Spatial Data in an Ontology for Research on Forest Resources

Scott Jensen, Michael Cox, David Bender, Miao Chen, Julie England, Beth Plale, and David Leake

scjensen, miecox, dcbender, miaochen, england, plale, leake @indiana.edu

In a close collaboration between social-ecological researchers and computer scientists, we have been developing an ontology that captures the semantics of social and ecological interactions regarding forest resources. Social-ecological researchers study the interactions between institutions, actors (those who participate in decisions regarding the utilization or governance of resources), and natural resources (such as forests, fisheries, or irrigation systems) in order to understand the outcomes generated by such interactions. Such a system of resources, institutions and actors is referred to as a social-ecological system (SES). Spatial relations between the resources, actors, and institutions have an impact on both the interactions that occur and the resulting outcomes.

Our ontological work on SESs starts with a database containing 15 years of data from on-going research by the International Forestry Resources and Institutions (IFRI) network, which has collected and maintains detailed longitudinal studies of forest management systems in 18 countries. Our research in building an ontology to describe the IFRI forestry data has both broader long-term goals and more immediate applications, both of which have strong spatial components. In developing an ontology to describe the IFRI data, the long-term goal is to tie the ontology to concepts in the social-ecological system (SES) framework of [4, 5]. This is a multi-tiered approach to analyzing SESs that considers resource systems, the resource units generated in those systems, users, governance, and the actions and outcomes within a system at a particular time and place. By defining our ontology for forest resources so that concepts can be related to the SES framework, as research on SESs increasingly applies the SES framework, the actions and outcomes witnessed in a forest system and represented in terms of the ontology can be compared to actions and outcomes in other resource systems such as fisheries and irrigation. An open question is whether by building on spatial concepts to describe the relations between resources, institutions, and actors, additional inferences can be drawn which would enable comparison of the actions and outcomes in SESs for different resources.

The goal of constructing the IFRI ontology is four-fold:

- To advance knowledge representation for SESs, starting with the detailed forestry data captured in IFRI to identify spatial and other concepts needed to describe SES relationships as an ontology.
- To support education, to train future social-ecological researchers, and as the basis of a graphical tool for explaining social-ecological concepts to K-12 students.
- To enhance query capabilities for researchers in the IFRI network, by enabling SPARQL queries to be executed against a semantic store populated from the IFRI master relational database with additional inferences enabled by linking to externally developed ontologies.
- To provide enhanced metadata for archiving, based on additional inferences obtained from the ontology.
The IFRI research instrument used to capture detailed data about SESs for forestry resources is very detailed, consisting of 10 forms totaling 180 pages. Additionally, when completing the research instrument during a site visit in the field, the research team completes many of the forms numerous times. For example, researchers frequently take detailed measurements of the vegetation at each of 30 plots within each forest at a site. In addition, the research instrument captures detailed data regarding the settlements that interact with the forest resources, the multiple user groups that harvest or otherwise use the forest (including for relaxation or religious purposes), the products that users generate from forest resources, governing or administrative bodies involved in managing or monitoring resource use, and the complex interactions between these groups and the resulting outcomes. As a component of the on-going IFRI research, changes in the forests are tracked over time across multiple visits.

There is a strong geospatial component to the IFRI data, reflecting the relationships between forest resources, users, settlements, and administrators. Some of the spatial relationships are expressed explicitly as latitude and longitude coordinates, such as the location of each forest in a site or the position of each detailed plot measurement in a forest, while other relationships are expressed as distances measured in kilometers. However, many important spatial concepts captured in research about SESs are not based on specific geographic coordinates; such as COVERS, CONTAINS, ADJACENT, or INTERSECT. As noted in [1], which focuses on interior and boundary spatial relationships such as “A” contains “B”, such relationships do not require distance metrics. These types of relationships can be expressed using existing GIS tools based on established standards [3], but the SES literature also often describes spatial relations using terms such as NEAR or FAR [2] where the semantics are less precise than for concepts such as “contains” or “intersects”. As with interior-boundary spatial relationships, fuzzy spatial concepts such as “near” and “far” are also likely to be relevant in many domains.

SES researchers also need to address spatial concepts that are expressed temporally, based on nonsymmetrical spatial relations, and others that use boundaries without fixed geospatial coordinates. In studying forest resources, the IFRI researchers capture spatial relationships such as the distance from user settlements to both forests and administrative centers that govern the forests as measured by the time required to walk the distance in both the dry and wet seasons. In studying irrigation systems, the nonsymmetrical upstream-downstream relationship between water users is important in that the impact of users upstream in the SES can be much greater than that of users downstream. Other SESs such as fisheries often have to deal with resource and administrative boundaries that are either fluid or migratory. Two open questions we are looking to address are whether such spatial relationships are specific to SES researchers or apply in other domains, and how an ontology can define and capture spatial relationships expressed temporally, for nonsymmetrical relationships, or based on non-fixed boundaries - allowing new inferences to be drawn regarding the research data collected.

As an educational tool, the IFRI ontology will provide a valuable tool for training both students and field researchers around the globe. The volume and detail of the IFRI data, along with the complex relationships it captures between resources, users, and institutions presents a challenge in training new researchers to both collect and use the data. At Indiana University, the Workshop in Political Theory and Policy Analysis runs a class every Fall semester to train
students on the research instrument in preparation for field research. Training sessions are also held at IU for researchers in the IFRI Collaborative Research Centers (CRCs) located around the globe. These participants are then responsible for training researchers in the CRCs in their home countries. By providing a graphical depiction of the relationships in the data, the ontology will help new researchers visualize data relationships. In addition to being used to teach new field researchers, the IFRI ontology could be combined with other graphical tools such as Google Earth to teach K-12 students about the complex interactions between people and the environments in social-ecological systems in a way that is easier for them to grasp. By linking the ontology, and particularly spatial concepts, to less technical external resources such as DBPedia, these tools could enable K-12 students to further explore ecological systems in different regions of the world.

The ontology will also benefit experienced IFRI researchers by providing the ability to automatically infer new relationships that may be implicit in the data and also by linking to external ontologies that provide further details regarding plant species, regions, user groups, or governance. As an example, for each forest visited, the scientists collect information on the species, height, and DBH (diameter at breast height) of each tree in the plots studied. By linking this data to external ontologies and taxonomies such as the Missouri Botanical Garden’s Tropicos database [6] which contains over a million scientific names for plant species, additional inferences can be drawn about the trees in a plot. For spatial data, an open issue is what additional relationships can be inferred, particular for “fuzzy” spatial relationships based on external ontologies.

We have an on-going research project to archive long-running longitudinal studies such as the IFRI data and describe the data with rich metadata that can be used over the long-term for data discovery, enabling further reuse of this valuable data source by future SES researchers. The same approach of linking to external ontologies that is used to infer new relationships for current researchers can also be used to enhance the metadata used for discovering archived data. In addition, the metadata for archived research data can be enhanced incrementally, so as new spatial ontologies are developed, additional inferences can be drawn to enhance the metadata describing the spatial relations such as distances between user groups, forest resources and forest administrators even in the archived historical data.

References: