Database preservation
DPC training course
Practical session (basic features)
Resolution
Export provided sakila database (MySQL sample database) to SIARD
Click here to start the process of create a SIARD file.
1. Select the DBMS on the left sidebar panel and fill up the connection form.

2. Test the connection to ensure that you have the right information.

3. Click Next to continue the process.
Click Next to continue the process. By default it will select all tables and document all views.
Custom Views

Custom views option allows you to aggregate data with SQL queries.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schema Name</td>
<td>Schema where the custom view should be integrated</td>
</tr>
<tr>
<td>Custom View Name</td>
<td>Name of the custom view</td>
</tr>
<tr>
<td>Description</td>
<td>Description of the custom view</td>
</tr>
<tr>
<td>Query</td>
<td>SQL query to obtain the data to populate the custom view</td>
</tr>
</tbody>
</table>

Click **Skip** to continue the process
Merkle Tree Filter

This filter module is used to generate a hash which represents the content of large data structures. This process aims to fulfill the requirement of completeness and correctness of the archival process, i.e. to ensure that no message is lost (not archived or not correctly archived).

Apply Merkle Tree Filter

Destination Folder
Path to save the merkle tree

Digest
- SHA-256
Define the type of integrity information. Supported types are: MD5, SHA-1 and SHA-256.

Explain
Flags to show the hashes of the whole tree (tables, rows, cells). Mainly for debug purpose.

Font Case
- Lower case
Define the type of font case for the message digest. Supported font case are: upper case and lower case.

Click **Skip** to continue the process.
### SIARD Export Options

Resumes the different options available to create a SIARD file.

<table>
<thead>
<tr>
<th>SIARD Version</th>
<th>SIARD 2</th>
<th></th>
</tr>
</thead>
</table>

#### Destination Folder

1. Select the destination folder

2. Choose compress checkbox (this will reduce the size of the SIARD file)

3. Hit **Next** to continue the process
1. Fill up metadata information about the SIARD file

2. Click **Create** to start the migration process
Wait for the process to finish, this may take a while, depending on the machine specs and total size of the database.
Archival date: 2020-08-03 01:00 (UTC+1)
Archiver: mguiamaraes
Archiver contact: mguiamaraes@keep.pt
Client machine: unspecified

Database product: MySQL 5.6.49
Data origin timespan: unspecified
Data owner: unspecified
Producer Application: Database Preservation Toolkit (version 2.9.2)

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: PRACTICAL_SESSION_1_SIARD
Size: 789.3 KB

Validation
See details about the validation process.
Status: Not Validated

Browsing
See details about the database status. Manage the database metadata by configuring table, columns and data.
Status: Pre-loaded
Edit the database name to your user name

Edit the description of column postal_code of table address
Click here to start the metadata edition process. This process will alter the SIARD file metadata content.
### Database Information

**Global information at database level**

<table>
<thead>
<tr>
<th><strong>Name</strong></th>
<th>mguimaraes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Archival date</strong></td>
<td>2020-08-03</td>
</tr>
<tr>
<td><strong>Archivist</strong></td>
<td>mguimaraes</td>
</tr>
<tr>
<td><strong>Archivist contact</strong></td>
<td><a href="mailto:mguimaraes@keep.pt">mguimaraes@keep.pt</a></td>
</tr>
<tr>
<td><strong>Client machine</strong></td>
<td>unspecified</td>
</tr>
<tr>
<td><strong>Product</strong></td>
<td>MySQL 5.6.49</td>
</tr>
<tr>
<td><strong>User</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Data origin time span</strong></td>
<td>unspecified</td>
</tr>
<tr>
<td><strong>Data owner</strong></td>
<td>unspecified</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>DPC Workshop practical session activities resolution</td>
</tr>
<tr>
<td><strong>Producer application</strong></td>
<td>Database Preservation Toolkit (version 2.9.2)</td>
</tr>
</tbody>
</table>
Click on **Save** to commit the changes.

The address table contains address information for customers, staff, and stores. The address table primary key appears as a foreign key in the customer, staff, and store tables.

