"Implementar el kit de Preservación de datos: una experiencia real desde el lado del usuario,\"
Implementation of the Database Preservation Toolkit: a real experience from the user's side

Zoltán Lux
National Archives of Hungary
lux.zoltan@mnl.gov.hu
http://www.mnl.gov.hu
Topics

• Why are relational databases (RDBMS) so important?
• The record management system of the Hungarian Prosecution Oficces - The first Hungarian Ingest of Relational Database
• Preservation of Relational Databases (Archival Process)
• Access of Relational Databases (Access Process)
• Consequences
Why are Relational Databases so Important?

Workshop of tools and experiences from European Union in electronic archiving and digital preservation.
4-5 October 2018, Cordoba General Archive
The first Hungarian Ingest of RDBMS

- The record management system of the Hungarian Prosecution Offices
- 2010-2011. The Wish.
- What does it consist of, how is data stored in it?
  - Application software written in Clipper 5
  - Data in dBASE IV files
- Documentation?

Workshop of tools and experiences from European Union in electronic archiving and digital preservation.
4-5 October 2018, Cordoba General Archive
The first Hungarian Ingest of RDBMS - Preservation

- In form of Information Packages (OAIS Model)
- RDBMS into SIARD file
  - But how? dBase couldn’t be exported into SIARD directly
  - Export into CSV
  - Load into live (Oracle) RDBMS
  - SIARD export
- Documentation
  - Created during the ingest process
  - Screenshots
  - Most common use cases
The First Hungarian Ingest of RDBMS – Preservation (The data)

Workshop of tools and experiences from European Union in electronic archiving and digital preservation.
4-5 October 2018, Cordoba General Archive
The first Hungarian Ingest of RDBMS - Preservation

Workshop of tools and experiences from European Union in electronic archiving and digital preservation. 4-5 October 2018, Cordoba General Archive
The first Hungarian Ingest of RDBMS - Access

• Load SIARD package into a live RDBMS
• Developed a reporting application in Oracle APEX

Workshop of tools and experiences from European Union in electronic archiving and digital preservation. 4-5 October 2018, Cordoba General Archive
The First Hungarian Ingest of RDBMS – Open Questions

• The understandability of the archived database
• Restructuring the original database in order to make it more transparent and searchable. The structure of the database in production phase is not optimal for search
• Application of other tools and methods both for archiving and access
• E-ARK project 2014-2017

Workshop of tools and experiences from European Union in electronic archiving and digital preservation. 4-5 October 2018, Cordoba General Archive
The Archival Process

Workshop of tools and experiences from European Union in electronic archiving and digital preservation. 4-5 October 2018, Cordoba General Archive
The Archival Process

• Documentation of the database objects (tables, fields, etc.) - Who and when? It can be checked!

Workshop of tools and experiences from European Union in electronic archiving and digital preservation.
4-5 October 2018, Cordoba General Archive
The Archival Process

- Database transformation
- The structure of the archived database must be understood – once
- Several manifestations (of the original database) within the AIP
The Archival Process – The Transformed Relational Model

Workshop of tools and experiences from European Union in electronic archiving and digital preservation.
4-5 October 2018, Cordoba General Archive
The Archival Process - DBPTK

- **General usage**: java [properties] -jar dbptk-app-x.y.z.jar [plugin]
  <importModule> [import module options] <exportModule> [export module options]
The Archival Process - DBPTK

**General usage:** java [properties] -jar dbptk-app-x.y.z.jar [plugin] <importModule> [import module options] <exportModule> [export module options]

- java is the java command, the full path may also be used
- [properties] may be omitted or replaced with special configurations that influence the conversion (more details)
- -jar dbptk-app-x.y.z.jar tells java to execute the dbptk-app-x.y.z.jar file (the file name must be adjusted to match the one you have)
- [plugin] is optional, and should be replaced with plugin configurations (if any)
- <importModule> should be replaced with the import module specification, e.g. -i mysql or --import=postgresql
- <exportModule> should be replaced with the export module specification, e.g. -e mysql or --export=postgresql
- [import module options] should be replaced with parameters to specify the behavior of the import module, e.g. --import-username=username --import-password="p4ssw0rd" (to specify source database username and password)
- [export module options] should be replaced with parameters to specify the behavior of the export module, e.g. --export-file=filename.siard --export-compress --export-pretty-xml (to specify the SIARD-2 export module behavior)

Oracle, MySQL, PostgreSQL, Microsoft SQL Server, Microsoft Access (only export) and JDBC (SIARD 1, 2, DK)
The Archival Process - DBPTK

- SIARD Package - Header

Workshop of tools and experiences from European Union in electronic archiving and digital preservation.
4-5 October 2018, Cordoba General Archive
The Archival Process - DBPTK

• SIARD Package - Content

Workshop of tools and experiences from European Union in electronic archiving and digital preservation.
4-5 October 2018, Cordoba General Archive
The Access Process

- Load SIARD Package into a live RDBMS by DBPTK

Workshop of tools and experiences from European Union in electronic archiving and digital preservation.
4-5 October 2018, Cordoba General Archive
The Access Process – Tools available

- DBVTK (Keeps Solutions)
- SOFIA MSSql based software (Danish National Archives)
- SIARD Suite (BAR)
- SQL Developer, SQL Navigator, etc. (Business software)
- Oracle APEX (Framework)
- DW, BI, etc.
The Access Process

Workshop of tools and experiences from European Union in electronic archiving and digital preservation.
4-5 October 2018, Cordoba General Archive
The Access Process

Workshop of tools and experiences from European Union in electronic archiving and digital preservation.
4-5 October 2018, Cordoba General Archive
The Access Process

Workshop of tools and experiences from European Union in electronic archiving and digital preservation.
4-5 October 2018, Cordoba General Archive
The Access Process

Workshop of tools and experiences from European Union in electronic archiving and digital preservation.
4-5 October 2018, Cordoba General Archive
Consequences

• Understanding what relational database is
• Driven by technology -> new challenges in archival and access processes
  – New objects that need to be archived
  – New tools and methods that support the use (access)
Thank you for your attention