



Preserving the future

Preservation through standardisation with DBPTK

Luís Faria

Research & Innovation Director

KEEP SOLUTIONS

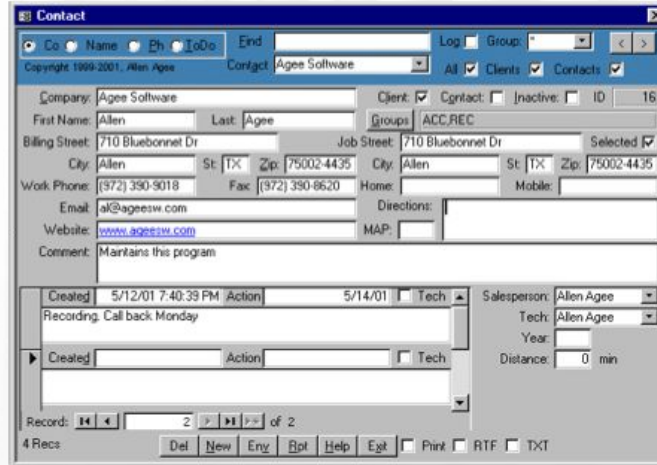
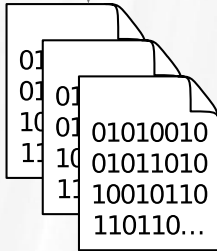
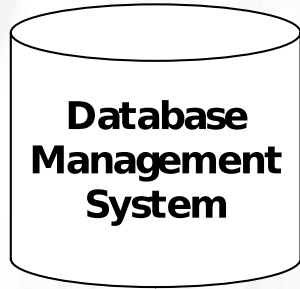
Databases

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

The diagram shows a table with the following data:

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

Annotations:

- An arrow points from the text "Cell" to the cell containing "1986-03-28".
- An arrow points from the text "Row" to the entire first data row (1, Mary, 1986-03-28, 2).
- An arrow points from the text "Column" to the "birth" column.

