



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Schweizerisches Bundesarchiv BAR
Ressort Innovation und Erhaltung

Archivierung von Daten aus relationalen Datenbanken im Schweizerischen Bundesarchiv

13. Tagung des Arbeitskreises "Archivierung von Unterlagen aus digitalen Systemen" vom 27./28. April im Staatsarchiv St. Gallen

28. April 2009 – Dr. Krystyna W. Ohnesorge



Zur Orientierung

- GEVER Programm des Bundes und die digitale Archivierung im BAR
- SIARD Format und SIARD Suite für die Archivierung von relationale Datenbanken
- Digitale Archivierung im BAR – Ausblick 2009

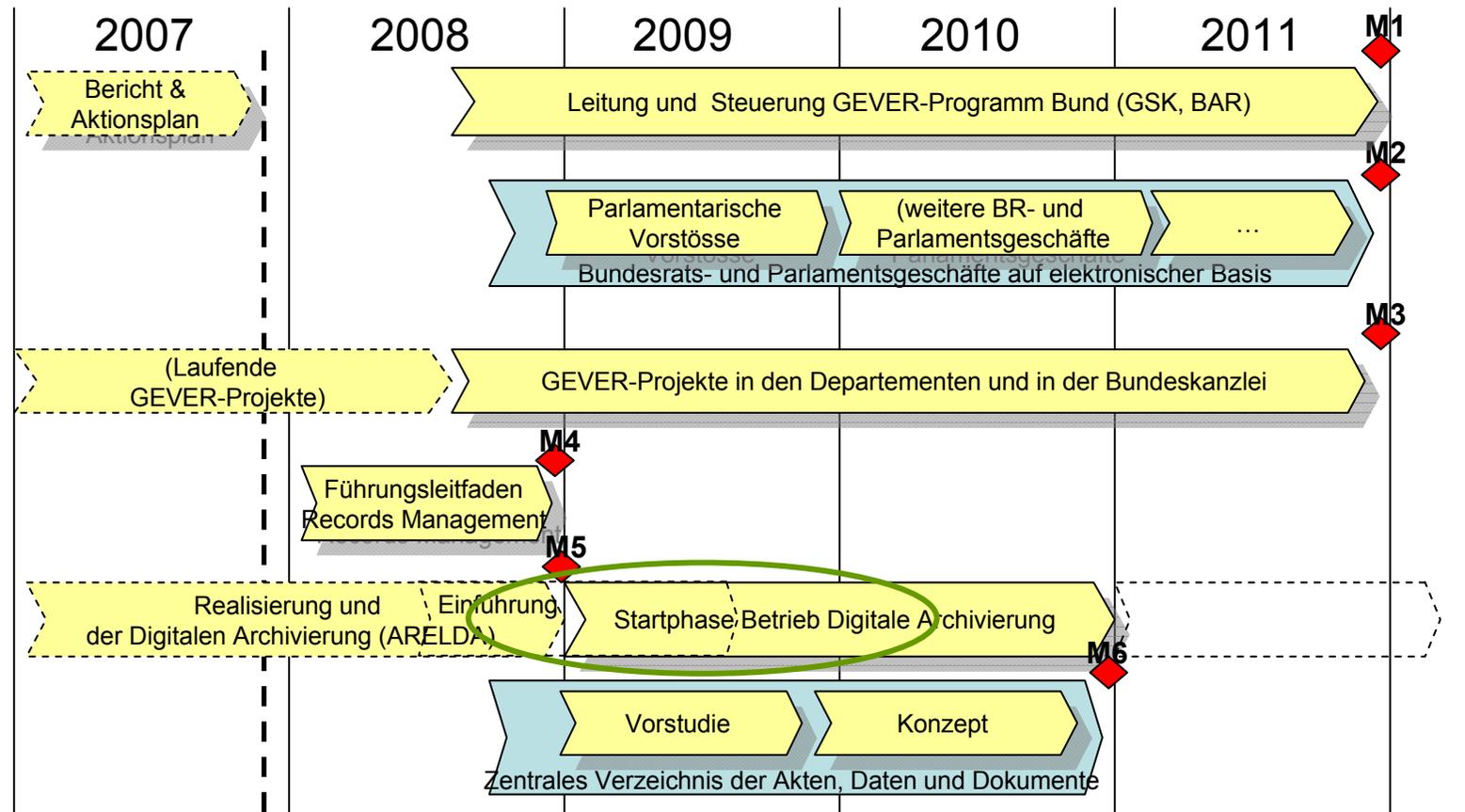


Zur Orientierung

- GEVER Programm des Bundes und die digitale Archivierung im BAR (Programm ARELDA)
- SIARD Format und SIARD Suite für die Archivierung von relationale Datenbanken
- Digitale Archivierung im BAR – Ausblick 2009



