MET OFFICE SATELLITE DATA VOLUME PLAN
by
Roger Saunders (Met R&D/SA)

Version: 7.0
Dated: 3 January 2008

List of Contents

1. Rationale and Scope

The routine processing of recently acquired global satellite datasets (particularly AIRS, ATOVS and SSM/I) has stretched the capabilities of some of our observation processing and database systems. However, these datasets are not large compared with those from some satellite instruments that the Office is planning to use over the next 5 years, in particular those from METOP launched in October 2006 and NPP due for launch in 2010.

The Met Office is committed to a very large investment in satellite systems through EUMETSAT, with the expectation of significant improvements in NWP and other forecast products resulting from effective use of these new data. It is therefore essential that we make adequate preparations to receive and process these data in all components of relevant infrastructure. Figure 1 summarises the problem not in terms of data volume but in terms of number of satellite instruments that the Met Office is expected to use data from for operational NWP.

The Satellite Data Volume plan (SDVP) quantifies the expected volumes of future satellite data and thus serves as an input for development projects on various components of Met Office infra-structure.

The SDVP only includes satellite data for which NWP is a customer. It relates to those data which are (or maybe in the future) assimilated into the NWP models (Global, NAE and UK 4/1.5km) and hence are stored in the MetDB. It does not include the satellite data received locally via AUTOSAT unless the product is subsequently archived in the MetDB for NWP use. The AUTOSAT and NIMROD archives are therefore not included in this plan. Only significant satellite datasets are included in the tables. There are also many ancillary datasets such as radiance bias correction files, satellite data monitoring files, etc. which are stored on the Linux servers but not included in the data volumes considered here. The SDVP does not attempt to estimate the load on computer systems of increasing amounts of satellite data recognising it is hard to anticipate improvements in computer technology over the next 5 years.

The SDVP summarises satellite datasets in terms of gross volumes entering the Met Office each day in MBytes/day and best guess start/end dates based on current estimates of launch dates from the satellite agencies. It also estimates the volumes of satellite data in the MetDB archive for several different archiving options. It does not address details of data format, transmission media, frequency of receipt, processing, etc.

The SDVP contains a summary of **past** and **current** satellite datasets used to serve NWP activities. This provides a baseline for assessing capacities required to handle future data. It contains estimates of **future** satellite datasets that are expected to be used in the same activities from the present to the year 2013. It will be updated annually to take account of revised launch schedules, more accurate estimates of data volume and new sources of data over the forthcoming 5 year timeframe and revised time scales for exploiting the data in NWP. The old estimates from previous year's plans are also kept to show how accurate they were and are a guide to the accuracy of future projections. The first version of the SDVP was issued in 1999 and so there is a history of projections made.

2. Definitions and assumptions for the Satellite Data Volume Plan

The SDVP includes all satellite datasets which the Met Office will, or may, use for operational (or other near real-time) activities. An entry in the SDVP does not imply an established business case to receive and process the data; it will provide one of the inputs to activities in which such business cases are considered. Some of the data are referred to as **operational NWP**, which means they are used in near real time for operational NWP or forecast products and so the loss will potentially impact the NWP index and hence Met. Office key performance targets. Other data referred to as **research** are either being experimented with prior to operational implementation or in some cases are just building a climate dataset (e.g. AATSR) so it is not critical to the NWP operations but nevertheless must be processed in near real time as the alternative of retrospective processing is complicated and expensive. Only research data for which there is a well-defined interest as set out in the NWP Research and Development Plans or Climate Prediction Programme are included here. For 3-5 years ahead it must be recognised that the assumptions for research data may change.

The SDVP also covers volumes at various points through the Met Office's processing systems, including: volume at entry point to the Office (referred to as **data received**), input to the MetDB (referred to as **MetDB observational data)**, and output from OPS/VAR as a merged file (referred to as **MetDB merged data)** defined in more detail below.

