

A FORMAL MODEL FOR HISTORICAL ATLASES AND HISTORICAL KNOWLEDGE

Until the end of the 20th century, history was mainly a social science, often closer to the humanities than to scientific rigor. Thanks to the arrival of information technology, this is changing. In this document we intend to examine the issue of historical knowledge from the perspective of historical atlases and to implement a formal model as a solution.

CONTENTS

A - HISTORICAL ATLAS DEFINITION AND ISSUES	2
A.1 - Main Issues in the Representation of Ancient Polities.....	2
A.2 - Fluctuating Entities and Fluctuating Criteria.....	3
A.3 - Polity Definition.....	4
A.4 - Issues Encountered with the EurAtlas Historical Atlas.....	5
B - A NEW APPROACH TO PROVIDE A SOLUTION	7
B.1 - A Formal Model for Geopolitical Knowledge.....	7
B.2 - Classes and Properties in a Formal Model	10
B.3 - A System Based on a Formal Model for Historical Knowledge.....	11
B.4 - Advantages and Defects Resulting from a System Based on a Formal Model	12
C - FORMAL MODEL CONCEPTION	14
C.1 - Design of the General Model.....	14
C.2 - Modern Legal Polities Catalogue and Personal Ties.....	14
C.3 - Attempts to Combine the Polities Catalogue with the General Model.....	16
C.4 - Necessity to Create a Generic Polities Catalogue.....	16
D - THE PROPOSED FORMAL MODEL	18
D.1 - General Formal Description of the Hierarchical-Cooperative-Convergent Model	18
D.2 - Description of Polity Types and Their Properties.....	21
D.3 - Description of the Relations and their Properties.....	21
D.4 - Territories and Territorial Control.....	22
D.5 - Automatic Derivation of the Territorial Control Transfer and Exclusivity.....	22
E - ADVANCED WORK	25
E.1 - Advanced Issues.....	25
E.2 - Missing Data and Probability.....	25
E.3 - Two Levels of Analysis.....	26
REFERENCES	26

A - HISTORICAL ATLAS DEFINITION AND ISSUES

Historical atlases are clear and informative. They provide an overview of political evolution (A1). However, since the definition of a polity varies through time and since historical narration is centered on specific places or countries, it is difficult to choose criteria allowing to determine the features to display (A2). It could be possible to draw a definition of a polity but the result would be vague and anachronistic (A3). The Euratlas historical atlas project is an attempt to provide a solution to these issues but, has still not reached the satisfactory levels we wish to achieve (A4).

A.1 - Main Issues in the Representation of Ancient Polities

Historical atlases are clear and informative because they offer an overview of cases that texts cannot provide. Furthermore, they allow one to consider and to compare changes in political entities over time and thus to attempt a synthesis of the historical evolution of a region or continent.

However, serious problems of anachronisms or confusions are emerging when political entities are represented at several historical moments separated by long time intervals. Indeed, the very definitions of a state, an administrative division or of a “dependent country” for instance, are not the same throughout history. In the 20th and 21st C., we generally accept that an independent state is a territory with a population living under a system of government which is fully sovereign. However, this definition is not unanimous. For instance, there is a trend aiming to link the notion of sovereignty with that of recognition by the international community. Moreover, the common law system differentiates between 'state' and 'sovereign state' while the continental law systems do consider that a 'state' is always sovereign. It should be recalled also that the concept of sovereignty was defined only gradually by Western jurists from 1648 on (Peace of Westphalia).



Cartographers use some usual symbols (thick borders, various colours) to represent different categories of polities. An example is the commonly referred to as “Charlemagne’s Empire”. This image seems to feature the extent of an empire, like a modern state. However, by reading the primary sources (Einhard & Frankish Annals), we learn that the polity ruled by Charlemagne was not a state as in today’s sense of the term and that its effective extent was mainly concentrated around the areas of Aachen, Paris, Mainz, Milan and maybe Rome. Also, Charlemagne power was in no way similar to modern public power. Note especially the Saxony and North Italy areas which could be considered as “dependent” but are not such according to modern criteria while Aquitaine, Gascony, Bavaria and the Spanish March may correspond to this definition if the sources were more accurate.

As an example of the difference between common law and continental law, simply compare the Wikipedia entries 'List of countries and dependencies by population' (English) and 'Liste des pays par population' (French) and you will notice that 203 entities are listed in the French version versus 244 in the English one.

A.2 - Fluctuating Entities and Fluctuating Criteria

Indeed, historical narration is generally centered on the present moment and on a particular country or group of countries. This means that there are a multitude of books, mostly written during the half century prior to reading, which specifically describe the history of a particular country. Thus, the historical narrative explains the temporal evolution of the focused country in the light of the political situation prevailing at the time of its elaboration. Moreover, it happens too often that the text seeks a justification for modern political situations in the events of the past.



Here is a map labelled "The Netherlands in the Time of the Roman Empire". Of course, the Netherlands came into existence only about 1500 years after the reported period and the red borders are valid only since 1830. Thus, the locations of the mentioned populations are totally anachronistic and only make sense to a modern reader. Similar maps are often presented with modern political aims. One will see, for instance, a map of a today state called "XXX at the time of its greater extent" with ancient borders superimposed over modern borders even if this specific territorial extent lasted, in fact, less than 10 years.

It follows that, if we try to factually synthesize the history of a continent or a cultural area, like Europe for example, we see some, more or less, immutable states - France, England, Portugal, Russia, Denmark - surrounded by an immeasurable number of political entities whose status is not clear. Generally, authors from one of these entities describe it as sovereign while external researchers are more cautious and try to accurately define the complex status of that territory. Strictly speaking, it is this definition that leads to anachronism or confusion.

As a matter of fact, historians are faced with the choice either to describe the old entity with modern standards and thus to fall into anachronism or to describe this entity according to criteria valid at the time and to fall into great complexity which may lead to confusion because these criteria were not the same in all regions and, in addition, the authorities involved - that is the neighbouring countries - did not agree on the nature or existence of a criterion, not to mention that our sources are often scarce in these subjects.

independence and so on. There are also various international organizations some of which - for instance, the European Union - having extensive supranational powers and, moreover, a number of political entities are practically independent - *de facto* - while they should not be such legally - *de jure*. As mentioned above, complications of this kind are not unique to the 21st century and already existed in ancient times. Thus, we could draw up a list of the various categories of political entities. Such a list would include:

- the confederation,
- the federation,
- the federated state,
- 1st level administrative divisions,
- the personal union,
- the unitary state,
- the dependent state,
- 2nd level administrative divisions etc.

Consequently, it would be possible to draw a historical atlas or create a history database by taking into account only the above categories and by specifying some of them. A confederation, for instance, could be a league or alliance or even a supranational *Staatenverbund*¹ and a personal union could also be a real union. It would then be possible to precisely define each of these categories however, as aforementioned, this list would correspond only to the current legal order, the one that has gradually been built mainly in Western Europe since 1648.

When one wants to describe older political entities, of the thirteenth or twelfth century or of Antiquity, the concepts of confederation, federal or unitary state become very vague and even anachronistic, not to mention that some old legal categories (e.g. "rule by right of conquest") have today totally disappeared. In the Middle Ages, the concept of a state entity separate from the person of the entity's ruler was very poorly understood or, in the best case, partially developed and justified by religious considerations.

A.4 - Issues Encountered with the Euratlas Historical Atlas

The Euratlas Periodis historical atlas, which has been published on the web since 2001², was realized in accordance with the preceding definitions. However, to lessen the impact of anachronisms, we had chosen to draw only 3 simplified categories of political entities:

- 1) sovereign states, that is the entities which are usually described as independent countries in textbooks or reference books,
- 2) administrative divisions, that is the provinces, counties, cantons, districts etc. generally classified in this group,
- 3) dependencies, that is a very large middle group which encompasses all the entities considered as neither totally independent nor totally integrated into an administrative structure.

This solution offers maps easy to read but does not make it possible to accurately cover all real situations and causes several problems. It is necessary, indeed, to add an extra *supranational entity* category to describe the Holy Roman Empire and the European Union. Moreover, several specific political situations cannot be represented in an acceptable way and need additional explanation: entities ruled by multiple authorities, various cases where the *de facto* situation does not fit the *de jure situation* as well as the personal or real unions which are, actually, a particular case of confusion between the state entity and its ruler as a person.

¹ System of multi-level governance in which states work closely together but retain their own sovereignty.

