Bauhaus-Universität Weimar
Faculty of Media
Degree Programme Computer Science and Media

Modelling Development of Countries in Historical Geographic Information Systems

Master's Thesis

Marcus Kossatz Matriculation Number 90487
b. 21.08.1989 in Spremberg, Germany

1. First Referee: Univ.-Prof. Dr.-Ing. habil. Volker Rodehorst
2. Second Referee: Prof. Dr. rer. nat. Sven Bertel

Submission date: 6. June 2016

Contents

List of Figures

List of Tables

Selbstständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Masterarbeit selbstständig und nur unter Zuhilfenahme der angegebenen Quellen erstellt habe.

Weimar, 6. June 2016 Marcus Kossatz

Acknowledgements

Helau

Chapter 1

Introduction

La Géographie n'est autre chose que l'Histoire dans l'espace,
de même que l'Histoire est la Géographie dans le temps.

Geography is nothing but History in space,
the same way as History is Geography over time.

– Élisée Reclus: "L'Homme et la Terre" (1908)

long and aspiring text about human life, the question of where, when, what and how?

time and space are everywhere, highly related to our lives and objects we perceive time personal time points of major life events, ...-¿ events, can trigger other events periods of studying, working, ... -¿ collection of events with similar characteristics world: every major issue has a time scale climate change (decades) climate tipping points (years) climate tipping points (years) economic meltdown (months) infectious diseases (weeks) disasters (days) -¿ time not easy to scale and to grasp space location of major life events (static) travel routes (dynamic) not always easy to grasp (exact location of monument is simple, but exact location of problematic area with fremdenfeindlichen hintergrund or historic countries are hard to set) event locations are sometimes not related to their consequences (e.g. Conferences of Tehran or Casablanca (1943) discussed how to deal with Germany after planned victory in WWII) motivation for spatio-temporal queries exploration of German history using historical maps of 1800, 1850, 1900 and 1950 each map has both temporal and spatial information in it but how to tell a story with that? more realistically, maps from 1871, 1919, 1933, 1945 and 1949, because of major events (founding of German Reich, end of WWI, beginning of Nazi dictatorship, end of WWII, founding of two German nations -¿ for one country might be suitable exploration of European or history would need a world map for each year how to see what has changed? -¿ inefficient how to know what is important? is that a reasonable way of storing information if one information set is with a high probability almost the same as the time point before? -¿ redundancy key problem model of historic maps at time points (-¿ snapshot model) given information at time point t1 and t2 How to know the status at time point t1 ¡ tm ¡ t2? -¿ It is impossible solution: away from snapshot based modeling of history to change-based modeling initial state ti, changes at point t1 and t2 How to know the status at time point t1 ¡ tm ¡ t2? -¿ it is ti + changes at t1 =¿ definition of each time point in history

research object of this thesis change over time of space of countries history of countries, their names and their borders and their relationships to each other visualize these changes and edit them (interface) Web-based historical geographic information system (WHGIS)

research questions of the thesis How to design and implement WHGIS? How to create an interface not just to explore historical changes? How to deal with uncertainty and fuzziness in history? Can researchers actually interact with such a system? How to design an interface that matches the mental model of a DH user of editing changes over time?

study of existing approaches, techniques and projects GIS: acquisition, management, analysis and presentation of spatial information handling of the spatial domain: extension to HGIS some systems allow presentations, but have very difficult interfaces no system that allows editing historical borders in time and space

domain: development of countries in time and space -¿ names and territories -¿ modelling, visualizing and !!! editing !!!

1.1 Motivation

research questions –¿ HGIS ¡– development of system historians / ME geographers (+) open source, direct manipulation, easy sharing and collaboration

1.2 Research Questions

1.3 Overview

Chapter 2

Basics

This chapter will lay the theoretical foundation of this Master's Thesis and will embed it into the context of current research. The title of this work is:

Modelling Development of Countries in
Historical Geographic Information Systems

It includes the domain (*history of countries*) and the system to acquire, model, manage and visualize data of the domain: *Historical Geographic Information Systems* (HGIS).

The first section of this chapter will define HGIS and related terms. Afterwards, concepts to model time and space in an information system are introduced. Data sources suitable for input into an HGIS are listed in the next section, followed by techniques to manage and analyse the data.

extension of Hivent model to actors

2.1 Historical Geographic Information Systems

> *"All human actions takes and makes place. The past is the set of places made by human action. History is a map of these places. The past thus exists not in time but in space."*

– Philip J. Ethington in [Eth07, précis]

An Historical Geographic Information System helps to answer research questions about how geographical phenomena have developed over time. To understand how it works, it is important to understand the four parts of the word: The research fields *history* and *geography* and the concepts of *information* an *systems*.

2.1.1 History

History is "an ideal field for thinking long and hard about important questions" [AHA]. The Greek word *Ιστορια / historia*, meaning "finding out, learning through research, narration of what is learned", is the origin [1] and it signifies the two main modern usage forms of the term: To research about and learning something and to tell a story. There are many different definitions of the word *history* [2]. The main goal of history is to study processes in the past to understand the situation in the present and make reasonable decisions for the future. The American Historical Association has developed the "five C's of historical thinking [that] together describe the shared foundations of [the] discipline" [AHA]:

Change over time The lives of people, their languages and their cultures are continuously changing. Describing these historical changes, triggered by historical events happed in the past, is a major goal of history. Snapshots in the form of historical maps or historical photography are used to tackle this task.

Context is an important element of historical thinking. The goal is to travel back in time to the moment of the event and recreate the world based on primary sources. The understanding of the historical context is crucial for the understanding of the event.

Causality The overall goal of each science is to answer the *why*-question concerning an event or a process. For historians that means to reasonably explain an historical event or process based on evidence. The problem is that history is not a science that can alter experimental conditions to extract new information, in a way that e.g. experiments in physics work. Historians have to focus on the interpretation of primary sources, which inherently yields multiple explanations for a single event.

[1] *History*, Dictionary.com, based on Random House Dictionary, 2015, URL: `http://dictionary.reference.com/browse/history`, last access: 23.10.2015

[2] *History*, Merriam Webster – an Encyclopædia Britannica Company, URL: `http://www.merriam-webster.com/dictionary/history`, last access: 23.10.2015

Contingency is a derived aspect from this problem. Each event has a whole network of prior conditions, because the world is highly interconnected. A slight change in one prior condition could have led to a completely different outcome of the event and a different state of the world.

Complexity The intrinsic human need for order conflicts with the complexity of history and their events and processes, because of its contingency. It is questionable if all details about events in the world are scientifically explainable.

Historical research is conducted by studying and interpreting primary sources, such as written documents, verbal texts, speeches, photographs, audio, video or historical maps. This signifies that most historical research is qualitative. The main organization principle in history is periodization: classifying events and processes to describe broader long-term changes and to explain complex phenomena [KH08, pp.4-7]. A special focus in this thesis is laid on historical maps as primary source to extract spatial information.

2.1.2 Geography

The term "geography" comes from Greek *γεωγραφια / geographia*, literally "describing the earth." [3] It is a science that studies the interplay between the landscapes and environments of the Earth (*physical geography*) on the one hand and the people, their cultures, societies and economies (*human geography*) on the other. That means geography is an interdisciplinary field between natural and social sciences [RGS].

Geographical research aims to understand where things are found, why they are there and how they developed over time. It focuses on the interconnectivity between elements of physical and human geography, which gets expressed in Tobler's First Law of Geography: "Everything is related to everything else, but near things are more related than distant things." [4]

Geographers use different technology and techniques to analyze geographic processes and to answer their research questions. The oldest and most important among those are maps. A map is a graphical expression of something that is not tangible: a part of the real world.

[3] *Geography*, Dictionary.com, based on Random House Dictionary, 2015, URL: `http://dictionary.reference.com/browse/geography`, last access: 23.10.2015

[4] "A computer movie simulating urban growth in the Detroit region", Waldo Tobler, 1970 Economic Geography, 46(2): 234-240.

A map shows the physical, environmental, political, economical or social properties of the Earth in order for the user of the map to get the most relevant information for his task, may it be orientation, learning or teaching. The "art and science of making maps" is the field of *cartography* [5]. Since maps visualize a model, they have a natural constraint: "No map can perfectly replicate the real world, since it inevitably generalizes, abstracts and approximates the complexity of the reality" [KH08, p. 181].

Comparison between geography and history Both research fields utilize maps for answering their research questions, which is the main commonality for the work of this thesis. However, the nature of both fields are also very different, illustrated in table 2.1.

geography	difference	history
where	dimension	when
exact, statistical	character	complex, fuzzy
mainly quantitative	research	qualitative
spatial proximity of conditions	causal explanation	temporal sequence of events
spatial differentiation	explanation	temporal differentiation
clustering	ordering	periodization
mostly visual (maps)	expression	mostly verbal (texts)
high (GIS)	digitalization potential	low (digital humanities)

Table 2.1: differences between history and geography [KH08, pp. 2-4]

Whereas geography answers the questions *where?*, history focuses on *when?* – but the ultimate goal for both sciences is to answer the question *why?*

2.1.3 Information

The terms "signs", "data", "information" and "knowledge" are sometimes used interchangeably and there is no coherent definition for any of them. However, all describe different concepts. This explanation seen in figure 2.1 is based on the work of [Dra].

[5] *History of maps and cartography*, James S. Aber, URL: `http://academic.emporia.edu/aberjame/map/h_map/h_map.htm`, last access: 24.10.2015

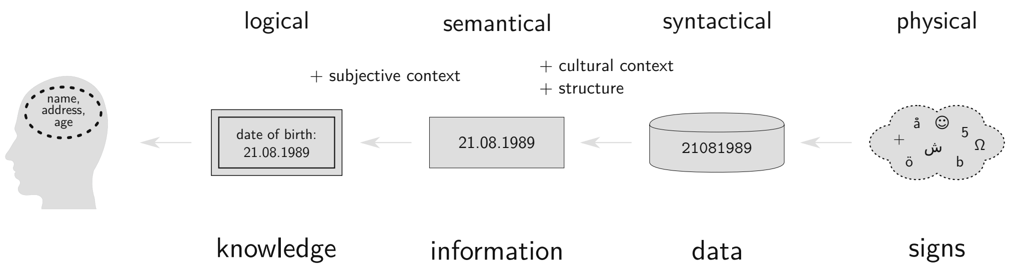


Figure 2.1: signs, data, information and knowledge

A *sign* is the physical representation of something in the real world. Since the real world is continuous, literally anything can be seen as a sign, so there are uncountably infinitely many different signs. *Data* is a subset of all possible signs and represents the syntactical level of what an information system deals with. Data itself does not have any meaning, but as soon as it is organized, it becomes *information*. However, information is sensitive to its cultural context. The string `14.07.1789` is useful and understandable for people in countries that use the date format `DD.MM.YYYY`. However, for people in Belize and the USA, that use the format `MM.DD.YYY`, this might just be a random string of numbers without any meaning, and therefore no information – although it is the same data. If information is visualized to and understood by a human and it can be integrated into his or her larger subjective context, it is *knowledge* [Nak]. The goal of a visualization is to present as much information as possible in a way that it can be transformed into knowledge by the viewer.

2.1.4 System

A *system* is an organized structure containing *elements* or *components* that are directly or indirectly *related* to and *interconnected* with each other. The elements and their relations form the whole of the structure. The surrounding of the system is its *environment*. There is an *internal state* at any point of the system's existence. This state only changes when it gets influenced by stimuli of its environment. *Emergent properties* characterize a system. They are independent from properties of the element of the system, e.g. water is liquid at room temperature, but the elements it consists of, hydrogen and oxygen, are a gas. Each system is both part of a larger system and can be decomposed into subsystems. Therefore, systems form a hierarchy.

A system has defined spatial and temporal boundaries. There are two types: *open systems* allow exchange of energy or information with their environment, whereas idealized *close systems* naturally do not interact with and are not influenced by its environment. Based on the black box principle the inner working of a closed system can not be seen from the outside [Bus].

2.1.5 Motivation

An *information system* (IS) is an application that is dealing with the acquisition, management, analysis and presentation of information. It is the unity of all its components and their interaction with each other [Zwa]. If the majority of the information in a system has a spatial relation to the Earth, its surface, its lithosphere, atmosphere or the social or economical structure of its habitation, it is a *geographic information system* (GIS). The data objects in the system are called *geo-objects* [Bol08]. If the information additionally has a temporal dimension, e.g. via time stamps or time spans, which enable to trace developments of geo-objects, it becomes an *Historical Geographic Information System* [GG14] or alternatively *Spatio-Temporal Information Systems* (STIS) [PTKT04].

HGIS react on the spatial turn of history: the integration of geographic methods in historical research. It aims to discover the power of cartographic representation: "The spatial turn in the humanities must [...] understand the role of space in human events" [BCH10]. At the same time, they are the product of the temporal run in GIS: the coexistence of space (where things are) and time (what has changed over time) [Sol14, p. 45]. With HGIS it is possible to analyze how "spatial patterns change over time in order to better understand large-scale Earth processes" [Peu99]. Since "the world never stands still", but "the retention of information relating to past events [is] an important element of human representation of the world", the dimension of time has to be integrated into a GIS [Peu99].

HGIS are rather recent tools and used mostly in *Digital Humanities* as a digital tool to answer research questions in the traditional fields of humanities: "situating history in its geographical context and using geographic information to illuminate the past" [KH08, p. 3]. Some interesting research questions that could be answered using HGIS could be:

- Did the European Union help to bring peace on the European continent? *(political)*
- Is there a coherence between life expectancy and fertility rate? *(social)*
- What is the effect of global warming on the melting of glaciers? *(physical)*
- What was the effect of Bismarck's foreign policy on peace in Europe? *(historical)*

Or on a more abstract level: Where and When has something changed and why did it change?

2.1.6 Components

Information systems in general are based on a data model — HGIS in particular on a *spatio-temporal data model*, introduced in section 2.3. On top of that, there are different components. One way to classify them is using the four-component model:

1. **Input**: Primary acquision of spatio-temporal data, i.e. historical events, historical and current countries and their territories.
2. **Management**: Physical storage and logical management of the data in a spatio-temporal database, using a structure that fits the spatio-temporal data model.
3. **Analysis**: Gaining spatio-temporal information by cleaning, transforming or combining the data in database.
4. **Presentation**: Visualization of information on different displays, e.g. a map and a timeline, transforming information into spatio-temporal knowledge.

2.1.7 Applications

"Today, operational temporal GIS does not exist". This quote summarizes the state of the art in this field. The main reasons are "the complexity of integrating space and time and the lack of standards" [Raz12, p. 5].

