# Operational strategies for a continuously growing public-cloud archive

Richard Hofmeister, Torben Keßler, Alexander Strecker, Arne Petersen, Quentin Saalfeld, Knut Bernhardt
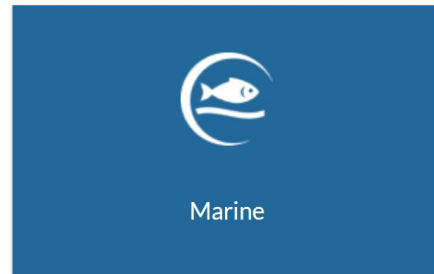
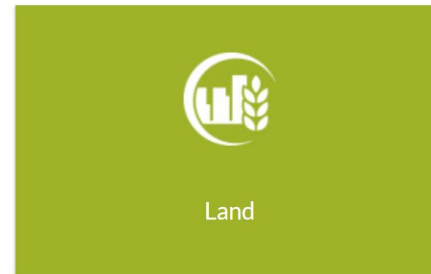[Werum Software & Systems AG]

Earth insights with Copernicus

Copernicus is the Earth Observation Programme of the EU,
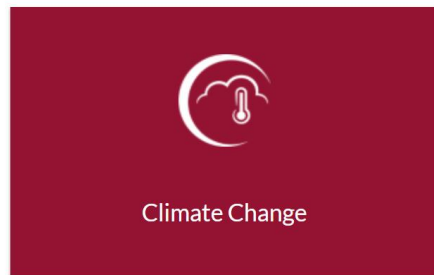to offer information services to all citizens of the EU member states
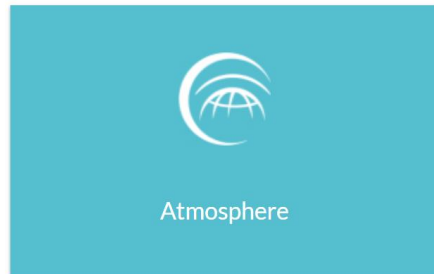
- satellite data
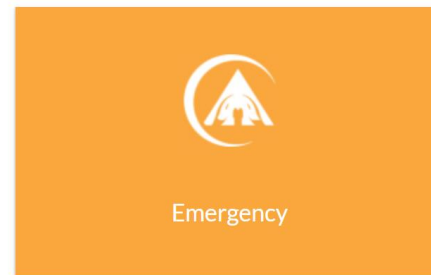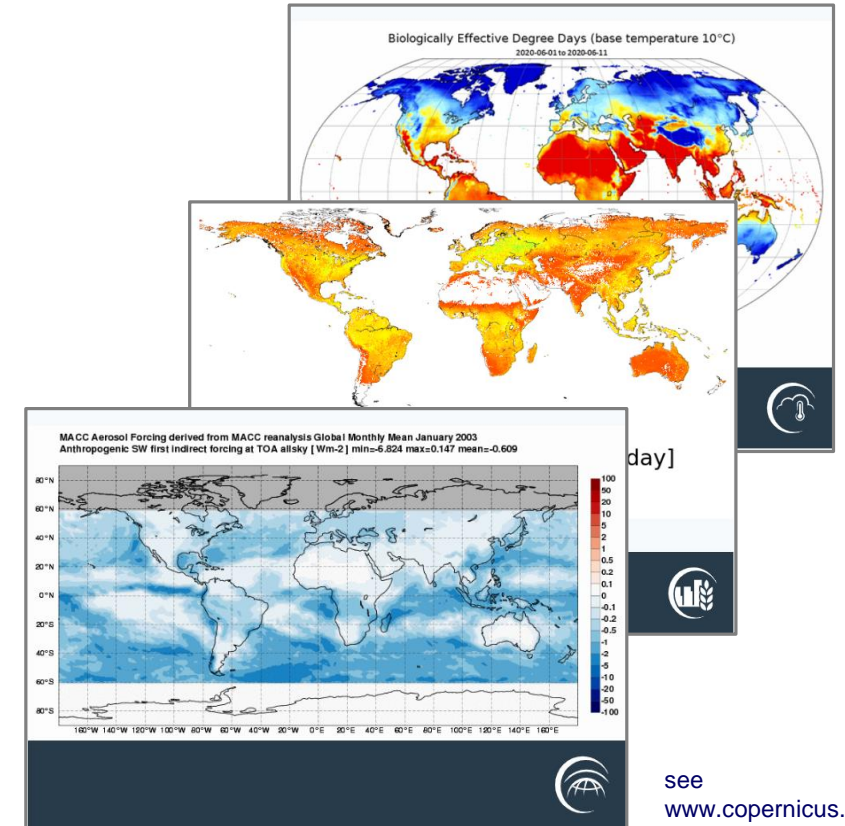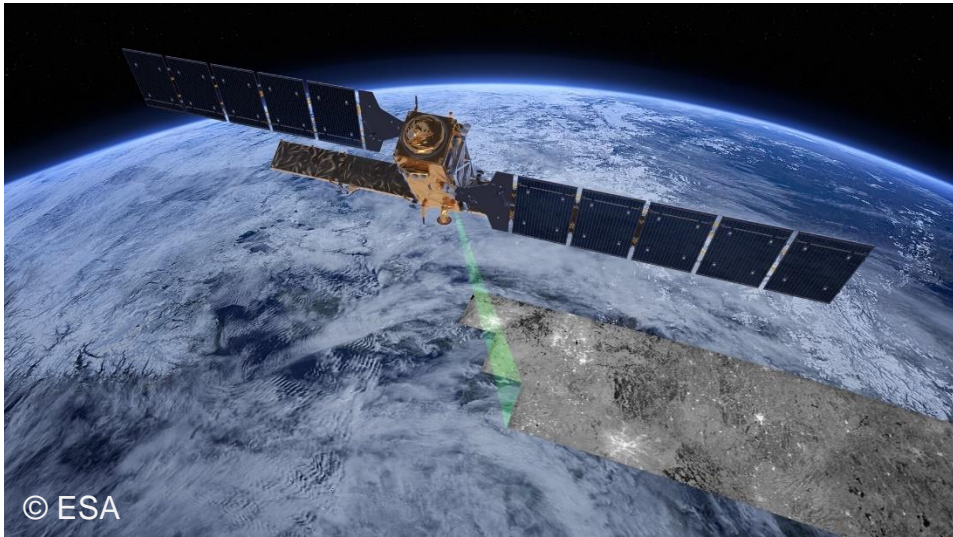- in-situ observations
- data from models & analysis
- assessment studies

