Universal Information Models for Environmental Management

B. Nick Rossiter, Computing Science, Newcastle University, NE1 7RU & Michael Heather, Sutherland Building, University of Northumbria at Newcastle, NE1 8ST, UK

Abstract

Software to manage the environment not only needs the usual structural requirements of the standard database systems but also very high–level global consistency. Database management systems to implement any schema based on the familiar network, hierarchical or relational models have rather too many limitations for complex heterogeneous data involving environmental biodiversity. Newer categorical models offer more promise because they are constructive, integrative and based on naturalness and universals. The constructions in category theory are very similar to those of the new theoretical models based on natural systems, enabling limits and chaos to be modelled. Work is continuing at Newcastle on developing a generalized categorical model which can be adapted to handle a variety of complex ecosystems.

1 Database Management Systems

There are a number of database models available for environmental software such as the hierarchical, network and relational. The network model is based on directed graph theory and the relational on functions and relations. The hierarchical is a special case of the network model with a restriction on the linking mechanism.

All these models are aimed at real–world data where there will be a small number of classes of data, each with many instances. However in practice there may not be very many exhibits in any one category. For the evidence in the example of an investigation or trial of a chemical disaster would be quite heterogeneous. Data might consist of items like shipping, aircraft, personnel, cargo, itineraries, charts, legal documents, technical checks, safety records, telephone logs, bank records, photographs, satellite readings, forged papers and so on. Each probably has too few examples to make use of the power of relational query languages.

Furthermore these basic models do not provide enough control to represent real–world data. Much of real–world data has a very complex structure as, for example, in text and graphical data. For these, it is necessary to develop a more generalized model often termed semantic models in contrast to the standard syntactic models. Some useful ideas for handling this kind of data is available from an object-oriented approach which offers the user a number of powerful abstractions to assist in rationalizing complex situations such as the ability to construct generalizations and compositions.

However, this approach at present lacks a formal model. This raises uncertainties about consistency and therefore reliability. The object–oriented paradigm must accordingly have a cloud of doubt hanging over its use for environmental software until the day it can be converted into some formal model.

1.1 Needs of Environmental Data

We need to examine the structural requirements and semantic factors of ecological systems in the context of generalized database management systems. Some needs are satisfied by standard classical database features like views, integrity and security, display formats, temporal management and retrieval languages, but which may still need to be greatly enhanced and developed in the ecological context.

Controlling ecological features requires the availability of timely accurate information on as many variables as possible. What is needed is a portfolio of software tools for monitoring the planet in real time both for long term legal procedures as well as to provide high quality immediate decision making for action to be taken by emergency forces of international environmental agencies. The requirement is to integrate information from the different sources in such a way that deductions and inferences can be made to give a true overall picture of objects such as the forest at any time, including future projections.

Some data is short term, some long term. The methodology of the data capture needs to be invariant of the times and conditions and of particular technological developments; however the methods themselves may change. The term dynabase has been used [1] for those data systems where all the information recorded or generated over a long period of time is to be seamlessly integrated. There is a major technical challenge to capture the large amount of knowledge in each domain and the dependencies between domains. Since data is likely to be held on different systems, the question of interoperability arises between different databases, where each has its own data model and inferencing mechanisms.

It will also be necessary in environmental information systems of the future to be able to model limits. This concept concerns a series of boundaries within which periodic transitions occur. Some boundaries will be very local, others more global, indicating the extent to which an entity may change its status or behaviour. If a system remains within such limits, environmental changes will be manageable by gradual adaption [2] and our system is relatively stable. However, environmental disasters are typified by a lack of stability caused by systems going off–limits into chaos [3]. In effect, therefore we require an information model that can handle both periodic oscillation and chaos.

1.2 Formal Database Methods

The universal nature of the problems involving the environment means that the information methods need to be universal. Universal means formal but we are searching for formal methods appropriate to environmental data rather than the simpler data constructions often illustrated in formal specification languages such as Z and VDM. These two languages are based on mathematical logic applied to sets as in functional programming.

However, a mathematics which is more expressive of the complex data types [4] and limits is necessary for environmental software. In theoretical computer science, the new subject of category theory [5,6] seems the best current method to work with. Instead of the set being thought of as the basic building block, category theory is founded on the morphism usually expressed by the concept of an arrow. Category theory also has a well–established notion of limit so that stability mechanisms can be investigated and conditions for the onset of chaos [7] predicted.

