



## **Archiving Relational Databases with SIARD Suite**

**Amir Bernstein,**

**Swiss Federal Archives**



# What is it all about?

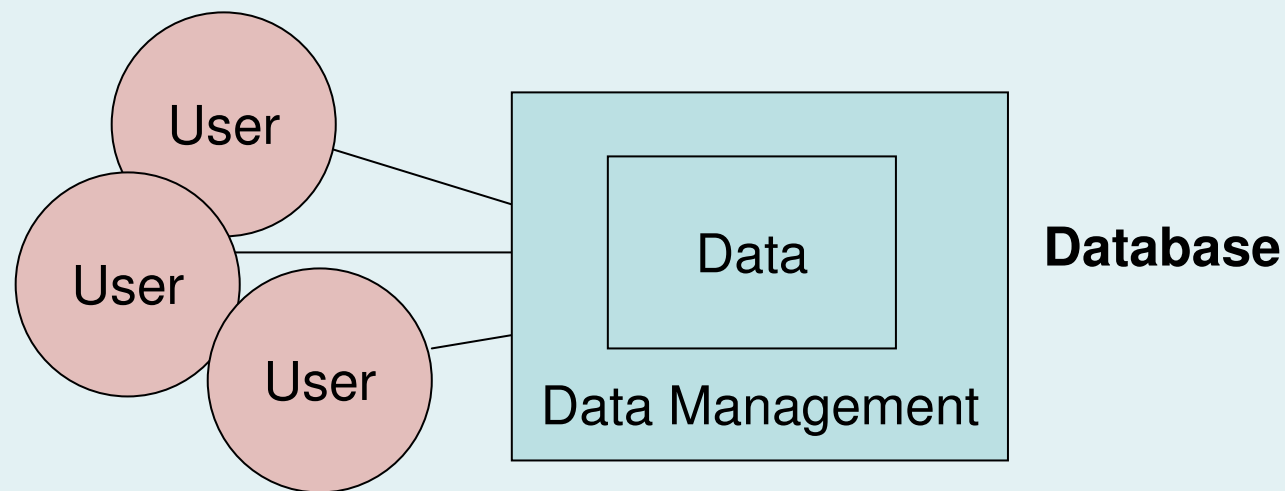
---

- Relational Databases at a glance
- A format suitable for long-term preservation / archiving
- Archiving of Relational Databases
- Demo – SIARD Suite 1.19



# Databases

---



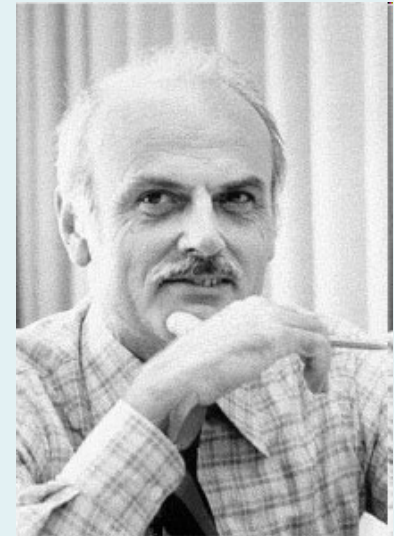
- Data: structured and permanent storage of records
- User: add, edit, curate the data
- Management: organize the data, regulate access



# The Relational Model

---

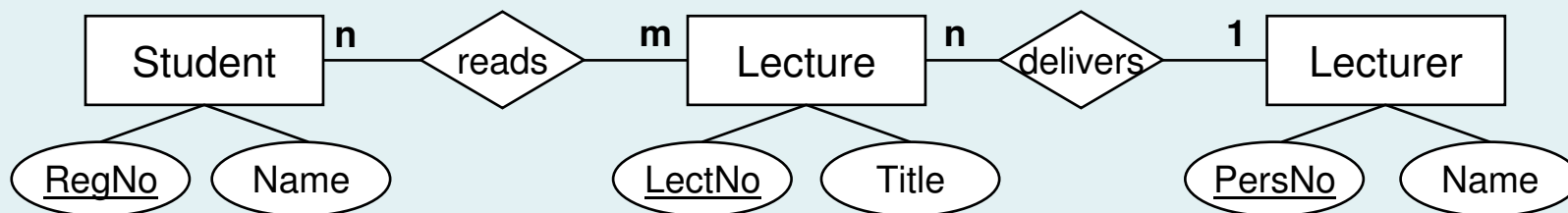
- Introduced by Edgar F. Codd around 1970
- Basic assumptions:
  - Data have a longer life than software, hardware or systems
  - Data must be independent of software, hardware or systems
  - A query language must be standardized
  - All queries must be treated equally



