



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI

Schweizerisches Bundesarchiv BAR
Ressort Innovation und Erhaltung RIE

Programm ARELDA II

METS und die Planung für die digitale Archivierung im Schweizerischen Bundesarchiv

Dr. Krystyna W. Ohnesorge

**3. Europäische Konferenz über EAD, EAC und METS
Berlin, 24.-26. April 2007**



Themen

- Kurze Übersicht strategisches Programm ARELDAII:
 - Zielsetzung und Vorgaben
 - Realisationsumfang
 - Aufbau und Transformation von archivischen Informationspaketen
 - Aktueller Projektstand

- Zwischenergebnisse digitale Archivierung in ARELDAII:
 - Projekte DAp und GAp
 - Statistik digitale Archivierung
 - Archiv-Dateiformate
 - Rolle von Standards, inkl. METS

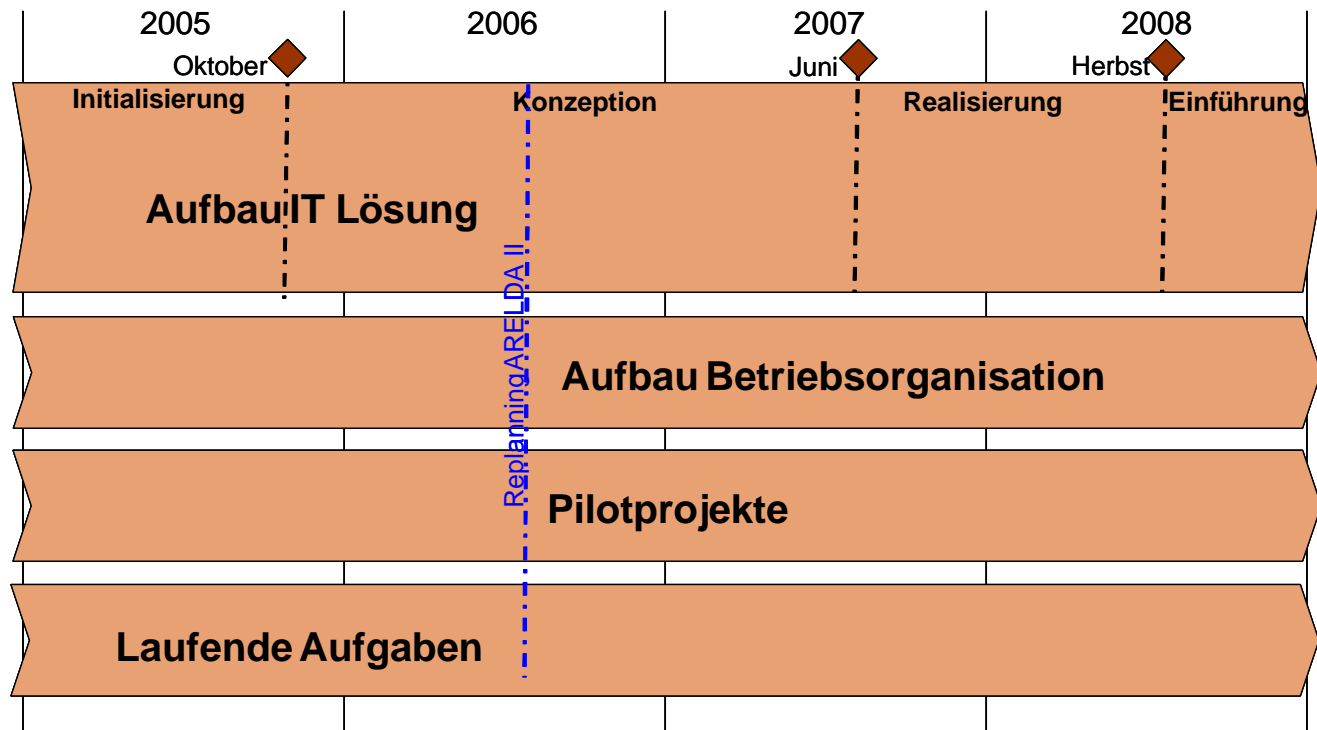


Strategisches Programm ARELDA II

- **Archivierung elektronischer Daten und Akten im Schweizerischen Bundesarchiv (BAR)**
- e-Government Schlüsselprojekt des Bundes
- Erste Projektetappe 2001 – 2004:
 - Exploration & Experimentieren
 - Grundlagenarbeit (Langzeitspeicher, SIARD)
- Zweite Projektetappe 2005 – 2008 (ARELDA II):
 - Archivprozess: End-to-End Betrachtung (Prozessmodell, Pilotprojekt DAp)
 - Grundlagenarbeit (KODIA, GAp)
 - Integrativer Ansatz: Organisatorische und technische Einbettung



Zielsetzung ARELDA II (1/2)





Zielsetzung ARELDA II (2/2)

- Aufbau eines IT Systems (Applikationen und Systeminfrastruktur) zur Archivierung digitaler Unterlagen
