Relational Database Preservation Standards and Tools

Workshop guide

Authors: Luis Faria, Bruno Ferreira, Marcel Büchler, Kuldar Aas

Here a list of the steps to be demonstrated will be presented to help attendants follow the demonstrations. The workshop will be divided in three themes, that are relevant in the archive preservation workflow: extraction from live system, validation upon reception, and access information.

Preparation

Do you have a laptop? Then:

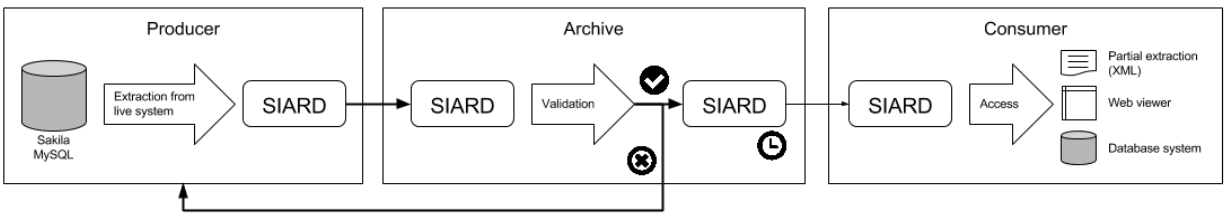
1. Request your pen drive with the software pack.
2. No pen drive?
   http://provided_address/pack.zip
   http://bit.ly/ipres16-db-pack
3. Unzip contents of software pack to your Desktop folder rename it to `WORKSHOP08`
4. Access the wireless “WORKSHOP08” with password “databases”

Scenario

The Sakila sample database will be used to demonstrate the features. This database is designed to represent a DVD rental store, with entities such as the stores, clients, staff, films, rentals and relationships between them.

We will be demonstrating three scenarios:

1. The extraction of a database from a live system into a SIARD file
2. Validating the SIARD file
3. Using a viewer to explore the contents of the SIARD file



1. Extraction from live system

A live MySQL database is provided in the system and a tool will be used to extract all information into a SIARD file. Two tools will be presented: Database preservation toolkit and SIARD Suite.

1.A. Database preservation toolkit

Execute the following steps:

1.A.1. Go to the software pack, and in the folder `01-db-visualization-toolkit` on your file explorer

1.A.2. Open a terminal

- **On Windows:**
  `SHIFT+Right Click` the folder and select “Open command window here”
- **On Mac OS:**
  `CMD+SPACE`, write “Terminal” and press `ENTER`, then change into the 01-db-visualization-toolkit directory. Example:
  `cd ~/Desktop/WORKSHOP08/01-db-preservation-toolkit`
- **On Linux**
  `Right Click` the folder and select “Open in terminal”

1.A.3. Execute the following command:

- **On Windows:**
  `migrate.bat provided_address`
- **On Mac OS:**
  `./migrate.sh provided_address`
- **On Linux:**
  `./migrate.sh provided_address`

EXERCISE #1

Rename the created SIARD file to “sakila-siard2.zip”, unzip it and open file “header/metadata.xml”:

#1.1. What is the database product version (i.e. MySQL version)?

#1.2. How many rows are in “actor” table?

1.B. SIARD Suite

1.B.1. Go to the software pack, and in the folder `02-SIARD-Suite-1.89` on your file explorer

1.B.2. Start SIARD Suite

- **On Windows (needs Java installed):**
  Doubleclick `siardedit.cmd`
- **On Mac OS and Linux (with Java installed):**
  Run `siardedit.sh` (either by doubleclick, or in a Terminal)
- **On Windows (without Java)**
  Open a Terminal (See 1.A.2 for instructions)
  Execute: `..\05-db-visualization-toolkit\jre\ windows\bin\java.exe -jar bin/SiardEdit.jar`
- **On Mac OS (without Java)**
  Open a Terminal (See 1.A.2 for instructions)
  Execute: `../05-db-visualization-toolkit/jre/ mac/bin/java -jar bin/SiardEdit.jar`
- **On Linux (without Java)**
  Open a Terminal (See 1.A.2 for instructions)
  Execute: `../05-db-visualization-toolkit/jre/ linux/bin/java -jar bin/SiardEdit.jar`

1.B.3. Import from Database:

1. Click File > New from database ...
2. Choose location to save SIARD file (e.g. Desktop) and enter filename (e.g. sakila-siard1), then click Save
3. Configure database connection
   Connection type: `MySQL`
   Connection Info: `provided_address/sakila`
   Database user: `ipres`
   Database password: `ipres`
   Click Open

1.B.4. Wait until the database export finished, then click Exit

- You can now view the file in SiardEdit and change Metadata (Click Apply beneath fields or File > Save to apply changes)
- You can rename the file to “sakila-siard1.siard.zip” and unzip it

---

EXERCISE #2

Browse the siard-file using SIARDEdit. Fill out the missing metadata fields “Data owner prior to archiving” and “Data creation time span” and save the changes. Add/change other descriptions.