2.1 Data Received

This refers to data that is received at the Met Office before any local pre-processing has been performed. It is often in BUFR but not always. The data are normally on the GTS but increasingly with satellite data they are being transferred by dedicated lines (e.g. Washington-Exeter link) or received from a direct broadcast satellite (e.g. EUMETCAST) or for research data via the internet. For future datasets best estimates for their volumes are made based on current data volumes, figures provided by the space agencies and in some cases only rough estimates can be made based on the data characteristics. Italics indicate estimated values in the data tables. Once received these raw data are often pre-processed locally, typically on LINUX servers, which can include several steps before the data are sent to the MetDB normally in BUFR or GRIB:

- ingest data from native or BUFR formats
- uncompress data if necessary
- pre-screen out data of no interest for application (e.g. obviously cloudy IR radiances, scatterometer winds over land etc.)
- combine data from various sources (e.g. remap AMSU-B and AMSU-A to HIRS grid)
- create product from raw satellite data (e.g. MSGRAD BUFR)
- encode the pre-processed data in BUFR (e.g. ATOVS level 1b to ATOVSG)

These steps can significantly reduce the data volume before it reaches the MetDB. For some data types the received data are sent on to other European centres either before or after pre-processing. For received data it is envisaged that only short term rolling archives of from 7-10 days are necessary. For older data, as a general rule, the originating centre should be able to provide the data on request.

2.2 MetDB observational data

The observational data files, as their name implies, contain only the observed data. These files are read by the OPS as part of the pre-processing for the real time NWP assimilation process. For observations that are compared with model fields these files may be transitory as they can be overwritten by the merged data files (see sec 2.3) which still contain the original observed data (however note archiving options below). In most cases the observational data files are retained for long term archive.

2.3 MetDB merged data

Once the data have been passed through the OPS and (optionally) assimilation system, retrieved products, the model equivalents and quality control flags can be appended to the observed data files to form a new merged data file. This can potentially lead to much bigger files (e.g. for ATOVS by a factor of 3) which are archived, instead of or in addition to the original observation files. Feedback from both the OPS and the VAR assimilation to the merged data file can occur. Note the VAROBS files that are temporarily stored (specifically for input to VAR) are not included in the data volumes presented here, as they are not archived long term. Note that AIRS and IASI radiances do not have a mergeback file stored in the MetDB but they do have an equivalent archived in MASS. For data assimilation experiments the observational data can either be read from the merged BUFR data file or more often from prepared OBSTORES which are temporary files set up for trial periods to ease the observation ingest load from the MetDB.

As part of the proposed logical data store (LDS) there are plans to rationalise the storage of mergeback data which will affect the data volumes but timescales are unclear at present. For this report it is assumed mergeback continues as is (i.e. we store both the original observation files and the observation + mergeback files). It should also be noted that Met R&D are currently reviewing their need for mergeback files. The observational data base (ODB) originally developed at ECMWF and now being ported to the UM systems by the Australian Bureau of Meteorology may also be adopted in the future and change the requirement for mergeback data.

2.4 MetDB Archive

Once the observed satellite data have been stored in the MetDB they are archived there either as an observation file or as a merged file for a fixed period currently defined in the MetDB Service Level Agreement (http://www01/Midas/General_docs/SPA_00-01_02.html) dated 16 Feb 2000. Once access is provided the transfer from disk/tape to memory should be < 1min for a 25 Mbyte file. The period the data is kept on-line varies according to data type (see Table in annex A of above link). Offline storage in MASS assumes the originally received satellite data is retained for 10 years and mergeback files for 12 months as agreed in MetDB UG minutes of meeting on 24/10/01 and subsequent meetings.

This plan gives accumulated satellite data volumes in the MetDB archive for three options. One assuming all the satellite data and model mergeback are retained indefinitely, one assuming all the observations are retained indefinitely but the merged data are only retained for 12 months and finally the retention periods currently agreed of 10 years for the observed data and 12 months for the mergeback files.

2.5 Assumptions about satellite data and their formats

The launch dates for future satellites assumed in this plan are set out in Table 1. They all err on being slightly later than the 'official' launch dates as experience has shown that it takes several months for near real time data to start to flow reliably and be ingested into the MetDB. ATOVS on METOP was exceptionally fast at 3 months. It is also assumed it takes a year before the data are assimilated operationally for a new data type with the exception of ATOVS and AMVs.

