Section 6 Technical issues: databases and pronunciation

Chapter 14 **Building a database for names management and index production for atlases**

Tjeerd Tichelaar

14.1 Introduction

The capstone of the commercial production of an atlas has for long been the compilation of an exhaustive index of geographical names. Falling outside the core discipline of the cartographer and typically subject to the toughest time pressure – ever occurring delays in the preceding cartographic process are rarely allowed to push backward deadlines imposed by the arrangements with the printing press – the production of the index of names was in pre-digital times known to be the most stressful, if not right-out dreaded part of the entire atlas production workflow.
Being both complicated and extremely labour-intensive, it was not uncommon for parts of the index production to be automated even before computers were called in to support the cartographic process. Whereas initial applications may have been limited to storage and sorting of names, it was the development of relational databases and SQL that would enable nearly complete automation of the editorial steps required for the production of a print-ready index of names.

14.2 The atlas and its index

The geographical contents of the interrelated maps constituting an atlas, labelled by geographical names, are unlocked in the average atlas by a number of search facilities. A list of contents is generally provided, along with a graphical index key to the maps. The most complete and detailed access portal to the objects presented in the maps is, however, the index of names. The names index of an atlas differs significantly from indexes provided in non-cartographic reference books and text books in general. Rather than to retrieve occurrences of text strings throughout the work, the entries in the atlas index are meant to localize named objects. The geographical name is not a concept one might be interested in for its own sake, but serves as a bridge to the object of our concern. We search the name to find the object it refers to. Once we localize it in geographical space, we have access to all the information pertaining to it in all the maps in which it occurs, which in the case of a school atlas or a national atlas may be a multitude of geographical and thematic maps at varying scales.

The consequence of the function of geographical names in the atlas, as explained above, is that we want just one index reference for each combination of object and name. As many names will occur in many different maps throughout the atlas, we must select the occurrence we believe to be most appropriate for localizing the object described by the name. This selection should be made in a consistent, systematic way, so that it is predictable to the user and automatable to the producer.

To effectively isolate the searched object, it may not always suffice to quote its name in the index: homonyms are quite common, and these must be made discernible. This can be done by adding object-specific information, which may be geographic (a country or region name) or generic (an object category or class).

The choices to be made to populate and structure the index of geographical names in an atlas and the requirements concerning data structure and definitions of the database employed can be illustrated by the example of the system employed for the Dutch high school atlas *De Grote Bosatlas* and its domestic and international spin-offs. This atlas has been published starting from the 19th century to support the geography curriculum at schools nationwide, with separate editions for primary schools and different levels of secondary education as well as foreign co-editions in Belgium, France, Denmark, Sweden, Switzerland and Canada. The cartographic bases of its constituting maps were reused wherever product specifications allowed. Production of maps and indexes was gradually digitized in the 1990s, and a strictly relational database for names management and index production was developed in 2004.

14.3 Index production: principles and rules

Traditionally, the Dutch school atlas presented two types of maps: oro-hydrographic overview maps providing a synoptic topographic reference, and thematic maps covering the themes dealt with in the contemporary geography curriculum. Due to the evolution of the demands of geography teaching over the last half century, the proportion of thematic maps was continuously increased. For each new edition, the thematic map content might be replaced, while the overview maps (the time-resistant 'backbone of the atlas'), would typically be updated and incidentally restyled. The geographical names in the overview maps were the first, and remained the most important to be listed in the index of names.
The compilation of the index of names for the Dutch school atlas has been given ample attention ever since it was first produced (1924). In its structure and composition, it follows long established conventions that were strictly maintained over the successive editions of the atlas. In the same way as a dictionary or a telephone directory, changes in composition or design can be expected to be unfavourably received by the customary

users: in this case the teachers in the schools. The rules to be followed are therefore meticulously defined.