see
www.copernicus.eu

Earth observation: **The Sentinel satellites**

Sentinel 1

© ESA

Sentinel 2

© ESA

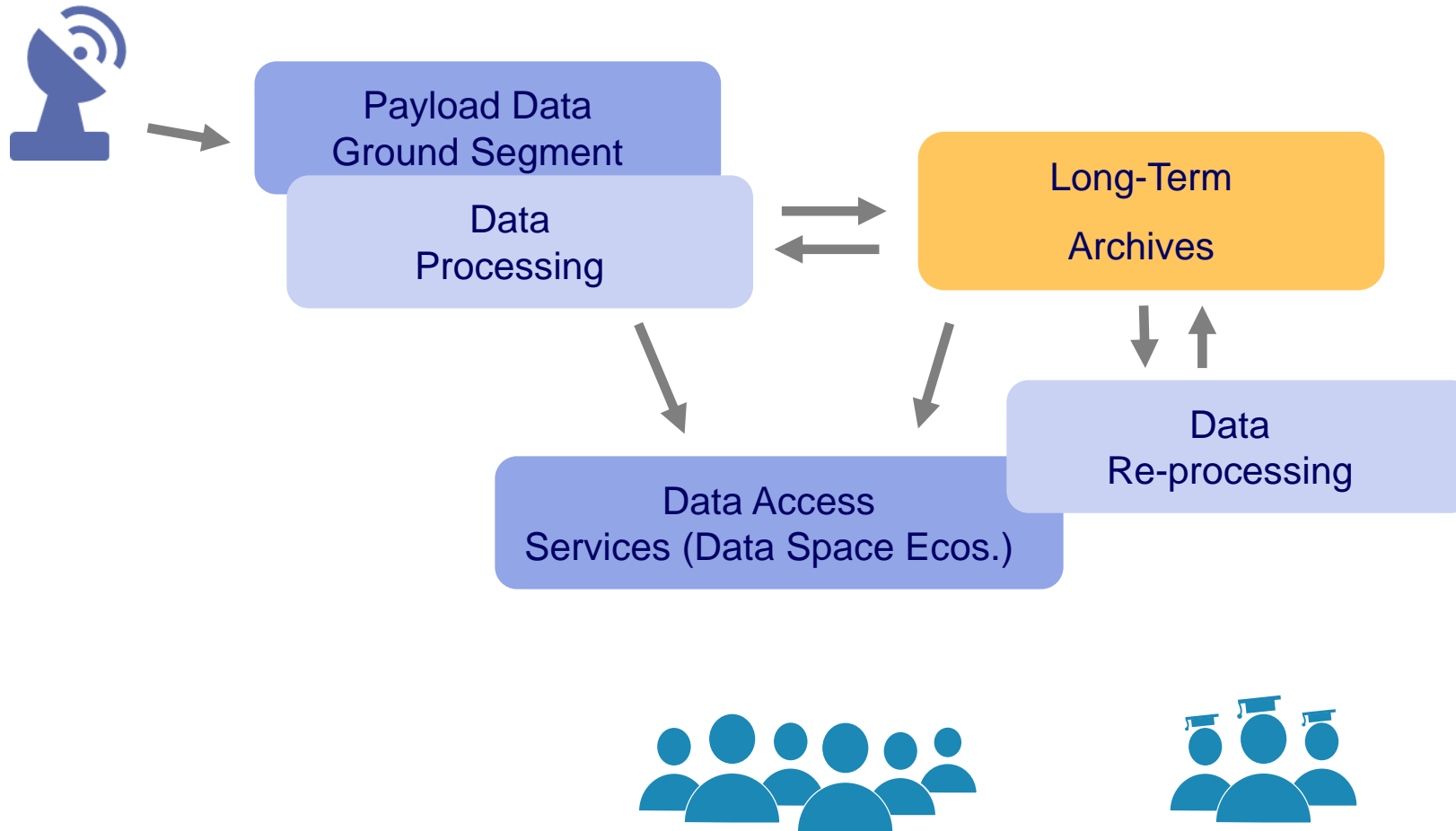
Sentinel 3

© DLR

Sentinel family

© ESA
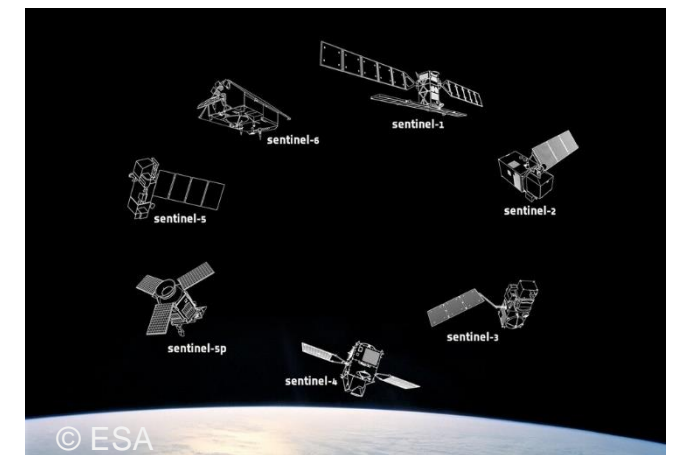
The Copernicus Long-Term Archiving

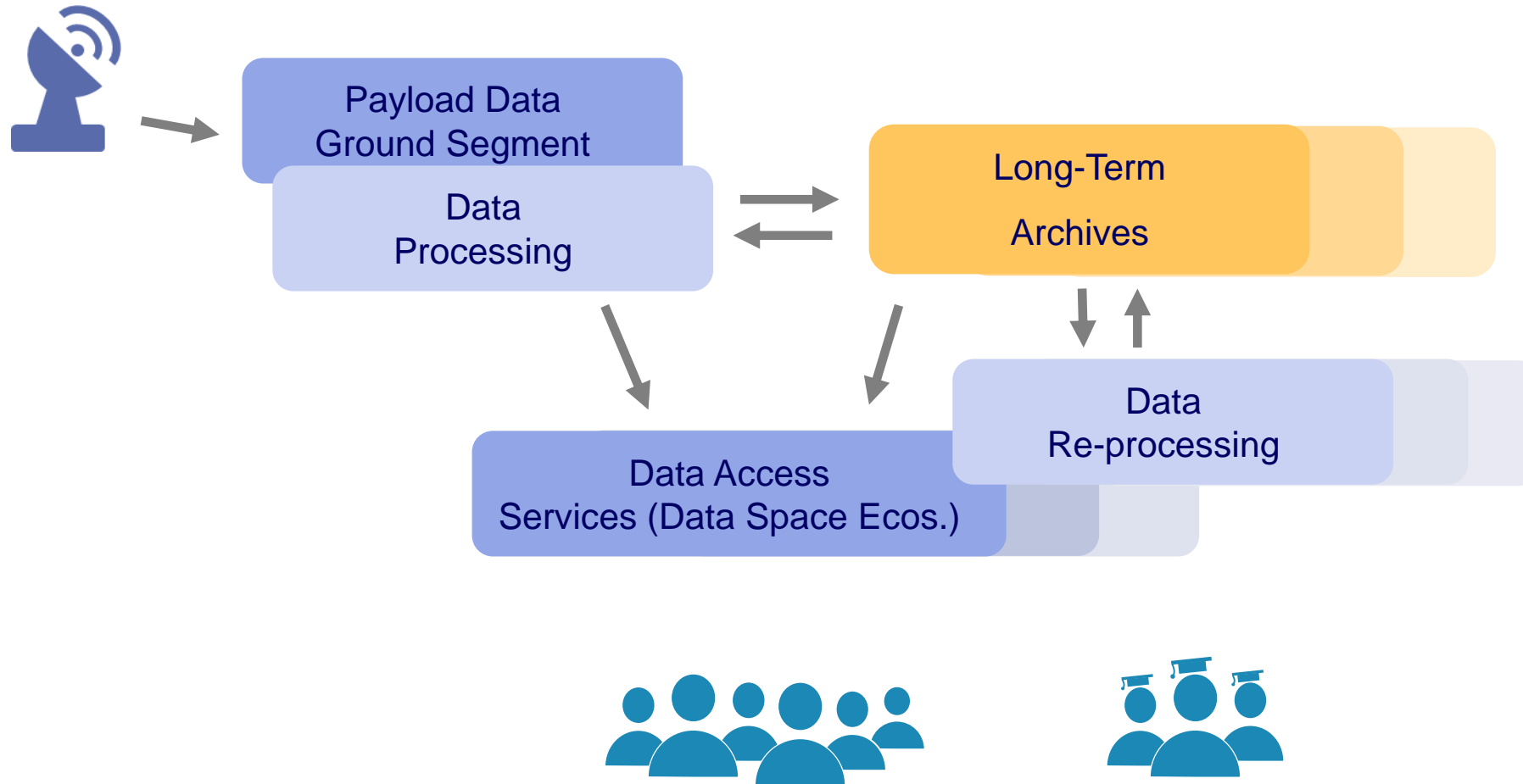**archive functionality**

- collecting data
- data verification
- data traceability
- meta-data catalogue
- request and lookup interface
- provide user download

The Copernicus Long-Term Archiving

Payload Data Ground Segment

Data Processing

Long-Term Archives

Data Access Services (Data Space Ecos.)