---

2. Validating the SIARD file

2.A. Using KOST-val (Windows only)

2.A.1. Go to the software pack, and in the folder `03-KOST-Val` on your file explorer

2.A.2. Perform validation using KOST-Val

- **GUI (Windows only):**
    a. Doubleclick `KOST-Val_en.exe`
    b. Click `Select File …` and choose *sakila-siard1.siard*
    c. Click `Validate`
    d. Click `Yes` to show detailed result in Browser
- **Command Line (Windows only):**
    a. Open a Terminal (See 1.A.2 for instructions)
    b. Execute command:
    ```
    resources\jre6\bin\java.exe -jar kostval_en.jar --format
    ..\sakila_siard1.siard
    ```
    c. Detailed logfile can be found in folder *logs*

EXERCISE #3

Validate the provided sakila_invalid.siard file with KOST-Val. What is wrong with this file?

---

3. Accessing information in the SIARD file

Two tools will be presented that aim to display information contained in a SIARD file: SIARDexcerpt and Database visualization toolkit.

3.A. SIARDexcerpt (Windows only)

3.A.1. Go to the software pack, and in the folder `04-SIARDexcerpt` on your file explorer

3.A.2. Perform extraction using SIARDexcerpt

- **GUI (Windows only):**
    a. Doubleclick `SIARDexcerpt_en.exe`
    b. Click Select File … and choose *sakila-siard1.siard*
    c. Enter Configuration: `country`
    d. Click Run
    e. Search (e.g. for `land`) (click Run, click Yes)
    f. Click on 3. Excerption
    g. Search for a certain Id (e.g. `91`) (click Run, click Yes)
- **Command Line (Windows only)**
    a. Open a Terminal (See 1.A.2 for instructions)
    b. Execute command to initialize with “country”:
    ```
    resources\jre6\bin\java.exe -jar siardexcerpt_en.jar
    ..\sakila-siard1.siard country --init
    ```
    c. Execute command, to search for “land”:
    ```
    resources\jre6\bin\java.exe -jar siardexcerpt_en.jar
    ..\sakila-siard1.siard country --search land
    ```
    d. Inspect file *Output\sakila_siard1.siard_land_SIARDsearch.xml* to identify Id of record to be extracted (e.g. 91), then execute command:
    ```
    resources\jre6\bin\java.exe -jar siardexcerpt_en.jar
    ..\path\to\sakila-siard1.siard country --excerpt 91
    ```
    e. Database Excerpt can be found in file *Output\sakila_siard1.siard_91_SIARDexcerpt.xml*

EXERCISE #4

Run initialisation without a configuration (use “(..)” for no configuration).
#4.1. Which table is automatically identified as the “main” table of the database?

Having initialised SIARDexcerpt with (..), search for “ven”:
#4.2. Who lives at “943 Johannesburg Avenue”?
#4.3. What about “23 Workhaven Lane”?
#4.4. Are they customers or staff?

3.B. Database Visualization Toolkit

OPTION 1
A server is available at http://provided_address:8080/ to explore a Sakila database loaded with additional descriptive metadata.

OPTION 2
To set up your own, execute the following steps:
3.B.1. Go to the software pack, and in the folder `05-db-visualization-toolkit`
3.B.2. Open a terminal
3.B.3. Execute the following command, and keep the terminal window open:

- **On Linux:**
  `./start.sh`
- **On Mac OS:**
  `./start.command`
- **On Windows:**
  `start.bat`

3.B.4. The server will be available at http://127.0.0.1:8080/ preloaded with Sakila database

---

EXERCISE #5
Using the searches and filters find out:
#5.1. How many movies have the word “amazing”?
#5.2. Find the rentals by “Eleanor Hunt” in August 2005. How many films did he rent?

OPTION 2 CLEANUP
3.B.5. If you’d like to add more databases, checkout the instructions of the 3.A.3 step output
3.B.6. To shutdown the server, execute the following steps:

- **On Linux**, execute the command:
  `./stop.sh`
- **On Mac OS**, execute the command:
  `./stop.command`
- **On Windows**:
  a. Close both terminal windows that opened when the server was started
  b. Execute the command:
  `stop.bat`
  (some Windows versions output errors when executing this command;
  the errors can be safely ignored)