

Database generated names indexes

for contemporary Dutch atlases

Tjeerd Tichelaar and Menno Bolder

Names indexes in atlases

Woensdrecht (gemeente)

240-241 B3

Woensdrecht (plaats) 42-43 C4

Woerden (gemeente)

240-241 C1

Woerden (plaats) 40-41 C4

Woeste Hoeve 40-41 E4

Woëvre 98-99 F2

Wognum 40-41 C2 D2

Wohlensee 112-113 C2

Woking 100-101 F6

Wold A 38-39 E5

Woldberg (Gelderland) 40-41 E3

Woldberg (Overijssel) 40-41 F1

Wolden, De 238-239 E3

Woldendorp 38-39 G2

Wolderwijd 40-41 E3

Woleai 146-147 K4

Wolf 182-183 D7

Wolfenbüttel 102-103 F4

Wolfgangsee 110-111 J2

X

Xaafuun 138-139 E7

Xaafuun, Raas 162-163 I3

Xaidulla 152-153 B4

Xaignabouri 146-147 C2

Xainza 152-153 D5

Xalapa 172-173 G8

Xam Nua 146-147 C1

Xankāndi 128-129 J8

Xanthi 120-121 H5

Xàtiva 116-117 E3

Xay 146-147 C1

Xiamen 152-153 J7

Xi'an 152-153 H5

Xiang 152-153 I6

Xiangfan 152-153 I5

Xianggang (plaats) 152-153 I7

Xianggang (provincie)

