



Database preservation

DPC training course

Introduction and basic features

Day 1, morning

Trainers: Luis Faria, Miguel Guimarães

Agenda for this morning

- 10:00 Welcome and introductions *by Jenny*
- 10:20 Database preservation archival workflows *by Luís*
- 10:45 Introduction to the SIARD format *by Luís*
- 11:10 Break
- 11:30 Tools for database preservation *by Miguel*
- 12:00 Case study: Testing SIARD 2.0 *by Brett Abrams, NARA*
- 12:30 Questions and discussion
- 13:00 Lunch

The background of the slide is an abstract composition of numerous vertical lines of varying thicknesses. These lines are rendered in a gradient of colors, ranging from deep black on the left to bright red on the right, with intermediate shades of dark red and maroon. The lines appear to have a slight 3D effect, with some showing highlights and shadows, giving them a sense of depth and movement. The overall effect is a dense, textured field of vertical strokes.

Database preservation archival workflows

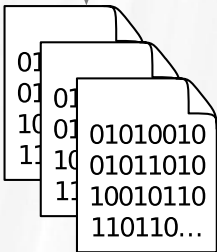
Databases

The **information** that supports institutions and businesses is usually **centralized on databases**. This information is of **great value** and needs to be **preserved for decades** due to strategic and legal reasons.

Databases

The systems that have this information are usually complex with **many software components** playing their part for supporting the **business-logic**, and the **submission** and **presentation** interfaces.

The information is usually laid out in an **organization specifically optimized for the database** and original business objectives, i.e. **not in a user-friendly** organization.



A screenshot of a 'Contact' application window. The window title is 'Contact'. It contains a form for 'Agee Software' with fields for 'First Name: Allen', 'Last: Agee', 'Billing Street: 710 Bluebonnet Dr', 'City: Allen', 'State: TX', 'Zip: 75002-4435', 'Work Phone: (972) 390-9018', 'Fax: (972) 390-8620', 'Email: a@ageesw.com', 'Website: www.ageesw.com', and 'Comment: Maintains this program'. There are also checkboxes for 'Client', 'Contact', 'Inactive', and 'ID'. The 'Created' date is '5/12/01 7:40:39 PM' and the 'Salesperson' is 'Allen Agee'. The window has a menu bar with 'File', 'Edit', 'View', 'Tools', 'Help', and 'Window'. The status bar at the bottom shows '4 Recs' and buttons for 'Del', 'New', 'Env', 'Bot', 'Help', 'Exit', 'Print', 'RTF', and 'TXT'.

Application

person			
<u>id</u>	name	birth	city_id
1	Mary	1986-03-28	2
2	Phillip	1974-11-08	3
3	Alison	1991-06-10	5
4	Barry	1979-09-14	2

Cell

Row

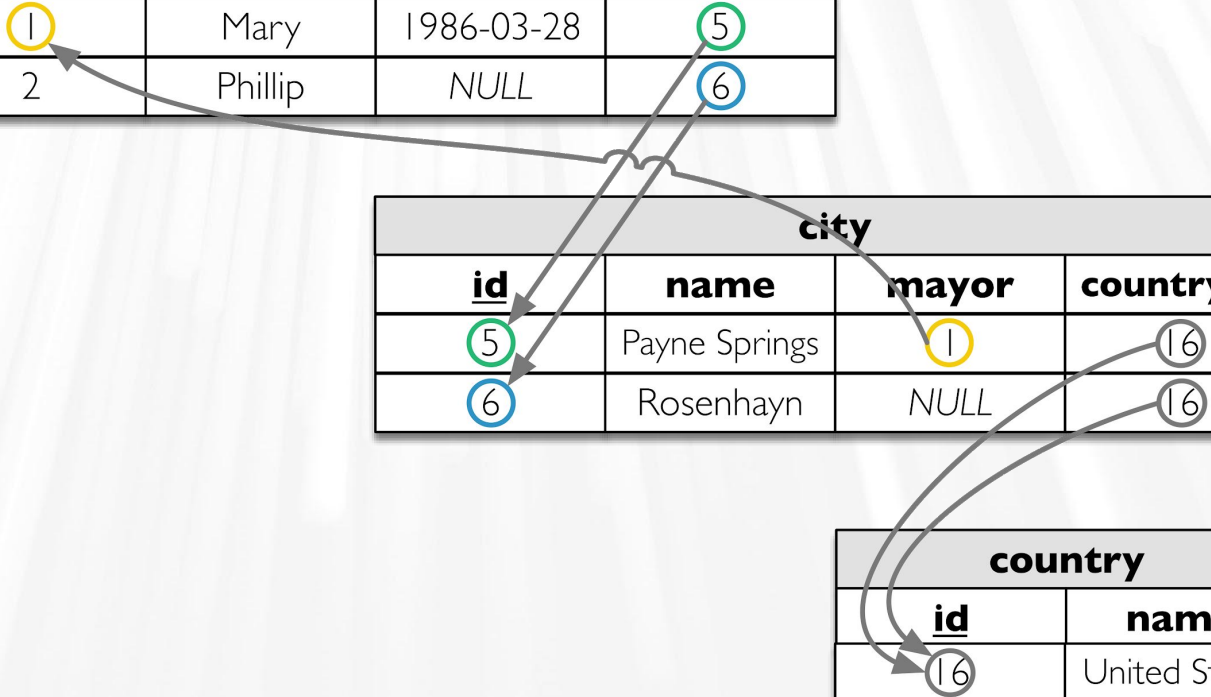
Column

The diagram illustrates a database table structure. The table is titled 'person' and has four columns: 'id', 'name', 'birth', and 'city_id'. The 'id' column is underlined, indicating it is a primary key. The table contains four rows of data. Annotations with arrows point to specific parts of the table: 'Cell' points to the 'birth' cell of the first row, 'Row' points to the entire first row, and 'Column' points to the 'birth' column.

person			
<u>id</u>	name	birth	city_id
1	Mary	1986-03-28	5
2	Phillip	NULL	6

city			
<u>id</u>	name	mayor	country_id
5	Payne Springs	1	16
6	Rosenhayn	NULL	16

country	
<u>id</u>	name
16	United States



Information in a relational database

Information in tables

Column data types

Relations and constraints

Projections (views)

Behaviour (triggers and routines)

Other (users, permissions, etc.)

System information not in the relational database

External resources (e.g. files in filesystem)

Submission forms

Presentation interfaces

Application logic and queries

Preservation strategies

Hardware and software museums

Emulation

File format migration

Encapsulation

Hardware and software museums

Preserve the whole technology stack needed to render the original content.

⊕ reproduction accuracy	⊖ great difficulty to maintain
	⊖ restrictions on the access to information
	⊖ need for users to understand how to operate long gone systems

Emulation

Use of a software system that allows to emulate the behaviour of an older hardware and/or software platform within a newer one.

⊕ reproduction accuracy	⊖ difficult to maintain
⊕ no need to maintain hardware	⊖ difficult to set up
	⊖ need for users to understand how to operate long gone systems

File format migration

Transfer of digital information from one hardware and software configuration into another.

Convert information encoded in a file format, tied into an obsolete technology stack, into another more current or better suited for long term preservation.

⊕ easier to use and reuse information	⊖ possible data loss during conversion (can be mitigated by quality assurance)
⊕ no need to maintain hardware	⊖ might need to migrate again in the future
⊕ no need to maintain software	

Encapsulation

Keep files together with all necessary documentation needed for future development of emulators, file format migrators or software renderers.

