



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of Home Affairs FDHA
Swiss Federal Archives SFA

iPres 2016 - Workshop on Relational Database Preservation Standards and Tools

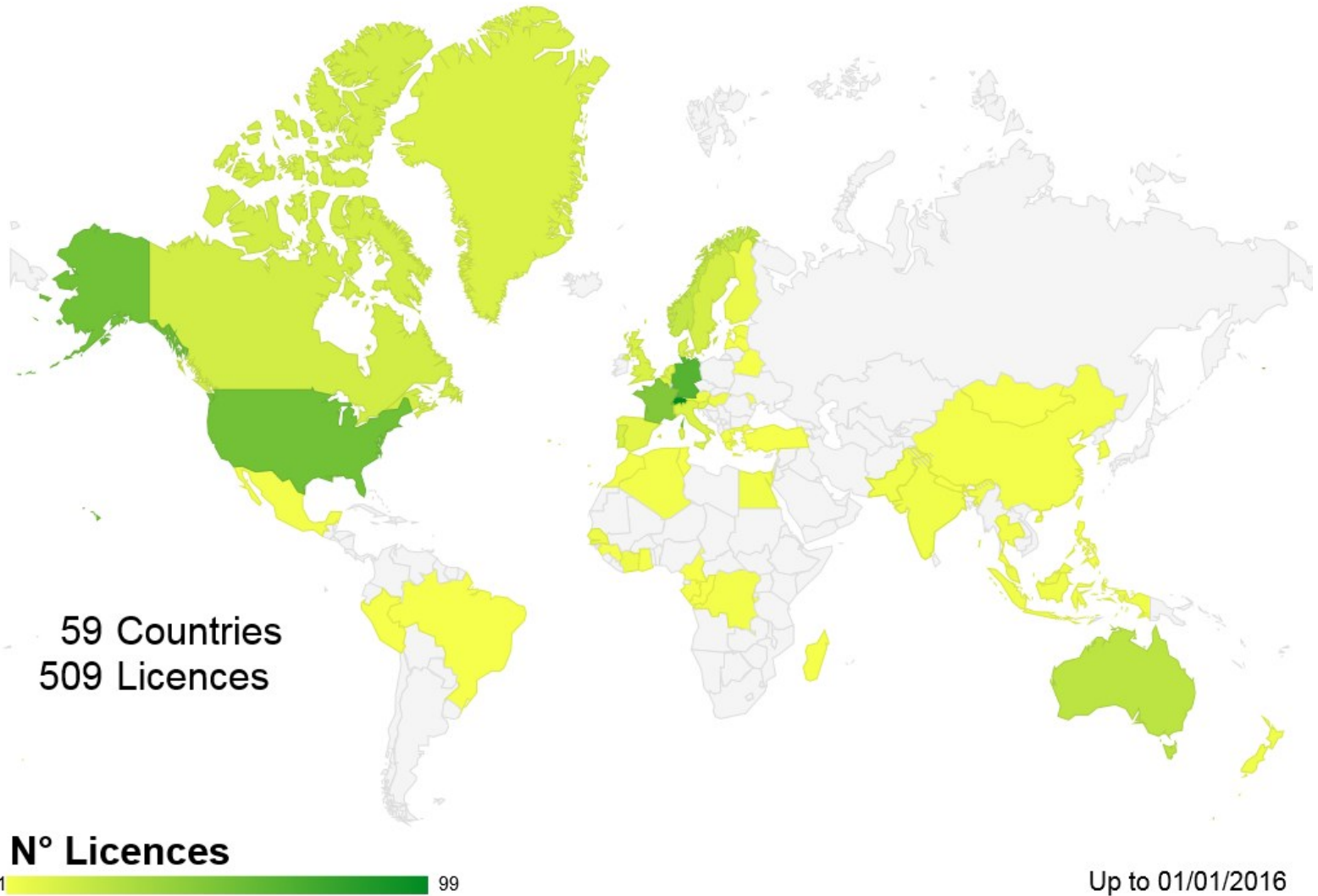
SIARD Format 2.0

06.10.2016

Marcel Bächler <marcel.buechler@bar.admin.ch>



SIARD usage distribution map





SIARD – what does it mean?

