Ontological Investigation of Ecosystem Hierarchies and Formal Theory for Multiscale Ecosystem Classifications

Alexandre Sorokine[1], Thomas Bittner[2], and Chris Renschler[3]

[1] Geography Department, University at Buffalo sorokine@buffalo.edu
[2] Institute of Formal Ontology and Medical Information Systems, Leipzig thomas.bittner@ifomis.uni-leipzig.de
[3] National Center for Geographic Information and Analysis (NCGIA) and Geography Department, University at Buffalo, rensch@buffalo.edu

Summary. This paper presents a formalized ontological framework for the analysis of multiscale classifications of geographic objects. We propose a set of logical principles that guide such geographic classifications. Then we demonstrate application of these principles on a practical example of the "National Hierarchical Framework of Ecological Units". The framework has a potential to be used to facilitate interoperability between such geographic classifications.

1 Introduction

Objects in the geographic domain are different compared to objects at surveyable scales like, for instance, table-top objects. Geographic objects are intrinsically tied to space and inherit from space many of its structural properties such as topological, mereological (part–whole relations) and geometric ones [1]. The goal of this paper is to investigate some of the properties of geographic objects using ecosystems as a representative example of a geographic object.

To achieve this goal we will develop a formalized framework for handling of the structure of ecosystem hierarchies. Our theory will be demonstrated using "National Hierarchical Framework of Ecological Units" [2] that is adopted by the United States Department of Agriculture Forest Service (USFS).

The framework that we are developing will be founded on theories of formal ontology including mereology, a theory of classification, and a theory of instantiation. Formal ontology as a tool for studying complex domains has already demonstrated notable successes in such areas of information technology as medical information systems [3, 4] and biological classifications [5]. A similar ontological methodology was applied to the concepts of surface hydrology in [6].

There are several purposes for this study. The study itself will allow to achieve better understanding of the existing ecosystem hierarchies and to provide useful hints for creating new and improved ones. Formalization of the framework and its specification as a small set of axioms is geared toward facilitating of computer–computer interaction and interoperability between various ecosystem hierarchies. Attempts of representing ecosystems and their components on maps and in the computer memory had resulted in a large variety of methods of delineating and classifying ecosystems [7]. The need for interoperating and comparing of ecosystem hierarchies and classifications requires to study basic principles behind the conceptualization of ecosystems and similar objects.

2 Conceptualization of an Ecosystem as a Geographic Object

The notion of an "ecosystem" had long been an integrating concept behind mapping of natural resources such as soils, lands, forests, vegetation and many others and implies close spatial and temporal interrelation between biotic and abiotic map attribute components. The term ecosystem was first introduced by Tansley in 1935 [8] and now is understood as:

> a community of plants and animals within a particular physical environment which is linked by a flow of materials through the non-living (abiotic) as well as the living (biotic) sections of the system... [9, "ecosystem"]

There are several aspects of the notion of an ecosystem that must be considered in order to understand its conceptualization. This includes but not limited to such issues as the question of ecosystem existence, spatial organization of ecosystems, individuation of ecosystems, the nature and the character of ecosystem boundaries, ecosystem dynamics and others.

In geographic space in ontological sense we can speak about two significantly ontologically distinct types of objects: bona-fide objects, i.e., ones that exist in the world and fiat objects, i.e., ones that result from human cognition [10]. Examples of the former are rivers, mountains, highways, the later are hemispheres, countries, property lines. In this sense ecosystems are bona-fide objects and they are not the products of human cognition or artifacts of map-making or management practices. The same is true about ecosystem hierarchy that is an emergent property of the process of ecosystem self-organization.

Ecosystems vary greatly in size and inherently posses a nested structure. One can think of the whole Earth and a drop of water as ecosystems even though those are not disconnected or even separate systems: for example, the flow of energy and matter never stops between the drop of water and remaining Earth. In this sense a drop of water is not a separate ecosystem but rather is a part of the larger Earth ecosystem. This is true also for other ecosystems such as oceans, gullies, lakes, mountain slopes, biomes, etc. that can satisfy the criteria of being of an ecosystem. As a result, ecosystems on the Earth surface form a very sophisticated nested structure with multiple levels of spatial organization [11].

Each ecosystem can be characterized by an infinite number of variables ranging from physical parameters of its environment such as temperature, humidity, pH to the composition of its biological species. Delineation of an ecosystem implies certain homogeneity of its properties within a region of space and time.

Ecosystems are dynamic objects that are individuated by a particular kind of its dynamics. These includes cyclical changes on the scale from hours (like in tidal zones) to centuries (like forest successions) or one-way (acyclic) dynamics like, for example, erosion. Spatial and temporal extents of ecosystems are highly correlated. Typically smaller ecosystems have shorter cycles.

3 Classifications of Ecosystems

United States Department of Agriculture Forest Service (USFS) has adopted “National Hierarchical Framework of Ecological Units” [2] for ecosystem classification and mapping. The main purpose of the framework is to provide support for ecosystem management at various geographic scales. This framework organizes terrestrial ecosystems into eight scale levels depending upon their size and delineation factors (Table 1). Each scale level is intended for a certain group of management and analytical purposes. The size of map units in each scale level typically falls within a certain range of sizes specific for that particular level. The units of each level are nested in the units of an upper level and exhaust them.

Delineation of ecological units at all scales of the hierarchy is based upon the similarity of patterns of a wide range of ecological factors. These factors include climate, lithology, hydrology, landforms and topography, soils, potential vegetation and prevailing natural processes. The combination of factors is specific for each particular scale level. Climate is the leading criteria for delineating of the ecological units at upper scale levels. Other factors such as topography, soils, potential vegetation become more important at the lower scale levels. At the bottom of the hierarchy (scale levels of landscapes and land units) ecological units are designed on the basis of local natural processes and ecological conditions.