SATELLITE NAME	**LAUNCH DATE** (*projected*)	**AVERAGE ALTITUDE** Nominal (km)
DMSP F-18	2008	~800
NOAA-N'	2009	850
NPP	2010	~833
METOP-B	2011	837
SMOS	2008	755
Aeolus/ ADM	2009	405
FY-3A	*2009*	836
NPOESS-C1	*2013*	~833

Table 1. Launch dates assumed for forthcoming satellites. Italics denotes this is a guess as no official dates are available yet.

There are a few critical assumptions made to determine the satellite volumes estimated here. They are itemised below:

⇒ Unless otherwise stated below the only compression of the data is to encode it in BUFR. This typically reduces data volumes by a factor of 4.

⇒ There are now ***six operational*** polar orbiters (4 NOAA, Aqua, METOP) and two DMSP polar orbiters providing global ATOVS, AIRS and SSM/I(S) radiances and local area ATOVS radiances. The projected use of data from ATOVS sensors is summarised in Figure 2, mergeback is required for the data used (green) but not the 'hot spare' (red) sensors. After the launch of a new satellite it is assumed a 6-month period where the new data is archived in *addition* to the operational data from the oldest satellite after which time one of the old satellites is removed from the data stream.

⇒ It is assumed mergeback files are only produced from operational model runs not assimilation trials

⇒ The mergeback from the stratospheric model has been included up to end 2005 (which doubles the ATOVS merged data volumes) and the NAE is assumed to require mergeback files from 2006 onwards.

⇒ No date for the planned termination of the NAE is assumed.

⇒ Both global ATOVS (ATOVSG) and locally received/EARS ATOVS (ATOVSL) are stored in the MetDB.

⇒ For ATOVS only ATOVS 1D (ATOVSG/L) and AMSU-B 1D (AMSUB) files are stored in the MetDB.

⇒ The original imagery data are available from all 5 geostationary satellites and are stored in GRIB as the TCR product, which has been modified to take into account SEVIRI data

from 2006.

⇒ ERS-2 scatterometer winds and other data will continue until end 2010.
⇒ ENVISAT continues until 2013.
⇒ Quikscat Seawinds data will cease end 2010.
⇒ Aqua will continue until 2015 providing AIRS, AMSU-A and AMSR-E data.
⇒ WINDSAT will be assimilated operationally from 2009 and be available until 2013.
⇒ AMSR-E will be assimilated operationally from 2009.
⇒ Only one METOP dataset will be archived and assimilated.
⇒ IASI global data will increase by a factor of 4 in 2010 (all fovs kept)
⇒ Only one SSM/I(S) radiance data stream will be received and assimilated.
⇒ MSG clear sky radiances will be stored in the MetDB both as EUMETSAT CSR and AUTOSAT MSGRAD BUFR radiance products.
⇒ CrIS, ATMS and VIIRS data will be received from 2010 but the full VIIRS dataset will not be stored in the MetDB.
⇒ CrIS data is assumed to have a similar data volume to IASI.
⇒ There will be no mergeback files for AIRS, AMSR-E, IASI and CrIS.
⇒ NPOESS-C1 will not be launched before 2013.
⇒ SMOS data will be received and archived.
⇒ Data from the Chinese FY-3 polar orbiter series will be stored for research from 2010.

3. The Format of the Satellite Data Volume Plan

The input data to the plan is presented in tabular form in Annex-A which lists the projected satellite data volumes for each year. Actual data volumes provided by the MetDB team are in normal type, estimated data volumes are in *italic type*. The data in these tables are also summarised in graphical format. Figure 3 provides an overall picture of actual and projected typical data volumes received at the Met Office in one day and the volume stored in the MetDB. Figure 4 shows the projections produced in previous years, which show how accurate our estimates were in previous years and also how they can change for various reasons. Figure 5 shows the volumes of different categories of data according to their status in the MetDB. Figure 6 shows the projected growth of the satellite data archive in the MetDB separated into the different satellite data types assuming an indefinite retention period. Figure 7 shows the archived data volumes assuming 3 different archiving options. The volumes plotted in Figures 6 & 7 assume that new satellite datasets start to be received on 1 Jan of the year ***after*** their nominal launch date. As the launch dates assumed in Table 1 are often subject to slippage the SDVP volumes presented here are the maximum possible data volumes required. In practice some of the datasets may be delayed or even cancelled due to funding cuts (e.g. NPOESS) or sensor failure (e.g. AMSU-A1 on NOAA-17, ADEOS).

