

# Welcome to this live webinar on DBPTK – an eArchiving solution for database archiving

---

Start 10:00 (CEST)

24 September 2020

## Ground Rules for the Live Webinar



Click on “Connect audio” to hear the presenters but please mute your microphone throughout the webinar.



Submit your questions in writing by using the Webex chat function. We will answer some questions live during the webinar and provide written answers to all (within the coming days).



**Please note that this webinar is recorded.**

# Agenda

---

10:00 – 10:05

## **Welcome**

Thomas Fillis – CEF Stakeholder Management Office – DG DIGIT

10:05– 10:30

## **CEF Telecom call 2020-2: How to prepare a successful proposal**

Adina Ratoi – CEF Telecom – INEA

10:30 – 11:15

## **DBPTK – an eArchiving solution for database archiving**

Luis Faria – Keep Solutions

11:15 – 11:30

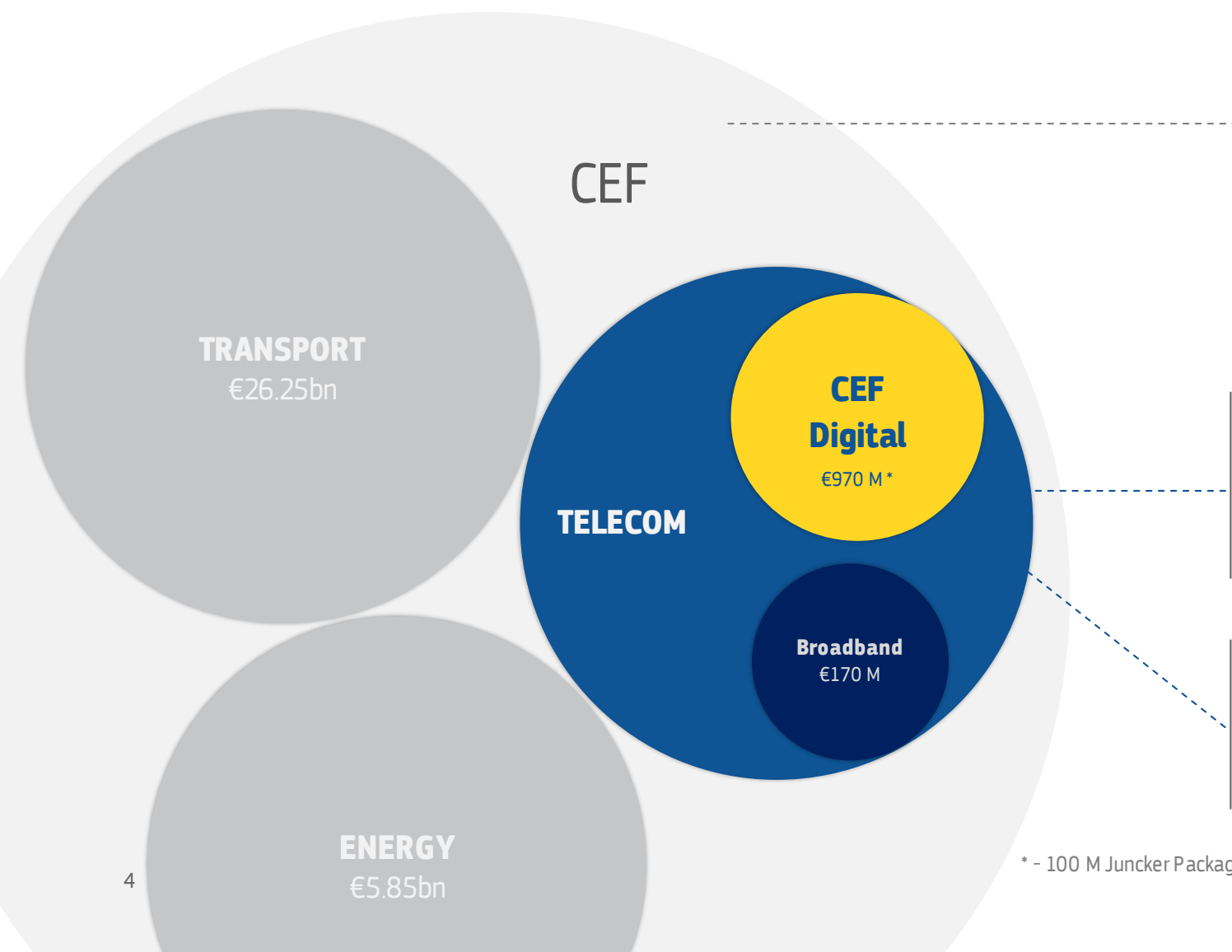
## **Q&A**

# Welcome to the Connecting Europe Facility (CEF) Building Blocks

Thomas Fillis  
CEF Stakeholder Management Office, DIGIT



# The CEF Building Blocks are funded by the Connecting Europe Facility



## CEF Regulation

Defines how the Commission can finance support for the establishment of trans-European networks to reinforce an interconnected Europe.

## CEF Telecom Guidelines

The CEF Telecom guidelines cover the specific objectives and priorities as well as eligibility criteria for funding of broadband networks and Digital Service Infrastructures (DSIs).

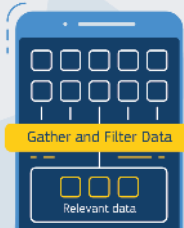
## CEF Work Programmes

Translates the CEF Telecom Guidelines in general objectives and actions planned on a yearly basis.



### Big Data Test Infrastructure

Explore and experiment with big data for improved performance and decision making



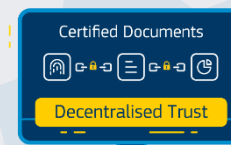
### Context Broker

Analyze, manage and share data, in real time, at the right time, throughout Europe



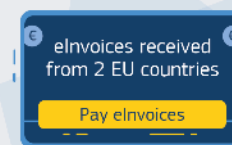
### eArchiving

Facilitates the preservation, migration, reuse and trust of your data



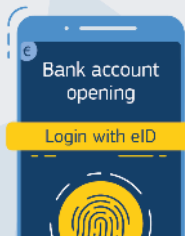
### European Blockchain Services Infrastructure

Harness the power of a European-wide network of blockchain services, increasing trust through data security, privacy and transparency



### eInvoicing

Promote the implementation of the European standard for electronic invoicing across borders



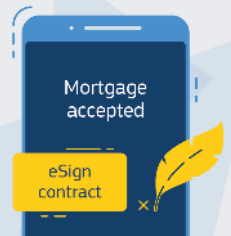
### eID

Allow citizens to prove who they are across borders, making it easier to access online services in another EU Member State



### eDelivery

Exchange online data and documents reliably and securely



### eSignature

Create and verify electronic signatures between businesses and EU citizens



### eTranslation

Offers machine translation to translate your documents and web content into any official EU language, Norwegian or Icelandic

---

# How does CEF support projects to use the Building Blocks?

It provides free services to help you implement them in your system. There are a range of services across the building blocks but services typically include training, sample software, testing services.

## Free services



**Training  
sessions**



**Sample  
software**



**Testing  
services**

### How to use a Building Block?

Build, buy or reuse the Building Blocks on your own.

Co-develop the solution or partner with other parties.

## Co-develop and partner

with other parties



**Build**  
The solution from scratch  
based on a European  
standard



**Buy**  
A compliant solution  
from the market



**Reuse**  
Sample software  
available on CEF website

## European Standards

# CEF Telecom call 2020-2: How to prepare a successful proposal

Adina Ratoi

Evaluation Manager, CEF Telecom – INEA – Unit R1





# READ: all call documentation

- See [call webpage](#) and consult:
  - [Work Programme \(Annex\)](#)
  - [Call text](#)
    - Take special note of the **Priorities & Objectives** (section 2.1) and **Results** (section 2.2) which provide specific information on what is expected from the proposals to achieve
    - Carefully read the **Award Criteria** (section 9) which explain how the proposal will be evaluated
  - **[Application forms](#)** – you must use the templates provided!
  - [Guide for Applicants](#)
  - [FAQs](#) – both general & specific
  - [Model grant agreement](#)

## 2020 CEF Telecom Call - eArchiving (CEF-TC-2020-2)

The **2020 CEF Telecom eArchiving** call makes an indicative €1 million of funding available for proposals in this area. To get started with your application, we suggest that you consult the list of relevant documents and information for the calls as provided below:

- read the call text and Work Programme sections relevant to your call
- review the application forms, Guide for Applicants and FAQs
- consult the background documents and other relevant information

Then you can apply online using the TENTec submission system.

For all questions related to the call, please check the call FAQ or contact the helpdesk (until 15 October).