The structure of the rest of this paper is first to introduce the concepts of category theory with explanations of category, functor, natural transformation and adjointness; then to discuss the architecture for an integrative approach; and finally to describe two examples to illustrate the modelling power of the categorical approach in its own right on specific environmental problems, followed by a discussion of the results.

2 Categorical Models

Category theory is based not on the set as a fundamental but on the concept of a morphism, generally thought of as an arrow and represented by $\longrightarrow$ [5]. The morphism can be regarded as an imperative arrow or as a relationship in computing. The arrow represents any dynamic operation or static condition and can cope therefore with descriptive/ prescriptive equivalent views.

The arrow is an effective representation of real–world phenomena. $A \longrightarrow B$ can represent an action from a state A to a state B, an interaction of A with B, for example a product of A with B, or a type change from type A to type B. $A \longrightarrow B$ may be a descriptive action or a prescriptive one. Alternatively it may be a probabilistic relationship. There may be any number of different arrows between the same objects.

The arrow can represent a more general relationship than the set-theoretic function. Much of scientific modelling is taken up with handling general relationships which exist between real world data. For example where there is more than one polluter for a given pollution, *polluter* is not a function of *pollutant*. The arrow can relate objects that are not sets like for instance bags and lists.

These simple categories are promising for modelling global real–world dynamic events as category theory is based on principles of naturalness that all the time relate universals with particulars [8,9]. Physical processes (interpreted widely to include chemistry and biology) are those that exist because they can be constructed. They are likewise universal in the sense that the laws of physics are invariant in all reference frames.

2.1 Functors and Transformations

An arrow between categories is termed a functor as shown in Figure 1. A functor provides the facility for transforming from one type of structure defined by one category to another type of structure defined by another category.

Functors are structure–preserving morphisms from one category to another. In Figure 1, the functor K assigns from each source object A, in category **A**, a target object K(A) to the object C in category **C**, and from each source arrow f, in category **A**, a target arrow K(f) to the arrow g in category **C**. Note that categories are given names in bold capitals. Functors really map structures. They carry across the high–level relationship as well as dealing fully in an integrated fashion with any lower–level relationship which needs to be constituted consistently within any higher–level mapping. This functorial character preserves the detailed information within transforming structures. In many models, the user has to flatten the structure when operating across different levels. Stochastic models are an example of this phenomenon where a structure may be collapsed into some statistical parameter.

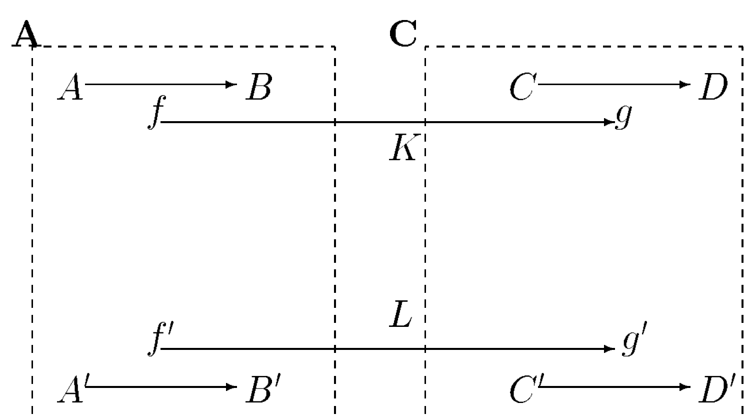


Figure 1: Functors compare Categories

A geophysical example could be the generation of earthquakes in fault regions. The aggregate activity is composed of large area faults in dynamic stability with smaller regional ones. A large earthquake could be initiated by local activity, or alternatively by large movements, or indeed by the interaction between small movements at both levels which in themselves would not be sufficient to cause a significant activity. Constructing functors to map from current inter–level geophysical structures to potential target geophysical structures is the first step in producing a model for investigating the limits and stability of target states as described later under the concept of adjointness.

