

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

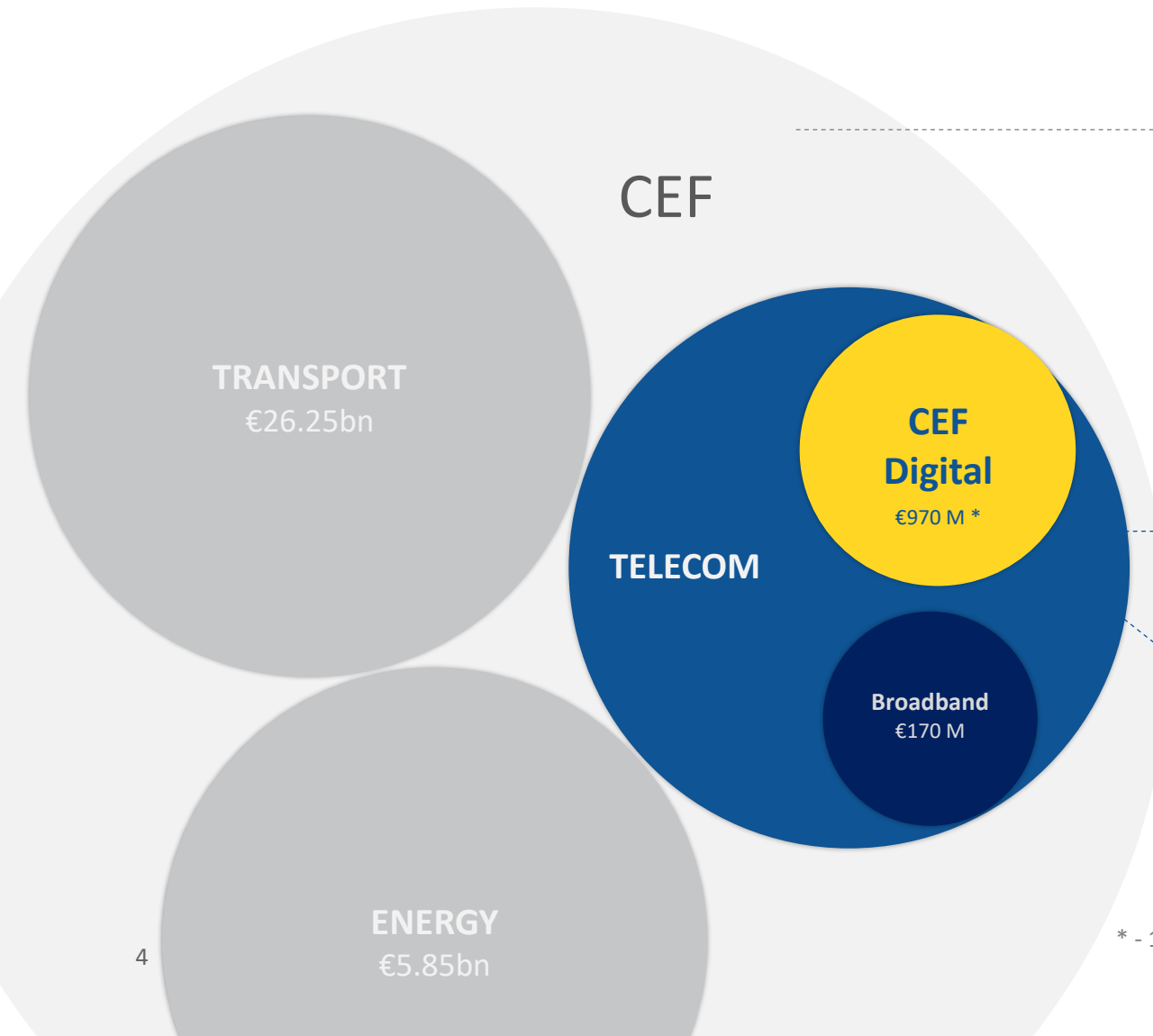
Welcome to the Connecting Europe Facility (CEF) Building Blocks

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.

* - 100 M Juncker Package



Big Data Test Infrastructure

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



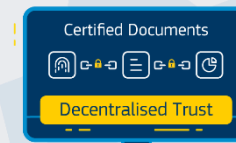
Context Broker

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



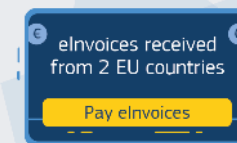
eArchiving

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



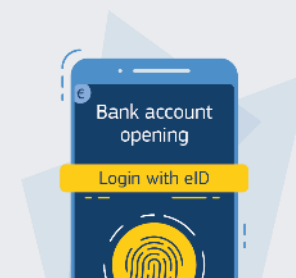
European Blockchain Services Infrastructure

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



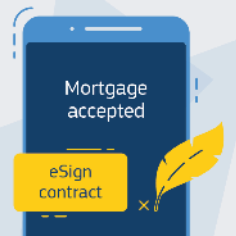
eID

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



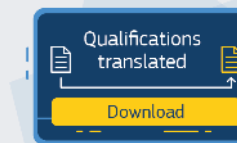
eDelivery

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 eSubmission system.

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

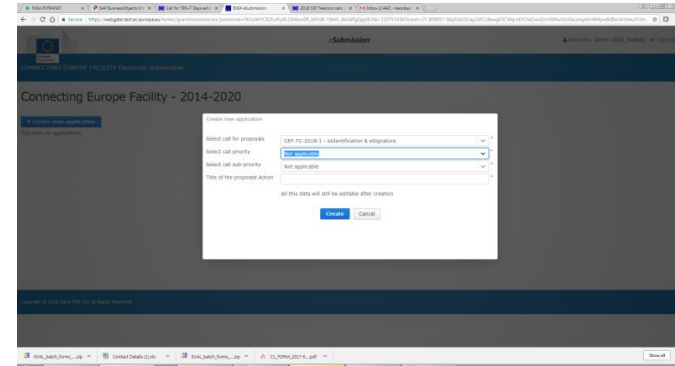
Watch the **2020-2 CEF Telecom call virtual Info Day**

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	
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



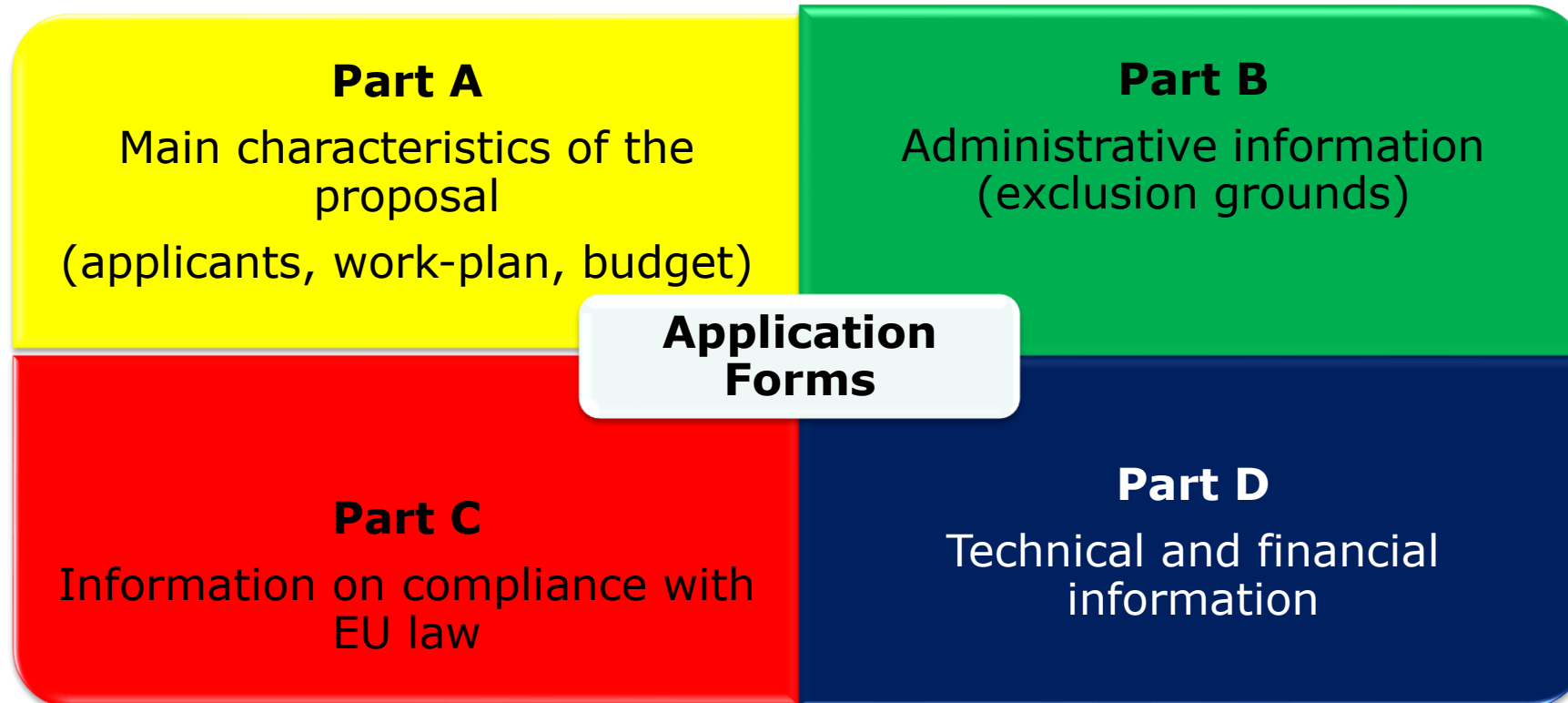
- **TENtec:** system used to manage the CEF projects during their entire lifecycle and which enables the electronic submission of proposals under the CEF calls
- Step-by-step instructions for the [TENtec eSubmission system](#) in the [Guide for Applicants](#)