Watch the **2020-2 CEF Telecom call virtual info day**

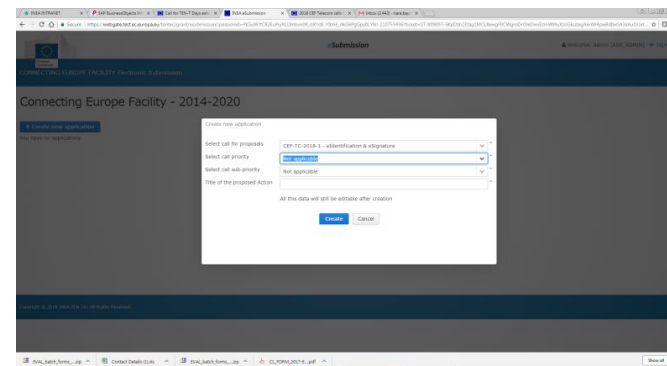
Looking for partners for your CEF Telecom consortium? Join this LinkedIn group

2020 CEF Telecom Calls

Indicative Call Timeline	
Call opening	30 June 2020
Deadline for submission	5 November 2020
Evaluation of proposals	November 2020 - January 2021
Consultation of the CEF Committee	March - April 2021
Information to the European Parliament	April 2021
Adoption of the Selection Decision	April 2021
Preparation and signature of grant agreements	Between May and August 2021

Work Programme	
EN 2019 - 2020 CEF Telecom Work Programme	Available here
Call for Proposals documents	
2020 CEF Telecom Call for Proposals - eArchiving	Call text

## READ: how to use TENtec



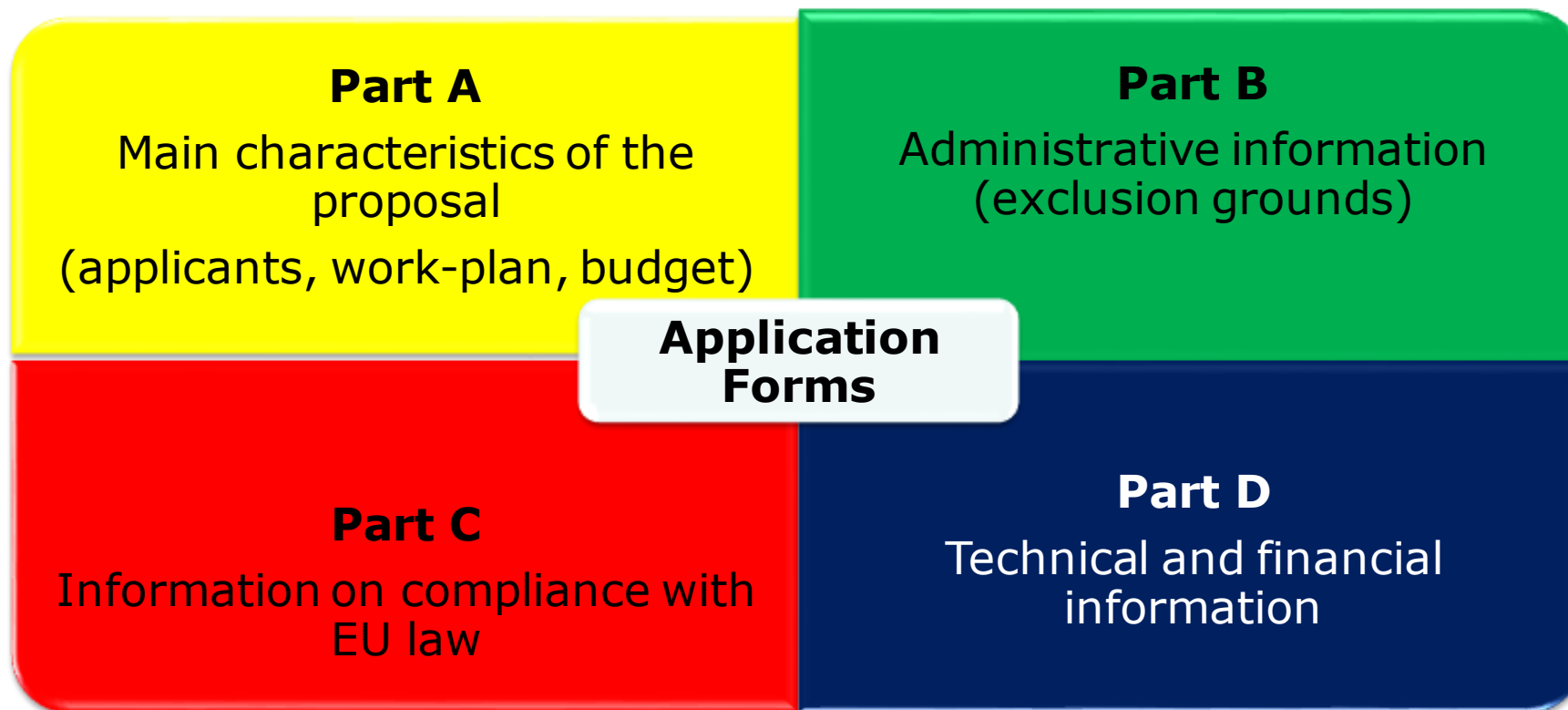
- **TENtec:** system used to manage the CEF projects during their entire lifecycle and which enables the electronic submission of proposals under the CEF calls
- Step-by-step instructions for the **TENtec eSubmission system** in the **Guide for Applicants**



**New feature:** if you encode an activity start date in the past or an activity end date before the results of the evaluation will see a **warning**

## READ: Application forms

All application forms and access to TENtec eSubmission module available via call page on INEA website



# Application form A

Essential information **on the applicants** and **on the proposal**: summary of the action, timing, activities and milestones, budget, breakdown of eligible costs and requested funding plus Member State approval

- **MUST** be encoded in the TENtec eSubmission module, .pdf will be generated automatically (Word version provided for reference on call webpage)
- Includes forms that require signature of the applicants (A2.2) and Member State validation (A2.3) – **upload separately**

**The descriptions of the proposed Action and activities will also be used for the grant agreement preparation: be complete, informative and precise!**

# Application form B

Administrative information on applicants to demonstrate compliance with **operational capacity** (*required by EU Financial Regulation*)

- Capacity of applicant to complete the proposed Action - complements information from application form part D2.3 (e.g. *activity reports, CVs, reports on similar projects, etc.*)
- **Certain types of applicants DO NOT** need to demonstrate operational capacity



**Each applicant** should **register** in the **Participant Register** before the call deadline and enter the **PIC number** in **Application Form Part A**.

**Complete all relevant parts of form B and upload to TENtec.**

# Application form C

- **Information on compliance with EU law on public procurement**
- **Information on other sources of EU financing** that may be received by the proposal (cannot receive grants from two EU funding sources)

# Application form D

## Detailed technical information describing the proposed Action

- **You must use the template available on the call page!**
- Order of the sections to be filled in reflects the **award criteria**
  - Address each point and subpoint in your application to ensure that your proposal contains all of the relevant information on which it will be assessed
- Part D: **30 pages** maximum
- Add **Gantt chart + other annexes**: please ensure that these are **readable and useful** for the evaluators!

# Award criteria

- **Defined in the Work Programme and call text**
  - Did you understand the priorities, objectives and expected results defined in the call texts?
  - Does your proposed Action address these points?
  - Can you justify why YOUR proposal should ultimately be selected for funding?

Relevance	Quality & efficiency of implementation	Impact & sustainability
<ul style="list-style-type: none"><li>• Alignment to DSI implementation objectives &amp; activities (WP)</li><li>• Alignment to EU/national policies, strategies and activities</li></ul>	<ul style="list-style-type: none"><li>• Maturity</li><li>• Coherence/effectiveness with work plan</li><li>• Quality of consortium/consortium members</li><li>• Support from national authorities/industry/NGOs</li><li>• Attention to security/privacy/inclusiveness/accessibility</li></ul>	<ul style="list-style-type: none"><li>• Quality of the approach to facilitate wider deployment/take-up of the proposed actions</li><li>• Capability of long-term sustainability without EU funding</li></ul>



# Consortium & approval requirements

- **Check the consortium requirements/requirements on types of applicants**
  - Do you have enough partners lined up to participate?
  - Who will serve as the consortium coordinator?
  - Does this organisation understand its role as a coordinator?
  - Can you provide evidence/justify that all applicants in the proposal meet the eligibility criteria?
- **Member State approval is necessary for all applicants and all applications to be eligible**
  - Do you understand how this approval process is done in your Member State?
  - Have you taken into consideration the time it will take to obtain the approval(s)?

## Consider...

- **Showing concrete evidence on how your proposed Action**
  - supports the objectives of the call
  - addresses the award criteria
  - mitigates any possible identified risks
  - incorporates a clear timetable and planning overview
- **Providing explanations/diagrams of IT solutions used, architecture, standards, etc.**
  - explain the work you will be undertaking
  - provide ample descriptions of your activities and milestones
- **Justifying costs (personnel, subcontracting, other costs)**
- **Including a business plan for sustainability**

# A good proposal...

- uses simple language
- provides clear descriptions on how the proposed activities/tasks will be implemented
- addresses **all** of the award criteria in sufficient detail
- is well-structured

**Evaluators must find the relevant information and evidence in the proposal in order to evaluate it – they will not make any assumptions!**

# REMEMBER: time flies...

- **Start NOW and don't forget about the deadline**
  - Completing an application is time consuming, especially for first time applicants
  - Member State endorsement and multi-applicant proposals take time
  - If the deadline passes and you haven't submitted your complete proposal, it will be declared inadmissible: it will not be evaluated!

# Answering your questions - FAQs

- **Helpdesk:** [INEA-CEF-Telecom-calls@ec.europa.eu](mailto:INEA-CEF-Telecom-calls@ec.europa.eu)
  - [General FAQs](#) and [specific FAQs](#)
  - **Deadline to submit questions: 15 October 2020**
  - **Deadline to publish answers: 29 October 2020**
- Questions on **TENtec eSubmission module?** Responses will be provided until the deadline
- **Visit your call webpage regularly to check for updates, sign up for our Twitter feed and FAQ notifications**

# One last step: make a final check before submitting your application

- Follow the steps as detailed in the Guide for Applicants
  - Use the [checklist](#) to ensure that you have all necessary forms
  - Upload **all** forms requiring signatures + make them clearly identifiable by their file name in English
  - Do not forget any supporting documents
  - Keep your originals – they may be requested later
- 
- Submit in TENtec **before the deadline 5 November 2020**(do not wait until the last minute!)

