



Københavns Universitet



From paper to bytes

Gammeltoft, Peder

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Gammeltoft, P., (2016). From paper to bytes: digitising geographical names records, No. W.P. 82/16, No. 29. UNGEGN Session, 7 p.

**Twenty-ninth session
Bangkok, 25 – 29 April 2016**

Item 16 of the provisional agenda

Activities relating to the Working Group on Geographical Names as Cultural Heritage

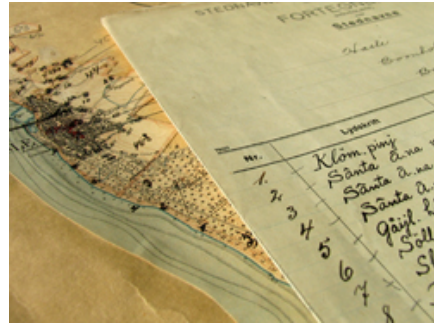
From paper to bytes – digitising geographical names records

Submitted by Denmark*

* Prepared by Peder Gammeltoft (Name Research Section, Department of Nordic Research, University of Copenhagen)

Resolutions concerned:

I/4, VII/5, VIII/9,
VIII/10, IX/4, X/3.



The author checks a recorded geographical name on the record list against a number on the map.

Summary

The Name Research Section at the Department of Nordic Research, University of Copenhagen in Denmark, houses a large corpus of collected geographical names records on minor or 'informal' names throughout Denmark, which currently serves as a research tool. By digitizing this archive, the data from this substantial name collection will be made available to a wider group of users.

The digitization of the Names Archive began in the spring of 2016. The project aims to convert all of the archive's collection of geographical names records maps and lists into spatially enabled geographical names data.. This will facilitate the use of this data with other spatial themes in GIS software as well as in wider areas of research.

Introduction

The Danish Name Research Section at the Department of Nordic Research, University of Copenhagen, is the only center for scientific research into geographical and personal names in Denmark. The section researches, publishes and gives advice on names and takes care of the dissemination of name research into a wider public, partly by publishing the series: *Place-Names of Denmark* (Danmarks Stednavne), and thus has a large archive on names, mainly names from Denmark and neighboring territories. In addition, it acts as a secretariat for the Place-Name Commission.

The entire geographical names archive is organized topographically with the parish as the smallest unit. This system is used throughout the Place-Name Research Section's archives. The topographical organization was originally devised by the Danish Folklore Archive, and which other research and cultural heritage institutions. The *Geographical Names Records Collection* is a central component of the Archive.

The contents of the *Geographical Names Records Collection*

The Geographical Names Records Collection dates back to 1921-22, when a large campaign was launched by the Danish Place-Name Commission to gather minor and 'informal' names of small settlements, fields, streams, waters, and wooded areas. Parish maps and lists were sent, primarily to local head teachers, who undertook the work locally. They were instructed to place a number on a map and note the number on the list and to give the name of the locality, and through primitive phonetics (Dania) what the pronunciation of the name was. A column for additional remarks was often use to describe various things, such as earlier owners, folk tales and archaeological remains and finds at the locality.

The inspiration for the 1921-22 campaign was, however, inspired by a similar, but smaller scaled campaign launched by the Danish Folklore Archive just after the turn of the century to collect local geographical names before they disappeared.

As a supplement to the original campaign, a new one was launched in the 1930's, where name researchers and specially trained scholars and students actively went out and collected names where there had been either no response from the head-teachers or where the phonetic rendering was deemed too imprecise. This campaign lasted until c. 1985.