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

- Alignment to DSI implementation objectives & activities (WP)
- Alignment to EU/national policies, strategies and activities

Quality & efficiency of implementation

- Maturity
- Coherence/effectiveness with work plan
- Quality of consortium/consortium members
- Support from national authorities/industry/NGOs
- Attention to security/privacy/inclusiveness/accessibility

Impact & sustainability

- Quality of the approach to facilitate wider deployment/take-up of the proposed actions
- Capability of long-term sustainability without EU funding

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
 - [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



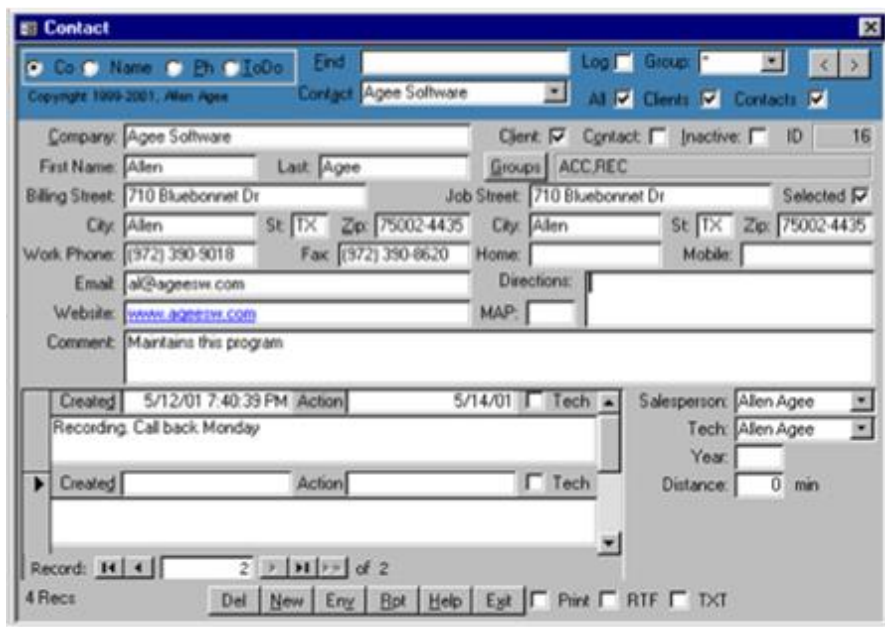
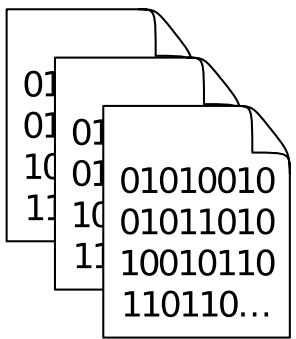
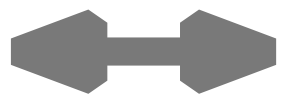
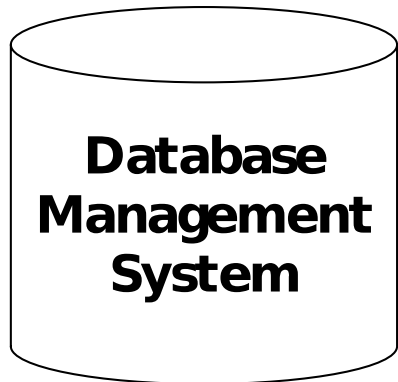
Databases

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.



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

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 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>

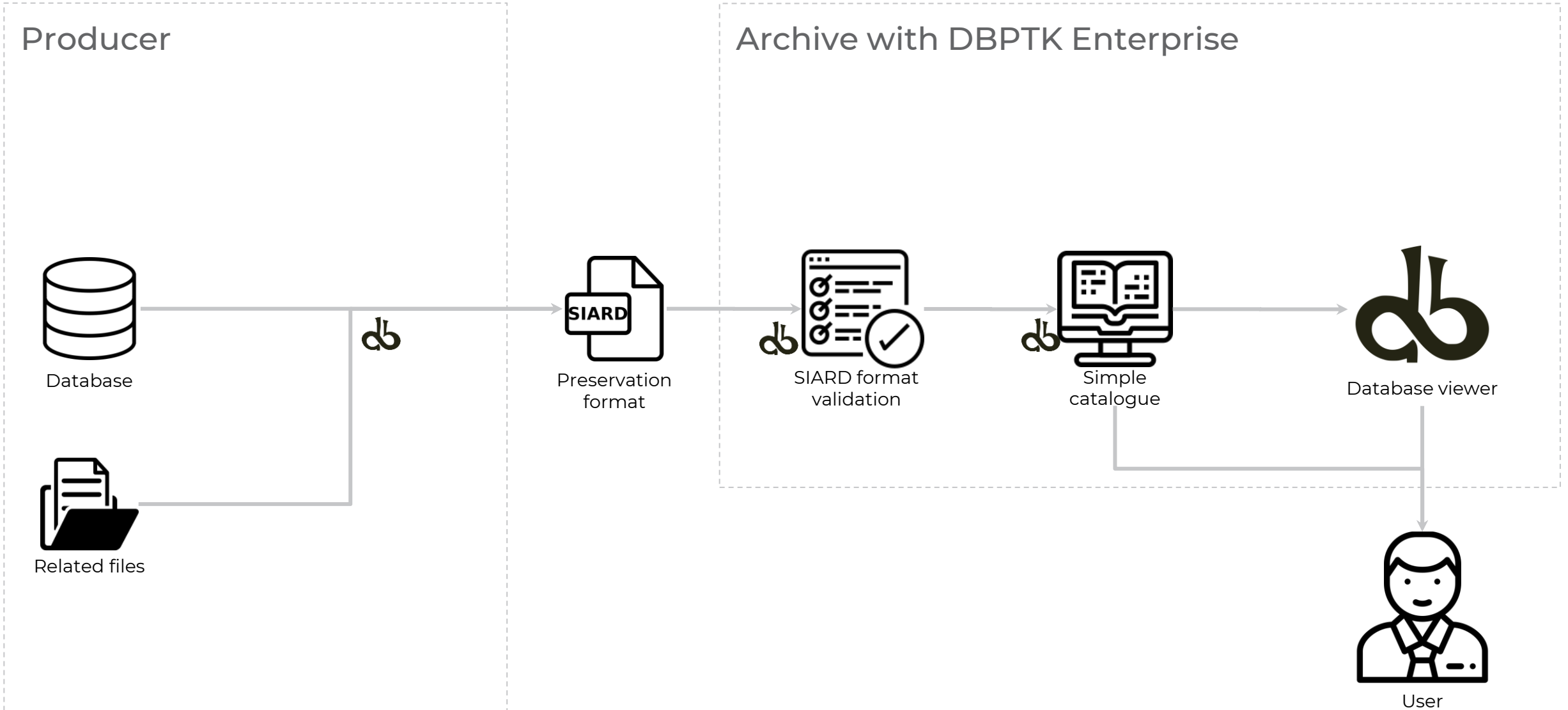
SIARD: Software Independent Archiving of Relational Databases

