A Survey of Geospatial Resources, Representation and Reasoning

Jerry Hobbs James Blythe Hans Chalupsky
Thomas A. Russ

University of Southern California
Information Sciences Institute

Contents

Executive Summary 1

1 Introduction 3
- 1.1 Motivations . . . 3
- 1.2 What is an Ontology? . . . 5
- 1.3 Overview . . . 7

2 Desired Coverage of a Common Geospatial Ontology 11
- 2.1 Topology . . . 11
- 2.2 Dimension, Direction, and Shape . . . 14
- 2.3 Natural and Geopolitical Regions . . . 17
- 2.4 Granularity . . . 18
- 2.5 Space, Time, Motion, and Processes . . . 19

3 Research in Spatial Representation and Reasoning 21
- 3.1 Fundamental Representational Choices . . . 21
- 3.2 Region Connection Calculus . . . 22
- 3.3 Point Set Topology . . . 22
- 3.4 Combined Qualitative and Quantitative Representations . . . 24
- 3.5 Directions . . . 25
- 3.6 Space, Time, and Motion . . . 26

4 Representative Datasets 29
- 4.1 Geographic Names Information Service . . . 29
  - 4.1.1 Domestic Names . . . 29
  - 4.1.2 Foreign Names . . . 30
- 4.2 Getty Thesaurus of Geographic Names . . . 31
- 4.3 MetaCarta . . . 33
- 4.4 Digital Elevation Model . . . 33
- 4.5 TIGER/Line Files . . . 34

4.6 The Seamless Data Distribution System . . . 34
4.7 Google Earth . . . 35

5 Geographic Information Systems **37**
5.1 ArcGIS . . . 37
5.1.1 Capture of Geographic Knowledge in a GIS . . . 38
5.1.2 Representational Primitives . . . 38
5.1.3 Feature Representation . . . 39
5.1.4 Spatial Relationships . . . 39
5.1.5 Thematic Layers . . . 39
5.1.6 Computation and Reasoning . . . 40
5.1.7 Summary with Respect to Analysis Dimensions . . . 41
5.2 Oracle Spatial . . . 44

6 Ontology Standards **45**
6.1 OpenGIS Feature Geometry (ISO 19107 Spatial Schema) . . . 45
6.2 ISO Standards . . . 46
6.3 Federal Geographic Data Committee . . . 48
6.4 Spatial Data Transfer Standard . . . 49

7 Large-Scale Research Ontology Efforts **51**
7.1 ResearchCyc . . . 51
7.2 SUMO . . . 59
7.3 SWEET . . . 60
7.4 DOLCE . . . 60

8 Uncertainty **63**
8.1 Scale and Precision . . . 63
8.2 Correctness . . . 64
8.3 Completeness . . . 65
8.4 Inherent Vagueness . . . 65
8.5 Provenance . . . 66
8.6 Reasoning about Uncertainty . . . 66
8.7 Visualizing Uncertainty . . . 67

9 Recommendations **71**
9.1 Introduction . . . 71
9.2 Semantic Interoperability . . . 72
9.3 Representation . . . 73
9.4 Context-sensitive Analysis . . . 74

A Geodetic Systems **75**

B ISO Standards **77**

C USGS GNIS Feature Classes **81**

References **85**

List of Figures

3.1 RCC-8 Relations Depicted in Two Dimensions 22
3.2 Egenhofer and Franzosa's 9 Intersection Relations Depicted in Two Dimensions 23
3.3 Topological Relations Are Ambiguous 24

4.1 Sample Results from the Getty Thesaurus of Geographic Names 32

7.1 Direct Descendants of `SpatialThing` 52
7.2 The Cyc collection hierarchy below `SpatialThing` down two levels. . . . 54

8.1 Uncertainty in Storm Location 68
8.2 Uncertainty of Storm Movement 68
8.3 Probability-based Wind Speed Forecast 69

List of Tables

4.1 GEOnet Names Feature Classes 30

6.1 The OpenGIS Feature Geometry (ISO 19107 Spatial Schema) Objects . . . 47

A.1 Comparison of Reference Ellipsoids 75

B.1 ISO Standards for Geospatial Information 77

C.1 USGS Gazetteer Feature Classes 81

Executive Summary

In this report we survey the current state of the art on research and practice in geospatial ontology. After motivating the need for geospatial ontologies, we describe a range of what one can mean by the term "ontology"—a standard vocabulary, a hierarchy of terms, or a full theory in first-order logic. In the survey we look at examples of all three, but we urge greater formalization as a path to more reliable interoperability.

We then describe the coverage we believe a widely-used geospatial ontology would have to have. It should include the possibility of rich descriptions of the topology of complex regions and three-dimensional structures. It should include a way of talking about direction, multiple frames of reference, shape, and size. It should have access to large compendia of natural, man-made, and geopolitical entities. One should be able to view geospatial fields and objects at different granularities. It should be possible to combine it with ontologies of time and of events and processes to produce ontologies of motion and change of varying complexity.

We examine the state of the art in spatial representation and reasoning in artificial intelligence, particularly with respect to qualitative topological information, hybrid qualitative-quantitative representations, directions, and combining space and time.

We then survey representative examples of several categories of geospatial resource, looking at their implicit or explicit ontology. The types of resource include geospatial datasets, such as the Getty Thesaurus of Geographic Names; geographic information systems, such as ArcGIS; geospatial ontology standards, such as the OpenGIS Feature Geometry; and large-scale research efforts on geospatial ontology, such as ResearchCyc and SUMO.

We discuss the problem of uncertainty in geospatial data, and list several different kinds of uncertainty. We briefly survey work on reasoning about and visualizing uncertainty in information.

We close with three principal recommendations:

- Encourage more reliable interoperability by more precise specifications of the semantics of geospatial standards.

- Encourage research in qualitative geospatial reasoning, hybrid representations, and combining geospatial theories with theories of time, events, and processes.

- Encourage development of geospatial methods of analysis that are sensitive to the context of the larger operation.

We believe all three of these challenges are tractable but not trivial, and consequently would benefit from concerted efforts at this time.

Chapter 1

Introduction

1.1 Motivations

Geospatial information is among the most important kinds of information that people use, and there are an immense number of geospatial resources that provide this information. Yet there are two problems that prevent it from being used more effectively.[1]

1. Lack of Interoperability: Resources often store the information in idiosyncratic and undocumented formats, and different resources sometimes conceptualize the domain from very different perspectives. One researcher we interviewed is building a system that utilizes many diverse resources, including satellite images, maps, and databases of addresses. He reports that the principal problem he faces is discovering how his resources represent the data in the first place, and then mapping the data from the different resources into a common representation. If a shared ontology were agreed upon and came into widespread use, this problem would largely go away.

   Another researcher tells about using a database about parcels of land. One of the fields was "Ownership Type", and the meaning and possible values were not explained, being viewed as self-esplanatory. The values she found in this field were "Public", "Private", "Unknown", and "1". Because of the lack of documentation on what things meant, she had no idea whether a value of "1" was significant or just noise. She had no option but to treat it as "Unknown". Had there been an explicit ontology that this database was linked to, she would not have faced this problem.

2. Qualitative Data: Most resources provide precise, quantitative data. But this is often difficult to link with the more qualitative ways humans understand the data, making it hard to acquire data from human reports and hard to communicate data to human

[1]We have greatly profited in putting together this survey from the advice and comments of Naicong Li, Snehal Thakkar, and Dan Goldberg.

> agents. Moreover, often precise data is simply lacking in situations where qualitative data would be adequate, and indeed, more efficient, for the task at hand. A better and broader understanding of qualitative geospatial concepts and an understanding of their relation to quantitative representations would greatly alleviate these problems.

Consider for example a driver heading up Highway 9. He calls 911 and reports that he has spotted a forest fire off to the left just after the creek. This report would seem to raise problems. It is imprecise. No latitude or longitude is given. How are we to relate "left" to the north-east-south-west system of coordinates. What is meant by "after"? How do we know what direction he was traveling, if he says he was driving "up" Highway 9? What is meant by "just"? This report is not in a format that could be entered into a database. Yet his directions are perfectly adequate for human firefighters.

Similar problems arise in the context of urban warfare. Soldiers making a sweep of a neighborhood would be immensely aided by very very detailed geospatial data. It would help them if they knew where the fences are, where the power lines are, what the lines of sight are, and even whether doors opened inward or outward. It could be that the same neighborhood was swept six months before by a different unit. They would have learned much of that information. But can it be captured for later use by others? Today soldiers returning from an operation are sometimes debriefed by geospatial information specialists, who enter some of that information into a database. But this is a difficult process of translating commonsense descriptions of an area into a quantitative schema, and important information that does not fit the schema is simply lost.

Smith and Mark (2001; 2003) call for a kind of naive or qualitative geomorphology. Scientific work in the area is done from a "field-based perspective", where every point on a grid has an elevation. The everyday view is in terms of an "object-based perspective", that has concepts like "mountain" and "valley". These concepts have no precise mathematical formulation. Yet these are the terms in which most users and certainly all casual users will understand the information. We need better theories of the qualitative concepts and better links between them and the quantitative, scientific concepts.

Progress in qualitative geospatial representation and reasoning would allow more of this information to be captured, and would make it easier to capture, since it deals with the terms in which the ordinary user couches the information.

The soldier is potentially the best sensor we have on the battlefield, and better understanding of qualitative geospatial information will enable us to utilize these "sensors" to the fullest.

Issues of uncertainty and provenance are at the heart of these problems. If there is one report at 3 p.m. that there is a barrier at a particular location and another, less reliable report at 4 p.m. that there isn't, how do we decide which is the case. More significantly, how do we represent and convey our uncertainty in the matter.

In this survey we investigate what is requuired to overcome these problems, and review the research and resources that point the way to solutions.

1.2 What is an Ontology?

Before going into deeper detail about various geospatial ontologies, we should say something about what ontologies are and why they are important.

An ontology is a specification of a vocabulary of concepts, or in logical terms, predicates, together with some indication of their meanings.

There is a range of levels of precision that are possible.

1. **Vocabulary List**: This is just a list of terms to be used, and their meanings. For example, should we call a political region at one level below the level of "nation" a "state" or a "province"? Should we call the nation-level concept a "nation" or a "country"? The vocabulary list should answer questions like this. The terms in the vocabulary should at the very least be accompanied by natural language descriptions of their meanings. These should be precise enough to allow us to decide difficult cases. For example, are the counties of England at the "county" level or at the "province" or "state" level?

2. **Hierarchy of Terms**: At this level, the vocabulary is organized into a hierarchy that captures subset and superset relations, and also typically captures other simple relations, such as "part of" relations and possession of certain types of attributes. This extra information can help resolve certain ambiguities and mismatches. For example, we could take the point of view that a country *is* a geographical region, or we could take the point of view that a country is a geopolitical entity and *has* a geographical region. Two geospatial resources may differ in this regard, and it is important to understand the difference if we are accessing information from both. A hierarchy of terms is usually represented in some variety of description logic, like OWL, and provides at least some rudimentary inference capabilities.

3. **Logical Theory**: This is a set of terms, or predicates, representing the concepts, and a set of axioms that define or characterize the predicates, interrelate them, and constrain their possible interpretations. An example of an axiom might be

$$\forall x.\text{province}(x) \Rightarrow \exists y.\text{country}(y) \land \text{part}(x, y) \tag{1.1}$$

that says that every province has a country that it is part of. In addition to pinning down meaning more precisely than a natural language description can do, a logical theory can also provide powerful inference capabilities. The usual way of specifying an ontology at this level is as a theory in some variant of first-order logic.

One might ask why the full machinery of a logical theory would be required. Why isn't good old English prose enough? The answer is that natural language descriptions are often ambiguous, vague or circular. Examples are often given, but these examples are rarely exhaustive, and don't help us decide the difficult cases. Consider an example from an early version of one large ontology. The prose description of the meaning of "AstronomicalBody" was

> The class of all astronomical objects of significant size, like planets, stars, and asteroids. Earth is an AstronomicalBody, but every region of Earth is a GeographicalArea.

Now we can ask whether Jupiter's Red Spot is an "AstronomicalBody". Astronomers discovered it, so it is astronomical, and its size is larger than the Earth and hence is significant. But it seems more like the regions of Earth than like stars, planets, and asteroids.

The definition of Continent was

> One of the seven largest land areas on Earth.

What about Eurasia? It's the largest named land area on earth, so it seems like it would be a continent. This description does not rule out overlapping land areas, so could we include Europe, Asia, and Russia in the list of seven? If so, Australia is no longer a continent. If not, Greenland *is* a continent.

Another example will illustrate the way axioms can constrain possible meanings. There were two axioms in this ontology concerning the predicate "near".

$$\forall xy.\text{near}(x, y) \Rightarrow \neg\text{connected}(x, y) \tag{1.2}$$

$$\forall xy.\text{near}(x, y) \Rightarrow \text{near}(x, y) \tag{1.3}$$

If x is near y then x and y are not connected, and if x is near y then y is near x. These axioms rule out the possibility of interpreting "near" as "connected"; that would violate axiom 1.2. They also rule out interpreting "near" as "bigger-than", because that would violate axiom 1.3; "bigger-than" is not a symmetric relation.

But we could still interpret "near" as "far"! If x is far from y, then they are not connected, and if x is far from y, then y is far from x. In order to rule out this interpretation, we would have to add more axioms. For example, we could introduce a predicate "far" and axioms 1.4 and 1.5 similar to axioms 1.2 and 1.3.

$$\forall xy.\text{far}(x, y) \Rightarrow \neg\text{connected}(x, y) \tag{1.4}$$

$$\forall xy.\text{far}(x, y) \Rightarrow \text{far}(x, y) \tag{1.5}$$

$$\forall xyz.\text{near}(y, x) \wedge \text{far}(z, x) \Rightarrow \text{distance}(x, y) < \text{distance}(x, z) \tag{1.6}$$

If x is far from y, then x and y are not connected, and y is far from x. Then we could relate the two concepts by adding axiom 1.6 that says that if y is near x and z is far from x, then the distance between x and y is less than the distance between x and z. Then "near" could no longer be interpreted as "far".

There are several reasons it is important to develop ontologies in geospatial and other areas, and to give them as great a precision and inferential power as possible. Artifacts, databases and other resources need to be able to communicate with each other, so we can leverage the strengths of each in complex systems. This is made easier if they already have a common language. Artifacts are getting smarter, and potentially very much smarter. Their functionality will be greatly enhanced if they have the relevant commonsense and expert knowledge in their area of expertise, and can reason with it. Finally, artifacts, databases, and other resources need to be able to communicate with people. For this, they need ontologies in which expert knowledge can be captured and linked to commonsense conceptions of the domain.

1.3 Overview

In this survey, we first consider the various topic areas one should be able to cover in a geospatial ontology. We have identified the following broad categories:

1. Topology, including connectivity, overlaps, topological shape such as "closed loop", boundaries and surfaces, holes, and knots. For example, one should be able to capture qualitative information about the structure of a complex warren of buildings or a cave.

2. Dimension and Orientation, including characterizations of the concept of "dimension"; projections from one representation to a system with more or fewer dimensions; qualitative notions of "about parallel" and "about perpendicular"; frames of reference, including earth-based (north-south), agent-based (left-right), vehicle-based (port-starboard), and force-based (upstream-downstream); transformations between frames of reference and coordinate systems; and qualitative trigonometry that imposes granularities on orientations.

3. Shape, including both two-dimensional and three-dimensional shape; qualitative descriptors like "round" and "tall"; the notions of "fits in"; symmetry; linking with various quantitative representations of shape, such as polygons; and the functionality of shape.

4. Size, including length, distance, area, and volume, both precise and qualitative measures and the relation between the two.

5. Locations and Regions, including the latitude, longitude, elevation system, and natural and geopolitical regions.

6. Spatial Aggregates, such as sets of regions and discontinuous regions.

7. Granularity, including, for example, viewing a city as a point, a region, or a volume; merging, splitting, and filtering on properties; and various levels of quantitative and qualitative precision.

8. Uncertainty, including both lack of information and inherent conceptual vagueness.

For each of these areas, we outline some basic ideas and mention research in the area where research has been done.

We next review in somewhat greater detail the research and concerns of some of the principal researchers in the area of geospatial representation and reasoning.

Then we examine representative geospatial resources in four categories and ask what seem to be the appropriate questions for each. In general, we assess each resource as to its quantitative vs. qualitative character, its theoretical basis, its degree of formality, the formalism used, the availability of a reasoning engine for the resource, the novelty of its approach, and the extent of its adoption in the various relevant communities.

The categories of resources, relevant questions, and representative examples are as follows:

1. Representative datasets:

   Questions: What is the implicit or explicit schema/ontology? How can its content, including metadata, be characterized?

   Resources: The NGA gazetteer Geonet Names; Getty Thesaurus of Geographical Names; Metacarta; the TIGER/Line files of the US Census Bureau; Digital Elevation Model; the Seamless Data Distribution System, Earth Resources Observation and Science (EROS) of the USGS; and Google Earth.

2. GIS systems, especially with a geospatial reasoning component:

   Questions: Is a dataset included? What dataset? What sort of geospatial reasoning is done? For example, does it do region calculus operations such as intersections and convex hull? Does it handle linear calculus operations such as distance and direction? Does it handle spherical geometry?

   Resources: Oracle 10g; ARCInfo.

3. Ontology standards:

Questions: What conceptual domains do they cover? How is meaning specified? What is the level of formalization?

Resources: ISO Standards 19107 (OpenGIS), 13249, 19108, and 19125; the FGDC Metadata Standard; the Spatial Data Transfer Standard (SDTS) of USGS.

4. Large-scale research ontology efforts:

   Questions: What conceptual domains do they cover? How is meaning specified? What is the level of formalization?

   Resources: ResearchCyc; SUMO; SWEET; DOLCE.

Finally, we conclude with a discussion of uncertainty and provenance, and several recommendations for future research.

Chapter 2

Desired Coverage of a Common Geospatial Ontology

2.1 Topology

Topology is, roughly, the study of properties that remain invariant under continuous deformations. We have all seen maps of the United States in which the size of the state corresponds not to its area but to its population. So New Jersey is larger than Montana. This is topologically the same as a normal map of the United States; all the connections are the same. Indiana still touches only Ohio, Kentucky, Illinois, and Michigan, even though the sizes are distorted. Other properties that remain the same under such a transformation, and hence are topological, are whether a path or surface is open or closed, whether one point is between another two on a path, whether a region is connected and whether it has a hole in it, and whether a path ties a knot in itself and if so, what class of knot.

Children learn topological relations before they learn metric ones, and topological relations retain their primacy for people all their lives. Few of us could say exactly how far our desks are from each of the walls, but we can all say whether it is against a wall and if so, which wall. It is only when we get into scientific or technical work that metric information begins to prevail, and even there, our first approach to a problem is likely to be topological.

One reason for having topological information is that it provides a simpler yet very functional account of the environment of interest, with a consequent gain in efficient computation. Suppose, for example, we would like to know whether one can get from California to Utah without going through Nevada. If we have the connectivities of states represented in the form of a graph, this is a trivial search for a path under constraints. If the only information we have is the boundaries of the states in a raster format, the calculations can be very complex indeed.

Another reason for topological representation is that often metric information is not available. Topological properties, in a sense, capture what we know in the presence of uncertainty, and moreover do so in the commonsense terms that the evolution of language and cognition have proven to be the most advantageous for people.

The fundamental objects of a spatial ontology are points, curves, surfaces, and volumes, having dimensions 0, 1, 2, and 3, respectively. These are the "inherent" dimensions of the object. They also have an "embedding" dimension. A straight line segment is a curve that is embedded in one dimension, a circle is embedded in a two-dimensional space, and a spiral in a three-dimensional space.

A point can be inside another object, outside it, or on the boundary, where the boundry of an object is of a lower dimension. One of the first questions one can ask is what are the possible intersections between two objects of various dimensions. For example, the intersection of a curve and a volume will be a set of segments or points on the curve (including the empty set, if they are disjoint).

A great deal of the research on spatial representation and reasoning in artificial intelligence has focused on this question. For exmaple, the Region Connection Calculus with 8 relations (RCC-8) developed by Cohn and his colleagues (Randell *et al.* , 1992; Cohn *et al.* , 1997), among others, is an algebra of relations like Disconnected, Exactly Connected (or tangential and outside of), Partially Overlapping, Equal, Tangential Proper Part, and Non-tangential Proper Part. A similar calculus of relations has been developed by Egenhofer and his colleagues (Egenhofer & Franzosa, 1995). Here the focus is on the intersections of the interiors, exteriors and boundaries of two different objects. See Section 3.2 for a review of this work.

One ought to be able to define a kind of "mean value" property at this level that says a curve from the interior of a region to the exterior of the region must pass through the boundary of the region.

Another area of interest in classical topology as well as in spatial representation and reasoning is on varieties of connectedness. Two points in the interior of an object are connected with respect to the object if there is a curve from one to the other such that every point interior to the curve is interior to the object as well. Then an object is self-connected if any pair of points interior to the object are connected with respect to the object. A self-connected object cannot be decomposed into two disconnected parts.

Self-connected surfaces can have holes in them, and a complete topological description of them will include information about the holes (Casati & Varzi, 1997). Essentially, there is a hole if a closed curve cannot be shrunk to a point entirely within the surface, although the mathematical formulation of this is surprisingly complex. We can characterize surfaces by the number of holes they have.