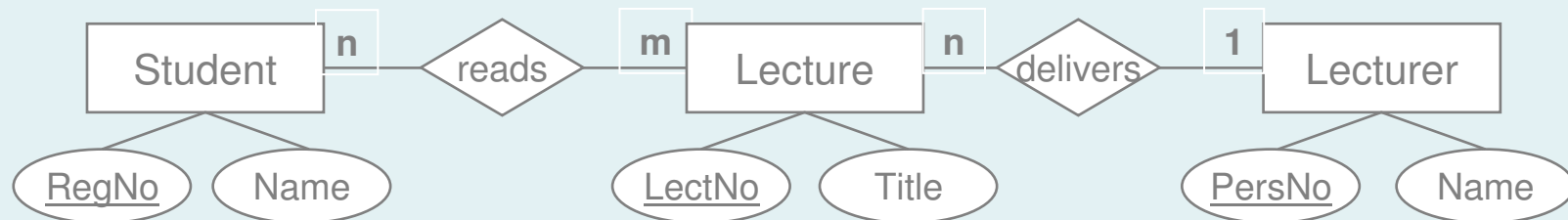
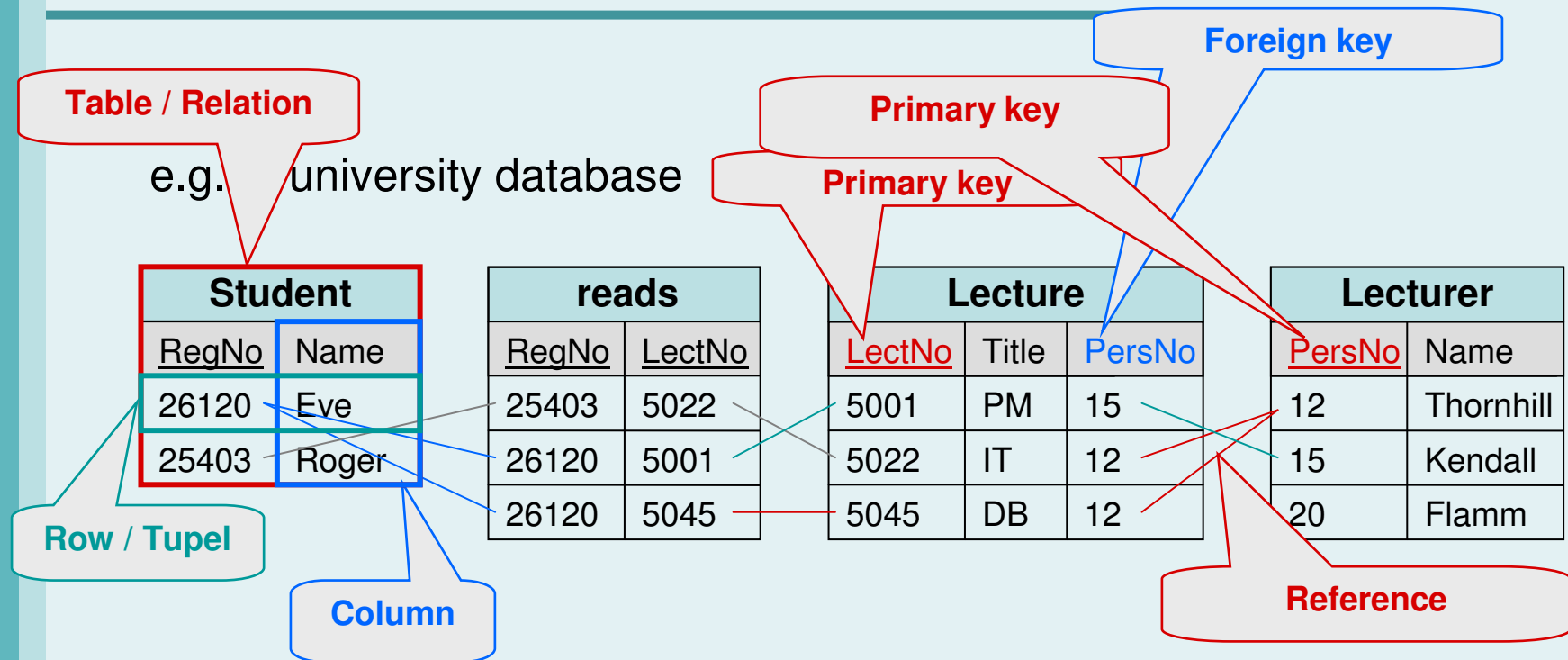
# Relational Databases