⊕ postpone actions that can be costly

⊕ no need to maintain hardware

⊖ may hinder timely access to information

⊖ difficult to gather documentation of complex or closed file formats

⊖ difficult to ensure quality and completeness without hindsight

The problem with databases

Every vendor has his data types and export formats

It is rare that information exported from one vendor's system works on another

Sometimes doesn't work on different versions of the same product

Need for a vendor-agnostic format based on standards

Preservation format criteria

Ubiquity	Stability	Complexity
Support	Ease of identification and validation	Interoperability
Disclosure	Intellectual Property Rights	Viability
Documentation quality	Metadata support	Re-usability

SIARD: Software Independent Archiving of Relational Databases

Database preservation format

More details later on

Based on international standards

For database data, structure and behaviour

Swiss national standard eCH-0165

Now managed by DILCIS board and the EU eArchiving building block

<https://dilcis.eu/content-types/siard>

<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eArchiving>

Simple database archive flow

Producer

Archive with DBPTK Enterprise



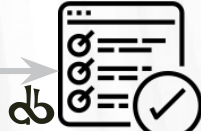
Database



Related files



Preservation
format



SIARD format
validation



Simple
catalogue

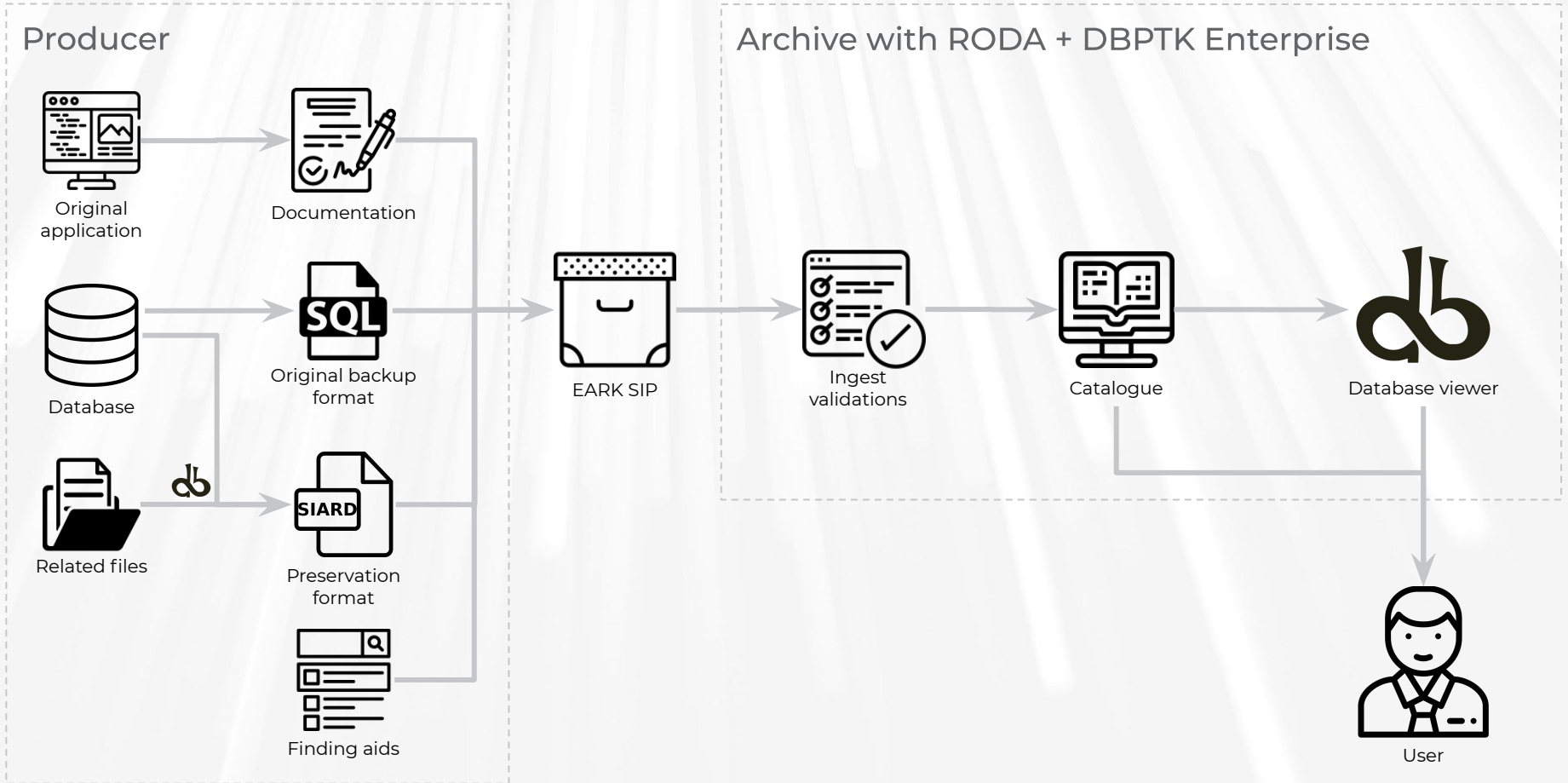


Database viewer



User

Full database archival flow



Preserving what is not in SIARD

Use original tools to create **backups in original backup format** (migration)

Complete **hard disk** or virtual machine **snapshot** (enable future emulation)

Image and video **recordings** of submission and presentation **interfaces** (documentation)

Interview original developers, operators or users (documentation)

Expert analysis report (documentation)

How to select the strategy?

User community **requirements**

Budget

Staff technical **skills**

NOTE: a different strategy can be taken for each part of the information.

Significant properties

The characteristics of digital objects that must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects, and their capacity to be accepted as evidence of what they purport to record.

The **significance** of properties **may be different** for each stakeholder, therefore their definition must be done in light of the **institutional objectives**.

Significant properties of a database

Property	Significant?	Preservation strategy
Information in tables	?	File format migration
Column data types	?	File format migration
Relations and constraints	?	File format migration
Projections (views)	?	File format migration, encapsulation
Behaviour (triggers and routines)	?	Encapsulation
Other (users, permissions, etc.)	?	Encapsulation
External resources (e.g. files in filesystem)	?	File format migration
Submission forms	?	Encapsulation
Presentation interfaces	?	Encapsulation
Application logic and queries	?	Encapsulation

Authenticity

Capability to prove (or vouch) that the digital object is according to the original.

The **credibility of the digital object** authenticity is endowed by the **trustworthiness** of the **digital repository and the institution** that supports it. This trustworthiness is a consequence of the institution **honourability** and **credibility** and is further improved on the repository by having **transparency on the mission, policies and procedures** in place for digital preservation, being **rigorous on their application** and being able to **prove, based on evidence**, that the defined policies and procedures are correctly followed.

Preservation policy and planning

Processes that define the **mission, drivers, objectives** and decision-making **processes** for the preservation of digital information so that **goals are met** with **minimal** operational **costs** and **maximal** (expected) content **value**.

1. Determine general and specific **needs**
2. Establish **priorities**
3. Identify **resources** for implementation
4. Define course of **action** and agenda
5. Document **compromises** (e.g. actions not taken)

The background of the slide is an abstract composition of numerous vertical lines of varying thicknesses. These lines are rendered in a gradient of colors, ranging from dark, almost black, on the left to a vibrant, saturated red on the right. The lines are slightly blurred, giving a sense of depth and movement. The overall effect is a dynamic, textured backdrop.

Introduction to the SIARD format

SIARD 2

Open format for archival of relational databases

Software Independent Archiving of Relational Databases

Joint effort from Swiss Federal Archives, E-ARK and eCH (eGovernment standards in Switzerland)

Based on **international standards**

Unicode, SQL:2008, XML, XML Schema, URI

Swiss national **standard** eCH-0165

Now managed by DILCIS board and the EU eArchiving building block

SIARD principles

Preserves information, not layout or interaction

Application and business-logic are not preserved

Preserves primary data, not code

Stored procedures, functions and other database features are documented by not transformed

Preserve **tables** with their **relations**, not dynamic data

Views are not preserved in SIARD, only documented

But views can be materialized as tables

Technical details

Data archived into a **single compressed file** (ZIP)

Optionally, it can also be in a folder structure

Folder **“header”** as information on database structure, data types, and behaviour

Kept in a single XML file

Folder **“content”** has the data from tables

One XML file per table

Binary objects can be stored in different ways

Inline, inside, outside

Internal structure

database.siard

```
├── header/
│   ├── siardversion/
│   │   └── 2.1/
│   ├── metadata.xml
│   └── metadata.xsd
└── content/
    ├── schema1/
    │   ├── table1
    │   │   ├── table1.xml
    │   │   ├── table1.xsd
    │   │   └── lob5
    │   │       └── record1.bin
    │   └── table2
    │       ├── table2.xml
    │       └── table2.xsd
```

Internal structure

database.siard

```
├── header/
│   ├── siardversion/
│   │   └── 2.1/
│   ├── metadata.xml
│   └── metadata.xsd
├── content/
│   └── schema1/
│       ├── table1
│       │   ├── table1.xml
│       │   ├── table1.xsd
│       │   └── lob5
│       │       └── record1.bin
│       └── table2
│           ├── table2.xml
│           └── table2.xsd
```

Version statement, so it can be easily detected by file format identifiers

Internal structure

database.siard

```
├── header/  
│   ├── siardversion/  
│   │   └── 2.1/  
│   ├── metadata.xml  
│   └── metadata.xsd  
└── content/  
    ├── schema1/  
    │   ├── table1  
    │   │   ├── table1.xml  
    │   │   ├── table1.xsd  
    │   │   └── lob5  
    │   │       └── record1.bin  
    │   └── table2  
    │       ├── table2.xml  
    │       └── table2.xsd
```

Information about the structure and behaviour of the database

Internal structure

database.siard

```
├── header/
│   ├── siardversion/
│   │   └── 2.1/
│   ├── metadata.xml
│   └── metadata.xsd
├── content/
│   └── schema1/
│       ├── table1
│       │   ├── table1.xml
│       │   ├── table1.xsd
│       │   └── lob5
│       │       └── record1.bin
│       └── table2
│           ├── table2.xml
│           └── table2.xsd
```

Embed XML schema to validate the metadata XML.