# CEF Telecom calls: for more information



**inea-cef-telecom-calls@ec.europa.eu**



**<https://ec.europa.eu/inea/en/connecting-europe-facility/cef-telecom/apply-funding/2020-cef-telecom-calls-proposals>**



**@inea\_eu**

**Thank you!**

# DBPTK – an eArchiving solution for database archiving

**Luis Faria**

Research & Innovation Director – KEEP SOLUTIONS





---

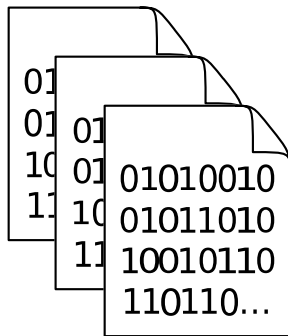
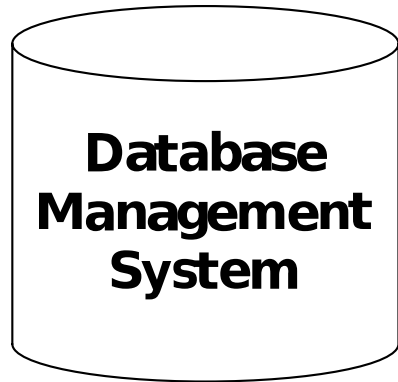
# Databases

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.

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 has a title bar "Contact" and a menu bar with "Co", "Name", "Ph", "ToDo", "End", "Log", and "Group". Below the menu bar is a status bar showing "Copyright 1999-2001, Allen Agee" and "Contact: Agee Software". The main area contains a form for a contact record. The form has fields for "Company", "First Name", "Last", "Billing Street", "City", "St", "TX", "Zip", "Job Street", "City", "St", "TX", "Zip", "Work Phone", "Fax", "Home", "Mobile", "Email", "Website", "Directions", "MAP", "Comment", "Created", "Action", "Tech", "Salesperson", "Tech", "Year", "Distance", and "Recording". The "Created" field shows "5/12/01 7:40:39 PM" and the "Action" field shows "5/14/01". The "Salesperson" field shows "Allen Agee" and the "Tech" field shows "Allen Agee". The "Year" field is empty and the "Distance" field shows "0 min". The "Recording" field shows "Recording: Call back Monday". The bottom of the window has a status bar showing "Record: 2 of 2" and "4 Recs". There are buttons for "Del", "New", "Err", "Bpl", "Help", "Exit", "Print", "RTF", and "TXT".

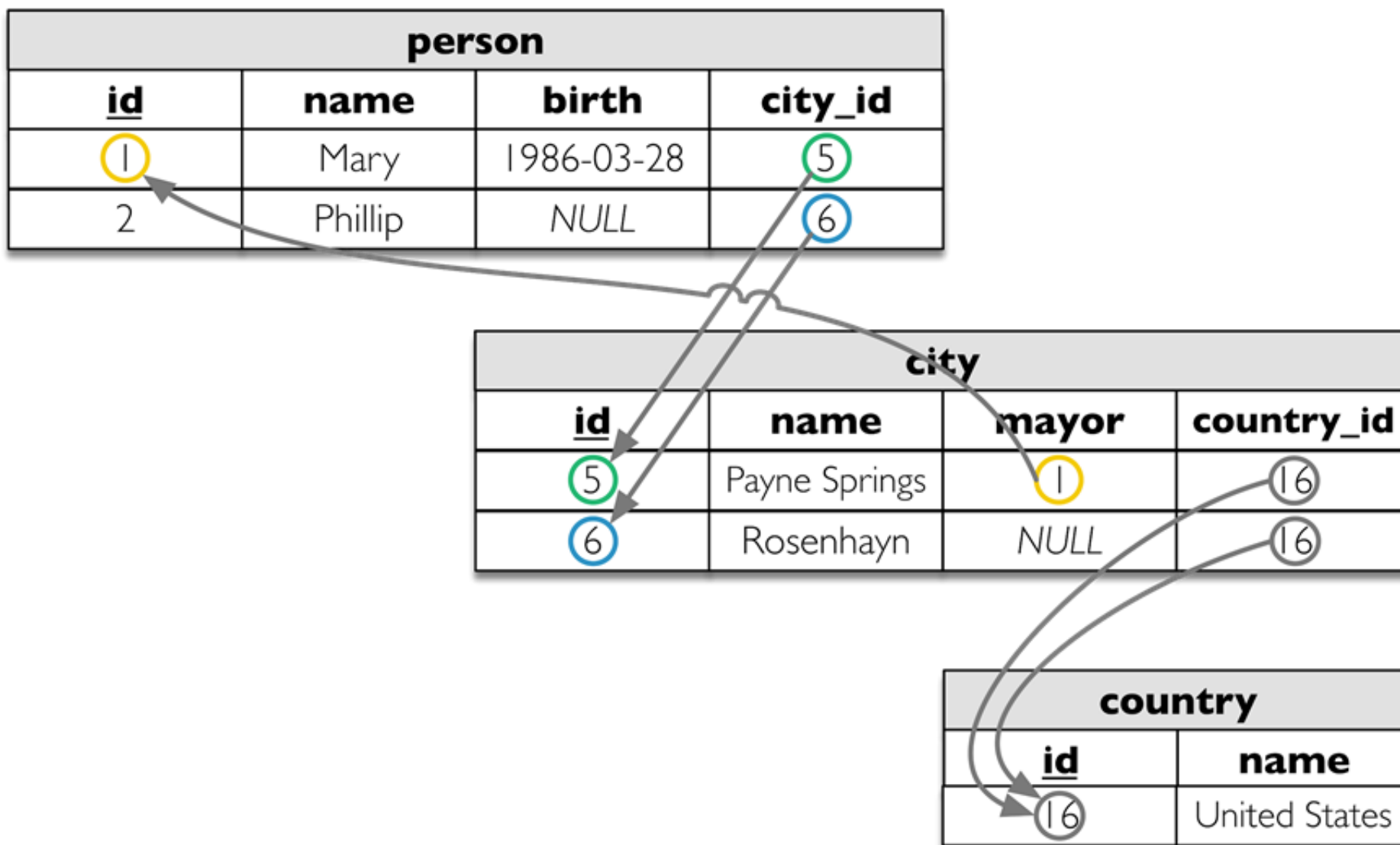
Application

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



---

# Information to preserve

## **Within the relational database:**

- Information in tables
- Column data types
- Relations and constraints
- Projections (views)
- Behaviour (triggers and routines)
- Other (users, permissions, etc.)

## **Outside 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 **emulates 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	⊖ may hinder timely access to information
⊕ no need to maintain hardware or software	⊖ difficult to gather documentation of complex or closed file formats
	⊖ difficult to ensure quality and completeness without hindsight

---

# The problem with preserving 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

<https://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>

---

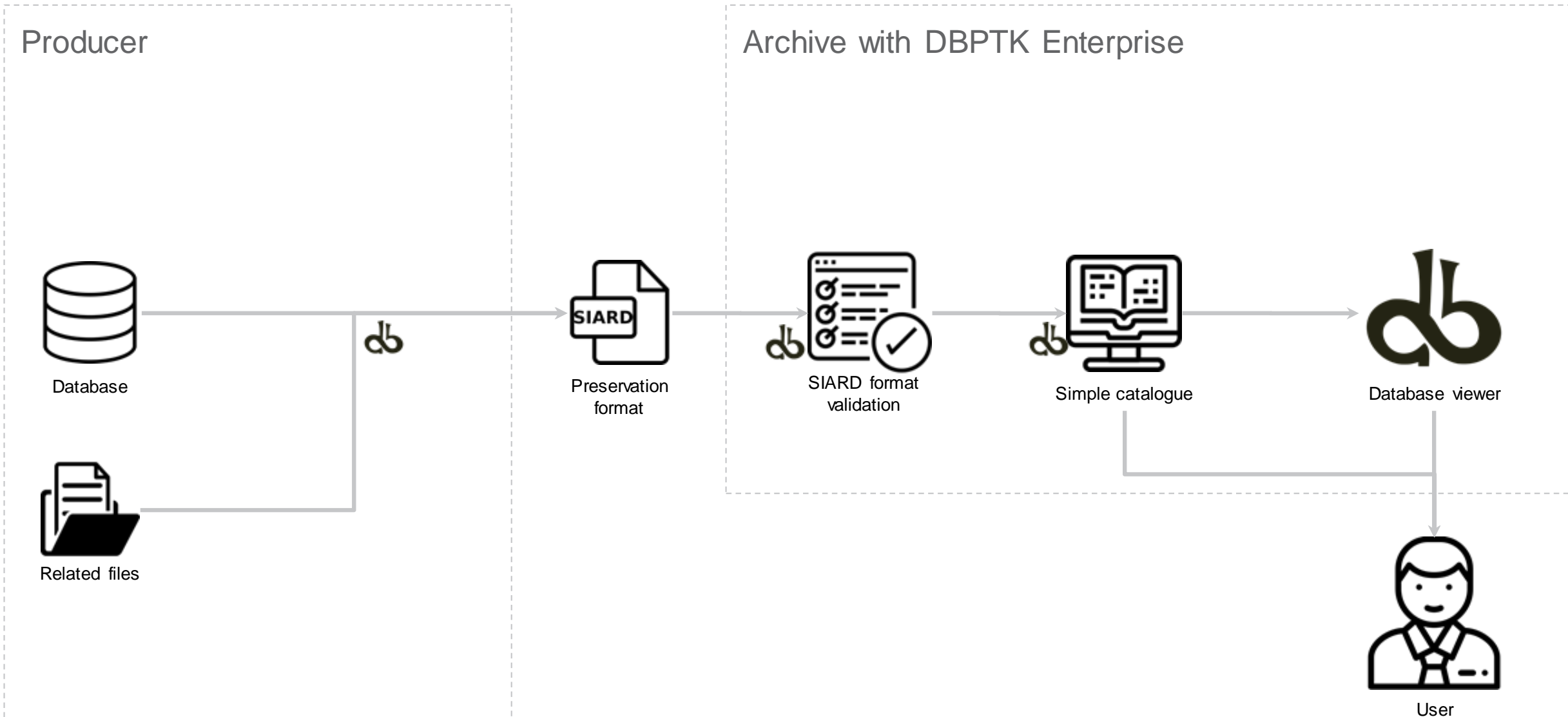
# SIARD: Software Independent Archiving of Relational Databases