152-153 I7 J7

Yibin 152-153 G6

Yichang 152-153 I5

Yichun 152-153 L2

Yidu 154E

Yıldızgebergte 120-121 I4 I5

Yiliang 152-153 G7

Yinchang 154E

Yinchuan 152-153 G4 H4

Yingcheng 152-153 I5

Yingkou 152-153 K3

Yingtān 152-153 J6

Yining 152-153 C3

Yio Chu Kang 149E B1

Yishun 149E B1

Yitulihe 152-153 K1

Yli-Kitka 106-107 I2

Yme 95 D2

Yobe 167C

Yogyakarta 146-147 D7 E7

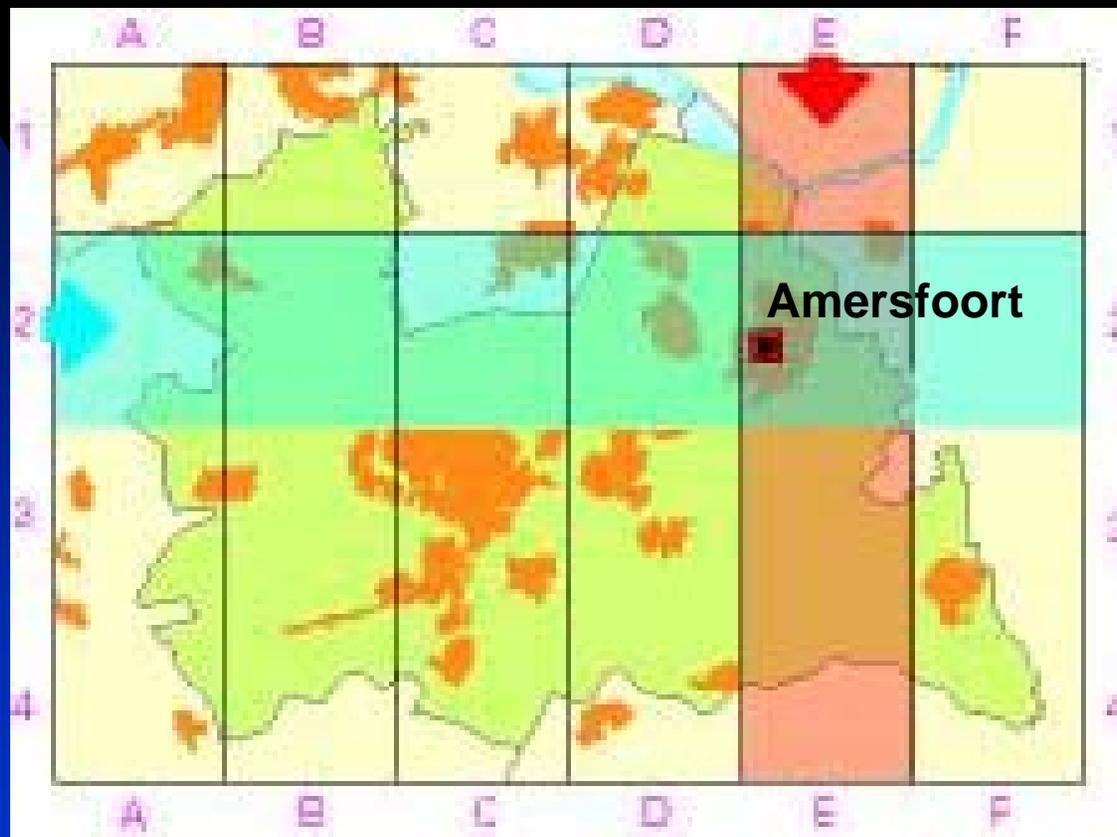
Yokkaichi 156 D3

Yokohama 156 D3

Names indexes in atlases

- The alphabetic names index is the most obvious key to localized geographic information to be found in the atlas.
- From the point of view of toponymy, the atlas index is instrumental in distributing geographical names: it is a bridge between names and map.
- Through cross references, the index also functions to lead the public from obsolete names or non-standard spellings to the preferred names, which in most cases should be official standard names.

Names indexes in atlases



Names indexes in atlases

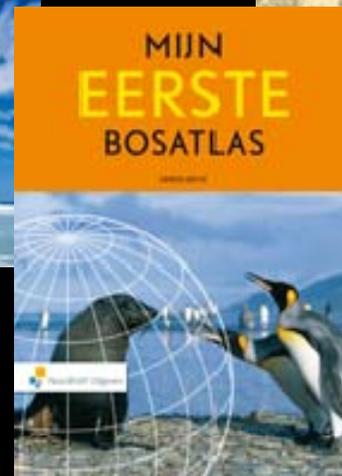
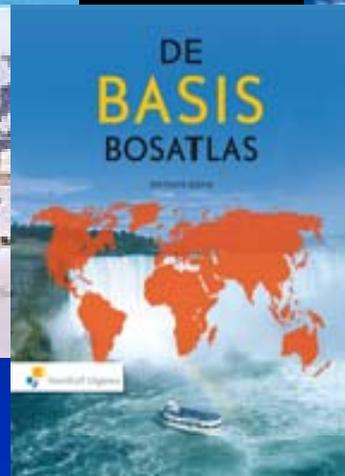
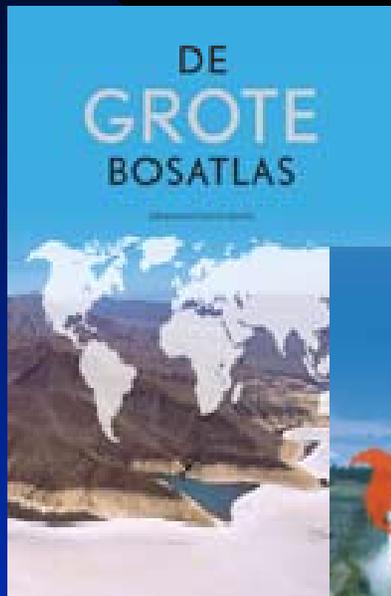
- Unlike a gazetteer, a school atlas index omits as much information as possible: nothing unnecessary should lengthen the journey from the name to the object's location in the map.
- An index in a reference atlas could be enriched with a gazetteer function by adding selected attribute data like coordinates, population or generic info.
- The only valid reason to add non-name info to a school atlas index is to avoid confusion: to keep homonyms apart.
- The index refers to the position of the name, not the object.