These rules can be summarized as follows:

1. Every name-object combination (a name referring to a unique object) is represented in the index with one reference only. When a name-object combination occurs in the atlas more than once, the map referred to in the index is defined by a set of selection rules.
2. All name-object combinations occurring in the atlas are included in the index in a similar way and unabbreviated, regardless of whether in the maps referred to they are parenthesized (secondary names) or abbreviated.
3. The names are listed alphabetically, following the currently customary sequence of the alphabet of the language in which the atlas is published.
4. When a name consists of a specific element preceded by a generic element, the generic element, including prepositions where they occur, is inverted in the index. In other words: names are listed by their specific elements, irrespective of their position in the name. *Example*: Negro, Rio (a river name in which Rio is a generic element) vs. Rio de Janeiro (a city name, in spite of its etymological meaning, Rio is not a functional generic element here).
5. When a name consists of a specific element preceded by an article or a preposition, the article or preposition, is in the index inverted as well. (Example: Hague,The)
6. To differentiate between homonyms, regional and/or generic specifications are added to the names of homonymous objects. *Example*: Victoria (city, Canada) - Victoria (city, Seychelles) - Victoria (state).

Cross-references may be included in the index, for instance to accommodate for recent name changes or frequently occurring alternative writings. These refer to the map name using an =-sign; they are inserted into the alphabetical order of the regular names. *Example*: McKinley, Mount = Denali.

Index production: the selection of entries
The principles by which index references are selected in the Dutch school atlas can be summarized as follows:

1. Priority is given to the overview maps above thematic maps. As an exception, for country names priority is given to political continent maps.
2. Priority is given to the map showing the entire object (for instance in the case of rivers or mountain ranges).
3. Priority is given to the maps in which the named object is included in the map's subject area. The subject area is generally described by the map's title: New York is in the subject area of the maps titled 'United States' and 'North America', but outside the subject area of the map titled 'Canada', even if because of this map's incidental layout, the object is also displayed there. Where the map's title doesn't suffice, the subject area is explicitly defined internally by the editorial staff.

Priority is given to the map with the largest map scale, except in the case of country names (see rule 1).

14.4 A database for names management and index production

A well-designed relational database can meet both the demand of efficient index production and a more comprehensive objective to provide the ability to store and query all possible information on geographical names (names management). The rationale behind the latter is the observation that knowledge of or even affinity with toponymy cannot be expected to be widespread within the editorial staff responsible for the maintenance of the atlas, while individual editors might still be expected to answer questions about decisions on the rendering of geographical names taken in the past. In the Dutch case, the pseudo-institutional status the Bosatlas historically acquired in the Netherlands – by many it is considered the most trusted work of reference about foreign geography, including the confusing subject of name orthography – gives rise to a continuous stream of questions forwarded to the publisher on the contents of its maps, and the bulk of these involve geographical names. Moreover, when for whatever reason a new name has to be added in any of the existing maps – or a new map is to be made – the editors must have access to information about the principles (including transliteration keys) and sources the existing names were derived by.
From a practical point of view, the database should ideally meet the following conditions:

1. It should be object-oriented.
2. The data structure should be relational, and normalized to the greatest possible extent: in other words, every piece of information should occur only once in the database. Violations of this condition expose the database to contamination: errors and inconsistencies will be imminent.

3. It should be portable, in the sense that only commonly available market software should be used.
4. The database should be able to communicate with other cartographic databases that may be employed to produce both paper and digital maps. Meanwhile, the names and indexing database should remain independent.
5. Its maintenance should require no special technical skills.
6. Data safety should be sufficient.
7. The character coding should be Unicode.
8. The result of the index production process should be an output file (txt) holding all the information required for automatic page layout with commonly available DTP software.

The index production application should incorporate the principles and selection procedures we choose to follow when preparing our index. It can be built into the database by adding a script and all the required queries to the necessary tables.

14.5 The data structure: core tables

The core of the database should consist of two separate tables, one holding the names and another the objects. A name can, in this definition, refer to only one object: in other words, homonyms are considered different names. An object, on the other hand, can be referred to by different names. The table NAME consists of a key field (name identifier) and a number of attributes directly dependent of the name. The table OBJECT consists of an object identifier and a number of attributes directly dependent of the object.