- Database preservation format
- 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>

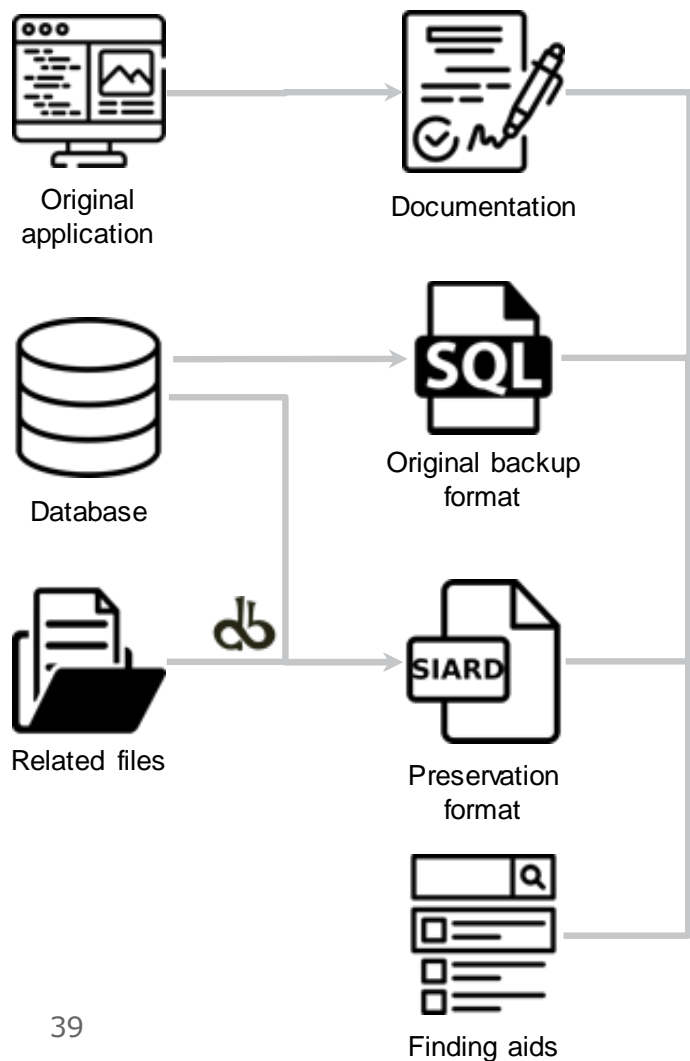
<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eArchiving>

# Simple database archive flow

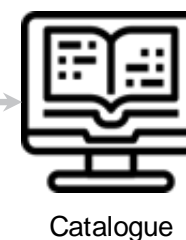
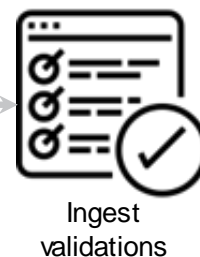
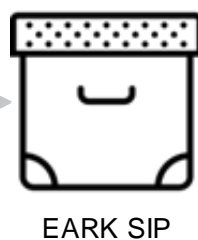