The Dutch school atlas

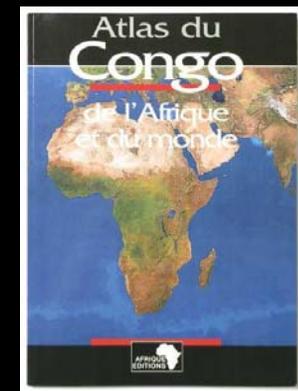
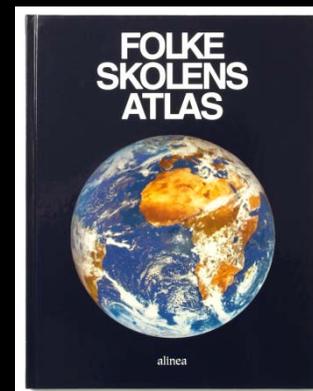
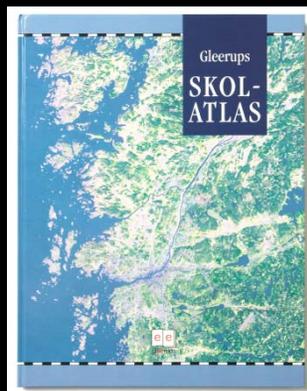
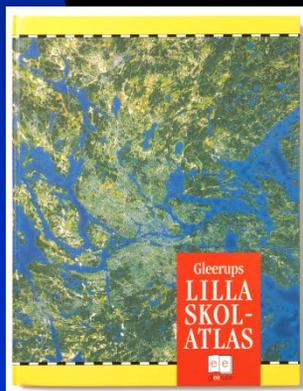
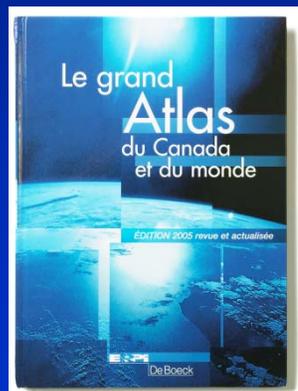
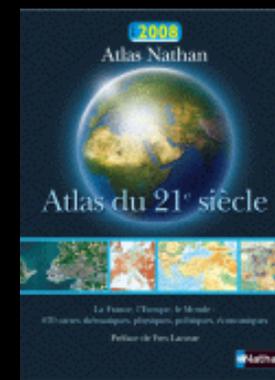
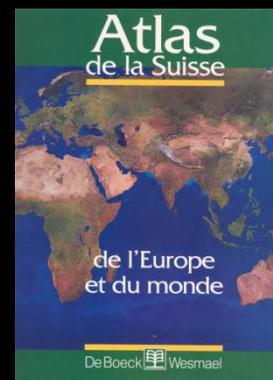
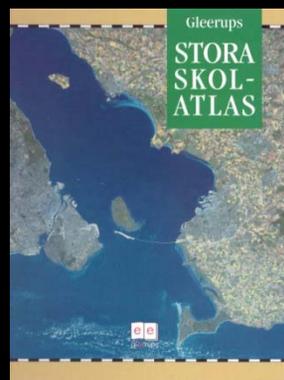
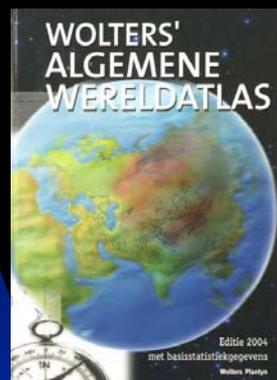
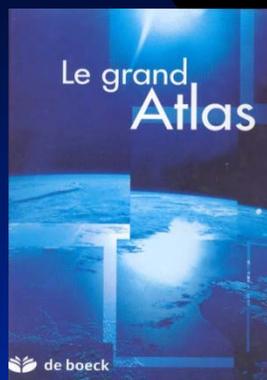
- In the Netherlands, the 'Bosatlas' has a long tradition. For 135 years it has followed the development of the geography curriculum in the different Dutch school types.
- The backbone of the atlas has always been a series of oro-hydrographic 'survey maps' covering all parts of the world on different scales.
- Through the years, the survey maps have been increasingly complemented with thematic maps: this too following the demands of school geography.



