

A conceptual modeling methodology based on niches and granularity

Sonia Berman and Thembinkosi Daniel Semwayo

Computer Science Department, University of Cape Town, South Africa
sonia.dsemwayo@cs.uct.ac.za

Abstract. This paper presents a methodology for conceptual modeling which is based on a new modeling primitive, the niche, and associated constructs granularity and reconciliation. A niche is an environment where entities interact for a specific purpose, playing specific roles, and according to the norms and constraints of that environment. Granularity refers to the relative level of power or influence of an entity within a niche. Reconciliation is a relationship from N entities onto one reconciled entity, and represents explicitly a situation where two or more different perspectives of the same entity have been reconciled, by negotiation, into a single consensus view. The methodology we propose provides a systematic method of designing conceptual models along with a process for normalising inappropriate relationships. Normalising is a prescriptive process for identifying and remedying inconsistencies within a model based on granularities. Drawing on a number of case studies, we show how niches and granularity make complexity easier to manage, highlight inaccuracies in a model, identify opportunities for achieving project goals, and reduce semantic heterogeneity.

1 Introduction

Conceptual modeling is a difficult task. It requires a group of people, typically from different fields of expertise, to understand a complex situation and to agree on an appropriate set of abstractions that captures correctly and completely the characteristics and interactions of the entities involved. While data models such as the ER model or UML provide many useful constructs to simplify this problem, there is little prescriptive guidance on how to develop a model, or how to evaluate one that has been designed. Errors and omissions in a conceptual model can have far-reaching consequences, as physical systems that are built from these models can be faulty; incorrect or incompatible data can be collected; and data can be integrated and analysed inappropriately leading to invalid conclusions about the real world.

We aim to improve on conceptual modeling by proposing a methodology that can address these problems, specifically:

- providing mechanisms to handle complexity so as to reduce inaccuracies and omissions in the model
- prescribing a sequence of steps to carry out, and constraints to check, so that the task is tackled in a systematic way
- reducing semantic heterogeneity – the phenomenon whereby different people use different terms for the same concept, or use the same term with subtly different meanings - as occurs particularly with multi-disciplinary teams
- where applicable, depicting the situation in a way that highlights where interventions can achieve project goals – i.e. where there is an opportunity for improving the real-world situation being modeled

In studying a number of projects where we acted as consultants, we realized that many of the modeling problems arose because entities and relationships were described in isolation, rather than in the context of the particular sphere in which they operated. As a result, much was omitted from the models altogether because certain interaction contexts weren't considered at all, and because the different roles that entities play in different contexts were not recognised. We accordingly introduce early in the methodology the notion of a *niche* (system context or sphere of operation) to ensure that such errors and omissions are minimized. We further observed that missing entities and roles were more likely to be noticed when the team was asked to consider the relative levels of influence (power, impact or status) of entities in each context. This often showed e.g. how an individual can have greater influence in one niche in which s/he operates than in another niche, because of different roles s/he plays in these different contexts. We have named this meta-property of an entity within a niche its *granularity* or granular level, and the methodology requires that designers assign granularities to entities after niches have been determined. Constraints on relationships involving entities of differing granularity (unequal relationships) are then checked: either they are modeling errors or else they are anomalies in the real world. Such anomalies highlight a vital aspect of the domain being modeled, namely interactions which are suboptimal. In many

domains, such as the socio-economic projects we have tackled, a key objective is to improve the status quo; interaction anomalies indicate precisely where improvements are possible. In addition to niches and granularity, we propose a *reconciles* relationship which serves to explicitly document where semantic heterogeneity has been identified and resolved. This indicates where and how differing views of the same entity have been replaced by a consensus view after negotiation among the design team.

In sections 2 and 3, the paper describes the new modeling constructs and the proposed methodology. Thereafter we discuss experiments in which the methodology has been applied, highlighting advantages and disadvantages noted during those workshops and case studies. Section 5 presents related work, and we conclude with a summary of the main ideas and some suggestions for future research.

2 New Modeling Primitive and Associated Constructs

In this section we introduce the niche modeling primitive and its associated constructs: granularity and reconciliation. We show why they are needed, how to recognise them and introduce them into a model, and when to do so. We also discuss what effects, if any, poor use of the new constructs would have.

2.1 Niches

When describing the entities in a domain and the relationships between them, it is important to bear in mind the different contexts in which interactions occur. In any project, it is important that no interaction context is forgotten, and that the entities and their relationships are accurately depicted by taking into account the conditions and norms that apply in that specific context. Drawing on the biological notion, we define a niche as a specific context in which entities interact for a specific functional purpose, favouring particular individuals or groups, and behaving according to the conventions and constraints of that particular environment. In a niche there is a stable pattern of behaviour and structuring of the community, and interactions take place accordingly. A niche frequently exhibits emergent properties, which are attributes that do not necessarily apply to the individual members but do apply to the niche as a whole.

A niche can be defined as a 7-tuple (S, T, M, N, H, A, P) where

S is a spatial location occupied by this niche

T is a time interval

M is a non-empty set of member entities present at location S for part of time interval T

N is a non-empty set of interactions between entities, the normal behaviour in that niche

H is a hierarchical structuring of the entities M based on relative impact/influence/power

i.e. H is a function mapping every m in M onto a granular level g i.e. $H(m) = g$

A is a set of attributes of the niche which aren't attributes of its members M

P is a possibly empty set of environmental parameters that hold at location S during time T

An example of a niche attribute is its functional purpose or its quality/performance. The environmental parameters describe domain conditions that make the niche viable – e.g. in an ecological domain these would be climate factors, whereas in a business domain these would be political and economic factors. We note further the following niche constraints:

If $m \in M$ then $H(m)$ is unique i.e. every entity has exactly one granular level in a niche.

If $m1 \in M$ and $m2 \in M$ and $m1$ is-part-of $m2$ then $H(m1) \leq H(m2)$.

If $m1 \in M$, $m2 \in M$, $R \in N$ and $R(m1, m2)$ then $\exists t \in T$ s.t. $\text{In}(S, t, m1)$ and $\text{In}(S, t, m2)$ i.e. for two entities to interact in a niche they must exist in that niche at the same time.

For all $m \in M$, $\exists t \in T$ s.t. $\text{In}(S, t, m) = \text{false}$ i.e. entities do not have to remain in a niche throughout its existence.

Considering the familiar example of an academic environment, it should be clear that individuals behave differently in a staff meeting (administration) than they would in a research meeting, because these are two different niches. An academic might play the role of subordinate in a staff meeting if she were a junior lecturer, the role of critic at a research paper presentation, and the role of student in a meeting with her thesis supervisor. In each of these niches we can identify relationships between academics, but the nature of these relationships varies according to the niche in which the interaction takes place. The functional purpose of the staff meeting,

the paper presentation and the supervision meeting differs, with interactions favouring the head of department, the speaker and the supervisor, respectively. Even in this simple example it is evident that, without considering these niches, modelers may well have used only one “meeting” relationship between academic staff entities, and failed to recognise the different roles that they play and the extra data that should be collected accordingly.