However, there are numerous project that use HGIS for one specific research question. A large collection them can be found in [KH08] and [GG14]. A famous visualization combining time and space Napoleons Moscow Campaign by Charles Minard from 1869 (see figure 2.2). The "best statistical graphic ever drawn" [6] shows the number of men in Napoleon's 1812 Russian campaign army, their movements, as well as the temperature they encountered on the return path [KH08, pp. 188-191].

[6] *The Visual Display of Quantitative Information* (p. 40), Edward R. Tufte, 2001

[7] *Minard.png* Charles Minard, 1869, URL: `https://commons.wikimedia.org/wiki/File:Minard.png`, last access: 03.11.2015,

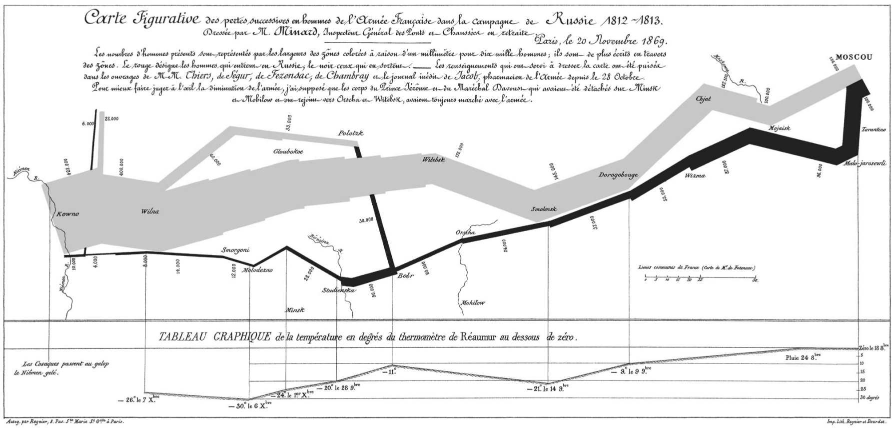


Figure 2.2: Napoleons Moscow Campaign [7]

The *Great Britain Historical GIS Project* (GBHGIS) [8] combines statistical data with territorial units of the United Kingdom, e.g. to analyze net migration in the districts in UK. The data is collected by the *British Ordnance Survey*, who automatically detect spatial changes to the geography and land use of the United Kingdom using aerial photography [9].

The *National Historical Geographic Information System* (NHGIS) [10] provides the digital boundaries of the United States of America and census data for each year since 1790. While the data in the system is extensive, the interface to analyze and use the data is very frustrating to use: A tutorial is necessary to go through the selection process. To download data, a user has to register, receive an email with a link to download a compressed file which has to be decompressed and then loaded into a GIS software to be visualized.

HGIS are not widely accepted in the humanities. One reason is the nature of the qualitative historical research: historic sources are subjective and biased, their content may be fuzzy and they are definitely incomplete. So the knowledge that can be extracted from a source bears the integral problem of *uncertainty*. Information systems on the other hand have a logical architecture and, to be as precise and accurate as possible. Analysis is based on mathematical functions – an information system is quantitative in its entire nature [KH08, p. 2].

[8] *Great Britain Historical Geographical Information System (GBHGIS)*, Ian Gregory & Humphrey R. Southall, University of Portsmouth, since 1994, URL: `http://www.port.ac.uk/research/gbhgis/`, last access: 02.11.2015

[9] *British Ordnance Survey*, URL: `https://www.ordnancesurvey.co.uk/education-research/research/automatic-change-detection.html`, last access: 02.11.2015

[10] *Welcome to NHGIS*, Minnesota Population Center, University of Minnesota, since 2007, URL: , last access: 02.11.2015

2.2 Time and Space

This section will explain ways to separately represent time and space in an information system. It will first explain the geospatial data model used in traditional GIS and then introduce maps as the representation of spatial information. In the second part of the chapter, models to represent the temporal dimension are introduced.

2.2.1 Model of Geographical Space

HGIS need to unambiguously locate geo-objects on, underneath or close to the Earth's surface using *geographic coordinates*. They express an object directly in the coordinate system of the Earth. To understand that, a model of the Earth has to be developed, the *geodetic datum*, that needs to fit the real shape of the Earth as accurately as possible.

The shape of the Earth measured in the field of *geodesy* is very complicated. In the Babylonian Empire ($\approx$ 2000-539 BC) the theory of the Earth being a flat disc surrounded by an infinite body of water evolved. The Greek scientists Pythagoras and Aristotle (340 BC) rejected this theory and proved the earth to be a three-dimensional spherical object. It took almost 2000 years until Sir Isaac Newton (1687) reasoned that due to the centrifugal forces of the rotating Earth the shape has to be flattened at the poles and is therefore better described as an *ellipsoid* with two radii: the polar radius (r_p) and the slightly larger equatorial radius (r_e) [Bol08, pp. 69-77].

However, the model disregards that the surface of the Earth is not flat but consists of deep oceanic trenches and high mountains. Therefore the gravitational field of the Earth is not homogeneous either: the actual *mean sea level*, the reference surface for the height of objects varies from 106 meter below to 85 meter above the uniform sea level of the ellipsoid model. These discoveries in the 20th century led to the complex *geoid* model (see figure 2.3). The latest and most accurate measurements for the shape of the Earth are the result of the GOCE satellite launched in March 2009 [Uot, Fra].

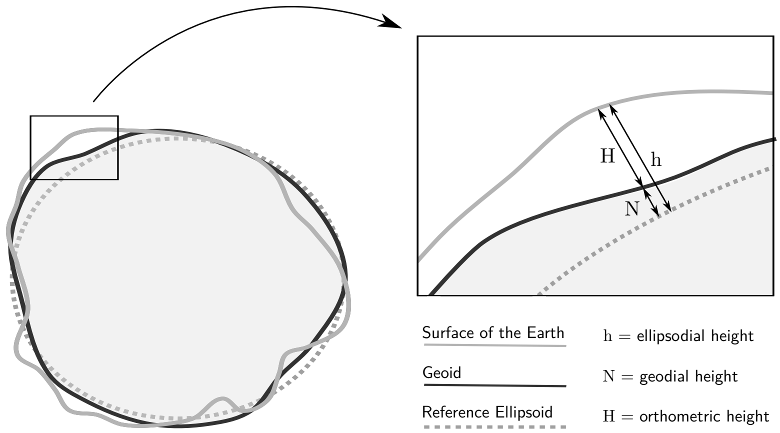


Figure 2.3: The geoid model, differences are exaggerated, [Bol08, Fig. 3-6, p. 75]

Geographic coordinate system The basis for the geospatial data model is the reference ellipsoid. It is represented in a three-dimensional *spherical coordinate system*. The *North* and the *South Pole* are defined as the two surface points closest to the Earth's center opposite to each other. The *Equator* is the line equidistant to the two poles and dividing the world in a *Northern* and *Southern Hemisphere*. Additionally, the *Prime Meridian* is defined as the line perpendicular to the Equator, running from the North to the South Pole. Since there are infinitely many lines like this, its definition is arbitrary, but by convention, the line running through Greenwich (London, United Kingdom) is used. Based on these two lines, each point in the spherical coordinate system can be unambiguously defined by [Bol08, pp. 26-28]:

1. The rotation angle along the Equator, defining its longitude: $\gamma = [-180° \text{ ... } +180°]$
2. The rotation angle along the Prime Meridian, defining its latitude: $\phi = [-90° \text{ ... } +90°]$
3. The distance to the origin: $r \in \mathbb{N}_0$

Lines of constant latitude are running horizontally and are called *parallels*, lines of constant longitude are *meridians* appearing in vertical direction. All parallels are circles with their center on the axis between the poles. No two parallels intersect. The longest parallel is the Equator (0° latitude). All meridians have the same length. Geographic coordinates are usually recorded either in degree-minutes-second (`DMS`, e.g. `50° 58’ 22’’`) or in decimal degree (`DD`, e.g. `50.973`) notation [Bol08, pp. 30, 79].

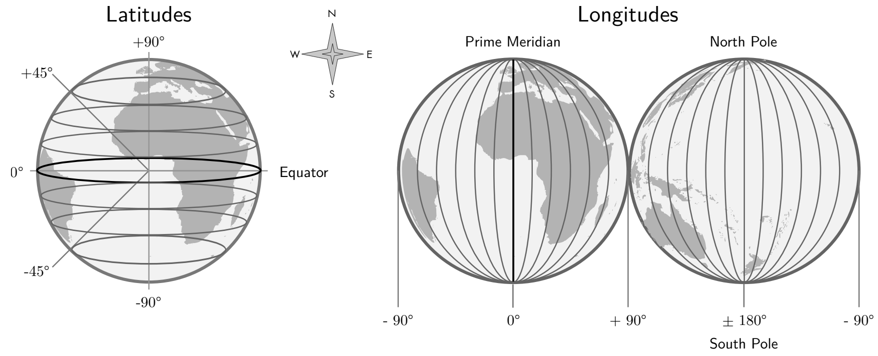


Figure 2.4: geographic coordinates using latitude and longitude

The Geodetic Datum is the digital model of the analogue Earth. It consists of two parts: The approximation of the Earth's surface in a the Cartesian coordinate system with the origin in the Earth's center and a set of reference points used to accurately locate a point.

Geodetic datums can be very accurate in one region of the world, i.e. the model fits the real geoid very well, but inaccurate in another region. This is the main reason why there are a lot of different geodetic datums used in the world. The same coordinates in two different geodetic datums define two different points on Earth. In order to be accurate is essential to know the geodetic datum of the coordinates [Bol08, p. 80]. The *World Geodetic System 1984 (WGS84)* is a model that found worldwide acceptance and is used in all major Web-based mapping services like *OpenStreetMap* and in the GPS unit of major mobile devices.

Raster and Vector Model The real world is infinite in detail, but storage in a computer system is finite. In order to model continuous geographical phenomena in an information system, a relevant subset of them has be sampled to create discrete spatial data. It can be represented in a raster or in a vector model.

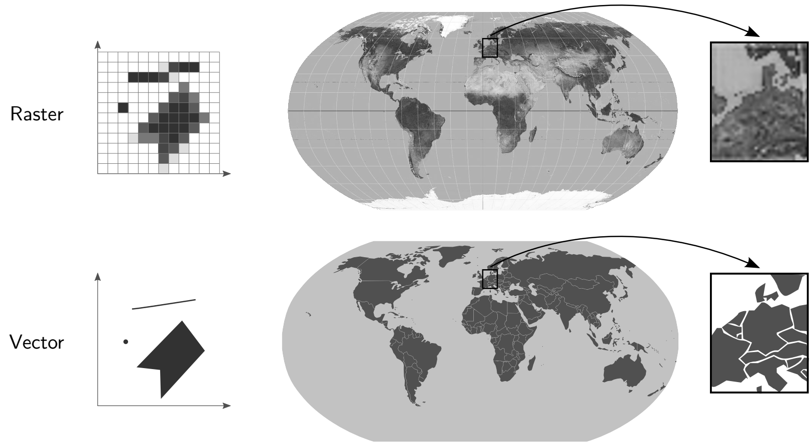


Figure 2.5: Comparison of the raster and the vector model

The *raster model* contains a regular grid with a fixed *cell dimension*. Each cell has a certain value, e.g. a color value. The model is simple and allows straightforward rendering: only affine transformations have to be applied in order to project two raster map layers on top of each other. The main disadvantage of the raster model is its fixed resolution: it can not be scaled up without losing quality [Bol08, pp.42-48]. Raster graphics are used for map tiles by most map engines, e.g. in OpenStreetMap or the satellite image by NASA in Google Maps.

In the two-dimensional *vector model*, each object is a mathematically described geometric primitive. All of them can be expressed by three basic primitives (figure 2.6):

0D A *point* is the fundamental object in vector geometry. It has no dimension, no size and is only defined by its position, specified in geographic coordinates. One point is independent from all others. Points can be used to represent the location of an event.

1D A *polyline* is constructed by an ordered set of points with at least one start and one end point. A border line can be expressed by a polyline.

2D A *polygon* is an ordered set of polylines creating a closed area. A polygon can be *simple*, *weakly simple* or *complex* (see figure 2.7). The territory of a country without islands can be described by a polygon. If a country does have islands or overseas territories, a *polypolygon* represents multiple separate polygons belonging to one logical entity.

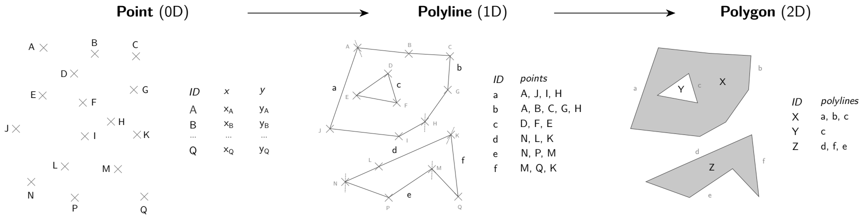


Figure 2.6: The basic geometric primitives point, polyline and polygon

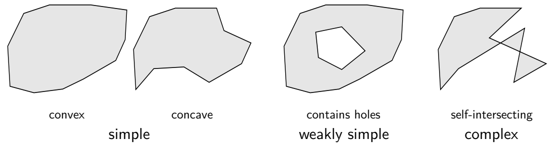


Figure 2.7: Different properties of polygons

Scale-independence is one of the biggest advantages of a vector model. The data model is more compact in comparison to the raster model. On the other hand, the model can become very complex. Since vector data has to be rasterized to be shown on the screen, the computational effort increases with complexity [Bol08, pp.33-42]. Vector models are suitable to represent phenomena that can easily to discretized, e.g. the boundaries of a country. Common file types for vector data with spatial reference are the open file format GeoJSON (`.geojson`) [11] or ESRI Shapefiles (`.shp`) [12].

2.2.2 Presentation of Geographic Space

The most common ways to present geographic space are two-dimensional *maps* and three-dimensional *globes*. The HGIS in this thesis will use a map to show the development of countries over time. A map is a discrete graphical expression of the geographical features of the continuous real world. The creation of a map is not just a scientific, but also a creative process: The form, function and interaction methods shall follow the purpose of the usage of the map.

[11] *GeoJSON*, IETF GeoJSON Working Group, URL: `http://geojson.org/`, last access: 30.10.2015

[12] *ESRI Shapefile Technical Description*, ESRI White Paper, July 1998, URL: `http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf`, last access: 30.10.2015

A map is typically structured according to the *layer* principle: Each layer is a transparent film showing one specific aspect, e.g. a physical layer showing coastlines, mountains or forests, a political layer showing international borders or a cultural layer showing cities or population densities. The layers are interchangeable, can be switched on and off and serve to serve a different visualization purpose. A *legend* including the scale bar and north arrow shall explain all symbols used on the map and give orientation. In interactive web based systems, there should be *menus* with different visualization options, e.g. panning and zooming on the map, switching map layers on and off or changing the color scheme of the map [Bol08, pp. 159-166].

