



Memorandum

About Migration assesment for georeferenced raster data

Date: 18. august 2020

Author Ann-Kristin Egeland, The Danish National Archives

1. Purpose.....	1
2. Stakeholders.....	1
3. List of production formats and significant properties of geo rasterdata	2
4.1 Production formats at map agency.....	2
4.2 Access formats at map agency.....	2
4.3 List of significant properties of geo rasterdata.....	2
4. Proposed solutions for preservation formats.....	4
5.1 Proposed solutions for preservation in existing preservation formats at DNA.....	4
5.2 Assessment of losses due to migration to existing preservation formats at DNA	10
5.3 Pros and cons of migration to existing preservation formats at DNA	10
5.3 Map agency's evaluation of DNA's proposed solution.....	11
5.4 Map agency's solution for a preservation format for geo rasterdata	11

1. Purpose

The purpose of this memorandum is to consider how georeferenced rasterdata (geo rasterdata) could be migrated to existing preservation formats (GML, JPEG 2000 and SIARD) in The Danish National Archives. Note that this is not a complete analysis of all significant properties.

2. Stakeholders

No stakeholders has been interviewed so far to reveal significant properties and structures of geo rasterdata (both digital images and grid files). This analysis is thus based on the existing knowledge of

the author. At least one map-agency should be interviewed. The DNA project manager may provide information of relevant stakeholders.

The interview should reveal the following:

1. List **original production formats** for geo raster data and describe their **significant properties**. Understand how geo raster data is **structured and stored** in the production system. (relation between digital images, grid files, databases, world-files and so on)
2. List **access formats** for geo raster data provided by map-agency.

Table 1: Stakeholders

Stakeholder	Type
	Creator

3. List of production formats and significant properties of geo rasterdata

4.1 Production formats at map agency

Should be filled out...

4.2 Access formats at map agency

Should be filled out...

4.3 List of significant properties of geo rasterdata

Table 2: Significant properties supporting use of geo rasterdata

Significant property	Description	Is property significant?
Raster object	Digital image or grid file	Significant
Coordinates	Vector coordinates attached or linked to the raster object (georeference).	Significant
Connection of georeference to raster image	How are the coordinates attached or linked to the raster object?	Significant
Coordinate and reference system	Coordinate and reference system compliant with the coordinates	Significant
Attributes	Attributes or local features describing the coordinate or cell information in a grid file	Significant



Reference between coordinates and attributes	The reference linking attributes to coordinates if attributes are stored apart from coordinates, eg. in a separate database.	Significant
Pixel size	See attributes? Pixel size of a digital image? Cell size of a grid file?	Significant
Rotation	What is this and what is it used for? See attributes?	?
Start and end date of content	See attributes?	?
Description of content of raster image	See attributes?	?
Scale	1:10.000 Relevant information for raster data? See attributes?	?
Text	Font, color, vertical... [NARA: These characteristics may be essential if the text displayed in GIS records, such as map legends or display headings, bears meaning through its formatting. The text itself is always essential, but the formatting may also be essential when it is evidence of how the maps were used or displayed by the creator.]	?
Color	Hue, saturation; brightness, contrast... [NARA: Even if exact colors cannot be made persistent, distinctions between colors may be essential to understand the attributes and overlays displayed as a result of a user query. All of these characteristics are ways of measuring and making distinctions between colors.]	?
Query	[NARA: Queries may be graphics-driven, spatial based (point and query) searches for objects and retrieval of the associated attribute data. They may also be data-driven, using data values to display the matching spatial features or the use of attribute values to determine shading pattern of the relevant spatial elements]	?



