



Archiving Relational Databases with SIARD Suite

Amir Bernstein,
Swiss Federal Archives

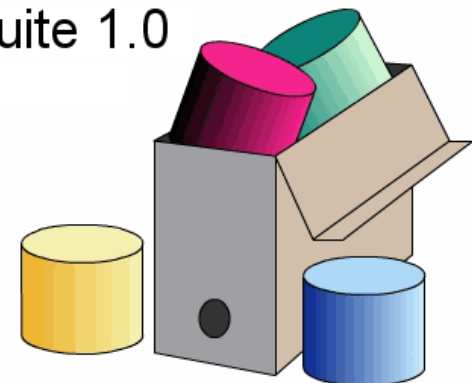


Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

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 1.17



Presentation, Demonstration & Hands-on

- ❑ Relational Databases: a brief introduction
- ❑ Archiving Relational Databases with SIARD
- ❑ Demonstration: SIARD Suite and command-line
- ❑ SIARD Suite hands-on: group exercise



Relational Databases: a Brief Introduction

- Databases, the basics
- Database history, the way to the relational model
- The relational model

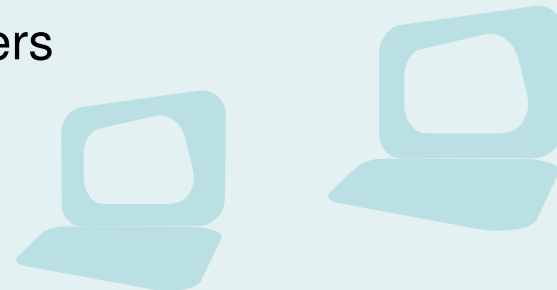


Database: The Basics

Database management system

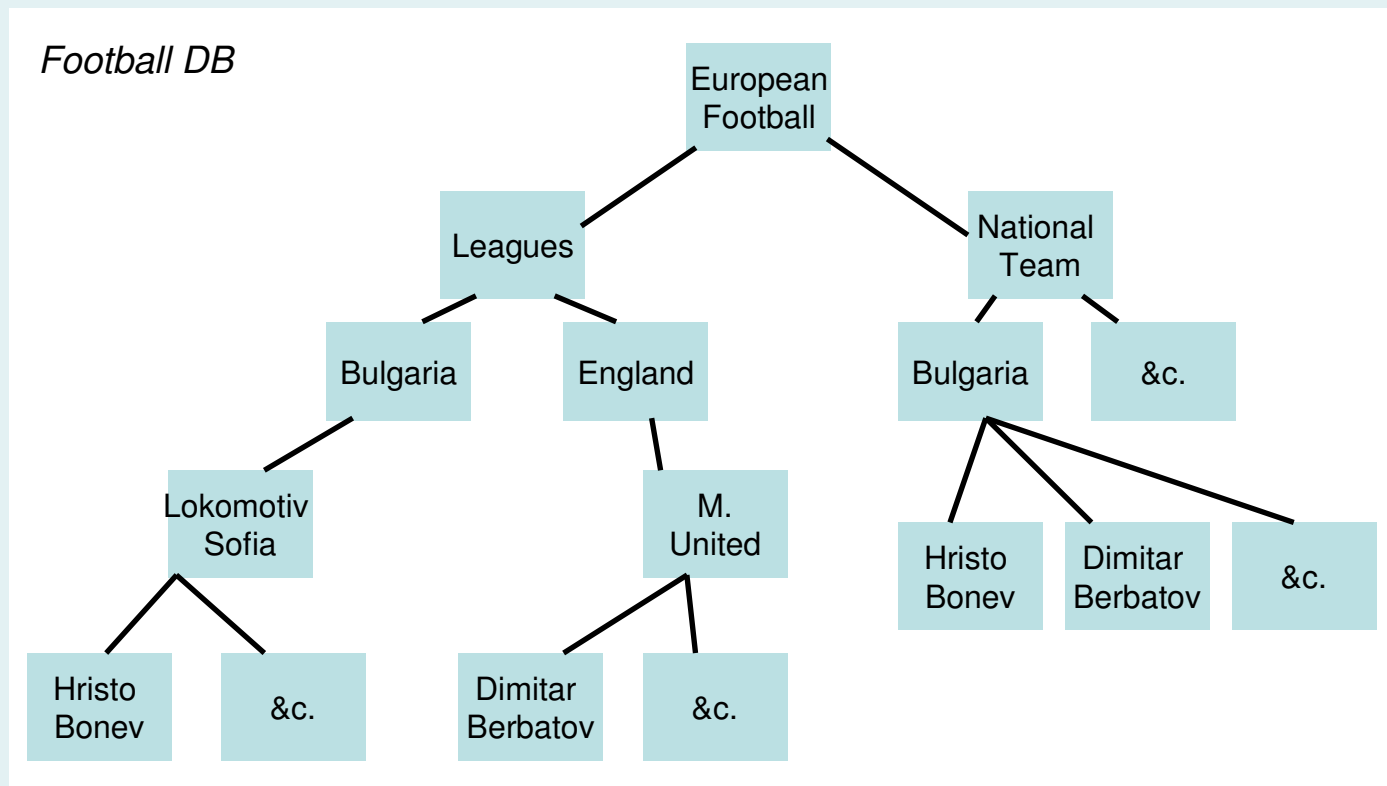


- A repository for a collection of computerized data files
- A database system consists of:
 - data
 - hardware
 - software
 - users



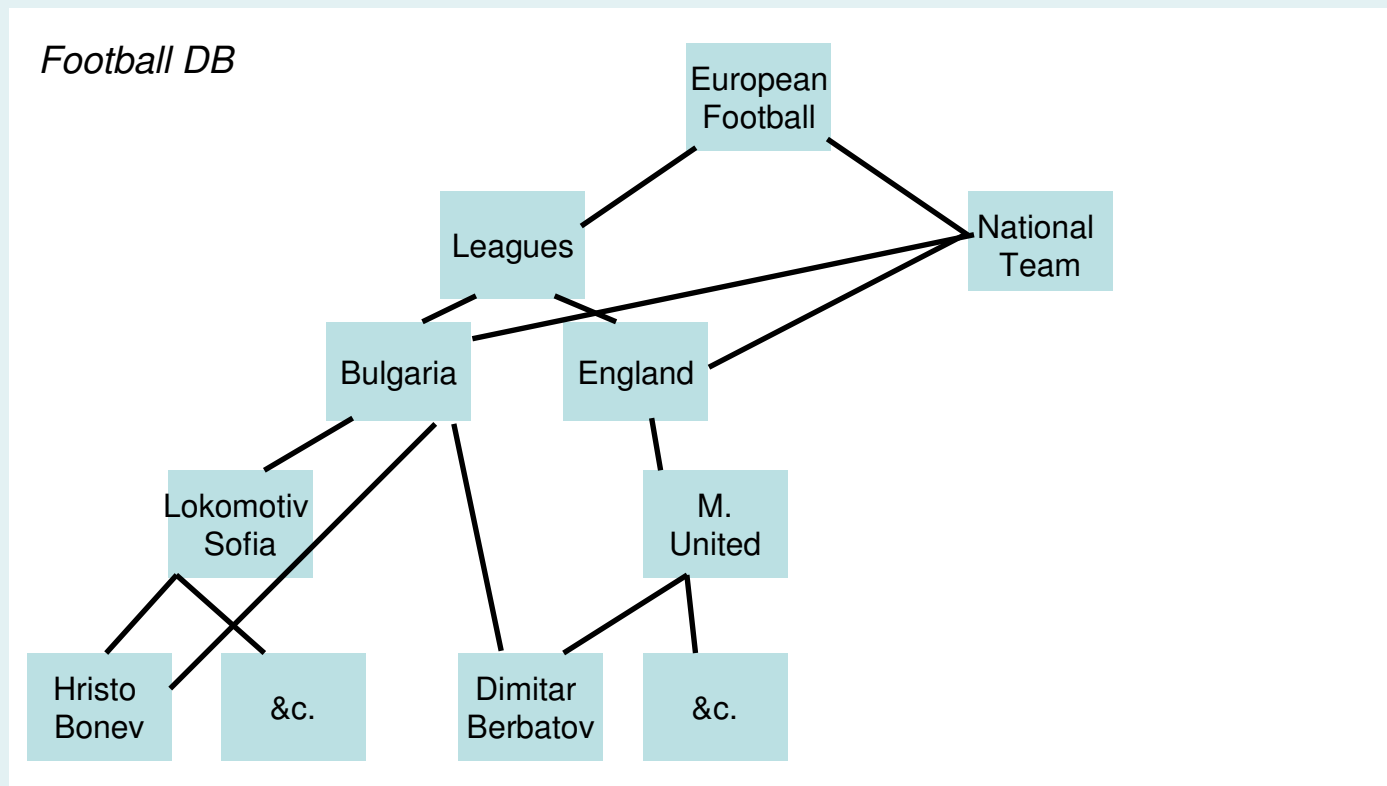
The Hierarchical Model (1960s)