Internal structure

database.siard

```
├── header/
│   ├── siardversion/
│   │   └── 2.1/
│   ├── metadata.xml
│   └── metadata.xsd
├── content/
│   └── schema1/
│       ├── table1
│       │   ├── table1.xml
│       │   ├── table1.xsd
│       │   ├── lob5
│       │   │   └── record1.bin
│       └── table2
│           ├── table2.xml
│           └── table2.xsd
```

Table content

Internal structure

database.siard

```
├── header/
│   ├── siardversion/
│   │   └── 2.1/
│   ├── metadata.xml
│   └── metadata.xsd
├── content/
│   └── schema1/
│       ├── table1
│       │   ├── table1.xml
│       │   ├── table1.xsd
│       │   ├── lob5
│       │   │   └── record1.b
│       └── table2
│           ├── table2.xml
│           └── table2.xsd
```

XML schema to validate table content
according to column data types

Internal structure

database.siard

```
├── header/
│   ├── siardversion/
│   │   └── 2.1/
│   ├── metadata.xml
│   └── metadata.xsd
├── content/
│   └── schema1/
│       ├── table1
│       │   ├── table1.xml
│       │   ├── table1.xsd
│       │   └── lob5
│       │       └── record1.bin
│       └── table2
│           ├── table2.xml
│           └── table2.xsd
```

Binary and large text objects that can be kept inside the archive

```
<siardArchive xmlns="http://www.bar.admin.ch/xmlns/siard/2/metadata.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
version="2.1"
xsi:schemaLocation="http://www.bar.admin.ch/xmlns/siard/2/metadata.xsd">
  <dbname>sakila</dbname>
  <description>Example database from MySQL</description>
  <archiver>Luis Faria</archiver>
  <archiverContact>lfaria@keep.pt</archiverContact>
  <dataOwner>MySQL Team</dataOwner>
  <dataOriginTimespan>2000-2010</dataOriginTimespan>
  <lobFolder>content</lobFolder>
  <producerApplication>
    Database Preservation Toolkit (version 2.8.2)
  </producerApplication>
  <archivalDate>2020-07-16+00:00</archivalDate>
  <clientMachine>unspecified</clientMachine>
  <databaseProduct>MySQL 5.6.49</databaseProduct>
```

Metadata about the context and archiving process

```
<table>
  <name>address</name>
  <folder>table2</folder>
  <columns>
    <column>
      <name>address_id</name>
      <type>SMALLINT</type>
      <typeOriginal>SMALLINT UNSIGNED</typeOriginal>
      <nullable>false</nullable>
    </column>
    <column>
      <name>address</name>
      <type>CHARACTER VARYING(50)</type>
      <typeOriginal>VARCHAR</typeOriginal>
      <nullable>false</nullable>
    </column>
    <column>
      <name>address2</name>
      <type>CHARACTER VARYING(50)</type>
      <typeOriginal>VARCHAR</typeOriginal>
      <nullable>true</nullable>
    </column>
    <column>
      <name>district</name>
      <type>CHARACTER VARYING(20)</type>
      <typeOriginal>VARCHAR</typeOriginal>
      <nullable>false</nullable>
    </column>
    <column>
      <name>city_id</name>
      <type>SMALLINT</type>
```

Metadata about the database structure

```

<primaryKey>
  <name>PRIMARY</name>
  <column>customer_id</column>
</primaryKey>
<foreignKeys>
  <foreignKey>
    <name>fk_customer_address</name>
    <referencedSchema>sakila</referencedSchema>
    <referencedTable>address</referencedTable>
    <reference>
      <column>address_id</column>
      <referenced>address_id</referenced>
    </reference>
    <deleteAction>NO ACTION</deleteAction>
    <updateAction>CASCADE</updateAction>
  </foreignKey>
</foreignKeys>
<candidateKeys>
  <candidateKey>
    <name>PRIMARY</name>
    <column>customer_id</column>
  </candidateKey>
</candidateKeys>
<triggers>
  <trigger>
    <name>customer_create_date</name>
    <actionTime>BEFORE</actionTime>
    <triggerEvent>INSERT</triggerEvent>
    <triggeredAction>SET NEW.create_date = NOW()</triggeredAction>
  </trigger>
</triggers>

```

Metadata about the constraints and behaviour

```

<view>
  <name>staff_list</name>
  <queryOriginal>CREATE ALGORITHM=UNDEFINED DEFINER=`root`@`%` SQL
SECURITY DEFINER VIEW `staff_list` AS select `s`.`staff_id` AS
`ID`,concat(`s`.`first_name`,`_utf8&apos; &apos;`,`s`.`last_name`)
AS `name`,`a`.`address` AS `address`,`a`.`postal_code` AS `zip
code`,`a`.`phone` AS `phone`,`city`.`city` AS `city`,`country`.`country`
AS `country`,`s`.`store_id` AS `SID` from (((`staff` `s` join `address`
`a` on((`s`.`address_id` = `a`.`address_id`))) join `city`
on((`a`.`city_id` = `city`.`city_id`))) join `country`
on((`city`.`country_id` = `country`.`country_id`)))</queryOriginal>
  <columns>
    <column>
      <name>ID</name>
      <type>SMALLINT</type>
      <typeOriginal>TINYINT UNSIGNED</typeOriginal>
      <nullable>>false</nullable>
      <defaultValue>0</defaultValue>
    </column>
    <column>
      <name>name</name>
      <type>CHARACTER VARYING(91)</type>
      <typeOriginal>VARCHAR</typeOriginal>
      <nullable>>true</nullable>
    </column>
    <column>
      <name>address</name>
      <type>CHARACTER VARYING(50)</type>
      <typeOriginal>VARCHAR</typeOriginal>
      <nullable>>false</nullable>
    </column>
  </columns>

```

Metadata about views

Metadata about routines

```
<routine>
  <specificName>sakila.film_in_stock</specificName>
  <name>film_in_stock</name>
  <body>
    CREATE DEFINER=`root`@`%` PROCEDURE `film_in_stock`(IN p_film_id
INT, IN p_store_id INT, OUT p_film_count INT)
    READS SQL DATA
    BEGIN
    SELECT inventory_id
    FROM inventory
    WHERE film_id = p_film_id
    AND store_id = p_store_id
    AND inventory_in_stock(inventory_id);

    SELECT FOUND_ROWS() INTO p_film_count;
    END
  </body>
</routine>
```

```
<users>
  <user>
    <name>root@localhost</name>
  </user>
  <user>
    <name>admin2@%</name>
  </user>
  <user>
    <name>admin1@%</name>
  </user>
  <user>
    <name>root@%</name>
  </user>
  <user>
    <name>admin3@%</name>
  </user>
  <user>
    <name>admin4@%</name>
  </user>
  <user>
    <name>admin5@%</name>
  </user>
</users>
```

Metadata about users

```
<?xml version="1.0" encoding="UTF-8"?>
<table
xsi:schemaLocation="http://www.admin.ch/xmlns/siard/2/schema1/table2.xsd
table2.xsd" xmlns="http://www.admin.ch/xmlns/siard/2/schema1/table2.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <row>
    <c1>1</c1>
    <c2>47 MySakila Drive</c2>
    <c4>Alberta</c4>
    <c5>300</c5>
    <c6></c6>
    <c7></c7>
    <c8>2014-09-25T21:30:27.000000Z</c8>
  </row>
  <row>
    <c1>2</c1>
    <c2>28 MySQL Boulevard</c2>
    <c4>QLD</c4>
    <c5>576</c5>
    <c6></c6>
    <c7></c7>
    <c8>2014-09-25T21:30:09.000000Z</c8>
  </row>
  <row>
    <c1>3</c1>
    <c2>23 Workhaven Lane</c2>
    <c4>Alberta</c4>
    <c5>300</c5>
    <c6></c6>
    <c7>14033335568</c7>
    <c8>2014-09-25T21:30:27.000000Z</c8>
```

Table content in standard formats

The background of the slide is an abstract composition of numerous vertical lines of varying thicknesses. These lines are rendered in a gradient of colors, ranging from dark, almost black, on the left to a vibrant, saturated red on the right. The lines are slightly blurred, giving a sense of depth and movement. Centered over this background is the text "Any questions?".

Any questions?



20-min break

Back at 11:30

GMT+1

The background of the slide is an abstract composition of numerous vertical lines of varying thicknesses. These lines are rendered in a gradient of colors, ranging from dark, almost black, on the left to a vibrant, deep red on the right. The lines are slightly blurred, giving a sense of depth and movement. The overall effect is a dynamic, textured backdrop.