To determine the niches in a domain, it is generally best to examine the life-cycle of the key entities and the value chain of that domain. The value chain concept was first used for business system modeling [1] but has since been applied in other contexts. It comprises the sequence of core, value-adding activities of a system. For example, in the business world, the niches would be purchasing, infrastructure, production, sub-contracting, management, marketing, etc. In education the niches would be application, registration, tuition, examination, qualification, administration, infrastructure, etc. Some niches will be sufficiently well understood by designers to be modeled directly, but most will be too complex and require decomposition into sub-niches, resulting in one or more hierarchies comprising niches nested within each other. Most conceptual modelers are able to identify the value chain for their project, to identify those nodes which require further decomposition, and to find appropriate sub-niches for these – simply because they are familiar with the domain and hence with its processes and sub-sections. What is the danger of choosing the wrong niches? If designers can articulate for each niche its functional purpose, favoured members and behavioural norms, then there is little chance that the niche is not a valid context or environment within the project domain. But even if a “wrong niche” should be included, it still gives the same modeling benefits: it forces designers to think about where their entities and relationships will be found, how this affects their behaviour, what exists around them, how these influence each other, which entities are more favoured, and for what purpose they interact; producing a better model as a result. It focuses the mind not only on what entities exist in a relationship, but also on where, why and how they interact. Including entity life-cycles and domain value chain diagrams ensures that all contexts - the full scope of the project - is modeled, reducing the risk of significant omissions.

Because of its usefulness in identifying niches, we include a value chain diagram (figure 1) as part of our model, where any node can be shown at the head of a hierarchy of sub-niches into which it has been divided.

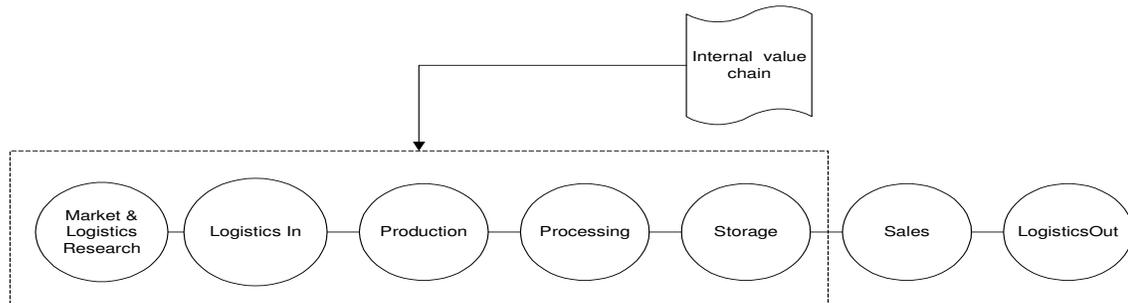


Fig. 1. Value chain diagram.

The model for each niche or sub-niche is drawn below the niche, as shown in figure 2.

2.2 Granularity

An important property of a relationship is the relative level of influence, power or impact of the entities involved. We call this the granular level or granularity of the entity, which is always given relative to a base entity – the main object in the project – which has granularity zero. Entities with similar impact, influence or power would thus also have granularity 0, those with less would have granularity of -1, those with even less have granularity of -2, and so on; while those with greater influence have granularity of +1, those with still more influence have +2, etc. Granularity is relative, it is not meaningful as an absolute measure but only in comparison with other entities, as it reflects the level at which an entity operates in a particular context relative to others in that niche.

For example, where humans interact, their relative power, status or influence indicates their granularity. In most institutions, status or ranking within the hierarchy is well-known, and interactions between individuals with very different granularity (status) rarely if ever occur. If we are to model a niche correctly, we must take cognisance of the granularity of entities in a relationship in order to determine whether the correct entities are being modeled in that context. To take the familiar example of a university model, a relationship between the Vice-Chancellor

and a student is contrary to the norms and constraints of the environment, unless that student is a particular student such as a StudentLeader who has higher status/granularity than his peers. In an agricultural example, rainfall has higher granularity (impact or influence) than a farmer, so if these two entities appear in a relationship there is again an anomaly that should be addressed – here the rainfall entity should be replaced by another that operates at the correct level of influence and scale for the farm context, such as a local water source like a dam or irrigation channel. Modeling with entities at the correct granular level for the niche/context reduces the chances of collecting incorrect data, such as rainfall figures for an area rather than information about farm dams and irrigation systems. Similarly, collecting data about students who meet with the Vice-Chancellor would be useless without information about the student body they represent. Thus we see that granularity aids the designer in finding the right abstractions to represent real-world phenomena, by making it far more likely that entities at the right level for the niche/context will be used.

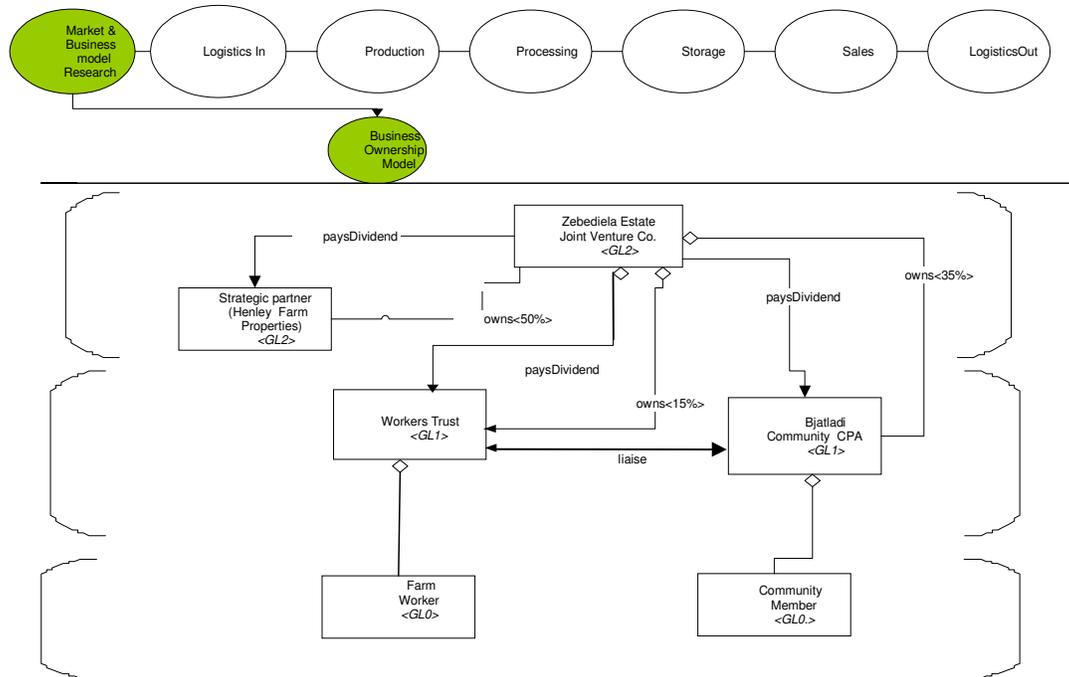


Fig. 2. Niche model showing business ownership of an agricultural estate property.

By annotating entities with their granularity (denoted GL for Granular Level and written after the entity name, as shown in figure 2) the relationships in a model can be studied to detect inequalities or anomalies, which imply that there is either a problem in the model or in the real world situation. Often modelers will realise in such cases that they have failed to recognise that an entity is playing a particular role in some context/niche, different from its other interactions in other niches. This in turn will highlight the need for either a generalisation hierarchy (e.g. MonopolySupplier *is-a* Supplier, or StudentLeader *is-a* Student) or an aggregation (e.g. Supplier *is-part-of* SupplierGroup, or Student *is-part-of* StudentCommittee) to be added to the model, and the entity in the relationship replaced with a new entity such as MonopolySupplier or SupplierGroup. If granularity is not taken into account in modeling, data is typically collected for the entities at differing resolutions, and so some of the data goes unused.

As granularity is a measure of gravitas or influence, it is only applicable to entities, not relationships; only to the actors or agents that are involved in interactions according to their power/impact/status. The granularity of an aggregate or composite entity cannot be less than the granularity of any of its component parts. For example, a Union entity representing an organised group of people cannot have lower granularity than its constituent union member entities. There is no corresponding constraint for *is-a* (subtype/generalization) hierarchies; if *S is-a T* then an *S* may have greater influence/power than a *T* entity, or it may have less, depending on the nature of the specialization. For example, StudentLeader *is-a* Student with higher granularity than Student in many niches, but StudentTeacher *is-a* Teacher with lower granularity than Teacher in many niches.

