EO Online Data Access in Big Data Era

**Big Data from Space
Turning Data into Insights**

Nino Pace
Director Consulting Services
19/03/2019

CGI
Experience the commitment®

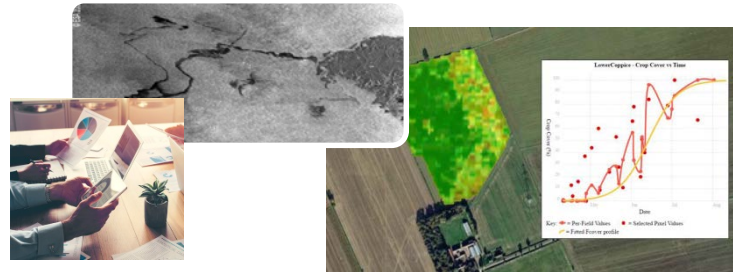# *Sarah is an EO data analyst*

CGI

Efficience and comfort

Rich data catalogue and easy to discover

Search by theme or data type

Efficient and reliable download

Whole products or subsets

Gerard is a programmer

Flexibility and standardisation

I need to …

perform **hosted processing**

run my own **Jupyter notebook**

browse/search data using a **simple REST API**

use **standard protocols**

```
1   package com.cgi.altipay.commons.entity
2
3   import com.cgi.altipay.commons.entity.*;
4
5   dialect "mvel"
6
7   rule "V_R_DIFFERENT_AMOUNT" @Controle("V") @Type("S") @Libelle("Different de R-1 en valeur") enabled tr
8       when
9           $ds : Ds(identificateur in (
10          // Liste des ID
11          "BONUS"),
12          // Executable uniquement si valeurs numériques
13          this.isRNum == true, this.isR1Num == true);
14          $param : Param(paramCode == "&V_DIFFERENT_AMOUNT_"+$ds.identificateur,
15          // Règle : Valeur absolue (Valeur R - Valeur R-1) >= paramètre
16          Math.abs($ds.valeurRNum - $ds.valeurR1Num) >= paramValueNum);
17      then
18          System.out.println("V_R_DIFFERENT_AMOUNT, " + $ds.matricule + ", " + $ds.valeurR1 + ", " + $ds.valeurR);
19      end
```

Build | Deploy

Deborah is the Data Service Manager

CGI

I want to have…

highest user impact

minimal changes on **legacy systems**

Maximise user impact
Cost effectiveness
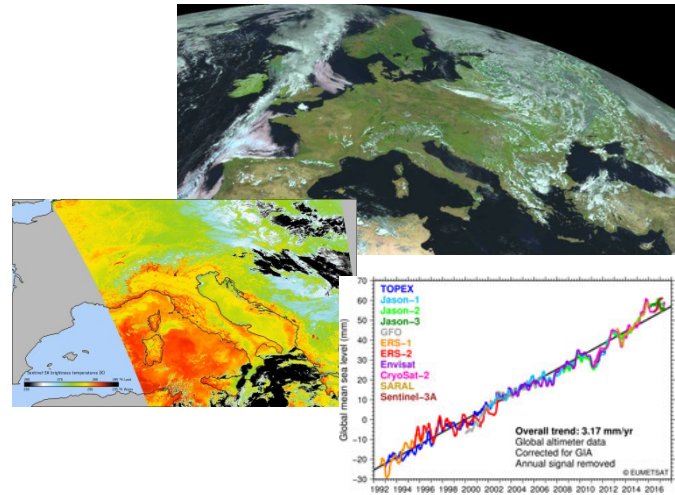
varying **service levels**

cost effective and **scalable** solution

heterogeneous data

… more challenges

- Common content delivery services set high expectation
- Unpredictable user data access load
- Scale while optimising resources cost
- Guarantee service levels

EUMETSAT Data Service Roadmap: OLDA Project

- EUMETSAT Launched Pathfinders projects within Data Service Roadmap, for *Next Generation data services*.
- CGI consortium was awarded On-Line Data Access (OLDA), where the current solution was developed