- 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>

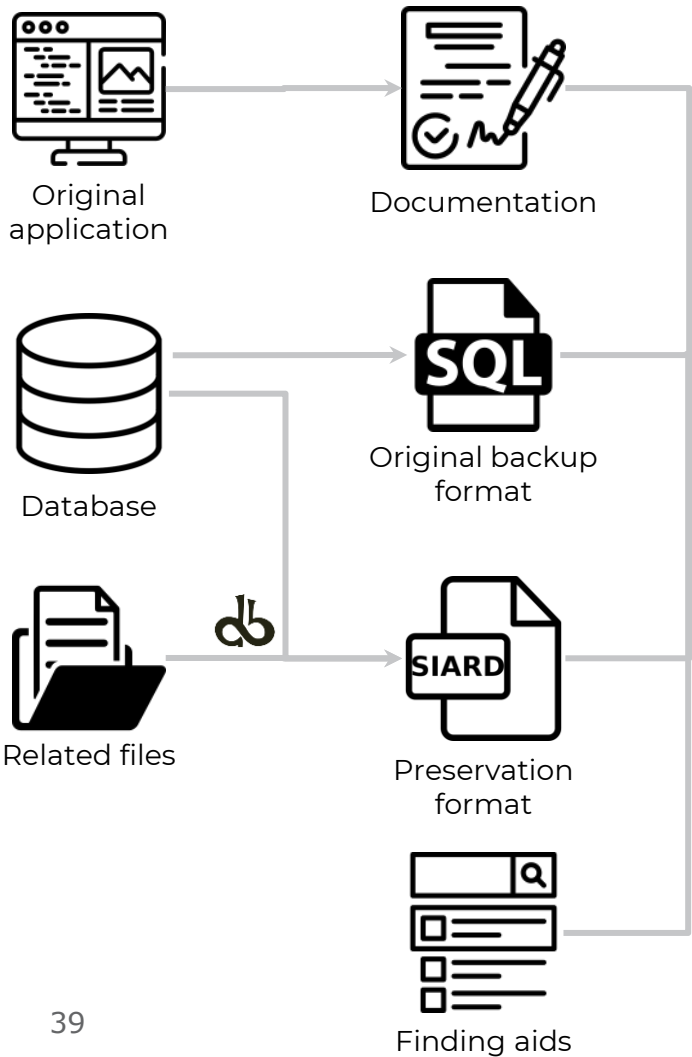
<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eArchiving>

