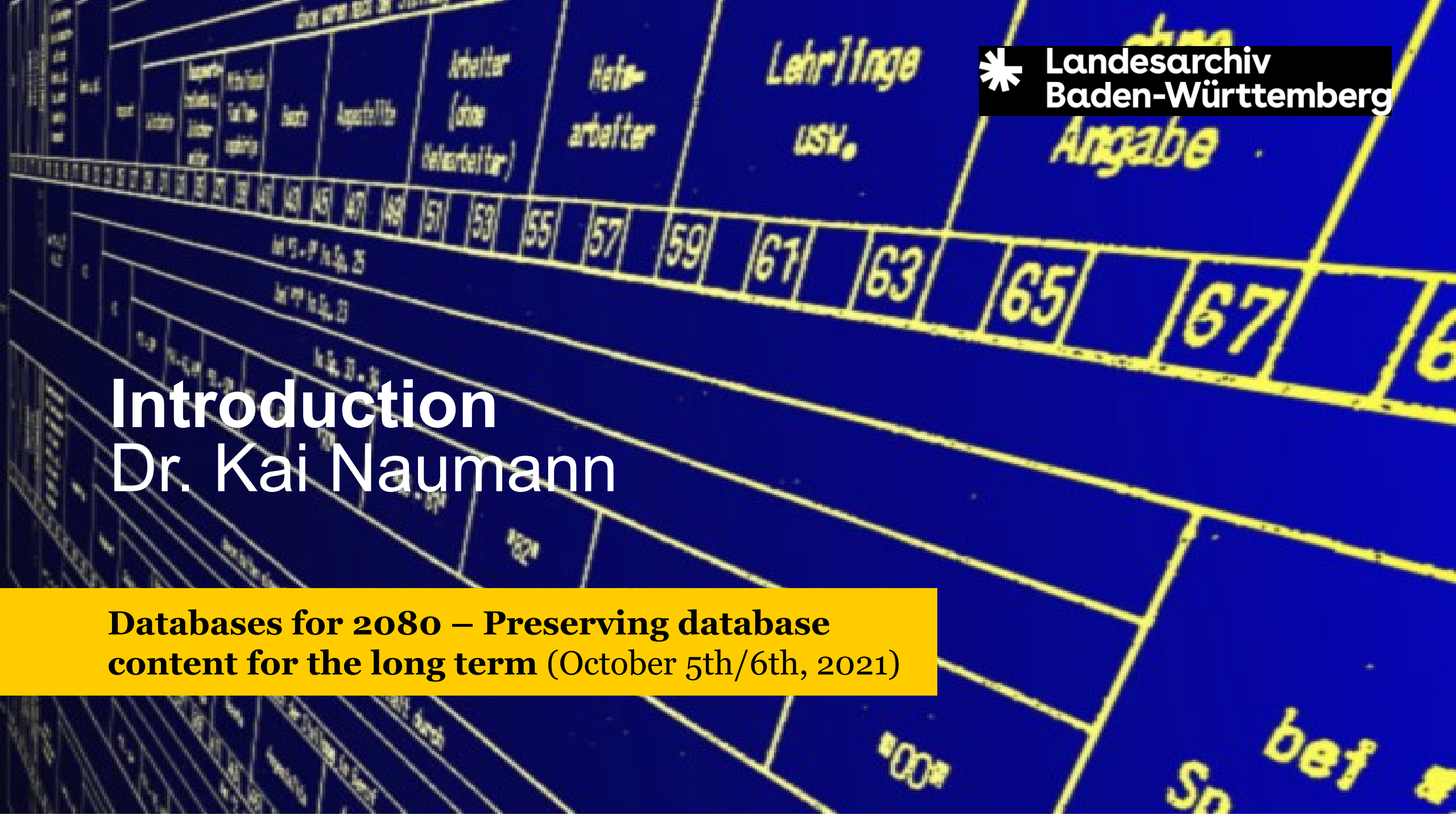



# Introduction

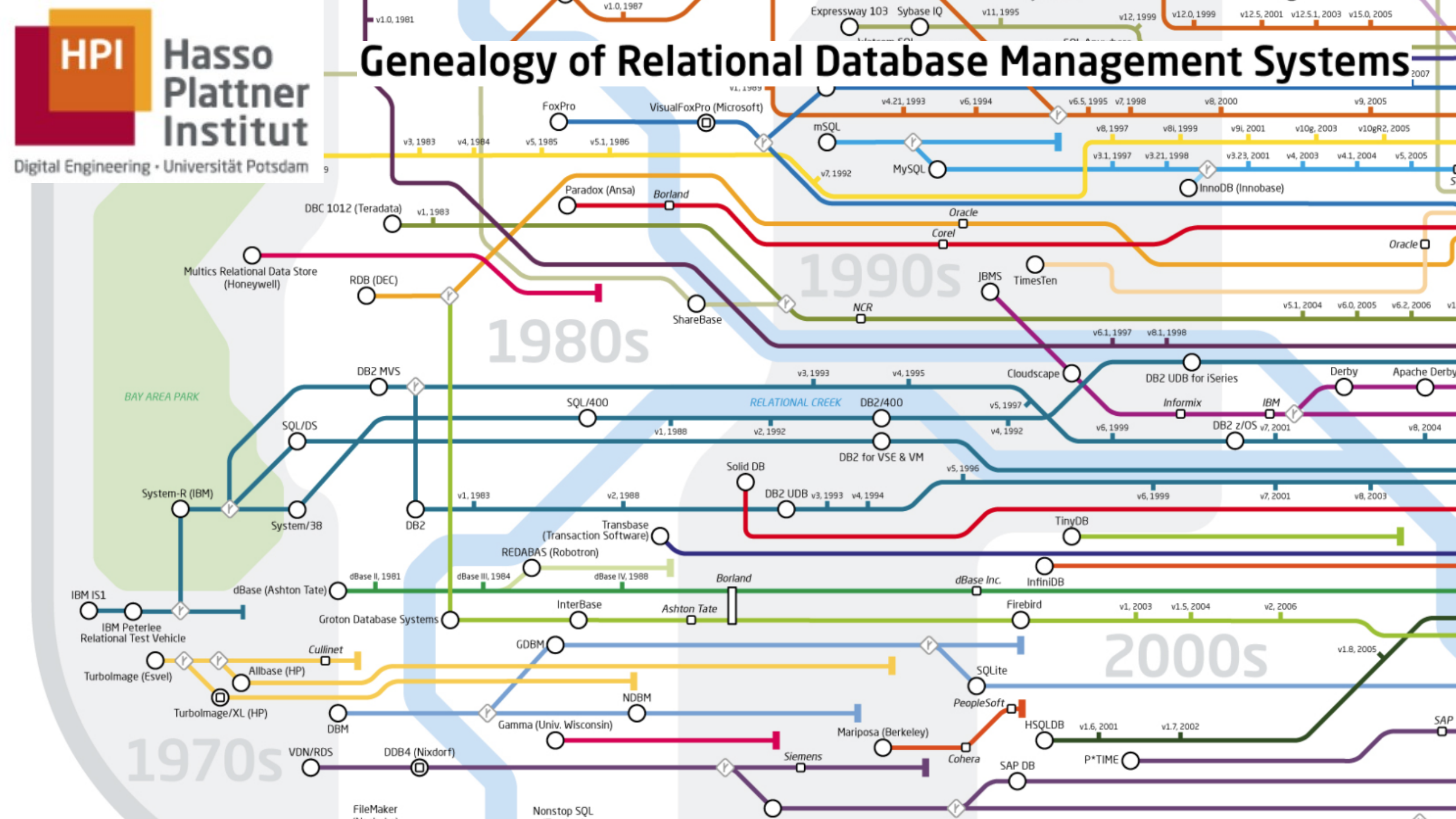
Dr. Kai Naumann

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

# What is a database?

- Graphical user interface layer (e.g. Web Browser, C# GUI library)
- Business logic layer (e.g. PHP, C#)
- Database management system layer (e.g. MariaDB, dBASE)
- Storage layer database (+ storage layer large objects) (e.g. Linux or CD-ROM file system)

# Genealogy of Relational Database Management Systems



# Research done so far (milestones)

[redacted]