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

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

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

Information 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

⊕ no need to maintain hardware or software

⊖ may hinder timely access to information

⊖ difficult to gather documentation of complex or closed file formats

⊖ difficult to ensure quality and completeness without hindsight

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

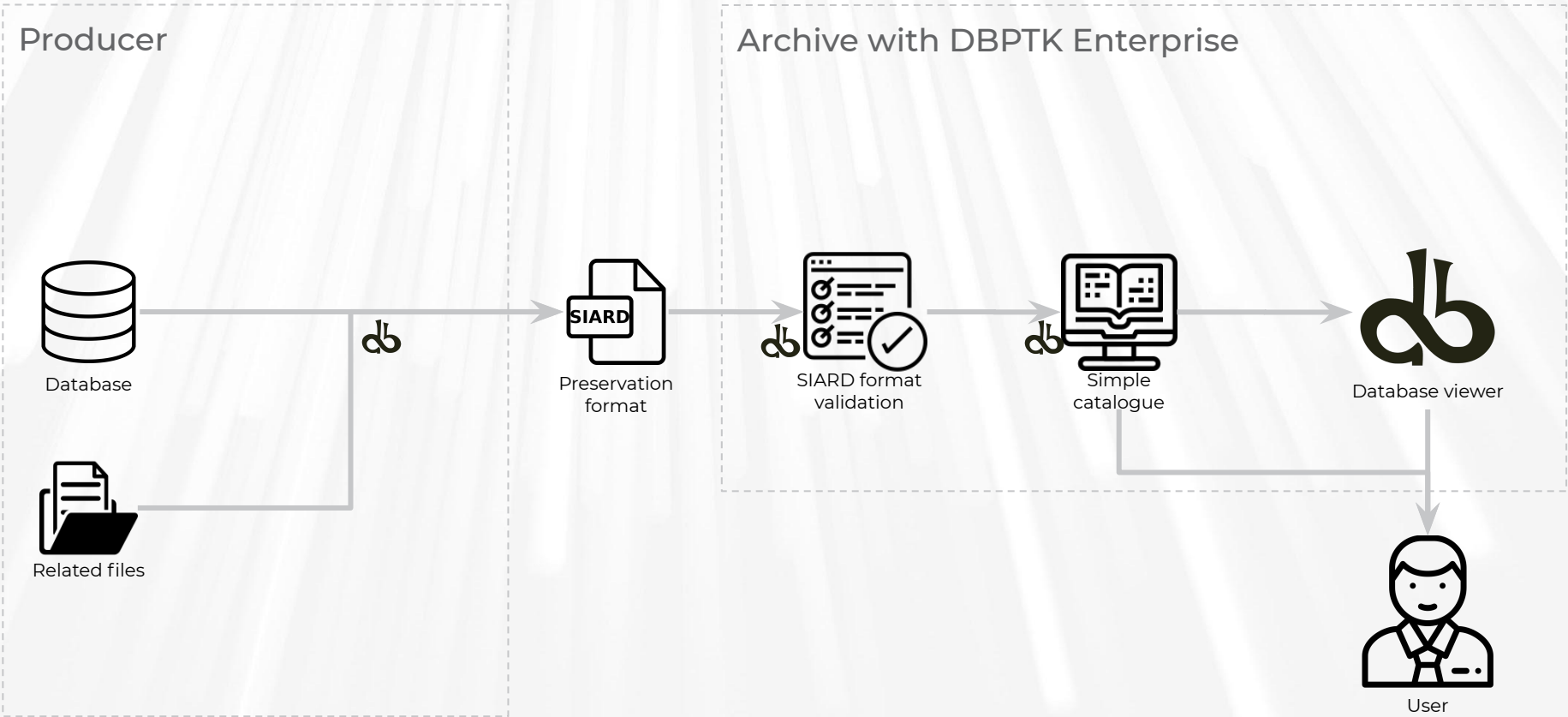
<https://dilcis.eu/content-types/siard>

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