Programm GEVER Bund - Zeitplan



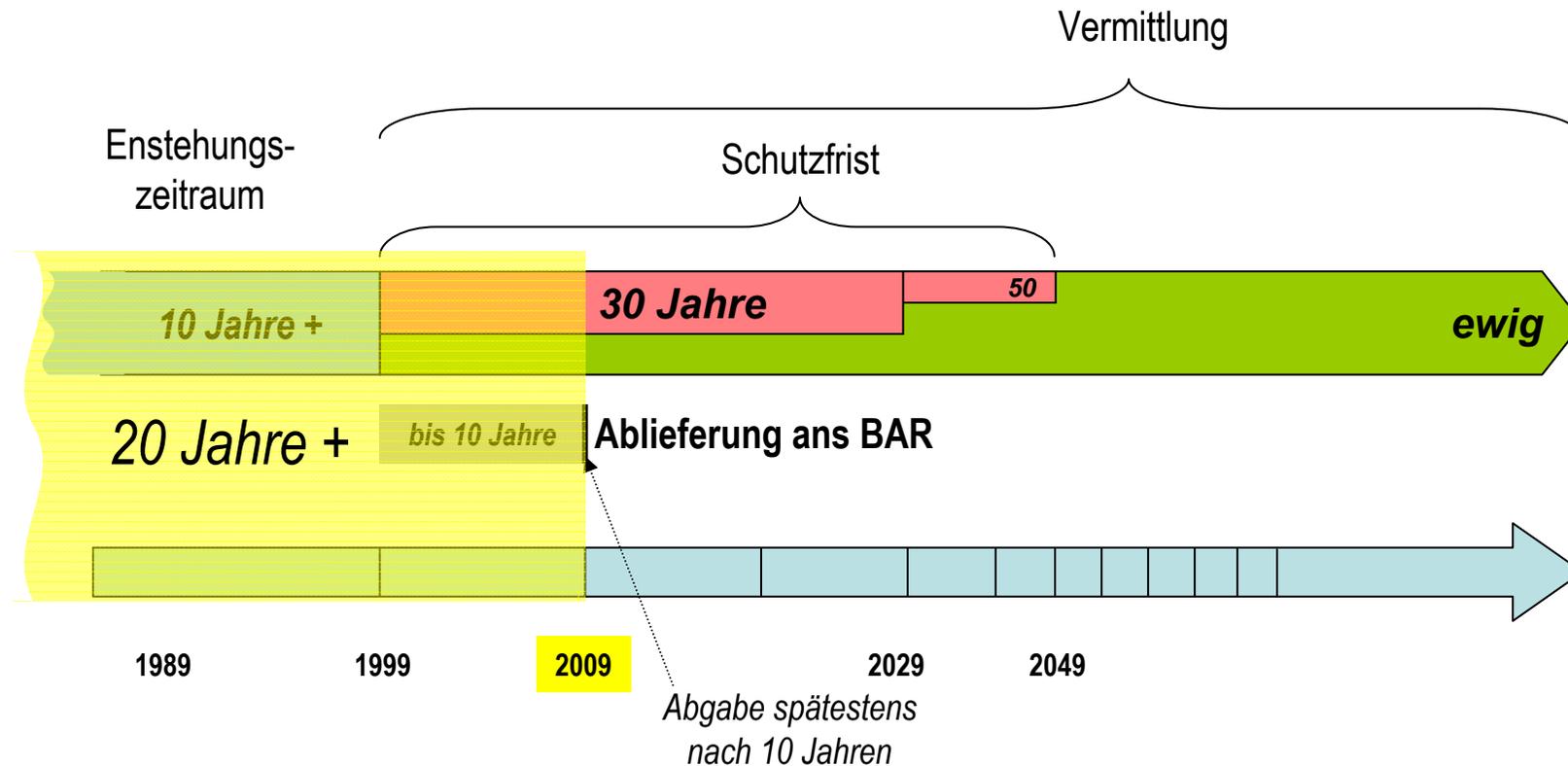


E-Archiv – Statistik per Ende 2008

- Grösse des digitalen Archivguts ~ 12.5 TB
- Anzahl digitaler Ablieferungen ~ 250 (davon 21 DB)
- Grösse der Ablieferungen
 - kleinste, grösste 122 KB, 3.5 TB
 - Median 15 MB
- Anzahl Files pro Ablieferung
 - kleinste, grösste 1, 400'000
 - Median 400