Leaflet.js is "an open-source JavaScript library for mobile-friendly interactive maps" [13] that offers functionality to embed a map with a chosen projection in on the client-side of a web based information system, use own map tiles, symbols and markers on the map and tools for zooming and panning.

Map projections Since the Earth is three-dimensional, but the map on the computer screen only two-dimensional, the model of the Earth has to be projected onto the map. But as previously discussed in subsection 2.2.1, the Earth is an inhomogeneous spherical object with a curved surface whereas the map is flat [Bol08, p.79]. That is why some features of the Earth will be distorted on the map: An *equivalent projection* preserves the area sizes of features on the map, whereas a *conformal projection* preserves angles and the shapes of objects. Every map projection that is area-preserving distorts shapes at the same time, and each shape-preserving map distrots areas to some degree. There is no perfect map projection. [Geo].

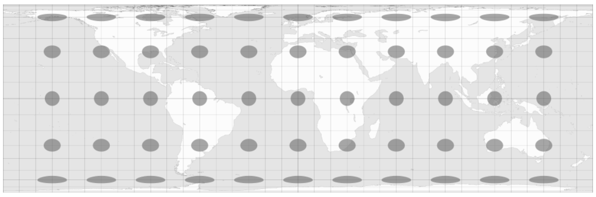
A compromise between preserving areas and shapes is the *Robinson projection*. It is neither conformal, nor equivalent, but provides a reasonable trade-off between both properties.

[13] *Leaflet - JavaScript library for interactive maps*, URL: `http://leafletjs.com/`, last access: 02.11.2015

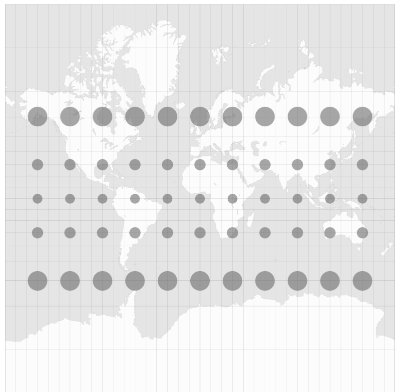
[14] *Tissot indicatrix world map equirectangular proj*, Eric Gaba / Sting (Wikimedia), June 2008 URL: `https://commons.wikimedia.org/wiki/File:Tissot_indicatrix_world_map_equirectangular_proj.svg`, last access: 28.10.2015

[15] *Logo of the United Nations*, Shizhao (Wikimedia), 13.06.2007 URL: `https://commons.wikimedia.org/wiki/File:Logo_of_the_United_Nations_(B%26W).svg`, last access: 28.10.2015, Comment: This work is excerpted from an official document of the United Nations prior to 17. September 1987.

[16] *Tissot indicatrix world map Robinson*, Eric Gaba / Sting (Wikimedia), June 2008 URL: `https://commons.wikimedia.org/wiki/File:Tissot_indicatrix_world_map_Robinson_proj.svg`, last access: 28.10.2015



(a) equivalent Lambert projection [14]



(b) conformal Mercator projection [15]

Figure 2.8: Comparison of equivalent and conformal map projections

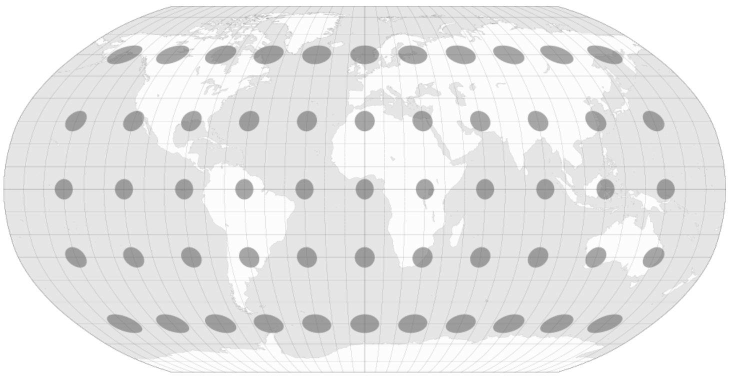


Figure 2.9: Robinson projection [16]

2.2.3 Model of Historical Time

Time is an abstract concept that "can be perceived only by its effects" [Lan89, p. 27]. Many philosophers and scientists have been developing models to work with time. In this case, the model needs to be appropriate to both represent time in an historical sense in interplay with geographical space.

A popular model is *Cartographic Time*, where time is seen as the "fourth cartographic dimension", is suitable for spatio-temporal information systems [Lan89, p. 28]. Whereas space is represented by geo-objects on a map, time may be seen as versions or states on a timeline, separated by events that change one state to another state. Unlike space, time knows only one dimension. The position of an event on the timeline is described by its date using a reasonable sampling unit like century, year, day, hour or millisecond [Lan89, p. 32].

Types of Time The simplest categorization is between a discrete *event* and a continuous *process*. Events can happen at a certain *time point* or like processes in a *time interval* or *time period*, defined by two time points. An information system that stores events with a significant outcome regarding the geo-objects in the system, is an *event-based historical geographic information system*. On the other hand, a *process-based historical geographic information system* models mainly processes as a series of events of one kind regarding a small set of geo-objects [Sol14, chapter 2, pp. 47-49].

The Taxonomic Model of Time by [Fra98] classifies time not only into discrete and continuous, but also by the *nature of time* or *time order*: a consecutive development on the time axis, defined by start and end, defines *linear time*. In a contrary, *cyclic time* has no predefined order and events reoccur on a regular cyclic basis. The other two types, *branching time* and *multi-dimensional time*, are more complex and not relevant for this thesis.

Temporal Topological Relations The topological relationship between two time points t_1 and t_2 is straightforward. Since they are discrete elements and therefore isomorphic to the number space of integers, there are three different order relations:

1. $t_1 < t_2$: the first event happens before the second event
2. $t_1 > t_2$: the first event happens after the second event
3. $t_1 = t_2$: the first and the second event happen at the same time

For time spans, there are six possibile temporal topological relations (table 2.2). Except for `equals`, each of them has an inverse, yielding a total of 13 different relations.

relation	symbol	visualization
X before Y	`X < Y`	X Y
X meets Y	`X m Y`	X Y
X overlaps Y	`X o Y`	X Y
X equals Y	`X = Y`	X Y
X starts Y	`X s Y`	X Y
X during Y	`X d Y`	X Y
X ends Y	`X e Y`	Y X

Table 2.2: Temporal relations of time spans, based on [All84]

2.2.4 Presentation of Historical Time

In contrast to space, time does not have an intrinsic representation. However, the most common form to visualize cyclic time is on a cyclic display, e.g. a time wheel or a clock. Linear time is very often visualized on a *timeline*. The purpose of a timeline is to show events as time points or processes as time intervals in chonological order. A timeline additionally shows time markers showing a certain date to support orientation. A timeline uses a certain time scale:

- On a *linear* timeline, the distance between any two time points is directly proportional to their actual temporal distance.
- A *logarithmic* timeline uses a logarithmic function to scale the depicted time. Relative to a reference point on the timeline, e.g. the timeline center, the further away a time point, the further away its position on the timeline – however, the distance between the time point and the reference point does not increasy linearily, but logarithmically. That means, events that are further away do not appear as far. This time scale accounts for logarithmic human perception: events that happened 20 years ago do not seem to be twice as long ago as events happening 10 years ago [DISP08].
- A timeline can also have an *irregular* scale, e.g. to have the same absolute distance of events on the timeline. This is useful if the distribution of the events on the timeline are far from homogeneous.

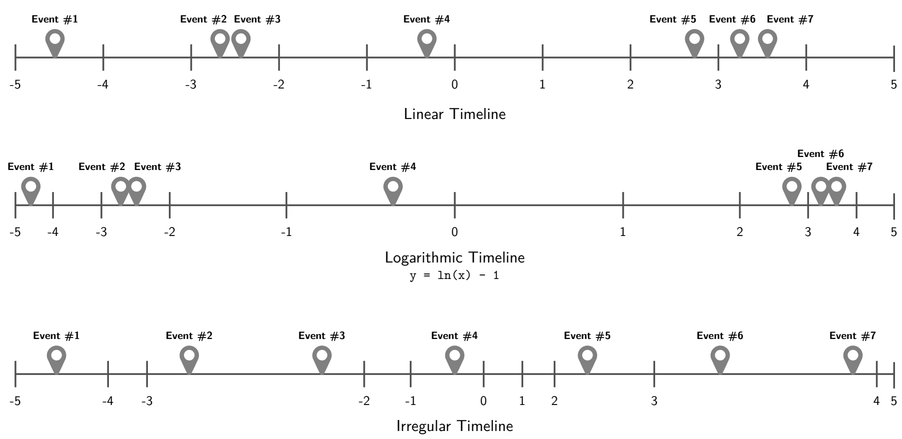


Figure 2.10: Comparison between a linear, a logarithmic and an irregular timeline

The visualization of time can be separate from the spatial dimension, according to the Triadic Framework, e.g. with a timeline. In another approach, space and time can also be coupled and displayed in the same presentation display, e.g. in a space-time cube [Häg70].

2.3 Spatio-Temporal Data Models

> *"Geography differs from geometry because*
> *in geography, space in indivisibly coupled with time"*

– Don Parkes & Nigel Thrift (1980)

A model tries to replicate a part of the real world. A data model abstracts a part of the world, identifies the most essential elements and their relation to each other. Historical Geographic Information Systems can be used to explain spatial-temporal phenomena in the real world. Therefore, it needs to handle the development of geo-objects and their attributes over time. Developments are driven by *changes* to the state of an object.

Based on the theory of the *Triadic Framework*, there are three components involved: space (3 dimensions), time (1 dimension) and attribute (multiple dimensions). All of these dimensions can change independently from each other [OS01, p. 53]. However, in order to trace spatial and attribute changes over time, the dimensions have to be related to each other.

Throughout the lifetime of a geo-object, it appears at some point, might undergo several changes and might disappear at some other time point. The data model has to be able to effectively and efficiently manage those changes. There are mainly two kinds: *Discrete changes* are based on the idea of a *state machine*: At any point in the lifetime, an object is in a certain state. It stays there until an event occurs that suddenly changes the object into a new state at a discrete time point. As an example, if an armistice agreement between two former war parties A and B contains a deal to cede parts of the territory of A to become territory of B, this territorial change is sudden. On a contrary, an object can gradually change according to a *continuous process*, e.g. the change of the coastlines of landmasses [Peu99].

The spatio-temporal data models developed in the previous 30 years differ mostly in the organizing dimension: In *location-based* models time is an attribute of a geo-object. On a contrary, *time-based* approaches handle events and processes that change objects suddenly or gradually. *Entity-based* models represent geo-objects as own entities. Spatial changes over time are related to these entities, but they are not attributes and therefore independent.

This section introduces different spatio-temporal data models to maintain relations between time and space of an entity.

2.3.1 Snapshot Model

One of the simplest, oldest and most frequently used models is based on the idea of *snapshots*: At a certain time point t_i, a new layer gets created. It stores the full picture of the current state of all geo-objects [Lan88].

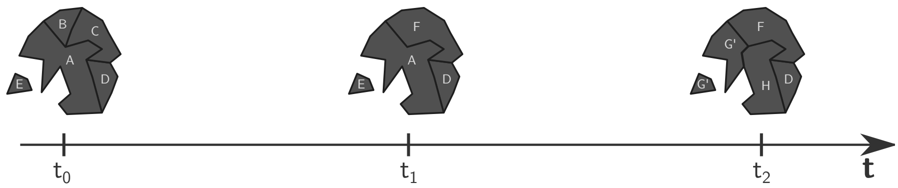


Figure 2.11: The Snapshot Model by [Lan88]

The model allows to retrieve the state of the system at a defined time point t_i. However, for all other time points $t \neq t_i$ that are not covered by a snapshot, it is impossible to retrieve the state of the system, because the data model does not record any changes. This is an integral problem of the model and can not be solved. The original model is also redundant, because objects that have not changed from one snapshot to the next one are duplicated. However, there have been improvements made, e.g. by [Arm92]. Historical maps are examples for snapshots: They show the state of the world at one point in history, e.g. Europe 1919 and Europe 1945. However, with no additional information, it is impossible to deduct how Europe looked like in 1939. Therefore, this model is not suitable for the domain of this thesis.

2.3.2 Simple Time-Stamping

This problem is solved by assigning a geo-object a period of existance by two additional attributes: at the *start date* t_{start} the object gets created and at the *end date* t_{end} it is ceased. If an object still exists its cessation date gets a special value, e.g. `NOW` [HW90].

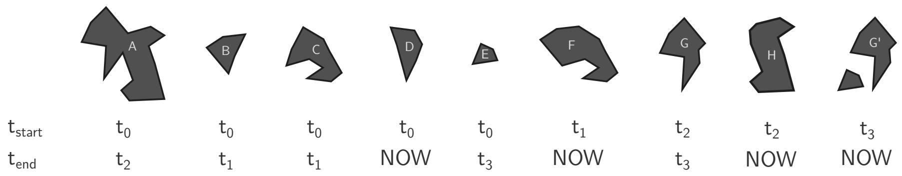


Figure 2.12: The Simple Time-Stamping method by [HW90]

The *Simple Time-Stamping* method is also location-based and tracks discrete changes of objects. Given full and integer information, the state of the system at each time point t_i can be retrieved: All geo-objects for which $t_{start} \leq t_i < t_{end}$ are active, all others are inactive. However, this retrieval is cumbersome, because without efficient data structures every time the date changes, it has to be checked for each geo-object if its state has changed.

Another problem of the model is that it does not allow for tracing the development of objects in different states. As an example, at time point t_1 the geo-objects B and C cease and G starts. Visually, G is a successor of B and C, but this historical relationship can not be deducted directly from the model. This shortcoming can be resolved by adding a reference to the predecessor and the successor of the object.

This model alone is not suitable for the domain of this thesis, because it is impossible to say what exactly has happened at a certain time point. Given the example above, it is unclear if two objects unified to a new one ($B + C \rightarrow G$) or if two are successors ($B \rightarrow G$) and one just stops to exist ($C \rightarrow \emptyset$). The model is also redundant: if a geo-object replaces another one ($B \rightarrow G$), then the end date of B is the same as the start date of G.

2.3.3 Event-Based Spatio-Temporal Data Model

A time-based approach addresses exactly those shortcomings: They explicitly represent events or processes in the data model and associate all objects that change according to them. One example of this approach is the *Event-Based Spatio-Temporal Data Model* (ESTDM) for geospatial raster data by [PD95].

