WEATHER CLIMATE WATER
TEMPS CLIMAT EAU

Managing large volumes of information

Julie Pérès (Météo-France)

Item 5.1

WMO OMM

World Meteorological Organization
Organisation météorologique mondiale

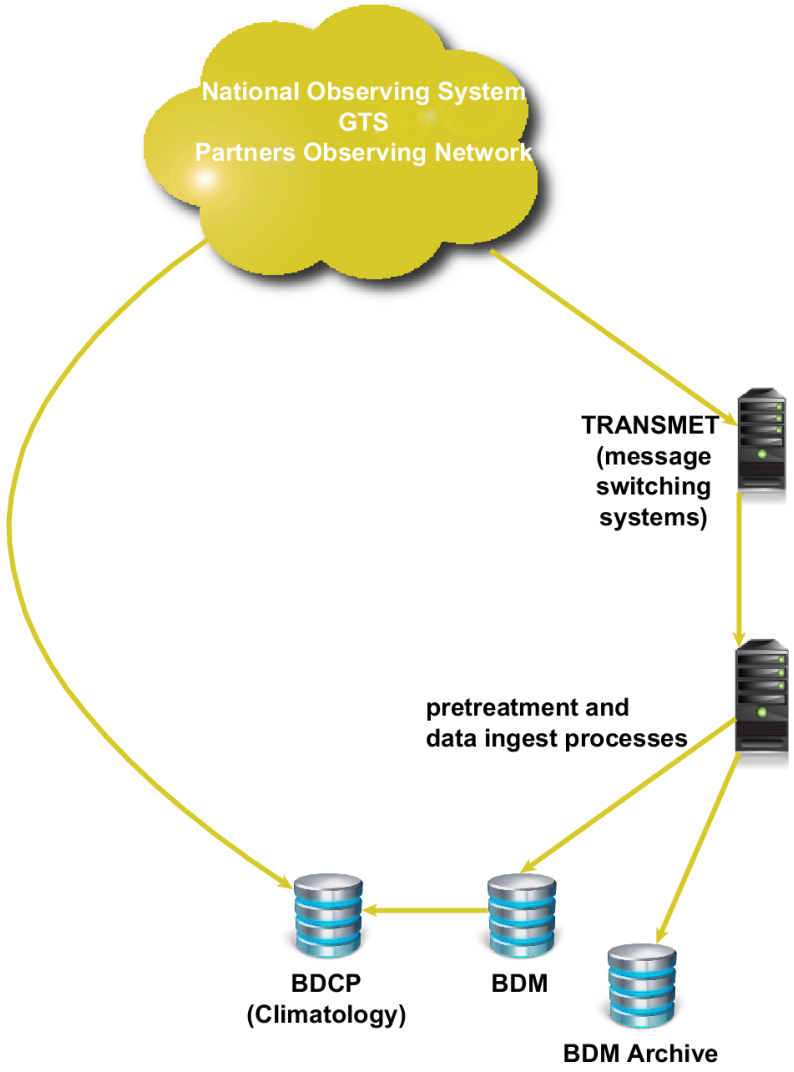
Plan

- Data acquisition and pre-processing
- Databases
- Archives and Storage
- Data delivery