4. Summary of SDVP analysis

An analysis of the plan provides the following insights into the volume of satellite data that is required to be received and archived at the Met. Office over the next 5 years given the above assumptions:

- With the launch of METOP-A in 2006 there has been a ***large increase in the volume of satellite data received*** from 4.4 Gbytes/day in 2006 to 23.6 Gbytes/day in 2007. After

2007 the data volumes received only increase slightly until 2011 when it increases to 33.6 GBytes/day due to NPP.

- With the launch of METOP-A in 2006 there has been ***an increase in the volume of satellite data to archived in the MetDB between 2006 and 2008 from 6.5 to 8.4 Gbytes/day.***
- The large volumes of received data provided by METOP and later NPP (i.e. global IASI and CrIS data) dictate that the pre-processing on ingest has to reduce the data volumes by ~90% for some data types to be manageable levels for the MetDB and assimilation systems. Currently in order to limit the data volume stored in the MetDB only 1% of IASI data is stored. **In order to make fuller use of the IASI data we should plan to increase this percentage significantly even with planned compression techniques.**
- The operational merged data files dominate the MetDB daily data volumes by about a factor of 2 up to 2006 (see Fig 5). After 2006 mergeback files occupy less than the observation files before mergeback. The decision has been made not to have mergeback files for the high volume datasets such as IASI and CrIS. ***It is recommended mergeback data is rationalised for all data types for the MetDB.***
- The *research data* volumes for a new dataset are large in the first year of a new satellite (e.g. 2006, 2009) but decline in later years as the data become *operational*. The AMSR-E, WindSat and NPP launches all provide datasets that will not be assimilated immediately but are planned to be after some testing. Typically research data occupy 20% of the MetDB.
- Polar orbiter radiance data is the dominating data type in terms of volume, providing 85% of the total data volume in 2007 and remaining the same percentage in 2013. Scatterometer winds provide about 6% of the data volume in 2007 but decrease to 3% by 2013. Atmospheric motion winds, ozone retrievals and GPS are insignificant in terms of data volume. Miscellaneous products comprise ocean and land surface products.
- The current **annual** satellite data volume estimates are 2.8 Tbytes in 2007 rising to about 4.4 Tbytes by 2013.
- By 2013 the **accumulated** satellite data archived in the MetDB will be 17.8 Tbytes if all the data since 1999 is retained (mergeback deleted after 1 year). If the obs files are archived **for only 10 years** the accumulated volumes by 2013 reduce to 14.1 Tbytes (see Fig. 7). **The MetDB users should now decide if the 10 year retention policy is to be invoked.**
- It is worth noting from Figure 4 that the data volume estimates made in 2003/4 for 2007 were slight overestimates of the actual data volumes for 2007. For this year's estimate beyond 2007 the trend is for similar increases in data volumes than those projected in previous years. This figure demonstrates that past years have shown **the projections from this plan are close enough to reality to be useful for planning purposes.**

5. Recommendations

Assuming the requirement for archiving satellite data in the MetDB continues the required storage for satellite data in 2008 will be over 5 TBytes assuming a 10 year retention of observation and 1 year retention of mergeback files. This will rise to 14 TBytes by 2013. As a result the following recommendations can be made:

- The data volumes assume that the advanced sounder data from IASI and CrIS are compressed. This is not the baseline for the space agencies at present and so they need to be urged to adopt data compression options for these data to reduce strain on the data

ingest and archive which Met Services have to cope with.