Simple database archive flow

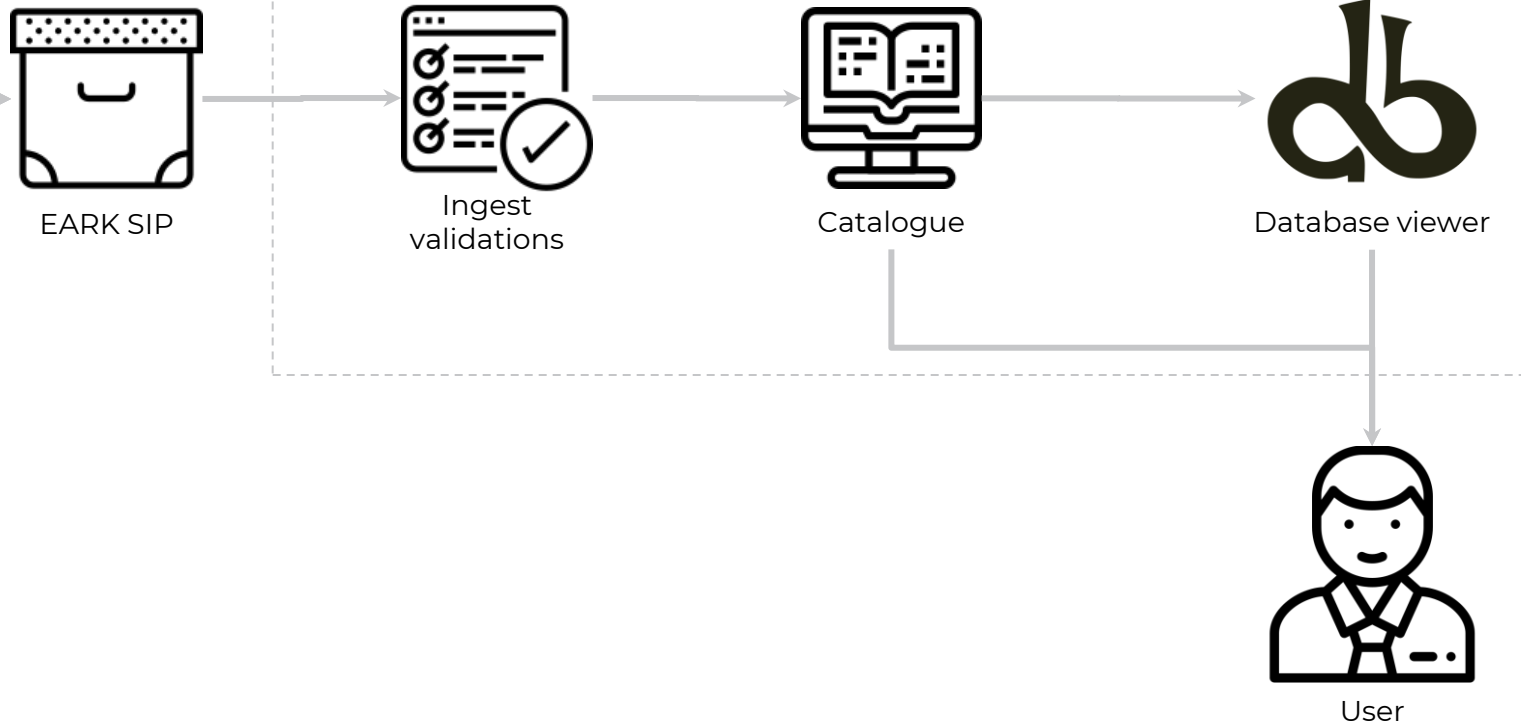


Full database archival flow

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>



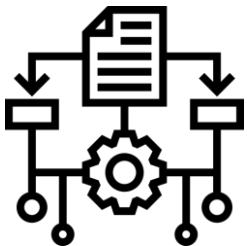
DBPTK Desktop

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

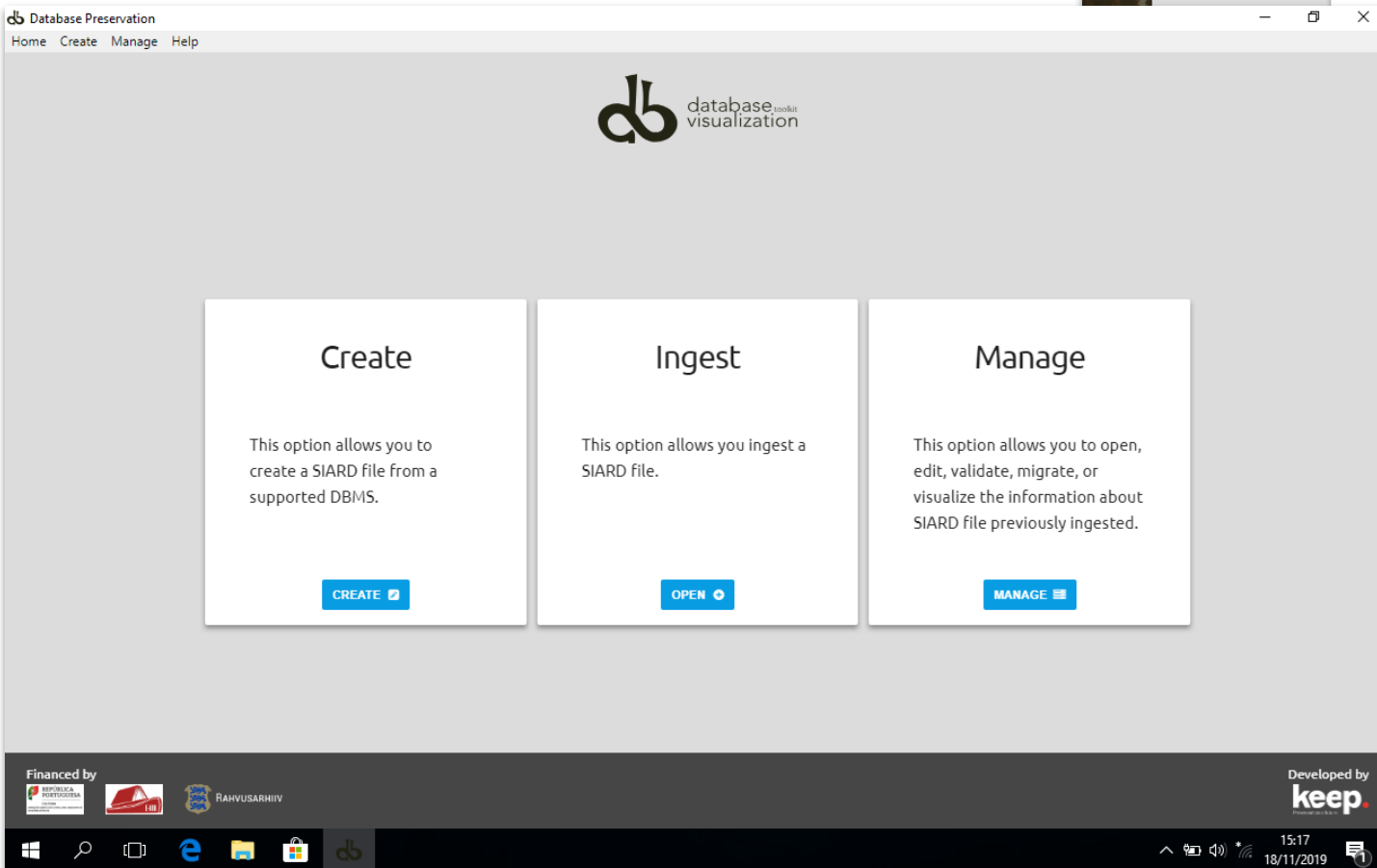
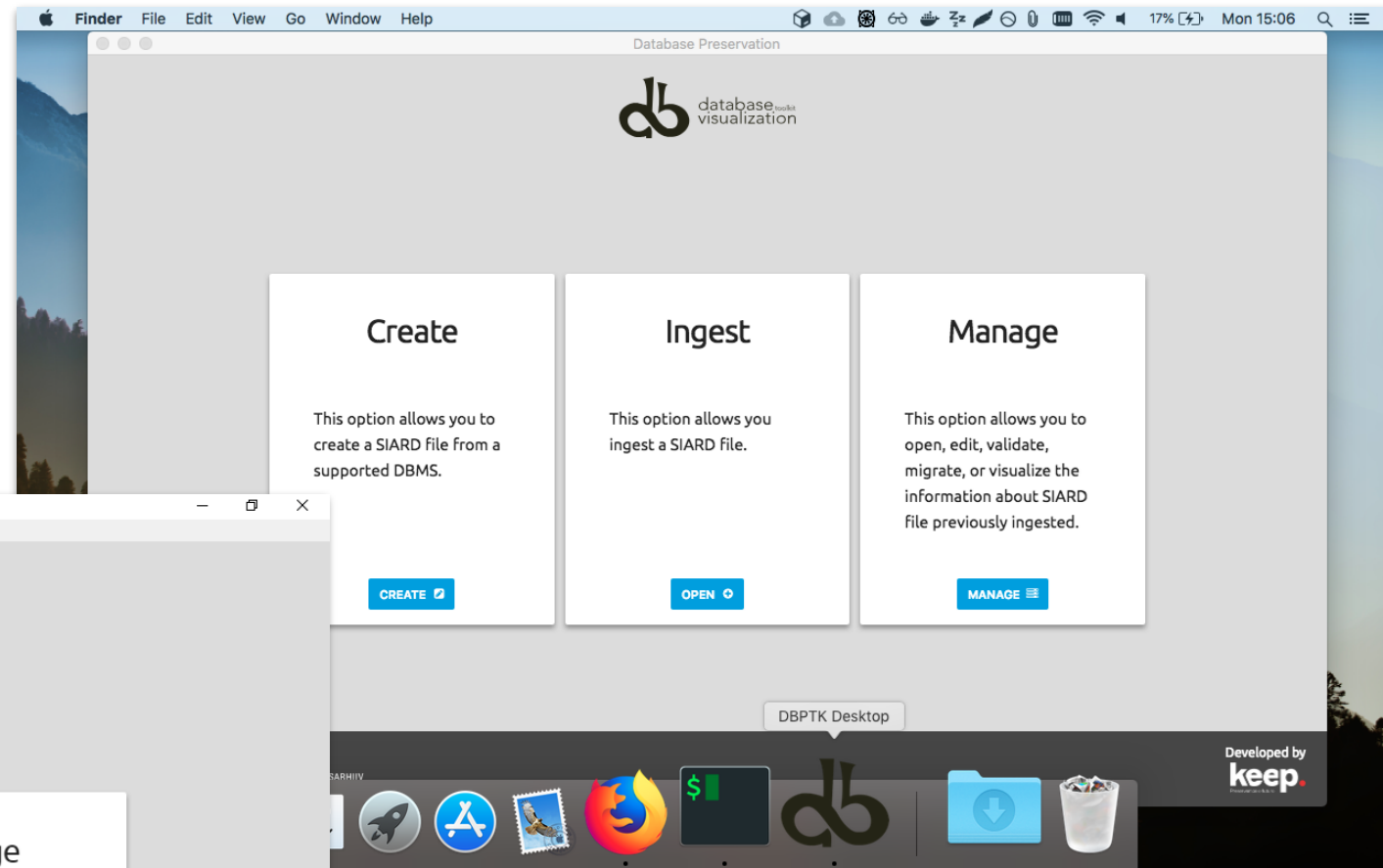
A command-line tool and development library for automation and system integration



DBPTK Desktop

Basic features

DBPTK Desktop features



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' interface for creating a connection. The breadcrumb path is 'Home > Create SIARD - Connection'. On the left, a list of DBMS options includes JDBC, Microsoft Access, Microsoft SQL Server, MySQL (highlighted), Oracle, PostgreSQL, Progress Openedge, and Sybase. The right pane shows the 'General' tab with the following configuration:

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 an unchecked checkbox for 'Disable Encryption' with the 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 configuration area. At the very bottom of the window, there are 'CANCEL' and 'BACK' buttons, and a partially visible 'NEXT' button.

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-lobs = false
- external-lobs-per-folder = 1000
- external-lobs-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)

DBPTK Desktop features

Edit SIARD metadata

Enrich archived database with descriptions

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

The screenshot displays the DBPTK Desktop interface. The top navigation bar includes 'Home', 'Create', 'Manage', and 'Help'. The breadcrumb trail shows 'Home > Databases > sakila > SIARD Edit Metadata'. A 'Filter sidebar' is visible on the left. The main content area is titled 'Database Information' and shows global information at the database level for 'sakila'. The information includes: Name (*), Archival date (*), Archivist, Archivist contact, Client machine, Product, User, Data origin time span (*), Data owner (*), Description, and Producer application.