Like many aspects of modeling, granularity assignment is subjective. It will require the design team interrogating the clients or domain experts about relative levels of influence, just as they would ask them about other aspects of relationships such as cardinality or participation constraints. Modeling a niche however reduces subjectivity significantly, as the context and scale of activity is defined. We note further that incorrect use of granularity in a model will not introduce new modeling errors. It would mean that entities in a relationship are depicted as having equal influence when in fact the one has less power/impact than the other. Essentially the model is then no worse than it would have been if granularity were not used at all – a modeling error or omission remains undetected because there has also been an error in granularity assignment. The chances of this occurring when a domain is well understood are slight however, because granularity is considered for every relationship, and again when checking the level associated entities across aggregation / composition hierarchies. Lastly we note that granularity will not add value to conceptual modeling if the application is so simple that multiple levels do not exist; while checking relative impact/status of entities in such systems may not provide new insights, however, it will cause developers to carefully consider design decisions they have made. Most systems are more complex than they appear to be however, so such simple systems will be rare. Granularity is useful in separating concerns when complexity is revealed even in a simple study – for example in a retail scenario, small-businesses and big-businesses differ in granularity and exhibit different roles, attributes and relationships in different contexts.

2.3 Reconciliation

Semantic heterogeneity is a well-known problem in modeling and design. It refers to two kinds of situation: one where different individuals use different terms without realising that they have the same real-world object in mind; the other where people all use the same term for a concept without realising that they have differing ideas of what exactly is meant by that term. Thus e.g. a botanist may speak of miombo scrubland and a geographer of tree-bush savanna without realising that they are referring to the same thing, since a botanist describes vegetation according to function and the dominant species in the area whereas a geographer provides a structural description according to the relative proportions of tree, bush and grass. Or two individuals may discuss “household” without realising that the economist means a unit with an income and location, while the sociologist has in mind a group of people living together with a range of ages, incomes, genders, etc.

When niches and granularity are used in modeling, we find that semantic heterogeneity is detected and resolved as a bonus side-effect. For example, when granularity is considered, the difference between a geologist’s and an engineer’s concept of a rock profile becomes evident because the geologist works at a much bigger scale or granularity than does the engineer. When designing in the context of a financial niche, the economists’ notion of a household as an income-generating object is the view shared by the whole team; and when modeling the sociological niche all designers see the household as a group of separate individuals with different ages, genders, etc. We represent explicitly in our models the reconciling of entities wherever this occurs: if e.g. Household is an entity with different attributes in different niche models, the reconcile diagram shows how a consolidated Household entity is derived from these, as shown in figure 3.

In this example, differences in granularity would have led the design team to realise that the household of the sociological niche was in fact a composition of people, one of whom is the head-of-household who was viewed as “the household” itself in the financial niche. By representing this in a reconcile diagram as well, the possibility of different interpretations is highlighted and the chances of incorrect future development or data collection are reduced. The utility of this construct lies therefore in explicitly including as part of a model the fact that people had different perceptions of an entity, to indicate the similarities and differences between these views, and to show how the differing views were unified. Without this, future model users will be unaware that there are different ways of looking at the entity, let alone how to relate each others’ perspectives to their own.

3 The Design Methodology

In our experience, conceptual modeling is typically done by groups of individuals who first discuss and agree on their goal and objectives as a project team, and then brainstorm the entities involved and the relationships between them. A complex picture emerges, which the team must then improve upon in order to remove errors, redundancy and inconsistency; to discover and remedy omissions; and to decide which abstractions to use in depicting the situation. As this is a complex process, a consultant or facilitator is typically engaged to manage and assist.

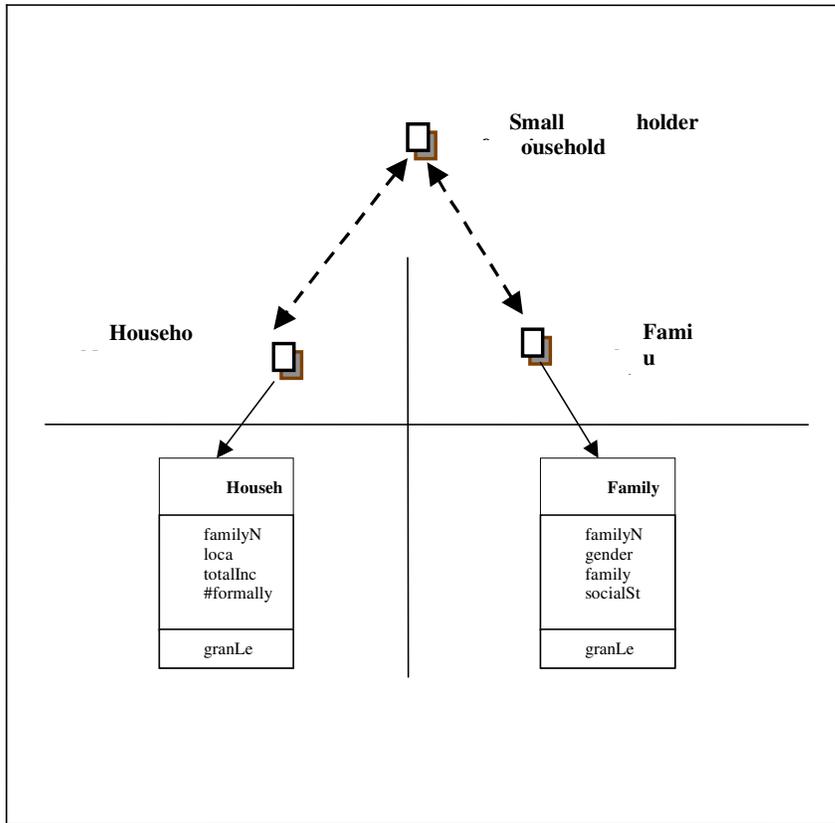


Fig. 3. Reconciliation example.

3.1 A Systematic Approach to Conceptual Modeling

We propose a methodology which requires a facilitator, skilled in conceptual modeling, to guide the team, and which commences in the normal way as described above. The steps of the methodology are as follows:

1. Define the goal of the project, and list the *objectives* to be met in order to achieve this goal.
2. Draw a *value chain diagram* depicting the main contexts or niches (spheres of operation) in which interactions take place.
3. Consider each node of the value chain, and draw an ER model for the entities and interactions that take place in this context. If a node is too large or complex, decompose it into sub-niches first and develop a model for each sub-niche. We call the resulting set of models the *niche models*.
4. Choose a key entity, that is central to the project goal and that occurs in most (if not all) niche models, and assign this a granularity of zero. For each niche model in which this key entity appears, assign *granularities* to all other entities, based on their relative level of influence/power – those with less influence being assigned negative granularities, etc. For any remaining niche model, assign granularities to entities in a similar way. Flag relationships involving entities of different granularity. An example is shown in figure 4.
5. *Normalize* flagged relationships: identify missing generalizations (role identification) or aggregations (composite entities), remedying the model and removing the relationship flag wherever these are found. For each relationship that is still flagged, consider whether this is an opportunity for achieving a project objective. If so, introduce the corresponding intervention in the model by adding dashed entities and relationships to represent these proposed changes to the status quo. Any relationships that are still flagged represent unsolved problems in the real world. The result of normalising the model of figure 4 is shown in figure 5. Here, specific operational staff for providing direct help to farmers have been identified through normalization.
6. Finally, create *reconcile* diagrams to record every situation where divergent views of the same entity were merged, or were resolved through generalisation, aggregation or composition.