- The policy of retaining satellite data for 10 years and mergeback data for 12 months should be reconsidered and reaffirmed by the MetDB user group so this decision is properly documented.
- Met R&D should review the use of mergeback data for different models with a view to having a consistent system across all data types.
- The IT development groups should remain aware of this document and provide feedback on the feasibility and costs of providing an archive for these data volumes.
- The satellite data volume plan will continue to be updated annually by SA to maintain accuracy and a 5 year outlook.

Figure 1: Number of satellite sensors for NWP

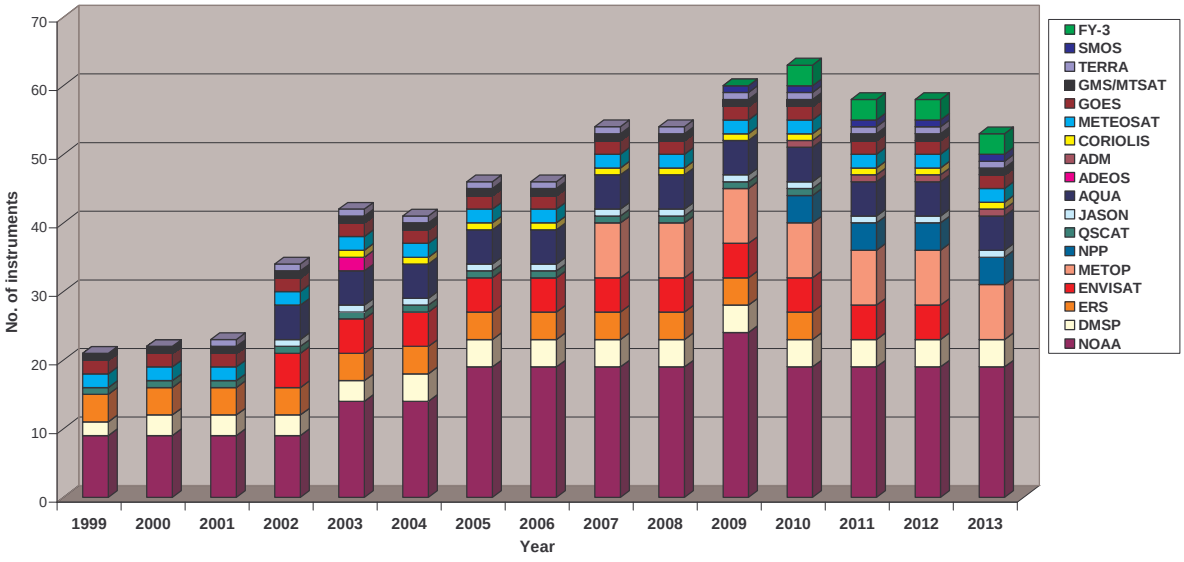


Figure 2: Projected ATOVS data stored in MetDB

Platform Eq X time	Sensor	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
NOAA-15 09LT	HIRS											
	AMSU-A											
	AMSU-B											
NOAA-16 14LT	HIRS											
	AMSU-A											
	AMSU-B											
NOAA-17 11LT	HIRS											
	AMSU-A											
	AMSU-B											
Aqua 13:30LT	HIRS											
	AMSU-A											
	AMSU-B											
NOAA-18 14LT	HIRS											
	AMSU-A											
	MHS											
NOAA-19 14LT	HIRS											
	AMSU-A											
	MHS											
METOP-A 10LT	HIRS											
	AMSU-A											
	MHS											
METOP-B 10LT	HIRS											
	AMSU-A											
	MHS											

