



AALBORG UNIVERSITY
DENMARK

Aalborg Universitet

Integrating ontologies

Stubkjær, Erik

Published in:

ScanGIS'2001 : Proceedings of the 8th Scandinavian Research Conference on Geographical Information Science, June, 25-27, 2001, Ås, Norway

Publication date:

2001

Document Version

Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Stubkjær, E. (2001). Integrating ontologies: Assessing the use of the cyc ontology for cadastral applications. In Bjørke, Jan Terje : Tveite, Håvard (eds.) (Ed.), ScanGIS'2001 : Proceedings of the 8th Scandinavian Research Conference on Geographical Information Science, June, 25-27, 2001, Ås, Norway (pp. 171-184). Norway.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- ? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- ? You may not further distribute the material or use it for any profit-making activity or commercial gain
- ? You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Integrating Ontologies: Assessing the use of the Cyc Ontology for Cadastral Applications

Erik Stubkjær

Dept. of Development and Planning, Aalborg University,
Fibigerstræde 11, DK 9220 Aalborg East, Denmark
est@i4.auc.dk

Abstract Most GISs uses system specific data models and databases. Although standardisation efforts have been made during decades, the integration of different data models is still a research issue. Ontologies are developed to provide an interface between diverse domains or universes of discourse. For example, ontological engineering is applied for establishing standards in the field of Geographic Information, as appears from the EuroConference on the Ontology and Epistemology for Spatial Data Standards, September 2000.

The paper introduces shortly ontological engineering as well as the Cyc ontology. The Cyc ontology is one of the few general-purpose or 'universal' ontologies that are publicly available. The Cyc ontology is reviewed from the point of view of a specific, yet multi-disciplinary knowledge domain, namely that of cadastral studies. The review provides the basis for an assessment of the usability of the Cyc ontology, and for the suggestion of further research.

1 Introduction

There are established (and ever developing) methods for the analysis and design of information systems, and more specifically for establishing data models to map the content of databases. Other descriptions of the content of databases exist in terms of catalogues of meta-data that describe the content, quality and accessibility of data. However, most data modelling approaches fail to capture all aspects of the information that is represented by the data. Often hidden assumptions and silent agreements as well as dependencies between different kinds of data are missed. Beyond this, the integration of different data models is still a research problem (Vekovski, 1998; Goodchild et al, 1999; Kottman, 1999).

Ontologies are developed to improve consistency and completeness of specifications, and to provide an interface between diverse domains or universes of discourse. For example, ontological engineering is applied for establishing standards in the field of Geographic Information, as appears from the EuroConference on the Ontology and Epistemology for Spatial Data Standards (EuroConference, 2000).

Ontological engineering developed during the 1990s from artificial intelligence and computational linguistics. Both fields depend on knowledge representation and knowledge engineering. A frequently quoted definition of ontology is the one given by Gruber (1993): "An Ontology is an explicit specification of a conceptualization." Recently, Bishr et al address the issue of interoperability between GISs. They suggest

the following definition: "A geospatial information community is a group of spatial data producers and users who share an ontology of real-world phenomena", and consider the ontology as "a meta-language situated above data models" (1999: 58). Two communities may have different ontologies, but in order to share information they must have a part of their ontologies in common.

It should be noted that the term ontology is used, also, by philosophers and by linguists. In "Semantics and Cognition" from 1983 Ray Jackendoff discuss how visual information, linguistic information and other peripheral information is mapped onto mental representations. Through an analysis of human perception, he arrives at the following list of ontological categories: THING, PLACE, DIRECTION, ACTION, EVENT, MANNER, and AMOUNT. The list is not meant to be complete. He claims, however, that "the total set of ontological categories must be universal: it constitutes one basic dimension along which humans can organize their experiences." (Jackendoff, 1983: 56; Stubkjær, 1994: 582).

Many ontology modelling languages have been developed during the past decade (Corcho & Gomez-Perez, 2000) that can be used to implement ontologies. However, not all of these languages are suited for real-life applications. Some languages are too expressive to support automatic reasoning and many rely on a specialised run-time environment. Recently, the world-wide web standardisation committee W3C has recommended languages for data and meta-data modelling. The proposed XML may be considered a candidate language for data modelling on the world-wide web. Stuckenschmidt and Wache (2000) propose the use of the language OIL (Ontology Interchange Language, 2000) in order to achieve semantic interoperability between heterogeneous information sources. The OIL integrates XML with a logical model that makes automated inference possible.