<table>
<thead>
<tr>
<th>Column name</th>
<th>Type name</th>
<th>Original type name</th>
<th>Nullable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>address_id</td>
<td>SMALLINT</td>
<td>SMALLINT UNSIGNED</td>
<td>NO</td>
<td>A surrogate primary key used to uniquely identify each address in the table.</td>
</tr>
<tr>
<td>address</td>
<td>CHARACTER</td>
<td>VARCHAR</td>
<td>NO</td>
<td>The first line of an address.</td>
</tr>
<tr>
<td>address2</td>
<td>CHARACTER</td>
<td>VARCHAR</td>
<td>YES</td>
<td>An optional second line of an address.</td>
</tr>
<tr>
<td>district</td>
<td>CHARACTER</td>
<td>VARCHAR</td>
<td>NO</td>
<td>The region of an address, this may be a state, province, prefecture, etc.</td>
</tr>
<tr>
<td>city_id</td>
<td>SMALLINT</td>
<td>SMALLINT UNSIGNED</td>
<td>NO</td>
<td>A foreign key pointing to the city table.</td>
</tr>
<tr>
<td>postal_code</td>
<td>CHARACTER</td>
<td>VARCHAR</td>
<td>YES</td>
<td>The postal code or ZIP code of the address (where applicable)</td>
</tr>
<tr>
<td>phone</td>
<td>CHARACTER</td>
<td>VARCHAR</td>
<td>NO</td>
<td>The telephone number for the address.</td>
</tr>
<tr>
<td>last_update</td>
<td>TIMESTAMP</td>
<td>TIMESTAMP</td>
<td>NO</td>
<td>The time that the row was created or most recently updated.</td>
</tr>
</tbody>
</table>
The address table contains address information for customers, staff, and stores. The address table primary key appears as a foreign key in the customer, staff, and store tables.

<table>
<thead>
<tr>
<th>Column name</th>
<th>Type name</th>
<th>Original type name</th>
<th>Nullable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>address_id</td>
<td>SMALLINT</td>
<td>SMALLINT UNSIGNED</td>
<td>NO</td>
<td>A surrogate primary key used to uniquely identify each address in the table.</td>
</tr>
<tr>
<td>address</td>
<td>CHARACTER</td>
<td>VARCHAR</td>
<td>NO</td>
<td>The first line of an address.</td>
</tr>
<tr>
<td>address2</td>
<td>CHARACTER</td>
<td>VARCHAR</td>
<td>YES</td>
<td>An optional second line of an address.</td>
</tr>
<tr>
<td>district</td>
<td>CHARACTER</td>
<td>VARCHAR</td>
<td>NO</td>
<td>The region of an address, this may be a state, province, prefecture, etc.</td>
</tr>
<tr>
<td>city_id</td>
<td>SMALLINT</td>
<td>SMALLINT UNSIGNED</td>
<td>NO</td>
<td>A foreign key pointing to the city table.</td>
</tr>
<tr>
<td>postal_code</td>
<td>CHARACTER</td>
<td>VARCHAR</td>
<td>YES</td>
<td>The postal code or ZIP code of the address (where applicable)</td>
</tr>
<tr>
<td>phone</td>
<td>CHARACTER</td>
<td>VARCHAR</td>
<td>NO</td>
<td>The telephone number for the address.</td>
</tr>
<tr>
<td>last_update</td>
<td>TIMESTAMP</td>
<td>TIMESTAMP</td>
<td>NO</td>
<td>The time that the row was created or most recently updated.</td>
</tr>
</tbody>
</table>
DPC Workshop practical session activities resolution

Archival date: 2020-08-03
Archiver: mguimaraes
Archiver contact: mguimaraes@keep.pt
Client machine: unspecified

Database product: MySQL 5.6.49
Data origin timespan: unspecified
Data owner: unspecified
Producer Application: Database Preservation Toolkit (version 2.9.2)

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: PRACTICAL_SESSION_1_SIARD
Size: 789.3 KB

Validation
See details about the validation process.
Status: Not Validated

Browsing
See details about the database status. Manage the database metadata by configuring table, columns and data.
Status: Pre-loaded

Click on Validate to start the validation process
Click on Confirm
**Validation**

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

**Database Name:** mguilaraes

- **Requirements that passed:** 27
- **Requirements that failed:** 0
- **Number of errors:** 0
- **Number of warnings:** 75
- **Number of skipped:** 11

**SIARD specification:** SIARD-21

**Additional checks specification:** OPEN

**Report:** OPEN

---

<table>
<thead>
<tr>
<th>Test Case</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>T_6.4.2</td>
<td>OK</td>
</tr>
<tr>
<td>T_6.4.2</td>
<td>OK</td>
</tr>
<tr>
<td>T_6.4.2</td>
<td>OK</td>
</tr>
<tr>
<td>T_6.4.2</td>
<td>OK</td>
</tr>
<tr>
<td>T_6.4.2</td>
<td>OK</td>
</tr>
<tr>
<td>T_6.4.2</td>
<td>OK</td>
</tr>
<tr>
<td>T_6.4.2</td>
<td>OK</td>
</tr>
<tr>
<td>T_6.4.2</td>
<td>OK</td>
</tr>
<tr>
<td>T_6.4.2</td>
<td>OK</td>
</tr>
<tr>
<td>T_6.4.2</td>
<td>OK</td>
</tr>
</tbody>
</table>

---

Click on **Back** and check the information about the validation.

