The Challenges of Remote Sensing Data Preservation and System Renewal during Continuous Operation

W. Wildegger, A. Huber, E. Mikusch, K.-D. Missling, P. Seifert, H.-J. Wolf

German Aerospace Center (DLR)
German Remote Sensing Data Center (DFD)
Oberpfaffenhofen, D-82234 Weßling, Germany and
Kalkhorstweg 53, D-17235 Neustrelitz, Germany
EMail: Wilhelm.Wildegger@dlr.de

ABSTRACT

At DLR/DFD the Data and Information Management System DIMS is used for handling and archiving remote sensing data. One task of DIMS is to support ongoing and upcoming new earth observation missions, the typical lifetime of an individual mission being 5 to 10 years. New missions (like TerraSAR-X or MetOp) tend to have additional requirements with respect to the functionality of the application software but also higher requirements with respect to the performance, reliability and availability of the whole system (hardware, Commercial Off The Shelf (COTS) software and application software). The other important task of DIMS is to carefully preserve the valuable remote sensing data beyond the mission's duration for future generations, since the earth's surface and climate continuously change and data gathered today cannot be regained later if lost. This task is commonly known as long-term archiving. Both the continuous support of missions and long term archiving require keeping DIMS in continuous operation; at the same time hardware and COTS software have to be renewed due to limited service lifetime and to meet new reliability and performance requirements. New versions of the application software have to be brought into operation as well. The challenge is renewing DIMS hardware and software components without interrupting operations for more than a few days. A similarly great challenge is preserving the remote sensing data currently archived in DIMS (ca. 150 TByte) by migrating them to new storage media while continuously providing access.

Keywords: Data Preservation, Long Term Archiving, System Migration, Data Migration, System Renewal Without Interrupting Operation, Data and Information Management

INTRODUCTION

The German Remote Sensing Data Center (DFD) of the German Aerospace Center (DLR) is concerned with the remote sensing of the earth's surface and atmosphere with sensors on board satellites and research aircraft. The areas of engagement are the complete payload ground segment and the long-term preservation of an extensive collection of earth observation data as a national resource. This comprises the reception of raw data from remote sensing satellites with receiving stations throughout the world, sensor specific processing of the raw data to basic products, and generation of value added products. Furthermore, it covers data management tasks common to all missions and sensors like production control, short and long term archiving including cataloging for all levels of products, user services offering the products via a Web interface (http://eoweb.dlr.de), and handling of orders including the delivery of products via FTP or on media. Whereas the basic processing of the sensor raw data and subsequent value adding processing require specific algorithms and systems, the data management tasks are unified for heterogeneous types of remote sensing data and different processing levels. To avoid re-inventing and implementing a data management system for every upcoming mission/satellite, DLR/DFD has developed and operates a multimission system, named Data and Information Management System (DIMS), which is capable of supporting many different missions and sensors in parallel and can easily be adapted to the requirements of new missions/sensors. DIMS is operated at the two DFD sites

Oberpfaffenhofen and Neustrelitz. Fig. 1 shows the DIMS components of the Oberpfaffenhofen site and associated processing systems (ingestion, processing, and post-processing systems).

