




# Arkivering af geodata i Danmark

Jeppé Klok Due  
It-arkivar  
Statens Arkiver  
Rigsarkivet

[www.sa.dk/ra](http://www.sa.dk/ra)

# Disposition

- Status for digital arkivering i Danmark
  - Det danske bevarings set-up
  - Den danske version af en SIP
  - GML
  - Geodata i en SIP
    - 3 forskellige modeller
  - Demonstration af en SIP med geodata
- 

# Arkivering i Danmark

- Første aflevering af digitale arkivalier 1973
- Data afleveres som en Submission Information Package (SIP)
- 4.000 SIP i bevaringslager
- Godkender 200-400 SIP årligt
- Regler for aflevering af geodata 2010
- Første aflevering af geodata 2011
  - Matrikelkortet/Jordregistret årgang 2011 fra KMS

# Det danske set-up

- Myndigheder anmelder nye it-systemer
- Rigsarkivet laver bevaringsvurdering
- Rigsarkivet bestemmer hvornår data afleveres
- Data afleveres som en SIP
- Omkostninger for konvertering til arkivformat påhviler arkivskaber

# Submission Information

## Packages

- SIARD (Software Independent Archiving of Relational Databases)
- Selvdokumenterende SIP
- Kun relationelle databaser
- Tabeller eksporteret til XML
- Få formater tilladt: TIF, JPG2000, MP3, MPG2 og MPG4 og GML

# Geodata i en SIP

- Geodata afleveres som GML-filer
- GML-filerne må ikke være større end 1 GB
- De nødvendige XML-skemaer til validering af GML-filerne skal indgå i SIP

*Byg*

# GML

## GML 3v3.DK - ISO 19136

- International standard
- Open Source – fuld systemuafhængighed
- Tekstbaseret, kan læses og forstås uafhængig af program
- Dataudvekslingsformat med bred anvendelse og udbredelse

# Strukturen i en SIP

- ▲ AVID.SA.18001.1
  - ▷ ContextDocumentation
  - ▷ Documents
  - ▷ Indices
  - ▷ Schemas
  - ▷ Tables

*Ulysses*



# Indices - strukturen i indholdet



Oplysninger om placering og indhold af filer i SIP

*BLA*

# Indices - strukturen i indholdet



Metadata om SIP:  
datering, indhold,  
arkivskaber

*Be:yer*

# Indices - strukturen i indholdet



Oplysninger om de dokumenter, der beskriver anvendelsen af driftssystemet og metadata

Be:yer

# Indices - strukturen i indholdet



Strukturen i dataindholdet. Metadata om tabelindhold, relationer mellem tabeller og relation mellem tabeller og dokumenter

