



Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen



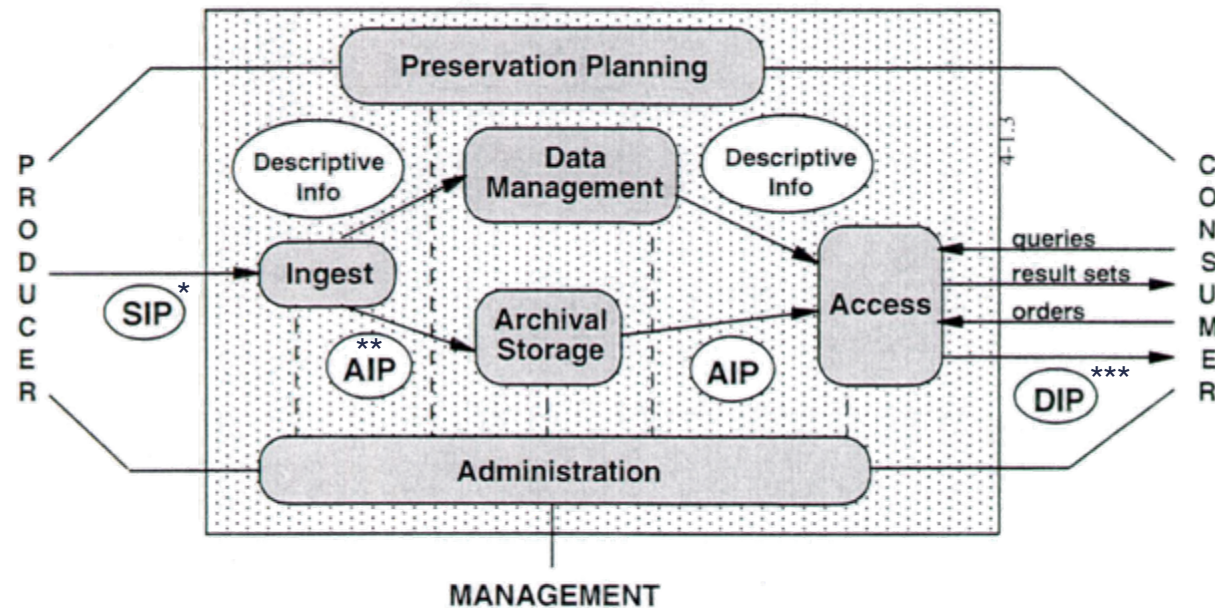
1. Das kopal Projekt: Eckdaten
2. Herausforderungen für ein Langzeitarchiv digitaler Informationen
3. OAIS Modell
4. Lösungskonzepte DIAS und Projekt kopal
5. Preservation Planning
6. Die Entwicklung eines universellen Objektformats
7. Projektziele
8. Die Partner

- Förderprojekt des Bundesministeriums für Bildung und Forschung (BMBF)
- Charakter: Forschungs- und Innovationsprojekt
- Fördervolumen: 4,2 Mio €, Laufzeit: 1.7.2004 – 30.6.2007
- Projektziel: Aufbau einer von Gedächtnisorganisationen nachnutzbaren technischen und organisatorischen Infrastruktur zur Sicherung der Langzeitverfügbarkeit elektronischer Publikationen
- Basis: DIAS (Digital Information Archiving System) entwickelt von IBM für die Koninklijke Bibliotheek, die Nationalbibliothek der Niederlande
- Projektpartner
 - Die Deutsche Bibliothek (DDB) (Gesamtprojektleitung)
 - Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB)
 - Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)
 - IBM Deutschland GmbH (IBM)

- Auflösung heutiger Engpässe bei der Archivierung bzgl. Anzahl und Art der Dokumente
- Rasanter Technologiewechsel
 - erschwert Zugriff auf ältere Datenformate
 - Zeitalter der Globalisierung verstärkt den Trend zum globalen Datenaustausch
 - Zwang zur Selektion durch die explosive Zunahme an Informationsmenge, versionierten Dokumenten, Annotationen an dynamischen Dokumentformaten
- Technische Aspekte: OAIS (Open Archival Information System)-Modell, Metadatenmodellierung, Authentifizierung