---

*Validation Details:

- T_6.4.2: Validation finish on path: content/schema1/table13/table13.xml
- 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
- 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
- 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
- T_6.4.2: The table file consists of row elements containing the data of a line subdivided into the various columns (c1, c2, ...).
- T_6.4.2: 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.
- T_6.4.2: 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.
Archival date: 2020-08-03
Archiver: mguimaraes
Archiver contact: mguimaraes@keep.pt
Client machine: unspecified

Database product: MySQL 5.6.49
Data origin timespan: unspecified
Data owner: unspecified
Producer Application: Database Preservation Toolkit (version 2.9.2)

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: PRACTICAL_SESSION_1_SIARD
Size: 789.3 KB

Validation
See details about the validation process.

Status: Valid
Details: 27 successes, 0 failures, 11 skipped
Warnings: There are 75 warnings
Validated at: 2020-08-03 10:29 (UTC+1)
Validate by: DBPTK developer 2.9.2 (SEE DETAILS)

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

Status: Pre-loaded
Import the SIARD file to DBPTK Enterprise
Welcome to Database Preservation Toolkit Enterprise!

A web application to browse and search the content of multiple large archived databases.

The Database Preservation Toolkit Enterprise is a lightweight web viewer for relational databases, especially if preserved in SIARD 2, that uses SOLR as a backend, and allows browsing, search, and export.

**Enterprise Architecture**
For large institutions with many databases and users
A web application that can be horizontally scaled to support many large-sized databases, accessed by hundreds of users simultaneously. Need help? Get support.

**Manage multiple databases**
Single system, multiple databases
Search through the databases, manage their status, enrich their metadata, validate them, make them ready for your end users.

**Browse and search**
Allow users to access database content on the Web
Allow them to search on a prepared, de-normalized and anonymised database content.

**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.

**SIARD validation**
Validate archived database
Validate SIARD against specification plus many additional checks for a thorough validation.

**Edit SIARD metadata**
Enrich archived database with descriptions
Add descriptions to database, tables and columns to better understand its contents.

**Advanced data transformation**
Transform content to answer useful questions
De-normalization and table/column hiding, to simplify browsing/search and allow anonymization of content.

**Single sign-on**
Support for multiple protocols
LDAP, Active Directory, Database, SAML, AD FS, OAuth, OpenID, Google, Facebook, Twitter, FIDO U2F, YubiKey, Google Authenticator, Authy, etc. Supports internal authorization definition or configurable external authorization.

Click on Login
Click on **Load SIARD file**
Drop the files to load or choose via file manager.
Load SIARD file
Upload is done. You can upload more files.

Drop files here
CHOOSE FILES...

Click Open SIARD
Export actors list into CSV/Excel
Click Browse
The actor table lists information for all actors. The actor table is joined to the film table by means of the Film_actor table.

<table>
<thead>
<tr>
<th>actor_id</th>
<th>first_name</th>
<th>last_name</th>
<th>last_update</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PENELope</td>
<td>GUNESS</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>10</td>
<td>CHRISTian</td>
<td>GABLE</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>100</td>
<td>SPENCer</td>
<td>DEPP</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>101</td>
<td>SUSAn</td>
<td>DAVIS</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>102</td>
<td>WALTer</td>
<td>TORN</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>103</td>
<td>MATThew</td>
<td>LEIGH</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>104</td>
<td>PENELope</td>
<td>CROWNy</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>105</td>
<td>SIDNEY</td>
<td>CROWE</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>106</td>
<td>GROUCHO</td>
<td>DUNST</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>107</td>
<td>GNA</td>
<td>DEGENERES</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>108</td>
<td>WARREN</td>
<td>NOLTE</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>109</td>
<td>SYLVESTER</td>
<td>DERR</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>11</td>
<td>ZERO</td>
<td>CAGE</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>110</td>
<td>SUSAn</td>
<td>DAVIS</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>111</td>
<td>CAMERon</td>
<td>ZELLWEGER</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>112</td>
<td>RUSSELL</td>
<td>BACALL</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>113</td>
<td>MORGan</td>
<td>HOPKINS</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>114</td>
<td>MORGan</td>
<td>MCDORMAND</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>115</td>
<td>HAMPSON</td>
<td>BAILE</td>
<td>2006-02-15 04:43:33</td>
</tr>
<tr>
<td>116</td>
<td>DAN</td>
<td>STREEp</td>
<td>2006-02-15 04:43:33</td>
</tr>
</tbody>
</table>
Select All rows to export
Include descriptions
Click Export
Do a query: How many addresses from Texas there are?
1. Fill up the red box with value ‘TEXAS’ and click Enter or the magnifying glass.