- We applied leading edge technologies for:
  1. Discovery search and access
  2. Storage and data management
  3. User management integration
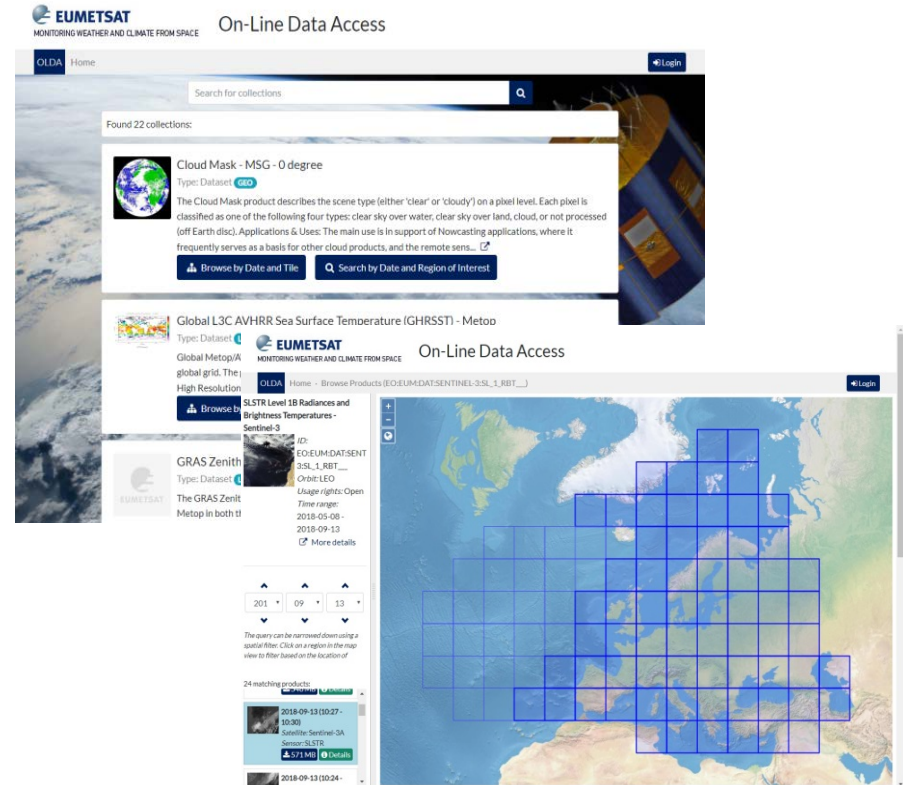 4. SLA scalability/reliability control

User experience – Discovery and Access from UI

- **Discovery:** selection of a Data Collection and presentation of:
  - Metadata
  - General description
  - Sample picture
- **Selection**: the user can navigate:
  - Equi7 Grid
  - AOI and TOI
  - Metadata
- **Download** :
 - Full product in original format
  - Product subset

User experience – API Discovery and Access

- The data access API, defined in OpenAPI (2.0/Swagger) is implemented by 3 components:
  - Browse API
  - Download API
  - OpenSearch-EO interface

- The Browse API associates the needed resources to predefined URL paths, allowing implementing browsing and navigation similarly to the UI use case