- Rasch wachsender elektronischer Publikationssektor
- Zunehmender Teil der kulturellen Aktivitäten im Netz
- Selektionskriterien entscheiden über die Tradierung unseren kulturellen Erbes
- Dieser „Nachlass“ ist – auch gesetzlich fixiert – Aufgabe von Nationalbibliotheken und vergleichbaren Einrichtungen
- Bestehende Archive erfüllen Anforderungen an „trusted digital repositories“ nicht

- Funktionale Anforderungen, auf denen DIAS beruht:
- Die relevanten Business Prozesse basieren auf dem OAIS Modell (=Reference Model for an „Open Archival Information System“ – ISO-Standard ISO 14721)
- OAIS dient als Framework für die DIAS-Lösung
- Der Kern von DIAS besteht aus 6 Prozessen, die mit Elektronischen Publikationen (EPs) in den drei Standardformaten arbeiten:
 - Ingest Process
 - Archival Storage
 - Data Management
 - Access
 - Administration
 - Preservation



* = SIP: Submission Information Package
** = AIP: Archival Information Package
*** = DIP: Dissemination Information Package

Figure 4-1: OAIS Functional Entities

- DIAS Core als generischer Kern der DIAS-Lösung bei der Nationalbibliothek der Niederlande

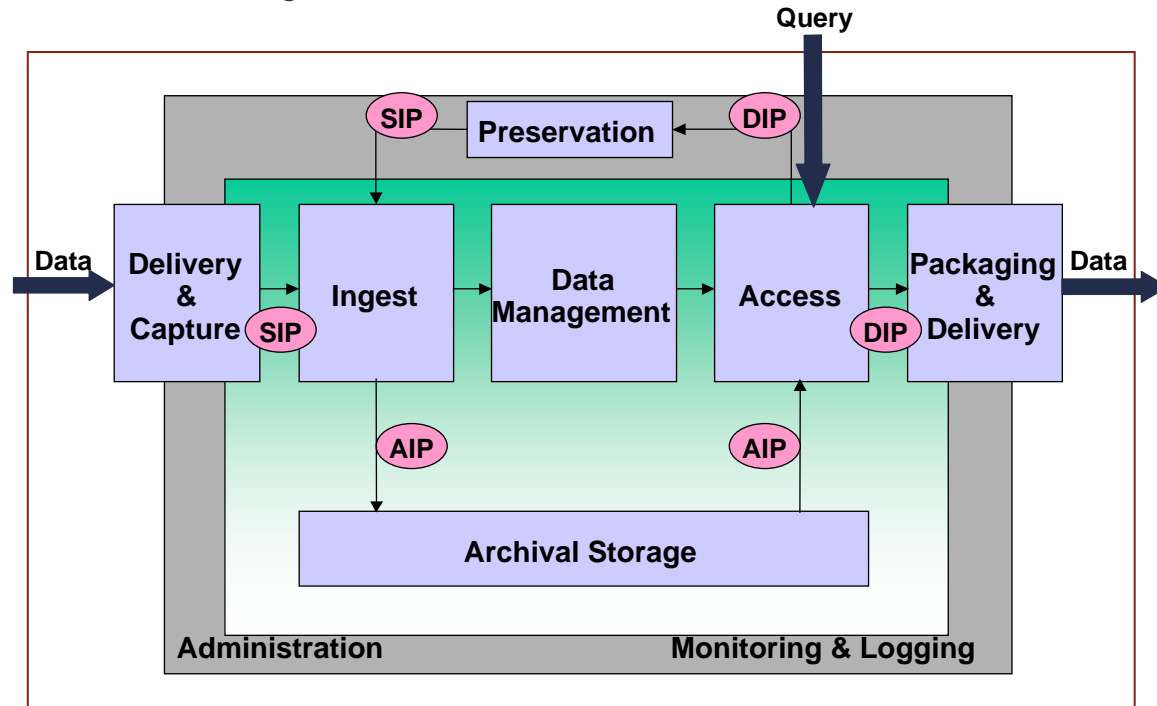
- DIAS Erweiterung

- Projektbeginn

- Mandantenfähigkeit
- Fernzugriff
- DIAS Upgrade auf Content Manager 8

- Projektdurchführung

- Funktionalitäten zu Preservation Planning
- METS (Metadata Encoding and Transmission Standard) Support