3.1 Parts of Ecosystems vs Kinds of Ecosystems

Figure 1 shows a fragment of the map of ecoregions of Alaska [12]. Ecoregions on Figure 1 belong to the “Polar Domain” and represent “Tundra” and

Table 1. National hierarchy of ecological units and corresponfing polygon sizes [2, pages 184 and 186, modified]

Planning and analysis scale		Scale Level	Ecological Unit	Approximate Unit Size, mi^2
Ecoregion				
	Global	1	Domain	10^6
	Continental	2	Division	10^5
	Regional	3	Province	10^4
Subregion		4	Section	10^3
		5	Subsection	$10^1 - 10^3$
Landscape		6	Landtype association	$10^0 - 10^1$
Land unit		7	Landtype	$10^{-1} - 10^0$
		8	Landtype phase	$< 10^{-1}$

"Subarctic" divisions. Relationship between the same ecological units can be presented schematically as shown on Figure 2. At one hand ecological units are involved in part-of relations. The Figure 2(a) shows an abstracted vision of the part-of relationship: if a smaller circle is contained in a bigger one than one is a part of another in the same sense as a smaller ecosystem is a part of a bigger one. It means that we can say that "Tundra Division" and "Subarctic Division" are parts of the "Polar Domain". This will be true in geographic sense due to collocation of these units. This also will be true in ecosystem sense because of particular character of energy and matter exchange between the ecosystem components. On the other hand, each patch of land that is categorized as "Tundra Division" at the same time has to be categorized as "Polar Domain". This relation of categorization we will call an "is-a" (kind-of) relation (Figure 2(b)).

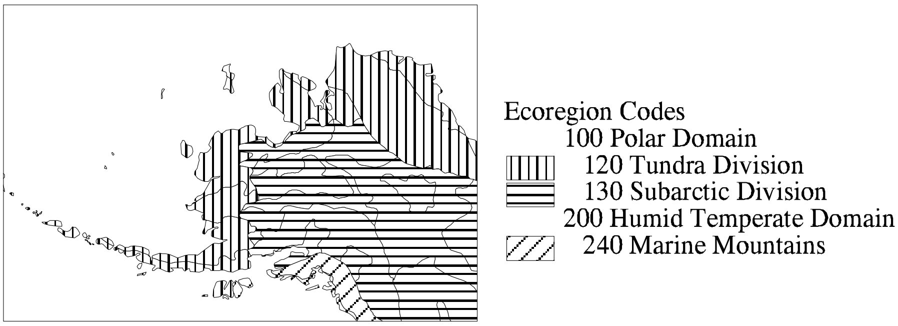
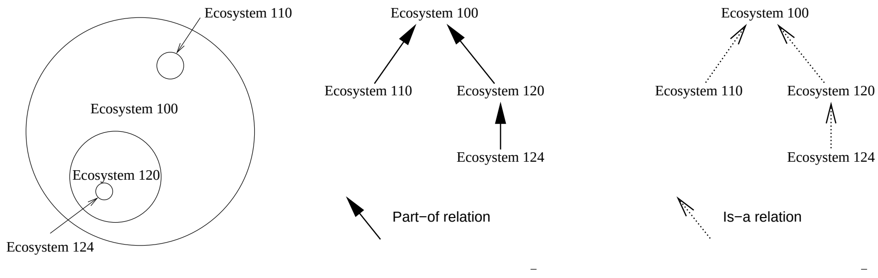


Fig. 1. Ecoregions of Alaska [12, modified]



(a) Part-of relations (b) Is-a relations

Fig. 2. Schematic representation of ecosystems shown on Figure 1

3.2 Classes and Individuals

It is important to make the distinction between a "*part-of*" and a "*is-a*" relation and how they are applied to ecological mapping and classification [13]. To improve understanding, among ecosystems we need to discuss the distinction between classes and individuals in the geographic context.

The distinction between individuals and classes is an obvious and simple in most non-geographic classifications: there is no doubt that my friend's cat Gertrude is a individual and is a very different entity from the biological species[4] "domestic cat" (*Felis catus*). In philosophical terms, individuals are particulars. Particulars can be abstract (like the number two) or concrete (like Gertrude). Concrete particulars are physically existing objects that have specific spatial and temporal locations. For details see for example [14].

Between individuals and classes the relation of instantiation holds. For example, Gertrude instantiates the class *Felis catus*, *i.e.*, Gertrude is an instance of the class *Felis catus*.

However, distinction between individuals and classes is not so obvious in many geographic contexts. For example, in the case of ecosystem classification many classes of ecological units at the upper scale levels have only on a single instance, *e.g.*, class "Polar Domain" has a single discontinuous region on the Earth surface as its instance.

In the special case where classes have only a single instance, the part-of relation, which holds between individuals, coincides with the is-a relation, which holds between classes as shown in the Figures 2(a) and 2(b). Since the is-a relation holds wherever the part-of relation holds the hierarchies induced by both relations are identical. This can be demonstrated using the National Hierarchical Framework of Ecological Units [2]. The units of each level of the framework are nested in the units of an upper level. Such nested structure is

[4] a biological species is a class in the sense the word "class" is used in this paper.

a manifestation of the part-of relationship. At the same time the patches of land that are categorized as units at a certain level of the framework also can be categorized as units of all higher levels. This represents an is-a hierarchy of the ecosystem classes.

Nevertheless the distinction between individuals and classes is still important also in the case of geographic classifications where classes have only a single instance. For example, there were periods in the Earth history when certain ecosystems did not exist. It is possible to imagine that after a few hundred years ice caps will cease to exist due to global warming. However, the class of ecosystems "Icecap Division" which would not have any instances then will still be needed to describe historical data.

There is a clear distinction between the notions of a class and an individual in geographic context at lower scale levels and this distinction is accepted in ecological literature [15, 16]. For example, gullies of a particular region can be classified into certain types using their physical character (deep vs shallow, v- vs u-shape), age, soils, etc. Each type of a gully will receive its name basing on the factors that were used for the classification. At the same time in most cases it would be possible to refer to each particular gully using its local or historical name.

4 A Formal Theory of Classes and Individuals

In this section we will introduce logical theories that are needed to formalize relations behind ecosystem classifications and demonstrate their application using the classification of the ecoregions of the United States (Table 2) as a running example. Formalization of the theories will be presented using first order predicate logic with variables $x, y, z, z_1, \ldots$ ranging over individuals and variables $u, v, w, w_1, \ldots$ ranging over classes. Predicates always begin with a capital letter. The logical connectors $\neg, =, \wedge, \vee, \rightarrow, \leftrightarrow, \equiv$ have their usual meanings: not, identical-to, and, or, 'if ... then', 'if and only if' (iff), and 'defined to be logically equivalent'. We write (x) to symbolize universal quantification and $(\exists x)$ to symbolize existential quantification. Leading universal quantifiers are assumed to be understood and are omitted. The definitions of these terms can be found in any introductory book on symbolic or mathematical logic, for example, [17].