At one defined time point t_b, a snapshot gets stored. This *base map* contains the current state of the map, i.e. the current value of each raster cell `(x,y)`. From that moment on, the system stores events that change the values of certain cells. Such an event has a time stamp (`t`) and a list of components associated with it. A component represents a new value (`v`) and knows which raster cells (`x, y`) change their value to `v`.

The method uses the following data structures: a header file contains information about the thematic domain, a pointer to the base map and to the first and last element of the event list. This doubly-linked list stores all events chronologically. Therefore, each event knows its preceding and succeeding event via a `prev` respectively `next` pointer.

If the time point of an event is reached, all its components are executed, i.e. the relevant raster cells change their value. The system follows the `next` pointer to know which event is waiting to be executed next. Since a change is relative to the previous change, not to the base map, change tracking is efficient.

The concept of the ESTDM suits the problem domain really well: An historical event changes the geometry of certain objects suddenly. The model explicitly represents these discrete changes. However, it does not work for vector data. The authors have explicitly stated that "the design of such a [vector-based] model is not seen as a straightforward task", because of the problem "how to maintain the integrity of spatial topology as it changes [...] The solution will require a more complex definition of components within individual events" [PD95, p. 21].

2.3.4 Three-Domain Model

An event-based STDM for vector geometry including lines and polygons has to answer the following questions: What uniquely identifies a geo-object? What kind of spatial, topological and attribute changes can happen to an object? Which of these maintain the identity and which create a new object? This problem is addressed in the *Three-Domain Model* by [Yua96a,Yua96b]. The model is based on abstract entities that represent a spatio-temporal object. It handles the three domains identity, space and time separately:

- The *semantic domain* holds an entity uniquely identifiable. An object in this domain corresponds to a human concept, e.g. a "country".
- The *spatial domain* represents geospatial objects in vector format, e.g. a polygon describing the territory of a country.
- The *temporal domain* stores all temporal objects, e.g. time points of an historical events, or time intervals of a war.

The model is not specific, but more a general abstract framework to handle space, time and identity. This makes the model very flexible, e.g. it can handle discrete and continuous changes, relative and absolute time, world and database time. One limitation of the model is that it only traces spatial attributes over time. In an alternative model by [CT95], the *thematic domain* is added to fully describe a spatio-temporal object and trace also aspatial attributes that can change over time, e.g. the name of an entity.

Since countries, their territories and attributes can change independently over time, the data model used in this thesis will be built on top of the Three Domain Model.

2.3.5 History Graph Model

Most of the data models introduced so far cover only static changes of geo-objects. [Ren96] identified three different types of temporal behaviour of changing objects:

- Dynamic objects that change continuously.
- Static objects that change according to events with duration (processes).
- Static objects that change according to sudden events.

Based on this observation he developed a data model that can handle all three kinds of temporal behaviour: The *History Graph Model*. It manages objects and events separetly from each other. An object can only be in three different states:

1. An object is *static*, if it currently does not change. This is called an *object version*. The version has an interval associated to it representing the duration of the object version, until it changes the next time. If the object is dynamic and changes continuously, the duration is zero.
2. If an object is currently *changing*, it is in an *object transition*. The transition has an associated interval as well, whose duration is zero if it is a sudden change. Additionally, a transition links the relevant objects to each other creating a historical predecessor-successor-relationship.
3. An object that is currently not active, is *ceased* and not visible on the map.

The history of a geo-object is a chronologically ordered set of versions and transitions, that can be visualized in a graph (see figure 2.13).

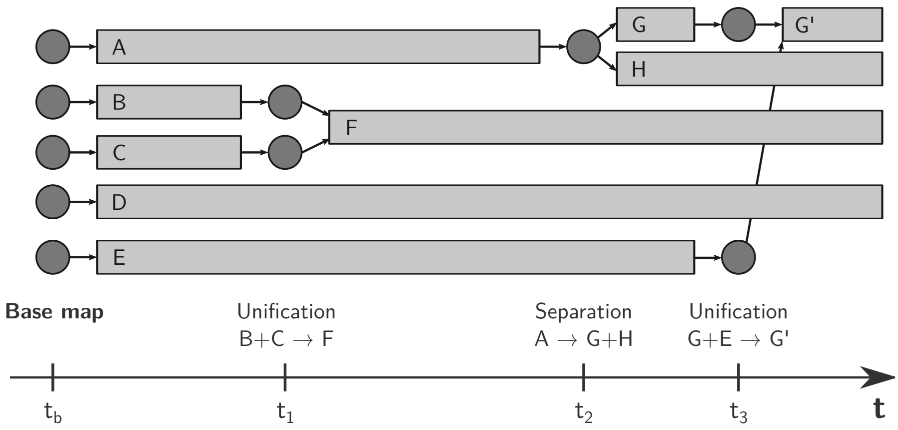


Figure 2.13: The History Graph model

The model defines six basic types of temporal changes that can happen (see figure 2.14):

- **Creation**: A new object is created.
- **Alteration**: A property of an object (e.g. geometry) changes.
- **Cessation**: An object is ceased.
- **Reincarnation**: An object that has previously been ceased is recreated.
- **Split/Deduction**: An object is divided into two or more new objects or one or more objects are deducted from an existing one.
- **Merge/Annexation**: Two or more objects are joined together to a new object or one or more objects are annexed to another object.

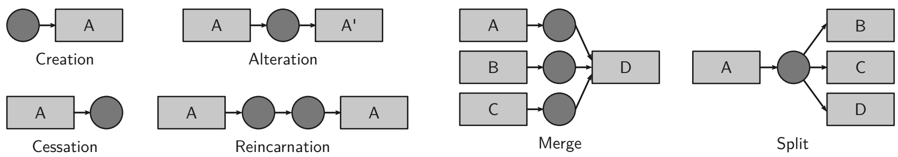


Figure 2.14: Types of changes in the History Graph model

The History Graph model can be seen as an extension to the ESTDM. It combines the advantages of event-based and entity-based spatio-temporal data models, supports discrete and continuous changes and relative and absolute time. The main improvement is that the historical development of a geo-object can directly be derived from the model, because objects are linked to their precedessors and successors — the History Graph Model can tell a story. This is the reason why this model is particularly suitable for the work of this thesis.

Other popular spatio-temporal data models that are not covered in this work, because they were not seen as relevant for the domain, include the *Space-Time Composite* model, the *Grid Model* and the *Amendment Vector Model*. Overviews about these and other spatio-temporal data models can be found in [Zha11], [PTKT04] and [Peu99].

2.4 Database Management Systems

Information systems use databases for managing the data. A *Database Management System* (DBMS) is a software system for the administration of data, mainly storage and retrieval. There are mainly two types of DBMS: the oldest and most common ones are *Relational DBMS*. *Object-Oriented DBMS* werde developed to adapt concepts of object-oriented programming into the database world. The combination of both approaches are *Object-Relational DBMS*.

2.4.1 Relational Database Management Systems

RDBMS are built upon the concept of *entitities*, e.g. an `HistoricalCountry`, with *attributes*, e.g. `name` and attribute values of a simple data type, e.g. the character string `"Germany"`. Entities are represented in a table with one row for each *tuple* and one column for each attribute. An entity has one attribute that unambiguously identifies each tuple, the *primary key*, mostly a contiguous number.

Entities can be related to each other in three different kinds of *relations*:

- `1:1` Direct attributional relation, e.g. one country has one head of state and vice versa.
- `1:n` One-to-many relation, e.g. one country can have many cities, but each city can belong to only one country.
- `m:n` Many-to-many relation, e.g. one country can have many rivers, but each river can also flow through multiple countries.

Entities and their relations are visualized in an *Entity-Relationship Model* (ER model). Data can be retrieved from and entered into a relational database using the *Structural Query Language* (SQL). The query to get the names of cities in `Germany` in alphabetical order is:

```
SELECT    city.id, city.name
FROM      (city JOIN country ON city.country = country.id)
```

```
WHERE       country.name = "Thüringen"
ORDER BY    city.name
```

The first RDBM developed was *Oracle*, released in 1979 [Ora]. Since then, the concept has been established as the state-of-the-art for databases. An example for a popular RDBMS used for Web-bases systems is *MySQL*, the "world's most popular open source database" [17].

2.4.2 Object-Oriented Database Management Systems

One problem with RDBMS is that attributes can only be assigned simple data types. Developers using object-oriented programming need to map the objects used in the application to tuples in the relational database – and vice versa data from the database needs to be transformed to objects in the application. This process can be cumbersome. OODBMS have been developed to address this problem and adopt the concepts of object-oriented programming for database management purposes [Dar].

- *Classes* are the structured representation of things in the real world of the same kind, with the same properties, e.g. a country, having a name and a territory. Classes in OODBMS relate to Entities in RDBMS.
- An *object* is an instance of a class, one specific thing with defined properties, e.g. the country of Germany with its territory. This correlates to a tuple in RDBMS.
  - The *attributes* of an object can not just be of a simple data type, but also instances of other classes, e.g. `country.territory` can be a `polypolygon` object. These complex data types are a major improvement compared to RDBMS.
  - Objects also have *methods* that can be called to do something with the object, e.g. `territory.getArea()` calculates the area size of a country.
- The internal state of an object can not be accessed from the outside. Methods are the only way to interact with an object. This is called *encapsulation* and maintains control over what can be done to and with an object and prevents corruption.
- According to the concept of *inheritance*, classes can be hierarchically structured, whereas the attributes and the methods of a *base class* are inherited to its *derived*

[17] *MySQL :: About MySQL*, URL: `https://www.mysql.com/about/`, last access: 31.10.2015

class. As an example, an `Area` has a `name`, a `territory` and the method `getArea()` associated to it. A `Country` can be derived from the `Area`, inheriting both attributes and the method. Additionally, it can get an attribute `head_of_state`, which is specific to `Country`, but not to `Area`. The class `Ocean` can just as well be derived from `Area`.

- An associated concept is *polymorphism*: The same function can be called on different objects and the return value will be of the same type. However, internally it might be calculated differently. As an example, consider the classes `Polygon` and `Polypolygon`, both inherited the method `getArea()` from their base class `Geometry`. Whereas a polygon calculates its area directly based on its geometry, a polypolygon internally calles the function `getArea()` on all its associated polygons and sums up their areas.

OODBMS support all those concept and allow to store the objects used in the application directly in the database. Additionally, objects from the database can be accessed directly, there is no need for an additional query language [Dar].

2.4.3 Object-Relational Database Management Systems

ORDBMS combine the advantages of both worlds. Internally, it uses the established and efficient relational database for the data storage. The database model and the interaction with the data happens in an object-oriented way while supporting all of the concepts mentioned in the previous subsection 2.4.2. The most popular ORDBMS example for Web-based systems is *PostgreSQL*, "the world's most advanced open source database" [18].

2.4.4 Spatio-Temporal Databases

Databases for HGIS have to deal with spatial, temporal and attribute data. According to the Triadic Frame, these aspects should be modeled separately from each other, as they can change independently.

Spatial data can easily become very large, because of the mass of very precise coordinate data. *Spatial databases* are specialized to work with spatial data: they process the data

[18] *PostgreSQL: About*, URL: `http://www.postgresql.org/about/`, last access: 31.10.2015

effieciently and to provide general data types, such as `Point` or `Polygon` and methods, e.g. to calculate the area or the distance between two points. *PostGIS* [19] is an extension for PostgreSQL that is especially utilized for handling spatial data.

Temporal data usually relates to events and processes. It is defined either by a time point or a time interval which is again defined by two time points. This is called a *bitemporal* element. A time point can be modeled in the database as an attribute of the complex `Date` type. For relational databases that only support simple data types, the date can be stored as a string or be expressed with a `long` integer (64 bit) $\forall n \in \mathbb{N} : n \in [-2^{63} \,..\, 2^{63} - 1])$ determining the number of milliseconds since 1st January 1970 (UNIX time) [Emi]. SQL was extended by features to handle time in a database, e.g. *SQL/MM* [Peu99, chapter 6].

Object-Oriented Spatio-Temporal Database Models The question is: How to implement the spatio-temporal data modles introduced in section 2.3 in a relational, object-oriented or object-relational database management system introduced in this section? While the implementation depends on the data model, there are common concepts and isses that have to be addressed.

When storing time related data, it is important to distinguish between the time that was true in reality (*valid time* or *world time*) and the time it was stored in the database (*transaction time* or *database time*). A property of spatio-temporal database models is whether valid and transaction time are supported.

Object-oriented concepts are more appropriate than the concept of relational databases, because of the complex nature of spatio-temporal data [PTKT04, section 3.9]. One of the first concepts was the concept of a *spatio-temporal object* combining geometrical and bitemporal properties in one object [WHM90].

A similar approach by [Raz12] is the *Spatio-Temporal Data Type* (STT). Time is not considered an attribute of space, but a separate class. They are aggregated in the `SpatioTemporal` class, using both spatial and bitemporal attributes. The model also provides spatio-temporal operators, e.g. `STT_intersects` returns `true` if two `SpatioTemporal` objects intersect in time and space, i.e. their geometries intersect and the time intervals in which they are active overlap. These operators are very helpful when analyzing spatio-temporal data or checking

[19] *PostGIS*, URL: `http://postgis.net/`, last access: 31.10.2015

for data integrity.

Version management An issue is how to perform retrospective updates. Given a database model that stores objects that are created, updated and destroyed by events. Object X is created at time point t_x. At time point t_y, X gets destroyed and replaced by object Y. If a new change that updates X to X' gets inserted at time point t_u in between, i.e. $t_x < t_u < t_y$, then the event at time point t_y is not correct anymore, because object X does not exist. The question is how to maintain data integrity on insertion, update and deletion from a spatio-temporal database? This issue has to be addressed using formal logic for temporal reasoning [Peu99, section 6].

2.5 HistoGlobe

Application: HistoGlobe A distributed *Web Information System*, consists of a remote server side, on which the storage and management of the actual data happens, and the client side on which the user communicates with the system. It hosts the user interface that is rendered in a Web browser.

map for spatial domain (x, y) timeline for temporal domain (t) -¿ 3D system

describe the components of the HG explicitally

ancestors successors layers of administrative units open to extension for additional attribute data (e.g. statistics)

requirements geographical knowledge contextualize / intersect historical sources accept imprecision prevent illusion of certainty

usable User Interface for both navigation and editing -¿ problem: all interfaces are trés horrible!

transition to concept chapter

Chapter 3

The Hivent Model

The task of this thesis is to model, visualize and edit the development of countries in time and space in a HGIS. Before the system itself can be developed, the underlying data model has to be fixed. This is the purpose of this chapter.

