Danmarks
Hukommelse

# Implementing DBPTK as the Access Tool for Born Digital Records at the DNA

DLM Forum Members Meeting
Brussels
May 27, 2024

Jan Dalsten
Danish National Archives

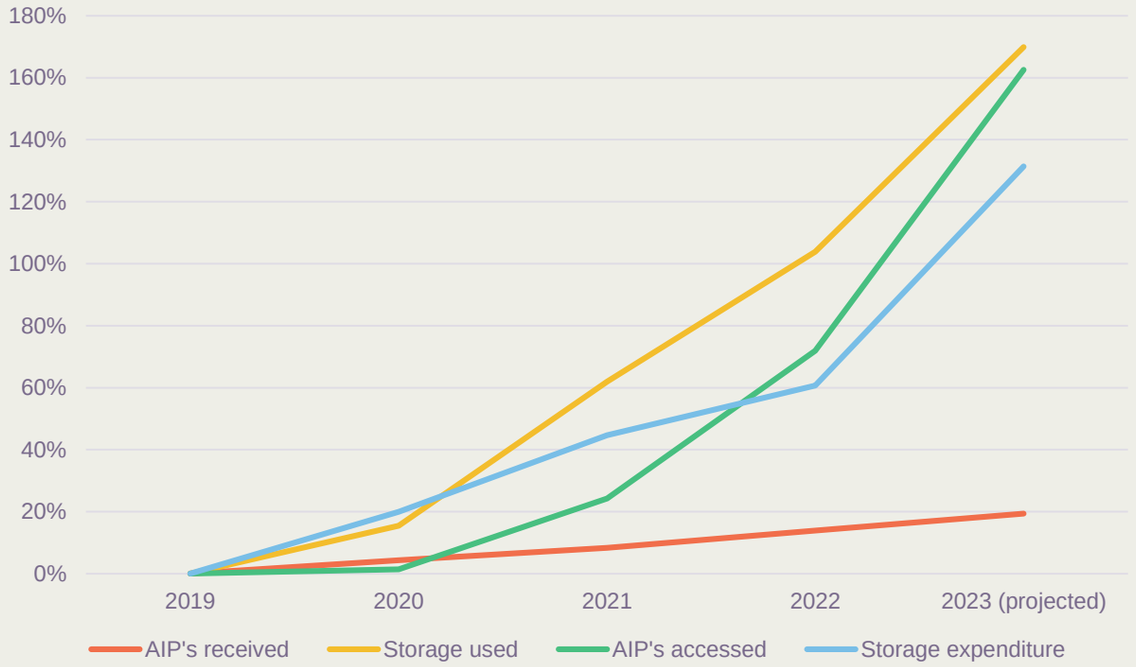
# Brief overview of our collection of born-digital data

- The Danish National Archives have collected digital records from Danish public administration since 1975
- Over 9.000 AIP's in our collection today:
  - 6.000 from (primarily) government agencies and private organizations
  - 3.000 research data sets
- 500 AIP's from government agencies and 2.200 research data sets are now publically available
- Of those approx. 750 are freely available as downloads

- Examples of born-digital data in our collection:
  - Records management systems from all over the public sector
  - Registries from public and private research centres
  - Election surveys from the 1970's and up
  - Gallup/Omnibus surveys from the 1960's and up