2. Check the number of results to answer the question.
Activities for the braves

Search on the activity log in DBPTK Enterprise
Welcome to Database Preservation Toolkit Enterprise!

A web application to browse and search the content of multiple large archived databases.

The Database Preservation Toolkit Enterprise is a Lightweight web viewer for relational databases, especially if preserved in SIARD 2, that uses SOLR as a backend, and allows browsing, search, and export.

Enterprise Architecture
For large institutions with many databases and users
A web application that can be horizontally scaled to support many large-sized databases, accessed by hundreds of users simultaneously. Need help? Get support.

Browse and search
Allow users to access database content on the Web
Allow them to search on a prepared, de-normalized and anonymized database content.

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

Advanced data transformation
Transform content to answer useful questions
De-normalization and table/column hiding, to simplify browsing/search and allow anonymization of content.

Manage multiple databases
Single system, multiple databases
Search through the databases, manage their status, enrich their metadata, validate them, make them ready for your end users.

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.

Edit SIARD metadata
Enrich archived database with descriptions
Add descriptions to database, tables and columns to better understand its content.

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.

Click on Login
Click on Activity log.
# Activity log

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

<table>
<thead>
<tr>
<th>Date</th>
<th>Component</th>
<th>Method</th>
<th>User</th>
<th>Duration</th>
<th>Address</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020-08-03</td>
<td>Activity log</td>
<td>Find</td>
<td>mguimaraes</td>
<td>93ms</td>
<td>81.84.255.101</td>
<td>Success</td>
</tr>
<tr>
<td>2020-08-03</td>
<td>Database</td>
<td>Find</td>
<td>mguimaraes</td>
<td>52ms</td>
<td>81.84.255.101</td>
<td>Success</td>
</tr>
<tr>
<td>2020-08-03</td>
<td>Database</td>
<td>Find</td>
<td>mguimaraes</td>
<td>45ms</td>
<td>81.84.255.161</td>
<td>Success</td>
</tr>
<tr>
<td>2020-08-03</td>
<td>Database</td>
<td>Find</td>
<td>mguimaraes</td>
<td>74ms</td>
<td>81.84.255.161</td>
<td>Success</td>
</tr>
<tr>
<td>2020-08-03</td>
<td>Job</td>
<td>Find</td>
<td>mguimaraes</td>
<td>56ms</td>
<td>81.84.255.101</td>
<td>Success</td>
</tr>
<tr>
<td>2020-08-03</td>
<td>Job</td>
<td>Find</td>
<td>mguimaraes</td>
<td>5ms</td>
<td>81.84.255.101</td>
<td>Success</td>
</tr>
<tr>
<td>2020-08-03</td>
<td>Job</td>
<td>Find</td>
<td>mguimaraes</td>
<td>6ms</td>
<td>81.84.255.161</td>
<td>Success</td>
</tr>
<tr>
<td>2020-08-03</td>
<td>Job</td>
<td>Find</td>
<td>mguimaraes</td>
<td>4ms</td>
<td>81.84.255.161</td>
<td>Success</td>
</tr>
<tr>
<td>2020-08-03</td>
<td>Job</td>
<td>Find</td>
<td>mguimaraes</td>
<td>5ms</td>
<td>81.84.255.161</td>
<td>Success</td>
</tr>
<tr>
<td>2020-08-03</td>
<td>Job</td>
<td>Find</td>
<td>mguimaraes</td>
<td>4ms</td>
<td>81.84.255.161</td>
<td>Success</td>
</tr>
<tr>
<td>2020-08-03</td>
<td>Job</td>
<td>Find</td>
<td>mguimaraes</td>
<td>6ms</td>
<td>81.84.255.161</td>
<td>Success</td>
</tr>
</tbody>
</table>

1. Fill up the red box with your user and click Enter or the magnifying glass
2. Click on a row to see details
Log details

Date: 2020-06-03 11:51:50 UTC+1
Duration: Sms
User: mguimaraes
Address: 81.84.255.161
Component:
Method: Find
Search parameters:
- Search is mguimaraes
Outcome: Success
Activities for the Braves

Create the SIARD file using the DBPTK Developer

Replace with the password assign to your user