In section 2.3, different spatio-temporal data models were introduced to solve this problem. The *Snapshot Model* was seen as unsuitable for the problem space. *Simple Time-Stamping* is helpful to link countries to their history, but the method does not explicitly model historical changes, which is desireable. For that purpose, the idea of the *Event-Based Spatio-Temporal Data Model* was developed, but since it only works for raster data, it is also not suitable for this thesis. This problem is solved in the *History Graph Model*. Additionally, the introduced temporal changes allow to represent historical changes and their influences on geographic entities directly in the model. Finally, the *Three-Domain Model* introduces a helpful concept to separate the spatial, temporal and thematic dimension of a spatio-temporal entity.

The *Hivent Model* developed in this thesis is constructed from components of some of these models: An Event-Based Spatio-Temporal Data Model organized according to the Three-Domain Model, using Simple Time-Stamping for the spatial entities, supporting vector data and visualization on a History Graph.

In the first section of this chapter, the main elements of the Hivent model are introduced. Afterwards, the preconditions are defined. The next section explains the Hivent Operations that are used to change countries in time and space. The database model is illustrated in the third section and the data structures used to organize the elements of the model in the last section.

3.1 Elements

The main elements of the the model are the *Hivent*, representing an historically significant happening, the *Area*, an abstract entity on a map with a name and a territory and the *Historical Change*, part of one Hivent and manipulating the history of one or more Areas.

Hivents are the main organizing element of the data model. The word is an acronym for ***Hi***storical e***vent***. It represents a significant happening in history at one specific time point, e.g. a treaty, bill or declaration. In this domain, the focus is on events that influence the geopolitical situation on Earth. That means, they introduce historical changes at their point in history. An Hivent has five different attributes:

1. The *name* of the Hivent.
2. A textual *description* of the topic of the Hivent.
3. The point in time, identified by the Hivent *date*.
4. The Hivent *location*.
5. The *historical changes* resulting from the Hivent.

Areas are the the visible entities on the map. They are an abstract representation of one identical current or historical country. The model can easily be extended to the history of states, provinces or regions. Therefore, from now on the term *political unit* instead of *country* is used to describe the object in the real world that is modeled by an Area.

Each Area has a *short name*, e.g. "Germany", and a *formal name*, e.g. "Federal Republic of Germany". The main spatial dimension of the data model is the Areas *territory*, represented by a polypolygon. It is a set of weakly simple polygons, because the model has to support enclaves and exclaves. The polylines of the polygons represent the borders of the political unit. A polyline consists of an ordered set of points, representing the border points.

An Area can change over time. Throughout its lifetime, it is created at some point, then its territory and short name can change multiple times and at some point it ceases. Since all changes in this model are sudden, there are only two possible states an Area can be in: It is *active*, if at the current time point it is historically existing and it is *inactive* if it does not. Each area is uniquely identified by its formal name. That means, the short name can change, but as soon as the formal name of an area changes (e.g. "German Empire" to "Federal Republic of Germany"), it is considered a "new" Area.

Historical Changes influence the development of an Area over time. They can create new Areas, update their territory or name and cease them. Each Historical Change belongs to exactly one Hivent, inheriting its time point at which the change happens. The actual historical change is defined by a set of five Hivent Operations introduced in section 3.3.

3.2 Preconditions

> *In the beginning God created the heavens and the Earth*
> *Now the Earth was formless and empty [...]*
> *And God said, "Let there be light" — and there was light.*

– Genesis 1:1, The First Book of Moses, Old Testament

The spatio-temporal system has to be initialized at some point. It is based on a set of trivial axioms stated in this section. Each Area in the system is located directly on the surface of the Earth. While the surface is curved, as explained in sections 2.2.1 and 2.2.2, it can be projected on a two-dimensional map.

Axiom 1 *The Earth's surface has an invariant area size, i.e. it does not change over time.*

This axiom sets the spatial foundation of the system: a constant dimension of the map. The basis of the temporal part of the system is introduced in the next three axioms:

Axiom 2 *The spatio-temporal system has an initial state at time point t_0. At this initial state, there exists exactly one Area, denoted by Ω and referred to as the* universe *Area. It has no name and its territory covers the whole surface of the Earth.*

Axiom 3 *At each time point $t_i \geq t_0$ an Historical Change can be introduced.*

Axiom 4 *At each time point $t_i \geq t_0$ each point on the surface of the Earth is covered by exactly one Area.*

As it has been defined in section 3.1, an Historical Change can create, manipulate and cease Areas on the Earth's surface. According to axoim 4, each change introduced in the system must maintain the spatial integrity on the map. That means, as soon as an Area is created in one territory on the map, the territory that was there before has to cease. Vice versa, if an Area ceases from the map, it has to be replaced with at least one new Area and these new Areas combined have to occupy exactly this same territory.

The first Historical Changes introduced in the system at time point t_0 are the creation of all bodies of water, including the oceans and lakes. They are created as Areas with their name and territory which is cut out of Ω. That means, at t_0, the map of the world is divided into a set of Areas representing water and Ω, representing land that can at each time point $t_i > t_0$ be occupied by an Area representing a political unit claiming the land as their territory. To simplify the data model, two assumptions are made:

Assumption 5 *The borders of a political unit are either* interior*, i.e. bordering another political unit, or a* coastline*, bordering a body of water.*

This assumption is a concious simplification of the reality: It assumes the territory of a political unit stops at the coastline. Juristically, this is not true, because in line with [Uni82], the territory extends in a range of 3 to 12 miles (5 to 20 kilometers) into international waters. This model disregards the sea territory of a political entity to keep the model simple.

In the real world, The name of a political unit changes according to sudden events, e.g. a declaration or a governmental bill. The territory can change either because of a geographical processes, e.g. the Sea Level Rise influencing the change of the coastline, or according to a historical event, e.g. a treaty.

Assumption 6 *The geographical conditions on Earth, especially the position of land and water and the coastlines have not changed over time.*

While this assumption is obviously wrong, it helps to keep the problem space clear and the data model simple: The Hivent model of this thesis focuses only on discrete historical changes and not on long-term geographical developments. However, it is designed to be open to future extensions to account also for geographic changes and international sea borders. In this data model, the temporal behavior of an Area can therefore be described as a *static object that changes according to sudden events*.

Earth is initially fully covered with one area: Ω 1) secession of water W from Ω =¿ rest = usable (unclaimed) land this part can change due to model of long-term continuous geographic changes 2) cede rest of Areas from Ω countries debated territories unknown land water

The data model only represents sudden changes of Areas, no processes, i.e. changes with duration. The model also assumes that coastlines never changed. Additionally to these two constraints, it is assumed that the surface of the Earth is divided into two types surfaces: *water* and habitable *land*. Land can at any point in time be either *claimed*, i.e. it is currently the territory of exactly one active Area, or on a contrary be *unclaimed*.

base: Newtons concept of absolute space?

=¿ topological rule: each border has exactly two neighboring Areas each Area has at least one neighboring Area

abstraction of NCH to ACH (attribute change) =¿ all 7 cases: ICH UNI INC SEP SEC BCH ACH

BCH can geographically also be reduced to combination of separation and unification, but that would historically be different

no support for hierarchies

no support for independent overlapping areas

=¿ one layer

3.3 Hivent Operations

transition to next chapter

-¿ Managing Vagueness, Uncertainty and Granularity in Spatial Information Systems (VUG) -¿ Karl Grasser (Diss. Santa Barbara) -¿ Fuzzy, Imprecice, proabilities vs. possibilities

big problem: why? intention and motivation of author? hard to find out... voice and perspective medieval maps: natural landmarks as border points =¿ inaccurate and imprecise perspective: who is making the map? (illiterates?) different names: US Civil War (North) vs. WWI (West) vs. Germanic War (Russia) WWII (West) vs. Great Fatherland War (Russia)

accepted uncertainty: date != exact timepoint, only D.M.Y location != exact location, only name of place

[Sol14, chapter 2, p. 51]

think about how to represent historical knowledge in geographic context degree of certainty -¿ ironically: that has to be exact as well in a database table =¿ reason: careful conclusions from historical maps

country borders coastlines interior disputed territories situation: n fully recognized countries and m non or partially recognized entities claim sovereignty over 1 territory territory is surrounded by disputed border question: does this disputed area claim sovereignty? [1]http://www.economist.com/blogs/economist-explains/2014/09/economist-explains-1 uncertain borders situation: n fully recognized countries commonly agree on a boundary between them, but the border is not clearly defined / fuzzy / uncertain states of borders planned agreed demarcated provisional valid vs. disputed

borders: complex model: different states of boundaries: draft -¿ proposal -¿ dispute -¿

However, the first question arises regarding the relevance of the location: While the exact position of the battlefield of Verdun or the place where John F. Kennedy was assassinated might very relevant to the event itself, the location of a governmental bill, a declaration of independence or a border convention might not play an important role and usually happens in a representative place, e.g. the parliament or the office of a president. In a lot of cases, it is much more important which territories an event actually influences instead of where it happened.

[1]\unskip\penalty\@M\vrulewidth\z@height\z@depth\dpff

Hierarchical Areas CTR -¿ STA -¿ CTY -¿ CIT each level one layer aggregate geometry upwards

Overlapping Areas e.g. war zone, independent layer

Chapter 4

Uncertainty

Every aspect of the concept (chapter **??**) and the development (chapter **??**) of this work is based on the prerequisite of full certainty of the data. That means both the Historical-Geographic Operations and the Hivent-Based Spatio-Temporal Data Model assume that the dates of the historical events, the names and territories of the historical and current areas and the historical relations between events and areas are accurate and reasonably precise (definitions see 4.2).

However, this assumption is far from valid. In historical research, uncertainty is one of the major problems (see 2.1.1) a historian has to deal with on a daily basis: sources, even primary sources, can be biased towards the author of the source, information can be imprecise or even inaccurate and information can be conflicting with other sources. This chapter explains problems with uncertainty in the domain of development of countries in time and space and develops approaches to deal with these problems.

4.1 Definition of a Country

The problem begins with the definition of a term that almost everybody in the world is familiar with: a “country”. Since countries are the domain of this historical geographic information system, it must be possible to decide for each current and historic territorial entity in the world if it is or was a country or not. Therefore a clear and non-conflicting definition of a country is necessary. However, this is impossible to do.

The Oxford Dictionary definition of a country reads as follows:

> "The *territory* of a *nation*; a *region* constituting an *independent state*, or a region, province, etc., which was once independent and is still distinct in institutions, language, etc." [1]

This definition includes many different concepts and terms: the territory or region that the country is on, a nation or state, a population and a culture of the territory in terms of institutions or languages. While nation and state are commonly used as synonyms for countries, their meaning varies from case to case, as it will be examined in this section.

To understand what a country really is, the United Nations as an intergovernmental organisation are a valuable source. It was found after World War II (October 1945) and promotes international peace keeping, security, protection of human rights or humanitarian aid to all its member states which should coincide with all the countries in the world. The committee currently has 193 full member states and two permanent obervers: The Holy See (Vatican City) and the State of Palestine [Unib]. But these 195 members in total do not cover all places in the world – and also a membership in the United Nations does not mean that the question of statehood can simply be answered.

4.1.1 Special Cases

Examining the list of the UN member states yields several interesting observations and special cases, which can be classified by their membership status in the United Nations and their degree of international recognition.

UN observer states The *Holy See* is the juridcal and spiritual entity representing the territory of Vatican City. It is a fully recognized and sovereign state but is not a full member of the UN, because it has never applied for it. It is the by far smallest sovereign state in the world (0.44 m^2), is an enclave inside the city of Rome with a population of only 800 people, including 30 women [Vat].

[1] *country, n. and adj*, Oxford English Dictionary, URL: `http://www.oed.com/view/Entry/43085?`, last access: 2016-04-25

The *State of Palestine* has a population of 4.8 million people [Pal, as of 2016] and is also an UN observer state. However, it is totally different in terms of sovereignty: While it consists of the territories of the West Bank, East Jerusalem and the Gaza Strip, their borders were drawn in the 1949 Green Line Armistice Agreement but were never intended to be used as international boundaries [Amn]. Since then, the ongoing and complex conflict with the State of Israel lead to a difficult situations regarding the sovereignty over the territories. Therefore, the state has no clearly defined territoy. Moreover, while 114 states officially recognize the Palestinian state, almost all current main economic powers do not, including the Canada, France, Germany, Italy, the United Kingdom and the United States. None of them even voted in favor of Palestine receiving an observing status in the UN [Unia]. That means, unlike the Holy See, Palestine is not a fully sovereign and recognized state.

UN non-members with limited recognition *Kosovo* is a state Europe and and declared independence from Serbia in 2008. It has a clearly defined territory and a permanent population and is recognized by 111 UN member states. In order for Kosovo to become a full member of the United Nations, all permanent members of the security council (United Kingdom, France, Russia, China and the United States) must agree. But since Russia and China strongly support the territorial integrity of Serbia, they would veto Kosovos membership in the United Nations. Therefore, Kosovo is not even an observer state of the United Nations, although having about the same degree of international recognition as Palestine [Peo].

The status of Taiwan is a very complicated issue. An overgeneralized description of the problem, which involves two territories and two political entities, is: There is the *People's Republic as China* (commonly known as China), with full control over mainland China, and the *Republic of China*, governing the island of Taiwan. However, both political entities claim each others land. That means, there are two states claiming the exact same territory. But, since 1971 the People's Republic of China is the representative of whole China in the United Nations, including the island of Taiwan. Because it is part of the Security Council, it successfully vetos membership requests of the Republic of China. Therefore, it can not be a member of the United Nations, although it operates like an independend country by international standards: They have an own jurisdiction, issue own passports and have unofficial diplomatic relations to most countries in the world. But officially, only 22 member states of the United Nations uphold diplomatic relations to Taiwan [Rep]. To all of these states the People's Republic of China does not have any diplomatic relations, which makes also them an only partially recognized state.

There are other non-member states of the United Nations which have not yet gained broad international recognition: the Sahrawi Arab Democratic Republic (recognized by 84 UN member states [Wes]), Abkhazia (6 [Glo]), South Ossetia (5 [BBCc]), the Turkish Republic of Northern Cyprus (1 [Leo15]), Nagorno-Karabakh Republic (0 [BBCa]), Transnistria (0 [Gut14]) and Somaliland (0 [BBCb]).

UN members with limited recognition In addition to the Republic of China, there are five other member states of the United Nations that are not fully recognized by all other UN members: Armenia (not recognized by Pakistan [Tod]), the Republic of Cyprus (not recognized by Turkey [Eur]), North and South Korea (officially Democratic People's Republic of Korea and Republic of Korea, mutual non-recognition [Dav]) and the State of Israel, which 32 UN member states do not recognize [Isr].

