Digitale Langzeitarchivierung wissenschaftlicher Daten bei DESY.

Quelle: http://www.tagesschau.de/inland/kabinett-vorratsdatenspeicherung-101.html

Steffen Bohr (ZEU-DV)

Technisches Seminar

Zeuthen, 03. November 2015

HELMHOLTZ GEMEINSCHAFT

DESY

Agenda

- Digitale Langzeitarchivierung
- Archivierung wissenschaftlicher Daten bei DESY
- Open Storage Manager (OSM)
- Fazit

Digitale Langzeitarchivierung.

- Grundlagen der digitalen Langzeitarchivierung
- Anforderungen
- Technische Grundlagen
- Herausforderungen

Grundlagen der digitalen Langzeitarchivierung

Übertragung klassischer Archivierung auf digitales Zeitalter

- Archivierung digitaler Datensätze auf sicheren Medien
- Digitale Medien leichter vergänglich als analoge Medien
- Archivierung von Medien und Technologie
- Zeiträume von $x \geq 10$ Jahren (Plan $x \geq 100$ Jahre)

Sicherstellung der guten wissenschaftlichen Praxis

- Reproduzier- und Validierbarkeit der Daten
- Gewinnung neuen Wissens aus alten Datenbeständen
- Als Grundlage für Lehre oder neue Experimente

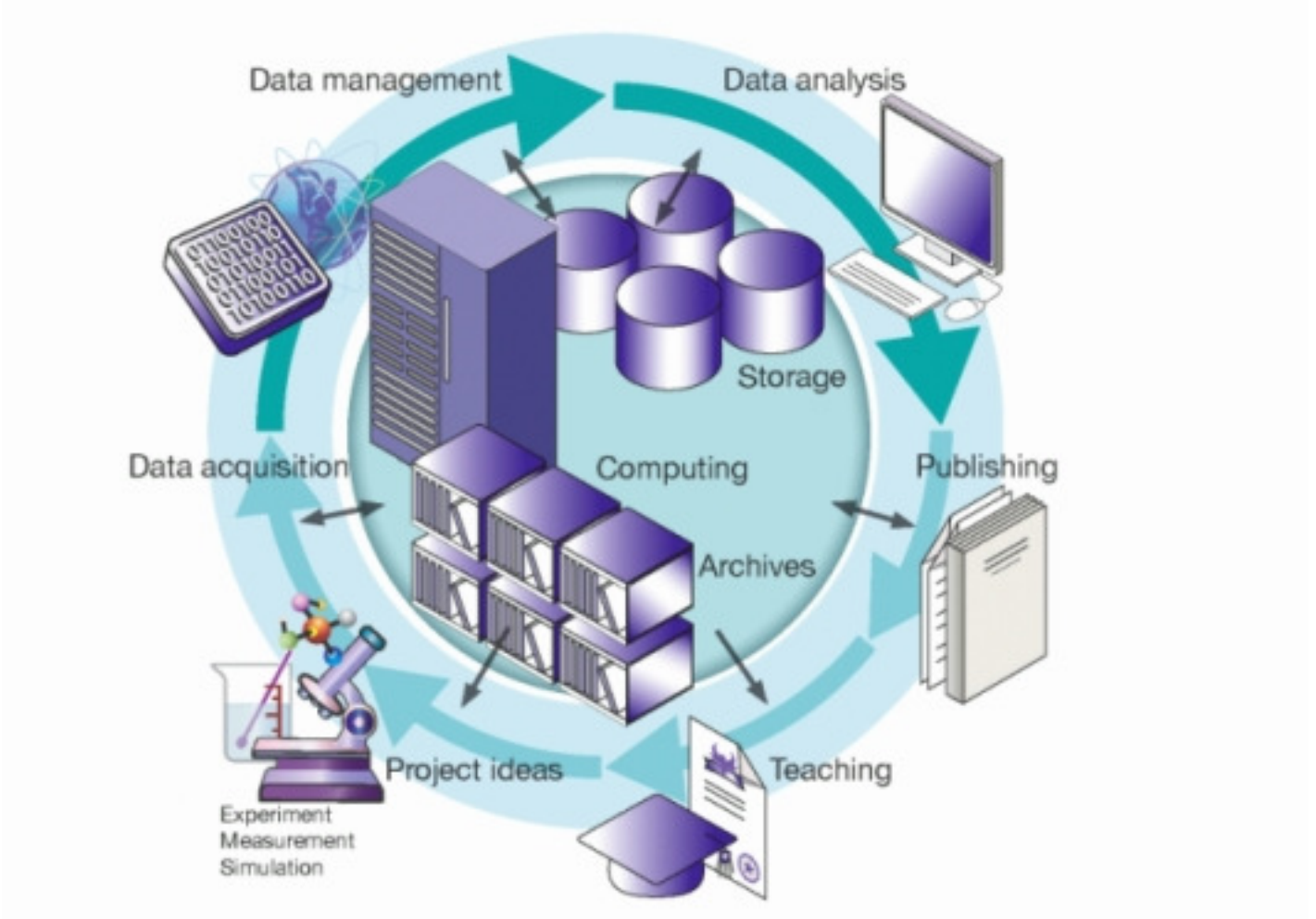

Abbildung: Information Lifecycle [1]