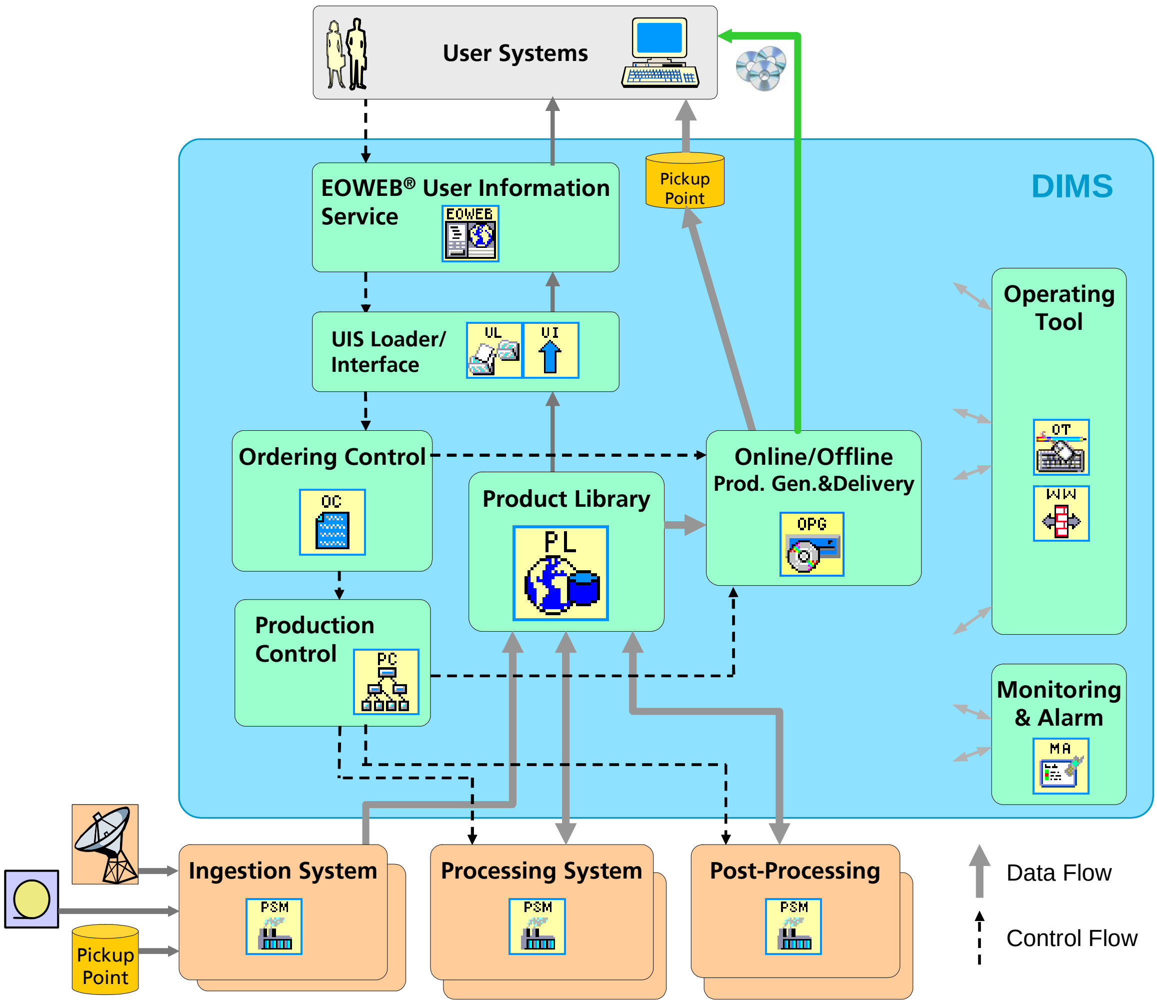

Fig. 1 DIMS Components and Main Communication Paths

NEED FOR NEW HARDWARE

The aim of supporting new missions with increasing amounts of data to be handled and stored as well as the ongoing task of long term archiving made it necessary to look for a new hardware platform for DIMS. The old DIMS hardware platforms dated from 1999 (some tape drives had even been in operation since 1995). It was obvious that most of the hardware components (servers, storage devices, tape drives) were approaching the end of their service life. This would have been reason enough to procure new hardware; however the type, performance and configuration of the new hardware was strongly influenced by a number of other issues:

The performance of the old hardware was no longer sufficient for the data rates and amount of data expected from upcoming new missions; therefore the new hardware had to offer **more performance**, the servers had to have higher processing speed and I/O, and the disks systems and tape drives faster data throughput. The disk storage space of the old hardware was exhausted; **more disk storage space** was required. The large amounts of data from the new missions required more archive space, i.e., **more space on tape media**. The previous concept of providing storage space using locally attached disks was considered out-of-date and inflexible; a **more** up-to-date and **flexible storage concept** for disk storage was to be introduced, providing an easy way of assigning storage space to individual servers/applications and the possibility to easily expand the overall storage capacity. The previous way of achieving data safety was too operator-intensive and conceptually imperfect: The new hardware had to achieve **higher data safety** in a reliable and efficient way. Finally, the upcoming missions asked for **higher availability** to an extent which was not possible with the old hardware concept.

New Hardware versus Old Hardware

Table 1 compares the old DIMS hardware with the new DIMS hardware for the component groups servers, disk storage and tape storage. It shows the Oberpfaffenhofen hardware configuration; the Neustrelitz hardware configuration is similar, although smaller versions are used with some components, (e.g., the new archive server is a V890 in Neustrelitz instead of V4900 in Oberpfaffenhofen, or the 2nd copy library is an i2000 with 400 slots in Neustrelitz instead of an i2000 with 700 slots in Oberpfaffenhofen. Fig. 2 and Fig. 3 below visualize the old and new DIMS hardware configuration.