- Aufbau und Einführung der Betriebsorganisation zur Archivierung digitaler Unterlagen
- Entwicklung der Kompetenzen für die digitale Archivierung
- Übernahme und Erschliessung ausgewählter Unterlagen im Rahmen von Pilotprojekten
- Erbringung eines anerkannten Nutzens für die Interessengruppen (Stake Holders)



ARELDAII – Vorgaben (Auszug)

- **Das Ziel des Programms ARELDAII ist schnell praktische Resultate zu liefern für die Archivierung von:**
 - Digitalen Dossiers (1. Priorität)
 - Relationalen Datenbanken (2. Priorität: Überarbeitung SIARD)
 - File-Sammlungen (3. Priorität)

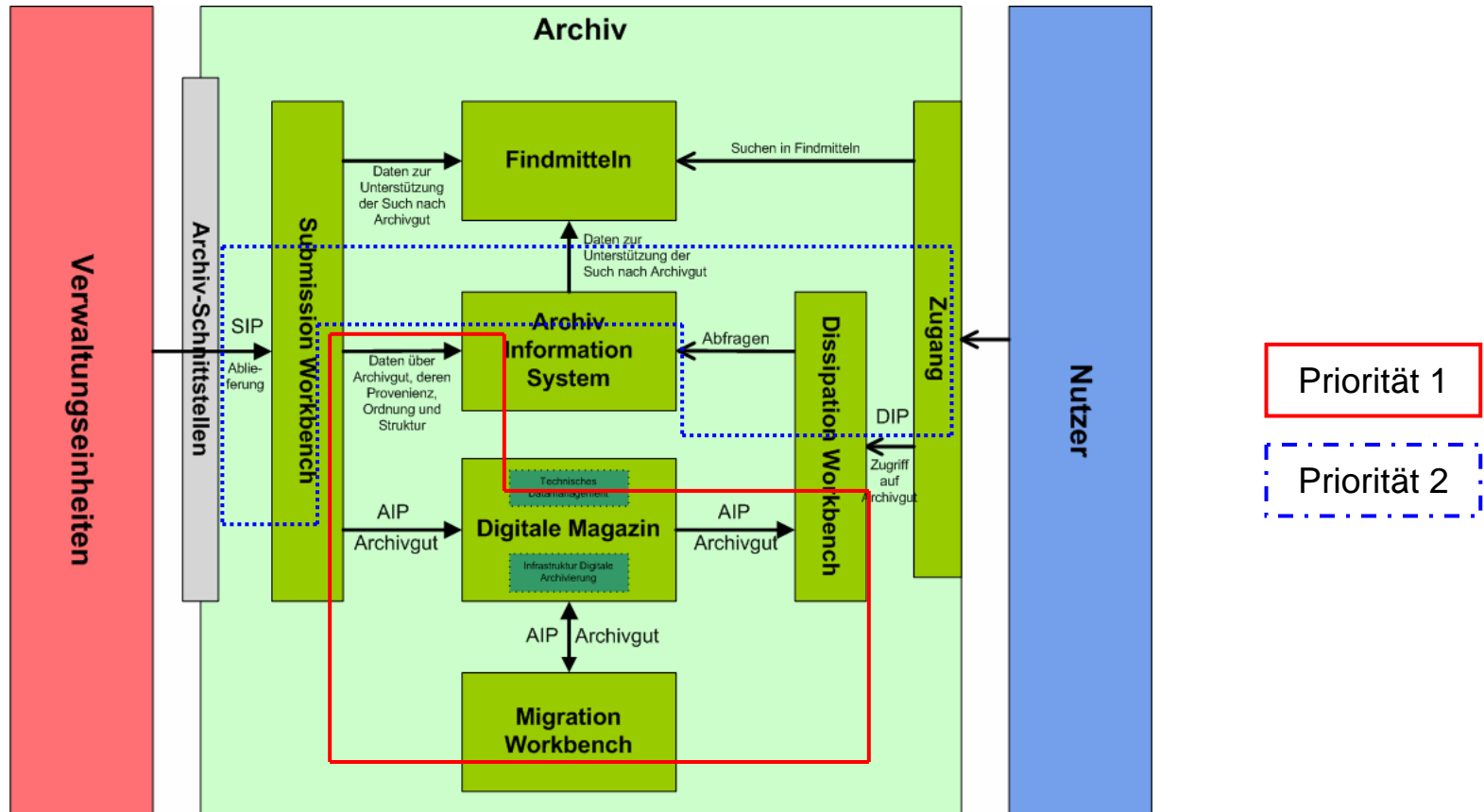
- **Die geltenden rechtlichen Grundlagen müssen erfüllt werden.**