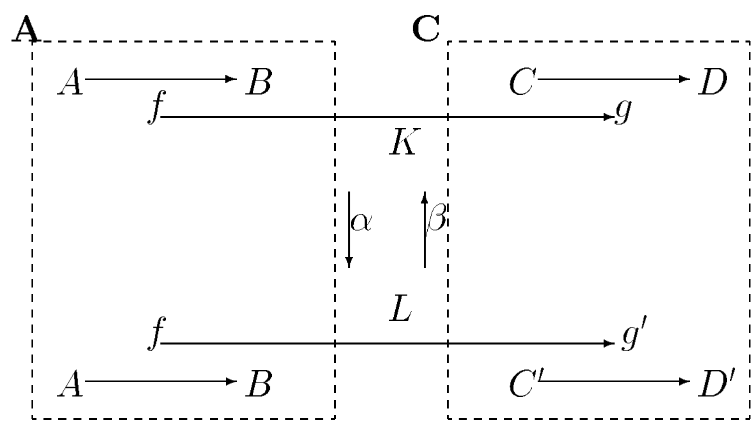


Figure 2: Natural Transformations compare Functors

An arrow between functors is termed a natural morphism (or transformation) as shown in Figure 2 where there is a natural transformation α from K to L, written $\alpha : K \longrightarrow L$

Natural transformations correspond to metalevel control mechanisms such as policy, political principles and ideals, regulators, audit programs, the principles in documents like the EMAS (Eco–Management and Audit Scheme) and the ISO standards (e.g. the British BS7750).

2.2 Adjointness

In dealing with the environment and biodiversity, we need to deal in universal properties. A very important universal concept which has emerged in the second half of the twentieth century is the principal of adjointness which can be expressed naturally in categorical models. Adjointness expresses an essential concept for formally handling complex environmental processes. Trying to integrate across levels is difficult with set theory and the more natural approach of category theory to express adjunctions is sufficient justification, in our view, for resorting to this new branch of mathematics.

Adjointness is particularly relevant to the environment for it relates to relative ordering which is the basis of natural balances and of the way that equilibria are controlled. Ormerod [10] shows how the order in nature is the result of very complex interactions, is very easily destabilized and need not be harmonious at all. By induction, he considers the same principles should apply to economics. Such apparent complexity can be modelled naturally by adjointness.

Adjointness between two categories **A** and **B**:

$$F \dashv U : \mathbf{A} \longrightarrow \mathbf{B}$$

has left and right components which specify how an arrow in category **A** is related to an arrow in category **B**. The left component is the free functor $F : \mathbf{A} \longrightarrow \mathbf{B}$ and the right component the underlying functor $U : \mathbf{B} \longrightarrow \mathbf{A}$. F is left adjoint to U and U is right adjoint to F; F may preserve colimits (sums) and U may preserve limits (products). There is a natural bijection between arrows which holds subject to the condition for all objects $A \in \mathbf{A}$ and all $B \in \mathbf{B}$ such that:

$$F(A) \longrightarrow B \text{ implies and is implied by } A \longrightarrow U(B)$$

F is a generalization of natural growth processes and evolution in an open–ended environment while U is the underlying genetic codes, laws of continuity, conservation, chemistry, thermodynamics, etc. With this condition there are two natural transformations or unit of adjunction:

$$\eta : 1_{\mathbf{A}} \longrightarrow UF, \quad \epsilon : FU \longrightarrow 1_{\mathbf{B}}$$

Adjunctions are universal descriptors for any kind of correspondence between systems whether in space or time, for example: thermodynamics stability, chemical equilibrium, biodiversity, radioactivity, and body temperature regulation.

3 Databases in Category Theory

The extra 'dynabase' advantage of the arrow in geometric logic can be seen by examining our preliminary database architecture, shown in Figure 3, with a dynamic categorical representation for use with real–world data. Each box represents a category and each arrow between categories is a functor. There is a pair of adjoint functors between each category. For ease of representation, only a functor in one direction is named.

This model provides the ability to integrate diverse models in a dynabase fashion. In effect it provides the 'glue' for linking together seamlessly the various models whether they are semantic (e.g. object–oriented, extended entity–relationship, Taxis, SDM, functional, etc), syntactic (e.g. relational, network, hierarchical) or simply stored files. The 'intelligence' in the integration comes from the functorial mappings MMt and MMt' between the categorical definition and the other models. The categorical definition then acts at the meta-meta level relating concepts and values in one model to those in another. The role of the categorical model can be compared to that of the ISO–standard Reference Model [11] which provides a standard reference point against which other models can be compared.

