



Preserving the future

# Database preservation

Database preservation: concepts, strategies and tools

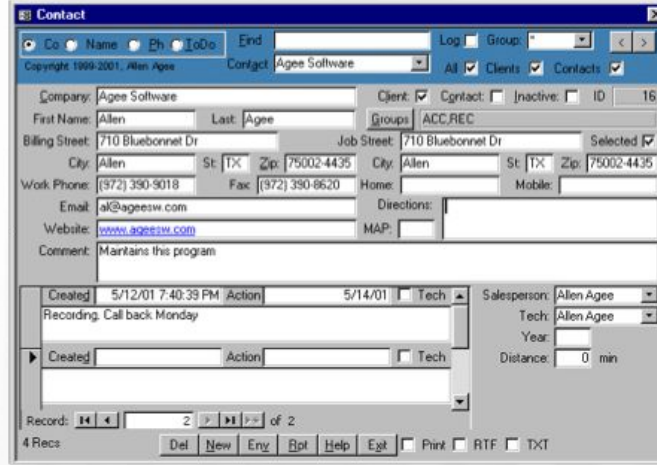
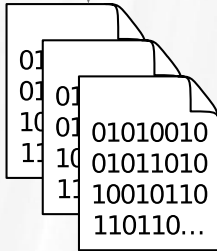
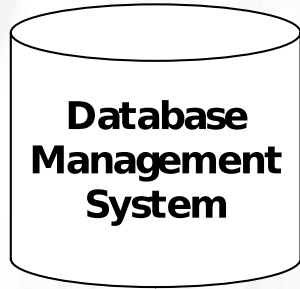
## Databases