- Metadaten müssen in das bestehende Archiv Information System des Bundesarchivs integriert werden (Kontinuität zum bestehenden analogen Archiv).

- Auf internationale und nationale Standards ist Rücksicht zu nehmen.



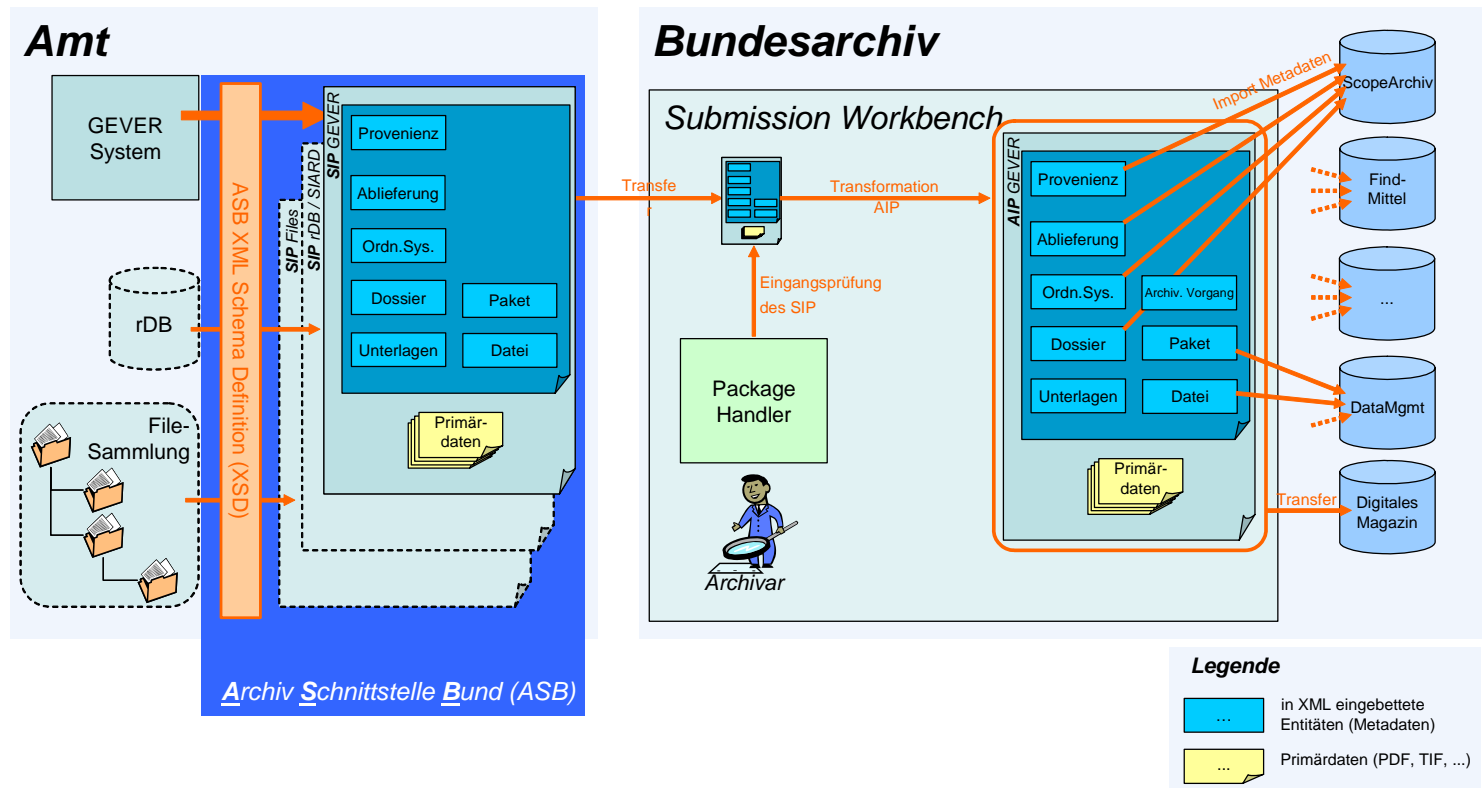
Realisationsumfang ARELDA II (1/2)