Strict distinction between classes and individuals is one of the cornerstones of our theory. In the sense of this distinction Table 2 can be viewed as either showing a hierarchy of classes or hierarchy between individuals. In our understanding ecoregions, *i.e.*, such entities as "Polar Domain", "Tundra Division" or, for example, "California Dry Steppe Province" play both the roles of an individual and a class. The same distinction is represented schematically on the diagram on Figure 3. Figure 3 shows two parallel hierarchical structures. Individuals (black circles on Figure 3) are connected with each other through

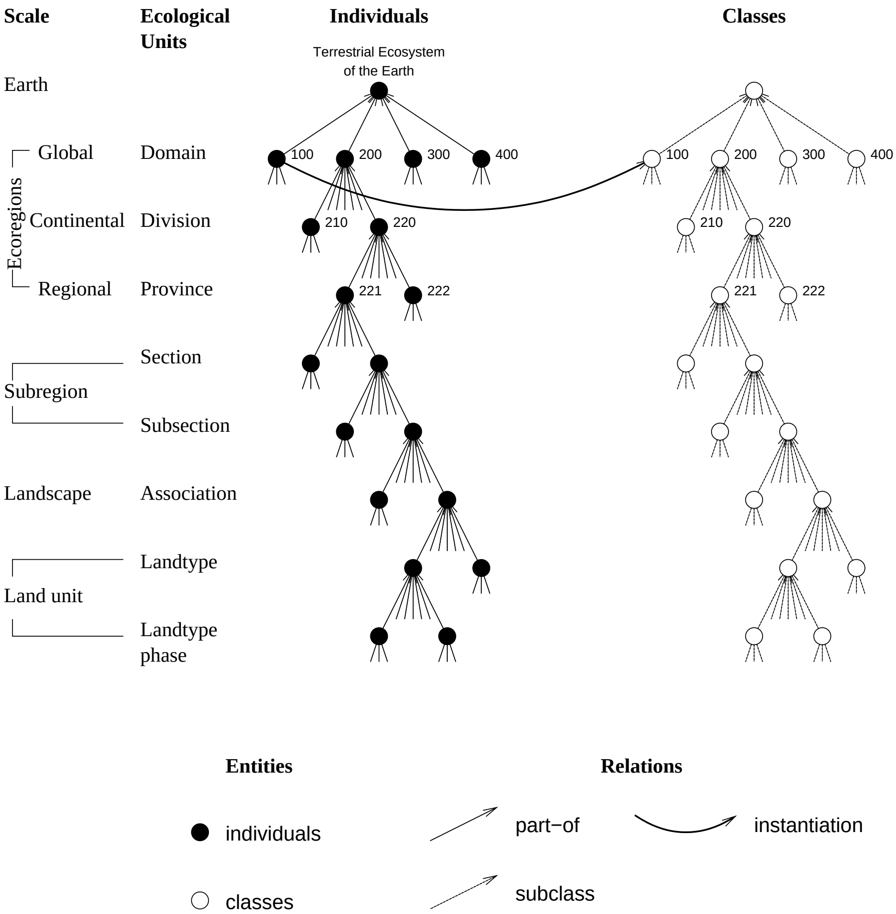


Fig. 3. Individuals and classes in ecoregion taxonomy

the part-of relations (shown as straight solid lines with arrows). This relation will be introduced in Section 4.3. Classes (empty circles on Figure 3) are connected with each other through subclass relation (shown as dotted straight lines). Subclass relation will be discussed in Section 4.1. Classes and individuals are connected between each other through the relation of instantiation (curved solid lines, only a single relation is shown not to overload the diagram). Instantiation will be considered in Section 4.2.

4.1 The Tree Structure of Classes

In this section we will describe properties of the relations between classes. General examples of classes are the class *human being* or the class *mammal*. In our particular study an ecoregion can designate either a class or an indi-