We can define a similar ontology of topological shape for volumes. Volumes can have tunnels through them, and we can describe this in part in terms of the number of holes in the surface of the volume. A limiting case is one in which there is only one hole in the surface, thus giving us an indentation or cavity. The normal case is that of a torus, where a single tunnel produces two holes in the surface of an object. However, we can imagine much more complex arrangements, such as a tunnel which branches and/or one that ties knots in itself. The topological nature of such tunnels can be captured in knot theory by specifying the number of "crossings". We conjecture that the essential topological shape of any complex volume can be characterized by the addition and subtraction of objects described in this vocabulary.

A lot of the literature in qualitative spatial representation and reasoning has concerned itself with what seem to us to be false questions, and the discussions of the problems have been black holes that suck in otherwise promising research efforts. We believe it is possible to avoid these black holes, but one has to recognize them. We believe there are several principles one must adhere to in designing a geospatial ontology.

Physical objects should be distinguished from their geometric realizations. If we try to identify a city, for example, with some part of space, then we run into problems when trying to align perspectives that view cities at different granularities. Is a city a 0-dimensional, 2-dimensional or 3-dimensional entity? This is a false question. A city is just a city. It *has* geometric realizations of different dimensions, but it is not identical with any of those realizations.

As Galton (1997) pointed out, a lot of false questions arise when we try to state identities cross-dimensionally. Is a point on the boundary between the United States and Canada a part of the United States or a part of Canada or both or neither? Are geographical regions open or closed sets, in the mathematical sense of the terms? In the OWL-Time ontology (Hobbs & Pan, 2004; Pan & Hobbs, 2004), we avoided this problem by positing instants and intervals, and *begins*, *inside*, and *ends* relations between instants and intervals, but remaining silent on whether an interval is *composed* of instants. We would advocate a similar approach in a geospatial ontology. A point on the boundary between the United States and Canada is simply on the boundary, it is not *inside* either country, and whether it is a *part* of one country or the other is a question the ontology should simply be silent about.

The ontology should also be as silent as possible on issues involving infinity. Are there points at infinity? In some ontologies it is convenient to assume there are, and in some it is more convenient to assume they are not. Generally, not much hangs on the issue, and the important content can be shared between the ontologies, while trigger conditions in the antecedents of selected axioms can protect each from the assumptions of the other. This is the approach that was taken in OWL-Time.

A great deal has been written about the ontology of holes (Casati & Varzi, 1997) that we believe is largely misdirected. The argument is that holes are somehow secondary entities, dependent upon their hosts for their existence. We believe that a more profitable way of looking at the issue is as follows. Objects occupy regions of space, and holes in those objects also correspond to regions of space. A moving object can be thought of as occupying a four-dimensional cylinder, conceived as a mapping from times to regions of three-dimensional space. Similarly, a hole in a moving object corresponds to a four-dimensional cylinder, the cross-product of times and the region that corresponds to the hole at that time. Such four-dimensional regions are first-class objects in a framework that combines a geospatial ontology with a time ontology. That some piece of space-time *is* a hole depends on the object it is a hole in; the term "hole" is a relation between a region in 3- or 4-space and an object. Whether or not the relation holds depends on the object, but the region that has that hole property exists independently, regardless of whether it is a hole or not. There are no secondary entities, only properties that may or may not hold.

Another false question concerns composite objects. Consider an object consisting of a cylinder with a line segment projecting from its side. Is it a 3-dimensional object? What is its boundary, the surface of the cylinder together with the entire line segment, none of it, or just its far endpoint? The truth of the matter is obvious. It is a composite object, consisting of a 3-dimensional object and a 1-dimensional object. We can talk about the boundaries of each of the components, but we have not defined "boundary" for such a composite object. The general geospatial ontology should be silent on this issue. If particular ontology developers find it necessary to define the boundaries of composite objects, they can do so, and conditionalize them on triggers that protect the general ontology from undesired conclusions and false puzzles.

2.2 Dimension, Direction, and Shape

Whereas time is one-dimensional, space is three-dimensional, and this fact raises a whole host of issues that don't have to be faced in building a time ontology.

The first problem is characterizing dimension. Linear algebra provides a mathematical, quantitative definition of independent dimensions. But dimension is also a qualitative concept. Qualitatively, the basic fact about multiple dimensions is independence. You can't predict the ordering of two entities on one dimension just from their ordering on another dimension. From the fact that San Francisco is west of Chicago, you can't predict anything about which is farther north. Two other properties of topological flavor are the fact that the boundary of an object is of a smaller dimension, and that to separate two objects of dimension n in general requires an object of dimension at least $n-1$.

In one-dimensional time, one can go either forward or backwards (in imagination or simulation), but those are the only two possible directions. In three-dimensional space, direction becomes an interesting concept.

In describing the relative orientation of lines, the four most common concepts are "parallel", "perpendicular", "acute" or "sharp", and "oblique" or "blunt". The first term is distinguished from the last three by whether or not the lines meet, and the last three are distinguished from each other by the relative sizes of the angles subtended.

The usual treatment of direction is in terms of frames of reference. Frames of reference have been a major focus in linguistic and psychological research on the semantics of the language of space (Talmy, 1983; Tversky, 2000; Tversky *et al.* , 1999). We have four principal ways of setting up roughly orthogonal coordinate systems, essentially, world-based, person-based, vehicle-based, and force-based. The world-based system gives us the cardinal directions north-east-south-west.

The person-based system is front-right-back-left. This system is attributed to inanimate objects as well, if the objects have an intrinsic orientation. For example, the front of a building is the wall of the building containing the main entrance. So the sentence "My car is parked in back of the building," will mean that the car is by the wall of the building opposite the wall with the main entrance. Inanimate objects without an intrinsic orientation can be given one from the perspective of the viewer. The sentence "The ball is in front of the tree," means that the ball is between the tree and the speaker, because the side of the object facing the observer is defined as the front. Even inanimate objects that have intrinsic orientation may be viewed from the speaker's perspective. If someone is facing the side of a building and says the ball is in front of the building, he could mean that it is near *the* front of the building or that it is between him and the side of the building he is facing.

The most common vehicle-based frame of reference is the one used on ships—"bow", "starboard", "stern", and "port". This is often viewed as mere tradition, but in fact the reason it is needed on ships and not in cars is that on ships a person's orientation and the vehicle's orientation are not generally the same, whereas in cars they are. On a ship, "right" is ambiguous; "starboard" is not. Sailing is an activity in which all four kinds of frames of reference are relevant. A force-based frame of reference is upwind-leeward-downwind-windward. Another force-based frame of reference gives us "upriver" and "downriver", and another "uphill" and "downhill".

In addition, there are special-purpose frames of reference. In theatres, one has the terms "stage right" and "stage left" to resolve an ambiguity in a situation where two groups of people are facing each other. The regions of football fields are labeled by the team that is defending it. In the Hawaiian language, there are prepositions meaning "toward the center of the island" and "toward the ocean", a natural language example of polar coordinates.

"Up" or "above" and "down" or "below" are the terms used for the third axis in all these frames of reference, but they can mean different things. When a person is lying on his back on the grass, the sentence "There's a bird above my head," is ambiguous. It could be with respect to the world-based frame of reference—there's a bird flying above his face—or with respect to the person's intrinsic orientation—there's a bird in the grass beyond his head on the axis defined by his body.

One task of a qualitative geospatial theory is to relate all of these frames of reference, for example, by saying that north and east are at right angles, and that when you are facing north, east is to your right.

Once we have a frame of reference, we can define various qualitative granularities on it, giving us a kind of qualitative trigonometry (Liu, 1998). The most common is perhaps the one that divides the world among the four cardinal directions. If 0 degrees is true north, then the north region would be from –45 degrees to +45 degrees. For a finer granularity, we can go to 8 directions—north, northeast, east, southeast, south, southwest, west, and northwest. The obvious way to divide up the world between the 8 directions would seem to be in equal sectors of 45 degrees each, centered on the true direction. But one common qualitative system introduces a bias, giving the four cardinal directions only a narrow slice around the true direction. So the north region would be almost true north and nothing more, similarly for the east region, and everything in between would be northeast. This is much like the straight-obtuse-right-acute complex of terms for angles. Probably the finest granularity that is in everyday usage has 12 directions—the hours on a clock when one is facing 12. We might note when we enter a traffic circle that the exit is at two o'clock. Our concern in each of these cases is to link it at least approximately with a quantitative coordinate system. In section 3.5, we discuss various proposals for dealing with directions qualitatively.

In a time ontology, there is no interesting concept of shape. Once we have multiple dimensions, then we can talk about the shape of objects. The standard way of representing shape in geographical information systems is by means of polygons, where the endpoints of the sides of the polygon are specified. A very coarse version of this is the use of bounding boxes—rectangles—around a country to indicate the region it covers. This has the advantage of yielding efficient computation. It has the disadvantage of being a very bad approximation in many cases. If we consider the United States to include American Samoa, then the bounding box around the United States contains Mexico.

Natural language has a rich array of everyday terms for describing prototypical shapes, including "round", "square", "straight", "curved", "wide", "narrow", "convex", "concave", "sharp" and "blunt". These are the terms that people are most likely to use for describing objects, so it is important for geospatial systems to be able to handle them, and to use contextual information to relate them to quantitative information. If something is identified as a round building, how will that be related to a geospatial resource that stores shapes as polygons? It can be done, but methods need to be worked out.

It is also important to be able to deal with relative shapes. One object can be rounder, or sharper, or more convex than another. We need to be able to deal with the concepts of two entities having roughly the same shape, and of one object "fitting into" a hole in another object, as a door fits into its frame. We need notions of symmetry, comparing the shape of one part of an object with the shape of another.

A geospatial ontology should also link with an event and process ontology to enable the statement of knowledge about the functionality of shape. In artifacts, the shape of components is almost always functional. It is no accident that wheels are round and not square, and it is no accident that doors fit in their doorframes. The artifacts wouldn't work right if the shapes were wrong. Wear on a component of an artifact frequently causes the artifact as a whole to fail because the component changes shape enough that it can no longer perform its function.

In natural objects, shape is not exactly functional, but it certainly can have consequences in how we interact with it. The slope of a hill is an important factor in trafficability, for example. The contour of the land influences lines of sight.

Size and shape are independent qualities, but we also need to accommodate everyday descriptions of size—of length, area, and volume. The mathematical treatment of these measures is well-worked-out. We need to develop context-dependent ways of relating everyday descriptors of the size of objects to more quantitative representations. For example, a reference to "the tall building" will be resolved by comparing the heights of contextually relevant candidates.

2.3 Natural and Geopolitical Regions

A geospatial ontology of wide applicability should have a number of descriptors of the most commonly referred to regions, including land masses like continents and islands; bodies of water like oceans, rivers and lakes; and terrain features like mountains, valleys, forests, and deserts. As Mark and Smith (2003) have noted, many of these objects do not define precise regions. They consider the case of mountains and valleys. When you climb Mount Whitney, at what point do you actually set foot on the mountain? When there are two peaks between you and the highest point on Mount Whitney, you are certainly not on the mountain. When you are at the highest point, you certainly are, and you certainly are when you are climbing up the slope that culminates in that point. But what about the ridge between Mount Whitney and the peak just south of it? Terms like "mountain" and "valley" are inherently imprecise, and invite description in terms of the so-called "egg yolk" theories (Cohn & Gotts, 1996), which identify regions of "certainly in" and "certainly not in" and a gray area in between.

A geospatial ontology also needs standardized terms for political regions, including countries; political subdivisions such as provinces and counties; municipalities including cities, towns, and villages; other districts such as Indian reservations, national parks, and regulatory zones; and residences and places of business with their addresses. Multiple grammars for addresses in different countries need to be supported.

A problem that arises here is uniformity of terminology. One nation's state is another nation's province, and counties are at the second level of division in England while being at the third level of division in America. Is Cornwall a state, a province, or a county. This is not a deep conceptual issue, but agreement has to be reached if different geospatial resources are to be interoperable.

In addition, a complete geospatial ontology will need to be able to deal with spatial aggregates, including sets of points and lines and their distributions in regions, to handle notions like "dense", as well as discontinuous regions and sets of regions.

2.4 Granularity

A city can be viewed as a point, a two-dimensional region, or a three-dimensional volume. When we are planning a trip, a road is a line. When we are driving down the road, it is a surface, and we have to pay attention to where on that surface we are. When we hit a pothole, we realize the road is a volume. These are differences in the granularity with which we perceive our world. Granularity considerations are pervasive. Adopting the best granularity is one way we make computations about the world tractable. Granularity has to be built into a geospatial ontology right from the beginning, and not added as an afterthought.

It is important to note that granularity is more than simply the scale of a map. It is rather a way of selecting what is important. If you are hiking in the Sierras, the trail you are walking on may only be two feet wide and yet it will show on the map, while a boulder thirty feet around that you walk past will not show on the map.

Imagine a map for trafficability for tanks that averages land type over ten-meter squares. An area of rice paddies will all be labeled as shallow water with a firm bottom, and hence trafficable. But what is missed is that every 30 or 40 meters, there is an earthen barrier two or three meters high between fields and these force the tanks into periodic moments of maximum vulnerability. Relevant obstacles have to be shown regardless of size.

Much work in qualitative representation and reasoning can be viewed as investigations of different granularities imposed on scalar quantities. Much qualitative physics (e.g., Forbus (1981; 1984); deKleer and Brown (1984)) has reduced measurement to three values – minus, zero, and plus. We want to know, for example, whether a component of a device is moving up or down or staying still.

Natural language typically divides scales into three primary, imprecisely demarcated values—large, medium, and small on a size scale, for example. ResearchCyc provides for this kind of division as well. But for this to be truly useful, we need also to be able to align these categories with a qualitative theory of distributions on the one hand and a theory of functionality on the other. With respect to distributions, if someone is characterized as tall, they are probably in the 70th percentile or higher, but this may depend on the shape of the distribution as well. With respect to functionality, often when we call someone tall, we mean that they are tall enough or too tall for some purpose. This will require linking with an ontology of events and processes. Hooks are needed to these other ontologies if we are to make maximum use of these qualitative concepts.

A certain amount of work has been done in AI on order-of-magnitude reasoning. Measures on objects or processes are specified to within an order of magnitude. The principal feature of this granularity is that normally one can ignore phenomena at lower orders of magnitude, greatly simplifying compuations. We weigh a letter to know how many stamps to put on it. But the stamps themselves add to the weight of the letter. We know we can ignore this extra weight. It is of a lower order of magnitude.

Hobbs and Kreinovich (Hobbs, 2000; Hobbs & Kreinovich, 2001) investigate a finer-grained version of this that they call “half orders of magnitude”. If we construct a logarithmic scale of a base of roughly 3 or the square root of 10, then we can do shallow, defeasible arithmetic on the categories of measurements. That is, we will usually be right in addition and multiplication if we don’t do too many of the operations. This is sufficient for many everyday purposes. Moreover, when we increase or decrease the size of something by a half order of magnitude, we generally change the way we have to interact with it. There is not a qualitative difference in the way someone lives in a house of 1200 square feet versus 1500 square feet, but a house of 5000 square feet forces one to change everyday strategies.

2.5 Space, Time, Motion, and Processes

The geospatial ontology will need to be linked to a time ontology, such as OWL-Time, in order to deal with the way locations and other geospatial properties change through time. These issues are discussed in Section 3.6.

Chapter 3

Research in Spatial Representation and Reasoning

3.1 Fundamental Representational Choices

One of the fundamental representational choices to be made concerns what the primitive item used for the representation of space should be. The main choice is between a point-centric or a region-centric representation. A point-centric approach uses points as the primitives and then defines regions as sets of points. A region-centric approach uses regions as the primitive and defines points in terms of regions.

Point-centric representations align well with raster-based datasets such as imagery and digital elevation models. This approach is also the basis of much of the mathematics-based reasoning about space. Region-centric approaches align well with qualitative reasoning research. Many researchers in the qualitative reasoning school of thought see regions as being more closely aligned with the way people think about space. A more in-depth discussion of this can be found in (Cohn & Hazarika, 2001).

We find most compelling a view expressed by Galton (1997), that both points and regions should be primitive and that relations between them (inside, onBoundary) should be defined, but that one should not try to construct one out of the other. To quote Galton (pp. 323–4), “while we want to be able to talk about volumes, surfaces, edges, and points, these entities are each *sui generis*. It does not make sense to combine entities of different dimensions; a surface is an utterly different *kind* of thing from a volume . . .” Many false paradoxes can be avoided in this way.

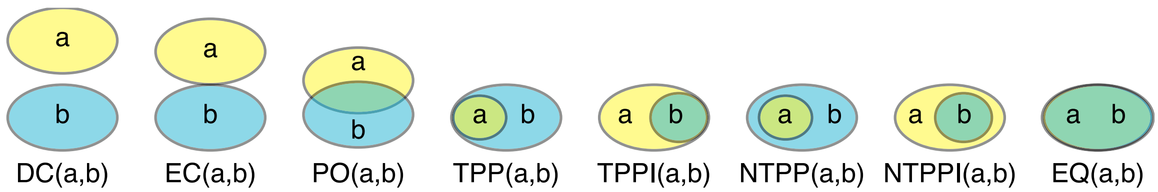

Figure 3.1: RCC-8 Relations Depicted in Two Dimensions

3.2 Region Connection Calculus

The dominant topological representational formalism is the Region Connection Calculus (Randell *et al.* , 1992; Cohn *et al.* , 1997). The basis of this representation is the use of regions as geometric primitives and a primitive connection predicate C. Regions have boundaries and interiors. A set of mutually exclusive, collectively exhaustive relations can then be defined based on this connection primitive and the boundaries and interiors of regions. Although there are several possible sets of such mutually exclusive and collectively exhaustive relations possible, the most popular formulation involves eight such relations. This is known as the RCC-8 set of relations. A two-dimensional graphical view of those relations is shown in Figure 3.1. The same relations are sufficient to handle three-dimensional figures as well, and in fact they correspond closely to the relations in Allen's temporal interval calculus. (The five extra relations there arise due to the linear ordering of time.)

Topological relations define the connectedness of regions. The RCC-8 is a rigorously-defined, mathematically based representational formalism. The definition is formally specified using first order logic formulae, and the collectively exhaustive nature of the collection of relations guarantees that all cases are covered. A substantial amount of research has been done on the computational complexity of reasoning in RCC-8 and on tractable variants of RCC-8.

The RCC-8 relations and their axioms have been incorporated into larger ontologies such as Cyc (Grenon, 2003) and SUMO (Niles & Pease, 2001; Pease *et al.* , 2002).

3.3 Point Set Topology

An alternate starting point for toplogical relations is point set theory (Egenhofer & Franzosa, 1991). In this formulations, two-dimensional regions have a set of points which form a boundary, and a set of points which form the interior. If one considers the relations defined by intersections between these various sets for two figures:

$$\text{boundary}(a) \cap \text{boundary}(b) \tag{3.1}$$

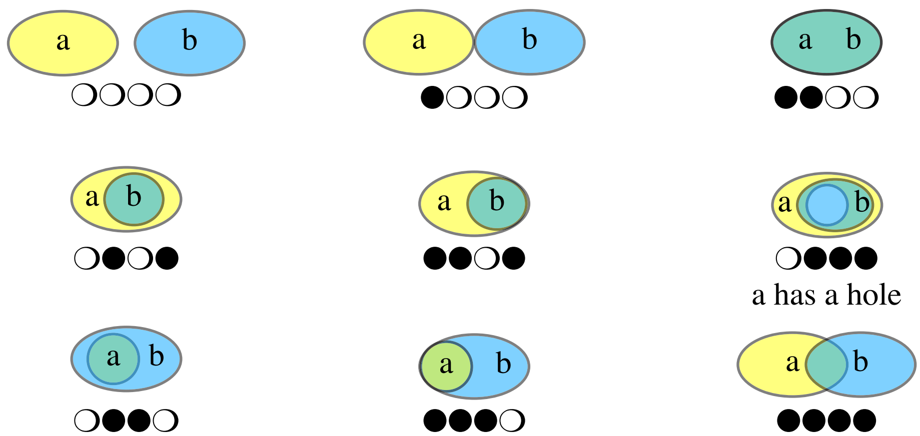

Figure 3.2: Egenhofer and Franzosa's 9 Intersection Relations Depicted in Two Dimensions. The relations are determined by considering intersections of region boundaries and interiors. Either the intersection contains points (coded by •) or it is the empty set (coded by ◦). There are four potential intersections: boundary(a) ∩ boundary(b), interior(a) ∩ interior(b), boundary(a) ∩ interior(b), and interior(a) ∩ boundary(b).

$$\text{interior}(a) \cap \text{interior}(b) \tag{3.2}$$
$$\text{boundary}(a) \cap \text{interior}(b) \tag{3.3}$$
$$\text{interior}(a) \cap \text{boundary}(b) \tag{3.4}$$

then one has $2^4 = 16$ potential relations. It turns out that of these 16, only 9 can be realized without resorting to degenerate forms or embedding figures in higher dimensional spaces. (The full details are in Egenhofer and Franzosa's paper). These nine are illustrated in Figure 3.2. The relations and the intersections have formal mathematical foundations and are defined using set algebraic operations. There is no semantic ambiguity.

The difference between the RCC-8 and Egenhofer intersection formalism is fundamentally based on a choice of primitive representation and on the relation used to generate the topological relations. Egenhofer ends up with nine relations versus the eight of RCC-8 because Egenhofer's formalism distinguishes two different cases of overlapping regions, where these are treated the same in RCC-8. Those two cases are labeled ◦••• and •••• in Figure 3.2. In the first case, regions a and b do not share boundary points; in the second case they do. These are subsumed by the partial overlap (PO) relation in RCC-8. The case where one region has a hole has overlapping regions with no intersection of the boundaries. The other difference is that Egenhofer assumes that regions are not scattered, and is thus a less general formulation than RCC-8 (Galton, 1997, p. 331).