Operational databases at Météo-France

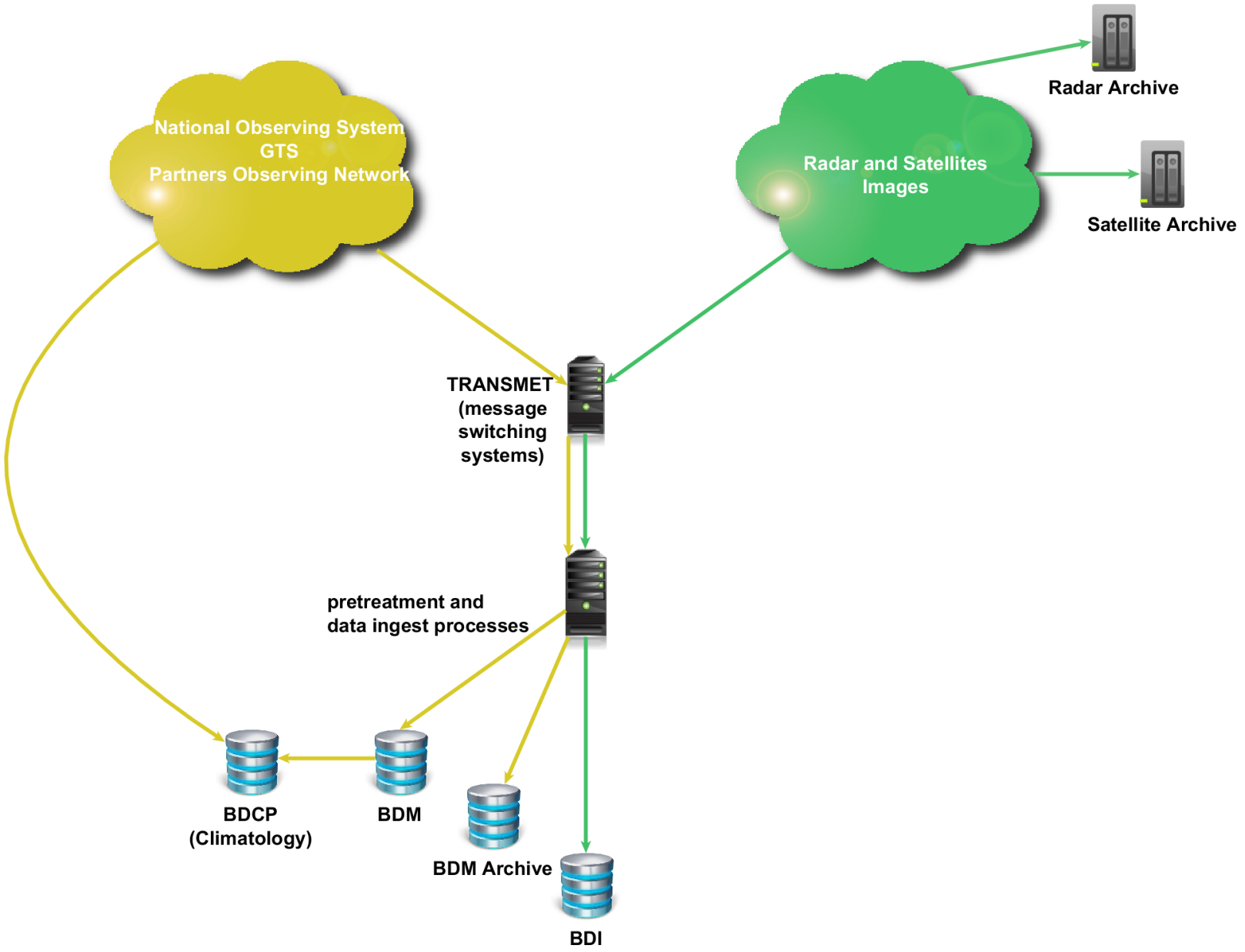
- Concentration of real-time data relevant to the central production
- Data processing
- Delivery (Forecast / Climatology / Client)
- What ?
  - MF, overseas and partners observation network
  - Radar, satellite data and images
  - Forecasting model