More about this on the next session

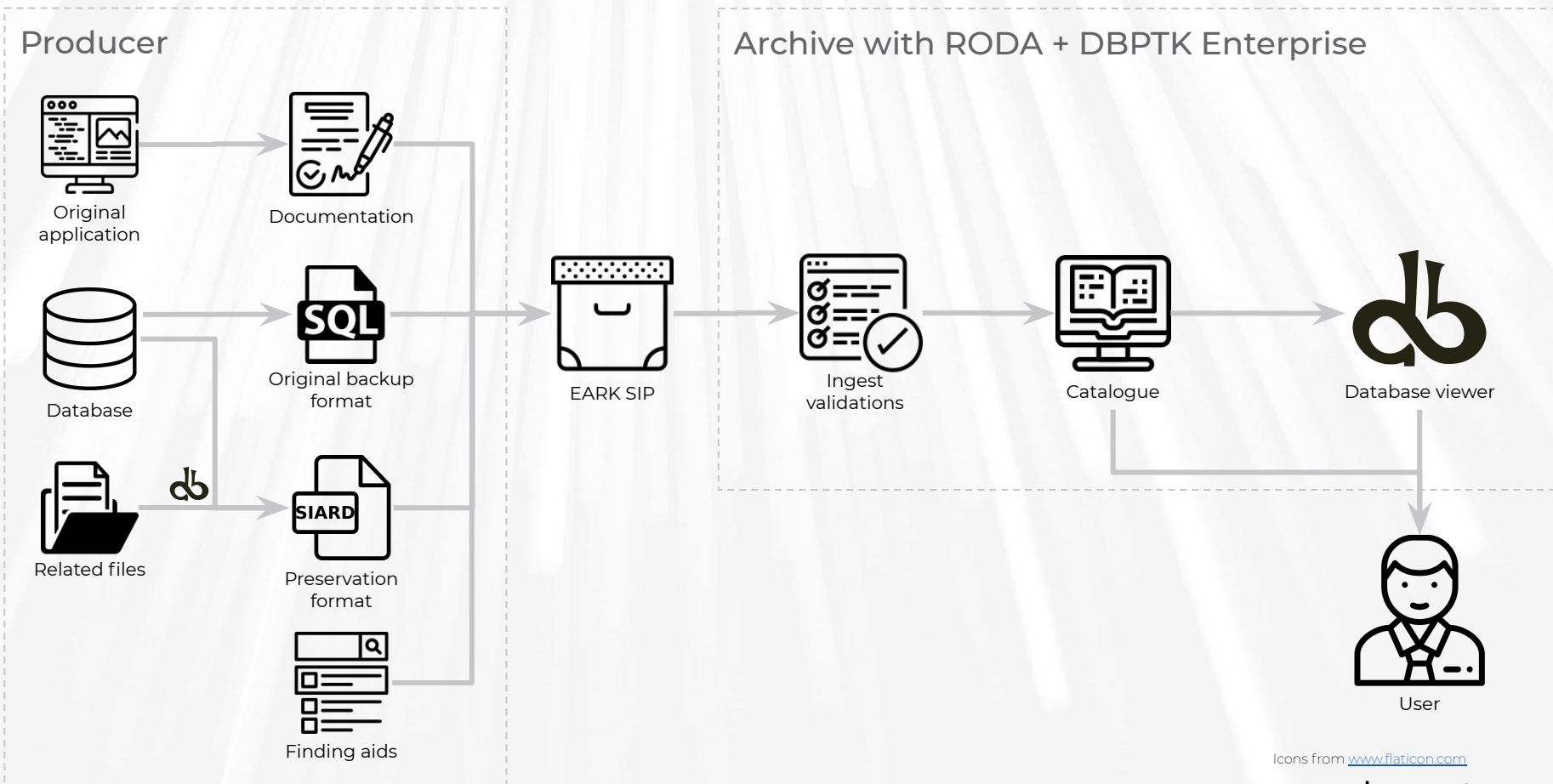
Kuldar Aas (National Archives Estonia): E-ARK standardisation efforts for databases

Simple database archive flow



Icons from www.flaticon.com

Full database archival flow



Icons from www.flaticon.com

www.keep.pt

DBPTK Database Preservation Toolkit

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



<https://database-preservation.com>



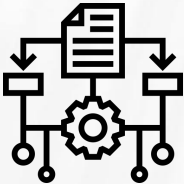
DBPTK Desktop

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



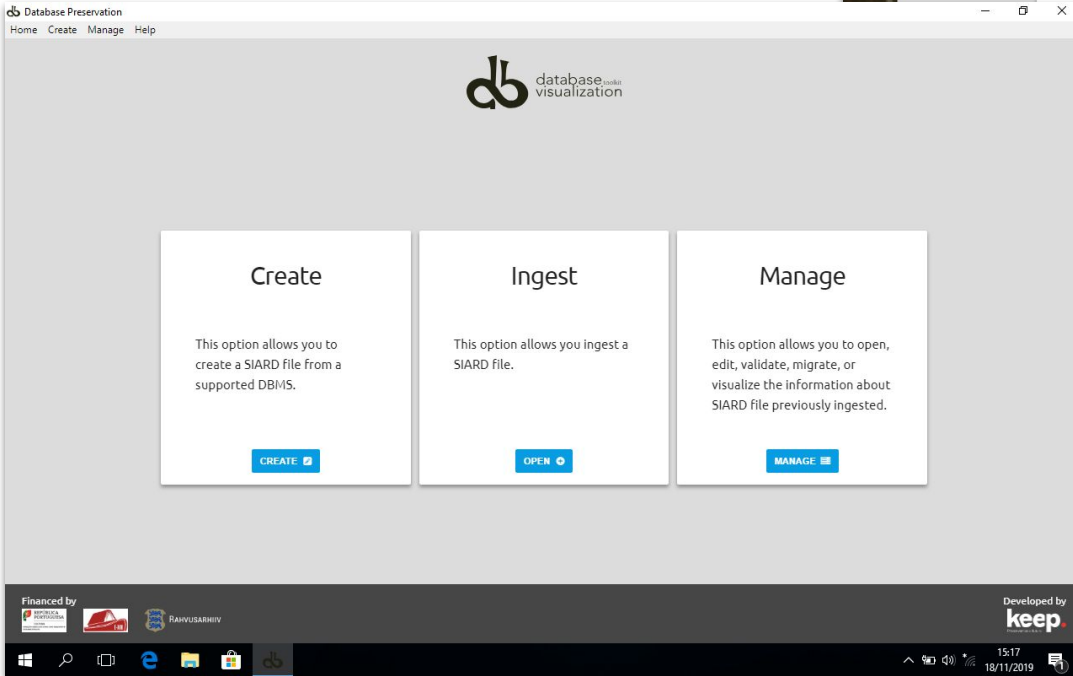
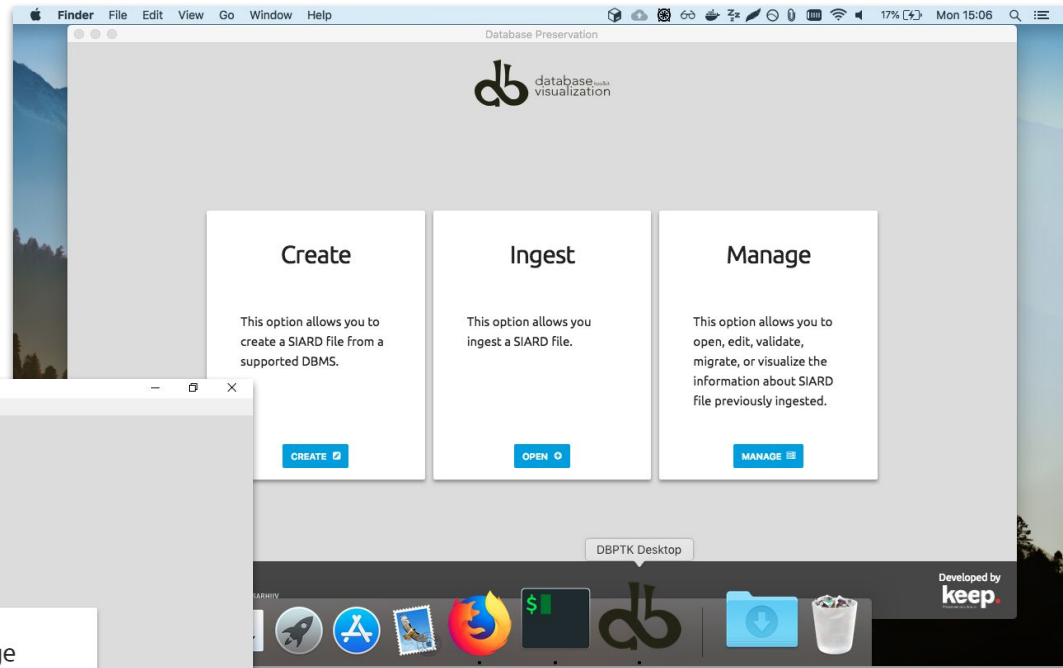
DBPTK Developer