- ❑ 1:1 or 1:n relations
- ❑ Redundancies



The Network Model (1960s)

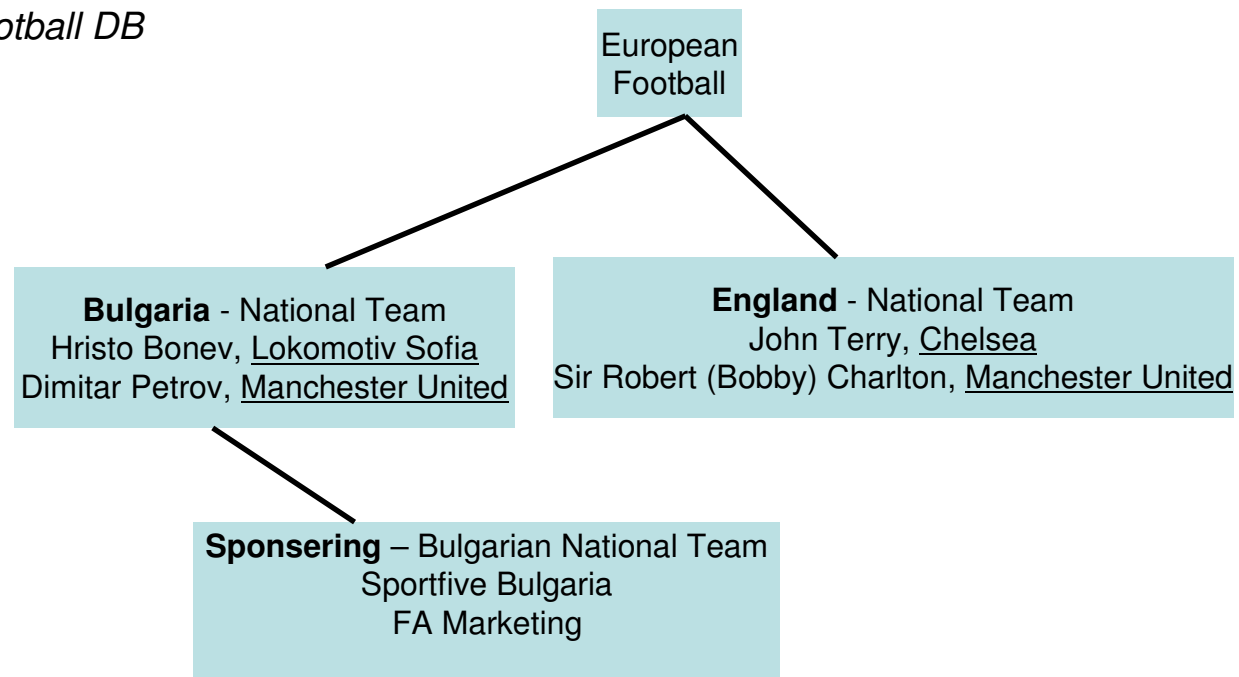
- ❑ No redundancies
- ❑ Complex relations (n:m)



Object-oriented Databases (1980s-1990s)

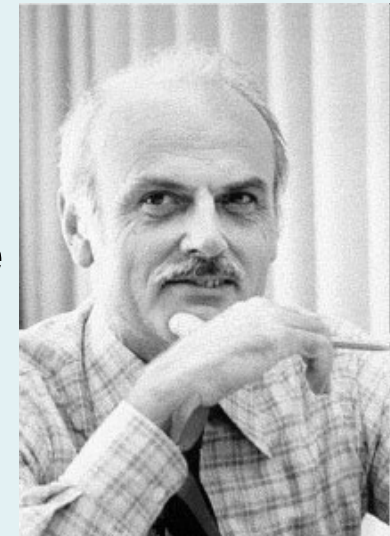
- ❑ Complex objects
- ❑ Code and data stored together