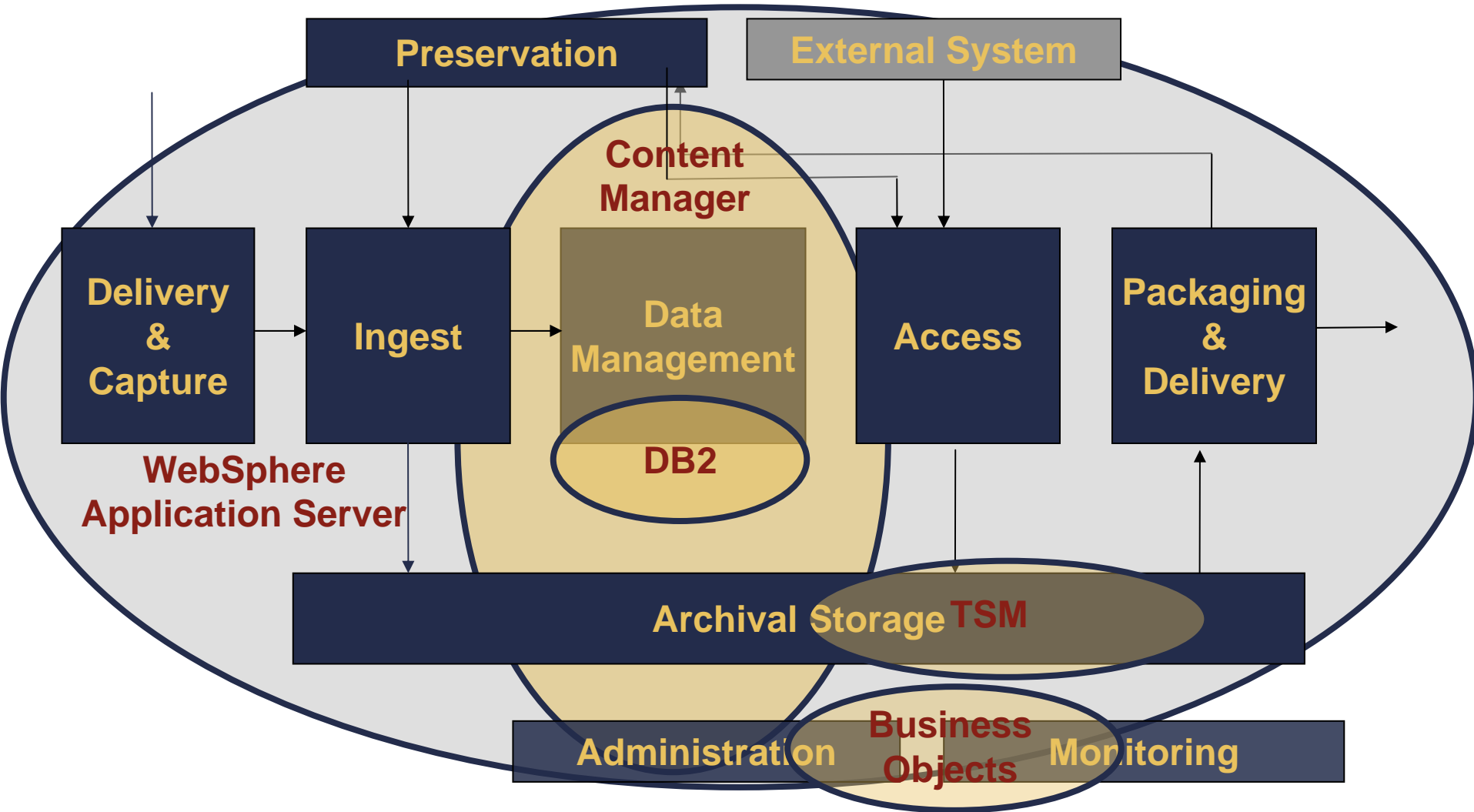


Quelle: NEDLIB

- Open-Source Framework: Entwicklung von Tools und Modulen zur funktionalen und technischen Einbettung von DIAS in den Anwendungskontext z.B.