From the environmental data viewpoint, all components of the architecture can be heterogeneous and of arbitrary complexity to represent any data available on the biosphere. However, we envisage that a categorical system can do much more than just act as 'glue' between other models. Some components of the environmental model could be modelled and implemented directly in a categorical system, a prototype of which has been produced at Newcastle [12], as an extension of a functional database model.

We conclude with two examples which show what individual components of the model might look like with categorical modelling. Each would be one component of the semantic modelling level in Figure 3. The first example looks at high–level balances in biology in geophysics, biology and humans; the second at a more detailed example of balances in carbon dioxide and nitrogen between vegetation, atmosphere and soil,

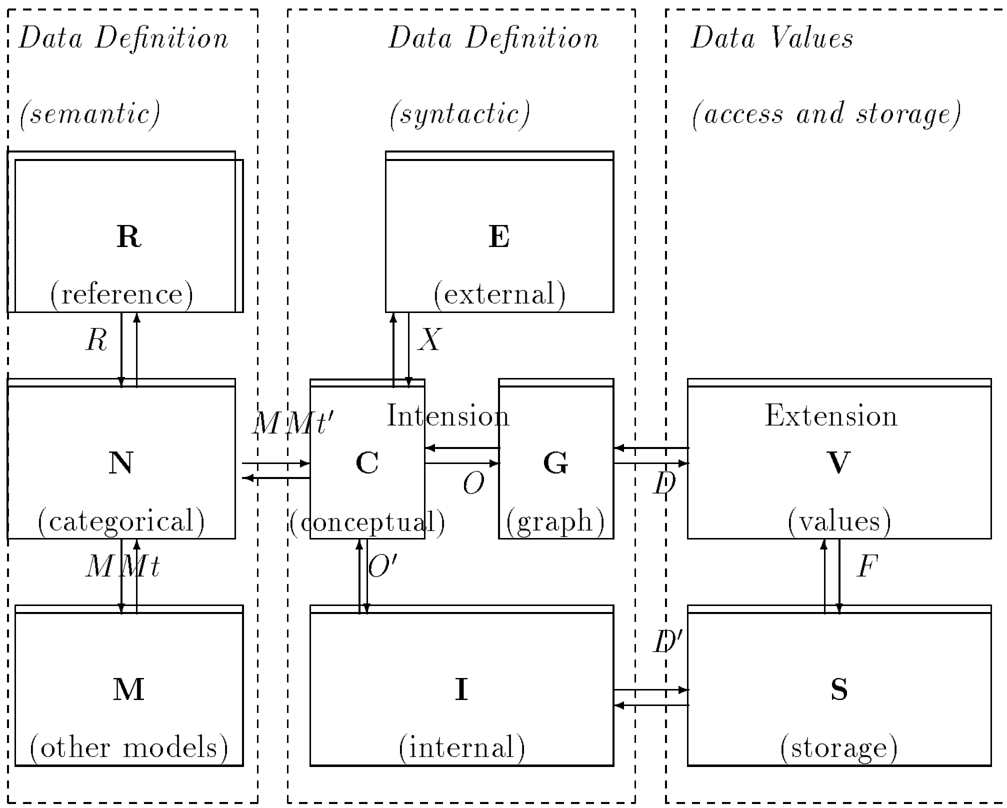


Figure 3: Integrative 'Dynabase' Architecture in Functorial Adjunction Terms

4 Global Categories

The environment is a very fine example of the adjunction between global categories. The environment is an arrow. The biosphere consists of a balance or equilibrium between the three categories shown below in Figure 4(a).

The adjointness between **B** and **H** represents the balance between the category **B** of biological species and their interrelationships and the category **H** of human activity and policy. The human category includes the axiom of choice in that we can choose to some extent the course of circumstances. The adjointness between **B** and **G** represents the balance between the category of biological species and their interrelationships **B** and the category of geophysics **G** including climate, minerals and weather and their interrelationships. This adjunction can represent the impact of climate change on food supplies [13]. The adjointness between **G** and **H** represents the balance between the category of geophysics **G** and the category of human activity and policy **H**. This adjunction can represent assessments of hazards by extra–terrestrial bodies [14].