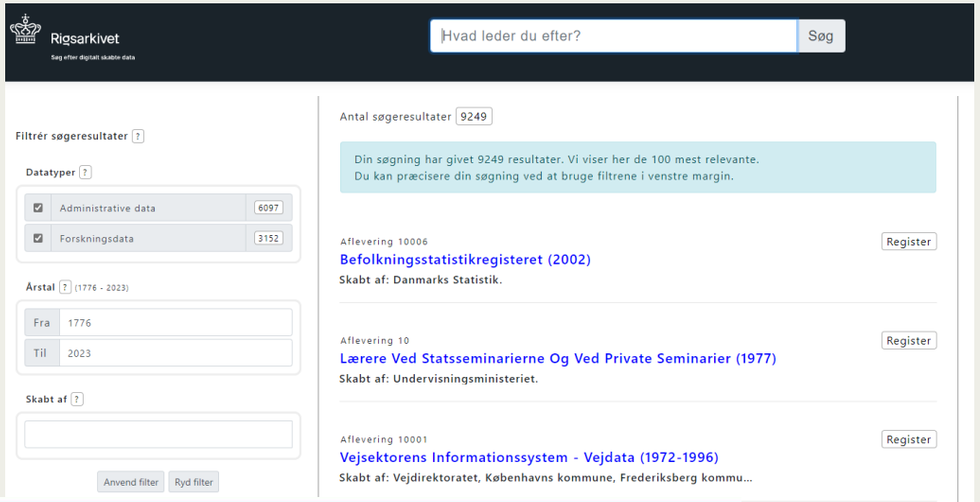
# Development in the use of born-digital data

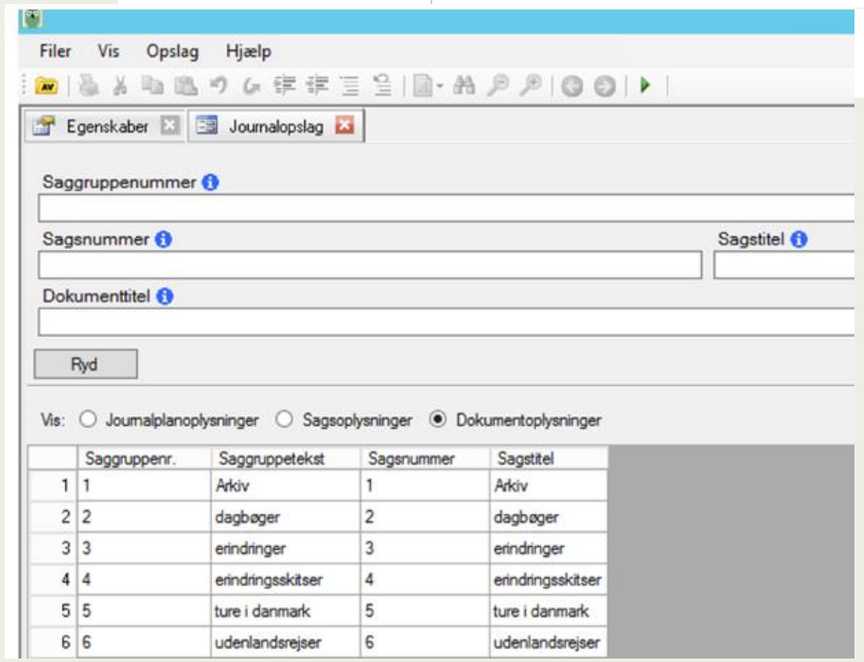
- The demand for access to our collection of born-digital data has historically not been very high
- Three recent developments have changed that:
  - The law of ”Egenacces” (own access)
  - The pension scheme ”Tidlig pension” (early retirement)
  - The agencies’ need for data they no longer access to in their own systems
- General developments:
  - Most of the archival material that we receive today is born digital
  - As our collection gets older, more data will become available to the general public



# Dissemination of born-digital data at the DNA

- Dissemination is a broad term used for all the ways in which we make our data available to our users
- Two main ways:
  1. Publication of metadata and freely available data (as download packages) on digidata.rigsarkivet.dk
  2. Conversion from AIP's to DIP's (relational databases) for use in our search and access system Sofia



