



STATENS ARKIVER

Sofia – a tool for access to digital born data and documents

Lone Smith Jespersen, The Danish National Archives



18. Tagung des Arbeitskreises Archivierung von Unterlagen aus digitalen Systemen
Weimar, 11.–12. März 2014



The Danish National Archives and digital records

- Transfer since 1970's and first dataset from 1961
- Media are monitored to secure migration to new media
- 2005-2008 format and structure conversion
- Plans for migration to new structure and formats

- Danish Archives Act: Authorization to lay down detailed regulations on transfer from state authorities and institutions



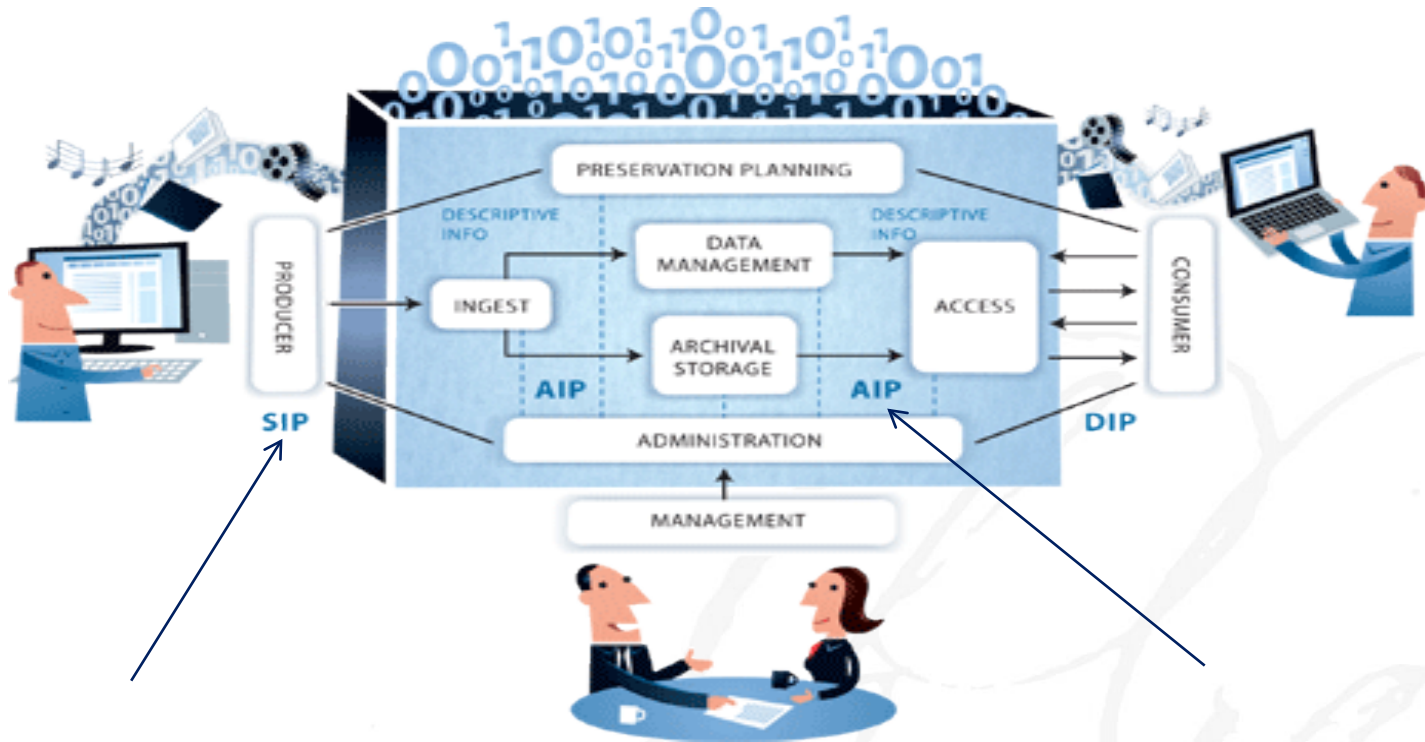
The Danish National Archives and digital records

- Now approx 5.000 archival versions (a package with data, documents and documentation)
- A very big part are not accessible without application and permission, also from the Data Protection Agency (information about private individuals)





Specifications for SIP / AIP



OAIS reference model (<http://digitalbevaring.dk/oais/>)