Ontology development

The first step in ontological engineering is the definition of the purpose the ontology is built for. This step is necessary, because the purpose always influences the way certain concepts are modelled. In knowledge engineering this phenomenon is known as the "Interaction Problem" (Bylander & Chandrasekaran, 1988). The purpose that is assumed in this paper, is information modelling. More specifically it regards cadastre as a special purpose information system that is used for legal and administrative purposes, as developed in cadastral studies.

Having specified the domain of interest, the second step of the ontological engineering process is concerned with the actual modelling of that domain. This step can be further subdivided into three activities (Ushold & King 1995):

1. Ontology Capture
2. Integration of Existing Ontologies
3. Ontology Coding

In the first step, important concepts and relations of the domain of discourse are identified and drafted in order to get an overview of the domain and ensure completeness of the resulting model. A contribution to this step has already been done by (Stubkjær 1996). The step of integrating existing ontologies is essential in order to reduce the modelling effort. The exploring of an existing ontology (cf. item 2 above)

was introduced by Stubkjær and Stuckenschmidt (2000) in order to identify concepts that could be part of an ontology of the cadastral domain. The following continues this paper by presenting - from a cadastral point of view - a rather complete review of the part of the Cyc ontology that is publicly available. The coding step (cf. item 3) will not be further developed here.

Closing the introductory section, it should be mentioned that the above approach is not the only one available for ontology development. To give but one example, Werner Kuhn at the above mentioned EuroConference suggested a method to derive ontologies from natural language texts describing the domain, cf. item 1 above (Kuhn, 2000).

2 The Cyc ontology

The company Cycorp, Inc. has published on the Internet the topmost constituents of the Cyc® ontology. The Cyc ontology is claimed to be "universal" in the sense that "every concept one can imagine can be correctly linked into the Upper Cyc® Ontology in appropriate places, no matter how general or specific, no matter how arcane or prosaic, no matter what the context (nationality, age, native language, epoch, childhood experiences, current goals, etc.) of the imaginer". For each of approximately 3,000 concepts or *constants* from the Cyc® Knowledge Base is listed

1. its Cyc® name,
2. an English comment on the intended meaning and use of the concept, and
3. a few of the taxonomic "links" which Cyc® uses to hierarchically order and interconnect its concepts.

By convention, the name of every Cyc® constant begins with the characters #\$.; for example, #\$RealEstate. A constant can represent a collection (such as the set of all people), an individual object (such as a particular person), a word in a natural language (such as the English word 'person'), a quantifier (such as 'there exists'), a relation (a predicate, function, slot, attribute, etc.), and so on.

Each entry starts with the name of the concept, off to the left and emboldened. Next comes an English comment, to help clarify the meaning and intended use of the constant. Then comes some hierarchical information provided as hypertext links: sets that this concept is an element of, supersets of it, etc. The following example is copied directly from the Cyc ontology (a discussion of the comment, etc. is deferred to a subsequent section):

#\$RealEstate

A collection of tangible objects. Each element of #\$RealEstate is either a parcel of land or a land-based property that can be bought, sold, or rented. This includes buildings and parts of buildings such as office suites or condominiums, as well as parcels of land. Some prominent examples: #\$GuantanamoNavalBase, #\$NewYorkHiltonAtBroadway, #\$WorldTradeCenter.

isa: #\$ProductType #\$ExistingObjectType

gens: #\$SolidTangibleProduct

some subsets: #\$Building #\$GroundsOfOrganization #\$RailroadStation-Physical

[#\\$ThreeStoryBuilding](#) [#\\$TwoStoryBuilding](#) [#\\$OneStoryBuilding](#)
[#\\$ModernShelterConstruction](#) [#\\$ModernHumanResidence](#) [#\\$SingleResidenceUnit](#)
 (plus 1 more public subset, 80 unpublished subsets)

Every Cyc@ constant is an element of one or more collections. #RealEstate, e.g., is a member of the set #ExistingObjectType. In each entry, there will be an isa: line, which lists some of the collections of which this constant is an element. To avoid clutter, generally only non-redundant series of collections are listed.

#RealEstate is itself a collection (the set of all units of real estate); it can therefore have supersets and subsets and elements. The gens: line of the #RealEstate entry lists a superset of #RealEstate, #SolidTangibleProduct. Other of Cyc's constants may have more generalisations, but again: generally only non-redundant series of collections are listed.