Hardware component	Old hardware platform	New hardware platform
Servers		
Servers for DIMS components with lower performance requirements	Sun E450	Sun Fire V490
Servers for DIMS components with higher performance requirements	Sun E450 / E3500	Sun Fire V890
Archive server	Sun E6500	Sun Fire V4900
Disk storage		
Connection of disk storage to servers	Locally attached at every server via SCSI	Via FC based SAN in redundant configuration with two brocade 4100 SAN switches (this SAN is called a disk SAN)
Disk storage other than archive disk cache *Size [TB]*	Locally attached (via SCSI) *ca. 1 in total*	Disk space within SAN attached Sun/HDS 9985 *ca 10*
Archive disk cache *Size [TB]*	Sun T3 (locally attached via FC) *ca. 1.5*	Disk space within Sun/HDS 9985 *ca. 12*
Fast access archive drives / media *Size [TB]*	Sony MOD (attached via SCSI) *ca. 0.7*	Disk space within Sun/HDS 9985 *1*
Tape storage		
Connection of tape drives to archive server	Locally attached at archive server via SCSI	Via FC based SAN in redundant configuration with two + two Brocade 4100 SAN switches (this SAN is called tape SAN)
Robot library, 1st copy *Number of. slots*	Quantum AML-2 *ca. 10,000*	Quantum AML-2 (equipped for new media) *11.000*
Tape drives / media, 1st copy *Capacity of one medium (GB)* *speed (MB/s)*	DLT4000/7000, Sony AIT-2 *20 / 35 / 50* *2 / 4 / 6*	Sun/StorageTek 9940B *200* *30*
Robot library, 2nd copy *Number of slots*	No separate 2nd copy library; same as 1st copy robot library *ca. 1,000*	Quantum i2000 *700 (expandable)*
Tape drives / media, 2nd copy *Capacity (GB)* *Speed (MB/s)*	DLT4000, DLT7000 *20 / 35* *2 / 4*	IBM LTO-3 *400* *up to 80*

Table 1 Comparison of Old and New Hardware

Fig. 2 and Fig. 3 show the DIMS hardware configuration in Oberpfaffenhofen before and after renewal. In addition to the hardware (in gray), the picture also shows the DIMS components from Fig. 1 as dashed red boxes to indicate which DIMS component is running on which server. In Fig. 3 two high availability constituents (Sun cluster and high availability server) are marked with orange boundary boxes. Both hardware configurations include examples of processing systems attached to DIMS.