# Full database archival flow

## Producer



## Archive with RODA + DBPTK Enterprise



# 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



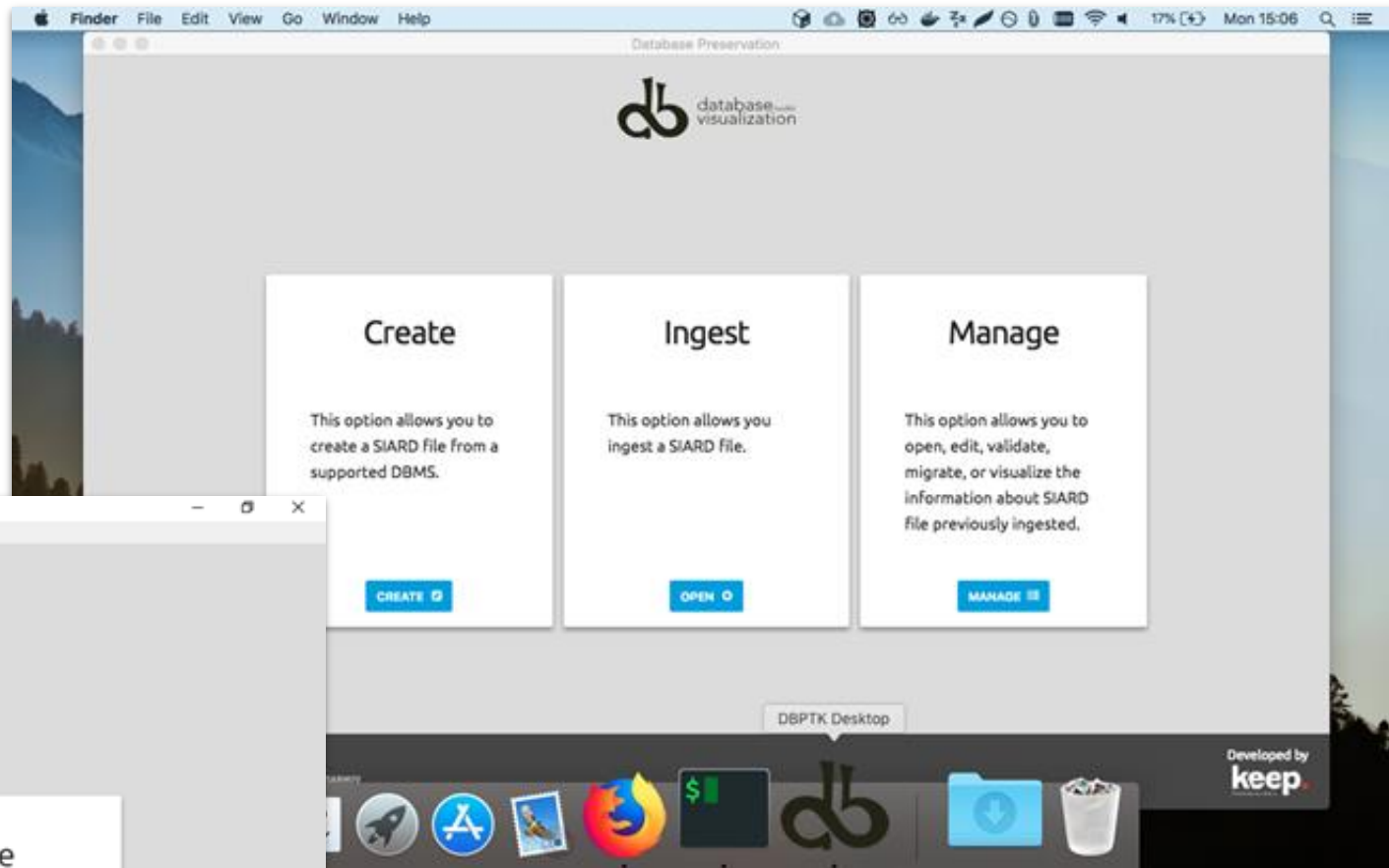
# DBPTK Desktop

Basic features

# DBPTK Desktop features

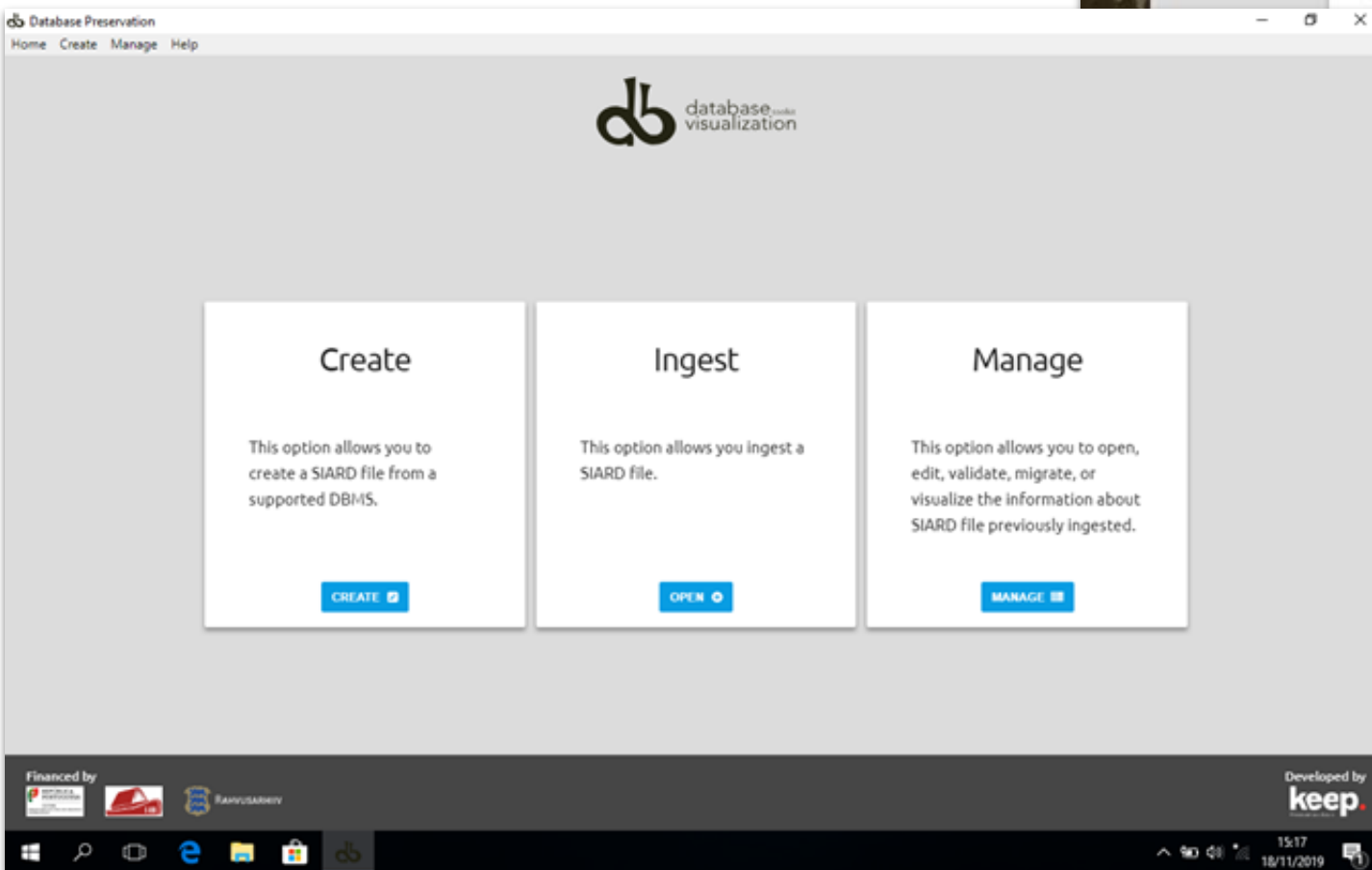


**Windows**



**macOS**

Also available on Linux



# DBPTK Desktop features

## 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

The screenshot shows the 'DBPTK Desktop' application window with the 'Create SIARD - Connection' dialog box open. The 'General' tab is selected, and the 'MySQL' database management system is chosen from the left-hand list. The configuration fields are as follows:

Field	Value	Description
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

Below the fields, there is a checkbox for 'Disable Encryption' which is currently unchecked, with a note: 'Use to turn off encryption in the connection'. A blue 'TEST CONNECTION' button with a lightning bolt icon is located at the bottom of the dialog. At the very bottom of the window, there are 'CANCEL', 'BACK', and 'NEXT' buttons.

# DBPTK Desktop features

## 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)

# DBPTK Desktop features

## 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 M... 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

# DBPTK Desktop features

## 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



# DBPTK Desktop features

## 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 is '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 area is titled 'Search all records' and features a search input field with the text 'dan' and a search button. Below the search bar, two tables are displayed: 'actor' and 'customer'. The 'actor' table shows three records with columns 'actor\_id', 'first\_name', 'last\_name', and 'last\_update'. The 'customer' table shows one record with columns 'customer\_id', 'store\_id', 'first\_name', 'last\_name', and 'email'.

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

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

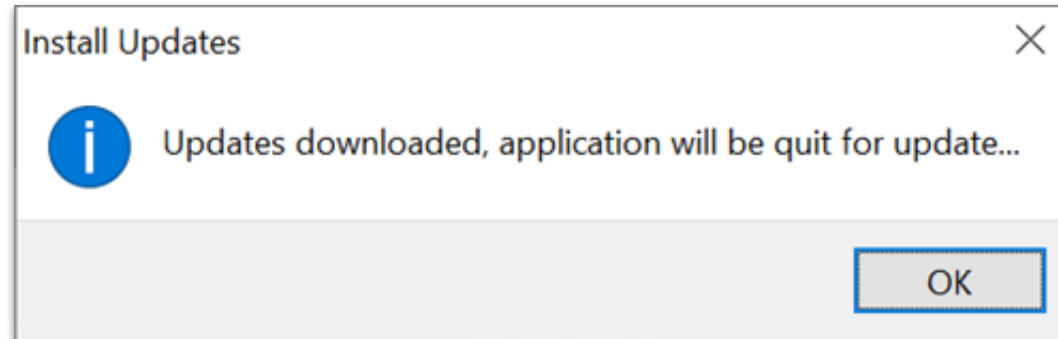
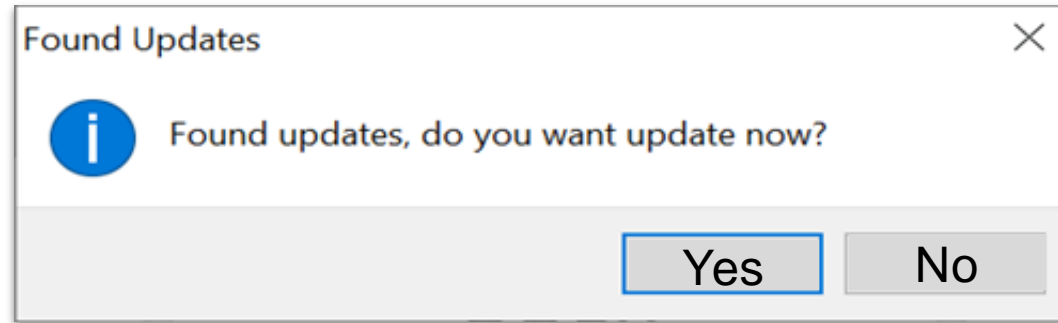