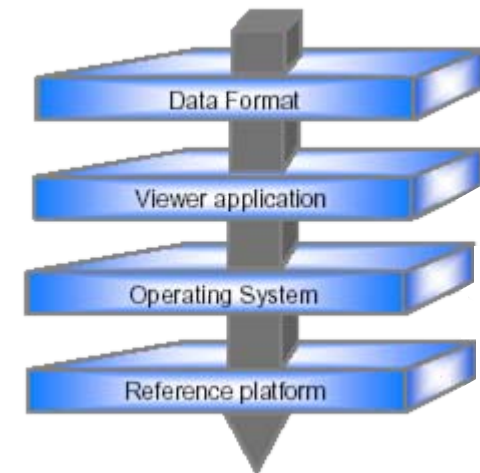
- Modul für Datenimport (Batch-Builder für unterschiedliche Eingangsformate der Verlage)
- Katalogintegrationen
- Zugriffsarten für archivierte elektronische Publikationen unterschiedlicher Art (z.B. Graphiken)

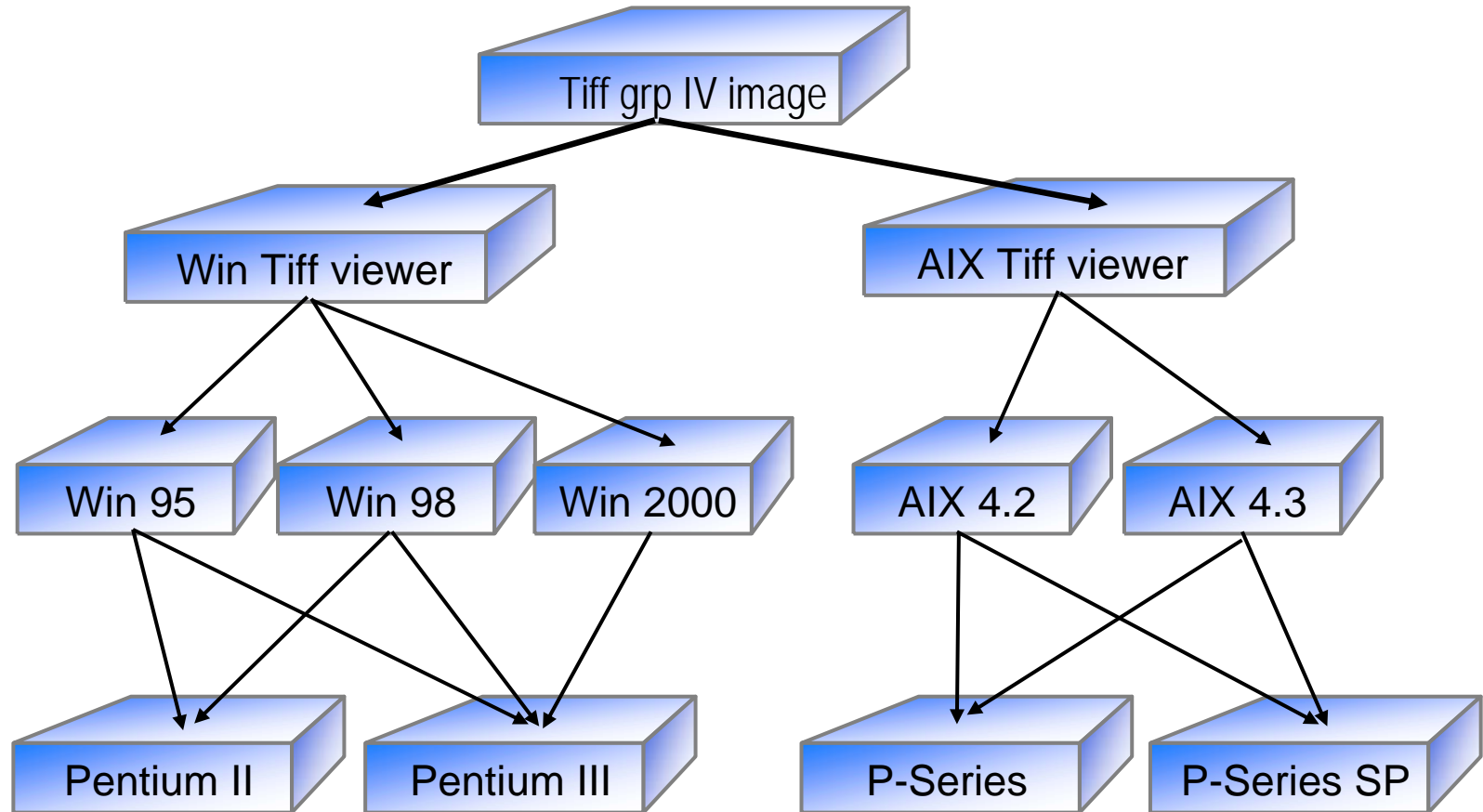
* = SIP: Submission Information Package
** = AIP: Archival Information Package
*** = DIP: Dissemination Information Package