² <https://www.euratlas.net/history/europe/index.html>



This is a detail from the Euratlas Periodis historical atlas, year 1700. Please note the purple line that demarcates the Holy Roman Empire supranational entity and the labels showing that England, Scotland and the Netherlands were a personal union in that very year. Furthermore, you can observe the unclear statuses of some smaller entities like Valais, the Danish part of Schleswig-Holstein-Gottorp or the western part of the Duchy of Bar, the so-called Barrois Mouvant which was simultaneously a dependency of France and of Lorraine.

These problems are not insignificant because, following the online publication of the Euratlas Periodis historical atlas, we regularly receive “correction requests” for the political status of a given entity in a particular century.

These uncertainties and approximations, present in the Euratlas historical atlas but also in the most famous historical atlases, do not only give an erroneous view of political history but also fuel more or less chauvinistic, pretentious, boastful and far from factual reality militant ideas.

Moreover, this issue has a significant impact on daily life since it is common for companies providing international online services - e.g. tourism, geographical or positioning - to be very embarrassed when qualifying some countries of the world, not to mention that many political movements justify their claims by wishful thinking or biased definitions of the concept of State.

Finally, since the aim is to build an accurate historical database, it is necessary to make an accurate distinction between the various kinds of existing or previously existing entities, otherwise the result will be only a vast, tremendous and inconsistent catalogue.

B - A NEW APPROACH TO PROVIDE A SOLUTION

It seems that a more general approach based on a formal model would offer an opportunity to manage these issues with geopolitical and historical knowledge. Thus, we first make the distinction between mapping and database, formalization and automated formal model (B1). Follows a short explanation of what is a formal model with an example (B2). Then we describe a formal model for historical knowledge management with an example from an institution for museum collections organization (B3). A computer-based formal system offers flexibility and automatic synthesis and may increase objectivity and consistency. Moreover, it facilitates work and achievement. (B4)

B.1 - A Formal Model for Geopolitical Knowledge

In order to overcome the issues of traditional historical atlases, we chose an approach which relies on the use of automated reasoning and information management techniques.

This involves firstly making a clean distinction between historical mapping, and knowledge base where the effective information is stored and structured. Historical mapping generally means the display of a map, a graph or whatever kind of image allowing the representation of different information pertaining to the field of history, like inhabited places, roads, monuments, languages, religions and, mainly, various kinds of political entities. A knowledge base offers the possibility to store and structure a huge amount of information about these very inhabited places, roads, monuments and different kinds of political entities.

In a second stage, it is necessary to formalize information. That means to build the knowledge base in such a way that the geopolitical facts can be described and stored with great accuracy. This implies establishing an accurate and well-defined vocabulary in which each term has to be used with one single exact and measurable meaning, like in mathematics or computing science. Such is formalization. However, this does not necessarily mean neglecting uncertainties, or even indeterminacies, that may be inherent to some concepts. Specific methods, like probabilistic³ approaches, are particularly well-suited to handle these issues. When using a formal method, uncertainty can be handled but this requires the precise identification of its perimeter.

Furthermore, such a preliminary distinction not only allows to create consistent maps but it also facilitates research. In fact, any investigation, even in humanities, has to meet strict scientific requirements in order to produce reliable results and to open the path towards new discoveries. See, for instance, the commentary "Why do We Need Mathematical History?" by Peter Turchin⁴. Thus, one of the main benefits of historical knowledge formalization would be that intellectual processes usually applied by researchers could be realized automatically by a computer.

Of course, there is still a long road to cover before historical knowledge will become formalized enough to allow computers to perform all complex reasoning as historians do. Moreover, it is almost impossible that some parts of reasoning could be approached for a relatively long time. However, this will certainly be done, step by step, collaboratively by researchers throughout the world. The more we advance in such processes, the more we will have access to richer, more consistent and more connected knowledge of history⁵.

To summarize, it is necessary to say that the proposed method does not primarily consist in map drawing but in the construction of a database that records historical information which will be, subsequently, used to generate maps. Even if the ultimate goal remains a sort of historical atlas, and

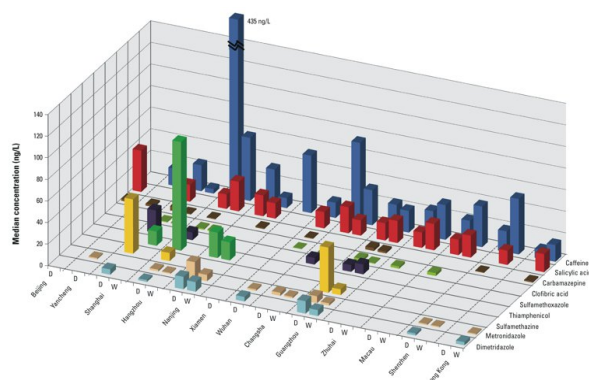
³ Model where there are multiple possible results, each having varying degrees of certainty or uncertainty of occurrence.

⁴ <http://peterturchin.com/cliodynamics/why-do-we-need-mathematical-history/> last viewed 27 October 2017.

⁵ See this case too: Brennan, Rob, Kevin Feeney, Gavin Mendel-Gleason, Bojan Bozic, Peter Turchin, Harvey Whitehouse, Pieter Francois, Thomas E. Currie, and Stephanie Grohmann. "Building the Seshat Ontology for a Global History Databank." In *The Semantic Web. Latest Advances and New Domains*, 693–708. Lecture Notes in Computer Science. Springer, Cham, 2016. https://doi.org/10.1007/978-3-319-34129-3_42

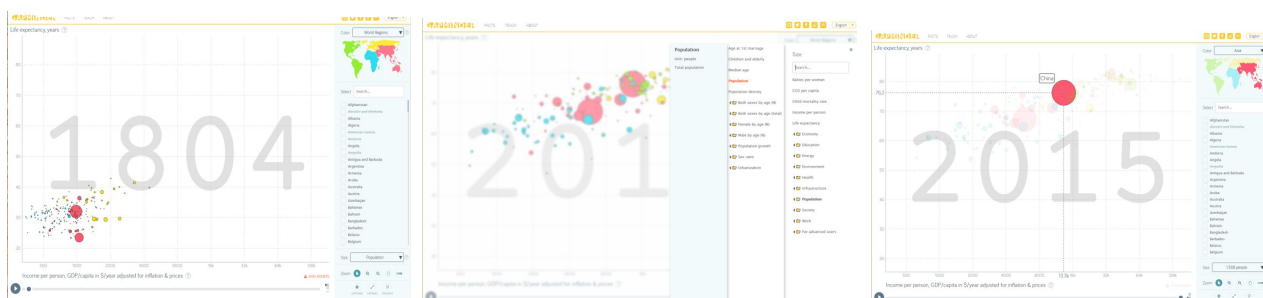
even if the focus is set on the representation of geopolitical and geo-historical knowledge, the overall structure will be different. Maps will not be a central element of the system but one possible representation, among others, of its content. This is the essence of the formal model we intend to build and is the core of our work.

Let us now add a few concrete illustrations to the above explanations. A person is limited by the capacities of human intellect. The speed of reasoning and the amount of information the brain can process are narrow while a machine can always exceed these limits when its computing power increases. A human can understand 2 or 3, maximum 4, correlated data blocks whereas a computer can process a very large number, depending on its processing power. This can be seen by visualizing any table summarizing economic, social or medical data. Most of these representations are in the form of two-dimensional or three-dimensional bar graphs and, sometimes, with a fourth component allowing to catch several images in parallel, but it is practically impossible to consider the synthesis of more data. A machine instead easily handles a larger number of data blocks, draw some conclusions, and thereafter present one or several syntheses as text or graphic, optionally in a table with several tabs. It is not even possible to exhaustively list all the possibilities that this opens because data processing is constantly evolving.



Here on the left is a common three-dimensional bar chart synthesizing 3 blocks of data. We can understand such an image, but it remains relatively difficult to draw conclusions from it.