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

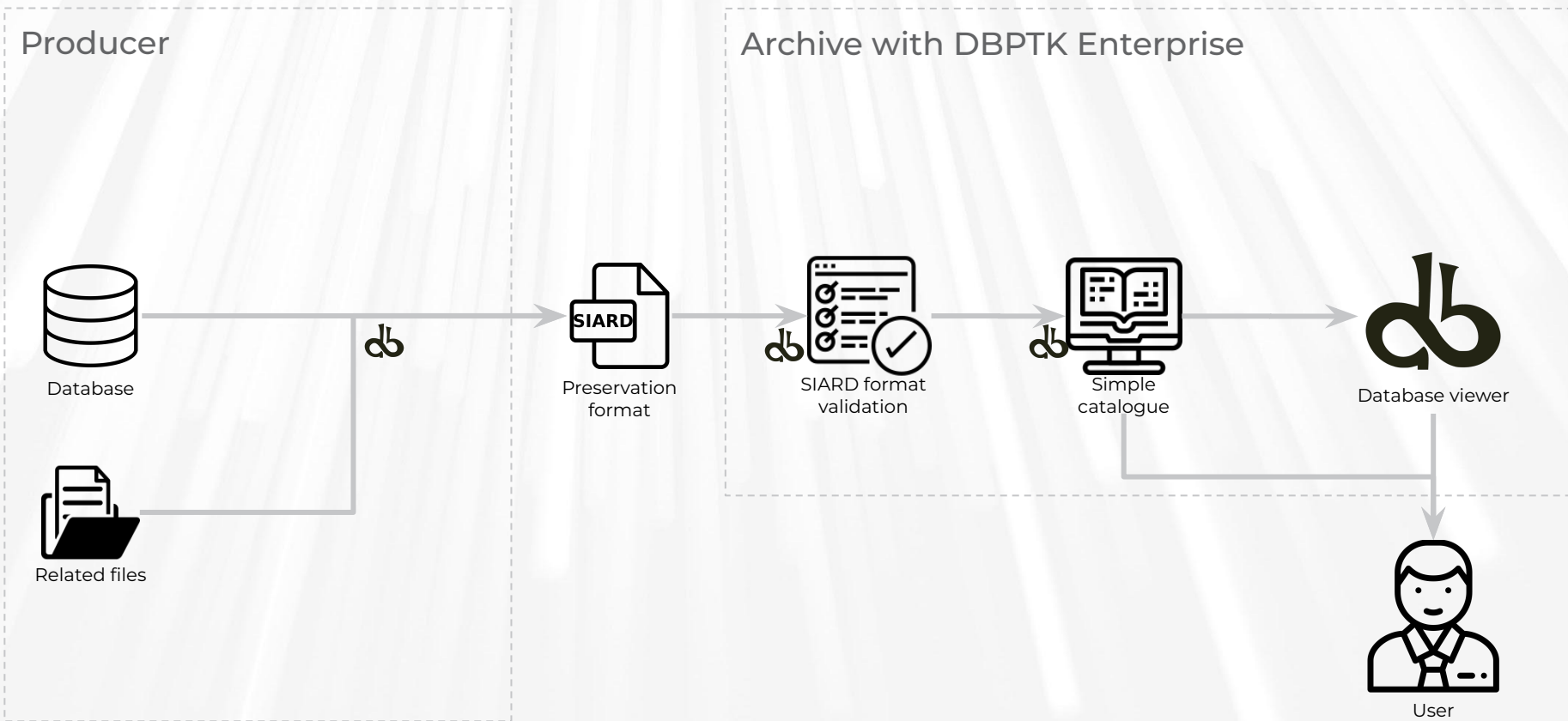
Swiss national standard eCH-0165

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>

# Simple database archive flow



# Full database archival flow

## Producer



Original application



Documentation



Database



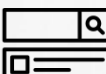
Original backup format



Related files



Preservation format



Finding aids

## Archive with RODA + DBPTK Enterprise



EARK SIP



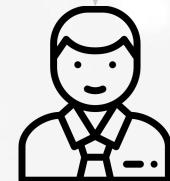
Ingest validations



Catalogue



Database viewer



User



# 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



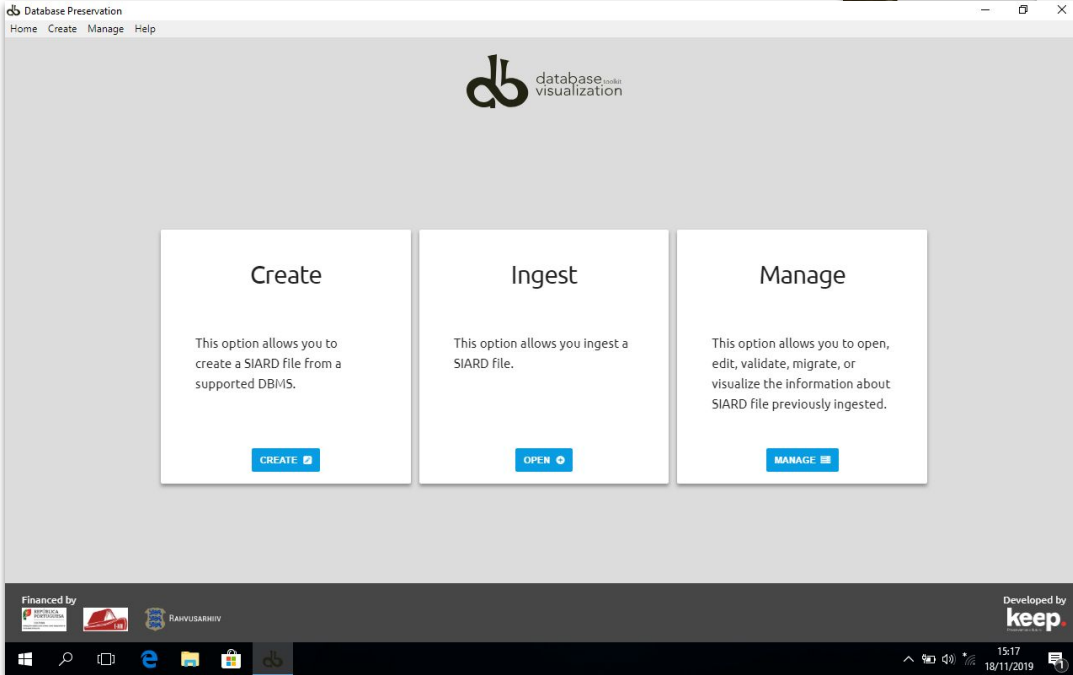
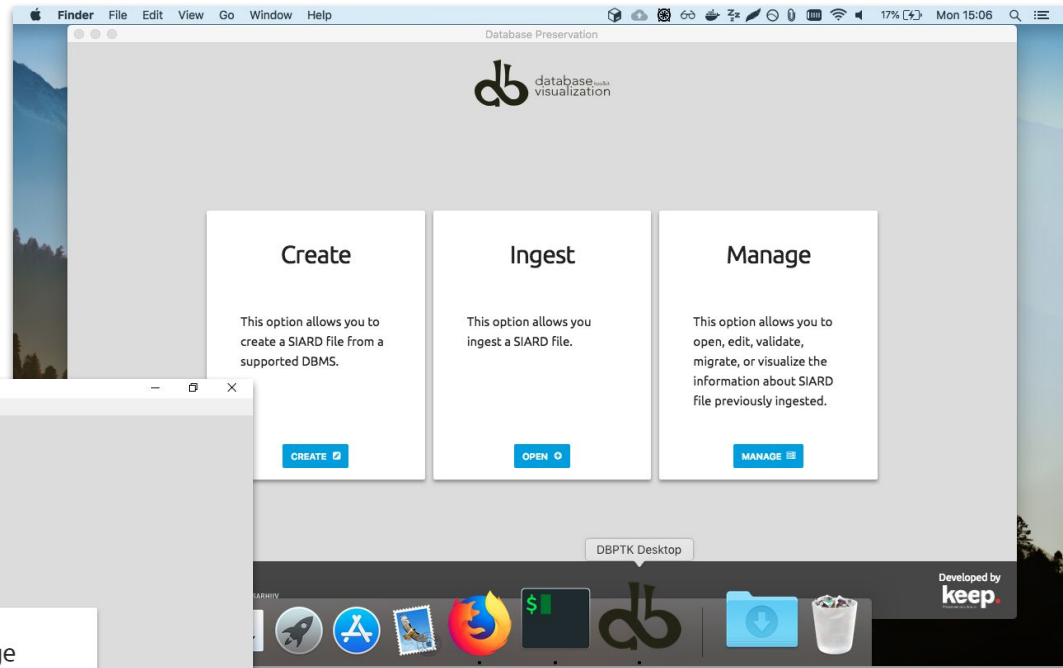
## **DBPTK Developer**

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

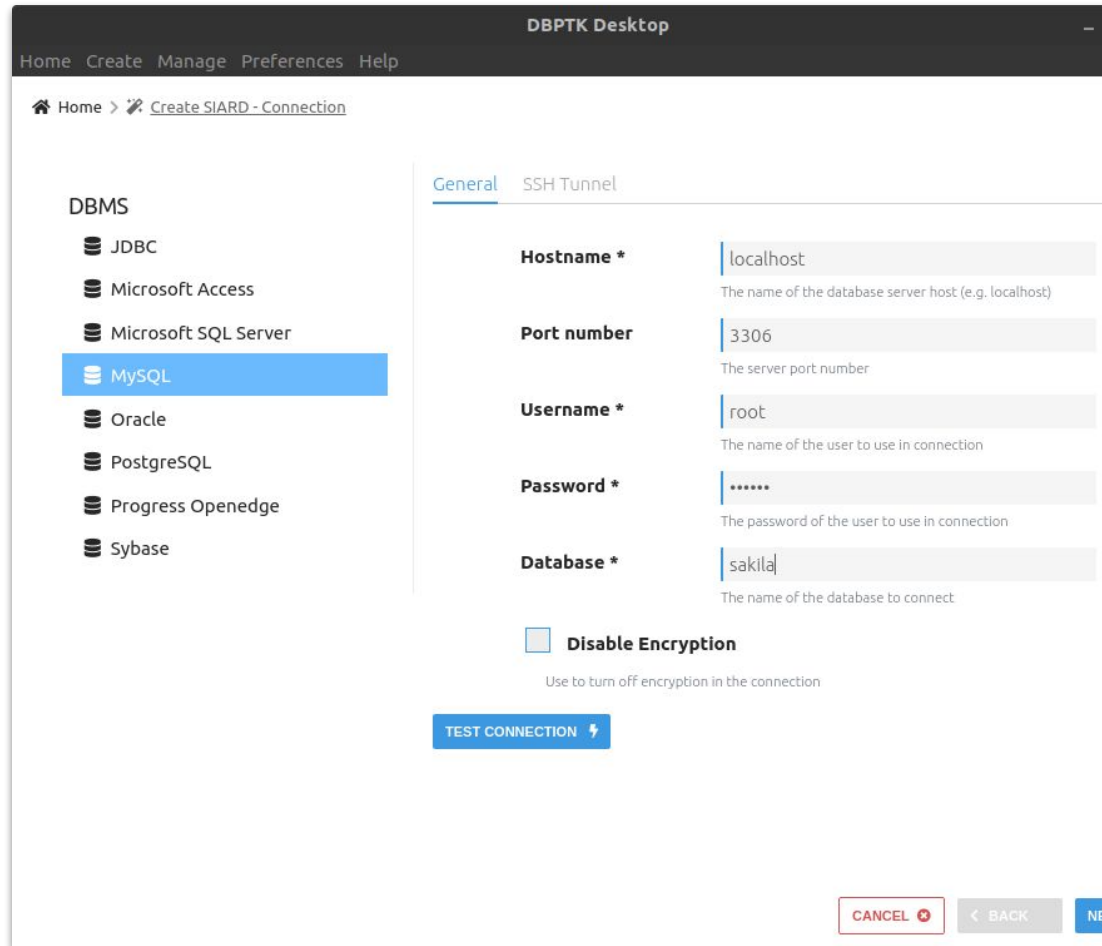
Export database to a preservation format

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

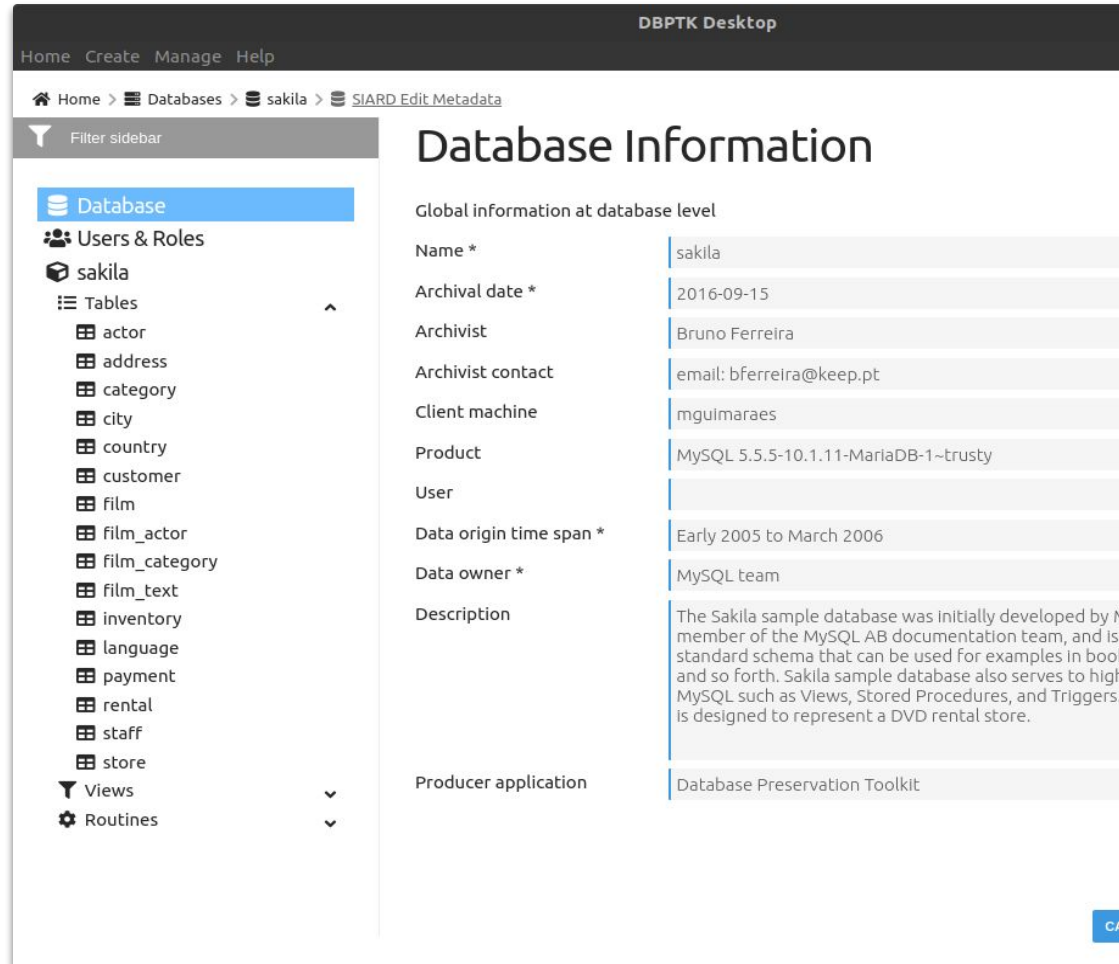
- Type conversion in import module: in sakila.address.address (format: schema.table.column) has original type VARCHAR and was converted to the standard type CHARACTER VARYING(50)
- Type conversion in import module: in sakila.address.district (format: schema.table.column) has original type VARCHAR and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



# 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 with the 'Validation' window open. The window title is 'DBPTK Desktop' and the menu bar includes 'Home', 'Create', 'Manage', 'Preferences', and 'Help'. The breadcrumb path 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.'

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

Below the summary table is a scrollable list of validation messages. A 'Scroll to the end' button is visible on the right. The messages 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 is titled "Filter sidebar" and contains a search icon and the text "Search all records". Below this, there are sections for "Information", "Saved searches", and "Tables". The "Tables" section is expanded to show 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 main content area is titled "Search all records" and has a search input field containing "dan". Below the search field, there are two table views. The first table is titled "actor" and has columns: actor\_id, first\_name, last\_name, and last\_update. It displays three rows of data for actors with the first name "DAN". The second table is titled "customer" and has columns: customer\_id, store\_id, first\_name, last\_name, and email. It displays one row of data for a customer with the first name "DAN".

Search all records

dan

actor

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

1-3 of 3

customer

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

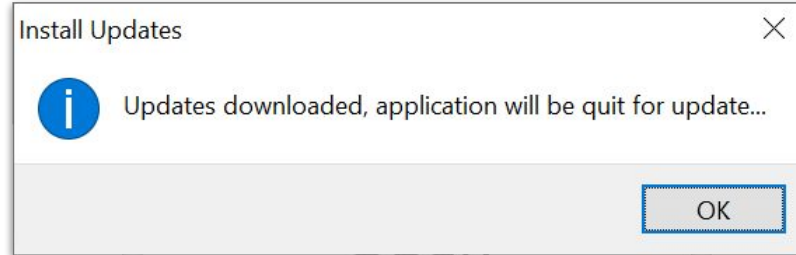
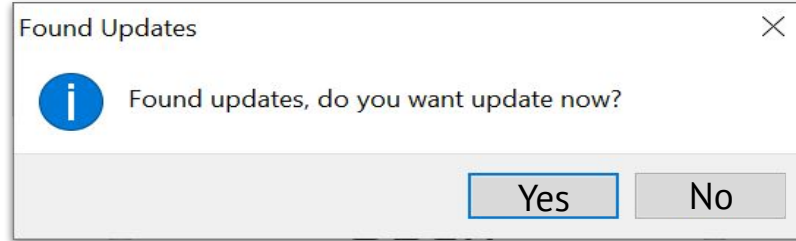
1-1 of 1