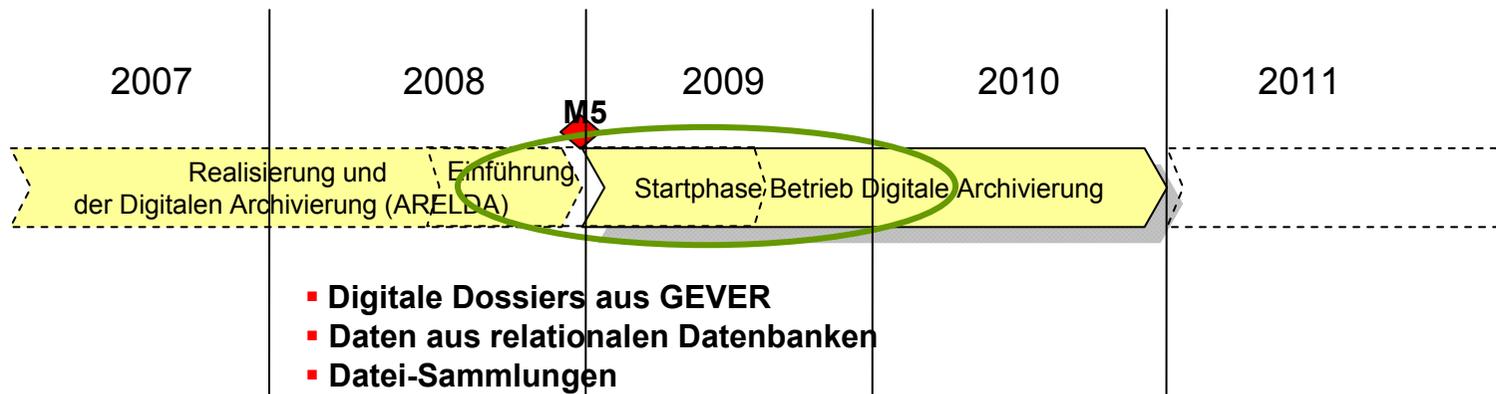


Information Data Life Cycle





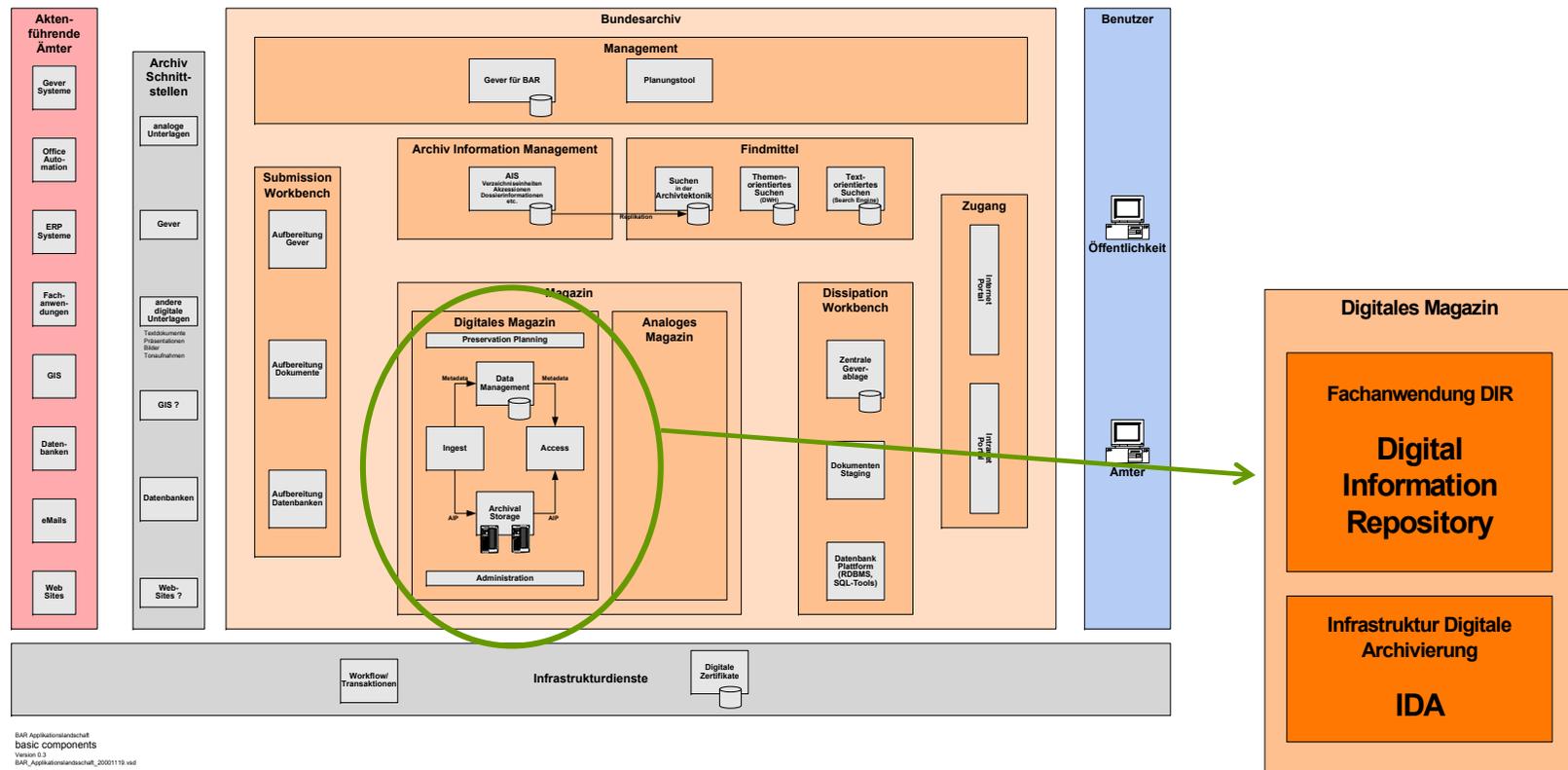
Strategisches Programm ARELDA



- Ab 2009 wird ein Standard-Archivierungsprozess für digitale Unterlagen etabliert
- Das BAR ist ein zuverlässiger Partner für die digitale Archivierung und den nachhaltigen Zugang zum Archivgut

