



CGMS-39, CMA-WP-11
Prepared by CMA
Agenda Item: GIV/5
Discussed in WG-IV

CMA Report on Long-Term Data Preservation

Summary of the Working Paper.

The CMA working paper reports the operational procedure for the long-term preservation of data. At NSMC/CMA the received multi-mission data are stored online, near-line, offline, respectively indicating the data archived on disk array, automatic tape library, and tapes preserved elsewhere in the NSMC operational building. All generated products (i.e. level 0, level 1, and level 2) are archived and preserved in two copies for long-term preservation. The oldest archived product traces back to 1983. CMA is interested in experience and practice of other Meteorological Satellite Operators regarding the data preservation and welcomes the proposal of establishing a common guidance for archiving.

CMA Report on Long-Term Data Preservation

1 Introduction

The National Satellite Meteorological Center (NSMC) has been receiving the meteorological satellite data since 1978. The NSMC Data Service Center located in the NSMC Building archives and stores the data from Fengyun missions, NOAA, Meteosat, GMS, MTSAT, GOES, EOS/TERRA, EOS/AQUA, and MSG. With FY-3 series satellite coming into operation in recent years, the volume of archive is growing tremendously fast, the total amount of 2PB raw data and derived products have been preserved.

2 Operational Procedure for Long-term Preservation of Satellite Data

At NSMC, data is stored online, near-line, and off-line. The data stored on disk array that constitutes the file system is the Online Archive. The Near-line Archive is the data stored in the automatic tape library. The Offline Archive is the data on tapes off the automatic tape library and preserved elsewhere in the NSMC operational building.

All generated products (i.e. level 0, level 1, level 2) are archived. Data is saved in the format of the program. The metadata is read out and stored in the metadata data base for the convenience of data management and service.

Every day the received data is first stored in the disk storage, for duration being one week of the shortest to 3 months the longest, varying according to the data category. The products of last 24 hours in the disk file system shall be archived onto tape and stored in respective pool of the tape library by category. Each product is saved on two tapes, one of which preserved at an offline site but in the same building at present. With the current operational procedure the online data on the disk is cleared off according to the saving cycle based on the data category; the near-line data on tape is transferred to the offline site according to the near-line data saving cycle.

The automatic tape library is the IBM device, IBM-3494, which has 28 tape drivers, maximum capacity 5,000 tapes, able of operation on tapes, for instance, re-spooling to avoid tape-sticking. Data on tapes is migrated to new media, usually at circumstance when higher density storage tape is introduced into the system or tape driver changes. The data that used to be stored on tape T6250, T3480, STK-9840 have been migrated on tape IBM-3592 recently.

3 Conclusion

NSMC preserves the received multi-mission data that can trace back to 1983. The archived data is made accessible on the NSMC website. Online and near-line data can be called back at user's request, but the request for large bulk data or offline data needs the manual service. 20,000 users have registered the NSMC data service system, and download approximately 200TB data a year.

NSMC is interested in experience and practice of other Meteorological Satellite Operators regarding data preservation, and welcomes the proposal of establishing a common guidance for archiving.