To allow the index production application to collect the information it requires, the database must include tables containing relevant data about the atlas, the maps it includes and the position of these maps within the atlas, and the names and their location (grid reference) occurring in the maps:

14.6 What is a name?

NAME	
Field	*Data Type*
Name identifier*	Number
Object identifier*	Number
Language identifier	Number
Script identifier	Number
Orthography identifier	Number
Name (main specific element) *	Text
Non-specific name element identifier*	Number
Grammatical fields (gender, number etc.)	Number
Is the name an official name?	Yes/No
Is the name a primary name?	Yes/No
Does the name contain a generic element?	Yes/No
Is this a long /complete version of the name?	Yes/No
Is this a short /abbreviated version of the name?	Yes/No
Is this a redundant or historical name?	Yes/No
Source identifier	Number
Adjective (masculine singular) form of the name	Text
Pronunciation (IPA notation)	Text
Syllable of primary emphasis	Number

Table 14-1 the NAME table

The table NAME could contain the following fields (fields marked * are mandatory, fields in bold type are required for the index production process).The table contains one row for every geographical name we recognize. Its key attribute is a name identifier. Note, that the attributes collected for every name presuppose a subjective definition of ‘name’. A name in the context of the database is a *written* code referring to one single geographical object. This may be self-evident in the light of the purpose of the database, which is to help produce a printed atlas, but it is neither the only possible nor the original definition of a name. As writing is a relatively new addition to the systems of oral communication we refer to as languages, the original description of a name would be ‘a sequence of sounds conventionally used to refer to a single person or object’ or, in database terms, ‘... a single instance of a certain entity-type’.

The decision to define ‘name’ as a written instead of a sonic code has far-reaching consequences, some of which may be considered undesirable. It implies, for instance, that different writings/alphabet versions or transliterations of a single name are different names in database-terms, even when they refer to the same object and their pronunciation is per definition the same. This decision thus leads to imperfect normalization, as some attribute values (for instance Pronunciation) are shared between different written versions of the name: it will be a challenging effort to address this imperfection, by grouping together those written forms of a name that are essentially (pronunciation-wise) the same. On the other hand, the French name Paris for the capital of France (IPA: paˈʁi) and the English name for the same object (IPA: ˈpærɨs), which happen to be similar in writing but are certainly not so in pronunciation, are to the database one and the same name. The addition of the attribute holding the name’s pronunciation in IPA notation, makes it an obvious step to consider ‘Paris’ in English and ‘Paris’ in French two different instances of the entity type Name, in which case the language of the first is English, not French. In other words, we would consider Paris in English a sonic exonym to the homonymous French endonym in its written form. This is a matter of case-to-case definition: by defining Paris in its English pronunciation a name in its own right, we recognize it as being by convention part of the English language, and not merely a mispronounced French name. Just as is the case with written exonyms, the number of recognized sonic exonyms should preferably be confined to some well-known names in frequent use in the host language.

The question of whether an exonym in a certain language should be defined as a Name in its own right not just involves names that in their written form are homonymous to the endonym. Sometimes non-homonymous exonyms are in their written form borrowed from another than the locally official language. A representative example of this kind of situation is the French name for the Italian city of Firenze (endonym). The French exonym Florence is borrowed by several other languages, among them English and Dutch. The pronunciation of the English name (IPA: /ˈflɒrəns/) clearly differs from the French one, so English Florence and French Florence might each be considered separate exonyms. The Dutch pronunciation however sticks to the Original French one, so in this case it is debatable whether Dutch Florence is to be considered a Dutch exonym and thus a separate Name in the database, or the Dutch language merely borrows the French name. Whether or not a graphically homonymous and synonymous name should be considered a separate name, might be decided after the rules of the normalized database. These define that it can only be the same name if the values of all attributes are identical. A typically discriminating attribute would in this example be Adjective. In the case of Florence, its French adjective is ‘florentin’. The English adjective is ‘Florentine’. Dutch uses an adjective form ‘Florentijns’. The value of this attribute thence requires Dutch Florence to be included in this database as a name separate from the original French exonym.

Taking this reasoning a step further, local or any other viable pronunciation variants deviating from the standard pronunciation (if existent) of the name in the official language might likewise be included in the database as separate name. The name(s) of the American city of New Orleans, in ‘standard’ American English pronounced /nuː ɔrˈliːnz/ or /nuː ˈɔrliənz/ but locally /nuː ˈɔrlənz/ or more commonly /ˈnɔːlənz/ (IPA) sets a fine example, those of the Brazilian city of Rio de Janeiro (English pronunciation /ˈriːoʊ deɪ ʒəˈnɛəroʊ/ or /ˈriːoʊ deɪ dʒəˈnɛəroʊ/, in modern European Portuguese /ˈʁi.u dʒi ʒaˈnejɾu/, but in the local Carioca dialect of Brazilian Portuguese /xiudʒiʒaˈnei̯ɾu/) another.