## 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

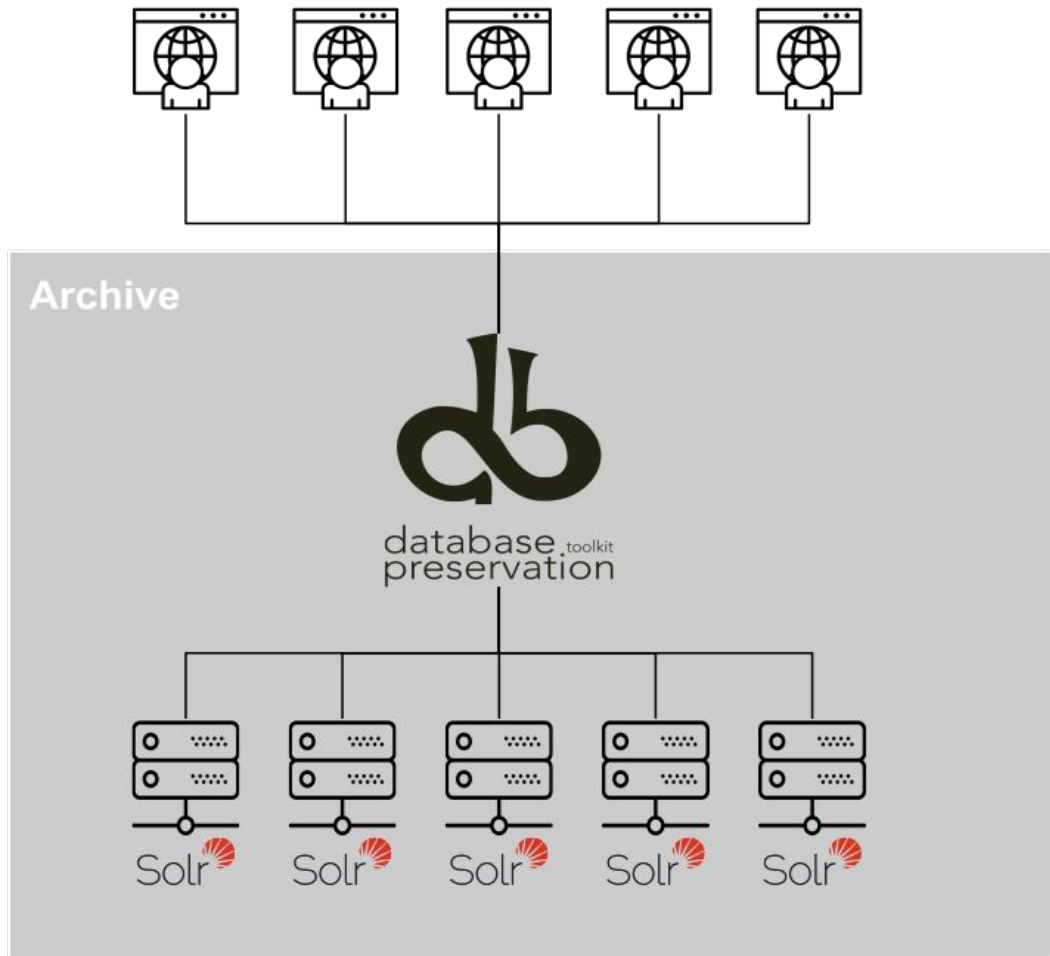
Top features

## 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

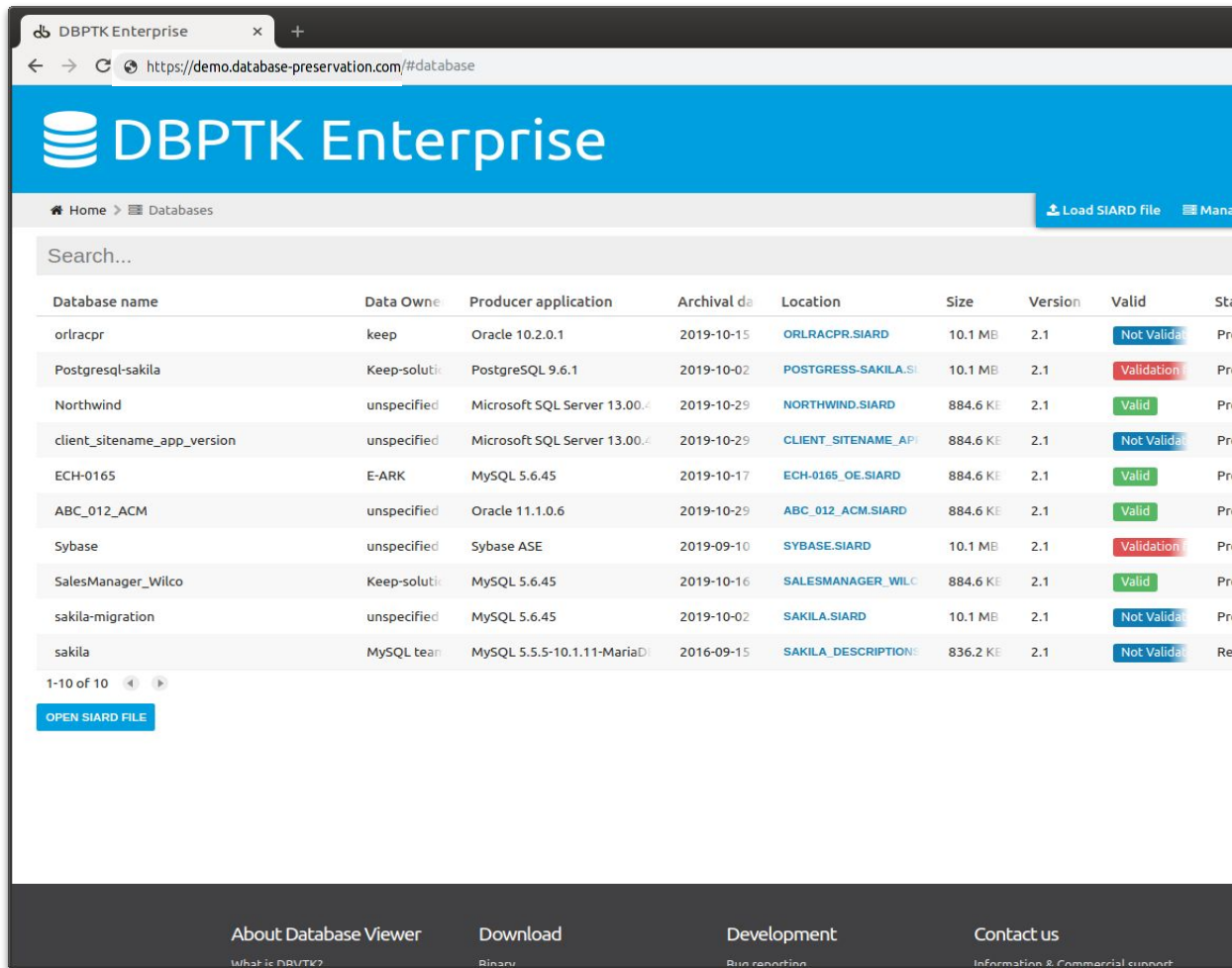


# DBTPK Enterprise features

## 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 Valid', 'Validation', 'Valid') 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	<a href="#">ORLRACPR.SIARD</a>	10.1 MB	2.1	Not Valid	Pr
Postgresql-sakila	Keep-soluti	PostgreSQL 9.6.1	2019-10-02	<a href="#">POSTGRESS-SAKILA.SI</a>	10.1 MB	2.1	Validation	Pr
Northwind	unspecified	Microsoft SQL Server 13.00.4	2019-10-29	<a href="#">NORTHWIND.SIARD</a>	884.6 KB	2.1	Valid	Pr
client_sitename_app_version	unspecified	Microsoft SQL Server 13.00.4	2019-10-29	<a href="#">CLIENT_SITENAME_AP</a>	884.6 KB	2.1	Not Valid	Pr
ECH-0165	E-ARK	MySQL 5.6.45	2019-10-17	<a href="#">ECH-0165_OE.SIARD</a>	884.6 KB	2.1	Valid	Pr
ABC_012_ACM	unspecified	Oracle 11.1.0.6	2019-10-29	<a href="#">ABC_012_ACM.SIARD</a>	884.6 KB	2.1	Valid	Pr
Sybase	unspecified	Sybase ASE	2019-09-10	<a href="#">SYBASE.SIARD</a>	10.1 MB	2.1	Validation	Pr
SalesManager_Wilco	Keep-soluti	MySQL 5.6.45	2019-10-16	<a href="#">SALESMANAGER_WILC</a>	884.6 KB	2.1	Valid	Pr
sakila-migration	unspecified	MySQL 5.6.45	2019-10-02	<a href="#">SAKILA.SIARD</a>	10.1 MB	2.1	Not Valid	Pr
sakila	MySQL team	MySQL 5.5.5-10.1.11-MariaDB	2016-09-15	<a href="#">SAKILA_DESCRIPTION</a>	836.2 KB	2.1	Not Valid	Re

# DBTPK Enterprise features

## Data transformation

Transform content to answer useful questions

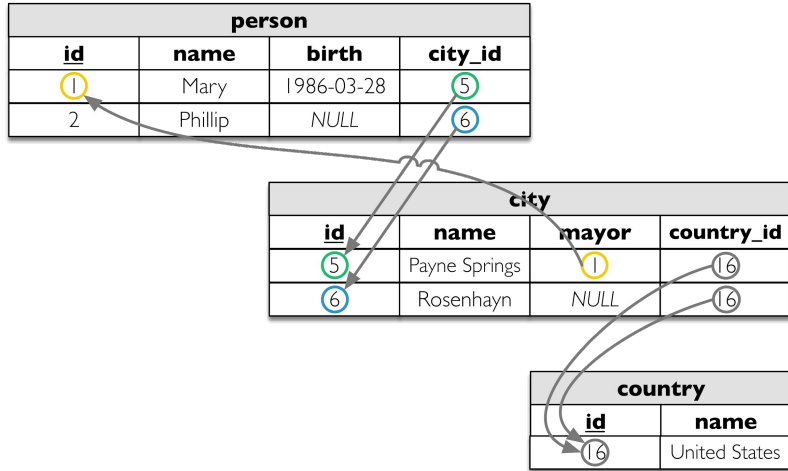
**De-normalization** and table and **column hiding**, to simplify browsing and allow **anonymization** of content

