

Langzeitarchivierung von 1988 bis jetzt am RZG

Hartmut Reuter
reuter@rzg.mpg.de

- Komponenten der Langzeitarchivierung
- Technologiewechsel
- Langzeit-Archive am RZG
- Empfehlungen

Langzeitarchivierung am RZG

- Keine Erfahrung mit
 - Portalen
 - Metadaten-Datenbanken
 - geeigneten Formaten für digitalisierte Dokumente
- RZG stellt nur ein einigermaßen sicheres Filesystem zur Verfügung
 - anwendungsunabhängig
 - beliebig erweiterbar (skalierend)
 - weltweit sichtbar
 - Technologiewechsel für Benutzer unsichtbar
- Benutzer sind selbst verantwortlich für
 - ihre Metadaten-Datenbanken
 - Konvertierung ihrer Daten auf neue Formate

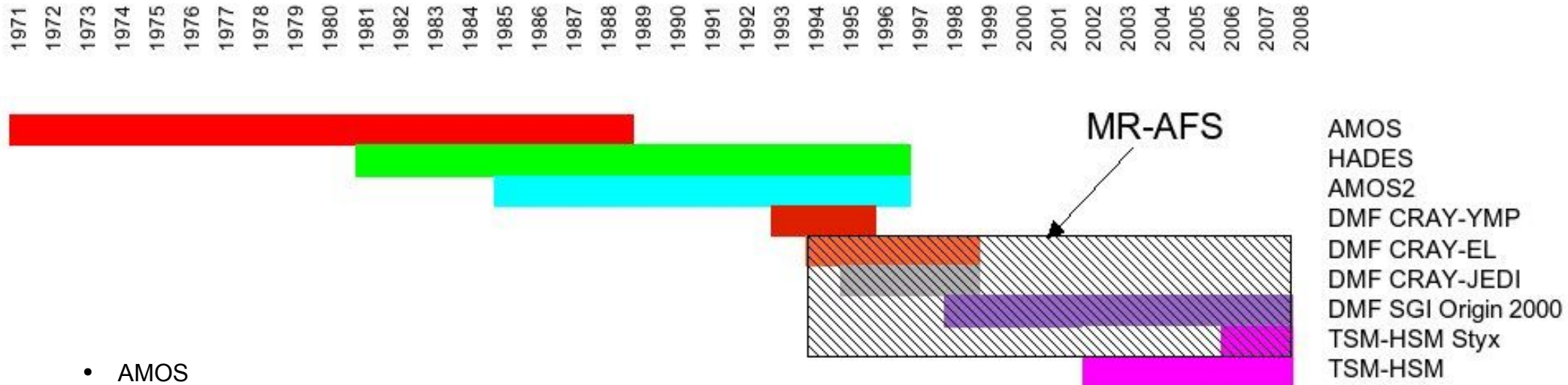
Was ist Langzeitarchivierung?

- heißt länger archivieren als Lebensdauer der Komponenten
 - Archivierungs-Hardware (Band-, Plattensysteme) 5 Jahre
 - HSM-Systeme (Software, Hardware) 10 Jahre
 - Filesysteme mit HSM Funktionen und/oder 20 Jahre
 - Metadaten-Server (Datenbanken) 20 Jahre
 - Dokumentformate (.tif, .pdf ...) ? Jahre
- Ersetzen von Komponenten kann Jahre dauern (Umkopieren)
 - deshalb kurzlebige Komponenten möglichst verstecken:
 - Hardware unter HSM-System
 - HSM-System unter
 - Filesystem mit HSM Funktionalität oder
 - Portal mit Datenbank
 - Gelegentlich müssen leider auch diese ersetzt werden!

Komponenten der Langzeitarchivierung am RZG

- Archivierungs-Hardware (Bänder+Platten)
 - Lebenszyklus < 5 Jahre
 - Wartungskosten, Kapazitätsentwicklung
 - IBM 3480 -> D2 -> D3 -> STK 9840 -> STK 9940 -> STK 9940B
- HSM (Hierarchical Storage Management) Systeme
 - erwartete Lebensdauer ~ 10 Jahre
 - In der Regel Plattform-abhängig
 - HADES -> DMF (Cray) -> DMF (SGI) -> TSM-HSM (IBM)
- Filesystem mit HSM-Funktionalität
 - erwartete Lebensdauer 15 – 20 Jahre
 - AMOS erreichte 18 Jahre, HADES 15 bzw. 21 Jahre
 - MR-AFS am RZG seit 1994 -> auf OpenAFS+OSD 2008 geplant
- Projekt-spezifische Metadaten Server (Datenbanken)
 - erwartete Lebensdauer > 20 Jahre
 - Oracle > 14 Jahre im Einsatz, bisher kein Technologiewechsel