Software Independent Archiving of Relational Databases

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



Why is that difficult to archive?

There exist several major database management systems, their storage formats are not compatible.

... And these systems and storage formats evolve over time!

Conclusions:

- we cannot archive proprietary storage formats
- today's storage files are almost certainly not readable with tomorrow's applications



SIARD principles

- Preserve Information, not layout or interaction
- Preserve primary data, not code
- Preserve tables with their relations

Functionality Preservation

Constraints

Archived databases are consistent when they are created from consistent databases. Since, once archived, they will not be changed anymore, preserving constraints is not mandatory.



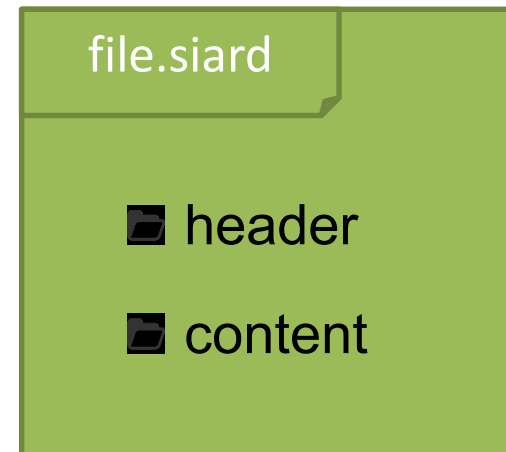
SIARD Format: Technical Details

- The SIARD format saves database-content in a **SIARD-file (SIARD-archive)**
- A SIARD-file is an uncompressed **ZIP-folder (ZIP64)** which contains several XML-files
- There is a single **XML-file** that documents all metadata for the database content, based on **SQL:1999**
- The remaining **XML-files** contain data from the tables (the actual database content)
- The format SIARD is based on **open standards:** SQL:1999, XML, XML Schema, UNICODE