The Cyc constants are organised into 43 topical concept files, which are named according to their content, e.g. Top vocabulary, Agent vocabulary, etc. Of these 43 files, concepts from the following 11 files are considered relevant for presentation here:

Top vocabulary	
Things	Events
Agent vocabulary	Time vocabulary Possession transfer vocabulary
Organizational vocabulary	
Social vocabulary	
Information vocabulary	
Space vocabulary	
Mathematical vocabulary	
Geography vocabulary	
Construction vocabulary	
Device vocabulary	

The following review of the files' content is structured into four sections, three of which concern time-invariant things: agents with organisations and professionals (section 3), information with communication (4), and the geospatial domain (5), while the dynamic aspects are addressed in section 6. It should be noted, however, that neither in the Cyc files, nor in the following presentation, the distinction between things and events is clear-cut. After the reviews, section 7 offers a discussion, and section 8 concludes the paper.

3 Agents, professionals, and owners

The Cyc ontology offers a graph to illustrate the generalisation relation (sub/superset) among entities, which can show individual action (see Figure x below). In the cadastal context, the #LegalAgent plays an important role. A legal agent is one that has some legal status in a particular legal system, e.g. being a citizen of a

specific country. "Instances of #LegalAgent include agents that may have property rights, may be taxed, may have a government identification number, may be sued, may have an address, etc."

Similarly, the set of #Professional is highly relevant. According to Cyc, "elements of #Professional are agents who spend a significant part of their waking hours doing activities that are characteristic of some occupation, skilled or unskilled... Typically their actions are performed for pay, but not always". This definition may be questioned, because often the criteria for a professional include that the professional has a certified or even higher education.

The Cyc definition of #IntelligentAgent mentions and includes intelligent machines. However, computers, etc. seem - strangely enough - not to be among the entities of the Cyc vocabulary.

In a cadastral context, organisations generally would be a subset of #LegalAgent, but Cyc specifies #Organization to be a collection, which "includes both informal and legally constituted organizations". Cyc defines an entity, #LegalCorporation, that is a subset of both #Organization and #LegalAgent, but this is not shown or referred to by Figure 1.

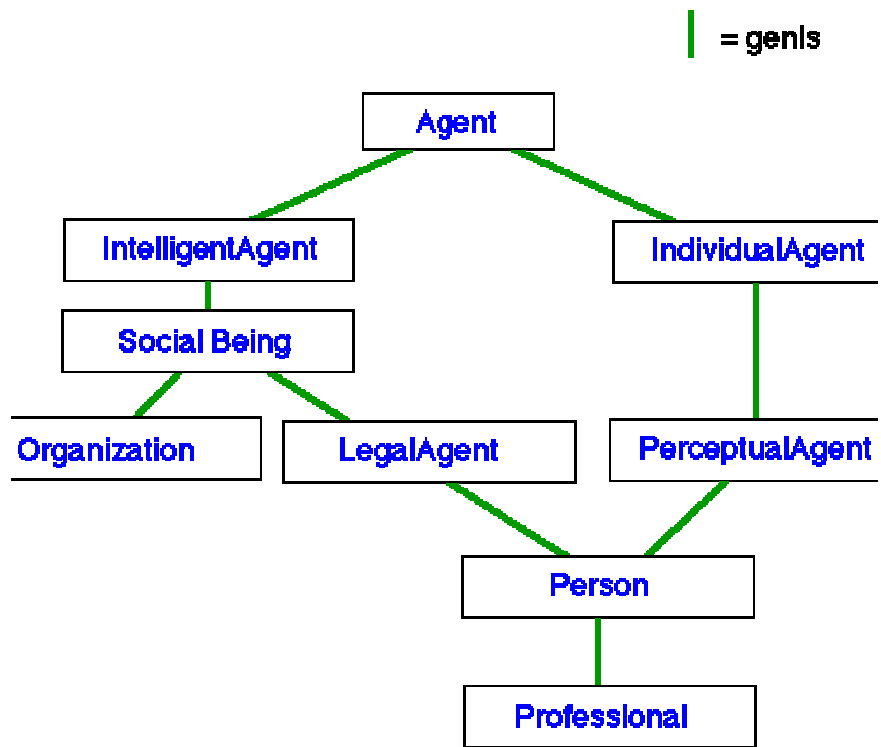


Figure 1: A hierarchical view of #Agent and its subsets
 (Source: Cycorp, Inc., www.cyc.com; Copyright© 1996, 1997).

