

# Consultancy services: Recommendation of preservation formats for raster geodata

## Case

The Danish National Archives (DNA) is looking for a long-term preservation format for georeferenced raster data (hereinafter referred to as “raster geodata” -referring to both georeferenced digital images and georeferenced grid files). The long-term preservation format is to be placed inside the structure of an E-ARK Information Package. Due to the use of the migration strategy, the contracting authority prefers open, well documented system independent preservation formats that can be validated, migrated and used in contemporary IT-systems (GIS) both now and in 500 years.

## Background

The DNA is an authority and research institution under the Ministry of Culture with 250 employees that resides in Copenhagen, Odense, Viborg and Aabenraa. For more than four decades, the DNA has performed digital archiving by collecting, preserving and giving access to data from the public and private sector.

Data are submitted from the state, regions and municipalities of Denmark as well as from private organisations, individuals and research institutions. Notification to and approval of new IT systems by the DNA is mandatory for state authorities. Submission of data to the DNA is mandatory for all state authorities.

Access to data is dependent on application, apart from those that are now publicly available. Data are submitted to the DNA in a system independent data format structured as relational databases with or without documents. The DNA also preserves huge amounts of scanned data, *digitized records* - i.e. photographic scans of traditional paper records.<sup>1 2</sup>

## Geodata

DNA preserve vector geodata in a system independent data format structured as relational databases with references to gml-files containing significant properties of vector geodata. DNA have not yet received raster geodata due to lack of a suitable preservation format. The present preservation formats includes JPEG2000 files, GML-files and the database preservation format SIARD. A future preservation format for raster geodata based on these existing preservation formats

---

<sup>1</sup> See *Strategy for archiving digital records at the Danish National Archives*:  
<https://www.sa.dk/wp-content/uploads/2014/12/Strategy-for-archiving-digital-records-2013.pdf>

<sup>2</sup> See *Executive Order on Information Packages*:  
<https://www.sa.dk/wp-content/uploads/2020/05/Executive-Order-on-Information-Packages128-1.pdf>

would ease the validation of the format. However, this is just an option to be explored, not a requirement.

### **Future expectations**

In the near future, we expect to embrace the E-ARK preservation formats including the E-ARK information package (IP) and the updated SIARD 2.2 format for databases. E-ARK has also developed a Content Information Type Specification (CITS) with recommendations for preservation of geodata (see Annex 1.1). However, with this call for tender the DNA wish to explore possible preservation formats for raster geodata in more detail before choosing one.

## **Required services**

The Consultant(s) is/are expected to:

- a) Provide a full list of significant properties of raster geodata that are important to preserve in order to be able to see and use the preserved data again in the future.<sup>3 4</sup>**

**This includes:**

- I. Exploration of how users use geodata and outline future significant properties of the raster geodata important for future use.
- II. Conduction of interviews with users using georeferenced digital images and georeferenced grid files. Record the interview with permission.
- III. Consideration of possible needs for searching and identifying specific georeferenced digital image or a group of images among many thousand images.
- IV. Name and provide examples of the original raster geodata files/formats) used for this analysis of significant properties

- b) Complete the migration assessment in Annex 1.3.**

**This includes:**

- I. Interview a map agency in order to enrich and complete the migration assessment. Record the interview with permission. Contact is provided by DNA.
- II. List of production formats and access formats used at a map agency.
- III. Conclude on whether significant properties of the grid file format is covered well enough by the proposed solution in the migration assessment.
- IV. Conclude on whether the losses due to migration to existing preservation formats are acceptable.
- V. Suggest a suitable migration tool for migration of the proposed preservation format for raster data to an access format (eg. GeoTiff).

- c) Analyse and suggest candidates for suitable long-term preservation formats for raster geodata.**

- I. Use the method described in Annex 1.4 and fill out the form in Annex 1.5
- II. Based on the aforementioned analysis (point “a”), argue why these candidates are suitable preservation formats for a preservation institution using the migration strategy.

---

<sup>3</sup> Examples of significant properties of raster geodata format is the measured coordinates, the coordinate and reference system used, additional tabular information, attributes describing the coordinates and so on.

<sup>4</sup> The InSPECT framework is a method that can be used to investigate significant properties of electronic content: <https://significantproperties.kdl.kcl.ac.uk/inspect-framework.html> (Knight, Gareth (2009): InSPECT Framework Report, The Centre for e-Research hos King’s College London and The National Archives [UK])

**d) Recommend the most suitable preservation format(s) for raster geodata.**

**This includes:**

- I. An inclusion and consideration of the recommendations for a preservation format for raster geodata from a relevant map agency, who is responsible for submitting raster geodata to the DNA. This information is available at the DNA.
- II. A description of how and where the significant properties of the raster geodata are placed inside the recommended preservation format(s). Provide an example of a file in the recommended preservation format and show/describe the content/information inside the file; where is the coordinate and reference system information placed in the file? Where is the array information from a grid file placed in the format and so on.
- III. An in-depth consideration of to what extent the recommended preservation format(s) is/are system independent.

**e) Describe how the recommended preservation format(s) should be validated**

- I. List some obvious requirements for the validation of the recommended preservation format(s).
- II. Suggest validation tools for the recommended preservation format(s).

**Product**

The aforementioned services are expected to be presented for the DNA project team in the form of numerous reports, presentations/meetings (such as examples of analysed geodata files in their original format vs. geodata files in the recommended preservation format, recordings of interviews, as well as oral transfer of knowledge) on a regular basis throughout the autumn/winter of 2020.

We expect all services to be delivered in close collaboration with the DNA project team. Thus, the Consultant(s) will get access to DNA's knowledge within the domain through this collaboration.

Delivery of the final product is only considered to have taken place once the Danish National Archives has approved the final report. If the project group has any questions to the final product, the Consultant(s) is obliged to answer these. The contracting authority thus suggest that that the final deliverable is handed over to the project team a week before the contract terminates.

**Time schedule**

The final products must be approved by DNA no later than December 1<sup>st</sup>, 2020. We thus suggest that the overall plan for the performance of the task allows one week of final review for approval. Hence, the Consultant(s) is/are expected to spend a maximum of 2 months from the end of September until the end of November.

**Place of work**

No physical presence in Denmark is required. Ongoing communications as well as potential presentations are to take place on platform(s) chosen by DNA.

## Requirements for the Consultant(s)

**Education:**

The Consultant(s) must have a university degree in the areas of *Computer Science, Geographic Engineering* or similar areas relevant to Digital Archiving and/or geodata.

**Positions:**

The Consultant(s) must be able to document relevant experience from positions in one or more of the following organisations: research institutes, archive institutions, governmental institutions working with raster geodata, software development companies, etc.

**Experience:**

The Consultant *must* have experience with:

- Raster geodata within digital preservation (appraisal, ingest, migration or access)
- Creation, management and/or use of raster geodatasets as well as how these are used in Geographical Information Systems (GIS)
- Converting raster geodata formats

The Consultant *should* have knowledge about:

- Digital Archiving
- The OAIS reference model
- Open standards and open source
- E-ARK preservation formats (SIARD 2.1 and Geodata CITS)
- Raster geodata formats (GeoTiff, JPEG2000 ...)
- GML version 3.1.1 and 3.2.1
- GeoPackage
- Tools for validating and converting raster geodata formats
- Raster geodata formats used by geodata creators in the public administration

**Language**

All communication and writing must be in English or Danish.