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

DBPTK Desktop

Top features

DBPTK Desktop features



Also available on Linux

DBPTK Desktop features

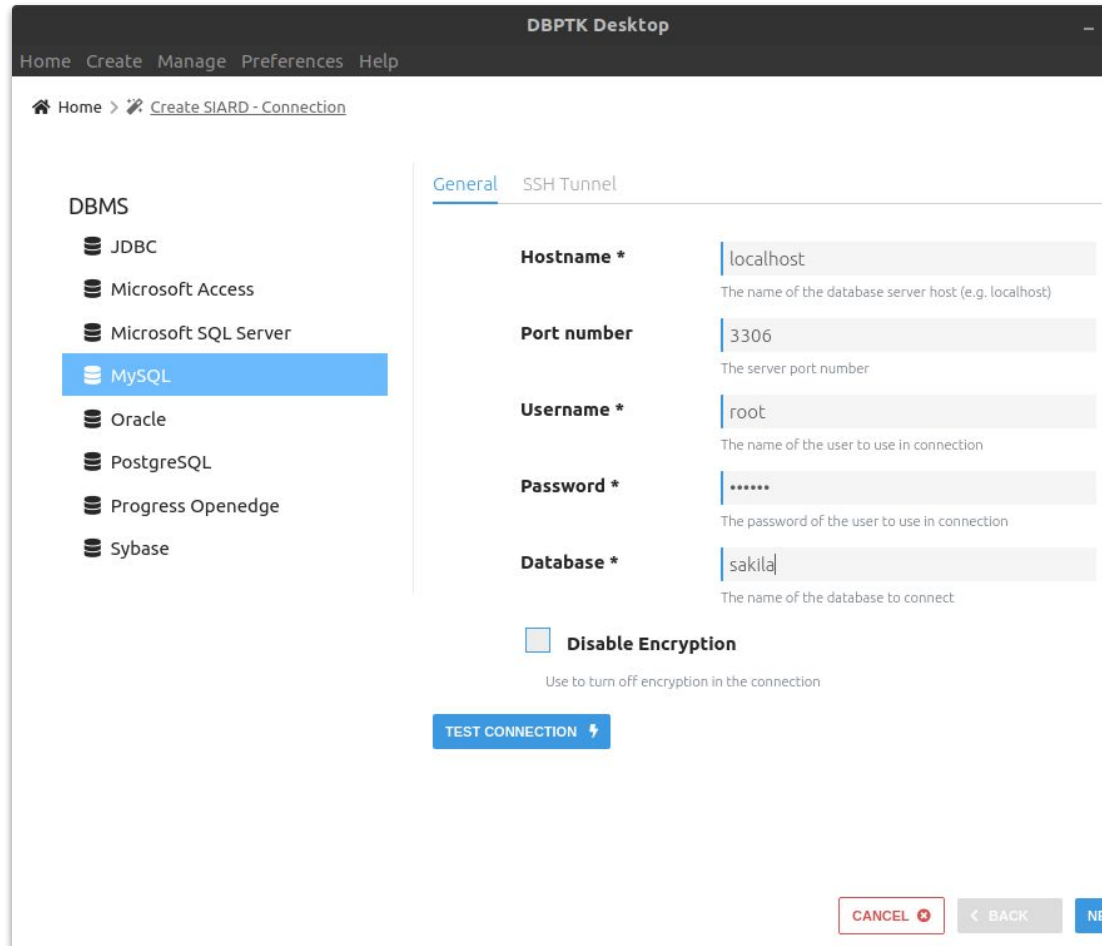
SIARD creation

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



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

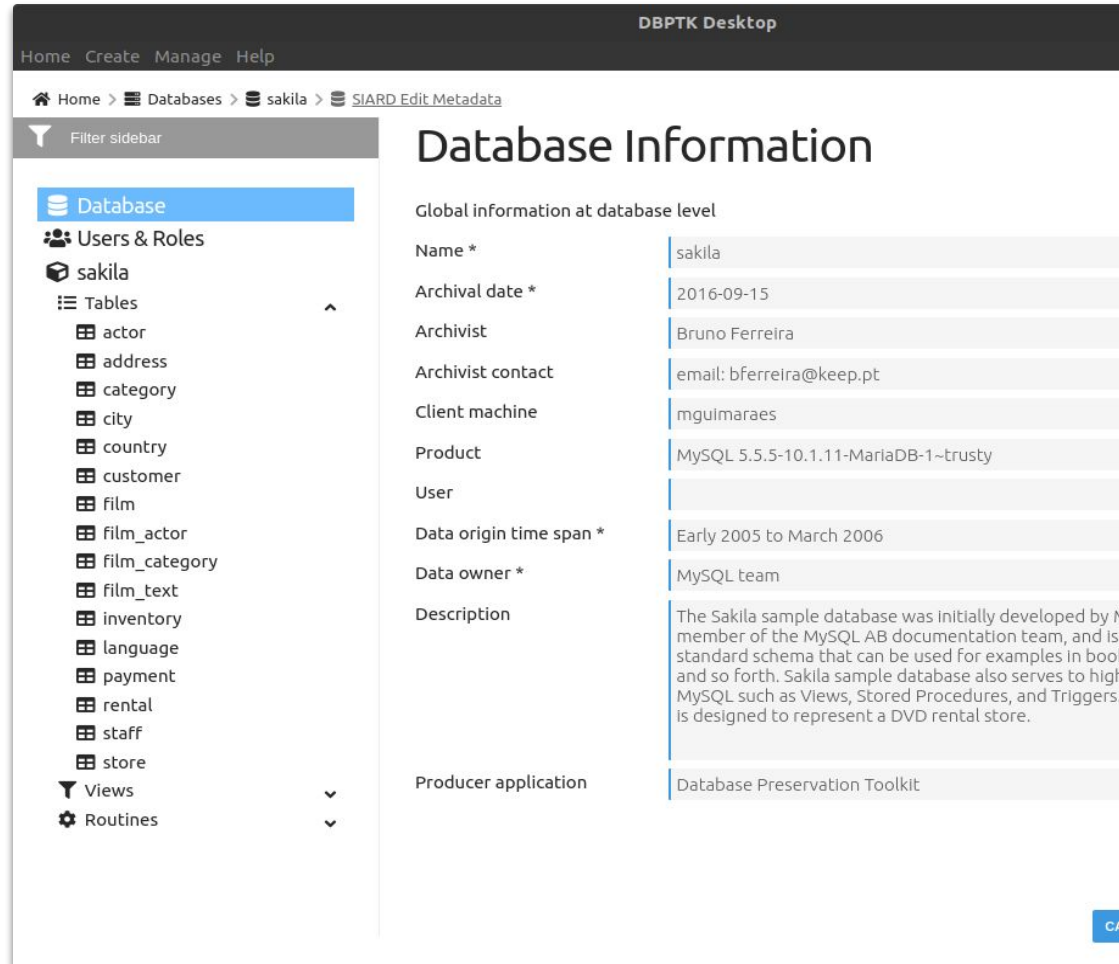
- Type conversion in import module: in sakila.address.address (format: schema.table.column) has original type VARCHAR and was converted to the standard type CHARACTER VARYING(50)
- Type conversion in import module: in sakila.address.district (format: schema.table.column) has original type VARCHAR and 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.address (format: schema.table.column) has original type VARCHAR and was converted to the standard type CHARACTER VARYING(50)

DBPTK Desktop features

Edit SIARD metadata

Enrich archived database with descriptions

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

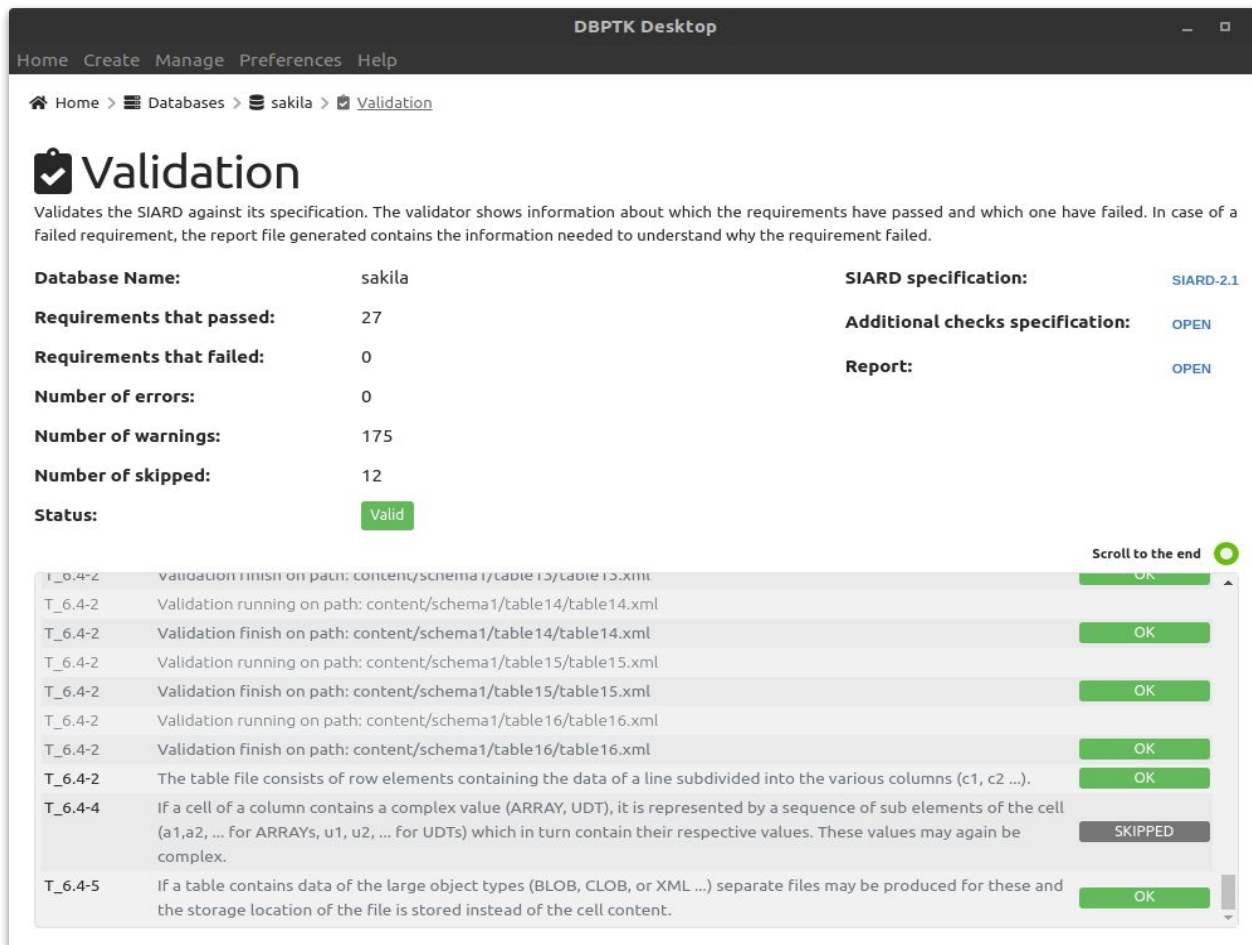


DBPTK Desktop features

SIARD validation

Validate archived database

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



The screenshot shows the DBPTK Desktop interface for the Validation feature. The breadcrumb navigation is 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.'