Data Re-processing

scalable services for a growing amount of data, users and use-cases

# The Copernicus Long-Term Archive Service
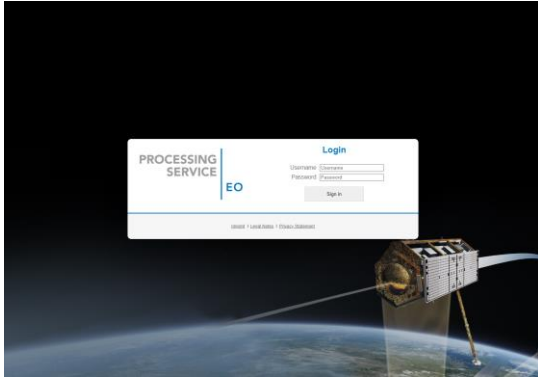
- Long Term Archive as a Service

- Self-funded software solution based upon processing orchestration service

- Next generation Copernicus interface point compliant (OData 4: AIP, PRIP)

- Cloud deployed

- DevOps approach

- All Sentinel-1, -2, -3 Level 0 and auxiliary data

- 125 TiB and more than 3 million products per month
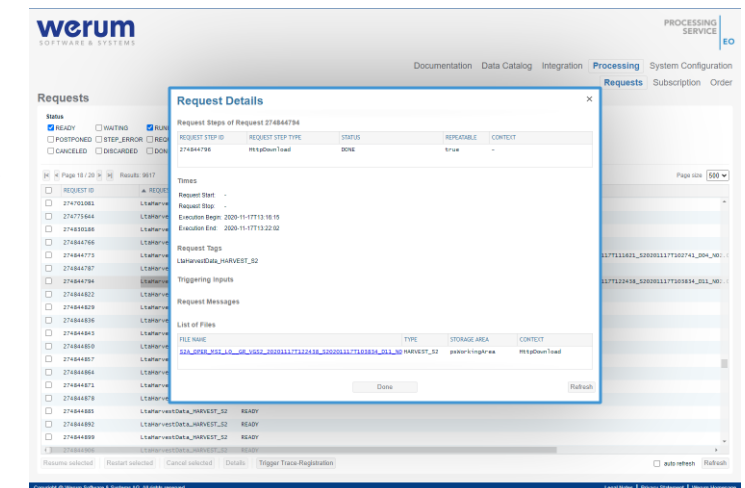
Copernicus Archive: **Data Set Characteristics**

- monthly data volumes
  - S1: 50 TiB, 65~73k products (25TiB since 2022-01; ~40-50k products),
  - S2: 55 TiB, 2,7-3,7 Mio products
  - S3: 20 TiB, 65-73k products

- very different product sizes (1 GB vs. 17 MB) and types (meta data, payload)
- additional historical dataset (>6 PiB), >155 Mio products in 6 months (250 TiB/week)
- product ordering and download (up to 10 TiB/day)
- Currently 266M products , 9.5 PiB
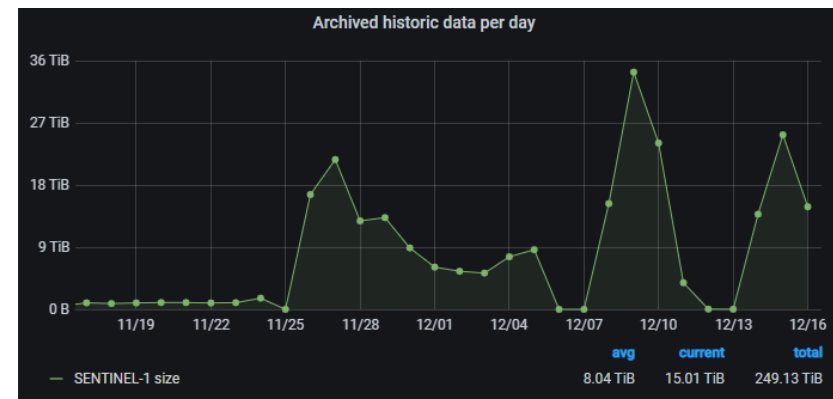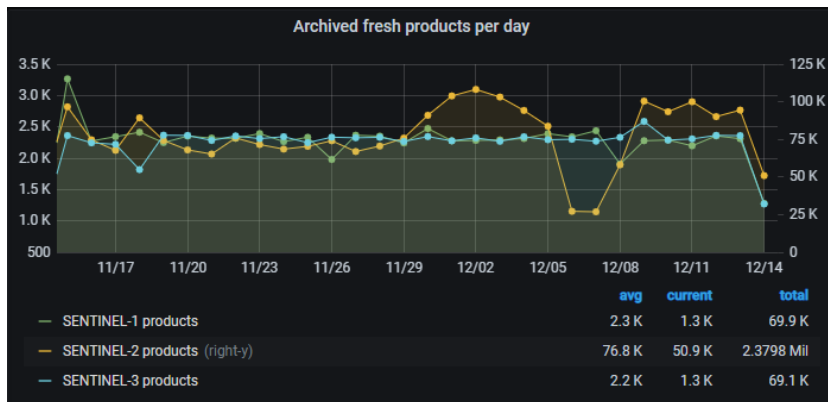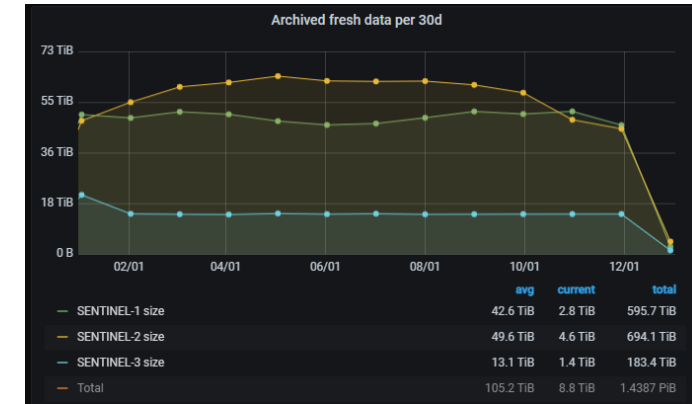
Operational strategies