## DBPTK Desktop features

### Auto-update

Automatic check of updates

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





# DBPTK Enterprise

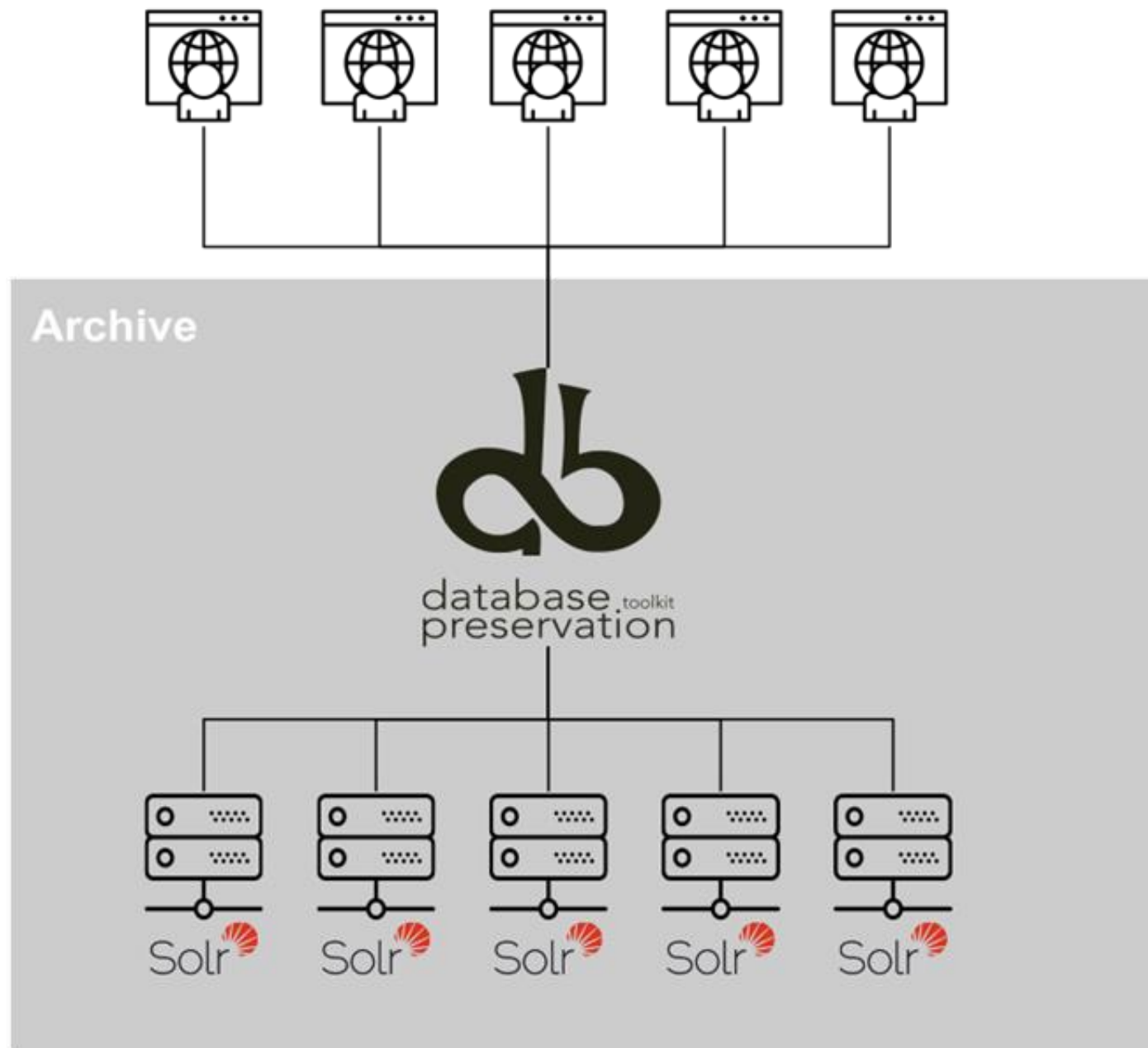
Basic features

## DBTPK Enterprise 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

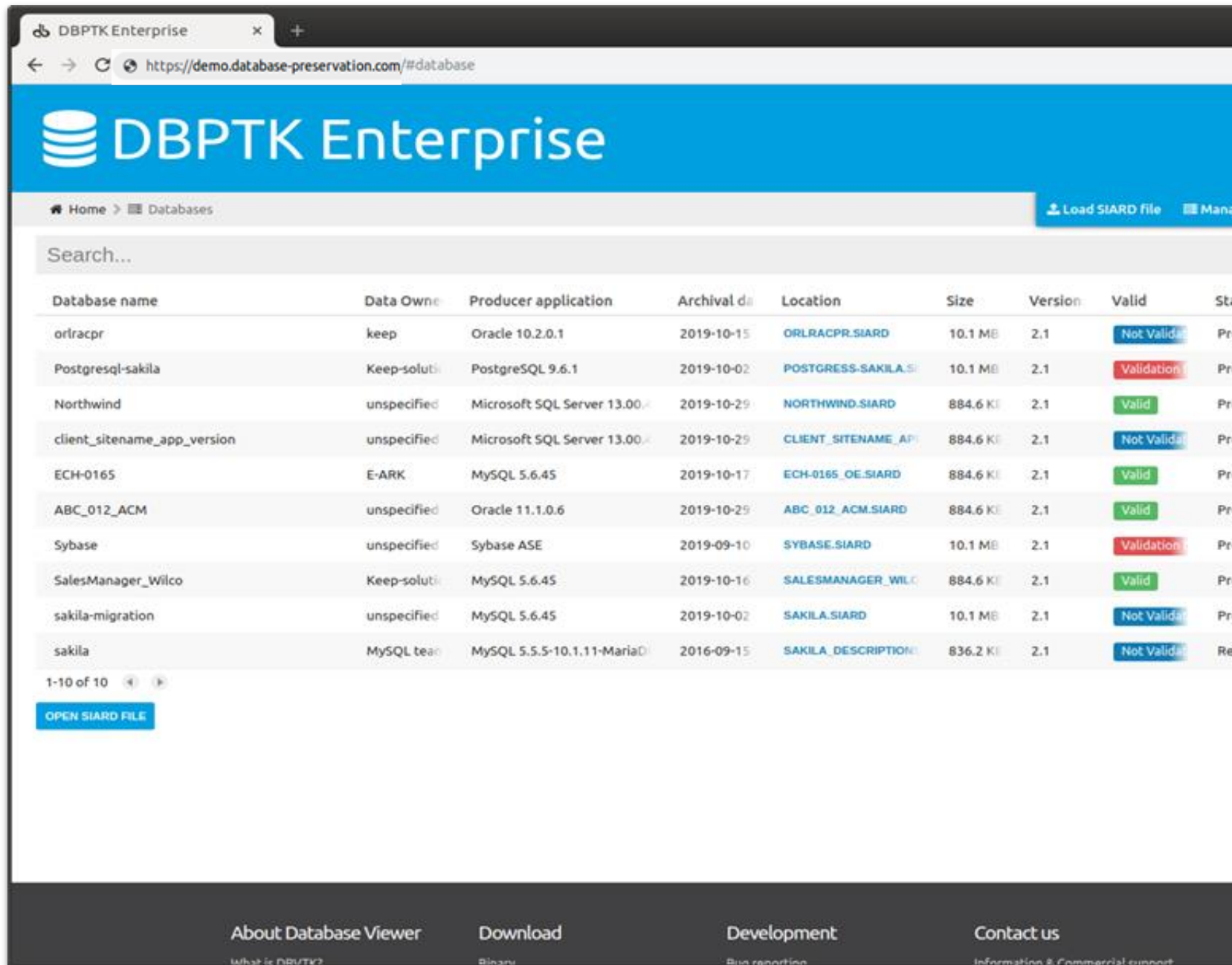


# DBTPK Enterprise features

## 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 displays the DBTPK Enterprise web application interface. The browser address bar shows the URL <https://demo.database-preservation.com/#database>. The page features a blue header with the DBTPK Enterprise logo and navigation links for Home, Databases, Load SIARD file, and Manage. A search bar is located below the header. The main content area contains a table listing various databases with columns for Database name, Data Owner, Producer application, Archival date, Location, Size, Version, Valid status, and Status. The table lists 10 databases, including orlracpr, PostgreSQL-sakila, Northwind, client\_sitenam\_app\_version, ECH-0165, ABC\_012\_ACM, Sybase, SalesManager\_Wilco, sakila-migration, and sakila. The Valid column shows status indicators like 'Not Valid', 'Validation', and 'Valid'. At the bottom of the table, there is a pagination control showing '1-10 of 10' and an 'OPEN SIARD FILE' button. The footer includes links for About Database Viewer, Download, Development, and Contact us, along with the European Commission logo.

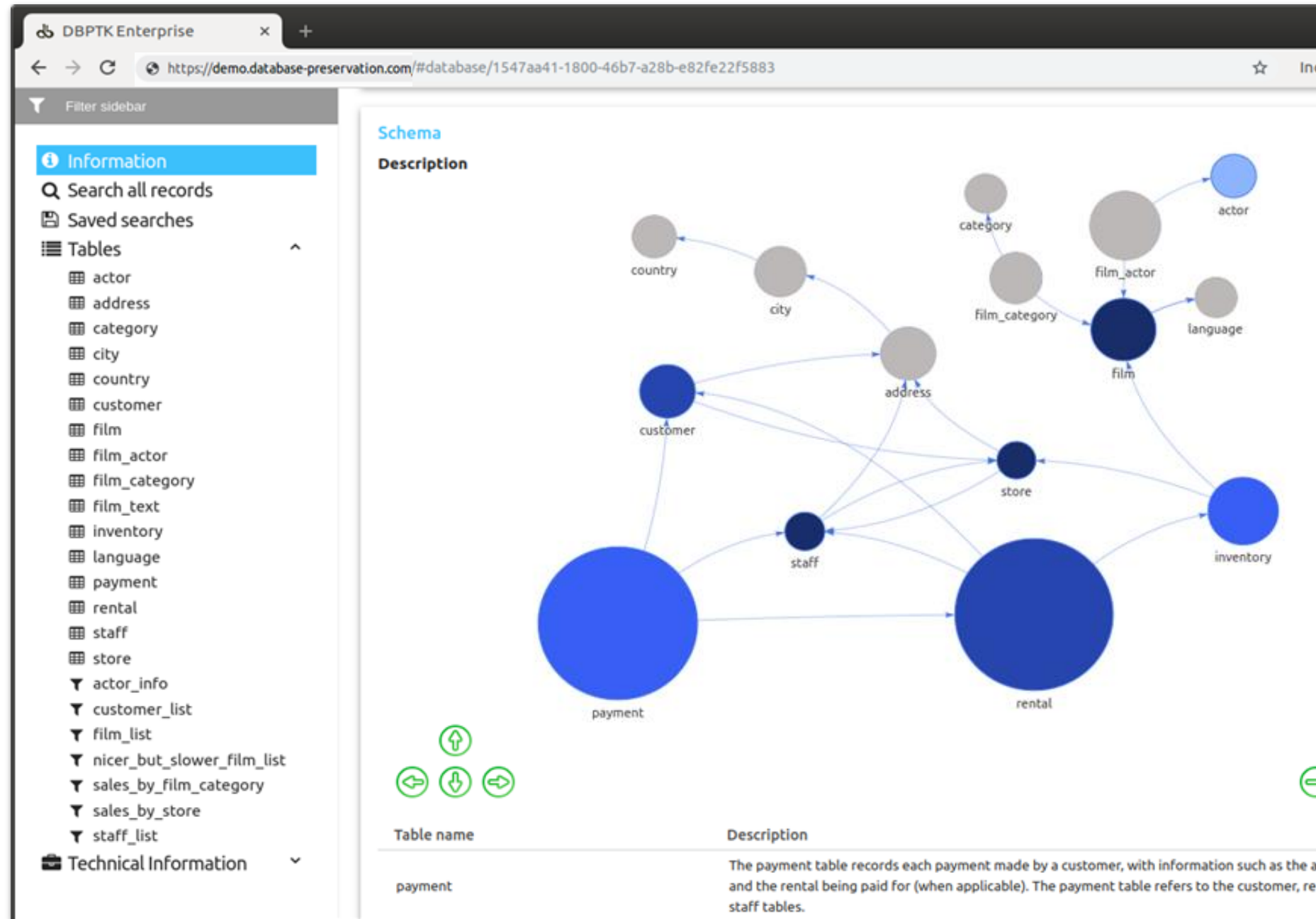
Database name	Data Owner	Producer application	Archival date	Location	Size	Version	Valid	Status
orlracpr	keep	Oracle 10.2.0.1	2019-10-15	<a href="#">ORLRACPR.SIARD</a>	10.1 MB	2.1	Not Valid	Pr
Postgresql-sakila	Keep-soluti	PostgreSQL 9.6.1	2019-10-02	<a href="#">POSTGRESS-SAKILA.SI</a>	10.1 MB	2.1	Validation	Pr
Northwind	unspecified	Microsoft SQL Server 13.00	2019-10-29	<a href="#">NORTHWIND.SIARD</a>	884.6 KB	2.1	Valid	Pr
client_sitenam_app_version	unspecified	Microsoft SQL Server 13.00	2019-10-29	<a href="#">CLIENT_SITENAME_AP</a>	884.6 KB	2.1	Not Valid	Pr
ECH-0165	E-ARK	MySQL 5.6.45	2019-10-17	<a href="#">ECH-0165_OE.SIARD</a>	884.6 KB	2.1	Valid	Pr
ABC_012_ACM	unspecified	Oracle 11.1.0.6	2019-10-29	<a href="#">ABC_012_ACM.SIARD</a>	884.6 KB	2.1	Valid	Pr
Sybase	unspecified	Sybase ASE	2019-09-10	<a href="#">SYBASE.SIARD</a>	10.1 MB	2.1	Validation	Pr
SalesManager_Wilco	Keep-soluti	MySQL 5.6.45	2019-10-16	<a href="#">SALESMANAGER_WILCO</a>	884.6 KB	2.1	Valid	Pr
sakila-migration	unspecified	MySQL 5.6.45	2019-10-02	<a href="#">SAKILA.SIARD</a>	10.1 MB	2.1	Not Valid	Pr
sakila	MySQL team	MySQL 5.5.5-10.1.11-MariaDB	2016-09-15	<a href="#">SAKILA_DESCRIPTION</a>	836.2 KB	2.1	Not Valid	Re