Football DB



The Relational Model (1970s)

- ❑ Introduced by Edgar F. Codd around 1970
- ❑ Basic assumptions:
 - Data have a longer life than software, hardware or systems
 - Data must be independent of software, hardware or systems
 - A query language must be standardized
 - All queries must be treated equally



The Relational Model - Advantages

- ❑ The model disconnects the schema (logical organization) of a database from the physical storage methods
- ❑ It allows the separation of content and media

External Level

User defined views



Conceptual Level

Logical view, „community user view“

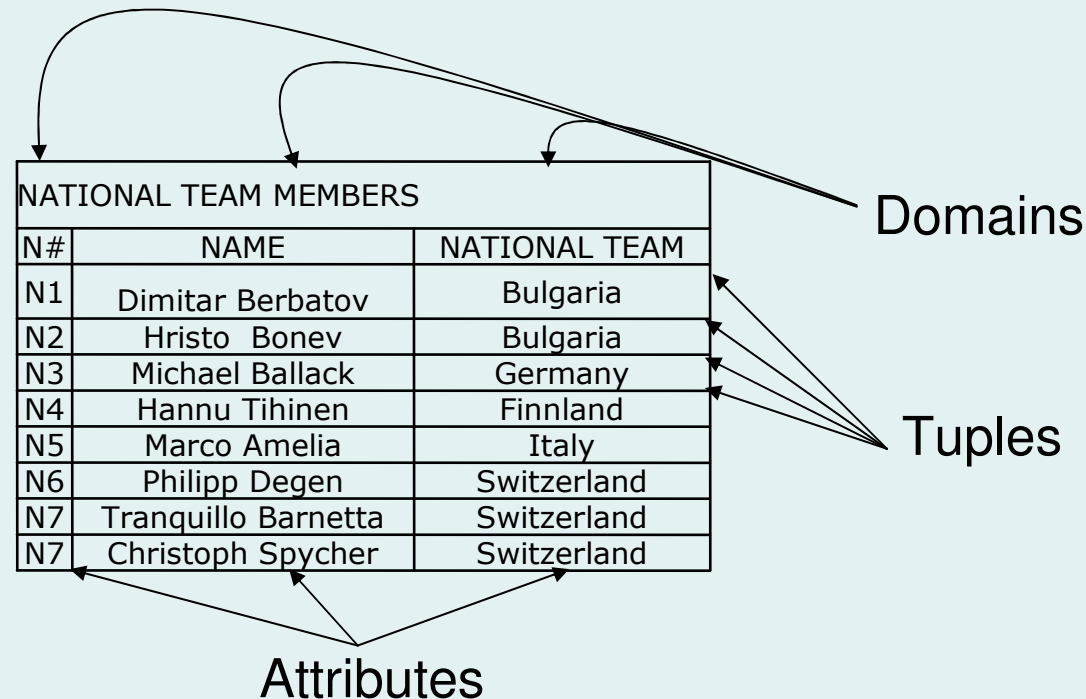
Internal Level

Physical description (blocks & pages), storage view



The Relation Model

- ❑ A simple table structure
- ❑ All information stored in tables

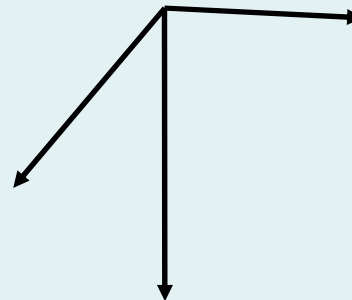


The Base Tables (Entities)

- ❑ Relations instead of redundancies

League	
L1	BVB
L2	Byer Leverkusen
L3	FCZ
L4	Chelsea
L5	Munchester United
L6	Livorno
L7	Lokomotiv Sofia
L8	Eintrach Frankfurt

Base Tables



Player	
P1	Philipp Degen
P2	Primin Schwegler
P3	Hannu Tihinen
P4	Michael Ballack
P5	Dimitar Berbetov
P6	Marco Amelia
P7	Hristo Bonev
P8	Christoph Spycher
P9	Kresimir Stanic