Special Territories Additionally to countries gaining for international recognition there are territories belonging to fully sovereign countries with a varying degree of sovereignty. For example Greenland is an autonomous country within the Kingdom of Denmark, but not a sovereign state and therefore not a member of the United Nations. The same applies to the Faroer Islands (part of Denmark) and numerous overseas territories of the United Kingdom, the French Republic and the Kingdom of the Netherlands in the Carribean, the Indian Ocean or the Southern Pacific Ocean. Moreover, there are five quasi-independent countries in a so called *Free Association*: Niue and Cook Islands are associated to New Zealand and not part of the United Nations. The Marshall Islands, the Federated States of Micronesia and Palau are associated to United States, but in contrast are full UN members [Won].

This incomplete and simplified list of special cases manifests the big problem that is associated with the terms "country", "state" or "nations": There is neither a *de jure* consistent definition nor a *de facto* consistent usage of these terms. Everything breaks down to two different concepts:

4.1.2 Declaratory vs. Constituitive Theory

The declaratory theory, established in the Montevideo Convention 1933 [Yal], gives each entity the right to declare a state if it matches all of the four requirements:

1. a clearly defined territory

2. a permanent population
3. a political representation / government
4. the *capacity* to enter diplomatic relations

These four requirements make sure that a state can exist physically and politically. However, it is worth noticing that this definition does not include any actual diplomatic relations to other states, but only the capacity to enter them. Therefore the existance of a state is independent from its recognition by other states. In other words: "A country is a country when it thinks it is a country."

In contrast, the constituitive theory requires exactly that: A state can only be considered as such if it is recognized by other states. However, it is not defined anywhere by how many other states [Law]. In short: "A country is a country when other countries think that country is a country." [CGP]

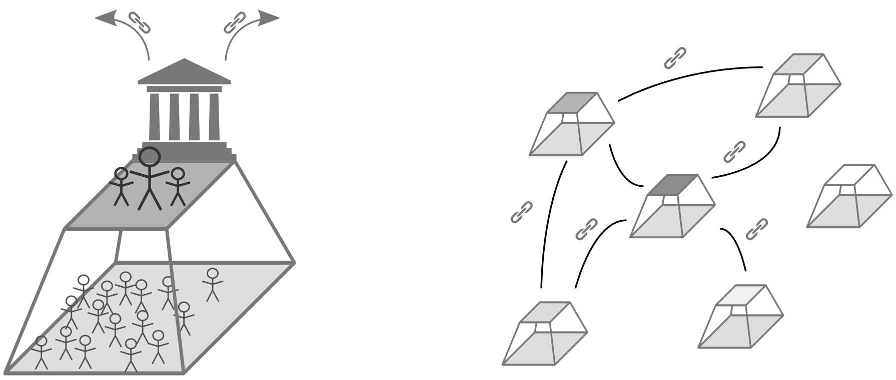


Figure 4.1: The Declaratory Theory (left) and the Constituitive Theory (right) of Statehood

Both theories have advantages and disadvantages, but the two main problems are:

1. Following the declarative theory, countries are self-classifying and potentially conflicting entities. The application of this measure would grant Kosovo, the Republic of China, Abkhazia or the Sarhawi Arab Democratic Republic full statehood. However, since their territories are contested, this would lead to overlapping territories with Serbia, China, Georgia and Morocco, which is impossible.

2. There is no superior organization that can judge if a country is a country or not. Even the United Nations fail to do so, because their membership requirements prevent states like Kosovo or the Republic of China from becoming full members. They also have no power to rule out problems regarding the independence of Transnistria or Somaliland.

Therefore it is impossible to objectivly classify an area as a country or not: nobody can say if Kosovo, the State of Palestine or Niue are countries or not. These theories have been introduced in the previous 80 years. For the time before that, a conflict-free decision of what is country is not just impossible, but also not justifyable because of a lack of jurisdiction.

That means, an historical geographic information system with the goal to visualize the development of the countries on Earth in time and space inevitably deals with uncertain information that certain parties see as wrong. Its data model can not perfectly fit self-classifying data and can not rely on an objective data source. The system has to contain approaches that deal with this problem.

4.2 Types of Uncertainty

In order to understand different types of uncertainty it is important to understand the concepts of *disagreement*, *precision* and *accuracy*.

The model in an information system tries to resemble the real world as good as possible and necessary – in this case the history of countries. If there is already a conflict in the real world, e.g. the Kasmir region which is claimed by both India and Pakistan as part of their territory, then this is a *disagreement* which also has to be proberly modeled as such in the system.

The better a model simulates the reality, the more *accurate* or correct it is. That means, the closer it gets to the target, the higher is the accuracy. *Precision* or exactness describes how similar the results are compared to each other, independent from the distance to the target. That means a precise model gets the same results over and over again (see figure 4.2).

If the border between the Principalities of Transsylvania and Wallachia is deducted from an historical map of 1600, the course of the border is is inaccurate to a certain degree, because the map does not show the real world correctly. However, it can be modelled in the system very precisely, because the coordinates of the border points are stored as floating point numbers in the data model. In contrast, there is currently no agreement upon territory of Palestine, although the different versions can be modelled very precisely. In order for the model to also be accurate in this case, it would need to support contested territories.

Hereafter the current data model introduced in section **??** is evaluated in terms of accuracy and precision.

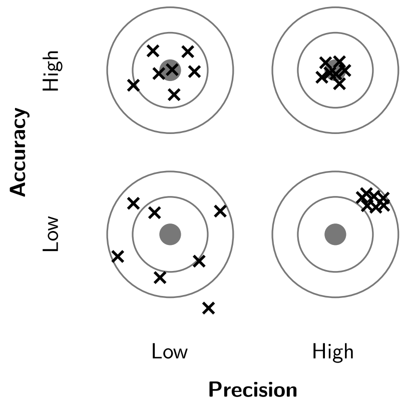


Figure 4.2: The difference between accuracy and precision

Hivents The model for historically significant happenings contains only the following meta information: name, date and location of the event. This has several shortcomings in terms of precision:

The name of an historical event can have different versions: a long, official version and a short common version. The commonly known "Treaty of Versailles" (1919) is officially called the "Treaty of Peace between the Allied and Associated Powers and Germany". Also the name is different in other languages affected by the treaty. Additionally, there can be different versions of the name from different perspectives, even within the same language, e.g. the "American Civil War" as it is known today was alternatively called "War Between the States", "War for Southern Independence" or "War of Northern Aggression" depending on the perspective. The `Hivent` model does not account for different languages and versions and is therefore not very precise.

The `Hivent.date` is supposed to represent the temporal dimension of an historical event. While an historical change itself is discrete and happens at exactly one time point, the historical event yielding this change might not. The "Congress of Vienna" which reordered the empires on the European mainland was one of the main historical events in modern European history. While the changes of the congress came into effect on 9. June 1815, the congress itself took place in Vienna from September 1814 until June 1815 which is also a timespan of interest. Another phenomenon becomes apparent in the "Convention for the Extension of Hong Kong Territory" (1898) which had a predefined length of 99 years. The treaty therefore has two dates in which historical changes happened: the date the treaty came into effect (Hong Kong becomes part of the United Kingdom) and the date it stopped being in effect (Hong Kong is handed over to China). Other interesting aspects are different calendar systems used in different parts of the world throughout history: the October Revolution in Russia (1917) happened in November in their Gregorian Calendar system, but in October in the Julian Calendar. Also timezones can play a crucial role: The German Instrument of Surrender ending World War II in Europe came into effect on 8. May 1945 at 23:01 Central European Time, so the 8. May is celebrated as the Victory Day in Western Europe. But in the Soviet Union and nowadays Russia that happened at 1:01 Moscow Time on 9. May 1945 which is why the celebration of the Victory Day there happens one day later. While the `Hivent.date` field in the data model works with timezones, it does not support different calendar systems or multiple dates associated with one `Hivent` which limits its precision.

The event location is represented by the `Hivent.location` name of a place, which can e.g. be a city, a battlefield or a region. The model is not very precise, because the actual geospatial location or region in which an historical event happened is not stored in the system. Additionally, it does not support names in different languages.

The even larger problem is an integral lack of accuracy: The whole nature of historical research is based on subjective interpretation of supposedly objective primary sources. But it is questionable if a source can actually be objective. Each bill, treaty or speech is written by somebody, each map was drawn by someone and has therefore a subjective note. Information in a primary source can be (un)intentionally incomplete, imprecise or inaccurate. The source can be biased towards the author, can contain secret passages not open to the public or its geographic information might be wrong. There are many problems involved in historical sources which makes the acquisition of objective historical data almost impossible. The further documents go back in time, the lower is the expected accuracy. Since all the information in the historical geographic information system is based on primary sources, the data in the system inherits these problems.

Areas Also the model of an abstract area, consisiting of a territory and a name, is problematic in terms of accuracy and precision. As it has been discussed in subsection 4.1 in detail, it is impossible to objectively model all areas free of conflicts. But the current model does not support the status of a territory as being contested Also, countries can be part of other autonomous (constituent) countries, like England is part of the United Kingdom of Great Britain and Northern Ireland or Greenland is part of Denmark. However, the data model does not support different levels of sovereignty, autonomy or international recognition.

The `AreaName` has the same problem than the `Hivent.name`: it differs among the languages or even among cultures using the same language. The model does not support that. But in one aspect it is more precise than for historical events, because it contains both the formal and the short name of a country.

More problematic is the `AreaTerritory`: Areas bordering international water have a constant coastline assuming that it has never changed. This is inaccurate, because coastlines gradually change all the time, therefore also the boundaries of the countries. The data model does support neither that nor international sea borders which are parts of a countries territory. The primary source for territories of countries are historical maps. They show the status of a country at one point in history or sometimes a territorial change. The process of extracting a boundary from an historical map is error-prone and yields to a loss of accuracy in each step on the way: digitizing, georeferencing and contour tracing. The level of inaccuracy depends on the resolution, the map projection and the colors used in the map. In the data model it is not possible to provide information about the expected accuracy of a territory.

Another problem is that the territory is stored as a whole polypolygon. Different parts of the border can have a different status, e.g. one part is a sea border, one is a well-established and demarcated border to neighboring country X and another part is a contested border to neighbour Y. The `AreaTerritory` data model does not account for these differences.

Accurately modelling contested territories is also problematic. It is based on the principle that there can not be overlapping territories at the same time. That means, a contested territory, for example China or occupied territories in the State of Palestine by the State of Israel can only exist once at the same time and therefore have to treated specially. But the data model does not support contested areas. To go even further, it is questionnable which areas should be included in the data model and which not. While it seems obvious to have Spain, Saudi-Arabia and Azerbaijan in the system, the question of whether or not to include the State of Palestine, Abkhazia, Somaliland or micronations like the Conch Republic in the Florida Keys is hard to answer.

Overall, the current data model poorly accounts for different levels of uncertainty in historical geographic information: imprecise and inaccurate sources, different viewpoints and interpretations, contested territories, changing coastlines or different languages. The question of the upcoming subsection is: How can the data model be extended in order to be more accurate and more precise?

4.3 Solution Approaches

In summary, the shortcomings of the current concept are:

1. General
   (a) only one language (English)
   (b) constant coastlines
2. Hivent
   (a) only one historical perspective on the Hivent name
   (b) only one discrete Hivent date
   (c) only one calendar system (Julian Calendar)
   (d) only location name, no connection to the map
3. Area
   (a) only one historical perspective on the Area name
   (b) all Areas on the same level (no dependencies)
   (c) no support for non-sovereign autonomous regions
   (d) no credibility of Areas existence (via international recognition)
   (e) only clear territories, no support for neutral zones or contested territories
   (f) no support for uncertain parts of a territory
   (g) no support for international sea borders

A higher accuracy in the data model usually leads to a higher complexity. This trade-off has to be thoroughly taken into consideration when supporting a new feature to make a model more accurate. This is why the following problems will be ignored in the rest of the thesis:

1b) Coastlines change continuously, therefore the Hivent-Based Spatio-Temporal Data Model is not suitable. A support for coastline changes would require another data model applied to coastlines. This is out of the scope of this thesis. One approach is to model international waters just like any other area with a name and a territory and change the boundaries according to an underlying continuous function. This way, the countries sharing that coastline as their international border would change likewise.

2a) The support for different historical perspectives on the same event, e.g. different names and descriptions or even different historical changes would create a research tool with great potential. It would enable the possibility for different versions of history based on alternative scenarios ("What if X would have (not) happened?"). However, this would significantly increase the complexity of the system and would also be very subjective.

2c) The introduction of different calendar systems would not increase the accuracy of the model significantly. The dates in the system must all stick to the Julian calendar, which is a reasonable requirement to avoid unnecessary complexity.

3a) see 2a)

3g) Currently each countries territory extends in a range of 3 to 12 miles (5 to 20 kilometers) [Uni82] into international waters. While this is important to accurately model a territory, it is complex, because not every country has signed the convention and each signing party can choose their range into international water. This would not just increase the complexity of the model but also create unfamiliar country territories.

In order to tackle the remaining shortcomings of the current concept, both the user interface and the data model have to be extended.

4.3.1 Extension of the Edit Mode

Two new operations (see figure 4.3) are introduced: `SCH` changes the status of an area and `REC` declares a new recognition, i.e. one country internationally recognizes another one.

Figure 4.3: Newly designed and extended buttons for edit operations.

Set New Territory Also the edit operation workflow gets changed. The second step (`SET_NEW_TERR`) defines the territory of the new area(s). Instead of drawing the whole territory as a set of polygons, the user draws one borderline at a time, geometrically as a polyline. This has the main advantage that each part of the border is treated separately.

The borderline is assigned a degree of certainty, in the interface controlled by a horizontal slider, in the model as a certainty value (`certainty` $\in\ ]0..1]$). Absolute certainty (1.0) creates a sharp and crisp line on the map. In case of uncertainty (`certainty` $\in\ ]0..1[$) three different visualization methods are introduced:

1. Blurred Border: The higher the uncertainty, the wider and more blurry the border.
2. Border Corridor: With increasing uncertainty, the offset around the actual border line extends. That creates a corridor in which the actual border is probably in.
3. Blurred Border Corridor: The combination of the first two approaches.

A simple model for the calculation of the blur factor, line width and offset distance is:

$$f(c) = -1 \cdot S \cdot ln(c) + I$$

where c is the certainty factor, $S > 0$ is a scaling factor and I is the inital value (for width: $1\ px$, for blur: 0, for offset: $0\ px$). In the example in figure 4.4 the scaling factor $S = 4$. In the Blurred Border Corridor method, the scaling factor for line width and the blur factor was halved. Further analysis and user testing are required in order to decide for one of the three approaches to be used in the system.