# DBTPK Enterprise features

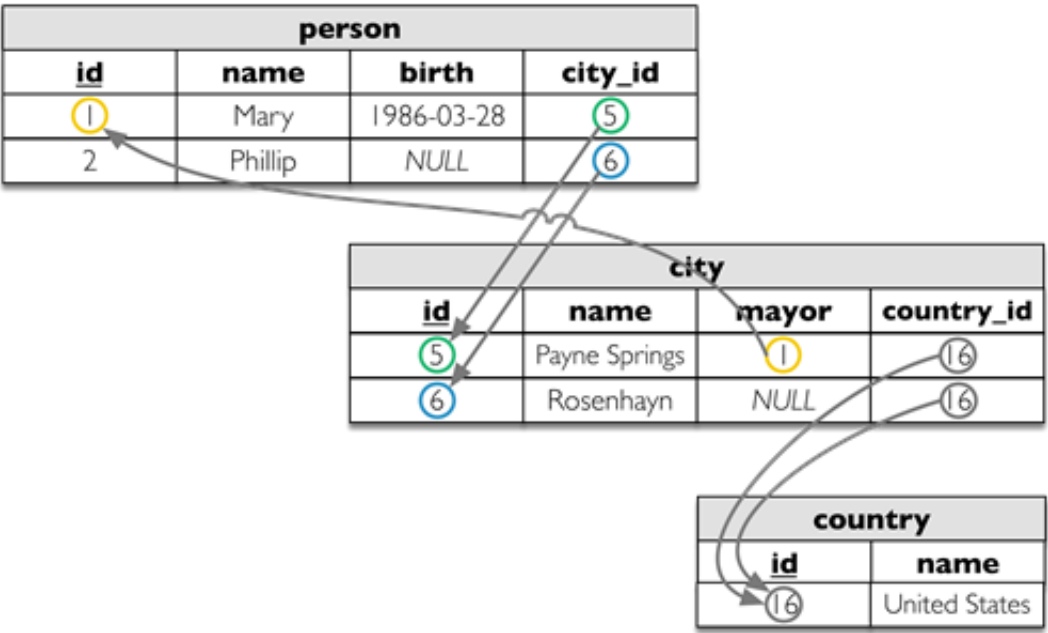
## Data transformation

Transform content to answer useful questions

- **De-normalization** and table and **column hiding**, to simplify browsing and allow **anonymization** of content



## Data transformation (aka denormalization)



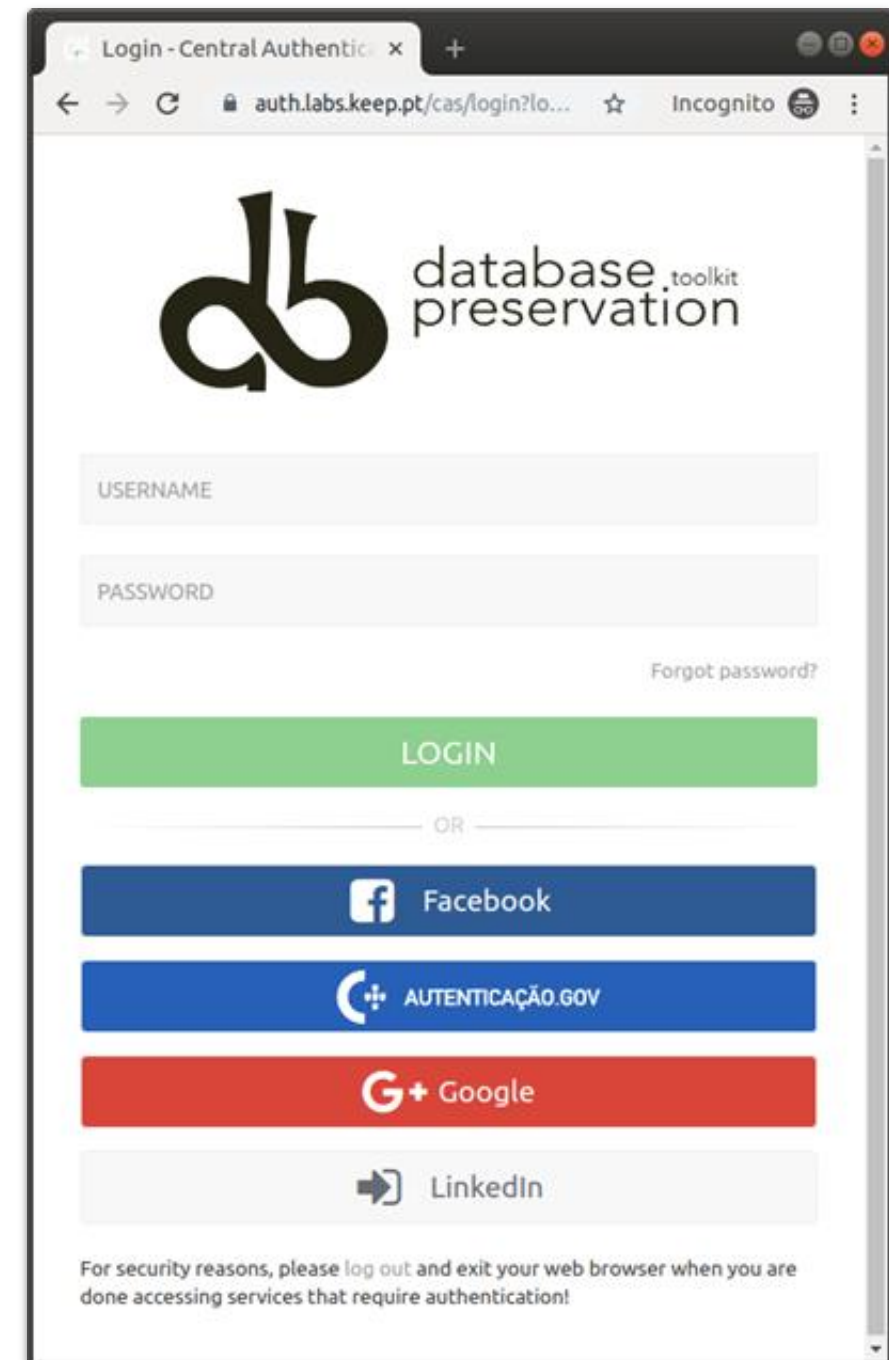
person				
Name	Birth	City name	Mayor	Country name
Mary	1986-03-28	Payne Springs	<u>Mary</u>	United States
Phillip		Rosenhayn		United States

# DBPTK Enterprise features

## 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



# DBPTK Enterprise features

## 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 DBPTK Enterprise web application interface. The browser address bar shows the URL: <https://demo.database-preservation.com/#table/1547aa41-1800-46b7-a28b-e82fe22f5883/cf931074-1079-4d1f-8212-6b519c78fa81/update>. The interface features a blue header with the 'sakila' logo and navigation links: Home, Databases, sakila, and film. A sidebar on the left lists various database tables, with 'film' currently selected. The main content area displays the 'film' table, including a description and a search interface. The search interface has a search bar and filters for 'release\_year' (set to 2006) and 'rating' (set to PG-13). Below the search bar are buttons for 'ADD SEARCH FIELD', 'CLEAR', 'SAVE SEARCH', and 'SEARCH'. A table of film records is shown below the search interface.

title	description	release_year	length	rating	special_features
The title of the film.	A short description or p...	The year in which the ...	The duration of the fil...	The rating assigned to ...	Lists which common sp...
AIRPLANE SIERRA	A Touching Saga of a t...	2006	62	PG-13	Trailers, Deleted Sce...
ALABAMA DEVIL	A Thoughtful Panorarr	2006	114	PG-13	Trailers, Deleted Sce...
ALTER VICTORY	A Thoughtful Drama o	2006	57	PG-13	Trailers, Behind the !
ANTHEM LUKE	A Touching Panorama	2006	91	PG-13	Deleted Scenes, Beh
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 !