e.g.: a university database

Student		reads		Lecture			Lecturer	
<u>RegNo</u>	Name	<u>RegNo</u>	<u>LectNo</u>	<u>LectNo</u>	Title	PersNo	<u>PersNo</u>	Name
26120	Eve	25403	5022	5001	PM	15	12	Thornhill
25403	Roger	26120	5001	5022	IT	12	15	Kendall
		26120	5045	5045	DB	12	20	Flamm



# Relational Databases



# Today's Databases

---

- Mostly **relational** (over 90%)
- Other database types...
  - Hierarchical databases
  - Network databases
  - Object oriented databases (OODB)are less common



# Archiving Relational Databases

---

## Goal:

- Long-term **preservation of the content**
- Ensure that we **understand** the content
- ... independent of any specific database product



# What to preserve?

---

- Only the essence...
  - Data (primary & meta) and relations
  - „Look and feel“ is lost



# How to preserve...

- Choose the right format



Know the alphabet  
and translate

Letter Name	Trans-literation	Early	Middle	Late	Modern
Aleph	A	א	א	א	א
Beyt	B	ב	ב	ב	ב
Gimel	G	ג	ג	ג	ג
Dalet	D	ד	ד	ד	ד
Hey	H	ה	ה	ה	ה
Vav	V	ו	ו	ו	ו
Zayin	Z	ז	ז	ז	ז
Chet	Hh	ח	ח	ח	ח
Tet	Th	ט	ט	ט	ט
Yud	Y	י	י	י	י
Kaph	K	כ	כ	כ	כ
Lamed	L	ל	ל	ל	ל
Mem	M	מ	מ	מ	מ
Nun	N	נ	נ	נ	נ
Samech	S	ס	ס	ס	ס
Ayin	Gh	ע	ע	ע	ע
Pey	P	פ	פ	פ	פ
Tsade	Ts	צ	צ	צ	צ
Quph	Q	ק	ק	ק	ק
Resh	R	ר	ר	ר	ר
Shin	Sh	ש	ש	ש	ש
Tav	T	ת	ת	ת	ת



Try to read these disks with a modern machine

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



## SIARD format – the essentials...

---

- A collection of XML files
- The XML files contain all the information from the archived database (e.g. tables, LOB, relations, &c.)
- Key element is the use of an open and published format, based on international standards



## SIARD format – in detail...

---

- SIARD format stores database content in a single [SIARD file / SIARD archive](#)
- A SIARD file is an uncompressed [ZIP container \(ZIP64\)](#), containing XML files
- A single [XML file](#) documents all the metadata from the database (based on SQL:1999)
- The database content (table content &c) is stored in [XML files](#)
- The SIARD format is based on [open standards](#): SQL:1999, XML, XML Schema, UNICODE



# Archiving with SIARD

---

Archiving Relational Databases with SIARD means:

- **Retain the relations:** archive all relevant tables from the database as a single unit
- **Retain the tables:** convert the fields' content in a format suitable for long-term preservation
- Further requirements:
  - No codes in the fields (e.g. Java script)
  - No encrypted content
  - The data type (in each column) must be suitable for long-term preservation



