Reflections on the big data thematic

Lothar Wolf

Competence Area Manager for Data Services

Lothar.Wolf@eumetsat.int

Abstract

- The "Big Data" topic is one of the most heard buzz words within today's IT and Data community. This presentation reflects on some aspects of the big data hype and addresses some key issues as they are perceived by the data providers and the users. An important aspect in this discussion is the notion that users often need data in order to get answers to their questions – which is not a new issue as such however, in the big data era the generation of such answers is also due to the exponential growth of the base data representing a significant challenge for creating the most important aspect: the value. Furthermore the simple existence of such amounts of data together with new science and tools, data mining, data analytics & correlations allow the generation of questions that were not even thought about in the beginning.

- The question is on how this could be supported and to what extend by the Data providers and others.

- An overview is presented on how EUMETSAT considers those aspects within the context of its existing operational data services but also their big data related evolutions. In this context the different models of data delivery (push – pull & Near-Real-Time – online/offline) are discussed and the bridge a data provider might build in order to make steps towards the user for easier access to the actual information the user requires.

EUMETSAT

EUMETSAT is an intergovernmental organisation with 30 Member States and 1 Cooperating State

30 YEARS 1986-2016

Member States

AUSTRIA	BELGIUM	BULGARIA	CROATIA
CZECH REPUBLIC	DENMARK	ESTONIA	FINLAND
FRANCE	GERMANY	GREECE	HUNGARY
ICELAND	IRELAND	ITALY	LATVIA
LITHUANIA	LUXEMBOURG	THE NETHERLANDS	NORWAY
POLAND	PORTUGAL	ROMANIA	SLOVAK REPUBLIC
SLOVENIA	SPAIN	SWEDEN	SWITZERLAND
TURKEY	UNITED KINGDOM		

Cooperating States

SERBIA

EUMETSAT

EUMETSAT headquarters

Lothar.Wolf@eumetsat.int

EUMETSAT
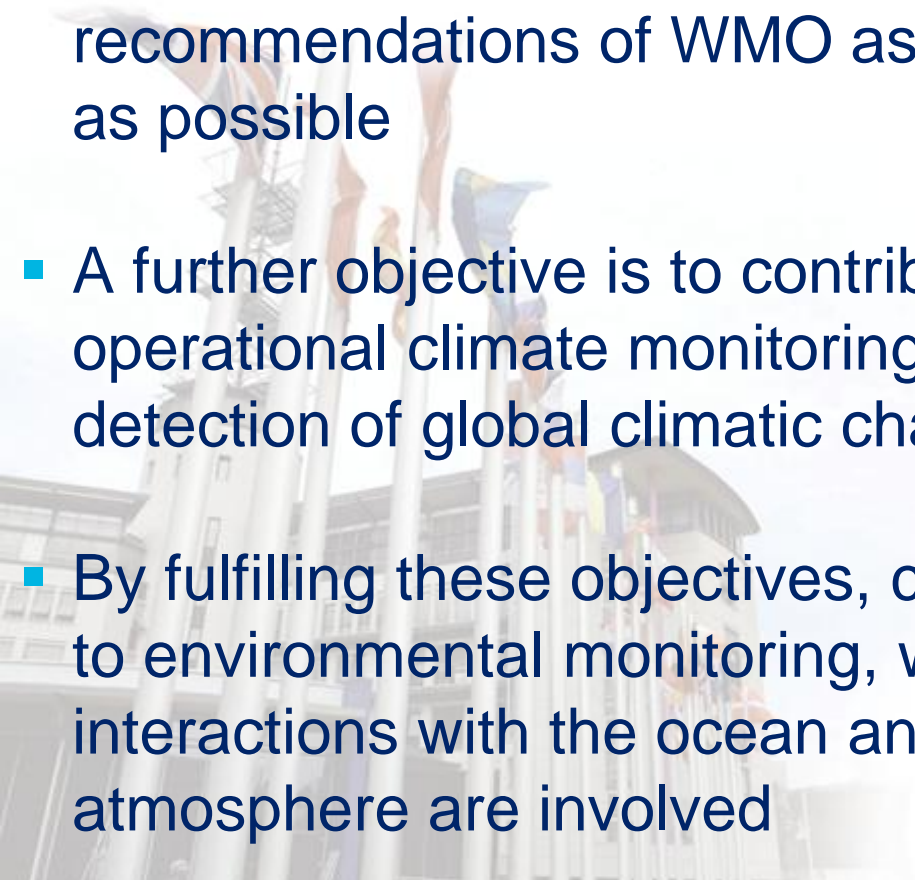
EUMETSAT Mission

- To establish, maintain and exploit European operational meteorological satellite systems, while considering the recommendations of WMO as much as possible

- A further objective is to contribute to operational climate monitoring and detection of global climatic changes

- By fulfilling these objectives, contribute to environmental monitoring, where interactions with the ocean and the atmosphere are involved

- Deliver cost-effective operational satellite data and products that satisfy the meteorological and climate data requirements of its Member States

- Encourage more users to benefit from the increasing range of EUMETSAT data and products

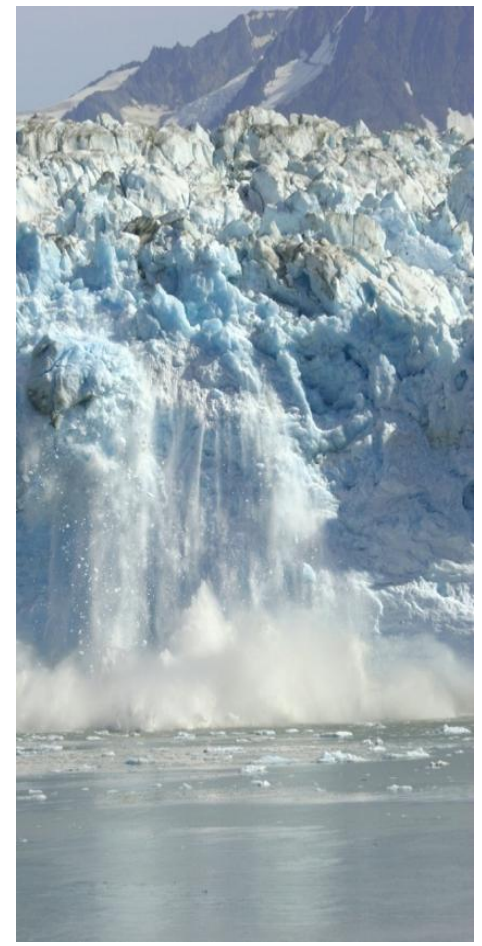EUMETSAT

Areas benefitting from weather forecasting

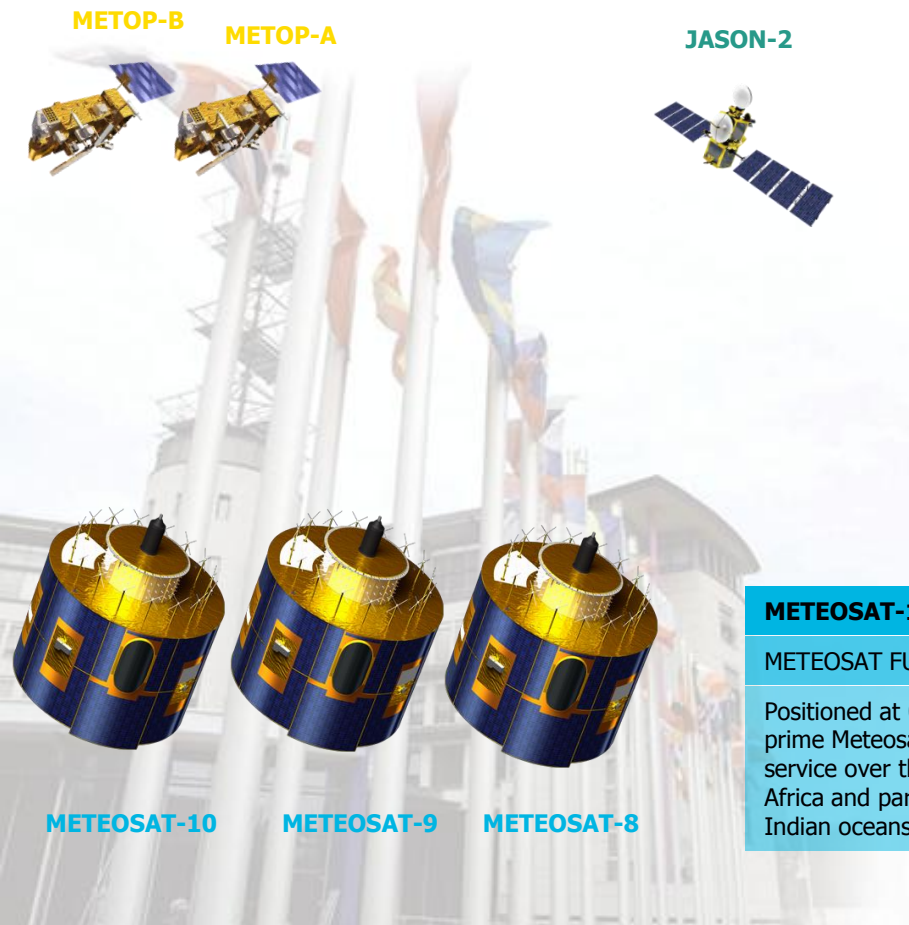**Safety of life, property and infrastructure**