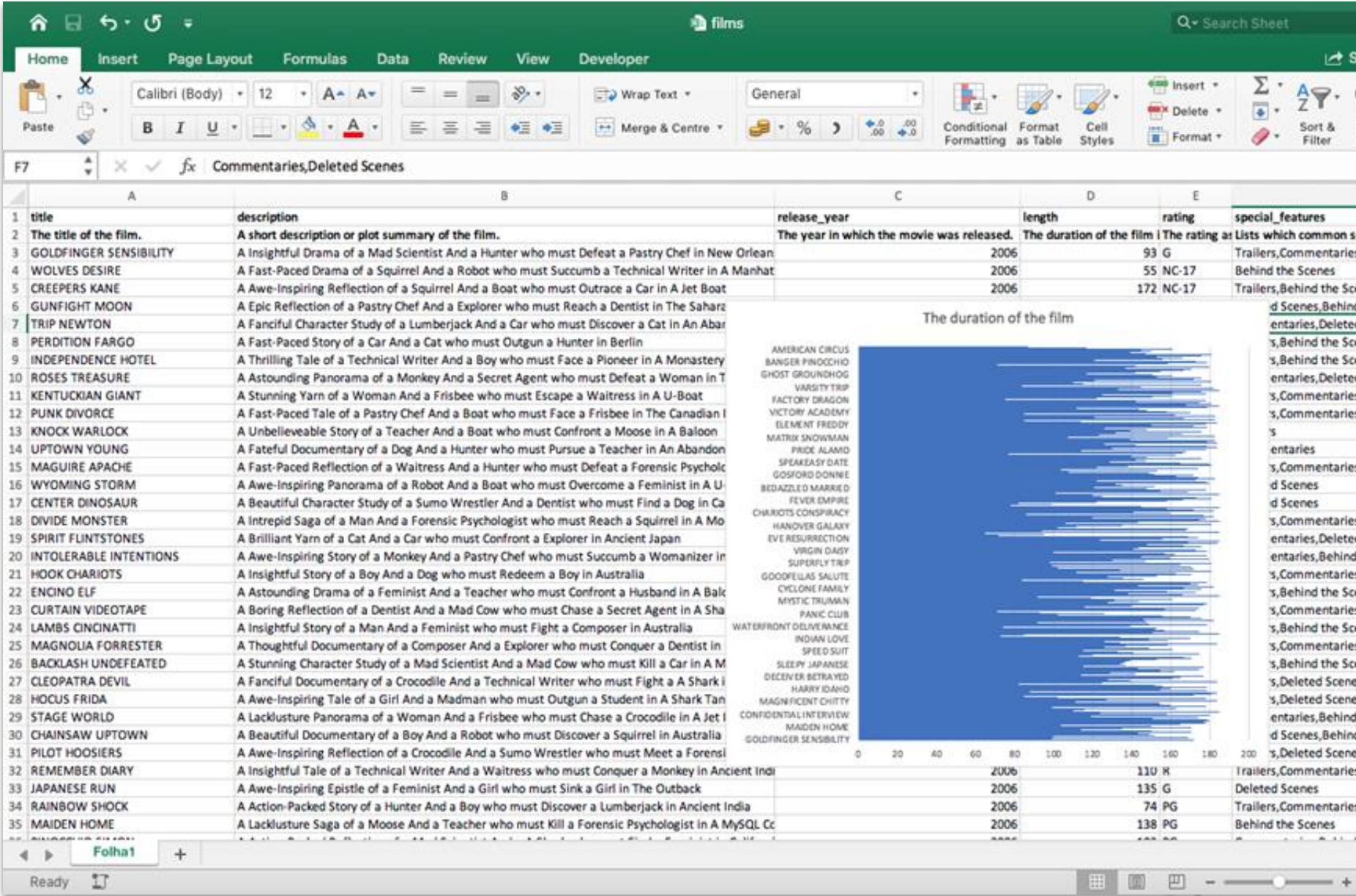


# DBPTK Enterprise features

## 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



# DBPTK Enterprise features

## Activity log

Audit every access

- Who has done what, when and from where.
- Requirement for ISO 16363 certification.

**DBPTK Enterprise**

Databases > Activity log

### 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...

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)

---

# DBPTK Enterprise & Desktop

## 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)



# DBPTK Developer

Basic features



# DBPTK Developer features

## 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 + [ ] [ ] Tlilx: 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

~ $ _
```

# DBPTK Developer features

## 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 artifact structure under the path `2.6.3`, including `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 set to "Maven" and the corresponding XML snippet:

```
<dependency>
  <groupId>com.databasepreservation</groupId>
  <artifactId>dbptk-model</artifactId>
  <version>2.6.3</version>
</dependency>
```

# DBPTK Developer features

## Open source

For custom development

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

The screenshot shows the GitHub repository page for `keeps/dbptk-developer`. The repository is under the `master` branch, has 10 branches, and 55 tags. It features 1,246 commits and 71 issues. The repository is described as a library and command-line tool for database preservation actions. The file list includes folders like `.github`, `.travis`, `code-style`, `dbptk-bindings`, `dbptk-core`, `dbptk-model`, `dbptk-modules`, `dbptk-plugin-example`, `doc`, `examples`, `scripts`, and `testing`, as well as files like `.gitattributes`, `.gitignore`, and `.grenrc.vml`. The right sidebar shows the repository's description, links to the website, a README, license, releases (latest is 2.9.2), and contributors.

Why GitHub? Team Enterprise Explore Marketplace Pricing Search Sign in Sign up

keeps / dbptk-developer Watch 15 Star 31 Fork 11

<> Code Issues 71 Pull requests 1 Actions Projects 1 Wiki Security Insights

master 10 branches 55 tags Go to file Code

hmiguim Setting version 2.10.0-SNAPSHOT ✓ 5ced0bc 17 hours ago 1,246 commits

.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
.grenrc.vml	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

Readme View license

Releases 55 Version 2.9.2 Latest 17 hours ago + 54 releases

Contributors 11

---

# And many more features

## For archiving databases:

- SSH Tunnel
- Selection of tables and columns
- Selection and materialization of views
- Custom views
- External files (files stored outside the DB)
- External files via SSH tunnel
- Automated quality assurance
- Save LOBs outside SIARD file
- Migrate from SIARD to SIARD
- Migrate from SIARD to live DBMS
- Convert ORACLE geodata

## For accessing archived databases:

- Configure visible tables
- Configure visible columns
- Set column name, description and order
- Binary columns advanced options
- REST API
- Load on access and auto-unload



# DBPTK

	Desktop	Enterprise	Developer
Save to preservation format	✓	✓*	✓
Quality assurance (merkle tree)	✓	✓*	✓
Validation	✓	✓	✓
Enrich descriptions	✓	✓	✓
Browse and search	✓	✓	✗
Transform (de-normalization)	✗	✓	✗
Export to live databases	✓	✓*	✓
Activity Log	✗	✓	✗
Authentication	✗	✓	✗
Number of users	one	many	one
Number of loaded databases	few	many	N/A

\* Enterprise feature done via the upload/download of SIARD and usage of related tools

***DEMONSTRATION***

---

# Database preservation

---

Real-world use cases

---

# Hospital legacy databases

## Context

Set of **database systems** created to support **specific hospital services** (cardiothoracic, neonatology and neutropenia)

They contain **crucial information** about the **history of some patients** that may be needed for **urgent interventions**

## Problem

- Databases were **replaced** by newer systems
- Information was **never migrated** to newer systems
- Original Database Management Systems are **obsolete**
- Original developers and submitters are **gone**
- **Not enough documentation** is available

---

# Hospital legacy databases

## Solution

- **Export** of all information into **SIARD**
- **Expert analysis** of original database and interfaces to create **documentation**
- Using **RODA** to keep documentation and **DBPTK Enterprise** to provide access
- Use table and column management and data transformation to make databases more **user-friendly** and **better documented**.

## Main software used

- DBPTK Desktop for export into SIARD
- RODA for catalogue and archiving representation information (documentation)
- DBPTK Enterprise for access to database content

## Main features used

- Custom views and materialized views
- SIARD metadata edition
- Table and column management
- Data transformation

---

# European Taxation and Customs Union: trader messages archive

## Context

New **EU service** that will provide a **centralized interface with customs authorities** for **thousands of economic operators** that bring the **goods into the European Union**.

All **transaction messages** will need to be **archived for a decade**.

## Problem

- Estimated **10 million messages per day**
- Production database needs to **offload to archive** daily and purge information
- Must **ensure no message is lost** or mangled in the archival process
- Archive process must **keep up with production**

---

# European Taxation and Customs Union: trader messages archive

## Solution

- Archive partial exports of database into SIARD (e.g. 1-hour timespans)
- Archive into RODA and load into DBPTK Enterprise when access is needed
- Continuous extraction, archive and validation workflow
- **Quality assurance is key**

## Main software used

- DBPTK Developer for continuous partial export to SIARD
- RODA for archival, search and load into DBPTK Enterprise
- DBPTK Enterprise to access on request and retrieve original message(s)

## Main features used

- DBPTK developer automation scripts
- Automated quality assurance

# Questions?

**Luis Faria**  
**Research & Innovation Director**

KEEP SOLUTIONS  
lfaria@keep.pt

## **E-ARK Programme**

LinkedIn: [www.linkedin.com/groups/8343650/](https://www.linkedin.com/groups/8343650/)

Twitter: #EARKProject

# Ready to get started?

**Find out more at:**  
[ec.europa.eu/cefdigital](https://ec.europa.eu/cefdigital)

**Contact us:**  
[cef-building-blocks@ec.europa.eu](mailto:cef-building-blocks@ec.europa.eu)

# Thank you!