It is a challenge to visualise the mutual relations among the Cyc entities. For example, it would be possible to visualise subsets of `#IntelligentAgent` that are not included in `#SocialBeing`, e.g intelligent machines, but a systematic rendering of all subsets of the entities of the figure would hardly be usefull. The Cyc ontology permits other structuring principles than generalisations, for example predicates and functions, and taking these into consideration would make the visual rendering a still harder task. Consequently, the following sections will refrain from using graphics, but render main generalisation relations through the relative location of cells of a table.

Ownership is modelled in Cyc through a predicate:

```
#$owns : <#$Agent> <#$SomethingExisting>.
```

The predicate (`#$owns AGENT OBJECT`) means that AGENT owns OBJECT -- that is, AGENT enjoys `#$FullUseRights` over OBJECT. Less than full ownership is modelled by the predicate:

```
#$hasOwnershipIn : <#$Agent> <#$SomethingExisting>.
```

or by a general predicate on rights:

```
#$userRightsRelation :
```

```
<#$Agent> <#$PartiallyTangible> <#$UserRightsAttribute>.
```

The `#$userRightsRelation` predicate relates an agent to the use rights, which that agent has over an object. The predicate (`#$userRightsRelation AGENT OBJECT URA`) means that AGENT enjoys the `#$UserRightsAttribute URA` over OBJECT. For example, if AGENT owns OBJECT, then AGENT has `#$FullUseRights` over it.

It should be noted that Cyc does not establish a specified relation between the general rights predicate: `#$userRightsRelation` and the more specific: `#$owns`. This may illustrate the need of specifying generalisation relations not only between sets, but also between predicates. In fact, Cyc includes a predicate-type, `#$MetaPredicate`, that is a collection of predicates. Each element of `#$MetaPredicate` can be used to define other predicates, but this potential seems not exploited.

Natural language provides us with shorthand notations, like owner, mortgagor, mortgagee, etc., for instances of the predicates that are mentioned above. A non-public Cyc[®] lexicon exists that currently spans over 14,000 root words in English, containing the usual sorts of linguistic information about words, plus their denotations as Cyc[®] constants and expressions. Details on the structure of such lexicon and the use of it in the ontology development process is not pursued further here.

4 Information carriers

The cadastral domain includes a lot of 'paper-work'. The paper carries information, and the corresponding root entity of the Cyc ontology is `#$InformationBearingThing`. A problem with the information vocabulary of the Cyc ontology is that the root entity has a long list of subsets, which are only weakly related. The list includes the following:

- #\$Map
- #\$TextualMaterial
- #\$StructuredInformationSource
- #\$VisualInformationSource
- #\$InformationBearingWavePropagation
- #\$InformationBearingObject
- #\$NonPublishedText
- #\$ReferenceWork
- #\$OrganizationalChart
- #\$Form-StandardizedIBO
- #\$Sculpture
- #\$Gesture

The various subsets are discerned by attributes, which for the present purpose may be categorised as follows:

- the structure of information
- the information extraction (reading) method, and
- the meaning of the information

The structure of information may be linear (#\$TextualMaterial) or non-linear (#\$Map); it may be formalised (#\$StructuredInformationSource) or non-formalised, propositional or non-propositional (#\$ArtObject, some of #\$VisualInformationSource), and finally simple (#\$VisualMarkers) or compound (#\$StructuredInformationSource).

The information extraction method discerns whether information is accessible by viewing it without using tools (except possibly spectacles), e.g. #\$VisualInformationSource and #\$HardcopyInformationBearingObject. Information in terms of #\$InformationBearingWavePropagation may be accessible without using tools, e.g. as sound or as images of visible light, while radio waves demands tools to be accessible.

Finally, the meaning of information is used to distinguish between #\$SymbolicObjects (e.g. related to nation or religion), #\$ArtObjects, #\$OfficialDocuments, and other information carrying objects.

The entries of the Information vocabulary of the Cyc ontology are too heterogeneous and mutually too loosely related to be applied for the cadastral domain. In the following, a relevant subset is presented that benefit from the general hierarchy of the Cyc ontology, and presumably can be formally specified:

<u>#\$InformationBearingThing</u>				
<u>#\$VisualMark</u> , e.g. boundary marks, geodetic point markers, street signs	<u>#\$StructuredInformationSource</u>			
	<u>#\$TextualMaterial</u>			
	<u>#\$Map</u>	<u>#\$OfficialDocument</u>	<u>#\$Form-Standardized IBO</u>	<u>#\$Computer Program</u>

It should be noted that the Cyc ontology includes a set of terms, which relate to the information content, rather than to the things that carries information. To give some examples, the collection #PropositionalInformationThing is an abstract entity with subsets including #IllocutionaryForce, #Request, #Offer, and #Promise. The collection #PropositionalInformationThing is itself a subset of #Microtheory. This part of the Cyc ontology ought to be explored to catch the details of #PurposefulAction of the cadastral domain, but the task is deferred here.

