Simplifies the process of finding and accessing geospatial data.
  - Facilitates interoperability across different systems.
  - Promotes the sharing and reuse of geospatial data.
  - Standardization (OGC API) facilitates integration with other tools and platforms.

STAC Core Format

JSON Format:

- STAC uses JSON and GeoJSON.
- JSON is easy to read and use with web technologies.
- JSON can be easily extended by any fields.

```json
{
  "type": "Feature",
  "id": "unique-id",
  "bbox": [-180, -90, 180, 90],
  "geometry": {
    "type": "Polygon",
    "coordinates": [[…]]
  },
  "properties": {
    "datetime": "2020-01-01T00:00:00Z",
    "title": "Example Item",
    "description": "A description of the example item."
  },
  "assets": {
    "image": {
      "href": "https://example.com/image.tif",
      "type": "image/tiff"
    }
  }
}
```

Core Metadata of a STAC item

Common Fields:

- id: Unique identifier for the item.

- type: Type of GeoJSON object (usually "Feature").

- bbox: Bounding box of the asset.

- geometry: Spatial geometry of the asset.

- datetime: Timestamp of the asset.

- properties: Additional metadata about the asset.

```
{
  "type": "Feature",
  "id": "unique-id",
  "bbox": [-180, -90, 180, 90],
  "geometry": {
    "type": "Polygon",
    "coordinates": [[…]]
  },
  "properties": {
    "datetime": "2020-01-01T00:00:00Z",
    "title": "Example Item",
    "description": "A description of the example item."
  },
  "assets": {
    "image": {
      "href": "https://example.com/image.tif",
      "type": "image/tiff"
    }
  }
}
```

STAC API

- The STAC API is a specification for web services that provides a standard way to search and retrieve STAC Items and Collections.

- Compliant with OGC API – Features (OAFeat)

- **Features**:

    - Filtering by spatial and temporal parameters.

    - Paging through search results.

    - Extensibility for additional query parameters.

- **Benefits**:

    - Enables dynamic querying.

    - Facilitates integration with other web services and applications.

GET /stac/search?bbox=[-180, -90, 180, 90]&time=2021-01-01T00:00:00Z/2021-12-31T23:59:59Z

```
{
  "type": "FeatureCollection",
  "features": [
    {
      "stac_version": "1.0.0",
      "stac_extensions": [],
      "type": "Feature",
      "id": "CS3-20160503_132131_05",
      "bbox": [],
      "geometry": {},
      "properties": {},
      "collection": "CS3",
      "links": [],
      "assets": {}
    }
  ],
  "links": [
    {}
  ],
  "context": {
    "returned": 1,
    "limit": 5,
    "matched": 314159
  }
}
```

STAC Landscape

The adoption of STAC is a major trend in the earth sciences and remote sensing industry!

- Adoption by major players to catalog and access earth observation data:
  - Governmental organizations
    - ESA - Copernicus Data Space Ecosystem (CDSE)
    - NASA - Common Metadata Repository (CMR)
    - USGS Astrogeology Provided Analysis Ready Data
  - Large number of private companies support and uses STAC
    - Commercial satellite imagery providers both Optical and SAR
    - Data analytics companies

- Examples of public catalogs: Planetary Computer (Microsoft), Earth Search (E84), Earth Engine Data Catalog (Google Earth Engine), Digital Earth Africa

- Beyond satellite data: environmental models, scientific citations, Machine Learning labeling

Benefits of STAC

Discoverability: Easier to find relevant geospatial data.

Interoperability: Standardization (OGC API) facilitates integration with other tools and platforms.

Scalability: Built for Extendability through Core and API extensions mechanisms

Community Support:

Growing community and ecosystem around STAC.

Increasing adoption by many organisations: large availability of public STAC catalogs

Community Driven:

Open source

Fostering open contribution

STAC Workshop

A first STAC Workshop was held in ESA/ESRIN & hybrid format on 6 June 2024 to take stock of the status of STAC across Copernicus and within the larger landscape.

20 presentations from diverse organisations (institutional & industry) provided the opportunity share status of current implementations and plans, benefits and perspectives.