# Primary and Metadata in SIARD

- **Primary data** is stored in the **content** folder in an XML format.
  - For each database table SIARD SUITE will generate automatically a separate XML file
- **Metadata** is stored in a single file, in an XML format: **metadata.xml**

```
<?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>
```

test

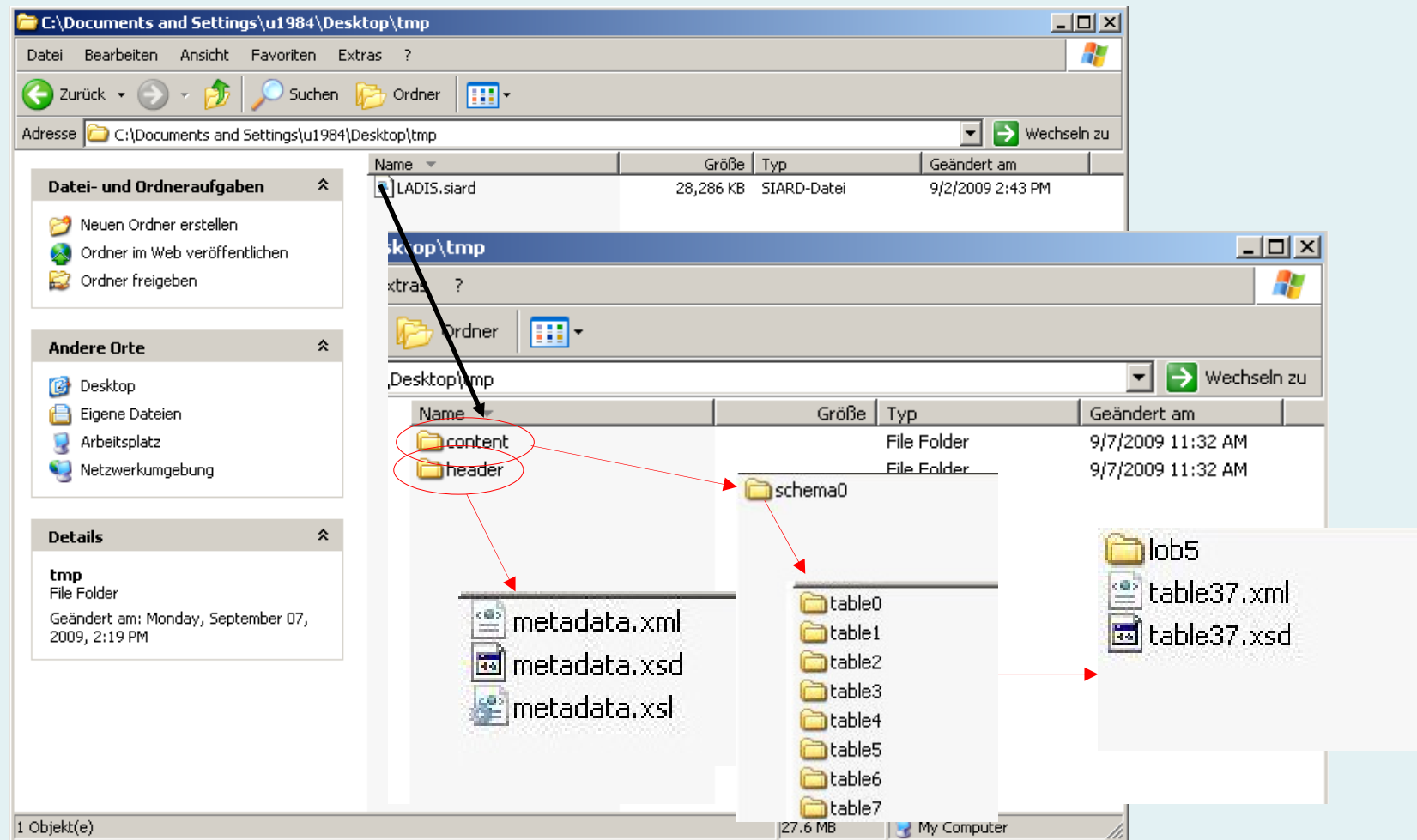
Name	test
Version	1.0
Description	
Archiver	
Archiver Contact	
Data owner	(...)
Data origin timespan	(...)
Archival date	2009-09-02
Message digest	MD51A45C60C5C17C814F07EA06FE5747ECD
Client machine	edixp1348.edi.intra.admin.ch
Producing application	
Database product	Oracle:Oracle9i Enterprise Edition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

