Towards Interoperability of Geopolitical Information within FAO

Caterina Caracciolo, Marta Iglesias Sucasas, and Johannes Keizer*
{name.surname}@fao.org

Food and Agriculture Organization (FAO) of the UN
v.le Terme di Caracalla, 1
00100 Rome
Italy

Abstract. This paper reports ongoing work on using an ontology as a mechanism to bridge various types of country-based information systems at the Food and Agriculture Organization of the United Nations (FAO). The type of geopolitical information addressed by this work include country international classifications, country names in the five FAO languages (Arabic, Chinese, English, French and Spanish), and other geographical information such as water bodies. Although the data required for the geopolitical ontology is already available, it is scattered across many information systems, which are often not clearly connected to one another. The expected advantage of using an ontology to achieve interoperability is that it can accommodate semantic relationships (between countries and geographical entities) that can be exploited for inference. Moreover, in virtue of the standardized semanticsoriented languages used to encode the ontology, it will provide a highly sharable and reusable resource for the international community.
This paper describes the geopolitical information to manage, presents the requirements imposed on the ontology and gives details about the ontology prototype. Finally, it discusses design issues and draws some preliminary conclusions.

1 Background and Motivations

Dealing with data about countries and regions is part of the day-to-day work of most international organizations. FAO is not an exception in this respect, as it manages and exchanges data about the subjects of its competence (i.e., agriculture and food security), regarding its 190[1] Member Nations [1] and other territories in the world.

The Food and Agriculture Organization of the United Nations leads international efforts to defeat hunger. Serving both developed and developing countries,[2]

---

* Authors are listed in alphabetical order.
[1] As at 11 Aprile 2006.
[2] The designations employed and the presentation of material do not imply the expression of any opinion whatsoever on the part of the Food and Agriculture Orga-

FAO acts as a neutral forum where all nations meet as equals to negotiate agreements and debate policy. FAO is also a source of knowledge and information, to help developing countries and countries in transition modernize and improve agriculture, forestry and fisheries practices and ensure good nutrition for all. FAO was founded in 1945, focusing special attention on developing rural areas, home to 70 percent of the world's poor and hungry people.

Managing information is crucial to FAO and that is reflected in Article 1 of its Constitution, which reads that "The organization must collect, analyze, interpret, and disseminate information relating to nutrition, food and agriculture and development."

Over the last decade, much effort has been put in the organization and management of geopolitical information such as matching different country classifications and code systems, streamlining the multilingual updates of official names and managing geographic and economic groups. In 1995, FAO established a support structure to foster the dissemination of agricultural information through its World Agricultural Information Centre (WAICENT) [2].Part of the mandate of WAICENT is to enable countries to make their own information available using electronic means to reach wider audiences and contribute to knowledge in agriculture worldwide.

In addition, in the last years, FAO started the Agricultural Information Management Standards (AIMS) initiative [3] to increase coherence among agricultural information systems, to create a clearing house for information management standards used to make existing or new agricultural information systems interoperable, and to share and promote the uptake of common methodologies, standards and applications.

Notwithstanding the substantial work carried out during the last years, the potential of exploiting the information that FAO generates by country or region remains at a very low level, circumscribed in general to single systems and with very limited interoperability. Figure 1 depicts most common organization of information systems within FAO where each application (managing statistics, maps or documents) accesses its own corresponding database. Although some information systems use XML DTDs and XML Schemas for exchanging data, this mechanism is not general enough to provide interoperability across applications and semantics are needed, given the numerous country-and region-based systems in the Organization.

Another source of common problems stems from legacy databases and corporate systems not built around geopolitical based information but using this information for cataloguing purposes. In those cases, the major problem is the management of historical records. For example, records catalogued years ago us-

---

nization of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries. In some cases, the designations "developed countries" and "developing countries" are intended for statistical convenience and do not necessarily reflect a judgement of the stage reached by a particular country or area in the development process.