# Indices - strukturen i indholdet



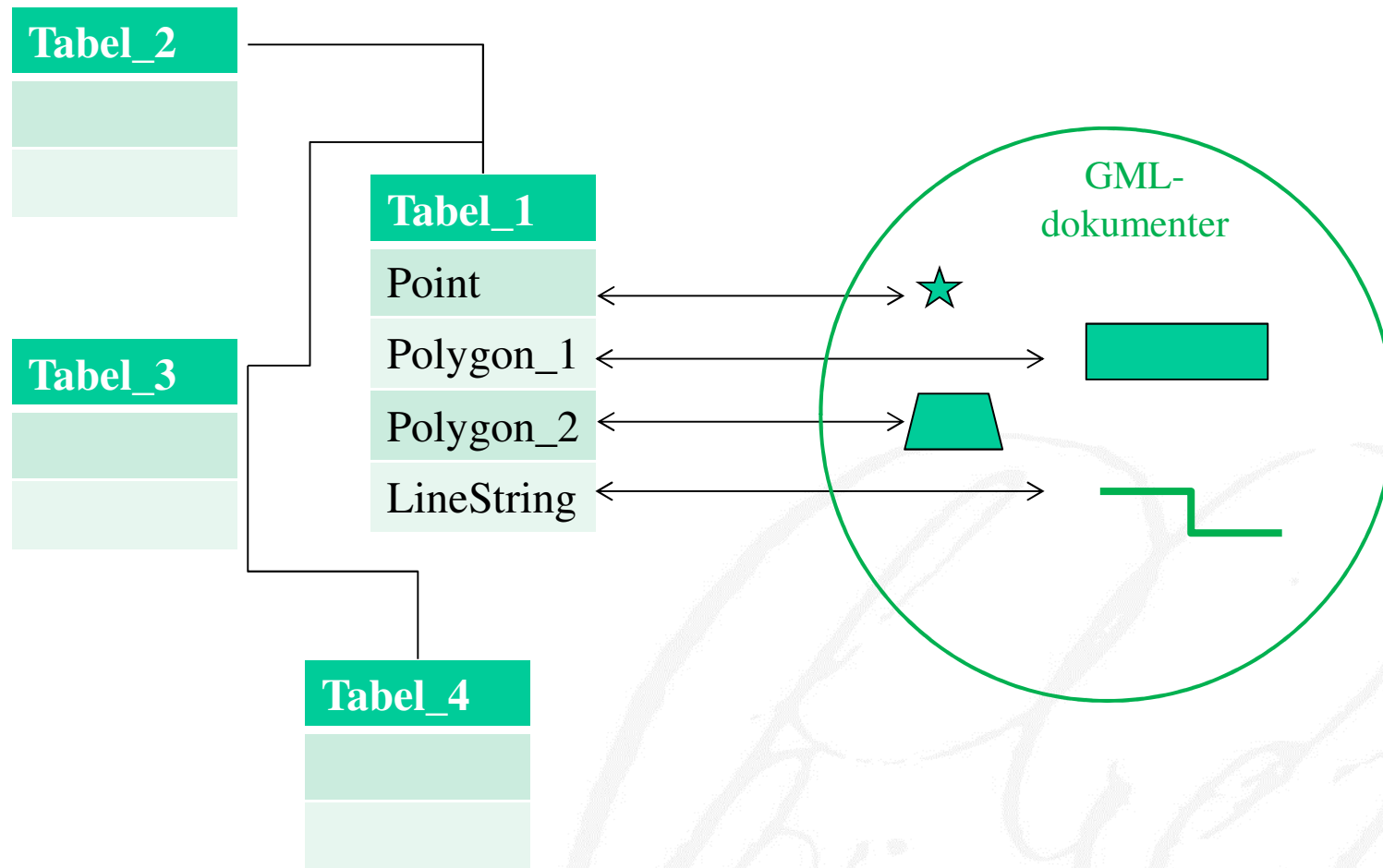
Oplysninger om dokumenterne i SIP: Formater, relation til dokumenthierarki og skemareference til GML-filer

# Forskellig datastrukturer i en SIP afhængig af driftssystemet

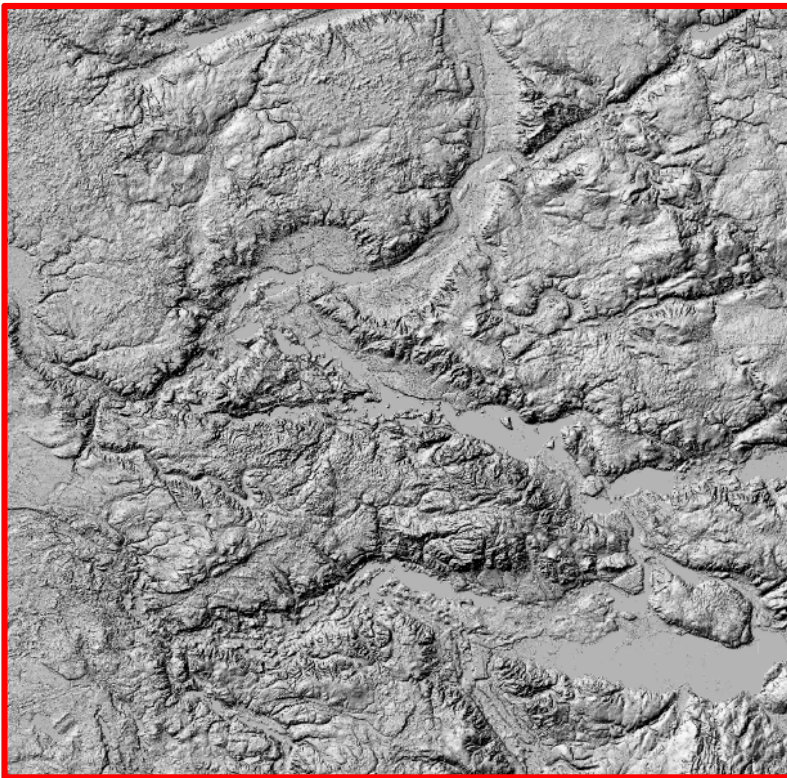
- En record i en database/tabel pr. GML-fil
- Geodatasæt uden database
- En GML-fil og en database

