

## **PART TWO**

# **BUILDING AND SUSTAINING DIGITAL SPATIAL INFRASTRUCTURES: CHALLENGES AND SOLUTIONS**



## Chapter 6

# PLACE-NAME DATABASES: A SPATIO-TEMPORAL MESS

Peder Gammeltoft\*

THE systematic USE of place-names is not a new thing. Ever since the first geographies, maps, and gazetteers, place-names have been a main source of localization. Until recently localization was usually only indirect and stated by association with a certain administrative unit or country. It was not until the so-called “spatial turn” that direct location—coordinate-based geodata—came to be considered important. Now direct location is of scientific and societal importance and in everyday use everywhere. In digital onomastics, however, geolocated place-name data are only just beginning to play a serious role as master data for cultural and historical research, as well as for research into linguistic data with a geographic profile. There are several reasons, but this is largely the result of what has been described as the “nature” of place-names: linguistic and locational complexity.

This chapter shows that place-names may have multiple references in terms of both *place* and *name*. No one has previously focused on this methodological and theoretical problem, although the examples given are common to geolocation-oriented digital onomastics. The solution is to introduce the notion of a unique place-name concept identifier to geographical datasets to enable the linking of the same place-name form across feature locations. This will enable coordination between multiple features with the same place-name origin, strengthen place-name standardization management, and, not least, more exactly represent place-name data across time, space, and domains of usage.

Place-names constitute one of the more important means of communication, as they furnish us with spatial “anchors” around which we construct our lives, identity, and history. Without the possibility of forming a mental map of reference to understand our relation with other places, people, and events, we would not be able to place ourselves in relation to the world in which we exist. Place-names give us a location framework with

---

\* Peder Gammeltoft is Scientific Manager of the Norwegian Language Collections (*Språksamlingane*), the home of three of the largest dictionaries and the most substantial dialect collections in Norway, as well as all the main place-name collections and many Old Norse resources. Previously Gammeltoft was Associate Professor in Name Research at the University of Copenhagen, focusing on place-names, in particular Scandinavian place-names in the British Isles, Normandy, and the North Atlantic. He is Norwegian representative of UNGEGN (United Nations Group of Experts on Geographical Names) and he is a driving force in place-name digitization in Denmark and Norway. Email: peder.gammeltoft@uib.no.

which to situate us in relation to what we learn, experience, and realize. It is, therefore, somewhat surprising that place-names do not loom very large in either linguistic research or in the field of human geography.

In linguistic research, place-names are generally seen as being uncomplicated bits of language belonging to the category of proper nouns that mainly function as “filler” in conversation. In human geography, place-names are often seen either as simple attribute data<sup>1</sup> or as fuzzy locators.<sup>2</sup> Nonetheless, place-names play a significant role in the humanities to express indirect location,<sup>3</sup> and the systematic utilization of place-names in geography has been in existence since the first geographies, atlases, maps, and gazetteers.

Within the emergent fields of computational humanities and spatial humanities, place-names are realized to be important geolocators, particularly within corpus research and socio-spatial text marking.<sup>4</sup> Here, a special research field dedicated to named entity recognition (NER) is in full development.<sup>5</sup> Names and textual references to people and places are subject to specialized mark-up to tease out the social and spatial framework of a text or group of texts. This field is still maturing, but place-names are recognized to be quite complex data elements. They are difficult to distinguish from other NER object types: personal names.<sup>6</sup> This is because place-names can be created from personal names, such as the state of *Victoria* in Australia (named after the British Queen Victoria), while personal names and personal designations may be based on place-names, such as the Duke of *Edinburgh*, *Milton Friedman*, and Ernest *Hemingway*. An even greater problem, however, is stability of writing—or, rather, the lack of it—over time.

## Place-Name Aspects and Data Integration

Our notion of language, particularly written language, is that of stability, guided by norms and conventions, often even national standards. An official standardized written norm as a concept generally harks back only to the late nineteenth century, however, and place-name standardization and dedicated bodies regulating the spelling of place-names are generally a twentieth-century phenomenon. Thus, place-name spellings may vary considerably over time, and a medieval written manifestation of a place-name may often differ incomprehensibly from its modern form. Although this has partly been remedied by means of the science of toponymy and the publication of place-name volumes containing individual place-names and all their temporal

<sup>1</sup> European Commission, “INSPIRE Data Specification,” 20.

<sup>2</sup> Jones and Purves, “Geographical Information Retrieval,” 216, 221.

<sup>3</sup> Ell, “GIS, e-Science and the Humanities Grid,” 148–49.

<sup>4</sup> See Dunn, this volume, 217.

<sup>5</sup> Won, Murrieta-Flores, and Martins, “Ensemble Named Entity Recognition (NER).”

<sup>6</sup> Ruokolainen and Kettunen, “Name the Name,” 152.

manifestations, the integration of place-names into digital spatial infrastructure has never been entirely successful.

What has been lacking in digital spatial infrastructures so far is a recognition that a “deeper” reading of place-names and their nature is essential. By using the concept of “deep mapping,”<sup>7</sup> it is possible to create a social and spatial narrative of the historical and contemporary human experience of a place-name through a limitless range of sources.<sup>8</sup> To achieve this, there are three aspects of place-names that must be combined correctly to achieve this: *location*, *reference*, and *time*. If these aspects are addressed correctly, we will achieve a major leap forward in the integration of spatiality and quantitative research into humanities—as we will finally be able to perform sophisticated space-time queries on large amounts of geographical, statistical, and literary data through dedicated digital spatial research infrastructures. For examples, see [Chapters 8 and 9](#) for instances of space-time integration in digital humanities infrastructures.<sup>9</sup>

When discussing spatial place-name databases and suitable data models, we must make sure that we have in advance taken the abovementioned core aspects of place-names (*location*, *reference*, and *time*) into account.

## Location

Before geographic information system (GIS) technologies became mainstream, place-name location was usually given as part of an administrative unit, or as labels on maps. Place-name location was thus usually indirect and did not give an exact reference to the locality, as expressed by Karen Kemp: “Direct locations are those that are stated in coordinates ... Indirect location uses references to other objects whose direct location is known. The two most used indirect references are place-names and addresses.”<sup>10</sup>

With the introduction of GIS technologies into the humanities, it became common to see direct location (geolocation) and indirect location combined in digital place-name data. Today geolocated place-names play an important role in digital humanities as master data for cultural, historical, and linguistic data with a geographic feature profile.

An untold premise of geolocation, however, is that it is direct-location-oriented, be it point, multi-point, line, or polygon. Geolocation furnishes a feature with an impression of certainty and definiteness through direct location. Even in research projects in which unknown or fictive places have been mapped in a GIS, geolocation embodies the results with this same sense of certainty. In this way, direct location is at odds with the general nature of place-names, which, although mono-referential in nature, are multifaceted and capable of signifying several referents at the same time.

<sup>7</sup> Bodenhamer, Corrigan, and Harris, “Introduction: Deep Maps and the Spatial Humanities,” 1.

<sup>8</sup> Ridge, Lafreniere, and Nesbit, “Creating Deep Maps and Spatial Narratives,” 177–84.

<sup>9</sup> See Dam, and Foka et al., both this volume.

<sup>10</sup> Kemp, “Geographic Information Science,” 44.