Summary statistics:

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		

A scrollable log of validation events is shown below, with a 'Scroll to the end' button on the right. The log entries include:

- 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 window. The title bar reads "DBPTK Desktop". The menu bar includes "Home", "Create", "Manage", and "Help". The breadcrumb navigation shows "Home > Databases > sakila > Search". The left sidebar contains a "Filter sidebar" with sections for "Information", "Search all records" (highlighted), "Saved searches", and "Tables". The "Tables" section lists various tables including actor, address, category, city, country, customer, film, film_actor, film_category, film_text, inventory, language, payment, rental, staff, store, actor_info, customer_list, and film_list. The main content area displays "Search all records" with a search input containing "dan". Below this, the "actor" table is shown with columns: actor_id, first_name, last_name, and last_update. The results are:

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

Below the actor table, the "customer" table is shown with columns: customer_id, store_id, first_name, last_name, and email. The results are:

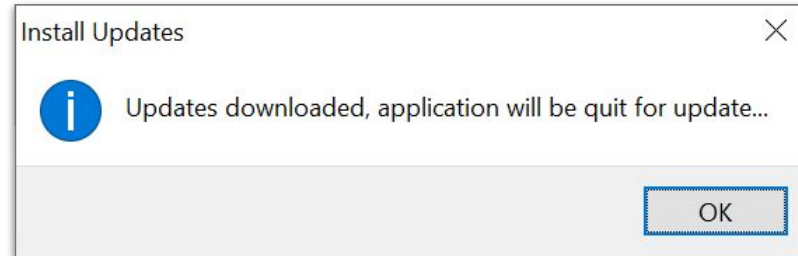
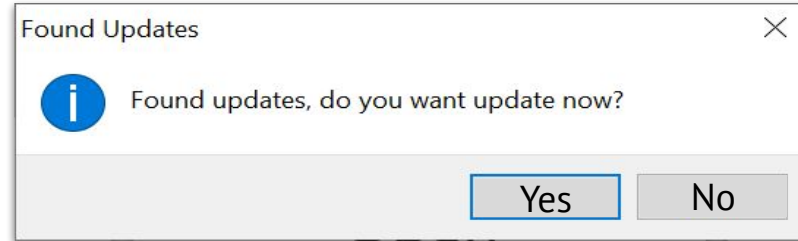
customer_id	store_id	first_name	last_name	email
477	1	DAN	PAINE	DAN.PAINE@sakilacus