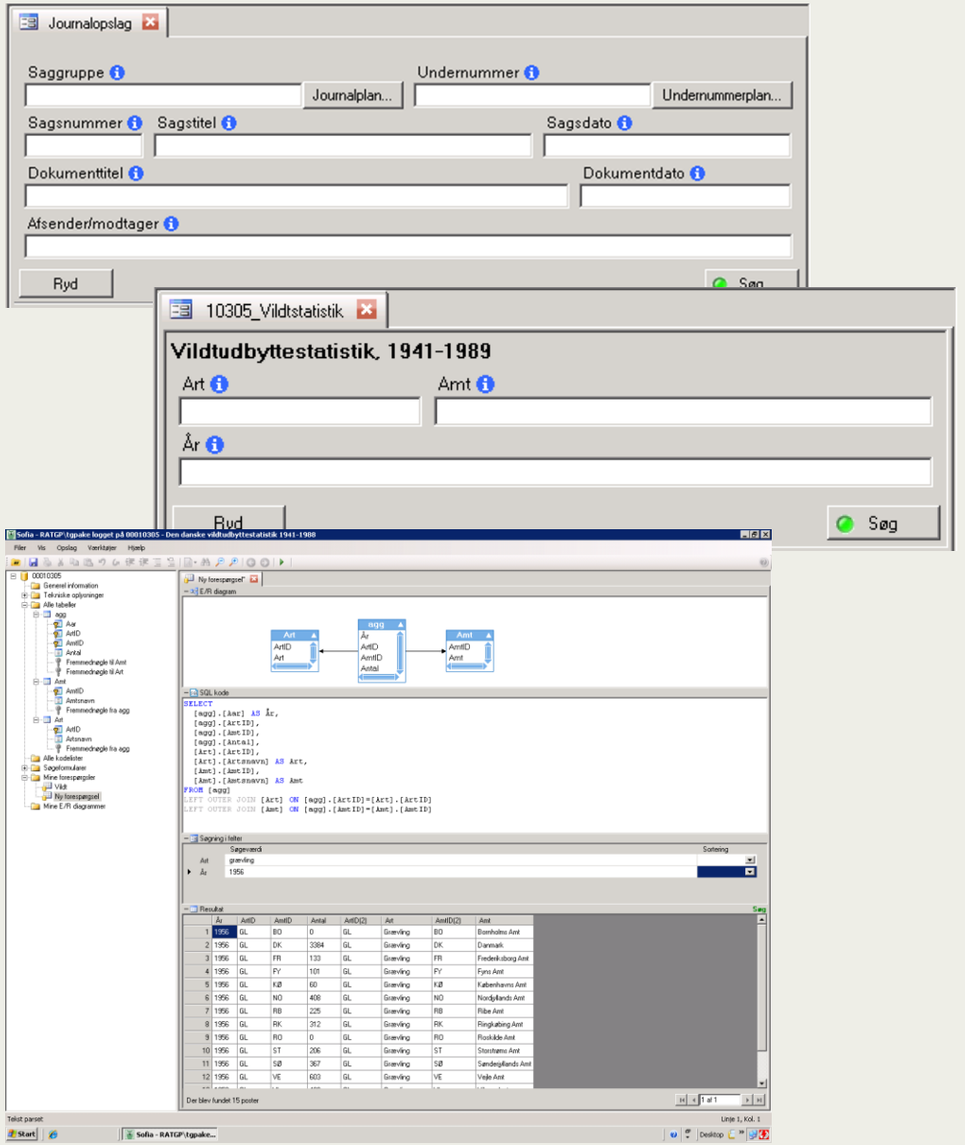
# Sofia

- Sofia is DNA's current access and search system for born-digital data and consists of an user interface and a backend SQL server, where the IPs are loaded by a conversion-tool
- Sofia was developed for accessing IPs submitted according to the regulations from 2005 and later adjusted (more or less) to handle IPs submitted after the subsequent regulations of 2010 and 2020
- The main function is to allow a user to access archived data and its development was defined by two questions:
    1. How do we make one system, that can provide access to many different kinds of systems?
    2. How do we create an access system, that archivists with no technical database knowledge can use? (Usability)
- In order for an user to access data and documents, an IP needs to be "prepared", where an employee connects one of three different predefined views (templates) to the IP and create relations to the documents stored on external servers outside of Sofia