Active interest from ~200 online attendees

Useful discussion on the current experiences and need for fostering further development and adoption.

A wealth of experience in diverse domains

Workshop resources at:
https://eoframework.esa.int/display/SSR/STAC+Workshop+-+June+6%2C+2024

Some key messages

STAC has very useful and powerful extension mechanisms to tailor the specification to support diverse data sets and use cases (well beyond legacy lowest common denominator interoperability approaches).

Rich set of Open Source resources and libraries provides low barrier for implementations

STAC goes beyond data discovery and matches well to providing interoperable data access for cloud native datasets

There is an active community shaping the evolution of the standard, ready to contribute to the implementation.

With great flexibility comes great responsibility

→ THE EUROPEAN SPACE AGENCY
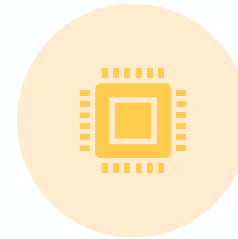
STAC & Zarr Workshop – follow up – 17 April 2025

A follow up STAC Workshop is proposed for 17 April 2025 – save the date.

Feedback on the "Proof of Concept" implementations

Update with synergy with Zarr data formats, adoption as I/F point also within the Copernicus Ground Segment

Explore if further "Best Practices" can facilitate interoperability within the heterogeneous data management contexts

Focus on key extensions of common interest to provider and user communities.

Discuss and advance use cases for simplified data access across missions and in federated access scenarios.

Creating a Collaborative Campus

Leveraging the European Data Spaces Guidelines and Earth Observation Framework,

the ESA DTE framework will serve as the first experiment to create a deployable, 'batteries included' environment to support essential building blocks such as:
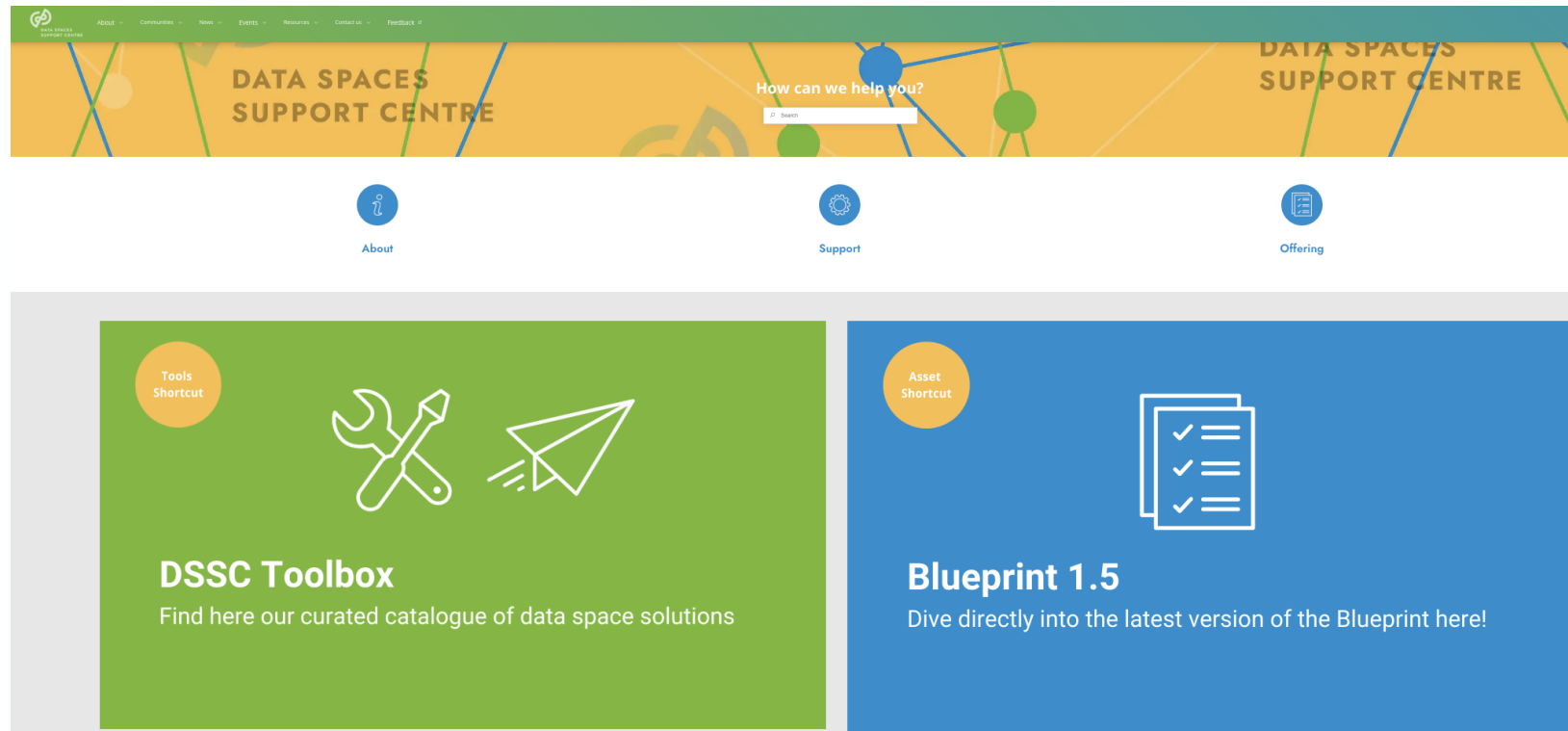
user management, infrastructure management, data and services management