- Service is based on reliable, automatic software

- Infrastructure layer managed externally in public cloud environment

- Agent based software is scalable to minimise costs and energy use (carbon footprint)

- Desaster recovery possible from tape backup

- DevOps allows for short communication lines and quick issue analysis

- DevOps enables efficient evolutions of the system:
  * configurability of scaling capabilities
  * changes induced by modifications
     of external services
  * optimisations
     (performance and anomaly management)
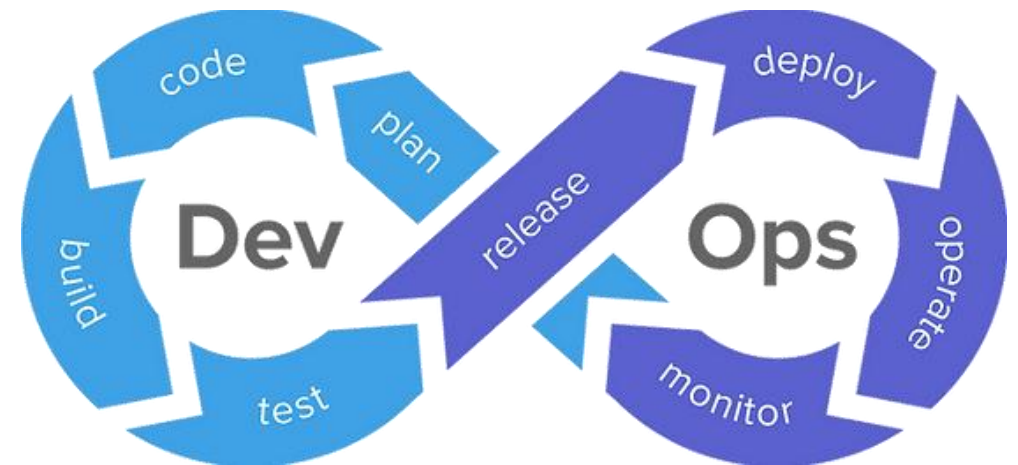
Operational strategies: Anomaly Management

- Service is based on reliable, automatic software
  -> only few anomalies expected per month
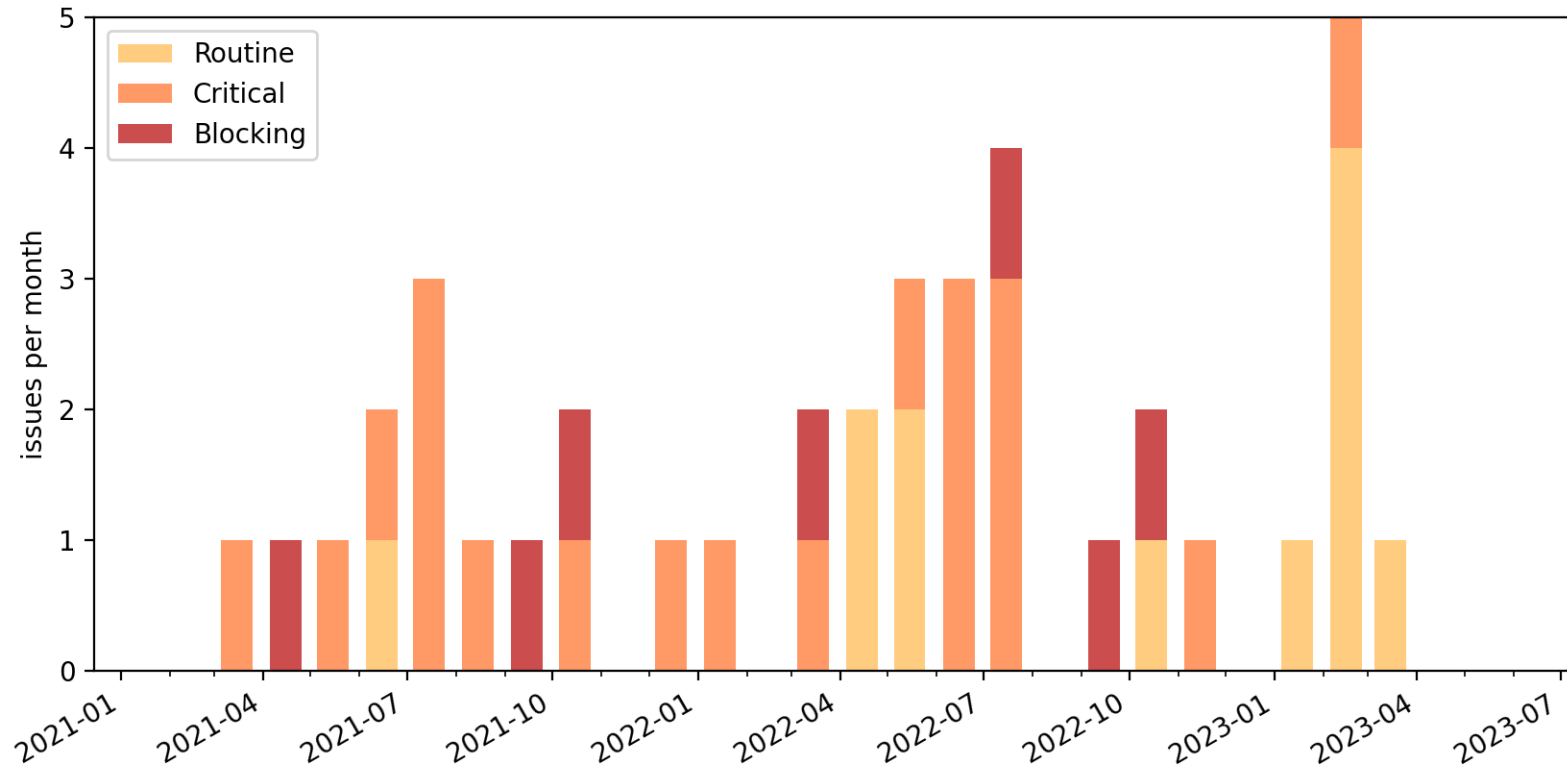
Operational strategies: Anomaly Management