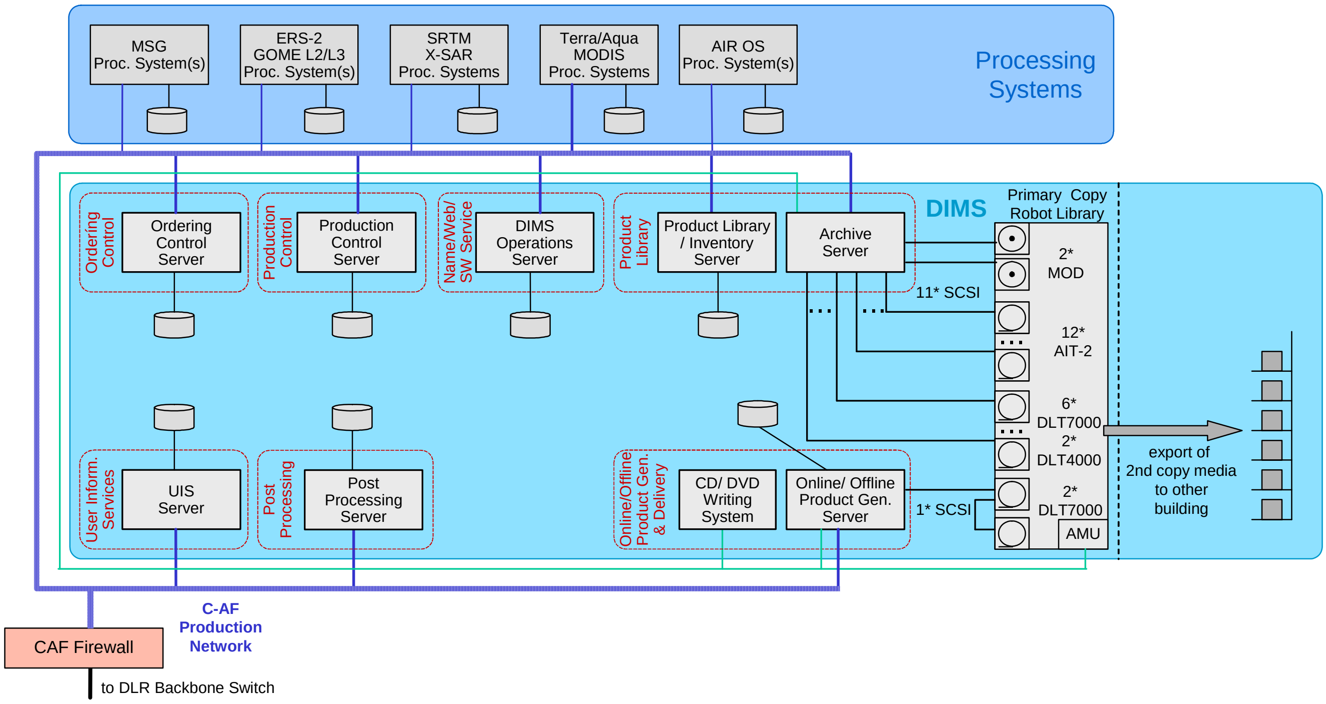

Fig. 2 Old DIMS Hardware in Oberpfaffenhofen

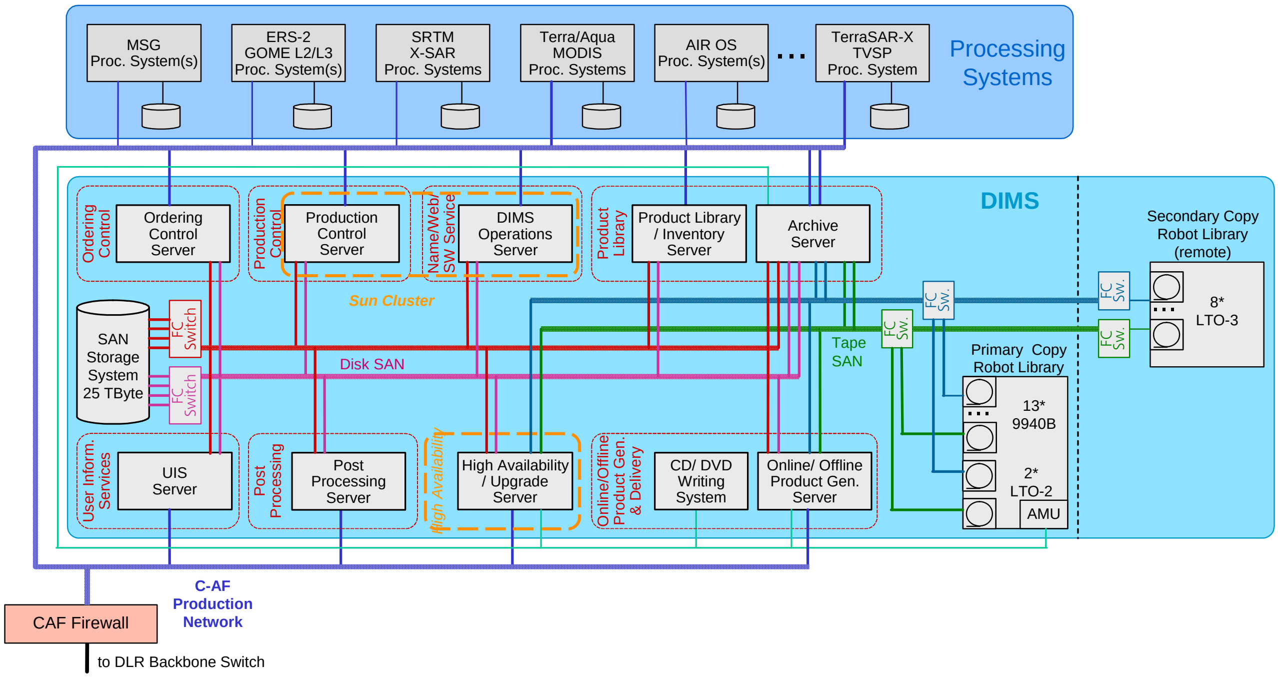

Fig. 3 New DIMS Hardware in Oberpfaffenhofen

Additional Information on New DIMS Hardware

Some important features of the new DIMS hardware which could not be visualized in Table 1 or Fig. 3 are described here:

- Processor throughput is ca. 10 times higher and main memory is ca. 10 times bigger than with the old servers
- The only disks directly attached to the servers are disks for the operating system (Solaris 10); the operating system disk in every server is mirrored with the help of the Sun Volume Manager
- The SAN disk storage system 9985 is fully redundantly connected to all servers with the help of two “disk” SAN switches. Servers with high data throughput like the archive server use more than one link per switch. The disk storage system is connected eight-fold to both disk SAN switches.
- Volumes are built on the disk groups of the 9985 with the help of the Veritas Volume Manager
- The archive server is connected to the 1st copy drives via two SAN switches, whereby half of the tape drives are connected to one switch and the other half to the other switch. If one switch fails, half of the tape drives are still accessible. (indicated in Fig. 3)
- The 2nd copy library with the LTO-3 drives is located in a separate building, some 100m away from building with the 1st copy library. There the drives are connected to a second pair of SAN switches. The connection between the local tape SAN switches and the remote tape SAN switches is established with long range (LX)/ single mode fiber links (also indicated in Fig. 3).