Egenhofer's 9 intersection relations are supported by international standards such as the OpenGIS Consortium's work on ISO-19107 (Herring, 2001, pp. 124–125).

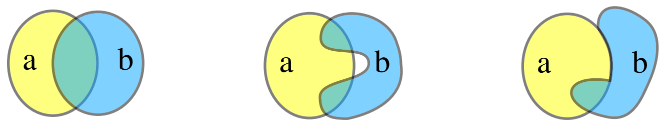

Figure 3.3: Toplogical Relations Are Ambiguous. Both RCC-8 and Egenhofer and Franzosa's 9 intersection relations are unable to distinguish between certain forms. The three sets of figures are all described by the same relation, namely PC or ●●●●, respectively.

One inherent characteristic of the topological relations in both RCC-8 and the Egenhofer and Franzosa formulation, is that they are fairly abstract. For example, the pairs of regions in Figure 3.3 are all described by the same relation. The relations do not take shape into consideration. More complicated extensions of both of these formalisms can make more distinctions, (Cohn *et al.* , 1997; Egenhofer & Franzosa, 1995), but any qualitative abstraction will necessarily lose some information.

3.4 Combined Qualitative and Quantitative Representations

In his early work in the 1980s on qualitative process theory, Forbus (Forbus, 1985) adopted a qualitative framework for space and motion that allowed distinctions, for example, between motion upwards, being stationary, and motion downwards. In more recent work, Forbus and colleagues (Forbus *et al.* , 1991) make the "poverty conjecture", that "there is no problem-independent, purely qualitative representation of space or shape". In other words, representations general enough to handle a variety of tasks must support both calculations and perceptual-like processing. This leads them to develop a hybrid representation called the "MD/PV model" after the metric diagrams and place vocabularies that it combines (Forbus, 1995). Metric diagrams provide enough quantitative information about locations to support calculations, for example, for navigation. Place vocabularies contain named, task-related contiguous regions of space, *e.g.* "inside the well", linking to the metric diagram.

Donlon and Forbus (Donlon & Forbus, 1999) describe an approach using a GIS system as a metric diagram to support joint quantitative and qualitative analysis, validated for the problem of trafficability of military vehicles on off-road terrain. The approach uses a set of formal rules for trafficability that can also be used to identify the relevant terrain features for a given query. These are used to compute the intersection of the polygons within the relevant layers of the GIS system, such that each polygon then has uniform values for every feature relevant to trafficability. Once a final qualitative result is computed, neighbouring polygons with the same label are merged to provide the final place names in the task-dependent vocabulary, which are contiguous regions with the same trafficability.

Kuipers' work on the *Spatial Semantic Hierarchy* (SSH) uses a similar model to support robot exploration and mapping, but with four separate levels (Kuipers, 2006). The "control level" is characterized by hill-climbing over a space governed by control laws. Stable attractor zones become distinctive states in this space. The "causal level" consists of a partially-known finite-state automaton over the distinctive states. The "topological level" groups these states into places, paths and regions and captures connectivity, order and containment. Finally the "geometrical level" adds metric properties such as distance and direction. The hierarchy follows some accounts of how spatial knowledge can arise from procedural navigation and mapping, and allows robust performance under some kinds of uncertainty, such as missing metric information.

3.5 Directions

There has been a certain amount of work on qualitative reasoning involving directions. The work that strikes us as the most intuitive is that of Frank (Frank, 1996). He examines systems with four and eight canonical directions, both precise and approximate, and defines operations of reversing one's self and composing two directions—going in one direction for a while and then going in another. For example, if you go north and then northeast, the composition is approximately north. If you go southwest, and then either north, northeast, or east, then you can't really say anything about the composition. Frank labels the result approximately zero.

A different kind of granularity on directions was proposed by (Freksa, 1992). Imagine an agent going from point **a** to point **b**. There are functionally six regions—the right and left regions in front of him at **b**, the right and left regions behind him at **a**, and the right and left regions between **a** and **b**. Add the axes of his motion and the perpendicular axes at **a** and **b**. One can then develop an algebra of possible combinations.

Moratz and his colleagues (Moratz *et al.* , 2000) introduce a set of basic "dipole relations" that can obtain between two vectors. The vectors can be parallel or perpendicular, and one can be right of, be left of or intersect the other. The combinations of these yield 24 relations. These relations can be composed in various ways and employed in qualitative navigation problems.

Liu (1998) develops a qualitative trigonometry. The length of two line segments can be compared and given one of five values—less, slightly less, equal, slightly greater, and greater. The angle between them can be classified as having one of five values—acute, slightly acute, right, slightly obtuse, and obtuse. (These are easy judgments for us to make because we are so good at detecting symmetry.) One can then solve qualitatively the kinds of problems one solves quantitiatively in traditional trigonometry. For example, given the qualitative values of two sides and an angle of a triangle, what are the qualitative sizes of the other two angles and the third side. Of course, certain ambiguities can be introduced, but Liu is able to describe the range of possible ambiguities. He develops an algebra for solving these kinds of problems efficiently.

3.6 Space, Time, and Motion

Given an ontology of space and an ontology of time, we can combine them into an ontology in which we can describe motion and other kinds of change.

The simplest kind of ontology of time that is common in artificial intelligence research is the situation calculus, in which there is a single agent in the world, and each of the agent's actions take the world from one state to the next. In this ontology of time, there is no time as such, only discrete changes. We see a combination of qualitative space and time in diagrams in papers on RCC-8 which show the possibly adjacent conditions that might hold. For example, if region **A** is at first a tangential proper part of **B** and there is a change to a new RCC-8 relation, then that relation must be "nontangential proper part", "partial overlap" or "equal". No other transitions are allowed.

Hayes (Hayes, 1985b; Hayes, 1985a) proposed the idea of "histories" as a model of events as objects move about through time. These are essentially 4-dimensional cylinders, the mapping from times, in whatever ontology of time one has, to the region of space that an or event occupies at that time, in whatever ontology of regions one has.

Galton (Galton, 1997) develops an ontology of motion essentially by taking the cross product of a topological ontology of time, loosely related to Allen's temporal calculus (Allen, 1984), and an ontology of space similar to RCC-8. He goes on to define what the notion of continuity would look like in such a framework, even in cases where space and/or time is conceived of as a discrete system. This is important when we know a process is continuous and want to know what constraints that places on the possible courses of events. Bennett and his colleagues (Bennett *et al.* , 2002) build on very similar ideas, formalizing the theory in modal logic.

Richer ontologies of time have been developed. For example, OWL-Time (Hobbs & Pan, 2004), enables a range of expressiveness from the purely qualitative to the highly quantitative. With a correspondingly broad spatial ontology, much richer ontologies of motion and change should be possible.

Yuan (1996) has argued that a truly rich ontology of space-time should incorporate ontologies of events and processes as well. There have been a number of ontologies or partial ontologies of events and processes developed recently, including those in ResearchCyc and in SUMO, the Process Specification Language (PSL) developed by Mike Gruninger and his colleagues (Schlenoff *et al.* , 2000; Grunninger, 2006), and the process specification language SPARK (Morley & Myers, 2004). The recent DTO-sponsored IKRIS project has focused on developing an "inter-theory" that will enable all of these ontologies to interoperate. This presents a unique opportunity for integrating this with a rich geospatial ontology to get characterizations of events with significant geospatial content.

Rich ontologies of events and processes have causal information such as preconditions and effects, and they enable one to specify subevent or subprocess structure. These are important in geospatial-temporal reasoning, because they allow us to infer what processes might have caused observed changes and what larger process an observed subprocess might be a part of.

Chapter 4

Representative Datasets

In this section we review representative geographic datasets. These span the range from geographic names and gazetteers, to sources of geographic information such as borders, imagery and other features that can be geo-referenced. These are not ontologies strictly speaking, but they do have an implicit ontology in the sense that they deal with certain types of entities, and these entities have certain classes of possible attributes. There is an enormous number of such datasets; for example, virtually every county in the United States has one. Here we discuss only the most widely used of these.

4.1 Geographic Names Information Service

There are several sources of geographic name information. The U.S. Government formed the U.S. Board on Geographic Names (BGN) to standarize geographic name usage throughout the federal government. The Geographic Names Information Service (United States Geological Survey, 1981) is the federal standard for geographic names inside the United States and in Antarctica. Responsibility for domestic names is handled by the United States Geological Survey.

4.1.1 Domestic Names

Domestic names are classified into one of 65 feature classes, listed in Table C.1 in the appendix. The other information associated with a name is the state and county where it is located, the latitude, longitude and elevation; and a reference to the USGS map which displays that feature. The data contains more than 1.8 million entries. The features that appear can be very detailed. For example, there are over 180,000 church names and more than 170,000 school names in the dataset.

Table 4.1: GEOnet Names Feature Classes

Code	Feature Classification
A	Administrative region
P	Populated place
V	Vegetation
L	Locality or area
U	Undersea
R	Streets, highways, roads, or railroad
T	Hypsographic
H	Hydrographic
S	Spot feature

Geographic names are represented in these datasets as point features. There is containment information with regard to county and state in which the features are found, but the containment is purely assertional, as none of the features in the dataset have any area associated with them.

4.1.2 Foreign Names

Names for places outside Antarctica and the United States are handled by the National Geospatial-Intelligence Agency (NGA). Those names are contained in the GEOnet Names Server. This data source contains approximately 4 million features with 5.5 million names. It is updated bi-weekly. Geographic coordinates are specified as latitude and longitude based on the WGS84 Geoid. The names are classified into categories of conventional, BGN standard, historic, provisional, variant and unverified. All names are associated with a particular geographic feature. The nine general classes of features are shown in Table 4.1. The GEOnet Names dataset identifies more than 640 specific features.

Geographic names are represented in these datasets as point features. There is some limited containment information, but this coverage is not complete. There is some evidence that the containment may in some cases represent administrative containment rather than geographic containment. Plotting the locations assigned by the GEONet Names Server database to administrative districts in Iraq, for example, will show that although most of the contained locations form a compact region, there are a few outliers that are not located in geographic proximity.

One nice feature of this dataset is the provision of alternate spellings for many of the features. For example, the capital of Iraq is listed as “Bagdad”, “Baghdad” and “Baghdād”. The set of alternative spellings and transliterations is, however, incomplete—perhaps necessarily so. One example that we encountered was discovering that the preferred spelling of Fallouja used by the Los Angeles *Times* was not in the dataset.

This dataset is available free of charge.

4.2 Getty Thesaurus of Geographic Names

The Getty Thesaurus of Geographic Names® (Getty, 2006a; Getty, 2006b) is a compilation of geographic names and locations covering the entire world. Among the sources for this compilation are the USGS and NGA geographic name lists described above. The Getty Thesaurus, though, has a much better developed containment hierarchy which has been compiled through the use of additional data sources. There is editorial guidance by the curators which requires such containment information to be provided for data entry points. There are also additional attribute descriptions which make this a much richer dataset for use in describing artifacts, natural features and geopolitical regions. There are over 1700 terms defined for features, covering common items like "volcano", more specialized concepts like "urban park" and some fairly obscure features like "voivodship".

The database records for the vocabulary have provisions for bounding box specification in addition to simple point representations. An example of the information provided for "Yosemite Village" is shown in Figure 4.1. There is a containment hierarchy and limited type information available.

Licensing terms for use of this dataset are negotiable with the J. Paul Getty Trust, owner of the copyright.

Common to all of the gazeteer representations is that the feature space is flat. Beyond some general feature classes, there is no hierarchical organization of the features based on meaning or other semantic relationships. For example, in GEONet churches are spot features, but are not further identified as, for example, buildings. In the Getty Thesaurus, there are various types of park: national, provincial, state, urban, but other than sharing the string "park" in their name, they are not otherwise related ontologically.

This can interfere with the usefulness of the lookup. For example, consider interaction with the Web search page[1] for the Getty Thesaurus. If one attempts to search for "Yosemite" and specifies the place type as "park", then no results are returned. One would have to specify the place type as "national park" instead, or manually select all of the potentially relevant park types from a pop-up menu, where the features are arranged alphabetically.

[1] http://www.getty.edu/vow/TGNSearchPage.jsp

Getty Thesaurus of Geographic Names® Online
Full Record Display

New Search Previous Page Help

Vernacular Display | **English Display**

Click the icon to view the hierarchy.

ID: 2785938 **Record Type: administrative**

Yosemite Village (inhabited place)

Coordinates:
Lat: 37 44 00 N *degrees minutes* Lat: 37.7333 *decimal degrees*
Long: 119 35 00 W *degrees minutes* Long: -119.5833 *decimal degrees*

Names:
Yosemite Village (**preferred**, C,V,N)
Yosemite (C,V,N)
Yosemite National Park (C,V,N)

Hierarchical Position:
World (facet)
.... North and Central America (continent)
.... United States (nation)
.... California (state)
.... Mariposa (county)
.... Yosemite Village (inhabited place)

Place Types:
inhabited place (**preferred**, C)

Sources and Contributors:
Yosemite.......... [VP]
.... USGS, GNIS Digital Gazetteer (1994) GNIS6039391
Yosemite National Park.......... [VP]
.... USGS, GNIS Digital Gazetteer (1994) GNIS6039391
Yosemite Village.......... [VP]
.... USGS, GNIS Digital Gazetteer (1994) GNIS6039391
Subject: ..... [VP]
.... USGS, GNIS Digital Gazetteer (1994) GNIS6039391

Figure 4.1: Sample Results from the Getty Thesaurus of Geographic Names. This shows the results returned for a search of "Yosemite Village" inside Yosemite National Park.

4.3 MetaCarta

MetaCarta is a Boston-based company that uses geographic entity extraction and information retrieval techniques to provide a map-based view of a set of unstructured documents. MetaCarta tools can be used to provide visualizations of a set of geographically related documents, or to provide watches over geographic regions for intelligence analysts. The entity extraction system makes use of a set of proprietary gazetteers of places and regions, with containment information. Each entry also contains information about alternative names and "disambiguation information", such as a city's population. Areas with higher population or economic activity are a priori considered more likely to be referred to in a document. The generic gazetteer contains nearly 10 million entries, including cities and towns, physical locations such as mountains or bodies of water, and counties and other regions. A specific gazetteer for the energy industry contains thousands of additional references for such locations as wells and oil or gas fields. The "geographic reference engine" assigns a confidence score to each entity, which is a measure of the probability that the entity is correctly labeled, based on a combination of heuristics and statistical data gathered on a corpus.

MetaCarta is a service that produces the references using a proprietary hardware processing device. The underlying gazetteer is not available for other uses.

4.4 Digital Elevation Model

Digital elevation models are matrices which specify the elevation of a land's surface at regularly spaced intervals. The primary attributes of such datasets are the area covered, the resolution (spacing) of the grid, and the accuracy of the elevation measurements.

The USGS makes several digital elevation models public. These datasets vary in resolution. Inside the conterminous United States, the National Elevation Dataset (NED)[2] has a resolution of 1/3 arc second (about 10m) and work on generating a 1/9 arc second (about 3m) resolution dataset is proceeding. Outside the United States the resolution is limited to 3 arc seconds (about 90m), at least for unclassified data. These data are available for download through the `http://seamless.usgs.gov/` Web site. (See Section 4.6.

In 2001 the Space Shuttle Radar Topography Mission (SRTM)[3] measured elevations with accuracy to 16m over most of the earth (between 60degN and 56degS latitude with 1 arc second (30m) resolution using rader interferometry. This was a joint mission between the National Geospatial-Intelligence Agency (NGA) and the National Aeronautics and Space Administration (NASA).

The GTOPO30[4] dataset at 30 arcsecond resolution (about 1km) with worldwide cov-

[2] `http://edc.usgs.gov/products/elevation/ned.html`

[3] See `http://srtm.usgs.gov/mission.html`

[4] `http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html`

erage of land areas is also available. This dataset is a compilation from several primary sources. The range of elevation is -407 to +8,732m. Ocean areas have no data. Small islands of less than 1 square kilometer are not represented. Accuracy varies depending on the part of the world and is source dependent, ranging from about 15m to 500m.

4.5 TIGER/Line Files

The TIGER/Line files are published by the U.S. Census Bureau.[5] They are a digital database of geographic features, such as roads, railroads, rivers, lakes, legal boundaries, census statistical boundaries, and so on, covering the entire United States. These files do not include demographic statistics. The database contains information about these features such as their location in latitude and longitude, the name, the type of feature, address ranges for most streets, the geographic relationship to other features, and other related information. They are the public product created from the Census Bureau's TIGER® database. The most recent version is the 2005 Second Edition released in June 2006.

The term TIGER stands for Topologically Integrated Geographic Encoding and Referencing. The system was developed at the U.S. Census Bureau to support its mapping needs for the Decennial Census and other programs. The goal was to provide automated access to and retrieval of relevant geographic information about the United States and its territories. These files are topologically structured (not graphic images of maps) and are intended to be used with a geographic information system (GIS). They are not in any proprietary GIS format, but some GIS vendors have translated some of them into their own format (for example, there is an ArcGIS version of the 2000 Census available from ESRI[6]).

4.6 The Seamless Data Distribution System

The Seamless system is an integrated browsing system for maps, digital elevation data, and other geo-referenced data products. The seamless system nominally covers the entire world, but most of the data products are limited to coverage of the United States. Using a Web browser, one can select an area and select from a range of geo-referenced data products covering that area. It is a tool for selecting combinations of geo-referenced datasets and making them available for download.

The Seamless Web site makes use of geographic coverage metadata to provide index information showing areas of coverage of different data products. The dataset overlays are shown in a preview window in the application. When the desired geographic region

[5] http://www.census.gov/geo/www/tiger

[6] http://www.esri.com/data/download/census2000_tigerline/

and selection of datasets is determined, the appropriate subset of the underlying data is extracted and made available for download.

The data is freely available for download from the http://seamless.usgs.gov/ Web site.

4.7 Google Earth

Google Earth is not so much a dataset as a way of presenting information from a wide range of datasets. This system is primarily a collection of satellite and aerial images organized into a tiling structure that enables real-time zooming in and out. This is augmented with a wide variety of information from various sources. The entities in this data are identified with points and lines in the images with moderate accuracy. (In fact, the lack of accuracy in many geospatial resources can be quite striking when one views, for example, the image of a tightly meandering river with the long segments of a polygonal line representing the river overlain several miles away.)

Google offers a range of products, from the free Personal Google Earth, up through the Professional and Enterprise versions. It is an enormously successful system; there have been around 100 million installs in the first year it has been available. These products enable users to build their own Web sites and services with a geospatial component. Several articles in the February 16, 2006, issue of *Nature* describe the system and typical uses, including its use in relief efforts after the hurricane Katrina and after the Pakistani earthquake.

On the other hand, Google Earth does not have any GIS analytic functions. For example, one cannot calculate buffer areas around roads, or overlay multiple layers of geospatial data and calculate composite values.

The developers of Google Earth say that their product is not so much a way of organizing geographical data, but a way of organizing data geographically. This is strikingly illustrated by a system developed by Anand and Swanson at IBM. They found that it was very difficult to navigate in large semantic nets in spring graph representations or other graph layout formats, because one quickly loses track of where one is relative to where one has just been. There is no familiar background or landmark to orient one's self with respect to. So they clustered the nodes in the semantic net and mapped the result onto Google Earth, with, for example, customer needs being in Africa and available products being in South America. Users very quickly begin to refer to the location of nodes in geographical terms, and it became easier for them to find their way around.

Chapter 5

Geographic Information Systems

5.1 ArcGIS

Geographic Information Systems (GIS) connect data with spatial location. GIS were in invented in Ottawa, Canada, in the 1960s by Dr. Roger Tomlinson for the purpose of conducting the Canada Land Inventory initiative (Foresman, 1998).

There are a variety of commercial and open-source GIS systems available today, including ArcGIS[1], MapInfo Spatial Ware[2], Intergraph's GeoMedia[3], the GRASS Open-Source GIS[4], Atlas GIS[5], and many others. ArcGIS is by far the most widely used, and we will use it as the representative for describing commercial-grade GIS technology available today.

ArcGIS is an integrated collection of GIS software products for building a complete GIS or GIS-based application. ArcGIS (latest version 9.2 at the time of this writing) is being developed and marketed by ESRI (Environmental Systems Research Institute) which is the largest vendor in the GIS software industry today. It deals with geo-referenced data.

ArcGIS is not a single application but a very large (and somewhat difficult to navigate) suite of integrated tools for a variety of purposes and delivery modes. For example, ArcGIS Desktop is a suite of GIS products for developing and delivering GIS on desktop computers. This suite contains tools such as ArcReader, ArcView, ArcEditor and ArcInfo that each expose progressively more GIS capabilities ending with the full-function ArcInfo tool. For server-based applications, tools such as ArcGIS Server, ArcIMS (Internet Map Server) and ArcGIS Image Server are available. ArcGIS Server is a central application server that is used to build server-side GIS applications that run in enterprise and Web computing frameworks. ArcIMS is a scalable Internet Map Server for publishing maps, data, and metadata

[1] http://www.esri.com/
[2] http://www.mapinfo.com/
[3] http://www.intergraph.com/
[4] http://grass.baylor.edu/
[5] http://www.rpmconsulting.com/atlas/

over the Web using standard Internet protocols. ArcGIS Image Server is a management, processing, and distribution platform for geospatial imagery. A variety of tools for mobile and Web-based services, such as ArcPad and ArcWeb, are available as well.