Specifications for SIP / AIP

- Since 2010 a Danish version of SIARD
- 2000 – 2010:
 - data stored in "flat files" with a file for each table in a relational database structure
 - structural and semantic metadata (xml markedup)
 - tiff documents (if RBMS-system)
 - documentation (tiff documents)
- archival versions from before 2000 have been converted to the specifications 2000-2010



STATENS ARKIVER

The Sofia tool

Developed to give access to AIP's according to the Danish specifications 2000-2010

Hopefully developed to handle SIARD (DK) AIP's in 2014



Digitalbevaring.dk

SQL Converter and log

SQL Converter

Filer Hjælp

AV-ID: C:\00010728\10728001

Start Indstillinger <<

Database Datakonverteringer

Indlæs

Struktur og generel information

Alt excl. DOKMAP-dokumenter

Hele arkivaliet

Forbindelse

Server: ITVSRV-03\TEST

Database: DB10728

Anvend ODBC

Servertype

Microsoft SQL Server

MySQL

Oracle

IBM DB2

Autorisation

Log på med Windows-godkendelse

Log på med brugernavn og adgangskode

Brugernavn: _____

Adgangskode: _____

Generér SQL-komandofil i mappe: _____

```
;  
; Hvis "MaxDataSize", DocumentDataSize eller "MaxLogSize" udelades, anvendes server default.  
;  
;  
; UseUnicode Angiver om der skal benyttes unicode, dvs. nchar, nvarchar og ntext datatyper  
; Værdier kan være (0=nej; 1=ja).  
;  
-----  
[System]  
SystemLog = 1  
Niveau = 2  
SADeveloper = 1  
IndexColumn = 1  
AutoConvertToString = 1  
ReservedSqlKeywords = "C:\TGP\SQLConverter\bin20090626\ReservedWord.txt"  
  
[OnlineRaid]  
; Path = "\\RATGPPSRV01\Depot1\MetaData"  
Path = "\\RATGPPSRV01\Depot1\AV"  
  
[SqlOutput]  
Path = ""  
  
[OnlineHelp]  
Titel = "BrugerVejledning"  
Fil = "Vejledning.pdf"  
  
[Server]  
ServerType = "MSSQL"  
DataSource = "RATGPSQL01\TGPSQL01"  
Security = "Integrated"  
GenInfoOnBLOB = 0  
DokmapOnBLOB = 0  
TextverOnBLOB = 0  
  
[MSSQL]  
TableNameLength = 128  
ColumnNameLength = 128  
MaxFields = 1024  
MaxRowSize = 8060  
DatabaseTimeout = 1800  
CommandTimeoutL = 30  
CommandTimeoutM = 60  
CommandTimeoutH = 600  
CaseSensitive = 1
```



The Sofia-system and SQL Converter

Sofia is based on a database with:

- all data from AIP's, converted to DIP's
- the archives administrative data, e.g. user access
- views created by the archives (or the user)

SQL Converter load all data from an AIP into the standard database (MS Sql Server)





Preparation for access

- Convert AIP to DIP using SQL Converter
- Specify accessibility for the DIP or part of it (0, 20 or 75 years)
- Specify access for a user to a specific DIP
- Create views (can be done by the user)





- 00010305
 - Generel information
 - Tekniske oplysninger
 - Alle tabeller
 - agg
 - Aar
 - ArtID
 - AmtID
 - Antal
 - Fremmednøgle til Amt
 - Fremmednøgle til Art
 - Amt
 - AmtID
 - Amtsnavn
 - Fremmednøgle fra agg
 - Art
 - ArtID
 - Artsnavn
 - Fremmednøgle fra agg
 - Alle kodelister
 - Søgeformularer
 - Statens Arkivers
 - 10305_Vildtstatistik
 - Mine
 - Mine forespørgsler
 - Mine E/R diagrammer

Egenskaber 10305_Vildtstatistik

E/R diagram



SQL kode

```

SELECT
  [Amt].[Amtsnavn],
  [agg].[Aar],
  [Art].[Artsnavn],
  [agg].[Antal]
FROM [agg]
LEFT OUTER JOIN [Art] ON [agg].[ArtID]=[Art].[ArtID]
LEFT OUTER JOIN [Amt] ON [agg].[AmtID]=[Amt].[AmtID]
  
```

Søgeformular design

```

<form>
  <sektion>
    <titel>Vildtudbyttestatistik, 1941-1989</titel>
    <linie>
      <element bredde="2" type="Tekst">
        <titel>Artsnavn</titel>
        <reference binding="[Art].[Artsnavn]" type="TEXT" typedb="STRING" />
        <beskrivelse>Navn på dyret</beskrivelse>
      </element>
      <element bredde="2" type="Tekst">
  