Generic elements

A field defining whether or not a name contains a functional generic element will be required by the index production application if for a certain atlas we choose to add a generic specification to all names in index entries where this might be useful. If such a specification is textual instead of graphical (an icon), it will be redundant and hence undesirable to do this where the name itself already shows which generic category the object belongs to: for instance ‘Vänern *(lake)*’ makes sense, but in ‘Michigan, Lake (lake)’ the addition is redundant, as well as ‘Bodensee (lake)’ – the latter presuming that the meaning of See in the German language is familiar to the public, or is explained in a list of geographical terms.

Grammatical fields

Some grammatical details may be useful to store in the table NAME as well. Although they will rarely be relevant to the rendering of the name in the map or in an index, they do become relevant when names are used in written text or oral discourse: for instance, when they carry a declinable adjective or an article, or when it needs to be inflected when occurring in a particular grammatical case. For instance, the possibility to store grammatical number (singular, plural) and grammatical gender (masculine, neutral, feminine) can be easily accounted for.

In the case of grammatical number, geographical names incidentally behave in a way not matching the elements they were etymologically constructed from, even when these elements were left unchanged, are clearly recognizable yet, and call for the use of a proper article in speech: ‘United States’ and ‘Netherlands’, for instance, are plural names that in English usage in speech, but not in the map, go with the article ‘The’ – their French equivalents carry the plural article Les (États-Unis, Pays-Bas). In English speech, however, they are treated as singular words whenever they carry verbs: ‘the United States is a big country in America, the Netherlands is a small country in Europe’. Such knowledge could be stored in the database.

Much relevant grammatical information is language-specific, for instance lexical roots of names in languages that practice inflexion depending on grammatical case (Latin, Lithuanian), or definite-indefinite distinctions in languages like Turkish, Romanian and Albanian. An inventory of such issues per (relevant) language would be required to decide which additional attributes might be desirable to complement the database.

Adjectives

Adjective forms of geographical names are in many languages different from the nouns. Quite often they are not formed in a systematic way, sometimes they are even derived from predecessors of the current name or a different language version (e.g. Latin in Romance languages). They are often historically defined and not seldom influenced by non-official regional or local language. Homonymous names may thus carry different adjectives: the adjective going with the French city name of Beaumont in the department Manche, for instance, is 'beaumontais', while the city of Beaumont in the department Vienne carries the adjective 'beaumontois'. The adjective of Beaumont-en-Véron is 'bellimontois', while what belongs to Beaumont-sur-Sarthe is known as 'belmontais'. Knowledge of the adjective forms of a geographical name may be relevant if the name is to be used in prosaic texts, and adjectives may also be used to make demonyms.

Pronunciation

IPA notation (or, if the database format can accommodate this, an audio file) and accents are useful to know when names need to be pronounced. If more than one pronunciation is encountered, which in real life occurs more often than not, a decision has to be made whether this should be allowed for in the database. If pronunciation is not officially standardized, and it is possible to store more than one IPA string in this field, a case per case decision must be made whether or not we are still talking of the same name.

Syllabic stress (accent) may or may not follow systematic rules in a language. In Finnish, the stress (primary stress in the case of compound names) always falls on the first syllable of the word. In Italian, it is commonly the penultimate syllable which receives the stress, but under the influence of regional languages there are many exceptions to this rule. In some languages (for instance Spanish), diacritics are mandatory to mark syllabic stress not following the general rule, in others (Italian) they are optional and most commonly not used. In the database, a numerical field may be reserved to store information about the syllable receiving the (primary) stress.

Additional name attributes for names selection

Some other attributes included in the exemplary table NAME are not required for index production either, but may be useful for the selection of the right name versions as the map is compiled or revised: for instance, language, script, orthography, and the 'flags' (on/off-switches, stored as Boolean) official name, primary name etc.