[1] Optimization of data life cycles · C. Jung, ...· Journal of Physics · DOI: 10.1088/1742-6596/513/3/032047

Anforderungen

Empfehlungen der Deutschen Forschungsgemeinschaft (DFG) [2]:

> *Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für **zehn Jahre** aufbewahrt werden.*

Bitkom Leitfaden zur GoBD konformen Archivierung (Auszug) [3]:

1. *Elektronische Archivierung ist technologieneutral*
2. *Die Archivierung [...] hat zeitnah zu erfolgen*
3. *Elektronische Archivierung muss Unveränderbarkeit sicherstellen*
4. *Archivierte Objekte müssen mit einem Index versehen werden*
5. *Elektronisch archivierte Objekte müssen les- und auswertbar bleiben*

[2] Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten · DFG · Januar 2009
[3] Elektronische Archivierung und GoBD · Bitkom · October 2015 · https://www.bitkom.org/Publikationen/...

Technische Grundlagen

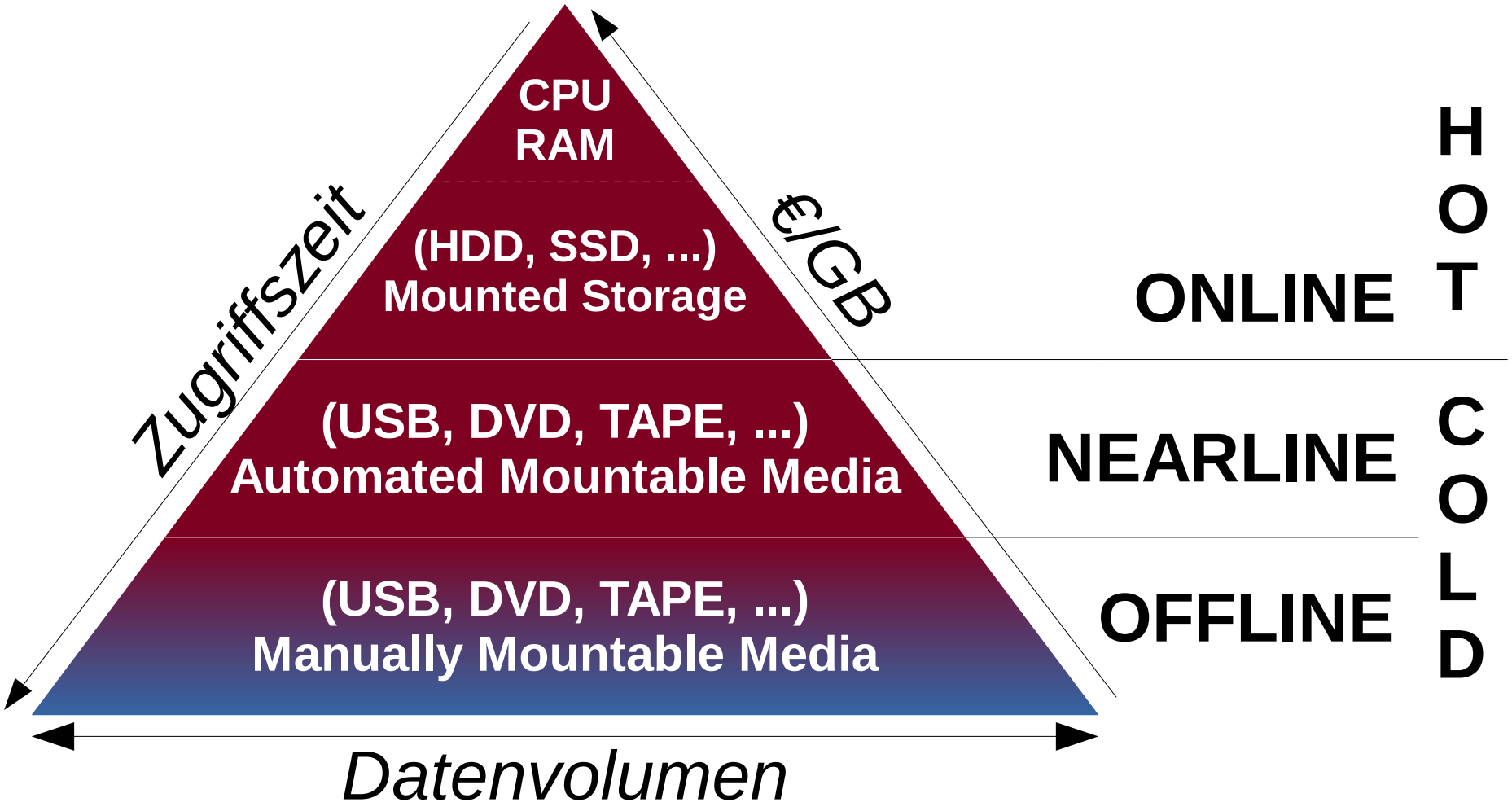

Abbildung: Speicherhierarchie

	Festplatten	Optische Medien	Magnetbänder
Beispiel	Seagate Archive	BD-R XL	LTO6
Kapazität (bis)	$8\,TB$	$100\,GB$	$2,5\,TB$
Preis/Einheit	$\approx 630€$	$\approx 15€$	$\approx 25€$
Preis/TB	$79€$	$150€$	$10€$
Datarate	12Gb/s SAS	6Gb/s SATA	8Gb/s FC
BitErrorRate	$1x10^{-15}$	$1x10^{-6}$	$1x10^{-17}$

Tabelle: Auswahl möglicher Archivmedien

- Managementsoftware wird unabhängig vom Medium benötigt
- Auswahl des Mediums vorwiegend ökonomisch basiert
- Break-Even Point für effektive Festplattennutzung überschritten
- Magnetbänder bei DESY Mittel der Wahl für Langzeitarchivierung

Herausforderungen

Technologieübergang und mechanische Abnutzung

Zyklisch neue Medien- und Laufwerksgenerationen

- Alle ≈ 3 Jahre neue Mediengeneration und Laufwerke
- Erhöhte Kapazität und Datenraten, neue Features
- Migration von alten Generationen auf neue Medien

Medien, Laufwerke und Roboter verschleißen

- Alle Operationen führen zu mechanischer Abnutzung
- Lesefehler oder kompletter Datenverlust möglich