Global information at database level	
Name *	sakila
Archival date *	2016-09-15
Archivist	Bruno Ferreira
Archivist contact	email: bferreira@keep.pt
Client machine	mguimaraes
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... standard schema that can be used for examples in boo... and so forth. Sakila sample database also serves to high... MySQL such as Views, Stored Procedures, and Triggers... 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

The screenshot shows the DBPTK Desktop application window. The title bar reads "DBPTK Desktop". The menu bar includes "Home", "Create", "Manage", "Preferences", and "Help". The breadcrumb navigation shows "Home > Databases > sakila > Validation". The main heading is "Validation" with a checkmark icon. Below the heading is a descriptive paragraph: "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." The interface displays the following summary data:

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		

Below the summary is a scrollable log of validation events. A "Scroll to the end" button is visible in the top right corner of the log area. The log entries are as follows:

ID	Message	Status
T_6.4-2	validation finish on path: content/schema1/table15/table15.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 interface. The top navigation bar includes 'Home', 'Create', 'Manage', and 'Help'. The breadcrumb path is 'Home > Databases > sakila > Search'. A search bar contains the text 'dan'. The left sidebar shows a 'Filter sidebar' with a list of tables: 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 'Search all records' option is selected. The main area displays two search results:

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

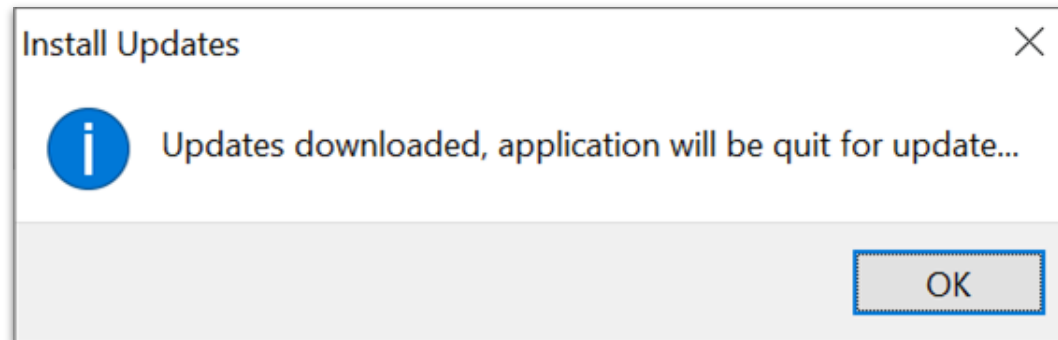
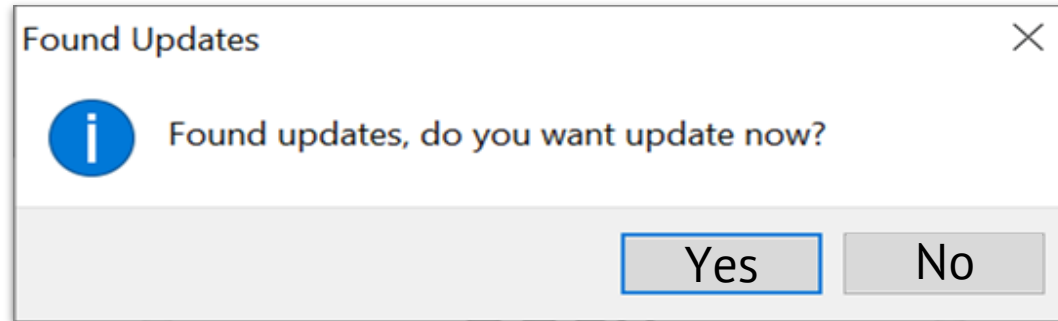
1-1 of 1