- Service is based on reliable, automatic software
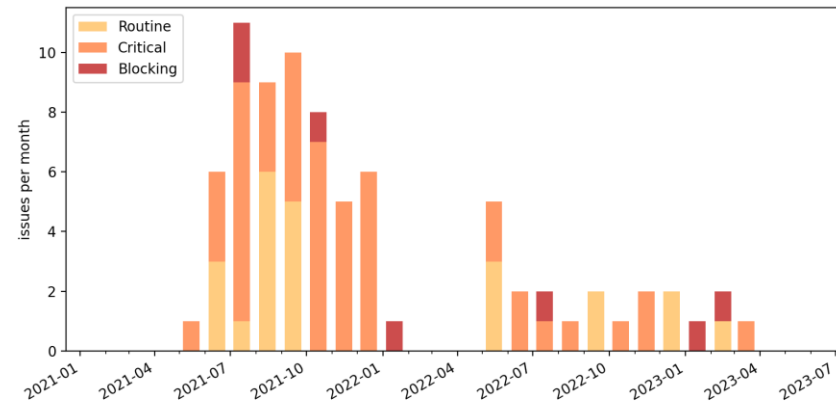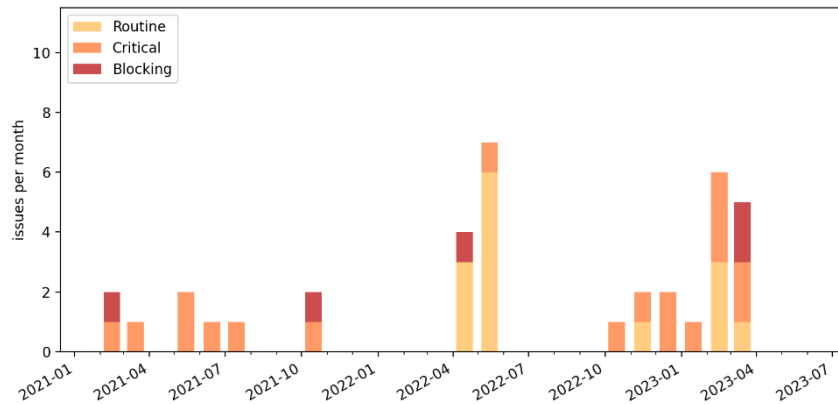