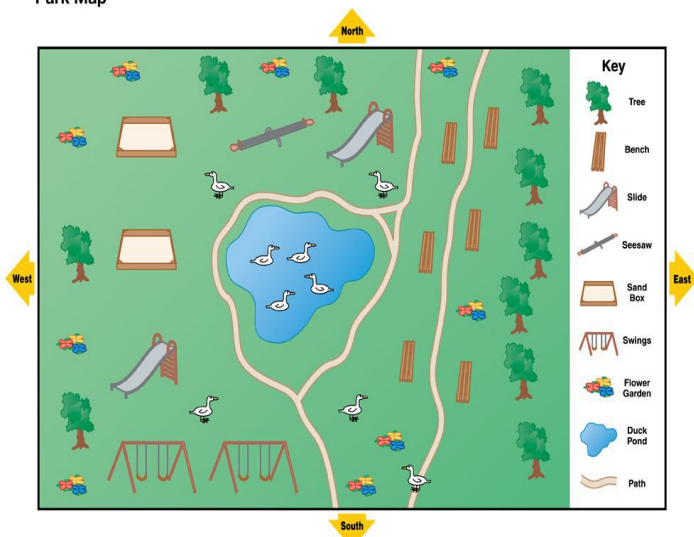
Compare it with the multi-dimensional charts drawn from the noteworthy Gapminder website, here below. These graphs are displaying data blocks on the abscissa (by column), on the ordinate (by line) while the timeline is moveable, and more data blocks are available in tabs to the right of the screen. For further examples, please see: <http://www.gapminder.org/>



Regarding the specific case of history, we can consider the fact that all historical events are interdependent and thus, it may be necessary to consider a vast number of different events in order to globally reason a specific situation. A human would always be limited in the number of interdependent factors he or she could process, while a computer is potentially able to take all reasoning dependent factors into account, provided the information has been formalized and entered in the system.

Respecting formalization, one can also make an analogy with cartography proper. Indeed, every map, geographical or historical, is a kind of small formal model since its conventional signs, generally listed on an insert or at the beginning of the work, constitute formal definitions of colours, lines, and symbols used on the map. Implicitly, these definitions determine categories of real-world objects and thus of some aspects of our knowledge. Somehow, a map is a representation of knowledge based on the formal model defined by conventional signs.

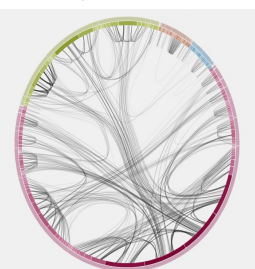
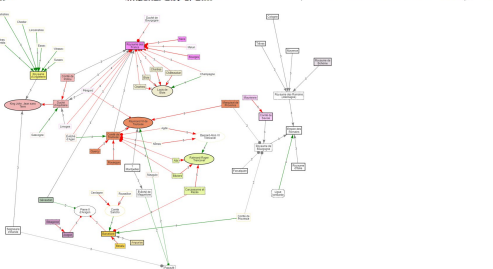
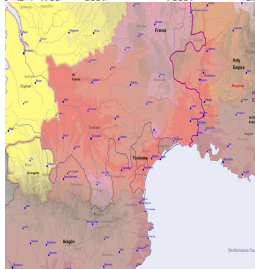
Park Map



Here is a simple plan of a park. Even if it seems trivial, we have to keep in mind that every map or plan constitutes a formal model of reality. The symbols listed on the right side of the image refer to actual objects while the colours are considered so evident that no formal definition is given. Note also that the scale definition is missing from this model.

However, maps are limited in the nature and quantity of information they convey otherwise they become too complex. You can use many graphical symbols or text fonts to show different kinds of information, but the more information you put using graphical symbols the more difficult they become to read. Moreover, using graphic symbols strongly reduces the type of information that can be represented. Thus, map makers are often forced to omit part of the information and in some cases, even omit some known and important element that could be useful to qualify a situation. On the contrary, a system based on a strong distinction between representation and knowledge permit, to both store the information with all of its known details, and to automatically derive various depictions of the data. In addition to the possibility of generating different map representations of the knowledge, this also makes it possible to display the data in the form of a graphic chart or any other kind of visual representation.

A	B	C	D	E	F	G	H	I	J
1	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
2	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
3	1700	4870	-14870	-14870	Q	1802	6084	1802	Q
4	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
5	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
6	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
7	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
8	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
9	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
10	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
11	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
12	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
13	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
14	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
15	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
16	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
17	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
18	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
19	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
20	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
21	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
22	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
23	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
24	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
25	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
26	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
27	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
28	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
29	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
30	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
31	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
32	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
33	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
34	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
35	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
36	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
37	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
38	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
39	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
40	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
41	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q
42	1700	6084	1802	6084_1802	Q	1802	6084	1802	Q



Even if not obvious, a geo-historical database can be converted into a historical map, a chart or any visualization graph.

B.2 - Classes and Properties in a Formal Model

There are different ways to create a formal model, but in this case, we use a formalism referred to as object-oriented which was developed in the framework of software engineering⁶. It was subsequently extended through the evolution of the semantic web, in other words, the development of a general framework to represent formal knowledge so that it can be interlinked and, thus, exploited by robots.

Specifically, the design of an object-oriented model consists of:

A - establishing classes, i.e. categories, of objects which can be represented in the model, e.g. the class of *Persons* or the class of *Places*;

B - defining properties that can be attributed to the different objects of each class, e.g. the objects in the class *Persons*, could have the properties *DateOfBirth* and *PlaceOfBirth*.

Note that the properties of an object may:

B1- refer to a numeric or textual value, like in the case *DateOfBirth*,

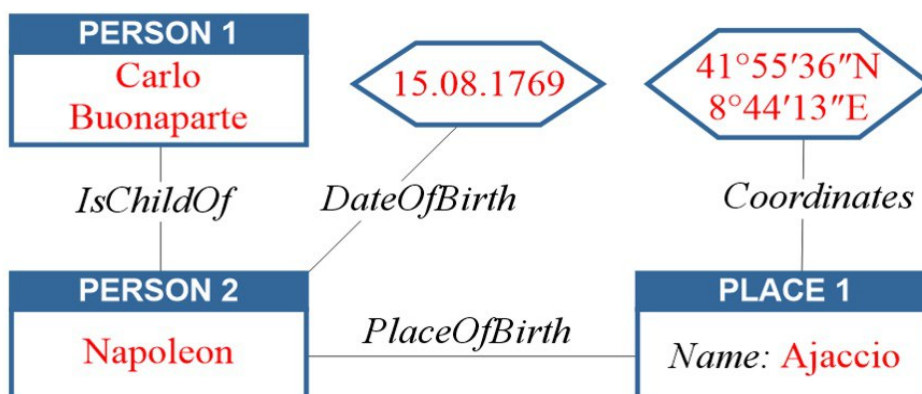
B2- refer to an object of another class, e.g. *PlaceOfBirth* may refer to an object of the class *Places*.



This is a class and property diagram.

Thus, the properties may induce relations linking objects of different classes or even of the same class, e.g. the property *PlaceOfBirth* links objects of the class *Persons* and objects of the class *Places*, another property *IsChildOf* would link two objects of the class *Persons*.

Once the model, i.e. the classes and properties, has been defined, it can be used to represent specific situations. Hence, we define some specific objects belonging to our classes and we set their properties, either by providing a value or by making a link with another object. Such depicted situations can be shown graphically, using a so-called object diagram.



This is a simple object diagram showing the 'Person' Napoleon with 'PlaceOfBirth' and other properties.

⁶ That is to say in the study of software development methods aiming to make them reliable, economic and optimal.

In the following section, we will explain how an object-oriented formalism can be applied to historical knowledge and historical cartography.

B.3 - A System Based on a Formal Model for Historical Knowledge

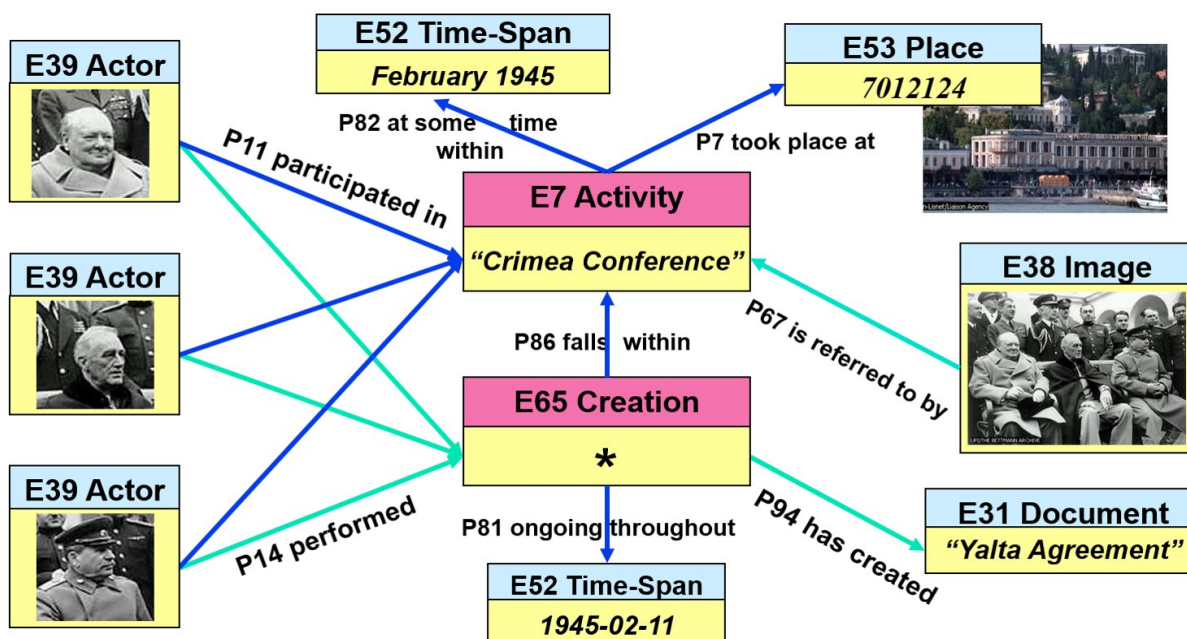
As mentioned, object-oriented formal models have been developed within software engineering and it has to be stated that such models cannot work without a computer system. That is why we are in the process of creating such a system.