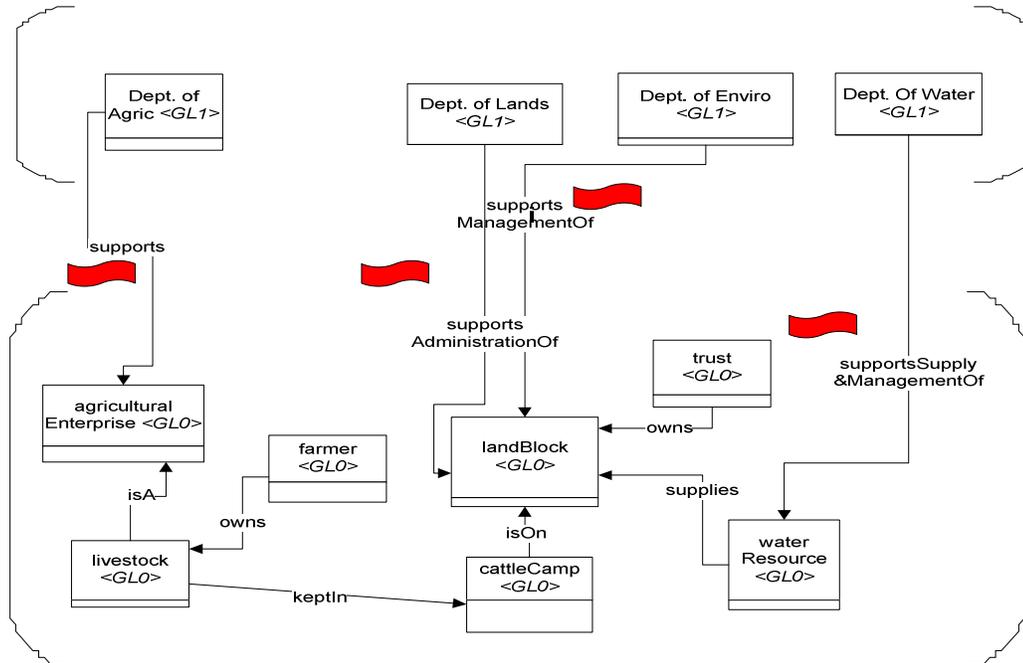


Fig. 4. Flags highlight cross-granular relationships in a niche model.

We distinguish sortals from other entities – these are ontological entities like “person” that inarguably exist independent of context, and have a stable set of attributes. Other entities in the model are aggregations, compositions or specialisation of sortals. Sortals are useful for obtaining cross-niche perspectives, since the same sortal will appear in several niches corresponding to different roles played in different contexts. An overarching view of the niche models is easily formulated by taking some key entity and looking at all the relationships in which it participates. If all occurrences of the subtypes of a sortal are considered across all niches, one gets a global picture of the interactions that such entities can be involved in. In one case study, for example, this highlighted the fact that an individual could be a community member, a leader, a shareholder and a director, and be financially rewarded repeatedly in each separate role.

3.2 Relationship Normalisation

Normalisation in step five of the methodology involves examining the entities in each flagged relationship to detect any missing specialisations or aggregations. Where these do not apply, this means the relationship is problematic in the real world, rather than being a modeling error. In this case, normalisation requires that the project team apply their mind to this opportunity or challenge that has just been highlighted. If they come up with a practical solution, they include it in the model as a possible intervention/strategy (denoted using dashed lines); otherwise the flags remain, as annotations that highlight anomalous relationships in the system. The designers thus leave a flag in place whenever they have checked that the correct entities are indeed being used and the relationship does in reality involve entities of unequal power/influence. In projects where a goal is to improve the system being modeled, the team would first look for ways of remedying the situation, and only leave the flag in place where these do not exist. When our methodology is not followed, intervention strategies suggested in such projects are often doomed to fail precisely because the niche is not taken into account; suggestions don’t address problems in the context of the sphere of operation, and the crucial factor of relative impact levels is not considered.

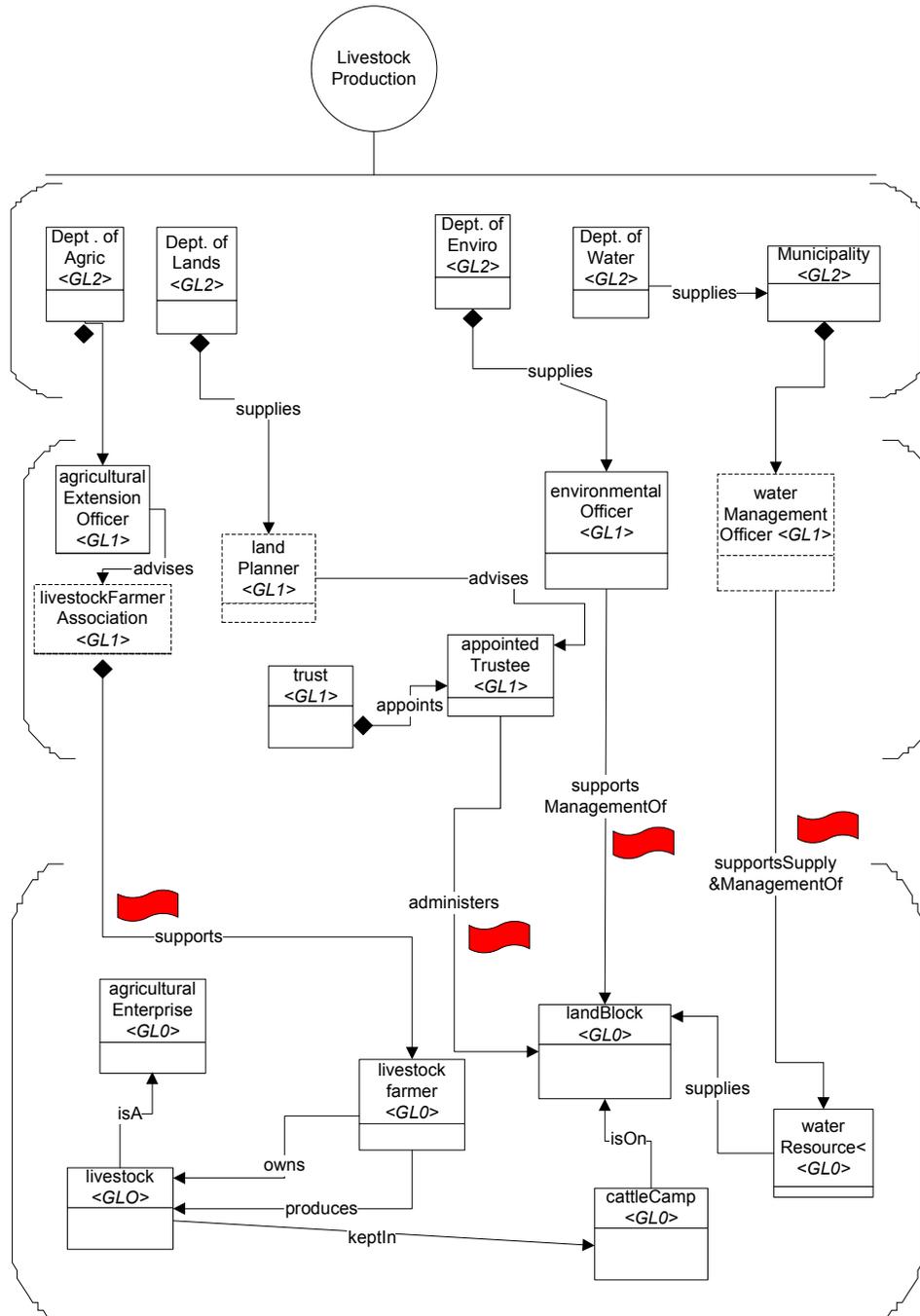


Fig. 5. Normalised relationships, with intervention opportunities/plans shown using dotted rectangles.