Used
Hot spare

Figure 3. Daily data volumes entering Met Office

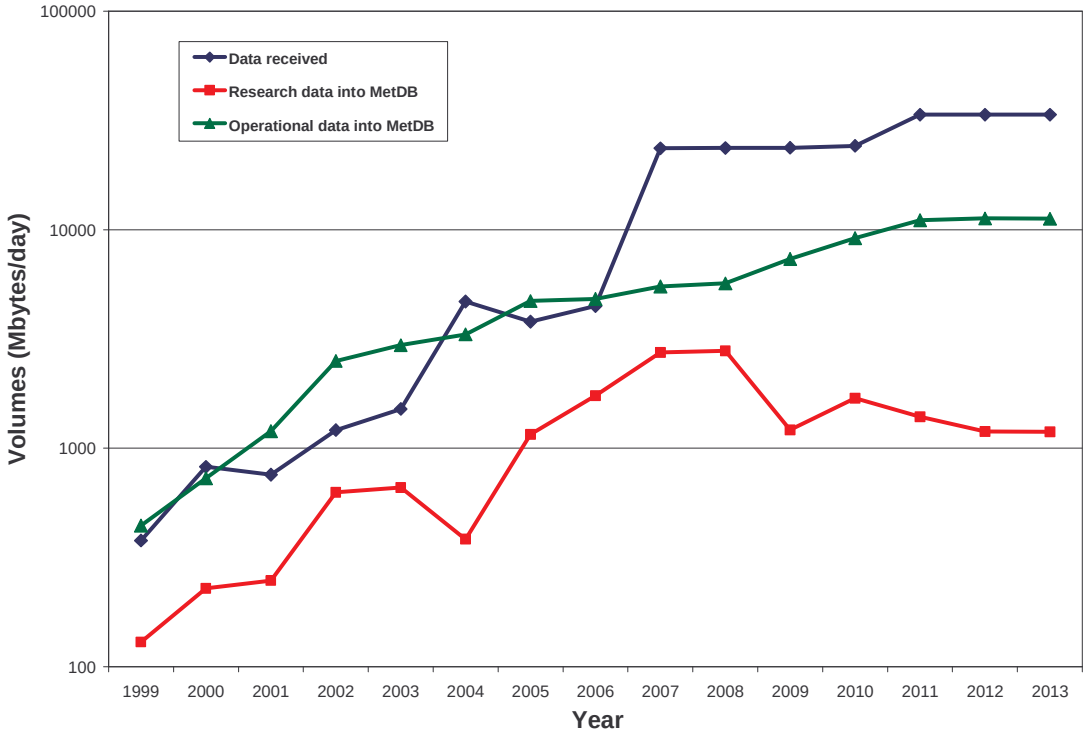


Figure 4: Projected satellite data volumes ingested into MetDB based on past and present best estimates (volumes based on input obs + mergeback + research files)

