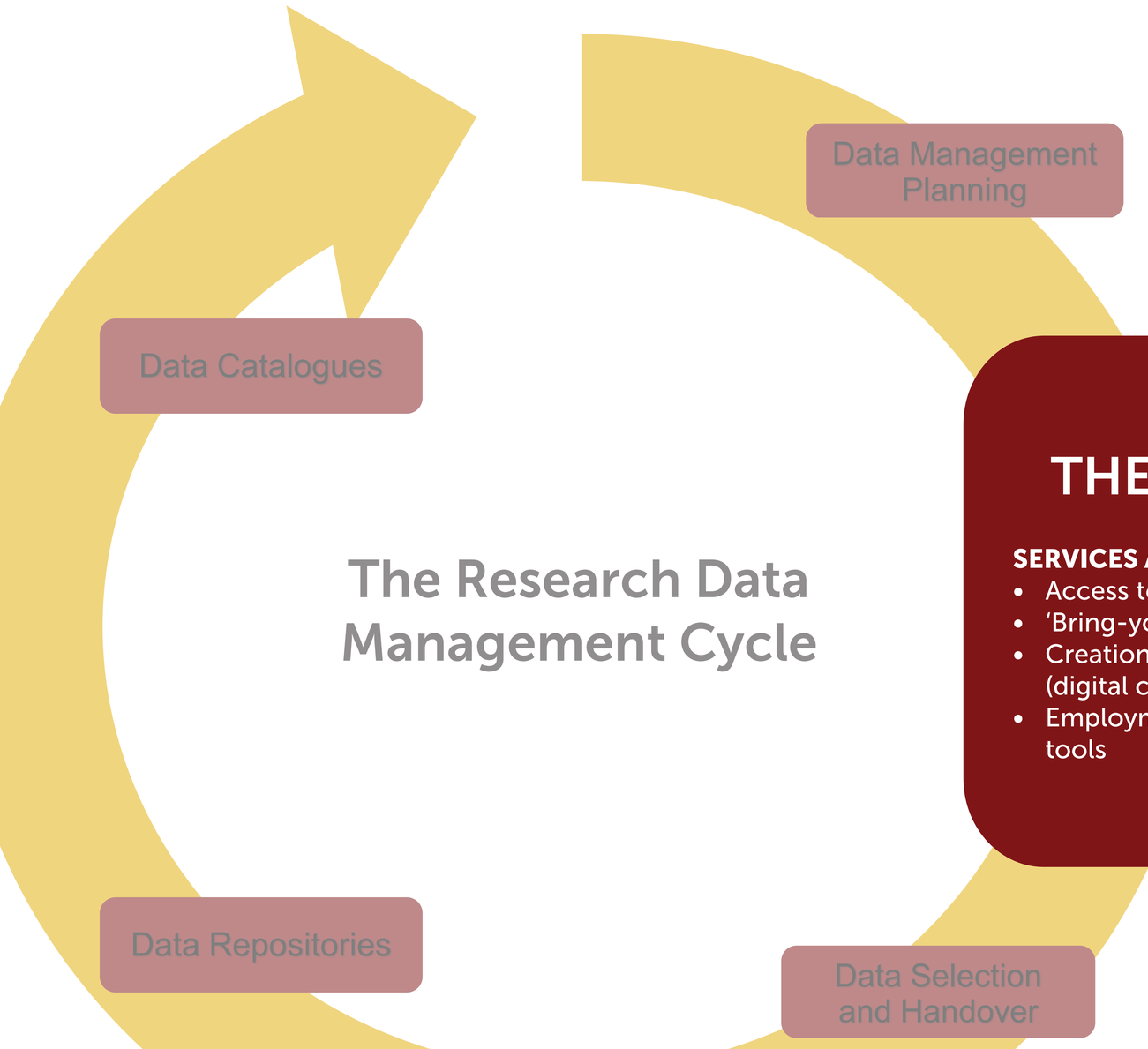
Opening Digital Archives for Research:

DEIC NATIONAL CULTURAL HERITAGE CLUSTER, STATE AND UNIVERSITY LIBRARY



MANAGING ACTIVE DATA IN THE CULTURAL HERITAGE CLUSTER

SERVICES AND TOOLS
- Access to research data and objects
- 'Bring-your-own-research-data'
- Creation of personalized collections (digital corpora)
- Employment of quantitative analytic tools

DEVELOPMENT IS BASED ON RESEARCH PROJECTS
- Probing a Nation's Web Domain
- Digital Footprints
- The Development of the Danish Language in Social Media
- - and other researchers' requests based on software tests

AVAILABLE DATA

Netarchive 2005-
Harvest of the Danish part of the internet
691 TB data – more than 20 billion objects

Radio/TV 2005-
Collection of all national television and radio channels
2.4 PB data – more than 2.6 million hours of broadcast

Digitized newspapers 1749-2013
Goal: 32 million pages by the end of 2017
Present: 203 TB data – 20,944,248 pages

ANALYTIC TOOLS

IBM BigInsights, Hadoop, Spark, Python, R

CHALLENGES TO BE MET

- Establish an interdepartmental organization
- Develop services covering the entire Data Management Cycle
- Develop facilities for collaborative research and data sharing
- Legal issues related to immaterial law and protection of personal data

WHAT IS IN IT FOR THE LIBRARY

- Archived data used for research
- Development of Library services relevant to researchers
- Promotion of the Library

THE PROJECTS

Probing a Nation's Web Domain. Aim: to analyze the development of the Danish internet from 2005, based on data from the Webarchive.

Digital Footprints. Aim: to analyze photographic images and related metadata mainly from Facebook. The data is collected by the project researchers.

The Development of the Danish Language in Social Media. Aim: to analyze this development, based on the digital collection of the State and University Library.

CONTACT

Filip Kruse - fkr@statsbiblioteket.dk
Jesper Boserup Thestrup - jbt@statsbiblioteket.dk

Kulturarvscluster DeiC DEff STATSBIBLIOTEKET STATE AND UNIVERSITY LIBRARY