

Archivace relačních databází

Možnosti, formát SIARD,
nástroje, tvorba, prohlížení, datové výstupy

Martin Rehtorik



Archivace relačních databází

1. Možnosti archivace relačních databází

2. Formát SIARD

- » Specifikace, verze a použití
- » Praktické fungování a možné využití
- » Aplikace pro práci s archivem SIARD
- » Datové výstupy

3. Výhody, nevýhody, shrnutí



Možnosti archivace relačních databází

» Chronos

» XML

› DeepArc

› SIARD



The image shows two screenshots. The top one is the 'database toolkit preservation' website, featuring a large white 'db' logo on a dark green background. Below the logo are two buttons: 'View on GitHub' and 'Download v2-beta5'. The bottom screenshot is the SIARD Suite 2.1.68 software page. It features the Swiss flag logo and text in four languages: German (Schweizerische Eidgenossenschaft), French (Confédération suisse), Italian (Confederazione Svizzera), and Romansh (Confederaziun svizra). Below this, it lists the responsible departments in four languages: Eidgenössisches Departement des Innern EDI, Département fédéral de l'intérieur DFI, Dipartimento federale dell'interno DFI, and Schweizerisches Bundesarchiv BAR. It also lists the archival codes: Archives fédérales suisses AFS and Archivio federale svizzero AFS. To the right of the text is an illustration of a cardboard box with three colored cylinders (yellow, pink, blue) inside, and two more cylinders (yellow and blue) outside.

database toolkit preservation

View on GitHub Download v2-beta5

 Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

SIARD Suite 2.1.68

Eidgenössisches Departement des Innern EDI
Département fédéral de l'intérieur DFI
Dipartimento federale dell'interno DFI
Schweizerisches Bundesarchiv BAR
Archives fédérales suisses AFS
Archivio federale svizzero AFS



Formát SIARD

- » Vznik na základě požadavku na programově nezávislé archivování relačních databází
- » SIARD (Software Independent Archiving of Relational Databases) – otevřený formát pro dlouhodobé archivování relačních databází
- » Cílem je trvale uložit data uložená v databázi
- » Verze SIARD 1.0, SIARD DK, SIARD 2.0, SIARD 2.1



Verze formátu SIARD

SIARD 1.0

- » neoperační časový snímek databáze
- » Metadata vyžadovaná standardem SIARD slouží především k popisu originálního řídicího databázového systému, k popisu relací, tabulek, pohledů atd.
- » Primární data relační databáze (tabulky) jsou uložena v datové části archivního souboru

SIARD DK

- » Je založen na pevné adresářové struktuře
- » Metadata jsou rozložena do 4 popisných souborů a to: tableIndex, archiveIndex, fileIndex, docIndex
- » Umožňuje archivovat velké objekty mimo tabulková data



Verze formátu SIARD

SIARD 2.0

- » Oproti SIARD 1.0 je rozšířen o možnost ukládat velké datové objekty uložené v databázi do samostatných souborů, a to uvnitř i vně archivačního souboru siard.
- » Struktura obsahuje všechna metadata uložená v jediném souboru metadata.xml, implementace standardu SQL:2008

SIARD 2.1

- » Zavádí nové atributy, které omezují chybovost konverze SQL jazyka do podoby XML popisu