All of these tools are built onto a central Geodatabase technology which provides a mixture of file-based and RDBMS-based storage to store and manage GIS data. ArcSDE (spatial data access engine) is an integral part of this concept and provides a uniform interface to various backend databases such as IBM's DB2 Universal Database and Informix Dynamic Server, Oracle, and Microsoft SQL Server.

ArcGIS tools also ship with a large initial collection of maps covering areas such as the United States, Canada, Mexico, Europe and the world as a whole. Many different kinds of information are available such as states, counties, cities, populated places, ZIP code points, population, lakes, rivers, cultural points, hospitals, as churches, with the most comprehensive coverage available for the United States.

5.1.1 Capture of Geographic Knowledge in a GIS

In a GIS such as ArcGIS geographic knowledge is captured in a variety of ways:

Geographic Datasets are file bases and databases of geographic information such as features, networks, topologies, terrains, surveys and attributes.

Data Models define schema, behavior and integrity rules for geographic data.

Maps and Globes provide interactive views into geographic data to answer questions and present results.

Metadata allows users to describe, organize, discover and access geographic knowledge.

Processing and Workflow Models are collections of geoprocessing procedures that provide tools for analysis and task automation.

5.1.2 Representational Primitives

The central primitives represented and described by a GIS are spatial objects called *features*. The International Standards Organization (ISO) defines a feature as an "abstraction of a real world phenomenon", and a feature attribute as a "characteristic of a feature".

The *spatial attribute* of a geographic feature is defined by its dimension and location, and is referred to as the feature's *geometry*. The *dimension* of a feature geometry describes its form in space using a geometric model centered around operations on various types of geometric shapes. Shapes represent geometric objects in two-dimensional planar space and can be 0-dimensional points, 1-dimensional lines (simple or complex), and 2-dimensional areas. Each shape has *spatial reference system* which describes the coordinate space in which the shape is defined.

The spatial reference system of a feature describes the *coordinate system* in which a shape is defined which is used to define its location. The coordinate system contains information about the number of ordinate values, the mathematical rules for projecting the geometry coordinates, and how the coordinate system is related to a datum on the earth's surface. Spatial data is often defined with different coordinate systems, and the associated spatial reference systems allow transformation between shapes in different coordinate systems as well as verifying the integrity of geometric calculations.

5.1.3 Feature Representation

In ArcGIS features and feature attributes are represented via two basic mechanisms:

Files using a variety of different formats and data formats such as *shape files* to describe vector data (features and their geometry) or raster data for representing imagery and digital elevation models.

Relational Database Tables to store very large datasets and allow sophisticated spatial querying. In ArcGIS each feature is represented by a unique ID and has a special *shape* attribute (or column) associated with it (maintained by ArcSDE) that points to a feature's geometry. Any number of other columns can be used to describe other attributes of a feature.

5.1.4 Spatial Relationships

An important part of any GIS database is the representation of spatial relationships such as *topologies* and *networks*. Topology is used to represent and manage common (or shared) boundaries between features, it defines and enforces integrity rules (e.g., “lines are not allowed to cross”), and it supports topological navigation and queries (e.g., whether two features are adjacent or connected). Topology is also used to create structured features from unstructured geometry (e.g., polygons are connected sets of lines).

Networks describe a connected path of GIS objects that can be traversed which is used for modeling pathways, navigation, hydrology, and so on.

5.1.5 Thematic Layers

GIS organize geographic data into a series of thematic layers and tables such as road networks, elevation, political boundaries, and land use. These layers are geo-referenced, building a connection between them and allowing sophisticated queries covering different aspects of geospatial data.

5.1.6 Computation and Reasoning

The many components of ArcGIS support a large number of spatial operators, computations and analysis. For example, a user might be interested in the straight-line or cost-weighted distance between two points, the area covered by a set of polygons, or more complex questions, for example, where water will flow or whether there are any hospitals within a five-mile radius from a toxic waste site. ArcGIS Spatial Analyst and ArcGIS 3D Analyst are two toolsets in the ArcGIS suite providing a large number of 2D and 3D analysis functions for GIS data to answer such questions. Spatial Analyst allows advanced spatial analysis for raster and vector data, such as density histograms, statistics, buffer zones, terrain analysis, spatial relationships, and suitability modeling. For example, using the map algebra analysis language of ArcGIS Spatial Analyst a user can take the mean between multiple rasters of the same area to assess changes in a region. 3D Analyst allows a user to create 3D views directly from GIS data, use cut and fill, line-of-sight computations, and terrain modeling, and visualize modeling or analysis results in three-dimensions.

The ArcSDE spatial query engine provides access to spatial SQL as defined in the ISO 13249-3 standard (ISO-13249-3, 2003), which defines a large number of spatial operators on geometric objects. ArcSDE does not implement its own version of these operators but instead relies on their implementation by the host database system such as Oracle. For example, binary operators such as `equals`, `contains`, `intersects`, `disjoint`, `overlaps`, `touches`, `crosses`, `within` take two geometric objects (line string, polygon, etc.) as input and compute whether the binary relationship holds or not. The operator `relate` can be used to find out the geometric relationship between two objects based on the Dimensionally Extended 9 Intersection Matrix (DE-9IM) (Clementini & Felice, 1996). Spatial analysis operators such as `union`, `intersection`, `buffer`, `convex hull`, etc. can be used to take two geometric objects as input and build a new object from them. Given these operators, one can formulate complex spatial queries such as “find all states on the banks of the Mississippi”, or, in spatial SQL:

```
SELECT States.Name FROM States
WHERE EXISTS
      (SELECT * FROM Rivers
                WHERE Rivers.Name = "Mississippi"
                AND Touches(States.ID, Rivers.ID));
```

The algorithms that support these various operators usually involve a combination of combinatorial and numerical computation and are susceptible to issues of robustness. For example, due to a round-off error an algorithm might fail to determine that two polygons touch or result in dimensional collapse, where a resulting object has a lower dimension than the correct result. Robust algorithms exist for many of these operators but are generally more expensive to compute.

ESRI is developing a new Shape Comparison Language (SCL) based on the Calculus-Based Method (CBM) (Clementini & Felice, 1995). CBM defines shape relationships by the intersection of their boundary, interior, and exterior, taking the dimensionality of the intersection into account. It has five basic relationships: `touch`, `overlap`, `in`, `cross` and `disjoint`, and for any two given shapes only one of these five relationships is true. The relationship `identical` is true when two shapes are geometrically identical and share correct topological integrity. SCL can be used to describe the relationship between two specific or, for queries, hypothetical shapes. For example, if we want to find all road segments that connect to one of the end points of a given segment `G1` we could use one of the following SCL queries:

```
G1 touch G2
G1.boundary intersect G2.boundary
```

5.1.7 Summary with Respect to Analysis Dimensions

Below is a discussion of ArcGIS and GIS in general with respect to a set of analysis dimensions. The main conclusion is that today's GIS focus primarily on *quantitative* spatial representation and reasoning and provide little or no support to perform qualitative reasoning with qualitative spatial representations.

Representation Formalism: In ArcGIS spatial data and associated attributes can be represented in many different formats such as relational database tables, shape files, raster files and many other file formats.

Topology, Connectivity: ArcGIS allows the explicit representation of geometric objects such as points, lines, and multi-lines. Topology such as adjacency or connectivity between objects is expressed by sharing (part of) the boundary of an object or by representing the polygons adjacent to an arc. Topology is primarily used to enforce topological constraints and allow the computation of certain spatial predicates. There is no explicit symbolic representation for connectivity, even though a language such as SCL could be used to represent it. Binary spatial predicates such as `touches`, `intersects`, and `overlaps`, can be computed via operators in spatial SQL, but these computations are primarily quantitative based on computational geometry and do not employ symbolic reasoning methods.

Shape: ArcGIS supports the representation of arbitrary 2D and 3D shapes based on a basic set of shape types such as `Point`, `PolyLine`, `Polygon` and `MultiPoint` that come in 2D and 3D varieties as well as an "M" variety that allows the association of some arbitrary measure with a shape. A special `MultiPatch` type allows the representation of complex 3D shapes such as buildings and hillsides, (ESRI, 1998). There is no symbolic shape vocabulary available for common shapes such as circle, ellipse, rectangle, and cube, but such shapes can be represented or at least approximated with the available shape type primitives.

Size: The dimensions and size of objects can be represented and ArcGIS also supports a variety of associated computations such as distance, area, and volume.

Metrics: ArcGIS supports a variety of different unit systems to specify metric information such as metric units, Imperial units, and U.S. survey units. These units can be used as the default units of a particular coordinate system. There are also various conversion utilities to convert between different units and unit systems.

Location, Coordinates: The specification of location information, coordinates and coordinate systems is at the core of any GIS. ArcGIS provides extensive support for specifying location information, many different coordinate systems, and utilities for operations such as projections and coordinate system transformations.

Direction, Orientation: There is no built-in support for viewer-based or object-based representation or querying of direction; however, direction can be explicitly represented via specialized tables as well as computed by more complex spatial SQL expressions. For example, to test whether a building is to the right of another relative to a particular spatial reference system, one could test the relationship of the minimum and maximum x-coordinates of their respective bounding rectangles. Alternatively, an approach as the one outlined in (Shekhar *et al.* , 1999) could be used to implement direction computation on top of SQL.

Mereology, Part-of: There is no direct support or representation for part-of relationships; however, certain aspects of part-of reasoning can be achieved quantitatively via the `contains` operators, for example, to test whether a building is part of the City of Los Angeles.

Path, Trajectory: ArcGIS supports the representation of complex *geometric networks* where geographic features with their associated geometry participate as junctions and edges (ESRI, 2004a). Each geometric network has an associated *logical network* that represents the actual connectivity information. A large number of solvers are available to compute paths (e.g., shortest, cheapest), trace paths until a certain condition is met, or compute related concepts such as network flow. User-defined solvers can be plugged in to perform arbitrary network analysis tasks.

Features: There is no built-in vocabulary for geographical features such as lakes or rivers, or geopolitical features such as cities or counties. Such features are primarily represented by their spatial attributes, and a user would associate the appropriate feature class as one of the attributes associated with the feature. However, ArcGIS comes with a large number of maps and coverages that do define instances of such features (e.g., all states in the U.S.), and a user can make use of those features as well as the feature classes used in these maps. Moreover, some of these features are defined in various data models for different application domains, developed by ESRI and the ArcGIS community.

Time, Change: Temporal GIS that integrate spatial and temporal data are an emerging capability in the GIS world. ArcGIS Tracking Analyst is one such tool that allows a user to define temporal events that have information about the date and time of the event, the location of the event, as well as any other attributes associated with the event (ESRI, 2004b; ESRI, 2004c). A variety of features are provided for visualizing time series, tracking change of attribute values, analyzing patterns, creating animations, offsetting data into the past or future, and so on. Temporal representation is quantitative using time stamps. Qualitative temporal information such as "event e1 occurred before event e2" without supplying explicit time values is not supported.

Uncertainty: There is no direct support for representing or computing uncertainty. However, uncertainty annotations are possible via specialized tables or user-defined metadata attributes.

Reasoning: ArcGIS supports 2D computations based on spatial SQL provided by the particular database backend used, as well as a large number of specialized analysis procedures in its many tools and plug-ins (some of them supporting 3D computations as provided by 3D Analyst). As mentioned above, reasoning is primarily quantitative based on the quantitative spatial attributes of features. Some non-quantitative reasoning such as computing paths in networks is supported as well.

General Characteristics: Representation of and reasoning with spatial information is primarily quantitative, that is, based on the specific coordinates and extent described in the spatial attribute of a feature. Topology such as the representation of what polygons are adjacent to an arc can be viewed as a qualitative representation, but such representation exists primarily to enforce topological integrity and support certain types of spatial queries. It is not intended to support the representation of features whose exact geospatial location and extent is unknown.

ArcGIS does not have a formal axiomatization of its spatial representation and associated computations, but it is based on a variety of ISO standards that provide careful descriptions of the spatial data model, topology, and spatial computations.

ArcGIS is a commercial product that is very widely used. For example, over 90% of GIS users in the U.S. Government use ArcGIS.

5.2 Oracle Spatial

Oracle Spatial and Oracle Locator are spatial extensions to the Oracle 10g database.[6] These allow geographic and location data to be managed as a native type in an Oracle database which supports applications such as GIS, location services and location-enabled e-business. For example, the Oracle 10g Spatial database is one of the spatial database backends that can be employed with ArcGIS to implement its geodatabase.

Oracle Locator (Oracle, 2005a) provides core location functionality needed by applications that support location-based services. It builds the core of the Oracle spatial engine and allows representation of points, lines and complex polygons with holes as well as a variety of associated operations (e.g., `contains`, `interacts`). Spatial objects are stored as a native datatype in the Oracle database, spatially indexed via R-trees and accessed via SQL statements. Oracle Locator uses a whole-Earth geometry model that takes into account the curvature of the Earth's surface when performing calculations on geodetic data. Different geoids/datums can be used to model the shape of the Earth. It also supports the most commonly used distance and area units.

Oracle Spatial (Oracle, 2005b) extends the core functionality provided by Oracle Locator with a variety of more advanced functions such as buffer generation, spatial aggregates, area and length calculations and linear referencing (i.e., associating measurement information with linear geometry). It supports a GeoRaster datatype for image and raster data, network and topology data models, geocoding and routing routines and spatial analysis and data mining functions. Oracle is a Principal Member of the Open Geospatial Consortium (OGC), and Oracle Locator complies with the OpenGIS Simple Features Specification for SQL.

Oracle Spatial is primarily a database backend technology that enables the development of GIS applications and GIS tools such as ESRI's ArcGIS (ESRI is one of Oracle's technology partners). Its strength is the scalability and enterprise readiness derived from its database platform, but it does not have the wide variety of features and user tools for map development and spatial analysis as provided, for example, by a more specialized system such as ArcGIS. However, some basic functionality of this kind is available. For example, Oracle's MapViewer can be used to display and render map information stored in an Oracle database. One interesting feature of Oracle (and Oracle Spatial) is its emerging support for managing RDF (Resource Description Framework) data, which is an important step towards storage of more sophisticated metadata to describe attributes of spatial data (Oracle, 2005c).

[6] http://www.oracle.com/technology/products/spatial

Chapter 6

Ontology Standards

6.1 OpenGIS Feature Geometry (ISO 19107 Spatial Schema)

The Open Geospatial Consortium (OGC)[1] has the mission of leading the development of standards for geospatial services. One of the key standards is the OpenGIS® Feature Geometry standard (Herring, 2001,) that is being developed in collaboration with the International Organization for Standardization (ISO) Technical Committee ISO/TC 211[2]. ISO/TC 211 develops standards for information associated with geographic locations.

This specification is of interest, since it specifies the content and semantics of geometric objects and ways of linking them to an underlying geographic coordinate system. The support of ISO and implementation in GIS systems are also key factors in making this an important standard to examine.

The OpenGIS Feature Geometry (also ISO 19107 Spatial Schema) is a specification with formal definitions of objects and data structures that use the Unified Modeling Language (UML)[3] (Object Management Group, 2005). In addition to the structural descriptions secified in UML diagrams, the standard also describes the semantics of geometric primitives by describing them in English text, but with reference to rigorous definitions that are grounded in mathematical set theory.

In addition to objects and data structures, the semantics and effects of derivative computations are also specified by the standard. This includes notions such as "length" and "distance", when the geometric objects are referenced in a suitable coordinate system.

The major areas that are covered by the standard are

- Geometry
  - Geometric primitives such as points, curves, solids, boundaries, and bearings.

[1] http://www.opengeospatial.org/
[2] http://www.isotc211.org/
[3] http://www.uml.org/ and also ISO/IEC 19501

  - Coordinate geometry to hold information for references to a coordinate system defined in accordance with ISO 19111. This allows geo-referencing of the geometric objects.
  - Geometric aggregates, which are arbitrary aggregations of geometric objects without any additional internal structure.
  - Geometric complexes are sets of primitive geometric objects with disjoint interiors.

- Topology
  - Topological primitive objects for defining topological collections
  - Derived topological relations using boolean set operations.

The stated purpose of the topological component is that often topological representations can accelerate computational geometry; hence, there are precise mappings between geometric and topological concepts. Another use of topology is to relate features when the geometry is unknown.

A complete listing of the object types can be found in Table 6.1.

The standard has support for derived topological relations that utilize Egenhofer's point-set intersection formalism as well as Clementini's topological relations.

6.2 ISO Standards

The International Organization for Standardization (ISO)[4] is an international organization that manages and publishes international standards in many fields. They have published more than 16,000 standards in areas such as agriculture, construction, engineering, medical devices and information technology. Agreed-upon standards, whether formal or informal, are essential for interoperation of engineered systems. ISO provides formal standards with international scope.

For the purposes of this report, we are most interested in the ISO standards that relate to geospatial information. The standards that are available cover a wide range of levels. At the simplest level, standards such as ISO 6709 merely specify the data format (syntax) for representing data items, in this case latitude, longitude and altitude. Such agreement is necessary for the interchange of data between computer systems, but there is very little ontological content. In the case of ISO 6709, there is a notion of latitude and longitude, but no specification in the standard itself of the meaning of those terms. The semantics are external to the standard.

[4]Website of the ISO: `http://www.iso.org/`. The abberviation "ISO" derives from the Greek word *isos*, meaning "equal" and is used to avoid having different abberviations in different languages.

Table 6.1: The OpenGIS Feature Geometry (ISO 19107 Spatial Schema) Objects

Geometric Objects		**Topologic Objects**
GM_Object	GM_Clothoid	TP_Object
GM_Boundary	GM_OffsetCurve	TP_Boundary
GM_ComplexBoundary	GM_Knot	TP_ComplexBoundary
GM_PrimitiveBoundary	GM_KnotType	TP_PrimitiveBoundary
GM_CurveBoundary	GM_SplineCurve	TP_EdgeBoundary
GM_Ring	GM_PolynomialSpline	TP_FaceBoundary
GM_SurfaceBoundary	GM_CubicSpline	TP_SolidBoundary
GM_Shell	GM_SplineCurveForm	TP_Ring
GM_SolidBoundary	GM_BSplineCurve	TP_Shell
GM_Primitive	GM_Bezier	TP_Primitive
GM_Point	GM_SurfaceInterpolation	TP_DirectedTopo
Bearing	GM_GenericSurface	TP_Node
GM_OrientablePrimitive	GM_SurfacePatch	TP_DirectedNode
GM_OrientableCurve	GM_PolyhedralSurface	TP_Edge
GM_OrientableSurface	GM_Polygon	TP_DirectedEdge
GM_Curve	GM_TriangulatedSurface	TP_Face
GM_Surface	GM_Triangle	TP_DirectedFace
GM_Solid	GM_Tin	TP_Solid
DirectPosition	GM_ParametricCurveSurface	TP_DirectedSolid
GM_PointRef	GM_GriddedSurface	TP_Expression
GM_Envelope	GM_Cone	TP_Complex
TransfiniteSet<DirectPosition>	GM_Cylinder	
GM_Position	GM_Sphere	
GM_PointArray, GMPointGrid	GM_BilinearGrid	
GM_GenericCurve	GM_BicubicGrid	
GM_CurveInterpolation	GM_BSplineSurfaceForm	
GM_CurveSegment	GM_BSplineSurface	
GM_LineString	GM_Aggregate	
GM_LineSegment	GM_MultiPrimitive	
GM_GeodesicString	GM_MultiPoint	
GM_Geodesic	GM_MultiCurve	
GM_ArcString	GM_MultiSurface	
GM_Arc	GM_MultiSolid	
GM_Circle	GM_Complex	
GM_ArcStringByBulge	GM_Composite	
GM_ArcByBulge	GM_CompositePoint	
GM_Conic	GM_CompositeCurve	
GM_Placement	GM_CompositeSurface	
GM_AffinePlacement	GM_CompositeSolid	

In contrast, there are more extensive standards such as ISO 19107 (Spatial Schema) and ISO 19108 (Temporal Schema) which incorporate a substantial description of the semantics of the items being standardized. ISO 19107 was discussed in Section 6.1, “OpenGIS Feature Geometry”. The most important geospatial standards from an ontological and semantic point of view are the following. A full list of ISO geospatial standards is presented in Appendix B.

ISO 13249–3 This standard describes spatial extensions to the SQL database query language. It is discussed more extensively in this report in Section 5.1.6.

ISO 19107 Defines geometric and topological primitives, providing the semantics for them and certain operations. Discussed in detail in Section 6.1.

ISO 19108 Defines a set of temporal entities for linking information to dates, specifying durations, and representing qualitative relations such as before, after, and during. Although not strictly speaking a geospatial standard, it does define the terminology needed to express movement and change, which have temporal component.

ISO 19125 This specifies an SQL schema supporting database management of simple geospatial feature collections. It defines terms for geographic information in connection with ISO 19107. It defines names and definitions for SQL geometry types and functions for geometry.

6.3 Federal Geographic Data Committee

The Federal Geographic Data Committee has developed a “Content Standard for Digital Geospatial Metadata” (Federal Geographic Data Committee (FGDC), revised June 1998). This standard is widely used by United States Government agencies for the description of geospatial data products. The metadata is organized into the following broad sections:

Identification Information

Data Quality Information A general assessment of the quality of the dataset. Recommends using quality information from the Spatial Data Transfer Standard (see below).