# Searching in Sofia

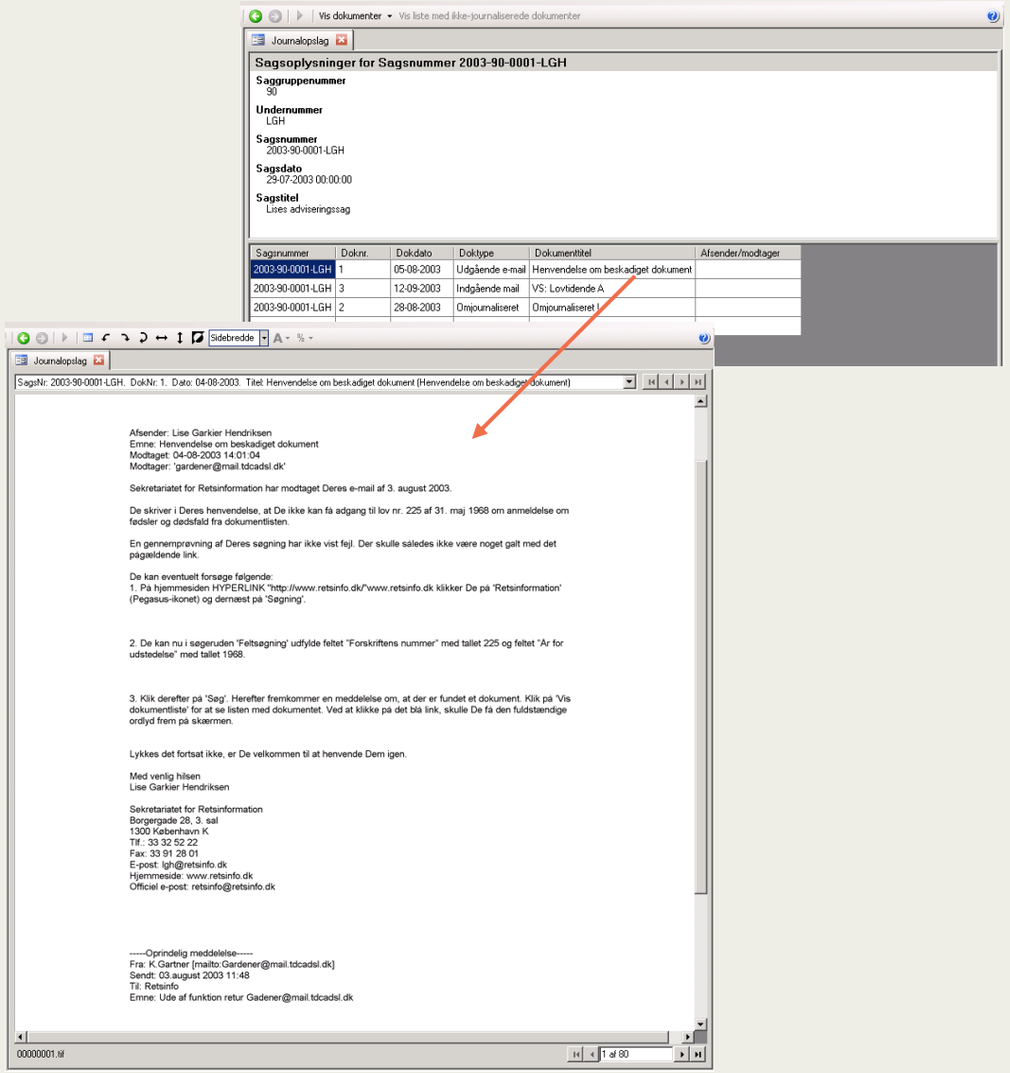
Sofia has three different search templates:

1. Template for records management systems for (IPs with documents)
2. Custom made templates for other kinds of digital records (IP's without documents)
3. SQL-queries