Achievement of the Required New Features

This section describes how the features required of the new DIMS hardware configuration were achieved.

More performance is achieved with new servers with faster CPUs and more main memory, with faster disks in the Sun 9985 disk storage system having high I/O performance, with fiber channel based SAN as interconnection, and with faster tape drives.

More disk storage space is provided by a SAN disk storage system with sufficient capacity (25 TB, expandable to 50 TB in Oberpfaffenhofen, and half that size in Neustrelitz)

More storage space on tapes is achieved by using newer tape / media technology (1st copy 9940B in AML-2: 10,000 slots à 200 GB = 2 PB)

More flexible disk storage is achieved by using a central SAN attached disk system, which allows more or less storage space to be assigned to the one or the other DIMS component.

Higher data safety is achieved by using a remote 2nd copy tape library and also by using enterprise tape technology (Sun STK 9940B). RAID 6 (two parity disks instead of one) is used for disk space except for the archive cache. (For the archive cache RAID 5 is considered sufficient, since shortly after data are written to disk, the two copies are produced on tape.)

Several measures contribute to the **higher availability** of the new system compared to the previous one. The first contribution is the use of a SAN (Storage Area Network) both for disk storage and tape drives. Every server is connected to the disk storage system via two independent paths: every server has two independent HBAs (Host Bus Adapters) whereby each one is connected to one of the two disk SAN switches; the disk storage system itself is also connected from two different controllers to the two disk SAN switches (eight-fold each). This results in two completely independent paths between every server and the storage system. So if a HBA of a server, or a disk SAN switch, or a port of a controller in the disk storage system fails, the storage partitions are still visible for the server, the performance of course being reduced, since normally load balancing is active between the two independent paths. The case is similar for the connection of the tape drives to the archive server: if a HBA of the archive server fails, all tape drives are still accessible (with reduced performance); if a tape SAN switch fails, half of the tape drives are still accessible.

The disk storage system 9985 has high availability through the use of internally redundant paths. For data safety we use RAID 6, i.e., two parity disks in a RAID set. (For the archive cache we only use RAID 5, since the copies to tape are made very soon after data have been written to disk). Furthermore, there are a number of hot spare disks.

The third measure is the use of Sun cluster software in the most sensitive area: the vital functions of DNS master/slave and CORBA name service do not run on the DIMS operation server but on a Sun cluster built with the DIMS operation server and the DIMS production control server. If the DIMS operation server fails, the DNS master, or slave, and the CORBA name service are automatically started on the DIMS production control server; this happens transparently for the DIMS applications, without operator intervention and within two seconds.

The fourth measure to ensure high availability is the use of a so called high availability (HA) server: This is a server which is always up and running in terms of hardware and operating system (OS) but not executing any DIMS application under normal conditions. The HA server is powerful enough to take over any DIMS component. (Some performance reduction is tolerated if it takes over the archiving function). If any application server fails or has to be upgraded with respect to operating system or application software, the corresponding DIMS component can be activated on the HA server. This does not happen automatically but is done by an operator using a written and tested procedure. Each DIMS component can be activated on the HA server within 15 to 20 minutes; for the archive server it takes a little longer: 30 to 40 minutes. These out-of-service times are absolutely compliant with the requirement of 98% availability relative to one month. An important fact is that absolutely no hardware change is necessary to activate the DIMS component on the high availability server, even if the archiving component is moved to it. Another absolutely necessary prerequisite is that the disk groups / partitions on the disk storage system are not dedicated to servers but to DIMS services. For example, there is a server normally running “ordering control” (ordering control server) which goes by the name of “Pluto.” A disk group is not dedicated to the “Pluto” server but to the DIMS component “ordering control,” which under normal conditions is connected to “Pluto.” If “Pluto” fails or its operating system has to be upgraded, the HA server can attract the “ordering control” disk group and after having been started on the HA server the DIMS ordering control component can continue its work exactly where it stopped on “Pluto.” A further prerequisite is that all important states of a DIMS component are stored in a transactional database (PSEpro) which is directly attached to every component and which makes the important states persistent, i.e., the states survive stop/crash of the component.