Tools for database preservation

DBPTK Database Preservation Toolkit

Set of tools to store relational databases
in a standard archival format.



<https://database-preservation.com>



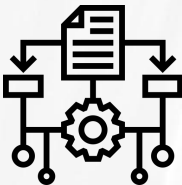
DBPTK Desktop

Desktop application to save database to preservation format, validate it, and browse and search the content



DBPTK Enterprise

Web application to browse and search on the content of multiple large preserved databases



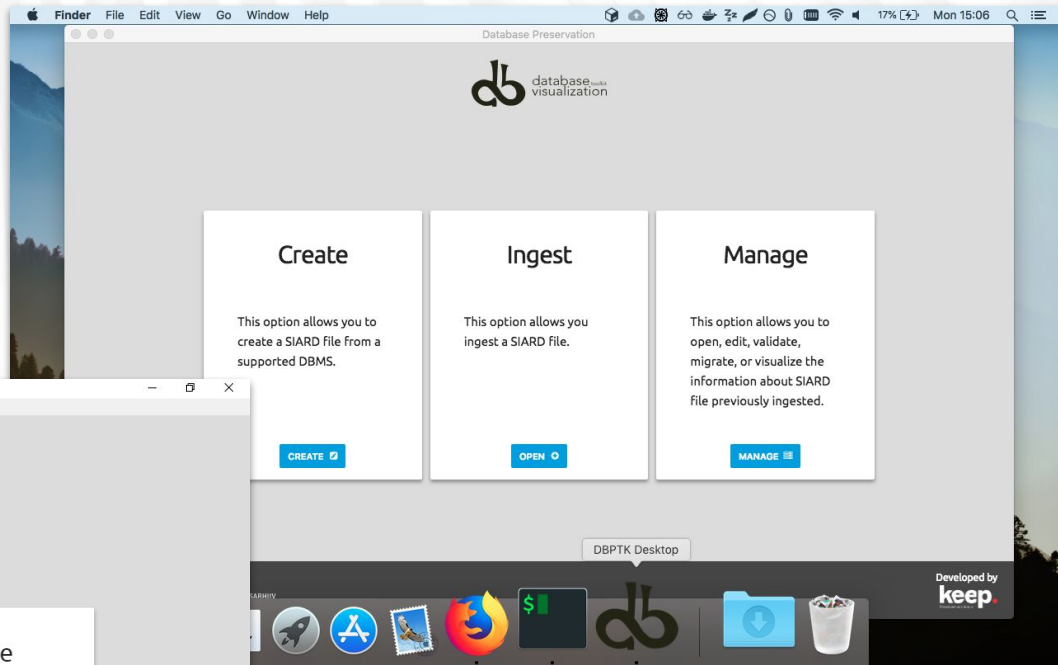
DBPTK Developer

A command-line tool and development library for automation and system integration

DBPTK Desktop

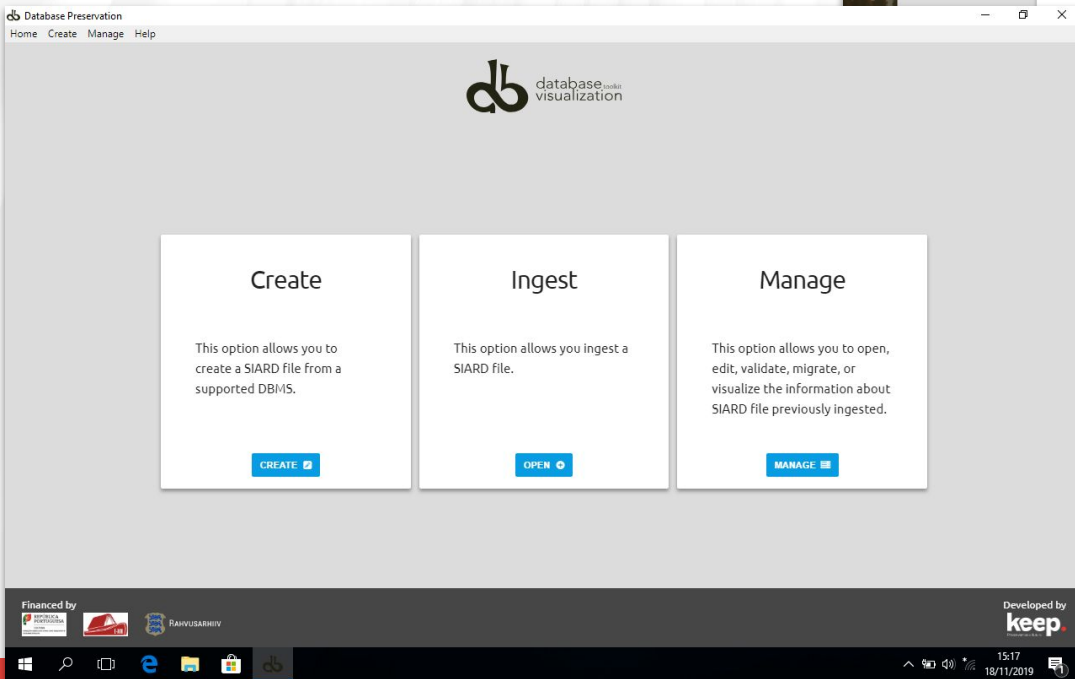
Basic features

keep.
Preserving the future



Also available on Linux

www.keep.pt



SIARD creation

Export database to a preservation format

Connect to a local or remote database and save all content into a preservation format like SIARD

Test connection will diagnose most common problems and provide you with helpful hints to solve them

Supported DBMS:

Microsoft Access
Microsoft SQL Server
MySQL / MariaDB
Oracle
PostgreSQL
Progress Openedge
Sybase

DBPTK Desktop

Home Create Manage Preferences Help

Home > Create SIARD - Connection

DBMS

- JDBC
- Microsoft Access
- Microsoft SQL Server
- MySQL**
- Oracle
- PostgreSQL
- Progress Openedge
- Sybase

General SSH Tunnel

Hostname * localhost
The name of the database server host (e.g. localhost)

Port number 3306
The server port number

Username * root
The name of the user to use in connection

Password *
The password of the user to use in connection

Database * sakila
The name of the database to connect

☐ **Disable Encryption**
Use to turn off encryption in the connection

TEST CONNECTION ⚡

CANCEL BACK NEXT

Migration report

Detailed report of migration changes and losses

All export and selection parameters are presented.

All column data types mapping to standard types are recorded.

All compromises are documented.

Database Preservation Toolkit (version 2.8.2) – Conversion Report

Parameters

Import module: mysql

- hostname = dpc.database-preservation.com
- database = sakila
- username = mguimaraes
- password =
- port-number = 3306
- disable-encryption = false

Export module: siard-2

- version = V2.1
- file = /home/mguimaraes/Desktop/sakila-dpc.siard
- compress = true
- pretty-xml = false
- external-lob = false
- external-lob-per-folder = 1000
- external-lob-folder-size = 0
- digest = SHA-256
- font-case = lowercase

Date: 2020-07-22

Details

- Type conversion in import module: in sakila.address.address (format: schema.table.column) has original type VARCHAR and was converted to the standard type CHARACTER VARYING(50)
- Type conversion in import module: in sakila.address.district (format: schema.table.column) has original type VARCHAR and was converted to the standard type CHARACTER VARYING(20)
- Type conversion in import module: in sakila.city.city (format: schema.table.column) has original type VARCHAR and was converted to the standard type CHARACTER VARYING(50)
- Type conversion in import module: in sakila.country.country (format: schema.table.column) has original type VARCHAR and was converted to the standard type CHARACTER VARYING(50)
- Type conversion in import module: in sakila.actor.actor_id (format: schema.table.column) has original type SMALLINT UNSIGNED and was converted to the standard type SMALLINT
- Type conversion in import module: in sakila.actor.first_name (format: schema.table.column) has original type VARCHAR and was converted to the standard type CHARACTER VARYING(45)
- Type conversion in import module: in sakila.actor.last_name (format: schema.table.column) has original type VARCHAR and was converted to the standard type CHARACTER VARYING(45)
- Information: check constraints is not yet supported for MySQL. But support may be added in the future
- Type conversion in import module: in sakila.address.address_id (format: schema.table.column) has original type SMALLINT UNSIGNED and was converted to the standard type SMALLINT
- Type conversion in import module: in sakila.address.address (format: schema.table.column) has original type VARCHAR and was converted to the standard type CHARACTER VARYING(50)
- Type conversion in import module: in sakila.address.address2 (format: schema.table.column) has original type VARCHAR and was converted to the standard type CHARACTER VARYING(50)
- Type conversion in import module: in sakila.address.district (format: schema.table.column) has original type VARCHAR and was converted to the standard type CHARACTER VARYING(20)
- Type conversion in import module: in sakila.address.city_id (format: schema.table.column) has original type SMALLINT UNSIGNED and was converted to the standard type SMALLINT
- Type conversion in import module: in sakila.address.postal_code (format: schema.table.column) has original type VARCHAR and was converted to the standard type CHARACTER VARYING(10)
- Type conversion in import module: in sakila.address.phone (format: schema.table.column) has original type VARCHAR and was converted to the standard type CHARACTER VARYING(20)