- Wechsel auf neue Medien, Generationen, Technologien

Wachsender Datenbestand führt zu Continuous Migration

Gewährleistung der Datenpfadkonsistenz

- Sicherstellung der Korrektheit von Pool bis Medium
- Fehlerquellen steigen proportional mit Anzahl der Instanzen
- Vergleichsweise einfach zu gewährleisten

  Softwareseitig $\Rightarrow$ Prüfsummen je Instanz (CRC, MD5, ...)
  Hardwareseitig $\Rightarrow$ integrierte Prüfsummen (T10/DIF, SHA-3, ...)

Gewährleistung der Datenintegrität auf dem Medium

- Sicherstellung der Integrität auf dem Medium
- Abnutzung und Einlagerung kann Fehler verursachen
- Gewährleistung der Integrität ohne Abnutzung nicht möglich

Archivierung wissenschaftlicher Daten bei DESY.

- Verantwortliche Gruppen und Projekte
- Anforderungen an ein Archivsystem
- Architektur des Archivsystems
- Realisierung der Datenarchivierung
- Archivnutzung

Verantwortliche Gruppen und Projekte

IT und DV Gruppen in Hamburg und Zeuthen

- Konzeptionierung und Realisierung von Archivsystemen
- Betrieb und Administration laufender Systeme

Data Preservation High Energy Physics (DPHEP)

- Konzeption und Standardisierung von Archivlösungen
- Vermutlich beendet (?)

Large-Scale Data Management and Analysis (LSDMA)

- Etablierung von Data Lifecycle Labs (DLCLs)
- Konzeption und Entwicklung angepasster Archivlösungen
- Endet 2015 $\Rightarrow$ vermutlich Nachfolgeprojekt

Anforderungen an ein Archivsystem

Allgemeine Anforderungen

- Offene Daten-, Archiv- und Medienformate
- Archivierung unabhängig von Supportzyklen
- Trennung von Kontroll- und direktem Datenpfad
- Automatisiertes Medien- und Devicemanagement

Warum keine proprietären Systeme anschaffen?

- Geschlossene Daten-, Archiv- und Medienformate
- Abhängigkeit von Herstellern und Supportzyklen
- Kontrolle und Zugriff durch externe Hersteller

Public Domain Software erfüllt Anforderungen nicht

- DESY/HEP benötigt spezielle Lösung(en)

