CGMS

CGMS-38 EUM-WP-38
v1A, 2 September 2010

Prepared by EUMETSAT
Agenda Item: WGIV/5

EUMETSAT DATA CENTRES AND ARCHIVE AND LONG-TERM DATA PRESERVATION

The importance of historic data and products derived from Meteorological Satellites, e.g. for Climate Monitoring and for improvements of numerical weather prediction models, has many implications for the Archives of the Meteorological Organisations regarding the easy access and the secure storage of data.

Easy access to the historic, archived meteorological satellite data, often in large amounts and across organisations, is necessary to fulfil the need of a growing number of users. Interoperability between partner organisations as well as efficient, standardised discovery, search, ordering and delivery of the data can help to meet this demand. Stronger cooperation between partners might also provide some leverage to the exponential growth rates of Archive ingestion and access.

Sharing global data sets between Archives could e.g. improve data access and increase data redundancy, the latter aspect being of great importance in the long-term preservation of Archive data. The historic data needs to be well preserved and maintained in order to provide and extend long Climate Data Records. Best practices need to be in place to ensure the availability and integrity of the archived data at an economic rate.

Action/Recommendation proposed:

Proposed Action 1
The CGMS members are invited to report on their measures and plans regarding interoperability and standardised online data access for archived datasets. Deadline CGMS-39.

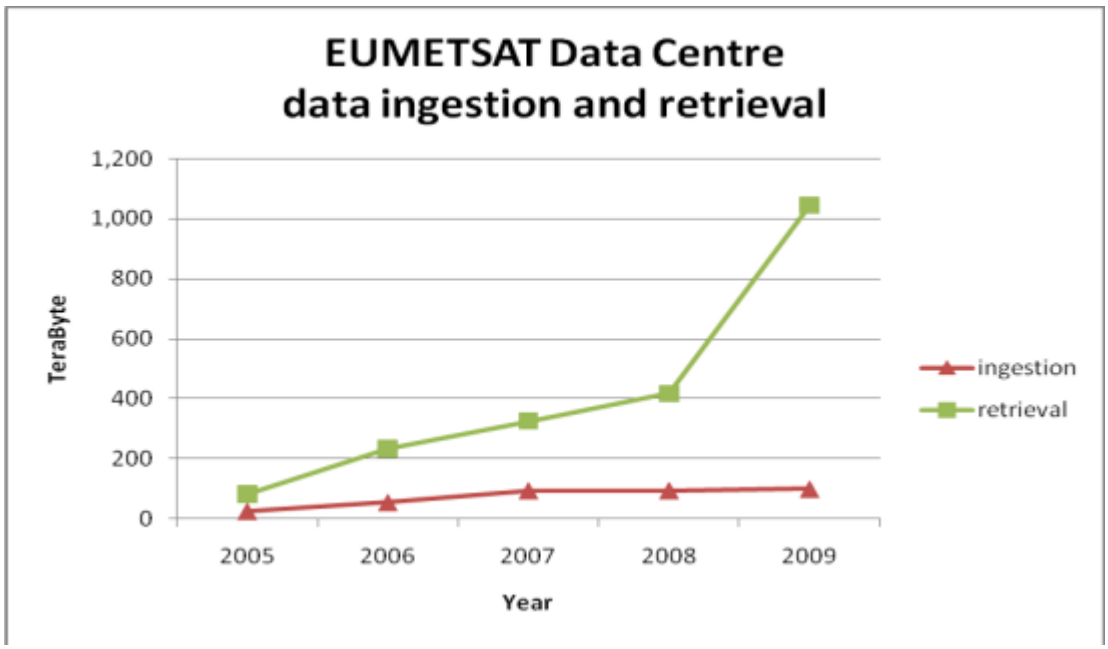
Proposed Action 2
The CGMS members are invited to report on the current measures taken in their Organisation for the long-term preservation of data and indicate if a future harmonized approach (e.g. common guidelines) would be helpful. Deadline CGMS-39.

EUMETSAT DATA CENTRES AND ARCHIVE AND LONG-TERM DATA PRESERVATION

1 INTRODUCTION

The EUMETSAT Data Centre is a large multi-mission facility located at EUMETSAT headquarters, which stores all the organisation's satellite data and derived products extending back to 1981 and helps users to access the archived data. Data Centre ordering is free of charge and amongst many other features allows the ordering of large amounts of data (bulk ordering) which is delivered online or on media. The EUMETSAT Data Centre is also the central node in a network of Archives encompassing the Satellite Application Facilities (SAF)

The Data Centre Network supports the data needs of external users such as the National Meteorological Services, research organisations, universities and commercial companies but also private individuals. It also supports internal users who need data for purposes such as research and reprocessing. Data retrieval is carried out in parallel with the constant ingestion of satellite data from EUMETSAT's satellites. User access to and orders for EUMETSAT data from the Data Centre are increasing continually and an exponential growth of the data to be ingested, archived and provided to the user community is expected.



This poses two main challenges:

1. From a user point of view, easy access to the historic data, often in large amounts and across organisations is expected. Efficient, standardised discovery, search, ordering and delivery of the data are very relevant in this context as well as interoperability between partner organisations.