Display Graph or Plot	[NARA: Features on one data layer are overlaid onto those of other data layers in order to show areas which have a certain combination of attributes: Single map; Multiple overlays; 3-dimensional display. The ability to graph or plot data is essential to the meaning of GIS map records. If there is no value to the map display, or no ability to plot, then the records could be handled much like databases.]	?
Display reports	[NARA: Reports from data tables]	?
Manipulation Functionality (Zoom)	[NARA: Includes but not limited to: Draw; Zoom; Animate (continuous and/or step-by-step progression); Contour; Pan; Enhance (smooth, simplify, merge, dissolve, rotate, invert). Depending on the software toolkit and data elements available to the user, a host of behaviors are possible that may be essential to the meaning or value of the records. Much GIS functionality concerns manipulation of the display, whether it be a plotted map or reported data from a query. The data elements that allow this functionality are a function of the data type and transfer format. If the records' value lies in how the creator manipulated map attributes and the utilities they used to do so, these behaviors will have to be identified and articulated at appraisal or transfer.]	?
		Not significant
		Maybe significant

4. Proposed solutions for preservation formats

5.1 Proposed solutions for preservation in existing preservation formats at DNA

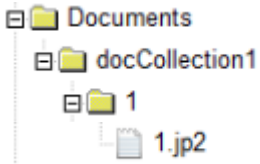
Table 4 describes and illustrates how geo rasterdata could be preserved in DNA's existing preservation formats and wrapped inside a Danish Archival Information Package (se figure 1).

The solution propose that the raster object is migrated to JPEG2000 format, coordinates and other significant properties and attributes related to the coordinates is migrated to GML 3.1.1 and additional

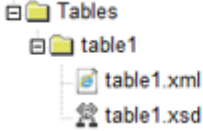
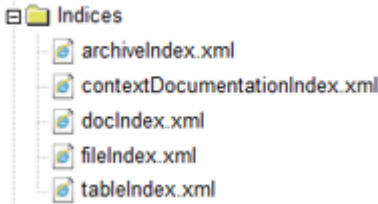
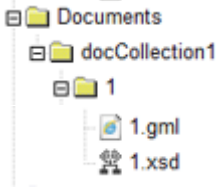


attributes and the reference combining the coordinate with the raster object is migrated to a database in SAIRD.DK format.

Table 3: Solution for preserving geo rasterdata in DNA' existing preservation formats.

Significant property	Preservation format	Is property significant?												
Raster object	<p>Raster objects (digital images) preserved in JPEG2000 format and stored as documents in <i>Documents</i> folder with documentID (1) in the IP:</p>  <p>Table in SIARD.DK holding the documentID and minimum one or two columns with information about the digital image/grid file, imageID or ImageDescription:</p> <table border="1" data-bbox="392 1081 991 1290"> <thead> <tr> <th>documentID</th> <th>imageID</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>32548</td> <td>Copenhagen north</td> </tr> <tr> <td>2</td> <td>32549</td> <td>Copenhagen south</td> </tr> <tr> <td>3</td> <td>67532</td> <td>Aarhus south</td> </tr> </tbody> </table> <p>table1.xml</p> <pre data-bbox="392 1361 743 1827"> <row> <c1>1</c1> <c2>32548</c2> <c3>Copenhagen north</c3> </row> <row> <c1>2</c1> <c2>32549</c2> <c3>Copenhagen south</c3> </row> <row> <c1>3</c1> <c2>67532</c2> <c3>Aarhus south</c3> </row> </pre>	documentID	imageID	Description	1	32548	Copenhagen north	2	32549	Copenhagen south	3	67532	Aarhus south	Significant
documentID	imageID	Description												
1	32548	Copenhagen north												
2	32549	Copenhagen south												
3	67532	Aarhus south												



	 <p>tableIndex.xml</p> <pre><table> <name>Orthophotos</name> <folder>table1</folder> <description>Orthophotos of Denmark 2010</description> <columns> <column> <name>documentID</name> <columnID>c1</columnID> <type>INTEGER<type> <typeOriginal> INTEGER </typeOriginal> <nullable>>false</nullable> <description>Name of folder in the IP containing the Orthophoto (raster image)</description> ...</pre> 	
Coordinates/ georeference	GML 3.1.1 in the GML-file  <pre><gml:Point> <gml:pos>711137.05 6181726.45 </gml:pos> </gml:Point></pre>	Significant
Connection of georeference to raster image	GML 3.1.1 - Local feature in the GML-file <pre><sa:documentID>1</sa: documentID ></pre> <p>GML 3.1.1 – Description of Local feature in the schema file (.xsd)</p>	Significant