Spatial Data Organization Information Describes the mechanism used to represent spatial information. This can be *direct* or *indirect*. Vector as well as raster data types are supported.

Spatial Reference Information Defines the frame of reference for coordinates used in the dataset. This can be linked to geographic references, but need not be. Resolution and units can be specified.

Entity and Attribute Information Details about the information content of the dataset. Includes specification of the domain from which they are drawn.

Distribution Information How the dataset can be obtained.

Metadata Reference Information Information about the currentness of the metadata and the responsible parties.

For the most part, the content of the metadata is directed mostly at human readers rather than automatic processing. Many of the metadata fields are specified as having text values, or reference to some external thesaurus to define the meanings of terms.

Some fields, however, are more structured. In particular, the metadata section for defining the geographic regions encompassed by the underlying dataset is quite rich. In addition to points and bounding boxes, it is possible to specify arbitrary polygonal regions. This metadata is used to display index frames in the Seamless Data Distribution System described in Section 4.6.

6.4 Spatial Data Transfer Standard

The Spatial Data Transfer Standard is essentially an ontology for geographic and cartographic repesentation, developed by the United States Geological Survey (2005) for transfer of earth-referenced spatial data between computer systems with no information loss. The standard has been adopted by a number of government agencies and private companies. It has been developed over the last 25 years in collaboration with academic, industrial, and government users of geospatial data.

One important piece of the standard is a set of basic spatial objects. There are zero-dimensional points, including points for labels, points representative of areas, and nodes defined by topological intersections. In one dimension, there are line segments, strings of connected line segments, arcs defined by mathematical expressions, topologically defined links, and several kinds of rings. In two dimensions, there are polygons defined by various sorts of line segments, pixels, grid cells, digital images, grids, layers, rasters, and graphs. The basic elements of three dimensions are voxels and voxel spaces. The meanings of these concepts are not elaborated in any kind of axiomatization, but they are described precisely in English and the mathematics behind the descriptions is for the most part well-understood.

In addition there is a large set of spatial features, including both human and natural landmarks. A random selection will give an idea of the range of this set: airport, beach, cave, demilitarized zone, earth surface, fault, grave, headwaters, iceberg, lake, missile site, offshore platform, place, quicksand, racetrack, shoreline, time zone, utility, valley, windbreak, and zone of occupation. There is also a large set of attributes or properties that can be attached to these entities. The terms in both of these sets are described by English phrases.

Chapter 7

Large-Scale Research Ontology Efforts

7.1 ResearchCyc

Cyc is a very large multi-contextual knowledge base and associated reasoning engine aimed at formalizing human commonsense knowledge and reasoning (Lenat & Guha, 1990; Guha & Lenat, 1990; Lenat, 1995; Panton *et al.* , 2006). The name derives from the original intention of encoding the knowledge contained in the definitions of a single-volume desktop encyclopedia as well as the knowledge necessary to understand such definitions. Cyc is a commercial product being developed by Cycorp, Inc. based in Austin, TX[1].

In this section we give a brief overview and evaluation of the parts of Cyc dealing with spatial representation and reasoning. This evaluation is based in part on documents available on the Cycorp Web pages[2], in particular, (Cycorp, 2002c; Cycorp, 2002a; Cycorp, 2002b). These documents describe vocabulary available in OpenCyc[3] which is the open-source version of Cyc. Unfortunately, these documents are somewhat out of date given that the latest version of OpenCyc 1.0 was released in July, 2006. They also do not describe any of the associated axioms. To compensate for that we also consulted a version of ResearchCyc 1.0 which includes all the non-proprietary parts of the Cyc knowledge base. ResearchCyc contains almost 3,000,000 assertions and is based on an ontology containing over 300,000 concepts (or collections/classes) and over 26,000 relations (or predicates).[4]

Given the very large size of ResearchCyc, a thorough analysis of even just the spatial subset of its ontology is beyond the scope of this report. For example, the terms categorized within Cyc's `Spatial-Topic` comprise about 70 collections and 130 predicates, there are about 90 predicates and 60 collections having to do with paths, and about 40 predicates and 75 collections are in the area of geography. These topic collections are

[1] http://www.cyc.com
[2] http://www.cyc.com/cycdoc/vocab/vocab-toc.html
[3] http://www.opencyc.org/
[4] http://research.cyc.com/

not complete and only describe the most important terms in a topic area (for example, `tangentialProperPartOfSpaceRegion` is not classified under any topic); therefore, the complete list of terms in these areas is actually significantly larger.

Given Cyc's mission to formale commonsense knowledge, it primarily focuses on representation and reasoning with qualitative spatial descriptions, forming an interesting counterpoint to the quantitative nature of spatial descriptions used in GIS. Cyc focuses primarily on objects and their location and interrelation in space, and not so much on ontologizing pure regions of space. At the root of Cyc's spatial collection hierarchy is `SpatialThing`, which describes objects, events, and regions located in space. The immediate descendants of `SpatialThing` are listed in Figure 7.1.

```
SpatialThing
   AmorphousThing
   AxisymmetricObject
   GeometricallyDescribableThing
   LocallyEuclideanSpatialThing
   ZeroDimensionalThing
   PositiveDimensionalThing
   FiniteSpatialThing
   SpatialThing-Localized
   SpatialThing-NonSituational
   SpatiallyContinuousThing
   SpheroidPatch
   SpheroidalSurface
   (CollectionDifferenceFn SpatialThing Situation)
   (CollectionUnionFn (TheSet AxisymmetricObject
                              BilateralObject))
   OptionsField-Packet
   PresentationProtocolDataUnit
   TransportProtocolDataUnit
```

Figure 7.1: Direct Descendants of `SpatialThing`

There are a few peculiarities in this list. For example, `AmorphousThing` is a direct descendant as opposed to being a descendant of `PositiveDimensionalThing` where its disjoint counterpart `ShapedObject` is located. It is unclear whether this is an oversight or intentional to also allow 0-dimensional points to be classified as amorphous. Unfortunately, there are no axioms associated with this class that would make this more precise. There are also a few classes describing aspects of network data packets that are viewed as having location in space. Since these are part of the empirically observable universe, it seems they should be specializations of `SpatialThing-Localized` instead. Note, that in Cyc spatial things might be tangible or partially tangible, such as a book, or completely intangible, such as the Earth's equator.

`SpaceRegion` is not an immediate descendant of `SpatialThing` but a specialization of `SpatialThingNonSituational` which can be seen in Figure 7.2. That figure shows the descendants of `SpatialThing` two levels down (non-atomic collections formed by collection-forming functions have been excluded to save space). `SpaceRegion`'s are intangible, immobile objects that purely serve as possible containers for other spatial objects and do not have much standing in and of themselves. Cyc's spatial ontology is primarily relational in the sense that it is concerned with the positional relationships of objects in space, but not space itself. This leads to some non-trivial ontological engineering issues when trying to fit a theory such as the Region Connection Calculus (RCC) (Randell *et al.* , 1992) into Cyc, since RCC takes a substantivalist point of view, that is, regions of space are substances that exist independent of the objects that occupy them (Grenon, 2003).

Representation Formalism: Cyc is a logic-based system that uses CycL as its representation language (Matuszek *et al.* , 2006). CycL is the language of a higher-order predicate logic with a Lisp-based syntax somewhat similar to KIF (Genesereth, 1991). For example, the following sentence says that the Eiffel Tower is in Paris:

```
(objectFoundInLocation EiffelTower CityOfParisFrance)
```

The following rule states that if a thing is near some location and on some other object, then the other object is also near the location (variables are indicated via '?' and are implicitly universally quantified):

```
(implies (and (near ?LOC ?THING)
              (on-Physical ?OBJ ?THING))
         (near ?LOC ?OBJ))
```

Topology, Connectivity: Cyc has a large number of topological and spatial relations. The `touches` relation is a specialization of `near` and specifies that two objects touch either directly or indirectly. They can touch indirectly if separated by a thin layer or substance and the distance between their surfaces is much less than the distance between their centerpoints. For example, a person's foot touches the ground even if they wear shoes, but two sheets of paper separated by another sheet do not touch. The relation `touches-Directly` should be used if objects are in direct contact. Note that this subtlety (like many others) is only expressed in the documentation string but not with formal axioms. There are various covering relations, such as `covers-Hairlike`, `covers-Skinlike`, and `covers-Ruglike`, that express specialized touch relationships. Cyc also has `adjacentTo`, `connectedTo` (which implies physical connection with limited relative motion) and variants such as `hingedTo`

- SpatialThing
  - AmorphousThing
  - AxisymmetricObject
    - BilaterallySymmetricObject
    - RadiallySymmetricObject
  - FiniteSpatialThing
    - AstronomicalObject
    - BiologicalLivingObject
    - ContainerIndependentShapedThing
    - EllipsoidalSolid
    - InanimateObject-NonNatural
    - LinearObject-Finite
    - SurfaceRegion-Finite
  - GeometricallyDescribableThing
    - Circle
    - DoubleHelix
    - EllipticalCurve
    - GeometricFigure
    - L-Shaped
    - OneDimensionalGeometricThing
    - OpenGISGeometry
    - Place-4D
    - Point
    - SectorShapedObject
    - Semicircle
    - Semiellipse
    - Spiral
    - SurfaceOfRevolution
    - ThreeDimensionalGeometricThing
    - Tube
    - TwoDimensionalGeometricThing
  - LocallyEuclideanSpatialThing
    - Cone
    - ConicalSurface
    - CruciformObject
    - Cylinder
    - CylindricalSurface
    - CylindricalSurface-WithoutTopAndBottom
    - CylindricalSurface-WithTopAndBottom
    - Ellipse
    - Ellipsoid
    - EllipsoidalSolid
    - EllipticalCurve
    - EllipticalRegion
    - HalfPlane
    - HemisphericalSolid
    - Line-Straight
    - LinearRing
    - LineSegment-Straight
    - Parallelepiped
    - Plane
    - Polygon
    - PolyhedralSurface
    - Pyramid
    - Wedge
  - OptionsField-Packet
  - PositiveDimensionalThing
    - ExtendedSpaceRegion
    - HomogeneousPositiveDimensionalSpaceRegion
    - InfiniteSpatialThing
    - NullSpaceRegion
    - OneDimensionalThing
    - PolyDimensionalThing
    - ShapedObject
  - PresentationProtocolDataUnit
  - SpatiallyContinuousThing
    - ConnectedSpaceRegion
    - ConvexThing
    - FreeSpaceContent
    - HumanShelterConstruction
    - LinearObject
    - MultiplyConnectedThing
    - Organism-Whole
    - Pile
    - ScrewShapedObject
    - SectorShapedObject
    - ShapedObject
    - SheetOfSomeStuff
    - SimplyConnectedThing
    - SpatiallyBoundedThing
    - SpatiallySemiBoundedThing
    - SpatiallyUnboundedThing
    - SurfaceRegion
  - SpatialThing-Localized
    - AboveGround
    - Afloat-Generic
    - Afloat-OnWater
    - Airborne
    - Artifact-AnimalCreated
    - AtSea
    - BilateralObject
    - Blackish
    - Bluish
    - Brownish
    - CavityInterior-Generic
    - CavityOrContainer
    - ColorlessThing
    - CustomarySystemOfLinks
    - EnduringThing-Localized
    - EnergyStuff
    - Grayish
    - Greenish
    - Indoors-IsolatedFromOutside
    - InformationBearingThing
    - InOuterSpace
    - InPublic
    - Orangeish
    - Outdoors-ExposedToWeather
    - Pinkish
    - Point-Empirical
    - Purplish
    - Reddish
    - Shadow
    - Situation-Localized
    - SpaceRegion-Empirical
    - StrategicTarget
    - Superstring
    - Trajectory
    - Trajectory-SweptSpace
    - Underground
    - Underwater
    - VisibleThing
    - Whiteish
    - Yellowish
  - SpatialThing-NonSituational
    - EnduringThing-Localized
    - SpaceRegion
    - StrategicTarget
  - SpheroidalSurface
    - Spheroid
    - SpheroidPatch
  - SpheroidPatch
    - Hemispheroid
  - TransportProtocolDataUnit
  - ZeroDimensionalThing
    - MultiPoint
    - Point

Figure 7.2: The Cyc collection hierarchy below `SpatialThing` down two levels.

and a variety of predicates describing the relative locational properties of two objects such as `inRegion`, `spatiallyContains`, `spatiallyDisjoint`, `spatially-Intersects`, `spatiallySubsumes`, `objectFoundInLocation`, as well as many "in"-relations: `in-ContOpen`, `in-ContClosed`, `in-Held`, `in-Spiked`, etc. This highly incomplete list gives a flavor of the richness of distinctions made in Cyc, but also the complexity it means for someone trying to use this ontology.

Surfaces, Portals, Cavities: This area describes aspects of surfaces such as `FlatPhysi-calSurface`, corners, `InteriorSurface-Tangible` vs. `EntireExternal-Surface-Tangible`, `ConvexTangibleObject` vs. `ConcaveTangibleObject`, `Tube`, `Valley`, cavities and crevices, `Portal`, as well as related relationships, such as `cavityHasWall` and `hasPortalToRegion`.

Shape: An extensive vocabulary of geometric 2D and 3D shapes is available such as `Sphere` (the 2D surface), `Ellipsoid`, `CubicalSurface`, `Plane`, `Polygon`, `Rect-angle`, `Square`, 3D shapes such as `SphericalSolid`, `Cube`, `Cone` and many others. Various qualitative structural shape types such as `PointyEnded`, `LongThin-Object` (one dimension exceeds the other two by at least a factor of three – again, no axiom formalizing this), `SheetShapedObject`, and `SharpEdged` exist as well. `AxisymmetricObject`, `BilaterallySymmetricObject` and `RadiallySymmetric-Object` describe different kinds of symmetric objects.

Size: Cyc supports a variety of size-related quantities such as `Distance`, `Angular-Distance`, `Area` and `Volume`. Relations such as `distanceBetween`, `areaOf-Object` and `volumeOfObject` can be used to specify these quantities qualitatively (e.g., as high or low) as well as quantitatively. Various specialized versions, such as `landAreaOfRegion` and `altitudeAboveSeaLevel` (including direction), are available as well. Quantitative values can be fixed or or one can specify a range as an interval. Various qualitative values, such as `Thick`, `Thin`, `PlanetSized`, and `ContinentSized`, are available, as well as rules that map a qualitative value onto a quantitative range. For example, a `ContinentSized` area is defined to be between 5,000,000 and 64,000,000 square miles. `fitsInsideObject` describes whether one object can fit inside another without destroying or significantly distorting it. Quantitative computations are supported as well, but there the coverage is less complete. For example, there are rules on how to compute the area of a square and parallelogram, but not a circle. Simple rules to compute geographic distance between two points based on their lat/long coordinates on an idealized sphere are available, but these do not take a particular geoid or datum into account. There is no support to compute the area of a polygon specified by a set of vertices or more complex computations supported by GIS such as the area of intersection of two polygons.

Metrics: Cyc supports a variety of different unit systems, such as `MetricUnitOfMeasure` and `USUnitOfMeasure`, as well as a large number of different units, represented as functions, such as `Meter`, `Foot`, `SquareKilometer`, and `CubicInch`. Conversion between these units is supported fairly flexibly via rules and `QuantityConversionFn`. For example, the following idiom would convert one square mile into the equivalent in square meters:

```
(evaluate ?result
   (QuantityConversionFn SquareMeter (SquareMile 1)))
```

Location, Coordinates: Location can be specified both qualitatively, e.g., that a chair is `between` a table and a wall, as well as quantitatively by providing coordinates of an object such as its `latitude` and `longitude`. There are a large number of qualitative relative locational predicates, some of which we already described above. For example, `above-Directly`, `between`, `northOf`, `relativePosition`, and many others. Cyc also supports a number of different coordinate system types such as `CartesianCoordinateSystem` or `PolarCoordinateSystem` which are specializations of `FrameOfReference`. These have different coordinate predicates such as, for example, `easting` for the `UTMCoordinateSystem` or `longitude` for the `GeodeticCoordinateSystem`. Conversion between coordinates in different systems, e.g., from UTM to lat/long does not seem to be supported directly. Some terrestrial coordinate systems based on a specific datum (or geoid) such as the `WGS84CoordinateSystem` are also available. Note that these have their own coordinate value predicates such as

```
(AltitudeAnglePredicateFn WGS84CoordinateSystem)
```

(where `AltitudeAnglePredicateFn` is a term or name-producing function). This means the more familiar `latitude` and `longitude` relations cannot be used if coordinates need to be specified with respect to a specific datum. These coordinate systems are also mapped to the spatial reference system ID used in the OpenGIS specification.

Direction, Orientation: Cyc has a rich vocabulary for qualitative directions and orientations. It supports the eight geographic directions N, NE, E, SE, S, SW, W and NW. These directions come in a direct variety such as `North-Directly`, and a general form such as `North-Generally` which means the set of directions within 22.5 degrees on either side of `North-Directly`. This definitional constraint is defined via an axiom. There are `Up-Generally` (within 45 degrees of vertical) and `Up-Directly` (and their down varieties), `Rightward` and `Leftward` (viewer-based directions ontologized as unit vector intervals) and `directionBetweenObjects` to describe the direction between two objects via a unit vector interval or geographic direction. `perpendicular` and `parallel` state that two objects are perpendicular or parallel to each other. There are also spatially oriented parts of objects such as `BottomSide`, `TopSide`, `LeftSide`, `RightSide`, `BackSide`, `FrontSide` as well as relations such as `objectRightSide` and `objectLeftSide` which relate an intrinsically `LeftAndRightSidedObject`, such as a person, to a right or left region. The `orientation` relation can relate an object to qualitative orientations such as `VerticalOrientation`, `HorizontalOrientation`, `SlantedOrientation`, `UpsideDown` or `RightSideUp`.

Mereology, Part-of: Cyc does support part-of representation and reasoning in a variety of forms. The basic relationship is `parts` which relates an individual to (some of) its parts which can be spatial, temporal, conceptual, members of groups, and so on. Various specializations exist, such as `properParts`, `physicalParts`, `anatomicalParts`, and `subEvents`, `groupMembers`.

Path, Trajectory: Cyc's ontology of paths, path systems, paths between objects and trajectories is very extensive and probably one of the most completely axiomatized areas in Cyc. Given the size of this area we refer the reader to (Cycorp, 2002b) for more information. This domain was very significant in application used for evaluation in the DARPA-sponsored High Performance Knowledge Bases project several years ago.

Features: An extensive set of several hundred natural and topographical features are available such as `Mountain`, `Valley`, `River`, `Atoll`, `Bay`, `Canyon`, and `Cave` to more esoteric ones like `Estuary` and `Ejecta`. Regions can be qualified by human use, such as `UrbanArea` and `HumanResidenceArea`, or by natural character such as `WildernessArea`. There are various geopolitical types, such as `City`, `County`, `State-Geopolitical`, `Province`, and `Country`, and a large number of instances. For example, there are over 200 countries, over 6,000 cities, 3,500 states, and 500 provinces.

Time, Change: Cyc has an extensive ontology of times and dates and their constituent parts, durations, seasons, calendars, holidays, temporal relations such as Allen's interval relations (Allen, 1984). The truth of a Cyc sentence (or proposition) can be temporally qualified via the `holdsIn` and `holdsSometimeDuring` modal predicates. Entities with temporal extent such as `Events` can be associated with start and end times. Temporal analysis, such as how certain properties of an object changed over time or detection of temporal patterns, is not supported.

Uncertainty: Cyc is a logic-based system primarily concerned with representing things that are (default) true or false. Logic systems do not generally excel at representing and reasoning with uncertainty. However, Cyc has some vocabulary to represent probabilities and some rudimentary facilities and hooks to connect to probabilistic reasoners, such as a Bayes Net, which could be used to model and reason with uncertain information.

Reasoning: Cyc's main mode of reasoning is logical deduction based on the assertions and rules in its knowledge base. Since logical rule-based inference is computationally complex and expensive, a large number of faster, specialized (though incomplete) reasoners (HL modules) implemented directly in procedural computer code are available. Cyc's reasoning with spatial concepts and relations is primarily qualitative, but some quantitative computations, such as computing distances, are available as well.

General Characteristics: As already pointed out above, Cyc's representation and reasoning with spatial information is primarily qualitative and on the opposite end of the spectrum compared to the kind of reasoning performed in a GIS. Cyc is based on a formal logical representation language, but the language is representationally promiscuous allowing modal operators, quantification over relations, defaults, reference to truth values, and so on, and is not based on a widely accepted and well-understood logical theory like first-order predicate logic. While still a work-in-progress, Cyc is a large and fairly mature system that has been used in a large number of research projects. Until recently, Cyc was not very widely available and visible to the outside community which obstructed its analysis and general acceptance. To date, it is still primarily used by Cycorp to build research prototypes as well as some commercial products. With the latest releases of OpenCyc and ResearchCyc, however, this could potentially change given their accessibility and extensive coverage.

Besides the various concepts described above, Cyc also has some OpenGIS terminology in its `OpenGISSimpleFeaturesMt` which suggests that a linking up with GIS terminology and possibly GIS systems has been underway. Given the complementary approaches taken by Cyc and GIS, this seems to be an interesting (potential) marriage. Combinations of logic-based systems with GIS to combine qualitative and quantitative spatial reasoning have been successfully attempted in the past, see, for example, (Donlon & Forbus, 1999).