```

Søgeformular

10305_Vildtstatistik

Vildtudbyttestatistik, 1941-1989

Artsnavn Amt

År

Ryd

Søg

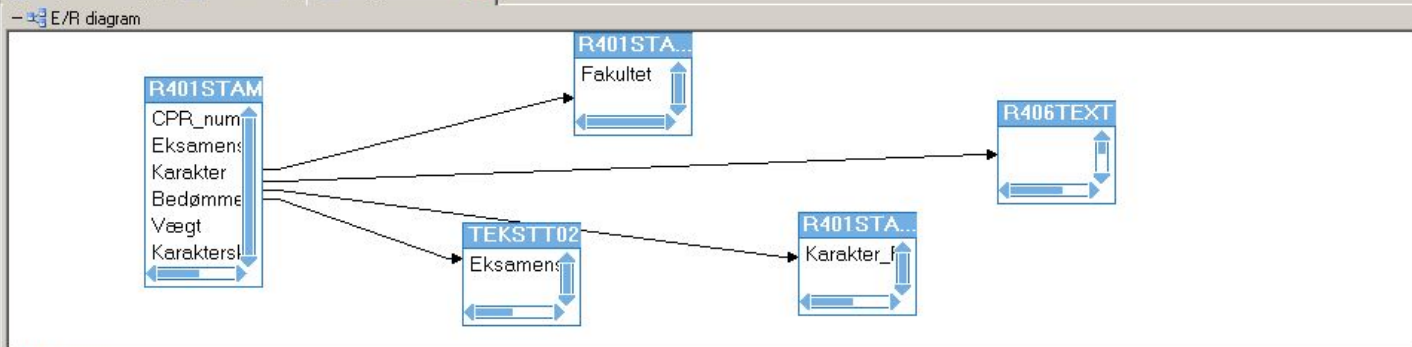
Resultat

	Amtsnavn	Aar	Artsnavn	Antal
1	Viborg Amt	1965	Grævling	293
2	Viborg Amt	1965	Ræv	7436



- 00000060
 - General information
 - Tekniske oplysninger
 - Alle tabeller
 - R401STAM
 - R404DISP
 - R406TEXT
 - NULREC
 - EKSAMT01
 - TEKSTT02
 - DELMGT03
 - EKRAVT04
 - FAGBLT51
 - FAGBLT52
 - VGFAKT06
 - STTIDT07
 - BTILMT08
 - KONVT09
 - EMNET10
 - LAERET11
 - LOKALT12
 - CNORMT13
 - PFORMT14
 - SLUTREC
 - Tabelgruppe 1
 - Tabelgruppe 2
 - Tabelgruppe 3
 - Tabelgruppe 4
 - Tabelgruppe 5
 - Tabelgruppe 6
 - Tabelgruppe 7
 - Tabelgruppe 8
 - Tabelgruppe 9
 - Tabelgruppe 10
 - Tabelgruppe 11
 - Alle kodelister
 - Søgeformularer
 - Statens Arkivers
 - 60_Karakterer
 - Kontrol
 - Mine
 - Mine forespørgsler
 - Mine E/R diagrammer

Egenskaber Kontrol 60_Karakterer



SQL kode

```

SELECT
[R401STAM_FAKU].[Forklaring] AS [Fakultet],
[R401STAM].[PENR] AS [CPR_nummer],
[R401STAM].[EKKO] AS [Eksamenskode],
[TEKSTT02].[TEXT] + ' ' + [R406TEXT].[TEXT] AS [Eksamensbetegnelse],
[R401STAM].[RESU] AS Karakter,
[R401STAM_RESU].[Forklaring] AS [Karakter_Forklaring],
[R401STAM].[BEDA] AS [Bedømmelsesdato],
[R401STAM].[VGTF] AS [Vægt],
[R401STAM].[KASK] AS [Karakterskala]
FROM [R401STAM]
LEFT OUTER JOIN [TEKSTT02] ON CONVERT(varchar, [R401STAM].[EKKO])=[TEKSTT02].[EKDI]
LEFT OUTER JOIN [R401STAM_FAKU] ON [R401STAM].[FAKU]=[R401STAM_FAKU].[FAKU]
LEFT OUTER JOIN [R401STAM_RESU] ON [R401STAM].[RESU]=[R401STAM_RESU].[RESU]
LEFT OUTER JOIN [R406TEXT] ON [R406TEXT].[PENR]=[R401STAM].[PENR] AND [R406TEXT].[EKKO]=[R401STAM].[EKKO] AND [R406TEX
    
```