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

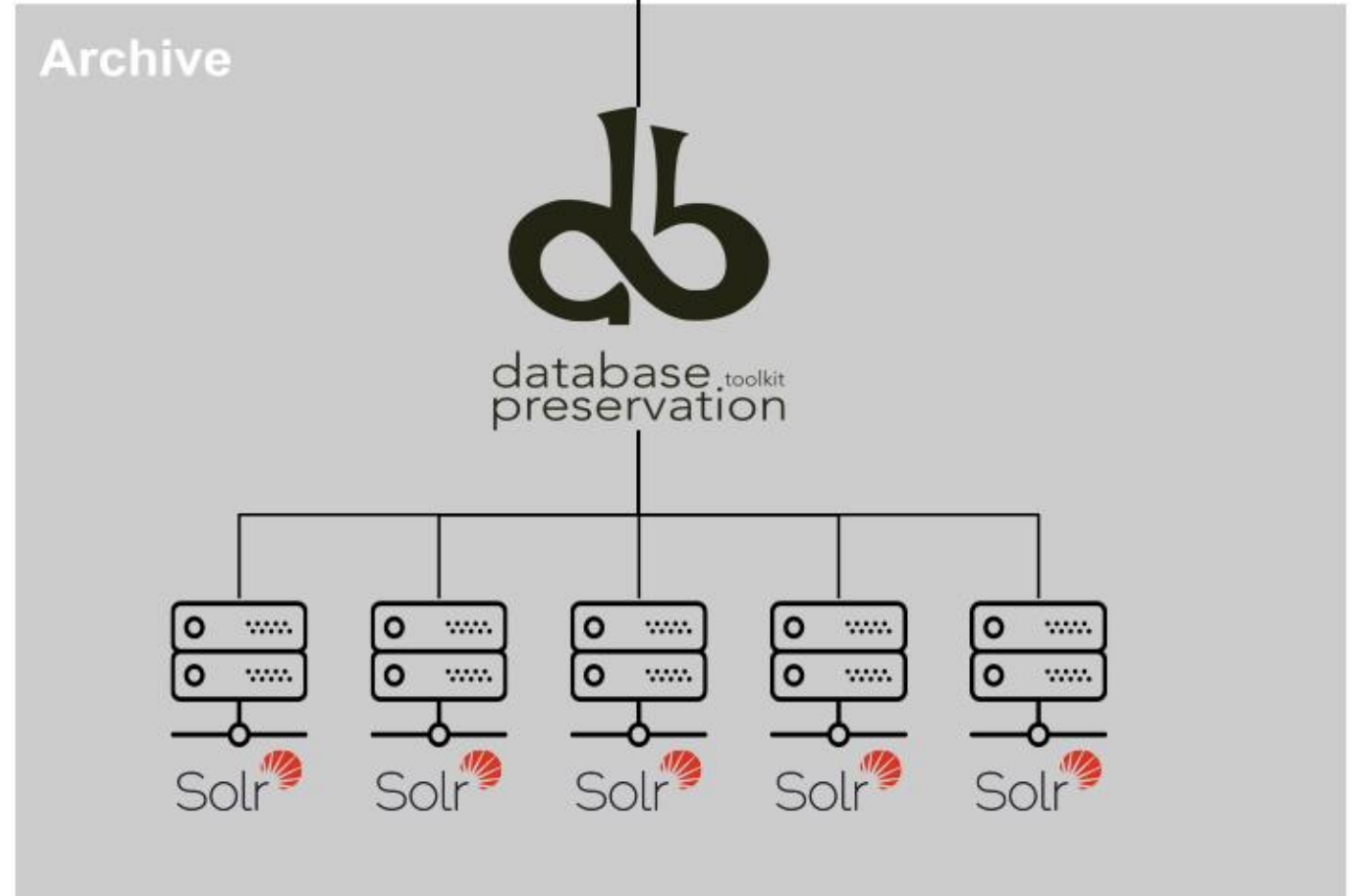
Basic features

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

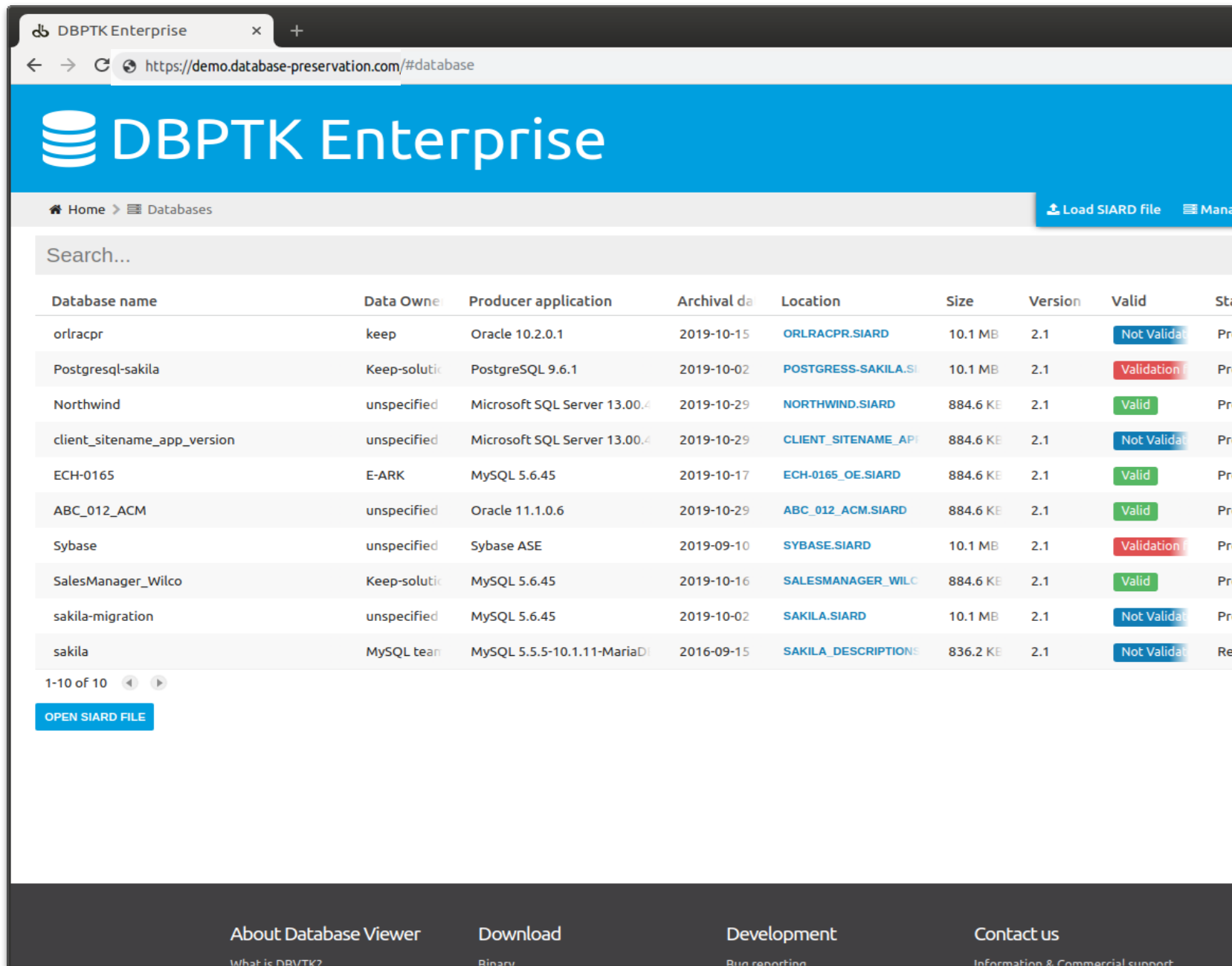


DBTPK Enterprise features

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 header features the DBTPK Enterprise logo and navigation links for Home and Databases. A search bar is present above a table listing various databases. The table columns include Database name, Data Owner, Producer application, Archival date, Location, Size, Version, Valid, and Status. The 'Valid' column uses color-coded buttons: blue for 'Not Valid', red for 'Validation', and green for 'Valid'. A pagination control shows '1-10 of 10' items, and an 'OPEN SIARD FILE' button is located below the table. 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-solutio	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_API	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-solutio	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

DBTPK Enterprise features

Data transformation
Transform content to answer useful questions

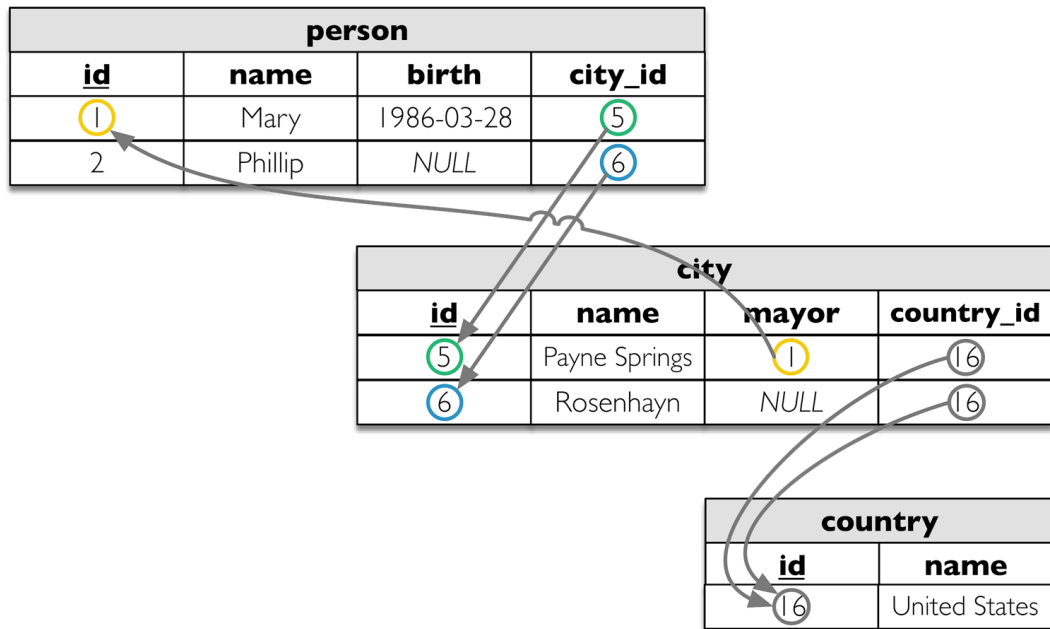
- **De-normalization** and table and **column hiding**, to simplify browsing and allow **anonymization** of content

The screenshot shows the DBTPK Enterprise web interface. On the left is a sidebar with a search bar and a list of 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, film_list, nicer_but_slower_film_list, sales_by_film_category, sales_by_store, and staff_list. The main area displays a schema diagram with nodes for country, city, address, customer, staff, store, payment, rental, film, film_actor, film_category, language, and actor, connected by arrows. Below the diagram are navigation icons and a table description for 'payment':

Table name	Description
payment	The payment table records each payment made by a customer, with information such as the a and the rental being paid for (when applicable). The payment table refers to the customer, re staff tables.

DBTPK Enterprise features

Data transformation (aka denormalization)