As an example, a relationship between a farmer and a bank would be flagged because the granularity (power/influence) of the bank is greater than that of the farmer. This can be remedied by an intervention such as the introduction of a farmer's union or grouping of some sort, which would have more influence than a single

individual. In other situations specialisation, rather than composition, is needed. A flagged relationship between Field and HighValueCrop is an example, where the latter has a higher granularity because it represents a crop which is sold on global rather than local markets. To handle this flagged relationship requires that the Field be a special type of field that can support a HighValueCrop, e.g. one which is irrigated and treated appropriately. Thus a specialisation of Field, called say HighValueField, is required here. Additional entities and relationships providing the extra inputs that such a field requires (irrigation etc.) have to be added to the model at the same time. This might be an intervention opportunity, or it might simply be rectifying a modeling error – if in fact the fields being modeled are already being irrigated and treated, this aspect may simply have been omitted in the model, despite being crucial to the production of a HighValueCrop. The flagging of cross-granularity relationships brings the problem to the modeler’s attention, thus increasing the chance of such an omission being remedied. Similarly, in the previous example, it is possible that a farmer’s union does already exist but was erroneously omitted from the model; the flagged relationship highlights the need to model that the bank is approached by a group representative rather than an individual. For the example in figure 5, relationships between high-level government departments and the lower-granularity entities that they support have a considerable gap in power/influence level. The intervention that emerges, viz. to consult the appropriate official in that department, is explicitly added to the model, but using dashed notation to indicate a future plan rather than an existing situation.

Note that these normalisation examples also illustrate how semantic heterogeneity is reduced as a natural side-effect of using our methodology. Often designers use the same term but have different subtypes in mind, e.g. all will speak of a “field”, but some will have in mind a HighValueField (irrigated, fertilised, etc.) and others not; or they will all use the word “farmer” but some will mean subsistence-farmer and others commercial-farmer. Our methodology brings such discrepancies to light and ensures that the generalisation hierarchy for “field” or “farmer” is added to the model. In step 6, reconcile diagrams are also produced, so that there is a record of the fact that conflicting definitions of “field” and “farmer” have been resolved, and how.

4 Experiments

While the model and methodology were still being developed, we applied it in a number of projects involving multi-disciplinary teams. In these initial experiments, which we call the workshops, we explained the new concepts, facilitated application of the methodology, and then pointed out the benefits of our approach. We evaluated the usability of the methodology through questionnaires and interviews. As a result of these three initial workshops the model, methodology and tools were improved considerably based on observations and feedback.

Then, after the model and methodology had been developed, we used them in a number of case studies for clients in the field. As it isn’t possible to “prove” that a methodology is beneficial, these case studies were done in an attempt to demonstrate by example the validity, applicability, and advantages of our methodology over conventional modeling practices. In each case study therefore, the initial design produced by the team was compared to the final model obtained after applying our methodology, in order to discover any differences and identify which benefits and shortcomings, if any, were introduced using our approach.

4.1 Initial Workshops

Our initial experiments were conducted while the model and methodology were being refined. By observing novices discussing, questioning and applying our approach we were able to evaluate its usability and identify ways in which it could be improved or extended. For each of the workshops we drew up a pro forma plan of our expectations beforehand, started the workshop by introducing the new concepts, facilitated application of the methodology by the team while a co-worker observed, and obtained feedback from participants through questionnaires and interviews.

The aims of the workshops were to:

1. determine whether participants appreciated the danger of inaccurate and incomplete models
2. discover whether participants felt that they knew how to solve such problems
3. establish the validity of niches, granularity and reconciles
4. assess the applicability of the methodology
5. evaluate the usability of our approach
6. elicit user perceptions of the model and methodology

7. gauge whether the methodology reduced semantic heterogeneity

At the first workshop a multi-disciplinary team from such diverse fields as computer science, engineering, ecology, geography, zoology, geology, biology and financial management, tackled the problem of integrated coastal zone management. In this workshop, as a first step, only granularity was introduced; the project was small enough to manage complexity by modeling within a single broad context, without niches. With the aid of the facilitator (ontological engineer), the team assigned granularities to entities and normalised cross-granular relationships; they were even able to identify interventions to improve the status quo. For example, a relationship between the DistrictTourismAuthority and the CommunityOrganisation, entities with different granularity, led to the team identifying a potential strategic partnership to remove this anomaly.

At the start of the workshop, nine of the fifteen participants were aware of the dangers of modeling errors. Of these nine however, only two realised that such errors would not be “automatically” corrected in software. After the workshop, all but one of the participants were convinced of the dangers of poor modeling. All but one were convinced that granularity was valid and useful, while 13 of the 15 felt it was applicable in their domain (2 unsure). On a rating scale from 1 to 5, two subjects chose 5 (our approach much better than other methodologies), 12 chose 4 (better), and one chose 3 (same as others). Comments noted i.a. the following benefits: “improved data classification”, “the modeling process (is) more specific”, “assist in focusing (the) project”, and “able to minimise risk”. From observations during the workshop, it was clear that normalisation involving specialisation hierarchies for different roles contributed significantly to a reduction in semantic heterogeneity. Clearly a key factor in avoiding semantic heterogeneity lies in distinguishing between entities and entity roles.

In the second workshop a team of engineers and geologists, amongst others, modeled a gold mining project using the full methodology, including niches and granularity, under the guidance of a facilitator. Granularity was particularly useful here because of the natural tendency for engineers to describe rocks and related objects at a far smaller scale than do geologists. Participants were able to model with niches, and found them particularly useful for defining emergent properties, i.e. properties of the group or niche that cannot be traced back to properties of the individuals in that niche. It was also noted that niches and granularity helped the team to define attributes at the correct scale or level of detail, because the context and impact level was taken into account. Usability results were similar to the first workshop: all but one were convinced that granularity was valid and useful. Five subjects felt the methodology was applicable in their field, three were unsure and one thought it was not applicable in his domain. On the rating scale of 1 to 5, one participant rated the methodology 3 (same as others), three people rated it 4 (better than others) and five participants rated it 5 out of 5 (much better than other methodologies).

The third workshop was a follow-up study in which the same group who had attended workshop one, set about refining and extending the model they had produced there. It was encouraging to see that they had assimilated the ideas to the extent that they required only a brief introduction to remind them how the methodology worked. They also found using niches helpful in handling the increased complexity of the extended task. Overall the workshops indicated that the model and methodology were usable and broadly applicable, and led to improvements not only in the methodology itself but also in the toolkit we used to support the process.

4.2 Modeling Toolkit

In the first workshop, we customised the Poseidon UML modeling tool [2] to suit our model, and the facilitator used this during the modeling exercise. It was discovered that participants who were not computer scientists struggled to understand these diagrams, which hampered rather than helped the process, because of its sophisticated interface and the richness of the diagrams and repositories it displayed. For the next workshop, we used a far simpler tool which was geared specifically to support our methodology. Surprisingly, the use of this simpler tool during the design was still problematic, with discussion often relating to the tool rather than the task at hand. For the final workshop we used a tablet on which designs were hand-drawn, and found that this worked far better. The facilitator would input these hand-drawn diagrams to our simple tool each night, and discuss this with the group the next morning. At that stage the team was not attempting to be creative but rather to understand the implications of what they had previously designed, and were able to assimilate lessons learned from inputting diagrams into the tool. Thus we conclude that a tool should not be used during creative design phases, but rather when a design is being reviewed, evaluated and refined.