*B. Petersen*

# En record i en DB pr. GML-fil



# Geodatasæt uden database



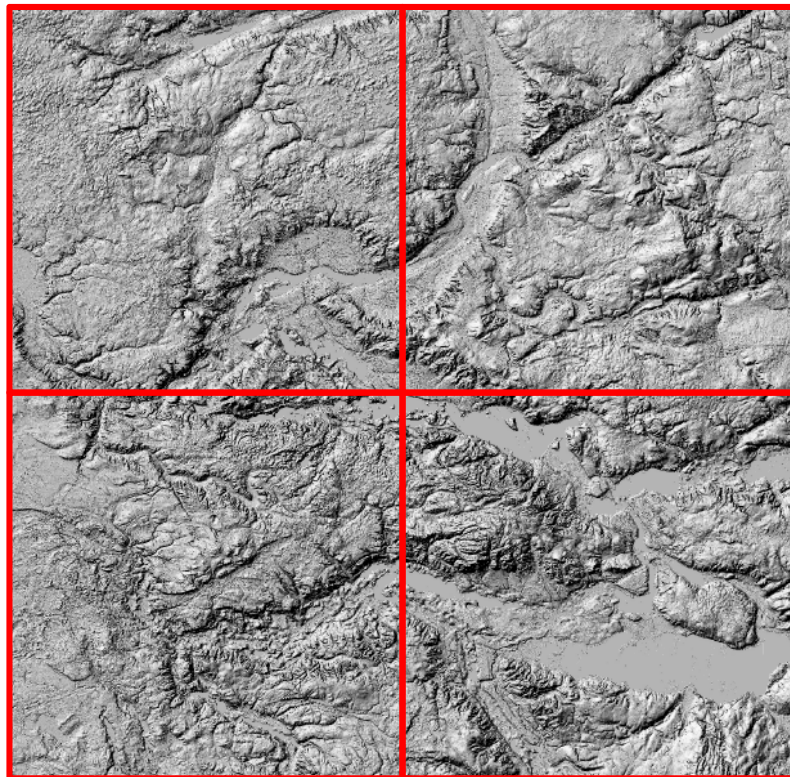
1.gml

Index-tabel		
ID	kommun	envelope
1		

Byggen



# Store geodatasæt uden database



1.gml

2.gml

Index-tabel		
ID	Kommun	envelope
1		
2		
n		

Big Data

# En stor GML-fil og en DB



## GML-fil

```
<LineString  
<Point  
<Polygon_1  
<Polygon_2  
<N...  
</>
```

## Tabel\_2

## Tabel\_3

## Tabel\_1

## Tabel\_4

Et objekt i GML-filen  
relaterer til en record i en  
databasetabel

# En stor GML-fil og en DB



## GML-fil

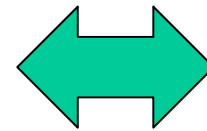
<LineString  
<Point  
<Polygon\_1  
<Polygon\_2  
<N...  
</>