Concretely, this system implies a dedicated computer structure including

- I. a database to store:
 - the historical information,
 - the meta-information⁷;
- II. a dedicated software:- to enable the input of information,
 - to allow exploration through maps or other forms of visualization.

The dedicated software plays an important role as it allows historians to interact with the model. Hence, it must offer an ergonomic and intuitive way for the users to enter formalized historical information. Also, it must be able to generate maps and other representations of the entered data. While it is no easy to build such software, it is essentially a technical task and has no strong implications on the data that will be created.

On the contrary, the design of the database, specifically of the formal model, is both the most difficult and also the most important part; as it will determine which information will be considered and how this information will be conceived. To design such a formal model, it is necessary to accurately define the classes of objects we want to represent, their properties and the links that they may have between them.



Here is an example from the International Committee for Documentation Conceptual Reference Model or CIDOC-CRM (an acronym for Comité International pour la **DOC**umentation - **CON**ceptual **REF**erence **MO**del). The CIDOC is an international institution in charge with a formal model for the International Council of Museums. The above example is relatively clear. Keep in mind that each letter E stands for 'entity' and letter P represents the properties. For more information please see: <http://www.cidoc-crm.org/>

⁷ Information about the source of the data.

Hence, the goal is to design a history database which would be based on a well fitted model to represent historical knowledge in the domain of geopolitical situations as well as possible. Thereafter, the information stored in it could be used to automatically generate maps or graphs according to rules determining how each type of information should be represented. Thus, it becomes conceivable to define a model allowing to store and use information with great accuracy, even though many details will never, or only with difficulty, be shown on a map.

Then, the maps will be used to represent some of the content of the database but, contrary to ordinary maps, they will be much more flexible and interactive and will allow easier searches. Firstly, it will be possible to create multiple maps from the same information, by varying, for example, the level of detail (granularity) of information, or even the type of information in the display. Secondly, there will also be the possibility to build interactive maps giving access to detailed contextual information that could not be easily shown otherwise. Finally, another great benefit of this approach will be the opportunity to integrate meta-information - i.e. clues about the information - in the data, such as bibliographical references, levels of confidence, or even more detailed explanations about the origin of the information. Usually, meta-information cannot be shown consistently, or only partially, on a map while it is extremely important for scientific use. In a general and well-fitted database, one can store such provenance information with as many details as needed and this can be done in a contextual manner. To reach the great benefits of such an approach, the database needs a complete and accurate structure. Hence, in this article, we propose a first attempt to develop such a formal model to represent the politics and their relations over a long period spanning more than 2000 years.

B.4 - Advantages and Defects Resulting from a System Based on a Formal Model

There are four main benefits in the use of a system based on the formalization of historical knowledge.

A - Flexibility of display

Multiple maps and different pictures or views of the recorded geopolitical knowledge can be automatically generated simply by varying the detail level on display or by highlighting some elements. There is the choice, for instance, to hide some minor geopolitical relations or to alternately show the legal - *de jure* - and effective - *de facto* - situations. Moreover, it is easy to make the maps interactive by using a dedicated software, allowing the user to adapt the view as needed or to access more information contextually, through a message box or a new tab. Finally, other kinds of views other than mere maps can be produced, for instance, graphs or charts showing the relations between the different entities.

B - Automatic synthesis

The information in the formal model can be used to perform automated reasoning. In such a case, an artificial intelligence module will apply logical rules on the data in order to deduce, for instance, new knowledge, like a relationship, or to detect logical inconsistencies, like any chronological divergence between related events⁸.

C - Achievement: omission of elements in a database is often unavoidable since objects are entered by users and, as previously mentioned, the human memory is not perfect. Thus, after automated treatment of the data in a formal model, it is possible to display the results in such a way that some hints will appear where some data seem to be missing. Actually, it is, for instance, possible to display

⁸ See an applied example of automatic historical deduction with CIDOC CRM: Holmen, J. and C.-E. Ore 2010. Deducing Event Chronology in a Cultural Heritage Documentation System, in: Frischer, B., J. Webb Crawford and D. Koller (eds.), *Making History Interactive. Computer Applications and Quantitative Methods in Archaeology (CAA)*. Proceedings of the 37th International Conference, Williamsburg, Virginia, United States of America, March 22-26 (BAR International Series S2079). Archaeopress, Oxford, pp. 122-129.

in red, at each time period, all the fields where a specific object should be present. Thereafter, the user can check if a blank field corresponds to a true absence or to an oversight.

D - Ease: the amount of time required to enter information in a database is always exhausting. However, since a formal model can be managed by automatic processing, it becomes possible to adapt the fields to the user's needs so that, for instance, only important and relevant fields will be on display. Moreover, it's easy to reduce the database to the objects or to the properties only, so that a user can, for instance, translate and double-check the data.

Note that automated processing and reasoning are often seen as alienating and enslaving for the humanities and inevitably recall novels such as *Brave New World* (Aldous Huxley, 1932) or *Nineteen Eighty-Four* (George Orwell, 1949). That is why we need to also list the main arguments with their advantages and defects resulting from a system based on a formal model compared to human reasoning.

1 - Objectivity may be the main issue. Especially in history and humanities, most research is influenced by the personal opinions of researchers and the need to attract the general public. It is commonplace to say, for instance, that the 'private life' of Charlemagne attracts more readers than the primary sources attesting the existence or the real power of Charlemagne. Thus, a formal model offers guarantees of impartiality unless the categories fixed by the definitions are biased. Therefore, it is necessary to give, as far as possible, measurable definitions and to explain reasons for each definition.

2 - Consistency means, in short, that two similar situations have to be treated similarly and that two different situations have to be marked in such a way that the difference is visible. The searcher must rely primarily on memory to treat similar and different situations but when a study covers several hundred cases, memory becomes useless. Moreover, everyone knows certain subjects better because they are close in time or in space and hence tend to treat them differently even if there is no reason to do so. Thus, a big advantage of a formal model is that the user only needs to fix initial formal rules and subsequently rely on automated treatment. Moreover, if it appears that the artificial intelligence module has mistreated a case, showing, for instance, two countries as non-sovereign while one should have been displayed as sovereign, the user is in a position to examine the relevance of the data entered in the system.

In conclusion, we must recognize that a formal model offers many advantages provided, of course, that the whole system remains transparent and that the human user maintains control over it.

C - FORMAL MODEL CONCEPTION

There are some basic requirements for a geo-historical formal model (C1). It would be possible to use modern definitions for the various types of polities, yet this would not be valid for ancient times (C2). Moreover, if we try to combine the modern polities definitions with the general model, this raises several issues (C3). A kind of solution is to take as a basic unit a generic polity and to specify it by adding properties (C4).

C.1 - Design of the General Model

Here are the main requirements that we identified for a geo-historical formal model:

- geopolitical situations, at various periods, of the territories and populated places,
- naming, alternative naming, and appellations of the polities and populated places,
- time duration or temporality of each geopolitical situation,
- the uncertainty of situations and naming,
- references and sources,
- additional explanations,
- physical geographical features (coastlines, rivers, mountains etc.)

They are many challenges involved in fulfilling these requirements, but it is certain that one of the most important pertains the model used to describe the relations or links between the polities. Since the model will have a strong impact on the type of information a historian will be able to enter in the system, it has to permit the description of geopolitical situations in a way that corresponds to the approaches commonly used by historians and researchers. So, for the time being, we focus mainly on this first aspect, that is to say, the representation of geopolitical situations.