Architektur des Archivsystems

Nutzung einer hierarchischen Speicherarchitektur bei DESY

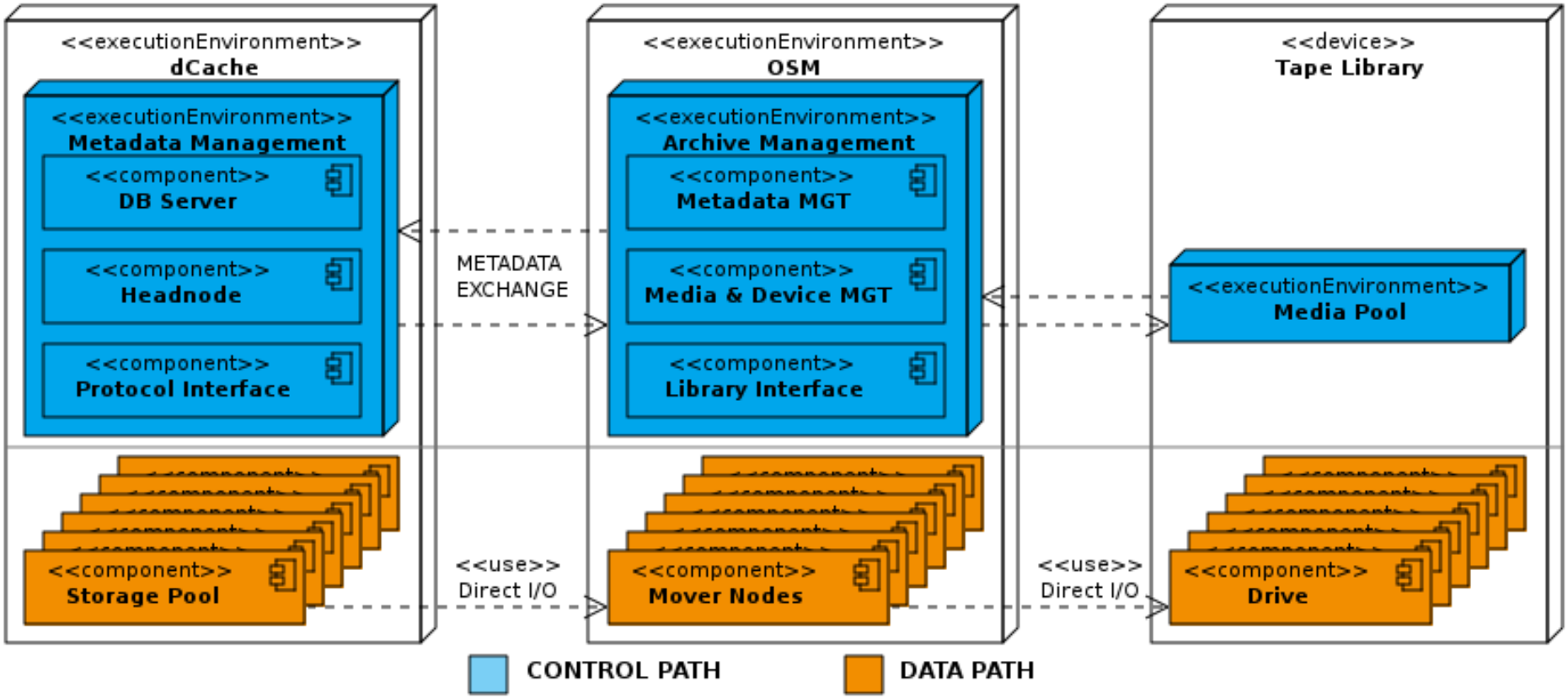

Abbildung: DESY Storage Infrastruktur

Realisierung der Datenarchivierung

Laufende Systeme zur Archivierung

- Open Storage Manager (OSM)
  - Archivierung von Rohdaten (IceCube, HERA, ...)
  - Analyse- und Simulationsdaten
- Tivoli Storage Manager (TSM) Archiver Komponente
  - Archivierung von Nutzerdaten und (ausgelaufenen) Accounts
  - Computer Aided Design (CAD) Projektdaten

Archivierung

- Automatische Archivierung aller aktuellen Datensätze
- Archivierung transparent vor dem Benutzer
- Daten können zusätzlich online vorgehalten werden