In traditional place-name publications, multifaceting is found where several feature types are given for one and the same name. For instance, in *Svenskt ortnamnslexikon*, the place-name *Kiruna* is given as “**Kiruna** kn, stad, Lappland,”<sup>11</sup> with *Kiruna* described as the name of a municipality (Swedish: *kn* = *kommun*), and at the same time a city (Swedish: *stad*) within a given part of Sweden (Lappland). Likewise, in the series *Danmarks stednavne*, the Danish place-name *Allerslev* is given as “**Allerslev**, lb., s.”<sup>12</sup> betraying the name as being used both of a village (Danish: *lb.* = *landsby*) as well as of a parish (Danish: *s.* = *sogn*).

Although both *Kiruna* and *Allerslev* are said to be referring to two distinct feature types—in each case a settlement and an administrative unit—this does not mean that localization is ambiguous, as the settlements are nested within the administrative units; see Figure 6.1. The boundaries of locations are ambiguous, but, given that localization is not direct (in the form of coordinates), this is not an issue. A precise determination of the exact location is purely contextual. It can be determined only from examining the communicative situation or the document in which the name occurs, through determination of the subject matter and text position near geographical references.

The function of indirect location is to refer to the area where the place-name is found. The type of locality of the place-name is normally implicit from the communicative situation and thus not directly stated. This function goes directly against the finiteness of direct location, which must have a specific geographical location and a specific feature type. In the cases where the place-name expression refers to only one geographical location, there is no problem with this model. The problem typically arises when administrative units assume the name of an existing local place-name expression. An example of this is the name of *Blåvandshuk*. Not only is it the name of the headland that forms the most westerly point of Denmark, it is also the short-form place-name of the local parish, and earlier also the name of the local municipality. In this case, “*Blåvandshuk*” is one place-name reference with three direct locations: a headland feature, a parish, and a historical municipality. Each direct location has its own distinct spatial extent and its own distinct feature type, and all have their own temporality. Only the communicative context may determine which exact location is meant by the linguistic expression “*Blåvandshuk*.”

Traditional geolocation cannot handle this ambiguity of localization in its data model, as localization must be definite. So, to represent reality, the traditional geolocation model must focus solely on direct location, establish three distinct entries, and consider everything else as being attribute data; see Figure 6.1. Each entry has its own feature type (one headland, one parish, and one historical municipality) and geometric object type (one point and two polygons, respectively), and shares same place-name attribute (*Blåvandshuk*). The combination of coordinates, feature type, and object type defines the

11 Wahlberg, *Svenskt ortnamnslexikon*, 176.

12 Jørgensen, *Danmarks Stednavne* nr. 26, 134.

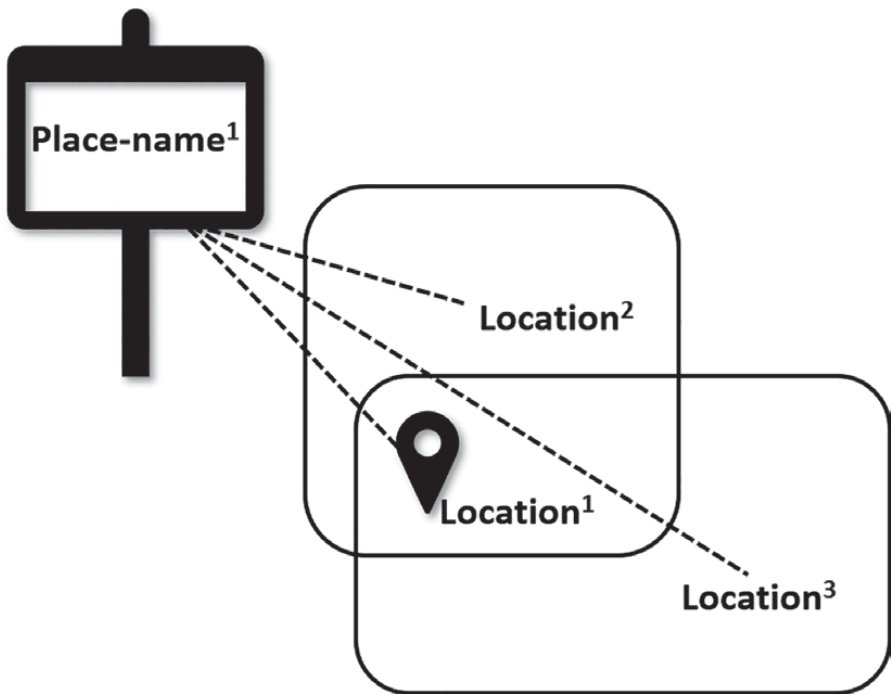


Figure 6.1. Schematic overview of indirect location as used in place-name research. The place-name expression is here used as the name attribute for three different locations. An example of this is the Danish place-name “Ringsted.” The name itself is an indirect location referring to three direct locations, as it is both the name of a town (Location 1), the short form of a municipality (Location 2), and a county (Location 3) in the period 1662–1748. In the first example, direct location is given as a point, whereas the second and third examples are administrative units represented by polygons. The third example, Location 3, represents a historical entity, whereas the other examples are current entities. Image by Peder Gammeltoft.

object and supplies it with “uniqueness”; the place-name expression is a mere attribute of no significant relevance to the direct location. In this model, the three occurrences of “Blåvandshuk” are individual entities, and the fact that they share the same place-name attribute form is purely coincidental. No effort is given to indicate a relationship of origin between the three Blåvandshuk localities, as this is not relevant to a direct location model of representation.

## Reference

Let us leave the problem of one place-name expression being able to refer to multiple locations for a moment, and concentrate on the concept of direct location, or reference. In its purest form, direct location is a set of coordinates and a set of attributes supplying us with the necessary information about the feature the object represents. The most

important attribute is probably one stating the feature type of the geometric object, followed by an official code, such as the official number 555 given to represent the above historical municipality of Blåvandshuk,<sup>13</sup> again followed by the place-name of the feature. This is a model that works well for most geolocations, particularly representations of administrative units. Since place-names are normally mono-referential—that is, a place-name refers to one feature only—this model is in principle on par with the nature of place-names. With language and geography, however, things are not as simple as that. Mono-referentiality does not exclude one feature from having several names, and nor does it exclude the same place-name expression to be used for several features. Depending on the cultural, linguistic, and ethnic make-up of the surrounding user groups, any feature may have several names. For instance, *Lomtjernet*, in Sør-Varanger in northern Norway, bears a Kven<sup>14</sup> name as well, *Kaakkurinlampi*. What is it we are seeing here?