Data acquisition and pre-processing

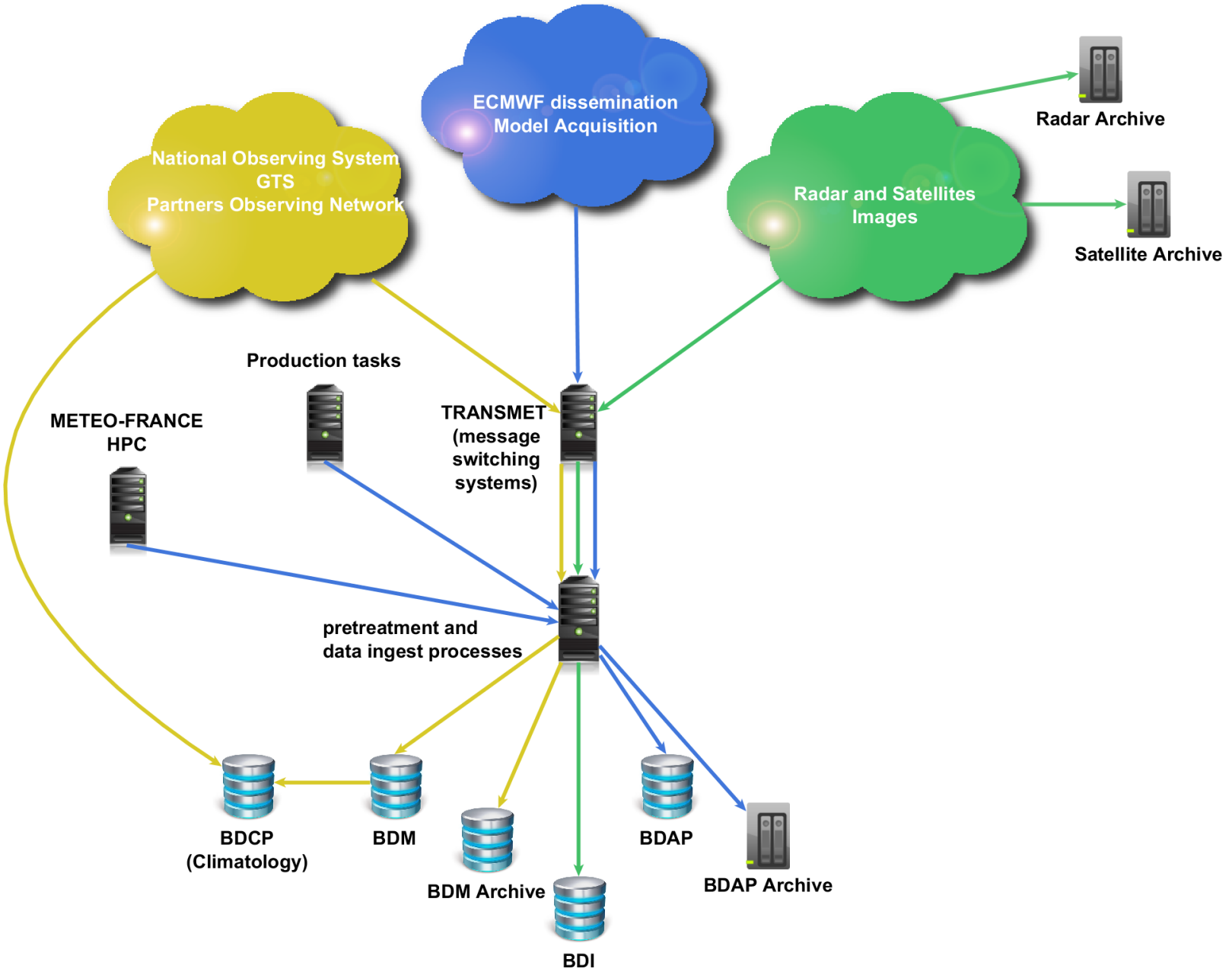


WMO OMM

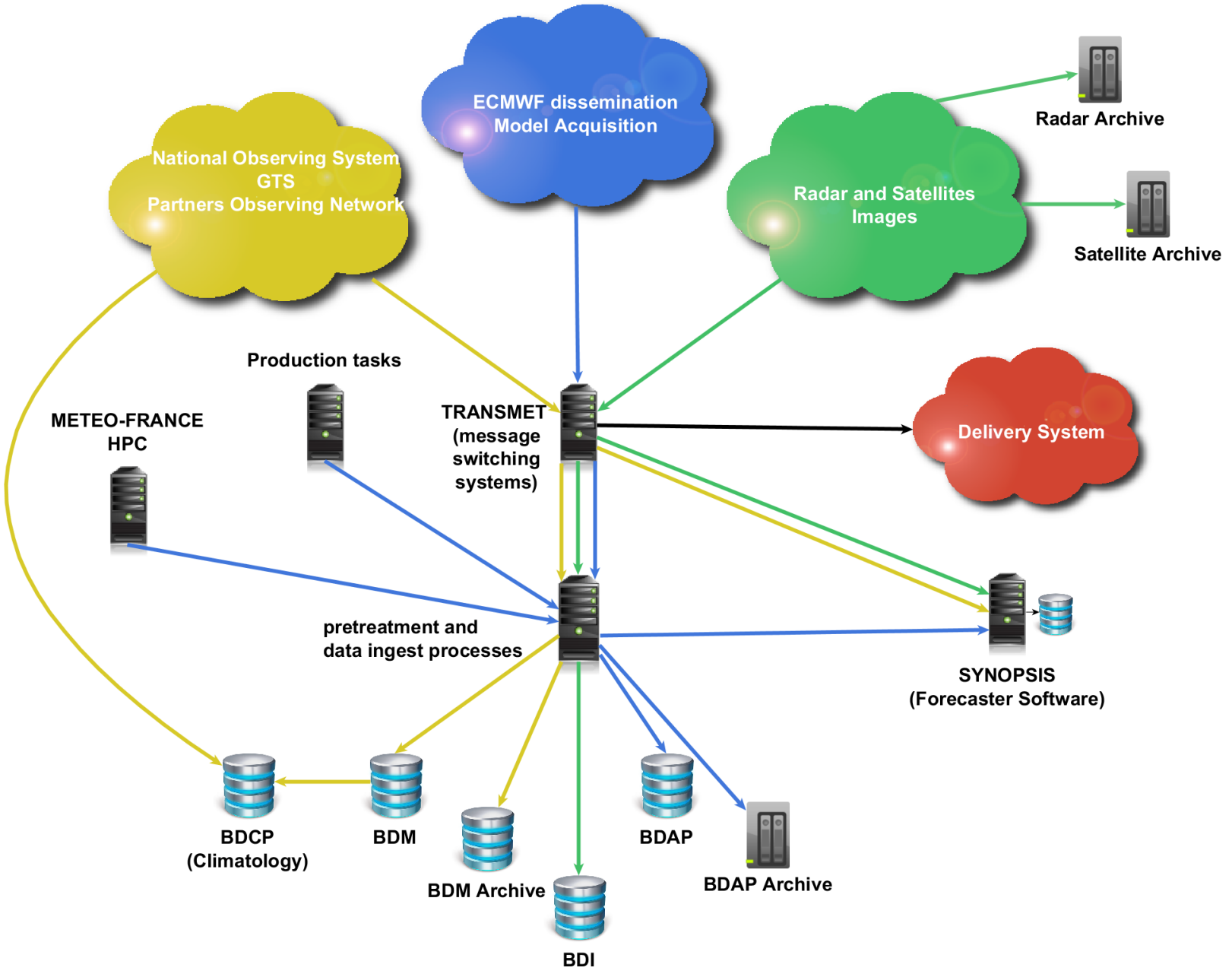
Data acquisition and pre-processing



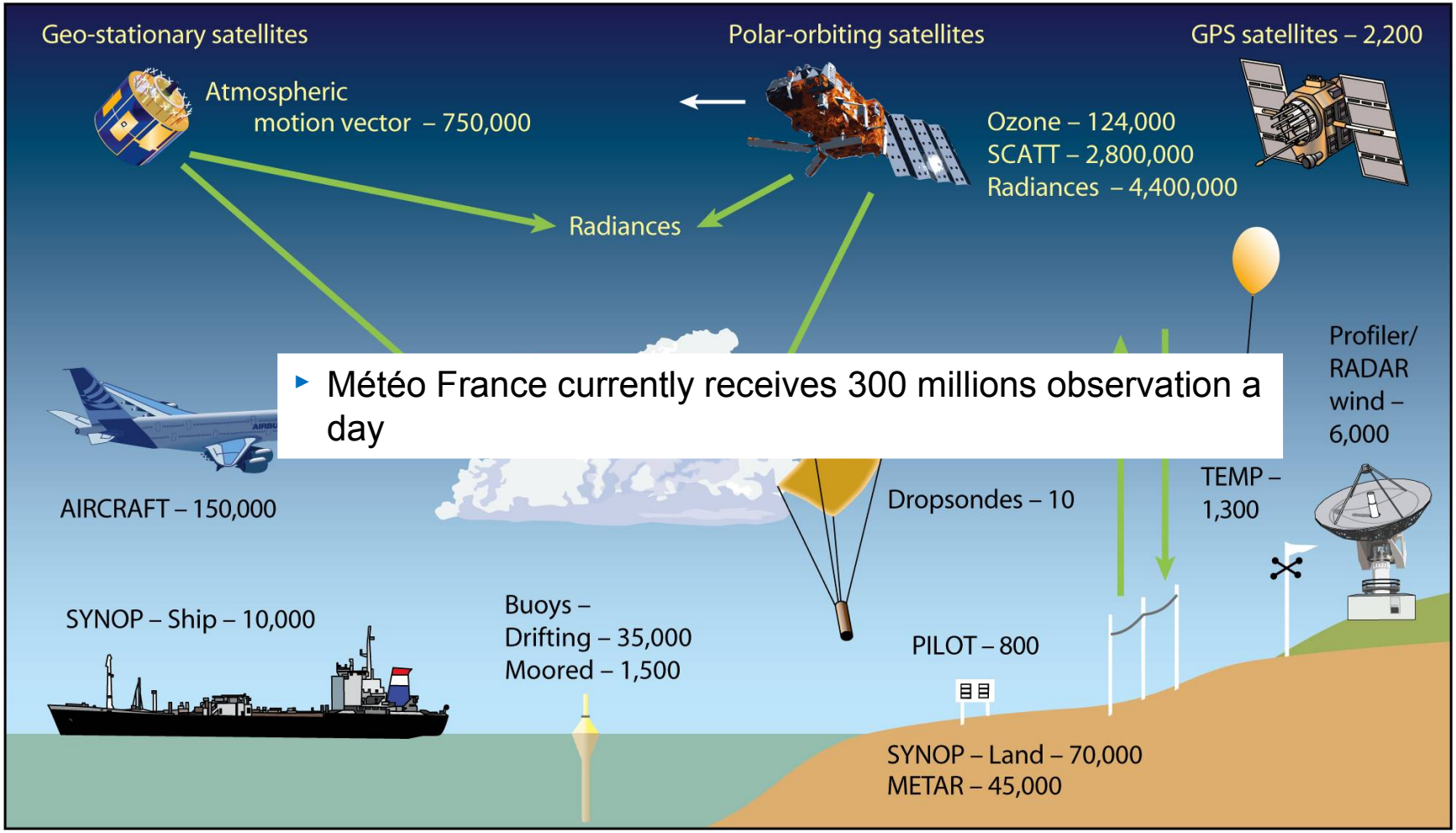
Data acquisition and pre-processing