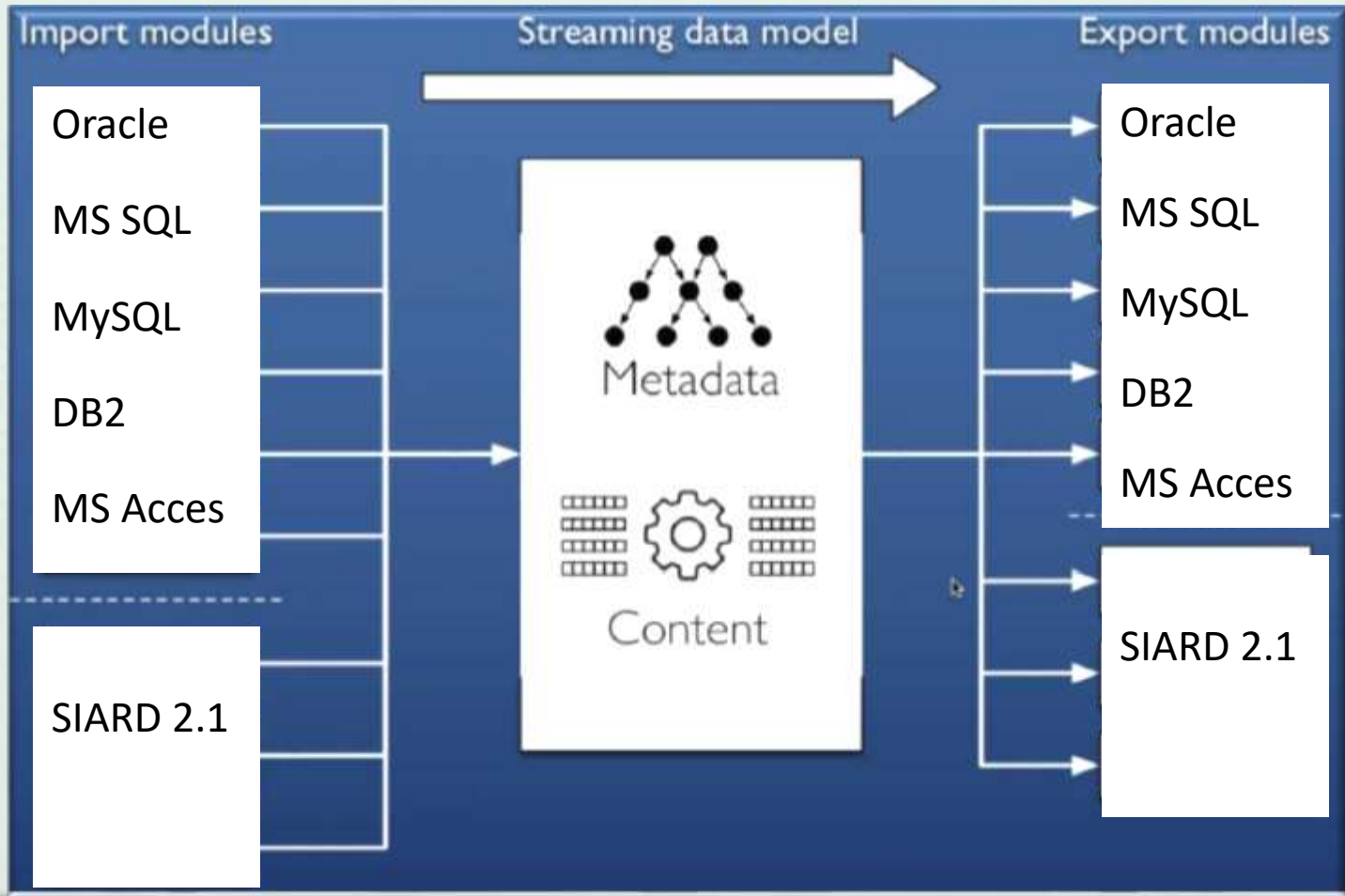
Použití formátu SIARD

- » Tvorba archivů je možná přes příkazovou řádku nebo v příslušné grafické podobě aplikace
- » Aplikace Database Preservation Toolkit a umožňuje archivaci do formátu SIARD V1.0, V2.0 a SIARD DK (uživatelská grafická podoba, tzv GUI je dostupná jen pro operační systémy Linux nebo OS X)
- » Pro SIARD 2.1 vyvinut (stále ve vývoji) SIARD Suite, uživatelsky příjemná aplikace s grafickou podobou
- » Pro databáze většího datového rozsahu je doporučeno používat příkazovou řádku