2. The satellite data needs to be stored securely for a very long time to enable the provision of long-term climate data records. Long-term preservation of the ingested data is therefore a critical issue that needs to be addressed by all Archive Operators.

This Working Paper summarises the activities of EUMETSAT in these fields.

2 THE ROLES OF THE EUMETSAT DATA CENTRE

2.1 The EUMETSAT-SAF Data Centre Network and Partner Interoperability

As the central node in the existing Data Centre network, the EUMETSAT Data Centre provides a central catalogue comprising all entries of products generated in the SAFs and the Central Application Facility at EUMETSAT. Central ordering of all products, whether hosted in the Archive at EUMETSAT or in the local SAF Archives, is possible from the EUMETSAT Data Centre and a user-transparent order delivery process is ensured.

The central product catalogue is part of the EO-Portal implementation, which is EUMETSAT's single point of online access to all of EUMETSAT data and services (**please refer to EUM-WP-36**). The EO-Portal is EUMETSAT's response to the growing needs and demands for facilitating web-based, standardised data access and interoperability with partners. Once completed, it will allow Portal users a centralised discovery, search, order and subscription to data/services available from EUMETSAT and will even allow the same at cooperating Partner Agency sites – often referred to as a "one-stop-shop" to data /services.

Regarding the Data Centre, the EO-Portal will additionally provide a standardised web service interface based on OGC and ISO standards for a programmatic access, however, with limited functionality of data access (e.g. a limited choice of product retrieval formats). This interface is added to be compliant with interoperability demands from INSPIRE and GMES. Similar work will be done for the WMO/WIS initiative. Standards employed include OAIS, ISO19115/19119, OGC WMS/WCS and various OGC HMA.

Data Dissemination from the EUMETSAT Data Centre has strongly increased in the last years. The presented interoperability and standard access methods as well as the availability of new satellite data and longer re-processed time series will add to the further growth of historic online data ordering. In this context improved interoperability and cooperation between CGMS partners can be helpful to meet the growing user demand and facilitate easy access at the same time, e.g. by sharing global data sets to improve data access and increase redundancy.

Proposed Action

The CGMS members are invited to report on their measures and plans regarding interoperability and standardised online data access for archived datasets. Deadline CGMS-39.

2.2 Long Term Data Preservation in the EUMETSAT Data Centre

The archived satellite data are ultimately the biggest and most valuable asset of EUMETSAT and other Satellite operators. Some records date back over 30 years and are an indispensable data source for e.g. the climate monitoring community. The historic data needs to be well preserved and maintained in order to provide and extend these long records. In EUMETSAT, the new satellite programmes fund the transcription of previous programme archives so that they 'inherit' the existing archive, thus ensuring long-term data preservation.

A number of principles are followed to facilitate data access to the archived data in an operational environment.

Regarding the architecture of the Archive solution, emphasis is put on selecting a well-established Hierarchical Archive Storage Management solution and by employing Commercial off-the-Shelf (COTS) software as far as possible.

In EUMETSAT, all generated product levels (i.e. level 0, level 1, level 2) are normally archived. Data is stored in the native format of the programme (e.g. EPS native) but is written in standard format to tape (TAR). In the rare case of facility loss, no special environment is needed to retrieve the archived data, apart from a standard tape drive.

The Archive is embedded in the security concept of EUMETSAT's Ground Segment. Integrity of data sets retrieved from all media can be checked against a checksum (using SHA2) stored in its meta-data.

In terms of redundancy, all archived mission data is stored on two physical copies of different storage media (currently T10000 and LTO-4). One "online" copy is stored at EUMETSAT and a second "offline" copy is created at the time of data ingestion and is stored at a physically remote site. A third copy is maintained on disk for popular datasets (e.g. MTP&MSG level 1.5) to reduce tape recycling and increase performance.

With the current operational procedure data is transferred weekly to the offline site. Approximately every five years, data on the tapes is migrated to new media. Every two years the offline tapes are re-spooled at EUMETSAT to avoid tape-baking.

EUMETSAT is currently revising its data preservation policy especially regarding the risk of data loss and securing the integrity of the archived data due to the advent of higher density storage tapes. There are also review activities in the European EO sector triggered by ESA concerning long-term data preservation. In this context it would be beneficial to obtain an overview of the data preservation activities of the CGMS members to possibly derive a set of common guidelines ("best practices") for data preservation.

Proposed Action

The CGMS members are invited to report on the current measures taken in their Organisation for the long-term preservation of data and indicate if a future harmonized approach (e.g. common guidelines) would be helpful. Deadline CGMS-39.

3 CONCLUSIONS

Easy access to the historic, archived meteorological satellite data, often in large amounts and across organisations, is necessary to fulfil the need of a growing number of users. Interoperability between partner Organisations as well as efficient, standardised discovery, search, ordering and delivery of the data can help to meet this demand. Stronger cooperation between partners might also provide some leverage to the exponential growth rates of Archive ingestion and access.

Sharing global data sets between Archives could e.g. improve data access and increase data redundancy, the latter aspect being of great importance in the long-term preservation of archived data. Historic data needs to be well preserved and maintained in order to provide and extend long Climate Data Records. Best practices need to be in place to ensure the availability and integrity of the archived data at the most economic rate.