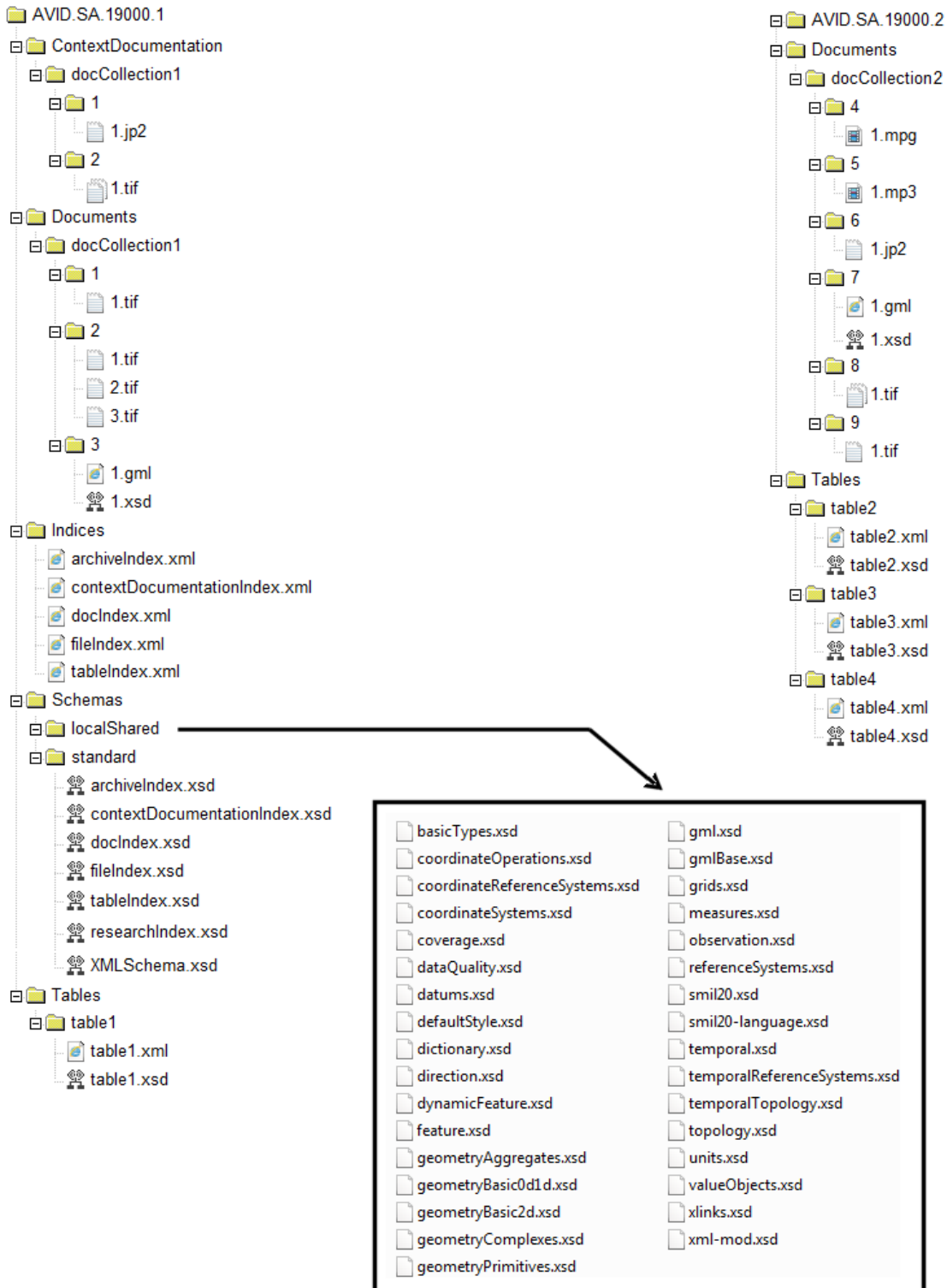
	<pre><element name="documentID" type="integer" nillable="false"> <annotation> <documentation>Raster object in the IP georeferenced by these coordinates (georeference) </documentation> </annotation> </element></pre>	
Coordinate and reference system	GML 3.1.1 in the GML-file srsName="EPSG:25832"	Significant
Attributes	<p>1. GML 3.1.1 - Local features in the GML file.</p> <pre><sa:flightphotoID>32548</sa:flightphotoID > <sa:location>Zealand</sa:location > <sa:dateOfShoot>2010-05-03</sa:dateOfShoot > <sa:cellSize>0,4</sa:cellSize ></pre> <p>The local features are described in annotation and documentation elements in for the local feature in the schema file (.xsd):</p> <pre><element name="dateOfShoot" type="integer" nillable="false"> <annotation> <documentation> The date the image was photographed from the flight. </documentation> </annotation> </element></pre> <p>Or</p> <p>2. SIARD.DK – stored as columns in a database</p> <p>Note: Both 1 and 2 is permitted – compliant with existing regulation concerning vector geodata.</p> <p>Note: If attributes are store as database information in SIARD.DK in the IP (2) the reference between coordinates and attributes should be documented (se below).</p> <p>The IP must always have a database with information about documentID (link to the raster obkect in the Document folder in the IP) even if the attributes are stored directly in the GML-file.</p>	Significant
Reference between	2.	Significant