National Team	
N1	Bulgaria
N2	Germany
N3	Finland
N4	Italy
N5	Switzerland



The Relation Tables (Relations)

Player	
P1	Philipp Degen
P2	Primin Schwegler
P3	Hannu Tihinen
P4	Michael Ballack
P5	Dimitar Berbetov
P6	Marco Amelia
P7	Hristo Bonev
P8	Christoph Spycher
P9	Kresimir Stanic

National team player	
P# Player	N# National team
P1	N5
P2	N2
P3	N3
P4	N2
P5	N1
P6	N4
P7	N1
Hristo Bonev / Bulgaria	
P8	N5
P9	N5

National Team	
N1	Bulgaria
N2	Germany
N3	Finland
N4	Italy
N5	Switzerland



Easy Queries

- ❑ All queries are possible
- ❑ Efficient search method

```
SELECT NATIONAL.PLAYER,  
NATIONAL.TEAM AS "NATIONAL TEAM",  
LEAGUE.TEAM as "LEAGUE TEAM"  
FROM NATIONAL, LEAGUE  
WHERE LEAGUE.PLAYER =  
NATIONAL.PLAYER;
```

PNL			
PNL#	Player	National Team	League
PNL1	Hristo Bonev	Bulgaria	Lokomotiv Sofia
PNL2	Dimitar Berbatov	Bulgaria	Manchester United
PNL3	Michael Ballack	Germany	Chelsea
&c...			



Archiving the Relational Model

- ❑ What do we have to archive?
 - At least all tables

- ❑ Attention!
 - Datatypes must be suitable for archiving
 - Database table must be archived in a format suitable for long-term preservation
 - Values in the files must also be suitable for long-term preservation
 - No codes
 - No encryption



The Goal: Preserving the Essence

- ❑ Data (primary & meta) and relations preserved
- ❑ „Look and feel“ is lost



Choosing the right Format

❑ Why format matters...



Know the alphabet
and translate

ursprüng- liche piktogra- phische Schrift Kulturzeit	Frühbabyl- onisch	Assyrisch	ursprüng- liche oder abgeleitete Bedeutung
			Vogel
			Fisch
			Ochse
			Sonne, Tag
			Korn, Getreide
			Garten
			pflügen, ackern
			Baumartung werfen, umwerfen
			stehen, gehen

„Shadrach gave 1 bushel of
barley to the temple...”



Try to read these disks with a modern machine

...10010100100...

Know the alphabet and translate

...23,010273,9300,00005...

See that it's a data base. Know the
language of that data base. Perform
some statements in this language

„At the cbot February 1989, the trade
limit for barley \$0.09 per bushel ...



The SIARD Format

- ❑ **Software Independent Archiving of Relational Databases**
- ❑ SIARD is a universal file format, facilitating
 - SIARD converts database content into a single SIARD file
 - A SIARD file is a ZIP file (ZIP64) containing XML files
 - The SIARD file format is based on open standards: SQL:1999, XML, XML Schema, UNICODE, ...



The SIARD Archive

❑ Primary data

- “content” folder with:
 - Folder for each table
 - All tables in xml format
 - LOB folders

```
<?xml version="1.0" encoding="utf-8" ?>
<table xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://
xsi:schemaLocation="http://www.admin.ch/xmlns/siard/1.0/schema0/table
- <row>
  <c1>1</c1>
  <c2>-1</c2>
  <c3>1</c3>
  <c4>1</c4>
  <c5 file="content/schema0/table37/lob5/record0.bin" length="108277" />
  <c6>759</c6>
  <c7>480</c7>
  <c8>738900</c8>
  <c9>548</c9>
  <c10>98500</c10>
  <c11>53</c11>
  <c12>489700</c12>
  <c13>22</c13>
  <c14>280800</c14>
  <c15>440</c15>
  <c16>479277</c16>
  <c17>299642</c17>
  <c18>838864</c18>
  <c19>73534</c19>
</row>
- <row>
```