- Preservation Layer Model (PLM)

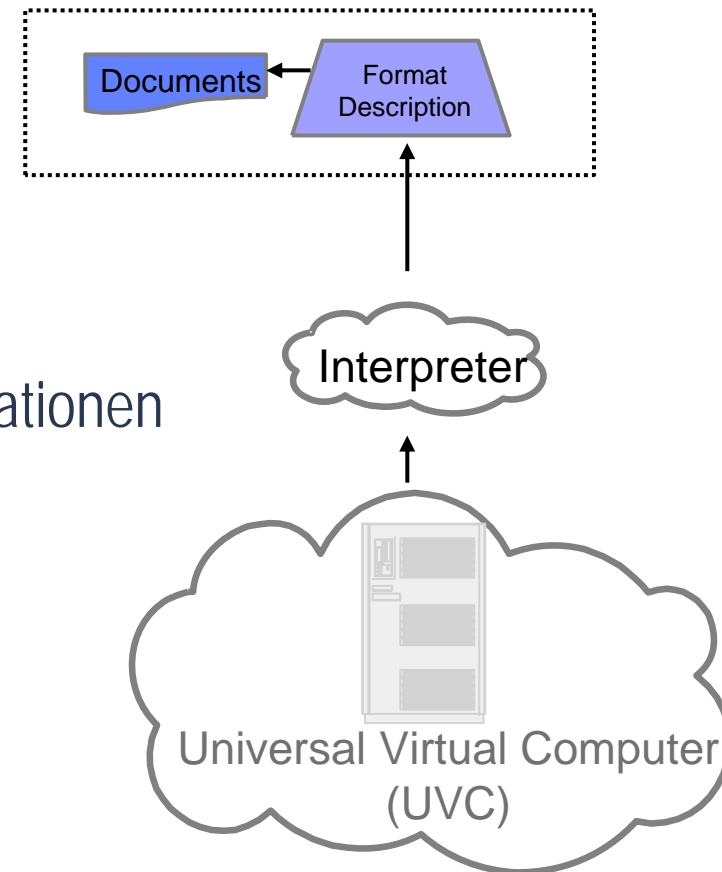
- Zugang unabhängig von technologischen Veränderungen
- basiert auf Beziehung zwischen Infrastrukturkomponenten (Zugang, Präsentation Elektronischer Publikationen)
- Entwicklung von ViewPaths für jedes archivierte Datenformat