The screenshot displays the DBTPK Enterprise web interface. On the left, a sidebar lists various database tables and views, 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 shows a 'Schema' diagram with nodes representing tables and their relationships. The 'payment' table is highlighted in blue. Below the diagram, a table description is provided for the 'payment' 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

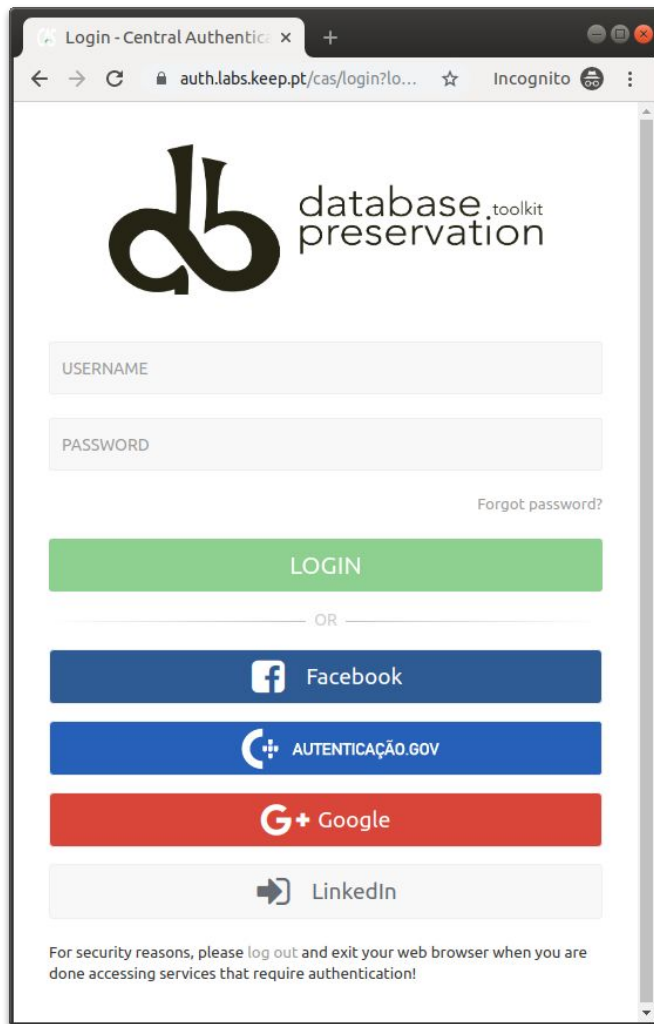
## DBPTK Enterprise features

### 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 interface includes a "Filter sidebar" on the left with a list of tables. The "film" table is selected and highlighted in blue. The main content area shows the "film" table 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."

A search interface is displayed with the following criteria:

- release\_year: 2006 to 2006
- rating: PG-13

Buttons for "ADD SEARCH FIELD", "CLEAR", "SAVE SEARCH", and "SEARCH" are visible.

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 data.

title	description	release_year	length	rating	special_features
AIRPLANE SIERRA	A Touching Saga of a F...	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 chart. The table has columns for title, description, release\_year, length, rating, and special\_features. The chart, titled 'The duration of the film', is a horizontal bar chart showing the length of various movies. The x-axis represents length in minutes, ranging from 0 to 200. The y-axis lists movie titles. The chart shows that most movies have a length between 80 and 140 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 Orlean	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	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 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 CHARLOTS	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 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 list of activity events 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 contains filters for Components and Methods.

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

1-20 of 2,972 [Show More](#)