coordinates and attributes	<p>GML 3.1.1 – a local feature in the gml-file holding the reference to attributes in a table in the IP.</p> <p>The reference is described in annotation and documentation elements in for the local feature in the schema file (.xsd) that holds the reference.</p> <pre><element name="documentID" type="integer" nillable="false"> <annotation> <documentation> The coordinates/georeference/raster object are linked from this local feature to attributes to in table1 (Orthephotos) column imageID (or reuse documentID) in the IP. </documentation> </annotation> </element></pre> <p>Note: If it is made mandatory that the name of this feature holding documentID in the GML-file is always documentID a future migration tool for migrating preservation format to access format (eg. GeoTIFF) is not dependent upon the table information in the SIARD format.</p>	
----------------------------	---	--



Figure 1: An example of the a Danish Archival Information Package with geodata





5.2 Assessment of losses due to migration to existing preservation formats at DNA

The loss during migration from original geo raster format to existing preservation formats for geodata is considered according to the identified significant properties. Table 5 illustrates the expected loss and access whether it is acceptable or not.

Conclusion...

Table 5: Assessment of loss during migration from original format to proposed preservation format

Significant property	Loss	Possible loss	No loss	Comment	Assessment
Raster object			X	Since JPEG 2000 is already a chosen preservation format for images a former decision has concluded that this migration is lossless.	Acceptable
Coordinates			X	Since GML 3.1.1 is already a chosen preservation format for vector geodata a former decision has concluded that this migration of coordinates is lossless and acceptable.	Acceptable
Connection of georeference to raster image			X	How is this connection obtained in the production formats? Other information about the connection lost?	Acceptable
Coordinate and reference system (CRS)			X	Since GML 3.1.1 is already a chosen preservation format for vector geodata a former decision has concluded that this documentation of CRS is lossless and acceptable.	Acceptable
Attributes			X	Since GML 3.1.1 and SIARD.DK are already a chosen preservation format attribute information this migration is lossless and acceptable. Often attributes are stored in databases in production formats.	Acceptable
Reference between coordinates and attributes			X	Since GML 3.1.1 is already a chosen preservation format for documenting references between coordinates in gml and attributes in SIARD.DK in the existing vector data preservation format this migration is lossless and acceptable.	Acceptable

5.3 Pros and cons of migration to existing preservation formats at DNA

Pros

- System independent preservation format
- The formats used has already been chosen as formats suitable for long term preservation



- Validation tools for GML, JPEG2000 and SIARD already exists at DNA and should only be further developed to validate any new significant properties of raster data.
- With this solution the DNA could receive, validate and preserve georeferenced raster data in the near future.

Cons

- A migration tool for migration of preservation format to access format (eg. GeoTiff) is required. Suggest a tool for this. Can FME be used for this migration?

5.3 Map agency's evaluation of DNA's proposed solution

Will be added by the DNA.

5.4 Map agency's solution for a preservation format for geo rasterdata

Will be added by the DNA.