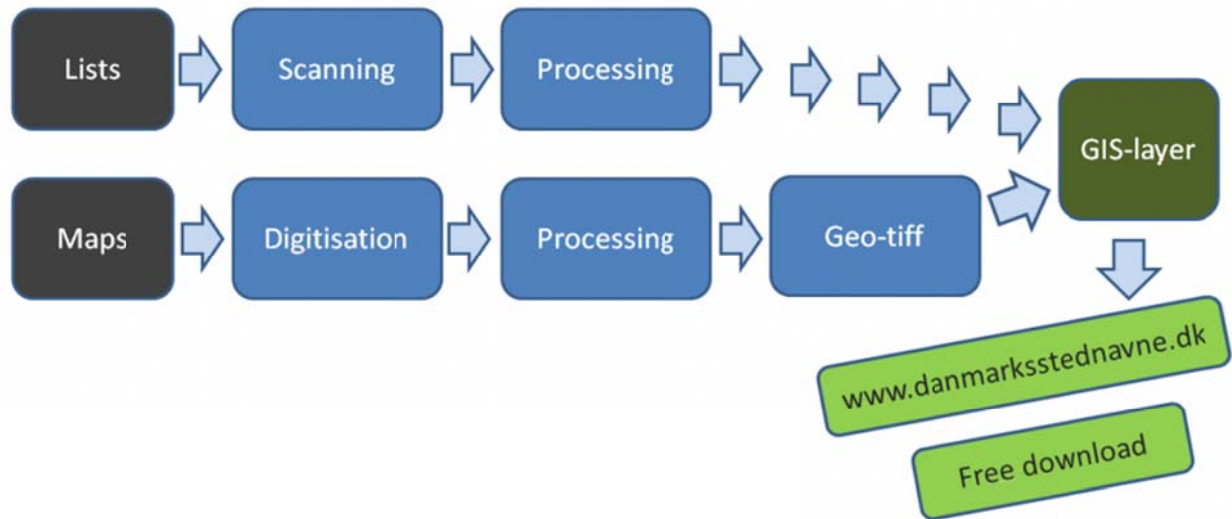
The Geographical Names Collection generally fulfills recommendation B (Collection of geographical names) of UNCSGN resolution I/4 (National standardization) and is a crucial tool for fulfilling the Place-Name Commission's obligations towards UNCSGN resolution VII/5 (National standardization based on local usage).

The digitization process

Before venturing into this digitization project, considerable time was spent on preparing the project. Digitization projects are sometimes thought too big at the outset and without eye for the usage of the digital result. Another thing which needs serious consideration is: What is the reason for digitization? Is the digitized object going to be part of the digital outcome, or is it just a medium for further digitization? If the object is part of the result, for instance a map or an illuminated manuscript, then the quality of the digital object is of prime importance. If the digitized object is just a medium, however, then quality is less important.

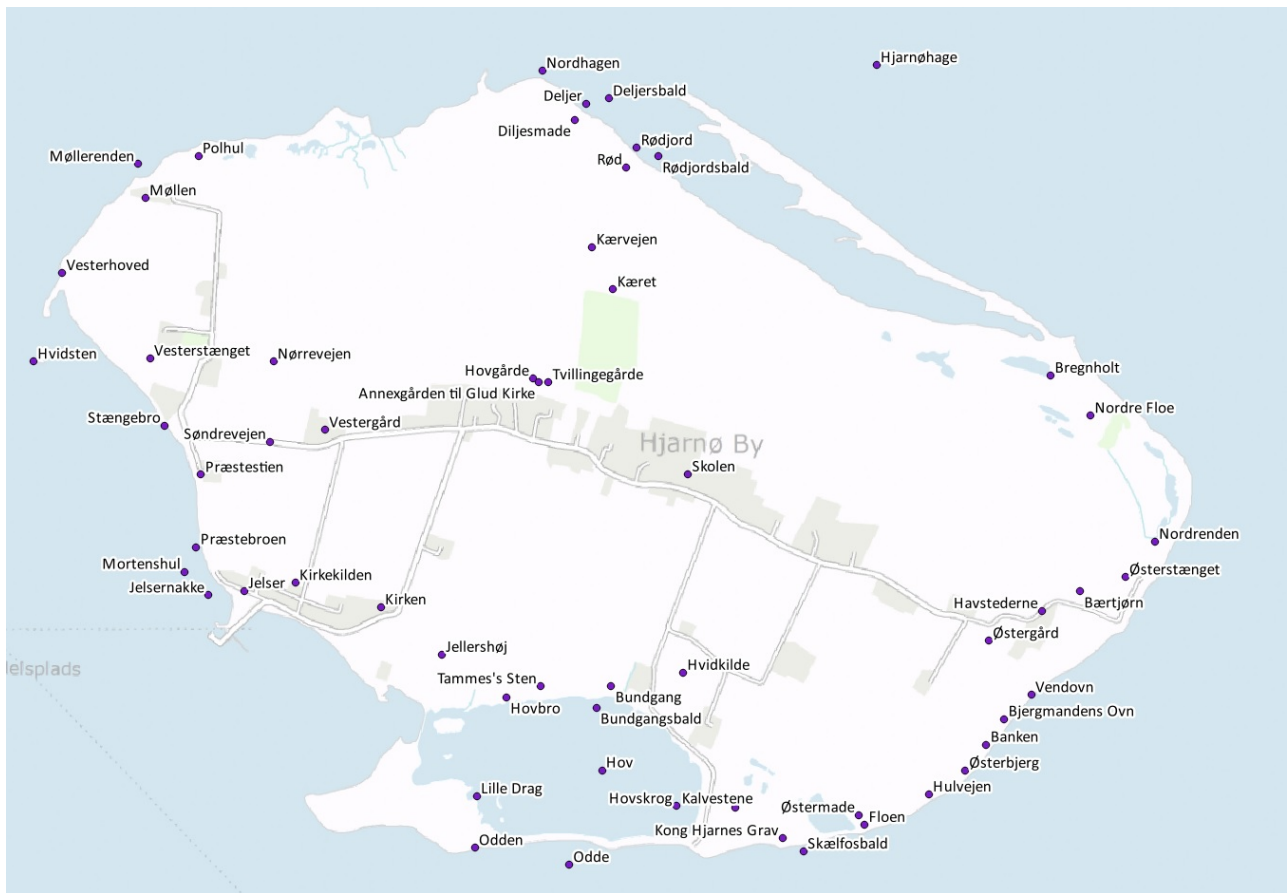
With the current project, the digitized objects are just a medium for a fully digital, geo-spatial dataset. However, the layout of the project means that the objects digitized – parish maps and record lists can be utilized for research purposes throughout all stages of digitization, thus reducing the risk of failure greatly.