In toponymic research, it is well known that one and the same feature may have several names. It may, as illustrated by Figure 6.2, be a result of several languages being spoken in the same region at the same time. This is probably the most common reason, but there are also other possibilities, such as independent naming of the same feature from different points of view, differences in user group focus, or naming because of differences in usage. The first possibility, independent naming, can be exemplified by the Norwegian mountain *Kråkvasstinden* in Oppdal, Trøndelag, which is also known as *Sandåhøa*. This mountain has a very steep face with a pronounced peak when observed from the northeast, but from the west and southwest it looks merely like a sloping hill. This has given rise to the parallel names of *Kråkvasstinden*, with the generic *-tinden* (“the peak”), and *Sandåhøa*, with the generic *-høa* (“the hill,” “elevation”). Both names are, incidentally, official names for this feature.<sup>15</sup> Differences in naming focus by different user group and differences in usage are most easily found in multicultural environments where the different cultural segments live by separate means of existence. If one cultural segment lives mainly by farming, another by fishing, and a third by transhumance, it is natural that naming will reflect what each cultural segment sees as important about a locality. An example of this is the three names of the north Norwegian settlement of *Mikkeli* or *Kersantti* in Nordreisa in Troms and Finnmark. The settlement’s Kven names reflect ownership of the settlement (once owned by Kersantin Mikkeli).<sup>16</sup> The Norwegian-language name of the locality, *Andsjøen*, is seemingly a name transfer from Trøndelag, presumably commemorating the “old

13 Nordic Group for Regional Analysis, *Regional Development in the Nordic Countries*, 139.

14 The Kven people constitute a Balto-Finnic ethnic minority in Norway. They are descended from Finnish peasants and fishermen who emigrated from the northern parts of Finland to northern Norway in the eighteenth and nineteenth centuries. Their language, Kven, is a mutually intelligible dialect of Finnish. In 2005 Kven received the status of a legal minority language in Norway within the framework of the European Charter for Regional or Minority Languages.

15 See <https://stadnamn.kartverket.no/fakta/861998> (accessed February 16, 2021).

16 See [www.kvenskestedsnavn.no/stedsnavn/view/8156](http://www.kvenskestedsnavn.no/stedsnavn/view/8156) (accessed February 16, 2021).

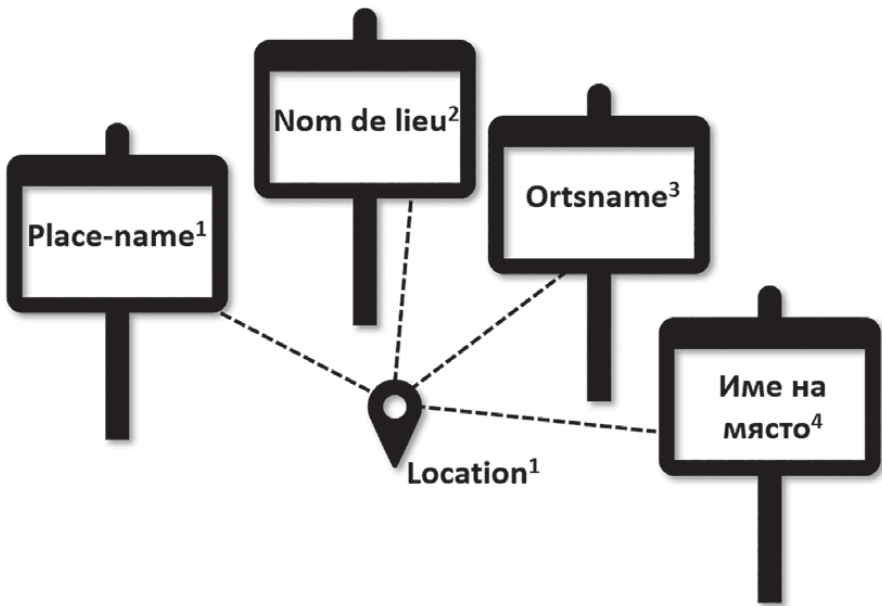


Figure 6.2. Example of direct location with multiple place-names in different languages. Direct location is here represented by a point location of a named feature. This direct location is known by multiple names, which is typical of multilingual areas. Here, Location 1 is known by four indirect locations in English (Place-name 1), French (Nom de lieu 2), German (Ortsname 3), and Bulgarian (Име на място 4). Image by Peder Gammeltoft.

home” of an earlier settler.<sup>17</sup> Here, only the Norwegian-language form has attained official status.<sup>18</sup>

The traditional way to handle this in geolocation is twofold. Either you can add as many localities to a feature as there are place-names, or, more often, additional attribute fields containing variant names are added. With either solution, there are several problems. The first, one location per place-name, creates a lot of feature doubling. The result is a geolocation system that is complicated and difficult to manage. It is impossible to ascertain if the place-names belong to one and the same feature, or if we are really dealing with several different features (with different place-names). In this case, the identification of several place-names to the same topographical feature relies on being able to equate coordinates and feature types, or through internal cross-referencing attributes. With the second possibility, there are no problems in establishing a unique object for a feature, as all variation is handled in the attribute fields. The problem here,

<sup>17</sup> Rygh et al., *Norske Gaardnavne*, 17:187.

<sup>18</sup> See <https://stadnamn.kartverket.no/fakta/929739> (accessed February 16, 2021).

however, is one of the retrievability and gradation of status of alternative forms: which are official, and which are not? How are several alternative forms handled: all place-name forms placed in one additional attribute field, or separate fields for each possible alternative form? In addition, placing alternative forms in a separate attribute field automatically establishes the entry in the main place-name form attribute field as the more important in relation to any other alternatives, even if they have co-official status. This second type of direct place-name field and “variants” field is a typical feature of, for instance, the GeoNames datasets, where, for example, the Moldovan place-name of *Chișcăreni*, in its “Alternate names” field, features the following alternatives: *Chiscareni*, *Chișcăreni*, *Kishkaren’*, *Kishkareni*, *Kishkareny*, *Kishkaren’*, *Kishkaryany*, *Kishkereni*, *Lazo*, *Лазо*.<sup>19</sup> It is clear that there are at least two distinct place-name forms, *Chișcăreni* and *Lazo*, each with multiple name forms. Whether these are place-name forms in different languages or different temporal manifestations, or represent different written standards, is entirely impossible to figure out. It is also impossible to see which forms are official place-name forms and which are not.

From the examples in the subsections above, it is clear that place-names can have multi-referentiality for both *location* and *reference*. No one has seemingly focused on this theoretical problem previously, although, to the user of geolocation-oriented digital onomastics, this is a constantly recurring challenge—and one threatening the sustainability of spatial onomastic infrastructures. In database terms we are dealing with a many-to-many relationship between location and reference—between place and name, so to speak. Fortunately, it is possible to overcome this with a well-structured relational database or web-based semantic data model. It is not the only challenge of spatial humanities data, though. We must also consider the time aspect: temporality.

## Time

In spatial humanities, the temporal aspect is extremely important: as soon as we have to query anything more than just a handful of decades old, we run into changes in language—changes that only accelerate the further back in time the study goes. A recent French digitization and XML mark-up project used NER technology to mark up geographical information from French sixteenth- and seventeenth-century texts. The main challenge was the problem of differences between seventeenth-century spelling and modern French.<sup>20</sup> The project’s solution was to modernize the text prior to mark-up.<sup>21</sup> The problem is, however: how do you know how to normalize earlier place-name spellings to modern spelling? The answer is, of course, only with great difficulty—unless you already have full knowledge of previous spellings of any given name. For languages with a long-established written norm or tradition (such as French and English), this is

19 See <https://download.geonames.org/export/dump/MD.zip> (accessed June 21, 2020).

20 Kogkitsidou and Gambette, “Normalisation of the 16th and 17th Century Texts,” 28–32.

21 Kogkitsidou and Gambette, “Normalisation of the 16th and 17th Century Texts,” 32.

less of a problem as compared with languages with short-duration written standards, a high degree of allowed (dialectal) variation, or no established written standard.