THE MIGRATION FROM THE OLD TO THE NEW HARDWARE PLATFORM

It was quite a challenge to migrate the DIMS components from the old hardware platform to the new hardware platform without interrupting DIMS operations for more than two or three days. Before the migration itself, several preparatory steps were taken while the DIMS software was still in full operation on the old hardware platform:

1. The 1st preparatory step was made in the old DIMS environment: the Informix database on the inventory server and all DIMS application software were upgraded to exactly the releases which we wanted to use on the new hardware platform. This was to avoid an application software switch together with the hardware platform switch. Not upgraded, however, was the operating system on the old servers: we stayed with Solaris 9, while planning to use Solaris 10 on the new platform.
2. The 2nd preparatory step was to completely set up the new hardware. (One exception was that in Oberpfaffenhofen only two new 1st copy drives were installed in the 1st copy library due to lack of physical space within the AML-2).
3. The 3rd preparatory step was to configure all network connections, disk partitions, access paths to disk partitions, access paths to tape drives and tape libraries, and to set up and configure the Sun cluster. Then we tested everything on the operating system level, i.e., without real applications: visibility of all disk partitions, writing to and reading from the disk partitions, including performance; visibility of tape drives, writing to and reading from tape drives, including

performance; handling of media within the 1st copy and the 2nd copy tape library. Using simple applications we checked the functionality of switching over to the 2nd server within the Sun cluster and back to the original server. We also tested whether disk partitions from all other servers and tape drives from the archive servers can be made accessible on the HA server, taking into consideration the two possible cases: original server still being operational and original server having crashed.

4. In the 4th preparatory step we installed the COTS software, among others the “Informix” database on the inventory server and the hierarchical storage management software “SAM-FS” on the archive server. Again, we tested the switchover from original server to HA server, especially for the inventory and archive servers.

5. The 5th preparatory step was to install the DIMS application software on all servers and to perform some tests with the application software. We also wrote down and tested the procedures for transferring an application from its original server to the HA server and later back to the original server.

After these preparatory steps came the big moment: no, not one single big moment, but rather a smooth procedure. Service by service we deactivated the DIMS services on the old hardware platform, copied all necessary data to the corresponding service partition on the new disk storage system, adjusted the application configuration to use new directory paths, and put the service into operation on the new hardware platform. To reduce the number of out-of service days, we selected two or three days when DIMS was allowed to be partially or fully out of operation and on each of these days moved one or two DIMS components. One prerequisite for the easy move was the DIMS name service, which tells all other DIMS services on which hardware server a service is running. The content of the name service is automatically updated whenever a DIMS service is started: it tells the name service: “Today ‘service abc’ is running on IP-address w.x.y.z.

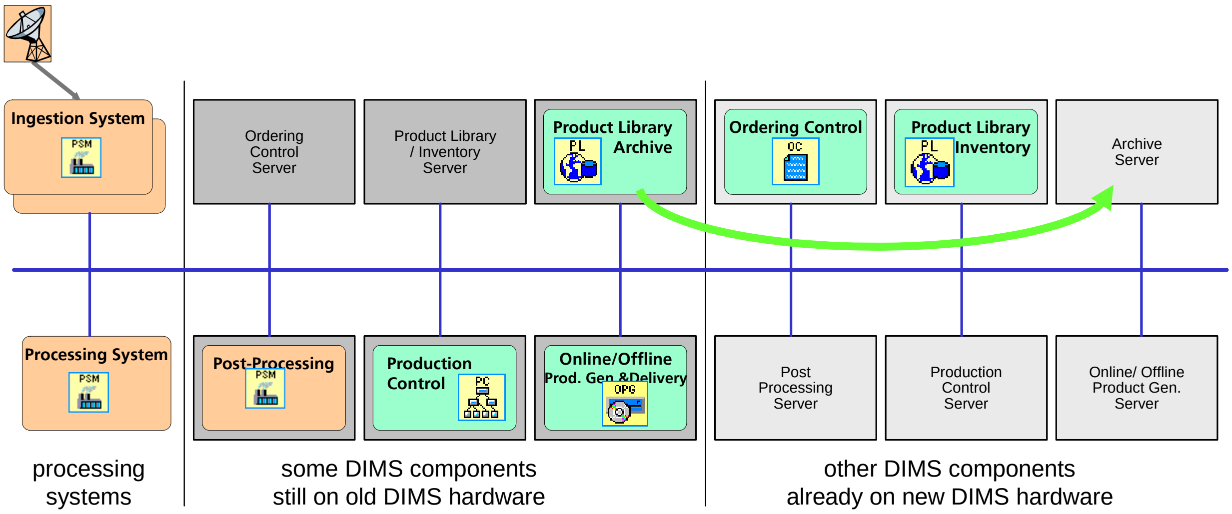

Fig. 4 The Migration of the DIMS Components (Application Software) from the Old to the New DIMS Hardware

The biggest transfer effort of course was the migration of the archive server. To migrate the archive application to the new server we had to request an out of operation period of three consecutive days. We had to physically de-install about half of the old 1st copy drives (AIT-2) and 2nd copy drives (DLT4000 and DLT7000), physically install the new 1st copy drives (9940B), perform a complete reconfiguration of the AML-2 library, connect the new 9940B drives to the tape SAN, configure the tape SAN and assure the visibility of the 9940B drives both on the archive server and on the HA server, connect the old AIT-2 drives, DLT4000 and DLT7000 drives which were kept in the AML-2 to the new archive server via SCSI interfaces and assure their visibility, transfer the SAM-FS media catalogs, dump the SAM-FS directory and i-node information of all SAM-FS file systems on the old archive server and restore it to the new SAM-FS file systems on the 9985, bring SAM-FS into operation on the new archive server, and finally activate the application software. After three hard days the archive service was operational on the new hardware.