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

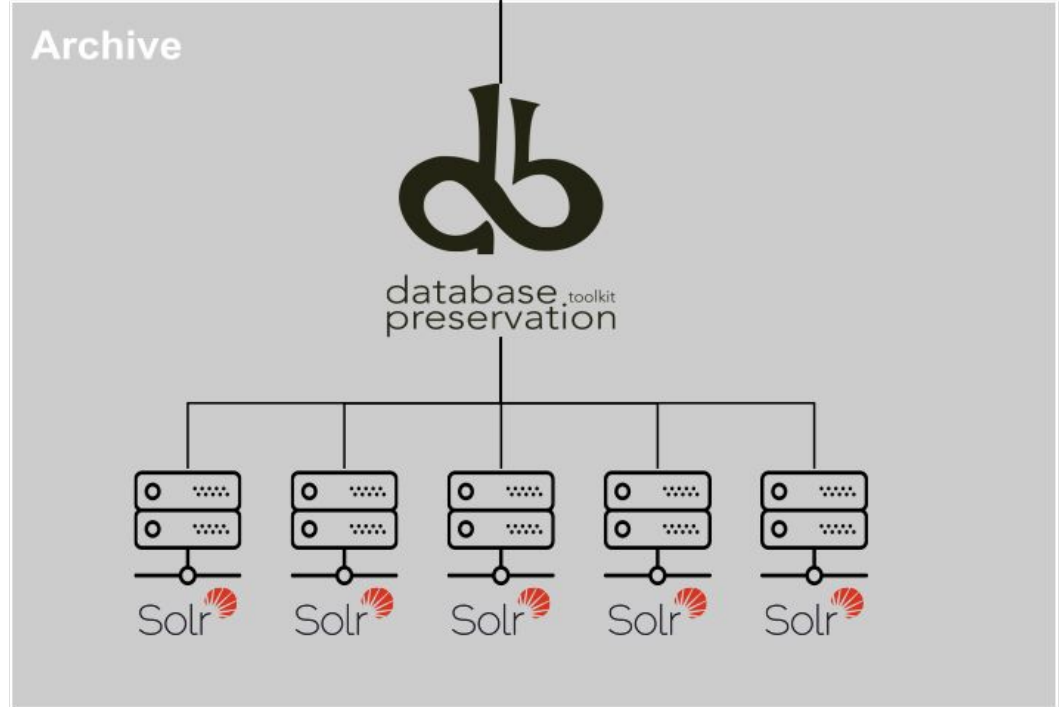
Top features

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

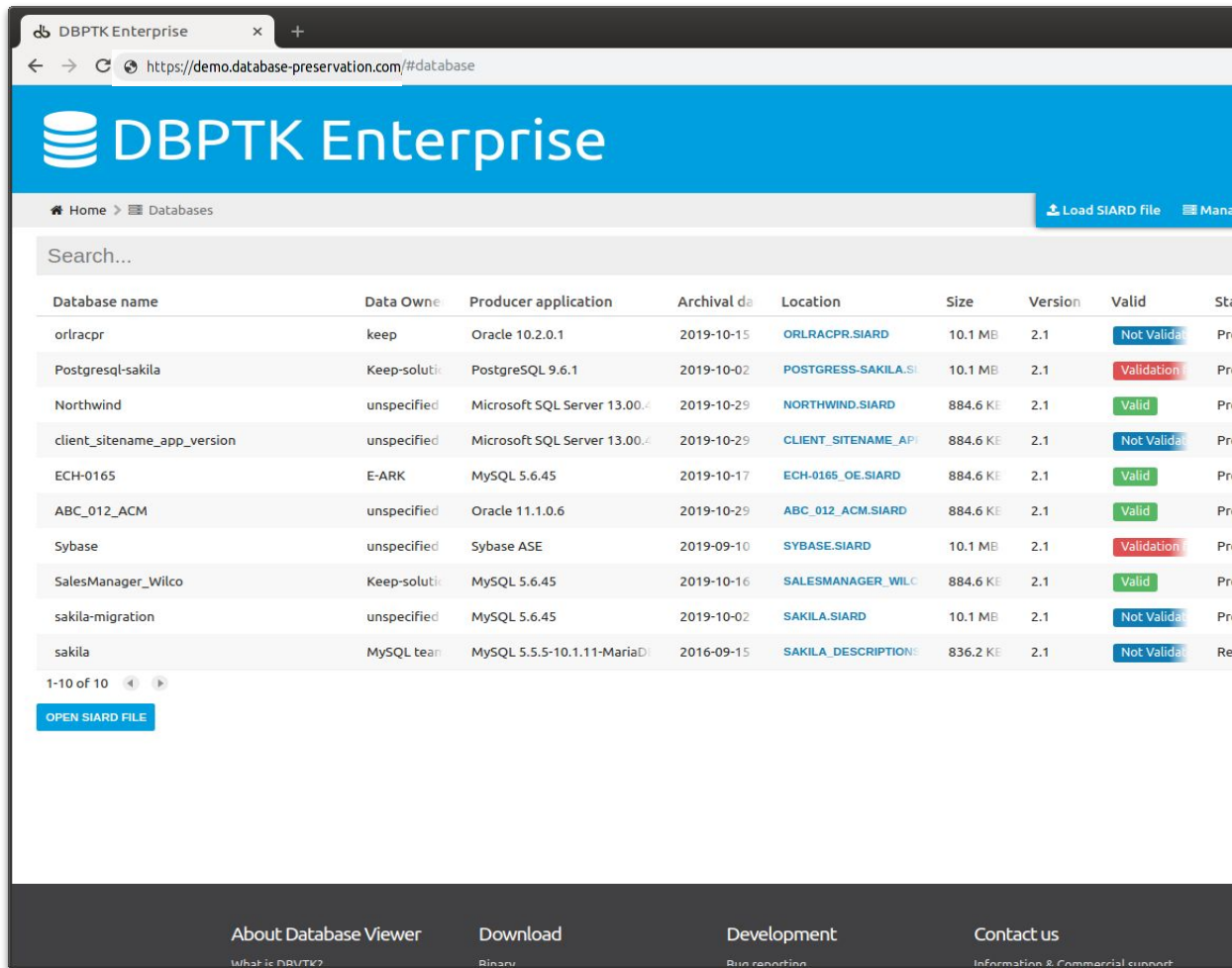


DBTPK Enterprise features

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 located below the header. The main content area contains a table listing various databases with the following columns: Database name, Data Owner, Producer application, Archival date, Location, Size, Version, Valid, and Status. The table lists 10 databases, including 'orlracpr', 'Postgresql-sakila', 'Northwind', 'client_sitename_app_version', 'ECH-0165', 'ABC_012_ACM', 'Sybase', 'SalesManager_Wilco', 'sakila-migration', and 'sakila'. Each row includes a 'Valid' status indicator (e.g., 'Not Validated', 'Validation', 'Valid', 'Not Validated') and a 'Status' column. A pagination control shows '1-10 of 10' and an 'OPEN SIARD FILE' button is visible 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 Validated	Pr
Postgresql-sakila	Keep-soluti	PostgreSQL 9.6.1	2019-10-02	POSTGRESS-SAKILA.SI	10.1 MB	2.1	Validation	Pr
Northwind	unspecified	Microsoft SQL Server 13.00.4	2019-10-29	NORTHWIND.SIARD	884.6 KB	2.1	Valid	Pr
client_sitename_app_version	unspecified	Microsoft SQL Server 13.00.4	2019-10-29	CLIENT_SITENAME_AP	884.6 KB	2.1	Not Validated	Pr
ECH-0165	E-ARK	MySQL 5.6.45	2019-10-17	ECH-0165_OE.SIARD	884.6 KB	2.1	Valid	Pr
ABC_012_ACM	unspecified	Oracle 11.1.0.6	2019-10-29	ABC_012_ACM.SIARD	884.6 KB	2.1	Valid	Pr
Sybase	unspecified	Sybase ASE	2019-09-10	SYBASE.SIARD	10.1 MB	2.1	Validation	Pr
SalesManager_Wilco	Keep-soluti	MySQL 5.6.45	2019-10-16	SALESMANAGER_WILC	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 Validated	Pr
sakila	MySQL team	MySQL 5.5.5-10.1.11-MariaDB	2016-09-15	SAKILA_DESCRIPTION	836.2 KB	2.1	Not Validated	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. The browser address bar displays the URL: <https://demo.database-preservation.com/#database/1547aa41-1800-46b7-a28b-e82fe22f5883>. The interface is divided into a left sidebar and a main content area.

Filter sidebar:

- Information
- Search all records
- Saved searches
- Tables
 - actor
 - address
 - category
 - city
 - country
 - customer
 - film
 - film_actor
 - film_category
 - film_text
 - inventory
 - language
 - payment
 - rental
 - staff
 - store
 - actor_info
 - customer_list
 - film_list
 - nicer_but_slower_film_list
 - sales_by_film_category
 - sales_by_store
 - staff_list
- Technical Information

Schema Description:

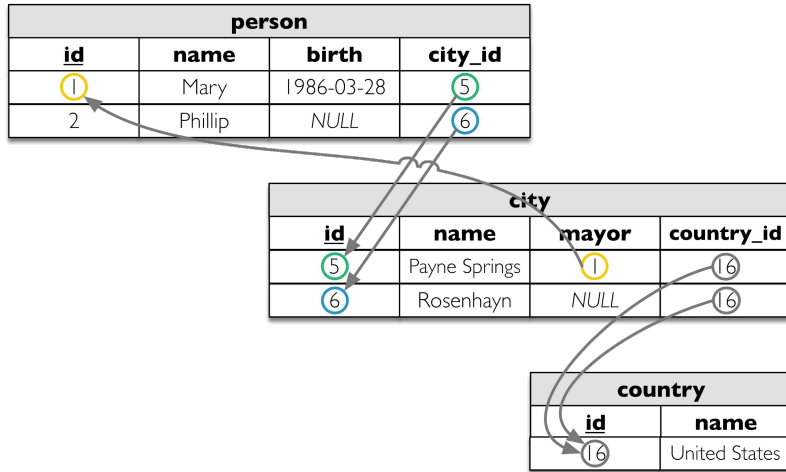
The schema diagram shows a network of tables represented by circles. The size of the circles indicates the number of records. The tables are: payment (largest, blue), rental (second largest, blue), customer (medium, blue), staff (small, blue), store (small, blue), inventory (small, blue), film (medium, dark blue), film_actor (medium, grey), film_category (small, grey), language (small, grey), address (medium, grey), city (small, grey), and country (small, grey). Arrows indicate relationships between tables.