19-02-2019

Storage and data management

- Data often span across multiple storage domains.
- Object Storage scales well, but normally does not allow subsetting
- Configurable rule engine dynamically associate data to different Object Storage Providers.
- Elasticity between on-premise and external clouds possibly with different service quality
- Data granularity:
  - Whole products in original format
  - Product subsets (range queries)
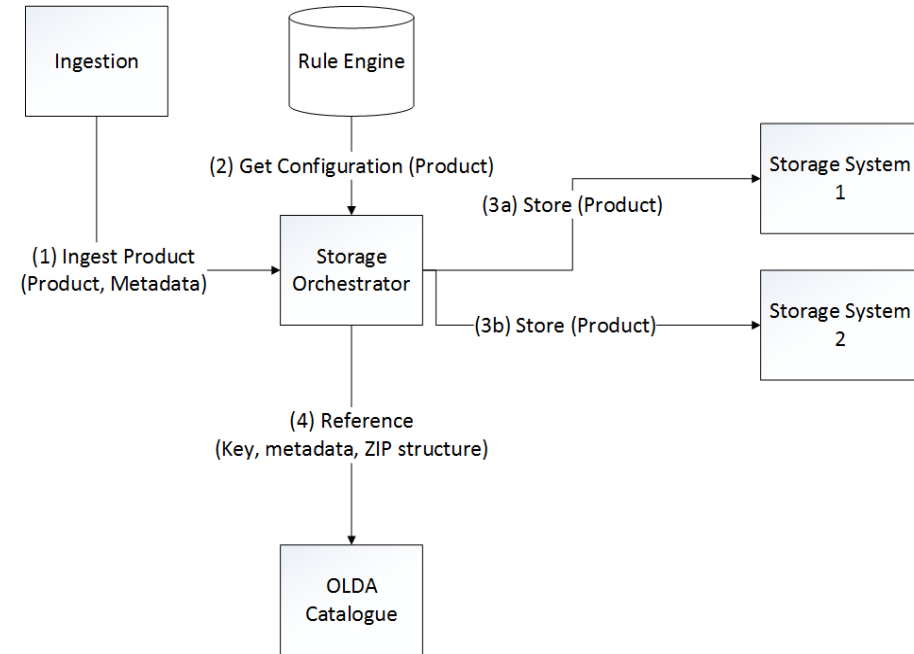
User management integration

- Challenges:
  - API access control, not just SSO !
  - Integrating legacy systems
- Solution:
  - based on API Gateway (WSO2 API Manager and Identity Server)
  - OAuth2
- Strengths:
  - simplify legacy systems integration
  - M2M authorisation using OAuth2, opening to virtual marketplaces
  - Enforces security with tokens' configurable duration and scope

19-02-2019
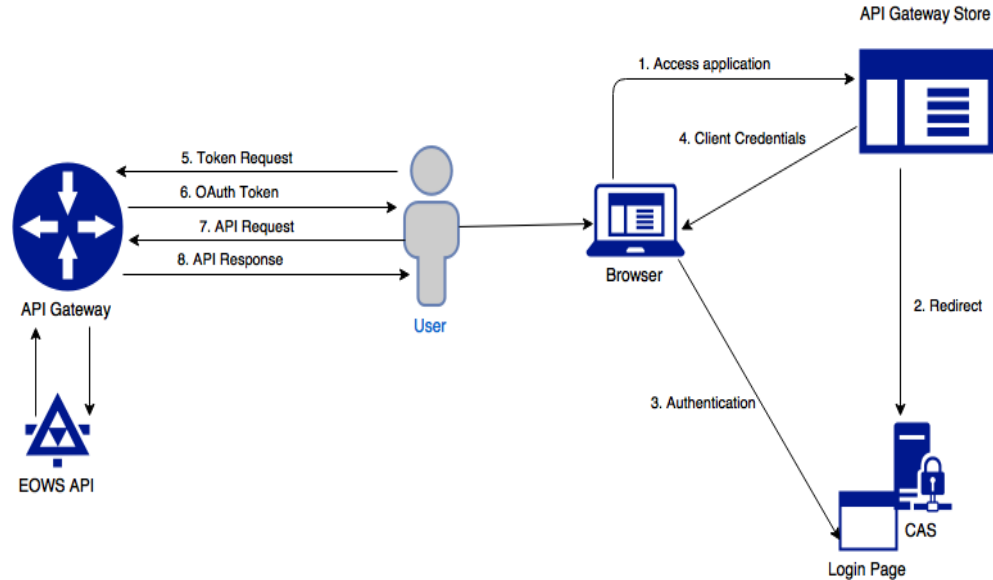
… and Kubernetes?