**Components**

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

**Methods**

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

## DBPTK Enterprise & Desktop

### Multiple languages supported

#### Interface translated into:

English, German, Estonian, Czech, Portuguese

#### Search stemming and stopwords support for:

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

# DBPTK Developer

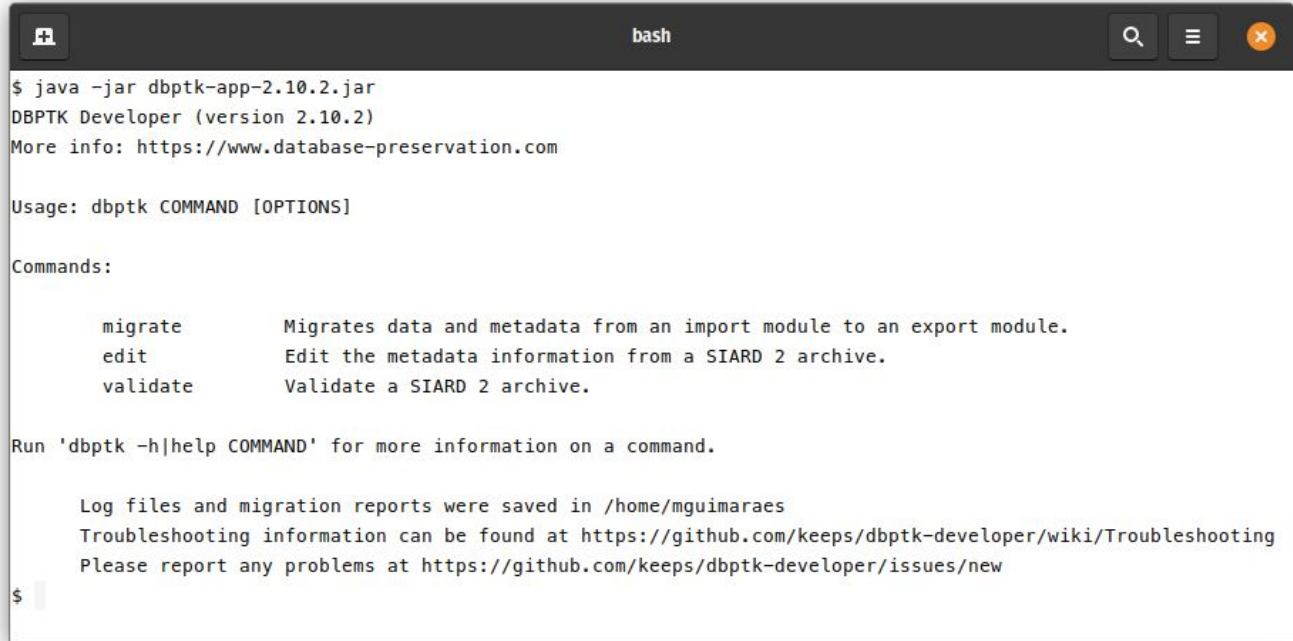
Top features

# 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.



```
bash
$ java -jar dbptk-app-2.10.2.jar
DBPTK Developer (version 2.10.2)
More info: https://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.

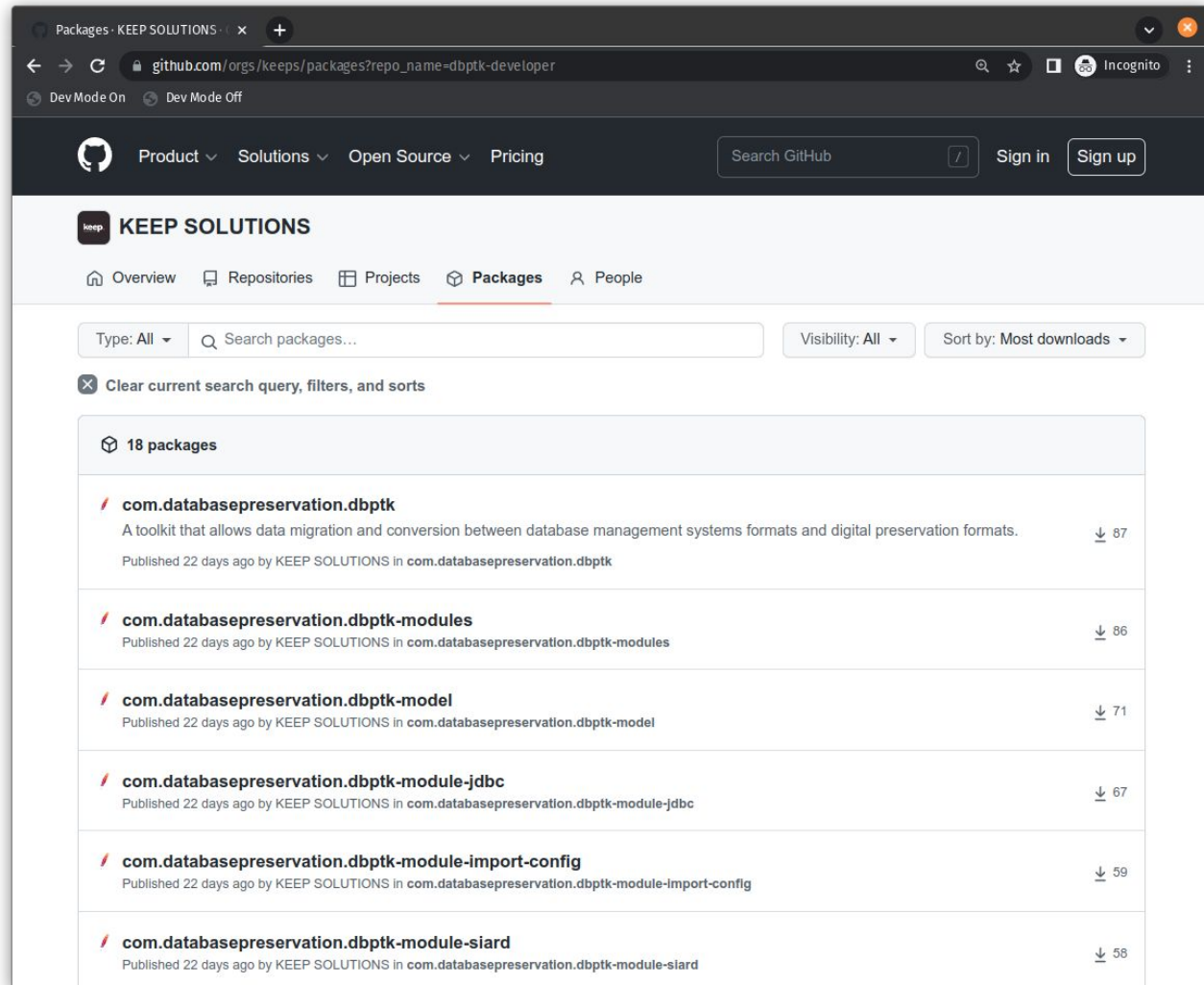
Log files and migration reports were saved in /home/mguimaraes
Troubleshooting information can be found at https://github.com/keeps/dbptk-developer/wiki/Troubleshooting
Please report any problems at https://github.com/keeps/dbptk-developer/issues/new

$
```

# DBPTK Developer features

## Systems integration Java library

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



The screenshot shows the GitHub Packages page for the organization KEEP SOLUTIONS. The page displays a list of 18 packages. The first package is `com.databasepreservation.dbptk`, which is a Java library for database migration and conversion. The page also shows the package's description, its publication date (22 days ago), and the number of downloads (87).

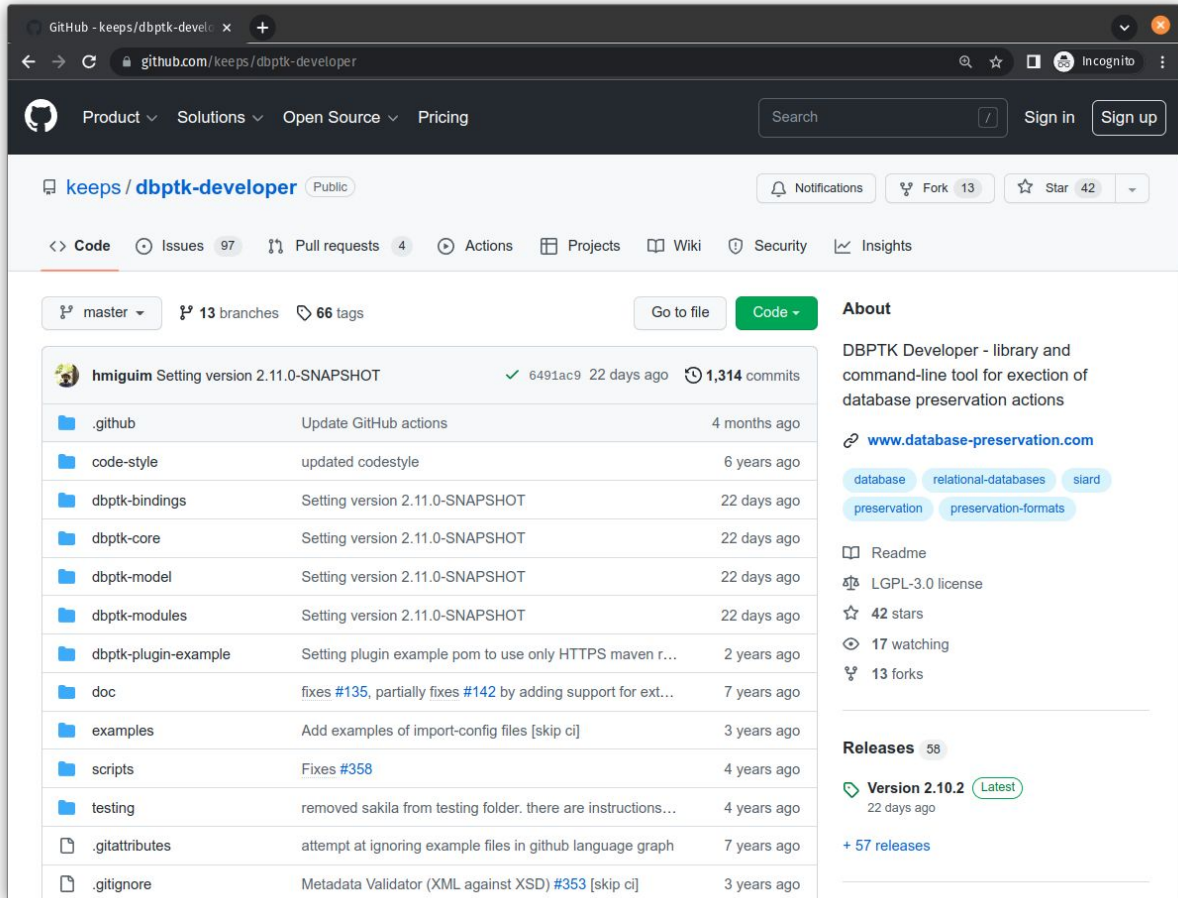
Package Name	Description	Downloads
<code>com.databasepreservation.dbptk</code>	A toolkit that allows data migration and conversion between database management systems formats and digital preservation formats.	87
<code>com.databasepreservation.dbptk-modules</code>		86
<code>com.databasepreservation.dbptk-model</code>		71
<code>com.databasepreservation.dbptk-module-jdbc</code>		67
<code>com.databasepreservation.dbptk-module-import-config</code>		59
<code>com.databasepreservation.dbptk-module-siard</code>		58

# 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.





## 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
- Permissions per database

# 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

# Demonstration



## *Sakila*

### **DVD rental store database**

The Sakila database is a nicely normalised schema modelling a DVD rental store, featuring things like films, actors, film-actor relationships, and a central inventory table that connects films, stores, and rentals.