Edit SIARD metadata

Enrich archived database with descriptions

Add descriptions to database, tables and columns to better understand its contents

DBPTK Desktop

Home Create Manage Help

Home > Databases > sakila > SIARD Edit Metadata

Filter sidebar

Database

- Users & Roles
- sakila
 - Tables
 - actor
 - address
 - category
 - city
 - country
 - customer
 - film
 - film_actor
 - film_category
 - film_text
 - inventory
 - language
 - payment
 - rental
 - staff
 - store
 - Views
 - Routines

Database Information

Global information at database level

Name *	sakila
Archival date *	2016-09-15
Archivist	Bruno Ferreira
Archivist contact	email: bferreira@keep.pt
Client machine	mgulmaraes
Product	MySQL 5.5.5-10.1.11-MariaDB-1~trusty
User	
Data origin time span *	Early 2005 to March 2006
Data owner *	MySQL team
Description	The Sakila sample database was initially developed by a member of the MySQL AB documentation team, and is a standard schema that can be used for examples in books and so forth. Sakila sample database also serves to highlight MySQL such as Views, Stored Procedures, and Triggers. It is designed to represent a DVD rental store.
Producer application	Database Preservation Toolkit

SIARD validation

Validate archived database

Validate SIARD against specification
plus many additional checks for a
thorough validation

DBPTK Desktop

Home Create Manage Preferences Help

Home > Databases > sakila > Validation

Validation

Validates the SIARD against its specification. The validator shows information about which the requirements have passed and which one have failed. In case of a failed requirement, the report file generated contains the information needed to understand why the requirement failed.

Database Name:	sakila	SIARD specification:	SIARD-2.1
Requirements that passed:	27	Additional checks specification:	OPEN
Requirements that failed:	0	Report:	OPEN
Number of errors:	0		
Number of warnings:	175		
Number of skipped:	12		
Status:	Valid		

Scroll to the end

T_6.4-2	Validation finish on path: content/schema1/table13/table13.xml	OK
T_6.4-2	Validation running on path: content/schema1/table14/table14.xml	
T_6.4-2	Validation finish on path: content/schema1/table14/table14.xml	OK
T_6.4-2	Validation running on path: content/schema1/table15/table15.xml	
T_6.4-2	Validation finish on path: content/schema1/table15/table15.xml	OK
T_6.4-2	Validation running on path: content/schema1/table16/table16.xml	
T_6.4-2	Validation finish on path: content/schema1/table16/table16.xml	OK
T_6.4-2	The table file consists of row elements containing the data of a line subdivided into the various columns (c1, c2 ...).	OK
T_6.4-4	If a cell of a column contains a complex value (ARRAY, UDT), it is represented by a sequence of sub elements of the cell (a1,a2, ... for ARRAYS, u1, u2, ... for UDTs) which in turn contain their respective values. These values may again be complex.	SKIPPED
T_6.4-5	If a table contains data of the large object types (BLOB, CLOB, or XML ...) separate files may be produced for these and the storage location of the file is stored instead of the cell content.	OK

Search records

Browse and search database content

Google-like search on the database content.

Drill down on specific tables and do advanced search for specific fields to find exactly what you are looking for.

The screenshot shows the DBPTK Desktop application interface. The top navigation bar includes 'Home', 'Create', 'Manage', and 'Help'. The breadcrumb trail indicates the current location: 'Home > Databases > sakila > Search'. The left sidebar, titled 'Filter sidebar', contains sections for 'Information', 'Search all records' (highlighted), 'Saved searches', and 'Tables'. The 'Tables' section lists various tables including 'actor', 'address', 'category', 'city', 'country', 'customer', 'film', 'film_actor', 'film_category', 'film_text', 'inventory', 'language', 'payment', 'rental', 'staff', 'store', 'actor_info', 'customer_list', and 'film_list'. The main content area is titled 'Search all records' and shows search results for the query 'dan'. The results are displayed in two tables: 'actor' and 'customer'. The 'actor' table shows three records with actor IDs 18, 56, and 116, all with first names 'DAN' and last names 'TORN', 'HARRIS', and 'STREEP' respectively. The 'customer' table shows one record with customer ID 477, store ID 1, first name 'DAN', and last name 'PAINE'. The email address is 'DAN.PAINE@sakilacus'.

DBPTK Desktop

Home Create Manage Help

Home > Databases > sakila > Search

Filter sidebar

Information

Search all records

Saved searches

Tables

- actor
- address
- category
- city
- country
- customer
- film
- film_actor
- film_category
- film_text
- inventory
- language
- payment
- rental
- staff
- store
- actor_info
- customer_list
- film_list

Search all records

dan

actor

actor_id	first_name	last_name	last_update
18	DAN	TORN	2006-02-15
56	DAN	HARRIS	2006-02-15
116	DAN	STREEP	2006-02-15

1-3 of 3

customer

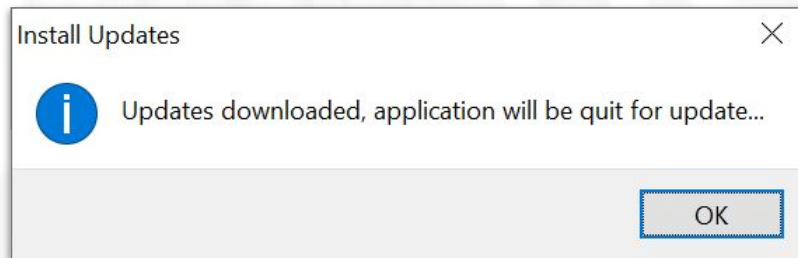
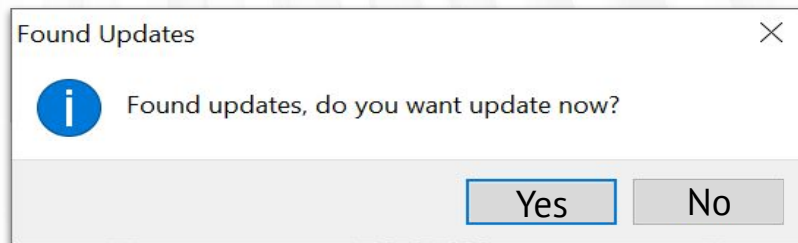
customer_id	store_id	first_name	last_name	email
477	1	DAN	PAINE	DAN.PAINE@sakilacus

1-1 of 1

Auto-update

Automatic check of updates

Stay up-to-date with automatic update check on startup and installation of new versions.