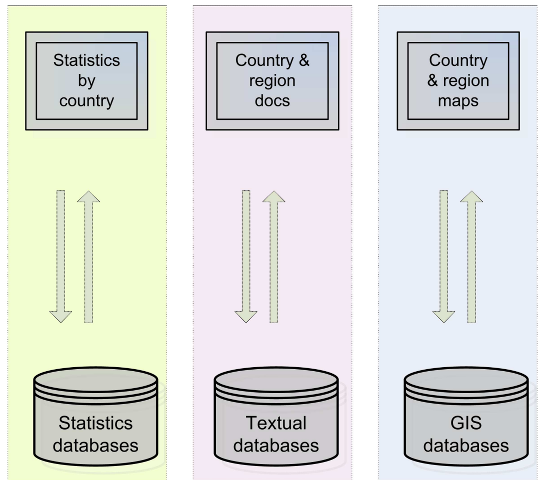

Fig. 1. A schematic representation of the current organization of information systems dealing with geopolitical information within FAO.

ing countries that no longer exist (e.g. Czechoslovakia, Yugoslavia, etc) are now difficult to retrieve or to find the relationship with established new countries, leading to the risk of increasing hidden knowledge.

Recent history shows that many changes in countries and territories status happened over the last 25 years (e.g. the union of West/East Germany, or the independence of former Soviet Republics, etc) or are still occurring (e.g. recent declared autonomy of Montenegro from Serbia and Montenegro). Managing and reflecting properly these changes (country codes and official names in various languages) is crucial for all information systems relying on this type of data.

Summarizing, the problems that moves our work is twofold: ensure the possibility of exchange of data between the various systems within FAO that use geopolitical information, and effectively represent and manage the dynamics of territories and their grouping, without having to re-engineer the complete business process.

The solution we are investigating, and about which we report in this paper, consists of using an ontology as an intermediate layer between the applications, as schematically depicted in Fig. 2. The ontology should contain the core of information necessary to for the management of both current and historical data about individual territories, groups of them, and relevant geography.

To the best of our knowledge, currently there are no similar attempts to build such a geopolitical ontology, both in terms of the information to be managed and on the relationships to be implemented.

The rest of this paper is organized as follows. In Sec. 2 we list the features the ontology should have and in Sec. 3 we present the prototype we have built in or-

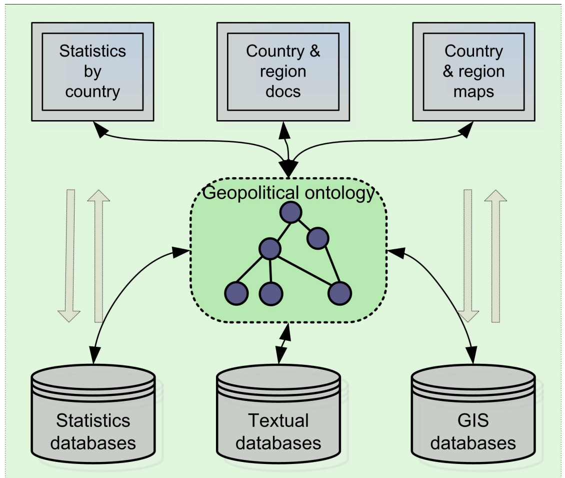

Fig. 2. A schematic representation of the future organization: the geopolitcal ontology will serve as a bridge to allow communication between the various systems.

der to meet these features. In Sec. 4 we discuss some design and implementation issues and in Sec. 5 we present current and future work.

2 Issues to Address

In this section we enumerate the types of pieces of information the ontology needs to manage in order to be of use within FAO.

We differentiate three types of objects to represent: individual territories,[3] groups of territories, and geographical information, such as water bodies. We pay special attention to the dynamics of these objects (especially territories and groups), as to allow the ontology to be used both by new information systems (representing and dissemination the status of the world as today) as well as legacy systems managing documents or statistics with historical data since FAO foundation.

2.1 National Level

In order to ensure interoperability between systems managing geopolitical information the following data should be managed.

[3] Subnational territories are out of the scope of this geopolitical ontology.

1. Types of territories. It should be possible to distinguish territories at least into the following types: self-governing territories, non-self-governing territories, disputed areas and other types.

Self-governing territories, most commonly known as states or countries, are all those that do not politically depend from others.[4] Non-self-governing territories are those that have some kind of dependency from another country, as for example Gibraltar depending on the United Kingdom or La Martinique depending on France. Since all non-self-governing territories depend on a self-governing one, we need to keep track of this dependency in the geopolitical ontology. Disputed territories are those reclaimed by more than one (self-governing) territory.

2. Names. All territories have names in several languages, and FAO manages territories' names in all the five languages of the Organization. FAO and other sister UN organizations coordinate the updates of the names of territories in agreement. For the scope of our work it is important to keep track of the names in the five languages of FAO, and moreover, manage different types of names such as the official name and the short name. As an example, Vietnam is the short name of The Socialist Republic of Vietnam (official name).

As countries always agree beforehand upon changes on country names, they are available at the same time in different language versions (for example, for all UN and FAO official languages). These changes have to be synchronously reflected in all information systems.