Realisationsumfang ARELDA II (2/2)

■ Ablieferungsüberblick





Informationspakete SIP, AIP und DIP (1/3)

- Die Ablieferung eines digitalen Dossiers an das Bundesarchiv erfolgt in Form eines SIP-Containers via eine archivische Schnittstelle, die auf Basis XML spezifiziert ist.
- Nach erfolgter BAR-interner Qualitätssicherung wird das SIP in ein AIP überführt. Mit Metadaten aus AIP werden verschiedene BAR-internen Applikationen beliefert und anschliessend wird das AIP an das digitale Magazin übergeben.
- Sobald ein Benutzer digitales Material anfordert, wird ein DIP mit dem angeforderten Inhalt erstellt und ausgehändigt.
- Die SIP Struktur wird für SIARD- und Filesammlungen-Ablieferungen weitgehend wie bei der SIP-Struktur für GEVER-Ablieferungen beschrieben.

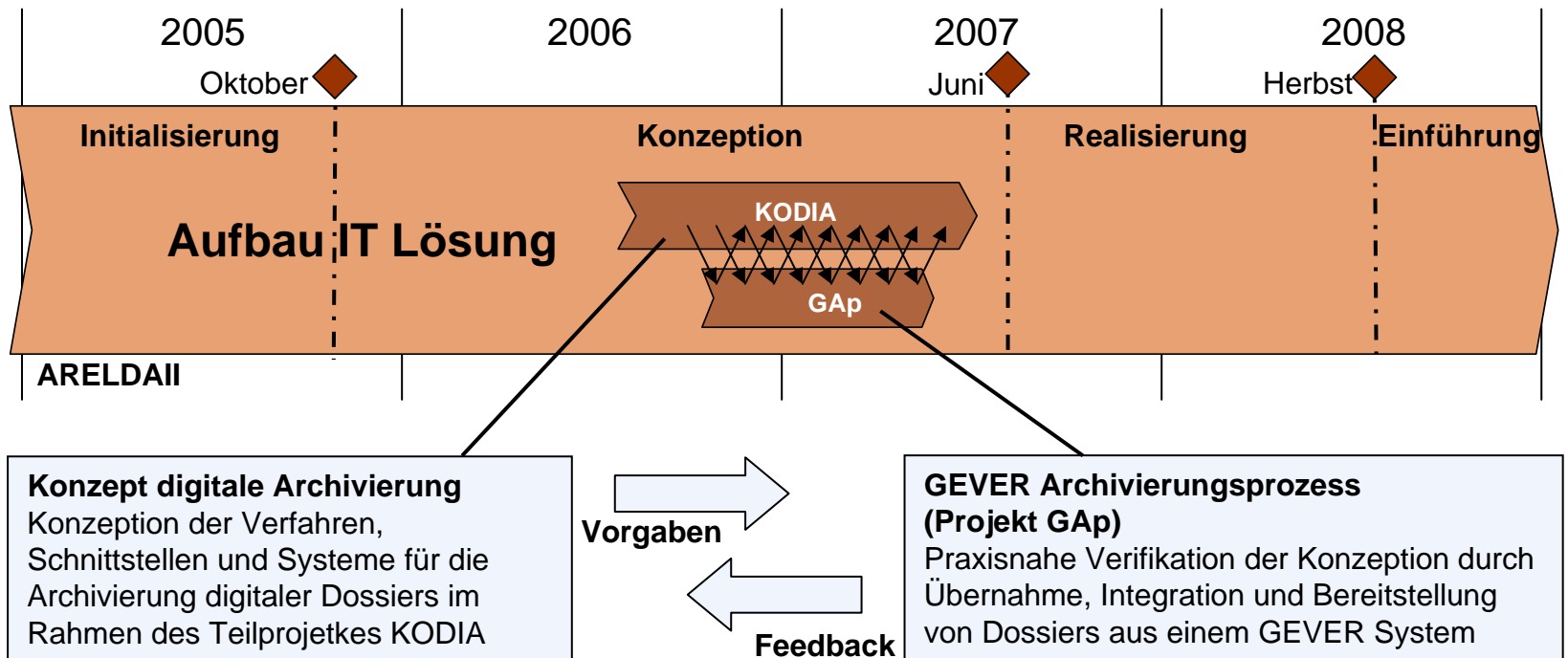


Informationspakete SIP, AIP und DIP (3/3)

- Das Datenmodell ist in ARELDAII stark hierarchisch strukturiert.
- Die Metadaten sind in XML-Dateien SIP, AIP und DIP mit Hilfe von XML Schema Definitionen spezifiziert.
- Ein SIP hat, analog zu METS, einen Header, trennt zwischen deskriptiven und administrativen Metadaten, etc.
- Ein SIP ist angepasst an bundesarchivischen Anforderungen und Tradition der Darstellung und Eingliederung von Informationen:
 - Im Ordnungssystem und seinen Positionen
 - In Dossier- und Subdossierstrukturen.