- 100 POLAR DOMAIN
  - 110 Icecap division
  - M110 Icecap Regime Mts
  - 120 Tundra Division
    - 124 Arctic Tundra Province
    - 125 Bering Tundra (Northern) Province
    - 126 Bering Tundra (Southern) Province
  - M120 Tundra Division - Mountain Provinces
    - M121 Brooks Range Tundra–Polar Desert Province
    - M125 Seward Peninsula Tundra–Meadow Province
    - M126 Ahklun Mountains Tundra–Meadow Province
    - M127 Aleutian Oceanic Meadow–Heath Province
  - 130 Subarctic Division
    - 131 Yukon Intermontane Plateaus Tayga Forest
    - 135 Coastal Trough Humid Tayga Province
    - 139 Upper Yukon Tayga Province
  - M130 Subarctic Division - Mountain Provinces
    - M131 Yukon Intermontane Plateaus Tayga–Meadow Province
    - M135 Alaska Range Humid Tayga–Tundra–Meadow
    - M139 Upper Yukon Tayga–Meadow Province
- 200 HUMID TEMPERATE DOMAIN
  - 210 Warm Continental Division
    - 212 Laurentian Mixed Forest Province
  - M210 Warm Continental Division - Mountain Provinces
    - M212 Adirondack-New England Mixed Forest–Coniferous Forest–Alpine Meadow Province
  - 220 Hot Continental Division
    - 221 Eastern Broadleaf Forest (Oceanic) Province
    - 222 Eastern Broadleaf Forest (Continental) Province
  - M220 Hot Continental Division - Mountain Provinces
    - M221 Central Appalachian Broadleaf Forest–Coniferous Forest–Meadow Province
    - M222 Ozark Broadleaf Forest–Meadow Province
  - 230 Subtropical Division
    - 231 Southeastern Mixed Forest Province
    - 232 Outer Coastal Plain Mixed Forest Province
    - 234 Lower Mississippi Riverine Forest Province
  - M230 Subtropical Division - Mountain Provinces
    - M231 Ouachita Mixed Forest–Meadow Province
  - 240 Marine Division
    - 242 Pacific Lowland Mixed Forest Province
  - M240 Marine Division - Mountain Provinces
    - M242 Cascade Mixed Forest–Coniferous Forest–Alpine Meadow Province
  - 250 Prairie Division
    - 251 Prairie Parkland (Temperate) Province
    - 255 Prairie Parkland (Subtropical) Province
  - 260 Mediterranean Division
    - 261 California Coastal Chaparral Forest and Shrub Province
    - 262 California Dry Steppe Province
    - 263 California Coastal Steppe, Mixed Forest, and Redwood Forest Province
  - M260 Mediterranean Division - Mountain Provinces
    - M261 Sierran Steppe–Mixed Forest–Coniferous Forest–Alpine Meadow Province
    - M262 California Coastal Range Open Woodland-Shrub-Coniferous Forest–Meadow Province
- 300 DRY DOMAIN
  - 310 Tropical/Subtropical Steppe Division
    - 311 Great Plains Steppe and Shrub Province
    - 313 Colorado Plateau Semidesert Province
    - 315 Southwest Plateau and Plains Dry Steppe and Shrub Province
  - M310 Tropical/Subtropical Steppe Division - Mountain Provinces
    - M313 Arizona-New Mexico Mountains Semidesert-Open Woodland–Coniferous Forest–Alpine Meadow Province
  - 320 Tropical/Subtropical Desert Division
    - 321 Chihuahuan Semidesert Province
    - 322 American Semidesert and Desert Province
  - 330 Temperate Steppe Division
    - 331 Great Plains-Palouse Dry Steppe Province
    - 332 Great Plains Steppe Province
  - M330 Temperate Steppe Division - Mountain Provinces
    - M331 Southern Rocky Mountain Steppe–OpenWoodland–Coniferous Forest–Alpine Meadow Province
    - M332 Middle Rocky Mountain Steppe–Coniferous Forest–Alpine Meadow Province
    - M333 Northern Rocky Mountain Forest-Steppe–Coniferous Forest–Alpine Meadow Province
    - M334 Black Hills Coniferous Forest Province
  - 340 Temperate Desert Division
    - 341 Intermountain Semidesert and Desert Province
    - 342 Intermountain Semidesert Province
  - M340 Temperate Desert Division - Mountain Provinces
    - M341 Nevada-Utah Mountains Semidesert–Coniferous Forest–Alpine Meadow Province
- 400 HUMID TROPICAL DOMAIN
  - 410 Savanna Division
    - 411 Everglades Province
  - M410 Savanna Division - Mountain Provinces
    - M411 Puerto Rico Province
  - 420 Rainforest Division
  - M420 Rainforest Division - Mountain Provinces
    - M423 Hawaiian Islands Province

Table 2. Ecoregions of the United States [18, 12]

vidual depending upon the context (for discussion see section 3.2). Classes are organized hierarchically by the *is-a* or the *subclass* relation in the sense that a male human being *is-a* human being and a human being *is-a* mammal or, in our example, any patch of land that can be categorized as an "Icecap Division" also can be categorized as a "Polar Domain" (Table 2).

In the present paper the is-a or subclass relation is denoted by the binary relation symbol $\sqsubseteq$ and we use symbol $\sqsubset$ for the proper subclass relation. We will write $u \sqsubseteq v$ to say that class u is a subclass of class v. Also we will call v a superclass of u if the relation $u \sqsubseteq v$ holds (Figure 4).

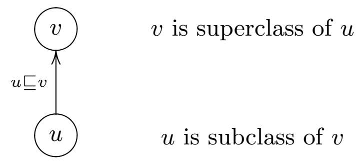


Fig. 4. Denotations of is-a (subclass) relationship

The proper subclass relation is asymmetric and transitive (ATM1–2). It very closely corresponds to the common understanding of the is-a (kind-of) relations:

$$\text{(ATM1) } u \sqsubset v \rightarrow \neg v \sqsubset u$$
$$\text{(ATM2) } (u \sqsubset v \wedge v \sqsubset w) \rightarrow u \sqsubset w$$

Axiom ATM1 postulates that if u is a proper subclass of v then v is not a proper subclass of u. Transitivity (ATM2) implies that all proper subclasses of a class are also proper subclasses of the superclass of that class. In our example (Table 2) class "Arctic Tundra Province" is a proper subclass of "Tundra Division" that in turn is a proper subclass of "Polar Domain". Due to the transitivity of the proper subclass relation we can say that class "Arctic Tundra Division" is also a proper subclass of the class "Polar Domain".

Next we define the relations of subclass $\sqsubseteq$ as $D_{\sqsubseteq}$. Unlike proper subclass, subclass relation allows for a class to be a subclass of itself:

$$D_{\sqsubseteq}\ u \sqsubseteq v \equiv u \sqsubset v \vee u = v$$

Then then can prove that the subclass relation $\sqsubseteq$ is reflexive, transitive, and antisymmetric, i.e., a partial ordering TTM1-3

$$TTM1\ u \sqsubseteq u$$
$$TTM2\ (u \sqsubseteq v \wedge v \sqsubseteq u) \rightarrow u = v$$
$$TTM3\ (u \sqsubseteq v \wedge v \sqsubseteq w) \rightarrow u \sqsubseteq w$$

Class overlap ($O_{\sqsubseteq}$) is defined as $D_{O_{\sqsubseteq}}$:

$$D_{O_{\sqsubseteq}} \quad O_{\sqsubseteq}\, uv \equiv (\exists w)(w \sqsubseteq u \wedge w \sqsubseteq v)$$

Classes overlap if there exists a class that is a subclass of both classes, *e. g.*, in Table 2 class "California Dry Steppe Province" overlaps with the class "Mediterranean Division".

We now add the definition of a root class. A class is a root class if all other classes are subclasses of it (D_{root}):

$$D_{root} \quad \text{root}\ u \equiv (\forall v)(v \sqsubseteq u)$$

In practice in many classifications root classes are not specified explicitly however their existence is implied. For example, Table 2 does not contain a root class but it can be inferred from the context that the root class is "All Terrestrial Ecosystems". We have added this class for consistency to the diagram on Figure 3.