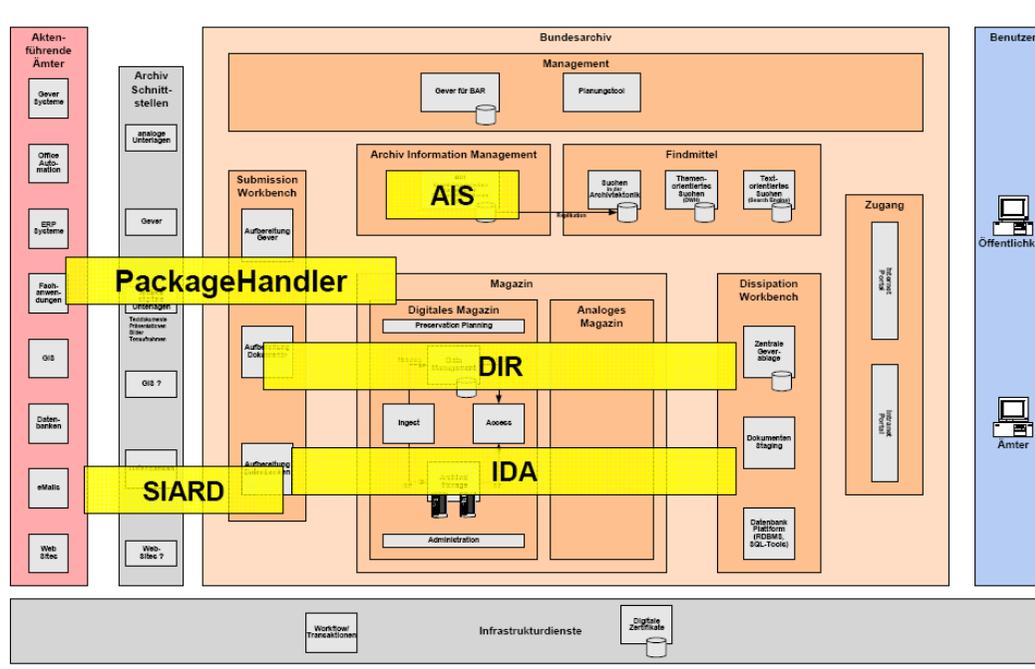


Schwerpunkt Realisierung ARELDA





Mit ARELDA per 31.12.08 erreicht



PackageHandler: Ein Werkzeug, um SIPs lesen, deren Metadaten editieren und im Fall von File Ablieferungen auch SIPs erstellen zu können.

DIR: Eine Anwendung zur Einlagerung, Verwaltung und Auslagerung von SIPs in einem digitalen Archiv.



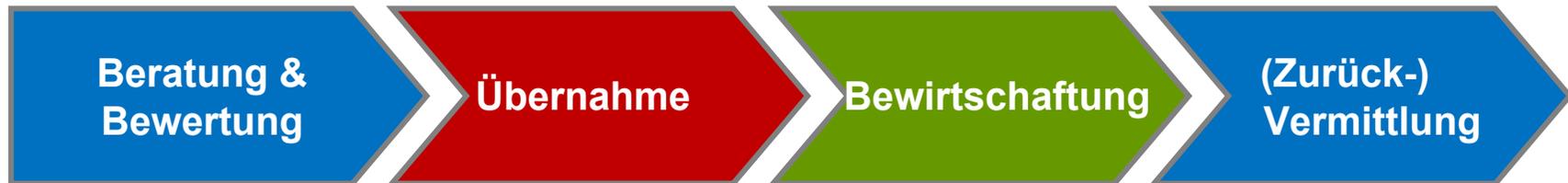
Archivtaugliche Dateiformate

- Die vom BAR als archivtauglich definierten Dateiformate
 - Textdaten (unstrukturiert) **TEXT**
 - Druckbare Dokumente **PDF/A**
 - Tabellarische Daten **CSV**
 - Relationale Datenbanken **SIARD**
 - Rasterbilder **TIFF**
 - Audiodaten **WAVE**
 - Videodaten **MPEG/4**

Merkblatt des BAR „Archivfachliche Dateiformate – Standards für die Archivierung digitaler Unterlagen“
(Stand: Juli 2007): <http://www.bar.admin.ch/dienstleistungen/00516/00517/index.html?lang=de>



Pilotprojekt Dap – 2005 / 2006



- **DAp**: Datenbank Archivierungsprozess
- Zusammenarbeit mit verschiedenen Ämter (Schwerpunkt EJPD)
- Einsatz des Software-Prototyps SIARD
- Archivierung von relationalen Datenbanken:
 - über 20 Datenbanken behandelt (inkl. Beratung) und davon 8 archiviert



Pilotprojekte GAp – 2008

- **GAp**: GEVER Archivierungsprozess
- Zusammenarbeit mit GS-EVD und Fabasoft
 - Design und Prototyp der Archivschnittstelle für Fabasoft
- Zusammenarbeit mit BIT:
 - Design der Archivschnittstelle für GEVER Office
- 2007: Proof of Concept Projekt: Zusammenarbeit mit ISB



Zur Orientierung

- GEVER Programm des Bundes und die digitale Archivierung im BAR
- SIARD Format und SIARD Suite für die Archivierung von relationale Datenbanken
- Digitale Archivierung im BAR – Ausblick 2009