❑ Metadata

- “metadata” folder with:
 - One XML file (metadata.xml)
 - Includes all metadata from all levels

test

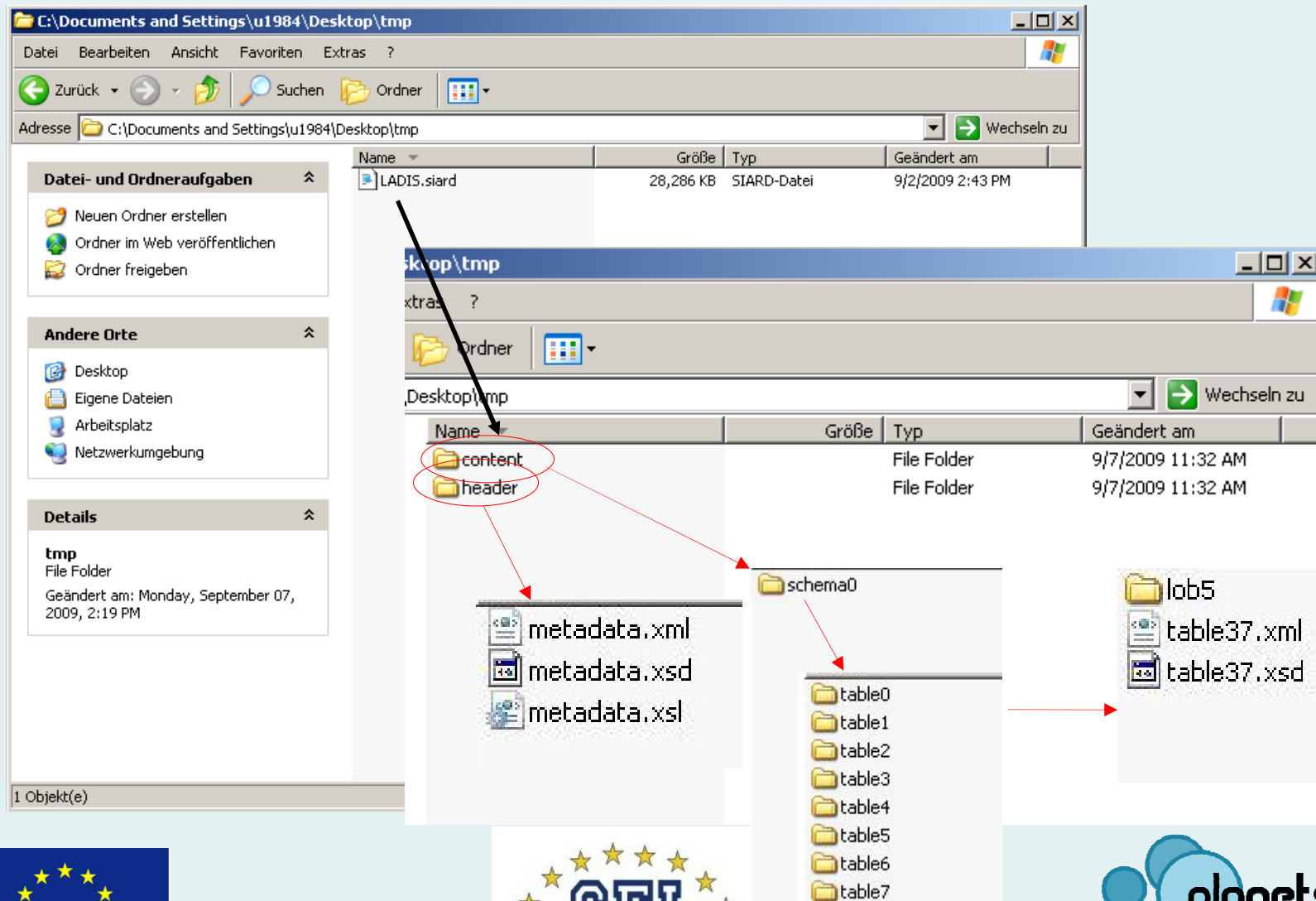
Name	test
Version	1.0
Description	
Archiver	
Archiver Contact	
Data owner	(...)
Data origin timespan	(...)
Archival date	2009-09-02
Message digest	MD51A45C60C5C17C814F07EA06FE5747ECD
Client machine	edixp1348.edi.intra.admin.ch
Producing application	
Database product	Oracle Oracle9i Enterprise Edition Release 9.2.
Connection	
Data base user	DIAS