The transfer of all DIMS services / applications to the new hardware platform was completed beginning of Oct. 2006. Thereafter, all new data were written to the new 1st and 2nd copy media, whereas the data written before that date could be read from the old media. That's why about half of the old drives were connected to the new archive server; however, they were used for reading only.

THE MIGRATION OF DATA FROM OLD MEDIA TO NEW MEDIA

Of course the final aim included getting rid of the old archive media and drives. To achieve this, all old data (about 120 TByte in Oberpfaffenhofen and about 50 TByte in Neustrelitz) had to be migrated to the new 1st and 2nd copy media. Of course they always had to stay accessible during migration of the data. Thanks to SAM-FS this was not a problem at all. The re-archiving capability which comes with SAM-FS is suitable for data migration, but is not very fast. Therefore, we used the Fast Media Refresh Toolkit (FMRT) of "HMK Computer Technologies" on top of SAM-FS. By parallelizing reading from old tapes and writing to new tapes, and by doing this in several parallel streams, migration is much faster than with SAM-FS alone. The FMRT also avoids having current data displaced from the archive disk cache.

The migration to new media has been completed both in Oberpfaffenhofen and Neustrelitz, taking about 3 months in Neustrelitz and about five months in Oberpfaffenhofen. After some additional checks the old drives will be disconnected from the new archive server and removed from the tape library. The old media will also be removed from the tape library and disposed in an orderly manner.

This will then be the end of the current migration of the DIMS components to the new hardware platform and of the data to new media. However, it's not the end of the migration story.

CONCLUSION

In a four year process comprising system design, fund raising, procurement, system set up, system migration and data migration we managed to migrate the system to new hardware and the data to new media.

The same reasons which made us renew the old hardware platform and migrate the data will again be valid four or five years from now. One has to be aware that the only way to preserve remote sensing data for a long time and be able to support upcoming missions is to again and again renew all the components of a data archive and management system like DIMS: hardware, media, COTS software (new versions) and application software. Due to the great speed with which new hardware and software versions come to market and old hardware and software reach the end of their service life, hardware renewal has to occur every five to seven years, maximum (software versions are upgraded about once per year). Taking into consideration the time needed to raise money, select new hardware, set up and evaluate the bidding, procure and install the new hardware, and finally perform the system and data migration a new cycle starts short after the last one has ended with the completion of data migration.

Long Term Archiving is a continuous sequence of system and data migrations.

Our next steps are already visible: it will be the renewal of the 1st copy library which has the end of service life announced for autumn 2009, most probably combined with the exchange of the 1st copy tape drives. This is also a prerequisite to be able to cope with the huge amounts of data of the TanDEM-X mission, which is expected to produce ca 1.8 PByte of data in the years 2009 to 2012, resulting in a total amount of data of ca. 2.4 PB in 2012 in the Oberpfaffenhofen DIMS archive. Fig. 5 shows the growth of the amount of data stored in the Oberpfaffenhofen archive from the past (1995) to the near future (2012).