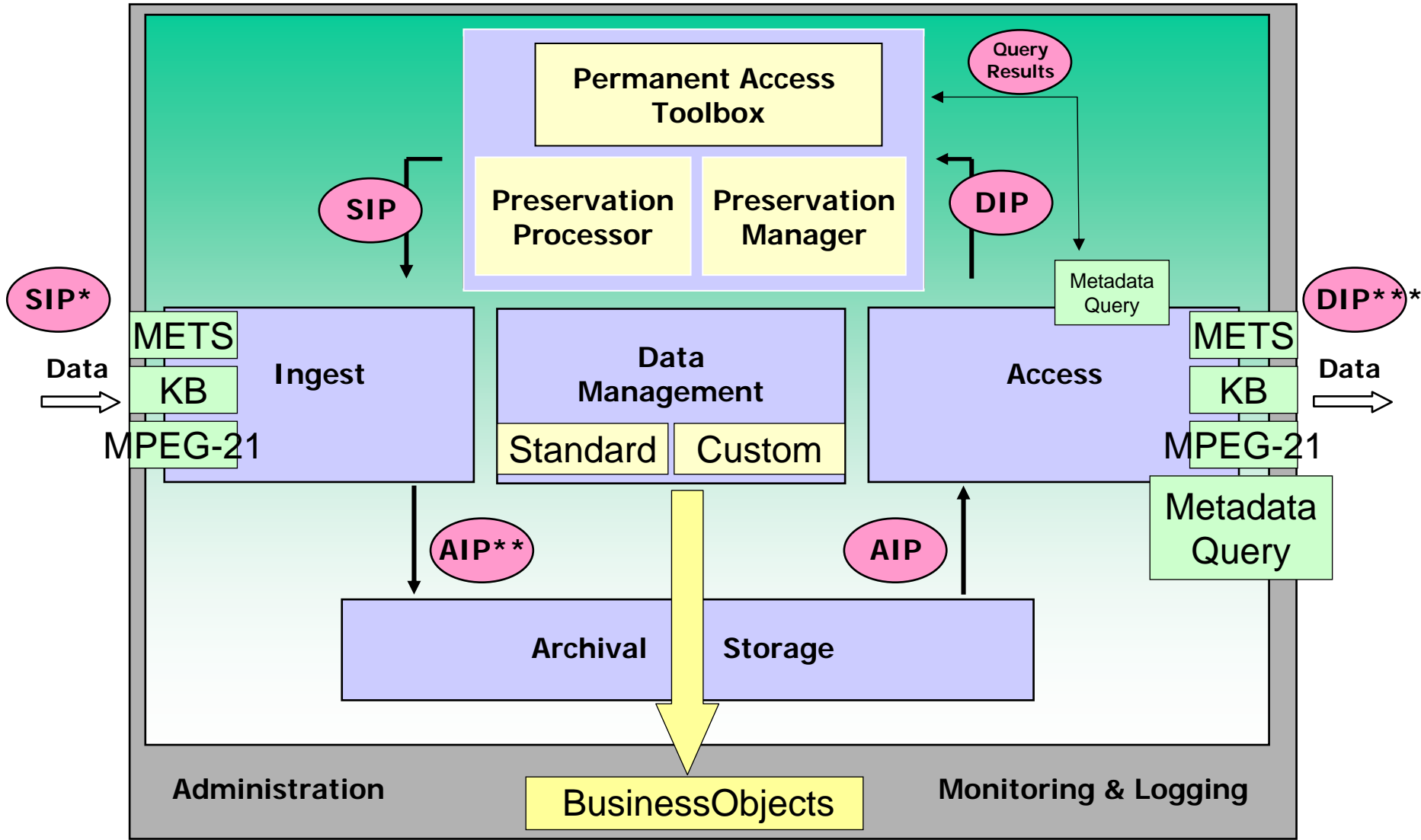


■ Universal Virtual Computer (UVC)

- Konzept eines flexiblen Emulators
 - Logische Analyse von Datenformate
 - Emulation von Präsentationsumgebungen
- Archivierung von Elektronischen Publikationen im Originalformat
- Viewer mit UVC Technologie garantiert Zugang
- Vorteile des UVC
 - Kein Zwang zur Festlegung von standardisierten Formaten
 - Nur Schreiben des UVC Interpreters bei neuen Architekturentwicklungen notwendig, keine Auswirkungen auf die archivierte Information
 - Gleichzeitige Allgemeingültigkeit und Einfachheit des UVC



- Die Entwicklung des universellen Objektformats basiert auf METS, dem in internationaler Kooperation entwickelten Metadata Encoding and Transmission Standard
- The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library' (Official METS Web Site)
- Eigenschaften von METS bzw. des universellen Objektformats:
 - XML-Format
 - Hülle für verschiedene Metadatensets
 - ermöglicht u.a. die Einbettung von Bitstreams