5 Space, terrain, and property boundaries

The Cyc ontology subsumes spatial aspects under the entity #SpatialThing. #SpatialThings divide into the subsets #PartiallyTangible and #Intangible. A subset of the latter is #GeometricalThing with further subsets #FrameOfReference and #TerrestrialFrameOfReference.

It is interesting to note the Cyc description of the subset #PartiallyTangible. The elements of this subset "have some tangible (i.e., material) part and also have a temporal extent (i.e., they exist in time). They may or may not also have an intangible part; e.g., a book is made of matter, has a temporal extent, and also has intangible content which is the information content of the text that the author wrote." This combination of tangible and intangible part applies very well to a unit of real estate, or - more precisely - to the borders of this unit, as they consist of an intangible part that corresponds to the information content of the book. To give an example: a hedge (tangible part) may or may not indicate the property boundary. Whether the hedge is actually a property boundary is derived from the information content of documents, application of legal rules, etc (intangible part).

An important subset of #PartiallyTangible is #Place. Every #Place "is stationary in the frame of reference of the current microtheory." In a cadastral context, a place would need a place name, but in Cyc #Place has a more abstract denotation and need not be confined to the surface of the Earth. Similarly, #Border refers to lines, planes, etc. that constitutes a boundary between two elements of #SpatialThing, that is an abstract entity, and there is no specification of the boundary between two elements of #RealEstate.

The binding to the surface of the Earth is achieved by #GeographicalRegion and its subsets. The subset #GeopoliticalEntity is interesting, as it has two generalisations (supersets): #GeographicalRegion and #LegalAgent. This would fit appropriately for #RealEstate, too, as a unit of real estate covers a piece of the Earth (#GeographicalRegion) and has an owner (#LegalAgent), but Cyc fails to recognise this fact. Moreover, neither #RealEstate nor its supersets #HumanOccupationConstruct and #HumanlyOccupiedSpatialObject are related to basic (that is: supersets of) geographical entities, and their relations to other supersets seems inconsiderate.

A better solution could be achieved by arranging #RealEstate as a subset of #GeopoliticalEntity. The latter is another term for jurisdiction and as one may conceive a real estate as a jurisdiction that is governed by a citizen, one arrives at the following generalisations: The set of real estates is contained within the set of

jurisdictions (or `#$GeopoliticalEntity`) that again is contained within `#$GeographicalRegion` and further within `#$Place` and `#$SpatialThing`. In this way we have repaired the set relation (genls) of `#$RealEstate`.

The isa-relation can be repaired as well: The Cyc-comment to `#$GeographicalRegion` (not rendered here) states that ".. Each element of `#$GeographicalRegion` is a `#$PartiallyTangible` that may be represented on a map of the Earth." ".. like `#$Texas-State` .." (in the note to `#$SpatialThing`). An element of `#$RealEstate` is an element of `#$GeographicalRegion` and thus also member of the `#$PartiallyTangible` collection. This fits with the cadastral conception of real estate as consisting of tangible land (and buildings), but the boundary of which may not be tangible.

Other subsets of `#$GeographicalRegion` are likewise insufficiently structured for cadastral purposes, although essential entities are provided by the Cyc vocabulary, see the table below. One entity is, however, missing in the geographical domain: The collection of markers in the terrain: Street signs, landmarks, boundary markers, markers of geodetic points, etc. Such markers are, however, represented in Cyc, but only in the context of visible information.

<u>#\$SpatialThing</u>		
	<u>#\$PartiallyTangible</u>	<u>#\$Intangible</u>
	<u>#\$Place</u> <u>#\$Border</u>	<u>#\$GeometricThing</u>
<u>#\$RealEstate</u> <u>#\$HumanlyOccupied</u> <u>SpatialObject</u> <u>#\$HumanOccupation</u> <u>Construct</u>	<u>#\$GeographicalRegion</u> with subsets: - <u>#\$GeopoliticalEntity</u> - <u>#\$LandTopographicalFeature</u> - <u>#\$ContactLocation</u> (post address)	<u>#\$FrameOfReferenc</u> <u>e</u> <u>#\$Terrestrial</u> <u>FrameOfReference</u>

The presentation of the table is incomplete, as far as mathematical/geometrical entities of Cyc are not shown: Line, angle, etc. Also, the table does not show interesting details. For example, `#$LandTopographicalFeature` appears as a subset of `#$PartiallyTangible`, while the description states that "each element of `#$LandTopographicalFeature` is a region of land distinguishable from surrounding regions by salient physical differences in surface texture, three-dimensional shape, or elevation." Provided that the physical discontinuities can be recorded, the elements of the subset are tangible rather than `#$PartiallyTangible`.