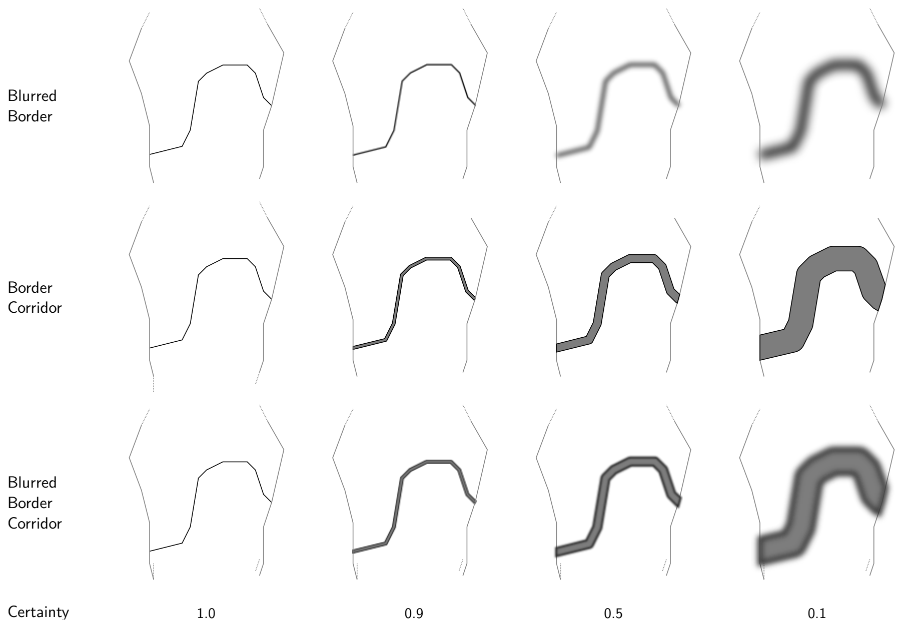


Figure 4.4: Three different methods to visualize uncertain courses of a border

Another advantage of the input of borderlines instead of territories is that once the model is further advanced, coastlines can be continuously changed according to an appropriate change model (see problem 1b). This can be applied solely to the coastlines without affecting the interior borders.

A new border point automatically snaps to an existing border point, if the mouse position is close enough to it (an appropriate threshold might be $5\ px$). This allows for a smooth workflow and is required to create closed polygons. In case the borderline is closed, it gets treated as a complete polygon and territory. When the user finished a territory by defining all surrounding polylines that create a closed ring, the polygon gets assembled. If a borderline meets another borderline at an interior node, the polyline gets split up into two parts so that each meeting point of borders is the start or end point of a polyline. This way integrity is maintained and each territory compounds of several polylines creating a set of closed polylines: a polypolygon.

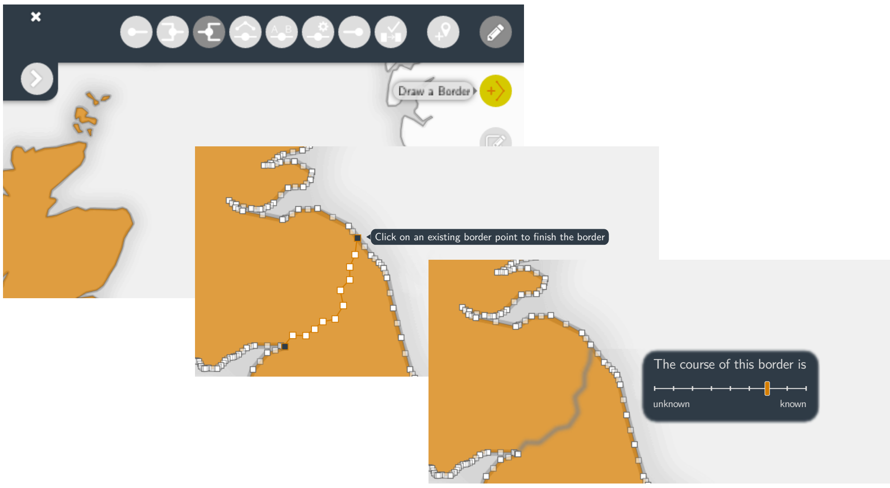


Figure 4.5: Drawing historical borders instead of full areas and definining a level of certainty.

If the created territory overlaps with an existing territory, its intersection will created a separate territory. In the next step, this territory can then be defined as a contested area or defined as a part of another area. If the step yields an empty territory that was claimed before, it can later be defined as a neutral zone or unclaimed land.

Set New Name When defining the name of an area, the user will get actual name suggestion. These result from a collection of current and historical countries from Wikipedia. That saves time for researching short and formal names of areas. In the long run, the system can be synchronized with Wikipedia or even be designed as an extension for Wikipedia articles about current or historical countries.

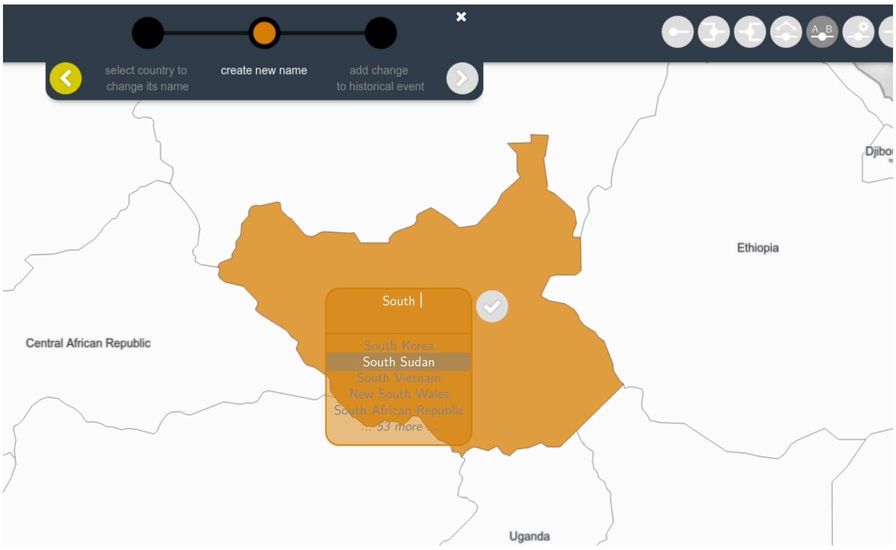


Figure 4.6: Getting suggestions for the name name from Wikipedia.

Set New Status To treat special areas differently, a new step in the edit operation workflow gets defined. After the territory and the name of a new area are defined, a special status can be assigned to it:

1. A *fully sovereign country* is a political entity with full sovereignty over its territory and people and significant international recognition, e.g. Estonia.
2. An *unclaimed land* is a territory that is not claimed by any political entity, e.g. currently Antarctica.
3. A *neutral zone* is often a buffer zone between two conflicting countries, e.g. the UN Buffer Zone in Cyprus.
4. A *contested territory* is claimed by at least two different political entities of the same hierarchical level, e.g. the Kashmir region between India and Pakistan. It is also suitable for areas that have claimed independence from a sovereign country but are not yet regonized as such, making their whole territory contested, e.g. Nagorno-Karabakh (see figure 4.7).

5. A territory can be a subordinate part of another country with a certain degree of autonomy ($\in [0..1]$). Fully subordinate parts of a country, like a US State or a German Bundesland have no autonomy (0). Autonomous countries within another country, like England to the United Kingdom or Greenland to Denmark, receive have certain a degree of autonomy ($\in]0..1[$). Full autonomy (1) would mean the territory is a fully sovereign country and the value can therefore not be set in the options.

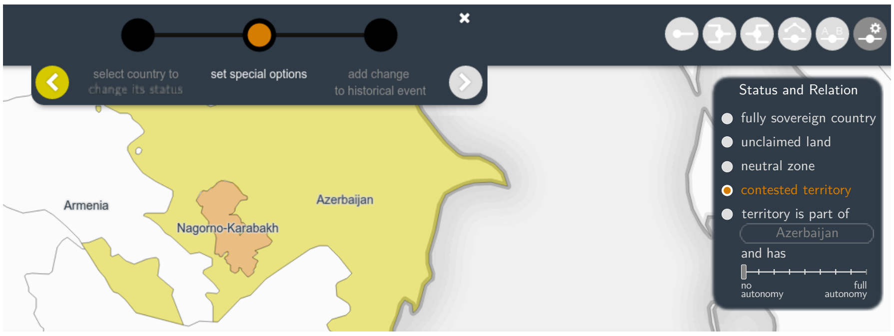


Figure 4.7: Defining a special status or relationship to a territory.

Add Historical Change The visualization of an Hivent gets split up into three parts:

1. An information section storing important meta data of the event location, the dates (timespan in which the event happened), a description and the lin kto the wikipedia article (if given).

2. A section storing all historical changes associated with that historical event. Each historical change is visualized and is assigned a date at which this event came into effect.

3. A multimedia section stores images, videos, audio files and documents and their sources associated to the historical event.

Similar to the extension of the area name step, also Hivent names can be chosen among a collection of Wikipedia articles describing historical events. Selecting a name from a wikipedia article automatically fills the information section and adds multimedia files from the wikipedia article. The historical change will automatically be entered in the section (see figure 4.8). With this separation, different historical changes at different dates can be associated with one historical events, largely increasing the `Hivent` data model.

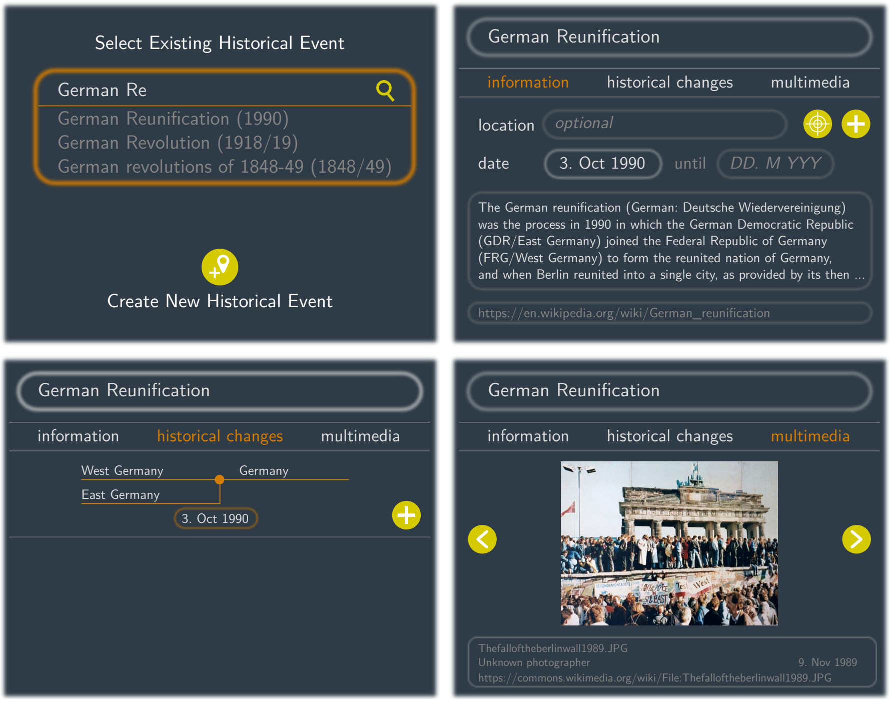


Figure 4.8: Creating a new Hivent and adding the newly created historical change.

New Area Recognition One new operation is to add the recognition of one country by another country. That is simply performed by selecting two areas on the map, whereas the first area recognized the second area. This is an historical change that can afterwards be attached to an Hivent.

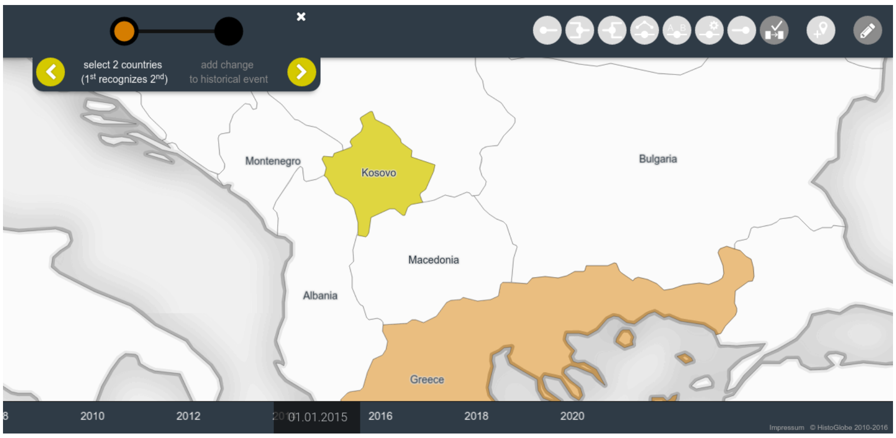


Figure 4.9: New edit operation: Recognition – sets up the recognition of one area to another.

Multi-language support In order to support different languages, a language selection is placed on the bottom right corner of the interface, on the timeline (see figure 4.10). This changes the language of the whole interface and loads the translations of the area names and the Hivent names, locations and descriptions in the newly created language. If a term is not defined in the language, the fallback language (English) is used instead.

Figure 4.10: Changing the language in the user interface.

4.3.2 Extension of the Data Model

To account for the changes in the interface, also the data model has to be adapted. The main changes to the original data model developed in section **??** are:

1. Creation of a `Multilang` entity to store a name of an Hivent, its location or an Area name in different languages.
2. Outsourcing of the `HiventLocation` into an own entity to identify a location with a name and a geospatial reference.
3. Creation of a `Multimedia` entity to manage multimedia files associated to an Hivent.
4. Attachment of a date to an `HistoricalChange`.
5. Inclusion of the `formal_name` into the `Area` model to emphasize it as the identifier of an area.
6. Creation of an `AreaBorder` with a `borderline`. A set of `AreaBorders` create one `AreaTerritory` which is associated to the Area. Each change of an `AreaBorder` creates one or two new `AreaTerritory`/ies.

7. Creation of an `AreaStatus` an an `AreaRelation` to account for special status of an area alone or in relation to another area with a certain level of autonomy.