Transport

Energy, agriculture, tourism

Climate policy and environment protection

EUMETSAT

Current EUMETSAT satellites

30 YEARS 1986-2016

METOP-B

METOP-A

JASON-2

METEOSAT-7

METEOSAT-10   METEOSAT-9   METEOSAT-8

METOP-A (98.7° incl.)

EUMETSAT POLAR SYSTEM

In nominal mid-morning sun synchronous orbit at 817km altitude ,as part of the EUMETSAT Polar System (EPS) .

METOP-B (98.7° incl.)

EUMETSAT POLAR SYSTEM

In orbit at 817 km altitude, the primary operational satellite of the EUMETSAT Polar System (EPS).

JASON-2 (66° incl.)

OCEAN SURFACE TOPOGRAPHY

In nominal non-synchronous low Earth orbit at 1,336km altitude, in support of the Ocean Surface Topography Mission.

METEOSAT-7 (57.5° EAST)

INDIAN OCEAN DATA COVERAGE

Operated in support of the Indian Ocean Data Coverage (IODC) mission, bridging an observational gap in this region.

METEOSAT-10 (0°)

METEOSAT FULL DISC IMAGERY

Positioned at 0° supporting the prime Meteosat full disc imagery service over the European continent, Africa and parts of the Atlantic and Indian oceans.

METEOSAT-9 (9.5° EAST)

RAPID SCANNING SERVICE (RSS)
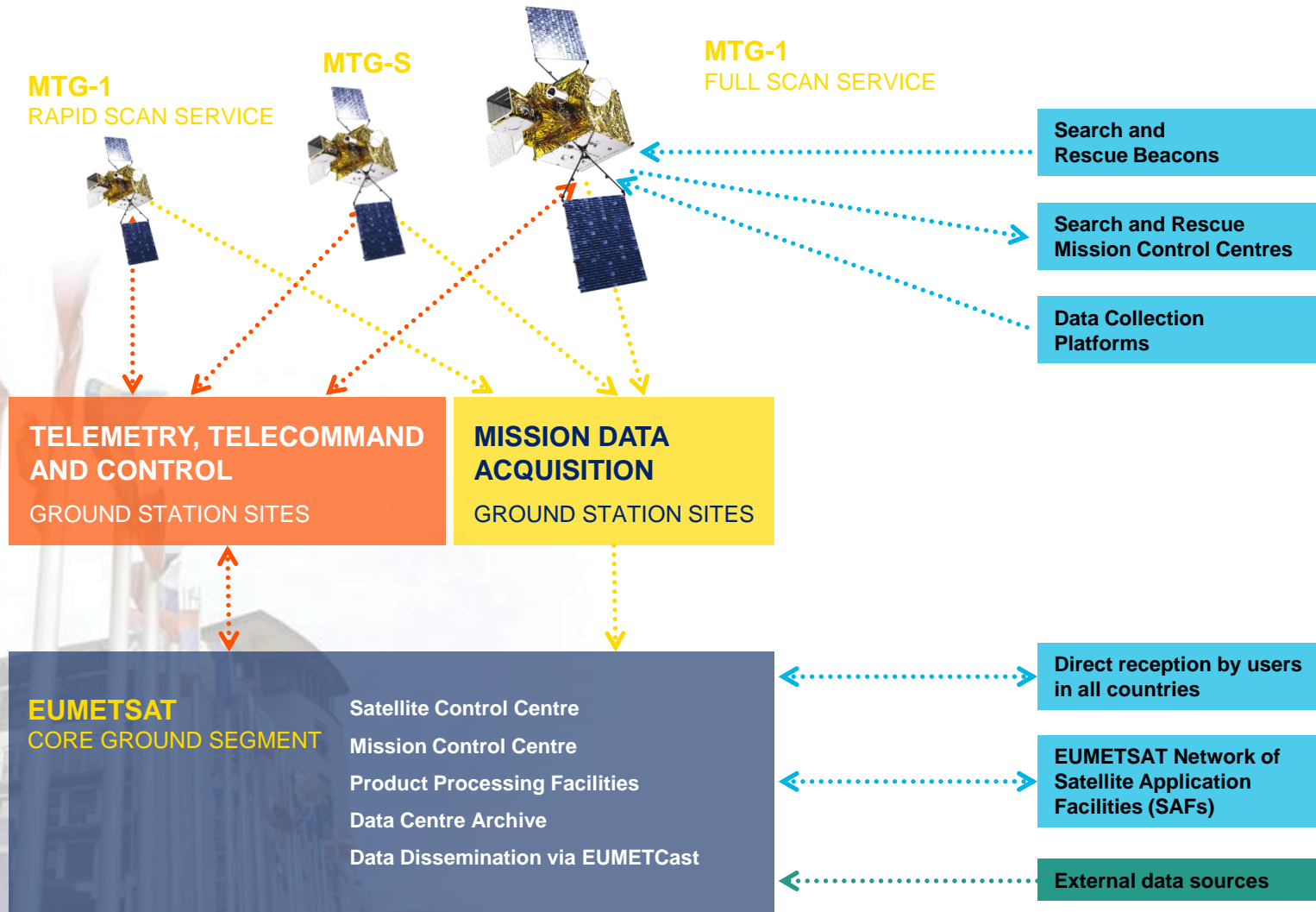
Provides the Rapid Scanning Service (RSS) delivering more frequent images every five minutes over parts of Europe, Africa and the adjacent seas.
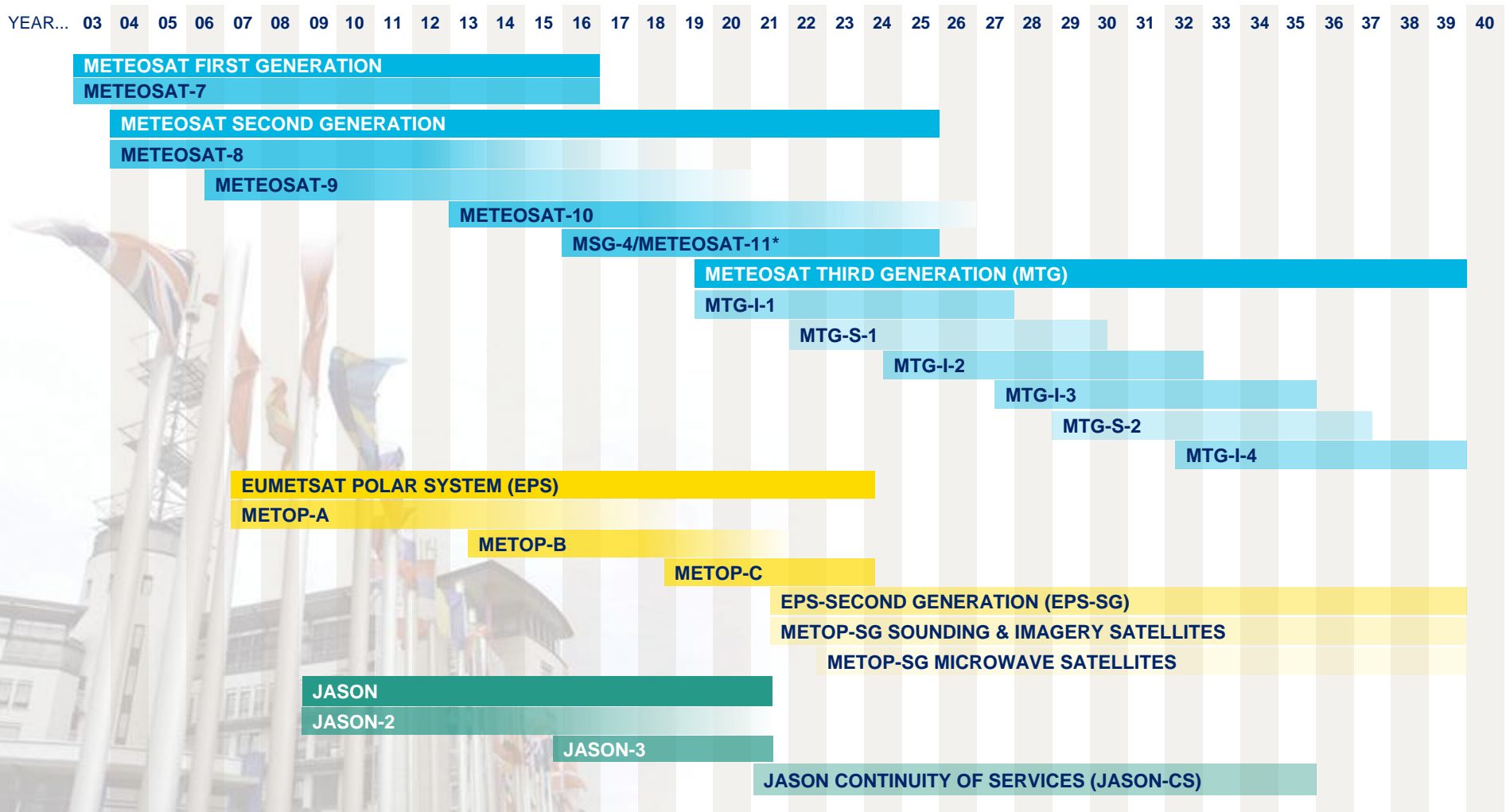
METEOSAT-8 (3.5° EAST)

