

## **AVAILABILITY AND USABILITY OF ARCHIVED SATELLITE DATA FOR CLIMATE STUDIES**

In response to CGMS action 33.05

It has been recognised and will become more and more urgent that the huge amount of remote sensing data contained in the archives of the various operational satellite agencies is made available for reprocessing. A prerequisite for this is that the quality of the data can be assured. In addition good calibration and intercalibration of sensor data is essential. The question of easy data access and delivery of data needs to be further pursued.

Therefore CGMS can contribute to the consolidation amongst operational satellite concerning these archive developments. A first step would be recommending a common data delivery format(s).

## **AVAILABILITY AND USABILITY OF ARCHIVED SATELLITE DATA FOR CLIMATE STUDIES**

### **1 INTRODUCTION**

The purpose of this paper is to give EUMETSAT's response to action 33.05 of the XXXIII CGMS meeting; namely satellite operators are requested to analyse the steps needed to make their archived satellite data usable for climate studies, and report to CGMS accordingly.

At present, all satellite operators offer similar types of services, an archive and retrieval service. However, the same satellite operators offer many different ways of accessing, describing, and retrieving their datasets.

For processing long time series for climate studies and optimisation of the user services, issues like data preservation, standardisation, and interoperability between the satellite operators need to be addressed to allow a more easy and effective processing of global datasets, in particular for creation of global datasets.

These efforts should be in line with the GCOS Climate Monitoring Principles, in particular with:

**No. 10.** Data management systems that facilitate access, use and interpretation of data and products should be included as essential elements of climate monitoring systems.

Therefore this paper briefly addresses some of the relevant issues:

- Data preservation and access methods
- Data quality and intercalibration
- Data Mining and access for third party algorithms
- Metadata

### **2 DATA PRESERVATION AND ACCESS METHODS**

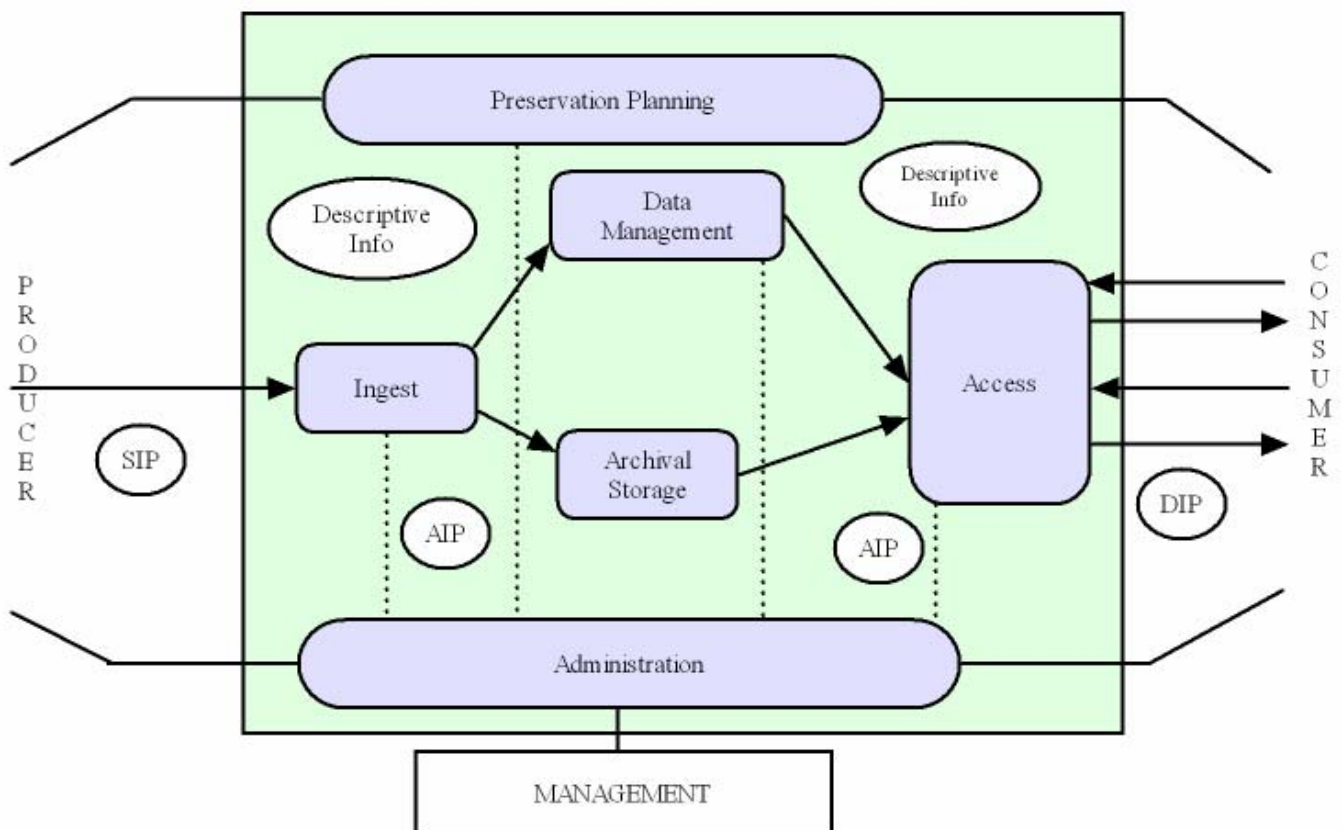
There are various means to achieve data preservation and these have been implemented differently at various archive centres successfully for many years, including EUMETSAT. However, a more standardised approach could be considered. Therefore EUMETSAT, amongst others, recommends the Open Archival Information System (OAIS) Reference Model as the basis for data preservation.