NL-produced school atlases



NL-produced school atlases



Rules of index composition

1. Every name-object combination (a name referring to a unique object) is represented in the index with one reference only.
2. When a name-object combination occurs in the atlas more than once, the map referred to is defined by a set of selection rules.
3. All name-object combinations are included in the index in a similar way and unabbreviated, regardless of whether in the maps they are parenthesized or abbreviated.

Rules of index composition

4. The names are listed alphabetically, following the sequence of the Dutch **alphabet**:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

5. When a name consists of a specific element preceded by a generic element, including preposition where it occurs, is in the index inverted:

Jawa

Jawa, Laut

Jawa Barat

Note: this implies that generic terms are recognized.

Rules of index composition

6. When a name consists of a specific element preceded by an article or a preposition, the article or preposition is for sorting purposes inverted as well. Again, **this implies that articles are recognized.**
7. To differentiate between homonyms, **regional and/or generic specifications** are added to the names of homonymous objects.
8. Cross-references may be included in the index, for instance to accommodate for recent name changes or frequently occurring alternative writings:

Cross references

Belarus = Wit-Rusland

Birma = Myanmar

Jogja = Yogyakarta

Leningrad = Sint-Petersburg

Pusan = Busan

Solo = Surakarta

Selection principles

1. Priority is given to survey maps above thematic maps.
2. The entire object should be showed on the map referred to.
3. Priority is given to the maps in which the named object is included in the map's 'core area'.
 - Core area typically defined by map title
 - Where map title doesn't suffice, the core area is explicitly defined by the editorial staff.

Selection principles

4. Priority is given to the map with the **largest map scale**.
5. For **country names**, the map scale is overruled and priority is given to the political continent map.

A relational database

- In 2004, a new relational database was built with the following aims:
 - To produce txt-output ready to load into a dtp application to produce an index of names;
 - To provide the ability to store and query all possible information on geographical names.
- As a **production database**, the new system replaced an older tailor-made application that had grown obsolete.

Selection principles

- As a **knowledge-base**, the system should enable individual editors to answer questions forwarded to the publisher about geographical names.
- The database should also give access to all the **rules** (e.g. transliteration) and **sources** the existing names were derived by, in case a new name has to be added to an existing map.

Database specifications

The new geographical names and indexing database had to meet the following specifications:

1. Portable: commonly available market software
2. Object oriented
3. Relational, maximally normalized
4. Able to communicate with the different cartographic databases employed, yet independent
5. Maintenance should require no special skills
6. Data safety sufficient
7. Character coding Unicode
8. Output of .txt-file for automated page layout in dtp₁₈

Database components

- Tables interconnected by key fields
- Knowledge base:
 - Table of geographical names
 - Table of geographical objects
 - Subsidiary data files
- Indexing system:
 - Table containing all occurrences of geographical names in all maps.
 - Tables listing all atlases and the maps they contain.

NAME

Field Name	Data Type
Name_ID	Number
Object_ID	Number
Language_ID	Number
Script_ID	Number
Transliteration_ID	Number
Name_Main	Text
Inversion_ID	Number
Gender_ID	Number
Number_ID	Number
Flag_Official	Yes/No
Flag_Primary	Yes/No
Flag_Definite	Yes/No

- One row for every name
- Name_ID: key attribute
- Name = 'written code referring to/defining one single geographical object'
- Many names predate writing
→ Name = 'a sequence of sounds conventionally used to refer to a definite person or object'
- Consequence: Amsterdam, Paris vs. London/Londen