Table of contents

- Schemas
 - [DIAS](#)
- Tables
 - [ABTEILUNGSTYP](#)



The SIARD Archive in a Glance:

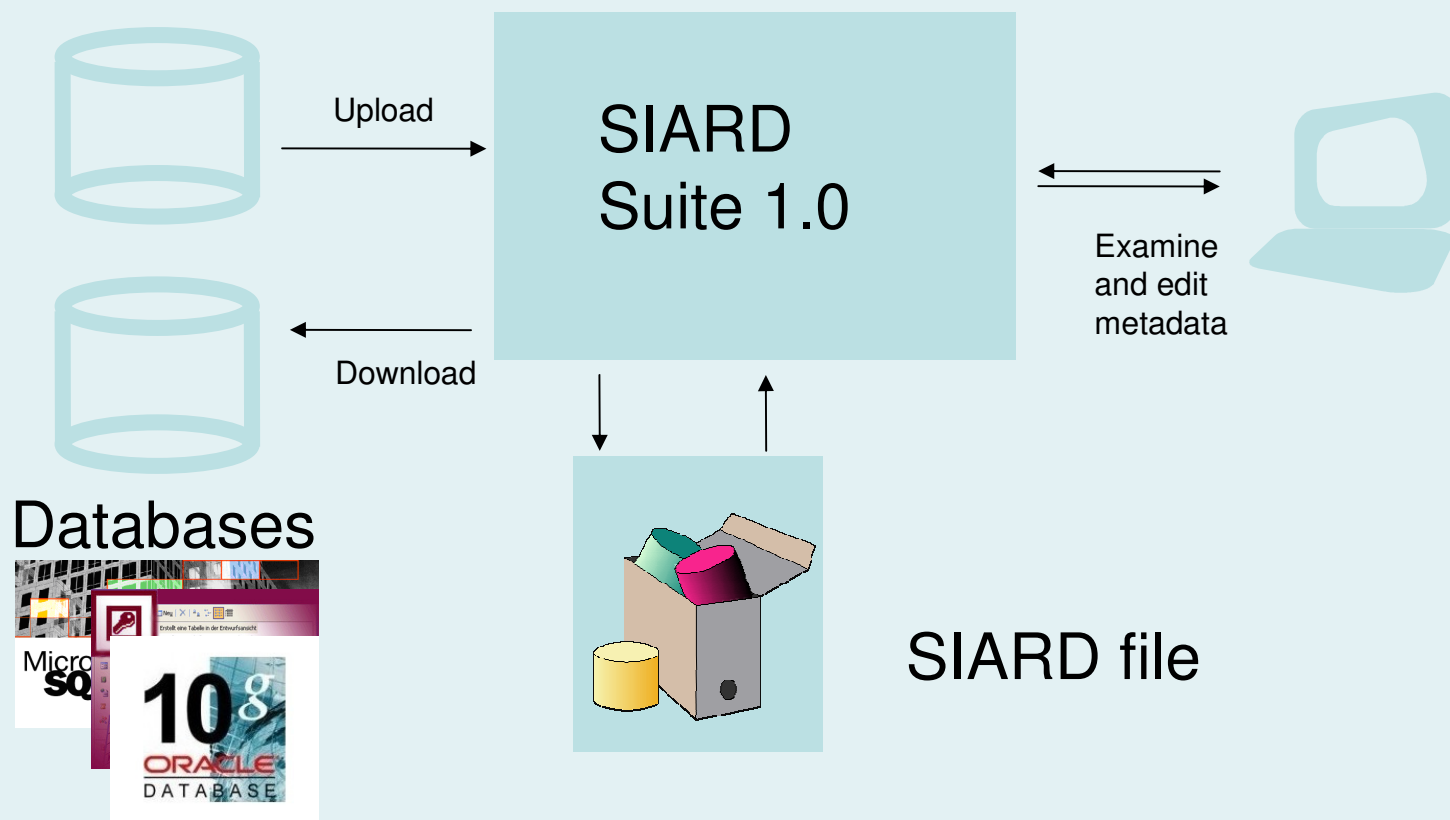


SIARD Archive – an Open Format

- ❑ Official Planets format for archiving databases
- ❑ Can be used free of charge
- ❑ Downloadable for the SFA website