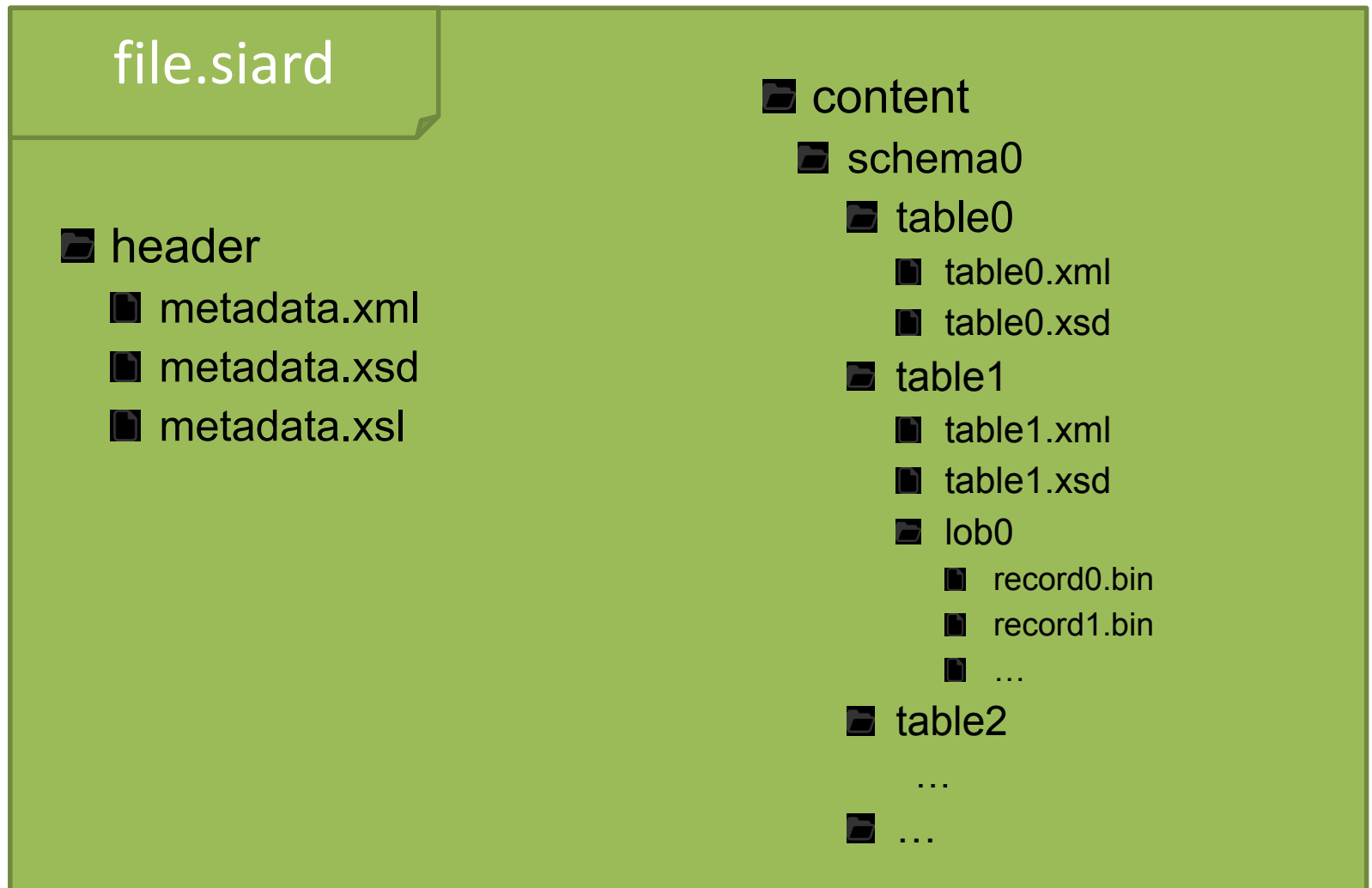
Structure of a SIARD file

- **Primary data** is saved as XML in the **content** folder
 - For each database table a separate XML file is generated
- **Metadata** is saved in the **header** folder within the **metadata.xml** file





Structure of a SIARD file





XML table data example

```
<?xml version="1.0" encoding="utf-8" ?>
- <table xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="http://www.admin.ch/xmlns/siard/1.0/schema0/table0.xsd"
  xsi:schemaLocation="http://www.admin.ch/xmlns/siard/1.0/schema0/table0.xsd table0.xsd">
- <row>
  <c1>1</c1>
  <c2>Boeing</c2>
  <c3>747-400</c3>
  <c4 file="content/schema0/table0/lob4/record0.bin" length="35964" />
</row>
- <row>
  <c1>2</c1>
  <c2>Airbus</c2>
  <c3>A330-300</c3>
  <c4 file="content/schema0/table0/lob4/record1.bin" length="13234" />
</row>
- <row>
  <c1>3</c1>
  <c2>Fokker</c2>
  <c3>50</c3>
  <c4 file="content/schema0/table0/lob4/record2.bin" length="44065" />
</row>
</table>
```