That is why, we intend to create a model comprising 3 kinds of objects: polities, territories, and relations. Once the model, i.e. the three classes of objects and their properties, has been defined, it can be used to represent a wide range of specific situations. We then have to define the specific objects belonging to these classes and set their available properties, either by providing a value or by associating the property with another object. Such depicted situations can be shown graphically, using a so-called object diagram (see B2 above). Thereafter, such a model will accept the presence of polities whose number will not be limited. Each territory will be linked to a polity but not all polities will be directly linked to a territory since the model's supporting structure consists of relations which may link a polity to another one or more. More details and graphical examples will be provided in the next chapter.

C.2 - Modern Legal Polities Catalogue and Personal Ties

As mentioned above (sect. A.3), to describe political history it is conceivable to define first which kind of polities should be represented. To do this, it is relatively easy to rely on major history textbooks and on public international law. Here below is a list the polities generally featured in the main reliable works. Please note that some sloppy books are mixing the representation of political entities with other features like religious or ethnic groups. It is, of course, necessary to disregard such publications.

1 - Confederation: a non-sovereign union of sovereign states by a treaty devolving some powers to joint agencies without the formation of another state. *E.g. Switzerland 1648-1798 and 1813-1848; United Netherlands 1588-1795; Livonian Confederation 1418-1561.*

2 - Staatenverbund: a non-sovereign supranational institution which may exercise some sovereign actions but cannot choose its field of competence. *E.g. the European Union since 2000; Holy Roman Empire 1274-1803.*

3 - Personal Union: non-sovereign union created by law, treaty or inheritance between two or more entities having the same ruler but remaining separate sovereign states without any joint government structure. *E.g. England-Scotland-Netherlands 1700.*

4 - Real Union: the union, by law, treaty or inheritance, between two polities having the same head of state and joint agencies, resulting in a loss of sovereignty of the weaker polity (quasi-virtual annexation). *E.g. Habsburg.*

5 - Federation or federal state: sovereign state consisting of autonomous polities whose status cannot be challenged by the central authority. *E.g. Germany 1949 to present; Switzerland 1848 to present; Belgium 1970 to present.*

6 - Unitary state: a sovereign state governed by a single authority and whose administrative divisions have only the powers delegated by this authority. *E.g. France, Greece or Italy in the 20th C.*

7 - League: a non-sovereign union of non-sovereign polities delegating some powers to joint agencies. *E.g. Lombard League 1167-1250; Décapole of Alsace 1354-1679.*

8 - Federated state: constitutional polity forming part of a federation and having a certain level of autonomy. *E.g. Swiss Cantons or German Länder in AD 2000.*

9 - Dependency: polity that does not have full political sovereignty but remains politically outside of the controlling state area. *E.g. Maine and Anjou in 1400; Valachia in 1500; Akrotiri and Dhekelia 2000.*

10 - Administrative division 1st level: administrative polity aggregating several lower level administrative divisions under the authority of an agent of the sovereign state. *E.g. Roman dioceses; Ottoman eyalets; French regions 2000.*

11 - Administrative division 2nd level: administrative polity included in a higher-level polity and under the authority of an agent of the sovereign state. *E.g. Roman provinces; English counties; Ottoman sanjaks; French départements.*

12 - Proto-polity: polity comprising a population living on a territory not clearly defined and without precise state organization. *E.g. Alamanni AD 400; Burgundians AD 400; Saxons AD 400.*

Although this list may be useful and consistent, we must admit that it is valid only for the early modern period, mainly from the 18th C. on. Indeed, such concepts as sovereignty, public power or competence have only been implemented gradually since the treaties of Westphalia in 1648 and the very international legal definition of a state dates back to 1933 (see above A.3). It follows that if we try to describe 14th or 15th C. Europe in accordance with these criteria, the result will be badly informative and anachronistic. Moreover, in the Middle Ages, other criteria derived from public law, from private law, from religion or local customary law, were taken into account to determine sovereignty, dependency or inclusion in a higher-level polity. Furthermore, the issue becomes even more complex because, after the collapse of the Roman Empire in the West in the 5th - 6th C., there was no longer a clean differentiation between public and private law and personal links were tending to replace real links between the polities, giving birth to a nested network of interpersonal relations commonly called feudalism.

During this period, however, law enforcement was weak, and feudalism was not defined as a general form of governance. We of course know of several charts of the 13th - 14th C. that attempted to set an ideal form of feudalism but even then, it appears that:

- the nature of the personal relations was different in each area of Europe,
- there was no differentiation between private personal relations and public personal relations,
- real relations, i.e. territorial links, still existed and superimposed on personal links.

In short, we must admit that the ideal feudalism system was a sort of hierarchical structure whose legal definition came when the system was already in sharp decline, that is after the 14th C. Nevertheless, certain notions borrowed from the vocabulary of feudalism are currently used nowadays, e.g. overlordship, vassalage etc. although they don't point towards true legally feudal structures. In fact, it remains possible to depict the medieval period by indicating the feudal ties but, still, the result will not be consistent since these personal ties were different in the British Isles from what they were in the Loire-Rhine area or south of the Loire, east of the Rhine, south of the Pyrenees or between the city of Rome and the Alps or south of Rome and so on, not to mention that the situation was completely different in east or south-east Europe.

C.3 - Attempts to Combine the Polities Catalogue with the General Model

To summarize the problems raised above, we can draw the following conclusions.

- The use of a modern polities catalogue would be easy, clear and rational but anachronistic for the periods before the 18th -17th C. that is to say for the Middle Ages, Antiquity or Prehistory because the definitions used do not correspond to the prevailing political situation during the ancient periods.
- The use of a system describing the personal ties - i.e. mainly feudalism - would be accurate and historically correct for the medieval period but neither valid nor consistent for all the geographic areas of a continent because of the wide variability of political and legal situations for each geographical area. In addition, such a system would generate highly complicated information and would not allow the creation of a database covering all the centuries.
- The use of a *sui generis* system - i.e. of its own kind or specific - for each period and area would be merely illustrative and would not allow to show, compare and study different periods in the same work, unless by adding long and often confusing explanations for each particular case. Moreover, it would largely be affected by subjective opinions since no comparison would be possible with other periods or geographical areas.

Given the necessity to fit facts and reality, the solution that comes to mind would be the creation of a big database with a great number of criteria and categories allowing to store specific data for each given period and giving the opportunity to operate a differentiation for every reduced cultural area. However, such a huge system would soon become a labyrinthine and accessible only to specialists. It even seems that only some parts of the database would be understood by specialists in a specific field. We chose to develop a formal system which remains simple and intuitive and almost open to everyone. This will be a guarantee of flexibility and objectivity.

C.4 - Necessity to Create a Generic Polities Catalogue

All these elements lead us to search for new possibilities in the creation of a geohistorical database. Firstly, it is necessary to find a definition of the polity which would be valid in all cases and all periods throughout history. The global consistency of the database should disregard the modern catalogue and find an absolute definition based on social and anthropological facts.

The minimal definitions below are generally accepted.

- A *state* is an organized community living in a particular territory under a system of government whose aim is in the public interest. There is a divergence as to whether the concept of the state implies sovereignty.
- *Sovereignty* is full right and power of a governing body to govern itself without any interference from outside sources or bodies.

However, this yet again raises a large number of questions: what is exactly is a community, what is public interest, what do we mean by right and power or which are the outside sources? Each question may deserve careful thought and may lead to purely subjective opinions, that is to say, irrelevant information.

To overcome these problems, it is logical to develop a multi-level model of representation, the first level being a general and relevant frame, the second level a more accurate and factual analysis and it should remain possible to also add other levels for more in-depth consideration. This could allow historians to represent well-known uncontroversial situations easily and quickly while more accurate information could be added later or for different cases.

Thus, there remains the possibility to create, by drawing inspiration from the modern legal polities catalogue (see C.2 above), a limited number of generic polities; not very specific polities which are determined by the subsequent addition of properties - actually links and coefficients - for each specific case.

D - THE PROPOSED FORMAL MODEL

Firstly, the description of the model comprising 3 classes of objects and 11 properties with graphical examples of the objects and properties is presented (D1). Then the definitions of the properties determining the type of each polity are given (D2). The properties *competences* and *effectiveness* are attributed to the relations between the polities and can get different possible values (D3). The property 'primary controller' attributed to the object *territory* determines which polity has direct control over this territory (D4). The exclusive territorial control is determined through an automatic reasoning procedure, which is explained with some graphical examples (D5).