The SIARD Suite



Prerequisites

- ❑ SIARD is platform independent
 - It operates in a JAVA environment (Java SE 1.5 or higher)
- ❑ SIARD can run on a single computer with a common GUI
- ❑ Installation
 - Click & install
 - or direct use from a USB stick



The SIARD Suite Components

❑ **SiardEdit**

- Edit your metadata
- Create a SIARD-Archive with a new set of metadata
- Match your metadata against those of a different archive
- Update and complete your existing set of metadata
- View and sort your primary data

❑ **SiardFromDb**

- Convert your database into a SIARD-Archive
- Create a full SIARD-Archive (with both metadata and primary data in the SIARD format), or:
- Generate an empty SIARD-Archive (i.e. containing no primary data)

❑ **SiardToDb**

- Facilitate your research within a given database
- Load your SIARD-Archive into a database instance (with tables, views etc.)
- Comfortably navigate and search within your database



SIARD Demonstration

- ❑ A stroll through a SIARD Archive (LADIS)
 - Using SIARD Edit
 - BLOBs in SIARD
- ❑ Archiving an Oracle DB with SIARD
- ❑ What's inside? A look at a SIARD file
- ❑ ODBC connection and archiving a local MDB



SIARD – Hands-on!

- ❑ Four work groups
 - Archiving a database with SIARD (local / server-based)
 - Upload a SIARD archive into a database instance
- ❑ Rapporteurs
 - Your opinion on SIARD Suite



Exercise I – Create a SIARD Archive

- ❑ Launch SIARD Suite
- ❑ Download an Oracle database (cf. the following page)
- ❑ Navigate through the Data base using the SIARD Suite Editor
- ❑ Try to:
 - Add metadata
 - Edit the primary the data
 - Find the added meta data
 - Retrieve data to an Excel Sheet
- ❑ Please report to the plenary session



Exercise I – Create a SIARD Archive

- ❑ Database password: crm

Load SIARD File from Database

Connection type: Choose database connection type

Connection Info: (DNS-Name):(Port):(SID) of the Oracle database (e.g. dbhost.enternet.ch:1521:SIARD)

Database user: Userid for the database connection

Database password: Password for the database connection

Query timeout limit: Query timeout limit

Load metadata only? ☐ Select to load an empty database

Recent connections

Connection type	Database	Database user
ODBC	crm	admin
Oracle	dbhost.enternet.ch:1521:SIARD1	crm
Oracle	dbhost.enternet.ch:1521:LADIS	DIAS
ODBC	accounting	admin
Oracle	dbhost.enternet.ch:1521:SIARD1	accounting
SQL Server	dbhost.enternet.ch:1433/ACCOUNTING	ACCOUNTING
ODBC	test	admin
ODBC	case study	admin
ODBC	si	admin



Exercise II – Create a SIARD Archive

- ❑ Download an Access database
 - Use the database „crm“ provided on the USB stick (folder: databases)
 - Create a ODBC connection (remember the connection name)
 - Create a SIARD archive using the ODBC connection you have defined
- ❑ Navigate through the Data base using the SIARD Suite Editor
- ❑ Try to:
 - Add metadata
 - Edit the primary the data
 - Find the added meta data
 - Retrieve data to an Excel Sheet
- ❑ Please report to the plenary session



Exercise III – SIARD Archive to DB

- ❑ Download an Access database
 - Locate the “accounting.siard” archive provided on the USB stick (folder: databases)
 - Create a new empty Access Database
 - Ensure you have read and write rights in this database
 - Create a ODBC connection for the database (remember the connection name)
 - Launch SIARD Suite.
 - Open the ***accounting.siard***
 - Upload the SIARD archive into your empty access databases using the ODBC connection you have created

- ❑ Navigate through the Data base using MS Access



Exercise III – SIARD Archive to DB

- ❑ Try to:
 - Add metadata
 - Edit the primary the data
 - Find the added meta data
 - Retrieve data to an Excel Sheet

- ❑ Please report to the plenary session



Any Questions?

- ❑ For further information please contact the Swiss Federal Archives:

For SIARD:

Amir.Bernstein@bar.admin.ch



Thank you ! / Благодаря!