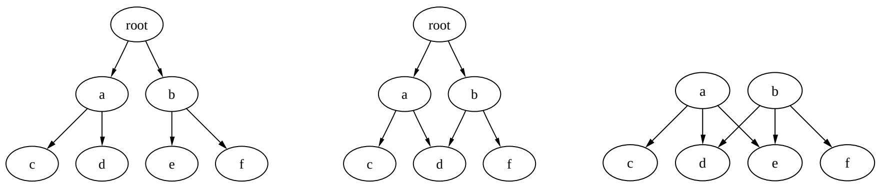
Axioms ATM1-2 ensure that the hierarchy formed by classes satisfies the laws of a partial ordering. Since the hierarchy formed by classes of ecosystems are the result of a scientific classification process we can assume that the resulting class hierarchy forms a tree. We are justified to assume that scientific classifications are organized hierarchically in tree structures since scientific classification employs the Aristotelean method of classification.

As [19] point out, in the Aristotelian method the definition of a class is the specification of essence (nature, invariant structure) shared by all instances of that class. Definitions according to Aristotle's account are specified by (i) working through a classificatory hierarchy from the top down, with the relevant topmost node or nodes acting in every case as undefinable primitives. The definition of a class lower down in the hierarchy is then provided by (ii) specifying the parent of the class (which in a regime conforming to single inheritance is of course in every case unique) together with (iii) the relevant differentia, which tells us what marks out instances of the defined class or species within the wider parent class or genus, as in: human = rational animal, where rational is the differentia. (See also [20] for more details.) This method can be illustrated with an ecoregions example too: "Tundra Division" is a superclass for "Arctic Tundra Province" and "Bering Tundra Province", in this case "Tundra" is the genus and "Arctic" and "Bering" are differentia.

At the formal level we now add axioms that enforce tree structures. These axioms will admit structures of the form shown in Figure 5(a) but will rule out structures shown in figures 5(b) and 5(c). Firstly, we demand that there is a root class (ATM4). Secondly, we add an axiom to rule out circles in the class structure: if two classes overlap then one is a subclass of the other (ATM5). This rules out the structure in Figure 5(b) and also it is very much true for our running example: all overlapping classes in Table 2 are subclasses of each other. We call ATM5 an instance of the *no-partial-overlap principle* (NPO). In more general case it is possible to encounter classifications that involve partial overlap of classes. As it is suggested in [21] in such cases it is often desirable to isolate subclassification which form proper trees. Example of such ecosystem classification was discussed in [22].

Thirdly, we add an axiom to the effect that if u is a proper subclass of v then there exists a class w such that w is a proper subclass of c and w and u do not overlap (ATM6). This rules out cases where a class has a single proper subclass or a chain of nested proper subclasses (ATM6). In the literature on mereology ATM6 is often called the *weak supplementation principle* (WSP) [23].

(ATM4) $(\exists u)\mathrm{root}(u)$
(ATM5) $O_{\sqsubseteq}\ uv \rightarrow (u \sqsubseteq v \vee v \sqsubset u)$
(ATM6) $u \sqsubset v \rightarrow (\exists w)(w \sqsubset u \wedge \neg O_{\sqsubseteq}\ wv)$



(a) Proper tree (b) Overlaps (c) Multiple roots

Fig. 5. Trees (a) and non-trees ((b) and (c)).

Upon inspection the classification in Table 2 violates axiom ATM6 because there are classes that have only a single subclass, for example, "Laurentian Mixed Forest Province" under "Warm Continental Division". We explain this as follows: Firstly, in Table 2 we are dealing only with a portion of a large hierarchy and this portion is constituted of ecoregions that represent only the territory of the U.S. Hierarchies for larger areas would be constituted of a larger number of ecoregions and the ecoregions that in our case have only a single subregions will have more subregions. Secondly, a reason for adding ecoregions with single subregions into a class hierarchy is to simplify navigation of the hierarchy and ensure consistency of the scale levels. For example, if "Marine Division" of "Humid Temperate Domain" is to be omitted then "Pacific Lowland Mixed Forest Province" would immediately follow "Humid Temperate Domain". This would lead to a gap in the scale level of Divisions.

After the introduction of the axioms ATM4–6 we can prove that there exists only a unique root class. We use the symbol $\mathcal{R}$ in order to refer to this class. This rules out the structure shown in Figure 5(c).

4.2 Classes and Individuals

Classes and individuals are related to each other by the relationship of instantiation. This is a binary relation which first parameter is an individual and which second parameter is a class. $\mathit{InstOf}\, xu$ then is interpreted as "individual x instantiates the class u". From our underlying sorted logic it follows that classes and individuals form disjoint domains, *i.e.*, there cannot exist a entity which is a class as well as an individual. Therefore instantiation is irreflexive, asymmetric, and trivially transitive. In terms of our theory each individual (an ecoregion) instantiates a class of the subsection hierarchy (Figure 3): for instance, individual "Temperate Steppe Division" instantiates a class of the same name.

The axioms (AI1+2) establish the relationship between instantiation and the subclass $\sqsubseteq$ relation. AI1 tells us that for all classes that have at least one instance it holds that u is a subclass of v if and only if every instance of u is also an instance of v.

$$\begin{array}{ll} AI1 & (\exists x)(\mathit{InstOf}\, xu) \rightarrow \\ & \quad (u \sqsubseteq v \leftrightarrow (x)(\mathit{InstOf}\, xu \rightarrow \mathit{InstOf}\, xv)) \\ AI2 & (\exists x)(\mathit{InstOf}\, xu \wedge \mathit{InstOf}\, xv) \rightarrow (u \sqsubseteq v \vee v \sqsubset u) \end{array}$$

AI2 guarantees that if two classes share an instance then one is a sub-class of the other.