3. Codes. Although information and knowledge systems display territory names for usability purposes, internally this information is managed in the corresponding databases by means of territory codes. Various international organizations maintain country/region classifications. The International Standards Organization (ISO) maintains the ISO-3166 ALPHA-2 [4] and ALPHA-3 [5], and the UN Statistical Division maintains the M49 [6].

In FAO, territory- and region-based information systems usually adopt one of the existing international classifications and then adapt it to their specific needs. The geopolitical ontology should be able to accommodate several sets of international classifications (including their variants) in order to meet the requirements imposed by the FAO information systems.

4. Dynamics. When territories change and undergo splitting (e.g., Czechoslovakia into Czech Republic and Slovakia), unification (e.g., West and East Germany into Germany), or change of status (e.g., Hong Kong from United Kingdom to China), also their associated codes and sometimes names will change accordingly: new codes and names are added, old ones cease to be valid. These types of changes must be managed and recorded, so that even if a country ceases to exist (as in the case of Czechoslovakia or Yugoslavia), the legacy data about it can still be retrieved. Moreover, it should always be possible to find the successor (or predecessor) country(ies) of no longer existing ones. For instance, we should be

[4] All FAO member countries and UN member states are self-governing territories.

able to link Czechoslovakia as predecessor of Czech Republic and Slovakia and Germany as successor of West Germany and East Germany.

5. Neighbor territories. Finally, it is important to manage information about territories' neighborhood. For example, Vietnam has borders with Cambodia, Laos, and China.

Information about neighboring territories is useful in a variety of applications. Information systems using the geopolitical ontology will be able to implement mechanisms to find information on a particular subject for a country and automatically compare it with the same information obtained for its neighbours (e.g. to compare forestry issues in Ecuador versus its neighbours Colombia and Peru). Also, additional knowledge could be inferred such as, territories with no neighbours defined are necessarily islands (although the opposite is not always true).

2.2 Grouping Territories

Most information systems within FAO have to deal with groups of territories of different nature, mainly for data dissemination purposes. In this section we introduce the most important groups for FAO.

6. Types of groups. FAO manages data relative to geographical and economic groups, as well as data relative to international organizations and groups of diverse nature, that we call special groups. Continents, such as Africa and Asia, are typical geographical regions; the Caribbean Community (CARICOM), the Union Economique et Montaire Ouest Africaine (UEMOA), and the Gulf Cooperation Council (GCC) are all example of economic regions.[5] UN and FAO itself are typical examples of international organizations, while the Low Income Food Deficit Countries (LIFDC) [8], the Small Island Developing States [9], or the Landlocked Developing Countries [10] are examples of special groups.

7. Dynamics of groups. While geographical groups tend to be quite stable (they only change when territories in it change), the other groups have a more intense dynamics.

2.3 Geographical Data

Finally, concerning the geographical data, we are interested in modeling water bodies, because of the great importance of water (both salty and fresh) for human life. In fact, water bodies are important for agriculture and fisheries, and in general they are crucial to food security and safety. We distinguish two types of water bodies, as follows.

[5] For a list of "Regional Economic Organizations" FAO works with, we refer to [7].

8. Sea. This group includes salty water bodies, such as seas and oceans, which are especially important in fisheries. The basic type of information to store in this case is the list of territories that have access to them.

9. Inland water. Given the international nature of FAO, the inland water bodies in which we are primarily interested are rivers and lakes that cross or have shore with more than one country.

3 Prototype Design and Implementation

For the current implementation we decided for an OWL-DL [11], manually edited with Protg [12]. Since the current country ontology is a prototype, no implementation has yet been made towards experimentation with existing information systems.

In the current modeling the class Territory consists of all territories in the world (instances), organized into four disjoint subclasses, i.e., Disputed, Self-Governing, NonSelfGoverning and Other. Figure 3 depicts an instance of Self-Governing territories. The class Territory (and its subclasses) has the following properties, inherited by its subclasses:

- isValidFrom (in years), isValidUntil (in years), to account for the time interval in which the territory exists.
- hasOfficialName (string), and hasShortName (string), with subproperties, to account for territories' official and short names in all languages.
- hasCode (any), with subproperties to account for codes in different international classifications and coding systems.
- isSuccessorOf (and inverse), to account for all possible types of dynamics of territories (splitting/joining/status change).
- hasBorderWith (domain/range: territories), to account for neighborhood between territories.
- dependsOn (domain: non-self-governing territories, range: self-governing-terriotires), to account for political dependencies.