Konstantes Wachstum über die letzten Jahre

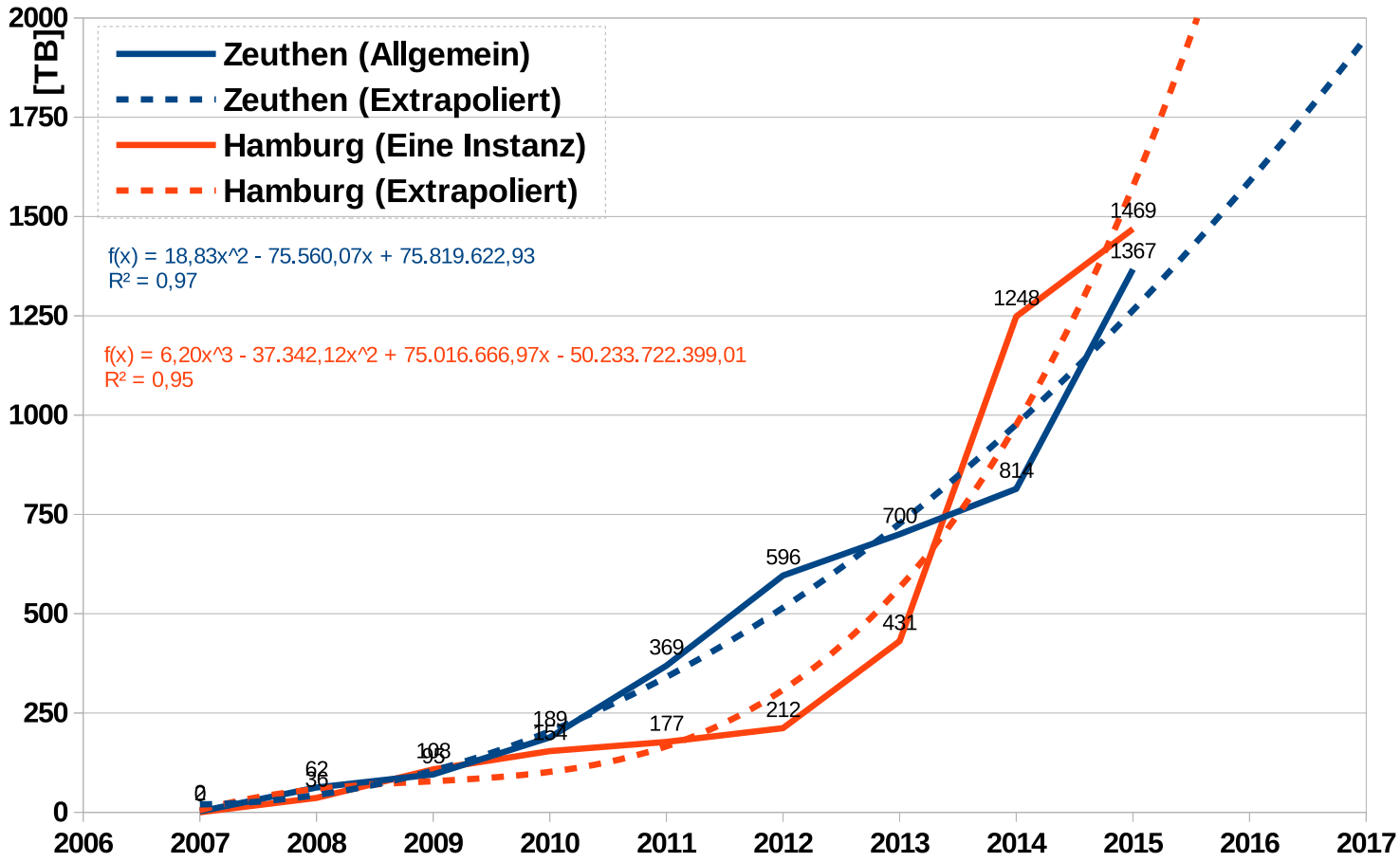

Abbildung: Datenbestände in dCache [4]

[4] dCache Billing Statistics · Stand: 2015-10-27 · ZN: puck · HH: dcache-se

Archivnutzung

Automatisierte Archivierung in Tape Libraries

- Hamburg: 2 Roboter mit 26 Laufwerken und 20.200 Slots
- Zeuthen: 1 Roboter mit 19 Laufwerken und 4.900 Slots

Aktueller Datenbestand auf Tape: $\approx 10,3$ PetaByte

- Hamburg $\approx 8,2$ PB auf 19.264 Medien
- Zeuthen $\approx 2,1$ PB auf 3.346 Medien

Archivierung auf Tapes unterschiedlicher Generationen

- Neue Daten ausschließlich auf LTO6 (bisher ≈ 200 TB in Zeuthen)
- Altbestände und Sicherung auf älteren Generationen ($\leq$ LTO4)
- Migration alter Bestände auf neue Tapes sehr zeitaufwendig

Statistisch 30x mehr Schreib- als Lesezugriffe

Open Storage Manager (OSM).

- Ausgangssituation
- Ausgangslage und Aufgabenstellung
- Entwicklungsarbeiten und Migration
- Mögliche zukünftige Ziele und Meilensteine

Ausgangssituation

Entwickelt von Lachman Technology Inc.