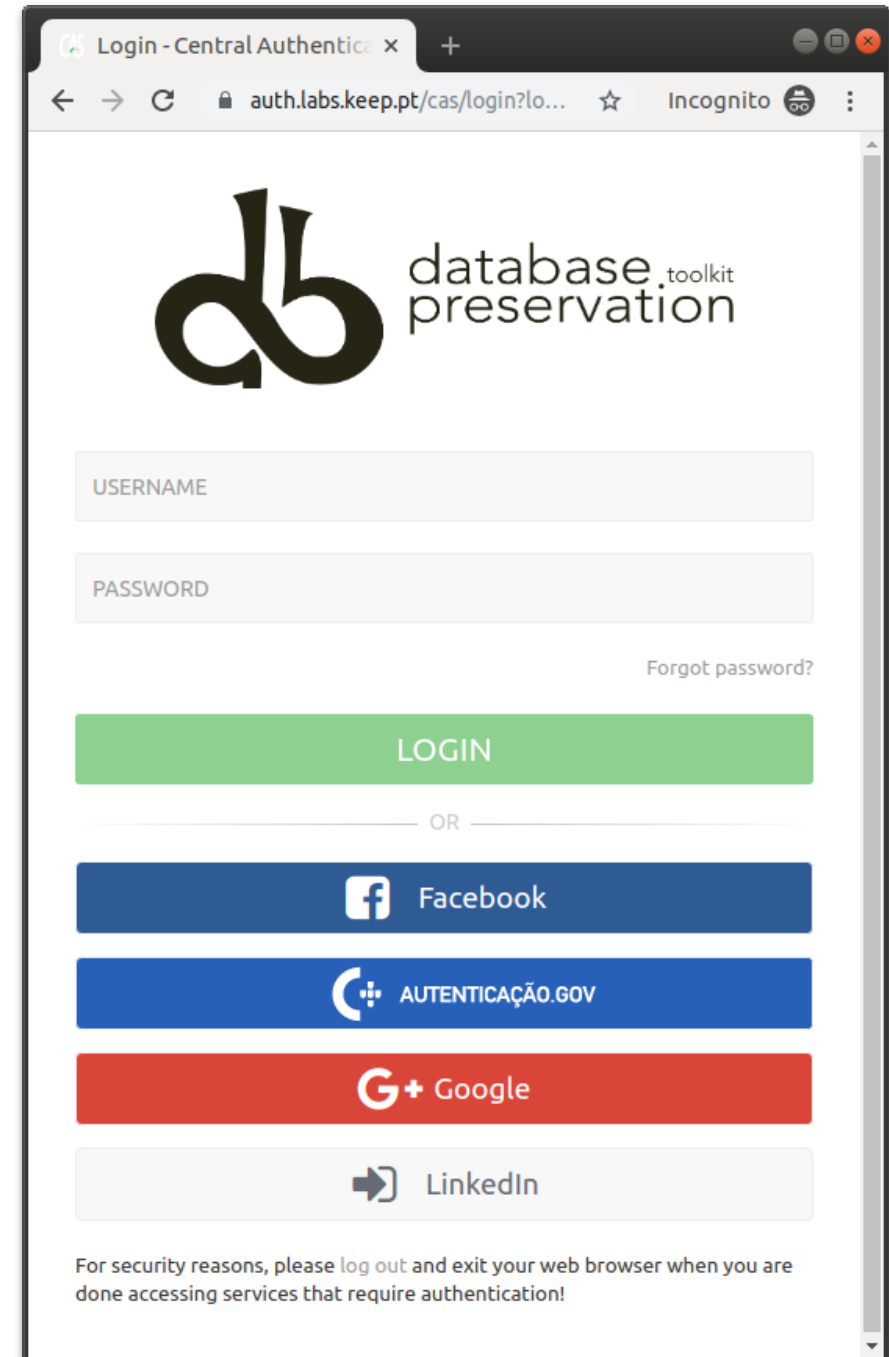
person				
Name	Birth	City name	Mayor	Country name
Mary	1986-03-28	Payne Springs	<u>Mary</u>	United States
Phillip		Rosenhayn		United States

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 shows the DBPTK Enterprise web interface for the Sakila database. 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 for 'Home', 'Databases', 'sakila', and 'film'. A 'Filter sidebar' on the left lists various database tables, with 'film' selected. The main content area displays the 'film' table description and a search interface. The search criteria are set to 'release_year' from 2006 to 2006 and 'rating' as PG-13. Below the search bar is a table of film records.

title	description	release_year	length	rating	special_features
AIRPLANE SIERRA	A Touching Saga of a H...	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,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 S...
ATTRACTION NEWTON	A Abundant P...	2006	87	PG-13	Trailers,Behind the S...

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

The screenshot shows a Microsoft Excel spreadsheet with a table of movie data. The table has the following columns: title, description, release_year, length, rating, and special_features. The data includes titles like 'GOLDFINGER SENSIBILITY', 'WOLVES DESIRE', 'CREEPERS KANE', etc. A chart titled 'The duration of the film' is overlaid on the data, showing a distribution of film lengths. The chart is a horizontal bar chart with a blue background and white bars. The x-axis represents the duration in minutes, ranging from 0 to 200. The y-axis lists the titles of the movies. The chart shows that most movies have a duration between 100 and 150 minutes.

title	description	release_year	length	rating	special_features
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
GOLDFINGER SENSIBILITY	A Insightful Drama of a Mad Scientist And a Hunter who must Defeat a Pastry Chef in New Orleans	2006	2006	93 G	Trailers,Commentaries
WOLVES DESIRE	A Fast-Paced Drama of a Squirrel And a Robot who must Succumb a Technical Writer in A Manhat	2006	2006	55 NC-17	Behind the Scenes
CREEPERS KANE	A Awe-Inspiring Reflection of a Squirrel And a Boat who must Outrace a Car in A Jet Boat	2006	2006	172 NC-17	Trailers,Behind the Sc
GUNFIGHT MOON	A Epic Reflection of a Pastry Chef And a Explorer who must Reach a Dentist in The Sahara				d Scenes,Behind
TRIP NEWTON	A Fanciful Character Study of a Lumberjack And a Car who must Discover a Cat in An Abar				entaries,Delete
PERDITION FARGO	A Fast-Paced Story of a Car And a Cat who must Outgun a Hunter in Berlin				s,Behind the Sc
INDEPENDENCE HOTEL	A Thrilling Tale of a Technical Writer And a Boy who must Face a Pioneer in A Monastery				s,Behind the Sc
ROSES TREASURE	A Astounding Panorama of a Monkey And a Secret Agent who must Defeat a Woman in T				entaries,Delete
KENTUCKIAN GIANT	A Stunning Yarn of a Woman And a Frisbee who must Escape a Waitress in A U-Boat				s,Commentaries
PUNK DIVORCE	A Fast-Paced Tale of a Pastry Chef And a Boat who must Face a Frisbee in The Canadian I				s,Commentaries
KNOCK WARLOCK	A Unbelievable Story of a Teacher And a Boat who must Confront a Moose in A Baloon				s
UPTOWN YOUNG	A Fateful Documentary of a Dog And a Hunter who must Pursue a Teacher in An Abandon				entaries
MAGUIRE APACHE	A Fast-Paced Reflection of a Waitress And a Hunter who must Defeat a Forensic Psycholc				s,Commentaries
WYOMING STORM	A Awe-Inspiring Panorama of a Robot And a Boat who must Overcome a Feminist in A U-				d Scenes
CENTER DINOSAUR	A Beautiful Character Study of a Sumo Wrestler And a Dentist who must Find a Dog in Ca				d Scenes
DIVIDE MONSTER	A Intrepid Saga of a Man And a Forensic Psychologist who must Reach a Squirrel in A Mo				s,Commentaries
SPIRIT FLINTSTONES	A Brilliant Yarn of a Cat And a Car who must Confront a Explorer in Ancient Japan				entaries,Delete
INTOLERABLE INTENTIONS	A Awe-Inspiring Story of a Monkey And a Pastry Chef who must Succumb a Womanizer in				entaries,Behind
HOOK CHARIOTS	A Insightful Story of a Boy And a Dog who must Redeem a Boy in Australia				s,Commentaries
ENCINO ELF	A Astounding Drama of a Feminist And a Teacher who must Confront a Husband in A Balc				s,Behind the Sc
CURTAIN VIDEOTAPE	A Boring Reflection of a Dentist And a Mad Cow who must Chase a Secret Agent in A Sha				s,Commentaries
LAMBS CINCINATTI	A Insightful Story of a Man And a Feminist who must Fight a Composer in Australia				s,Behind the Sc
MAGNOLIA FORRESTER	A Thoughtful Documentary of a Composer And a Explorer who must Conquer a Dentist in				s,Commentaries
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
CLEOPATRA DEVIL	A Fanciful Documentary of a Crocodile And a Technical Writer who must Fight a A Shark i				s,Deleted Scene
HOCUS FRIDA	A Awe-Inspiring Tale of a Girl And a Madman who must Outgun a Student in A Shark Tan				s,Deleted Scene
STAGE WORLD	A Lacklustre Panorama of a Woman And a Frisbee who must Chase a Crocodile in A Jet I				entaries,Behind
CHAINSAW UPTOWN	A Beautiful Documentary of a Boy And a Robot who must Discover a Squirrel in Australia				d Scenes,Behind
PILOT HOOSIERS	A Awe-Inspiring Reflection of a Crocodile And a Sumo Wrestler who must Meet a Forensi				s,Deleted Scene
REMEMBER DIARY	A Insightful Tale of a Technical Writer And a Waitress who must Conquer a Monkey in Ancient Indi	2006	2006	110 R	Trailers,Commentaries
JAPANESE RUN	A Awe-Inspiring Epistle of a Feminist And a Girl who must Sink a Girl in The Outback	2006	2006	135 G	Deleted Scenes
RAINBOW SHOCK	A Action-Packed Story of a Hunter And a Boy who must Discover a Lumberjack in Ancient India	2006	2006	74 PG	Trailers,Commentaries
MAIDEN HOME	A Lacklustre Saga of a Moose And a Teacher who must Kill a Forensic Psychologist in A MySQL Cc	2006	2006	138 PG	Behind the Scenes

DBPTK Enterprise features

Activity log Audit every access

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

The screenshot displays the DBPTK Enterprise Activity Log page. The header shows the DBPTK logo and the text 'DBPTK Enterprise'. Below the header, there is a navigation bar with 'Databases > Activity log' and user information 'lfaria Administration English'. The main content area is titled 'Activity log' and includes a descriptive paragraph about event logs. Below this is a search bar and a table of activity log entries. The table has columns for Date, Component, Method, User, Duration, Address, and Outcome. The entries show various actions like 'Find', 'Login', and 'Cas Login' performed by users 'lfaria' and 'mguimaraes'. A sidebar on the right lists 'Components' and 'Methods' with checkboxes and counts. At the bottom of the table, it says '1-20 of 2,972' and there is a 'Show More' button.