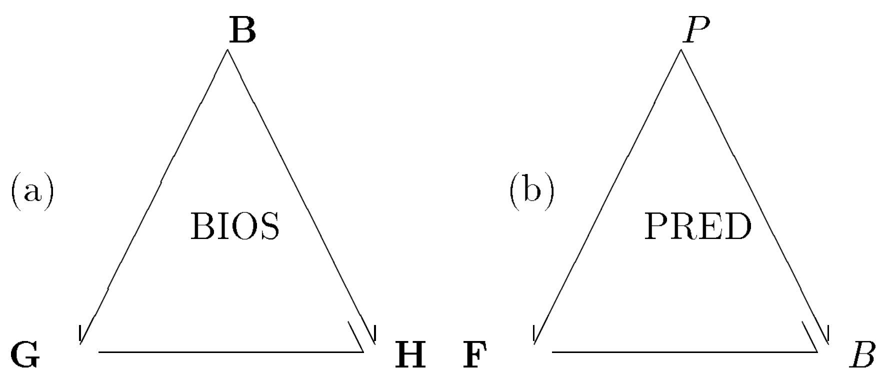


Figure 4: Global Balances between (a) Geophysics (**G**), Biology (**B**) and Humans (**H**); (b) Birds of Prey (P), Other Birds (B) and Food in General (**F**)

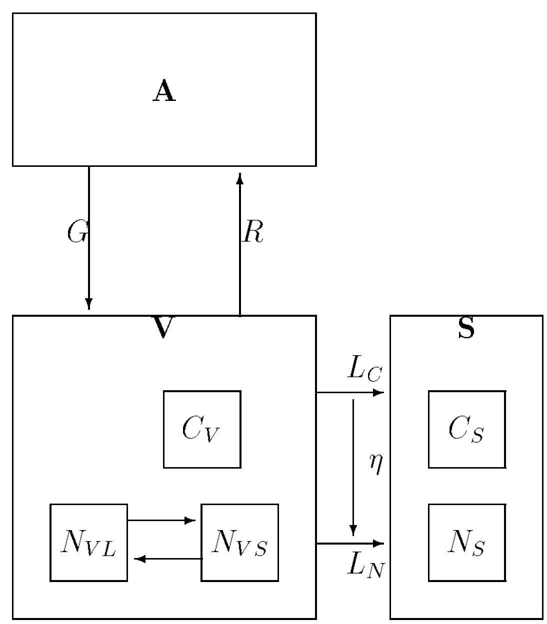
These adjunctions are shown in the Figure 4. The whole of this figure can be viewed as a category of categories **BIOS** representing the biosphere. This type of construction is termed a topos in category theory.

The categories $\mathbf{B}, \mathbf{G}$ and **H** have their own complex internal structure, involving a number of local balances in addition to the top–level ones in Figure 4(a). For example, category **B** could be defined as a collection of adjoint triangles, each the shape of Figure 4 and representing a particular balance in nature. For instance, consider the relation between birds of prey (represented by the object P), birds in general (the object B) and food in general (the category **F**, as shown below in Figure 4(b). Note that this is a fractal of the more general diagram $BIOS$.

5 Categorical Models for Natural Systems

Another example of natural equilibrium, to augment those of Figure 4, is shown below in conventional terms in Figure 5(a) as reported by [15]. This is a more detailed example looking at balances in carbon dioxide and nitrogen between vegetation, atmosphere and soil, Our equivalent categorical model is a direct representation of the balances with categories for local systems, functors relating local systems either by structure preserving or by structure transformation, and adjoints maintaining an equilibrium between two categories connected by functors in both directions. This model is shown in Figure 5(b). Note the adjointness between the atmospheric carbon dioxide category **A** (for ACD) and the vegetation category **V** (for VEG). Here we see represented two separate equilibria: $G : \mathbf{A} \longrightarrow \mathbf{V}$ and $R : \mathbf{V} \longrightarrow \mathbf{A}$.

The first is a free functor G (for GPP) assigning objects in the category **A** to objects $C_V \in \mathbf{V}$. The second is a forgetful functor R (for R_A) assigning part of the structure in **V**, that is C_V, to the category **A**. There is another intra–category adjointness between objects N_{VS} and N_{VL} in **V** where N_{VS} is nitrogen in vegetation in the structural pool and N_{VL} in the labile pool.