This part of the Cyc vocabulary demonstrates one of the difficulties by establishing a cadastral ontology, namely finding the balance between emphasis on correspondence with more general ontologies like the Cyc ontology, or specifying a 'local'/cadastral optimum. The former strategy would develop the level of `#$Place` and `#$Border` as far as possible, and accept a certain departure from 'local' jargon. The latter strategy presumably would include the level of `#$Place` and `#$Border` in a cast that as far as possible uses the vocabulary of the scientific community concerned, and structures the generalisations/specialisations to achieve a kind of local optimum.

6 Events - The dynamic dimension

Important activities within the cadastral domain include measurement, communication, and the transfer of real property rights, respectively. However, within the Cyc ontology, there are no simple relationships between these activities, as we shall see from the following.

Communication, in Cyc terminology: *Communicating*, stressing the dynamic aspect, is a subset of `#$InformationTransferEvent` that again is a subset of `#$GeneralizedTransfer` (see the table below). `#$GeneralizedTransfer` includes the subsets `#$OrganizationalTransfer` (in the table a subset of `StatusChange`), and `#$ChangeInUserRights` (a subset of `#$Transaction`), but it cannot serve as the only or dominant generalisation (superterm), as `#$Transaction` refer to `#$SocialOccurrence` as well as to `#$PurposefulAction` as generalisations.

The entities `#$GeneralizedTransfer`, `#$SocialOccurrence`, and `#$PurposefulAction`, respectively, are supersets for the entities `#$WatchingSomething`, `#$InformationTransferEvent`, etc. shown in the tables below. The generalisation relations may be inferred from the relative position of the cells of the two tables. A detailed rendering of these relations would be complex and hardly informative.

<u>#\$Event</u>				
<u>#\$GeneralizedTransfer</u>		<u>#\$HumanActivity</u> <u>#\$PurposefulAction</u>		<u>#\$SocialOccurrence</u>
<u>#\$WatchingSomething</u>	<u>#\$InformationTransferEvent</u>	StatusChange	<u>#\$Transaction</u>	<u>#\$HostileSocialAction</u>
<u>#\$InformationRecordingProcess</u>	<u>#\$Communicating</u> <u>#\$IBTGeneration</u>	<u>#\$ObtainingPermission</u> (gnls: Transaction) <u>#\$OrganizationalTransfer</u>	<u>#\$TransferringPossession</u> <u>#\$ChangeInUserRights</u> <u>#\$ExchangeOfUserRights</u> <u>#\$TransferringOwnership</u>	<u>#\$DisputeEvent</u> <u>#\$Trial</u>

One might thus consider whether the structuring of the Cyc ontology is optimal for the present purpose. Firstly, it appears that the entities `#$Transaction`, `#$TransferringPossession` and `#$TransferringOwnership` are not well related to `#$ChangeInUserRights` and `#$ExchangeOfUserRights`. Probably, the two sets of entities have been developed by different working parties. Moreover, Cyc seems not to include the general notion of `StatusChange`, which is needed to describe the transformation of a hedge into a property boundary, or the transformation of a parcel lot into a (new) unit of real property, cf. the baptizing or naming of a child. Cyc does include entities that relate to `Status` or `StatusChange`. For example, Cyc includes

#\$SocialStatusAttributeType that is described as follows: "Each #\$SocialStatusAttributeType is itself a coherent collection of attributes that pertain to rank/status along some particular `dimension' related somehow to `status in society'." Also, #\$SocialRitual relates to a status change. The Cyc entities, which most directly relate to StatusChange are #\$ObtainingPermission and #\$OrganizationalTransfer. The former is a subset of #\$Transaction, and the latter of three supersets: #\$GeneralizedTransfer, #\$SocialOccurrence, and #\$PurposefulAction.

The surveyor's measurement is somehow represented in Cyc by #\$WatchingSomething and #\$InformationRecordingProcess. Cyc's societal activities includes #\$Trial, a reminder that the cadastral domain encompasses more than smooth running activities.