NAME

Field Name	Data Type
Name_ID	Number
Object_ID	Number
Language_ID	Number
Script_ID	Number
Transliteration_ID	Number
Name_Main	Text
Inversion_ID	Number
Gender_ID	Number
Number_ID	Number
Flag_Official	Yes/No
Flag_Primary	Yes/No
Flag_Definite	Yes/No

- Differently transliterated forms of the same name or different alphabet versions are considered different names in terms of the database.
- Similarly written exonyms were initially considered one name, even when pronunciation differs:
Florence (French/English)

LANGUAGE

Field Name	Data Type
Name_ID	Number
Object_ID	Number
Language_ID	Number
Script_ID	Number
Transliteration_ID	Number
Name_Main	Text
Inversion_ID	Number
Gender_ID	Number
Number_ID	Number
Flag_Official	Yes/No
Flag_Primary	Yes/No
Flag_Definite	Yes/No

- In the database, every name belongs to a language.
- A language is a set of names belonging to a certain language or dialect in a historical context:
 - Specific dialects, e.g.: Transitional High German (South Franconian), North Straits Salish (Lummi)
 - General categories: Occitan (local dialect), Pre-Columbian Native American (unspecified)
 - A category 'unknown'

LANGUAGE

Field Name	Data Type
Name_ID	Number
Object_ID	Number
Language_ID	Number
Script_ID	Number
Transliteration_ID	Number
Name_Main	Text
Inversion_ID	Number
Gender_ID	Number
Number_ID	Number

- Name vs. language
 - Los Angeles (name of city in California) is an English name.
 - Los Ángeles (original name of city in California, currently exonym) is a Spanish name.

<u>Name_ID</u>	<u>Object_ID</u>	<u>Language_ID</u>	<u>Script_ID</u>	<u>Name_Main</u>	<u>Inversion_ID</u>
14282	69150	English (Modern)	Roman	Los Angeles	
14282	69150	Spanish (Castilian)	Roman	<u>Ángeles</u>	Los

NAME

Name_Main contains the specific element of the name:

Name_ID	Object_ID	Language_ID	Script_ID	Transliteration_ID	Name_Main
222317	90554	Sundanese	Aksara Sunda		ᮊᮧᮒᮧᮒ
106762	90554	Malay (Bahasa Indonesia)	Roman	Ejaan Van Ophuijsen	Bandoeng
2277	90554	Malay (Bahasa Indonesia)	Roman	Ejaan Yang Disempurnakan	Bandung

Name_ID	Object_ID	Language_ID	Script_ID	Transliteration_ID	Name_Main
221951	100043	Tagalog	Roman		Haba
213050	100043	Latin (unspecified)	Roman		lauva Insula
213051	100043	Latin (unspecified)	Roman		lava maior
11119	100043	Dutch	Roman		Java
11124	100043	Malay (Bahasa Indonesia)	Roman	Ejaan Yang Disempurnakan	Jawa
141094	100043	Javanese (Ngoko)	Roman	Phonetic	Jåwå
222318	100043	Javanese (Ngoko)	Dentawiyanjanya (Carakan)		ꦗꦮ

SCRIPT

Field Name	Data Type
Name_ID	Number
Object_ID	Number
Language_ID	Number
Script_ID	Number
Transliteration_ID	Number
Name_Main	Text
Inversion_ID	Number
Gender_ID	Number
Number_ID	Number
Flag_Official	Yes/No
Flag_Primary	Yes/No
Flag_Definite	Yes/No

- By storing names in their original script version in the database we are able to re-transliterate names in case UN-recommended transliteration keys are replaced.
- Unicode fonts are used to store and visualize names in all known scripts.

TRANSLITERATION

Field Name	Data Type
Name_ID	Number
Object_ID	Number
Language_ID	Number
Script_ID	Number
Transliteration_ID	Number
Name_Main	Text
Inversion_ID	Number
Gender_ID	Number
Number_ID	Number
Flag_Official	Yes/No
Flag_Primary	Yes/No
Flag_Definite	Yes/No

- The table Tbl_Transliteration includes:
 - Systems of transliteration
 - Systems of transcription
 - Orthographic standards
 - Optional or unofficially adapted variants like 'accentuated' or 'vocalized'
- A category 'unknown' is included.
- The target script is included as an attribute.