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

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 + [ ] [ ] Tmux: 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.

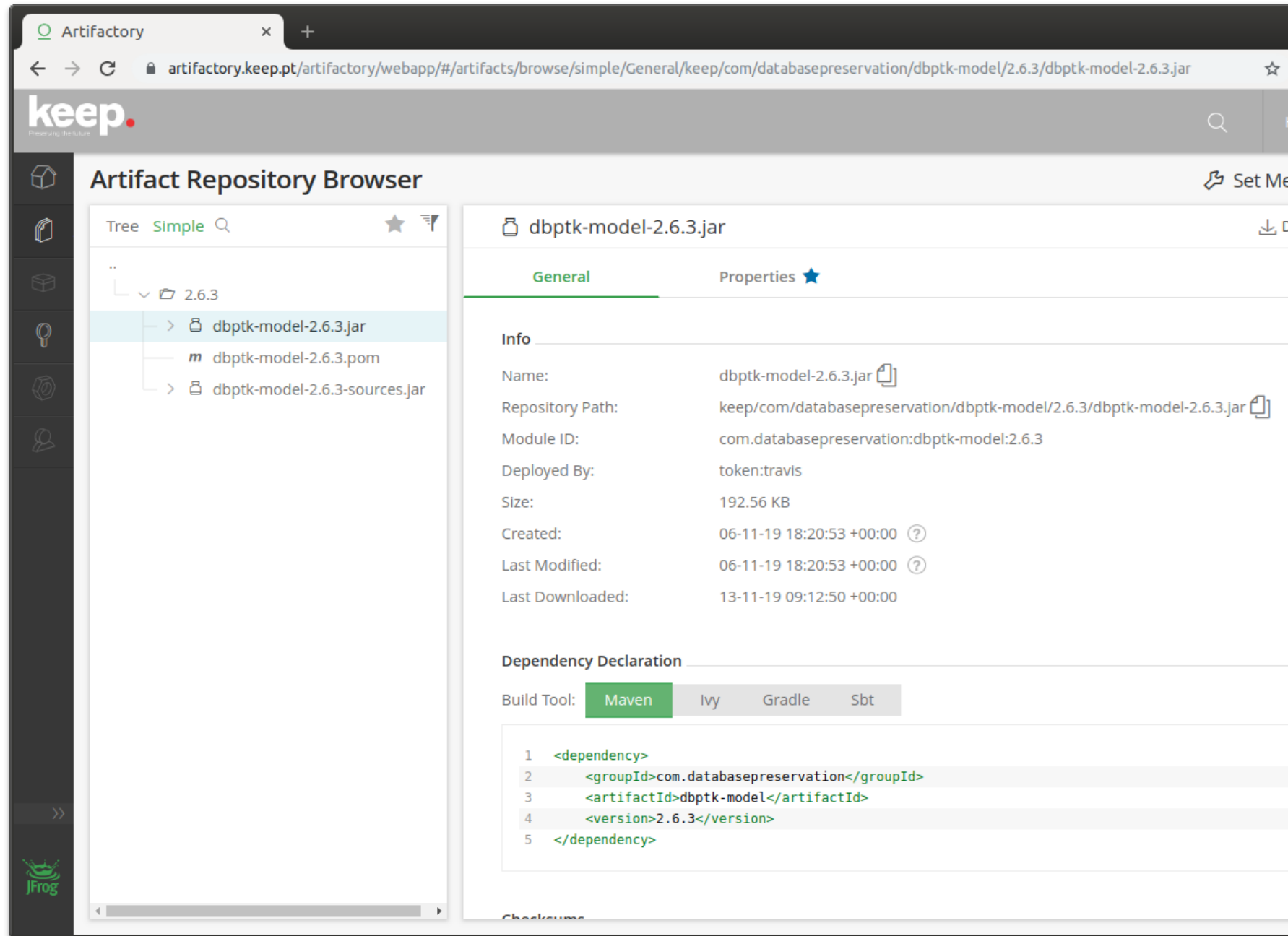
    Log files and migration reports were saved in /home/mguimaraes
    Troubleshooting information can be found at http://www.database-preservation.com/#troubleshooting
    Please report any problems at https://github.com/keeps/db-preservation-toolkit/issues/new

~ $ _
```

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". The left sidebar shows a tree view of the repository structure, with the following items listed under the "2.6.3" folder:

- `dbptk-model-2.6.3.jar` (selected)
- `dbptk-model-2.6.3.pom`
- `dbptk-model-2.6.3-sources.jar`

The main content area shows the details for the selected artifact, `dbptk-model-2.6.3.jar`. The "General" tab is active, displaying the following information:

Info	Value
Name:	<code>dbptk-model-2.6.3.jar</code>
Repository Path:	<code>keep/com/databasepreservation/dbptk-model/2.6.3/dbptk-model-2.6.3.jar</code>
Module ID:	<code>com.databasepreservation:dbptk-model:2.6.3</code>
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 "General" tab, the "Dependency Declaration" section is visible, showing the build tool configuration for Maven:

```
1 <dependency>
2   <groupId>com.databasepreservation</groupId>
3   <artifactId>dbptk-model</artifactId>
4   <version>2.6.3</version>
5 </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'. At the top, there are navigation links for 'Why GitHub?', 'Team', 'Enterprise', 'Explore', 'Marketplace', and 'Pricing', along with a search bar and 'Sign in'/'Sign up' buttons. The repository name 'keeps/dbptk-developer' is displayed, with statistics for 'Watch' (15), 'Star' (31), and 'Fork' (11). Below this, there are tabs for 'Code', 'Issues' (71), 'Pull requests' (1), 'Actions', 'Projects' (1), 'Wiki', 'Security', and 'Insights'. The main content area shows a commit by 'hmiguim' titled 'Setting version 2.10.0-SNAPSHOT' with 1,246 commits. Below the commit is a file tree listing various folders and files, including '.github', '.travis', 'code-style', 'dbptk-bindings', 'dbptk-core', 'dbptk-model', 'dbptk-modules', 'dbptk-plugin-example', 'doc', 'examples', 'scripts', 'testing', '.gitattributes', '.gitignore', and '.grenrc.vml'. On the right side, there is an 'About' section with a description: 'DBPTK Developer - library and command-line tool for execution of database preservation actions', a link to 'www.database-preservation...', and tags for 'preservation', 'database', 'relational-databases', 'siard', and 'preservation-formats'. There is also a 'Releases' section showing 'Version 2.9.2' as the latest release, and a 'Contributors' section with 11 contributors.

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

	Deskto p	Enterpri se	Develope r
Save to preservation format	✓	✓*	✓
Quality assurance (merkle tree)	✓	✓*	✓
Validation	✓	✓	✓
Enrich descriptions	✓	✓	✓
Browse and search	✓	✓	X
Transform (de-normalization)	X	✓	X
Export to live databases	✓	✓*	✓
Activity Log	X	✓	X
Authentication	X	✓	X
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/
Twitter: #EARKProject

Ready to get started?

Find out more at:
ec.europa.eu/cefdigital

Contact us:
cef-building-blocks@ec.europa.eu

Thank you!

