ICA
ROMA 2022

ARCHIVES: BRIDGING THE GAP
ARCHIVES: COMBLER LES FOSSÉS

September 19th-23rd, Italy

SECURING OUR DIGITAL MEMORY – OR THE ENTANGLED WORK OF LONG-TERM PRESERVATION OF DIGITAL RECORDS

ARCHIVES: BRIDGING THE GAP

9th Annual Conference of the International Council on Archives

19-23 September 2022 | Rome, Italy

Christian Larsen
Birgitte Vedel-Troelsen
The Danish National Archives

INTRODUCTION

Digital record systems in Denmark since the 1960s

The Danish National Archives has received information packages (archive packages) since 1975.

9,397 archive packages in an archive-defined structure with data from 1964-2022 – 491 terabyte.

Intensive digitalization of the public administration, new technologies and new types of data during the last two decades.

Changes in the reuse of data – from historical research to the history of climate changes, or documentation of abuse in foster care homes.

WHAT HAVE WE DONE?

New practice-oriented appraisal guidelines 2021/2022

Related to our 2017 strategy for selecting the data that provide the most accurate reflection of Danish society.

Moving towards Cook's paradigm of «identity archiving» documenting citizens as much as the state.

New guidelines to function in a digital world with a focus on the *properties* of the data and *not* the system in which they exist.

And an explanation for our colleagues in the documentation and access department what we preserve – and do not preserve

FOUR PRINCIPLES

4 principles or guidelines:

1) Data content
2) Data population
3) Data's actor
4) Data coverage and availability

Each principle is divided into several criteria, and with a method for assessing what to emphasize, which data we normally will take in and which data we normally discard.

Using the principles to appraise 1,500 government IT systems from the Archives' 2020 governmental survey on IT systems.

WHAT TO DO NOW?

Developing archival science as a part of the National Archives' new research strategy.

And perhaps – within some years – a new Archives Act focusing on data and not archives?

LEGISLATION ON INFORMATION PACKAGES, 2004-2022

Until 2000: Electronic records were submitted on punch cards, magnetic tapes etc.

2000 and 2002: Circular letters No 25 and no. 4 regarding government authorities' submission of electronic archives systems to the National Archives

2004: Executive Order No 342 on Information Packages of data that are determined to preservation from electronic archives systems

2010: Executive Order No 1007 on Information Packages (based on SIARD.dk)

2020: Executive Order No 128 on Information Packages (minor adaptions)

NEW CHALLENGES

Increased amounts of data
Many tera bytes
Tables with millions of rows
Records systems with millions of files.

How to include new data formats
Danish Concept Model
See more: https://github.com/the-danish-national-archives/concept-model

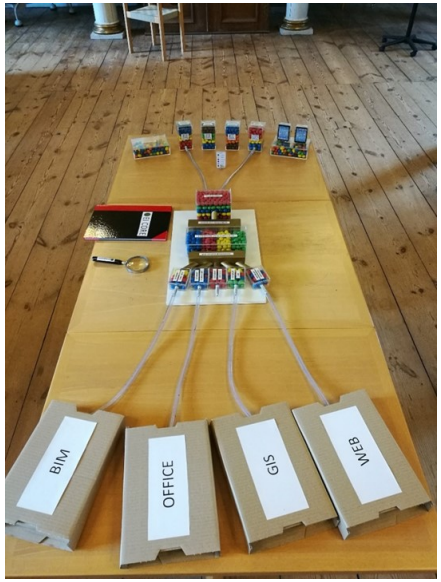
Test
Spreadsheets
Genom data

ARTIFICIAL INTELLIGENCE, DATA LAKES AND ANALYSIS PLATFORMS

Problems
How to document the background of the decisions?

In a system of independent long-term preservation, how to preserve the code?

How do we appraise and preserve, when a specific data set is only put together for a very short period of time?



Acquisition & Government Services is learning about data warehouses with the use of M&M chocolates

BIG DATA IN A HANDS-ON PERSPECTIVE

1 single IT platform, called «Aula»,
used by 1,700 public schools and 2,500 daycare facilities

3,045 TB in 5 years: 250,000 emails each month, and 10,000 posts from institutions and schools to parents and pupils

Three models
Model 1: All messages, all posts, all photographs from all schools and all daycare facilities – from ONE DAY

Model 2: All messages, all posts, all photographs from ALL YEAR from ONE school and ONE daycare from each of the 98 municipalities

Model 3: A combination of model 1 AND model 2.

CONCLUSION

Long tradition for long-term preservation of digital-born data in Denmark.

An increased interest in using our data in new and alternative ways

New challenges, consisting of, amongst others,
- Vast amounts of data
- Data formats not suited for submission in the traditional preservation formats
- IT systems with artificial intelligence and data analysis platforms with short-term data collections

Sollutions
- New approaches to appraisal
- Beginning to analyze how to include new and more complex data into our collections.