* = SIP: Submission Information Package
** = AIP: Archival Information Package

*** = DIP: Dissemination Information Package

- Langzeitarchivierung digitaler Informationen (>100Jahre)
- Entwicklung einer standardisierten Lösung
 - nachnutzbarer Transfer für andere Institutionen
 - vereinfachte Kooperation zwischen Bibliotheken
- Erweiterung des IBM Assets DIAS-Core um Open-Source basierte Komponenten
 - flexibler Einsatz durch Modularität der Lösung
- Entwicklung einer erweiterbaren User Group (KB, DDB, SUB...) zur kooperativen Weiterentwicklung der Lösung
 - Niedrigere Kosten
 - Zeitlich schnellere Weiterentwicklung
 - Stärkere Standardisierung möglich
- Prozessintegration z.B. Automatisierung der Einlieferrountinen, Optimierung für Massenverfahren

Projektziel / Motivation

- Einbringen von Expertise und Technologien für den Aufbau einer standard-basierten Langzeitarchivierungslösung für digitale Medien. Arbeiten mit innovativen Kunden, um die Lösung eng an den fachlichen Anforderungen auszurichten.

Kompetenzen

- spezifische Expertise durch Aufbau von DIAS bei der Nationalbibliothek der Niederlande
- Technologien für Langzeitarchivierung als Produktschwerpunkt (z.B. Universal Virtual Computer als Basis für Emulations- / Migrationsstrategien von elektronischen Publikationen)
- Service- und Projektkompetenz

Aufgaben

- Lieferung von DIAS-Core und IBM Content Manager als Basis zur Langzeitarchivierung
- Funktionale und technologische Weiterentwicklung von DIAS

Projektleitung IBM

- Irmgard Kurth, ikurth@de.ibm.com

Projektziel / Motivation

- Entwicklung einer tragfähigen Lösung für die Langzeitarchivierung der stark anwachsenden Zahl deutscher Netzpublikationen

Kompetenzen

- Zentrale Archivbibliothek Deutschlands
- Maßgeblich an der Entwicklung internationaler Standards beteiligt
- Umfangreiche Sammlung deutscher elektronischer Dissertationen und Netzpublikationen

Aufgaben

- Gesamtprojektleitung (R. Altenhöner)
- Konzeption technischer Anforderungen
- Einbringen der Erfahrungen bei Sammlung und Archivierung von Netzpublikationen
- Internationale Zusammenarbeit

Projektleitung DDB

- Tobias Steinke, steinke@dbf.ddb.de

Projektziel / Motivation

- Digitalisierte Wissens- und Informationsbestände im langfristigen Zugang für Wissenschaftler und Lernende bereitstellen

Kompetenzen

- Serviceorientierter Informationsdienstleister
- Innovationsorientierung durch mehr als 20 Projekte zur Digitalen Bibliothek
- GDZ als bundesweites Kompetenzzentrum für Digitalisierung und Metadatenhandling
- LRC für medienneutrales Publizieren

Aufgaben

- Erarbeitung der METS basierten Objektstruktur(en)
- Erstellung / Programmierung von Import- und Exportroutinen (TIFF, TeX)
- Entwicklung von Workflows
- Öffentlichkeitsarbeit

Projektleitung SUB

- Frank Klaproth, klaproth@sub.uni-goettingen.de

Projektziel / Motivation

- Langfristige Datenarchivierung ist eine Kernaufgabe der GWGD
- Langzeitarchivierung als künftig wichtige Dienstleistung der GWGD
- Großes Interesse an LZA auf Seiten beider Gesellschafter

Kompetenzen

- Rechen- und Kompetenzzentrum für Universität Göttingen und Max-Planck-Gesellschaft
- Backup- und Archivsystem mit räumlich getrennten Standorten der eingesetzten Bandbibliotheken
- Betreiber Hochgeschwindigkeitsnetz GÖNET (1 Gbit/sec) mit 622 Mbit/sec Anbindung an das Gigabit-Wissenschaftsnetz (G-WiN)

Aufgaben

- Hardware: Beschaffung + Pflege
- IT-Infrastruktur
- Serveradministration
- Speichermedienverwaltung

Projektleitung GWGD

- Dagmar Ullrich, dullric@gwdg.de