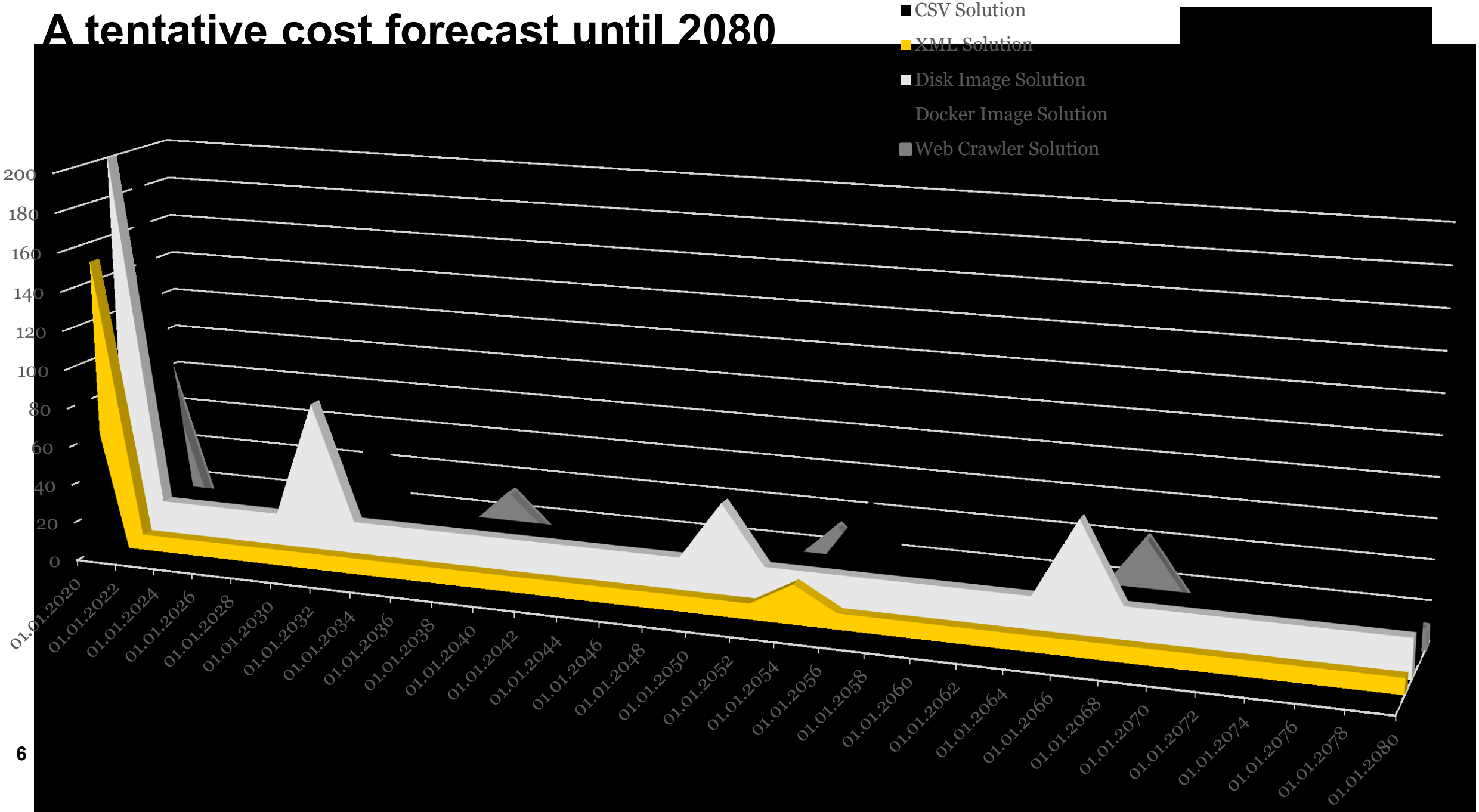
| Year | Event |
|---|---|
| 1983 | • Klopprogge and Lockemann on “Information Preserving Databases: Consequences of the Concept of Time”, in: VLDB '83: Proceedings of the 9th International Conference on Very Large Data Bases, S. 399–416. |
| 2003 | • XML-Archiving of DB dumps (Zeller/Herbst/Kemper, https://dl.gi.de/handle/20.500.12116/30094) |
| 2007 | • SIARD Format and Tool 1.0 released<br>• First International Workshop on Database Preservation (PresDB'07), https://doi.org/10.1145 |
| 2013 | • SIARD Format 1.0 a Swiss national standard (updated to 2.1 in 2018) https://www.ech.ch/de/standards/60647<br>• Proof of concept: database preserved through emulation at Land Survey Authority New Zealand: https://openpreservation.org/blogs/migrating-windows-2000-database-server-virtualized-and-emulated-hardware/ |

# Outcomes of the challenge approach

[redacted]

- Challenge published in April 2020 (Naumann, WADL 2020, http://hdl.handle.net/10919/99569)
- Two responses with clear directions, also literature and some advice given during conferences
  - CSV Solution
  - XML Solution
  - Disk Image Solution
  - Docker Image Solution
  - Web Crawler Solution

# A tentative cost forecast until 2080



# Day 1: Options and Strategies

| | |
|---|---|
| 13:20 | Brigitte Mathiak (GESIS): Sustainability strategies for digital humanities systems |
| | Raja Appuswamy (EURECOM): Storing and reviving databases on DNA |
| 14:20 | Coffee break |
| 14:40 | Florian Hartl (CSP GmbH & Co KG) and Albin Brandl (EfficientNodes GmbH): Database preservation for industry customers |
| | Klaus Rechert (University Freiburg i.Br.): Emulation options for legacy databases |
| | Damir Bulic (Spectral Core Ltd): Database preservation through conversion technology |
| 16:10 | Coffee break |
| 16:30 | Luis Faria (KEEP Solutions, LDA): Preservation through standardisation with DBPTK |
| | Kuldar Aas (National Archives Estonia): E-ARK standardisation efforts for databases |
| 17:30 | Short wrapup |
| 19:30 | Video chat (location to be announced) |

# Day 2: Standardisation, application, discussions

[redacted]

| | |
|---|---|
| 08:30 | James Doig (National Archives Australia): Transfer and Preservation of Databases at the National Archives of Australia: Problems and Directions |
| | Markus Schmalzl (Directorate General of State Archives Bavaria): A tool for the preservation of data from governmental and research databases |
| 09:30 | Coffee break |
| 09:45 | Torbjørn Aasen (Intercommunal Archive Møre og Romsdal): Database ingest and use at public archives with the SIARD format |
| | Martin Rechtorik (National Archives Czech Republic): DB archiving and use at Czech authorities |
| 10:45 | Parallel discussion groups (A) Standards, software, deployment and (B) Strategies, efficiency, documentation, see Cryptpad |
| 11:45 | Plenary with outcomes of parallel groups, discussion of contentious issues, outlook |
| 12:45 | Closing remarks |