From (AI1) it follows that two non-empty classes are identical if and only if they are instantiated by the same individuals (TI1).

$$\begin{array}{ll} TI1 & ((\exists x)(\mathit{InstOf}\, xu) \wedge (\exists x)(\mathit{InstOf}\, xv)) \rightarrow \\ & \quad (u = v \leftrightarrow (x)(\mathit{InstOf}\, xu \leftrightarrow \mathit{InstOf}\, xv)) \end{array}$$

It is important to notice that in our framework it is possible to have classes that do not have any instances (empty classes). For example, such classes of ecosystems as "Icecap Division" or "Arctic Tundra Province" may cease to exist due to global warming but our framework will still be able to accommodate them. Empty classes in our theory do not collapse into something like the empty set in set theory, *i.e.*, there does not exist a class which is a sub-class of all classes. This is important since in we need to be able to distinguish the classes "Icecap Division" and "Arctic Tundra Province" even when they are empty.

We say that the class u is the *most specific* class which instantiates the individual x, $\mathit{MSC}\, ux$, if and only if x is an instance of u and every other class which has x as instance has u as sub-class.

$$\begin{array}{ll} D_{MSC} & \mathit{MSC}\, ux \equiv \mathit{InstOf}\, xu \wedge (v)(\mathit{InstOf}\, xv \rightarrow u \sqsubseteq v) \\ TI2 & \mathit{MSC}\, ux \wedge \mathit{MSC}\, vx \rightarrow u = v \end{array}$$

We then can prove that every individual has exactly one most specific class (TI2). This means that we are justified to say that if $\mathit{MSC}\, ux$ holds then u is *the* most specific class of x.

If u is the most specific class of x then we say that x is a *direct instance* of u. For instance, individual "Arctic Tundra Province" instantiates a class of the same name. This class is its most specific class. Also "Arctic Tundra Province" instantiates classes "Tundra Division" and "Polar Domain" that are not its most specific classes. In general case any most specific class can be instantiated by several individuals, in the example of ecoregions some of such classes are instantiated strictly by a single individual.

4.3 The Mereology of Individuals

The properties of the part-of relation between individuals (or instances) can be described using extensional mereology [23]. We use the binary predicate $P\ xy$ in order to say that the individual x is a part of individual y. The relation of parthood can hold between individuals only. In our theory the relation of parthood is primitive, *i.e.*, we do not define it further. It closely corresponds to the intuitive understanding of parthood like in the statements "a hand is a part of a human body" or "California is a part of the U.S". In the example of ecoregions, we can say for instance that "California Dry Steppe Province" is a part of "Mediterranean Division".

We then add axioms to the effect that parthood is reflexive, antisymmetric, and transitive (AM1-3), *i.e.*, a partial ordering.

$$AM1\ P\ xx$$
$$AM2\ (P\ xy \wedge P\ yx) \rightarrow x = y$$
$$AM3\ (P\ xy \wedge P\ yz) \rightarrow P\ xz$$

The first two axioms are rather obvious. Axiom AM1 states that every individual is a part of itself. AM2 suggests that if individual x is a part of individual y and at the same time individual y is a part of individual x then both of these individuals are identical. The third axiom (AM3, transitivity) implies that if x is a part of y and y is a part of z then x is also a part of z. AM3 can be illustrated by the following example: "California Dry Steppe Province" is a part of "Mediterranean Division" that in turn is a part of "Humid Temperate Domain" thus it can be stated that "California Dry Steppe Province" is a part of "Humid Temperate Domain".

Next, in terms of parthood we define the relations of proper parthood and overlap: x is a proper part of y if and only if x is a part of y and x and y are not identical (D_{PP}):

$$D_{PP} \quad PP\ xy \equiv P\ xy \wedge \neg(x = y)$$

Two entities overlap if they share at least one part (D_O):

$$D_O \quad O\ xy \equiv (\exists z)(P\ zx \wedge P\ zy)$$

In our example "California Dry Steppe Province" overlaps with "Mediterranean Division" because they both share a part that is "California Dry Steppe Province".

We then can prove that proper parthood is irreflexive, asymmetric, and transitive, and that overlap is symmetric and non-transitive.

The axioms AM1-3 allow for models where an entity has a single proper part or a finite or infinite sequence of nested proper parts. Such models have to be ruled out because there are no plausible examples that can justify there existence. If we say, for example, that a hand is a proper part of the body than there is always a reminder that is a sum of all other parts of the body.

In order to rule out those models we add an axiom stating that if x is a proper part of y then there exists a z such that z is a proper part of y and x and z do not overlap (AM4). This is another instance of the weak supplementation principle (WSP).

$$AM4 \quad PP\ xy \rightarrow (\exists z)(PP\ zy \wedge \neg O\ zx)\ WSP$$

Finally we define that an entity is the universe of some domain if and only if it has all entities in this domain as parts (D_U). In our case universe is the terrestrial ecosystem of the Earth. We then add an axiom to the effect that such an entity exists (AM5).

$$\begin{array}{ll} D_U & U\ x \equiv (y)(P\ yx) \\ AM5 & (\exists x)U\ x \end{array}$$

We then can prove that this root-entity is unique. We use $\mathcal{U}$ in order to refer to this entity.

4.4 Interaction Between Classifications and Part-Of Structures

At this point it is important to see that the part-of relation is not necessarily a tree structure. For example, the left half of my body and the right half of my body are parts of me which properly overlap, *i.e.*, they share a part (the upper left part of my body) without being parts of each other. However, as we observed above in the case of ecosystems, under certain circumstances the part-of structure of a domain does form a tree. In the case of ecosystems this is due to the fact that there is a quite intimate relationship between the subclass relation among classes, the part-of relation among individuals, and the instance-of relation which connects both domains. To make this relationship explicit we add the following axioms.

First we demand that every individual has a most specific class (AEC1).

$$AEC1 \quad (\exists u)(MSC\ ux)$$

In our context it means that we constrain our domain to individuals which are instantiated by one of the classes of the category tree formed by the ecoregions classification, *i.e.*, each individual instantiates at least one most specific class. From TI2 it then follows that every individual instantiates exactly one most specific class. For example, individual "Tundra Division" instantiates a class with the same name and a class"Polar Domain". In this case class "Tundra Division" is the most specific class of the individual "Tundra Division".