- Original Codebasis aus den 1980er Jahren (K&R C)
- Integrierte Partitionierung und Medienmanagement
- Proprietäre Embedded Objektdatenbank (Raima)
- Unterstützte Systeme: SunOS/Solaris, AIX, IRIX, HP-UX

Von DESY als Source-Code Lizenz gekauft

- Erlaubt individuelle Anpassungen an DESY Anforderungen
- Verbessertes Scheduling, unterstützte Medien, Limits, ...
- ≈ 300.000 Zeilen Code in ≈ 3.700 Dateien

Ausgangssituation

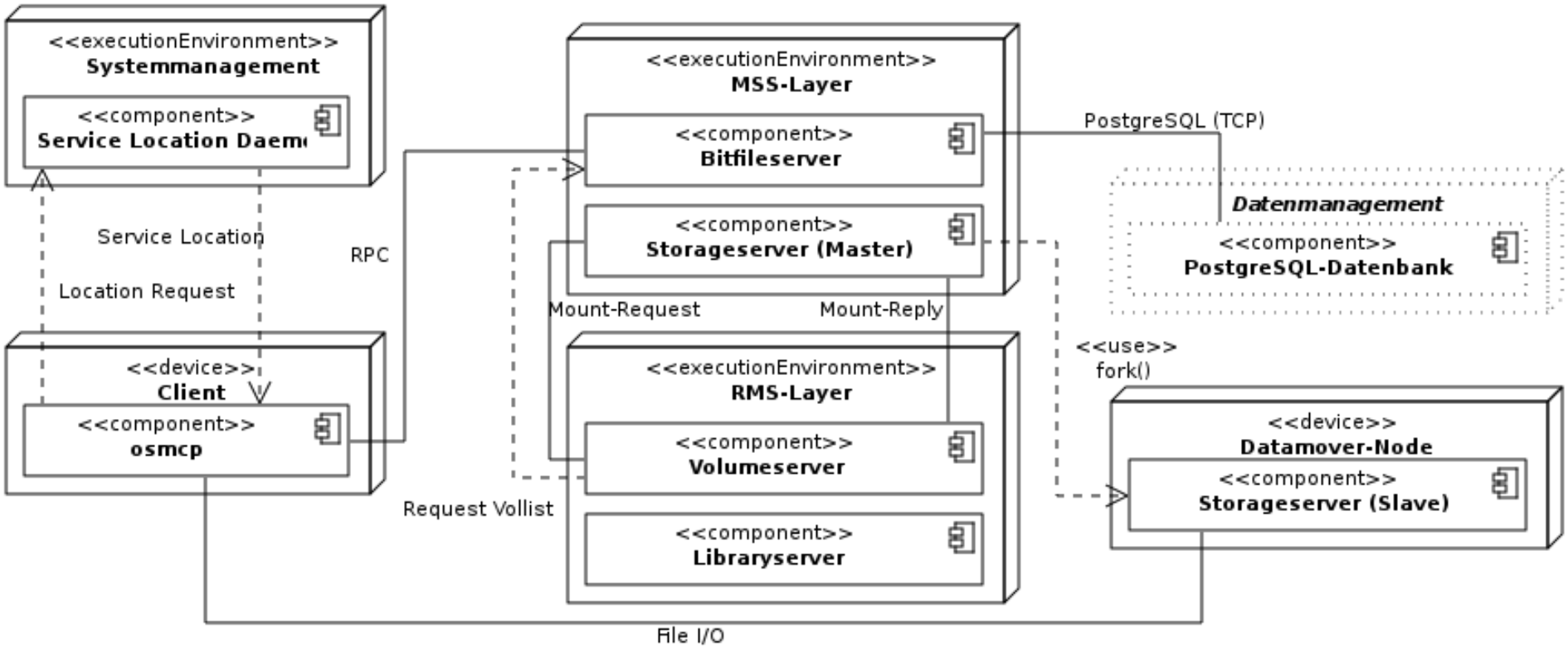

Abbildung: Softwarearchitektur OSM

Ausgangslage und Aufgabenstellung

Probleme und Limitierungen im Originalsystem

- Überalterte Codebasis aus den 1980er Jahren
- Unterstützte ausschließlich SunOS/Solaris
- Maximale Dateigröße 2GB $\Rightarrow$ nach Anpassungen 1TB
- Maximale Anzahl der Requests auf 1024 limitiert
- Maximale Größe der Datenbank 2GB

Aufgabenstellung

- Portierung von Solaris x86 auf GNU/Linux x64
- Austausch von Raima gegen PostgreSQL Datenbank
- Anpassung der administrativen Anwendungen

Entwicklungsarbeiten und Migration

Anpassungen am System

- Entwicklung eines neuen Datenbankmodells
- Verbesserte Fehlerbehandlung und Bookkeeping
- Bereinigung und Modernisierung der Codebasis
- Austausch des Kommunikationssystems, Polling, ...
- Entwicklung zusätzlicher Tools und Dienste
- Aktuell: ≈ 253.000 Zeilen Code in ≈ 3.600 Dateien

Migration auf neues OSMng

- Produktivinstanz in Zeuthen seit 15.04.2015
- Migration in Hamburg in weiten Teilen abgeschlossen

