



DB archiving and use at Czech authorities

Martin Rechterik (National Archives Czech Republic, Prague)

Martin Rechterik is historian and archivist and has been employed in the National Archives of the Czech Republic for the last four years. He is member of the digital archives methodological team and focuses on database preservation. He has almost 10 years of experience as record management specialist in private sector too.

Rechterik's presentation had three parts – theory, sharing a practical experience and then looking at how to present data to users. On theory, he asked four questions: What does it mean to preserve a database? Which database do we choose to be preserved? How do we preserve it and at what stage (sometimes periodically across DB lifetime)? How do we present the data, depending on the legal question of who is allowed to access it (open or closed data)?

He then applied these questions (cf. figure 34: What do we need to preserve a DB?). What output is necessary for the future? Do we need queries/views? What is of real value and importance for the future? The presentation of databases is important. Can we present SIARD files? Rechterik saw a need to select the right tool for presentation.

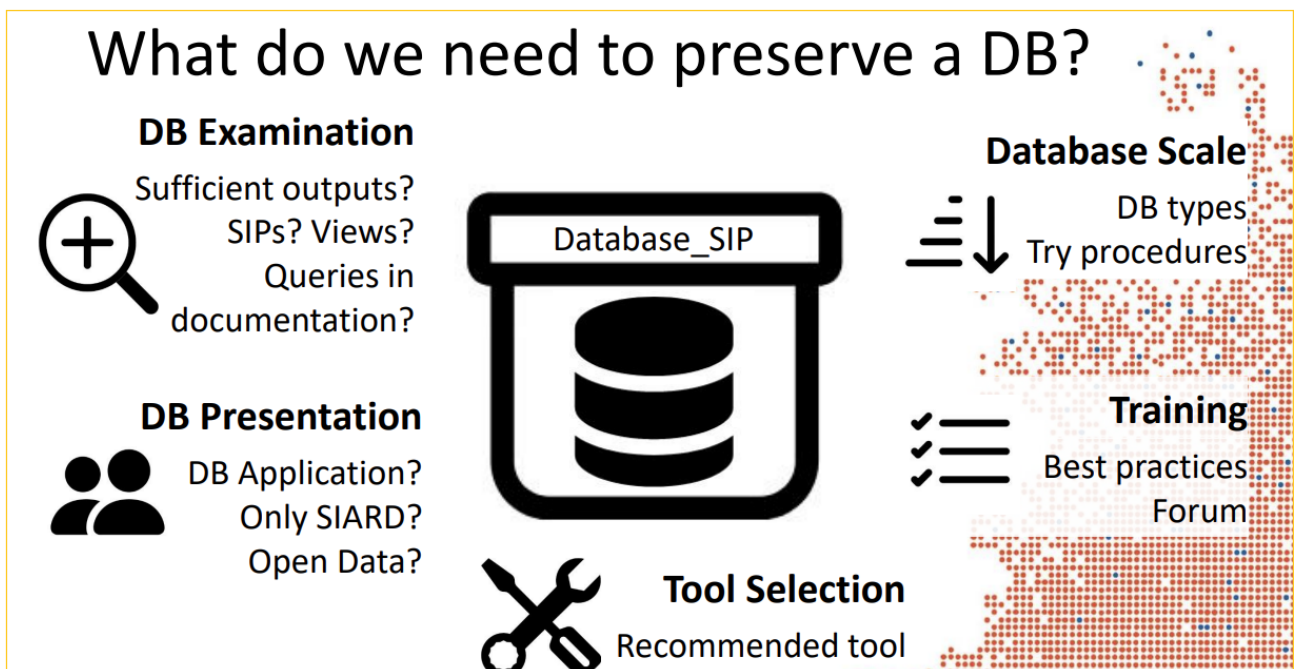


Figure 34: What do we need to preserve a DB?



Archives will have to do this at scale and to cover different types of databases – at least those in use in the public sector. They must also share this knowledge with colleagues across borders. The European Union wants member states to establish databases with similar outputs to be sent to EU authorities – this should be seen as an incentive to member states to cooperate on solving those problems.