Written standards emerged only relatively recently. For most countries, written standards usually do not go further back than a century or so. Prior to this people had a more free or open attitude to writing—one that relied on several socio-cultural factors, fashions, and personal preferences, combined with factors such as personal background and personal written competence. In the sixteenth and seventeenth centuries, for instance, writing was often ornate and opulent, adding as many characters to a word as at all acceptable. For example, the Danish manor of *Overgård*, Randers Kommune, had no fewer than fourteen different spellings of its name over a 150-year period from the middle of the sixteenth century to the end of the seventeenth, ranging from forms such as *Owergaard*, *Ofuergord*, *Ofvergaardt*, and *Offuergaardt* to *Offuergaardtt* and *Offvergaardtt*.<sup>22</sup> Similarly, in Norway, the island and farm of *Io* is historically spelled *Hia*, *Iha*, *Ju*, *Hin*, *Iden*, *Ien*.<sup>23</sup> The current, regulated form is now *Io*. Since 1886, however, when Norwegian place-name normalization began in earnest, it has had regularly occurring written variant forms: *Io*, *Ien*, and *Iden*.

These examples show, that to understand and use place-names in a temporal dimension, it is necessary to abandon our modern notion of linguistic stability. Historical linguistics has taught us that language is in continuous change and is always evolving.<sup>24</sup> But what is not relayed to a great extent is how individual tastes and general socio-cultural trends also influence what we see in historical documents.<sup>25</sup> Place-names seem to be much exposed to variety, perhaps even more than the ordinary word stock. At the same time, we also must abandon the notion of written stability over time; this is entirely a modern feature, derived from the notion of a standard written form; see Figure 6.3. It is clear in the above Norwegian example of *Io* that place-names both adhere and do not adhere to standard written form even after a written standard is established. Some sources render the correct authorized form (*Io*), others use earlier established forms (*Iden* and *Ien*). Instead of viewing spelling as a time span feature, it is better to see spelling as individual time instances, almost as small linguistic “flashes,” reliant on the individual writer’s preferences, present norm, traditions, and source-internal factors (such as reliance on earlier sources, type of commission and payment for the assignment), among others. This also means that every place-name may have written forms, historical as well as present, that have hitherto not been collected. The more forms collected for each name, however, the greater the certainty in assigning correct location to a place-name.

<sup>22</sup> Overgård, Randers Kommune. Not yet published in *Danmarks Stednavne*: <http://danmarksstednavne.navneforskning.ku.dk/?deeplink=8be2494cdbae452d9790257bb3889f32> (accessed February 16, 2021).

<sup>23</sup> Rygh et al., *Norske gaardnavne*, 11:379.

<sup>24</sup> Hauser, Chomsky, and Fitch, “The Faculty of Language.”

<sup>25</sup> See Petrulevich, this volume, 190–98.

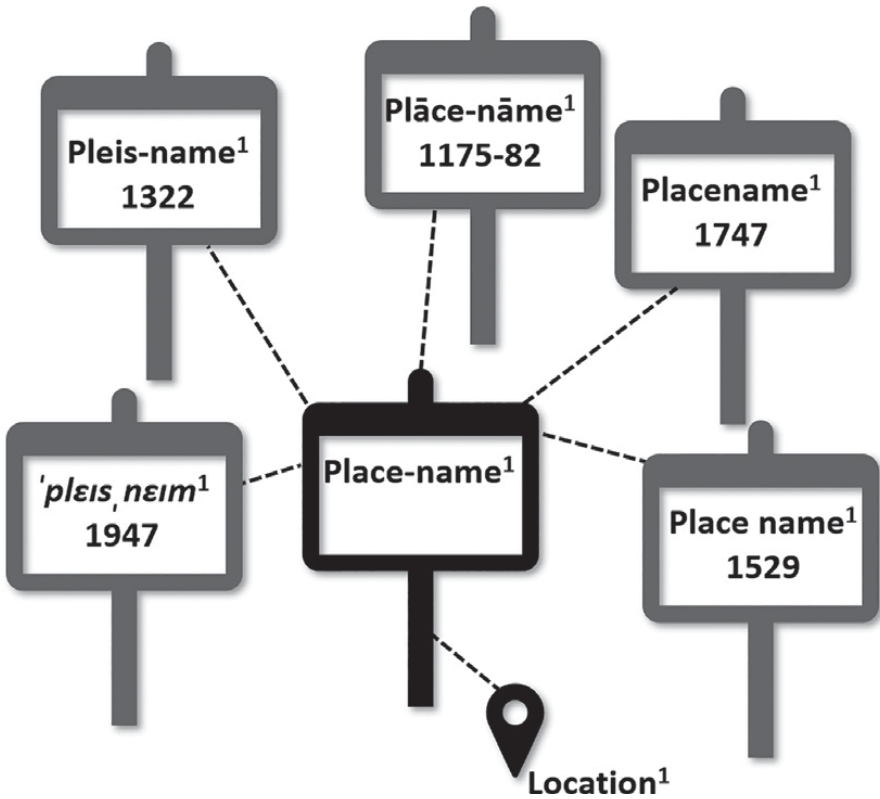


Figure 6.3. Schematic example of current and previous spellings of a place-name. In this example, direct location (Location 1) has one indirect location in the form of a current place-name, here represented by the black road sign, Place-name 1. This place-name has a number of temporal manifestations associated from 1175–1182, 1322, 1529, 1747, and 1947, here represented by road signs. These temporal manifestations are all expressions of one and the same indirect location but associated with a direct location (Location 1) though the relation of the direct location to a current indirect location (Place-name 1). See also Table 6.4 for a schematic representation. Image by Peder Gammeltoft.

### Location, Reference and Time: Can They Be Handled Together?

As the previous sections have shown, place-names have very complex space-time profiles. As illustrated, one place-name form may refer to one or more localities, and, similarly, one locality may be bestowed one or more place-names. In addition to all this, a place-name may also have varying spellings over time, occasionally of great variation. Not only must many-to-many relations be handled between location and place-name; when adding the temporal factor to spatial place-name databases, the complexity increases to having to handle many-to-many-to-many relations.

