Workshop on Relational Database Preservation Standards and Tools

Luis Faria
KEEP SOLUTIONS
Rua Rosalvo de Almeida 5
4710 Braga, Portugal
lfaria@keep.pt

Marcel Büchler
Swiss Federal Archives
Archivstrasse 24
3003 Bern, Switzerland
marcel.buechler@bar.admin.ch

Kuldar Aas
National Archives of Estonia
J. Liivi 4
50409 Tartu, Estonia
kuldar.aas@ra.ee

ABSTRACT

This 3-hour workshop focuses on presenting the current state-of-the-art relational database preservation standards and tools used by major national archives and other institutions.

It presents **SIARD 2**, a new preservation format for relational databases. It also presents the current tools for harvesting information from live database management systems[1] into SIARD format and back, namely **SIARD Suite**[2] and the **Database Preservation Toolkit**[3]. Furthermore, two tools to access and view the information preserved in SIARD-files are presented: the **E-ARK database viewer**[4] and **SIARDexcerpt**[5].

The workshop includes live demonstration of the tools and prompts the participants to use them on their own laptops using the demonstration databases provided.

This workshop closely relates to a tutorial on relational database preservation guidelines and use cases, that focuses on the operational concerns of database preservation and relevant real-world use cases.

Keywords

Preservation; Archive; Relational Database; Ingest; Access; SIARD; Tools; E-ARK

1. INTRODUCTION

Databases are widely used to store and manage information. That is why archives have to ask themselves how they should preserve this type of information so that the databases will still be understandable in 20 or 50 years time. Furthermore, the database content needs to be stored independent of its specific Database Management System (DBMS), because a DBMS usually uses proprietary formats whose specifications are not freely accessible. In 2007 the Swiss Federal Archives (SFA) developed and standardised such an open format named SIARD (Software Independent Archiving of Relational Databases). Since 2008 SIARD has been actively used in the Swiss Federal Administration. Subsequently, the format has spread around the world, and is now used in many archives. In 2013 the SFA and the Swiss Coordination Agency for the Preservation of Electronic Files (KOST-CECO) specified the SIARD format as

an eCH standard (Swiss E-Government Standards). working together with the E-ARK project (European Archival Records and Knowledge Preservation) the version 2.0 of the SIARD format was developed in 2015. To convert databases into SIARD files there are two existing tools: SIARD Suite (developed by the SFA) and the db-preservation-toolkit (developed by KEEP SOLUTIONS from Portugal).

2. OUTLINE

The workshop starts with a brief general introduction to the topic of database preservation, to familiarise the participants with the motivation and challenges of this topic.

In order to understand the possibilities and limitations of database preservation using the SIARD 2.0 some knowledge of the SIARD format is needed. Therefore we explain how the SIARD 2.0 format is based on four internationally recognised standards (XML, SQL:2008, UNICODE and ZIP64), and how SIARD files are structured.

Equipped with this theoretical background, workshop participants will be presented with some tools for database preservation. The Database Preservation Toolkit is open source software for conversion of live or backed-up databases into preservation formats such as SIARD. A demonstration database will be provided, so that the participants can experience themselves how to use the software. In order to make the preserved information accessible again, the E-ARK Database Viewer will be then demonstrated. The E-ARK Database Viewer allows the rapid ingest of SIARD files and provides a web application that allows on-line browsing and search of the database content. It also enables the printing and export of parts of the database into usable formats such as text and spreadsheet (e.g. Word and Excel).

After a break a second preservation tool is presented. SIARD Suite is a freeware reference implementation of the SIARD format developed by the SFA. Apart from database import/export, it also provides a basic viewer that can be used for user-friendly metadata enrichment of the SIARD files. Again, workshop participants are encouraged to try the software themselves on the demonstration database. Additionally, SIARDexcerpt, an open source viewer developed by the Swiss Coordination Agency for the Preservation of Electronic Files (KOST-CECO) is demonstrated. SIARDexcerpt allows users to search and select data directly from a standardised SIARD-file with a simple GUI and some configuration options. Neither a re-import into a database nor special know-how are necessary. The selected data can be represented in a human-readable form.

The workshop concludes with an open discussion and Q&A.

---

[1] e.g. Oracle, MySQL, Microsoft SQL Server
[2] https://www.bar.admin.ch/bar/en/home/archiving/tools/siard-suite.html
[3] http://www.database-preservation.com
[4] https://github.com/keeps/dbviewer
[5] https://github.com/Chlara/SIARDexcerpt

Table 1: Workshop timetable

Topic and duration	Presenter
Why preserve databases? Preservation strategies and problems (10 min)	Kuldar Aas
SIARD format 2.0 (20 min)	Marcel Büchler
Live demo and hands-on: Database Preservation Toolkit and E-ARK Database Viewer (60 min)	Luis Faria
Live demo and hands-on: Database Preservation Toolkit and E-ARK Database Viewer (60 min)	Luis Faria
Live demo and hands-on: SIARD Suite and SIARDexcerpt (60 min)	Marcel Büchler
Discussion and Q&A (30 min)	Luis Faria

See Table 1 for details on topics and timetable.

3. LEARNING OUTCOMES

At the end of the workshop the participants will have learned:

- The main problems and strategies to preserve relational databases

- Details on the SIARD relational database preservation format version 2.0

- How to use current state-of-the-art tools to preserve relational databases by harvesting them into SIARD format 2.0

- Some possibilities for how the preserved databases can be accessed, viewed and used

4. AUDIENCE AND REQUIREMENTS

This workshop targets digital preservation practitioners with some technical abilities and with interest in relational database preservation. Although not required, it is recommended to bring a laptop to test tools hands-on. The tools require Java version 7 or earlier to work. Access to demonstration databases will require wireless network access to be provided on-site.

5. PRESENTERS

Below are short biographies of the main presenters of this workshop:

Kuldar Aas

Deputy director of the Digital Archives of the National Archives of Estonia, working at the archives since 2002. Kuldar has participated in developing a set of national standards, requirements and guidelines in the areas of records management, metadata, transfer of electronic information to long-term archives, description and preservation of relational databases and Linked Open Data. He has also taken part in developing the Estonian proactive digital preservation and reuse environment at NAE. He is representing the national archives in Estonian semantic interoperability and Linked Data task forces.

Marcel Büchler

After finishing his studies in Computer Science, Marcel worked as a Database Engineer for a software development company. In 2015 he joined the Swiss Federal Archives where he is responsible for the SIARD format and the SIARD Suite.

Luis Faria

Innovation Director at KEEP SOLUTIONS, Portugal, and involved in European research projects focused on digital preservation, in particular SCAPE, 4C, E-ARK and VeraPDF (PREFORMA). Luis is taking his Ph.D. in digital preservation at the University of Minho. He qualified as a Systems and Informatics Engineer from the University of Minho. Luis was part of the original development team of RODA (Repository of Authentic Digital Objects) and has engaged in R&D tasks dedicated to systems design, platform development, format migration services and database preservation.

6. ACKNOWLEDGEMENTS