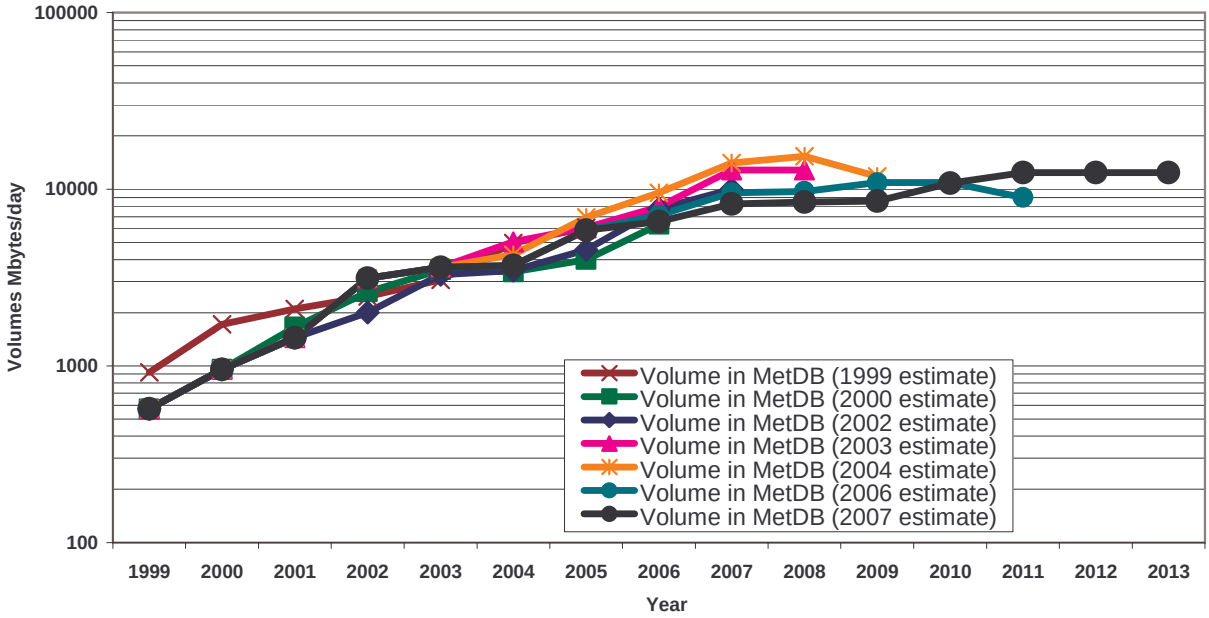


Figure 5: Different categories of data stored in MetDB

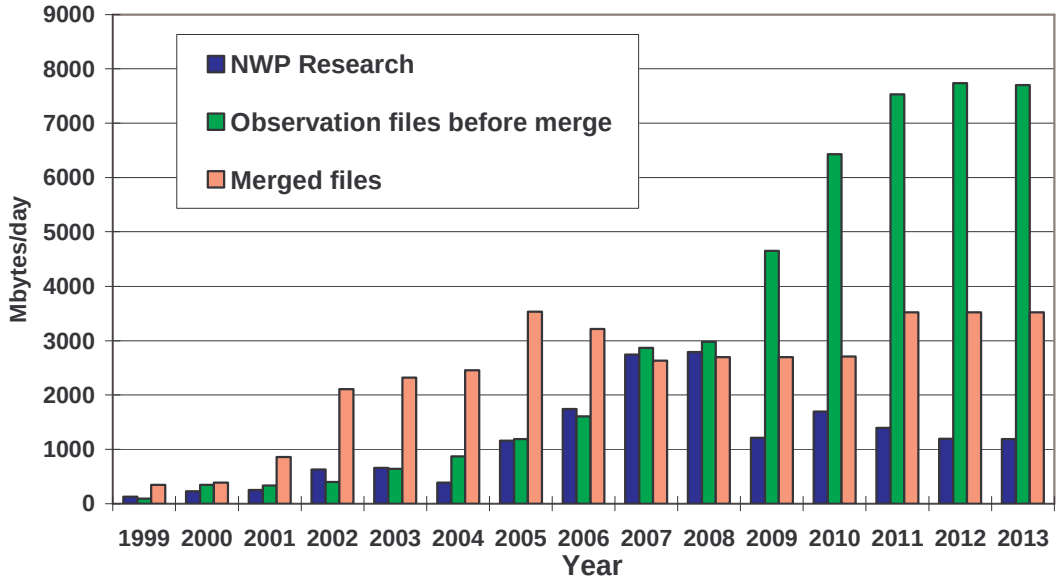
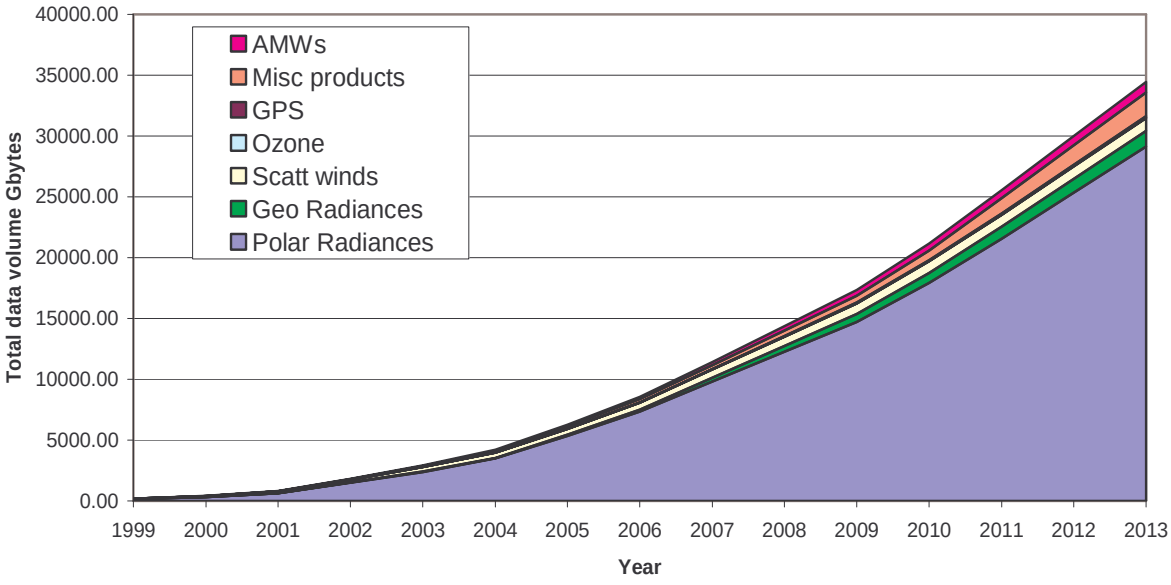