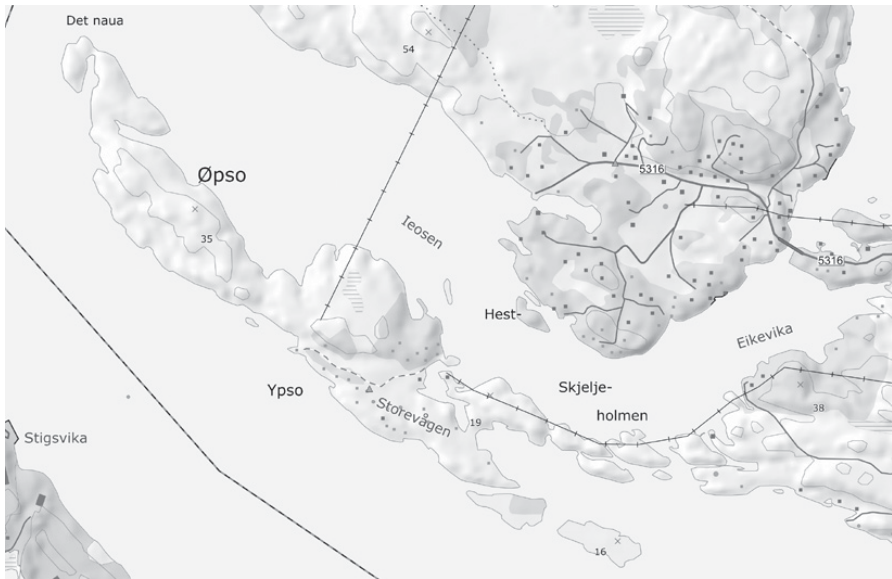


Figure 6.4. Map showing result of not stating a same-origin place-name relationship between localities/feature types. If a place-name geodata data model does not have an association between localities of the same name origin, the possibility that official spelling updates are not synchronized between all same-origin place-name expressions is extremely likely, as in this case. The island name “Øpso” and the settlement name “Ypso” are different modern manifestations of one and the same indirect location, now represented with six individual direct locations—Ypso, a settlement area, and Øpso, an island—as well as a four single settlements (shown as orange dots). This mistake has recently been corrected. Image ©Kartverket, CC-BY 4.0.

So far, this problem has never been solved nor addressed satisfactorily. There may be several reasons for this. First, spatial place-name databases are not very common, and precise localization is not very important for them. Similarly, spatial infrastructures of official organizations, such as national mapping agencies, are usually not concerned about the origin or history of their location attribute data, including place-names. By not paying heed to the place-name stock in a spatial infrastructure, however, you end up in a situation where you run the risk of re-authorizing the written norm for one place-name locality without setting the same norm for what is clearly the same name, but of another locality (and with a different feature type). A recent example of this was seen with the name of an island on the west coast of Norway, *Øpso* (which, by the way, is the neighbouring island to *Io*, mentioned above). On the official online map of Norway, Norgeskart<sup>26</sup> (see Figure 6.4), the name of the island, *Øpso*, is written in the

<sup>26</sup> See [www.norgeskart.no](http://www.norgeskart.no) (accessed February 16, 2021).

upper centre of the page.<sup>27</sup> Also on the map, in the lower centre, is the name *Ypso*, which was previously the official name of the main settlement area of the island.<sup>28</sup> Zooming in reveals that one farm settlement (also a cadastral unit),<sup>29</sup> and three individual farm units, go under the name of *Øpso*.<sup>30</sup> When checking with the Statens kartverk, it turns out that the place-name form *Ypso* was incorrect. It was an earlier authorized form, but had been re-authorized as *Øpso* in 2004, alongside all other localities with this name. It was promptly corrected by the Statens kartverk.

Why the settlement area name was left out of the 2004 authorization is irrelevant. What is important is to recognize that, by not having control of the localities that share identical place-name concepts, it is very easy to miss place-name forms that should have been re-authorized. In other words, place-name management becomes very accident-prone. Mistakes such as this one will inevitably lead to confusion and go against best practices for place-name standardization. According to the principles set up by the United Nations Group of Experts on Geographical Names (UNGEGN), and expressed in resolution UNCSGN I/4,<sup>31</sup> standardization should be in adherence to spelling rules and variant forms should be avoided.<sup>32</sup>

The way to solve this is, in principle, simple. An additional attribute needs to be assigned to the spatial dataset—i.e., a place-name-expression-specific ID. When dealing with attribute IDs, it is necessary to know what they stand for. In principle, there are two kinds of IDs available. There are identifiers that represent, or signify, an eternally constant value; and then there are identifiers that signify a concept, whose appearance, size, or name may vary over time or from dataset to dataset. In other words, the first type identifies a constant, whereas the second type identifies a variable merely acting as a placeholder. Both types are unique values, but their application varies. Coordinates are usually considered to be constants; should a set of coordinates be changed by a location being moved, the coordinates automatically gain a new value. Whatever the coordinates represent is constant for as long as the values are unmodified. In reality, however, a set of coordinates is usually also represented by its own attribute ID, so as to keep control of the feature the coordinates represent. The ID for an administrative unit is constant, but the object—or concept—it signifies is not. There will be constant revisions to its extent, and possibly also to its name label(s). But any changes to the signified will not have any

27 See <https://stadnamn.kartverket.no/fakta/6927> (accessed February 16, 2021).

28 See <https://stadnamn.kartverket.no/fakta/287108> (accessed February 16, 2021); corrected to *Øpso* after April 1, 2021.

29 See <https://stadnamn.kartverket.no/fakta/703656> (accessed February 16, 2021).

30 See <https://stadnamn.kartverket.no/fakta/97652>, <https://stadnamn.kartverket.no/fakta/218626>, and <https://stadnamn.kartverket.no/fakta/703655> (all accessed February 16, 2021).

31 UNCSGN Resolutions, I/4, [https://unstats.un.org/unsd/ungegn/documents/RES\\_UN\\_E\\_updated\\_1-11\\_CONF.pdf](https://unstats.un.org/unsd/ungegn/documents/RES_UN_E_updated_1-11_CONF.pdf) (accessed February 16, 2021).

32 See also UNCSGN Resolutions, IV/2. [https://unstats.un.org/unsd/ungegn/documents/RES\\_UN\\_E\\_updated\\_1-11\\_CONF.pdf](https://unstats.un.org/unsd/ungegn/documents/RES_UN_E_updated_1-11_CONF.pdf) (accessed February 16, 2021).

effect on the ID, as it functions as a placeholder. Even though we think of IDs as signifying constants, most IDs act as placeholders for concepts capable of variation.

Place-names, as discussed above, are also variable in time. So, any place-name ID must be seen as a time-variable ID placeholder. The dynamics of location-specific ID variables and place-name specific ID time variables are explored in the following sections.

Traditional spatial place-name databases, whether from national mapping agencies or from research institutions, are generally of the type “one feature, one place-name.” In recent years, however, coinciding mainly with growing recognition of the rights of linguistic minorities, there has been a general awareness that one locality may have different written expressions (see above). Despite this awareness, it has seemingly escaped notice that the same name expression can be used for different localities and widely different feature types. The evidence of this is seen in the fact that spatial place-name databases always have unique identifiers for the locality itself, and usually for the feature type too. There is no identifier for the named expression, however—and certainly not one that transcends the individual geographical locality.

Since place-names are indirect locators,<sup>33</sup> the actual referent a place-name expression signifies is “fuzzy.” It is meant only to create a cognitive connection in the mental map of the name users. The “nature” or “being” of place-names is thus not to say exactly where the locality is or what it is but, rather, to establish a spatial framework within which to conceptualize what is being conveyed during communication. The fact that one place-name expression may refer to several different kinds of localities, or features, is irrelevant for the communication, and, if specification is needed, then additional markers may be supplied—such as “farm,” “village,” or “parish”—to point out a specific locality. The mechanisms behind this are related to the notion of *metonymy* (association by proximity), and, more specifically, *polyonymy* (association to multiple distinct, but related referents).

To illustrate this problem, let us venture to the small island of *Frøya* just off the coast of western Norway. On this tiny island we find several examples of the same name occurring for different localities and feature types; see [Table 6.1](#).