HSM-Systeme am RZG

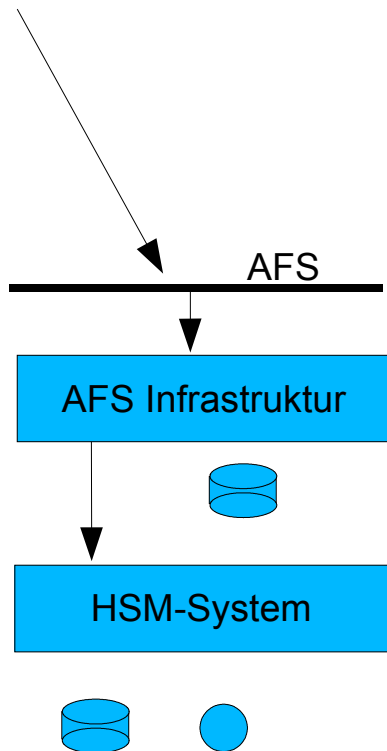


- AMOS
 - Eigenentwicklung des IPP: Betriebssystem für IBM Mainframe
 - Bänder und CDC-Massenspeicher (16 GB Gesamtkapazität.)
 - von 1971 – 1989 im Einsatz
- HADES
 - Weiterentwicklung von AMOS, OS unter VM oder Subsystem für MVS
 - Bandroboter, Netto-Datenvolumen ~ 1 TB
 - von 1981 – 1996 am RZG im Einsatz (In Heidelberg 20 Jahre)
- AMOS2
 - Neuentwicklung des IPP für Experiment ASDEX-Upgrade
 - von 1993 – 1995 im Einsatz
- MR-AFS (Multiple-Resident-AFS)
 - HSM-Erweiterung von AFS-Fileservern (nicht open-source)
 - seit 1994 am RZG im Einsatz
- OpenAFS+OSD
 - Erweiterung von OpenAFS, nutzt Object Storage Technik auch für HSM
 - Soll 2008 MR-AFS ersetzen

Totaler Technologiewechsel 1995

- Altes System HADES auf IBM Mainframe Computern
 - EBCDIC Zeichensatz, Filesystem mit Recordstruktur
- Neues System MR-AFS auf UNIX (damals AIX, SUNOS und UNICOS)
 - ASCII Zeichensatz, Filesystem ohne Recordstruktur (bytestream)
- Benutzer mussten ihre Daten selbst transferieren!
- FTP-Interface von HADES ermöglichte Konversion “on the fly”
 - automatische Zeichensatz-Konversion für Text-Files
 - spezielle Konversion für W7AS-Schussfiles, möglich durch selbstbeschreibende Files mit Text, Integer- und Floatingpoint-Blöcken.

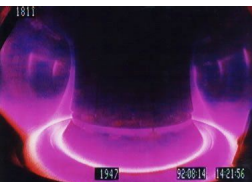
Warum AFS für Langzeitarchivierung



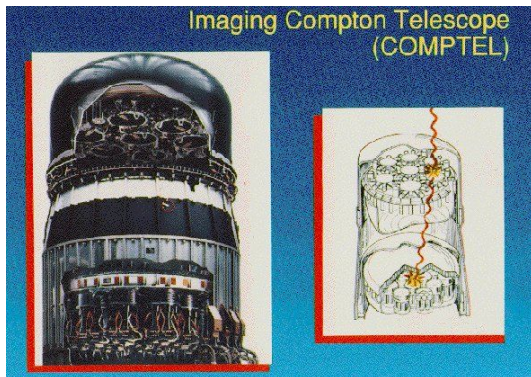
- AFS gibt es auf (fast) allen Plattformen (Linux, Unix, Windows, MacOS)
- Verteiltes weltweites Filesystem
- Ausreichende Sicherheit durch Kerberos Authentifizierung
- Granulare Vergabe von Rechten durch ACLs
- Volle Kontrolle über den Source-Code
 - OpenAFS ist open source
 - MR-AFS wurde vom RZG gepflegt und weiterentwickelt
 - OpenAFS+OSD wird am RZG entwickelt
- Skaliert problemlos:
 - Hinzufügen von Servern im laufenden Betrieb
 - Umverteilung von Daten im laufenden Betrieb
- MR-AFS und AFS+OSD Server bieten HSM-Funktionalität
 - können beliebige HSM-Systeme benutzen
 - Transparenter Zugriff auf ausgelagerte Daten
- **HSM-Technologiewechsel für Benutzer unsichtbar**