Cyc's ontology makes many careful and fine-grained distinctions, but most of these are still only specified in prose in documentation strings; we gave some examples above. There are a large number of formal axioms, but many of them are non-definitional and simply used in some application context. A significant number do no more than place constraints on the arguments of a predicate. One of the main claims to fame of Cyc is its extraordinary large size and coverage. To date, this has also been its biggest Achilles heel, since it is very difficult to navigate which makes it very difficult to use and assess. In particular, there is not much documentation of Cyc's knowledge base available apart from the documentation strings associated with terms. It is especially difficult or even impossible to access coherent clumps of axioms that together formalize a particular topic such as some aspect of spatial reasoning. Axioms for a single term are also spread over many different microtheories, with complex hierarchical structure, and are usually not documented which makes it difficult to assess what motivated them or their particular formulation and what connects them to other related axioms.

Summing up, Cyc has a good and very extensive collection of spatial concepts and relations linking the mathematical with the everyday. A large number of geographic concepts and instances is available as well, but a precise characterization of coverage is not available. The focus is primarily on qualitative representations related to human commonsense reasoning. Cyc is especially strong in areas where substantial AI research has preceded it; it is more scattered and uneven in formal specification elsewhere.

7.2 SUMO

SUMO, or the Suggested Upper Merged Ontology, (Niles & Pease, 2001; Pease *et al.* , 2002) is an ontology developed largely by Adam Pease and Ian Niles, begun when they were at Teknowledge and continued by Adam Pease at Articulate Software. At first it resulted from merging several existing ontologies, but it has been given extensive subsequent independent development. There is a mapping from WordNet synsets to SUMO concepts. The coverage of SOMO is very large. A basic ontology explicates concepts in such areas a set theory, graph theory, arithmetic, mereotopology (the "part" relation and holes), and units of measure, among others. In addition, a number of "mid-level" ontologies have been built on top of this base ontology. All parts of SUMO are downloadable for free. The Web site for SUMO is `http://www.ontologyportal.org/`.

The mereotopology domain, for example, contains an axiomatization of the "part"relation that says that it is a partial ordering. It gives a number of specializations of the "part" relation, including "component" for artifacts and other complexes, "member" for sets, and "subRegion" for regions. Then there is a very large set of specific axioms involving "part", such as that wholes inherit connectedness from parts, that if something freezes then part of it must be a liquid, that steeples are part of buildings, and that Massachusetts is part of New England.

Of the mid-level ontologies, the most relevant one is Geography. This includes a number of terms and facts from the *CIA World Fact Book*, as well as rich set of other geographical terms. It provides a vocabulary, together with a number of instances, for talking about subregion relations, such as Europe being a subregion of the Eastern Hemisphere; orientation relations, such as Europe being north of Africa (8 directions are possible); the "borders on" relation; measures of distance, area, angles, and so on, and their relation to latitude and longitude; coastlines and maritime claims; climate zones and weather phenomena; terrain types, such as "MountainousTerrain", and attributes, such as slopes; elevations; a large class of natural resources such as minerals; land use types, such as arable; natural hazards like earthquakes; environmental phenomena like air pollution; astronomical bodies (at the time of writing, Pluto was classed as a planet); land forms and bodies of water like mountains and lakes; specific facts about which countries touch which bodies of water; currents and tides; vegetation types; and a large instance set of countries and cities.

The meanings of the predicates or concepts are defined or constrained by a large set of axioms. In addition, the intended meanings are described in English prose. The axioms do not completely pin down the meanings, but that is impossible in any case with most commonsense concepts, and there is a richer set of axioms than in most ontologies that have been developed.

7.3 SWEET

The Semantic Web for Earth and Environmental Terminology, or SWEET, is a set of ontologies developed by Rob Raskin at the Jet Propulsion Lab. At its base is an ontology of numbers, or arithmetic, and this serves as the basis for ontologies of space and time and units of measure. Thus, the spatial concepts of "above" and "north of" and the temporal concept of "later than" are specializations of the arithmetic "greater than" relation. This would seem to make this set of ontologies more appropriate for scientific applications than for applications in which qualitative and quantitative concepts are mixed. On top of this basis are built ontologies for physical properties and processes, living and non-living substances, the Earth and Sun realms, natural phenomena, and human activities. The ontologies are hierarchies of terms stored in a Postgres DBMS and convertible into an OWL XML format. SWEET is being used in a number of NASA-sponsored efforts. The Web site for SWEET is `http://sweet.jpl.nasa.gov/`.

The Space ontology is particularly rich in various coordinate systems and projections. It also includes concepts of latitude, longitude, and altitude; directions like north; spatial scales; 0-, 1-, 2-, and 3-dimensional spatial objects and regions; and geopolitical regions like countries. The Earth realm ontology is very extensive. There are polygons for representing spatial regions, which can be associated with properties and events; vertical layers in the Earth realm, such as the mantle and the stratosphere; and horizontal features such as coastal regions, fjords, and mountains as regions.

The ontologies encode hierarchical or subclass relations, and some attribute relations. The terms are not described in English, although their meanings for the most part are self-evident. There is no accompanying rich axiomatization of the concepts to constrain their possible interpretations and eliminate possible ambiguities.

7.4 DOLCE

DOLCE, or "Descriptive Ontology for Linguistic and Cognitive Engineering", is a relatively small upper-level ontology that aims to capture ontological categories underlying natural language and human common sense. Objects in DOLCE are divided into "endurants", which are wholly present at any time point they are present, and "perdurants", which have proper temporal parts that may not be present at each point when the object is present, such as a play consisting of several acts (Masolo *et al.* , December 2003). The subcategory of physical endurants must have a physical location, while a perdurant's location is indirectly given by its associated endurants. DOLCE takes a multiplicative approach to representation, where different objects may be co-located in space and time, distinguished by their essential properties. For example, a mountain may have the same location as the set of rocks of which it is composed.

Spatial and temporal locations are treated the same way as other physical qualities, such as color, where the quale of a physical object's location is a region in physical space. This representation is independent of the properties of the geometric space used, allowing different approaches to be used as extensions of the ontology. Indeed, Bateman and Farrar (June 2005) note while discussing potential ontological frameworks: "For space, an excellent starting point appears to be offered by the DOLCE framework … this is precisely because it says very little about how space is to be captured."

While DOLCE contains minimal commitments on the representation of space and time, we note two efforts that extend it. Masolo and Vieu (1999a) are contributors to DOLCE who describe a formalism for space that could be incorporated into the DOLCE top-level, showing how to characterize the notion of divisibility of space. Probst (2005) describes an experiment using DOLCE to integrate six geospatial web services that request and provide wind speed data, by augmenting DOLCE with several domain ontologies, including geospatial information as well as meteorological and measurement concepts.

Chapter 8

Uncertainty

8.1 Scale and Precision

The general topic of uncertainty encompasses a number of phenomena that should be distinguished for greater clarity. The first of these is the issue of the scale and precision of the geospatial data. Obviously, the scale of a map is a limit on its utility; you can't find your way around Arlington, Virginia, if you only have a world map. Scale is not so much a form of uncertainty as a limit to the precision with which objects can be localized in a geographic dataset. For example, if the granularity or resolution of a data set only contains data with a resolution of 10 meters, then it will not always be possible to distinguish the locations of items that are closer than 10 meters to each other. A dataset that is discrete, as most imagery data, imposes a grain size on what are often continuous variables in the real world. This quantization effect causes anomalies when applied to distances that are small compared to the grain size. For example, edge effects near the discrete boundaries can occur.

Within a given scale, uncertainty is introduced by the imprecision of the data. For example, the positional accuracy of a data source may only be "within 10 meters of" with no guarantee of accuracy below that, even though the scale of the map goes below that. We can learn this by comparison of the data source with a data source of greater reliability. It is also important to know how the data items were arrived at. For example, how many significant digits were used to record measurements, and what was the round-off error in calculations done to produce secondary data. The imprecision may be caused by limitations in the technology used to localize information. For example, Global Positioning System readings are accurate only to within 5-8 meters 50% of the time.

The precision will generally not be uniform for the entire resource. For example, it is common for the accuracy to be better in urban areas than in rural areas in many types of resource. Even where the resource reports its accuracy, it is highly unusual for this variability to be reported. Resources will typically report their best accuracy, rather than the full range.

When dealing with areal cover or other two-dimensional attributes, data sources should ideally have uniform coverage. But the information is sensitive to the density of observations. An insufficient number of observations may not provide the level of resolution, related to scale, that is required. The digitizing methods that are used will introduce some level of imprecision. For areal coverage, satellite imagery is surrogate data in the sense that a satellite does not "see" vegetation, only certain types of digital signatures with associated assumptions and limitations.

One source of variation in the accuracy of geospatial data is the fact that it is derived from multiple sources or multiple persons. Different sources may have different levels of precision. Other sources for variation and imprecision are faulty observations, biased observers, or multiple observations using different equipment.

Standards for what positional accuracy to report and how to calculate it have been published by the Federal Geographic Data Committee (1998).

8.2 Correctness

Positional imprecision is one source of uncertainty. The information is there, but it is only approximately correct; the values in the data are close but not equal to the "true value". Another source of uncertainty is the fact that some of the data may simply be wrong. This issue is often raised with respect to attribute accuracy. Are the discrete label values given for such items as land use type in fact correct? Is this parcel that is classified as residential in fact residential? How accurately is data classified into appropriate categories? This should be but rarely is reported as the percentage of items in the data that are actually correct. Often we would like a combined measure of precision and completeness, such as that 98% of the entities are located within 10 meters of their true locations.

A common source of incorrect data is stale data. We need to know the age of the data. Data sources may be too old to be useful or relevant, and past collection standards may be unknown, nonexistent, or not currently acceptable. The world changes, and geospatial resources often don't track the changes in as timely a fashion as they should. Temporal coordinates are often only implicit such as a timestamp. Even when we know the date of the resource, this is often inaccurate, because it characterizes the database as a whole rather than each individual item.

Classification problems can arise. Often discrete categories are required where the set of categories is really a discretization of a continuous measurement. Defining appropriate class intervals can be a subjective process.

Logical inconsistency is a very good indication that something in incorrect somewhere. For this reason, it is good to have a description of the fidelity of relationships encoded in the data structure, such as permissible values, number of occurrences (no duplications), and constraints in relationships (e.g., intersections, overshoots, and undershoots). For example, sliver polygons or "virtual data" may result from overlaying multiple layers of data. Taking data from two different county databases may result in illegitimate overlaps or gaps.

Honest disagreements and conflicts sometimes occur. Reports may differ on whether a particular parcel is being used for residential or agricultural purposes. Where this is the case, it would be good to have means for indicating this, so that the data source is not forced to lose information or give misleading information.

8.3 Completeness

Correctness is a measure of the reliability of what is present in the resource. Completeness is a measure of the reliability of what is absent from the resource. If some entity is not in the resource, is it because it doesn't exist, or is it because we lack knowledge of it?

This can be a very important issue. In the Pakistani earthquake, it was very important to know where *all* of the villages were in order to know whether any were buried under landslides.

Soldiers on patrol will want to know all the potential ambush sites. If a hill next to a curve in the road does not show on a map that is assumed to be complete, serious consequences could occur.

Geospatial resources should characterize their level of completeness as precisely as possible. For example, a map may cover only cities with population greater than 100,000.

Attribute completeness (United States Geological Survey, 2005) can be a particularly tricky issue. Suppose a set of attributes is defined for some particular class of objects. If some of the values for those attributes are missing, how are we to interpret it. Does it mean they are unknown, does it mean they have the more negative value, or does it mean they have the default, or the non-default, value?

8.4 Inherent Vagueness

Some concepts are simply inherently vague. No amount of further research could make them more precise. For example, what exactly are the boundaries of Mount Whitney? A mountain is not a crisp object in the sense that its boundaries can be precisely located. There are points that everyone would certainly identify as being part of the mountain and points that everyone would certainly identify as not being part of the mountain. But there is a large gray area in between.

Similar problems arise for attributes of regions. At what point does desert become savannah? Where exactly does landcover change from forest to agricultural? What are the precise boundaries of a hurricane?

Many commonsense notions in geography are like this. Nature does not partition the world neatly into categories for us. Smith and Mark (2001; 2003) have discussed this issue at length.

8.5 Provenance

Perhaps the most important information one can have for judging the reliability of the data in a geospatial resource is its provenance.

Provenance could be interpreted minimally as the source of the data—who said so, where and when? More generally, the provenance includes the method used for collecting the component data and the algorithm, procedure, or projection that produced the final data from the components.

Still more generally, a theory of provenance is a theory of information-producing processes. Groth et al. (2004; 2005) develop a theory of provenance in terms of agents communicating by messages. At the lowest level, only the fact of a message is recorded. At the next level, causal relations are recorded, such as the one message is a reply to another message. At the next level, the functional transformation that an agent performs on messages to produce new messages is recorded. At the highest level, there is all this information plus information about the internal state of the agent, such as what machine was used, what version of what algorithm was employed, and so on.

Ideally, provenance should be associated with individual items of data rather than with the resource as a whole, since there could be great variability in the reliability of the items in a database.

A very thorough account of varieties of uncertainty and how to document the influence of provenance on the reliability of geospatial data has been published as part of the Spatial Data Transfer Standard (United States Geological Survey, 2005).

8.6 Reasoning about Uncertainty

There has been a great deal of research in artificial intelligence and related fields on reasoning with uncertain information. The seminal work in this area is on Bayesian networks or belief nets by Pearl (1988)).

Work on uncertainty specific to spatial reasoning includes that of Cohn (1996) and of Worboys and his colleagues (Worboys & Clementini, 2001). Worboys and Clementini survey different kinds of uncertainty in GIS systems, including vagueness, imprecision and multi-resolution. They develop a lattice of topological relations between regions with broad boundaries, or egg-yolk regions, and investigate its use in a framework for integrating uncertain geospatial data. In egg-yolk theories, there is an inner region in which a property certainly holds and an outer region in which it may hold.

Duckham and colleagues (Duckham *et al.* , 2006) develop a qualitative framework for reasoning about inconsistency based on description logics and use it for integrating geographic information at different granularities. The core approach annotates RCC-5[1] with a notion of connectedness of geographic features. For example, $C(\text{woodland}, \text{forest})$ indicates that the concepts "woodland" and "forest" have some overlap. They then represent qualitatively the assumption that ground cover and other geographic features might be expected to change continuously rather than suddenly. By using a tractable description logic, the authors aim to defer resolving inconsistencies between sources until the information is used. When integrating information from heterogeneous sources, the connectedness relation is applied to geographic features in the different sources, as a form on ontology alignment.

One of the principal reasons for using qualitative information is the uncertainty in or lack of quantitative data, so in a sense all research done on qualitative representation and reasoning is research on reasoning about uncertainty. Moreover, the categories one finds most natural in human language and cognition are the categories that have proved to be most functional over the millennia. There is a strong argument that they have located the optimal trade-off between precision and functionality, or the optimal trade-off between the value of more precise information and the cost of obtaining it.

8.7 Visualizing Uncertainty

Uncertainty in the location of geo-referenced items is a concern. Some forms of uncertainty are due to deficiency in the data collection process, some forms are due to differences in scale or resolution, and sometimes, especially with forecasts, the uncertainty is inherent in the process.

Take, for example, the forecast of tropical storm and hurricane tracks. There is an inherent uncertainty in the future locations and the exact track of the eye of the storm. The Natinal Oceanic & Atmospheric Administration (NOAA) publishes a standard rule governing the uncertainty in its forecasts, known as the "Mariner's 1-2-3 Rule". The 1-2-3 Rule describes the uncertainty in the forecast locations produced by the National Hurricane Center. The errors are of 100-200-300 nautical miles at 24-48-72 hours into the future. Figure 8.1 shows a map depicting error circles of this nature.

A more sophisticated graphical representation uses an envelope around the storm track, to emphasize the contiuous nature of the track, and also eliminates gaps in the area caused by high speed movement of the storms. Figure 8.2 shows several examples of this representation. The NOAA example uses a smoothed envelope with center locations. The

[1]This is similar to the RCC-8 formalism, but with fewer distinctions. No distinction is made between the tangential and non-tangential notions of overlap or containment.

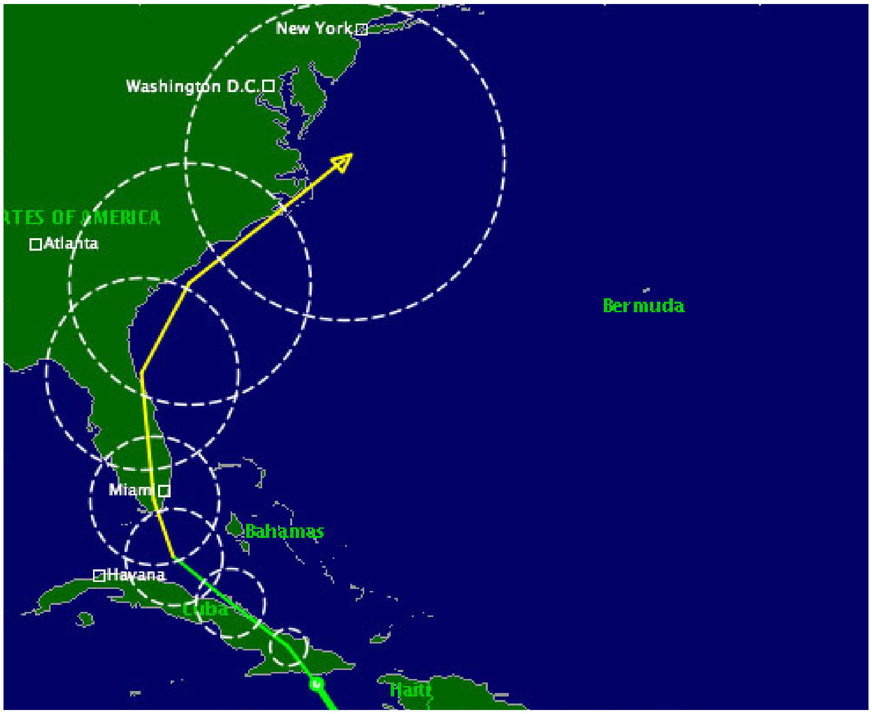

Figure 8.1: Uncertainty in storm center locations of tropical storm Florence, depicted as circles around future locations. The further into the future the forecast extends, the larger the error circle. This map, dated September 1, 2006, is from the University College of London's Tropical Storm Risk website.

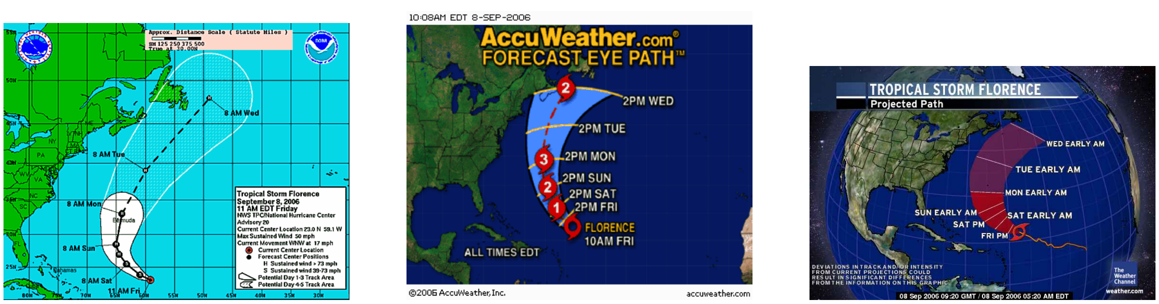

Figure 8.2: Uncertainty of storm movement for Tropical Storm Florence expressed as a smoothed envelope surrounding the forecast storm track. Different sources use different depictions of the envelope, with emphasis on either the central predicted lcoation, or the limits of expected travel. Maps are from September 8, 2006.

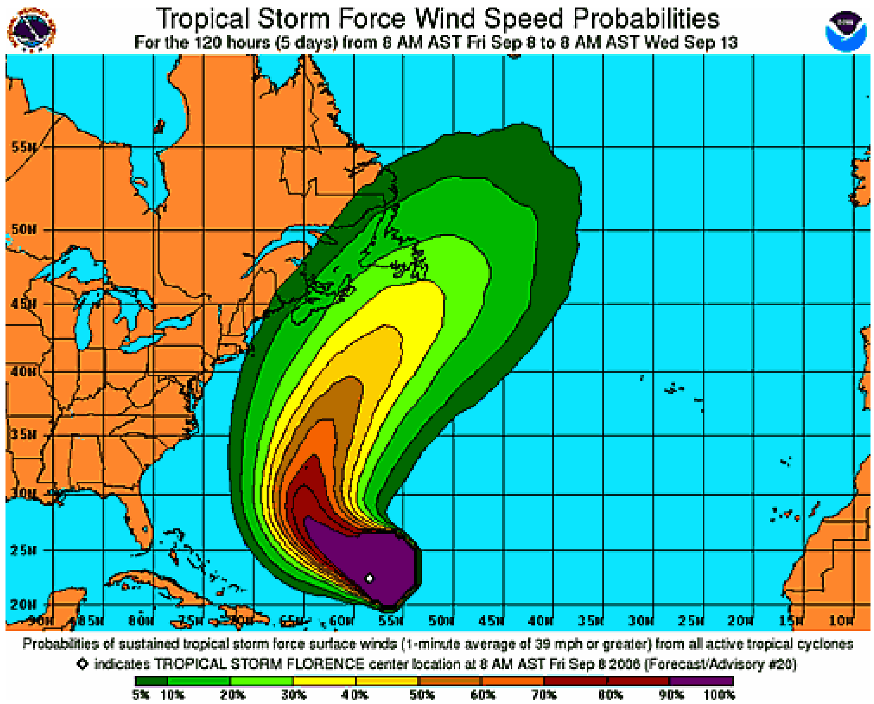

Figure 8.3: Probability-based forecast map showing the chance that tropical storm force winds will be present in a given area during the next five days. This NOAA map also shows the effects of the uncertainty in the path of tropical storm Florence as well as the likely weakening of the storm as it moves further north. The map is from September 8, 2006.