8. Creation of an `AreaRecognition` to account for international recognition of one area to another one.

9. Adaption of the `AreaChange` entity to model a change of each possible property of an area.

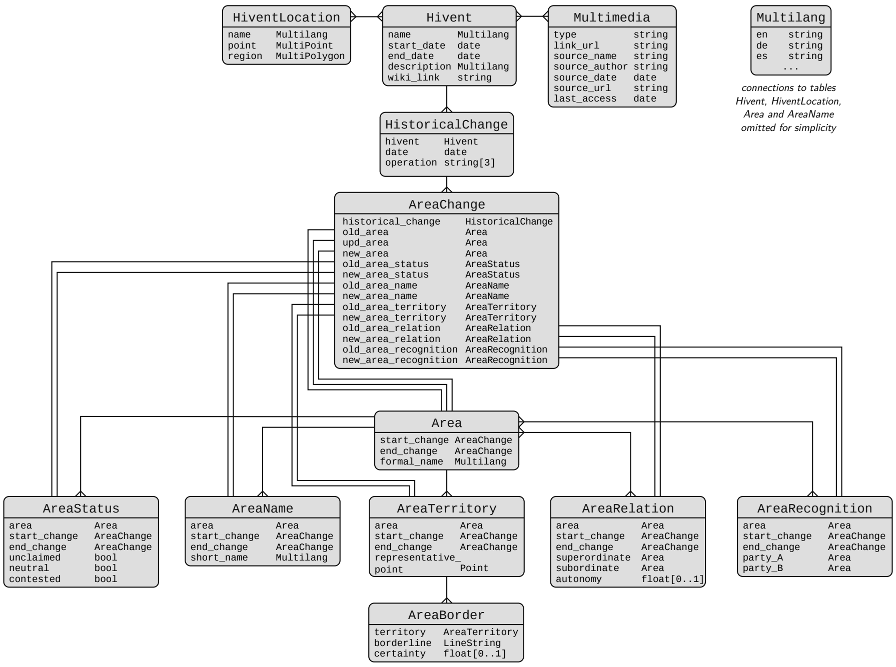


Figure 4.11: The new data model to support the developed approaches regarding uncertainty

Chapter 5

Summary

5.1 Results

Research Questions

5.2 Problems

5.3 Future Work

step further: temporal GIS to narrative GIS

idea: explain history with spatial narratives geographically contextualize events and interactions organizing principle: time

extend the pure presentational purpose of st data to analytical purpose, e.g. where have most border changes take place in previous 200 years?

Another problem for historians is that they do not necessarily need a tool to better visualize existing knowledge (e.g. historical maps), but to generate new knowledge by analyzing spatio-temporal coherences or distributions in historical data. Spatio-temporal reasoning is still an open field and not easily possible with existing HGIS

[KH08, p. 268], [GG14, p. xii]. space-time premise by Gaddis 2002 time and space equal importance event what significantly has happend and by whom? (singularity!) process how something has happened? (event+activity =¿ trigger of process) change driven by process spatiotemporal data defines all above three

extend area model to hierachies (country -¿ states -¿ counties/cities)

Bibliography

[AHA] AHA - American Historical Association (Andrews, Thomas and Burke, Flannery). *What Does It Mean to Think Historically?* URL: http://historians.org/publications-and-directories/perspectives-on-history/january-2007/what-does-it-mean-to-think-historically [cited 23 10 2015].

[All84] James F. Allen. Towards a general theory of action and time. *Artif. Intell.*, 23(2):123–154, 1984.

[Amn] Amnon Sella. *Custodians and Redeemers: Israeli Leaders' Perceptions of Peace, 1967-79*. URL: http://www.jstor.org/stable/4283114?seq=1#page_scan_tab_contents [cited 26 04 2016].

[Arm92] C. Armenakis. Estimation and organization of spatio-temporal data. *Proceedings of the Canadian Conference on GIS92*, 1992.

[BBCa] BBC News. *Nagorno-Karabakh profile*. URL: http://www.bbc.com/news/world-europe-18270325 [cited 26 04 2016].

[BBCb] BBC News. *Somaliland profile*. URL: http://www.bbc.com/news/world-africa-14115069 [cited 26 04 2016].

[BBCc] BBC World News. *Russia recognises Georgian rebels*. URL: http://news.bbc.co.uk/2/hi/7582181.stm [cited 26 04 2016].

[BCH10] D.J. Bodenhamer, J. Corrigan, and T.M. Harris. *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Spatial humanities. Indiana University Press, 2010. URL: https://books.google.de/books?id=8VkhPkIRM4EC.

[Bol08] P. Bolstad. *GIS Fundamentals: A First Text on Geographic Information Systems*. Eider Press, 2008. URL: https://books.google.com/books?id=2rjYPQAACAAJ.

[Bus] BusinessDictionary. *What is a system? definition and meaning*. URL: http://www.businessdictionary.com/definition/system.html [cited 04 11 2015].

[CGP] CGP Grey on YouTube. *How Many Countries Are There?* URL: https://www.youtube.com/watch?v=4AivEQmfPpk [cited 26 04 2016].

[CT95] Christophe Claramunt and Marius Thériault. Managing time in gis: An event-oriented approach. In *Proceedings of the International Workshop on Temporal Databases: Recent Advances in Temporal Databases*, pages 23–42, London, UK, UK, 1995. Springer-Verlag. URL: http://dl.acm.org/citation.cfm?id=647207.719698.

[Dar] Dare Obasanjo. *An Exploration of Object Oriented Database Management Systems*. URL: https://www.cs.umd.edu/class/spring2002/cmsc434-0101/MUIseum/applications/odms.html [cited 14 May 2016].

[Dav] David Scofield, AsiaTimes. *Seoul's double-talk on reunification*. URL: http://www.atimes.com/atimes/Korea/GA04Dg01.html [cited 26 04 2016].

[DISP08] Stanislas Dehaene, Véronique Izard, Elizabeth Spelke, and Pierre Pica. Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures. *Science (New York)*, pages 1217–1220, 2008.

[Dra] Nicolae Dragulanescu. Knowledge Map of Information Science: Data, Information, Knowledge. URL: http://www.success.co.il/is/dik.html [cited 05 11 2015].

[Emi] Emil Mikulic, unix4lyfe. *Time*. URL: https://unix4lyfe.org/time/?v=1 [cited 15 May 2016].

[Eth07] Philip J. Ethington. Placing the Past: 'Groundwork' for a Spatial Theory of History. *Rethinking History*, 11:465–493, Dec 2007.

[Eur] European Parliament. *Turkey and the problem of the recognition of Cyprus*. URL: http://www.europarl.europa.eu/meetdocs/2004_2009/documents/nt/553/553930/553930en.pdf [cited 26 04 2016].

[Fra] Fraczek, Witold (ESRI). *Mean Sea Level, GPS, and the Geoid*. URL: http://www.esri.com/news/arcuser/0703/geoid1of3.html [cited 27 10 2015].

[Fra98] Andrew U. Frank. Different Types of Times in GIS. *Spatial and Temporal Reasoning in Geographic Information Systems*, pages 40–62, 1998.

[Geo] Geokov. *Map projection - types and distortion*. URL: http://geokov.com/education/map-projection.aspx [cited 27 10 2015].

[GG14] I.N. Gregory and A. Geddes. *Toward Spatial Humanities: Historical GIS and Spatial History*. The Spatial Humanities. Indiana University Press, 2014. URL: https://books.google.com/books?id=JYQxAwAAQBAJ.

[Glo] Global Security. *Georgia mocks Nauru's recognition of Abkhazia*. URL: http://www.globalsecurity.org/military/library/news/2009/12/mil-091215-rianovosti11.htm [cited 26 04 2016].

[Gut14] Perparim Gutaj. The foreign policy of counter secession: Preventing the recognition of contested states by james ker-lindsay. oxford: Oxford university press, 2012. 224pp., £50.00, isbn 978 0 199 69839 4. *Political Studies Review*, 12(2):278–278, 2014. URL: http://EconPapers.repec.org/RePEc:bla:pstrev:v:12:y:2014:i:2:p:278-278.

[Häg70] Torsten Hägerstrand. What about people in regional science? *Papers of the Regional Science Association*, 24(1):6–21, 1970. URL: http://dx.doi.org/10.1007/BF01936872, doi:10.1007/BF01936872.

[HW90] Gary J. Hunter and Ian P. Williamson. The development of a historical digital cadastral database. *International journal of geographical information systems*, 4(2):169–179, 1990. doi:10.1080/02693799008941538.

[Isr] Israel Ministry of Foreign Affairs. *Israel's Diplomatic Missions Abroad*. URL: http://www.mfa.gov.il/mfa/abouttheministry/pages/israel-s%20diplomatic%20missions%20abroad.aspx#note [cited 26 04 2016].

[KH08] A.K. Knowles and A. Hillier. *Placing History: How Maps, Spatial Data, and GIS are Changing Historical Scholarship*. ESRI Press, 2008. URL: https://books.google.com/books?id=VN1v7rzhSQEC.

[Lan88] G. E. Langran. A framework for temporal geographic information systems. *Cartographica*, 25, 1988.

[Lan89] G. E. Langran. Time in geographic information systems. 1989. UMI Order No: GAX90-00269.

[Law] Law Teacher. *Declaratory And Constituitive Theories Of State*. URL: http://www.lawteacher.net/free-law-essays/constitutional-law/declaratory-and-constitutive-theories-of-state.php [cited 26 04 2016].

[Leo15] Leon Hadar, AntiWar.com. *In Praise of 'Virtual States'*, 2015. URL: http://antiwar.com/hadar/?articleid=8042 [cited 26 04 2016].

[Nak] Frieder Nake. Data, Information and Knowledge. URL: http://viola.informatik.uni-bremen.de/typo/fileadmin/media/DataInformKnowledge.pdf [cited 05 11 2015].

[Ora] Oracle. *1970s — Defying Conventional Wisdom*. URL: http://www.oracle.com/us/corporate/profit/p27anniv-timeline-151918.pdf [cited 14 May 2016].

[OS01] T. Ott and F. Swiaczny. *Time-Integrative Geographic Information Systems: Management and Analysis of Spatio-Temporal Data*. Number v. 1 in Time-integrative Geographic Information Systems: Management and Analysis of Spatio-temporal Data. Springer Berlin Heidelberg, 2001. URL: https://books.google.com/books?id=0Yvg_PErYRkC.

[Pal] Palestinian Central Bureau of Statistics. *Estimated Population in the Palestinian Territory*. URL: http://www.pcbs.gov.ps/Portals/_Rainbow/Documents/gover_e.htm [cited 26 04 2016].

[PD95] Donna Peuquet and Niu Duan. An event-based spatiotemporal data model (estdm) for temporal analysis of geographical data. *International Journal of Geographical Information Systems*, 9(1):7–24, 1995. URL: http://dblp.uni-trier.de/db/journals/gis/gis9.html#PeuquetD95.

[Peo] People of Kosova. *Who Recognizes Kosova?* URL: http://www.kosovothanksyou.com [cited 26 04 2016].

[Peu99] D. J. Peuquet. Time in GIS and geographical databases. *Geographical information systems*, pages 91–102, 1999.

[PTKT04] Nikos Pelekis, Babis Theodoulidis, Ioannis Kopanakis, and Yannis Theodoridis. Literature review of spatio-temporal database models. *Knowl. Eng. Rev.*, 19(3):235–274, September 2004. URL: http://dx.doi.org/10.1017/S026988890400013X, doi:10.1017/S026988890400013X.

[Raz12] Ale Raza. Working with spatio-temporal data type. *International Archives of the Photogrammetry, Remote Sensing and Sptial Information Sciences*, XXXiX-B2:5–10, 2012.

[Ren96] A Renolen. History Graphs: Conceptual Modeling of Spatio-temporal Data. *GIS Frontiers in Business and Science*, 2, 1996.

[Rep] Republic of China. *Ministry of Foreign Affairs*. URL: http://www.mofa.gov.tw/en/default.html [cited 26 04 2016].

[RGS] RGS - Royal Geographical Society. *What is geography?* URL: http://www.rgs.org/geographytoday/what+is+geography.htm [cited 23 10 2015].

[Sol14] A.C. Solana. *Spatio-Temporal Narratives: Historical GIS and the Study of Global Trading Networks (1500-1800)*. Cambridge Scholars Publisher, 2014. URL: https://books.google.com/books?id=oqIxBwAAQBAJ.

[Tod] Today.az. *For Azerbaijan Pakistan does not recognize Armenia as a country*. URL: http://www.today.az/news/politics/30102.html [cited 26 04 2016].

[Unia] United Nations. *General Assembly Votes Overwhelmingly to Accord Palestine 'Non-Member Observer State' Status in United Nations*. URL: http://www.un.org/press/en/2012/ga11317.doc.htm [cited 26 04 2016].

[Unib] United Nations. *Member States*. URL: http://www.un.org/en/member-states/index.html [cited 26 04 2016].

[Uni82] United Nations. *United Nations Convention on the Law of the Sea*, 1982. URL: http://www.un.org/depts/los/convention_agreements/texts/unclos/unclos_e.pdf [cited 28 04 2016].

[Uot] Uotila, Urho A. (Britannica.com). *geoid — geology*. URL: http://www.britannica.com/science/geoid [cited 27 10 2015].

[Vat] Vatican City State. *Population*. URL: http://www.vaticanstate.va/content/vaticanstate/en/stato-e-governo/note-generali/popolazione.html [cited 26 04 2016].

[Wes] Western Sahara Online. *Proclamation Of The Saharawi Arab Democratic Republic*. URL: https://wayback.archive.org/web/20080502035045/http://www.wsahara.net/sadr.html [cited 26 04 2016].

[WHM90] MICHAEL F. WORBOYS, HILARY M. HEARNSHAW, and DAVID J. MAGUIRE. Object-oriented data modelling for spatial databases. *International Journal of Geographical Information Systems*, 4(4):369–383, 1990.

[Won] WonderWhy on YouTube. *How many countries are there in the world?* URL: https://www.youtube.com/watch?v=4vsPB_lbo94 [cited 26 04 2016].

[Yal] Yale Law School. *Convention on Rights and Duties of States*. URL: http://avalon.law.yale.edu/20th_century/intam03.asp [cited 26 04 2016].

[Yua96a] May Yuan. Modeling Semantical, Temporal, and Spatial Information in Geographic information Systems. *Geographic Information Research: Bridging the Atlantic*, pages 334–347, 1996.

[Yua96b] May Yuan. Temporal gis and spatio-temporal modeling. In *Proceedings of Third International Conference Workshop on Integrating GIS and Environment Modeling, Santa Fe, NM*, 1996.

[Zha11] Ziliang Zhao. Design and Implementation of an Object-Oriented Space-Time GIS Data Model. *Master's Thesis, University of Tennessee*, 2011. URL: http://trace.tennessee.edu/utk_gradthes/1043/ [cited 07 11 2015].

[Zwa] Vladimir Zwass. information system. URL: http://www.britannica.com/topic/information-system [cited 05 11 2015].

Stuff