Aktueller Projektstand



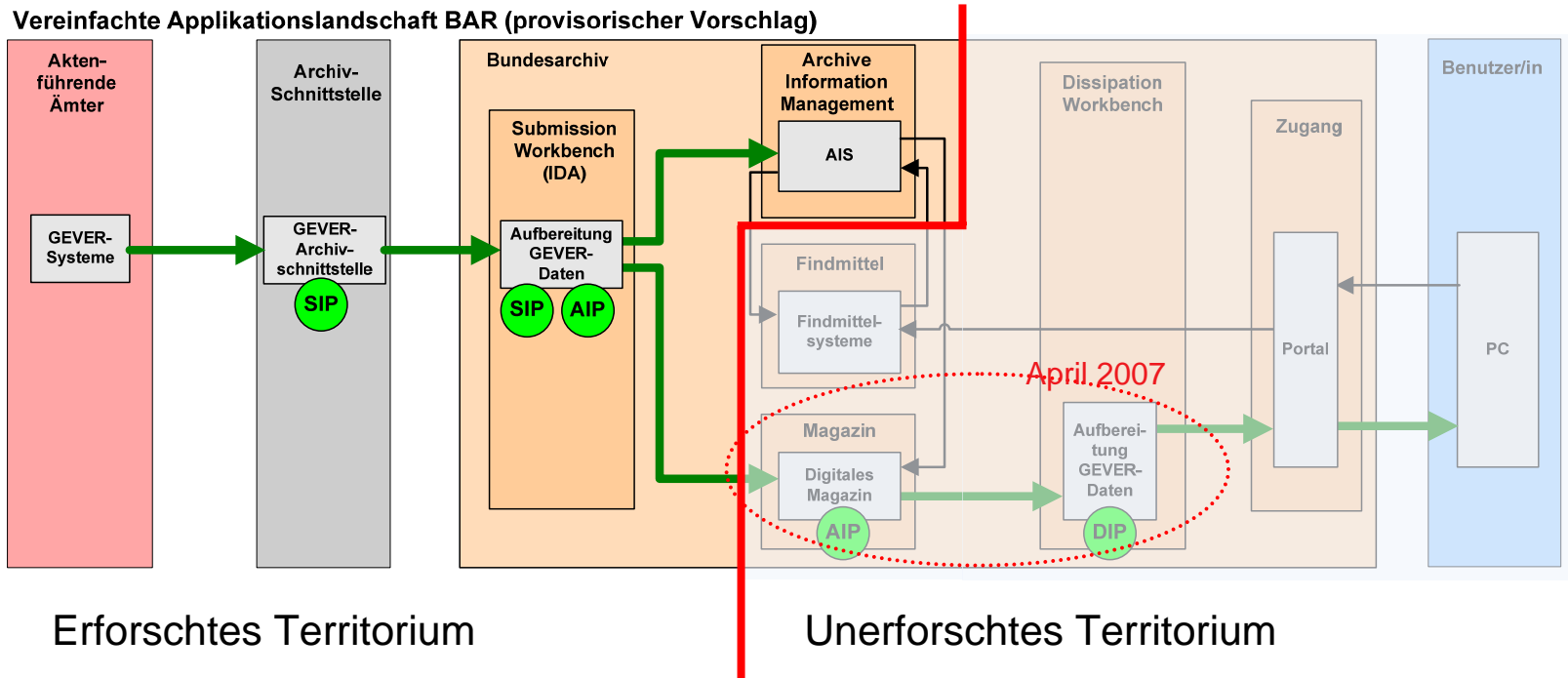


Proof of Concept durch das Teilprojekt GAp

- Ablieferung der digitalen Dossiers:
 - Kooperation mit einem Amt („echte“ digitale Dossiers / keine Testdaten)
 - Fabasoft eGov Suite als GEVER System
 - Archivschnittstelle durch ARELDA II definiert (XML)
 - Mapping der Schnittstelle in Zusammenarbeit mit Fabasoft Schweiz
- Integration und Bereitstellung der digitalen Dossiers:
 - Erschliessung der Dossiers im Archiv Information System des BAR auf Basis XML Technologie
 - Showcase des Zugriffs und der Bereitstellung digitaler Dossiers



Datenfluss im GAp





Erkenntnisse aus GAp

- Die Umsetzung der Schnittstellenvorgaben (GEVER ASB) ist möglich.
- Übernahme der Dossiers aus Fabasoft eGov Suite ist möglich.
- Die Erschliessung der übernommenen Dossiers ist möglich.
- Bereitstellung einzelner Dossiers oder Unterlagen ist möglich
- Kennzahlen der Übernahme:
 - Insgesamt knapp 1'000 Dossiers inkl. Subdossierstruktur
 - Ca. 10'000 Unterlagen übernommen
- Die Vorgaben aus der Konzeption konnten erfolgreich verifiziert werden!
- Erkenntnisse fliessen noch in den Abschluss der Konzeption ein.



Relationale Datenbanken

- Das Pilotprojekt „Datenbank-Archivierungsprozess“ (DAP) wurde 2006 durchgeführt, um einen konkreten Nutzen von ARELDAII aufzuzeigen. Zu diesem Zweck wurde der gesamte Archivierungsprozess für relationale Datenbanken definiert und implementiert.
- Innerhalb des DAP wurden in 2006 insgesamt 8 Datenbanken vollständig archiviert und 13 weitere behandelt (Bewertung, konzeptionelle Beratung, teilweise Übernahme, etc.).
- In 2007 sind weitere 3 relationale Datenbanken übernommen worden.



