



CloudFerro

Federated data access for the digital model of the Earth.

EODC Forum, Vienna, 9.05.2023

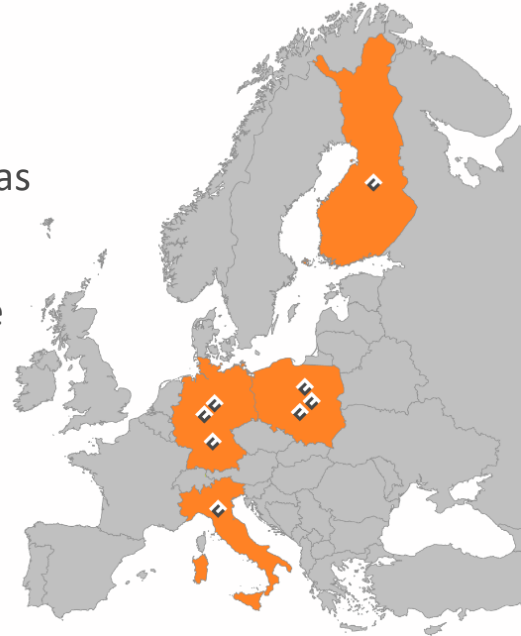
Michał Bylicki

mbylicki@cloudferro.com

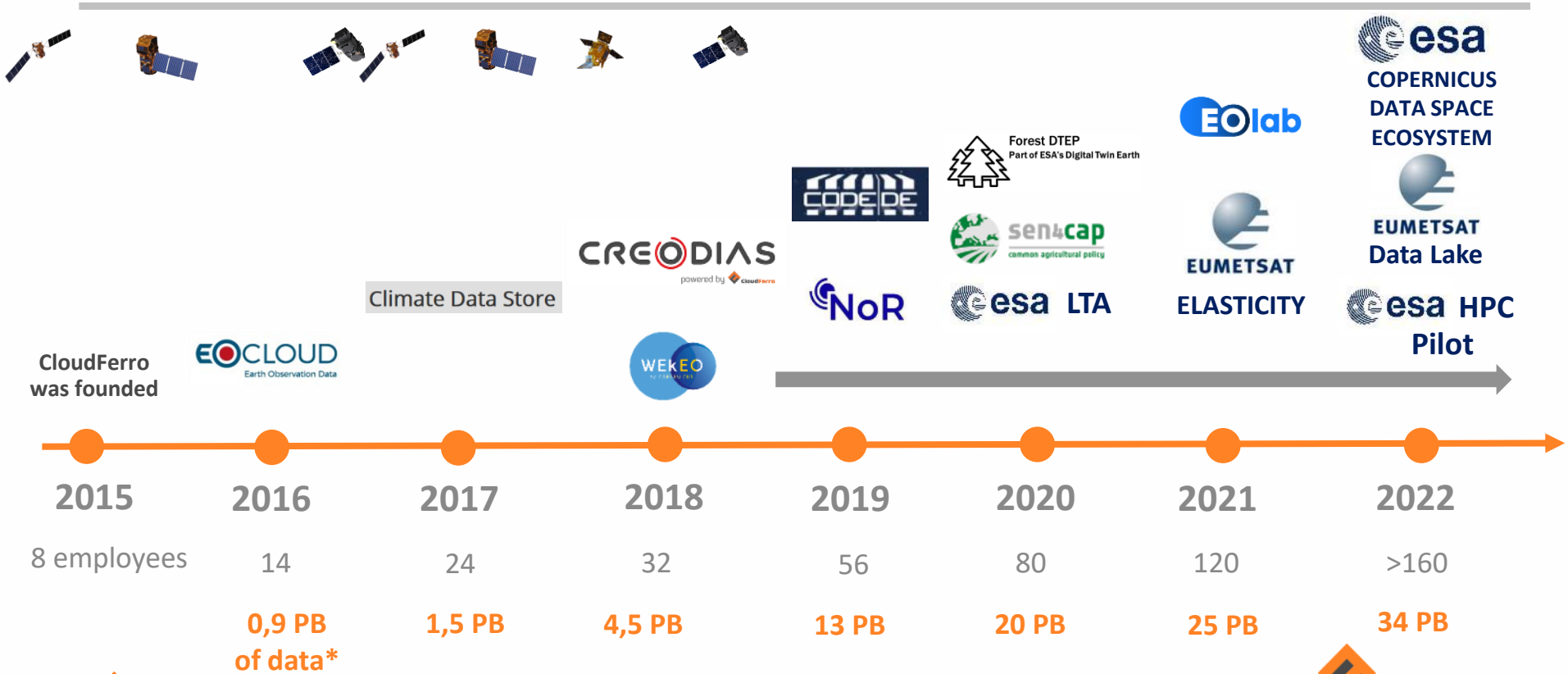


About CloudFerro

- **European, private, technological company** founded in 2015
- Provider of **dedicated cloud computing services**
- Delivers and operates cloud platforms for **demanding markets**, such as the European space sector, climate research and science
- Specialized at storing and processing **big data sets**, like multipetabyte repositories of EO data
- Provides **flexible, open-source** based, tailor made cloud solutions, matching best technical and organizational practices to business requirements
- Achieved **technological autonomy** and guarantees reliability and independence of delivered services thanks to full control of technology stack



Our journey



* Amount of EO data publicly available for users

Public (or private) Cloud



Computing Services – public cloud environments providing Virtual Machines (VM), Virtual Machines with local storage (VM.local), Dedicated server virtual machines (DS), GPU, Kubernetes