Table description table:

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

DBPTK Enterprise features

Data transformation (aka denormalization)



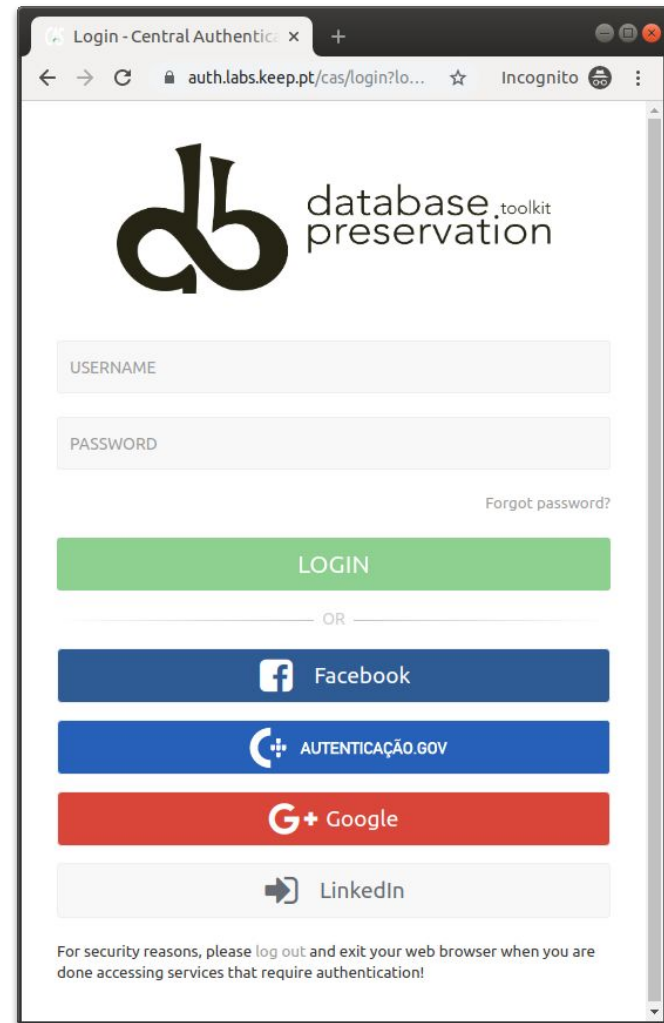
person				
Name	Birth	City name	Mayor	Country name
Mary	1986-03-28	Payne Springs	<u>Mary</u>	United States
Phillip		Rosenhayn		United States

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 Sakila database web interface. The browser address bar displays the URL: <https://demo.database-preservation.com/#table/1547aa41-1800-46b7-a28b-e82fe22f5883/cf931074-1079-4d1f-8212-6b519c78fa81/update>. The page title is "sakila" and the breadcrumb navigation is "Home > Databases > sakila > film".

The left sidebar contains a "Filter sidebar" with a list of tables. The "film" table is selected and highlighted in blue. Other tables listed include actor, address, category, city, country, customer, film_actor, film_category, film_text, inventory, language, payment, rental, staff, store, actor_info, customer_list, film_list, and nicer_but_slower_film_list.

The main content area displays the "film" table. It includes a description: "The film table is a list of all films potentially in stock in the stores. The actual in-stock copies of each film are represented in the inventory table. The film table refers to the language table and is referred to by the film_category, film_actor, and inventory tables." Below the description is a search interface with a search bar and filters. The search criteria are: release_year from 2006 to 2006, and rating PG-13. There are buttons for "ADD SEARCH FIELD", "CLEAR", "SAVE SEARCH", and "SEARCH".

The search results are displayed in a table with the following columns: title, description, release_year, length, rating, and special_features. The table contains 10 rows of film records.

title	description	release_year	length	rating	special_features
AIRPLANE SIERRA	A Touching Saga of a P...	2006	62	PG-13	Trailers,Deleted Sce...
ALABAMA DEVIL	A Thoughtful Panoram	2006	114	PG-13	Trailers,Deleted Sce...
ALTER VICTORY	A Thoughtful Drama o	2006	57	PG-13	Trailers,Behind the S...
ANTHEM LUKE	A Touching Panoram	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 Touching Drama o	2006	92	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 displays the Microsoft Excel interface with a table of movie data and a bar chart. The table has columns for title, description, release_year, length, rating, and special_features. The bar chart shows the duration of each film in minutes, with a title 'The duration of the film' and an x-axis ranging from 0 to 200 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	93	G	Trailers,Commentaries
WOLVES DESIRE	A Fast-Paced Drama of a Squirrel And a Robot who must Succumb a Technical Writer in A Manhattan	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	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 Abare				ntaries,Deleted
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 CINCINNATI	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 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 Anci	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	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	74	PG	Trailers,Commentaries
MAIDEN HOME	A Lacklustre Saga of a Moose And a Teacher who must Kill a Forensic Psychologist in A MySQL C	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 interface. The page title is "DBPTK Enterprise" and the URL is "dpc.database-preservation.com/?locale=en#activity-log". The interface shows a table of activity logs with the following columns: Date, Component, Method, User, Duration, Address, and Outcome. The Outcome column shows "Success" for all entries. A sidebar on the right lists various components and methods with checkboxes for filtering.

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	mguiaraes	15ms	81.84.255.161	Success
2020-07-24 11:45:50	Database	Find	mguiaraes	9ms	81.84.255.161	Success
2020-07-24 11:45:40	Database	Find	mguiaraes	9ms	81.84.255.161	Success
2020-07-24 11:45:30	Database	Find	mguiaraes	17ms	81.84.255.161	Success
2020-07-24 11:45:20	Database	Find	mguiaraes	10ms	81.84.255.161	Success
2020-07-24 11:45:10	Database	Find	mguiaraes	10ms	81.84.255.161	Success
2020-07-24 11:45:00	Database	Find	mguiaraes	10ms	81.84.255.161	Success
2020-07-24 11:44:50	Database	Find	mguiaraes	9ms	81.84.255.161	Success
2020-07-24 11:44:40	Database	Find	mguiaraes	12ms	81.84.255.161	Success
2020-07-24 11:44:30	Database	Find	mguiaraes	11ms	81.84.255.161	Success
2020-07-24 11:44:20	Database	Find	mguiaraes	13ms	81.84.255.161	Success
2020-07-24 11:44:10	Database	Find	mguiaraes	10ms	81.84.255.161	Success
2020-07-24 11:44:00	Database	Find	mguiaraes	20ms	81.84.255.161	Success
2020-07-24 11:43:50	Database	Find	mguiaraes	10ms	81.84.255.161	Success
2020-07-24 11:43:40	Database	Find	mguiaraes	12ms	81.84.255.161	Success
2020-07-24 11:43:30	Database	Find	mguiaraes	13ms	81.84.255.161	Success
2020-07-24 11:43:20	Database	Find	mguiaraes	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

