



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of Home Affairs FDHA
Swiss Federal Archives SFA

Akz: 145.24 - SIARD

Save your databases using SIARD!



Chiara Marciani
Roger Kaufmann
Swiss Federal Archives, SFA



What does SIARD mean?

Software Independent Archiving of Relational Databases

- **Nature of data:** databases
- **Type:** relational databases
- **Convert content** into archivable format
- **Detachment** of data from concrete applications



Proprietary database management systems cannot be preserved for the long term





The SFA developed the format SIARD for archiving the contents of relational databases



Database normalisation using SIARD





Emulation - Migration - Normalization

- **Emulation:** Keep a live instance of every version of every database system available for the users of the archive.
- **Migration:** Keep a live instance of every relevant database system as a container of the archived data.
- **Normalization:** Software-independent representation, which preserves the essence and is standardized.



SIARD principles

- **Preserve the information**, not its representation or interaction
- **Preserve the primary data**, not code
- **Preserve the tables and their relations**

PRESERVING FUNCTIONALITY

Constraints

Archived databases presumable are consistent, because they were downloaded from consistent originals. They will never be changed any more, so the constraints are superfluous.

Triggers, Stored Procedures etc.

Not archived if not possible in a standard that runs on every RDBMS



SIARD Format Basics

- **Single ZIP64 File** containing XML and possibly large objects (CLOBs or BLOBs)
- The character set for the XML files and text files is **UTF-8**
- One metadata file describing all objects according to the **SQL:1999 standard**
- Several **XML files** containing the primary data (one file for each table)



SIARD Format vs. SIARD Suite

SIARD Format is a **standardized** and durable **format** based itself on open standards

SIARD Suite is a **reference implementation** of an **application** software which is maintained and updated. Runs on Java (1.6 and above), platform-independent. Currently supports five of the most common database management systems over JDBC or ODBC:





SIARD Format is a swiss e-gov standard

eCH: Association for eGovernment Standards

 **eCH** E-Government Standards

[eCH Share](#) [Sitemap](#) [Agenda](#) [Kontakt](#) DE | FR

[Suchen](#)


[Start](#) [News](#) [Standards](#) [Dokumente](#) [Projekte ffO](#) [Gremien](#) [Über eCH](#)

eCH-0165: SIARD-Formatspezifikation

Version 1.0, Genehmigt, 21.03.2013

Dieses Dokument enthält die Spezifikation des SIARD-Dateiformats. SIARD steht für Software-Independent Archival of Relational Databases. Das Format wurde vom Schweizerischen Bundesarchiv entwickelt. Es handelt sich um eine normative Beschreibung eines Dateiformats für die langfristige Erhaltung von relationalen Datenbanken.

Dokumente

Hauptdokument
 STAN_d_DEF_2013-03-21_eCH-0165_V1.0_SIARD-Format.pdf

Kategorie: Standard

Version	Status	Publiziert
1.0	Genehmigt	21.03.2013

Fachgruppe: digitalearchivierung

Themenbereiche:



SIARD Format ensures comprehensibility, independently from a specific database product

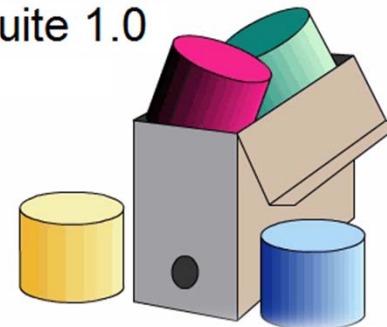


Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Département fédéral de l'intérieur DFI
Dipartimento federale dell'interno DFI
Schweizerisches Bundesarchiv BAR
Archives fédérales suisses AFS
Archivio federale svizzero AFS