Archivierung von relationalen DB

- Heutzutage werden in der schweizerischen Bundesverwaltung Fachanwendungen hauptsächlich mit relationalen Datenbanken eingesetzt
- Das Schweizerische Bundesarchiv hat eine normative Beschreibung eines Datenformat für die langfristige Erhaltung von relationalen Datenbanken entwickelt – **SIARD Format**
- Das SIARD Format wurde im Mai 2008 als das offizielle Format des EU-Projektes Planets angenommen: www.planets-project.eu



Das SIARD Format

- Das SIARD Format ist nur das Langzeitspeicherformat für eine spezielle Sorte von Datenbanken – von **relationalen DB**
- Es ist unabhängig von Paketstruktur wie SIP (Submission Information Package), AIP (Archival Information Package) und DIP (Dissemination Information Package) des OAIS Modells
- Eine DB im SIARD Format kann als Teil eines Archivpakets zusammen mit anderen Unterlagen (Dokumentation, Datenmodell, etc.) archiviert werden



Das SIARD Format

- Das SIARD Format beruht im Wesentlichen auf den ISO Standards SQL:1999, XML und UNICODE sowie dem Industriestandard ZIP
- Eine im SIARD Format archivierte relationale DB wird als eine einzige Datei abgelegt – als ein einziges unkomprimiertes ZIP-Archiv (ZIP)
- Sämtliche DB-Inhalte werden in einer Kollektion von XML-Dateien gespeichert. Einzige Ausnahme sind grosse BLOB- und CLOB-Daten, die in separaten binären Dateien gespeichert aber in den XML-Dateien referenziert werden

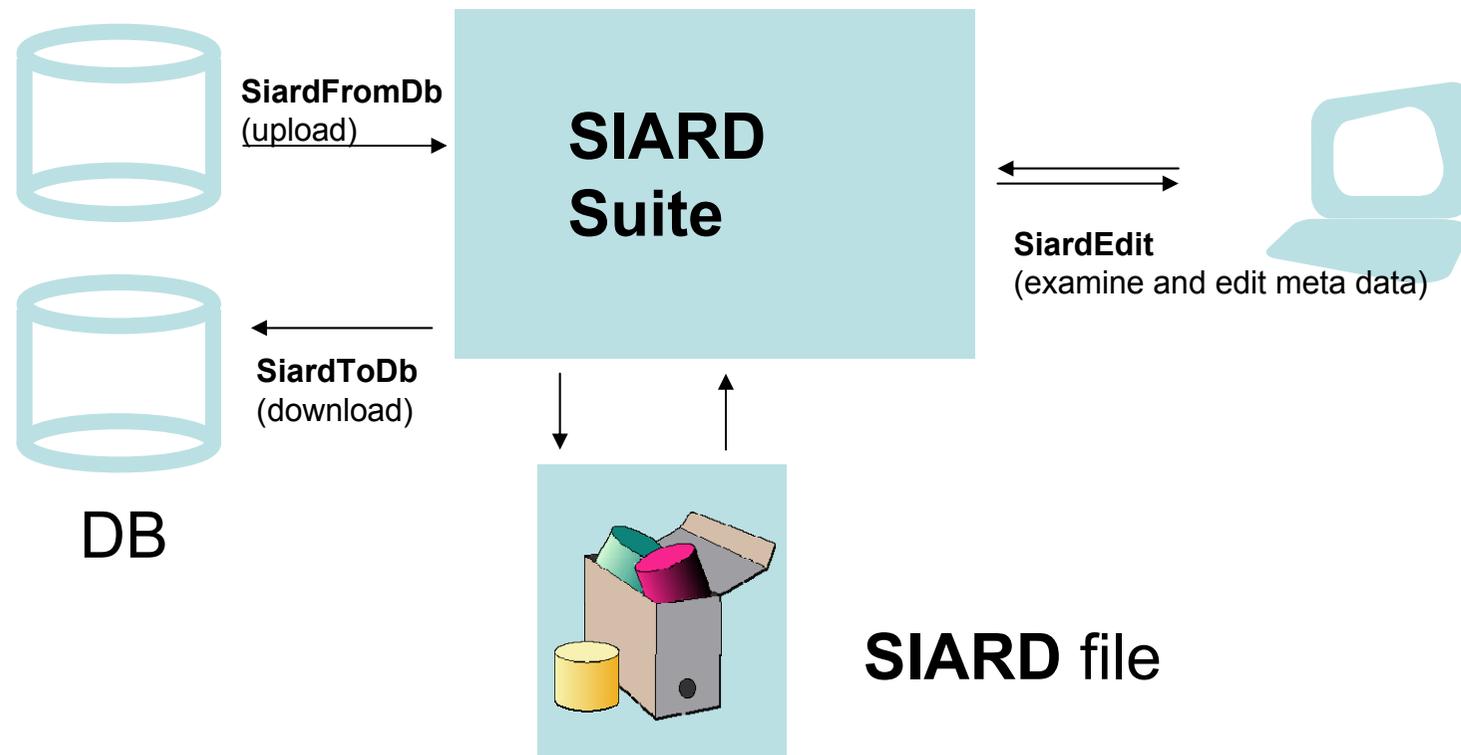


Archivierung von relationalen DB

- Wenn Struktur und Inhalt einer DB ins SIARD-Format übersetzt werden, wird es später jederzeit möglich sein, auf die DB-Daten zuzugreifen, selbst wenn die ursprüngliche DB-Software nicht mehr verfügbar oder nicht mehr lauffähig sein wird
- Im Rahmen des Projektes ARELDA wurde die Anwendung **SIARD Suite** entwickelt, welche Inhalte relationaler Datenbanken archiviert
- Das offene SIARD-Datenbankformat kann unabhängig von der Anwendung SIARD Suite verwendet werden



Das SIARD Format und die SIARD Suite





Voraussetzungen

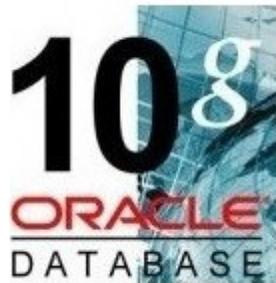
- **SIARD Suite ist platformunabhängig**
 - Lauffähig unter **Windows, Linux, Mac...**
 - Korrekt installiertes **JAVA 1.5** und höher
- **Installation**
 - SIARD Suite kann jederzeit ab **USB Stick, CD-ROM** oder **Netzwerklaufwerk** gestartet werden
 - Ohne spezielles **Installationprogramm**





Voraussetzungen

- SIARD Suite ist vorzugsweise für die folgenden Datenbanktypen konzipiert



Microsoft
SQL Server 2000
Standard Edition

& weitere...



Beschränkungen

SIARD Format unterliegt folgenden Beschränkungen:

- Die Grösse der SIARD Datei kann ca. 18 ExaBytes nicht überschreiten (ZIP64-Beschränkung)