BACKUP SERVICE

Serves as a back-up to both the Meteosat-9 and -10 spacecraft for full disc imagery and rapid scanning.

EUMETSAT

MTG – Overall system configuration

MTG-1
RAPID SCAN SERVICE

MTG-S

MTG-1
FULL SCAN SERVICE

Search and Rescue Beacons

Search and Rescue Mission Control Centres

Data Collection Platforms

TELEMETRY, TELECOMMAND AND CONTROL
GROUND STATION SITES

MISSION DATA ACQUISITION
GROUND STATION SITES

EUMETSAT
CORE GROUND SEGMENT

Satellite Control Centre
Mission Control Centre
Product Processing Facilities
Data Centre Archive
Data Dissemination via EUMETCast

Direct reception by users in all countries

EUMETSAT Network of Satellite Application Facilities (SAFs)

External data sources

EUMETSAT mission planning

| YEAR... | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |

METEOSAT FIRST GENERATION
METEOSAT-7

METEOSAT SECOND GENERATION
METEOSAT-8
METEOSAT-9
METEOSAT-10
MSG-4/METEOSAT-11*

METEOSAT THIRD GENERATION (MTG)
MTG-I-1
MTG-S-1
MTG-I-2
MTG-I-3
MTG-S-2
MTG-I-4

EUMETSAT POLAR SYSTEM (EPS)
METOP-A
METOP-B
METOP-C

EPS-SECOND GENERATION (EPS-SG)
METOP-SG SOUNDING & IMAGERY SATELLITES
METOP-SG MICROWAVE SATELLITES

JASON
JASON-2
JASON-3

JASON CONTINUITY OF SERVICES (JASON-CS)

Only the full operational phase of each mission is represented, excluding commissioning.
* MSG-4/Meteosat-11 will be stored in orbit, before replacing Meteosat-10

EUMETSAT

Copernicus

- Copernicus is an EC programme aiming to achieve an autonomous, multi-level operational Earth observation capacity

- PART OF THE COPERNICUS MANDATE IS TO USE DATA FROM METEOSAT, METOP AND JASON-2 OPERATIONAL SATELLITE SYSTEMS ESTABLISHED, MAINTAINED AND OPERATED BY EUMETSAT.

- OPERATION OF COPERNICUS-DEDICATED MISSIONS ON ATMOSPHERE AND OCEANS FALLING WITHIN EUMETSAT'S MANDATE I.E. THE SENTINEL-3 MARINE MISSION AND THE JASON-3 AND JASON-CS MISSIONS OF THE COPERNICUS HIGH-PRECISION OCEAN ALTIMETRY (HPOA) ACTIVITY.

- PLANNING, DEVELOPMENT AND INTEGRATION INTO FUTURE EUMETSAT SYSTEMS OF COPERNICUS MISSIONS DEDICATED TO ATMOSPHERIC CHEMISTRY (SENTINEL-4 AND MTG; SENTINEL-5 AND EPS-SG), AND THEIR EXPLOITATION IN FULL SYNERGY WITH EUMETSAT'S OWN MISSIONS

Copernicus

- MAKING OPTIMAL USE OF THE EUMETSAT MULTI-MISSION INFRASTRUCTURE IN THE COPERNICUS FRAMEWORK, INCLUDING THE EUMETCAST REAL-TIME DATA DISSEMINATION SYSTEM, EUMETSAT EARTH OBSERVATION PORTAL AND ITS ARCHIVES.

- DELIVERY TO THE COPERNICUS PROGRAMME OF DATA AND PRODUCTS AVAILABLE FROM AND AGREED WITH EUMETSAT PARTNERS IN THE UNITED STATES, CHINA, INDIA AND JAPAN.

EUMETSAT

EUMETSAT Ground Segment
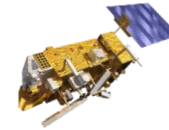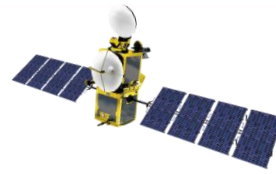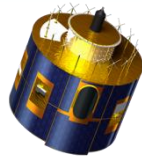
INITIAL JOINT POLAR SYSTEM
METOP   NOAA SATELLITES

CONTROL AND DATA ACQUISITION	FLIGHT OPERATIONS

PRE-PROCESSING
EUMETSAT HEADQUARTERS

DATA CENTRE
EUMETSAT HEADQUARTERS

METEOROLOGICAL PRODUCT EXTRACTION
EUMETSAT HEADQUARTERS

Central processing and generation of products

SATELLITE APPLICATION FACILITIES
WITHIN EUMETSAT MEMBER STATES

Distributed processing and generation of products

REAL TIME DISSEMINATION OF DATA AND PRODUCTS VIA EUMETCAST

APPLICATIONS GROUND SEGMENT

USERS

EUMETSAT

Reflections ...