D.1 - General Formal Description of the Hierarchical-Cooperative-Convergent, or HCC, Model

On the basis of the preceding considerations, we have designed a formal model which is probably not the only possible one, but which offers fair answers to our requirements. Below is a summary presentation.

First, we have set three classes of objects:

- 1 - polities, that is to say, political entities,
- 2 - territories, that is polygons representing mere topographical units,
- 3 - relations that are political ties linking the polities.

Said objects can be represented graphically.

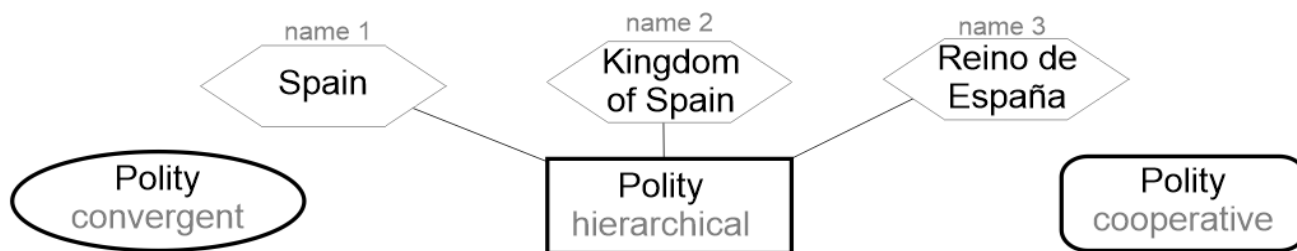


And there are 11 main essential properties that can be attributed to the objects. It is possible to summarize the objects and properties with 2 tables, here below. Note that some properties - in *italic* no. 3, 10 and 11 - are automatically generated, or derived, by an algorithm according to the other attributed properties. More explanation will be given later.

	Object	Description	Available properties
1	polity	any political entity	name, type, <i>exclusivity</i>
2	territory	polygon representing a topographical unit	geometry, primary ctrl, exclusive ctrl, non-exclusive ctrl
3	relation	any kind of link between two polities	source, target, competences, effectiveness

	Property	Applied to	Value	Cardinality ⁹	Attributed by
1	name	polity	text	multiple	user
2	type	polity	[cooperative - convergent - hierarchical]	1	user
3	exclusivity	polity	[yes - no]	1	automatic
4	source	relation	polity	1	user
5	target	relation	polity	1	user
6	competences	relation	[regular - intermediate - irregular]	1	user
7	effectiveness	relation	[effective - partial - nominal]	1	user
8	geometry	territory	topographical polygon	1	user
9	primary controller	territory	polity	1	user
10	exclusive controller	territory	polity	multiple	automatic
11	non-exclusive controller	territory	polity	multiple	automatic

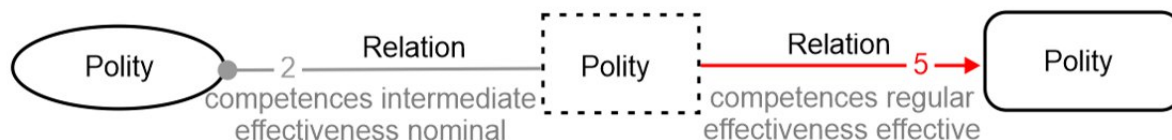
In order to facilitate the drawing of diagrams representing specific situations, we have defined some specific graphic conventions. The *type* of a polity is indicated by its shape and the arrows representing the relations vary according to the value of *competences* and *effectiveness*. The following examples illustrate the model by means of these conventions.



On the upper line the property 'name' for a polity appears. This property allows multiple text values, that is why 3 names are mentioned. On the lower line, 3 polities of 3 different types are shown, that is why the blocks have different shapes.



Here is an example of the properties 'source' and 'target' attributed to relations. The central polity is the source of 2 relations and the polities on the hedges are, each, target of one relation. Note that these properties affect the objects 'relation', that is why they are unique: each relation has only one source and target while the polities may have multiple incoming or outgoing relations.



Example of the 'competences' and 'effectiveness' properties attributed to the objects 'relation'. On the left side relation, the 'competences' property is intermediate and the 'effectiveness' property is nominal while, the right-side relation has a property 'competences' regular and 'effectiveness' effective. The 'competences' combined with 'effectiveness' value is also reflected through the arrow colour and shape of its end. Numbers '2' and '5' marked on the arrows result from the quality of their properties; see D.3 below for details.

⁹ Maximum possible number of elements.

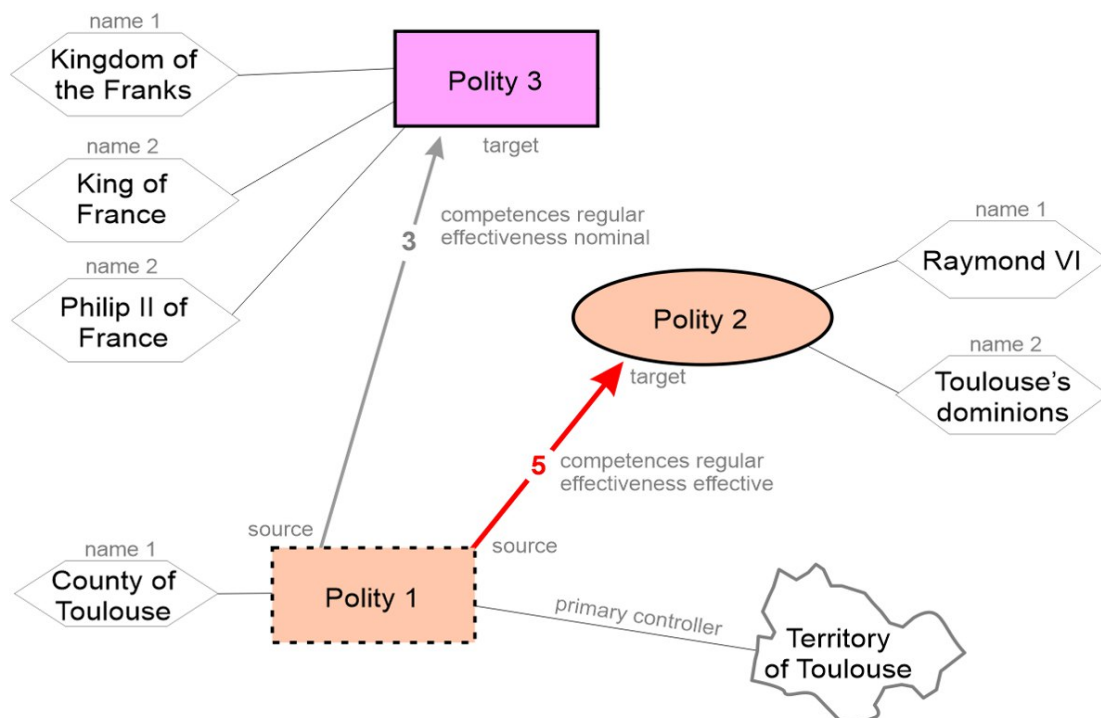


The 'primary controller' property simply marks to which first polity a territory is attached and the property 'geometry' determines the topographic shape of that territory.

As a result of the properties 2, 6 and 7 defined above, it must be clearly stated that:

- 1 - each polity is marked in a summary form and can belong to one of 3 different types.
- 2 - polities may be linked by relations of different *competences* and different effectiveness values.

In addition, and this will be explained further, the lowest polity in a chain possesses a primary control over a territory. In fact, we postulate that such a primary control corresponds to exclusivity, a notion that is more or less equivalent to sovereignty, and derives from the political relations between the polities. Thus, we accept that exclusivity transfer is not set by the user but automatically calculated by an algorithm depending of the relation properties. However, users will be allowed to set the transfer threshold according to their own valuation.



Here is a partial example of the objects with their main properties applied to the real case of Toulouse-France in the year 1200. Polities surrounded by a plain line are exclusive or sovereign while a dashed line marks a non-exclusive polity. We are only representing the situation of the territory of Toulouse whose primary controller is Polity 1. The colours denote that the exclusive control over this territory is transferred to Polity 2, that is to say, to count Raymond VI. This is due to a transfer threshold set to coefficient 4. Further explanation of this feature will be given below.

D.2 - Description of Polity Types and Their Properties

A polity has one of the 3 types listed below. This means for instance that in the model processing, a user can create a polity and choose its type by assigning one of these 3 values, to the property *type* attributed to the object *polity*.

I - Hierarchical: used for the polities whose members are organized according to subordination principles. In the modern polities catalogue, this would be valid for the unitary states, for all levels of administrative divisions and, presumably but not necessarily, for the dependencies and proto-polities. Of course, the hierarchical type is not limited to the entities listed in the above catalogue. Throughout history, a lot of different polities holding their members in strict subordination have existed.