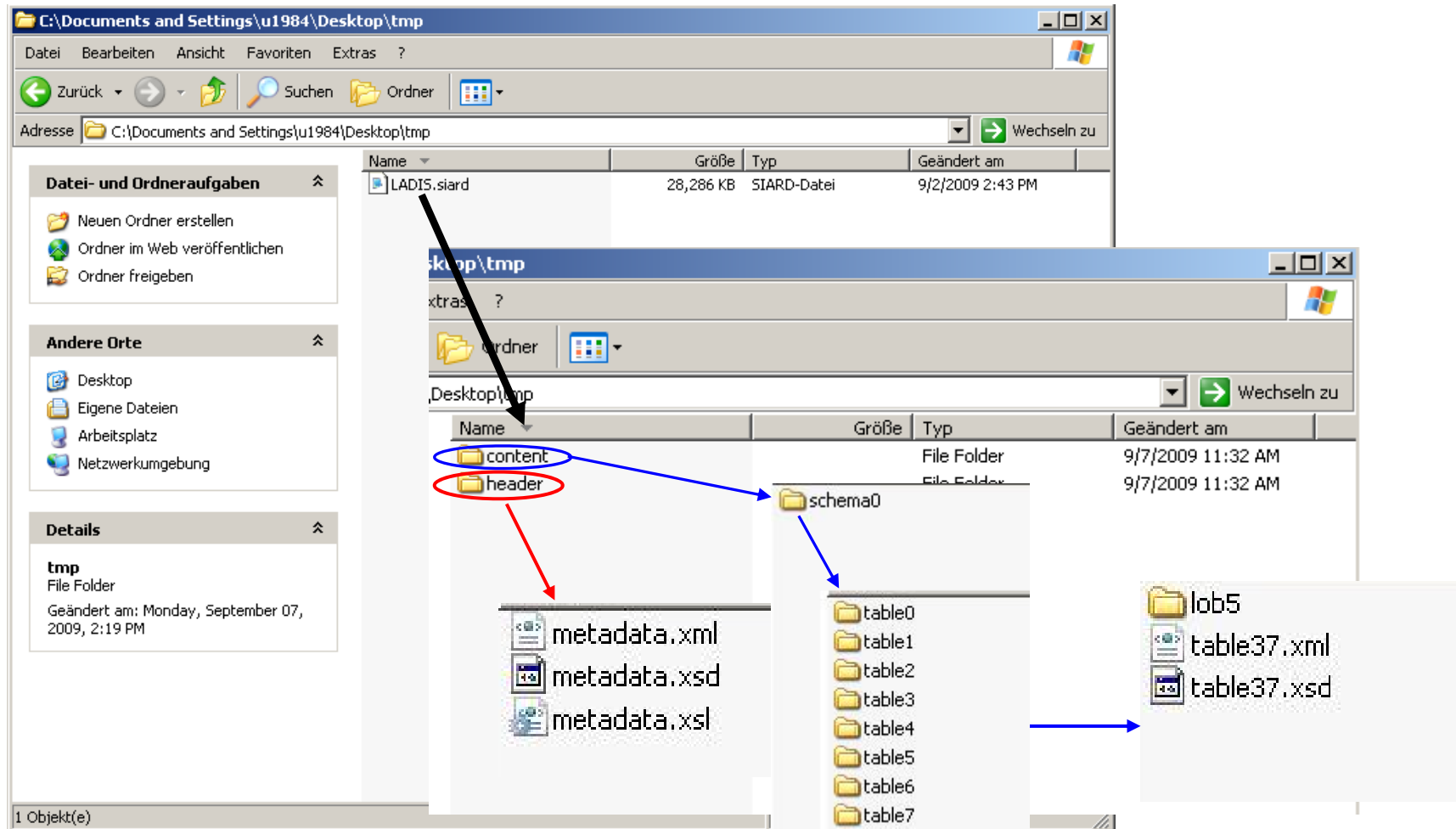
SIARD Suite 1.0

SiardEdit 1.70



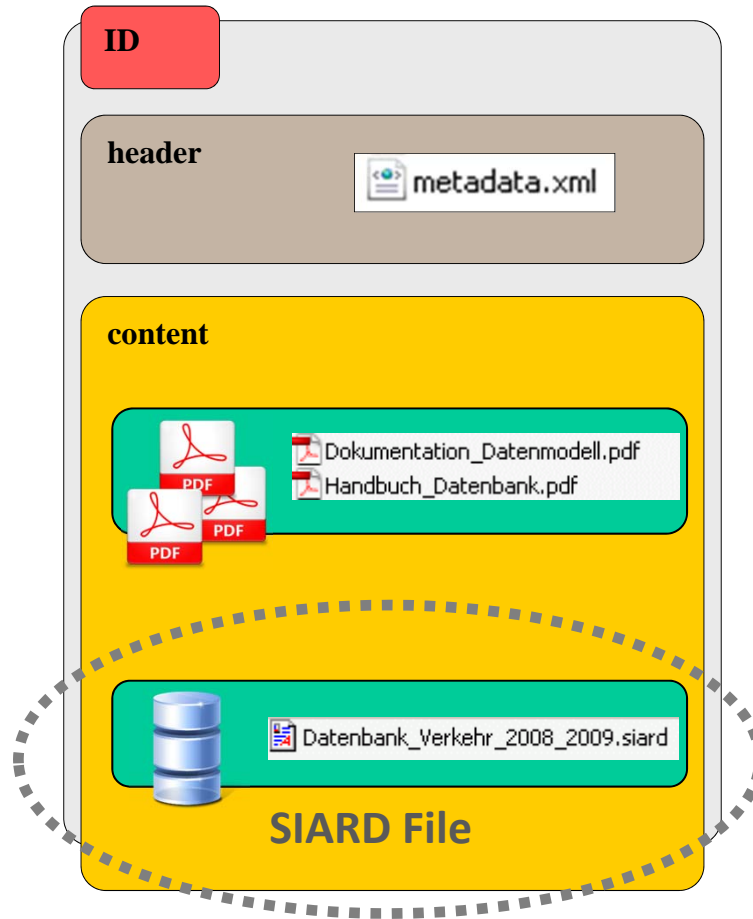


Composition of a SIARD File



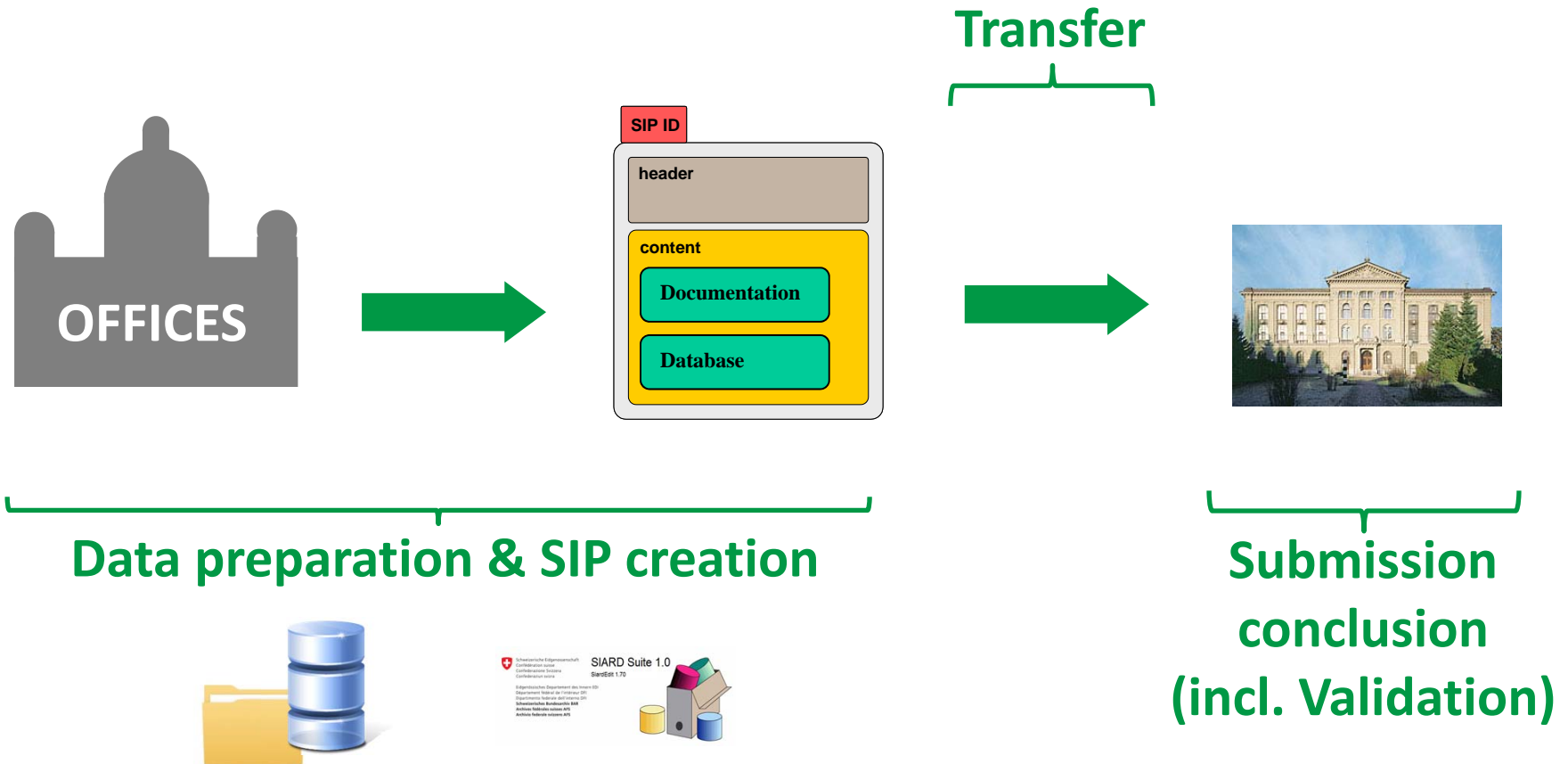


Submission Information Package (SIP)





SFA Submission Procedure





Archival strategies

Important questions prior to archiving

- **RELEVANCE OF DATA**
which tables, all records or sampling or deletions, original schema or more simple archival schema etc.
- **LEGAL ISSUES**
- **TECHNICAL CONSTRAINTS**
- **CONTEXTUAL DOCUMENTATION**

Examples of archival strategies

Yearly deletions, e.g. release from military service
Schema migrations previous to archiving
Preservation of structured flatfiles



Example: Visa Applications

- Archiving of (old) database for Visa applications because of migration to new data model
- SFA deemed archiving of all tables and records necessary because of socio-historic significance

Good contextual documentation is essential!

- ✓ System specification
- ✓ Schema migration description
- ✓ Entity-Relationship-Diagram
- ✓ Description of data, code set
- ✓ Frontend screenshots, manuals etc.



SIARD at the SFA

STATISTICS

- **Amount**
41 databases // 96 SIARD // 120 GB (up to 08.2014)
- **Period**
1960 (flatfiles) – 2014
- **Source**
Statistical Office, Army, Justice, Police and many others
- **Content**
Military Personnel, Wages, Customs Reports, Fruits