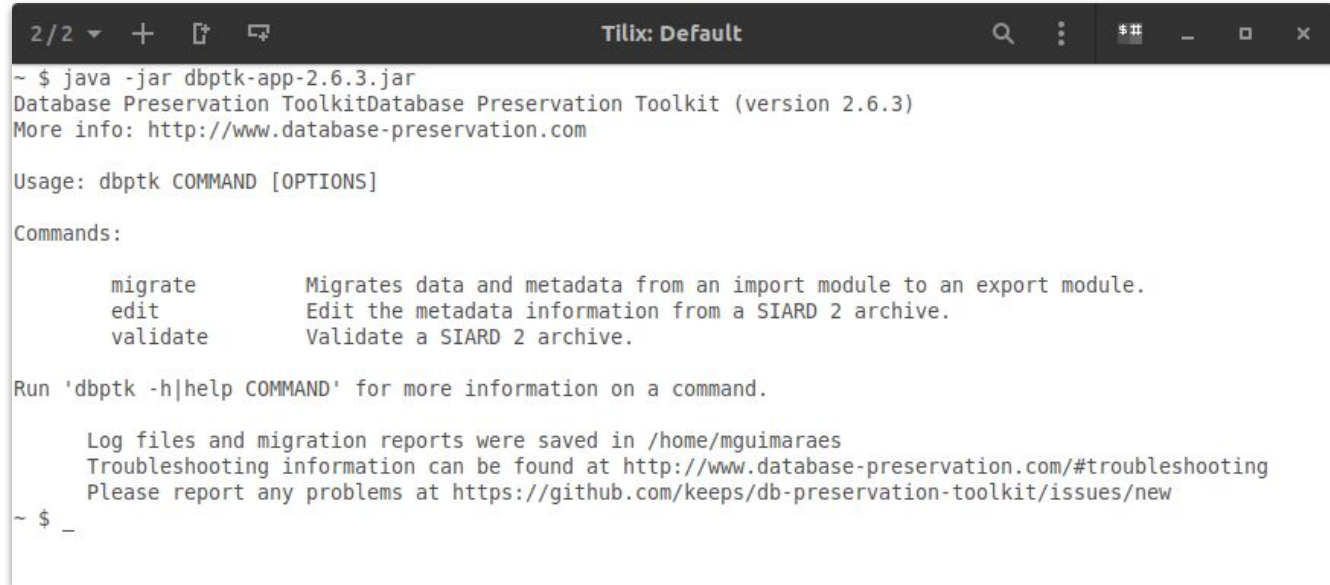
Top features

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

Usage: dbptk COMMAND [OPTIONS]

Commands:

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

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

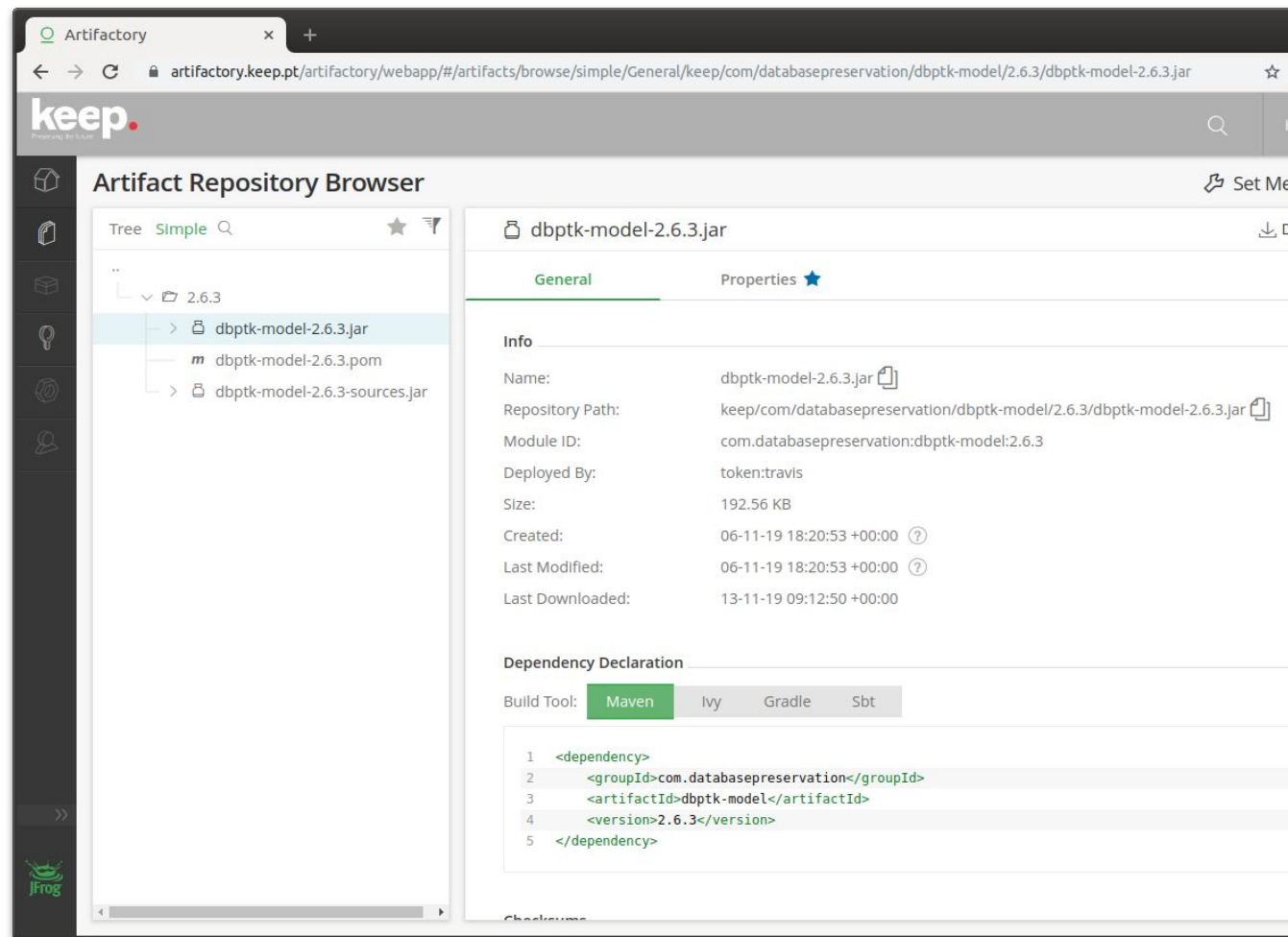
    Log files and migration reports were saved in /home/mguimaraes
    Troubleshooting information can be found at http://www.database-preservation.com/#troubleshooting
    Please report any problems at https://github.com/keeps/db-preservation-toolkit/issues/new

~ $ _
```

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:

- 2.6.3
 - dbptk-model-2.6.3.jar
 - dbptk-model-2.6.3.pom
 - dbptk-model-2.6.3-sources.jar

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

Info	Value
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

The "Dependency Declaration" section shows the Maven build tool selected, with the following XML snippet:

```
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.

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
qrenc.vml	Fixes #358	2 years ago

About DBPTK Developer - library and command-line tool for execution of database preservation actions

www.database-preservation.com

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
Integration with repositories	✗	✓	N/A
Embeddable in Web portals	✗	✓	N/A

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

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



Preserving the future

Questions?

Luis Faria
Research & Innovation Director

KEEP SOLUTIONS
lfaria@keep.pt

<https://database-preservation.com>