4.3 Case Studies

Using our methodology we have developed models for the following projects: a system for monitoring sustainability in land reform projects; a citrus farm development study; an eco-tourism investigation; a knowledge base for the Department of Water and Forestry; a feasibility study for the establishment of a land and water management regional database; a small business linkages system for the City of Cape Town and for the provincial government; and a strategy development tool for the South African Land Reform Programme. In this section we briefly describe aspects of one case study to illustrate some of the benefits of using our approach.

The Eastern Cape Development Corporation (ECDC) exists to improve the standard of living of a South African rural community, the Dwesa Cwebe. The community, represented by a Dwesa Cwebe Land Trust (DCLT), has a joint venture company with the ECDC - called the Operational Structure (OPS) - each owning 50% of the shares. OPS owns the local Haven Bashe hotel which is operated by a private entrepreneur. It is hoped that in the long run the ECDC will relinquish ownership in OPS.

4.3.1 Initial Conceptual Model

The multi-disciplinary project team first used the conventional approach of stating project objectives, brainstorming to identify pertinent entities, and then drawing up a conceptual model indicating these entities and the relationships between them. We first outline this initial modeling that took place. Two objectives were identified to meet the project goal:

1. Create employment opportunities linked to the Haven Bashe hotel
2. Create employment opportunities linked to the Dwesa Cwebe nature reserve lodges and chalets.

Brainstorming was then done as a team, to share ideas and explore the problem collectively, mind-mapping the thoughts that emerged. From this mind-map a conceptual model was drawn as shown in figure 6.

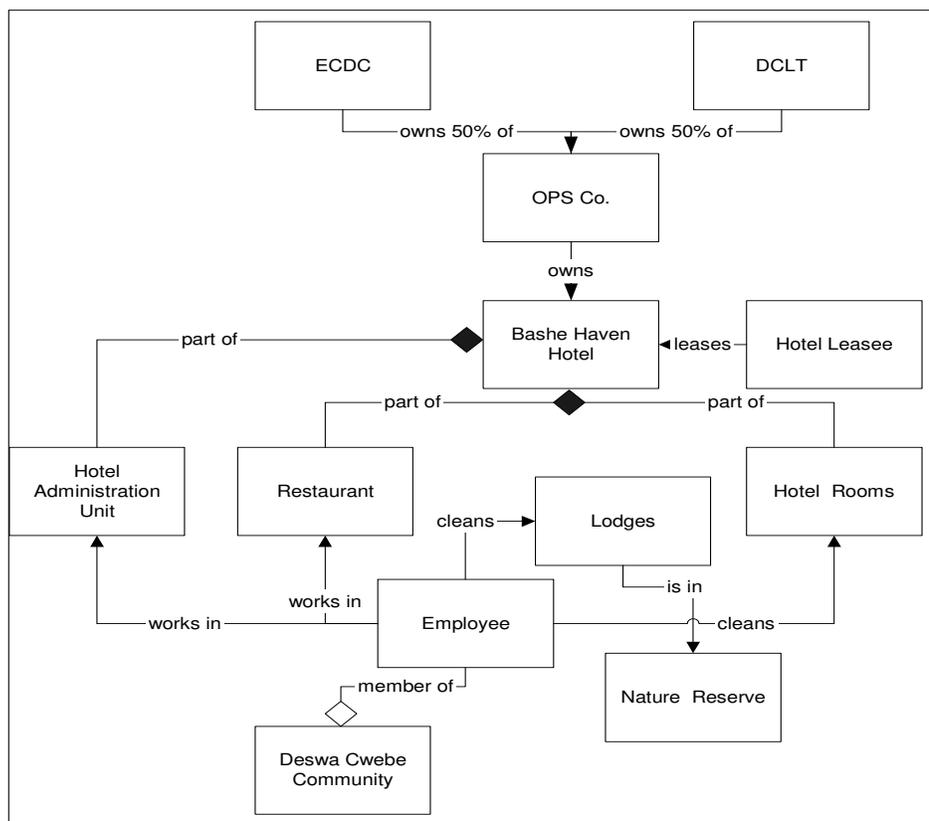


Fig. 6. Original conceptual data model, before applying our methodology.

It was observed that

- No more than 50 persons are employed by the hotel and surrounding Nature Reserve. The community has a population of 2382 households. Unemployment is therefore rife.
- Although the DCLT owns 50% of the hotel, it has not been able to leverage the full potential offered by the tourism industry.

The model provided a static, given, hopeless situation from the community’s point of view.

4.3.2 Applying the Methodology

At this stage our methodology was applied by the team, with the author as facilitator. First, the value chain, niches and sub-niches were identified for the project, as shown in figure 7.

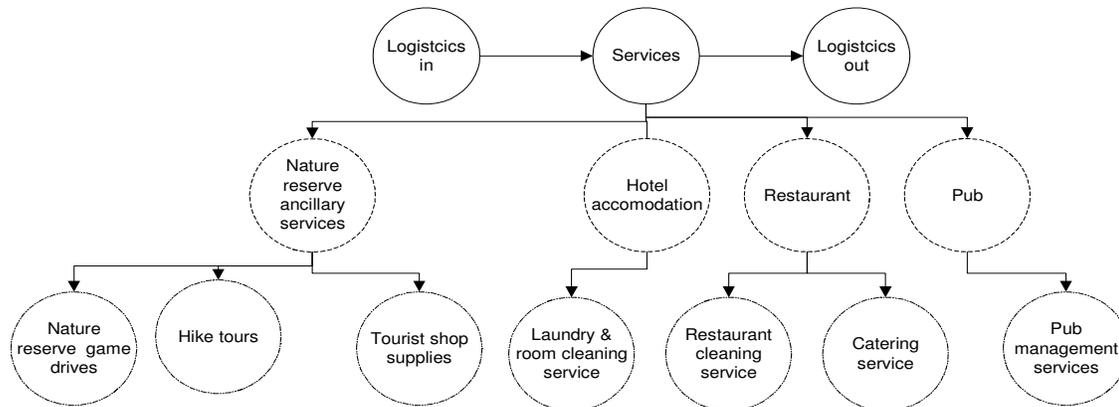


Fig. 7. Haven Bashe Project value chain and niches

It was clear that the use of the value chain and its decomposition into sub-niches was considerably broadening the team’s view of the problem, and bringing new aspects to bear that had not emerged in the earlier brainstorming or modelling sessions. Models were then drawn up for each sub-niche, granularities assigned to entities, and cross-granular relationships flagged. The resulting model for the Catering Service niche is shown in figure 8.

After normalising the flagged relationships, the model was refined to obtain that of figure 9. In so doing, intervention opportunities were identified. An idea emerged of forming a catering company part-owned by the chef and other restaurant employees. This effectively gave greater influence/power to those community members working in the restaurant. As a direct result of distinguishing the service niche from the broader project context, a strategy of forming separate enterprises serving the tourism industry was born, thus better leveraging opportunities offered by the hotel and nature reserve; and at the same time giving the workers from the community part ownership of these service enterprises.

There were several other situations where benefits arose from identifying and exploring specific niches and examining the granularity of the entities interacting there. For example, it was established that there was a relationship between the hotel lessee and the ECDC (nepotism) and this was a source of friction with the DCLT, resulting in an environment so fraught with suspicion and non-cooperation that it almost paralysed the project. As another example, the granularities of the “equal partners” in the joint venture company differed, since the ECDC, being a quasi-government organization, were in a privileged position (with access to information such as future plans, etc.) and were more highly skilled than the members of the Trust. Relations between the ECDC and the DCLT were particularly acrimonious with all parties seeing the DCLT as a “poor relation”. Another strategy emerged in trying to resolve this difference in granularity: it was decided that the ECDC and the DCLT form a consortium that owned the OPS, so that they could be a single entity working together, rather than being seen as two rivals.