- "Big Data"
- Hype cycle
- Use cases
- Push
- Pull
- Standards
- Interoperability
- Discovery
- Search
- Registration
- Fees
- Cost
- Resources
- Information
- Large Data

- User
- Provider
- Added value
- Question
- Result
- Infrastructure
- Evolution
- Access
- Subscription
- Archive
- Replication
- Data Object Identifier
- Processing
- Data mining

- User algorithms
- Added value service
- Public Service
- Catalogues
- Operations
- Research
- Delivery
- Retrieval
- Diversity
- Scalability
- Reliability
- Mission
- Bridge
- Value

Emerging Technology Hype Cycle
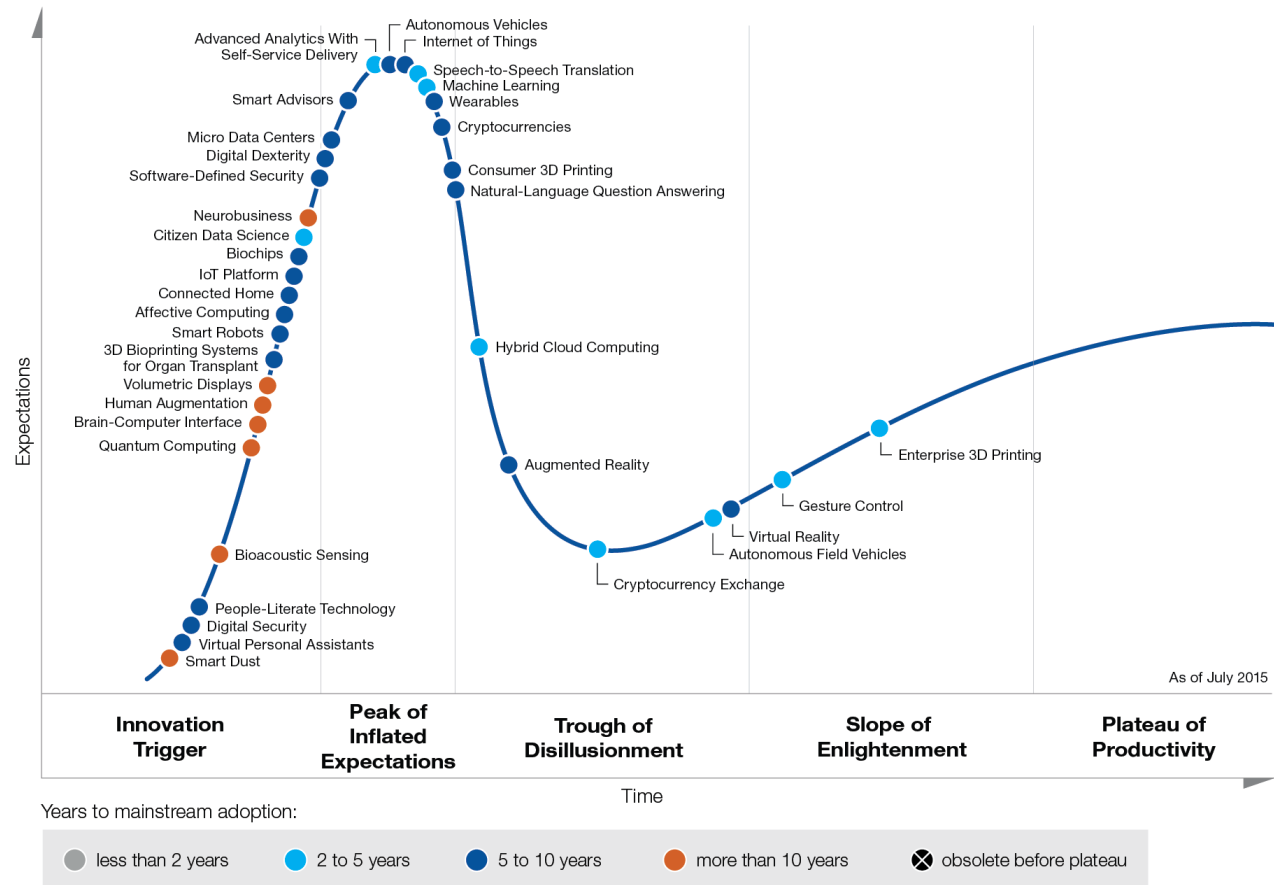
Projection of EUMETSAT satellite data rates
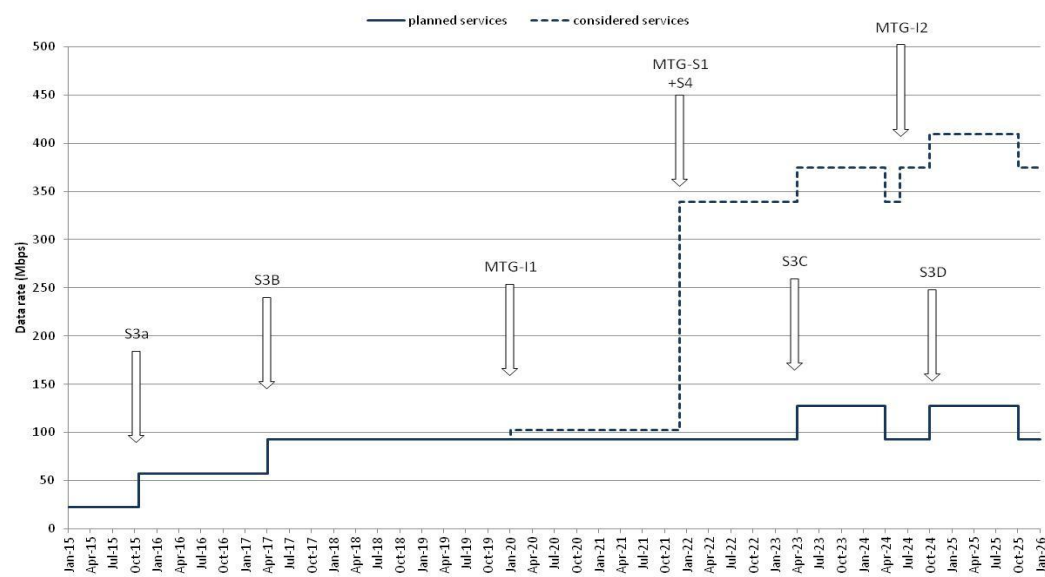