Entwicklungsarbeiten und Migration

Datenbankmodell

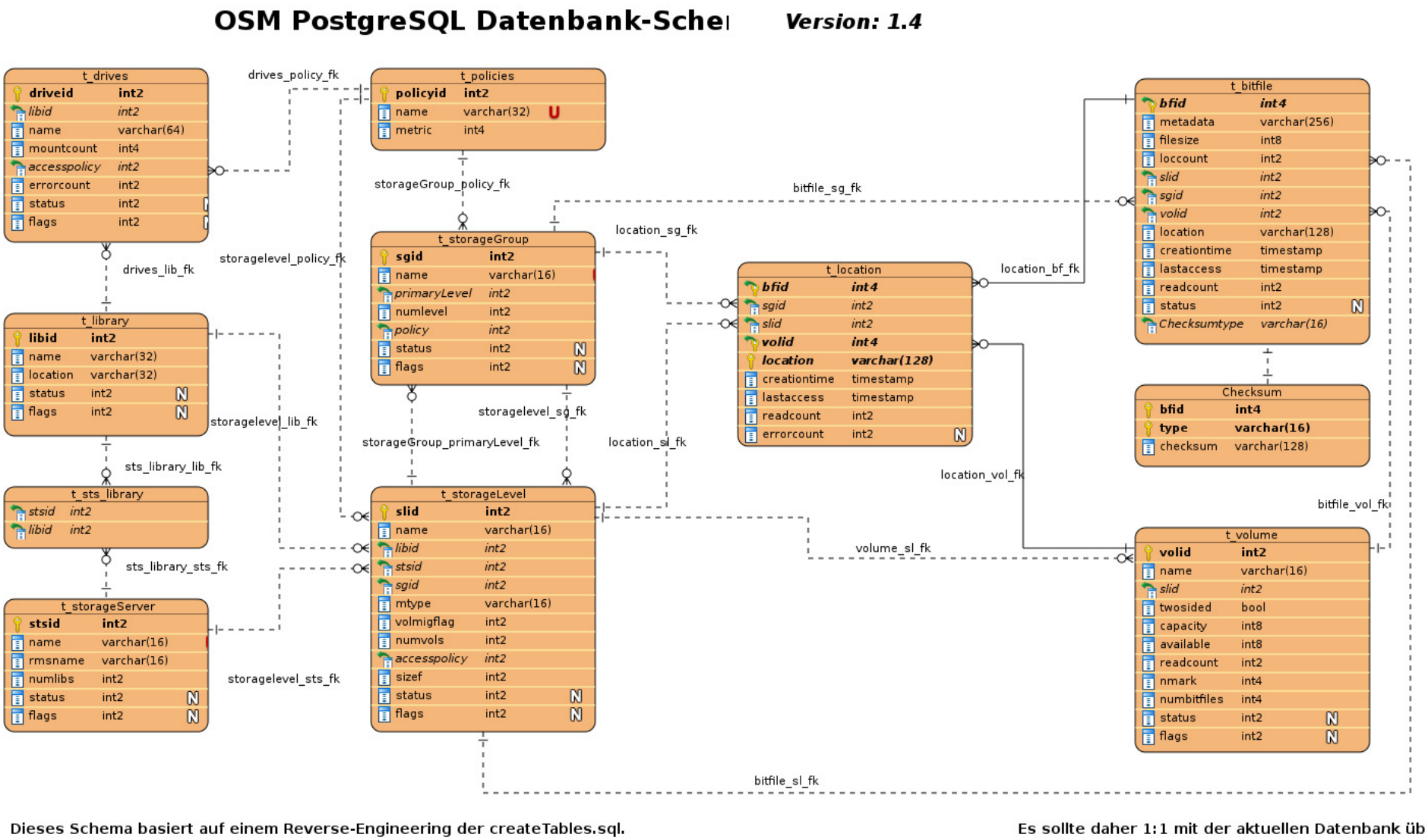

Abbildung: Neues OSM Datenbankmodell

Mögliche zukünftige Ziele und Meilensteine

Anpassung an moderne Standards und Formate

- Unterstützung für zukünftige Medien und Formate
- Linear Tape File System (LTFS) als Datenformat

Automatisierung von Prozessen

- Automatische Migration, Reklamation und Cloning
- Gewährleistung der Datenpfadkonsistenz via Hardware
- Integrierte und automatisierte Integritätsprüfungen