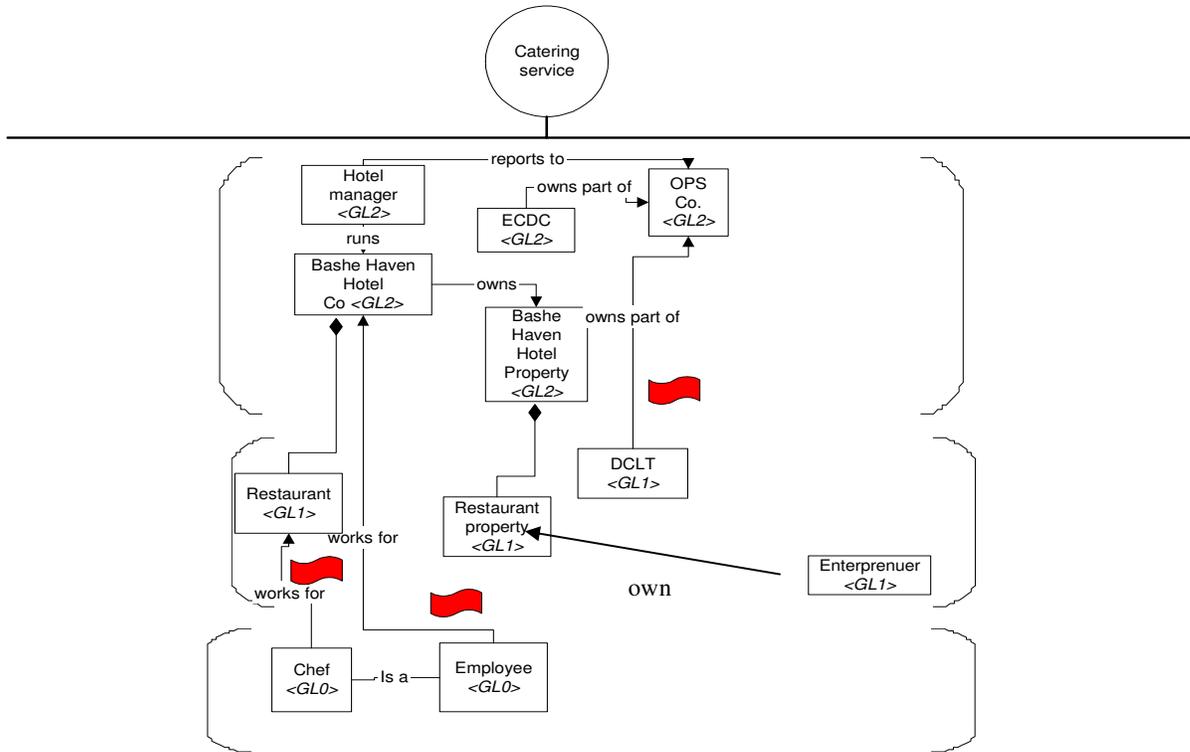


Fig. 8. Catering Service niche model, with cross-granular relationships flagged.

As a result of studying granularities and identifying intervention opportunities such as those illustrated above, the objectives of the Haven Bashe project were re-stated as follows:

1. Participate in the opportunities provided for the Haven Bashe hotel as employees and owners
2. Participate in the tourism industry by providing tourist related services inside and outside the Dwesa Cwebe nature reserve.

This represented a significant shift in the expected strategic outcomes compared to the original approach. From this and many other case studies, our methodology was seen to be highly useful in developing better models and in identifying ways of achieving project goals.

5 Related Work

In a previous paper [3] we proposed using niches in GIS system design. This paper follows on that work, describing a model and methodology for using niches, granularity and reconciliations that we have used extensively in many real-world projects. We are not aware of any other model that incorporates such modeling constructs, nor of any comparable methodology. In ecological modeling the focus is on spatial granularity, the cartographic scale of representation, at the expense of other important variables such as time, context and functional granularity. For example, [4] emphasise the importance of describing ecological entities at the relevant spatio-temporal scale, and state that “There is no reason to assume that concepts, theories or components, defined on some scale of resolution, will necessarily be applicable to ecosystem phenomena at some other scale level” [5]. In geographical information system modeling, recent thinking [6] is that the traditional approach of identifying separate layers or themes (e.g. vegetation, rainfall, slope) is inadequate because ecosystems are not disjoint but are connected wholes. [5] proposes that the rate of change within an ecosystem be used to group ecological processes into functional hierarchies. For example, bio-chemical processes such as photosynthesis occur within short time spans, while tree growth takes years, soil erosion decades, and denudation hundreds of years.

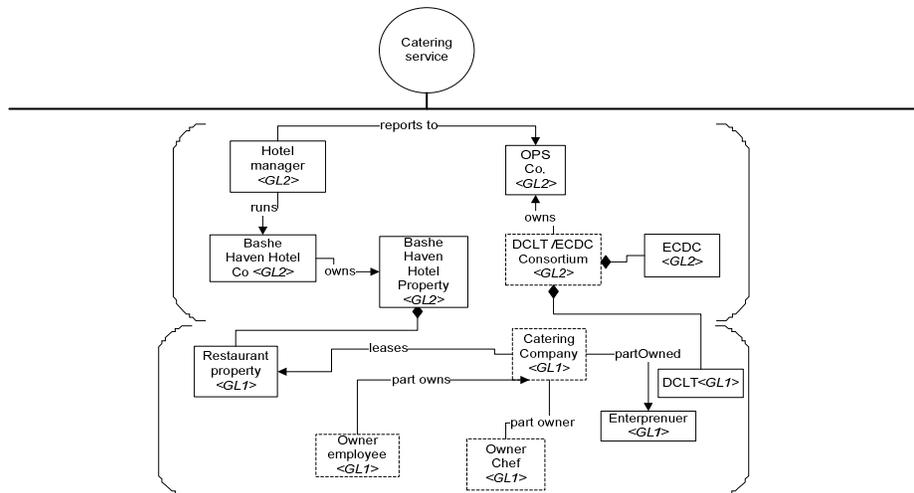


Fig. 9. Revised model, after normalisation had yielded some innovative interventions.

In [7] an ecosystem is defined as a system describing the relationships of organisms to their physical environment and to one another. A niche (also called a habitat, ecotope, biotope or micro-landscape) is an ecological whole with its own meta-properties, containing members that interact under specific environmental conditions, and favouring particular members or groups of members [7]. They refer to a niche as “a location in space that is defined additionally by a specific constellation of ecological parameters such as degree of slope, exposure to sunlight, soil fertility, foliage density, and so on”.

The suggestion that the functional niche be used as a fundamental unit of investigating ecological systems appears in [8,9]. Smith [7] defines an ecological niche as a containment space bounded by spatial and/or temporal parameters favouring specific groups of members. That entities interact with each other according to the ecosystem or environment in which they exist is also recognised by [9] who refers to niches as socio-economic units, and physical behavioural units. While these researchers advocate that niches are used in studying ecological phenomena, they offer no suggestions for incorporating this into conceptual models, nor do they propose any method for using them in the design process. They also give no indication as to whether they have put their idea into practice when tackling real problems. This paper therefore takes their work further by presenting a way of including niches in a model and methodology, and by reporting on its successful usage in a number of workshops and case studies.