Concluding, and taking Cyc as it is, it provides us with terms of high level of abstraction. Cyc informs us that the cadastral domain is concerned with the transfer of information (by humans or computers or other measurement equipment) and the transfer of rights, and that the transfer occurs in a societal setting (and not within the context of, e.g. engineering or biology).

Activities within the cadastral domain include a substantial creative element that is not depicted by the above extract from the Cyc ontology. #\$Transactions are, in Cyc, the collection of events performed by #\$Agents co-operating (willingly) under some #\$Agreement. Mostly, the #\$Agreement are statute law and derived rules, which often need a certain interpretation. Also, the steps of the individual #\$Transactions can be sequenced differently, to benefit from specific circumstances.

The extract does not intend to address the changes of the sets of rules (laws, agreements) that governs #\$Transactions. Also, the power games that mostly come into play in case of changes of rules are not covered.

7 Discussion

The Cyc ontology has provided as a source and reference for presenting the entities of which a core cadastral ontology can be made. However, the potential of the software to assist in the provision of consistency among definitions and relations has not been tested. As it appeared from the previous sections, the Cyc ontology does provide a substantial base for further developments in the sense that maybe 50-70% of a core cadastral ontology has been established in the previous sections. However, it appeared too that much still is left to be done, not only in naming the basic entities, but more critically in structuring them in an adequate way.

This is by no means surprising: An ontology is a new literary and intellectual genre, which appeared during the last decade of the 1900s. It can be compared with the encyclopaedias of the late 1700s, which like an ontology claim to cover a field of human knowledge in a complete way. Also, it can be compared with the axiomatic systems of David Hilbert and others, which like an ontology aim at presenting the knowledge in a coherent way, and without contradictions. It seems likely, that first generation global ontologies like Cyc will benefit from ontologies of more restricted domains during the next decade or so, and also from machine translation and theory and applications of computational linguistic.

The Cyc ontology includes information, which may appear useful for further development of a cadastral ontology, but which has been passed over, or mentioned only briefly. Firstly, Cyc is not reviewed here from the point of view of ontology development, but rather from a cadastral, application oriented view. This means that notions of micro-theory, of predicates and functions, and other contents of the Cyc Fundamental vocabulary file, are not referred here. Also, potential information on, how to combine sets, elements, predicates, etc. in an effective way is not extracted. Secondly, the Cyc vocabulary file on Possession transfer includes material that should be exploited in the context of a more detailed treatment of that field. This applies, finally, also to Cyc's notion of information content that is not related to a material object, e.g. #PropositionalInformationThing.

In the four sections above a number of more specific observations were made, observations that could be included in further work on developing a core cadastral ontology, and in fact call for further research.

One of the Cyc vocabularies (topical concept files) was supplemented with a graphical rendering of relations between the entities of the ontology. It would be an important task to review the means and methods for rendering conceptual relationships, and assess the applicability of current technology for ontology development.

Probably even more important is the relation between the (formalised) ontology and the corresponding natural language terms. The Cyc ontology claims to be "formalized common sense". The problem is not that Cyc fails to be "common sense". The problem is, what especially the presentation of the geospatial domain demonstrated, that a "common sense" vocabulary will not be coherent and consistent. Theoretical terms have the benefit of being detached from the specific context and various connotations, which paste to common sense terms, and they are tested for their consistency (in principle, at least). Similarly, the professional language of a scientific community is moulded and cut to pinpoint the peculiarities of the branch. Thus it seems a justifiable strategy to develop a specialised ontology on the basis of theory, as far as existing, and to encode the constructs of the theory into the ontology.

One could question a theoretical approach from the point of view of usability: Theory is often hard to apprehend, and couching it in an ontology might make the use of theory still harder. A response might be that ontology development could take place in dialogue with university teaching of the specific parts of the knowledge to be represented in the ontology, and benefit from the shown pedagogical ingenuity. Reciprocally, the university teaching could benefit from the more explicit structuring of the subject matter.

Concluding the discussion, the Cyc ontology is a large-scale investment from which lessons can be learnt. If theory development is difficult, then this is not less the case with ontology development, as it has been asserted above that an ontology should be based on available theory. Maybe, the eliciting of an ontology is just another name for theory development within the 'softer' sciences.

8 Conclusion

Recently, the development of ontologies has been suggested as a means for achieving interoperability among collections of geospatial datasets. The notion of ontology is presented, and one of the early and comprehensible ontologies, the Cyc ontology, is introduced.