Data acquisition and pre-processing



Observation data



Observation data

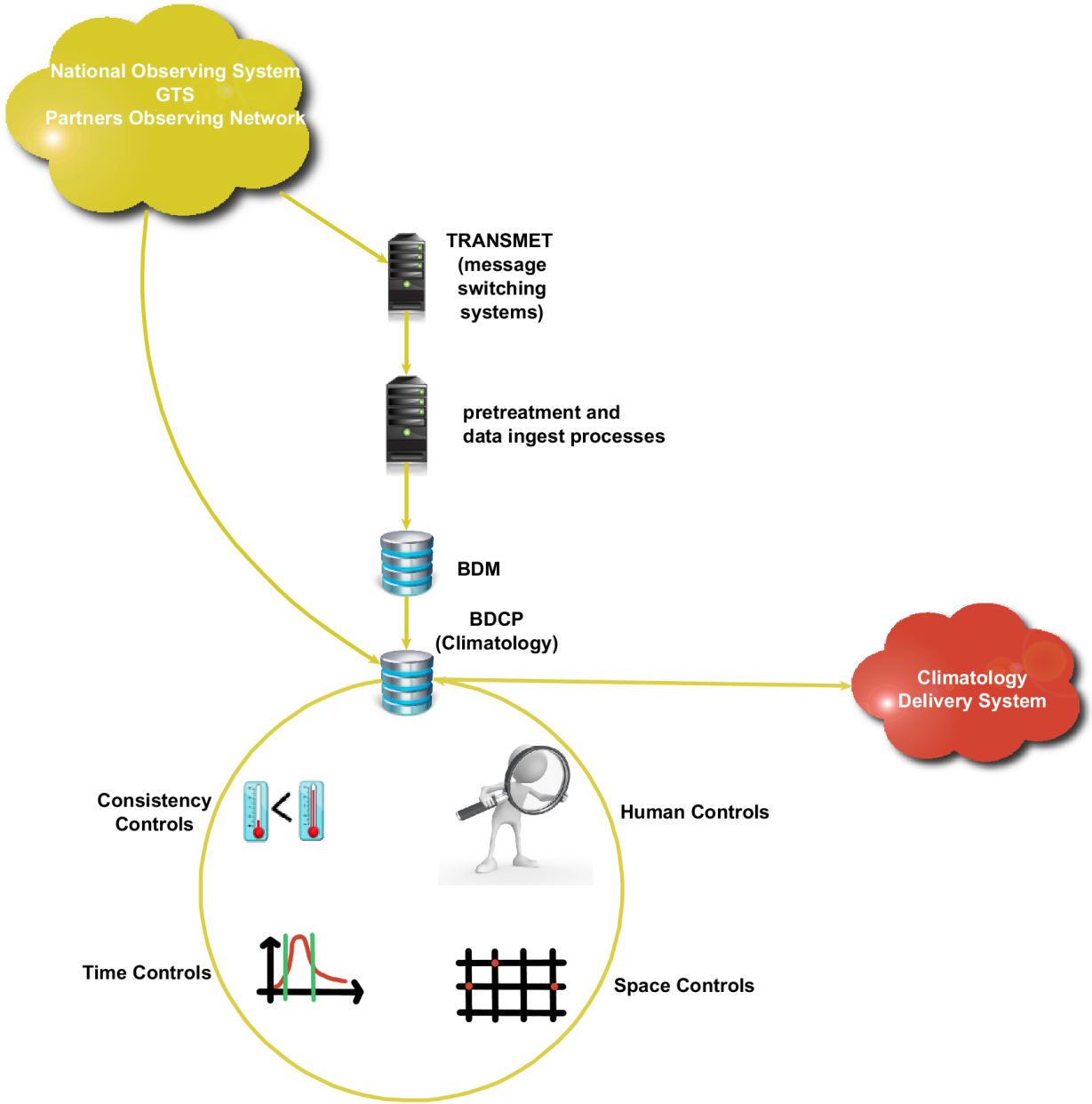
- More than 60 data types :
  - Aircraft data (ACAR, AMDAR), Buoy data, Ship data, Cyclone data, Satellite data, Radar data, Ground data (RADOME, SYNOP, …), Altitude data (TEMP, PILOT, …)
- Messages format :
  - BUFR, ASCII, JSON, NETCDF
- Pretreatment
  - Transcoding messages
  - Controls
  - Adding meta-data or calculated parameters
  - Formating for database
- Database : BDM (PostgreSQL)
- Volumetry : 200 GB
- Retention : 5d

