

Tutorial on Relational Database Preservation: Operational Issues and Use Cases

Krystyna W. Ohnesorge

Swiss Federal Archives
Archivstrasse 24
3003 Bern, Switzerland
Tel. +41 58 464 58 27
krystyna.ohnesorge@bar.
admin.ch

Marcel Büchler

Swiss Federal Archives
Archivstrasse 24
3003 Bern, Switzerland
Tel. +41 58 469 77 32
marcel.buechler@bar.
admin.ch

Anders Bo Nielsen

Danish National Archives
Rigsdagsgaarden 9
DK-1218 Copenhagen K
+45 41 71 72 35
abn@sa.dk

Phillip Mike Tømmerholt

Danish National Archives
Rigsdagsgaarden 9
DK-1218 Copenhagen K
+45 41 71 72 20
pmt@sa.dk

Zoltan Lux

National Archives of Hungary
Bécsi kapu tér 2-4
1250 Budapest, Hungary
+36 1 437 0660
lux.zoltan@mnl.gov.hu

Janet Delve

University of Brighton
CRD, Grand Parade
Brighton, BN2 0JY, UK
+44 1273 641820
J.Delve@Brighton.ac.uk

Kuldar Aas

National Archives of Estonia
J. Liivi 4
50409 Tartu, Estonia
+372 7387 543
kuldar.aas@ra.ee

ABSTRACT

This 3-hour tutorial focuses on the practical problems in ingesting, preserving and reusing content maintained in relational databases. The tutorial is based on practical experiences from European national archives and provides an outlook into the future of database preservation based on the work undertaken in collaboration by the EC-funded E-ARK project¹ and the Swiss Federal Archives².

This tutorial relates closely to the workshop: “Relational database preservation standards and tools” which provides hands-on experience on the SIARD database preservation format and appropriate software tools.

Keywords

Relational database, database preservation, SIARD2, E-ARK

1. INTRODUCTION

With the introduction of paperless offices, more information than ever is being created and managed in digital form, often within information systems which internally rely on database management platforms and store the information in a structured and relational form.

Preserving relational databases and providing long-term access to these is certainly a complex task. Database preservation methods cannot concentrate only on the preservation of the data

itself, but must also address multiple administrative and technical issues to provide a complete solution, satisfying both data providers and users.

Technically the *de facto* standard for preserving relational databases is the open SIARD format. The format was developed in 2007 by the Swiss Federal Archives (SFA) and has since then been actively used in the Swiss Federal Administration and also internationally. However, the administrative and procedural regimes, as well as the practical implementation of the standard, can vary to a large degree due to local legislation, best-practices and end-user needs.

2. OUTLINE

The tutorial starts with an overview of the problems and challenges in preserving relational databases. Following this introduction we present three national use cases (Switzerland, Denmark and Hungary) which highlight specific national aspects and best-practices. Finally, the tutorial will present database de-normalization and data mining techniques which are being currently researched and applied within the E-ARK Project.

Throughout the tutorial participants will have the possibility to contribute to an open discussion on the issues and solutions in relational database preservation.

Table 1: Tutorial overview

Topic and duration	Presenter
Introduction to preserving relational databases (30 min)	Kuldar Aas
National Use Case: Switzerland (30 min)	Marcel Büchler, Krystyna Ohnesorge

¹ <http://www.eark-project.eu>

² <http://www.bar.admin.ch>

National Use Case: Denmark (30 min)	Anders Bo Nielsen, Alex Thirifays
National Use Case: Hungary (30 min)	Zoltan Lux
Advanced Database Archiving Scenarios (30 min)	Janet Delve
Open discussion (30 min)	

2.1 Introduction to Preserving Relational Databases

This presentation sets the scene for the rest of the tutorial by highlighting some of the main issues in database preservation:

- How to convince data holders about the importance of database preservation?
- Which preservation method to use (i.e. migration vs emulation)?
- How to minimize the amount of work needed to be undertaken by system developers, data owners and archives especially during pre-ingest and ingest?
- How to ensure that the archiving process does not hinder the original data owner in providing their own services?
- How to ensure that the archived database is appropriately accessible to the users?

2.2 National Use Cases

2.2.1 Swiss Federal Archives

The Swiss Federal Archives (SFA) have been actively dealing with preserving relational databases for more than a decade. Most notably, they took charge in the early 2000's to develop the original SIARD format and the accompanying SIARD Suite software tools.

In the Swiss Federal Administration, there is a relatively large amount of freedom and flexibility in the DB design, therefore a variety of DB-models is encountered. In this use case SFA will focus on the coordination with data owners and explain the steps which need to be made in applying the SIARD approach in agencies.

2.2.2 Danish National Archives

The Danish National Archives (DNA) already started archiving government databases in the 1970's and has by now archived more than 4,000 databases. Over the decades they have gathered widespread experience in both administrative and technical issues.

In this use case DNA will focus on the most common issues which they have encountered while implementing their database preservation regime across the Danish public sector, and will provide an overview of the most common issues encountered during the ingest processes.

2.2.3 Hungarian National Archives

The use case of the National Archives of Hungary (NAH) will concentrate on the access and use of preserved databases. The specific problem addressed is the access to databases which originally have a complex data structure and are therefore impossible to be reused without in-depth knowledge about data modelling and structures.

The use case will demonstrate how to simplify archival access by de-normalizing the original database during ingest, and reusing it with the help of the Oracle Business Intelligence (Oracle BI)³ platform and APEX⁴ application.

2.3 Advanced Database Archiving Scenarios

As the last presentation the tutorial features an overview of the work undertaken in the E-ARK project on advanced database archiving and reuse.

More explicitly the presentation will cover the solutions for creating de-normalized AIPs ready for dimensional analysis via Online Analytical Processing (OLAP) and other tools. Essentially this is adding value to the AIPs by making them more discovery-ready, allowing complex analysis to be carried out on data within an archive, or even across several archives from different countries.

The main aim of the presentation is to discuss whether such an approach is possible to be automated to an extent that it would be applicable in preservation institutions with limited technical knowledge about relevant tools and methods.

3. AUDIENCE AND OUTCOMES

The tutorial targets equally preservation managers and specialists who need to establish routines for ingesting, preserving and reusing relational databases.

Participants will gain a broad overview of the problems and solutions which have been set up across Europe.

4. ACKNOWLEDGEMENTS

Part of the tutorial is presenting on work which has been undertaken within E-ARK, an EC-funded pilot action project in the Competitiveness and Innovation Programme 2007-2013, Grant Agreement no. 620998 under the Policy Support Programme.

³ <https://www.oracle.com/solutions/business-analytics/business-intelligence/index.html>

⁴ <https://apex.oracle.com/en/>