II - Cooperative: used for the polities composed of members conforming to cooperation principles commonly expressed by sayings such as *Unity makes strength* or *One for all, all for one*. In the modern polities catalogue, this would be valid for the confederation, Staatenverbund, federal state, league and for some federated states and dependencies. The case of the proto-polity remains open since a group of proto-polities was often, but not always, a league or a confederation. Here again, the polities catalogue does not set a limit: there were several kinds of cooperative polities throughout history.

III - Convergent: used for the polities composed of whatever other types of polities but all having the same unique ruler, king or supreme authority. In the modern polities catalogue, this would correspond to the personal union or real union but, like in the previous instances, would not be limited to those, since some cases of commonwealth¹⁰ could be included in this category.

D.3 - Description of the Relations and their Properties

In theory, every polity could exist alone for itself without any contact with the neighbouring polities. In fact, this may be true only for some proto-polities and, when we read historical records, we see that nearly every polity had and still has a big amount of links with the other polities of its environment. Some ties are purely commercial or cultural while other links are so intense that the polity becomes really part of another, greater or most powerful polity. Moreover, some links are merely nominal while others are effective.

Thus, in this model, as previously mentioned, there are 2 properties - *competences* and *effectiveness* - that can be attributed to the object *relation*. Each of these 2 properties may have 1 out of 3 values of intensity.

A - Competences that is to say as regards to the extent of the competences they affect, and they may be regular, intermediate or irregular,

A1 - *regular*: for relations covering all the competences belonging to the source,

A2 - *intermediate*: for relations covering a part of the competences belonging to the source,

A3 - *irregular*: for relations covering only a small part of the competences related to the source.

B - Effectiveness that is to say as regards to their factual existence,

B1 - *effective*: for factual and actual relations,

B2 - *partial*: for unstable and inconstant relations,

B3 - *nominal*: for theoretical or purely nominal relations.

The polities may be linked between them, without limitation, by one or more relations, starting from the lower polity or first source, towards a higher polity or target. Note that a polity having no incoming or

¹⁰ In French "communauté politique".

outgoing relation is regarded as hierarchical and we also consider that a polity without any incoming relation is always the primary controller of a territory.

The user has to determine the intensity of the *competences* and *effectiveness* properties for each relation. Next, the combination of these properties will calibrate the strength of the relation, sized by a coefficient, as marked in the table here below.

COMPETENCES	EFFECTIVENESS		
	effective	partial	nominal
regular	5	4	3
intermediate	4	3	2
irregular	3	2	1

These coefficients are useful to fix the compared value of political links between several polities having multiple relations and also to create a coherent whole of the various polities across a continent for example.

D.4 - Territories and Territorial Control

We are assuming that the smallest describable polity, or first source polity in a chain of relations, has, by a kind of legal fiction, a set of powers, that can be considered as an “exclusive control”, consisting in full domination over a territory. Actually, this characteristic corresponds to a material reality because, the distances being short, in every small territorial unit and at all times, the authority has the effective possibility to exercise its power over the said unit.

In fact, we do not have the ambition to set an absolute legal definition of sovereignty because this would go beyond our model. Thus, we are instead using this notion of exclusivity that offers some fair assets in the formal model. It is relatively easy to determine which polity has an exclusive and effective control over a territory and, thereafter, to assign some rules allowing or forbidding the transfer or inheritance of such an exclusivity towards higher polities.

D.5 - Automatic Derivation of the Territorial Control Transfer and Exclusivity

Furthermore, according to some systematic rules, this set of powers may be transferred to a higher polity if there is a sufficiently strong relation with it. The upper polity is then substituted for the lower one and thus, acquires this capacity of exclusivity. Accordingly, it is understood that the lower polity loses its exclusivity for this same reason and that the transfer operation can be repeated as often as necessary to another higher polity. Note that if a polity has several outgoing relations above the transfer threshold - i. e. justifying a transfer - thence exclusivity will be correspondingly shared between the multiple target polities.

Thus, an automated reasoning module will calculate if exclusivity has to be transferred and ceded - with further possible transfer - or not transferred nor ceded to a higher polity. However, a human user will always be allowed to set the exclusivity transfer threshold since the competences and effectiveness properties are sized by a coefficient, from 1 (lowest) to 5 (highest), and one can decide at which level the transfer must be calculated by the machine.

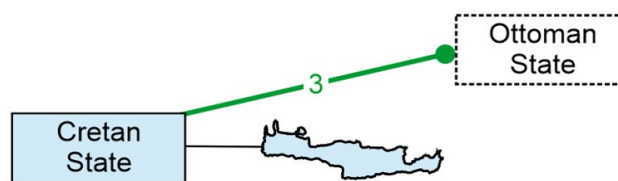
This feature is extremely important since it prevents partiality and personal opinions. Indeed, when you set the transfer threshold, for instance at level 3 or 4, that means that all the relations in the whole database - that is in all generated maps - having a strength of 3 will automatically transfer exclusivity to the higher polities and, as a result fewer sovereign-like polities will appear on the maps. Thus, a program will show, based on user-defined relations, the existence of so-called *exclusive polities* which

are similar to the *sovereign* or *independent* states that we know of. In short, the user defines the polities and the relations they maintain in the system. From this information, the program automatically deduces the exclusivity of each polity by means of the exclusive control transfer method described above.

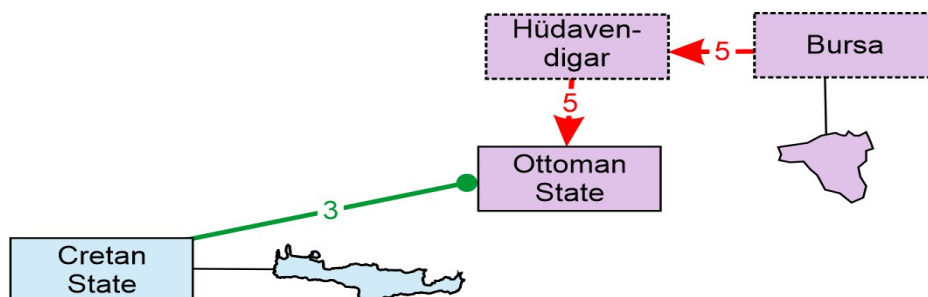
See now, as an example, the case of Crete in the year 1900.



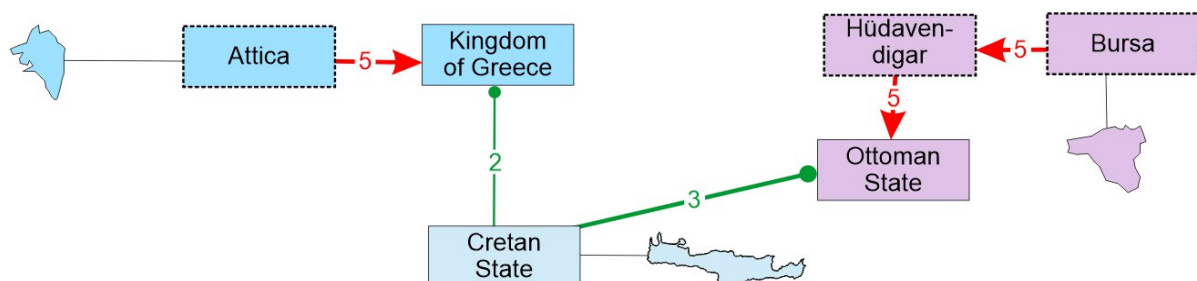
This is the 'Cretan State' polity in the year 1900 with the Cretan territory attached to it. Note that the rectilinear polygon denotes a hierarchical polity. At this stage, we can see that the Cretan State is the first controller of a territory and, therefore, has exclusivity over this territory. However, the political reality was somehow different.