According to this design, "new" territories are added, but no former territory is ever deleted, allowing to trace back to any point in time the status of the territories of the world just by querying the ontology.

Groups are modeled according following a similar approach. Geographic, economic and special groups as well as organizations are modeled as subclasses of the class Groups, and any group is an instance of one of these classes, e.g., Africa is an instance of GeographicRegion, Europe15, Europe25, and the CARIFORUM are instances of EconomicRegions, Low Income Food Deficit Countries is a SpecialGroup while FAO and UN are instances of Organization.

Besides the properties concerning names and codes (in analogy with Territories), groups are given the following properties:

- hasMembers (object, range: Territories), to account for the territories that are part of the group.

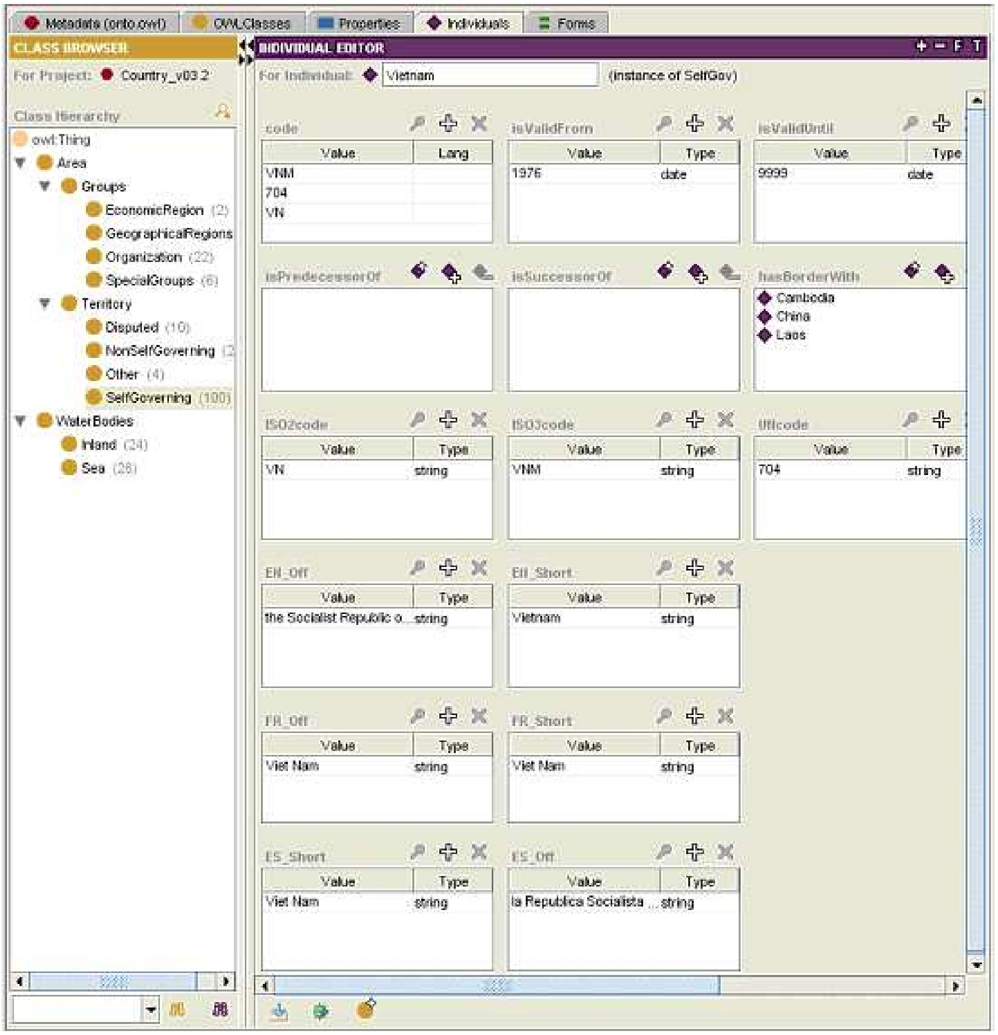

Fig. 3. A screenshot from Protégé showing the instance Viet Nam of the class Self-Governing.

– Year (date) to account for the year in which the group was created or started. Every time there is a change in the members of the group a new instance is created, with the date corresponding to the year in which that change took place and all members as of that date.

Finally, seas and inland water bodies are also modeled as subclasses of Water Bodies, and as such they have the property hasShoreWith.