Využití formátu SIARD

Schéma fungování a využití



Nástroje prohlížení archivu SIARD

SIARD 1.0

- »SIARD suite, SIARD Excerpt

SIARD 2.0

- »Database Visualization Toolkit
- »SIARD Excerpt
- »Předpokládá se přenos do živého databázového prostředí
- » SIARD DK
 - »SOFIA – „search and find in archives“, nelze získat
- » SIARD 2.1
 - »SIARD suite, SIARD Excerpt
 - »Předpokládá se přenos do živého databázového prostředí



Datové výstupy

Archiv SIARD = hierarchicky strukturovaný systém dat v souborech XML s příslušným schématem XSD, dále obsahuje soubory velkých datových objektů, veškerá metadata se ukládají do jediného souboru metadata.xml (neplatí do SIARD DK)

Výstupem může být:

- » celý archiv databáze
- » tabulka
- » datový záznam
- » datový objekt
- » metadata



Výhody formátu SIARD

- » Archiv SIARD = neoperační časový snímek databáze
- » data jsou ukládána ve stejné struktuře jako v jiných zemích, formát SIARD se používá po celém světě
- » Dokumentace k formátu a aplikaci bude dostupná v českém jazyce
- » Komunita, která se podílí na dalším vývoji (GitHub)
- » Archiv SIARD = zip, lze „procházet“ i bez aplikace
- » V podstatě jediná možnost trvalé archivace dat z informačních systémů bez datového výstupu
- » Archivace do strukturované sestavy souborů XML je v souladu s vyhláškou č. 259/2012 sb. o spisové službě, XML je výstupní formátem dle §23 odst. (6) a (7)



Nevýhody formátu SIARD

- » Archiv SIARD = neoperační časový snímek databáze
- » IMPORT a EXPORT databázi závisí na kvalitě a dostupnosti ovladačů k uvedenému systému a vzhledem ke značným rozdílům mezi jednotlivými databázovými systémy (implementace standardu jazyka SQL, dotazy v proprietární podobě), není běžně možné migrovat data a popisy mezi jednotlivými databázovými systémy.
- » Vývoj aplikace pro práci se SIARD archivem zaostává za rychlostí vývoje SQL serverů a standardů pro komunikaci s SQL servery. SQL standard je z roku 2016 a SIARD 2.1 implementuje standard z roku 2008.



Nevýhody formátu SIARD

- » Není zpětná kompatibilita verzí formátu a novějšími aplikacemi s výjimkou aplikace SIARD Excerpt
- » Databáze (IS) a tabulková data jsou normalizována adresářové struktury s jednotlivými tabulkami a jejich definičními schémata, normalizací dojde ke „zmrzačení“ databáze
- » „Procházení“ adresářové struktury se soubory XML a XSD je bez aplikace prakticky nevyužitelné
- » Verze SIARD 1.0 umožňuje ukládat velké objekty pouze v polích tabulky
- » Samostatné uložení velkých objektů snižuje využitelnost aplikace SIARD Excerpt



Závěrečné shrnutí

- » Databáze (IS) jsou z hlediska archivního zákona dokument (nahrazují různé dříve vedené úřední knihy) a není pochyb o nutnosti jejich archivace
- » Formát SIARD prošel vývojem v Evropě, přes SIARD 1.0 (Švýcarsko), SIARD-DK (Dánsko), SIARD 2.0 (Nizozemí, Německo, Portugalsko aj.), nejnověji SIARD 2.1 (Švýcarsko). Z hlediska způsobu ukládání dat je vhodný k archivaci relační databáze na konci životnosti
- » Základní podmínkou je podpora prostředí
- » NAČR na formát SIARD nahlíží jako nouzovou možnost archivace dat databáze, pokud informační systém nemá možnost strojově čitelného výstupu (SIP, xml, csv)



Děkuji za pozornost

martin.rechtorik@nacr.cz