(a) (b)

Figure 5: Global Balances for Carbon Dioxide and Nitrogen in (a) conventional terms, (b) Categorical Terms

If the adjointness $G \dashv R : \mathbf{A} \longrightarrow \mathbf{B}$ holds, in the sense that local limits and colimits are preserved by the functors, there is a balance in nature. If it does not hold, examination of the morphisms involved may establish stability within wider limits enabling the system to be viewed as an oscillating dynamism within broad boundaries. Total failure to find any global limits within the model would mean that the system was in a chaotic region. There are then a very large number of possibilities as regards transitions in the system: optimal handling of this problem is part of current work at Newcastle University on handling uncertainty in information systems.

We also see two functors from category **V** to the category **S** (for SOIL) called L_C and L_N respectively. A natural transformation between these functors η compares the mapping from source to target of the two functors and enables us to derive equations which must hold if stability is to be achieved. If the natural transformation holds, there is a balance in nature.

6 Concluding Remarks

The constructions in category theory are very similar to those of the new theoretical models based on the natural system approach. Category theory, with its rigorous mathematical basis, therefore offers a promising route as a basis for underpinning multi–level eco–models and as a tool for constructing universal diagrams representing the various balances and adaptions. Work is continuing at Newcastle on developing a generalized categorical model which can be adapted to handle a variety of complex ecosystems including chaotic ones.

References

1. L.Press, Personal Computing: Emerging Dynabase Tools, *CACM* **37**(3) 11–19, 1984.

2. R.M.Sibly & R.H.Smith, Behavioural Ecology : ecological consequences of adaptive behaviour, 25th Symp British Ecological Society, Oxford, 1985 (620pp).

3. F.Cramer, *Chaos and Order : the complex structure of living systems*, Weinheim, New York, 1993 (249pp).

4. B.N.Rossiter & M.A.Heather, *Database Architecture and Functional Dependencies Expressed with Formal Categories and Functors*, Comp. Sci. Tech. Rep. no.432, University of Newcastle (30pp), 1993.

5. M.Barr & C.Wells, *Category Theory for Computing Science* , Prentice–Hall, 1990.

6. P.J.Freyd & A.Scedrov, *Categories, Allegories* , North–Holland Mathematical Library **39**, 1990.

7. P.Berge, Y.Pomeau & C.Vidal, *Order within chaos : towards a deterministic approach to turbulence*, Wiley, New York, 1986.

8. M.A.Heather & B.N.Rossiter, An Interoperable Theory of Higher Level Logic and Integrable Law for the Environment, in: *Towards a Global Expert Systems in Law*, Florence 1993, Bargellini, G, & Binazzi, S, (edd.), Cedam, Padua **II** 1995.

9. M.A.Heather & B.N.Rossiter, Global Theories for Global Problems, *IV International Conference towards the World Governing of the Environment* , Venice, Italy, June 1994.

10. P.Ormerod, *The Death of Economics*, Faber, London, 1994.

11. Information technology – *Reference Model of Data Management*, Standard ISO/IEC 10032 (1993).

12. Nelson, D A, & Rossiter, B N, Prototyping a Categorical Database in P/FDM, in: *ACM Workshop ADBIS*, Moscow, June 1995, **I** 247–258 (1995).

13. C.Rosenzweig & M.L.Parry, Potential Impact of Climate Change on World Food Supply, *Nature* **367**(6459) 133–138, 1994.

14. C.R.Chapman & D.Morrison, Impacts on the Earth by asteroids ans comets: assessing the hazard, *Nature* **367**(6458) 33–40, 1994.

15. J.M.Melillo, A.D.McGuire, D.W.Kicklighter, B.Moore III, C.J.Vorosmarty & A.L.Schloss, Global Climate Change and Terrestrial Net Primary Production, *Nature* **363**(6426) 234–240, 1994.

Universal Information Models for Environmental Management

B.Nick Rossiter, Computing Science, Newcastle University, NE1 7RU & Michael Heather, Sutherland Building, University of Northumbria at Newcastle, NE1 8ST, UK

Keywords:

- adjoint functors.
- category theory,
- database models,
- ecosystem equilibrium,
- environmental balances,
- limits and chaos,
- management of environment,