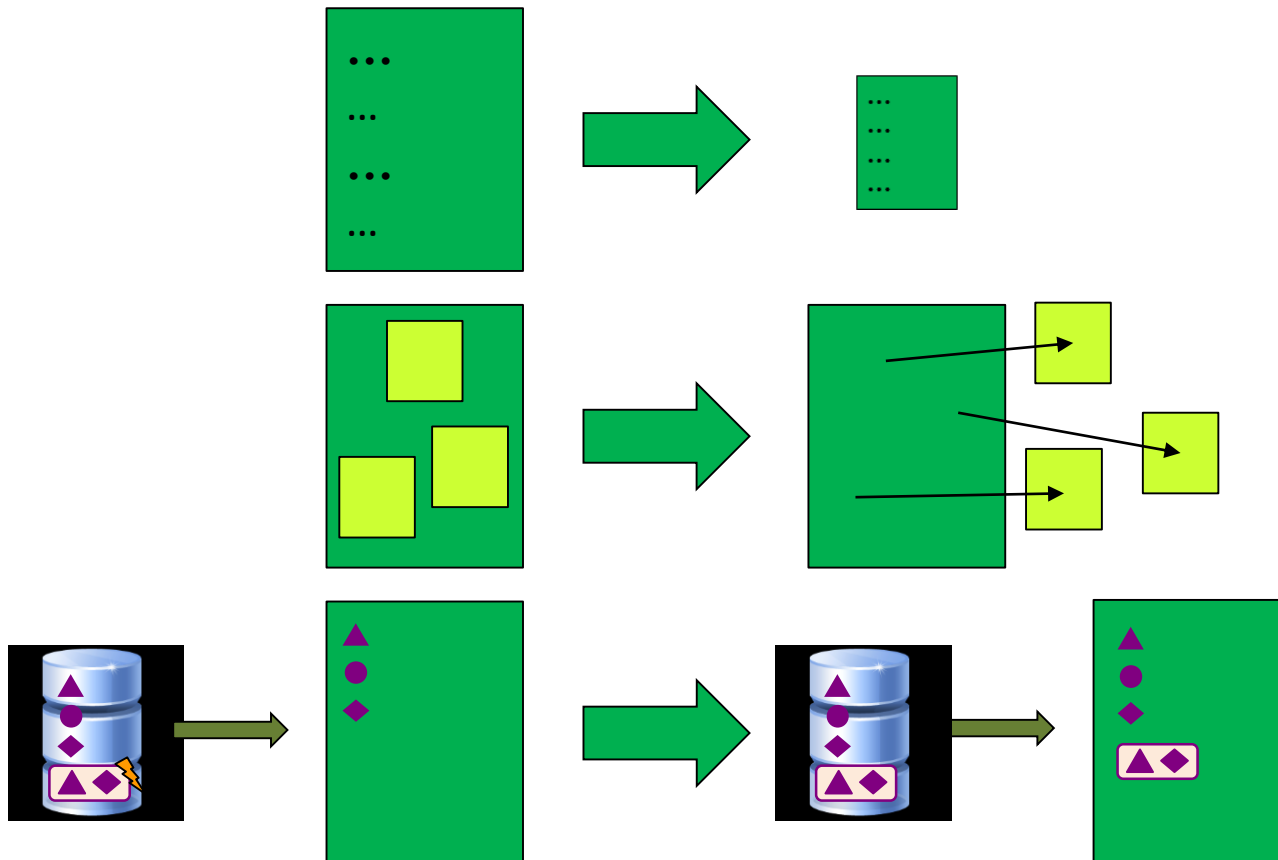


New challenges

- Databases can get huge, which means the SIARD XML Files get huge too
- Everything that lived in the database is stuffed into XML-files, this makes control and preservation harder than necessary
- SIARD Format 1.0 is based on SQL:1999, which, by IT standards, one could call ancient



Our solutions





SIARD 2.0 – Changes

- Upgrade of SQL:1999 to **SQL:2008**
- Support for:
 - all SQL:2008 types, in particular **user-defined data types** (UDTs)
 - **binary or character large objects** (BLOBs and CLOBs) outside of the SIARD file using “file:” URIs
 - **deflate** as a compression mechanism