We then add an axiom to the effect that u is a subclass of v if and only if for non-empty classes u it holds that for every direct instance of x of u there exists a direct instance y of v such that x is a part of y (AEC2).

$$AEC2 \quad u \sqsubseteq v \leftrightarrow ((\exists x) MSC\, ux \rightarrow (x)(MSC\, ux \rightarrow (\exists y)(MSC\, vy \wedge P\ xy)))$$

AEC2 ensures if a class is instantiated by some entity x then each of its superclasses is instantiated by an entity y which has x as a part. Consequently, if a class is non-empty then all of its superclasses are non-empty. Moreover, if for a non-empty class u all of its instances are parts of instances of some class v then u is a subclass of v.

This can be illustrated by the following example: individual "Arctic Tundra Province" instantiates a class "Arctic Tundra Province" (its most specific class) and all its superclasses ("Tundra Division" and "Polar Domain"). At the same time individual "Arctic Tundra Province" is a part of the individuals "Tundra Division" and "Polar Domain".

In the case of ecosystem classification classes of ecological units at the upper scale levels (Global, Continental, Regional and Subregional scales in Table 1) have only a single instance, *e.g.*, class "Polar Domain" has a single discontinuous region as its instance. For upper scale levels of the ecoregion classification it holds that every class contains at most one most specific individual (AEC$_{UL}$).

$$AEC_{UL}\ MSC\, ux \wedge MSC\, uy \rightarrow x = y$$

AEC$_{UL}$ is not necessarily true for lower levels of ecosystem hierarchy in Table 1 or other ecosystem classifications.

Using AEC2 and AEC$_{UL}$ we then can prove that the sub-class structure among classes corresponds to the part-of structure among individuals, *i.e.*, if u is the most specific class for x and v is the most specific instance for y then u is a subclass of v if and only if x is a part of y.

$$TEC1\ MSC\, ux \wedge MSC\, vy \rightarrow (u \sqsubseteq v \leftrightarrow P\ xy)$$

We are now able to show that under the assumption that AEC1-2 and AEC$_{UL}$ hold, the part-of hierarchy at the level of individuals forms a tree. To see this consider the following. If we compare the axioms for the tree structure of the class hierarchy imposed by ATM1-6 (tree-mereology) and the axioms of general extensional mereology (AM1-5) then we see that AM1-3 correspond to TTM1-3, *i.e.*, both structures are partial orderings. AM4 in mereology and ATM6 in tree mereology are both instances of the weak supplementation system. The axiom ATM4 ensuring the existence of a root-class corresponds to axiom AM5 ensuring the existence of the maximal individual, the universe. Using AEC1, ATM5, and TEC1 we then can prove that if two entities overlap then they are parts of each other (TEC2).

$$TEC2 \quad O\ xy \rightarrow P\ xy \vee PP\ yx$$

TEC2 theorem is another instance of the no-partial-overlap principle and corresponds to axiom ATM4 of tree mereology.

Consequently, under the assumptions AEC1-2 and AEC$_{UL}$ the subsumption relation $\sqsubseteq$ among classes and the part-of relation among individuals satisfy equivalent sets of axioms. It follows that both form tree structures. It now remains to study the relationships between those tree structures.

We start by proving that the root class in the class hierarchy, $\mathcal{R}$, instantiates the maximal individual $\mathcal{U}$ (TEC3 proved using AEC1 and AEC2).

$$TEC3 \quad MSC\, \mathcal{RU}$$

We now introduce a mapping h which maps every individual to its respective most specific class.

$$D_h \quad h\ x = u \equiv MSC\, ux$$

We can define h in this way since: (i) from TI2 we know that every individual has *at most one* most specific class; (ii) from ACE1 we know that every individual has *at least one* most specific class. Consequently, h is as mapping well defined. For example $(h\ Tundra\_Division) =$ "$Tundra\_Division''$.

Given the definition of the mapping h then AEC$_{UL}$ tells us that the inverse of h, h^{-1}, which returns an instantiating individual when given a class as input is a (possibly partial) mapping too. For example $(h^{-1}$ "$Tundra\_Division'') = Tundra\_Division$. h^{-1} is possibly partial since there might be empty classes for which h^{-1} cannot yield an instantiating individual. An example of an empty class for which h^{-1} cannot return anything would be the class "Icecap Division" during the geologic periods when Earth surface was warm enough to prevent this form of glaciation. For non-empty classes, however, AEC$_{UL}$ ensures that there is exactly one instance which then is returned by h^{-1}.

From theorem TEC1 it then follows that if the individual x is a part of the individual y, $P\ xy$, then the most specific class of x, $(h\ x)$, is a subclass of the most specific class of y, $(h\ x)$. This tells us that the mapping h is is an order-*homomorphism*, *i.e.*, a mapping which preserves the partial ordering structure: if $P\ xy$ then $(h\ x) \sqsubseteq (h\ y)$.

We now distinguish two cases: (i) all classes have instances and (ii) there are empty classes in the class hierarchy. In the former case it follows from TEC1 that h is an order-*isomorphism*. This means that it is does not only hold that the mapping h is an order-homomorphism but also its inverse, h^{-1}. The mapping h^{-1} takes a class as input and returns its single direct instance, which by assumption (i) always exists. That h^{-1} is an order-homomorphism then means that if the class u is a subclass of v, then the direct instance of u is a part of the direct instance of v, *i.e.*, if $u \sqsubseteq v$ then $P\ (h^{-1}u)(h^{-1}v)$. Consequently, if we ignore the distinction between a class having a single instance and the instance itself then both structures are indistinguishable.

In the second case (ii), where there are empty classes, h is a *partial* order-isomorphism since the inverse of h, h^{-1}, is not defined for empty classes. If we, again, ignore the distinction between a class having a single instance and the instance itself then the tree formed by the instances is a subtree of the tree formed by the classes.

Existence of h and h^{-1} homomorphisms is likely the main reason why in many geographic classifications of global and regional scales the distinction between classes and individuals is not clearly outlined. In most cases mixing hierarchies of classes with hierarchies of individuals does not prevent particular geographic classifications from achieving their goals. However, distinguishing these two hierarchies as separate entities allows for better understanding of the process of geographic classification and has a potential for improving interoperability between classifications, scale levels, regions and datasets.

