



CGMS-39 EUM-WP-38  
v1, 7 July 2011  
Prepared by EUMETSAT  
Agenda Item: WGIV38.5  
Discussed in WGIV

## **EUMETSAT REPORT ON LONG-TERM DATA PRESERVATION**

In response to CGMS action/recommendation 38/50

### Working Paper Abstract

EUMETSAT as an operational Meteorological Satellite data provider has a Long Term Data Preservation policy in operation since 1995, in order to maintain and provide access to its historic satellite data records. These records are maintained in the EUMETSAT Data Centre, a large multi-mission facility located at EUMETSAT headquarters which operationally archives all the organisation's satellite data and derived products.

Most importantly, the new EUMETSAT satellite programmes fund the transcription of previous programme archives so that they 'inherit' the existing archive, thus providing the means for long-term data preservation.

A number of principles exist and are followed to ensure a secure archiving of the ingested data in an operational environment and allowing the long-term access to the archived data. These principles are revised from time to time, e.g. due to the advent of new and larger storage media.

EUMETSAT is interested to share experiences and practices with other Meteorological Satellite Operators in view of establishing common guidelines for archiving.

## **EUMETSAT Report on 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 allows users to access the archived data. The EUMETSAT Data Centre is also the central node in a network of Archives encompassing the Satellite Application Facilities (SAF)

The archived satellite data, in total currently ca. 550 TB with an exponential growth rate prediction, 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.

### **2 Long Term Data Preservation in the EUMETSAT Data Centre**

A number of principles exist and are followed to ensure a secure archiving of the ingested data in an operational environment and allowing the long-term access to the archived data.

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 generally 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 tapesticking.

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. The assessment includes a ranking of the archived data to determine the level of preservation according to originality and importance of data sets for Operation and Users.

The preliminary evaluation suggests maintaining a minimum of three copies of the most valuable "vital" data, in order to avoid irrecoverable data loss in case of a disaster combined with a single failure in the recovery process.

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.

### **3 CONCLUSIONS**

EUMETSAT as an operational Meteorological Satellite data provider has a Long Term Data Preservation policy in operation since 1995 in order to maintain and provide access to the archived data records dating back to 1981.

A number of principles exist and are followed to ensure a secure archiving of the ingested data in an operational environment and allowing the long-term access to the archived data. These principles are revised from time to time, e.g. due to the advent of new and larger storage media.

EUMETSAT is interested to share experiences and practices with other Meteorological Satellite Operators in view of establishing common guidelines for archiving.