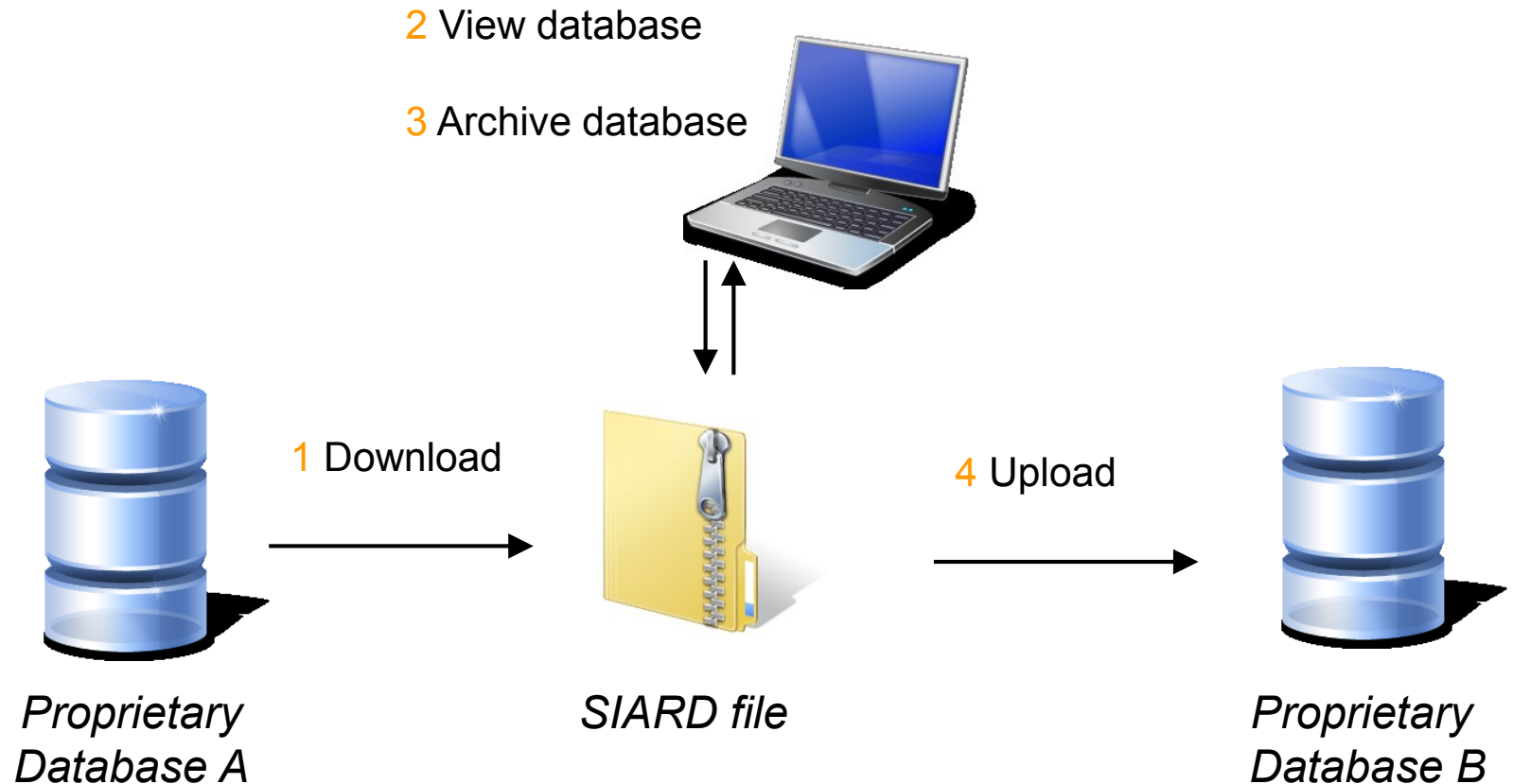


SIARD 2.0 – Benefits

- BLOBs or CLOBs can be stored outside of the XML as self-contained files. They can thus be indexed and accessed directly and more easily migrated to another format. It would also be easier to mandate specific formats (given a policy that requires external storage of binary data).
- Deflate compression minimises (expensive) storage usage.
- With user-defined data types modern databases are supported as they are; no data type conversions are needed when archiving.



Archiving Databases with SIARD





SIARD Format is an Open Standard

- eCH-Standard (eCH-0165: SIARD-Formatspezifikation)

The screenshot shows the eCH website header with the logo and navigation links. The main content area displays the title 'eCH-0165: SIARD-Formatspezifikation' and its version 'Version 1.0, Genehmigt, 21.03.2013'. A descriptive paragraph follows, explaining that SIARD stands for Software-Independent Archival of Relational Databases and was developed by the Swiss Federal Archives. A 'Dokumente' section lists the main document as 'STAN_d_DEF_2013-03-21_eCH-0165_V1.0_SIARD-Format.pdf'. On the right, a sidebar provides additional details: 'Kategorie: Standard', a table with 'Version 1.0', 'Status Genehmigt', and 'Publiziert 21.03.2013', 'Fachgruppe: digitalearchivierung', and 'Themenbereiche:'.

eCH E-Government Standards

eCH Share Sitemap Agenda Kontakt DE | FR

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 Soft-ware-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:

- eARK Specification



Thank you – Merci!

Contact and questions:

Archival Custody and Solutions Services
Swiss Federal Archives SFA

marcel.buechler@bar.admin.ch

www.bar.admin.ch



www.facebook.com/Bundesarchiv.Schweiz



www.twitter.com/CH_Bundesarchiv



www.youtube.com/bundesarchiv