EUMETCast Europe

EUMETCast Africa

EUMETCast Terrestrial
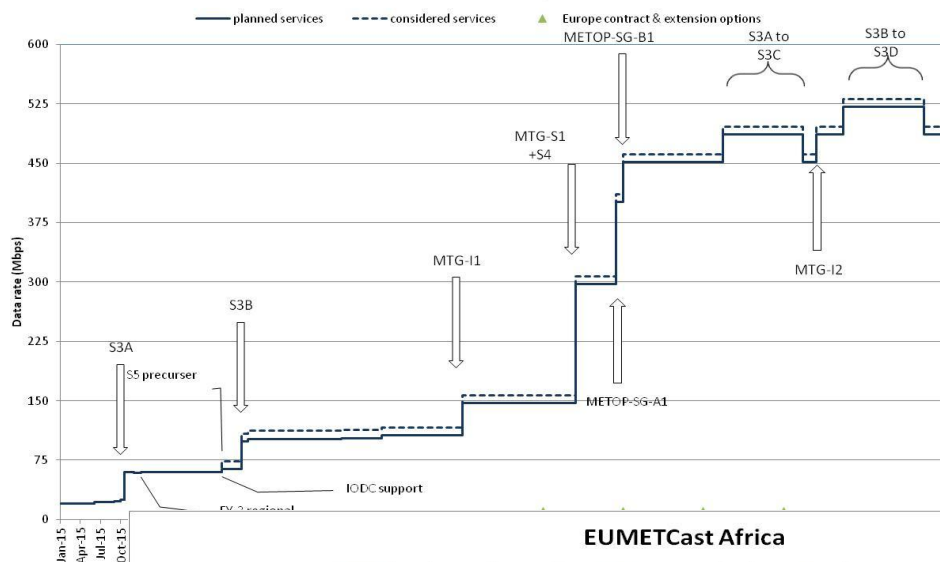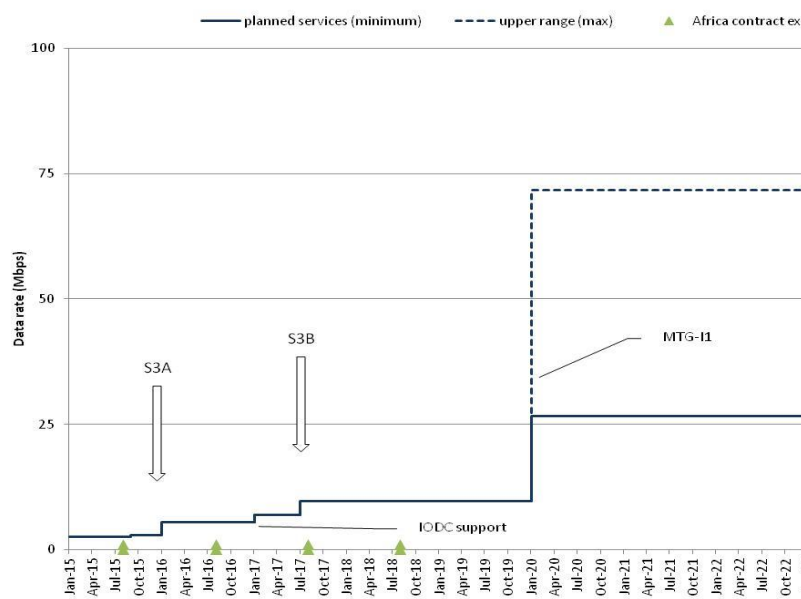
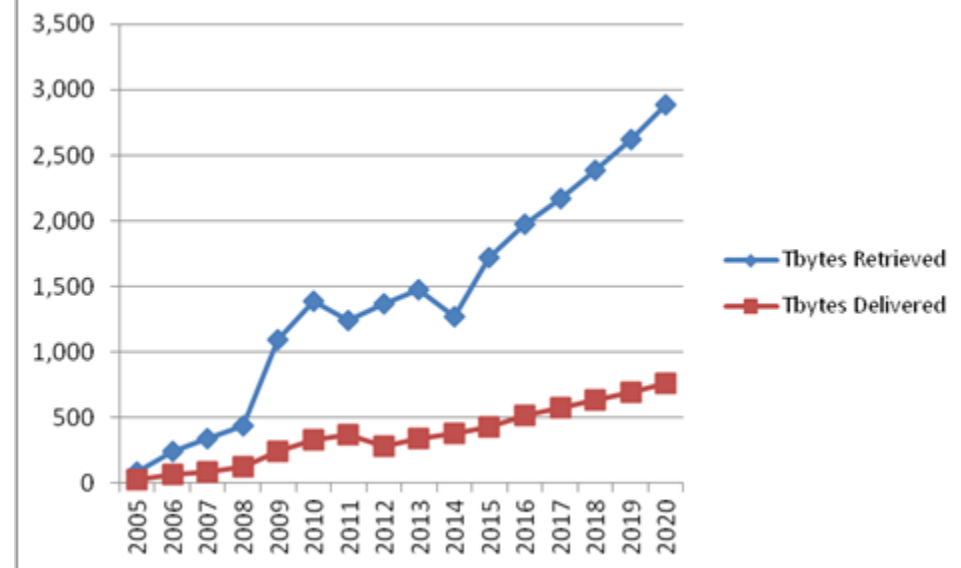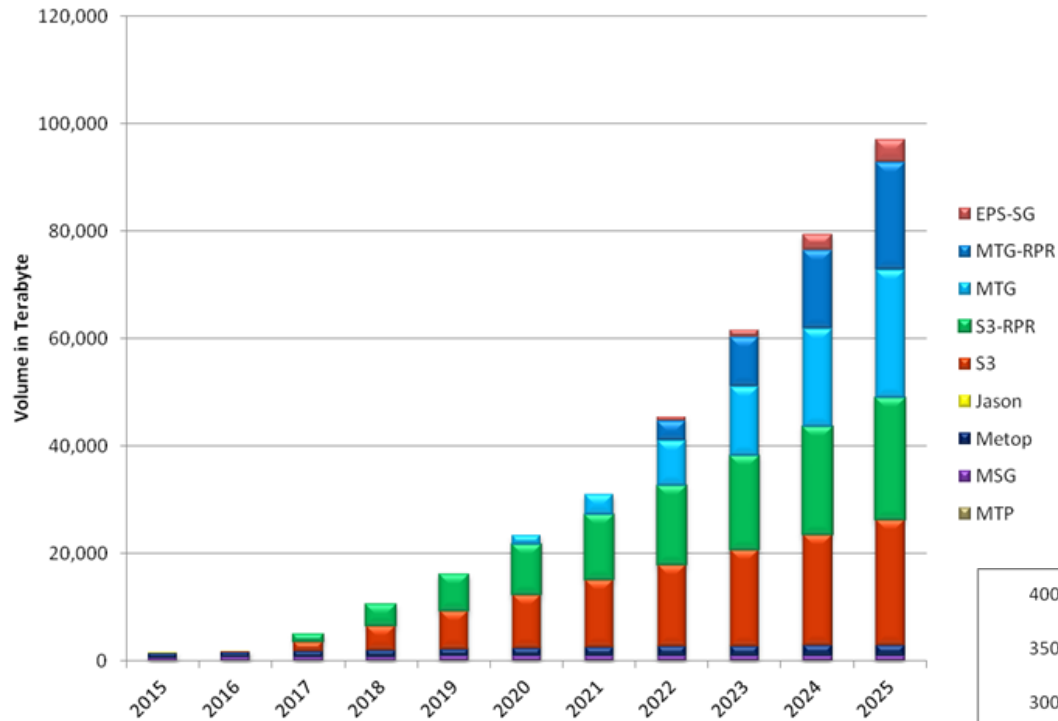**Total system output per day:**
- Currently over 4000 NRT users world wide
- Today's NRT output is approx. 60Mbit/s = approx. 646GB/day per user
➜ approx. 256 TB/day aka 90PB p.a.
- With MTG in 2020 this would be approx 150Mbit/s = 1.6TB/day/user
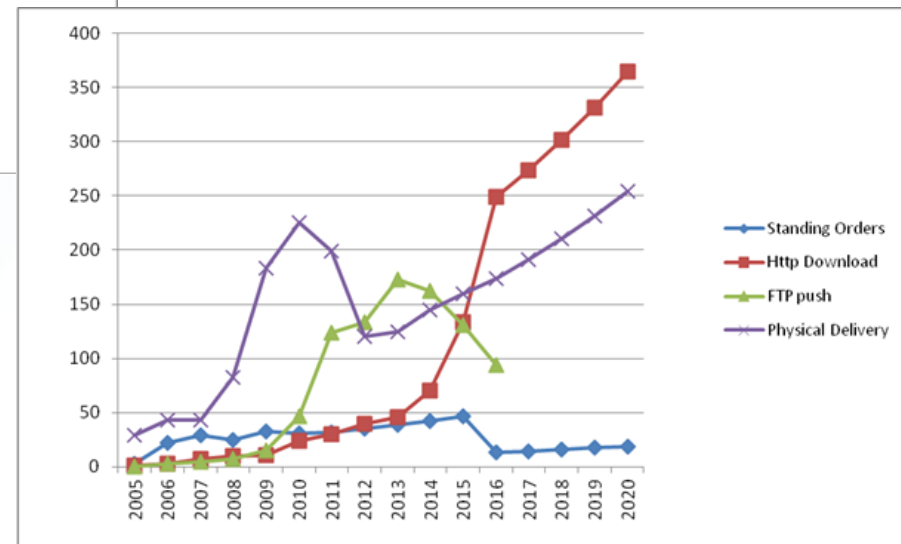➜ approx. 6.48PB/day

Projection of EUMETSAT Data Centre growth

Data Centre growth (prognosis) related to missions

-Currently approx 7000 LTA users (pa) world wide
-In 2016 2.5PB data delivery

Is The situation a "new Problem"?