Our methodology differs from other conceptual modeling methods not only in its use of niches, granularity and reconciliation, but also in the starting point of design. Booch[10] begins with a description of the system’s function and structure, as do others such as OPM[11] and Insyde[12]. Many UML-based methodologies [13,14,15,16,17] commence with a problem description and a set of use cases, and thus also focus on processes rather than on the value chain and the environments in which interactions occur. Most modeling methodologies offer little guidance to designers, with steps indicating what to do when (e.g. identify entities, then subtypes, then relationships [18], or specify process-events, then objects, then constraints[12]) but not how to do so, with little if any help in detecting errors or omissions. A few like [12,18] mention the problem of semantic heterogeneity, but fail to include this explicitly in the model. A methodology that offers more guidance than most is TCM[19] which uses noun-phrase analysis to detect possible entities, verb-phrase analysis to detect potential associations, Chen’s English Sentence Structure to determine attributes [20] and class category lists[21,22,23,24,25] to validate entities. TCM gives elimination rules to for entities and for associations. It also recommends that class categories (a given list of entity kinds such as person-role, event, etc.) be used as a checklist to detect omissions. However it does not take context or sphere of operation into account, and does not provide guidelines for discovering when inappropriate entities or entity roles have been used.

6 Conclusion

Conceptual modeling requires dealing with complexity, handling semantic heterogeneity (where different individuals use different vocabularies or attach subtly different meanings to terms), and choosing appropriate abstractions to represent the real world. A systematic, prescriptive method of undertaking such a difficult task is clearly needed. This paper presented such a methodology for conceptual modeling. It comprises the following six steps: setting objectives; identifying niches; designing niche models; assigning and checking granularities; normalising relationships; and recording reconciliations.

We introduced the notion of a niche as a basis for the methodology because of the importance of taking immediate context into account during design. A niche is an environment with a particular functional purpose that favours specific entities, in which interactions take place according to set norms and constraints. Within each niche we model not only entities, attributes and relationships, but also the granularity of entities. This we define as the relative level of influence, power or impact of the entity compared with others in the environment. This is an important tool for detecting errors and omissions in a model, because during the normalization process all relationships involving entities of differing granularity are systematically checked and remedied where necessary. We show how these remedial actions lead to the identification of missing aggregates or composition entities, or indicate the need for new specializations/roles for an entity. Without the use of granularity and normalization, such modeling inaccuracies typically go undetected. We further show that normalization can identify intervention opportunities where the status quo in the real world can be improved, because it highlights anomalies that exist and requires designers to think creatively about such problems.

Our methodology has been applied in a number of case studies and shown to be valid, applicable and useable. It improved the quality of models produced and also greatly reduced semantic heterogeneity. By virtue of modeling in the context of a specific niche, the chances of seeing a situation with discipline-specific biases are vastly reduced. Furthermore, normalizing cross-granularity relationships causes two different views of an entity to be recognized as different roles that the entity plays in different niches, and this is then resolved in the corresponding generalisation hierarchy. As a final step, the methodology requires that all instances in which semantic heterogeneity was resolved be explicitly recorded using a simple reconciles relationship that we developed for this purpose.

In future, we plan to investigate the potential of our model and methodology across a more diverse set of case studies. While we have applied them in many different situations, these have thus far all been socio-ecological projects. We need to explore the methodology in other domains, particularly in the business world. Another aspect that needs further work is that of defining more precisely the role of the facilitator (ontological engineer). Until now, the author has acted as facilitator, and questions remain as to how a facilitator should best be trained for such work and, once trained, how best s/he should proceed before, during and after each modeling session. A further important part of the research is the creation of domain ontologies from the models produced with our methodology. We have already done this for socio-economic upliftment projects, where we are now re-using our ontology in new projects quite successfully. However the general problem of converting models that incorporate niches, granularity and reconciliations into a suitable ontology needs further exploration.

References

1. Porter, M.E., *Competitive Advantage: Creating and Sustaining Superior Performance*, Free Press, 1985.
2. Poseidon for UML, Community Edition, <http://www.gentleware.com>.
3. Semwayo D.T., Berman S., 2004, Representing ecological niches in a conceptual model, *Proceedings of the first ComoGIS Workshop on Conceptual Modeling for GIS*, in *Conceptual Modeling for Advanced Application Domains*, eds. S Wang, K Tanaka et al, pp. 31-42, Springer LNCS Volume 3289, ISBN 3-540-23722-4.
4. Cheng T., Molenaar M., 1997, A process-oriented spatio-temporal data model to support physical environmental modeling, *Proceedings of the 8th International symposium on spatial data handling* pp.418-430
5. O'neill R.V., DeAngelis D.L. et al. *A hierarchical concept of ecosystems*. Princeton, N.J : Princeton University Press, c1986
6. Smith, B, Varci A.C., *The Formal Structure of Ecological Contexts*, in *Modeling and using Context*. *Proceedings of the Second International and Interdisciplinary Conference*, Berlin and Heiderberg: Bouquet P., Brezillon P., Serafini L., (eds), Springer-Verlag, 1999, pp. 339-350.
7. Smith B. *Objects and their Environments: From Aristotle to Ecological Ontology*, in *Communications of the ACM*, February 2002, volume 45, # 2. pp 79 – 9.
8. Raper J. *Defining Spatial Socio-Economic Units: Retrospective and Prospective*, in *Life & Motion of Socio-economic units*, Taylor Francis, 2001 *Multidimensional Geographic Information Science*. Taylor Francis, 2000.
9. Frank A., Raper J., and Cheylan (eds) *Life and motion of socio-economic units*, Taylor Francis, 2001.

10. Booch, G., Object-Oriented Analysis and Design with Applications, 2nd ed., Benjamin Cummings, 1994.
11. Liu, H. and Gluch, D.P., Conceptual Modeling with the Object-Process Methodology in Software Architecture, Journal for Computing Sciences in Colleges, pp. 10 – 21, January 2004.
12. King, R. and McLeod, D., A Database Design Methodology and Tool for Information Systems, ACM Trans. on Office Information Systems, Vol. 3 no. 1, January 1985, pp. 2 – 21.
13. Maciaszek, L.A., Requirement Analysis and System Design: Developing Information Systems with UML, Addison-Wesley, 2001.
14. Rosenberg, D. Use Case Driven Object Modeling with UML: A Practical Approach: Addison-Wesley, 1999.
15. Rumbaugh, J., Blaha, M. et al, Object-Oriented Modeling and Design, Prentice-Hall, 1991.
16. Siau, L, Unified Modeling Language: Systems Analysis, Design and Development Issues, Idea Publishing Group, 2001.
17. Stevens, P. and Pooley, R., Using UML: Software Engineering with Objects and Components, Addison-Wesley, 1999.
18. Teorey, J.J., Yang, D. and Fry, J.P., A Logical Design Methodology for Relational Databases Using the Extended Entity-Relationship Model, Computing Surveys, vol. 18. no. 2, pp. 197 – 222, June 1986.
19. Song I.-Y., Yano, K., Trujillo, J. and Mora, S.L., A Taxonomic Class Modelling Methodology for Object-Oriented Analysis, Proc. EMMSAD, 2003.
20. Chen, P.P., English Sentence Structure and Entity-Relationship Diagrams, Information Sciences, 1983, pp. 127-149.
21. Bahrami, A. Object-Oriented Systems Development, McGraw-Hill, 1999.
22. Larman, C., Applying UML and Patterns, 2nd edition, Prentice-Hall, 2001.
23. Richter, C. Designing Flexible Object-Oriented Systems with UML, Macmillan Technical Publishing, 1999.
24. Ross, R.G., Entity Modeling: Techniques and Applications, Database Research Group Inc., 1988.
25. Starr, L. Executable UML: How to Build Class Models, Prentice Hall, 2001.