BDM Archive

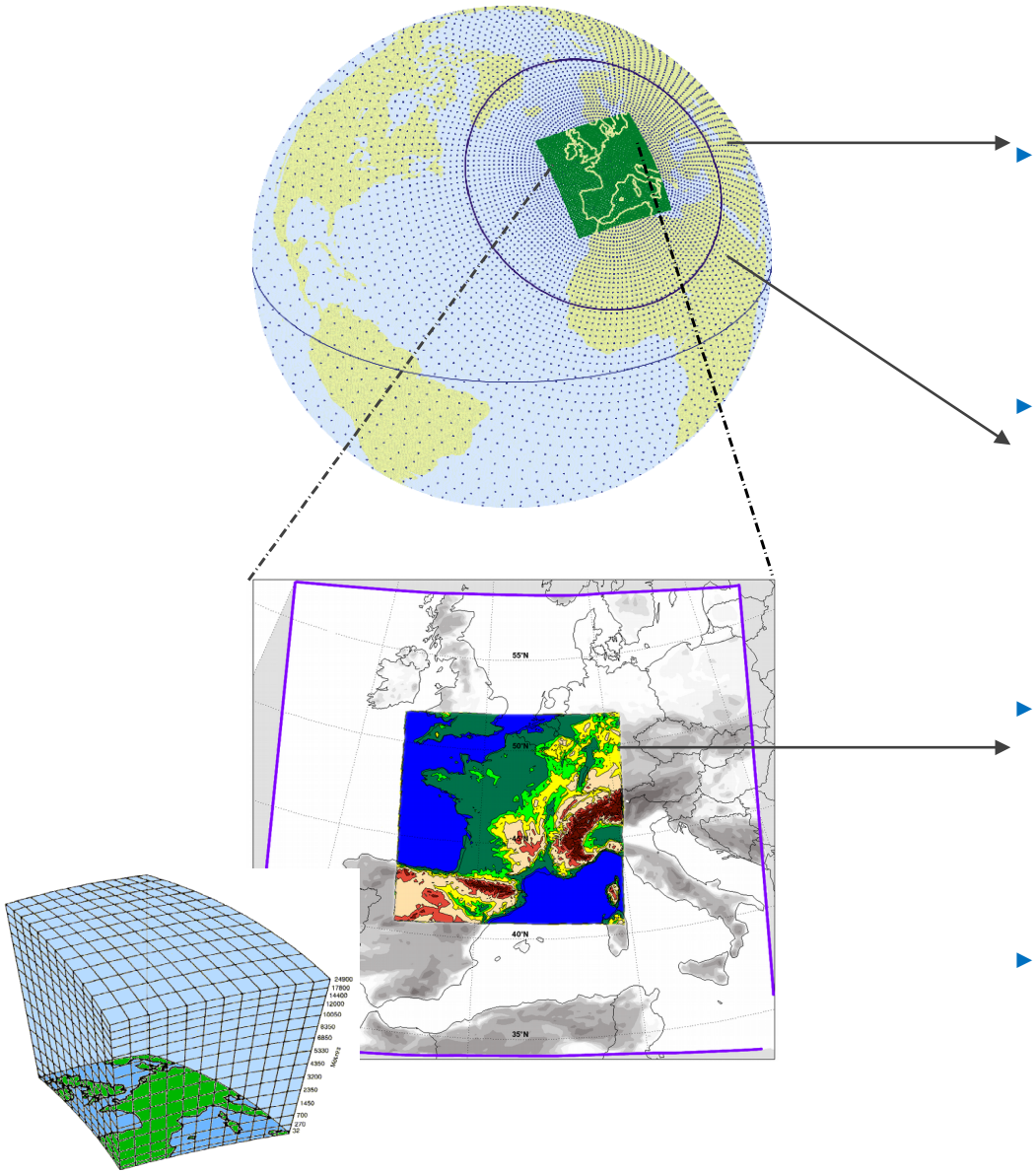
- BDM-Archive : another database (also PostgreSQL)
- Not accessible by operational tasks
- Retention is variable (from 3 months to eternity)
- Volumetry : 19 TB

Climatology

- Data :
  - minutes, hours, daily, monthly, normal, hydrological, stations metadata
  -
- Volumetry : 700 GB



Operational forecasting system



- Resolution 7,5 to 36 km
  Global model ARPEGE,
  Forecast up to 5 days

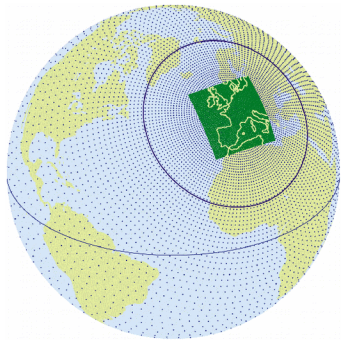
- Resolution 7,5 km
  Regional model ALADIN Europe,
  Forecast up to 5 days

- Resolution 1,3 km
  Mainland France model AROME,
  Forecast up to 36 hours

- Doubling resolution = x16 compute
  at constant scientific complexity

Operational forecasting system

- **Additional LAM outside Europe (overseas French territories)**
- **Ensemble forecast (EPS):**
  - Carrying out multiple runs of the model by introducing each time a slight modification of its initial state
  - Global PEARP twice daily 35 members, 16 km over France, to 4 days ahead PE-AROME :
  - France PE-AROME, twice daily, 12 membres, 2.5 km



- Ocean waves: 4 time per day
  - Global: 4 days ahead at 50 km with atmospheric forcing from IFS
  - Global: 4 days ahead at 50 km with atmospheric forcing from ARPEGE
  - European seas: 3 days ahead at 10 km
- Seasonal forecast: once a month
  - 41-members, 312 km, to 7 months ahead
- Air pollution transport forecast: MOCAGE
  - 10 km over France to 4 days ahead