? ... **Data** ...
Algorithm ...
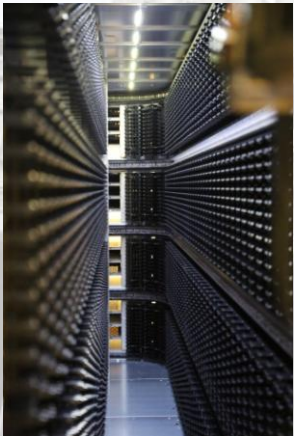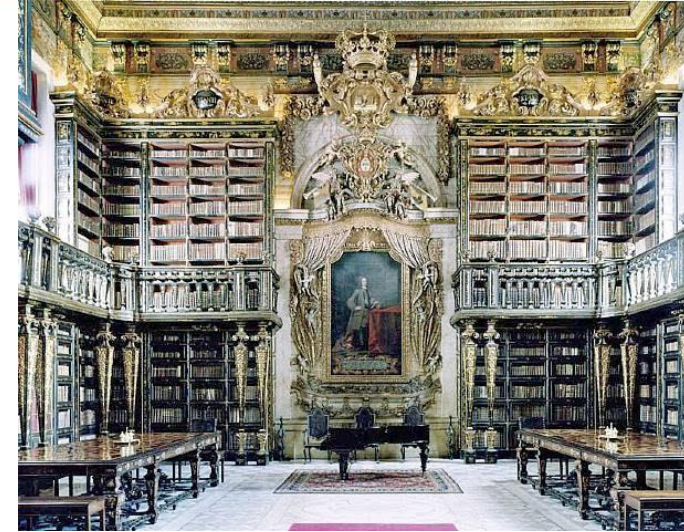**Information** ...
Answer ...
Value

- **Distributed data sources**
- **Heterogeneous data access functions & services**
- **Data delivery as well as data retrievals**
- **Large data volumes**
- **Information generation at the user end**

EUMETSAT
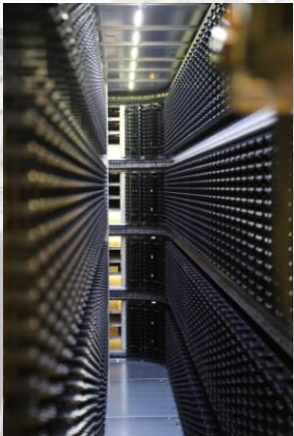
When does Data become "Big Data"?

? ... Data ...
Algorithm ...
Analytics ...
Correlation ...
Information ...
Answer ...
Value

DAR – Discovery Access Retrieval

Is it the main "job" of a user to search for data?
➢ Search and Discovery are just the first step …

➢ One-stop shop user experience is key
   Exploitation of the individual strengths of each data provider by
   systematic use of interoperability standards for Meta data,
   Search & Discovery

Interoperable Service Layer (based on standards)	DAR Portals
Individual Strengths & Contributions	Data Provider 1 · Data Provider N
Internal Services & Functions	Processing … · Data delivery … · Archive … · Processing … · Archive … · Data delivery …

Bringing Data to the User – Delivery Models ...
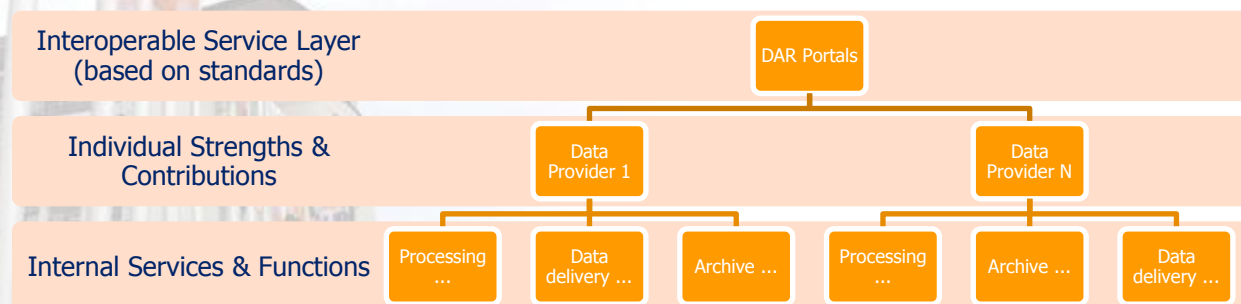
Pull model
- Off- & On-line, interactive, time series
- Bespoke and/or user defined data areas of interest
- Typically without guaranteed service levels
- User infrastructure dependent

Push model
- Data driven
- To the doorstep
- Continuous delivery of new data
- Typically guaranteed service levels

Bringing Data to the User ... The element of timeliness

Near Real Time world
➢Push model, Fast, current, to the doorstep
➢Operational users
➢Continuous delivery of new data & products & data sets
➢Typically guaranteed service levels

The "grey zone" in between
➢Pull model
➢Faster and easier than offline
➢interactive
➢Added value & bespoke hosted functions/processing (i.e. sub-sectioning, re-formatting etc.)
➢Service interface

Offline world
➢Pull model, Orders, time series
➢Specific formats
➢Non-time critical and typically without SLA
➢ Media delivery

EUMETSAT

Bringing Data to the User – general aspects

Some users want an <u>answer to a question</u> rather than the data
- Information & Analytics instead of pumping base data
- Benefit: Users won't require infrastructure to handle large amounts of data

Many projects => many data islands
- New projects create more islands

Information Location is key for answering questions
- Responsiveness to evolving user needs
- Post-processing services cannot afford to retrieve data from islands
- Copying of data has impact on network, storage and compute resources
- Very similar needs for internal and external users

=> Bring the users to the data <u>in addition</u> to traditional data-to-users
    (which is what information centric organisations have done)

Changes in the environment ...
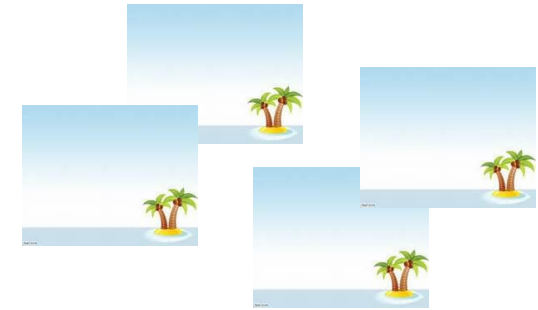
New and emerging needs and opportunities:

→Change in user needs & expectations – more frequent, localised, personalised, integrated, cross-sectoral information and services
→Next generation satellite missions and sensors
→Interoperable infrastructure and services
→New available technology

Bringing Users to the data ... VALUE generation

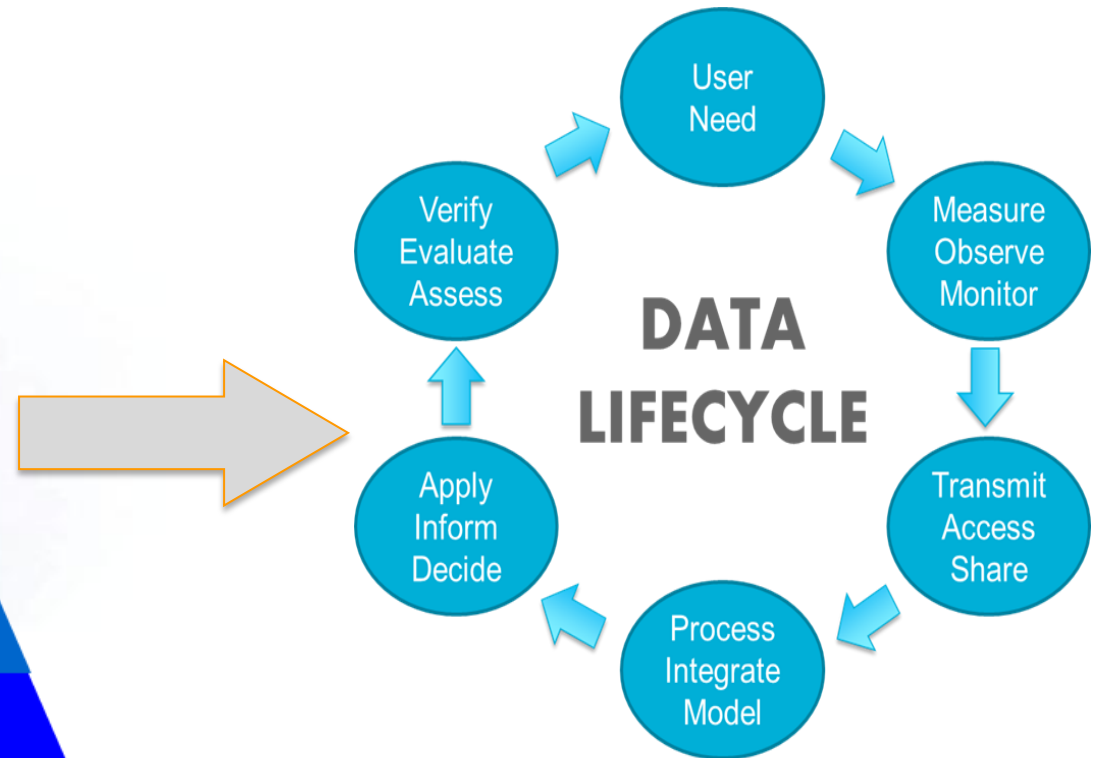Question → Data → Algorithms → Analytics & Correlation → Information → Answer → VALUE