Three of the above six examples are settlements that have gained their name from a nearby natural feature by means of metonymy, be it an island (*Frøya*), a beach (*Stranda*; literally, “the Beach”) or a bay (*Ånnevik*). Thus, all six examples are distinct localities of different feature types, but there are only three distinct name expressions. Given the size of the island, barely 5.5 km by 7 km in size, it should be relatively simple to retain similar spellings for different features. Nevertheless, when operating datasets with some 1 million named localities and feature types from different data sources created, maintained, and used by different government and local government agencies, the issue quickly becomes evident: it is impossible to keep control of spellings. The above example

---

**33** Bucher, Tiainen, Ellet, Acheson, Laurent, and Boissel, “Data Linking,” 12.

Table 6.1 Examples of toponyms from Frøya, Norway, with similar place-name expressions across different feature types. The different direct locations are represented by individual “LocalityIDs” (10001 and 10002).

| LocalityID | Expression | Feature_Type |
|------------|------------|--------------|
| 10001      | Frøya      | Island       |
| 10002      | Frøya      | Settlement   |
| 10003      | Stranda    | Smallholding |
| 10004      | Stranda    | Beach        |
| 10005      | Ånnevika   | Farm         |
| 10006      | Ånnevika   | Bay          |

of *Øpso/Ypso* shows how the management of place-names can swing out of control if the entire place-name material of a given area is not considered.

The problem outlined is not just a trivial onomastic conundrum dreamed up in an academic setting. This issue relates to how we manage place-name datasets and how we secure and implement a uniform means of place-name management—and, ultimately, gain a uniform means of standardization. If we do not have an overview of if or when a place-name expression occurs as the name of different feature types—or possibly even in another dataset—how can we be certain that we are dealing with the right place-name and the right feature? The answer is: we cannot, unless we start looking at place-name expressions as attribute data on par with other attribute variables, such as administrative units and statistics. The solution is, in principle, simple.

# The Solution

The answer to this issue is to add a unique identifier (UID) to the geodata as well as a UID for the place-name concept variable (here: “NameID”). If this is done, then it is possible to control and monitor the place-name inventory. The “NameID” functions across time, space, and expression and acts purely as a placeholder. The “NameID” acts as a UID of a place-name as a conceptual unit, not as an ID for the named expression of the actual geographical locality. If there is a need to distinguish the written expression of one feature type from another feature type with the same “NameID,” this is simply done by combining the UID’s “LocalityID” with the “NameID.” In this way, 10005\_CCC is distinct from 10006\_CCC; see Table 6.2. (Similarly, should the need arise to distinguish across time or datasets, this can be accomplished by adding the combination of “Source” and “Year” [of the source expression] to the ID string; see Table 6.4). In addition, if certain feature types are being investigated, then this is sorted simply by including attribute data on the feature type in the query.

The advantage of having individual ID values for different variables in a geodata dataset lies in the transformability and scalability of data. Each dataset and individual

Table 6.2 Concept table featuring “LocalityID” and “NameID.” The different direct locations are represented by individual “LocalityIDs,” whereas different indirect locations are represented by individual “NameIDs.” Because the expression “Frøya” is one indirect location with two direct locations, it is given one “NameID” (AAA) and two “LocalityIDs” (10001 and 10002).

| LocalityID | NameID | Expression | Feature_Type |
|------------|--------|------------|--------------|
| 10001      | AAA    | Frøya      | Island       |
| 10002      | AAA    | Frøya      | Settlement   |
| 10003      | BBB    | Stranda    | Smallholding |
| 10004      | BBB    | Stranda    | Beach        |
| 10005      | CCC    | Ånnevik    | Farm         |
| 10006      | CCC    | Ånnevik    | Bay          |

data entry can exist in its own right, at the same time as becoming a subset of other datasets. The level of detail and the focus of information—location, reference, time, and source, or a combination of these—is determined by the way IDs are combined to make up the subset. If a focus on location is needed, then “LocalityID” must be included. If, on the other hand, the place-name is in focus, then the “NameID” is the guiding ID in the combination of new subsets. In this way, the same data can be used for displaying place-name geodata, either as used by national mapping agencies, when the focus is on displaying correct location, or for place-name specific data, as used in applications and database systems aimed at traditional onomastic research into the origin of place-names.

Not everything is perfect in the model, however. It cannot be used without insight into onomastics as a discipline and to understand the “nature” of toponyms, as outlined above. The consequence of this inclusion is that onomastics moves into the realm of geodata management for good; *name* has finally found its *place* in geodata, so to speak. The inclusion of a place-name concept ID (“NameID”) makes data transformation easier, but it does have a trade-off, namely the added complexity of determining which “NameIDs” belong together internally and across datasets.

The benefits of adding an ID for the place-name concept are considerable, however. They do not stop with the ability to control the spelling of the same name concept across localities or feature types. The “NameID” attribute can be extended to forming part of other toponyms whose names derive, fully or in part, from other place-name concepts. In the island of *Frøya*, there is an abundance of examples where name concepts form part of other name concepts. To take two of the previous name concept examples, *Frøya* occurs as the specific of *Frøya kyrkje*, *Frøyadalen*, *Frøynes*, *Frøyaskjera*, *Frøyagrunnene*, *Frøya-Skorpeflua*, and *Frøysjøen*, whereas *Ånnevik* features as part of *Ånnevikholmen*. By adding information on the composition of a toponym, it is possible to make the relationship between place-names and derived compound place-names explicit. This

Table 6.3 Concept table showing how “NameID” can be managed with toponyms compounded with place-name concepts of other toponyms. The indirect location Frøya (AAA) occurs as the specific in, e.g., the compound place-name “Frøyadalen” (“NameID” EEE), which is stated in the “NameID\_Composition” column. The individual compound elements are highlighted in bold.

| LocalityID | NameID | Expression       | Feature_Type | NameID_Composition                           |
|------------|--------|------------------|--------------|----------------------------------------------|
| 10001      | AAA    | Frøya            | Island       |                                              |
| 10002      | AAA    | Frøya            | Settlement   |                                              |
| 10007      | DDD    | Frøya kyrkje     | Church       | <b>AAA</b>   kyrkje                          |
| 10007      | AAA    | Frøya            | Church       |                                              |
| 10008      | EEE    | Frøyadalen       | Valley       | <b>AAA</b>   <b>dal</b> + def.art. sg.       |
| 10009      | FFF    | Frøyenes         | Promontory   | <b>AAA</b> /root   <b>nes</b>                |
| 10010      | GGG    | Frøyaskjera      | Sea-rock     | <b>AAA</b>   <b>skjer</b> + def.art. pl.     |
| 10011      | HHH    | Frøyagrunnene    | Shallows     | <b>AAA</b>   <b>grunn</b> + def.art. pl.     |
| 10012      | III    | Frøya-Skorpeflua | Shallows     | <b>AAA</b>   ØØØ                             |
| 10013      | JJJ    | Frøysjøen        | Fjord        | <b>AAA</b> /root   <b>sjø</b> + def. art. sg |
| 10005      | CCC    | Ånnevik          | Farm         |                                              |
| 10006      | CCC    | Ånnevik          | Bay          |                                              |
| 10014      | LLL    | Ånnevikholmen    | Islet        | <b>CCC</b> /root   <b>holm</b> + def. art.sg |

relationship is here conceptually shown in the column “NameID\_Composition”; see Table 6.3. When an element in a compound place-name is derived from another place-name, this is recorded by adding the “NameID” of the original place-name expression.