Specific and non-specific name elements

If we want to be able to invert generic elements and/or articles and prepositions before alphabetizing the index output for certain atlases, these should be stored in a separate table in those cases where they precede the specific element of a name. In Dutch or French school atlases, for instance, the river name Río de la Plata is rendered as Plata, Río de la – while the town name Rio de Janeiro is simply rendered as Rio de Janeiro, because in this name the element Rio is not a functioning generic element even when it was derived from one. In Dutch school atlas indexes, articles are inverted as well, but in some Dutch-language Belgian primary school atlases they are not. The table holding the non-specific name elements should contain the following fields:

NON-SPECIFIC NAME ELEMENT	
Field	*Data Type*
Name element identifier*	Number
Name element (description)*	Text
Element type identifier	Number

Table 14-2 Non-specific name elements

The element type identifier should connect to a small table listing the options, for instance 'Generic', 'Article' and 'Article connected': the latter needs to be discernible by the indexing application because in some cases the element should not be separated from the rest of the name by a blank if it is not inverted in print (examples: elements ending in a hyphen or an apostrophe like the article al- in transliterated Arabic names or the contracted article l' in French; suffixes in Irish or transliterated Hebrew directly connected to the specific name like An t or Ha).

Language

According to the principles of the database, every Name belongs to a Language. The definition of Language in the sense of the database is not necessarily merely linguistic-taxonomical: a Language here may be a set of names belonging to a certain language or dialect in a historical context. The instances of the entity-type Language must suffice to store any geographical name we get hold of, allowing also for a lack of specific knowledge where it occurs. So, the table may include both very specific dialects like 'North Straits Salish (Lummi)', more general categorical instances like 'Occitan (local dialect)', 'Pre-

Columbian Native American (unspecified)' or 'British Indian (unspecified)', and even 'unknown'.

LANGUAGE	
Field	*Data Type*
Language identifier*	Number
Sorting system*	Number
Language (description)	Text
ISO-code fields	Text
Year of origin	Number
Year of end	Number

Table 14-3 Language table

The table LANGUAGE allows us to store information we might require when selecting the names that should appear in a map: the more we know, the better. It also plays a role in index production, as next to the table NAME, the table ATLAS also refers to it in connection with the addition of geographical or generic specifications. Moreover, the language of the atlas must be known in order to select the alphabetization rules applied in the output phase (sorting system), as these are language-specific. In Swedish indexes for instance, Å, Ä and Ö are inserted into the alphabet as separate letters after Z, while Ü is considered the same as Y and W the same as V; in the Danish alphabet, the sequence is Z-Æ-Ö-Å.

ISO codes of the languages and dialects may be included to facilitate communication with external parties. Data about the historical range of earlier language phases (the approximate years of origin and end according to the definition we apply) may be useful in the editorial phase of the compilation of a historical maps series, to avoid anachronistic names.

Name vs. Language

The starting point that any name can be assigned to one single language, as commanded by the data model of the database, poses theoretical problems that must be addressed in an unambiguous way. Debatable as it may be, for the practical use the database will be maintained for, the relationship should for each single name be explicitly defined by the Language and Orthography attributes assigned. Some examples are listed below to demonstrate both problems and solutions.

The language of the names of the American cities of Los Angeles, California (originally Spanish) and Terre Haute, Indiana (originally French) is English. They are the names by which the speakers of the official language of the United States of America, English, refer to them. In spite of their clearly recognizable Spanish, resp. French origin, in spite also of the fact that many inhabitants of the city of Los Angeles are till this day Hispanic, the English rules of pronunciation apply to them (IPA: lɒs ˈændʒəˌləs). The original Spanish name now functions as an exonym, insofar as Spanish is not considered a local language in the town: in the database, it will be a separate name (IPA: los ˈaŋxeles). In writing it differs by the acute accent on the A. Another difference is, that the element Los in

Name identifier	Object identifier	Language	Script	Orthography	Name (main specific element)	Non-specific name element
14282	79150	English (Modern)	Roman		Los Angeles	
14282	79150	Spanish (Castilian)	Roman		Ángeles	Los

Figure 14-1

the Spanish name is an article that might be inverted in the index, while in the English name it no longer functions as such (see figure 14-1): Likewise, the name by which the Romans referred to the settlement growing around their army camp where at present the Dutch city of Nijmegen stands, Noviomagus, may in the database be defined as Classical Latin because it was quoted by this name by sources written in Classical Latin language (figure 14-2):

Name identifier	Object identifier	Language	Script	Orthography	Name (main specific element)
206038	232601	Latin (Classical)	Roman		Noviomagus

Figure 14-2

Or it might, as historical linguists might advise, be defined as a Celtic name rendered in a Roman Latinised way (figure 14-3). The latter can be specified in the field 'Transliteration':

Name identifier	Object identifier	Language	Script	Orthography	Name (main specific element)
206038	132601	Celtic (Gallic-Transalpine Gaulish)	Roman	Roman Latinised	Noviomagus

Figure 14-3 language definition

Script

Unicode fonts are available to store and visualize names in all known writing systems. The description of the script should be stored in a separate table. Again, the instances maintained should accommodate for any name we encounter, also when specific knowledge as yet falls short. The table may consist of just a key field and a textual description. The purpose of including different script versions of names in the database, even when the atlases we produce won't show them in print, is that it allows us to store the original writings of names that we transliterate: we might need these in case official of UNGEGN-promoted transliteration keys are replaced and we need to re-transliterate.