Figure 6: Projected growth in data volumes for various types of data assumes combined obs/mergeback and obs research files archived indefinitely



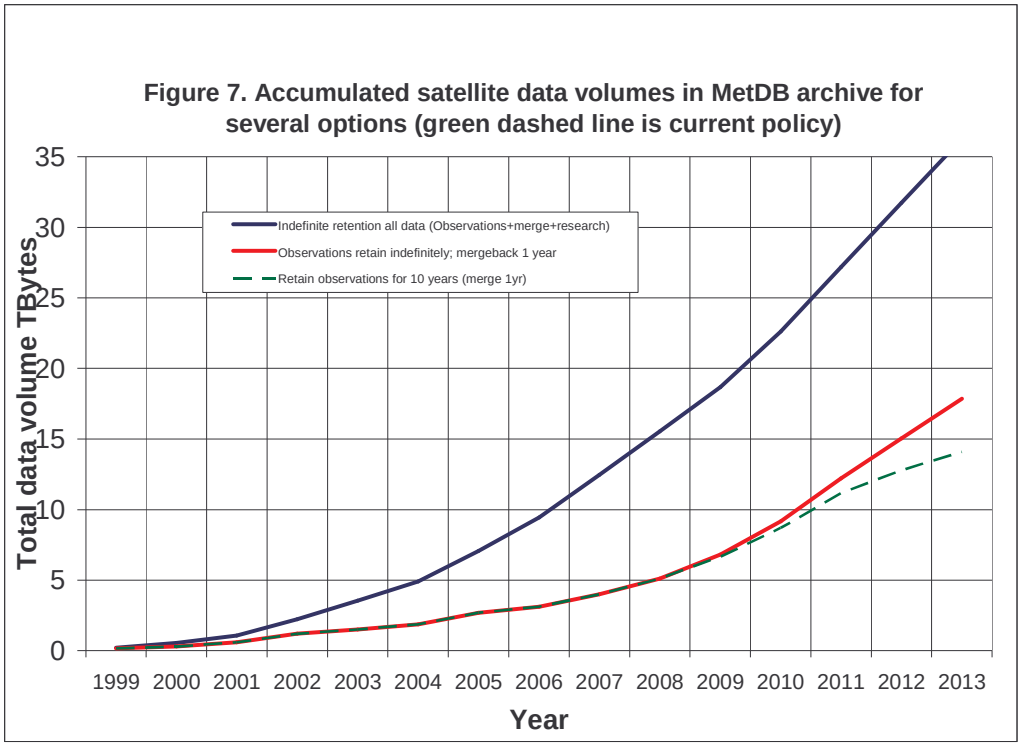


Figure 7. Accumulated satellite data volumes in MetDB archive for several options (green dashed line is current policy)
Total data volume TBytes
35
30
25
20
15
10
5
0
Indefinite retention all data (Observations+merge+research)
Observations retain indefinitely; mergeback 1 year
Retain observations for 10 years (merge 1yr)
1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013
Year