- **Velocity**: evolving quickly, while staying available;
- **Scalability**: # service replicas supports auto-scaling;
- **Abstraction** from the infrastructure: applications can be easily transferred between environments;
- **Efficiency**: applications can be co-located on the same node without impacting the application themselves.
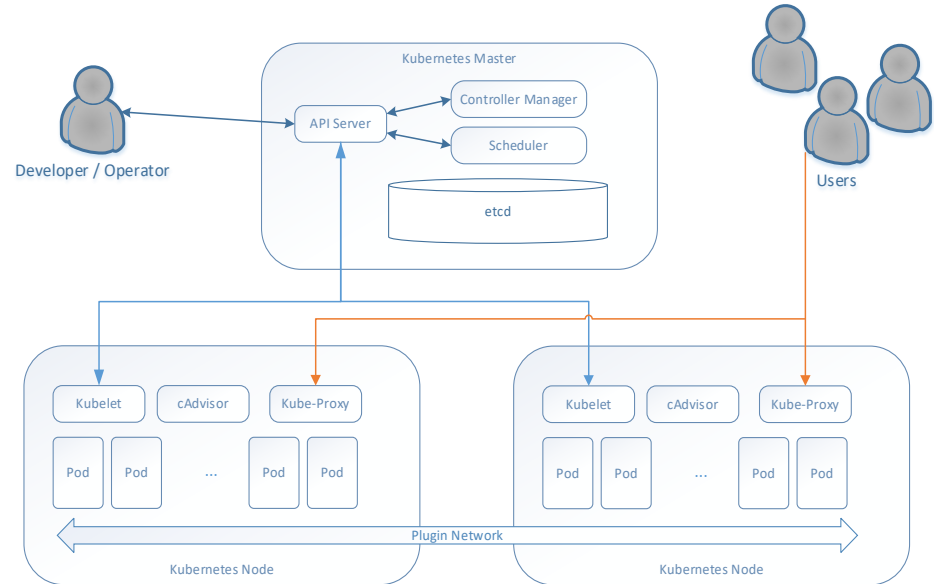
… future developments

- Kubernetes allows building self-healing systems, managing how to reach the desired state.

- *Prometheus* gathers service performance metrics (e.g. calls/sec) to be used with Horizontal Pod Autoscaler to dynamically control the number of active replicas depending on the service load (e.g. scaling pods when the load > 100 calls per second).

- Health checks in Kubernetes can be implemented using liveliness probes, i.e. agents used to know when a container should be restarted.

- Using Istio as a *service mesh* to support:
  - Fine grained traffic control, load balancing
  - Security and service control: rate limiting, quotas, access control
  - Observability: logs, metrics, etc.

EUMETSAT OLDA – summary and conclusions

- A flexible data storage mechanism allows getting data efficiently and with granularity without replication

- Kubernetes technology implements elasticity by autoscaling resources as needed

- A simple modern UI navigates the user providing the needed information at the right time

- API access uses a standard OAuth2 protocol which can be extended to replace legacy identity server with any other identity providers (e.g. a Google or Facebook account)