enabling seamless operation and interoperability with the ESA managed ecosystem, European and national initiatives,
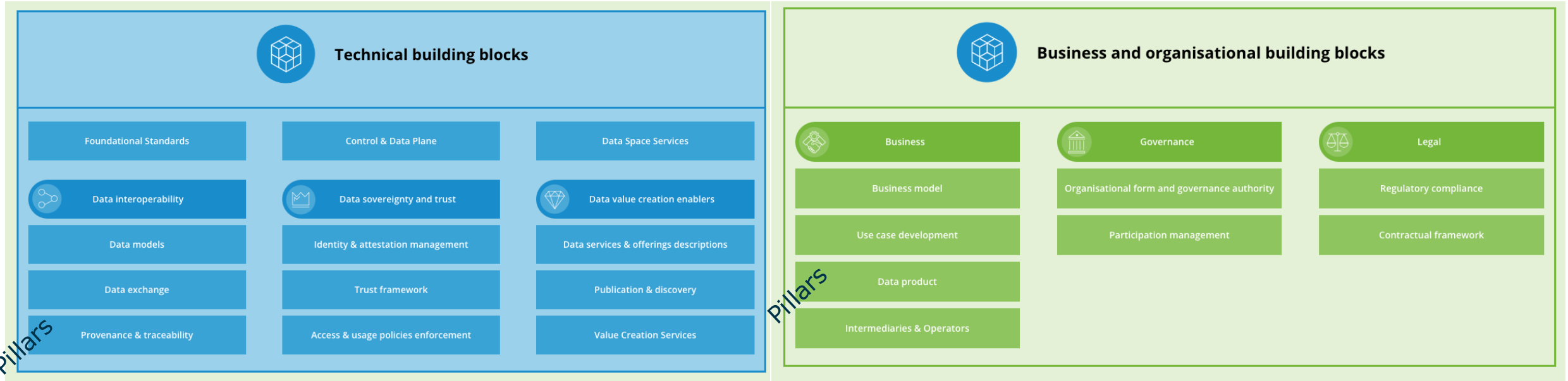
the integration of complementary community services and open-source Collaborative and DTE elements,

while addressing the challenges of the "Platform Jungle".

Data Spaces Support Centre

Many activities on-going to adapt and verify compliance to DSSC Blueprint

- https://dssc.eu/space/bv15e/766061169/Data+Spaces+Blueprint+v1.5+-+Home

Building blocks

Building Block is a basic unit or component that can be implemented and combined with others to achieve the envisaged functionality

First specification to be made available for review at next 2025 EOF Workshop (checkpoint)