## Tabel\_2

## Tabel\_3

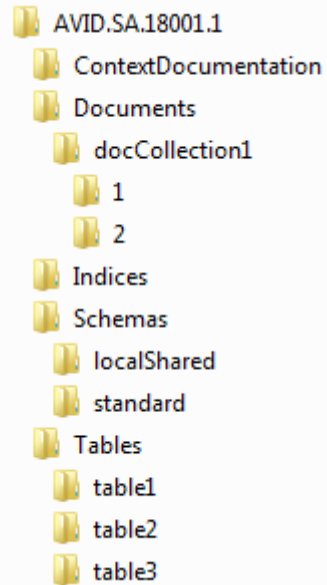
## Tabel\_1

## Tabel\_4



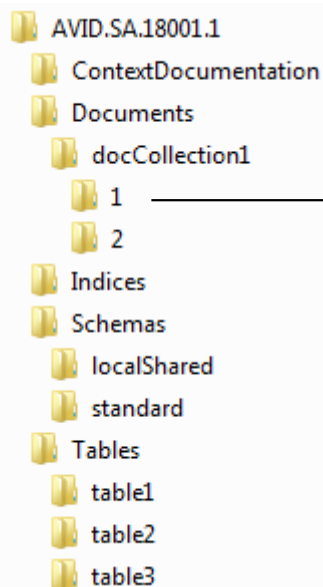
Relation defineret i  
tekstdokument  
ID i GML-fil og i DB

# Gennemgang af SIP med geodata



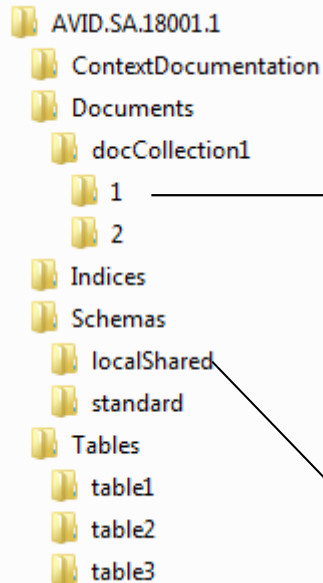
*Oliver*

# GML-fil 1



```
<?xml version="1.0" encoding="utf-8"?>
<gml:FeatureCollection xmlns:gml="http://www.opengis.net/gml" xmlns
  <gml:boundedBy>
  <gml:Envelope srsName="EPSG:25832" srsDimension="2">
    <gml:lowerCorner>441676.00 6049831.00</gml:lowerCorner>
    <gml:upperCorner>892787.00 6402201.00</gml:upperCorner>
  </gml:Envelope>
</gml:boundedBy>
<gml:featureMember>
  <sa:Paragraf_Syv_Arkiv gml:id="ID_1">
    <sa:arkivNavn>Ballerup Stadsarkiv</sa:arkivNavn>
    <sa:kommune>Ballerup</sa:kommune>
    <sa:kommuneNummer>151</sa:kommuneNummer>
    <sa:adresse>Hold-an Vej 7, 2750 Ballerup</sa:adresse>
    <gml:location>
      <gml:Point>
        <gml:pos srsName="EPSG:25832" srsDimension="2">
          711137.05 6181726.45</gml:pos>
        </gml:Point>
      </gml:location>
    </sa:Paragraf_Syv_Arkiv>
</gml:featureMember>
<gml:featureMember>
  <sa:Paragraf_Syv_Arkiv gml:id="ID_2">
    <sa:arkivNavn>Bornholms Ø-arkiv</sa:arkivNavn>
    <sa:kommune>Bornholm</sa:kommune>
    <sa:kommuneNummer>400</sa:kommuneNummer>
    <sa:adresse>Pingels Allé 1, 3700 Rønne</sa:adresse>
    <gml:location>
```

# Skema til GML-fil 1



```
<?xml version="1.0" encoding="utf-8"?>
<schema xmlns="http://www.w3.org/2001/XMLSchema" xmlns:sa="http://www.sa.dk/sa" xi
  <import namespace="http://www.opengis.net/gml" schemaLocation="http://schemas
  <element name="Paragraf_Syv_Arkiv" type="sa:saArkiv" substitutionGroup="gml:_
  <attribute name="arkivID">
    <annotation>
      <documentation>Unik arkivID som defineres fortløbende.
      Værdien af elementet gml:id relaterer til kolonnen arkivID (c1)
      i table3, og der kan etableres en relation mellem de enkelte
      objekter og en tuple i xml-filen table3</documentation>
    </annotation>
  </attribute>
  <complexType name="saArkiv">
    <complexContent>
      <extension base="gml:AbstractFeatureType">
        <sequence>
          <element name="arkivNavn" type="string" nillable="false"/>
          <element name="kommune" type="string" nillable="false"/>
          <element name="kommuneNummer" type="integer" nillable="false"/>
          <element name="adresse" nillable="false"/>
          <element ref="gml:location"/>
        </sequence>
      </extension>
    </complexContent>
  </complexType>
</schema>
```