5 Conclusions

In this paper we have demonstrated how principles of formal ontology can help to clarify the semantics of ecosystem hierarchies. We applied those principles in an analysis of ecosystem hierarchies of the "National Hierarchical Framework of Ecological Units" [2]. Our analysis showed that it is important to strictly distinguish between the notion of a class and the notion of an individual. Our formal framework was created around three types of relations: subclass (kind-of, is-a), part-of and instance-of. Subclass relation can be defined between classes and is reflexive, antisymmetric, and transitive. Classes are organized into classifications that form trees. Part-of relations can be defined between individuals and it is also reflexive, antisymmetric, and transitive. We have demonstrated that in some cases trees of classes and individuals can be indistinguishable.

Even though most of the operations in our approach would seem obvious for geographers and ecologists, such operations have to be outlined explicitly if the goal of interoperation of two datasets is to be achieved or the information contained in classifications is to be communicated to non-experts in the area. For example, to interoperate ecosystem classifications for two different regions one must first mark out classification trees according to the rules outlined in Sect. 4.1. Then the trees from these two classifications have to be combined into a single tree that also must satisfy axioms ATM1–6. Strict distinction between the trees of classes and the trees of individuals in our theory would allow to accommodate changes in the environment and classification systems. The theory presented in this paper was successfully tested against other kinds of geographic classifications such as regional ecosystem hierarchies and soil taxonomy [22, 24].

Acknowledgements

Support for the second author from the Wolfgang Paul Program of the Alexander von Humboldt Foundation and from the National Science Foundation Research Grant BCS–9975557: Geographic Categories: An Ontological Investigation, is gratefully acknowledged. The authors also acknowledge the helpful suggestions of the anonymous reviewers of this paper.

References

1. David Mark, Barry Smith, and Barbara Tversky. Ontology and geographic objects: An empirical study of cognitive categorization. In C Freksa and David Mark, editors, *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science*, number 1661 in Lecture Notes in Computer Science, pages 283–298. Springer, 1999.
2. David T. Cleland, Peter E. Avers, W. Henry McNab, Mark E. Jensen, Robert G. Bailey, Thomas King, and Walter E. Russell. National hierarchical framework of ecological units. In Mark S. Boyce and Alan W. Haney, editors, *Ecosystem Management Applications for Sustainable Forest and Wildlife Resources*, pages 181–200. Yale University Press, New Haven, CT, 1997.
3. Alan Rector and Jeremy Rogers. Ontological issues in using a description logic to represent medical concepts: Experience from galen. In *IMIA WG6 Workshop: Terminology and Natural Language in Medicine*, Phoenix, Arizona, 11 1999. International Medical Informatics Association.
4. Barry Smith and Cornelius Rosse. Is_a and part_of: the role of formal definitions in the alignment of biomedical ontologies. *in press*, 2003.
5. Barry Smith. The logic of biological classification and the foundations of biomedical ontology. In *Invited Papers from the 10th International Conference in Logic Methodology and Philosophy of Science*, Oviedo, Spain, 2004. Elsevier-North-Holland.
6. Chen-Chieh Feng, Thomas Bittner, and Douglas Flewelling. Modeling surface hydrology concepts with endurance and perdurance. In *GIScience 2004*, 2004.
7. M F Goodchild, B O Parks, and L T Steyaert, editors. *Environmental Modeling with GIS*. Oxford University Press, New York, 1993.
8. A G Tansley. The use and abuse of vegetational concepts and terms. *Ecology*, 16(3):284–307, July 1935.
9. Susan Mayhew. *A Dictionary of Geography*. Oxford University Press, oxford reference online edition, 1997. 13 September 2003 `http://www.oxfordreference.com/views/ENTRY.html?subview=Main\&entry=t15%.001014`.
10. Barry Smith and Achille C Varzi. The formal ontology of boundaries. *The Electronic Journal of Analytic Philosophy*, 5(5), spring 1997. `http://ejap.louisiana.edu/EJAP/1997.spring/smithvarzi976.html`.
11. Chris S Renschler and Jon Harbor. Soil erosion assessment tools from point to regional scales — the role of geomorphologists in land management research and implementation. *Geomorphology*, 47:189–209, 2002.

12. Robert G. Bailey. Description of the ecoregions of the united states. Technical report, USDA Forest Service, March 1995. `http://www.fs.fed.us/land/ecosysmgmt/ecoreg1_home.html`.
13. Gregory Nowaki and Alexandre Sorokine. The limitations of applying single-resource taxonomies to ecological partonomies. In Randy Davis and Curtis Talbot (co chairs), editors, *National Cooperative Soil Survey Conference Proceedings, Committee 2: Ecological Interpretations and Principles*, pages 17–18, Plymouth, MA, June 2003. USDA, National Resources Conservation Service, Supplement Report.
14. R Casati and A C Varzi. The structure of spatial localization. *Philosophical Studies*, 82(2):205–239, 1995.
15. Gerhard Wiegleb. Explanation and prediction in vegetation science. *Vegetatio*, 83:17–34, 1989.
16. J Stan Rowe and John W Sheard. Ecological land classification: a survey approach. *Environmental management*, 5(5):451–464, 1981.
17. Irving M Copi. *Symbolic Logic*. Pearson Education, 5 edition, 1979.
18. Robert G. Bailey. *Ecosystem geogrpahy*. Springer, 1995.
19. Barry Smith, Jacob Khler, and Anand Kumar. On the application of formal principles to life science data: A case study in the gene ontology. In E Rahm, editor, *Database Integration in the Life Sciences (DILS 2004)*. Springer, Berlin, 2004. forthcoming.
20. Harold P Cook and Hugh Tredennick. *Aristotle: The Categories, On Interpretation, Prior Anlytics*. Harvard University Press, Cambridge, Massachusetts, 1938.
21. Thomas Bittner, Maurene Donnelly, and Barry Smith. Individuals, universals, collections: On the foundational relations of ontology. In *International Conference on Formal Ontology in Information Systems*, Torino, Italy, 11 2004.
22. Alexandre Sorokine and Thomas Bittner. Understanding taxonomies of ecosystems: a case study. In *Spatial Data Handling*, pages 560–572, 2004. (in press).
23. Peter Simons. *Parts: A Study in Ontology*. Oxford University Press, October 2000.
24. Chris Renschler, Alexandre Sorokine, and Thomas Bittner. Soil mapping and ontology. 2004. forthcoming.