The Sakila sample database was initially developed by Mike Hillyer, a former member of the MySQL AB documentation team. It is intended to provide a standard schema that can be used for examples in books, tutorials, articles, samples, and so forth. The Sakila sample database also serves to highlight features of MySQL such as Views, Stored Procedures, and Triggers.

<https://dev.mysql.com/doc/sakila/en/>

## Archiving Sakila

The DVD store is to be **decommissioned**, but the **content of the database** is to be **kept** for historical purposes.

The **producer** will extract the content into **SIARD** using **DBPTK Desktop** and transfer it to the **archive**.

The **archivist** will **validate** it and **incorporate** into the repository. Later on, the **designated community** would like to make **specific requests** on the content (like a list of films with cast).

## **Producer workflow**

- 1) Prepare the database: read-only user, archival view (optional)
- 2) Extract the database to SIARD: select tables, select views, materialize views, custom views, external files, other options.
- 3) Validate the SIARD (optional): validate against the specification, manual validation of content with browse
- 4) Send the SIARD and additional documentation to the archive

## **Archivist workflow**

- 1) Receive the SIARD and documentation from producer
- 2) Validate against specification
- 3) Browse and validate content
- 4) Add to archive
- 5) Provide access to the designated community
- 6) Ensure designated community specific requests are fulfilled

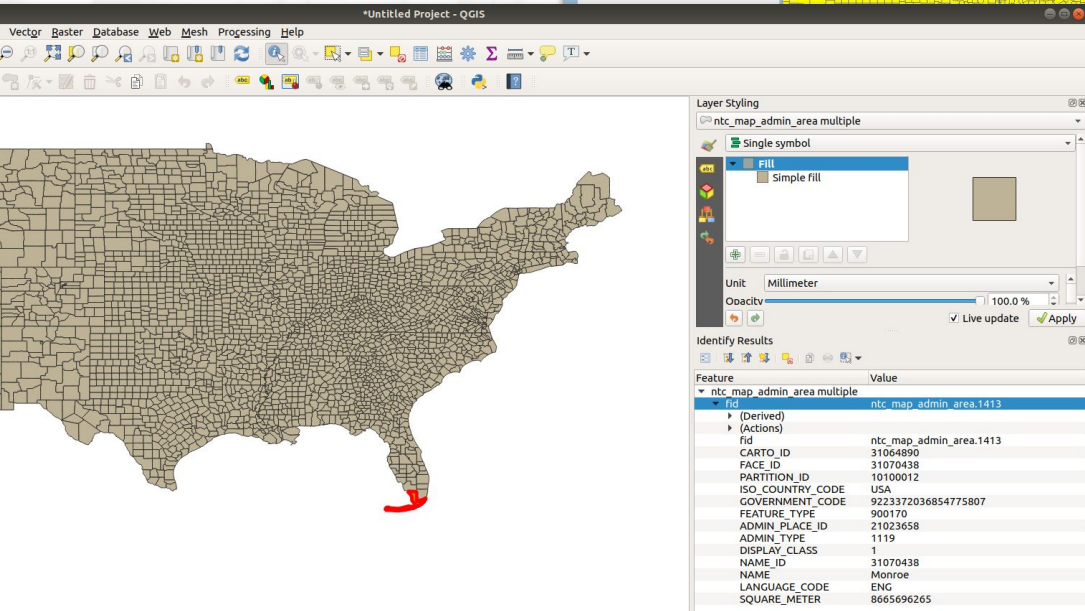
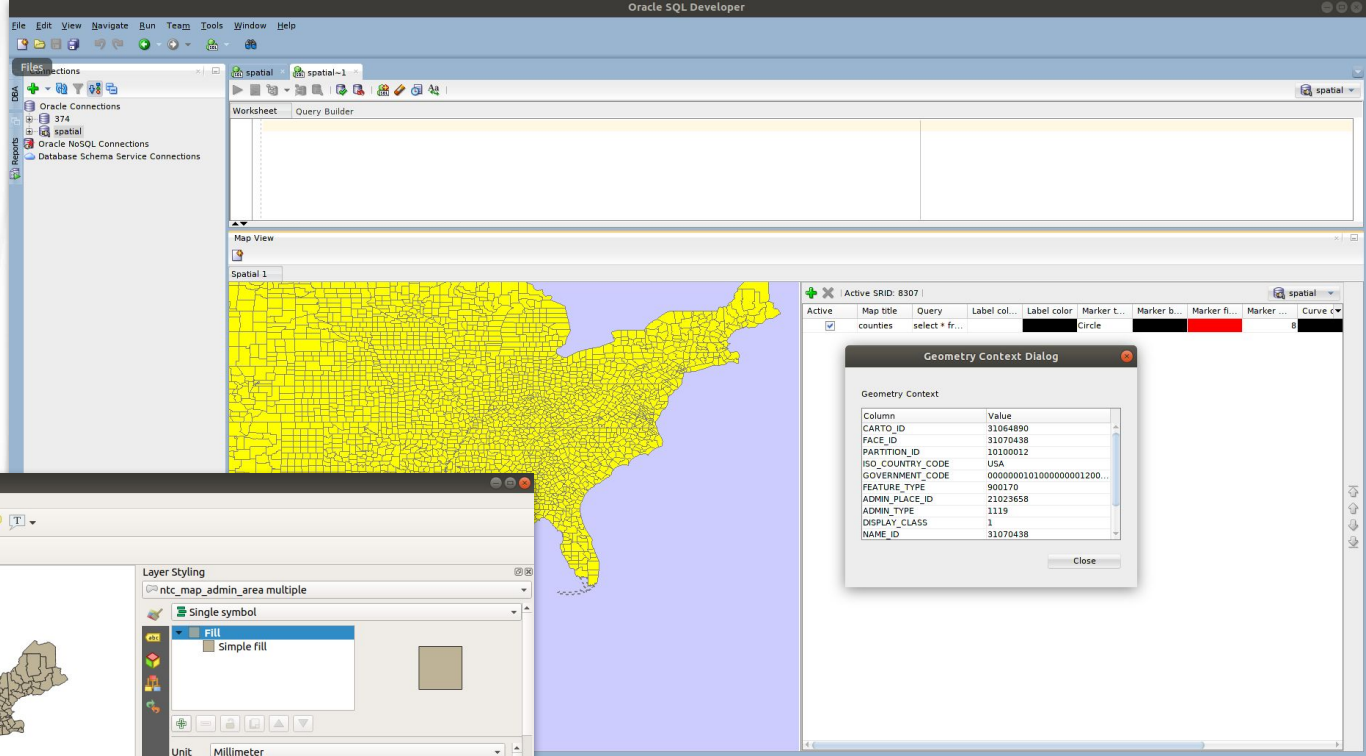
# **Advanced features**

We didn't have time for everything

# Convert Oracle geodata

Export Oracle geographical information

Closed specification only supported in specific Oracle license types



## Into GML

An open international standard

Ready to be used with open-source desktop or Web tools.