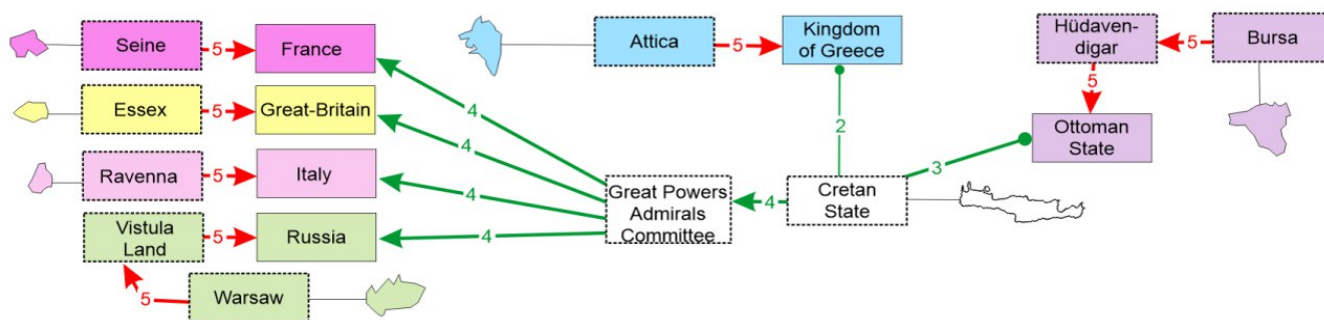
In fact, the Cretan State was autonomous but still part of the Ottoman Empire, so it is necessary to mark a relation with properties 'competences intermediate' & 'effectiveness partial' (i.e. strength 3) towards the Ottoman State. Note that the transfer threshold is fixed to 4 or higher and, here, the strict automated reasoning has not yet detected any territory attached to the 'Ottoman State' polity. Accordingly, the 'Ottoman State' polity is not shown as exclusive on this graph. Thus, we need to link, with the Ottoman State, at least one polity having exclusivity over a territory.



That's now done. The Sanjak of Bursa having exclusivity over its territory is linked with a relation of strength 5 towards the vilayet of Hüdevendigar which has itself a relation of strength 5 towards the target 'Ottoman State' polity. As a result, Bursa's exclusivity is transferred to the Ottoman State (i.e. Bursa -> Hüdevendigar -> Ottoman). Let's recall in this example, that exclusivity is transferred only by relations of strength 4 or higher. This also explains why the polity 'Cretan State' retains its own exclusivity. However, in 1900, other polities had claims or power on the Cretan State, mainly Greece.



In fact, Prince George of Greece, son of King George I of Greece, had been appointed as High Commissioner of the autonomous Cretan State but his power was rather weak so that only a relation with 'competences irregular' and 'effectiveness partial' properties (i.e. strength 2) is linking Crete to Greece. Indeed the real power over Crete was in the hands of a joint Great Powers Admirals Committee.



So, we now see on this image that the Cretan State is the source of a relation of strength 4 targeting the 'Great Powers Admirals Committee' and that this Committee is linked to the great powers, namely France, Great Britain, Italy, and Russia, by similar relations of strength 4. It appears in short that the Cretan State has lost its exclusivity which has been transferred to the Great Powers. Keep in mind that the user can fix the exclusivity transfer at strength 3 or 5 for instance and that the colours above are a mere visual help. In the final map or graph display, another kind of visualization may be set. Note too that administrative polities, like 'Seine', 'Essex', 'Ravenna' and so on, linked to the exclusive polities are shown for illustration purposes and also to give them the quality of exclusivity with the automatic reasoning system.

For the time being everyone can access our 1st level working model online and even make a draft by sending us a query.

Euratlas Historical Polities: <https://hpolities.euratlas.net/web>

E - ADVANCED WORK

Three issues remain to be addressed: metadata, missing data and decision tables (E1). Reliable sources being often scarce, this problem will be treated by a probability tree in a future stage (E2). In addition, a table decision tree has to be implemented to facilitate the choices of objects and properties and to reduce ambiguities and personal opinions susceptible to create inconsistencies or discrepancies (E3).

E.1 - Advanced Issues

With the general frame of our model having been settled, there are still some issues that need to be addressed in more detail. In fact, the formal model works fluently with the core features described above but there are still certain problems that needs to be explained in order to meet scientific accuracy.

Firstly, some fields for metadata properties will be added, particularly about the provenance of the information, such as sources, bibliography, textual justifications of the choice made, or even the level of confidence in the information. There is no great difficulty here, but this is a very important aspect that needs to be handled in order to produce transparent scientific data.

Next, it will be appropriate to address the missing data issue, as well as to handle the possible indeterminacy.

Finally, it will be necessary to specify a complete process allowing each user to choose, step-by-step and easily, but with accurate criteria and as little ambiguity as possible, the correct type of each polity and each relation, according to the specific geopolitical situation.

Of course, the topography of the territories is not addressed by our model since there are already a lot of open mapping databases and mapmaking software that allow the drawing of maps.

E.2 - Missing Data and Probability

Historical primary sources are infrequent for the vast majority of cases for the periods, before about the 16th - 17th C. Indeed, the more one moves back in time, the more writings are scarce, with some exceptions by regions or by periods. For instance, we have more eastern than western primary sources for the second part of the 1st millennium C.E. and, globally, we have more primary sources for the first part of the 1st millennium than for the 2nd part of it.

As a result, since the human brain is able to make deep analysis with missing data or rare data available (see Kahneman, Thinking, Fast and Slow, 2011), it appears that most of the historical records are made up of second hand or third-hand sources and much more of commentaries and interpretations over these kinds of documents. As a matter of fact, about 80% - 90% of historical records are glosses and commentaries, 10% -15% are second and third-hand sources and the primary sources represent only about 2-5% of this grand total.

Obviously, we have to conclude that uncertainty and undecidability constitute a wide part of our historical knowledge. Most historians, however, learned to manage such indeterminacies. When the facts are unclear for instance, it is usual in history works, to widen the time range or the wiggle room by using generic characterizations to describe a political situation. Similarly, where there is controversy, historians feel hence justified in defining the instance with generic terms able to describe a multitude of different cases, like 'vassalage', 'overlordship', 'dominion' or 'rights'. Such answers do not meet the requirements for a formal model, especially as historical research is expanding every year: we discover new sources and archaeology - almost a hard science -, which constantly provides new information.

Thus, a history database has to:

- accurately incorporate known scientific facts,
- make an understandable difference between certainty and uncertainty,
- indicate undecidable points,
- leave a door open for the integration of new scientific reliable information.

Our aim is to implement, at an advanced stage, probability system that would allow us to cope with indeterminacies and missing data.

E.3 - Two Levels of Analysis

Ideally, the HCC formal model should be understandable to everyone, but its basic features allow only historians or specialists to enter data in it. Indeed, someone who doesn't have a fairly in-depth knowledge of a specific historical period can hardly determine if a polity was hierarchical or cooperative and if the relations linking it with the neighbouring polities were regular-effective, irregular-effective, intermediate-partial or intermediate-nominal. Moreover, it often happens that even specialists have different appreciations of the same situation. That is why we intend to create several simple decision tables on measurable facts, as yes-no questionnaires, through which it will be possible to decide:

- the type of each polity,
- the properties of each relation.

This will reduce discrepancies due to personal opinions and, if discrepancies remain, conflicting opinions will have to be justified.

Thereby the answers given to each question will automatically lead the user towards the correct type of object and the correct property and build a working model showing the polities and their relations. Subsequently, it will be possible to use the working model in order to display a map or a table chart.

Marc-Antoine Nüssli & Christos Nüssli
February 2017

REFERENCES

Bibliography

- DUBY, Georges & al. - *Atlas historique Larousse*, Larousse 1978
- ENGEL, Josef & al. - *Grosser Historischer Weltatlas II: Mittelalter*, Bayerischer Schulbuch-Verlag 1979
- HILAIRE, Jean - *La construction de l'État de droit dans les archives judiciaires de la cour de France au XIIIe siècle*, Dalloz 2011
- KAHNEMAN, Daniel - *Thinking Fast And Slow*, Penguin Books 2012
- REYNOLDS, Susan - *Fiefs and Vassals. The Medieval Evidence Reinterpreted*, Oxford Univers. Pr. 1994
- WATZLAWICK, Paul & al. - *The Invented Reality: How Do We Know What We Believe We Know?*, W. Norton & Company 1980
- WHALEY, Joachim - *Germany and the Holy Roman Empire, vol. 1: 1493 - 1648*, Oxford Univers. Pr. 2012
- ZIEGLER, Andreas R. - *Introduction au droit international public*, Schulthess 2015

Articles

- Francesco Beretta, Claire-Charlotte Butez. The SyMoGIH project and Geo-Larhra: A method and a collaborative platform for a digital historical atlas. *Digital Humanities 2014*, Jul 2014, Lausanne, Switzerland. EPFL, Lausanne / UNIL, Lausanne, pp.117-119, 2014, Digital Humanities 2014.
- Southall, H. 2012, 'Rebuilding the Great Britain Historical GIS, part 2: a geo-spatial ontology of administrative units' *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, vol 45, no. 3, pp. 119-134. DOI: 10.1080/01615440.2012.664101