➢Data providers awareness of the information potential and science that is perceived to be obtained from the base data
➢Provision of added services (algorithm processing) on the "base" data to allow easier information extraction
➢Interconnect Data access across different Providers via Interoperability as the base for data analytics & correlation of diverse data sets
➢Manage and interact with the user communities
➢Develop services in the "grey zone"

EUMETSAT

Bringing Users to the data ... VALUE

- Value is the most important 'V'
  - → Unused or unusable data delivers little value but still incurs costs

- Generation of value due to the existence of big data, new science, new tools, data mining, data analytics and correlations, allowing to generate previously unknown questions.

- Value lies in fitness-for-purpose
  - → To meet new user needs – products, scale, timeliness, detail, content
  - → To ensure our information and services are useful, useable and used
  - → To support new policy and research goals
 - → To verify and demonstrate that user needs are being met

(Example: Tee Shop)

the Value Pyramid & Data Lifecycle

Big challenge and big opportunity of data
→ to extract value through insight from data

Value

Wisdom
Knowledge
Information
Data

User Need

Measure Observe Monitor

Transmit Access Share

Process Integrate Model

Apply Inform Decide

Verify Evaluate Assess

DATA LIFECYCLE

EUMETSAT
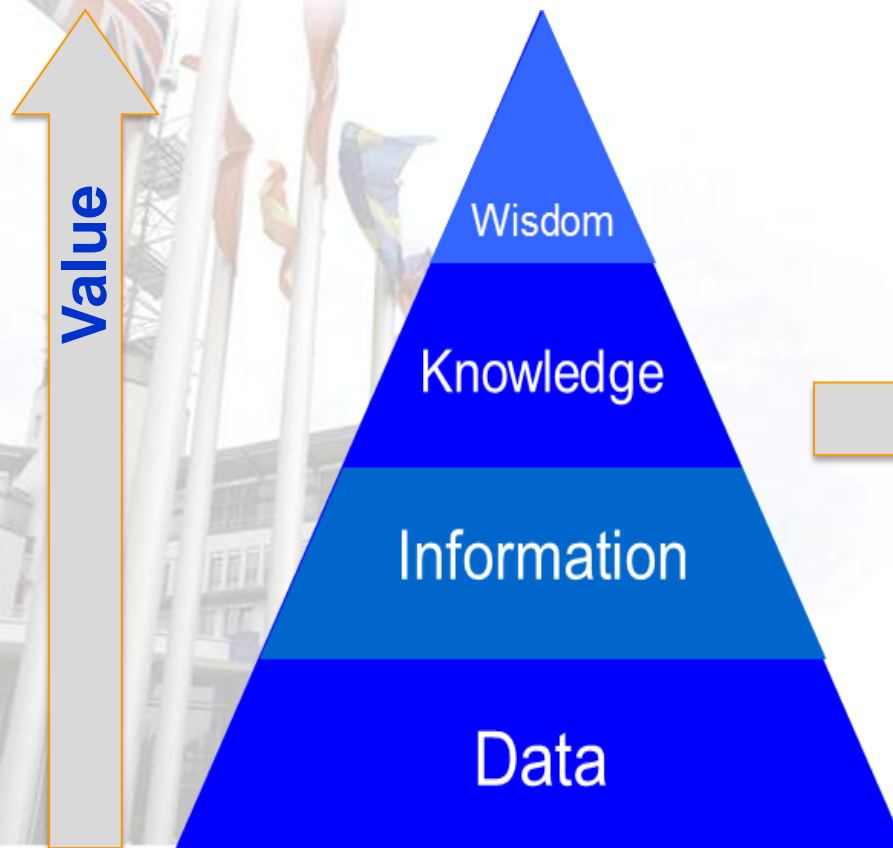
From linear to Cyclic Service Evolution ...

- Maintaining pace with user expectations and capabilities
  → Demand for more timely, more detailed, more interactive information and products
- Changing paradigms
  → Collaboration and partnership
  → Evolving from a linear production model to a cyclic interacting model
  → "*" as a Service aka general service orientation and interoperable services including constant user feedback

Opportunities & Building the "bridge" ...
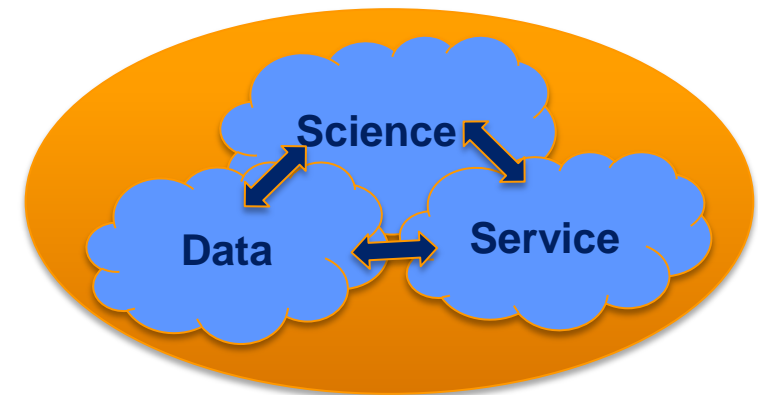
Data Provider builds a "Bridge" for the User by:

→ Understanding what data means to your organisation
→ Capturing opportunities
→ Managing risks
→ Making well-informed choices
→ Using WMO, EC, CBS, ... resources
→ Learning from experiences of others
→ Looking around and ahead – at trends in technology and those who use it well

EUMETSAT

# Opportunities & Building the "bridge" ...

- Highly Interactive User communities engagement process
- Enabling easy and homogeneous access to the data
- Preparing the data
- "Processing @ the data concepts
- Diversity of delivery mechanisms
- Harmonised access to push & pull models
- Efficient use of supporting technologies
- Understanding and translation of a variety of data policies

EUMETSAT

EUMETSAT Data Services
Recognising the scope → new Data Services Strategy

➤One fits all is not the right approach - push & pull is needed as well as "bringing data to the users" & "bringing users to the data"

➤Fully service oriented approach based on cyclic user interaction

➤Big Data and associated technologies raise also questions on: Privacy, IPRs, User credentials, IT Security
  ➤"Safe harbor" concepts are booming
  ➤Data as a Service

➤Fast developing technology supporting the big data concepts

➤"Free to the user" however represents for data providers potentially substantial cost
  ➤New business models are possible (public services, downstream services)

➤Users to the data services to avoid data islands

EUMETSAT

Definition of a new Data Services Strategy

- Implementation via a set of interconnected Pathfinders
- Strong involvement of industrial support
- 2017 Pathfinder implementation → 2018 User Validation Cycle → 2019 Operational level implementation