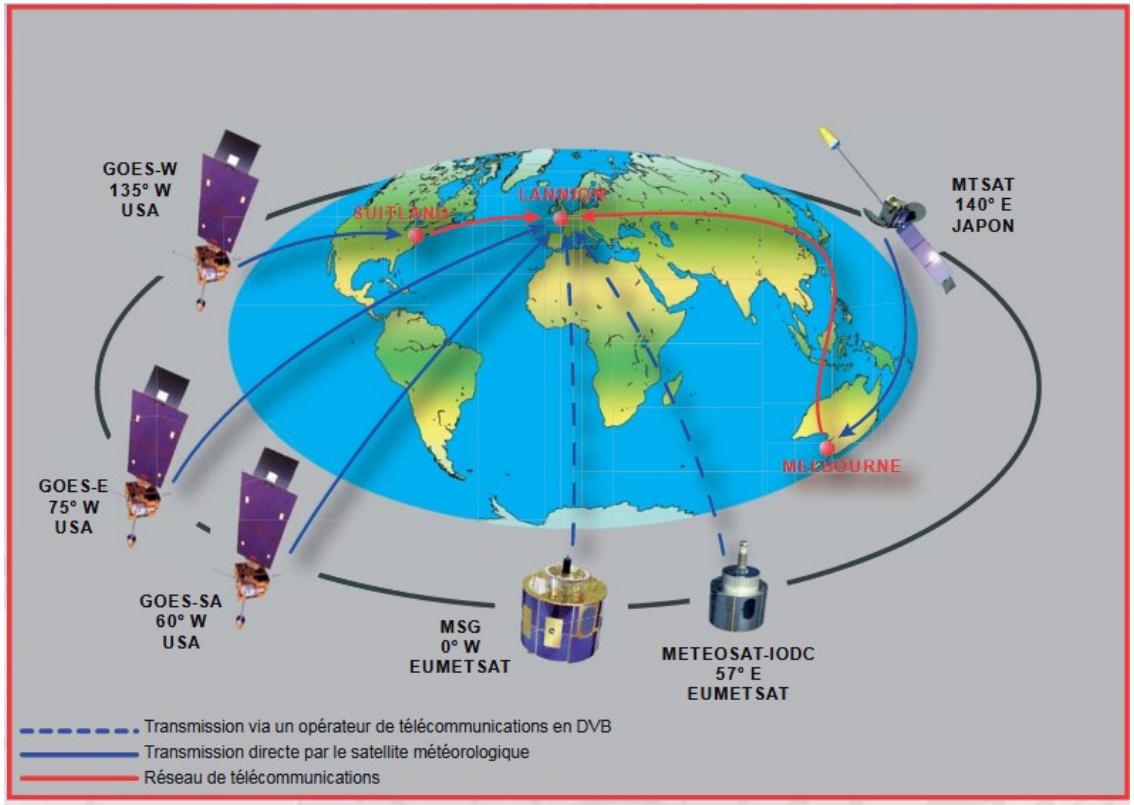
Models data

- Messages format :
  - GRIB edition 1 and GRIB edition 2 in database
- Pretreatment
  - Change of descriptor
  - Calculated parameters
  - Transformation of grids
  - Transcoding from GRIB1 to GRIB2
  - Compression
  - Dissemination to our forecasting client (Synopsis)
- Database : BDAP (PostgreSQL)
- Volumetry : 5 TB (900 GB/jour)
- Retention : variable

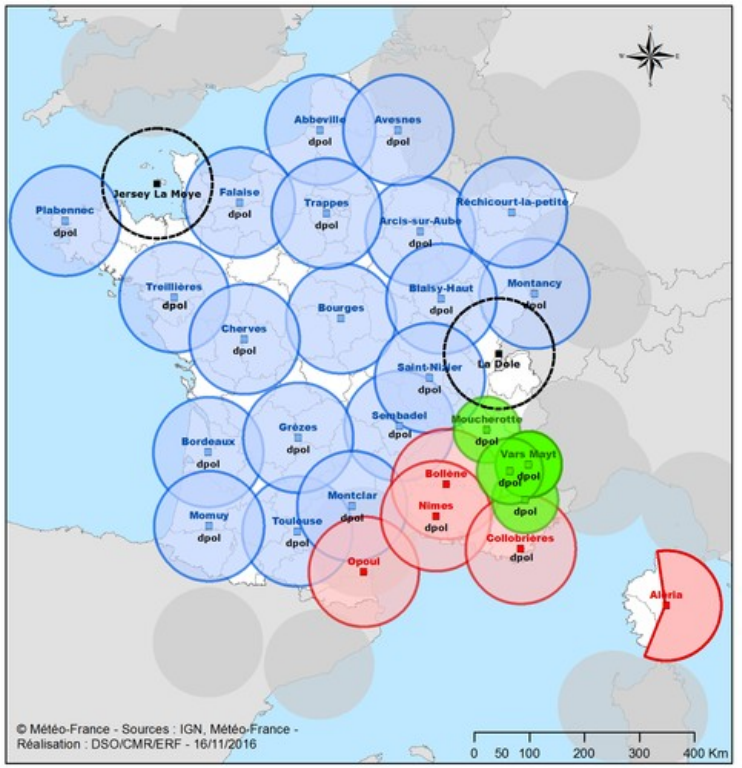
BDAP-Archive

- Not in database → Grib Files stored in our storage system, under particular trees
- Not accessible by operational tasks
- About 500 TB
- Not optimized for data retrieve