AccuWeather and Weather Channel maps use a curved limit bar inside the envelopes. One can also see the use of different map projections as well.

Although the envelope and circles show the uncertainty, they do leave the visual impression that all of the area inside the envelope is equally likely. Other representations of uncertainty use shading to indicate the relative likelihood of a particular event becoming true. Keeping with the meterological example, Figure 8.3 shows the probability that different areas will experience winds of at least tropical storm force over a five-day forecast period. This particular visualization provides a nice, compact presentation of key information derived from a number of different input factors.

Chapter 9

Recommendations

9.1 Introduction

In any discipline the best research talent should be directed, insofar as possible, toward problems that are tractable but not trivial. The trivial problems, if they are of economic or scientific interest, will be solved anyway. Many if not most commercial developments are a matter of solving relatively simple problems and applying the solution at a large scale. The intractable problems, where we can recognize their intractability, should wait until we have the tools to address them.

In this survey, three problems that fall between the trivial and the intractable stand out. The first is difficult for sociological reasons, the others for technical reasons.

The first is that the exploitation of the huge number of diverse geospatial resources that are potentially available would be greatly facilitated if they could interoperate—if they could share data and if the results derived from one resource could be utilized by other resources. This is difficult because there are so many resources, and for that matter, so many competing standards. It is tractable because if a standard ontology has easy mappings to existing standards and resources, an internal elegance, and the backing of the U.S. government, people will be strongly motivated to use it.

The second problem is the development of qualitative and hybrid representations of space and composite representations of space and time. Natural language is full of qualitative terms, and people in their everyday lives continually link qualitative and quantitative judgments. A framework for combining qualitative and quantitative geospatial information would allow geographic reasoning systems to incorporate qualitative observations from humans into situation assessment based on quantitative data, and to present summaries and recommendations in ways that are easier for humans to absorb while other tasks compete for their attention. But it is the quantitative concepts that yielded first to computational treatment, and it was not obvious how the qualitative could be dealt with in a computational framework. By now, however, there has been a substantial amount of research on qualitative concepts, and some on hybrid concepts, so that the time is ripe for a concerted effort on this problem. Similarly, we now have a number of ontologies of space and time, and the time is ripe to explore the problems in combining them

The third problem that suggests itself as ready for a concerted assault is the problem of using geospatial and other contextual factors to analyze situations in a context-sensitive fashion.

We discuss each of these challenges below in a little more depth.

9.2 Semantic Interoperability

Most of the standarization process has involved the definition of syntax for describing data and metadata. Although this allows the interchange of information at the syntactic level, it still does not solve the problems of determining the semantic content. This is true even for such seemingly simple areas like ISO-6709 "Standard representation of latitude, longitude and altitude". The standard provides a syntax for specifying the latitude, longitude and altitude, but does not have any provision for specifying the geodetic reference frame (See Appendix A). Altitude is specified as being in meters, and so is well-defined.

That means that the semantics of exactly how the latitude and longitude are interpreted must be agreed by systems that are communicating using some mechanism outside of the standardization process. This can be a significant issue. The difference in location can be several hundred meters, which can be a substantial problem if a combat unit and its supporting artillery are using different reference data. (National Geospatial-Intelligence Agency, 2006, "Why should I care?").

The problem of making different GIS datasets interoperable arises mostly becasue the semantics of the attributes of the datasets—the meanings of the field names and the meanings of the values in the fields—is not well defined. Sometimes the description is completely missing, because the developers think that the English-like field names are self-explanatory, e.g., "LUCode" for land use code in planning data), and when the description is present, it is in natural language and subject to each individual's interpretation. As a result, it is nearly always a nightmare to merge two datasets dealing with the same subject matter but created by different agencies.

One finds a similar weakness in the FGDC standard for specifying metadata. The framework provides only a skeleton to use for interchanging or declaring metadata. Many of the metadata fields are plain text fields. These are therefore aimed at human readers of the metadata rather than at computer systems. Other areas are amenable to using some form of controlled vocabulary, but the specification of that vocabulary lies outside the FGDC standard iteslf. The standard acknowledges the need for agreeing on common terminology. Section 1.6.1.1 of the standard explicitly points to the need for "a formally registered thesaurus or a similar authoritative source of theme keywords" for specifying the semantic content of the description of the underlying data.

The role that ontologies need to play is to provide the semantic content to enable these systems to operate and exchange information with each other. Therefore, our recommendation is that research should be supported in making more precise and formal the specifications of ontologies underlying such widely accepted standards as OpenGIS and that the relevant communities should be strongly encouraged to adopt them.

9.3 Representation

Geographic Information Systems use quantitative representations as their fundamental data structure. This provides a high degree of precision in the represented entities, but also limits the expressive power. One is forced to make very precise statements. To some extent this can be ameliorated by substituting regions for points (as shown in the storm track examples in Section 8.7).

Geographic Information Systems also support the overlaying of qualitative descriptions, based on the details of the underlying quantitative model. ISO draft standard 19107, for example, supports the use of topological relations, such as those in RCC-8, in GIS systems. (Herring, 2001) These topological relations seen as computational aids which can make certain types of reachability calculations proceed more smoothly.

But the potential benefits of increasing the computational efficiency is only part of the usefulness of such qualitative relations, and the kinds of qualitative relations that have been incorporated so far only scratches the surface of what is possible.

We recommend that research should be encouraged into more complex areas of qualitative reasoning and representation. For example, we should develop richer models of the topology of complex three-dimensional structures. This will enable us to capture, represent, and communicate imprecise but very useful information about such structures. Further research should be encouraged on qualitative trigonometries and their applications to questions of shape. There should be investigations of various granularities that can be imposed on individual dimensions, and how they can combine to produce useful granularities on higher-dimensional spaces.

We recommend that research should be encouraged on hybrid qualitative-quantitative representation and reasoning. There should be a smooth transition between the two, so that the two can reside comfortably in the same system, and so that systems can move easily between the two as appropriate for the task at hand.

We recommend a research emphasis on combining spatial and temporal ontologies, to give us a good range of ontology types for motion and change. Included here should be research on combining geospatial models with rich ontologies of events and processes. Very often when we see differences between two snapshots taken at different times, we want to know more than simply that a change occurred. We want to know what caused the change. The causal nature of rich event and process ontologies can help us discover and characterize the processes that brought the change about. The subevent and subprocess knowledge in the ontologies can help us abduce the larger picture in which this change should be understood.

A good start has been made in all of these areas, so that a concerted effort at this time will certainly produce substantial results.

9.4 Context-sensitive Analysis

Many of the tasks that are faced in battlefield command, intelligence preparation of the battlefield and in evaluating courses of action are expressed as qualitative descriptions. Although they derive from the underlying details of terrain, these descriptions embody higher-level concepts used in planning and decision making.

For example, "mobility corridors" can be computed based on underlying information provided by GIS and geographic decision support systems. But the determination of whether such a mobility corridor could constitute a potential "avenue of advance" depends on the context of the analysis. It requires some knowledge of the higher-level plan, since an avenue of advance depends on the desired direction of movement. Similarly, the question of whether a river constitutes an obstacle depends on whether one is trying to cross it or move parallel to it. And if the river is a barrier, this can be either good or bad depending on whether one is attacking or defending in that area.

To a limited extent, some work on incorporating mission-related elements has been done already. One fairly simple example involves computing terrain masking for aircraft paths, which requires knowledge of the desired path of travel. Tools for that purpose exist (FM 3-34.230, 2000, Appendix B, figure B-24), but still provide only part of a full solution. In AI research, there is Donlon's (1999) work on trafficability analysis. SRI (SRI International, 1995) and CMU (Grindle *et al.* , 2004; Glinton *et al.* , 2004), among others, have also developed prototype systems for intelligence preparation of the battlefield. These use a more integrated approach which considers the mission of forces as well as drawing on geographic data sources.

A good area for further research is to develop ways of allowing the mission to guide the analysis of the underlying geospatial data. This will allow more focused products to be built. This effort can extend some of the existing research efforts as well as leverage earlier DARPA work in CoA analysis.

Appendix A

Geodetic Systems

A geodetic system is a reference frame for surveying and navigational coordinates. Latitude and longitude are defined relative to a particular World Geodetic System. For more details, consult http://en.wikipedia.org/wiki/WGS84.

Geodetic reference frames are defined as ellipsoids (elliptical-shaped solid figures). A "geoid" is a mathematical model of the surface of the earth that corresponds to a global mean sea level. More details on geoids and their relation to ellipsoids can be found at http://www.ngs.noaa.gov/GEOID/geoid_def.html.

The key parameters used to describe an ellipsoid is the semi-major axis a, and the inverse flattening f^{-1}. The semi-major axis is the longest axis and the flattening is the fractional amount by which the semi-minor axis differs from the semi-major axis.

Table A.1: Comparison of Reference Ellipsoids

Reference Ellipsoid Name	a (m)	f^{-1}
Airy 1830	6377563.396	299.3249646
Australian National	6378160	298.25
Bessel 1841—Ethiopia, Indonesia, Japan and Korea	6377397.155	299.1528128
Bessel 1841—Namibia	6377483.865	299.1528128
Clarke 1866	6378206.4	294.9786982
Clarke 1880	6378249.145	293.465
Everest—Brunei and E. Malaysia	6377298.556	300.8017
Everest—India 1830	6377276.345	300.8017
Everest—India 1956	6377301.243	300.8017
Everest—Pakistan	6377309.613	300.8017
Everest—W. Malaysia and Singapore 1948	6377304.063	300.8017
Everest—W. Malaysia 1969	6377295.664	300.8017
Geodetic Reference System 1980	6378137	298.257222101
Helmert 1906	6378200	298.3

Comparison of Reference Ellipsoids (cont.)

Reference Ellipsoid Name	a (m)	f^{-1}
Hough 1960	6378270	297
Indonesian 1974	6378160	298.247
International 1924	6378388	297
Krassovsky 1940	6378245	298.3
Modified Airy	6377340.189	299.3249646
Modified Fischer 1960	6378155	298.3
South American 1969	6378160	298.25
WGS 1972	6378135	298.26
WGS 1984	6378137	298.257223563

Source: (National Geospatial-Intelligence Agency, 1997)

Appendix B

ISO Standards

There are a number of international standards dealing with geospatial data. A summary of those International Organization for Standardization (ISO) standards is given in Table B.1. A number of these standards are concerned with important, but mundane issues of interoperability. Others address semantic content and are of more relevance to this report.

Table B.1: ISO Standards for Geospatial Information

Number	**Content**
6709	Standard representation of latitude, longitude and altitude for geographic point locations
13249–3	Information technology — Database languages — SQL multimedia and application packages — Part 3: Spatial
19101	Geographic information — Reference model
19101–2	Geographic information — Reference model — Part 2: Imagery
19103	Geographic information — Conceptual schema language
19104	Geographic information — Terminology
19106	Geographic information — Profiles
19107	Geographic information — Spatial schema
19108	Geographic information — Temporal schema
19109	Geographic information — Rules for application schema
19110	Geographic information — Methodology for feature cataloguing
19111	Geographic information — Spatial referencing by coordinates
19111 rev	Geographic information — Spatial referencing by coordinates
19113	Geographic information — Quality principles
19114	Geographic information — Quality evaluation procedures
19115	Geographic information — Metadata
19115–2	Geographic information — Part 2: Extensions for imagery and gridded data

ISO Standards for Geospatial Information (cont.)

Number	Content
19116	Geographic information — Positioning services
19117	Geographic information — Portrayal
19118	Geographic information — Encoding
19119 Amd. 1	Geographic information — Services — Amendment 1
19120	Geographic information — Functional standards
19121	Geographic information — Imagery and gridded data
19122	Geographic information/Geomatics — Qualification and certification of personnel
19123	Geographic information — Schema for coverage geometry and functions
19124	Geographic information — Imagery and gridded data components
19125–1	Geographic information — Simple feature access — Part 1: Common architecture
19125–2	Geographic information — Simple feature access — Part 2: SQL option
19126	Geographic information — Profile - FACC Data Dictionary
	Deleted due to lack of progress
19127	Geographic information — Geodetic codes and parameters
19128	Geographic information — Web Map Server interface
19129	Geographic information — Imagery, gridded and coverage data framework
19130	Geographic information — Sensor data models for imagery and gridded data
	Deleted due to slow progress
19131	Geographic information — Data product specifications
19132	Geographic information — Location Based Services — Reference model
19133	Geographic information — Location-based services — Tracking and navigation
19134	Geographic information — Location-based services — Multimodal routing and navigation
19135	Geographic information — Procedures for item registration
19136	Geographic information — Geography Markup Language
19137	Geographic information — Core profile of the spatial schema
19138	Geographic information — Data quality measures
19139	Geographic information — Metadata — XML schema implementation

ISO Standards for Geospatial Information (cont.)

Number	Content
19140	Geographic information amendment process
19141	Geographic information — Schema for moving features
19142	Geographic information — Web Feature Service
19143	Geographic information — Filter encoding
19144–1	Geographic information — Classification Systems Part 1: Classification system structure
19144–2	Geographic information — Classification Systems Part 2: Land Cover Classification System LCCS
19145	Geographic information — Registry of representations of geographic point locations
	Geographic information — Amendment to ISO 19113:2002 Geographic information - Quality principles and ISO 19115:2003 Geographic information — Metadata

Appendix C

USGS GNIS Feature Classes

The US Geological Survey server classifies named features into 65 classes of features. They are listed here in Table C.1.

Table C.1: USGS Gazetteer Feature Classes

Name	**Feature Description**
airport	manmade facility maintained for the use of aircraft.
arch	natural arch-like opening in a rock mass.
area	any one of several areally extensive natural features not included in other categories.
arroyo	watercourse or channel through which water may occasionally flow.
bar	natural accumulation of sand, gravel, or alluvium forming an underwater or exposed embankment.
basin	natural depression or relatively low area enclosed by higher land.
bay	indentation of a coastline or shoreline enclosing a part of a body of water; a body of water partly surrounded by land.
beach	the sloping shore along a body of water that is washed by waves or tides and is usually covered by sand or gravel.
bench	area of relatively level land on the flank of an elevation such as a hill, ridge, or mountain where the slope of the land rises on one side and descends on the opposite side.
bend	curve in the course of a stream and (or) the land within the curve; a curve in a linear body of water.
bridge	manmade structure carrying a trail, road, or other transportation system across a body of water or depression.

USGS Gazetteer Feature Classes (cont.)

Name	**Feature Description**
building	a manmade structure with walls and a roof for protection of people and (or) materials, but not including church, hospital, or school.
canal	manmade waterway used by watercraft or for drainage, irrigation, mining, or water power.
cape	projection of land extending into a body of water.
cave	natural underground passageway or chamber, or a hollowed out cavity in the side of a cliff.
cemetery	a place or area for burying the dead.
channel	linear deep part of a body of water through which the main volume of water flows and is frequently used as aroute for watercraft.
church	building used for religious worship.
civil	a political division formed for administrative purposes.
cliff	very steep or vertical slope.
crater	circular-shaped depression at the summit of a volcanic cone or one on the surface of the land caused by the impact of a meteorite; a manmade depression caused by an explosion.
crossing	a place where two or more routes of transportation form a junction or intersection.
dam	water barrier or embankment built across the course of a stream or into a body of water to control and (or) impound the flow of water.
falls	perpendicular or very steep fall of water in the course of a stream.
flat	relative level area within a region of greater relief.
forest	bounded area of woods, forest, or grassland under the administration of a political agency (see "woods").
gap	low point or opening between hills or mountains or in a ridge or mountain range.
geyser	eruptive spring from which hot water and (or) steam and in some cases mud are periodically thrown.
glacier	body or stream of ice moving outward and downslope from an area of accumulation; an area of relatively permanent snow or ice on the top or side of a mountain or mountainous area.
gut	relatively small coastal waterway connecting larger bodies of water or other waterways.
harbor	sheltered area of water where ships or other watercraft can anchor or dock.

USGS Gazetteer Feature Classes (cont.)

Name	Feature Description
hospital	building where the sick or injured may receive medical or surgical attention.
island	area of dry or relatively dry land surrounded by water or low wetland.
isthmus	narrow section of land in a body of water connecting two larger land areas.
lake	natural body of inland water.
lava	formations resulting from the consolidation of molten rock on the surface of the Earth.
levee	natural or manmade embankment flanking a stream.
locale	place at which there is or was human activity; it does not include populated places, mines, and dams.
mine	place or area from which commercial minerals are or were removed from the Earth; not including oilfield.
military (historical)	place or facility formerly used for various aspects of or relating to military activity.
oilfield	area where petroleum is or was removed from the Earth.
other	category for miscellaneous named entities that cannot readily be placed in the other feature classes listed here.
park	place or area set aside for recreation or preservation of a cultural or natural resource and under some form of government administration; not including National or State forests or Reserves.
pillar	vertical, standing, often spire-shaped, natural rock formation.
plain	a region of general uniform slope, comparatively level and of considerable extent.
Post Office	an official facility of the U.S. Postal Service used for processing and distributing mail and other postal material.
Populated Place	place or area with clustered or scattered buildings and a permanent human population.
range	chain of hills or mountains; a somewhat linear, complex mountainous or hilly area.
rapids	fast-flowing section of a stream, often shallow and with exposed rock or boulders.
reserve	a tract of land set aside for a specific use (does not include forests, civil divisions, parks).
reservoir	artificially impounded body of water.

USGS Gazetteer Feature Classes (cont.)

Name	**Feature Description**
ridge	elevation with a narrow, elongated crest which can be part of a hill or mountain.
school	building or group of buildings used as an institution for study, teaching, and learning.
sea	large body of salt water.
slope	a gently inclined part of the Earth's surface.
spring	place where underground water flows naturally to the surface of the Earth.
stream	linear body of water flowing on the Earth's surface.
summit	prominent elevation rising above the surrounding level of the Earth's surface; does not include pillars, ridges, or ranges.
swamp	poorly drained wetland, fresh or saltwater, wooded or grassy, possibly covered with open water.
tower	a manmade structure, higher than its diameter, generally used for observation, storage, or electronic transmission.
trail	route for passage from one point to another; does not include roads or highways.
tunnel	linear underground passageway open at both ends.
valley	linear depression in the Earth's surface that generally slopes from one end to the other.
well	manmade shaft or hole in the Earth's surface used to obtain fluid or gaseous materials.
woods	small area covered with a dense growth of trees; does not include an area of trees under the administration of a political agency (see "forest").

References

AGARWAL, P. 2005. Ontological Considerations in GIScience. *International Journal of Geographical Information Science*, **19**(5), 501–536.

ALLEN, JAMES F. 1984. Toward a general theory of action and time. *Artificial Intelligence*, **23**(2), 123–154.

BATEMAN, JOHN, & FARRAR, SCOTT. June 2005. *Spatial Ontology Baseline*. Tech. rept. SFB/TR 8: I1, deliverable D2. University of Bremen, Germany. This is a draft report and labeled "not for citation", but since we consulted it, we wanted to make sure credit was given.

BENNETT, B. 2001. A Categorical Axiomatisation of Region-Based Geometry. *Fundamenta Informaticae*, **46**, 145–158.

BENNETT, B. 2003. The Role of Definitions in Construction and Analysis of Formal Ontologies. *In: Sixth Symposium on Logical Formalizations of Commonsense Reasoning*. AAAI Spring Symposium.

BENNETT, B., & GALTON, A.P. 2004. A Unifying Semantics for Time and Events. *Artificial Intelligence*, **153**, 13–48.

BENNETT, BRANDON, COHN, ANTHONY G., WOLTER, FRANK, & ZAKHARYASCHEV, MICHAEL. 2002. Multi-Dimensional Modal Logic as a Framework for Spatio-Temporal Reasoning. *Applied Intelligence*, **17**(3), 239–251.

BITTNER, T., & STELL, J. G. 2002. Vagueness and Rough Location. *GeoInformatica*, **6**, 99–121.

CASATI, ROBERTO, & VARZI, ACHILLE C. 1997. Spatial Entities. *In:* (Stock, 1997).

CLARAMUNT, CHRISTOPHE, PARENT, CHRISTINE, & THERIAULT, MARIUS. 1997. *Design Patterns for Spatiotemporal Processes*.

CLEMENTINI, E., & FELICE, P. DI. 1995. A Comparison of Methods for Representing Topological Relationships. *Information Sciences*, **3**, 149–178.

CLEMENTINI, E., & FELICE, P. DI. 1996. A Model for Representing Topological Relationships Between Complex Geometric Features in Spatial Databases. *Information Sciences*, **90**(1-4), 121–136.

COHN, A G, & HAZARIKA, S M. 2001. Qualitative Spatial Representation and Reasoning: An Overview. *Fundamenta Informaticae*, **46**(1-2), 1–29.

COHN, ANTHONGY G., & GOTTS, NICK M. 1996. The 'Egg-Yolk' Representation of Regions with Indeterminate Boundaries. *Pages 171–187 of:* BURROUGH, P., & FRANK, A. M. (eds), *Proceedings, GISDATA Specialist Meeting on Geographical Objects with Undetermined Boundaries*. Francis Taylor.