First, it is important to state the linguistic relationship between different name concepts, especially the role that one place-name concept has as part of another compound place-name concept. In Table 6.3, the compound divide is marked by a vertical line, or pipe. Both *Frøya* and *Ånnevik* occur as part of other place-name compounds in the table, forming the specific element of the new place-name constructions. The role of a specific is to point out what is special about a locality. When a place-name concept occurs as the specific, it is to signal the proximity of the compounded place-name to the place-name forming the specific element. The other part of a place-name concept, usually called the generic, is usually a word stating the type of locality carrying the place-name. Occasionally, place-names are compounded solely with place-names. An example of this is *Frøya-Skorpeflua*, which is the result of the compounding of two place-name concepts, *Frøya* (“NameID”: AAA) and *Skorpeflua* (“NameID”: ØØØ). Again, the specific element is used to state its proximity to the locality *Frøya*. The reason for this is to single out the name from a nearby—and similarly named *Skorpeflua*, usually compounded *Botne-Skorpeflua* for distinction.

It will often be necessary to include information about how a name concept forms part of another and how it is declined, etc. Depending on the type of language, the internal relationship between compound elements may be stated in different ways and may be expressed in the field “NameID\_Composition,” if needed. Place-name compounds such as *Frøyenes* and *Ånnevikholmen* contain only the root form of the specific element. This is marked with a suffixed */root* in the table. Likewise, the definite article is expressed as part of the generic element in *Ånnevikholmen* by the expression *+def.art.sg.*

Since this model does not have a specific focus, it also allows for several name forms for the same feature and to describe their internal relationship. In Table 6.3, “LocalityID” 10007 has two “NameIDs” and thus occurs twice. The primary place-name concept of “LocalityID” 10007 is *Frøya kyrkje* (“NameID”: DDD). It is what is usually termed an institutional name of a locality. As an institution, the church on this island is known both under its shorthand form, *Frøya* (“NameID”: AAA), as well as under its full institutional name, *Frøya kyrkje* (“NameID”: DDD). The latter is itself a compound of “NameID”: AAA and the Norwegian term for a church, *kyrkje*. The compound relationship is described in its “NameID\_Composition” field and readily allows for a greater understanding of the occurrence of two similarly “competing” place-name concepts for the same place-name locality.

So far, the temporal aspect has not been mentioned. Although temporal manifestations often relate to both a name and an object, it is important to be aware that the nature of historical sources does not always make it possible to establish the exact direct location that the historical form describes. Therefore, the important thing—as in traditional historical linguistic name research—is to align a temporal manifestation (also called source form and historical form) to the name, rather than any current location. Thus, it is best only to add a “NameID” to temporal manifestations and forms giving linguistic information, such as pronunciation; see Table 6.4. This allows temporal manifestations to be analyzed under one place-name expression, as well as to be related to one or more direct locations. This is possible through the association of a “LocalityID” with a “NameID.” Should a historical form be able to be associated with a direct location, it is always possible to assign it the “LocalityID” of the direct location in question, to hard-code it to its direct location.

The reason why it is important to align a temporal manifestation to the “NameID” is to ensure that the model is as close as possible to reality and to secure the scalability of the model. Since all place-name expressions are indirect locations with the possibility of referring to more than one direct location, the model needs to be scaled to accommodate this possibility. Therefore, all place-name expressions, be they current, historical, sound recordings, or phonetic renderings, should be related to a direct location (“LocalityID”) only by means of association with an indirect location (“NameID”). This ensures the usability of the data model for geodata purposes as well as catering for toponymic research. Most of all, though, it avoids the doubling of data in the database, as historical forms do not need to be entered multiple times, owing to multiple direct locations.

Table 6.4 Concept showing historical forms in the dataset. The different direct locations are represented by individual “LocalityIDs,” whereas different indirect locations are represented by individual “NameIDs.” Because the expression “Frøya” is one indirect location with two direct locations, it is given one “NameID” (AAA) three two “LocalityIDs” (10001, 10002, and 10007). The column “Status” states whether the expression is current, historical, or a pronunciation. Note that historical forms and pronunciation are related only to the “NameID,” and not also the “Locality ID.” This relation is given through the association of the “NameID” with a “LocalityID.” Date and source are given only for historical forms and pronunciation forms, as this information is not strictly necessary for current name forms.

| LocalityID | NameID | Expression   | Feature_Type | Status        | Date      | Source       | NameID_Composition |
|------------|--------|--------------|--------------|---------------|-----------|--------------|--------------------|
| 10001      | AAA    | Frøya        | Island       | Current       |           |              |                    |
| 10002      | AAA    | Frøya        | Settlement   | Current       |           |              |                    |
| 10007      | DDD    | Frøya kyrkje | Church       | Current       |           |              | AAA    kyrkje      |
| 10007      | AAA    | Frøya        | Church       | Current       |           |              |                    |
|            | AAA    | Frøien       |              | Historical    | 1919      | Rygh NG      |                    |
|            | AAA    | frøi'na      |              | Pronunciation | 1919      | Rygh NG      |                    |
|            | AAA    | Frøen        |              | Historical    | 1723      | Mat.         |                    |
|            | AAA    | Frøenn       |              | Historical    | 1603      | Jb.          |                    |
|            | AAA    | Ffrøø        |              | Historical    | 1516–1521 | NRJ. II 134. |                    |
| 10005      | CCC    | Ånnevika     | Farm         | Current       |           |              |                    |
| 10006      | CCC    | Ånnevika     | Bay          | Current       |           |              |                    |
|            | CCC    | Aannevik     |              | Historical    | 1919      | Rygh NG      |                    |
|            | CCC    | å'nevika     |              | Pronunciation | 1919      | Rygh NG      |                    |
|            | CCC    | Annevig      |              | Historical    | 1723      | Mat.         |                    |
|            | CCC    | Onneugen     |              | Historical    | 1667      | Mat.         |                    |
|            | CCC    | Arreugen     |              | Historical    | 1608      | Jb.          |                    |
|            | CCC    | Arneueigh    |              | Historical    | 1563      | Jb.          |                    |

## Conclusion

It is my hope that this chapter has displayed the clear benefits of implementing place-name concept variables to geodata as individual “NameIDs.” The “NameID” gives place-name datasets unprecedented control, transformability, and scalability of place at the same time as liberating place-name data from the straightjacket of geolocation. Granted, such an implementation presupposes knowledge of the “nature” of place-names, but the extra effort in terms of the acquisition of onomastic knowledge is easily offset by the increased usability of datasets across domains, time, and foci.

Some fields will be able to reap considerable benefits, such as place-name management and place-name standardization. By introducing onomastic principles to a geodata model, it is possible to move away from a strictly geo-oriented view of toponymic geodata to finally see them for what they are: multifaceted gems of location and communicative information.