Images data



RADAR network in 2016



Légende

- C band - radar limitrophe
- X band
- S band
- C band

Dpol : dual polarization

WMO OMM

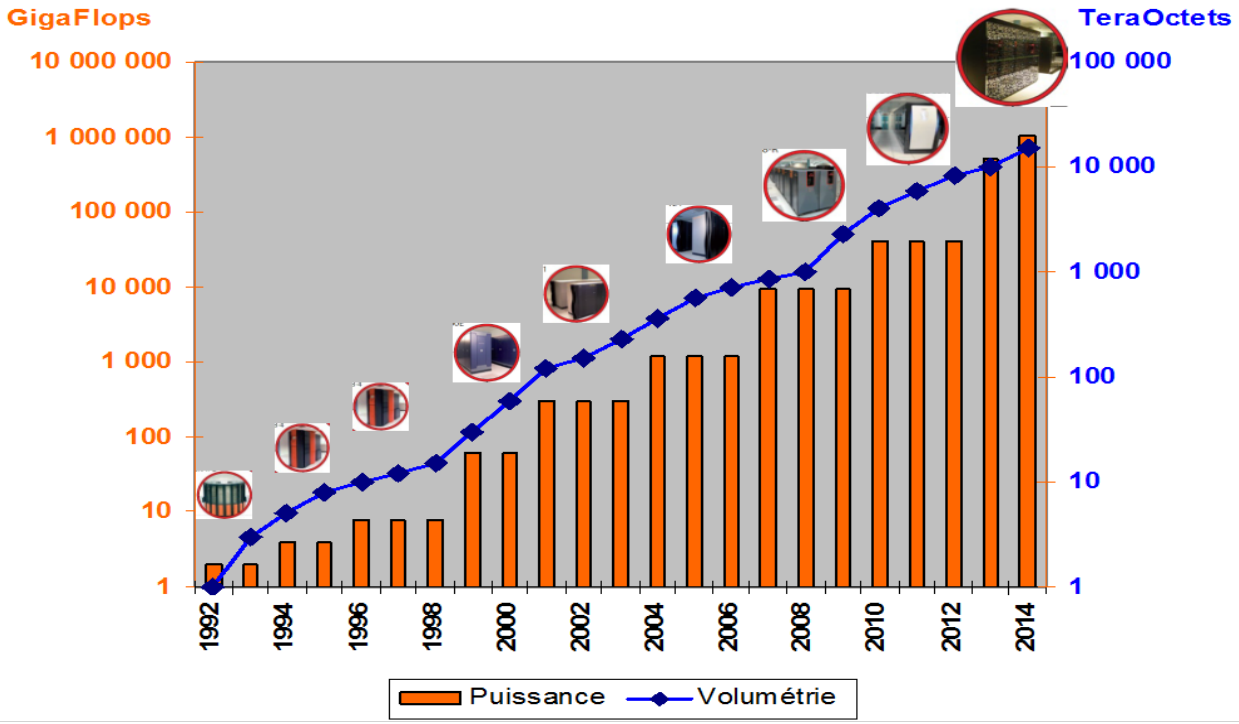
Images data

- Messages format :
  - GRIB, BUFR, TIFF , GEOTIFF
- Database : BDI (PostgreSQL)
- Volumetry : 55 GB (15 GB/d)
- Retention : 3d

Growth of the archive

- MF archive grows exponentially (60% increase per year)
- The daily amount of data added to the archive grows exponentially at the same time

- A large part of research data is kept forever
  - For many studies, a data becomes useful once enough data has been accumulated
  - Deleting old data in an exponentially growing archive is meaningless

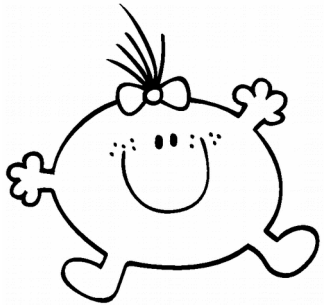


Currently

- Many small files BUT tape storage more efficient with large files
- Only FTP interfaces
- Free organization of data for each user → non-trivial sharing
- Few access services

In the future …

- MF decided to install MARS (Meteorogical Archive and Retrieve System) by 2020
  - Ease of access via directive files (the user does not care about the location)
  - Scalability
  - Optimization of the global flow of requests
  - Optimization of the tape reading
  - Rationalization of the storage space with control of the number of files



Data delivery evolution

- MF is continuously evolving its data delivery methods and services to cater for the new user requirements
- Interoperability is key
  - Follow standards and governance to enable interoperability
  - OGC, INSPIRE, ISO 19xxx series, NetCDF-CF, WMO Information System, GEOSS
- Provide high-level services on the data
  - Data portals for data discovery
  - Web services, REST APIs for data retrieval and manipulation
- Close-to-the-data processes
  - The era of pushing the data to the user is coming to an end
  - ….. The volumes involved are too large

WEATHER CLIMATE WATER
TEMPS CLIMAT EAU

Thank you
Merci

WMO OMM

World Meteorological Organization
Organisation météorologique mondiale