- Die Anzahl der Dateien kann ca. 4 Milliarden nicht überschreiten (ZIP64-Beschränkung)
- Die Gesamtheit der Metadaten der DB muss im JAVA-Speicher Platz haben



Zur Orientierung

- GEVER Programm des Bundes und die digitale Archivierung im BAR (Programm ARELDA)
- SIARD Format und SIARD Suite für die Archivierung von relationale Datenbanken
- **Digitale Archivierung im BAR – Ausblick 2009**



Digitale Archivierung – Ausblick 2009

- Projekt ARELDA wird im Sommer 2009 abgeschlossen → Betriebsorganisation „Dienst Digitale Archivierung“ etabliert
- Stabilisierung und Weiterentwicklung der IT Lösung und Ausbau des Vermittlungsteils
- September 2009: Durchführung der 1. Ausbildung digitale Archivierung für die BV: <http://www.bar.admin.ch/dienstleistungen/00516/00714/index.html?lang=de>
- Mai - Dezember 2009: Durchführung einer Veranstaltungsreihe zum Thema digitale Archivierung: <http://www.bar.admin.ch/aktuell/00568/00781/index.html?lang=de>
- Vorstudie Archivierung von GIS-Daten (Zusammenarbeit mit Swisstopo; Abschluss im Oktober 2009)



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Schweizerisches Bundesarchiv BAR
Ressort Innovation und Erhaltung

Danke für die Aufmerksamkeit!