NAME_MAIN and INVERSION

Field Name	Data Type
Name_ID	Number
Object_ID	Number
Language_ID	Number
Script_ID	Number
Transliteration_ID	Number
Name_Main	Text
Inversion_ID	Number
Gender_ID	Number
Number_ID	Number
Flag_Official	Yes/No
Flag_Primary	Yes/No
Flag_Definite	Yes/No

- Generic elements, articles and prepositions are separately stored in TBL_INVERSION where they precede the specific element of a name.
- Required by index application: in Dutch indexes, generics (but not false generics) and articles are inverted, in French just generics, in Swedish none.

GRAMMATICAL INFORMATION

Field Name	Data Type
Name_ID	Number
Object_ID	Number
Language_ID	Number
Script_ID	Number
Transliteration_ID	Number
Name_Main	Text
Inversion_ID	Number
Gender_ID	Number
Number_ID	Number
Flag_Official	Yes/No
Flag_Primary	Yes/No
Flag_Definite	Yes/No

- Become relevant when names are used in written text or oral discourse:
 - when they carry an **adjective** or an **article**;
 - when it needs to be inflected when occurring in a particular **grammatical case**.
- Thus far: grammatical number (sing/pl) and gender (m/f/n).

GRAMMATICAL NUMBER

Field Name	Data Type
Name_ID	Number
Object_ID	Number
Language_ID	Number
Script_ID	Number
Transliteration_ID	Number
Name_Main	Text
Inversion_ID	Number
Gender_ID	Number
Number_ID	Number
Flag_Official	Yes/No
Flag_Primary	Yes/No
Flag_Definite	Yes/No

- Geographical names incidentally behave in a way not matching the elements they were etymologically constructed from, even when these were left unchanged, are clearly recognizable still and even call for the use of a proper article in speech: 'The United States is a big country in America, the Netherlands is a small country in Europe'²⁹

FLAGS

Field Name	Data Type
Name_ID	Number
Flag_Official	Yes/No
Flag_Primary	Yes/No
Flag_Definite	Yes/No
Flag_Generic	Yes/No
Flag_Long	Yes/No
Flag_Short	Yes/No
Flag_Old	Yes/No
IPA	Text
Source_ID	Number
Onoma_ID	Number
Adjective	Text
Accent	Number

Filled to flag whether or not:

- the name is currently **official**
- the name is a **primary** name
- the name is **definite**
- the name includes a **generic element**
- the name is a **short** form or a **long** form of the customary name appearing on the map
- the name is old (historical, obsolete) or still viable

ADJECTIVE

Field Name	Data Type
Name_ID	Number
Flag_Official	Yes/No
Flag_Primary	Yes/No
Flag_Definite	Yes/No
Flag_Generic	Yes/No
Flag_Long	Yes/No
Flag_Short	Yes/No
Flag_Old	Yes/No
IPA	Text
Source_ID	Number
Onoma_ID	Number
Adjective	Text
Accent	Number

- Adjective forms are in many languages different from the nouns.
- 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.
- Homonymous names may carry different adjectives.

Field Name	Data Type
Object_ID	Number
Display_Name_ID	Number
ObjType_ID	Number
Country_ID	Number
State_ID	Number
Region_ID	Number
Object_Parent	Number
Object_Longitude	Number
Object_Latitude	Number
Year-Origin	Number
Elevation_m	Number
Length_km	Number
Flag_1stOrd_High	Yes/No
Flag_2ndOrd_High	Yes/No
Flag_1stOrd_Cap	Yes/No
Flag_2ndOrd_Cap	Yes/No

OBJECT

- The separation of attributes dependent of objects from those dependent of names is the most important step of normalization applied to this database.
- In the indexing system, object attributes are used to add specifying information to index entries in the case of homonyms.