COHN, ANTHONY G., BENNETT, BRANDON, GOODAY, JOHN, & GOTTS, NICHOLAS M. 1997. Representing and Reasoning with Qualitative Spatial Relations About Regions. *In:* (Stock, 1997).

CRISTANI, MATTEO, & COHN, ANTHONY G. 2002. SpaceML: A Mark-up Language for Spatial Knowledge. *Journal of Visual Languages and Computing*, **13**(1), 97–116.

CYCORP. 2002a. *Geography*. Cycorp, Inc., Austin, TX. Available from http://www.cyc.com/cycdoc/vocab/geography-vocab.html.

CYCORP. 2002b. *Paths and Traversals*. Cycorp, Inc., Austin, TX. Available from http://www.cyc.com/cycdoc/vocab/path-traversal-vocab.html.

CYCORP. 2002c. *Spatial Relations*. Cycorp, Inc., Austin, TX. Available from http://www.cyc.com/cycdoc/vocab/spatial-vocab.html.

DE KLEER, JOHAN, & BROWN, JOHN SEELY. 1984. A Framework for Qualitative Physics. *Pages 11–18 of: Proceedings of the Sixth Annual Conference of the Cognitive Science Society*.

DONLON, CPT JAMES J., & FORBUS, KENNETH D. 1999 (June 6–9). Using a Geographic Information System for Qualitative Spatial Reasoning about Trafficability. *In: Proceedings of the 13th International Workshop on Qualitative Reasoning*.

DUCKHAM, M., LINGHAM, J., MASON, K., & WORBOYS, M. F. 2006. Qualitative Reasoning about Consistency in Geographic Information. *Information Sciences*, **176**(6), 601–627.

EGENHOFER, MAX. 1991. Reasoning about Binary Topological Relations. *Pages 143–160 of:* GÜNTHER, O., & SCHEK, H. (eds), *Proceedings of the Second Symposium on Large Scaled Spatial Databases*. Berlin: Springer.

EGENHOFER, MAX J., & FRANZOSA, ROBERT D. 1991. Point Set Topological Relations. *International Journal of Geographical Information Systems*, **5**, 161–174.

EGENHOFER, MAX J., & FRANZOSA, ROBERT D. 1995. On the Equivalence of Topological Relations. *International Journal of Geographical Information Systems*, **9**(2), 133–152.

EHLSCHLAEGER, CHARLES R., & SHORTRIDGE, ASHTON. 1996. Modeling Elevation Uncertainty in Geographical Analyses. *Pages iB.15–9B.25 of: Proceeding of the International Symposium on Spatial Data Handling*.

ESRI. 1998 (July). *ESRI Shapefile Technical Description*. ESRI, Redlands, CA. http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf.

ESRI. 2004a. *The ArcGIS Network Model*. ESRI, Redlands, CA. http://edndoc.esri.com/arcobjects/9.1/ArcGISDevHelp/TechnicalDocuments/Network/ArcGISNetworkModel/ArcGISNetwork.htm.

ESRI. 2004b. *ArcGIS Tracking Analyst*. ESRI, Redlands, CA. http://www.esri.com/trackinganalyst.

ESRI. 2004c. *Using ArcGIS Tracking Analyst*. Redlands, CA: ESRI.

FEDERAL GEOGRAPHIC DATA COMMITTEE (FGDC). 1998. *Geospatial Positioning Accuracy Standards, Part 3: National Standard for Spatial Data Accuracy*. FGDC-STD-007.3-1998, available from http://www.fgdc.gov/standards/projects/FGDC-standards-projects/accuracy/part3/chapter3.

FEDERAL GEOGRAPHIC DATA COMMITTEE (FGDC). revised June 1998. *Content Standard for Digital Geospatial Metadata*. Tech. rept. FGDC-STD-001-1998. Federal Geographic Data Committee, Washington, D. C.

FM 3-34.230. 2000 (3 August). *Topographic Operations*. Department of the Army, Washington, DC. Available from http://www.globalsecurity.org/military/library/policy/army/fm/3-34-230/index.html, visited September 18, 2006.

FORBUS, KENNETH. 1995. Qualitative Spatial Reasoning: Framework and Frontiers. *In:* GLASGOW, J., H. NARAYANAN, H, & CHANDRASEKARAN, B. (eds), *Diagrammatic Reasoning : Computational and Cognitive Perspectives*. AAAI Press.

FORBUS, KENNETH, NIELSEN, P, & FALTINGS, BOI. 1991. Qualitative Spatial Reasoning: The CLOCK Project. *Artificial Intelligence*, **51**(October), 1–3.

FORBUS, KENNETH D. 1980. Spatial and Qualitative Aspects of Reasoning about Motion. *Pages 170–173 of: Proceedings of the First National Conference on Artificial Intelligence (AAAI-80)*.

FORBUS, KENNETH D. 1981. Qualitative Reasoning about Physical Processes. *Pages 326–330 of:* HAYES, PATRICK J. (ed), *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*. Vancouver, BC, Canada: William Kaufmann.

FORBUS, KENNETH D. 1984. Qualitative Process Theory. *Artificial Intelligence*, **24**(1-3), 85–168.

FORBUS, KENNETH D. 1985. The Role of Qualitative Dynamics in Naive Physics. *In:* (Hobbs & Moore, 1985).

FORESMAN, T. W. (ed). 1998. *The History of GIS: Perspectives from the Pioneers*. Prentice Hall.

FRANK, ANDREW U. 1996. Qualitative Spatial Reasoning: Cardinal Directions as an Example. *International Journal of Geographical Information Science*, **10**(3), 269–290.

FRANK, ANDREW U. 1997. Spatial Ontology: A Geographical Information Point of View. *In:* (Stock, 1997).

FRANK, ANDREW U. 2003. Ontology for Spatio-temporal Databases. *Chap. 2, pages 9–77 of:* KOUBARAKIS, MANOLIS, SELLIS, TIMOS, FRANK, ANDREW U., GRUMBACH, STÉPHANE, GÜTING, RALF HARTMUT, JENSEN, CHRISTIAN S., LORENTZOS, NIKOS, MANOLOPOULOS, YANNIS, NARDELLI, ENRICO, PERNICI, BARBARA, SCHEK, HAND-JÖRG, SCHOLL, MICHEL, THEODOULIDIS, BABIS, & TRYFONA, NECTARIA (eds), *Spatio-Temporal Databases: The CHOROCHRONOS Approach*. Lecture Notes in Computer Science, vol. 2520. Springer Verlag.

FREKSA, CHRISTIAN. 1992. Using Orientation Information for Qualitative Spatial Reasoning. *Pages 162–178 of:* FRANK, A. U., CAMPARI, I., & FORMENTINI, U. (eds), *Theories and Methods of Spatio-temporal Reasoning in Geographic Space*. Berlin: Springer.

GALTON, ANTHONY. 1997. Space, Time and Movement. *In:* (Stock, 1997).

GENESERETH, M.R. 1991. Knowledge Interchange Format. *Pages 599–600 of:* ALLEN, JAMES, FIKES, RICHARD, & SANDEWALL, ERIC (eds), *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*. San Mateo, CA, USA: Morgan Kaufmann Publishers.

GETTY. 2006a (28 March). *Getty Thesaurus of Geographic Names: Editorial Guidelines*. The Getty Vocabulary Program. Available from http://www.getty.edu/research/conducting_research/vocabularies/editorial_guidelines.html#tgn.

GETTY. 2006b. *Getty Thesaurus of Geographic Names Online*. Web page http://www.getty.edu/research/conducting_research/vocabularies/tgn/. Visited August 8, 2006.

GLINTON, ROBIN, OWENS, SEAN R., GIAMPAPA, JOSEPH ANDREW, SYCARA, KATIA, GRINDLE, CHARLES E., & LEWIS, C. MICHAEL. 2004 (July). Terrain-Based Information Fusion and Inference. *In: Proceedings of the Seventh International Conference on Information Fusion*.

GRENON, PIERRE. 2003. Tucking RCC in Cyc's Ontological Bed. *Pages 894–899 of:* GOTTLOB, GEORG, & WALSH, TOBY (eds), *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03)*. Acapulco, Mexico: Morgan Kaufmann.

GRINDLE, CHARLES, LEWIS, MICHAEL, GLINTON, ROBIN, GIAMPAPA, JOSEPH ANDREW, OWENS, SEAN R, & SYCARA, KATIA. 2004. Automating Terrain Analysis: Algorithms for Intelligence Preparation of the Battlefield. *Pages 533–537 of: Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*. P.O. Box 1369, Santa Monica, CA 90406-1369: Human Factors and Ergonomics Society.

GROTH, PAUL, LUCK, MICHAEL, & MOREAU, LUC. 2004 (September). Formalising a Protocol for Recording Provenance in Grids. *In: Proceedings of the UK OST e-Science Second All Hands Meeting AHM'o4*.

GROTH, PAUL, LUCK, MICHAEL, & MOREAU, LUC. 2005 (October). *Architecture for Provenance Systems*. Technical Report. University of Southhampton.

GRUNNINGER, MIKE. 2006. *PSL Ontology*. Web page http://www.mel.nist.gov/psl/ontology.html. Visited September 30, 2006.

GUHA, R.V., & LENAT, D.B. 1990. CYC: A midterm report. *AI Magazine*, **11**(3), 33–59.

HARROWER, MICHAEL. 2003 (November 18–20). Representing Uncertainty: Does it Help People Make Better Decisins? *In: UCGIS Workshop: Geospatial Visualization and Knowledge Discovery Workshop*. invited whitepaper.

HAYES, PATRICK. 1985a. Naive Physics I: Ontology for Liquids. *In:* (Hobbs & Moore, 1985).

HAYES, PATRICK. 1985b. The Second Naive Physics Manifesto. *In:* (Hobbs & Moore, 1985).

HERRING, JOHN R. (ed). 2001. *The OpenGIS Abstract Specification: Topic 1: Feature Geometry (ISO 19107 Spatial Schema)*. Open GIS Consortium. Draft International Standard.

HOBBS, JERRY R. 2000 (April 11). Half Orders of Magnitude. *Pages 28–38 of:* OBRST, LEO, & MANI, INDERJEET (eds), *Proceeding of the Workshop on Semantic Approximation, Granularity, and Vagueness, A Workshop of the Seventh International Conference on Principles of Knowledge Representation and Reasoning KR'2000*.

HOBBS, JERRY R., & KREINOVICH, VLADIK. 2001 (July). Optimal Choice of Granularity In Commonsense Estimation: Why Half-Orders of Magnitude. *Pages 1343–1348 of: Proceedings, Joint 9th IFSA World Congress and 20th NAFIPS International Congress*.

HOBBS, JERRY R., & MOORE, ROBERT C. (eds). 1985. *Formal Theories of the Commonsense World*. Norwood, New Jersey: Ablex Publishing Corp.

HOBBS, JERRY R., & PAN, FENG. 2004. An Ontology of Time for the Semantic Web. *ACM Transactions on Asian Language Processing (TALIP): Special Issue on Temporal Information Processing*, **3**(1), 66–85.

ISO-13249-3. 2003. *ISO/IEC 13249-3:2003: Information technology – Database languages – SQL multimedia and application packages – Part 3: Spatial*. 2nd edn. International Organization for Standardization.

KEUKELAAR, J. H. D., AHLQVIST, O., & OUKBIR, K. 2000 (December 10-13th). Representing Uncertainty in Geographical Data. *In: The 12th Annual Colloquim of the Spatial Information Research Centre*. University of Otago, Dunedin, New Zealand.

KEUKELAAR, JOHANNES, AHLQVIST, OLA, & OUKBIR, ARIM. 1998 (November 16-19th). Using Rough Classification to Represent Uncertainty in Spatial Data. *In: The 10th Annual Colloquim of the Spatial Information Research Centre*. University of Otago, Dunedin, New Zealand.

KUIPERS, BENJAMIN. 2006. An Intellectual History of the Spatial Semantic Hierarchy. *In:* JEFFERIES, MARGARET, & YEAP, ALBERT (WAI-KIANG) (eds), *Robot and Cognitive Approaches to Spatial Mapping*. Springer Verlag.

LENAT, D. 1995. CYC: A Large Scale Investment in Knowledge Infrastructure. *Communications of the ACM*, **38**(11), 32–38.

LENAT, D.B., & GUHA, R.V. 1990. *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*. Reading, Massachusetts: Addison-Wesley.

LIU, JIMING. 1998. A Method of Spatial Reasoning Based on Qualitative Trigonometry. *Artificial Intelligence*, **98**(1-2), 137–168.

MARK, DAVID M. 1999. Spatial Representation: A Cognitive View. *Chap. 7, pages 81–89 of:* LONGLEY, PAUL A., GOODCHILD, MICHAEL F., MAGUIRE, DAVID J., & RHIND, DAVID W. (eds), *Geographical Information System: Principles, Techniques, Applications and Management*, 2 edn., vol. 1. Wiley.

MARK, DAVID M., & SMITH, BARRY. 2004. A Science of Topography: From Qualitative Ontology to Digital Representations. *Pages 75–100 of:* BISHOP, MICHAEL P., & SHRODER, JOHN F. (eds), *Geographic Information Science and Mountain Geomorphology*. Chichester, England: Springer-Praxis.

MASOLO, C., & VIEU, L. 1999a (August). Atomicity vs. Infinite Divisibility of Space. *In:* FREKSA, C., & MARK, D. (eds), *Proceedings of Spatial information theory - Cognitive and computational foundations of geographic information science*.

MASOLO, C., & VIEU, L. 1999b (August). Atomicity vs. Infinite Divisibility of Space. *In:* FREKSA, C., & MARK, D. (eds), *Proceedings of Spatial information theory - Cognitive and computational foundations of geographic information science*.

MASOLO, C., BORGO, S., GANGEMI, A., GUARINO, N., & OLTRAMARI, A. December 2003. *WonderWeb Deliverable D18: Ontology Library*. Tech. rept. Laboratory for Applied Ontology - ISTC-CNR.

MATUSZEK, C., CABRAL, J., WITBROCK, M., & DEOLIVEIRA, J. 2006 (March). An Introduction to the Syntax and Content of Cyc. *In: Proceedings of the 2006 AAAI Spring Symposium on Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering*.

MCMASTER, ROBERT B., & USERY, E. LYNN (eds). 2004. *A Research Agenda for Geographic Information Science*. CRC.

MORATZ, REINHARD, RENZ, JOCHEN, & WOLTER, DIEDRICH. 2000. Qualitative Spatial Reasoning about Line Segments. *Pages 234–238 of: Proceedings, European Conference on Artificial Intelligence (ECAI 2000)*.

MORLEY, DAVID, & MYERS, KAREN. 2004 (July). The SPARK Agent Framework. *Pages 712–719 of: Proceedings of the Third International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS-04)*.

NATIONAL GEOSPATIAL-INTELLIGENCE AGENCY. 1997 (July 4). *Department of Defense World Geodetic system 1984, its Definition and Relationship with Local Geodetic Systems*. 3rd edn. National Geospatial-Intelligence Agency. NIMA Technical Report TR8350.2, updated June 23, 2004.

NATIONAL GEOSPATIAL-INTELLIGENCE AGENCY. 2006 (June 16). *Horizontal Datums*. Web page http://earth-info.nga.mil/GandG/publications/horizdatum.html. Visited September 30, 2006.

NILES, IAN, & PEASE, ADAM. 2001. Toward a Standard Upper Ontology. *Pages 2–9 of:* WELTY, CHRIS, & SMITH, BARRY (eds), *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*. New York, NY, USA: ACM Press.

OBJECT MANAGEMENT GROUP. 2005 (July). *Introduction to OMG's Unified Modeling Language™ (UML®)*. Web page http://www.omg.org/gettingstarted/what_is_uml.htm. Visited September 30, 2006.

ORACLE. 2005a (August). *Oracle Locator: Location-Enabling Every Oracle Database — Technical White Paper*. Oracle. Available from http://www.oracle.com/technology/products/spatial.

ORACLE. 2005b (August). *Oracle Spatial 10g — Technical White Paper*. Oracle. Available from http://www.oracle.com/technology/products/spatial.

ORACLE. 2005c (July). *Oracle Spatial RDF Manual*. Oracle. Available from http://www.oracle.com/technology/products/spatial.

PAN, FENG, & HOBBS, JERRY R. 2004. Time in OWL-S. *Pages 29–36 of: Proceedings of the AAAI Spring Symposium on Semantic Web Services*.

PANTON, K., MATUSZEK, C., LENAT, D.B., SCHNEIDER, D., WITBROCK, M., SIEGEL, N., & SHEPARD, B. 2006. Common Sense Reasoning – From Cyc to Intelligent Assistant. *Pages 1–31 of:* CAI, Y., & ABASCAL, J. (eds), *Ambient Intelligence in Everyday Life*. LNAI, vol. 3864. Springer.

PEARL, JUDEA. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann Publishers, Inc.

PEASE, ADAM, NILES, IAN, & LI, JOHN. 2002 (July 28–August 1). The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. *In: Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*.

PROBST, FLORIAN. 2005. Investigating the Applicability of DOLCE as Foundation for Service Descriptions. *In: W3C Workshop on Frameworks for Semantics in Web Services*.

RANDELL, DAVID A., CUI, ZHAN, & COHN, ANTHONY. 1992. A Spatial Logic Based on Regions and Connection. *Pages 165–176 of:* NEBEL, BERNHARD, RICH, CHARLES, & SWARTOUT, WILLIAM (eds), *KR'92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*. San Mateo, California: Morgan Kaufmann.

SCHLENOFF, CRAIG, GRUNINGER, MICHAEL, TISSOT, FLORENCE, VALOIS, JOHN, LUBELL, JOSHUA, & LEE, JINTAE D. 2000. *The Process Specification Language (PSL): Overview and Version 1.0 Specification*. NIST Internal Report (NISTIR) 6459. National Institute of Standards and Technology, Gaithersburg, MD.

SHEKHAR, S., LIU, X., & CHAWLA, S. 1999. An Object Model of Direction and Its Implications. *GeoInformatica*, **3**(4), 357–379.

SMITH, BARRY, & MARK, DAVID M. 2001. Geographic Categories: An Ontological Investigation. *International Journal of Geographical Information Science*, **15**(7), 591–612.

SMITH, BARRY, & MARK, DAVID M. 2003. Do Mountains Exist? Towards an Ontology of Landforms. *Environment and Planning B: Planning and Design*, **30**(3), 411–427.

SRI INTERNATIONAL. 1995. *Intelligence Preparation of the Battlefield*. Web page http://www.ai.sri.com/~gister/ipb.html. Visited September 30, 2006.

STELL, J. G. 2003a. Granularity in change over time. *Pages 95–115 of:* DUCKHAM, M., GOODCHILD, M., & WORBOYS, M. (eds), *Foundations of Geographic Information Science*. Taylor and Francis.

STELL, J. G. 2003b. Qualitative Extents for Spatio-Temporal Granularity. *Spatial Cognition and Computation*, **3**, 119–136.

STELL, J. G. 2004. Part and Complement: Fundamental Concepts in Spatial Relations. *Annals of Artificial Intelligence and Mathematics*, **41**, 1–18.

STOCK, OLIVIERO (ed). 1997. *Spatial and Temporal Reasoning*. Boston: Kluwer Academic Publishers.

SUMO. 2006. *Suggested Upper Merged Ontology (SUMO)*. Web page http://www.ontologyportal.org/. Visited September 29, 2006.

TALMY, LEONARD. 1983. How Language Structures Space. *Pages 225–282 of:* PICK, JR., HERBERT L., & ACREDOLO, LINDA P (eds), *Spatial Orientation: Theory, Research, and Application*. New York: Plenum Press.

TVERSKY, BARBARA. 2000. Levels and Structure of Cognitive Mapping. *Pages 24–43 of:* KITCHIN, R., & FREUNDSCHUH, S. M. (eds), *Cognitive Mapping: Past, Present and Future*. London: Routledge.

TVERSKY, BARBARA, MORRISON, JULIE B., FRANKLIN, NANCY, & BRYANT, DAVID J. 1999. Three Spaces of Spatial Cognition. *Professional Geographer*, **51**, 516–524.

UCGIS. 2002. *2002 Research Agenda*. Web page http://www.ucgis.org/priorities/research/2002researchagenda.htm. collection of white papers, visited September 8, 2006.

UNITED STATES GEOLOGICAL SURVEY. 1981. *19810501, U.S. Geographic Names Information System (GNIS)*. U.S. Geological Survey, Reston, VA.

UNITED STATES GEOLOGICAL SURVEY. 2005. *View the SDTS Document*. Web page http://mcmcweb.er.usgs.gov/sdts/standard.html. Visited September 30, 2006.

WORBOYS, M. F., & CLEMENTINI, E. 2001. Integration of Imperfect Spatial Information. *Journal of Visual Languages and Computing*, **12**, 61–80.

WORBOYS, MICHAEL. 2005. Event-oriented Approaches to Geographic Phenomena. *International Journal of Geographic Information Systems*, **19**(1), 1–28.

YUAN, MAY. 1996 (January 12-26). Temporal GIS and Spatio-Temporal Modeling. *In: Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling*. Santa Barbara, CA: National Center for Geographic Information and Analysis, Santa Fe, NM. CD, available from http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/main.html.

ZIMMERMAN, KAI, & FREKSA, CHRISTIAN. 1996. Qualitative Spatial Reasoning using Orientation, Distance, and Path Knowledge. *Applied Intelligence*, **6**, 49–58.