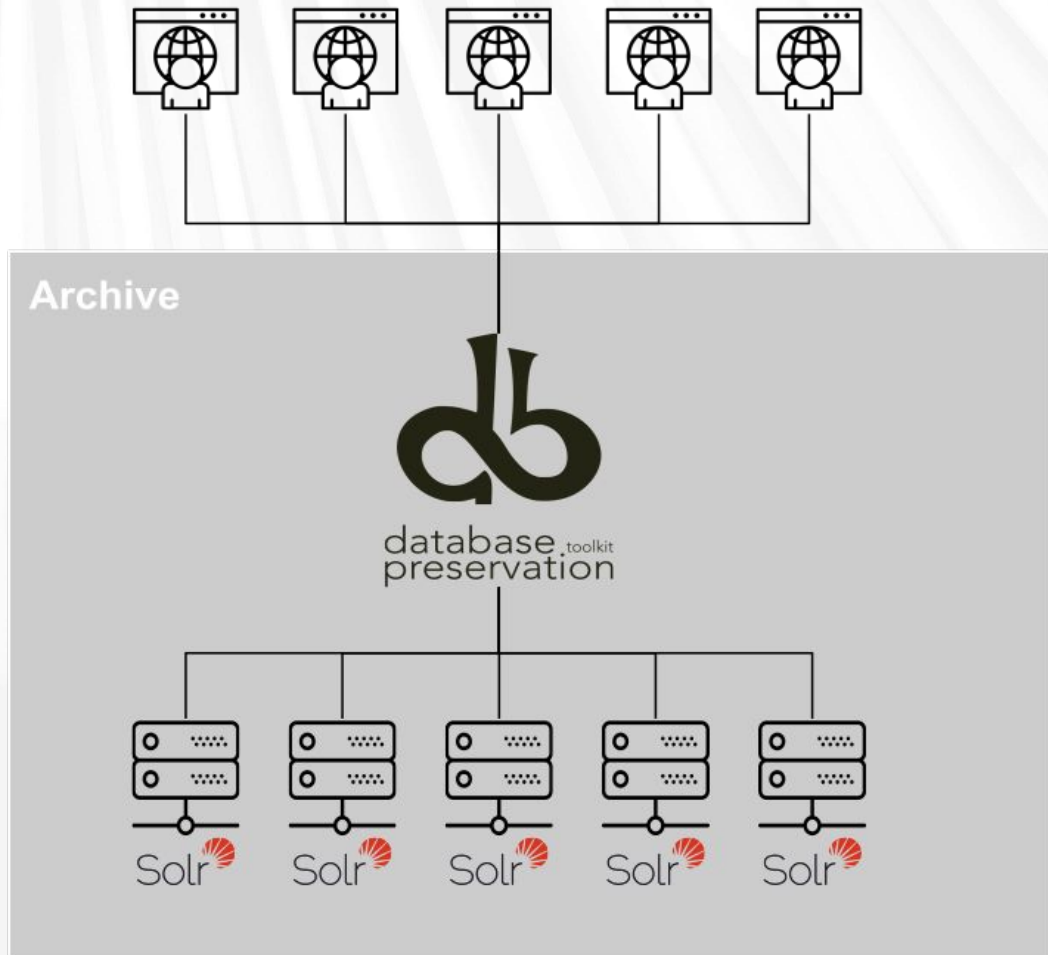
DBPTK Enterprise

Basic features

Enterprise architecture

For large institutions with many databases and users

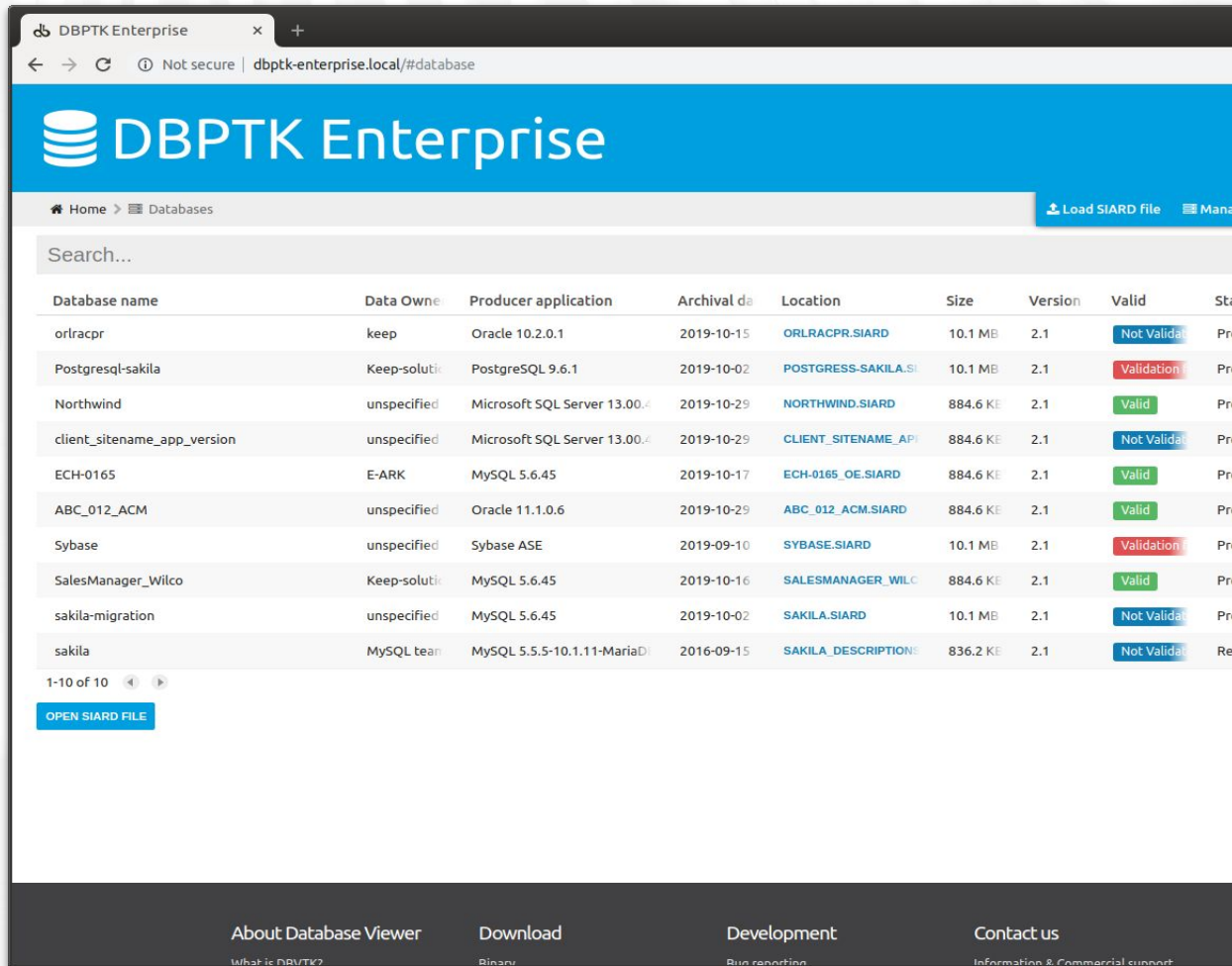
A web application that can be horizontally scaled to support many very large databases being accessed by many users



Manage multiple databases

Single system, multiple
databases

Search through the databases,
manage their status, enrich their
metadata, validate them, make them
ready for users to search.



The screenshot shows the DBPTK Enterprise web interface. The browser address bar indicates the URL is `dbptk-enterprise.local/#database`. The page has a blue header with the DBPTK logo and the text "DBPTK Enterprise". Below the header, there is a navigation bar with "Home" and "Databases" links, and a "Load SIARD file" button. A search bar is present above the table. The table lists various databases with columns for Database name, Data Owner, Producer application, Archival date, Location, Size, Version, Valid status, and Status. The "Valid" column contains buttons like "Not Valid", "Validation", and "Valid". At the bottom of the table, it says "1-10 of 10" and there is an "OPEN SIARD FILE" button. The footer contains links for "About Database Viewer", "Download", "Development", and "Contact us".

Database name	Data Owner	Producer application	Archival date	Location	Size	Version	Valid	Status
orlracpr	keep	Oracle 10.2.0.1	2019-10-15	ORLRACPR.SIARD	10.1 MB	2.1	Not Valid	Pr
Postgresql-sakila	Keep-soluti	PostgreSQL 9.6.1	2019-10-02	POSTGRESS-SAKILA.SI	10.1 MB	2.1	Validation	Pr
Northwind	unspecified	Microsoft SQL Server 13.00.4	2019-10-29	NORTHWIND.SIARD	884.6 KB	2.1	Valid	Pr
client_sitename_app_version	unspecified	Microsoft SQL Server 13.00.4	2019-10-29	CLIENT_SITENAME_APP	884.6 KB	2.1	Not Valid	Pr
ECH-0165	E-ARK	MySQL 5.6.45	2019-10-17	ECH-0165_OE.SIARD	884.6 KB	2.1	Valid	Pr
ABC_012_ACM	unspecified	Oracle 11.1.0.6	2019-10-29	ABC_012_ACM.SIARD	884.6 KB	2.1	Valid	Pr
Sybase	unspecified	Sybase ASE	2019-09-10	SYBASE.SIARD	10.1 MB	2.1	Validation	Pr
SalesManager_Wilco	Keep-soluti	MySQL 5.6.45	2019-10-16	SALESMANAGER_WILCO	884.6 KB	2.1	Valid	Pr
sakila-migration	unspecified	MySQL 5.6.45	2019-10-02	SAKILA.SIARD	10.1 MB	2.1	Not Valid	Pr
sakila	MySQL team	MySQL 5.5.5-10.1.11-MariaDB	2016-09-15	SAKILA_DESCRIPTIONS	836.2 KB	2.1	Not Valid	Re

1-10 of 10

[OPEN SIARD FILE](#)

[About Database Viewer](#) [Download](#) [Development](#) [Contact us](#)

What is DBPTK? Binary Bug reporting Information & Commercial support

Single sign-on

Support for multiple protocols

LDAP, Active Directory, Database, SAML, ADFS, OAuth2, OpenID, Google, Facebook, Twitter, FIDO U2F, YubiKey, Google Authenticator, Authy, etc.