Søgeformular design

```

<form>
  <sektion>
    <titel>Karakterer ved Københavns Universitet, 1969-1991</titel>
    <linie>
      <element bredde="1" type="Tekst">
        <titel>CPR-nummer</titel>
        <reference binding="[R401STAM].[PENR]" type="NUM" typedb="NUM" />
        <beskrivelse></beskrivelse>
      </element>
    </linie>
    <beskrivelse>Beskrivelse af sektionen (kan udelades)</beskrivelse>
  </sektion>
  <beskrivelse>Beskrivelse af formularen (kan udelades)</beskrivelse>
</form>
    
```



STATENS ARKIVER

Two applications – "inhouse" Sofia

Developed 2006-2008

Can only be used by employees and users in the reading rooms (not very often)

Security is a keyword (for both applications)



Digitalbevaring.dk



STATENS ARKIVER

Two applications – web-Sofia

Web-Sofia

Developed 2012-2013

March 2014: 30 DIP's accessible for all

Expected to be more than 200 in 2014 – not data, but metadata and documentation

A user with permission can be given access by using his NemID (individual identification and authentication, used by public authorities and financial institutions)



sofia.sa.dk



The future

Sofia.sa.dk for public authorities to access data and documents from their own old IT-systems (instead of keeping them alive)

A wish from many authorities to save money

The Data Protection Act!

Sofia can't be used for quantitative / statistical research

A future tool to support researchers?

EU-project E-ARK (European Archival Records and Knowledge Preservation) – read more on eak-project.eu





STATENS ARKIVER

Thank you for listening

<https://sofia.sa.dk>

www.sa.dk

Email: lsj@sa.dk





Søg i Statens Arkivers digitale arkivalier

Åben arkivalie

Klik på det arkivalie du ønsker at arbejde med

Tilgængelige arkivalier

00000075	Vejdirektoratet: Trafiktællingssystem	1976-01-01 - 1976-12-31
00000078	Vejdirektoratet: Hastighedsmålinger	1971-01-01 - 1981-12-31
00000092	Direktoratet for Toldvæsenet: Momsregister	1975-01-01 - 1975-12-31
00000115	Københavns Havnevæsen: Skibsregister ohsa	1985-01-01 - 1985-12-31
00000238	Statens Regnskabsdirektorat: Statens Budgetsystem	1992-01-01 - 1992-12-31
00000249	Statshavnsadministrationen i Frederikshavn: Havnesystem (hs), Frederikshavn Havn	1988-01-01 - 1988-12-31
00000305	Arbejdsmarkedets Tillægspension (ATP): Arbejdsgiverregister	1969-01-01 - 1970-12-31
00000333	Hvidt, Kristian: Register med udvandrerdatabaser (datagrundlag: Politiets udvandringsprotokoller)	1868-01-01 - 1899-12-31
00000397	Civilretsdirektoratet: Fonds- og foreningsregistret	1991-01-01 - 1991-12-31
00000398	Civilretsdirektoratet: Fonds- og foreningsregistret	1992-01-01 - 1992-12-31
00000494	Statshavnsadministrationen i Frederikshavn: Havnesystem (hs), Frederikshavn Havn	1989-01-01 - 1989-12-31
00000601	Statshavnsadministrationen i Esbjerg: Hydrometriske målingers database	1976-08-01 - 1976-12-31
00000938	Statshavnsadministrationen i Esbjerg: Hydrometriske målingers database	1977-01-01 - 1977-12-31
00000939	Statshavnsadministrationen i Esbjerg: Hydrometriske målingers database	1978-01-01 - 1978-12-31
00000940	Statshavnsadministrationen i Esbjerg: Hydrometriske målingers database	1979-01-01 - 1979-12-31
00000941	Statshavnsadministrationen i Esbjerg: Hydrometriske målingers database	1980-01-01 - 1980-12-31
00003006	Statshavnsadministrationen i Frederikshavn: Havnesystem (hs), Helsingør Havn	1989-01-01 - 1989-12-31
00010305	Landbrugsministeriets Vildtforvaltning: Den danske vildtudbyttestatistik 1941-1988	1941-04-01 - 1989-03-31
00010361	Told- og Skattestyrelsen: Selskabsskatteregistret, SSR	1984-01-01 - 1984-12-31
00010468	Danmarks Statistik: Varestatistik for industri	1986-01-01 - 1986-12-31
00010801	Danmarks Statistik: Det erhvervsstatistiske registersystem, Momsstatistik	1990-01-01 - 1991-12-31
00012739	Henning Eriksen: Henning Eriksens privatarkiv	1991-01-01 - 2009-12-31