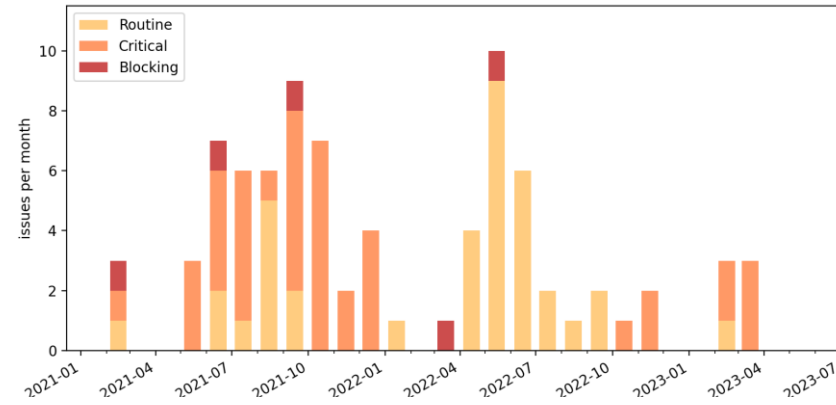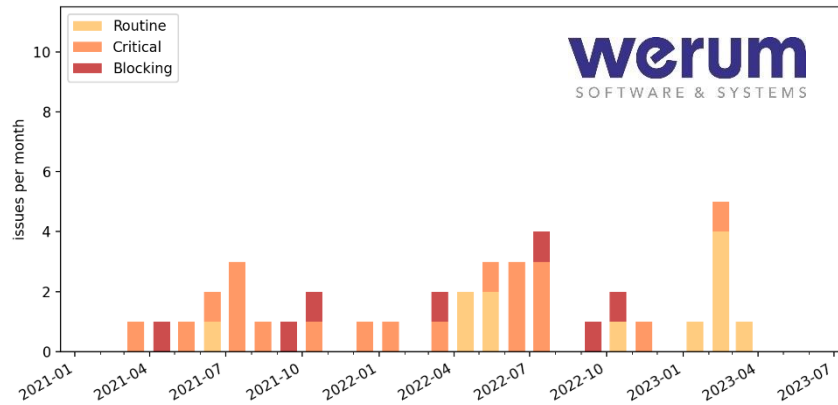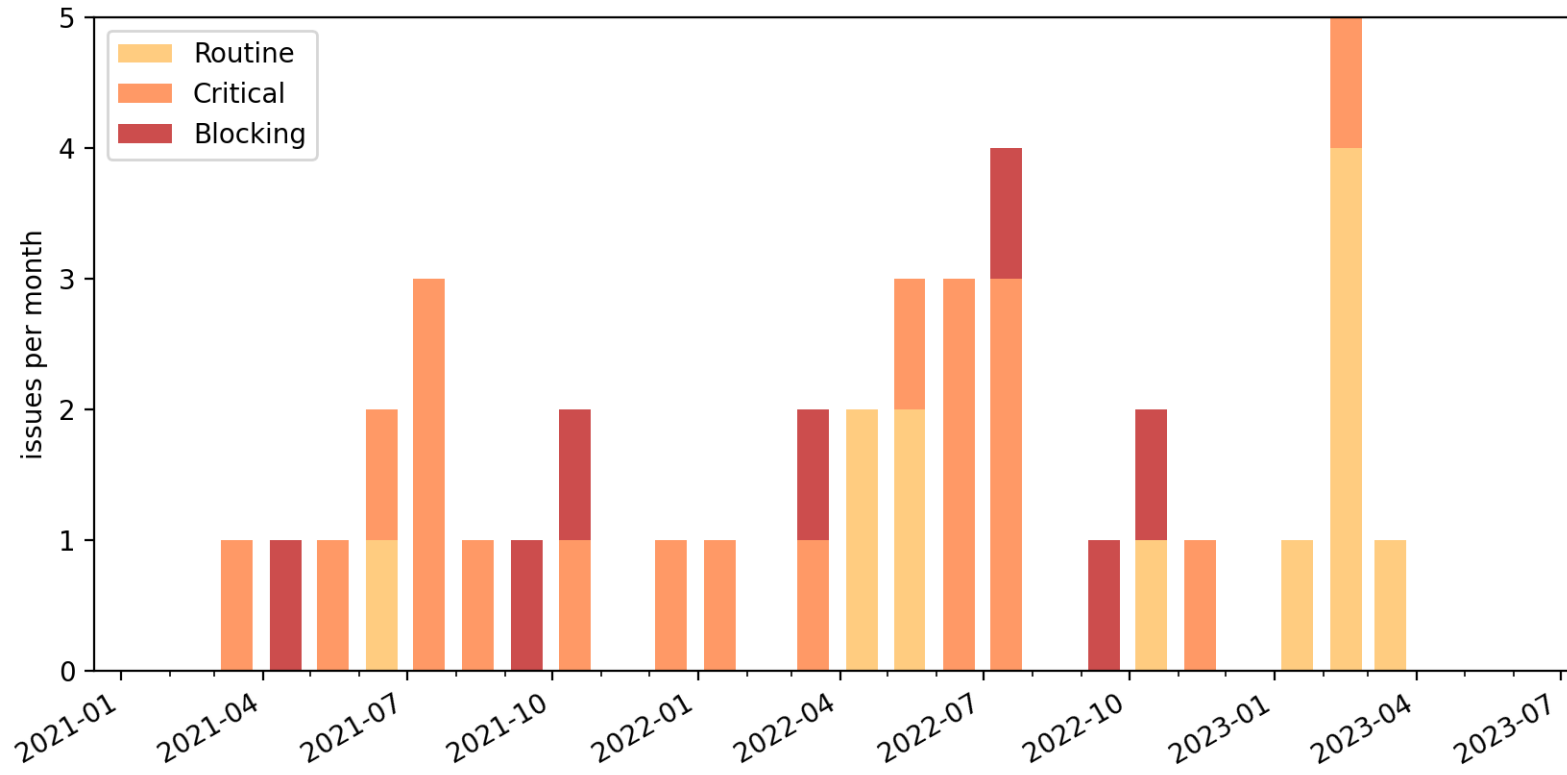 -> only few anomalies expected per month