Backup and long term Storage – cost efficient storage in secure remote locations targeted for backup, archive and dark storage



Network Storage – object, block and file storage with different performance and availability tiers dependent on storage media (NVMe, SSD, HDD), replication and location



Internet Access – Public Network access with geographical and carrier redundancy; direct access to GEANT academic network

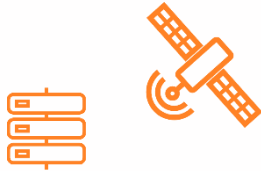


Local Storage – local NVMe and HDD drives for Virtual (VM, VM.local) and dedicated computing instances (DS, BM)

Cloud for EO

CloudFerro products and services for Earth Observation sector are based on a consistent cloud computing system delivered to customers in a service model (IaaS). Customers receive services or ready to use cloud platform.

CloudFerro IaaS for EO



EO DATA Storage

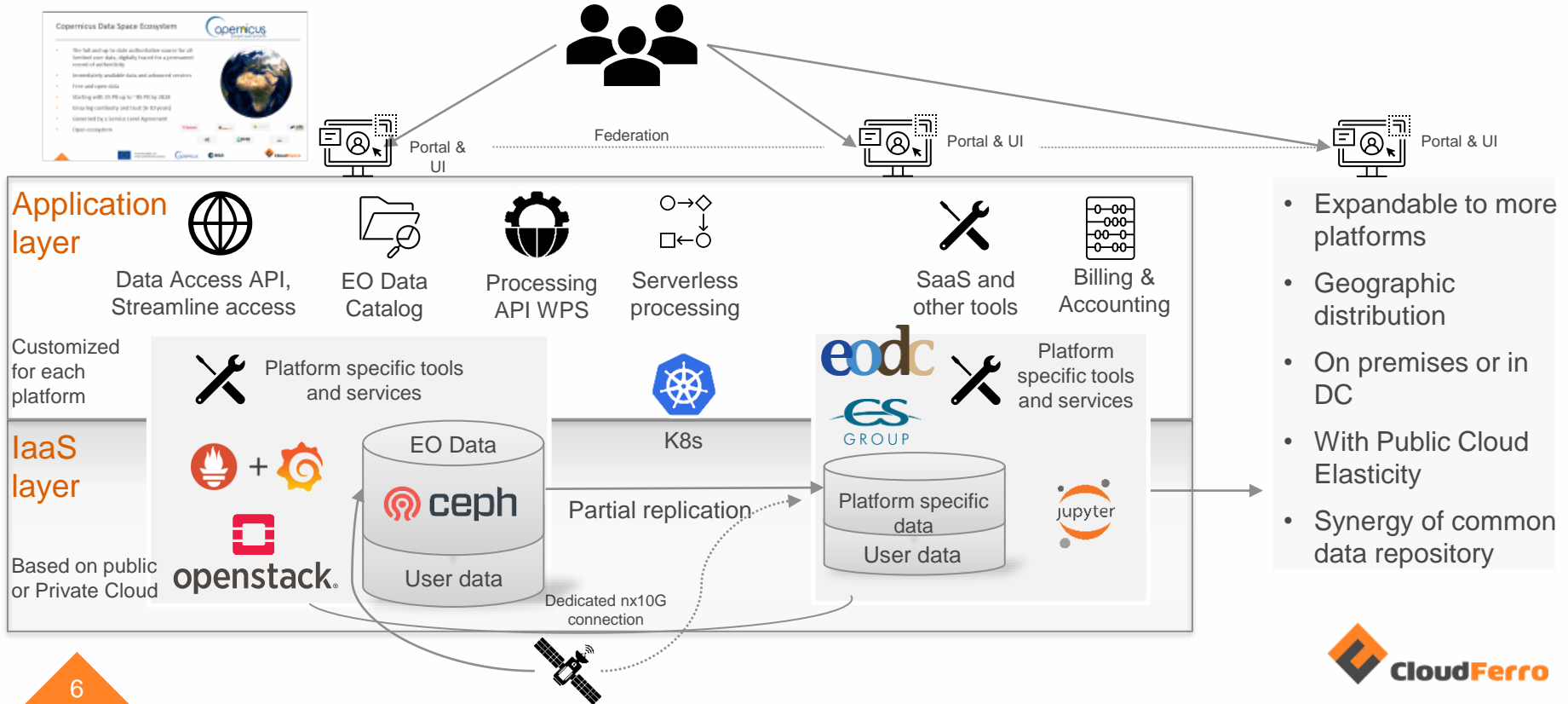
storage of high volume (tens of Petabyte) EO data in cloud related object storage; can be used as LTA or working area for cloud processing