Information Centric Service Infrastructure

DaaS – Data as a Service, SaaS – Software as a Service, Paas – Platform as a Service, IaaS – Infrastructure as a Service
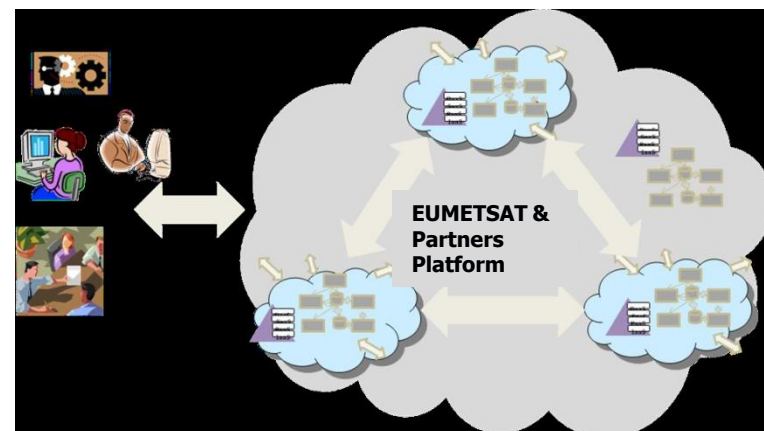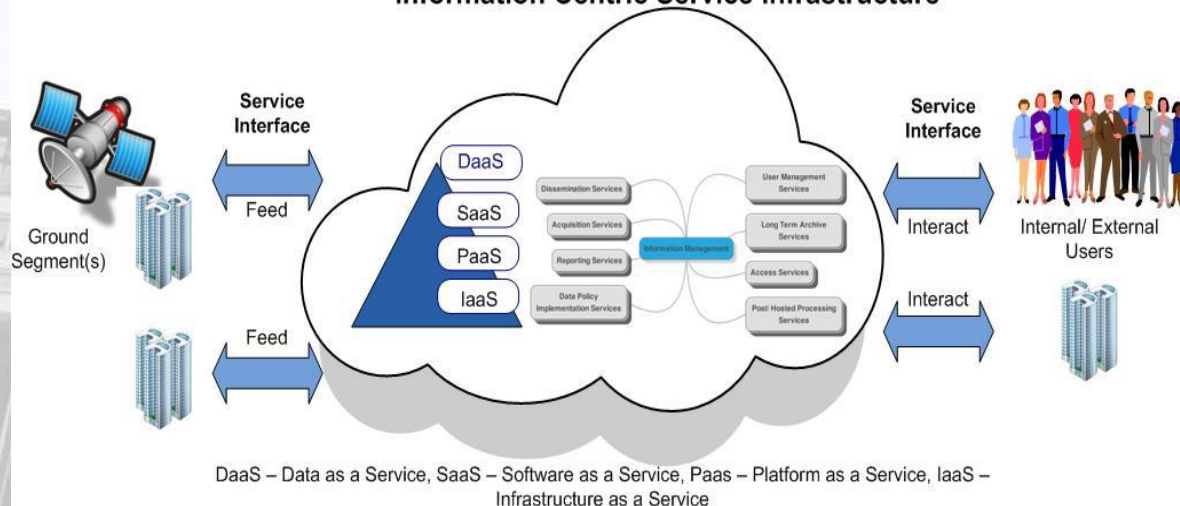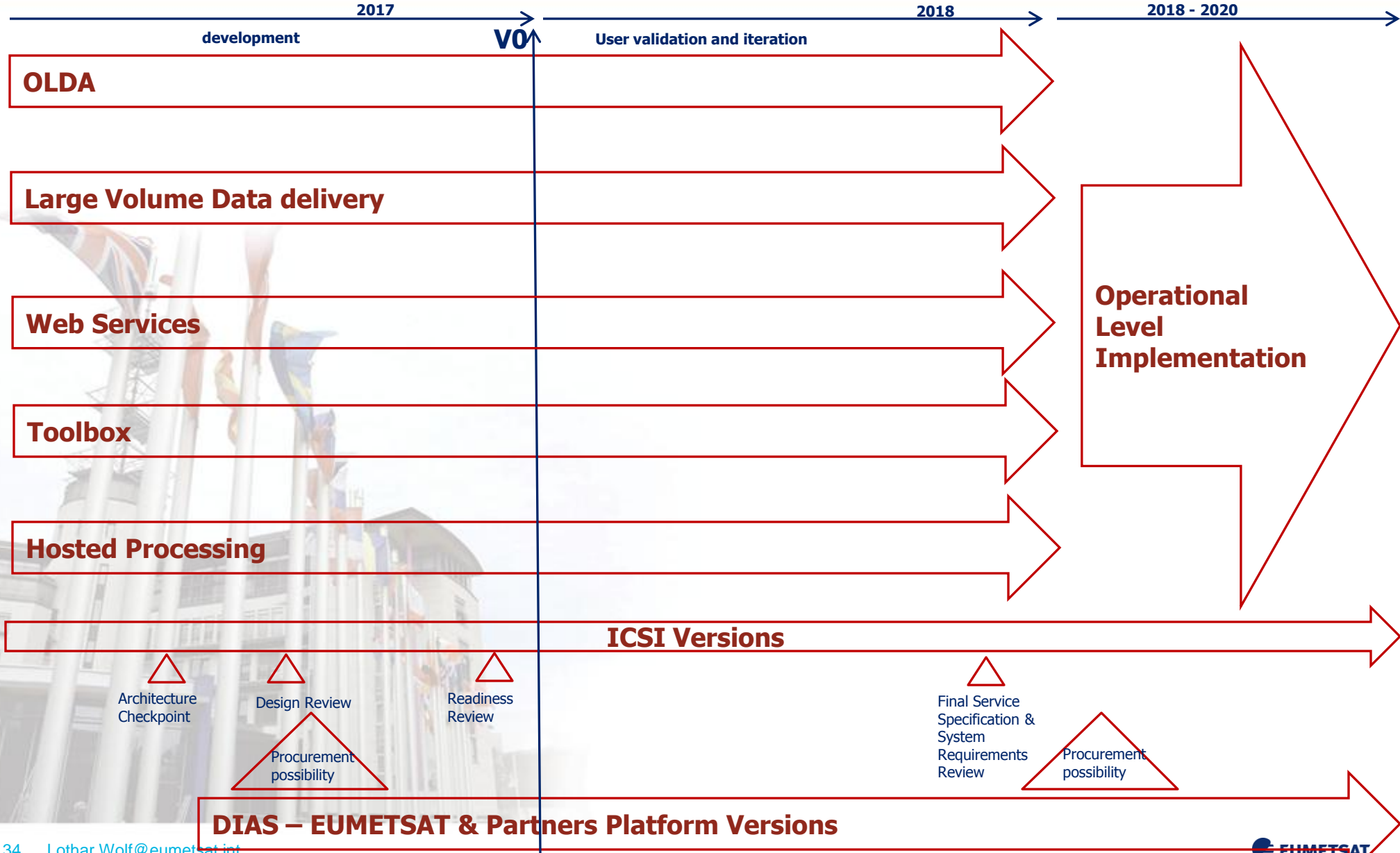
EUMETSAT Data Services Strategy – Implementation via Pathfinder projects

1. Online Data Access Services
2. Large Volume Data Dissemination via multicast
3. Web Services (including WMS)
4. Data Format Toolbox
5. Hosted Processing
6. DIAS - EUMETSAT & Partners Platform in the context of Copernicus

→ Pathfinder projects allow highly iterative exploration in new methods, concepts and technologies in order to develop the final future Data Service portfolio specification and the associated system level architecture and requirements

→ The overall activity of all contributing projects is managed via a Framework management structure

EUMETSAT

EUMETSAT Pathfinders

➢Formalized and cyclic user interaction processes
➢Interoperable Service interfaces
➢Re-use and integration of existing capabilities including cooperation with specialised centres

➢Interoperable Data Discovery, Access & Retrieval
➢Diverse Delivery mechanisms addressing diverse use cases (NRT, Online, Offline) with equal service delivery to users
➢Data as a Service (Data to the user & User to the data)

➢Hosted processing & platform services
➢Online Data Access Services
➢Interactive Web Services (WMS etc.)

➢Evolution of the related data policies
➢European partner cooperation
➢Support of value generation

EUMETSAT Implementation Logic & Timeline

2017   2018   2018 - 2020

development   **V0**   User validation and iteration

OLDA

Large Volume Data delivery

Web Services

Operational Level Implementation

Toolbox

Hosted Processing

ICSI Versions

Architecture Checkpoint

Design Review

Readiness Review

Procurement possibility

Final Service Specification & System Requirements Review

Procurement possibility

DIAS – EUMETSAT & Partners Platform Versions

Questions?