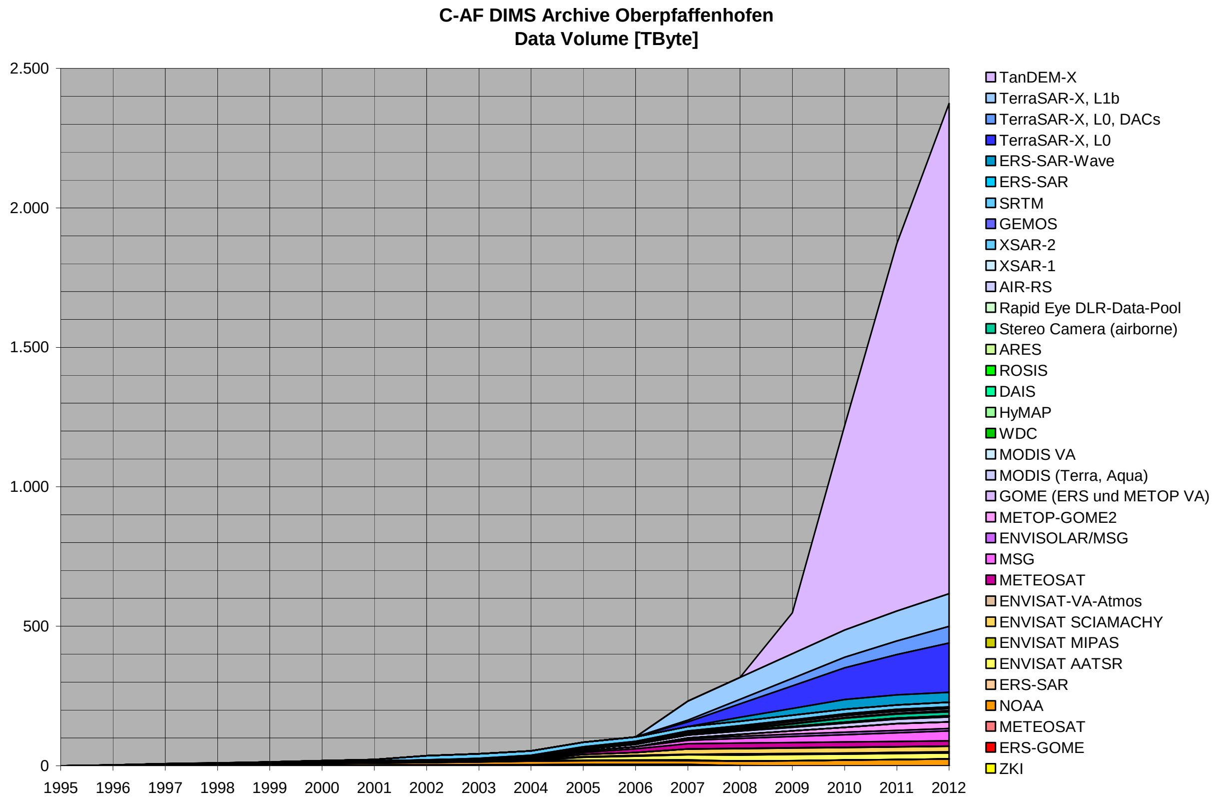

Fig. 5 Data in the DIMS Oberpfaffenhofen Archive from 1995 to 2012

REFERENCES

[1] - Mikusch, E., Diedrich, E., Göhmann, M., Kiemle, S., Reck, C., M., Reißig, R., Wildegger, W., Wolfmüller, M.: Data Information and Management System for Earth Observation Products based on CORBA and Java. Earth Observation (EO) & Geo-Spatial (GEO) Web and Internet Workshop 2000, London, 17.-19.04.2000, Commitee on Earth Observation Satellites, (2000).

[2] - Mikusch, E., Wolfmüller, M., Diedrich, E., Dietrich, D., Böttcher, M., Management of "Future Products" in the Distributed Payload Ground Segment for TerraSAR-X. The DASIA 2003 DAta Systems In Aerospace Conference, Prague, 2.-6.6.2003, EUROSPACE, ESA, SP-532, (2003)

[3] - Seifert, P., Kiemle, S., Mikusch, E., Wildegger, W., Böttcher, M., Wolf, H.-J., Maass, H., Missling, K.-D., Infrastructures to Improve User Access to EO Data, Ensuring the Long Term Preservation and Adding Value to the Scientific and Technical Data, PV 2004, Frascati, 5.-7.10.2004, ESRIN, ESA, WPP-232, pp. 127-134, (2004)

[4] - Kiemle, S., Schroeder-Lanz, A.-K., Reck, C., Object Query Language – Enabling Service for Earth Observation Product Processing, Access and Dissemination, Ensuring the Long Term Preservation and Adding Value to the Scientific and Technical Data, PV 2004, Frascati, 5.-7.10.2004, ESRIN, ESA, WPP-232, pp. 17-24, (2004)

[5] - Pinna, G-P., Mikusch, E., Bollner, M., Puin, B., Earth Observation Payload Data Long Term Archiving - the ESA's Mulit-Mission Facility Infrastructure. Ensuring Long-term Preservation and Adding Value to Scientific and Technical Data, PV 2005, Edinburgh UK, 21.-23.11.2005, Digital Curation Center Conference Proceedings, The Royal Society Edinburgh, (2005)

[6] - S. Kiemle, E. Mikusch, C. Bilinski, B. Buckl, D. Dietrich, S. Kröger, C. Reck, A.-K. Schroeder-Lanz, M. Wolfmüller, Data Information and Management System for the DFD Multi-Mission Earth Observation Data. Ensuring Long-term Preservation and Adding Value to Scientific and Technical Data, PV 2005, Edinburgh UK, 21.-23.11.2005, Digital Curation Center Conference Proceedings, The Royal Society Edinburgh, (2005)