⇒ **Verwendung von OSM als Object Store**

- Als Grundlage für eines erweitertes DESY Archivsystems

Mögliche zukünftige Ziele und Meilensteine

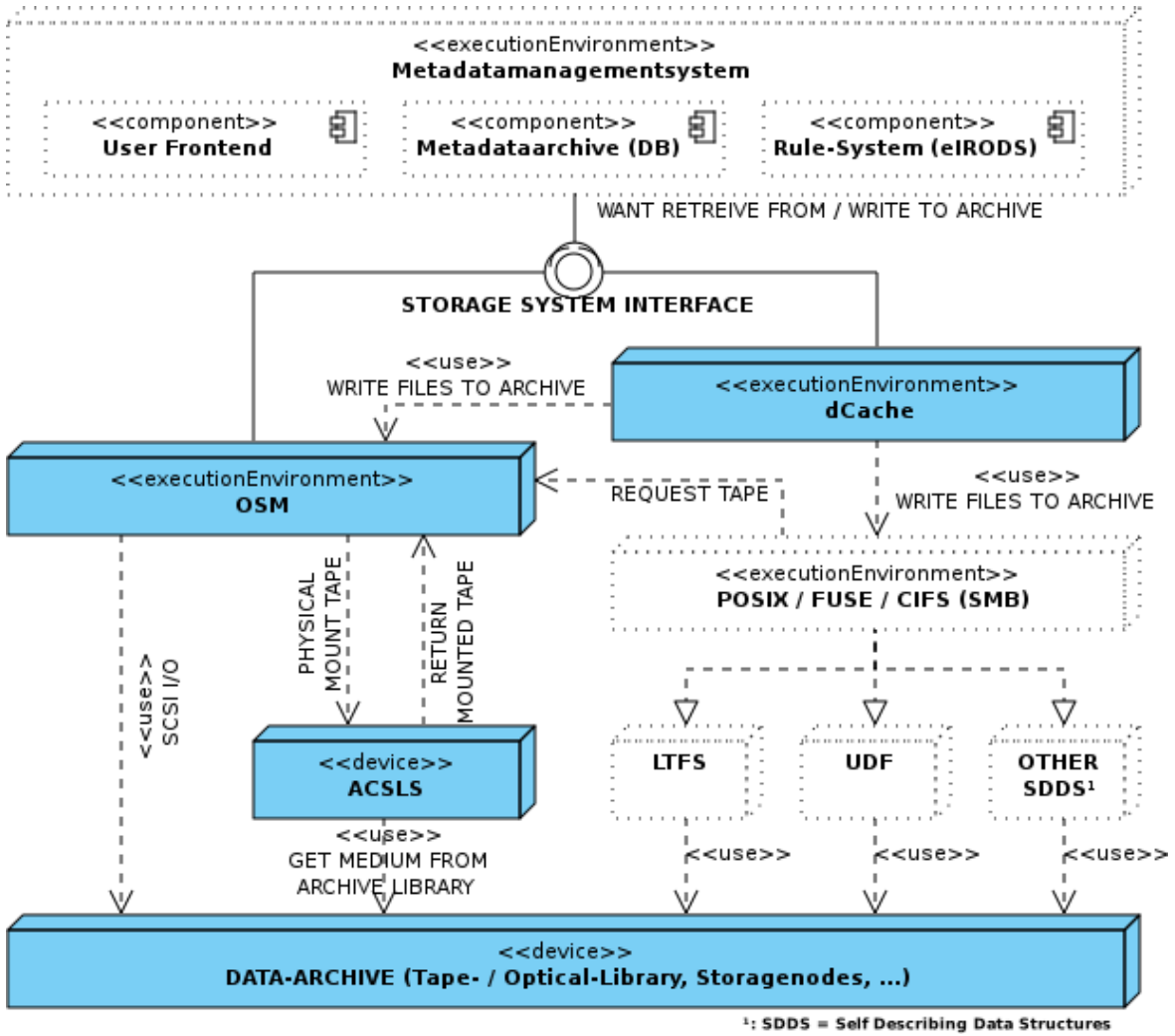

Abbildung: Konzept eines erweiterten DESY Archivsystems

Fazit.

- Zusammenfassung
- Zukünftige Entwicklungen

Zusammenfassung

Datenarchivierung

- Allgemeingültige Lösung existiert nicht
- Archivsystem abhängig von Anforderungen und Policies
- Nachhaltigkeit nur über offene Formate zu gewährleisten

Open Storage Manager

- Aktuelle Anforderungen weitestgehend implementiert
- Zusätzliche Features teilweise realisiert
- Stabiles System für die nächsten 5 bis maximal 10 Jahre

Zukünftige Entwicklungen

Herausforderungen bei der Archivierung

- Exponentiell steigendes Datenvolumen
- Verwaltung großer Datenmengen wird komplexer
- Zukünftige Experimente (CTA, XFEL, ILC, ...)

Entwicklungen bei Speichertechnologien

- Sony / Panasonic Archival Disc (≥ 1TB[5])
- Steigende Kapazitäten bei Magnetbändern (≥ 200TB[6])
- Festplatten als „Cold-Storage“ (3EB @ Facebook)[7]

[5] Archival Disc standard formulated · http://www.sony.net/SonyInfo/...
[6] IBM und Fujifilm stellen neuen Speicherdichte-Rekord auf · `http://heise.de/-2598090`
[7] Under the Hood: Facebook's cold storage System · https://code.facebook.com/posts/...

Vielen Dank für die Aufmerksamkeit.

Fragen!