Alle OGC-skemaer  
og eventuelt andre globale  
skemaer



# Relation mellem GML-fil og DB

- AVID.SA.18001.1
  - ContextDocumentation
  - Documents
    - docCollection1
      - 1 →
      - 2
  - Indices
  - Schemas
    - localShared
    - standard
  - Tables
    - table1
    - table2
    - table3 →

```
<gml:featureMember>  
  <sa:Paragraf_Syv_Arkiv gml:id="ID_1">  
    <sa:arkivNavn>Ballerup Stadsarkiv</sa:arkivNavn>  
    <sa:kommune>Ballerup</sa:kommune>  
    <sa:kommuneNummer>151</sa:kommuneNummer>  
    <sa:adresse>Hold-an Vej 7, 2750 Ballerup</sa:adresse>  
    <gml:location>  
      <gml:Point>  
        <gml:pos srsName="EPSG:25832" srsDimension="2">  
          711137.05 6181726.45</gml:pos>  
        </gml:Point>  
      </gml:location>  
    </sa:Paragraf_Syv_Arkiv>  
  </gml:featureMember>
```

```
<?xml version="1.0" encoding="utf-8"?>  
<table xmlns="http://www.sa.dk/xmlns/siard/1.0/s">  
  <row>  
    <c1>ID_1</c1>  
    <c2>Ballerup</c2>  
    <c3>0151</c3>  
    <c4>Ballerup Stadsarkiv</c4>  
    <c5>Hold-an Vej 7, 2750 Ballerup</c5>  
    <c6>stadsarkiv@balk.dk</c6>  
    <c7>NN</c7>  
    <c8>nn@balk.dk</c8>  
    <c9>1</c9>  
    <c10>2</c10>  
  </row>  
  <row>  
    <c1>ID_2</c1>
```

# Relation mellem GML-fil og DB

- AVID.SA.18001.1
  - ContextDocumentation
  - Documents
    - docCollection1
      - 1
      - 2
  - Indices
  - Schemas
    - localShared
    - standard
  - Tables
    - table1
    - table2
    - table3

```
<?xml version="1.0" encoding="utf-8"?>  
<schema xmlns="http://www.w3.org/2001/XMLSchema" xmlns:sa="http://www.sa.dk/sa" ;  
  <import namespace="http://www.opengis.net/gml" schemaLocation="http://schema  
  <element name="Paragraf_Syv_Arkiv" type="sa:saArkiv" substitutionGroup="gml:  
  <attribute name="arkivID">  
    <annotation>  
      <documentation>Unik arkivID som defineres fortløbende.  
      Værdien af elementet gml:id relaterer til kolonnen arkivID (c1)  
      i table3, og der kan etableres en relation mellem de enkelte  
      objekter og en tuple i xml-filen table3</documentation>  
    </annotation>  
  </attribute>  
<complexType name="saArkiv">
```

Sammenhæng beskrevet i  
tableIndex.xml, i definitionen af  
det felt i den database-tabellen



# Metadata om featuretypes

```
<gml:featureMember>  
<matrikel:OptagetVej gml:id="dk.kms.mat.optvej.48001843">  
<matrikel:featureID>48001843</matrikel:featureID>  
<matrikel:featureCode>9022</matrikel:featureCode>  
<matrikel:featureType>Optaget vej, udstykningsvej</matrikel:featureType>  
<matrikel:ejerlavsnavn>Østerby By, Hals</matrikel:ejerlavsnavn>  
<matrikel:landsejerlavskode>590252</matrikel:landsejerlavskode>  
<matrikel:matrikelnummer>25a</matrikel:matrikelnummer>
```

- Datatyper defineret i xml.skema
- Indholdet defineret i et dokument i contextDocumentation

# Dokumentation af geodata i SIP

- contextDocumentationIndex
  - Relation mellem database/tabeller og GML-fil
  - Beskrivelse af opdeling af filer, hvis >1 GB
  - Administrativ brug af data
  - Features beskrives
- Schemaer
  - Alle OGC-skemaer i localShared
  - Lokale skemaer sammen med GML-fil eller i localShared

# Spørgsmål?



Jeppe Klok Due  
IT-arkivar  
Statens Arkiver, Rigsarkivet  
jkd@sa.dk