Landesarchiv Baden-Württemberg

Discussions on
How to archive a DB for **60 years**?

Databases for 2080 – Preserving database content for the long term (October 5th/6th, 2021)

Document the original environment

– Manuals, specifications

– Screenshots of GUI

– Video screencasts of GUI

– Follow standard like Data Documentation Initiative (DDI), PREMIS, emerging E-ARK standard

– Make reviving the whole package a self-explaining „bootstrap" routine

Document the information package

– Make reviving the whole package a self-explaining „bootstrap" routine

– Follow standard like Data Documentation Initiative (DDI), PREMIS, emerging E-ARK CSIP SIARD standard

Option A: Select essential entities

████████

- *Only use existing reporting data*
  - Statistical reports (documents, spreadsheets)
  - Statistical microdata (CSV, other)
  - OLAP extracts (CSV, other)
  - Reporting data (subject-specific XML, JSON, etc.)
  - Geodata extracts (GML, GeoPackage)
- *Create archival reports or modify existing reports*
  - CSV export(s)
  - SQLite
  - SIARD XML
  - other XML, JSON, etc.

Option B: Preserve DBMS performance

– *Make sure you don't miss some parts without notice! (see below)*

– *Options*

- Complete SIARD XML dump

- Complete SQL 2008 dump

- Complete SQLite dump

- Complete proprietary dump (e.g. Oracle)

Option C: Preserve complete performance

– Call for emulation services

   – Emerging packaging standards for OS, DBMS, GUI

   – Specialized companies

   – Maybe use less intricate strategies as a benchmark in parallel?

And from then on …

– Keep package in a persistent, replicated, secure storage

– Keep access methods

– Wait
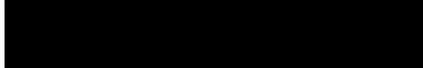
– And monitor regularly

BUT

You want to know the obstacles?

– *Understaffed DB admins*

– *Too little standards / software support*

– *Copyright and data protection legislation (lots of unknown rightholders)*

– *Proprietary extensions*

  – Geodata

  – Binary Large Objects (BLOBs)

  – Character Large Objects (CLOBs)

  – NOSQL technology

– *Unusual size*

  – XML > 200 MB doesn't parse on most computers

  – CSV > 1 million rows doesn't open in MS Excel

– *Floodings, fires, earthquakes, wars, ...*

Discussion rooms

- See GotoMeeting links on agenda on Landesarchiv workshop homepage

- CryptPad will be shared via GotoMeeting

- We choose minute takers who write into the cryptpad

- We answer existing questions, ask new questions, mark contentious issues

- Please follow advice of discussion hosts and don't hop back and forth too much ☺

- Group A Standards, software, deployment with Kuldar Aas (National Archives Estonia) and Carl Wilson (Open Preservation Foundation)

- Group B Strategies, efficiency, documentation with Annette Strauch-Davey (University of Hildesheim) and Kai Naumann (Landesarchiv Baden-Württemberg)

- Final plenary at 12:00 (see also link on agenda)

Landesarchiv Baden-Württemberg

Let's improve!

Dr. Kai Naumann

Archival Policies Department

Landesarchiv Baden-Württemberg

Stuttgart – Germany

kai.naumann@la-bw.de