The Cyc ontology is reviewed from the point of view of a specific application domain, and the outcome discussed. While the Cyc ontology claims to be "formalized common sense" the review suggest that a "common sense" vocabulary will not be coherent and consistent, and that, consequently, ontologies ought to be based on theory, as far as existing.

Acknowledgements

The presentation of the Cyc ontology was based on information made public available by Cycom, Inc., who hold copyright to The Cyc® Ontology Guide. The introduction to ontological engineering of the present paper draws upon the contribution by Heiner Stuckenschmidt to the paper by Stubkjær and Stuckenschmidt (2000). Both contributions are gratefully acknowledged.

References

- Bishr, Yaser A, Pundt H, Kuhn W & Radwan M (1999) Probing the concept of information communities - A first step toward semantic interoperability. Pp 55 - 69 in: Goodchild et al (1999) Interoperating Geographic Information Systems. Kluwer, Boston. 509 p.
- Bylander & Chandrasekaran (1988) Generic tasks in knowledge-based reasoning: The right level of abstraction for knowledge acquisition. In B. Gaines and J. Boose, editors, Knowledge Acquisition for Knowledge Based Systems, Vol. 1, pp. 65-77. Academic Press, London.
- Corcho, O and Gomez-Perez A. (2000) A Road Map on Ontology Specification Languages. Proceedings of the workshop on Applications of Ontologies and Problem-Solving Methods at ECAI 2000, Berlin.
- Cyc® Knowledge Base <http://www.cyc.com/products2.html#kb>
- EuroConference (2000) on Ontology and Epistemology for Spatial Data Standards <http://www.geoinfo.tuwien.at/events/Euresco2000/gdgis.htm>
- Goodchild Michael, Egenhofer Max, Fegeas Robin & Kotman Cliff (1999) Interoperating Geographic Information Systems. Kluwer, Boston. 509 p.
- Gruber, T. R. (1993) A Translation Approach to Portable Ontology Specifications *Knowledge Acquisition*, 5, 199-220.
- Guha, R. V., D. B. Lenat, K. Pittman, D. Pratt, and M. Shepherd (1990) Cyc: A Midterm Report. *Communications of the ACM* Vol. 33 , no. 8

- Jackendoff, Ray (1983) *Semantics and Cognition*. Cambridge, MA. MIT Press.
- Kottman, Clifford A (1999) The Open GIS Consortium and progress toward interoperability in GIS. Pp 39 - 54 in: Goodchild et al (1999) *Interoperating Geographic Information Systems*. Kluwer, Boston. 509 p.
- Kuhn, Werner (2000) How to produce ontologies: An approach grounded in texts. in Stephan Winter (Ed) *Geographical Domain and Geographical Information Systems Geo-info series* Vol. 19. Institute for Geoinformation. Vienna University of Technology. Pp. 63 - 71.
- Lenat, D. B. (1995) Cyc: A Large-Scale Investment in Knowledge Infrastructure. *Communications of the ACM* Vol. 38, no. 11
- Ontology Interchange Language (2000) <http://www.ontoknowledge.org/oil>
- Stubkjær, Erik (1994) Employing the Linguistic Paradigm for Spatial Information. In: *Advances in GIS research: Proceedings, Spatial Data Handling Conference, Sept. 1994, Edinburgh / Waugh, Thomas C; Healey, Richard G (Eds)*. Edinburgh, University of Edinburgh. Vol 1, pp 572-587.
- Stubkjær, Erik (1996) A Theoretical Basis for Cadastral Development. In: *ELIS' 94 European Land Information Systems. Proceedings, Third Seminar of the European Co-operation Network for Education and Research in Land Information Systems (EUROLIS), Sept. 1994. Delft; Delft University of Technology, 1996*. Pp. 0.9-0.23.
- Stubkjær, Erik and Stuckenschmidt, Heiner (2000) Ontological engineering for the cadastral domain. Paper presented at UDMS 2000, Delft. <http://www.tzi.de/~heiner/public/UDMS-00-2.pdf>
- Stuckenschmidt and Wache (2000) Context modelling and Transformation for Semantic Interoperability. *Proceedings of KRDB 2000*.
- Uschold, M and King, M (1995) Towards a Methodology for Building Ontologies. *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing at IJCAI-95, Montreal*.
- Vckovski, A (1998) *Interoperable and distributed processing in GIS*. Taylor & Francis, London.