Orthography

The definition of Orthography in the context of the database may include both transliteration, transcription, orthographic standards and optional or unofficially adapted variants like accentuated, vocalized or simplified spellings. Many instances may be unspecified and possibly unofficial/non-standardized transliterations and transcriptions matching the pronunciation-to-writing conventions of a certain language, but ISO-norms and transliterations recommended by the UNGEGN Working Group on Romanization should be included as well. An instance 'unknown' may be used temporarily until knowledge is gained. To avoid confusion, the target script may be included as an attribute.

ORTHOGRAPHY	
Field	*Data Type*
Orthography identifier*	Number
Orthography (description)*	Text
Target script identifier	Number

Table 14-4 Orthography table

Objects

The separation of attributes dependent of objects from those dependent of names is the most important step of normalization to be applied to the database. The indexing application will require, object attributes to add generic or geographical specifications to index entries in the case of homonyms. Regional identifiers specifying the country, state/province and possibly lower administrative units in which the object is located should connect to separate tables holding the descriptions and other relevant data about these units. Other attributes might be useful for specific index production purposes or other applications, for instance an electronic atlas (coordinates), cartographic presentation (capital underlining, representation of highest elevations per country or state), or optional additions to index output.

OBJECT	
Field Name	*Data Type*
Object identifier*	Number
Display name identifier	Number
Object type identifier*	Number
Regional identifiers*	Number
Latitude and longitude fields	Number
Years of object origin and end	Number
Elevation	Number
Flags required for printing the names in the map	Yes/No

Table 14-5 Object table

For the ease of database maintenance, one of the names under which the object is known may be selected from the Name table, so that we don't need additional queries to know what objects we are handling with.

A separate table listing the recognized object types should allow us to assign each object instance to an object class or category, like for instance 'populated place', 'independent first order area', different types of second, third and fourth order administrative areas, mountains, lakes, rivers, etc. When building the indexing application, the level of detail should suffice to keep homonymous objects apart in the indexes.

It should be noted that there is no universal way to cut up geographical reality into object categories. If the system must be able to produce atlas indexes for atlases in different languages and for different markets, the list of object types should therefore accommodate for distinctions made in each of these. *Example*: for specifying additions to index entries in some French language atlases, a distinction may be required between streams discharging into another stream or a lake (French: *rivière*) and streams discharging into an ocean or sea (French: *fleuve*), a distinction English or Dutch practice does not make. Likewise, distinctions between certain kinds of water bodies (straits, sounds, channels, bays, gulfs, creeks, coves) or mountain features (mountains, hills, rocks, peaks, mountain ranges, massifs, ridges) are language and society-dependent.

Objects vs. Names

One geographical object can be referred to by many different names. Different languages, scripts, transcriptions and transliterations, long and short forms, official and colloquial, old and new names may all be applied to one single object. In some cases, objects are subjectively defined by names: this holds, for instance, for the complete object class of streams as well as those of open water bodies.

14.7 Localizing the names

A table LOCATION is required to bring together the name, object and all relevant data connected to them with the name's grid reference in the map and other map-specific data:

LOCATION	
Field	*Data Type*
Location identifier*	Number
Name identifier*	Number
Map identifier*	Number
Does the named object occur in the map's subject area? *	On/Off
Does the name figure as a primary name?	On/Off
Grid reference fields*	Number and Text

Table 14-6 LOCATION table

This is the longest but also narrowest table of the database: it contains one row for every single name occurrence in every single map in every single atlas we produce, but apart from the usually one-letter grid references it includes no text strings. Essential to the names selection procedure of the indexing system is the field in which we define whether the name is part of the subject area (the main area of focus of the map) or not. In a map titled Indonesia, for instance, the names of all objects within the national territory of Indonesia are within the subject area, but the Malaysian capital Kuala Lumpur is not. In a map titled South-East Asia, the names of all objects located in South-East Asian countries are in the subject area. To qualify for index reference, maps in which the named object's position is within the subject area should receive priority above maps in which it is not.