Table of contents

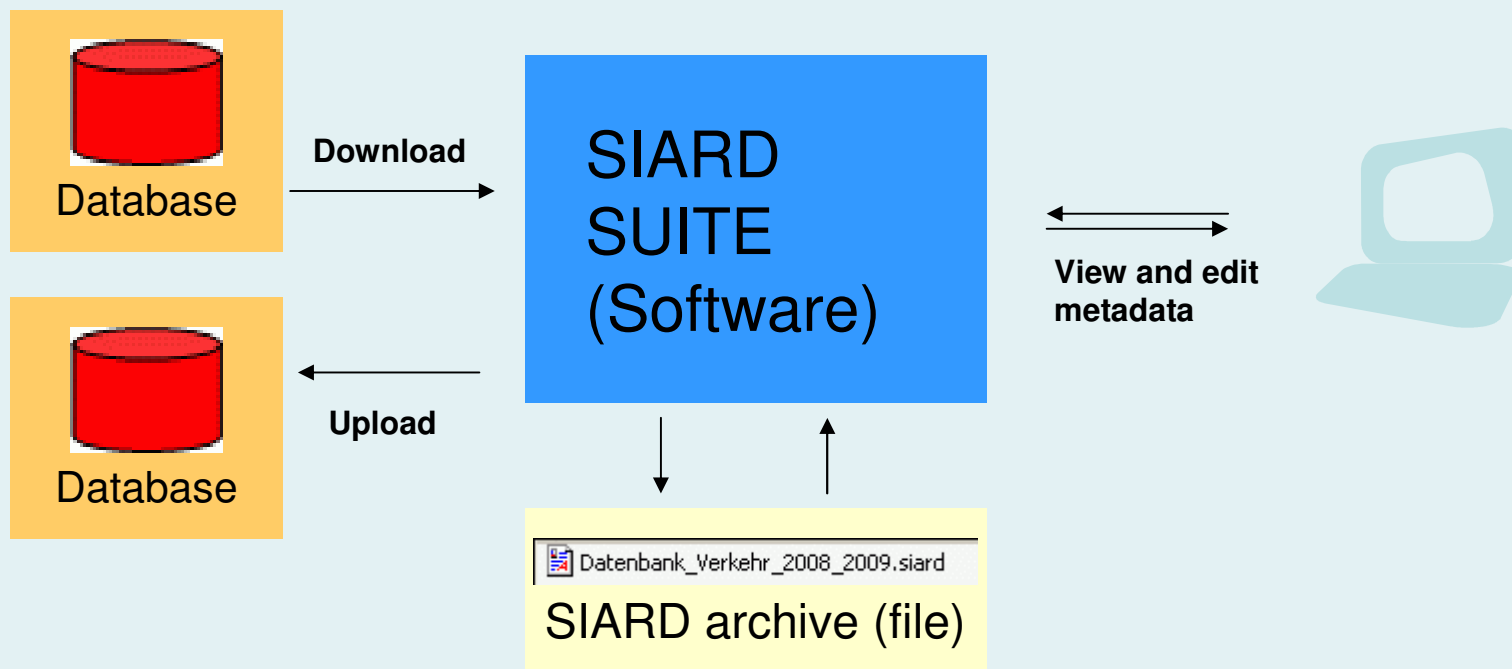
- Schemas
  - [DIAS](#)
- Tables
  - [ABTEILUNGSTYP](#)



# A SIARD archive at a glance



# Archiving a Database with SIARD SUITE



# Prerequisites for SIARD SUITE

---

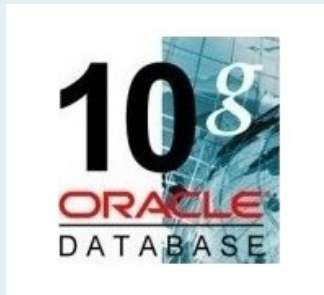
- SIARD SUITE is **platform independent**
- Runs in a **JAVA** environment (min. Java 1.5): Windows, Linux, Mac...
- **Installation:** SIARD SUITE can be used directly from a USB stick or installed per mouse click



# Databases supported by SIARD SUITE

- SIARD Format and SIARD SUITE support at present the following database products:

Oracle, MS SQL, MS Access



## Getting your Hands on the Software...

---

- SIARD command-line is integrated in the Planets framework
- SIARD format is an open format
- SIARD Suite
  - Downloadable from the SFA website ([www.bar.admin.ch](http://www.bar.admin.ch))
  - Can be used free of charge



# SIARD Demonstration

---

- Archiving a MS SQL DB with SIARD
- LOBs in SIARD
- What's inside? A look at a SIARD archive
- Uploading a SIARD archive into a DB



# Thank you for your attention !

---

Amir.Bernstein@bar.admin.ch