# Accessing documents (TIFF-files)

- A large part of an archivist's workflow in Sofia is to identify information stored in one or more case files from archived ERMS's
- The document viewer is an essential part of the system
- Documents is found by searching with case information in the ERMS template, where identified documents can be accessed



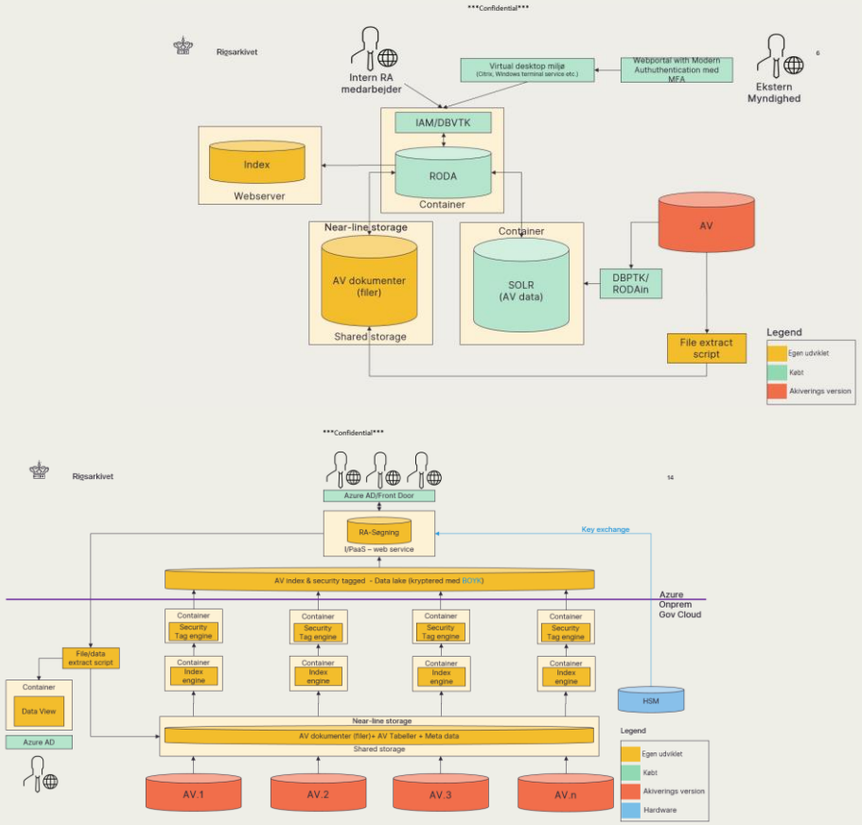
# Project ”Nyt tilgængeliggørelses-system” (New Access System)

- Launch spring 2022
- Purpose: to find an replacement for our current access and search system Sofia
- Milestones:
  - User needs’ assesment (June 2022)
  - System requirements (December 2022)
  - System choice (June 2023)
- Deliverables 2024:
  - Revised system requirements
  - System design
  - Tender
  - Implementation and closure of Sofia

# System choice

- One of our most important demands for a new system is the option for external access, which in turn raises the security demands
- Accenture was hired in the spring of 2023. Accenture are experts within the fields of IT security, system architecture, and more
- Goal: create a number of possible technical solutions for a new system, to qualify our descision on system choice
- On this basis, a solution based on DBPTK was decided upon