Supports internal authorization definition or configurable external authorization

The screenshot shows a web browser window with the address bar displaying 'auth.labs.keep.pt/cas/login?lo...'. The page title is 'Login - Central Authentica...'. The main content area features the 'database toolkit preservation' logo at the top. Below the logo are two input fields: 'USERNAME' and 'PASSWORD'. A link for 'Forgot password?' is located to the right of the password field. A large green 'LOGIN' button is positioned below the input fields. Underneath the login button is a horizontal line with the word 'OR' in the center. Below this line are four social media login buttons: Facebook (blue), AUTENTICAÇÃO.GOV (blue), Google (red), and LinkedIn (light gray). At the bottom of the page, a security notice reads: 'For security reasons, please log out and exit your web browser when you are done accessing services that require authentication!'.

Browse and search

Allow users to access database content on the Web

Allow them to search on a prepared, user-friendly and anonymized database content

The screenshot displays the Sakila web application interface. The browser address bar shows the URL `dbptk-enterprise.local/#table/1547aa41-1800-46b7-a28b-e82fe22f5883/cf931074-1079-4d1f-8212-6b519c78fa81/update`. The application header features the Sakila logo and navigation links: Home, Databases, sakila, and film. A filter sidebar on the left lists various database tables, with 'film' selected. The main content area shows the 'film' table description and a search interface. The search results table lists films with columns: title, description, release_year, length, rating, and special_features.

title	description	release_year	length	rating	special_features
AIRPLANE SIERRA	A Touching Saga of a P...	2006	62	PG-13	Trailers, Deleted Sce...
ALABAMA DEVIL	A Thoughtful Panoram	2006	114	PG-13	Trailers, Deleted Sce...
ALTER VICTORY	A Thoughtful Drama o	2006	57	PG-13	Trailers, Behind the S...
ANTHEM LUKE	A Touching Panorama	2006	91	PG-13	Deleted Scenes, Behi...
APOLLO TEEN	A Action-Packed Refle	2006	153	PG-13	Trailers, Commentar...
ARACHNOPHOBIA RO	A Action-Packed Refle	2006	147	PG-13	Trailers, Deleted Sce...
ARGONAUTS TOWN	A Emotional Epistle of	2006	127	PG-13	Trailers, Commentar...
ATTACKS HATE	A Fast-Paced Panoram	2006	113	PG-13	Trailers, Behind the S...

Export features

Export data into
tabular data

Allow users to save search
results in Microsoft Excel
or other spreadsheet
software format for easy
analytics and diagrams

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	
	title	description	release_year	length	rating	special_features
1	The title of the film.	A short description or plot summary of the film.	The year in which the movie was released.	The duration of the film	The rating as	Lists which common s
2	GOLDFINGER SENSIBILITY	A Insightful Drama of a Mad Scientist And a Hunter who must Defeat a Pastry Chef in New Orleans	2006	93 G	Trailers,Commentaries	
3	WOLVES DESIRE	A Fast-Paced Drama of a Squirrel And a Robot who must Succumb a Technical Writer in A Manhat	2006	55 NC-17	Behind the Scenes	
4	WOLVES DESIRE	A Fast-Paced Drama of a Squirrel And a Robot who must Succumb a Technical Writer in A Manhat	2006	55 NC-17	Behind the Scenes	
5	WOLVES DESIRE	A Fast-Paced Drama of a Squirrel And a Robot who must Succumb a Technical Writer in A Manhat	2006	55 NC-17	Behind the Scenes	
6	GUNFIGHT MOON	A Epic Reflection of a Pastry Chef And a Explorer who must Reach a Dentist in The Sahara			Trailers,Behind the Sc	
7	TRIP NEWTON	A Fanciful Character Study of a Lumberjack And a Car who must Discover a Cat in An Aban			d Scenes,Behind	
8	PERDITION FARGO	A Fast-Paced Story of a Car And a Cat who must Outgun a Hunter in Berlin			entaries,Delete	
9	INDEPENDENCE HOTEL	A Thrilling Tale of a Technical Writer And a Boy who must Face a Pioneer in A Monastery			s,Behind the Sc	
10	ROSES TREASURE	A Astounding Panorama of a Monkey And a Secret Agent who must Defeat a Woman in T			entaries,Delete	
11	KENTUCKIAN GIANT	A Stunning Yarn of a Woman And a Frisbee who must Escape a Waitress in A U-Boat			s,Commentaries	
12	PUNK DIVORCE	A Fast-Paced Tale of a Pastry Chef And a Boat who must Face a Frisbee in The Canadian I			s,Commentaries	
13	KNOCK WARLOCK	A Unbelievable Story of a Teacher And a Boat who must Confront a Moose in A Baloon			s	
14	UPTOWN YOUNG	A Fateful Documentary of a Dog And a Hunter who must Pursue a Teacher in An Abandon			entaries	
15	MAGUIRE APACHE	A Fast-Paced Reflection of a Waitress And a Hunter who must Defeat a Forensic Psycholo			s,Commentaries	
16	WYOMING STORM	A Awe-Inspiring Panorama of a Robot And a Boat who must Overcome a Feminist in A U			d Scenes	
17	CENTER DINOSAUR	A Beautiful Character Study of a Sumo Wrestler And a Dentist who must Find a Dog in Ca			d Scenes	
18	DIVIDE MONSTER	A Intrepid Saga of a Man And a Forensic Psychologist who must Reach a Squirrel in A Mo			s,Commentaries	
19	SPIRIT FLINTSTONES	A Brilliant Yarn of a Cat And a Car who must Confront a Explorer in Ancient Japan			entaries,Delete	
20	INTOLERABLE INTENTIONS	A Awe-Inspiring Story of a Monkey And a Pastry Chef who must Succumb a Womanizer in			entaries,Behind	
21	HOOK CHARLOTS	A Insightful Story of a Boy And a Dog who must Redeem a Boy in Australia			s,Commentaries	
22	ENCINO ELF	A Astounding Drama of a Feminist And a Teacher who must Confront a Husband in A Bal			s,Behind the Sc	
23	CURTAIN VIDEOTAPE	A Boring Reflection of a Dentist And a Mad Cow who must Chase a Secret Agent in A Sha			s,Commentaries	
24	LAMBS CINCINNATI	A Insightful Story of a Man And a Feminist who must Fight a Composer in Australia			s,Behind the Sc	
25	MAGNOLIA FORRESTER	A Thoughtful Documentary of a Composer And a Explorer who must Conquer a Dentist in			s,Commentaries	
26	BACKLASH UNDEFEATED	A Stunning Character Study of a Mad Scientist And a Mad Cow who must Kill a Car in A M			s,Behind the Sc	
27	CLEOPATRA DEVIL	A Fanciful Documentary of a Crocodile And a Technical Writer who must Fight a A Shark i			s,Deleted Scene	
28	HOCUS FRIDA	A Awe-Inspiring Tale of a Girl And a Madman who must Outgun a Student in A Shark Tan			s,Deleted Scene	
29	STAGE WORLD	A Lackluster Panorama of a Woman And a Frisbee who must Chase a Crocodile in A Jet i			entaries,Behind	
30	CHAINSAW UPTOWN	A Beautiful Documentary of a Boy And a Robot who must Discover a Squirrel in Australia			d Scenes,Behind	
31	PILOT HOOSIERS	A Awe-Inspiring Reflection of a Crocodile And a Sumo Wrestler who must Meet a Forensi			s,Deleted Scene	
32	REMEMBER DIARY	A Insightful Tale of a Technical Writer And a Waitress who must Conquer a Monkey in Ancient Indi	2006	110 K	Trailers,Commentaries	
33	JAPANESE RUN	A Awe-Inspiring Epistle of a Feminist And a Girl who must Sink a Girl in The Outback	2006	135 G	Deleted Scenes	
34	RAINBOW SHOCK	A Action-Packed Story of a Hunter And a Boy who must Discover a Lumberjack in Ancient India	2006	74 PG	Trailers,Commentaries	
35	MAIDEN HOME	A Lackluster Saga of a Moose And a Teacher who must Kill a Forensic Psychologist in A MySQL Co	2006	138 PG	Behind the Scenes	

A bar chart titled "The duration of the film" is visible on the right side of the spreadsheet, showing the duration of the films in minutes. The x-axis ranges from 0 to 200 minutes, and the y-axis lists the film titles. The bars are blue and represent the duration of each film.

Activity log

Audit every access

Who has done what, when
and from where.

Requirement for ISO 16363
certification.

DBPTK Enterprise

dpc.database-preservation.com/?locale=en#activity-log

Databases > Activity Log

lfaria Administration English

Activity log

Event logs are special files that record significant events that happen in the application. For example, a record is kept every time a user logs in, when a download is made or when a search is made. Whenever these events occur, the repository records the necessary information in the event log to enable future auditing of the system activity. For each event the following information is recorded: date, involved component, system method or function, target objects, user that executed the action, the duration of action, and the IP address of the user that executed the action. Users are able to filter events by type, date and other attributes by selecting the options available in the right side panel.