Langzeitarchive für IPP Experimente

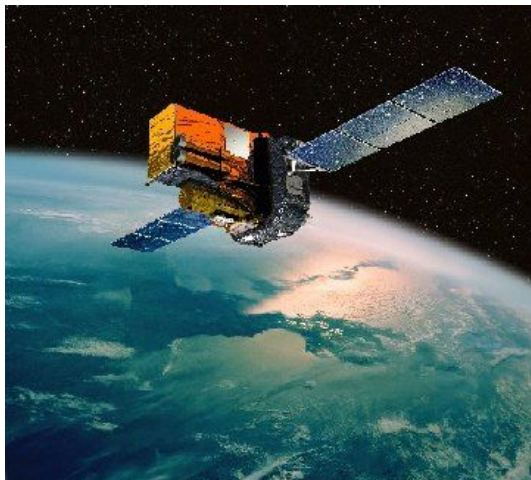
- Experiment Asdex des IPP (~1980-1990)
 - ca. 33500 Schussfiles kleiner 5 MB, insgesamt ~83 GB
 - von Bändern Ende der 90er in MR-AFS kopiert
- Experiment W7AS des IPP (1988-2002)
 - ca. 60000 Schussfiles 1.5 bis 90 MB, insgesamt ~2.2 TB
 - 1988-1995 in VM/CMS auf IBM Mainframes erzeugt und in HADES gespeichert
 - 1995-2002 in VMS oder UNIX(AIX) erzeugt und in MR-AFS gespeichert
- Experiment Asdex Upgrade am IPP (seit 1993)
 - ca. 20000 Schüsse mit jeweils vielen Diagnostikfiles, insges. ~26 TB
 - 1993-1995 in AMOS2 auf IBM Mainframe gespeichert
 - seit 1995 in MR-AFS



Gamma-Ray-Astronomy MPE



- COMPTEL (1991-2000)
 - knapp 2 TB in container files von ~ 100 MB Größe
 - 1991-1995 in HADES gespeichert, dann -> MR-AFS
 - 1995-2000 in MR-AFS, jetzt alles auf Bändern
- EGRET Energetic Gamma Ray Experiment (1991-1996)
 - im AFS ~ 56 GB archiviert
- INTEGRAL (2001-2010 oder länger)
 - bisher >5 TB in kleinen Files in AFS, alles auf Platten



Magic Projekt (MPP u.a.)



- Gamma-Strahlen-Teleskop auf Kanarischen Inseln
- seit 2003 > 53 TB in MR-AFS

Langzeitarchive geisteswissenschaftlicher Max-Planck-Institute

- Videos und Tondokumente des MPI für Psycholinguistik in Nimwegen
 - seit 2004, bisher ca. 1.1 TB
 - in Nimwegen direkt ins AFS kopiert
- Fototek der Biblioteca Hertziana in Rom
 - seit 2004, bisher ca. 800 GB
 - bisher von verschickten USB-Platten am RZG ins AFS kopiert
 - jetzt in Rom direkt ins AFS
- Fototek des Kunsthistorischen Instituts Florenz
 - seit 2004, bisher ca. 3.3 TB
 - bisher von verschickten USB-Platten am RZG ins AFS kopiert
 - in Zukunft von der GWDG direkt ins AFS

Empfehlungen (1)

- **Redundanz:** Je mehr Kopien eines File existieren, desto weniger wahrscheinlich sein Totalverlust
 - möglichst mehrere unabhängige Kopien an verschiedenen Orten
 - Beispiel: KHI und MPL kopieren Files zur GWDG **und** zum RZG
 - Kopien innerhalb einer Organisation (auch als Backup): AFS readonly Volumes mit automatischem nächtlichen Abgleich.
 - Mehrere Bandkopien, möglichst auch räumlich getrennt.
- **Trotzdem kommen Fehler vor**
 - Automatismus, der zweite Bandkopie erzeugt, hatte versagt.
 - Band ging kaputt.
 - Band wurde ans Labor eingeschickt, um die Daten zu retten.
 - Labor wurde wegen Umstrukturierung der Firma (STK->SUN) aufgelöst.
 - Band nicht mehr auffindbar -> **Datenverlust**

Empfehlungen (2)

- **Kontrollsummen:** (checksums) ermöglichen Verifizierung des Inhalts.
 - KHI liefert zu jedem File md5sum in getrenntem File.
 - RZG baut Datenbank mit md5-Checksums für alle Files auf Band auf.
 - AFS+Object Storage als Nachfolgesystem für MR-AFS wird das im Filesystem integriert haben.
- **Selbstbeschreibende Files:** Wenn möglich,
 - innere Struktur der Files im File selbst beschreiben
 - dadurch auch automatische Versionskontrolle möglich.
 - Metadaten, die in der Datenbank stehen, auch im File selbst halten.

Danke für Ihre Aufmerksamkeit