Operational strategies: Anomaly Management

- Service is based on reliable, automatic software
 -> only few anomalies expected per month

Operational strategies: Anomaly Detection

- Log data is analysed with deep-learning LSTM (long short-term memory) model
- Indicator time series is compared to expected behaviour
- Time window and threshold can be configured for anomaly alarming
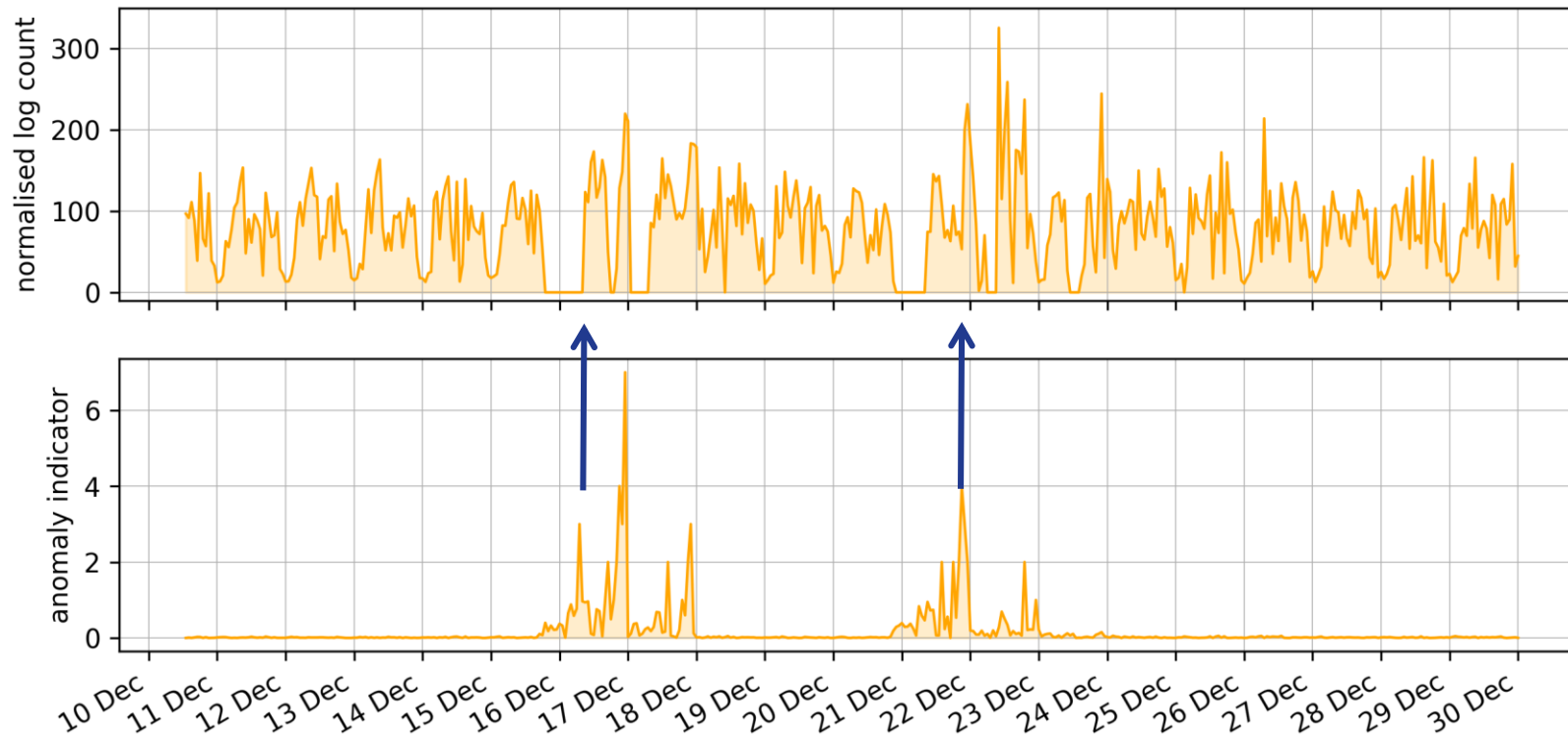
Summary

- Successful transformation of Copernicus data archive(s) into cloud service(s)

- Growing amount of data organised with adjustable service infrastructure

- DevOps is key concept for service evolution

- Anomaly management main uncertainty

- Newly developed anomaly detection implemented and tested in operational environment

- Extensions of anomaly detection possible,

    e.g. * refine detection parameters/training

        * to watch other entities in the

            data management ecosystem
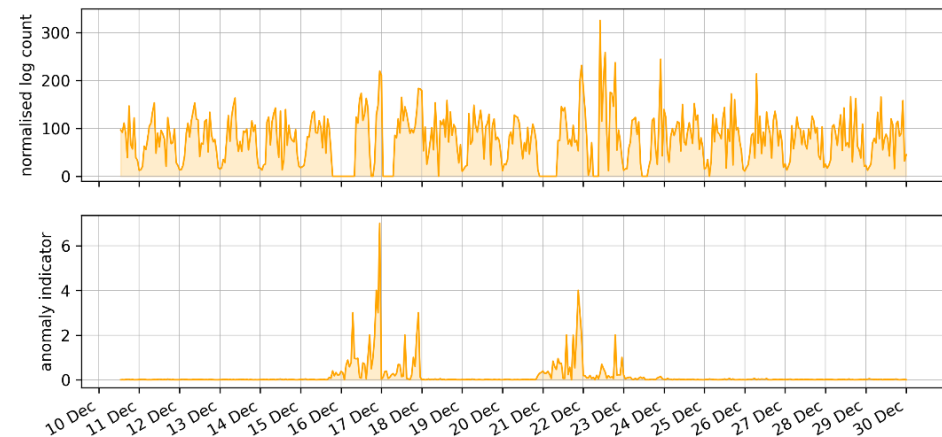
Operational strategies for a continuously growing public-cloud archive

Richard Hofmeister

richard.hofmeister@werum.de

Werum Software & Systems AG