Digitising the Geographical Names Report Collection is interesting, in as much as there are many processes to get right for the project to be a success. First of all, it requires a merger of two different types of media (record lists and maps) to create a third type of media (geo-spatial research data). This means that the transformation process is not just a digitization it is also a vectorization and geo-referencing. The total process looks something like this:



The record lists need to be scanned and processed before vectorized. In reality this means that all lists are divided into the smallest topographical unit possible, i.e. according to the above mentioned topographical system used at the Name Research Section. The same is the case with maps, they have to be associated the same topographical system, as it is then possible to merge the two datasets in the creation of a vectorized GIS-layer.

There is an additional layer of work with maps, as compared with the record lists, on the path to a vectorized and geo-referenced dataset. This is due to the fact that the individual maps have to be placed maps geographically correct in GIS (Geographical Information System), whereby vectorisation of the data from the maps – i.e. the hand-written numbers written on the parish maps – is made possible. The finished result should look something like this:



The numbers from the report maps are geo-referenced and the related data from the report lists is added to form a joint dataset. The advantages will be that there is no need to consult both a map and a list in order to obtain data. In addition, the data will be available for all to use, free of charge.

The plan is partly to merge the vectorized data with Name Research Section's online application, [Danmarks Stednavne](#), and partly let the full dataset itself be available for free download through a tailored web user interface. A Web Feature Service (WFS) interface is also under consideration. The goal is to complete the project before 2020, by means of crowdsourcing. When finished, the digitized Geographical Names Records Collection will mean that the Name Research Section will be able to put into practice UNCSGN resolution VIII/10 (Toponymic data collection procedures).

Project Status

Having started in early 2016, the project is far from concluded. However, progress has been rapid and scanning of record lists and digitization (digital photography) of the record maps is concluded. This amounts to approximately 50,000 pages of record lists and just under 2,500 record maps. Each map contains anything from a handful of numbers to in excess of 1,000, each one referring to a local geographical names. It is not known exactly how many names the collection houses but it is believed to have in the region of 500,000 geographical names.

The processing process is currently ongoing – approximately a third of the record lists and maps have been processed. In addition, the most viable means of geo-referencing record maps is also being tested as this report is being written.

All record maps are planned to be georeferenced by the end of 2016, after which the actual data vectorization may begin. Hitherto, the project has been entirely financed inhouse, and the fulfillment of the project cannot be done without a substantial amount of crowdsourcing.