Statistik per Ende 2006

- Anz. digitaler Ablieferungen: ~ 140
- Gesamtheit aller Ablieferungen:
 - Grösse in TB total: ~10 TB
 - Anzahl Files total: 1 M
- Grösse der Ablieferungen:
 - kleinste, grösste 200 KB, 3.5 TB
 - Durchschnitt, Median 75 GB, 250 MB
- Anzahl Files pro Ablieferung:
 - kleinste, grösste 10, 400'000
 - Durchschnitt, Median 6600, 20



Archiv-Dateiformate (1/2)

- **Akzeptierte Formate (Stand 2007)**
 - Text, unstrukturiert (ASCII, ISO-8859-1/15, UTF-8, UTF-16)
 - TIFF
 - WAVE
 - SIARD RDB (Eigenentwicklung für die Archivierung relationaler DB)
 - CSV (comma separated values)

- **In Evaluation (de facto akzeptiert)**
 - PDF/A

- **Bedarf**
 - Container-Formate (ZIP)



Archiv-Dateiformate (2/2)

- **Tools (in Evaluation)**
 - DROID (Identifikation)
 - JHOVE (Identifikation und Verifikation)
 - PDF/A (spezialisierte Softwareprodukte)

- **Format Registry (in Evaluation)**
 - Wissens- und Referenz-Datenbank für Formate (Identifikation, Beschreibungen, Standards, Referenzen, etc.)
 - Prinzipielle Struktur für Tabelle, eindeutige Format-ID
 - Noch offen PRONOM, GDFR



Rolle von Standards (1/2)

- Die Überprüfung von Standards (u.a. ISAD(G), OAIS, METS, EAD, Dublin Core, etc.) für die Konzeption wurde auf dem Stand 2006 eingefroren.
- Die Konzeption der digitalen Archivierung im Bundesarchiv „lehnt sich“ – wie METS – stark an den OAIS Standard (mit SIP, AIP und DIP) an.
- Die Ideen des METS-Standard haben zwar das Vorgehen beeinflusst, die Einhaltung des Standards war aber keine prioritäre Vorgabe des Projekts.
- Die Vorgabe war, ein einheitliches, einzelnes Schema für die Beschreibung der Ablieferung zu definieren.
- Die Überprüfung des METS für die Beschreibung der einzelnen Unterlagen ist sinnvoll. Sie wird in das Aufgaben-Portfolio aufgenommen.



Rolle von Standards (2/2)

- Wesentliche Differenzen zwischen dem Vorgehen des BAR und Vorgaben des METS Standards:
 - Beim Aufbau und Transformation der archivischen Informationspakete, werden zusätzlich zu Primärdaten vorarchivische (komplexe) Zusammenhänge abgebildet. Ein SIP, AIP und DIP enthält immer diese Kontextinformationen.
 - Das BAR definiert eine gemeinsame Sicht bezüglich Metadaten für analoge, digitale und hybride Dossiers gemäss spezifischen Anforderungen des Bundesarchivs.

- Das BAR arbeitet im EU-Projekt **Planets** (www.planets-project.eu) mit. Im Teilprojekt **Characterisation** setzt man sich mit Standardsfragen auseinander. Das BAR unterstützt diese Arbeiten.



Bundesarchivische Abkürzungen

- **BAR:**
Schweizerisches **B**undesarchiv
- **ARELDAII:**
Archivierung **e**lektronischer **D**aten und **A**kten, Strategisches Programm im BAR
- **SIARD:**
Software-**I**nvariante **A**rchivierung aus **R**elationalen **D**atenbanken, BAR
- **DAp:**
Datenbank-**A**rchivierungs**p**rozess, Pilotprojekt im Rahmen des ARELDAII
- **GEVER:**
Elektronische **G**eschäfts**v**erwaltung (Record Mgmt + Workflow + Controlling)
-
- **KODIA:**
Konzept **d**igitale **A**rchivierung, Projekt im Rahmen des ARELDAII
- **GAp:**
GEVER Archivierungs**p**rozess, *Proof of Concept* Projekt im Rahmen des ARELDAII