Field Name	Data Type
Object_ID	Number
Display_Name_ID	Number
ObjType_ID	Number
Country_ID	Number
State_ID	Number
Region_ID	Number
Object_Parent	Number
Object_Longitude	Number
Object_Latitude	Number
Year-Origin	Number
Elevation_m	Number
Length_km	Number
Flag_1stOrd_High	Yes/No
Flag_2ndOrd_High	Yes/No
Flag_1stOrd_Cap	Yes/No
Flag_2ndOrd_Cap	Yes/No

OBJECT TYPE

- Differentiates classes of objects like 'populated place', 'independent first order area', different types of 2nd, 3rd and 4th order areas, mountains, lakes, rivers, etc.
- Enhanced to meet the requirements of foreign publishers: e.g. 'fleuve' vs. 'rivière'.

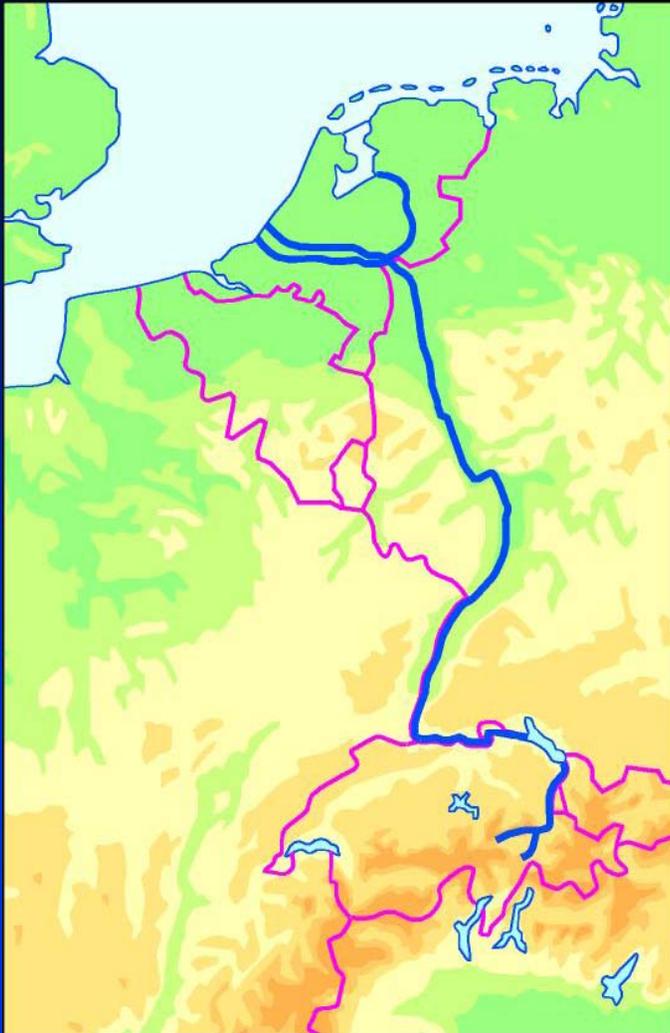
Objects vs. Names

- One geographical object can be referred to by many different names
- In some instances, objects are subjectively defined by names: this holds, for instance, for the complete object class of streams.

Objects defined by Names



Objects defined by Names



Objects defined by Names



LOCATION

Field Name	Data Type
Location_ID	Number
Name_ID	Number
Map_ID	Number
Area_ID	Number
Secondary_ID	Number
Location_BingoA	Text
Location_Bingo1	Number
Location_BingoB	Text
Location_Bingo2	Number

- The table Tbl_Location contains one row (Location_ID) for every name (Name_ID) occurring in every map (Map_ID).
- The field Area_ID specifies whether the name is part of the main area of focus of the map (**core area**) or not'.