#### **2.1 OAIS Reference Model**

The OAIS Reference Model is an ISO standard. It has a number of mandatory responsibilities, these include:

- Negotiating and accepting information
- Obtaining sufficient control of the information to ensure long-term preservation
- Determining the "designated community"
- Ensuring that information is **independently understandable**, i.e. without the assistance of those who produced it
- Following documented policies and procedures
- Making the preserved information available

The OAIS reference model showing the functional flow through an archive is shown in Figure 1. A further advantage of the OAIS reference models is that it can be implemented in various ways depending on each organisation's requirements.



AIP: Archive Information Package  
 SIP: Submission Information Package  
 DIP: Dissemination Information Package

*Figure 1 OAIS Reference Model*

## 2.2 Standardisation & interoperability

The term *interoperability* has been talked about by satellite operators for sometime now, with no real actual progress, from anyone group. The main issues, as far as the archives are

concerned, are the standardisation of metadata, the quality of the data, access to the data, the delivery formats and how the data are finally delivered.

EUMETSAT recently conducted a user survey concerning the use of its archive. This survey contained questions on performance of the archive and what the users expect from it in the future. Nearly all users responded that they would like to get data in a single data format from the archive (e.g. NetCDF). It would be very much appreciated if a common standard delivery format would allow users worldwide to retrieve the data from the operational archives, independently of what agency or what original format the data was archived originally.

As a contribution towards this goal, the Hierarchical Data Format (HDF5) has been evaluated within EUMETSAT. However, it has been found to be unsuitable for large datasets (e.g. MSG level 1.5, AVHRR, IASI etc.), because the time necessary to perform the format conversion, the relatively small amount of compression of the datasets and problems with backward compatibility of previous versions of HDF. This led EUMETSAT to the conclusion that HDF5 might not be the final answer.

### **3 DATA QUALITY AND INTERCALIBRATION**

The quantity of data contained in an archive does not mean anything unless the quality of that data can be considered good. Therefore reprocessing of datasets which are known as being of inferior quality should be done with high priority. This includes that meaningful quality indicators should be annotated to the metadata.

Another issue of high relevance for operators of archives which are meant to contribute to climate monitoring is that the calibration and intercalibration of the various datasets has to be taken care of. This was stressed at the twelfth session of the GCOS/WCRP Atmospheric Observation Panel for Climate (Geneva, 3-7 April 2006) where it was particularly recommended that:

- Satellite operators are responsible for intercalibration of level 1.5 data (navigated and calibrated) data (visible to microwave e.g. SSM/I) (Recommendation 6)
- Satellite operators to evaluate existing archives to determine if they are data worthy for climate studies. The goal is to establish the current quality with respect to GCOS requirements and re-process (including navigation and calibration) archived satellite data, if required. (Recommendation 7)

In support to the above and more general principles, EUMETSAT is supporting the Implementation plan for a Global Space-Based Inter-Calibration System (GSICS). It is considered as important that availability and access to specialised calibration datasets is facilitated within the archives of operational satellite operators. These datasets should contain in addition to the instrument data, some ground reference data, model output (for computation and validation) and data from other available satellite systems.

### **4 DATA MINING AND ACCESS FOR THIRD PARTY ALGORITHMS**

Especially for generation of global datasets (but not limited to this) some third party algorithms from external users should be able to run at the regional archives to generate products there.

The third parties should be asked to leave their products in the various archives after processing of the data locally. This will add additional information to those archives and thereby make their content more valuable. The overall aim is to turn the data into information at various levels.

Instead of running product generation at a centralised place a distributed infrastructure for this purpose could be envisaged. CGMS could play a coordination role for the operational agency's archives through the coordinated formulation of a set of guidelines and recommendations.

Access to the data for on-site processing should be encouraged with qualified end users running their algorithms and contributing to the creation of global datasets. As a requirement the archive operators could ask these users to leave also their algorithms in place to create the latest product versions when necessary.

## **5 METADATA**

There is currently a CGMS Permanent action No. 10 which concerns the standardisation of metadata. There are amongst many, two promising alternatives; the WMO initiative based on the ISO 19115 standard and the ESA led Heterogeneous Missions Accessibility (HMA). Both have different advantages and further study still has to be done to ascertain the best way forward.

## **6 CONCLUSION**

It has been recognised and will become more and more urgent that the huge amount of remote sensing data contained in the archives of the various operational satellite agencies is made available for reprocessing. A prerequisite for this is that the quality of the data can be assured. In addition good calibration and intercalibration of sensor data is essential. The question of easy data access and delivery of data needs to be further pursued.

Therefore CGMS can contribute to the consolidation amongst operational satellite concerning these archive developments. A first step would be recommending a common data delivery format(s).