One thing I have not touched upon, and which is outside the scope of this chapter, is its applicability across database system. The data model concept devised here is mainly inspired by web-based semantic data models, but it can equally well be used either with a traditional relational database system or in a hierarchical database environment. In both models the addition of another ID variable is a simple operation, but the introduction of a cross-feature unique place-name concept variable to geolocated place-name data is an effective way to create linkage across features—a feature that is much needed in computational humanities,<sup>34</sup> as this will enable linkage and coordination between multiple features with the same name origin and more exactly represent place-name data across time, space, and usage. Most importantly, however, it will enable a more flexible integration of place-name data into other data infrastructures—spatial or non-spatial.

## Bibliography

- “Andsjøen.” Faktaark, Statens kartverk. <https://stadnamn.kartverket.no/fakta/929739> (accessed February 16, 2021).
- Bodenhamer, David J., John Corrigan, and Trevor M. Harris. “Introduction: Deep Maps and the Spatial Humanities”. In *Deep Maps and Spatial Narratives*, edited by David J. Bodenhamer, John Corrigan, and Trevor M. Harris, 1–6. Bloomington: Indiana University Press, 2015.
- Bucher, Bénédicte, Esa Tiainen, Thomas Ellet, Elise Acheson, Dominique Laurent, and Sylvain Boissel. “Data Linking by Indirect Spatial Referencing Systems,” workshop report. Leuven: EuroSDR, 2019.
- Ell, Paul S. “GIS, e-Science and the Humanities Grid.” In *The Spatial Humanities: GIS and the Future of Humanities Scholarship*, edited by David J. Bodenhamer, John Corrigan, and Trevor M. Harris, 143–66. Bloomington: Indiana University Press, 2010.

---

**34** See Foka et al., this volume, 158–64.

- European Commission. "INSPIRE Data Specification on Geographical Names: Guidelines." Brussels: INSPIRE Thematic Working Group Geographical Names, 2009. [https://inspire.ec.europa.eu/documents/Data\\_Specifications/INSPIRE\\_DataSpecification\\_GN\\_v3.0.pdf](https://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_GN_v3.0.pdf) (accessed April 16, 2021).
- GeoNames. "Moldova Dataset." <https://download.geonames.org/export/dump/MD.zip> (accessed June 21, 2020).
- Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch. "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" In *The Evolution of Human Language: Biolinguistic Perspectives*, edited by Richard K. Larson, Viviane Déprez, and Hiroko Yamakido, 14–42. Cambridge: Cambridge University Press, 2010.
- Jones, Christopher B., and Ross Purves. "Geographical Information Retrieval." *International Journal of Geographical Information Science* 22 (2008): 219–28.
- Jørgensen, Bent, ed. *Danmarks Stednavne*, vol. 26, *Stednavne i Roskilde Amt*. Copenhagen: Copenhagen University, 2013.
- Kemp, Karen K. "Geographic Information Science and Spatial Analysis for the Humanities." In *The Spatial Humanities: GIS and the Future of Humanities Scholarship*, edited by David J. Bodenhamer, John Corrigan, and Trevor M. Harris, 31–56. Bloomington: Indiana University Press, 2010.
- Kogkitsidou, Eleni, and Philippe Gambette. "Normalisation of the 16th and 17th Century Texts in French and Geographical Named Entity Recognition." In *GeoHumanities 2020: Proceedings of the 4th ACM SIGSPATIAL Workshop on Geospatial Humanities*, edited by Ludovic Moncla, Patricia Murrieta-Flores, and Carmen Brando, 28–34. New York: Association for Computing Machinery, 2020.
- "Kråkvassstinden"/"Sandåhøa." Faktaark, Statens kartverk. <https://stadnamn.kartverket.no/fakta/861998> (accessed February 16, 2021).
- Kvenske stedsnavn. "Mikkeli." [www.kvenskestedsnavn.no/stedsnavn/view/8156](http://www.kvenskestedsnavn.no/stedsnavn/view/8156) (accessed February 16, 2021).
- Natural Resources Canada. *Resolutions Adopted at the Eleven United Nations Conferences on the Standardization of Geographical Names: 1967, 1972, 1977, 1982, 1987, 1992, 1998, 2002, 2007, 2012, 2017*. Ottawa: Natural Resources Canada, 2019. [https://unstats.un.org/unsd/uneggn/documents/RES\\_UN\\_E\\_updated\\_1-11\\_CONF.pdf](https://unstats.un.org/unsd/uneggn/documents/RES_UN_E_updated_1-11_CONF.pdf) (accessed February 16, 2021).
- Nordic Group for Regional Analysis. *Regional Development in the Nordic Countries: NOGRAN Periodic Report 1994/95*. Copenhagen: Nordic Council of Ministers, 1995.
- Norgeskart. [www.norgeskart.no](http://www.norgeskart.no) (accessed February 16, 2021).
- Overgård. <http://danmarksstednavne.navneforskning.ku.dk/?deeplink=8be2494cdbae452d9790257bb3889f32> (accessed February 16, 2021).
- Ridge, Mia, Don Lafreniere, and Scott Nesbit, "Creating Deep Maps and Spatial Narratives through Design". *International Journal of Humanities and Arts Computing* 7 (2013): 176–89.
- Ruokolainen, Teemu, and Kimmo Kettunen. "Name the Name: Named Entity Recognition in OCR'd 19th and Early 20th Century Finnish Newspaper and Journal Collection Data." In *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference, Riga, Latvia, October 21–23, 2020*, edited by Sanita Reinsone, Inguna Skadiņa, Anda Baklāne, and Jānis Daugavietis, 137–56. Aachen: Aachen University, 2020.

- Rygh, Oluf, et.al. *Norske Gaardnavne: Oplysninger samlede til Brug ved Matrikelens Revision*. 18 vols. Kristiania (Oslo): W. C. Fabritius & sønners bogtrikkeri, 1910–1911.
- Vasardani, Maria, Stephan Winter, and Kai-Florian Richter, “Locating Place Names from Place Descriptions.” *International Journal of Geographical Information Science* 27 (2013): 2509–32.
- Wahlberg, Mats, ed. *Svenskt Ortnamnslexikon*. Uppsala: Institutet för språk och folkminnen, 2016.
- Won, Miguel, Patricia Murrieta-Flores, and Bruno Martins. “Ensemble Named Entity Recognition (NER): Evaluating NER Tools in the Identification of Place Names in Historical Corpora.” *Frontiers in Digital Humanities* 5 (2018). DOI: 10.3389/fdigh.2018.00002 (accessed February 16, 2021).
- “Ypso.” Faktaark, Statens kartverk. <https://stadnamn.kartverket.no/fakta/287108> (accessed February 16, 2021).
- “Øpso (1).” Faktaark, Statens kartverk. <https://stadnamn.kartverket.no/fakta/6927> (accessed February 16, 2021).
- “Øpso(2).” Faktaark, Statens kartverk. <https://stadnamn.kartverket.no/fakta/703656> (accessed February 16, 2021).
- “Øpso(3).” Faktaark, Statens kartverk. <https://stadnamn.kartverket.no/fakta/97652> (accessed February 16, 2021).
- “Øpso(4).” Faktaark, Statens kartverk. <https://stadnamn.kartverket.no/fakta/218626> (accessed February 16, 2021).
- “Øpso(5).” Faktaark, Statens kartverk. <https://stadnamn.kartverket.no/fakta/703655> (accessed February 16, 2021).