An example is the ARIS (Automatic Budget Information System) of the Czech central government. It contains accounting and financial data of all state institutions (from the presidential office, ministries, municipalities to kindergartens) and covers the years 1997-2009. ARIS and its sibling systems RARIS and ARISweb were intended as paperless – an all-digital reference system. The contents are mostly open data. The Czech National Archives wanted to emulate the whole system because of the ARISweb interface, which was used for public access, provided a lot of functionality. This plan turned out as too optimistic. The point of failure was licences for the Informix DBMS that were too costly. The National Archives resorted to a SIARD normalisation approach but also transferred CSV files for easier access. The Ministry of Finance first provided the database files as PostgreSQL dumps, while the company Keep LDA was engaged to help with the SIARD conversion. ARIS/ARISweb, an interface showing all financial data sheets, was shut down in September 2021.

Rechtorik also explained the way in which the ARIS data is now used at the National Archives. To present data in their reading room, the National Archives use the dbDIPview tool ([p. 47](#)). ARIS, ARISweb, and RARIS are now consultable through the following services:

- one part of the data is available as CSV exports divided into seven browsing packages
- the other part, a SIARD representation of the RARIS module was exported using the table filter feature from original SIARD file.



Figure 35: Service selection screen at the digital reading room at NACR



Popis zobrazení: 1-RARIS - Registr organizací systému ARIS

RARIS
Registr organizací systému ARIS - RARIS vznikl pro potřebu řízení zpracování dat v Automatizovaném rozpočtovém informačním systému ARIS. Registr RARIS obsahuje množinu informací potřebných pro zpracování dat pro všechny organizace, které mají povinnost předávat data do informačního systému ARIS. Tuto povinnost ukládá Vyhláška Ministerstva financí č. 16/2001 Sb. ze dne 22. prosince 2000. RARIS v eBadatelně lze využít zejména k vyhledání IČO organizace, které je potřebné pro badání v datech informačního systému ARIS

Objekt: MŠ Opatovice nad Labem
Název: Mateřská škola Opatovice nad Labem,
IČO: 70986436
Sídlo: Opatovice nad Labem
Ulice: Ke Hřišti 163
PSČ: 53345
Okres: PARDUBICE
Typ: organizace zřízené obecním úřadem
Druh: příspěvková organizace (PO)
Řízení: Místně řízené organizace
Výkon státní správy: organizace nevykonává státní správu
Zdroj financí: školy a školská zařízení financovaná ze státního rozpočtu (zřizovatel ÚSC)
Funkce: Preprimární vzdělávání
Ekonomická činnost: Předškolní vzdělávání
Odvětví: Předškolní výchova
Platnost údajů k datu: 2005-03-30
Změna: aktivní organizace nebo hist.výskyt, po němž existuje výskyt s vyšším DATZ
Způsob odměňování: PLAT podle části 6, hlava III. ZP a z.č. 143/1992 Sb.



Figure 36: RARIS database record of a kindergarten in its archival rendition at NACR (the picture being an illustration only)

The packages are enhanced with several simple tables to better understand the meaning of the data. RARIS can be used to find the necessary identification numbers of reports. The dbDIPview tool can handle CLOB (variable-length character large object string) or BLOB data easily regardless of whether they are inside or outside the SIARD file.

Rechterik's talk was about a proof of concept for standardising the DB transfer process. The NACR has also accepted other DB systems in CSV and ACCDB formats, but now wants to deploy preservation and presentation of databases in a scalable and standardised way. An eminent goal is to provide means of periodic comparison of certain datasheets and variables over time.

To conclude, Rechterik pointed out some doubts. DB preservation needs a plan for future generations. Will a description of columns be enough in 20 years? Will government agencies be satisfied with a SIARD file without a working environment? He was skeptical, seeing SIARD as a great standard but feeling that it must be extended to fit into more diverse use cases. He saw the necessity to share experiences with creation of SIARD and presentation tools for SIARD packages and archived DBs in general.

Questions and discussion

Kai Naumann named an approach that had not been mentioned during the workshop: to hop from database instance to instance over decades or longer. If an Informix environment is outdated, a service could move the data to Oracle, for which a new working environment is to be deployed. From there, two decades later, the data might travel on to MariaDB and another user interface. If there were mechanisms to prove that the jump from, e.g., Oracle to MariaDB was controlled and that nothing significant was lost in the transition, the gap of 60 years could be bridged, too.