Vildtudbyttestatistik, 1941-1989

Artsnavn  Amt 

År 

1955

Ryd

Amtsnavn	Aar	Artsnavn	Antal
Bornholms Amt	1955	Grævling	0
Danmark	1955	Grævling	3369
Frederiksborg Amt	1955	Grævling	178
Fyns Amt	1955	Grævling	79
Københavns Amt	1955	Grævling	83
Nordjyllands Amt	1955	Grævling	246
Ribe Amt	1955	Grævling	195
Ringkøbing Amt	1955	Grævling	223
Roskilde Amt	1955	Grævling	0
Storstrøms Amt	1955	Grævling	223
Sønderjyllands Amt	1955	Grævling	376
Vejle Amt	1955	Grævling	580
Viborg Amt	1955	Grævling	495
Vestsjællands Amt	1955	Grævling	205
Århus Amt	1955	Grævling	486
Bornholms Amt	1955	Ræv	614
Danmark	1955	Ræv	42974
Frederiksborg Amt	1955	Ræv	2617
Fyns Amt	1955	Ræv	1863
Københavns Amt	1955	Ræv	1395
Nordjyllands Amt	1955	Ræv	6403
Ribe Amt	1955	Ræv	2522
Ringkøbing Amt	1955	Ræv	4459
Roskilde Amt	1955	Ræv	0
Storstrøms Amt	1955	Ræv	1534
Sønderjyllands Amt	1955	Ræv	3430
Vejle Amt	1955	Ræv	4741
Viborg Amt	1955	Ræv	4812
Vestsjællands Amt	1955	Ræv	17673
Ryd	1955	Ræv	4011

Rigsarkivet, IT-ARK
Rigsdagsgården 9
1218 København K

Systembeskrivelse og administrativ beskrivelse

Formål

Hydrometriske målingers database er en opsamling af vind- og vandstandsmålinger ved Esbjerg Havn. Databasen indeholder målinger pr. kvarter siden 1. august 1976, og de danner grundlag for udarbejdelse af en lang række vind- og vandstandsstatistikker.

Disse statistikker anvendes primært ilbm. pejle- og oprensingsopgaver, samt i udbudsmateriale for etablering og vedligeholdelse af havnekonstruktioner.

Databeskrivelse

De hydrometriske målinger er i princippet i en index sequentiell fil, hvortil der er udviklet et behandlings- og statistikprogram i 4. generationssproget Today under operativsystemet OpenVms.

Inden de opsamlede målinger er lagret i databasen, har de gennemgået en omhyggelig kontrol. Dette har medført sletning af fejlregistreringer, opdatering og justeringer af måleresultater, hvorfor databasens indhold er af størst mulige kvalitet.

I det følgende beskrives de enkelte data:

Datoen er registreret på formen CCYYMMDD, hvor CC=århundrede, YY=år, MM=måned og DD=dag. Der er p.t. målinger for perioden 1. august 1976 til 31. december 1999.

Klokken er en angivelse af tidspunktet for målingen på formen HHMM, hvor HH=timer og MM=minutter. Målingerne af fortaget med et kvarters interval på minuttallene 00, 15, 30 og 45.

I perioder med hel eller delvis udfald på målingerne er dato og klokken altid registreret. Det vil sige, at der ubrudt er registreret en post for hvert kvarter i hele perioden.

Vandstanden på Grådyb Barre er registreret med 2 decimaler i meter. Målingen forefindes i MSL, som er en forkortelse for 'Middel Springtids Lavvande'. Denne kote ligger 70 centimeter under DNN, som er en forkortelse for 'Dansk Normal Nul'. Således vil en registrering på 1,55 m MSL være det samme som 0,85 m DNN.

Den givne vandstandsmåler er placeret ca. 12 km. vest for Esbjerg Havn og syd for indsejlingen ved Grådyb Barre.

Har målingen manglet eller været ubrugbar, er værdien 9,99 registreret som invalid måling.