4 Discussion

In this paper we presented our effort towards achieving interoperability of geopolitical information within FAO, but we believe this effort can also be fruitful for other organizations dealing with the same type of information. The ontology we presented is meant to collect together in a unifying framework all data necessary to access FAO information systems dealing with geopolitical data. If this objective is reached, it would also be an important resource for other organizations with similar needs.

The design hinted so far successfully addresses all issues presented in Sec. 2, although the modeling of the dynamics of membership to groups somewhat leads to duplication of instances, because every time a country joins or leaves a group, a new instance of the group is created, with the date of change, the same group name but different members. In practice, the date of begin of the membership is modeled as a property of the group instance only (as opposed to a property of the country joining the group), and the difference between the organization from one year to the other can only be inferred by looking at the countries present (or not present) in the two lists.

An alternative design would consist in attaching the information about membership directly at the country/territory level, but this would require adding a new property (one for each group) to each territory, with the consequence that every time a new group is created or added, there would be a need for a new property for territories. This solution does not seem to be efficient and neither manageable.

Our decision to use OWL-DL was determined by its accepted status in the Semantic Web community, and also by the reasoning possibilities provided by the Description Logics underpinning it. In fact, an expected side results of this implementation could be the possibility of reasoning over the relations and properties modeled in the ontology. For example, information about neighboring territories (as well as about territories having coastline with the same water body, territories in the same special group and in the same continent, and so on) could be easily used to compare or aggregate data about them.

5 Current and Future Work

Current work focuses on testing alternative design options that could make the ontology more efficient when in actual use, and also easily populated in an automatic way from existing repositories (such as lists of territories, names and codes). Next, we plan on expanding the coverage of the ontology (i.e., other types of geographical information), designing the model of interaction between the ontology and the underlying systems, and analyzing the reasoning possibilities offered by our current implementation.

Our future agenda includes the inspection of theoretical issues such as ontology modularization, which will be critical as soon as the data managed by ontology grows. For example, two separated yet connected modules extracted from the geopolitical ontology (geographical data on the one hand, political data on the other) could be easier to maintain than one single large ontology.

Other planned future work includes a feasibility study on managing subnational and specially georeferenced data within the ontology. If this demonstrates to be manageable, the potential of the geopolitical ontology could grow exponentially allowing to easily incorporating FAO's world-wide thematic information within Web2.0 applications like Google maps [13] or Yahoo maps [14].

References

1. UN: FAO mambership. (2006) http://www.fao.org/unfao/govbodies/membernations3_en.asp [last accessed: Aug. 6th 2006].
2. FAO: World Agriculture Information Center (WAICENT). (2006) http://www.fao.org/waicent/ [last accessed: Aug. 6th, 2006].
3. FAO: Agricultural Information Management Standards (AIMS). (2006) http://www.fao.org/aims/ [last accessed: Aug. 6th, 2006].
4. International Standard Organization: ISO 3166 ALPHA-2. (2006) http://www.iso.org/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html [last accessed: Aug. 6th, 2006].
5. International Standard Organization: ISO 3166 ALPHA-3. (2006) Not freely made available by ISO.
6. United Nations: UN Code (M49). (2006) http://unstats.un.org/unsd/methods/m49/m49alpha.htm [last accessed: Aug. 6th, 2006].
7. FAO: Regional Economic Organizations. (2006) http://www.fao.org/tc/spfs/orgs_en.asp. [last accessed: Aug. 6th, 2006].
8. FAO: Low-Income Food-Deficit Countries (LIFDC). (2006) http://www.fao.org/countryprofiles/lifdc.asp?lang=en [last accessed: Aug. 6th, 2006].
9. FAO: Small Island Developing Countries (SIDC). (2006) http://www.fao.org/SIDS/ [last accessed: Aug. 6th, 2006].
10. UN: Landlocked Developing Countries. (2006) http://www.un.org/special-rep/ohrlls/lldc/default.htm [last accessed: Aug. 6th, 2006].
11. The World Wide Web Consortium: Web Ontology Language (OWL). (2004) http://www.w3.org/2004/OWL [last accessed: Aug. 6th, 2006].
12. Stanford University School of Medicine: Protégé Ontology Editor. (2006) http://protege.stanford.edu [last accessed: Aug. 6th, 2006].
13. Google: Google Maps. (2006) http://maps.google.com [last accessed: Aug. 6th, 2006].
14. Yahoo: Yahoo Maps. (2006) http://developer.yahoo.com/maps/ [last accessed - Aug. 6th, 2006].