Grid reference data may be stored in different fields containing the letters and numbers of either the grid section in which we find the first letter of the name, the grid sections where we find the first and the last letters of the name, or the grid sections that hold both ends of the named object. In the last case, coordinate information from the OBJECT table may also be used. Usually the location of the name is preferred because it can easier be found in the map visually, but more importantly because it is not always obvious where to logically locate beginning and end of areal objects.

The question whether a name is primary or secondary in a map is relevant in the cartographic process only, to decide whether a name should be parenthesized or printed in smaller type. If we choose to apply this field in a more sophisticated way, we could make it numerical and distinguish whether a name is primary, an exonym, a secondary, tertiary or quaternary name, whether it is preferably translated or not, etc.

14.8 The Atlas table

ATLAS	
Field	*Data Type*
Atlas identifier*	Number
Language identifier*	Number
Composition identifier*	Number
Atlas name	Text
Meta-information fields	Text

Table 14-7 ATLAS table

Of the atlas-dependent information that needs to be stored, the index production application requires the language of publication to select the proper text strings to be added to names in case of homonyms, as well as index composition details. The latter entail specifications about the types of name elements that will be inverted in the index, and whether or not both inverted and non-inverted index references will be included. For maintenance purposes, the name or title of the atlas and other meta-data (publisher, year of publication, edition number etc.) may be included.

14.9 Maps and Products

When a map is included in an atlas, we consider it a Product: within the atlas, it receives a title and possibly a subtitle, one or two-page numbers and possibly a map letter or number. Once we make clear for which atlas an index should be produced, the very first action of the index production application will be to consult the table PRODUCT and collect all the maps composing the atlas. The same table is also required to add page numbers and letters to the index entries in the end.

The MAP and PRODUCT tables should contain the following fields:

MAP	
Field	*Data Type*
Map identifier*	Number
Map description*	Number
Map type identifier*	Number
Map scale*	Number

Table 14-8 MAP table

PRODUCT	
Field	*Data Type*
Product identifier*	Number
Map identifier*	Number
Atlas identifier*	Number
Product title*	Text
Product subtitle	Text
Page reference fields*	Number
Map letter or number	Text or Number

Table 14-9 PRODUCT table

These tables are required to extract the name locations of all the names included in the atlas for which the index is being produced. The map type identifier refers to one of a set of distinguished map types, the descriptions of which are stored in a separate table MAP TYPE. It is required if in the selection procedure of our indexing system priority is given to occurrences in certain types of maps (in the Dutch case: the oro-hydrographic survey maps) above occurrences in other maps (for instance thematic maps).

The map's scale can be most simply stored as the denominator of the fraction of 1, at prevailing atlas map scales divided by 1000 (for instance: 6000 at a scale of 1/6,000,000). This value too is essential in the selection process of index production, where for the names of most categories of objects priority is given to the largest scale if the name occurs within the subject area of several maps of the same type.

14.10 Specifier Tables

SPECIFIER TABLES	
Field	*Data Type*
Regional or thematic addition identifier*	Number
Country, administrative unit or object type identifier *	Number
Language identifier*	Number
Text string to be added*	Text

Table 14-10 Specifier table

In order to allow the index production application to select the proper textual additions to homonyms, the database should contain one table for generic (object type) specifications, and several for geographical (regional) specifications.

14.11 Additional tables not used for index production application

Apart from the tables required by the indexing process, the database can be augmented with a number of tables to store information used in the editorial stage of cartography, or for referential purposes only. Dated population figures can be kept in a table that may include several rows per object, for different years and different object definitions (cities proper vs. metropolitan areas): these are required when population size determines the map symbol or type font or size cities and towns are represented with in the map. Object or other identifiers connect these tables to the core tables of the database.

The biggest advantage of maintaining a maximally normalized relational database for names management in an atlas producing environment is, that every piece of information having any relevance to the indexing or editorial process is stored and can be mutated at one place only within the structure of interconnected database tables, which reduces the chance of error because of unintentionally introduced inconsistencies to a minimum.

14.12 Reference:

Rennau, Günter – Register von Karten und Atlanten. Gotha: Haack 1976