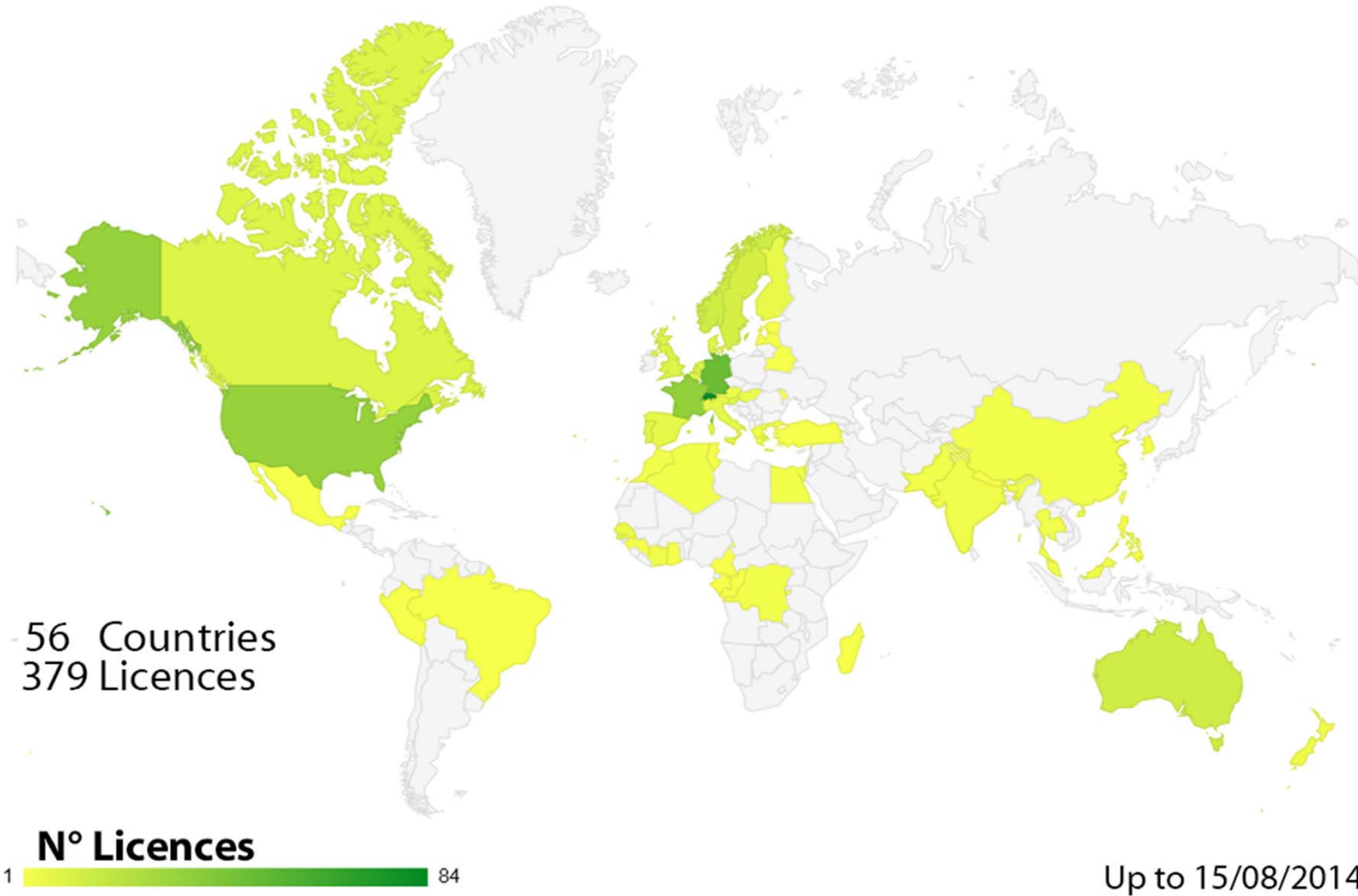


Future of SIARD

- The following changes & additions to the SIARD format are planned:
 - ✓ **Size and splitting:** Recommendation for allowing SIARD ZIP file splitting for very large files
 - ✓ **Data compression:** Permission for simple ZIP data compression (Deflate)
 - ✓ **SQL:1999 Compatibility:** The SIARD specification is to be extended to achieve full SQL:1999 compatibility. User-defined types, row type and array data type should be implemented in SIARD
- Current RFC are listed on the eCH [SIARD Webpage](#)



Who else is using SIARD ?



THANK YOU FOR YOUR
ATTENTION *AND DON'T*
QUESTIONS?
FORGET TO ARCHIVE
YOUR RELATIONAL
DATABASES WITH
SIARD!