EO DATA Processing (mass processing)

processing in the cloud model “bring processing to the data”, in CSC Ground Segment, third party processors

Outline of multi-platform approach



Copernicus Data Space Ecosystem

- The full and up-to-date authoritative source for all Sentinel user data, digitally traced for a permanent record of authenticity
- Immediately available data and advanced services
- Free and open data
- Starting with 35 PB up to ~85 PB by 2028
- Ensuring continuity and trust (6-10 years)
- Governed by a Service Level Agreement
- Open ecosystem



T Systems

CloudFerro

SINERGISE

vito
integrated sensing



ACRI
ST

RHEA



PROGRAMME OF
THE EUROPEAN UNION



esa



Best Practices and Lessons Learned

Bring processing to data

This is big data!

- Carrier-grade, scalable, redundant Internet access is mandatory for both acquisition and dissemination
- Redundant, scalable storage at 10-s of PB scale
- Provide scalable processing power to enable easy, repeatable, fast, large scale on-demand product generation
- Pre-generate useful datasets up-front to boost their usage
- Avoid multi-step pipelines and bottlenecks in the systems architecture



Federate users and data access

- Users often need to **combine datasets from different sources**
- Considering data sizes, it is **ineffective to keep more copies** of big data than necessary for redundancy
- **Federate** to provide users with transparent access to a large number of datasets
- Provide **homogenous interfaces, high bandwidth and low latency** (requires inter-operator cooperation)
- Ideally, keep a **common catalogue with references** to multiple data sources



Distribute data in standard, easily accessible formats

- Different users need different access methods:
 - ✓ HTTP/ object (S3) for remote,
 - ✓ filesystem for local access (NFS, S3FS)
 - ✓ OGC WMS/WMTS for tiled access
- Store data in uncompressed (unzipped) formats, use cloud-optimized geotiffs for fast sub-granule access
- Avoid unnecessary steps – download, copy, decompress
- Provide fast, homogenous catalogue tool with API and GUI



Summary of best practices and lessons learned

- Bring processing to data
- Distribute data in standard, easily accessible formats
- Federate users and data access
- And store (almost) everything you have observed or generated!





www.cloudferro.com

Michał Bylicki

mbylicki@cloudferro.com



facebook.com/cloudferro



linkedin.com/company/clfr



twitter.com/CloudFerro