- The OLDA will be opened to EUMETSAT member state users in the course of 2019 via a pilot service phase and then scaled into operations
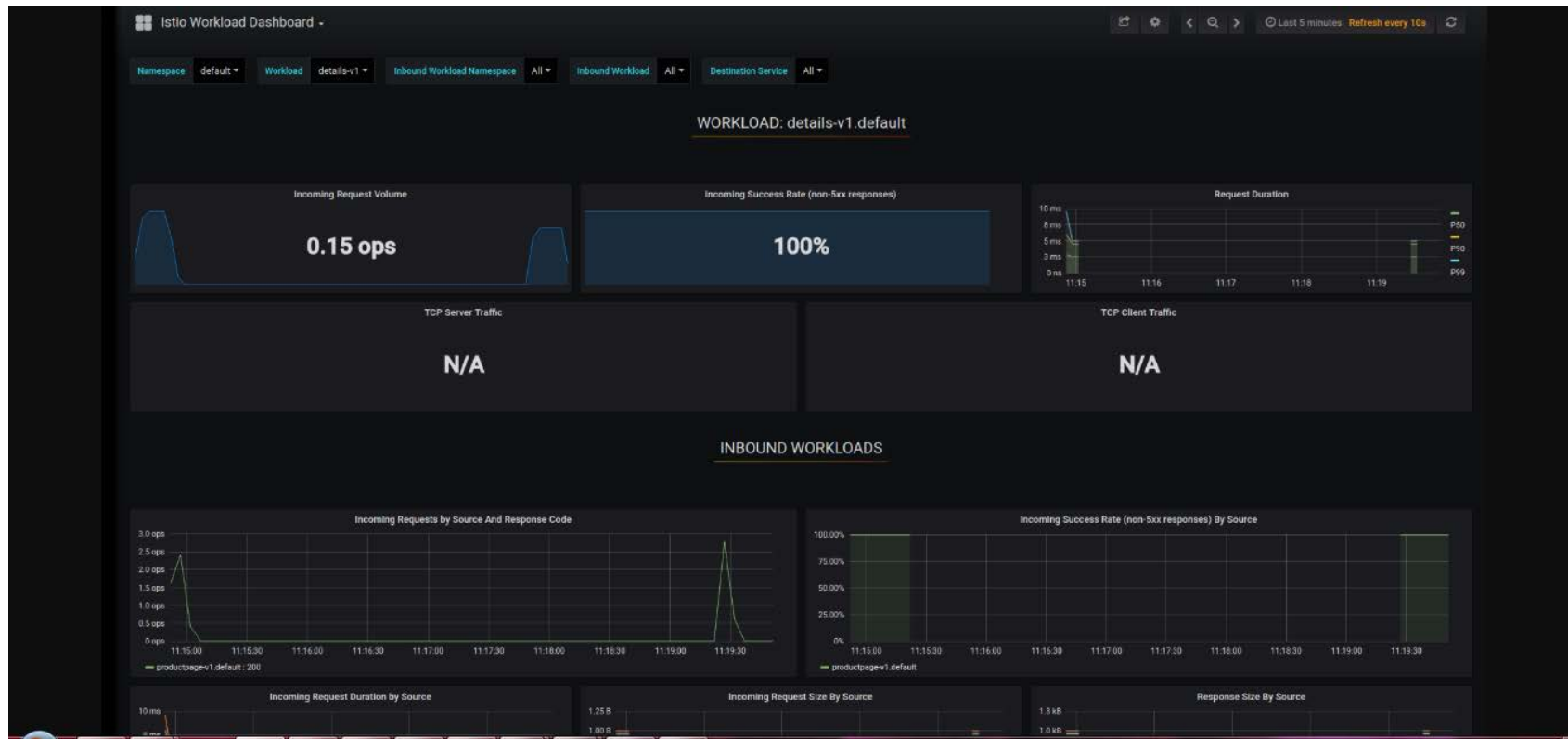
Thank you

Nino Pace - CGI

Extra slides

The *Istio* dashboard

Istio features

Istio it is an open source service mesh. It uses an extended version of a high-performance C++ proxy, to mediate all inbound and outbound traffic. Istio main features are:

- Traffic Management:
  - Automatic load balancing for HTTP, gRPC, WebSockets and TCP traffic
  - Fine-grained control of traffic behaviour with rich routing rules, retries, failovers, and fault injection.
- Security:
  - A pluggable policy layer and configuration API supporting access controls, rate limits and quotas.
  - Secure service-to-service communication in a cluster with strong identity-based authentication and authorization.
- Observability
  - Automatic metrics, logs, and traces for all traffic within a cluster, including cluster ingress and egress.

Let's see Istio in action: https://www.katacoda.com/courses/istio/deploy-istio-on-kubernetes