# Permission per database

Allow only authorized roles to browse a specific database.

**db DBPTK Enterprise**

Databases > sakila-teste Administrator Administration English

## sakila-teste

(af314267-00f3-4360-b665-01c7a1eb9fd0)  
unspecified

**Archival date:** 2019-10-15 01:00 (UTC+1)  
**Archiver:** unspecified  
**Archiver contact:** unspecified  
**Client machine:** gbarros-HP

**Database product:** MySQL 5.6.45  
**Data origin timespan:** unspecified  
**Data owner:** unspecified  
**Producer Application:** Database Preservation Toolkit (version 2.6.1)

### SIARD

See details about the archived format and manage the metadata of tables, columns and other technical elements of the database.

**Version:** SIARD standard version 2.1 (SEE SPECIFICATION)  
**Path:** SAKILA.SIARD  
**Size:** 10.1 MB

[EDIT METADATA](#) [DELETE](#)

### Validation

See details about the validation process.

**Status:** Not Validated

[VALIDATE](#)

### Browsing

See details about the database status. Manage the database metadata by configuring table, columns and data.

**Status:** Ready

[BROWSE](#) [CONFIGURATION](#) [DELETE](#)

### Permissions

See the roles that are authorized to browse this database.

**Groups:** Médicos Operadores users Arquivistas

[EDIT PERMISSIONS](#)

[BACK](#) [DELETE](#)

**keep.**

Preserving the future

**Questions?**

<https://database-preservation.com>



## **Usual issues and questions**

Ensure security when extracting information to SIARD

Database security restriction are tight: remote access disallowed, hard to setup access users and minimal access permissions

Lack of features or issues connecting (integrated security, multiple collations, database extensions that require additional libraries to connect and extract.

Missing support for specific DBMS

Disk space required for extracting very large databases

# Issues and help desk

## Issues

### Get information about the issue

DBPTK Desktop, menu Help > Logs

DBPTK Enterprise, docker logs

How to **reproduce** the issue



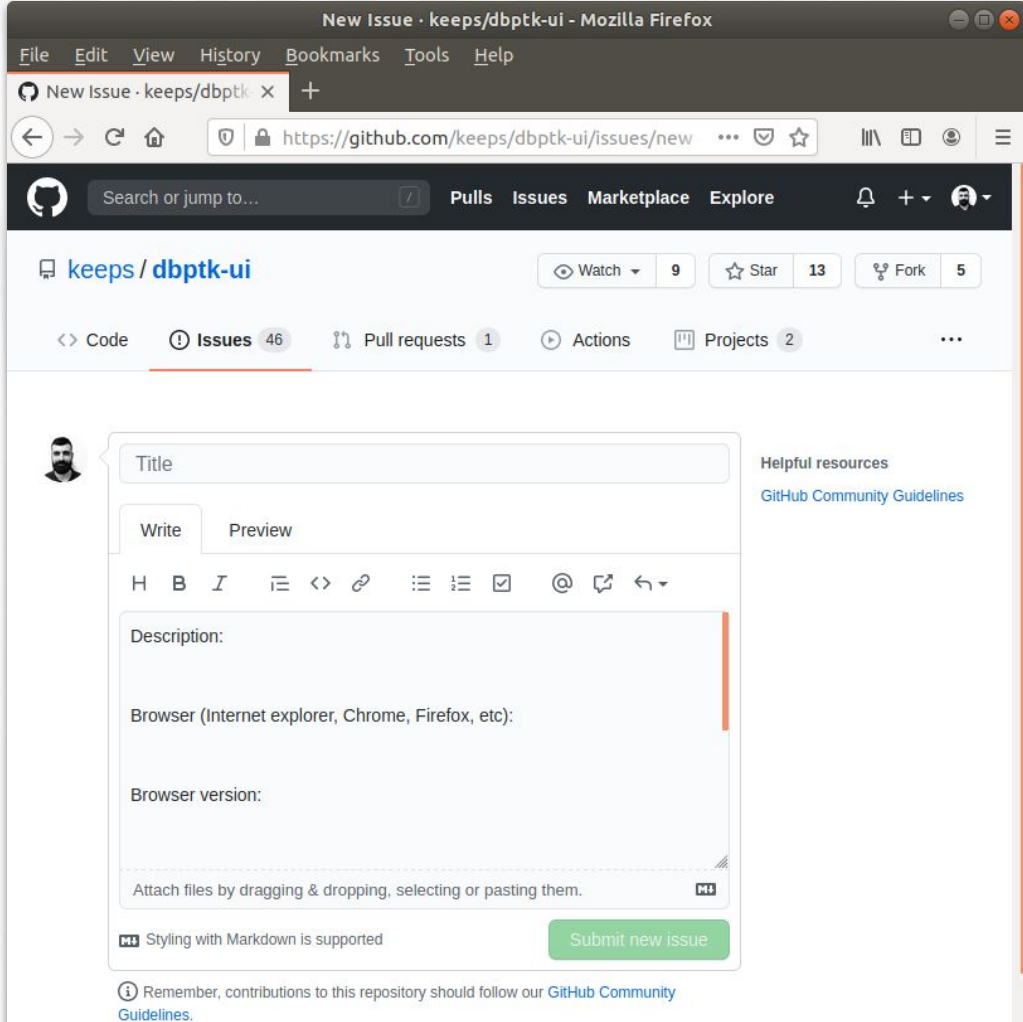
## Issues (free option)

### Submit the issue

Open-source support

Register free GitHub account

<https://github.com/keeps/dbptk-ui/issues/new>



## Issues (commercial service)

### Submit the issue

Commercial support

Contact [sales@keep.pt](mailto:sales@keep.pt)

The screenshot shows a web browser window with the URL `redmine.keep.pt/projects/helpdesk-dbptk-dpc/issues/new`. The page title is "New issue - Helpdesk | DB". The interface is in Portuguese and includes a navigation bar with "Home", "My page", "Projects", and "Help". The user is logged in as "Ifaria" and can access "My account" or "Sign out". The main navigation menu includes "Overview", "Activity", "Issues", "New issue" (selected), "Spent time", "Documents", "Wiki", and "Settings".

The "New issue" form contains the following fields and options:

- Tracker:** Support (dropdown menu)
- Subject:** Text input field
- Description:** Rich text editor with formatting options (B, I, U, C, H1, H2, H3, bulleted list, numbered list, pre, code, image, link, unlink)
- Status:** New (dropdown menu)
- Priority:** Low (10 days) (dropdown menu)
- Assignee:** Dropdown menu
- Parent task:** Search input field
- Start date:** 24/07/2020 (calendar icon)
- Due date:** dd/mm/aaaa (calendar icon)
- Estimated time:** Input field followed by "Hours"
- Files:** "Select file" button (Maximum size: 488 MB)
- Watchers:** Checkboxes for Gabriel Barros, Miguel Ferreira, Luís Faria, and Miguel Guimarães. A "Search for watchers to add" link is also present.

At the bottom of the form, there are three buttons: "Create", "Create and continue", and "Preview".