accenture



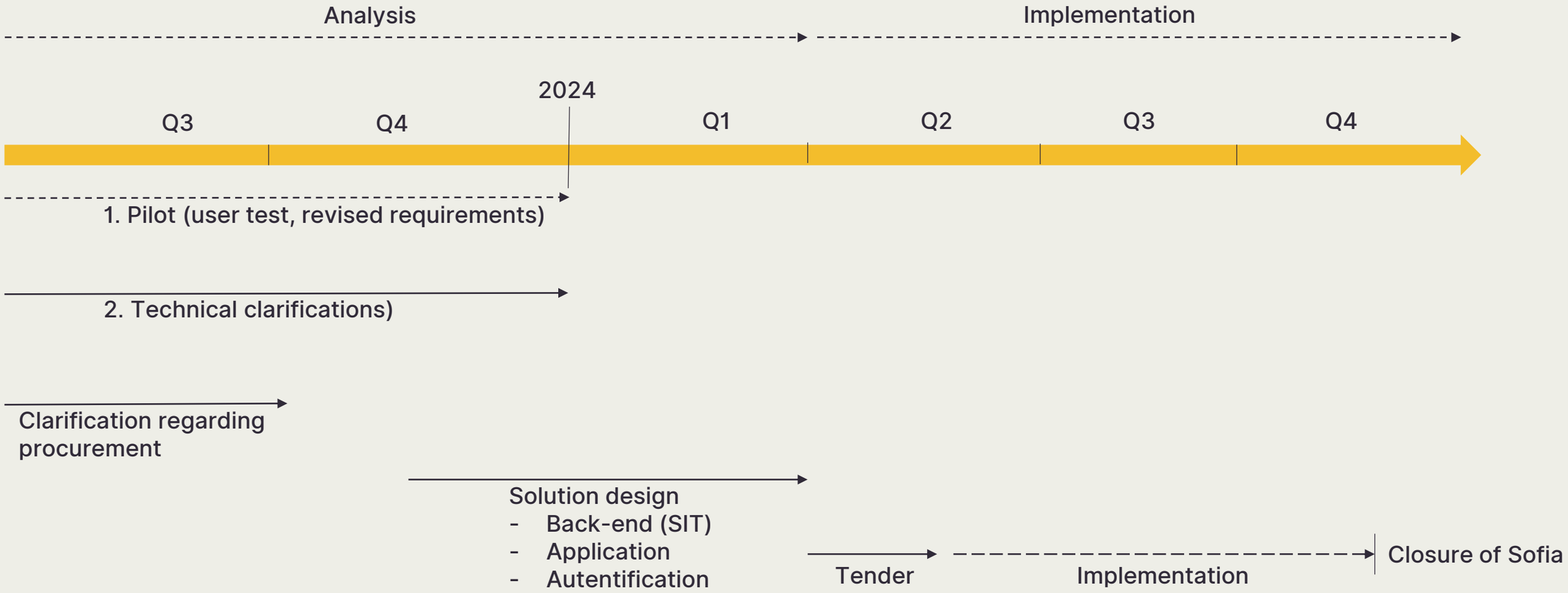
# Database Preservation Toolkit (DBPTK)

- DBPTK is a collection of multiple tools developed in relation to the E-ARK project
- DBPTK consist of two primary tools: a SIARD-creation tool and a SIARD-viewer, where it is possible to search through data in a ”Google-like” manner
- DBPTK exists in three editions:
    1. Desktop – creation and validation of SIARD-files as well as local viewing and browsing of data
    2. Enterprise – web application to browse and search through multiple large databases. Supports multiple users simultaneously
    3. Developer – a command-line tool for scheduling, automation and systems integration


database toolkit
preservation

# Preliminary assesment of DBPTK

- DBPTK fulfills ≈ 80 pct. of our demands:
  - Better search options
  - Faster dissemination
  - More options for customization
  - Better export options
- Missing features:
  - Full implementation of SIARD-DK
  - Lokalisation (Danish translation)
  - Integrated document viewer
- Needs clarification:
  - Search forms
  - Access and autentification


Analysis
Implementation
2024
Q3
Q4
Q1
Q2
Q3
Q4
1. Pilot (user test, revised requirements)
2. Technical clarifications)
Clarification regarding procurement
Solution design
- Back-end (SIT)
- Application
- Autentification
Tender
Implementation
Closure of Sofia

# How do we help each other in the archival community?

- User group for archiving of relational databases?
- OPF?
- The main idea behind using an open source standard solution is to lower own development costs - do we benefit enough from the possibilities that we have?