Search... advanced

Date	Component	Method	User	Duration	Address	Outcome
2020-07-24 11:46:06	Database	Find	lfaria	10ms	81.84.255.161	Success
2020-07-24 11:46:06	Database	Find	lfaria	12ms	81.84.255.161	Success
2020-07-24 11:46:05	Login	Cas Login	lfaria	1ms	81.84.255.161	Success
2020-07-24 11:46:00	Database	Find	mguimaraes	15ms	81.84.255.161	Success
2020-07-24 11:45:50	Database	Find	mguimaraes	9ms	81.84.255.161	Success
2020-07-24 11:45:40	Database	Find	mguimaraes	9ms	81.84.255.161	Success
2020-07-24 11:45:30	Database	Find	mguimaraes	17ms	81.84.255.161	Success
2020-07-24 11:45:20	Database	Find	mguimaraes	10ms	81.84.255.161	Success
2020-07-24 11:45:10	Database	Find	mguimaraes	10ms	81.84.255.161	Success
2020-07-24 11:45:00	Database	Find	mguimaraes	10ms	81.84.255.161	Success
2020-07-24 11:44:50	Database	Find	mguimaraes	9ms	81.84.255.161	Success
2020-07-24 11:44:40	Database	Find	mguimaraes	12ms	81.84.255.161	Success
2020-07-24 11:44:30	Database	Find	mguimaraes	11ms	81.84.255.161	Success
2020-07-24 11:44:20	Database	Find	mguimaraes	13ms	81.84.255.161	Success
2020-07-24 11:44:10	Database	Find	mguimaraes	10ms	81.84.255.161	Success
2020-07-24 11:44:00	Database	Find	mguimaraes	20ms	81.84.255.161	Success
2020-07-24 11:43:50	Database	Find	mguimaraes	10ms	81.84.255.161	Success
2020-07-24 11:43:40	Database	Find	mguimaraes	12ms	81.84.255.161	Success
2020-07-24 11:43:30	Database	Find	mguimaraes	13ms	81.84.255.161	Success
2020-07-24 11:43:20	Database	Find	mguimaraes	11ms	81.84.255.161	Success

1-20 of 2,972 Show More

Components

- ☐ Activity log (4)
- ☐ Collection (549)
- ☐ Database (1688)
- ☐ File (37)
- ☐ Job (126)
- ☐ SIARD (484)
- ☐ Login (84)

Methods

- ☐ Cas Login (73)
- ☐ Create (33)
- ☐ Create Collection (3)
- ☐ Create Denormalize Configuration F
- ☐ Create SIARD File (36)
- ☐ Delete (33)
- ☐ Delete Collection (2)
- ☐ Delete Validation Report (7)
- ☐ Export L O B (4)
- ☐ Export Single Row To C S V (1)
- ☐ Find (1531)
- ☐ Find Rows (107)
- ☐ Find Saved Searches (7)
- ☐ Get Collection Configuration (129)

Multiple languages supported

Interface translated into:

English, German, Estonian, Czech, Portuguese

Search stemming and stopwords support for:

English, Arabic, Bulgarian, Catalan, Czech, Danish, German, Greek, Spanish, Estonian, Basque, Persian, Finnish, French, Irish, Galician, Hindi, Hungarian, Armenian, Indonesian, Italian, Latvian, Dutch, Norwegian, Portuguese, Romanian, Russian, Swedish, Thai, Turkish, Japanese (using morphological analysis), CJK bigram (Chinese, Japanese, and Korean languages)

The background of the image is an abstract composition of numerous vertical lines of varying thicknesses. These lines are colored in a gradient from dark grey/black on the left to a vibrant red on the right. The lines are slightly blurred, creating a sense of depth and movement. In the center, the text 'DBPTK Developer' is displayed in a bold, white, sans-serif font. A solid red horizontal bar is positioned at the bottom left corner of the image.

DBPTK Developer

Command line interface

Automation of periodic preservation tasks

Command line interface allows easy automation of periodic tasks like saving database to preservation format, validating, and editing metadata.

```
2/2 + [ ] [ ] Tilix: Default
~ $ java -jar dbptk-app-2.6.3.jar
Database Preservation ToolkitDatabase Preservation Toolkit (version 2.6.3)
More info: http://www.database-preservation.com

Usage: dbptk COMMAND [OPTIONS]

Commands:

    migrate      Migrates data and metadata from an import module to an export module.
    edit         Edit the metadata information from a SIARD 2 archive.
    validate     Validate a SIARD 2 archive.

Run 'dbptk -h|help COMMAND' for more information on a command.

Log files and migration reports were saved in /home/mguimaraes
Troubleshooting information can be found at http://www.database-preservation.com/#troubleshooting
Please report any problems at https://github.com/keeps/db-preservation-toolkit/issues/new

~ $ _
```

Systems integration Java library

Library to allow integration of production systems to directly use database preservation features.

The screenshot displays the Artifactory web interface. The browser address bar shows the URL: `artifactory.keep.pt/artifactory/webapp/#/artifacts/browse/simple/General/keep/com/databasepreservation/dbptk-model/2.6.3/dbptk-model-2.6.3.jar`. The page title is "Artifact Repository Browser". On the left, a tree view shows the directory structure: `2.6.3` containing `dbptk-model-2.6.3.jar`, `dbptk-model-2.6.3.pom`, and `dbptk-model-2.6.3-sources.jar`. The main panel shows the details for `dbptk-model-2.6.3.jar`. The "General" tab is active, displaying the following information:

Info	
Name:	dbptk-model-2.6.3.jar
Repository Path:	keep/com/databasepreservation/dbptk-model/2.6.3/dbptk-model-2.6.3.jar
Module ID:	com.databasepreservation:dbptk-model:2.6.3
Deployed By:	token:travis
Size:	192.56 KB
Created:	06-11-19 18:20:53 +00:00
Last Modified:	06-11-19 18:20:53 +00:00
Last Downloaded:	13-11-19 09:12:50 +00:00

Below the info section, the "Dependency Declaration" section shows the build tool configuration. The "Maven" tab is selected, displaying the following XML snippet:

```
1 <dependency>
2   <groupId>com.databasepreservation</groupId>
3   <artifactId>dbptk-model</artifactId>
4   <version>2.6.3</version>
5 </dependency>
```


Open source

For custom development

Code base that allows custom development of new features or special support for new or legacy database systems.

The screenshot shows the GitHub repository page for 'keeps/dbptk-developer'. The repository is in the 'keeps' organization. It has 15 watchers, 31 stars, and 11 forks. The repository is in the 'Code' tab, showing 71 issues, 1 pull request, 1 action, 1 project, and 1 wiki. The repository is on the 'master' branch, with 10 branches and 55 tags. The repository is owned by 'hmiguim' and has 1,246 commits. The repository is a 'Setting version 2.10.0-SNAPSHOT'.

File	Description	Time
.github	fix typos	4 years ago
.travis	Testing travis conditions	11 months ago
code-style	updated codestyle	4 years ago
dbptk-bindings	Setting version 2.10.0-SNAPSHOT	17 hours ago
dbptk-core	Setting version 2.10.0-SNAPSHOT	17 hours ago
dbptk-model	Setting version 2.10.0-SNAPSHOT	17 hours ago
dbptk-modules	Setting version 2.10.0-SNAPSHOT	17 hours ago
dbptk-plugin-example	Setting version 2.0.1	2 years ago
doc	fixes #135, partially fixes #142 by adding support for e...	4 years ago
examples	Add examples of import-config files [skip ci]	5 months ago
scripts	Fixes #358	2 years ago
testing	removed sakila from testing folder. there are instruction...	2 years ago
.gitattributes	attempt at ignoring example files in github language gr...	4 years ago
.gitignore	Metadata Validator (XML against XSD) #353 [skip ci]	11 months ago
qrencr.yml	Fixes #358	2 years ago

About
DBPTK Developer - library and command-line tool for execution of database preservation actions
www.database-preservation...
preservation database relational-databases siard preservation-formats

Releases 55
Version 2.9.2 Latest
17 hours ago
+ 54 releases

Contributors 11

The background of the slide is an abstract composition of numerous vertical lines of varying thicknesses. These lines are colored in a gradient from dark grey/black on the left to a vibrant red on the right. The lines are slightly blurred, giving a sense of depth and movement. The word "Demonstration" is centered in the middle of the image in a white, bold, sans-serif font.

Demonstration



Case study: Testing SIARD 2.0

by Brett Abrams, NARA

The background of the slide is an abstract composition of numerous vertical lines of varying thicknesses. These lines are colored in a gradient from dark grey/black on the left to a vibrant red on the right. The lines are slightly blurred, creating a sense of depth and movement. Centered over this background is the text 'Questions and discussion' in a bold, white, sans-serif font. A solid red horizontal bar is located at the bottom left corner of the slide.

Questions and discussion

Lunch

Practical session starts at 14:00

GMT+1